Using Statistical Models and Sentiment Analyses to Better Understand User Engagement with YouTube Music Videos

PhD Thesis

Miaomiao Chang

Department of Computer and Information Sciences
University of Strathclyde, Glasgow

August 15, 2025

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Contents

Abstract

The increasing prevalence of social media platforms has transformed the way users consume, interact with, and share content. Platforms such as YouTube and Twitter serve as key media for music video dissemination, fostering dynamic user engagement and shaping audience sentiment. Despite their widespread use, limited research has examined the nuanced differences in user behaviour, emotional responses, and engagement patterns across these platforms. Addressing this gap, this thesis investigates user engagement, sentiment dynamics, and content strategies related to music videos across two major platforms, YouTube and Twitter.

This research explores how textual, numerical, and time-related features, as well as user sentiment and platform-specific dynamics, influence engagement on YouTube and Twitter. Using machine learning models, alongside sentiment analysis techniques, the study evaluates the effectiveness of statistical models in predicting engagement, examines feature impacts, and assesses sentiment consistency and heterogeneity across platforms. Specifically, a mixed-method approach, combining machine learning and sentiment analysis, was used to analyze two datasets: 1,538 YouTube music videos and 76,171 comments from Twitter hyperlinks, and 2,119 YouTube music videos and 40,754 comments from YouTube channels. Feature combination strategies were tested to identify key predictors of user engagement, with model performance varying across datasets and feature configurations. While BDTR (Bagged Decision Tree Regression) demonstrated strong and consistent performance in several settings, it did not consistently outperform other models such as Gradient Boosting or Random Forest. Sentiment analysis conducted using the VADER lexicon-based approach, along with topic modelling via Latent Dirichlet Allocation (LDA), provided further insights into the emotional and

topical dynamics of user discussions. By integrating these analyses, the research provides insights into platform-specific user behaviour and strategies to enhance content engagement.

The findings highlight significant differences in how users engage with content creators on each platform. On YouTube, channel branding and creator influence play a crucial role in shaping user engagement, while Twitter interactions are more influenced by the emotional tone and topicality of content. Positive videos on both platforms rely on high-quality content to drive engagement, whereas negative videos generate broader discussions and higher interaction rates due to their controversial nature. Additionally, while YouTube exhibits consistent emotional engagement across video categories, Twitter demonstrates greater emotional variability, reflecting the platform's fast-paced and interactive nature.

This research contributes to the growing body of knowledge on cross-platform user behaviour and engagement by integrating sentiment analysis, topic modelling, and predictive modelling. Specifically, it offers insights into how user interactions with music videos differ between YouTube and Twitter, shedding light on platform-specific engagement drivers such as emotional tone, topicality, and creator influence. These findings provide a framework for content creators, marketers, and platform managers to develop more effective strategies tailored to each platform's unique dynamics. For music video audiences, this study helps clarify how engagement behaviours shape content visibility and influence which videos gain traction, potentially informing strategies that align with user interests and viewing patterns. However, this research has limitations. The focus on music video content may limit the generalizability of findings to other domains such as news or education. Additionally, the reliance on lexicon-based sentiment analysis introduces potential biases in interpreting emotional tones. Future research should explore a broader range of content types, examine emerging platforms such as TikTok or Instagram, and incorporate advanced sentiment analysis techniques, such as transformer-based models.

In conclusion, this thesis enhances the understanding of user engagement dynamics and cross-platform content diffusion across YouTube and Twitter. By providing a com-

parative analysis of cross-platform dynamics, it advances knowledge on digital media consumption and informs content creators about audience interaction patterns in the evolving social media landscape.

List of Figures

List of Tables

Preface/Acknowledgments

First and foremost, I want to express my deepest gratitude to my beloved family. Through every challenge, every moment of doubt, and every step of this journey, their unwavering love and support have been my greatest strength. **Mr. Kaiguo Chang, my father**, is a man of few words, someone who doesn't easily express his love out loud. Every time I video call my family, he simply takes a quick glance at me, making sure I'm doing well, before quietly passing the phone to my mother or sister. But I've always known, through all these years, that in his own way, I have always been his pride. **Ms. Ruifang Zheng, my mother**, is the person who has influenced me the most. *Like mother, like daughter*, from her, I have inherited resilience, an unyielding spirit, and the courage to face challenges. These qualities have carried me through the long and often difficult journey of living in a foreign country, helping me take one step at a time to reach where I am today. Of course, our similar personalities sometimes lead to small disagreements, but the love between us is strong, enduring, and everlasting. **Miss. Tingting Chang, my younger sister**, who, having studied abroad herself, understands my situation more than anyone else. In moments of breakdown and loneliness, she has always been there to listen patiently to my frustrations, offering me comfort and encouragement. She is not just my family but also my confidant and companion in this life's journey. **Mr. Song Chang, my younger brother**, the rebellious one who often exasperates me yet somehow still listens to my words. He, too, is not good at expressing emotions, but I know that he is always supporting me in his own way. **Ms. Yueming Huang, my sister-in-law**, is a kind and caring woman who quietly supports our family with love and patience. Thank you for everything you do. **Haoxuan Chang, my first nephew**, I will always love you. **Jiaxuan Chang,**

my second nephew, I will always love you. To my entire family, thank you for your unwavering love and support. Just like the family photo we took on my birthday, I will treasure it forever, just as the song goes: "Loving can hurt sometimes, but it's the only thing that makes us feel alive when it gets hard. We keep this love in a photograph, we made these memories for ourselves, where our eyes are never closing, hearts are never broken, and time's forever frozen still".

Once again, I would like to extend my heartfelt gratitude to my supervisors. My primary supervisor, Prof. Crawford Revie who stepped in midway, has demonstrated exceptional professionalism and research expertise. His insightful guidance has helped me refine and deepen my research in ways I could not have achieved alone. His dedication and support have been invaluable in shaping my understanding of my work. Likewise, my second supervisor, Prof. Gobinda Chowdhury, has complemented this research team perfectly, providing steadfast support and invaluable feedback that has significantly strengthened my study. Without their mentorship, I would not have gained such a profound understanding of my research. Their encouragement and expertise have been instrumental in this journey, and I truly appreciate their guidance.

To my dear friends, **Chen Li, Dr. Yanan Chen, Dr. Shiyun Zhao, Dr. Yan Wang**, you have been my strength and refuge throughout this journey. No one has witnessed my struggles, doubts, and moments of despair more closely than you, just as no one has celebrated my growth, resilience, and breakthroughs as wholeheartedly. You stood by me when I felt lost, offered warmth when I needed comfort and provided honest insights, encouragement, and the push I needed to persevere. You were there to shake me awake with tough love when my spirit faltered and to embrace me when I needed reassurance. Together, we have navigated challenges, shared both sorrow and joy, and in the end, gathered around the table to relish the fruits of our journey. Your presence has made this experience richer and more meaningful, thank you for being a part of it.

Finally, I want to thank myself. Because the day was good, because the day was bad, because the day was just right, because in every moment of chasing my dreams, I shone brilliantly.

Preface/Acknowledgments

Chapter 1

Introduction

This chapter sets the foundation for the thesis by outlining the scope and objectives of the research. It begins with an introduction that acts as a gateway to the study, highlighting the key terms and relevant themes that are explored in the thesis. It outlines the road taken by the researcher to accomplish the primary objectives of the research and what the user is attempting to establish from the outset in order to attain the end outcomes of the approach employed.

1.1 Background

Over the past decade, social media has emerged as a crucial platform for users and companies to express opinions, share experiences, connect with others, and enhance marketing efforts, thereby fostering stronger relationships between customers and products, brands, and companies (He et al., 2013; Laroche et al., 2013). Social media is an umbrella term for a variety of online platforms that enable users to create and share content. Based on its characteristics, social media sites can be divided into the following categories: Social networks (e.g., Facebook and LinkedIn), blogs (e.g., Blogger and WordPress), microblogs (e.g., Twitter and Tumblr), social news (e.g., Digg and Reddit), social bookmarking (e.g., Delicious and StumbleUpon), media sharing(e.g., Instagram and YouTube), question-and-answer sites (e.g., stack Overflow and Quora) and review sites (e.g., Yelp, TripAdvisor) (Barbier and Liu, 2011; Gundecha and Liu,

2012).

Watching music videos online has become one of the most popular activities on the Internet, especially after the appearance of YouTube, one of the media-sharing sites which refers to the sharing of a variety of media on the Internet, including video, audio, and photos (e.g., YouTube, Flickr, Instagram). Users are equally likely to use YouTube as a source of information on a wide variety of topics, and the platform now reaches approximately 92% of UK internet users[1]. Along with these trends, digital streaming platforms (such as Google Play Music, Apple Music, Spotify, and YouTube) are now more prevalent than digital downloads, which represents a significant shift in the music industry. YouTube has dominated the online multimedia market and distributed a substantial quantity of music through music videos and lyric videos released in conjunction with albums, as compared to other music streaming service platforms. At the same time, tweets, and retweets of content on Twitter can result in information cascades; viewers of YouTube videos leave comments and other traces that, in turn, alter the metrics of popularity and communicate the value of content to future viewers as well as to algorithms that generate search results or recommend content (Thorson et al., 2013).

What's more, the widespread adoption of those social media tools (i.e., Twitter and YouTube) has yielded an abundance of visual and textual data containing hidden knowledge that businesses can use to gain a competitive advantage. Plus, with the fast growth of social media, individuals now have more interactive and interconnected ways to engage with their favourite artists and music, allowing for engagements that were previously unimaginable. The most studied areas across all social media platforms involve marketing (He et al., 2013; Roma and Aloini, 2019; Smith et al., 2012; Wang and Gao, 2019), politics (Shevtsov et al., 2023; Thorson et al., 2013), and the entertainment industry (Hudson and Hudson, 2013; Rothschild, 2011). In particular, in a networked environment characterized by information overload, it is difficult for users to find the information they need in a timely manner. (Bawden and Robinson, 2009; Eppler and Mengis, 2008; Jones and Teevan, 2007).

[1] https://theplatformlaw.blog/2023/10/18/youtube-should-be-designated-under-the-uks-forthcoming-dmcc-regime/ Accessed: 01/10/2024.

1.1.1 Prediction Study in YouTube Videos

An increasing number of studies have been conducted to predict the popularity of online content. Multiple studies have analysed and predicted the most popular YouTube videos. As summarised by Mishra (Mishra, 2019), recent popularity prediction models can be divided into two categories based on the type of data being modelled: The first characterises user actions as discrete events in continuous time (e.g., tweets). The second is based on aggregate user actions or event volume metrics (for example, the number of daily views). Existing methods for predicting popularity based on similarity measures can be categorised as feature-based, RNN-based, and GNN-based (Ji et al., 2023). As for popularity prediction based on YouTube video studies specifically, it can be broadly categorized as a study of cross-platform data and single platform, i.e., just data from YouTube. Early popularity prediction systems based on YouTube focused on user-generated item attributes by using linear/nonlinear regression analysis in the daily volume of shares, view counts and watch time (Figueiredo, Almeida, Benevenuto and Gummadi, 2014; Mekouar et al., 2017; Nisa et al., 2021; Wu et al., 2018). Recent developments in social media and online sharing techniques have led to multiple heterogeneous sources driving attention to various categories of online content, such as YouTube videos and news articles. These sources include microblogs and traditional media coverage (Mishra, 2019). This is especially evident with the emergence of viral or "burst" videos (Ratkiewicz et al., 2010). In response to this, several researchers have developed models for predicting popularity on social media by combining data from YouTube and other platforms. Some key approaches include combining data from multiple social platforms (Crane and Sornette, 2008), integrating data from YouTube and microblogs (Yu et al., 2014), and exploring cross-platform data for popularity prediction (Ji et al., 2023).

1.1.2 Cross-platform Sharing Behaviours Studies

Existing research on the analysis of engagements across various social networks based on both YouTube and Twitter is utilized for various purposes. They mainly focus on user engagement and user sharing behaviour for marketing, branding and social

impact activities. A first essential stream of research has investigated the influence of user-generated content across social media platforms, specifically online consumer reviews and online word-of-mouth (eWOM) on consumers' purchasing decisions. There is an abundance of research within marketing literature on User-Generated Content (UGC) on Twitter and YouTube that analyses the communities that form on social media and the impact that the user content generated on these platforms has on a product or brand, thereby providing additional marketing ideas for the products (Jernigan and Rushman, 2014; Thomas et al., 2015) or brands (Ahmad et al., 2020; Natarajan et al., 2014; Roma and Aloini, 2019; Smith et al., 2012; Vallet et al., 2015).

In recent years, researchers have also paid attention to the role of users' sharing behaviours on different social media platforms, especially the relevance brought by some social events and activities. Burgess and Matamoros-Fernández (Burgess and Matamoros-Fernández, 2016) collected Twitter data containing the hashtag #gamergate and YouTube video data containing the keyword gamergate, to show how the methods of controversy analysis and issue mapping can be used to study socio-cultural controversies on social media. Specifically, users, media objects and topics on Twitter and YouTube have different characteristics and roles, reflecting the diversity and complexity of controversies (Brady et al., 2017; Kümpel et al., 2015). At the same time, Park et al., (Park et al., 2015) conducted a comparative study of the information diffusion patterns of Twitter and YouTube networks in the context of the "Occupy Wall Street" movement. They used a large dataset of tweets and videos related to the movement, and applied network analysis and statistical methods to examine the structure, dynamics, and influence of the two networks. Recently, Shevtsov et al. (Shevtsov et al., 2023) investigated political content and engagements across two social networks, Twitter and YouTube, during the 2020 U.S. election. Their analysis focused on user tendencies, engagement patterns, sentiments, and the homogeneity of users and communities on both platforms. The study revealed that Twitter users from the same community frequently shared links to YouTube videos that had comments from users within the same community, demonstrating a strong alignment in content consumption and engagement between the two platforms.

A relevant stream, more closely related to this study, has been recent efforts toward building models to predict the popularity of online content using various techniques and exploring the patterns of video viewership and sharing on these platforms based on the data from Twitter and YouTube. Abisheva et al., (Abisheva et al., 2014) compare characteristics of YouTube videos, such as topic, popularity, and polarization, with characteristics of Twitter users, like demographics, interests, and behaviours. Additionally, they look at the variation in the amount of time between the making of the video and the Twitter sharing event. Yu et al., (Yu et al., 2015) proposed a novel representation in which phase information was used to predict future popularity, and this study believes this method outperforms prediction methods that rely solely on view-count representations. In this study, phase information refers to the distinct stages a YouTube video goes through in its lifespan, particularly regarding its popularity and engagement. Based on similar data collection phrases, Wu et al. (Wu et al., 2018) use the Twitter API to extract tweets on related topics, find YouTube links in the tweets and locate links to YouTube videos shared on Twitter to get YouTube video details and related data. The researchers observed in a large YouTube video study that engagement measures are steady and predictable based on video context, topics and channel information. These findings affect video promotion and budget allocation. These studies provide some background on the engagement between Twitter and YouTube networks, the dynamics of information diffusion, and models for predicting content popularity, all of which are essential for comprehending user behaviours and forming content promotion and resource allocation strategies.

1.2 Thesis Statement

This thesis argues that user engagement with music videos on social media platforms can be effectively predicted through the integration of diverse data features, including textual, temporal, and contextual variables. By leveraging machine learning models across different datasets, this study investigates how various content and metadata characteristics contribute to engagement metrics such as likes, replies, and views.

Furthermore, the thesis proposes that engagement behaviours and sentiment ex-

pressions are not uniform across platforms. Instead, they reflect distinct social and algorithmic environments, which can be uncovered through comparative analysis. In particular, by examining user interactions on YouTube and Twitter, this research aims to reveal platform-specific dynamics and potential heterogeneity in audience responses to music content.

In sum, this thesis positions itself at the intersection of computational social science and media analytics, and seeks to advance a data-driven understanding of how content features and platform contexts influence user engagement in the domain of music videos.

1.3 Research Aims and Questions

1.3.1 Research Aims

Despite the widespread prevalence of music videos in computer-related activities, it remains an understudied topic, and the literature on user engagement across social media platforms is quite limited. To the researcher's knowledge, there is limited research in terms of comparing user engagement under pop music video cross-platforms. To advance understanding of the user's engagement and sentiment analysis across platforms, this study aims to reveal the connection between music videos and users based on the link between the YouTube and Twitter communities. This research takes an approach to the study of engagement in social media by statistically modelling the impact of textual and numerical content features on user engagement while controlling for YouTuber and context-related features.

1.3.2 Research Questions

This chapter also includes the key research questions that the researcher aims to answer, as well as the contributions made by the experts in the field of social media-based recommendation systems and how this approach can contribute greatly to revealing the community differences based on YouTube and Twitter.

The primary objective of this study is to explore how machine learning methods can be utilized to predict the factors influencing user engagement across various social

networking platforms in relation to YouTube music videos. This study specifically focuses on understanding the emotional responses elicited by these videos among users from different social networks, acknowledging that each platform may foster a unique engagement dynamic with YouTube content. By analyzing these emotional responses, this research aims to discern the underlying reasons why different platforms encourage distinct user engagements with the same music videos.

RQ1: Which machine learning methods demonstrate the strongest performance in predicting user engagement based on textual and numerical content features?

Building on the answer to question 1, there are also,

RQ2: What is the impact of time-related features (e.g., publish date, comment date) and context-related features (e.g., comment content, title, tags) on user engagement with music videos?

RQ3: Which is the most suitable machine learning method for predicting user engagement based on two datasets, which are different data sources, respectively?

RQ4: What are the factors (textual factors, context factors) that influence user video engagement on each of the two different platforms?

RQ5: Does the comparison of sentiment expressed by users' comments on similar topics on Twitter and YouTube reveal consistency or differences in user sentiment across platforms?

RQ6: If heterogeneity exists, is there an association with the categories used by the creators when they are uploading their music videos?

Selecting appropriate features and removing irrelevant features is a key issue in machine learning, and in particular, an important component of many classification and regression problems. Some data will have the same effect, some will be misleading, and some will have no effect on classification or regression. Before an inductive algorithm can move beyond the training data and generate predictions for unique test instances, it must decide which features to include and which to exclude (Cai et al., 2018; Haury et al., 2011). Intuitively, it is important for the researcher to adopt those attributes that are relevant to the intended concept, i.e., it is useful to select the optimal and minimum feature size.

Hence for RQ1 and RQ2, this study conducts experiments based on the global dataset, i.e., T&Y-video dataset, which is a full-domain dataset based on the merging of T-video (YouTube videos referenced in Twitter, henceforth referred to as "T-video") and Y-video (YouTube videos contained in the YouTube dataset, henceforth referred to as "Y-video"), relevant details are discussed further in Chapter 5 for model performance validation and finally draw conclusions. Different features are also analyzed in depth in the data capture and methodology chapters of Chapter 3 and the extended descriptive analysis of data chapters of Chapter 4, to establish a database for better answering these questions in Chapter 5.

By constructing regression models, this study can quantitatively analyze the relationship between these features and the level of user engagement and predict the popularity of video content in different scenarios (T-video and Y-video). Hence, for RQ3 and RQ4, this study conducts model validation based on the T-video and Y-video datasets, respectively, in Chapter 6, and also draws conclusions based on the final model performance. Sentiment analysis is crucial for understanding and leveraging the vast data produced on social media platforms. By evolving and integrating advanced analytical methods, researchers can better capture and interpret the complex web of user sentiments across diverse digital environments. In this study, RQ5 and RQ6 focus on sentiment analysis, and they will be answered in Chapter 7.

1.3.3 Intended Outcomes of the Research

This research is particularly pertinent for social media marketers and content creators, who continuously seek strategies to enhance user engagement and increase the visibility of their content. The findings will offer crucial insights into the most effective methods for engaging users across different social media platforms, helping marketers tailor their approaches based on the unique characteristics and engagements of each community.

Additionally, this study delves into the content characteristics that are most appealing to users, providing valuable information that can aid in the development of targeted content creation strategies. This aspect of this research is critical as it addresses the

pressing need among marketers and influencers to optimize their content for better engagement outcomes, specifically aiming to attract more likes and followers. Moreover, by exploring the varied needs of users from different communities when interacting with YouTube videos on similar topics, those findings enhance the understanding of audience segmentation within social media. This deeper insight allows for the expansion of potential engagement strategies, enabling content creators and marketers to craft experiences that are more engaging and resonant with diverse user groups.

Overall, this study not only contributes to academic knowledge in the field of social media analytics but also provides practical, actionable strategies for marketers and content creators striving to succeed in the increasingly competitive landscape of social media marketing. Through an examination of user engagements and sentiment analysis, this study aims to equip stakeholders with the tools necessary for more effective audience engagement and community building on social media platforms.

1.4 Structure of the Thesis

Chapter 1 serves as an introduction, briefly outlining the structure of the entire thesis and guiding the reader through the content organization of subsequent chapters. It also provides the research questions used throughout the thesis to give the reader basic information about the study.

Chapter 2 provides a review of the literature, grounding the study in the areas of user studies, information behaviour, cross-social media studies and sentiment analysis.

Chapter 3 describes the data sources and data capture strategy for the study, as well as the model selection and experimental design framework.

Chapter 4 provides the results of analysing the captured data, including the global data, i.e. the T&Y-video, the T-video dataset and the Y-video dataset, as well as an in-depth comparative analysis of the different features (textual, numerical, and temporal data) within each dataset.

Chapter 5 answers RQ1 and RQ2. This chapter focuses on the presentation of the results of the experimental design based on the T&Y-video, as well as on the exclusion of some useless features combination strategies (including text data, temporal data, and

numerical data) based on the global dataset (T&Y-video) for the purpose of answering the questions related to the independent datasets T-video and Y-video in Chapter 6.

Chapter 6 answers RQ3 and RQ4. This chapter focuses on presenting the results of the experimental design based on the T-video and Y-video datasets (including text data, temporal data, and numerical data), respectively, and in building on the conclusions of Chapter 5, this study simplifies the presentation of the results in this chapter by removing a number of features combining strategies that clearly did not play a predictive role.

Chapter 7 answers RQ5 and RQ6. This chapter focuses on concentrating on the textual data in the T-video and Y-video datasets, performing sentiment analysis on the user replies in the two datasets based on this textual data, visually presenting the results of the sentiment analysis, and finally extracting the topics in each dataset based on the textual data in the two datasets and presenting the results.

Chapter 8 discusses the findings, their implications, and directions for future research.

1.5 Summary

This chapter presents the background of the study and identifies the importance and urgency of the research. A review of the existing literature has identified important gaps in the body of knowledge. The main objective of this thesis is to fill these gaps through empirical research, and the specific research methodology and the analyses will be developed in the following chapters. Through this research, this study expects to provide new insights into the theory and practice of predicting social network users' engagements and contribute to the development of research on user behaviour across social platforms.

Chapter 2

Literature Review

The purpose of the literature review in this chapter is to explore and evaluate existing research on user groups of different social platforms, with a particular focus on theoretical and empirical research on user groups of different social networks based on YouTube music videos. This not only provides a solid theoretical foundation for the current study but also reveals key research trends and notable knowledge gaps within the field. By systematically analysing and comparing the work of different scholars, this review will demonstrate the complexity and multidimensional nature of user group research on different social network platforms and provide the necessary prior knowledge to address the questions posed in this study - user group engagement prediction and user group sentiment analysis of different social networks focused on YouTube music videos.

2.1 Introduction

YouTube is one of the most influential sharing platforms in the world, with a huge impact on the distribution of music content (Burgess, 2018; Cayari, 2011). Music videos, as a core part of user-generated content (UGC) on the platform, have attracted the attention of hundreds of millions of users around the world. YouTube provides music creators with a wide range of audiences and engagement opportunities, as well as great business value for the platform and advertisers. However, with the rapid increase in

the amount of content on the platform, how to effectively predict user engagement and recommending appropriate music videos to users has become an urgent problem. User engagement prediction not only helps to improve the accuracy of the recommendation system, but also enhances user engagement, increases platform retention, and provides important data support for content creators and advertisers.

This literature review aims to provide the theoretical foundation and synthesise existing approaches to predicting user engagement with YouTube music videos. In particular, it focuses on how different data sources—such as video metadata, user comments, and social engagement metrics—are used in predictive modelling. Special attention is given to the application of regression analysis and machine learning methods, evaluating their strengths, limitations, and suitability in capturing diverse aspects of user behaviour.

The scope of this chapter focuses on research related to user engagement prediction for YouTube music videos. Specifically, this review will focus on the impact of different data sources on user engagement prediction. These data sources include but are not limited to metadata of videos (e.g., video length, upload time, tags, etc.), user behavioural data (e.g., clicks, viewing time, likes, comments, shares, etc.), social engagement features, and user-generated content, such as sentiment analysis. The literature review will include studies that use machine learning models (e.g., Random Forest, Gradient Boosted Decision Tree, Deep Learning, etc.) and regression analyses (e.g., Linear Regression, Logistic Regression, Multiple Regression, etc.), and explore the performance of these techniques with different datasets.

In the field of user engagement prediction, particularly in research on sentiment analysis and user behaviour prediction for music videos, affective tendencies, the content of comments, and users 'interactive behaviours (e.g., sharing and liking) are key predictor variables. A study by Thelwall et al. (Thelwall et al., 2010) explored how sentiment analysis can be used to understand users' responses to video content; this study identified that affective strengths, in the context of short textual comments, played a significant role in the prediction of user engagement. Meanwhile, Figueiredo et al. (Figueiredo et al., 2011), by analysing the dynamics of user engagement behaviour on

the YouTube platform, found that social engagements such as liking, commenting and sharing were key factors influencing the popularity of videos, which provides an important basis for the design of predictive. These studies reveal how social engagement characteristics and user behaviour influence the number of views and popularity of music videos.

In addition, the literature review will cover studies based on regression analyses, including the use of hybrid models, based on regression and deep learning techniques, such as those proposed by Abu-El-Haija et al. (Abu-El-Haija et al., 2016) to analyse video popularity on the YouTube platform. The study captured changes in user viewing behaviour through a time series model, demonstrating that the accuracy of predicting user engagement can be significantly improved by combining different data sources. This lays a theoretical foundation for user engagement prediction models based on diverse data sources.

In summary, through a systematic review of user engagement prediction with YouTube music videos, this chapter aims to shed light on how different data sources affect the predictive power of user engagement and provide new insights for future research.

2.2 YouTube Music Videos and User Engagement Behaviours

The popularity of YouTube music videos not only reflects global interest in musical works, but also offers deep insights into audience consumption behaviour, the underlying mechanics of the platform's content distribution systems, and the virality driven by social media engagement. Analysing these popular music videos allows us to uncover which content features—such as thumbnails, titles, user interactions, and recommendation cues—drive widespread dissemination (Shen, 2024). It also provides an understanding of how platform-specific algorithms, social endorsement signals, and user participation jointly shape the success of music videos (Figueiredo, Almeida, Gonçalves and Benevenuto, 2014).

By performing social network analyses that show the distribution path of YouTube

music videos in social networks, several studies have revealed key factors that influence popularity.

Views: The number of views is a fundamental indicator of the popularity of a YouTube music video. A high number of views usually means that the video has been widely distributed and may have further expanded its reach through YouTube's recommendation system (Burgess, 2018). Globally, videos with more than 100 million views usually represent cross-cultural appeal and a broad user base.

Engagement Rate: In addition to the number of views, the engagement rate is also an important measure of popularity. This includes the number of likes, comments, shares, etc., of a video. Videos with high engagement rates typically receive higher priority in YouTube's recommendation algorithm, which further increases their exposure. Research has shown that active viewer engagement enhances the viral effect of a video (Khan, 2017).

Social Media Sharing: Sharing behaviour on social media is also an important factor in determining the popularity of YouTube music videos. Users share YouTube music videos through platforms such as Facebook, Twitter, Instagram, etc., which not only expands the dissemination of the video, but also promotes multiple viewings and discussions of the video (Trilling et al., 2017). This cross-platform dissemination mechanism can make a music video rapidly popular in a short period of time.

Trending and Topicality: The popularity of music videos is also influenced by trends and topicality. For example, videos related to current events, pop culture, or hot topics are more likely to attract viewers' attention and sharing, and YouTube's "Trending" page displays the most popular videos, making them easier for users to discover and thus increasing their popularity (Cha et al., 2007; Kavoori, 2011).

2.2.1 YouTube Music Videos

Music videos on the YouTube platform have a different pattern of user engagement compared to other video genres due to their unique emotional expression and viewing motivation (Vernallis, 2013). The content of music videos is usually closely connected to users' emotions, which makes them capable of triggering emotional responses and mo-

tivating them to interact. Understanding these unique behavioural patterns is critical to improving the accuracy of recommendation systems.

1. Music videos and user engagement

The emotional expression of music videos plays a unique role on the YouTube platform. Music, as an art form, can trigger strong emotional resonance through melodies, lyrics, etc., thus influencing users 'viewing motivations and behaviours. North and Hargreaves show that music enhances users' emotional resonance, which has a significant impact on users' viewing engagement and interactive behaviours (North and Hargreaves, 1999). On YouTube, users' motivations for watching music videos are not limited to entertainment needs, but may also involve emotional connection, nostalgia, or personal emotional experiences, which makes music videos particularly prominent in eliciting interactive behaviours (e.g., liking, commenting, sharing).

The emotional appeal of music videos drives users to engage in more social engagements. When users watch music videos, they usually express their emotions because of their emotional resonance. This is in contrast to the engagement patterns of other types of videos, where users of music videos tend to convey their emotional experience through comments or shares. For example, when users are emotionally impacted by a particular song, they are more inclined to share personal stories or feelings in the comments section. Schäfer et al. (Schäfer et al., 2013) highlights the impact of music on emotional expression and notes that this emotionally driven behaviour is crucial in user engagement prediction.

2. Comparison with other video genres

Compared to music videos, other types of videos (e.g., movie trailers and educational videos) show significant differences in user behaviour patterns. Movie trailer viewing behaviour is typically time-sensitive, with users' motivations for watching such videos being primarily related to the upcoming release of a film. As a result, viewing of movie trailers usually peaks around the time of the film's release and is highly dependent on external factors such as star power or advertising (Hennig-Thurau et al., 2007). In contrast, music videos are capable of triggering long-term viewing and repeated engagements due to their emotionally resonant qualities.

The viewing behaviour associated with educational videos tends to be functional and purpose-driven. Users primarily watch these videos to acquire information or develop specific skills, leading to more structured and goal-oriented interaction patterns. A study by Guo et al. (Guo et al., 2014) found that the viewing duration and interactive behaviour for educational content are closely aligned with learners' needs, rather than being driven by emotional engagement. As a result, traditional metrics such as viewing duration and completion rate are often sufficient to assess user engagement in this context. By contrast, music videos frequently evoke emotional responses and stimulate social interactions, making it necessary to complement traditional engagement metrics with sentiment analysis and social engagement data for a more comprehensive prediction of user behaviour.

In addition, music videos often exhibit short-term explosive viewing behaviour. For example, when a new song is released or a music video becomes popular, its viewership tends to climb rapidly in a short period of time. This burst of viewing behaviour contrasts with the long-term accumulation of viewing of movie trailers or educational videos, and Hennig-Thurau et al. (Hennig-Thurau et al., 2007) showed how the viewing of movie trailers is influenced by time and external factors, whereas the emotional resonance of music videos makes their viewing behaviour more emotional and volatile, challenging traditional user engagement prediction models.

In summary, music videos exhibit different patterns of user behaviour on the YouTube platform when compared to other video genres. Due to the significant influence of sentiment expression on users' viewing motivation and social engagement behaviours, user engagement prediction models for music videos need to incorporate sentiment analysis and social engagement data to improve accuracy. Compared to movie trailers and educational videos, emotion-driven viewing behaviours of music videos are more difficult to predict with traditional behavioural data, so research should further explore how to use sentiment data to improve recommender systems.

2.2.2 User Engagement and Interaction Behaviours

In studying YouTube users' engagement and interaction behaviours, the focus is on analysing how users watch videos and interact (e.g., like, comment, share) on the platform, and how these behaviours affect community engagements and the platform's ecosystem.

1. Viewing Behaviour Analysis

Users' viewing behaviour on YouTube typically involves several factors, including viewing duration, viewing frequency, and viewing motivation. These behaviours are influenced by the user's personalised recommendation system, the appeal of the video content, and the user's immediate needs or interests. (1) Duration and frequency of viewing: Research has shown that the length and frequency of users' viewing are closely related to the type of video content, video duration, and the platform's recommendation algorithm. YouTube's recommendation system pushes videos that may be of interest to users based on their historical viewing history and behavioural patterns, thus extending their stay on the platform (Covington et al., 2016). This personalised recommendation not only increases the length of time users spend watching, but also improves user 'stickiness' on the platform. (2) Viewing Motivation: Users are motivated to watch for a variety of reasons, including entertainment, education, information acquisition, and social connection (Burgess, 2018). It has been found that different motives affect users' viewing behaviour. For example, users motivated by entertainment may be more inclined to watch short videos, whereas users motivated by learning may be more inclined to watch educational content of long duration. (3) Interactive Behaviour: Users' interactive behaviours include liking, commenting, and sharing, which not only reflect users' attitudes towards the video content, but also contribute to video dissemination and community engagement (Rotman et al., 2009). Liking and sharing behaviours are usually a simple endorsement of the content by the user, whereas commenting is a more in-depth form of interaction that involves communication and exchange of views between users.

The comment section is an important place for YouTube users to express their opinions, participate in discussions, and build a community. Users interact with video

creators and other viewers through comments, and these engagements not only enrich the content ecosystem of the video but also promote community activity.

2. Comments and engagements

The comment section is an important place for YouTube users to express their opinions, participate in discussions, and build a community. Users interact with video creators and other viewers through comments, and these engagements not only enrich the content ecosystem of the video, but also promote community activity. (1) Role of comments: The comment feature allows users to share their opinions, ask questions, or discuss related topics with others below the video (Snelson, 2011). The activity of the comment section usually reflects the influence and popularity of the video. An active comment section attracts more viewers to participate in the discussion, further increasing the number of views and frequency of engagement with the video. (2) Impact of comments on community building: comment sections are not only a place for users to express their personal opinions, but also an important way for users to build a sense of community belonging. Research has shown that positive comment engagements can enhance users' sense of community belonging and loyalty (Westenberg, 2016). This community-building effect is particularly significant on YouTube, especially under videos of creators with a large number of followers, where fans express their support and identification with the creator through comments.

By analysing users' viewing and engagement behaviours on YouTube, as well as their activity in the comments section, it is possible to better understand how users interact with the platform and other users. These behaviours not only reflect users' content engagement and motivations but also influence community activity and the platform's overall ecosystem.

2.2.3 User-Generated Content and Creative Behaviour on YouTube

On a social media platform such as YouTube, User-Generated Content (UGC) is a core component of the platform's ecosystem. By examining how users create and upload content, as well as their motivations and challenges, this study can better understand how YouTube maintains its position as the world's largest video-sharing platform.

1. User Generated Content (UGC)

UGC refers to content that is created, uploaded and shared by platform users themselves, rather than produced by professional content producers or media companies. On YouTube, UGC comes in a variety of forms, including video blogs (vlogs), tutorials, reviews, game videos, music covers, and more. This content enriches the diversity of the platform and attracts a large audience. Users generate content for a variety of motives, which usually include expressing themselves, sharing knowledge, seeking social recognition, and gaining financial rewards (Burgess, 2018). For example, many users upload videos to express their opinions or showcase their talents as a way to build their personal brand or influence.

Although YouTube offers an open platform for users to create and share content, the production of UGC still faces several challenges. First, there are technical barriers, such as the need for video editing skills, access to special effects tools, and the availability of appropriate equipment and software (Chau, 2010). Second, the high level of competition—particularly in saturated content categories—requires creators to continuously innovate and differentiate their content in order to attract and retain viewers.

Despite these challenges, UGC has had a profound impact on the YouTube platform ecosystem. It significantly enhances content diversity, enabling the platform to appeal to a wide range of audiences with varying interests and backgrounds (Kim, 2012). Moreover, UGC fosters interactive communities where users engage through viewing, commenting, and sharing. This sense of community, in turn, encourages further content creation and reinforces user participation on the platform.

2. Creative Motivation and Reward Mechanisms

YouTube's content creators (YouTubers) are motivated to create not only by self-expression and social recognition but are also driven by the financial reward mechanisms offered by the platform. For many content creators, YouTube is a platform for self-expression. It is a platform where they can showcase their talents, share their life experiences or express their personal opinions. Through content creation, users can not only build their personal brands but also gain audience attention and social recognition, which further enhances their motivation to create (Westenberg, 2016).

YouTube incentivises creators through several profit models such as ad-sharing, membership revenue and brand partnerships. Creators can earn revenue through YouTube's ad share program, which makes video creation not just a hobby but a full-time job (Postigo, 2016). In addition, YouTube offers other incentives such as Super Chat and members-only content, all of which provide creators with a diverse source of income. Creators enhance their connection with viewers by actively participating in platform engagements (e.g., posting videos regularly, responding to comments, and interacting with viewers). Such engagements not only increase viewer loyalty, but also boost revenue through increased views and ad clicks (Van Dijck, 2013). YouTube's algorithm also tends to recommend channels that are regularly updated and have frequent engagements with viewers, which further incentivises creators to keep creating.

2.2.4 Social Communication and Content Sharing on YouTube

On platforms such as YouTube, UGC is a fundamental driver of the platform's ecosystem. UGC transforms users from passive consumers to active producers, shaping the diversity, volume, and cultural relevance of online video content (Burgess and Green, 2009). Prior studies have shown that users create and share videos for a variety of motivations, including self-expression, social interaction, community building, and reputation enhancement (Jönsson and Örnebring, 2011). These motivations not only sustain the continual inflow of content but also contribute to YouTube's ability to maintain its dominance as the world's largest video-sharing platform. Furthermore, UGC plays a key role in audience engagement and content dissemination, with user participation—through uploading, commenting, and sharing—being directly linked to the visibility and popularity of videos (Cha et al., 2007).

Social sharing is an important mechanism that drives the wide distribution of YouTube content. Through social networks, users not only share videos of interest to them, but also promote the dissemination and discussion of these videos on a wider scale. Users share YouTube videos for a variety of motives, including conveying information, expressing opinions, entertaining others, or strengthening social connections (Benevenuto, Rodrigues, Cha and Almeida, 2009). Social sharing not only increases

the exposure of a video, but also promotes its viral spread. Users share videos with friends or groups through social networks, further extending the reach of the content. Videos shared via social networks typically spread faster and reach a wider audience (Khan, 2017). The act of sharing not only directly increases the number of times a video is viewed but may also trigger wider discussion and secondary distribution. This type of word-of-mouth marketing (WOM) (Berger and Milkman, 2012) has been shown to be very effective in increasing video exposure and attracting new viewers, especially when the video content is able to inspire emotional resonance or spark social discussions.

In addition to sharing within the YouTube platform, users also share YouTube videos through other social media platforms (e.g., Facebook, Twitter, Instagram), and this cross-platform sharing further enhances the dissemination of videos (Benevenuto, Rodrigues, Cha and Almeida, 2009). Cross-platform sharing makes the content not only limited to dissemination within one platform, but also reaches out to different social media user groups, increasing the cross-platform impact of the video.

Users have a variety of motivations for choosing to share YouTube music videos, and these motivations affect not only how quickly a video spreads, but also its reach and the diversity of its audience base. Here are some common motivations for sharing:

- Self-expression: Many users express their personality, interests, or emotions by sharing music videos. This expression can be achieved by sharing their favourite songs or videos on social media, reflecting users' musical tastes and personality traits (Berger and Milkman, 2012). For example, when users discover a song that expresses their current mood, they may choose to express those emotions by sharing it.

- Social Interaction: Sharing music videos is also a form of social interaction where users connect with friends and family by sharing music. Sharing videos can spark discussion, get feedback, or serve as a starting point for social interaction. By doing so, users not only enhance their social relationships with others, but may also increase the number of times a video is viewed and its impact (Multisilta et al., 2012).

- Dissemination of valuable information: Users may also share music videos for the purpose of disseminating information. For example, if a song's lyrics or video content addresses a social issue, public interest theme, or cultural phenomenon, users may view it as information worth spreading and expand its reach through sharing.

Sharing behaviour has a significant impact on the spread of YouTube music videos, especially when shared across platforms:

- **Viral spread:** The nature of social media allows music videos to spread quickly among users. Users "like", "comment", "retweet" and so on, which can make the video "viral" quickly, especially in some hot videos. This phenomenon is especially obvious in the dissemination of some hot videos (Khan, 2017). Once a video has gained enough attention on social networks, it is likely to be featured on YouTube's "popular" page, further expanding its influence.

- **Expanding audience reach:** By sharing across platforms, YouTube music videos can reach a wider audience. For example, a user sharing a YouTube link on Facebook may attract the interest of users on other platforms to watch and share the video. This cross-platform distribution mechanism not only enables videos to break through the limitations of a single platform, but also increases their reach across different user groups (Cha et al., 2007).

- **Enhanced interaction and engagement:** Sharing behaviours can also enhance the interactivity and engagement of a video. When a video is shared frequently across multiple social platforms, it tends to attract more comments and discussions, and these interactive behaviours further enhance the video's exposure and user engagement. Especially when the video content involves controversial or popular topics, the interactive behaviours are more active and the video's dissemination effect is more significant (Kümpel et al., 2015).

Modern social media platforms often provide convenient sharing features that enable users to easily share YouTube video links to platforms such as Facebook, Twitter, and

Instagram. This simplicity has made video sharing a part of users' daily interactions and has greatly increased the efficiency of video dissemination (Cha et al., 2007). When a music video is widely shared on social networks, other users are more likely to view it as "worth watching". This social acceptance helps to further extend the reach of the video and leads to more users engaging in sharing and discussion (Berger and Milkman, 2012).

2.3 The Role of Different Data Types in Predicting User Engagement

In user engagement prediction, a diversity of data sources is crucial. By combining multiple data sources (e.g. video metadata, user social engagement data, sentiment analysis in user comments, etc.), predictive models can capture the complexity of user behaviour and engagement changes more comprehensively. This section explores the impact of video metadata, social engagement data, and sentiment analysis on user engagement prediction.

However, whether such categorisation is associated with users' emotional responses—particularly in comment sentiment—remains underexplored. Empirical evidence linking content categories and sentiment dynamics is still limited.

2.3.1 The Role of Video Metadata

Video metadata, such as the length, upload time, title, and tags of a video, can provide crucial clues for predicting user engagement. Metadata is a fundamental attribute of videos that directly affects their presentation in search results and ranking in recommender systems (Zhou et al., 2010). This information provides strong support for predicting users' interests before they have even clicked to watch. By analysing metadata, platforms can effectively identify which videos are most likely to attract users 'attention and optimise recommendation algorithms to improve users' viewing experience and content discovery rate.

1. The Impact of Video Length

Video length is a key metadata element that directly influences viewing duration and completion rates. Empirical studies indicate that shorter videos often generate higher engagement, particularly in the current era of rapid content consumption, where users prefer short-form content during fragmented periods of attention (Guo et al., 2014; Joachims, 2002). This tendency is highly relevant for music videos, as audiences typically favour moderate durations that provide immediate emotional gratification and entertainment value. In the context of MOOCs, Guo et al. found that short videos (under six minutes) achieved significantly higher completion rates and user satisfaction, a finding that is transferable to entertainment media: concise music videos of approximately three to five minutes align well with users' consumption patterns, enhancing both the viewing experience and the effectiveness of recommendation algorithms.

2. The Impact of Upload Time

Upload time is an important factor that affects video exposure and user interaction. Users' viewing behaviour usually shows time dependency, i.e., users are active at different times of the day. For example, users are more likely to watch videos in the evening after work or on weekends, and the amount of user interaction is higher during these times. Platforms can optimise recommendation strategies based on upload times, enabling videos to gain more exposure and interaction during active user hours (Covington et al., 2016). Research has shown that there is a significant correlation between exposure and viewership of videos and upload time, and that optimisation of upload time can increase the number of times a video is viewed during popular times (Jiang et al., 2020).

In the social media environment, the time factor is also related to users' short-term engagement. Certain content (e.g., news, current event videos, or pop culture-related content) may become popular quickly during a specific time period, but its lifecycle is usually short. For music videos, while upload time does not significantly affect the lifecycle as it does for news videos, the right upload time in the early stages of promotion can maximise the video's initial exposure and user interaction engagement. This is particularly critical in real-time popular video recommendation, where recommender systems can better optimise the timing of video pushes by analysing patterns of upload

time and viewing behaviour (Zhou et al., 2010).

3. The Role of Video Tags and Titles

Video tags and titles are also an important part of metadata; not only do they help categorise and index video content, but they also have a direct impact on user search behaviour. The choice of video tags determines the visibility of the video in search engines and whether the video matches the user's interests. In a study by Zhou et al. (Zhou et al., 2010), video tags were shown to be a key factor affecting video search ranking and recommendation system matching. Proper use of tags can make videos more discoverable, especially in recommender systems, and tags can help improve the match between videos and user interests.

At the same time, video titles are critical in attracting users to click. Studies have found that videos with attractive titles tend to get higher click-through rates. Users often use titles to initially judge whether a video matches their interests, especially when recommendation algorithms display multiple videos, and Joachims (Joachims, 2002) pointed out that users are more likely to choose titles that are intuitively clear and convey a clear value or emotion when clicking on a selection. For music videos, titles that include the names of well-known artists, popular songs or albums tend to increase click-through rates, which in turn increases viewing and interaction.

When predicting user engagement, metadata modelling can effectively improve the performance of recommendation systems. By analysing metadata, the platform can construct a personalised recommendation model based on video length, upload time, tags, title, and other information, and it can predict in advance the user's possible viewing behaviour and interest tendencies. For example, based on the video length and the user's previous viewing history, the recommender system can speculate the user's engagement for videos of a specific length; while the combined use of upload time and tags can optimise the video's push during the user's most active hours (Covington et al., 2016). By introducing metadata, the system can determine a user's potential interest based on this basic information before the user has interacted with the video, providing a basis for further content pushes.

2.3.2 User Social Engagement Data

User social engagement data (e.g., likes, comments, shares, etc.) is a direct manifestation of users' active expression of reactions to video content on social media platforms. Different from passive behaviours (e.g., number of views or video duration), social engagement data can reflect users' emotional feedback and engagement with the video in a more in-depth manner, and is an important source of data for predicting the popularity of the video and user engagement. On social platforms, user engagement behaviours not only help videos gain greater distribution, but also provide feedback signals to the recommender system, further driving video exposure.

1. The Role and Limitations of Likes

Likes, as an easy way for users to express positive comments, are often considered a key indicator of a video's popularity. Studies have shown that the number of likes has a significant positive correlation with the popularity of a video. By clicking on the 'Like' button, users provide positive feedback to the platform's recommendation algorithm, driving the platform to push the video to other potential viewers (Figueiredo et al., 2011). An increase in the number of likes typically improves a video's ranking in the recommender system and promotes its visibility to other users. However, despite the fact that likes provide a strong signal of positive user reactions to a video, they still have limitations as a single indicator.

First, although liking behaviour can reflect users‘ positive evaluations of a video, it does not capture users' deep emotional responses or detailed feedback on the video content. Users may like videos for a variety of reasons, such as personal engagement, social pressure, or group influence, but these behaviours do not necessarily indicate a deep understanding or strong emotional resonance with the video content (Napoli, 2011). Therefore, relying solely on likes as a predictive basis for user engagement may lead to bias, especially in diverse user groups. Covington et al. (Covington et al., 2016) also point out that users sometimes like videos even without watching the full video, which makes the number of likes not always an accurate reflection of the actual quality of the video or the user's long-term engagement.

2. Emotional depth and impact of comments

Comments are a detailed form of user feedback on video content, which not only expresses the user's opinion but also further extracts the user's emotional tendency through sentiment analysis. The number of comments measures the user's interest in the video content to a certain extent, and the content of the comments can reflect the user's specific views on the video. Unlike simple likes, comments express richer emotions and opinions through textual forms, which can reveal users' deeper reactions to the content. Research has shown that videos with frequent comments tend to have higher discussion and visibility, and the more interactive the comments are, the more likely the video is to receive recommendations and more exposure from the platform (Thelwall et al., 2010).

Sentiment analysis of comments can categorise a user's affective tendencies as positive, negative or neutral through natural language processing (NLP) techniques, and this sentiment data provides additional information for predictive models. For example, Thelwall et al. (Thelwall et al., 2010) showed that the number of positive comments was positively correlated with the popularity of a video, while an increase in negative comments could signal a decrease in user experience. This type of sentiment analysis not only helps platforms understand users' attitudes towards videos, but also provides feedback to content creators to help them adjust their content strategies. However, the quality and quantity of comments are not always proportional, and sometimes a large number of short or meaningless comments (e.g., 'nice' or emoticons) may not be effective in improving the judgement of content quality, so it is important to analyse the sentiment and content depth of comments (Paltoglou and Thelwall, 2012).

Another limitation of comments is that although they can reveal user attitudes, the group of users who engage with them tends to be a relatively niche group (Lange, 2007). Not all viewers will actively comment, especially if the content is highly complex or involves sensitivity to personal views. As a result, the number of comments and the results of sentiment analyses may not be fully representative of the attitudes of all viewing users. This limitation means that when making user engagement predictions based on comments, it is important to incorporate other data sources (e.g., viewing behaviour and liking data) to gain more comprehensive insights.

3. Diffusion effects of sharing behaviour

Sharing behaviours represent a high level of user approval of the video content and are usually a direct indication of the user's willingness to proliferate the video content to users' social networks. Sharing behaviour not only increases the exposure of the video in the user's social circle but also helps the video gain additional exposure in the wider social network. Unlike likes or comments, sharing behaviours are often seen as a signal of high approval of video content, as users not only take the time to watch it, but also actively recommend the content to others. Zhou et al.'s (Zhou et al., 2010) study showed that the number of shares of a video is a key variable in predicting its popularity and long-term impact. The more sharing behaviours, the wider the distribution path of the video, and the number of times the video is viewed and the interaction rate increases.

However, the impact of sharing behaviour also depends on the structure of the user's social network. The diffusion effect of sharing is often limited by the size and engagement of the user's social network. For example, a video shared on a large social platform (e.g., Twitter or Facebook) may quickly gain widespread exposure, whereas the impact of sharing may be more limited in smaller or more closed networks. Thus, the predictive effect of sharing behaviour relies on the size of a user's social network and the strength of social ties (Bakshy et al., 2012). Nonetheless, sharing behaviour is still a valid indicator that can help recommender systems determine which videos have the potential to spread over a wider area.

Combining social interaction data such as likes, comments and shares in user engagement prediction models can significantly improve the accuracy of predictions. These data not only reflect users' behaviour, but also reveal their emotional inclination and depth of engagement. By integrating multiple interaction metrics, the recommender system can better capture the user's interest and the potential popularity of the video. Covington et al. (Covington et al., 2016) analysed the association between social interaction data and viewing behaviours through a deep learning model, and found that combining these data can effectively improve the effectiveness of the recommender system. Meanwhile, with the development of technology, more studies have begun to use

machine learning and deep learning techniques to further mine users' social interaction behaviour patterns and provide more accurate support for personalised recommendations.

2.3.3 Sentiment Analysis and User Comments

Sentiment analysis techniques are increasingly used in user engagement prediction, especially on social media platforms by analysing user comments, which can help identify users' attitudes towards videos and predict their future behaviour (Pang et al., 2008). Sentiment analysis uses natural language processing (NLP) techniques to categorise the emotional tendencies in comments, which are usually classified as positive, negative and neutral. Through this classification, platforms can better understand users' emotional responses to videos, thereby optimising recommendation systems and improving the accuracy of predictions of video popularity.

Emotional expressions in comments are usually aligned with the user's overall evaluation of the video; therefore, by analysing emotions in user comments, platforms can capture users' engagement, interests and possible interactive behaviours. For example, Thelwall et al. (Thelwall et al., 2010) showed that sentiment intensity (i.e., the degree of positive and negative sentiment in comments) is significantly correlated with users' viewing, liking, and sharing behaviours. Positive comments with strong emotions usually imply that the user highly approves of the video, which may lead to more liking and sharing behaviours, thus increasing the popularity of the video (Thelwall et al., 2010). Negative comments can also provide important feedback to help platforms identify negative user reactions and adjust video recommendations.

Moreover, the results of sentiment analysis can provide richer inputs to video recommendation algorithms. Traditional recommendation systems often rely on user viewing history and simple behavioural data (e.g., length of viewing, number of likes), which only provide limited information about user engagement. By incorporating sentiment analysis, the system can gain a deeper understanding of the user's emotional response and optimise the recommendation algorithm with this sentiment data. For example, when the system identifies that a certain type of video usually triggers a large amount

of positive emotional feedback, it can prioritise recommending similar types of videos to relevant users (Pang et al., 2008).

Ahmed et al. (Ahmed et al., 2013)further showed that combining sentiment analysis with other data sources such as video metadata and social interaction data can significantly improve the performance of predictive models. Sentiment tendencies in comments not only reflect users' attitudes towards videos, but are also closely related to users' social interaction behaviours (e.g., liking and sharing). By combining the emotional tendencies in user comments with video metadata (e.g., the subject, length, and upload time of the video), the combined utilization of these data can significantly improve the accuracy of the recommender system. For example, if a user leaves a positive comment under a music video and the video receives a large number of likes and shares at a particular time, the system can recognize that such videos may trigger more positive feedback and recommend similar videos to other users.

Sentiment analysis is particularly useful for emotionally driven content such as music videos. Music videos often evoke strong emotional responses from users, who are more likely to express their emotional reactions to the music in the comments. For example, users may express their favourite songs or personal experiences related to the music under the music video, and the sentiment data in these comments provides valuable additional information to the recommender system (Paltoglou and Thelwall, 2012). By analysing these affective responses, the system is able to more accurately capture the user's interests and emotional engagement, and thus optimise the recommended content.

Despite the important role of sentiment analysis in user engagement prediction, its application faces several challenges. Firstly, natural language processing (NLP) techniques still have limitations when dealing with unstructured text data. User comments often contain complex sentiment expressions and non-standard language, such as slang, sarcasm, irony, etc., which may complicate sentiment categorisation (Cambria et al., 2013). These linguistic features may not be accurately captured by traditional sentiment analysis models, thus affecting the accuracy of predictions.

In addition, the content of comments may vary significantly across platforms and

video types. For example, music video comments on YouTube may differ in sentiment expression from comments on news videos, educational videos, and other types of videos. Therefore, sentiment analysis models need to be customised according to the video content type to ensure that they can accurately capture users' emotional responses. For example, Pang and Lee (Pang et al., 2008) point out that sentiment analysis models for different domains require specific training data and vocabularies to perform optimally in their respective contexts.

Future research directions could further improve the accuracy of sentiment analysis by combining the video content itself (e.g., audio and visual features) with the sentiment data in the comments on the basis of multimodal analysis. For example, by analysing the relationship between the pitch and rhythm in music videos and the sentiment in the comments, the system can gain a deeper understanding of the user's emotional response and make accurate predictions.

In conclusion, combining multiple data sources in user engagement prediction significantly improves the accuracy and complexity of prediction models. Video metadata (e.g., video length, upload time, tags, etc.) provides basic structured information for recommender systems, which play an important role in initially screening user interests and optimising content recommendation strategies. However, metadata can only provide limited information about the content and cannot comprehensively capture users' behaviour and emotional responses.

Moreover, while prior studies have examined the predictive value of individual metadata elements such as tags or upload time, few have systematically explored how time-related and context-related features collectively influence user engagement. Understanding the distinct and combined impact of features like publish date, comment timing, video title, and tag structure remains an open research question. This gap motivates the second research question of this study.

2.4 Application of Regression Analysis Models and Machine Learning in User Engagement Prediction

Both traditional regression analysis models and modern machine learning techniques play an important role in user engagement prediction. Different technological approaches are able to capture the complex relationship between user behaviour and video popularity and provide key data to support recommendation systems. By analysing user interaction behaviour, video metadata and sentiment feedback, predictive models are able to identify the core factors that influence user engagement. The following section describes the application of regression analysis models and machine learning techniques in this area and explores how hybrid models can further improve the accuracy of predictions.

2.4.1 Regression Analysis Models

Regression analysis models have a long history of application in user engagement prediction, mainly including linear regression, logistic regression and multiple regression. The core goal of regression models is to find the relationship between independent variables (e.g., user behavioural characteristics) and dependent variables (e.g., number of video views, number of likes, number of shares, etc.). However, different regression models have certain limitations and advantages in dealing with user behaviour data.

1. Linear regression models in user engagement prediction

Linear regression models are often used to deal with scenarios where linear variation between the independent and dependent variables is assumed. For example, there may be a linear relationship between certain behavioural characteristics of a user (e.g., length of time spent watching a video) and the popularity of the video. Montgomery et al. (Montgomery et al., 2021) mentioned that the main advantage of linear regression is that it is highly explanatory, and the coefficients can be estimated to clearly show the impact of each variable on the results, thus helping researchers to understand the underlying patterns of user behaviour. In addition, linear regression is less computationally complex and is suitable for initial exploration of the relationship between user

behaviour and video popularity.

However, linear regression also faces significant limitations. In particular, when the relationship between user behaviour and video popularity does not conform to the assumption of linearity, the predictive effectiveness of the model can be drastically reduced. Gareth et al. (Gareth et al., 2013) point out that user behaviour on social media platforms often exhibits non-linear patterns, such as popularity trends due to emotional fluctuations, social trends, or sudden events, which are difficult to effectively capture in simple linear models. Furthermore, linear regression assumes that there is no multicollinearity between independent variables, but in practice, features in user behaviour data are often highly correlated, which limits the applicability of linear regression (Montgomery et al., 2021).

2. Logistic regression in user engagement prediction

The main application of logistic regression in user engagement prediction is to deal with binary classification problems, such as predicting whether a user will like, comment or share a particular video. Logistic regression can provide probabilistic predictions of user engagement by mapping the independent variables to a probability value. The advantage is that the model structure is simple and suitable for binary decision-making problems such as 'whether a user will click on a video' (Peng et al., 2002). For example, logistic regression can be a good solution for predicting binary behaviours such as whether to like or not, whether to share or not, and can provide researchers with the key factors behind user behaviours by interpreting the regression coefficients.

However, the application scenario of logistic regression is limited to binary classification problems. When users' decisions exhibit complex multi-level structures, the predictive effect of logistic regression appears insufficient. For example, users may make decisions based on multiple factors such as content quality, duration, and personal emotions when watching a video, and these complex non-linear relationships are beyond the modelling capability of logistic regression. Menard (Menard, 2002) points out that logistic regression, while easy to use, scales poorly when dealing with multi-categorical and continuous prediction tasks. In addition, logistic regression assumes a limit of covariance between independent variables, which can also be an obstacle when

dealing with social platform data, as multidimensional features of user behaviour are often highly correlated.

3. Multiple regression models in user engagement prediction

Multiple regression models can handle the effects of multiple independent variables at the same time and are therefore widely used in prediction tasks involving multiple features (e.g., video length, upload time, historical user behaviour, etc.). Gareth et al. (Gareth et al., 2013) emphasise that multiple regression is able to capture the interaction effects between different variables, providing researchers with detailed information about the impact of multidimensional features on user behaviour. This approach is particularly suitable for complex datasets, such as when predicting video viewing behaviour by considering not only the video length, but also by incorporating the user's historical behavioural patterns and social interaction data.

However, a major challenge for multiple regression models is the problem of high-dimensional data and multicollinearity. In high-dimensional data, the interactions and dependencies between different independent variables are complex and difficult to control, which increases the complexity and instability of the model (Montgomery et al., 2021). The multicollinearity problem may lead to instability of the estimated regression coefficients, which in turn affects the prediction accuracy of the model. In addition, as the number of features increases, the model is prone to overfitting, i.e., the model performs well on training data but poorly on new data. This makes multivariate regression, despite its powerful modelling capabilities, require careful handling of feature selection and dimensionality approximation in practical applications.

From a critical thinking perspective, different types of regression models have different applicability scenarios and limitations when dealing with user engagement prediction. Linear regression, despite its simplicity and ease of interpretation, performs poorly when confronted with non-linear complex behaviour. In contrast, logistic regression provides an efficient solution to the binary classification problem but is also limited by the scalability of its model. While multiple regression is capable of handling multiple independent variables, its performance in high-dimensional data is vulnerable to multicollinearity.

To overcome these limitations, researchers can consider using regression models in conjunction with other techniques. For example, by applying regularisation techniques such as ridge regression or Lasso regression, the problem of multicollinearity in multiple regression can be effectively addressed and the robustness of the model can be enhanced (Tibshirani, 1996). Moreover, with the rise of machine learning and deep learning, more and more studies have shown that the combination of regression models with these techniques can further improve the accuracy and adaptability of user engagement prediction (Zhang et al., 2019). This will provide new directions for future applications of regression models, especially in the processing and explanatory enhancement of multidimensional user behaviour data.

2.4.2 Machine Learning in User Engagement Prediction

With the increase in data size and complexity of user behaviour, machine learning techniques have become a key tool in user engagement prediction. Compared with traditional regression models, machine learning techniques are able to handle complex nonlinear patterns and a large number of features, and in particular, techniques such as Random Forest, Gradient Boosted Decision Tree (GBDT), Support Vector Machines (SVM), and Deep Learning show good prediction performance. However, different machine learning methods have their own strengths and weaknesses when dealing with user engagement prediction, and their performance is often closely related to data features and task complexity. In addition, integrated learning techniques such as bagging are emerging as important tools to improve model robustness and performance.

1. Bagging algorithm

Bagging (Bootstrap Aggregating) is an ensemble learning approach that aims to improve the accuracy and robustness of predictions by reducing model variance. The core idea is to train multiple models on randomly sampled subsets of the original dataset (with replacement) and to aggregate their predictions through averaging (for regression) or majority voting (for classification). In this way, Bagging helps reduce overfitting and enhances the model's generalisation ability.

Random Forest is an integrated learning algorithm based on Bagging technology. It

makes predictions by training multiple decision trees and taking the average or majority vote of their results, which is particularly suitable for coping with the complexity and heterogeneity of user behaviour data. Breiman (Breiman, 2001) emphasises the superiority of Random Forests in dealing with high-dimensional data and capturing non-linear relationships, pointing out that it has a strong resistance to overfitting. In user engagement prediction, Random Forest can effectively combine data from different sources, such as video metadata and user engagement data, to improve the accuracy of the recommendation system.

Another application scenario of Bagging is to combine with Support Vector Machine (SVM). SVM performs well when the data size is small, but as the data size increases, the stability and computational efficiency of SVM will be affected. By combining Bagging, it is possible to train multiple SVM models and vote on their results, thus improving the performance of SVM on large-scale datasets. Dietterich showed that the combination of Bagging and SVM effectively reduces the variance of the model and improves the ability to handle complex data, especially in classification tasks (Dietterich, 2000). In user engagement prediction, Bagging can reduce the error of individual SVM models and enhance the stability of the prediction.

In addition, Bagging is widely used in regression tasks. In video recommendation systems, platforms need to predict users' future viewing behaviour or the popularity of videos. By combining Bagging with regression models, platforms are able to improve the prediction accuracy of continuous variables (e.g., viewing duration, video popularity, etc.). Breiman (Breiman, 1996) suggested that Bagging can effectively reduce the variance in regression models, especially when dealing with high-dimensional data, and exhibits strong robustness. For user engagement prediction on the YouTube platform, the Bagging technique can effectively improve the reliability of the prediction results, especially when faced with complex user behaviour patterns, as it can capture a wider range of data features and enhance the stability of the prediction.

2. Random Forests and Bagging

Random Forest as a representative of the Bagging technique, Random Forest can effectively deal with high-dimensional data and complex feature engagements by con-

structing multiple decision trees and taking the average or majority vote of their results to make predictions. Random forests are particularly well suited to cope with the complexity and heterogeneity of user behavioural data. Breiman points out that Random Forests not only possess strong resistance to overfitting, but also handle non-linear relationships in the data, which makes them perform well in user engagement prediction (Breiman, 2001).

Although Random Forest and Bagging can significantly improve the accuracy and stability of the model, the problem of interpreting its results remains a challenge. Especially when models generate large numbers of decision trees, researchers and platform operators may have difficulty understanding which features play a key role in the final prediction results (Louppe et al., 2013). This can lead to problems for scenarios where explanatory recommendations need to be provided, and thus need to be combined with model interpretation techniques to improve transparency.

3. Gradient Boosted Decision Tree (GBDT)

Gradient Boosted Decision Tree (GBDT) is another powerful integrated learning method that reduces prediction error by progressively optimising multiple weak learners. Compared to Random Forest, GBDT focuses more on the portion of prediction error by adjusting the weights of each learner in a targeted manner, and therefore performs more accurately in fine-grained prediction tasks. Chen and Guestrin (Chen and Guestrin, 2016) showed that GBDT is effective in reducing bias and outperforms traditional regression models when dealing with non-uniform data. For example, in YouTube user engagement prediction, GBDT effectively improves recommendation accuracy by combining users' historical behaviour, social engagement data and video metadata.

However, unlike Bagging, GBDT tends to increase bias to reduce variance during the optimisation of the model. As a result, GBDT may be overfitted on certain datasets, especially those with higher noise (Ke et al., 2017). Nonetheless, GBDT still performs superiorly on many platforms, for example, in YouTube's recommender system, GBDT significantly improves recommendation accuracy by combining users' historical behaviours, social engagement data, and video metadata.

4. Deep Learning and the Black Box Effect

Deep learning, especially neural network-based models, has excelled in handling large-scale data and complex pattern recognition in recent years. Covington et al. proposed a deep learning model in the YouTube recommender system, showing how non-linear patterns and long-term dependencies in user behaviour can be captured by deep neural networks (DNNs) (Covington et al., 2016). The deep learning model has strong nonlinear fitting capabilities and is able to perform complex feature extraction on user behaviour data through multi-layer neural networks, which in turn generates highly accurate user engagement predictions.

However, one of the main drawbacks of deep learning is the 'black box effect. Due to the complex structure of deep neural networks, the internal decision-making processes of the models are often difficult to interpret. LeCun et al. (LeCun et al., 2015) noted that this interpretive problem can be a significant drawback in certain application scenarios, especially when it is necessary to provide a basis for recommendations for users or platforms. In addition, deep learning's reliance on large amounts of data, as well as its higher computational cost, makes it perform less consistently than other machine learning methods when there is insufficient or noisy data.

2.4.3 Hybrid Model in User Engagement Prediction

In order to combine the explanatory nature of regression models with the predictive power of machine learning techniques, researchers have developed a variety of hybrid models, i.e., combining the advantages of traditional regression methods with machine learning algorithms. By integrating the advantages of different algorithms, these hybrid models are able to improve prediction accuracy, especially when dealing with complex user behaviour data. Hybrid models not only take advantage of the intuitiveness and interpretability of regression models but also leverage machine learning's ability to process large-scale, high-dimensional data, which in turn improves prediction results.

1. Regression + Deep Learning

Regression + Deep Learning is a common hybrid model combination that uses the results of regression analyses as inputs to a deep learning model to improve the accuracy of predictions while maintaining the interpretability of the model. This approach

typically starts by using traditional regression models (e.g. linear regression or multiple regression) to identify key features that influence user engagement, and then feeds these features into a deep neural network (DNN) to further capture the complex non-linear relationship between user behaviour and video popularity.

For example, when predicting the popularity of YouTube videos, researchers can use regression analysis to identify key independent variables such as video length, upload time, and user engagement behaviour, and then use a deep learning model to process the complex interactions between these features. Zhou et al.'s (Zhou et al., 2010) study demonstrated that a deep learning model can capture more complex patterns through multilevel feature extraction that improves prediction accuracy, especially in big data environments where deep learning has a strong ability to fit complex behavioural patterns. Such combined models are widely used in recommendation systems on large platforms (e.g., YouTube) to optimise user recommendations and content ranking.

However, despite the high prediction accuracy provided by deep learning, the "black box effect ' remains a major limitation. Deep learning models often lack interpretability, making it difficult for researchers and platforms to understand the specific reasons for model decisions. By combining with regression models, hybrid models can alleviate this problem to some extent, as regression analyses provide clear explanations of the effects of key variables. This interpretability is not only useful for researchers, but also provides users with a basis for personalised recommendations (LeCun et al., 2015).

2. Regression + Random Forest

Regression + Random Forest is another common hybrid model combination that utilises the explanatory nature of regression analysis and the non-linear processing power of Random Forest. In this approach, regression analysis can first identify important features that influence user behaviour, simplify the complexity of feature selection, and then process the non-linear engagements between these features through a random forest model to provide the final prediction.

Random Forest, as an integrated Bagging-based learning algorithm, is able to provide stable predictions by constructing multiple decision trees and averaging or majority voting on their results. By combining it with regression analysis, researchers are able

to first filter and interpret features through a regression model, and then use Random Forest to deal with the complex relationships and engagement of the features. Breiman notes that Random Forest is able to effectively deal with high-dimensional data and has a strong resistance to overfitting (Breiman, 2001). This combination of methods is particularly suitable for contexts with complex user behaviours and large amounts of data, such as user engagement prediction on social media platforms.

While this combination offers significant advantages in terms of prediction accuracy, it also increases the computational complexity of the model. Since Random Forest makes predictions by generating a large number of decision trees, the computational cost increases significantly as the amount of data increases. However, compared to deep learning, random forest models have better interpretability because variable importance analyses can be used to identify which features play an important role in the final prediction results (Louppe et al., 2013).

3. Combination of Bagging with other models

Bagging (Bootstrap Aggregating) is an integrated learning technique to improve model robustness and prediction accuracy. By back-sampling the dataset playfully, Bagging generates multiple models, reduces the variance of a single model, and improves generalisation by voting or averaging the results. While the combination of Bagging with Random Forests is the most well-known application, it can also be combined with other models such as Support Vector Machines (SVMs), Neural Networks and Decision Trees to further improve model performance.

(1) Bagging + SVM

Support Vector Machines (SVMs) are excellent at handling small datasets and complex classification problems, but scale poorly on large datasets and are computationally expensive. By combining with Bagging, multiple SVM models can be generated and their performance on large-scale datasets can be improved by voting or averaging the results.

Kim et al., (Kim et al., 2002) state that Bagging can significantly improve the performance of SVM on large-scale datasets, especially in tasks such as user behaviour prediction. SVM models incorporating Bagging not only reduce the variance of individ-

ual models, but also better capture complex nonlinear relationships in user behaviour. In user engagement prediction, the combination of SVM and Bagging helps to improve the accuracy of capturing different user behavioural features, thus improving prediction accuracy.

(2) Bagging + Neural Networks

Neural networks are excellent in handling large-scale data sets due to their powerful nonlinear modelling capabilities, but they are sensitive to data noise and prone to overfitting. Bagging can reduce the variance and overfitting problems of the models by training multiple neural network models and integrating their predictions, thus improving the robustness of the models.

Liu and Yao (Liu and Yao, 1999) showed that the combination of Bagging and neural networks can significantly improve model stability, especially when dealing with complex datasets containing noise. For recommendation systems on platforms such as YouTube, Bagging is able to integrate multiple neural network models, which in turn improves the accuracy of the recommendation system in capturing user engagement. This approach not only improves the generalisation ability of neural networks but also effectively reduces the sensitivity to individual anomalous data.

(3) Bagging + Decision Trees

The combination of Bagging with a single decision tree is also widely used. Single decision trees, although easy to interpret, are prone to suffer from their overfitting problem to the training data. With Bagging, predictions from multiple decision tree models can be integrated, which improves the robustness and stability of the model. Random Forest is actually an integrated decision tree model based on Bagging, which improves the overall performance of the model by constructing multiple decision trees and integrating their prediction results.

Ho (Ho, 1998) emphasised that Bagging can effectively improve the predictive power of decision trees, especially when dealing with data with a large number of feature engagement. Bagging can better capture complex patterns in the data. Compared to a single decision tree, Bagging combinations can effectively reduce overfitting to training data, thus improving generalisation performance to new data. For the task of

user behaviour prediction, the combination of Bagging and Decision Trees can provide highly stable and accurate prediction results, especially when user engagements are complex and data dimensionality is high.

Thus, the Bagging technique significantly improves the robustness and prediction accuracy of machine learning models by integrating the prediction results from multiple models. However, this improvement is also accompanied by a certain increase in computational cost, especially in the processing of large-scale datasets, where it is more expensive to train multiple models. In addition, although Bagging can improve the overall performance of a model, the gain of Bagging may not be sufficient to offset its computational cost on certain datasets. Therefore, future research could further optimise the efficiency of Bagging, especially when dealing with large-scale data, by improving sampling and model combination strategies to reduce computational complexity.

Beyond this, existing studies often focus on a single algorithm or model type, with limited comparison across diverse machine learning approaches. Moreover, the influence of different content feature types—such as textual versus numerical attributes—on model performance remains underexplored. To address these gaps, this study first investigates which machine learning methods demonstrate the strongest performance in predicting user engagement across various content feature dimensions.

2.5 Sentiment Analysis for YouTube Music Videos

2.5.1 Lexicon-Based Approach for Sentiment Analysis

There are two primary approaches to Sentiment Analysis (SA): lexicon-based approaches and those using machine learning techniques (Drus and Khalid, 2019). One type of unsupervised learning approach is the lexicon-based approach. The lexicon technique depends on a dictionary and does not require any training data. Most of the research using this approach adopts the TF-IDF or Sentiwordnet technique for sentiment analysis. This method is based on computing the frequency of keywords in the text data with other positive or negative words in pre-existing polarity lexi-

cons such as Sentiwordnet (Agarwal et al., 2015). The TF-IDF method is a statistical technique computed using the term frequency-inverse document frequency formula. It measures the importance of a word within a specific document relative to a corpus by assigning numerical weights based on its frequency and rarity (Das and Chakraborty, 2018). Unlike advanced embedding techniques such as Word2Vec or GloVe, TF-IDF does not capture semantic relationships between words but instead represents documents as sparse vectors of term importance. The effectiveness of the whole method is critically dependent on the quality of the lexical resources, which serve as the foundation of the procedures. Its premise is based on the idea that the polarity of the words within a written text can be determined through lexical analysis. However, due to the complexity of natural languages, this method has limitations and is not designed to address every nuance of language, such as slang, irony, or negation (Khan et al., 2016).

Some lexicons, such as LIWC (Linguistic Inquiry and Word Count) and Gthis study (General Inquirer), categorise words as positive or negative based on their context-free semantic orientation. LIWC contains about 4,500 words divided into 76 categories, including 905 terms in two categories specifically linked to sentiment analysis. Hutto and Gilbert (Hutto and Gilbert, 2014) generate and then experimentally validate a gold-standard sentiment lexicon that is particularly well-suited to microblog-like circumstances using a combination of qualitative and quantitative methodologies. LIWC was well-established and validated via over a decade of investigation by sociologists, psychologists, and linguists (Pennebaker et al., 2015). Despite its widespread use for sentiment analysis in social media material, LIWC excludes acronyms, initialisms, emoticons, and slang, all of which are essential components in sentiment analysis (Bonta et al., 2019). Other lexicons, such as ANEW (Affective Norms for English Words), SentiWordNet, and SenticNet, do, however, correlate sentiment intensity valence scores. SentiWordNet has 1,47,306 synsets annotated with three sentiment scores: positive, negative, and objective (Baccianella et al., 2010)

When analysts use a lexicon-based approach for sentiment analysis, some lexicon-based sentiment analysis tools have emerged. As illustrated in Asghar et al.'s study (see Figure 2.1), sentiment analysis on YouTube can be divided into event classifica-

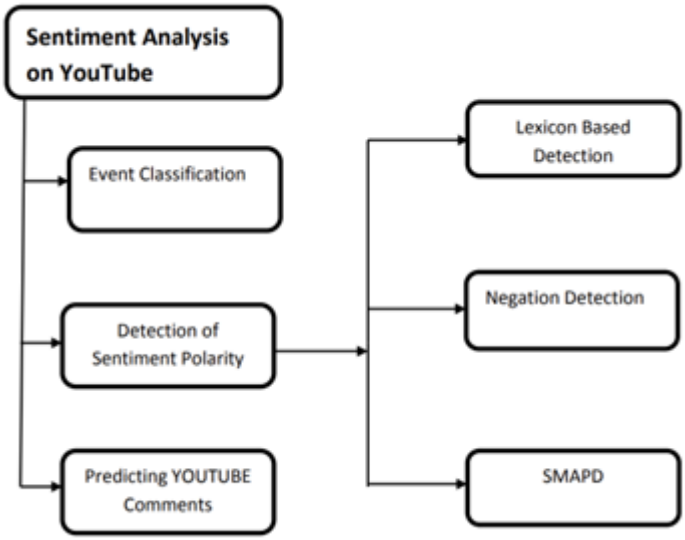


Figure 2.1: Sentiment analysis on YouTube (Asghar et al., 2015)

tion, detection of sentiment polarity and predicting YouTube comments based on the study's aims (Asghar et al., (Asghar et al., 2015), where SMAPD is shorthand for Social Media Aware Phrase Detection. While previous studies have proposed various sentiment analysis frameworks, this thesis specifically focuses on sentiment polarity detection. To support this, the Natural Language Toolkit (NLTK), an open-source Python library developed in 2001 at the University of Pennsylvania, is used for natural language processing tasks. It offers a user-friendly interface, more than 50 corpora, lexicon resources like SentiWordNet, and a variety of text processing tools for tokenisation, semantic reasoning, and classification. The sentiment score in NLTK is derived from SentiWordNet, which is composed of the polarity score of each WordNet synset with three sentiment numerical scores: positivity, negativity, and sum (1.0) for each synset. TextBlob: A Python module called TextBlob is used to process textual data for typical natural language processing (NLP) activities. It offers a standardised API [1]. Textblobs are similar to Python strings. VADER is a rule-based lexicon and sentiment analysis tool (Valence Aware Dictionary and sEntimentReasoner). When it comes to social networking, VADER Lexicon excels. The advantages of conventional sentiment lexicons

[1] https://textblob.readthedocs.io/en/dev/

such as LIWC (Linguistic Inquiry and Word Count) are preserved in VADER. It is more expansive, easily observable, comprehensible, rapidly implemented, and extensible. Human validation has assured the gold-standard quality of the VADER sentiment lexicon. VADER sets itself apart from LIWC by being more broadly favourable in other sectors and sensitive to sentiment expressions in social media contexts (Bonta et al., 2019). Therefore, this study adopts VADER as a tool for sentiment categorisation.

However, sentiment language alone is insufficient. Certain of the issues are that certain words have varied meanings depending on the context, that certain phrases with sentiment words may not express any sentiment, and that numerous sentences without sentiment words may also suggest an opinion (Akter and Aziz, 2016). The lexicon-based approach does, however, have several advantages of its own, including ease of use in counting positive and negative words, adaptability to many languages, and quickness of analysis completion.

2.5.2 Machine Learning Technique for SA

Supervised learning within the general machine learning set of approaches requires training data. The SVM and Naïve Bayes models are frequently used supervised learning techniques in the domain of sentiment analysis. Although support vector machines do well with low-dimensional datasets, Naïve Bayes works well with well-formed text corpora (Hassan et al., 2017). However, the machine learning approach fails on Facebook when users write at random lengths and with a lot of spelling errors (Belinkov and Bisk, 2017). To adjust the approach, a large number of training samples are needed because the size and quality of the output are influenced by the size of the dataset (Mahtab et al., 2018). Moreover, if training is needed, machine learning analysis can take hours in the sophisticated machine learning model (Chekima and Alfred, 2018). Although the procedure is quicker when the training sample is smaller, the classification accuracy suffers (Dhaoui et al., 2017).

Interestingly, according to researchers, the accuracy of both analytic techniques is somewhat comparable. To forecast the sentiment direction, two methods, lexicon-based sentiment classification with a sentiment scoring function and Naïve Bayes multinomial

event models, can be combined. Studies have demonstrated that combining both techniques offers higher efficiency than depending only on one (Medhat et al., 2014; Mullen and Collier, 2004). Combining the two techniques is therefore advised to enhance the result because they will complement one another, and the results are better than employing one approach alone. Combining methods can help to recognise a phenomenon (Dhaoui et al., 2017). Unstructured data management can also be improved by it (El Rahman et al., 2019).

These algorithms are capable of automatically picking up a wide range of features for classification via optimisations. Frequently, the sentiment classifier is ineffective when applied to different domains because it was trained on labelled data from one domain. Lexicon-based techniques are suggested as a solution to the issue (Bonta et al., 2019). In addition, machine learning analysis is time-consuming; hours can pass when using a sophisticated model, particularly if training is necessary (Chekima and Alfred, 2018). A smaller training dataset can speed up the procedure, but the classification accuracy suffers as a result (Dhaoui et al., 2017).

2.5.3 Sentiment Analysis on YouTube Comments

While the majority of research studies tend to concentrate on analysing comments made on social media, there are a few scholarly publications that specifically try to examine comments on the YouTube platform.

Khan et al.investigate the challenges of sentiment analysis in complex natural language, using YouTube comments as their primary data source. The study highlights the limitations of traditional techniques in handling nuances such as slang, irony, and negation. By analysing YouTube comments, the authors demonstrate the need for combining linguistic resources and machine learning approaches to improve sentiment classification accuracy in unstructured and informal text data (Khan et al., 2016).

In her study, Yee et al. (Yee et al., 2009) demonstrate the integration of user-generated comments into the search index, resulting in a notable enhancement in the precision of search results. In the absence of manually annotated data, these algorithms depend on automatically generated approximations of sentiment in comments. For

instance, Siersdorfer et al. (Siersdorfer et al., 2010)concentrate on utilising user rating counts, such as 'thumbs up/down' indicators provided by other users, in YouTube video comments. They employ these indicators to train classifiers that can predict the level of acceptance within the community for new comments.

Al-Tamimi, Shatnawi, and Bani-Issa (Al-Tamimi et al., 2017) focus on Arabic sentiment analysis of YouTube comments, addressing the challenges posed by the complexity of the Arabic language. The study develops a machine learning-based approach, utilising features specific to Arabic, such as morphological normalisation and dialect handling, to classify comments into positive, negative, or neutral sentiments. The findings highlight the effectiveness of customised preprocessing techniques for Arabic text and demonstrate the potential of sentiment analysis in understanding user engagement on YouTube.

YouTube videos are often ranked in search results based on traditional metrics such as the number of views or likes. However, these metrics sometimes allow irrelevant or low-quality videos to rank higher, leading to a suboptimal user experience. Bhuiyan et al. (Bhuiyan et al., 2017) propose a sentiment analysis method using NLP techniques to analyse user comments and address this issue. By incorporating sentiment analysis, their approach evaluates user feedback to identify the most relevant, popular, and high-quality videos more effectively. Data-driven experiments demonstrate that the proposed method enhances the accuracy of retrieving pertinent videos, ensuring better alignment with user expectations and supporting creators in improving their content quality.

Muhammad et al. (Muhammad et al., 2019) performed SA on instructional YouTube videos in Indonesia. Both NB and SVM were used to classify the gathered data into positive and negative groups. In other words, NB was used to convert the words into vectors that the SVM algorithm later employed and to calculate the likelihood of each word occurring. The suggested model obtained 83% recall, 87% f1 score, and 91% precision score.

However, most existing studies focus on sentiment analysis within a single platform, and little is known about whether users express similar sentiments on the same topic

across platforms such as YouTube and Twitter. This limits our understanding of cross-platform sentiment dynamics.

2.5.4 Existing Datasets for User Engagement Prediction

A number of datasets have been developed to facilitate research on online content popularity and user engagement. The *YouTube-8M* dataset (Abu-El-Haija et al., 2016) is one of the most widely used resources in this domain. It contains millions of video-level examples with pre-extracted visual and audio features, but it lacks detailed user interaction data such as comments, likes, and temporal information about engagement behaviour. Similarly, the *YouNiverse* dataset (Biel and Gatica-Perez, 2013) provides multimodal features for vlogs on YouTube but is limited in scope and does not focus on music video content.

For sentiment and engagement analysis on social platforms, many studies rely on data from the Twitter Streaming API (e.g., Jansen et al., 2009; Stieglitz and Dang-Xuan, 2013), which offers access to real-time tweets. These datasets are commonly used for event detection or political discourse analysis, but rarely focus on entertainment domains like music or support comparative studies across platforms. In addition, most Twitter datasets are either too domain-specific or anonymised to a level that restricts detailed cross-referencing with video content.

Another limitation is that existing datasets tend to focus on a single platform, either YouTube or Twitter, and thus do not allow for cross-platform analysis of user behaviour or sentiment consistency. Moreover, there is a lack of datasets that combine textual content (e.g., comments, titles), numerical features (e.g., views, likes), and time-related variables (e.g., upload date, comment date) in a structured format suitable for engagement prediction tasks.

Given these limitations, there is a clear need to construct a new dataset that integrates multimodal and cross-platform data related to music video content. Such a dataset can support more comprehensive analyses of engagement behaviour, facilitate comparative studies across social media platforms, and enable the evaluation of machine learning methods using richer, context-aware features.

2.6 Gaps and Challenges in Existing Research

Research on user engagement prediction on platforms such as YouTube has made significant progress, but there are still challenges in data integration, prediction accuracy, and model generalisability. One of the aims of the current study is to discover suitable feature combinations and prediction models to address these shortcomings in the context of music videos on YouTube and test different regression models.

2.6.1 Limitations of Existing Studies

1. Limitations of Multi-Data Source Integration and Feature Selection

Current user engagement prediction studies often rely on a single type of data (e.g. video metadata or user behavioural data), which limits the performance of predictive models, especially in the music video domain, where user engagements are influenced by multiple factors, including emotional resonance, video content, and social engagements. McAuley et al. (McAuley et al., 2015) note that the integration of multiple data sources is crucial in social media analytics. Integrating multimodal data (e.g., textual, audio, visual features) can improve predictive accuracy, but its complexity increases the computational cost and difficulty of model development. Their study was able to capture different dimensions of user behaviour more comprehensively and optimise prediction accuracy by exploring different feature combination strategies such as user comments, video metadata and social engagement data.

2. Challenges of model generalisability and long-tail content

Existing recommender systems usually tend to push popular content while ignoring user demand for cold content. Celma (Celma Herrada et al., 2009) points out that the handling of long-tail effects is difficult in existing models, especially in music videos, where user interest in niche music or cold content is more difficult to capture by mainstream recommendation algorithms. By testing different regression models combined with multi-feature combination strategies, the research can better accommodate the needs of personalised and long-tail users, especially in the prediction of cold content, which can enhance the generalisation of the model and improve the coverage of diverse

user groups.

3. Computational complexity and real-time issues

As the amount of data on the platform grows, existing prediction models often struggle to meet real-time requirements due to excessive computational complexity when dealing with large-scale data. Covington et al. (Covington et al., 2016) show that deep learning methods, despite their excellent performance when dealing with complex data, demand for computational resources makes real-time applications difficult, especially when dealing with large amounts of user behaviour data. By testing the computational efficiency of regression models (e.g., linear regression, ridge regression, etc.), this study was able to identify models that perform more efficiently with large-scale data, thereby reducing computational overhead and improving real-time responsiveness while maintaining high prediction accuracy.

Most studies focus on improving prediction accuracy using enhanced features or algorithms, but few address the scalability or interpretability of these models in real-world settings.

Furthermore, most existing studies train and evaluate predictive models on a single platform—such as YouTube or Twitter—without considering how model performance may vary across different types of datasets. This lack of cross-platform comparison limits our understanding of model adaptability and generalisability across heterogeneous data sources.

Building on this, another underexplored area is the potential link between user sentiment and the content categories assigned by video creators. While platforms often rely on these categories to structure content and drive visibility, their relationship with audience emotional reactions has received little empirical attention. This motivates further investigation into whether content categorisation influences the sentiment heterogeneity observed across platforms.

2.6.2 Future Research Directions

While existing research has compared user engagement patterns across platforms like YouTube and Twitter in political or crisis-related contexts, relatively little atten-

tion has been paid to cross-platform differences in user engagement with music-related content, particularly music videos. The unique characteristics of music consumption and community interaction on different platforms remain underexplored, highlighting the need for comparative studies in this domain.

1. Optimising multi-data source integration and feature combination strategies

Future research should further optimise the integration of multiple data sources, especially in fusing sentiment analysis, social engagement data and video metadata, and explore the optimal combination strategy of multimodal features. Zhang et al. (Zhang et al., 2019) investigated how the combination of multimodal features can improve the performance of recommender systems, and pointed out that the combination of different features can enhance the predictive ability of the model. A similar integration strategy can be used for music video recommendation to further improve the accuracy of user engagement prediction.

2. Personalisation and long-tail user prediction

To address the issues of personalisation needs and long-tail user engagements, future research can explore how to dynamically adjust feature combinations through more flexible models to adapt to the personalisation needs of different users. Jannach and Adomavicius (Jannach and Adomavicius, 2016) pointed out that personalised recommender systems need to introduce more detailed feature modelling when dealing with long-tailed content, and in particular, by dynamically analysing user behaviours and interest changes, the recommendation accuracy for long-tail content can be improved. In this study, the combination of features tested with different regression models can provide a more customised recommendation scheme for personalised needs.

3. Application of real-time prediction and online learning

Future recommender systems should be dynamically responsive and update user models in real-time. Hoi et al. (Hoi et al., 2021) proposed that online learning algorithms are able to adjust model parameters on the fly to improve the relevance and real-time performance of recommended content, especially in environments where user engagements change rapidly. Combined with the testing of multiple regression models

in this study, the combination of online learning techniques with regression analysis can be further explored in the future to improve the real-time responsiveness of music video recommendation systems.

2.7 Summary

Through the literature review, this study clearly shows the key role of different data sources and prediction methods in YouTube music video user preference prediction. First, the integration of multiple data sources (including video metadata, user comments, social engagement data, etc.) plays a central role in capturing the complexity of user behaviour. Studies have shown that a single data source (e.g., viewing duration or number of likes) is often insufficient to accurately predict user preferences, especially in the music video domain, where emotional resonance and social engagements become important complementary data sources. Through feature combination strategies, research can better capture the diverse behavioural patterns of users in viewing, interacting and sharing.

Secondly, regression models provide powerful tools in user preference prediction, especially when dealing with complex relationships in user behaviour data. Different regression methods (e.g., linear regression, ridge regression, Lasso regression, etc.) have their strengths and weaknesses in capturing the relationship between user behavioural characteristics and video popularity. For simple linear relationships, traditional regression models perform well, but hybrid regression models and machine learning methods (e.g., deep learning, random forests) demonstrate stronger generalisation capabilities when faced with multidimensional complex engagements. Future research directions will continue to work on improving the performance of user preference prediction models and further integrating diverse data sources. First, with the increasing abundance of social media data, future research should enhance the integration of multimodal data, such as combining video metadata, user behaviour data and sentiment analysis data more closely, so as to construct a more comprehensive user portrait.

Finally, personalised recommendations and prediction of long-tail content remain a major challenge for future research. By further developing hybrid models, especially

hybrid methods that combine regression analysis and deep learning, researchers can better capture personalised needs and improve recommendation accuracy when dealing with long-tail content. Personalised recommender systems need to adapt to users' diverse interests, especially preferences for niche and cold content, through accurate feature selection and dynamic adaptation.

In summary, future research should continue to conduct an in-depth exploration of data source integration, model performance enhancement and personalised recommendation, to build a more intelligent and efficient recommender system to provide users with more accurate and personalised content recommendations.

Against this background, the present study seeks to address several key gaps identified in the literature—namely, the limited comparative evaluation of machine learning methods, the underexplored role of time- and context-related features, and the lack of cross-platform analysis of user engagement and sentiment in the music video domain. By investigating these issues through a multi-perspective empirical approach, this research aims to contribute new insights into user behaviour modelling and predictive system design in social media environments.

Chapter 3

Data Collection and Methods

3.1 Introductions

This chapter describes how this study obtained the tweets, YouTube videos, and YouTube video-sharing events on Twitter in detail.

In evaluating user engagement across various social media platforms, a multitude of metrics can be employed, including but not limited to the number of likes, the volume of positive user comments, and the frequency of retweets. Such metrics, inherently platform-specific, offer a nuanced understanding of user engagement and the engagement of content, particularly videos. This analysis aims to quantify these dimensions by considering key indicators: the total number of video views, likes, comments, and the aggregate of comments received. Here, the number of views serves as a primary indicator of user interest and engagement with the content. In parallel, the volume of comments acts as a proxy for the level of user engagement, thereby providing insights into the extent of engagement with the content. This section elucidates the methodologies employed in capturing tweets, YouTube videos, and instances of YouTube video sharing on Twitter, thereby framing the chosen approach to assessing user engagement and sentiment analysis within these digital environments.

YouTube serves as the global online video-sharing and social media platform through which millions of videos are uploaded and billions of people watch, upload and comment on music videos. Basically, the idea of YouTube data collection here depends on the

topic which is related to music videos. Using a variety of human coding and computer-aided methods via API, this study identified a range of popmusic topic-related music videos from YouTube to set the datasets, and at the same time by looking for instances in which URLs to videos were shared on Twitter, as well as the related content created for those popmusic topic-related videos through the use of shared hashtags to setup the dataset. The data collection includes: (1) the results of a search for pop music-related channels on YouTube, based on which this study extracted its music videos; and (2) YouTube videos extracted from tweets that are relevant to the same set of pop music-related keywords and hashtags on Twitter. By using NLP techniques, the thematic keywords contained in the dataset of all YouTube music videos extracted from Twitter,i.e, T1 (N = 2,330) are merged with the keywords identified in the dataset of YouTube music videos crawled from the YouTube channels,i.e., Y1 (N = 3,360) to create a link between the two datasets. A detailed explanation of the dataset construction process will be provided in the following sections of this chapter. The dataset developed for this study is considered one of its key contributions, as it addresses the lack of publicly available datasets focusing on cross-platform engagement with music video content.

In this chapter, this study focuses on two parts: the data collection strategies, as well as the algorithms that have been applied in Chapters 5, 6, and 7. First the Data collection strategies are presented for learning how to build the datasets. Then this study described different algorithms in natural language processing (term frequency, TF-IDF and Word2Vec) and predictive modelling algorithms including Random Forest (RF), Gradient Boosting (GB), and Bagging Regression (BR), which will be subsequently applied to each type of model in Chapters 5 and 6, and the sentiment analysis methods introduced in Chapter 7, respectively.

3.2 Data Collection and Preprocessing Strategies

In this section, the data collection and data processing strategies are introduced, which include Twitter and YouTube data collection and the data cleaning process.

3.2.1 Twitter Data Collection and Preprocessing Strategy

(1) Twitter data collection strategy

To keep the collected data from Twitter and YouTube data topically relevant, the relevant keywords (i.e., pop music, popmusic, #popmusic #pop music) were used as search index and hashtag within Twitter.

Based on the policy and limitations of the Twitter API, the data collection code requires manual modification of the collection time and runs every seven days. Hence, as Figure 3.1 presents, the data collection work is conducted every 7 days with related content, location, username, and the number of likes of each tweet. The data collection spanned from 3 December 2022 to 11 January 2023, which means the collection work was run 5 times during this period.

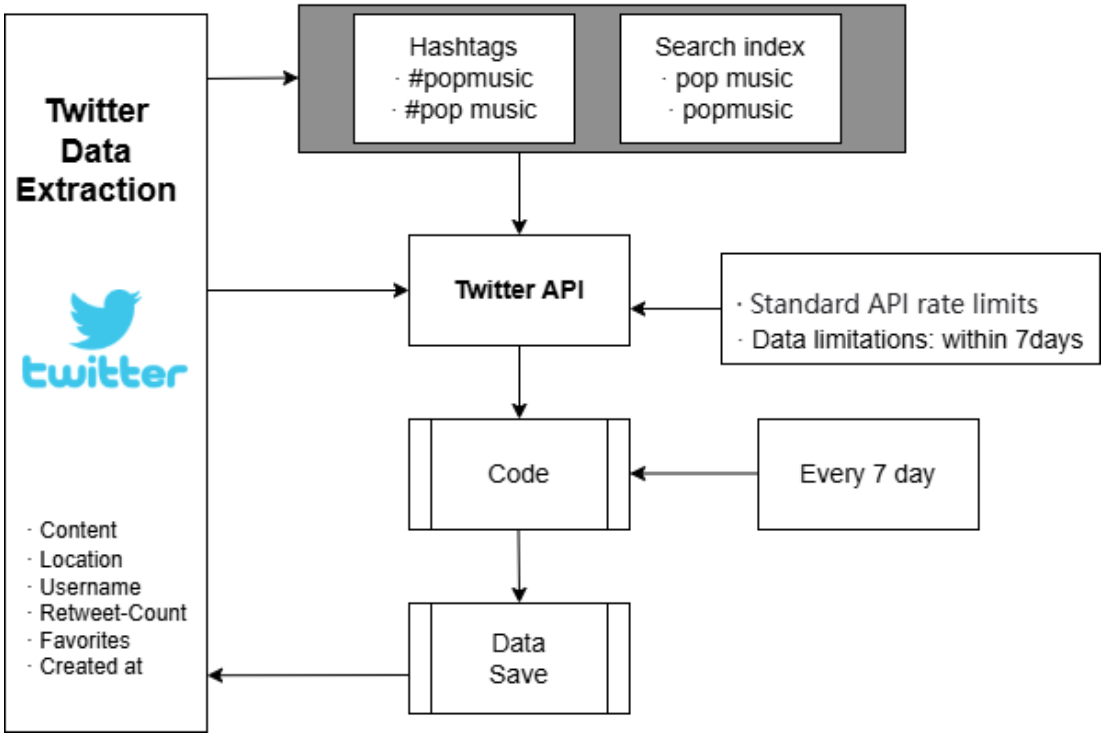


Figure 3.1: Twitter data extraction flowchart

The parameters collected are shown in Table 3.1. As a result, this study saved a dataset consisting of 101,427 rows and 8 columns of tweets. This study focuses on whether the content of the tweet is relevant to this research topic and whether the

variables	Num.	describe
Content	101,427	The text content, containing the main information of the tweet.
Location	101,427	The geographic location of the published content, if available.
Username	99,694	The name of the user who posted the content.
Favorites	101,427	The number of times the content has been liked by other users.
ID	44,774	Unique identifier of the content, can be the ID of the tweet, etc.

Table 3.1: The variables information of the tweets dataset

tweet contains a link that leads to a YouTube video, and then the data processing part of this study is and will be focused on the processing of the content of the tweet.

(2) Twitter data preprocessing strategy

The Twitter data preprocessing strategy based on **Content** columns mainly processes textual data by identifying and qualifying hyperlinks in the textual data so as to collect YouTube links from tweets pointing to them. The whole process is shown in Figure 3.2, which includes removing duplicate lines, removing mentions and hashtags from the tweets, identifying hyperlinks in the tweets, confirming the type of hyperlinks (where they are pointing to), and finally confirming hyperlinks pointing to YouTube, and storing the corresponding tweet information.

It is worth noting that this study only uses the tweet information in the crawled Twitter dataset for collecting hyperlinks pointing to YouTube, to further collect data related to YouTube music videos.

3.2.2 The Video dataset collection strategy based on Tweets (Tweet video links dataset, T1)

Tweet video dataset is a dataset consisting of YouTube videos from the collection of YouTube links pointing to YouTube videos contained in tweets. As shown in Figure 3.3, the links come from the collected tweets, which are contained within the Content column in the Twitter dataset.

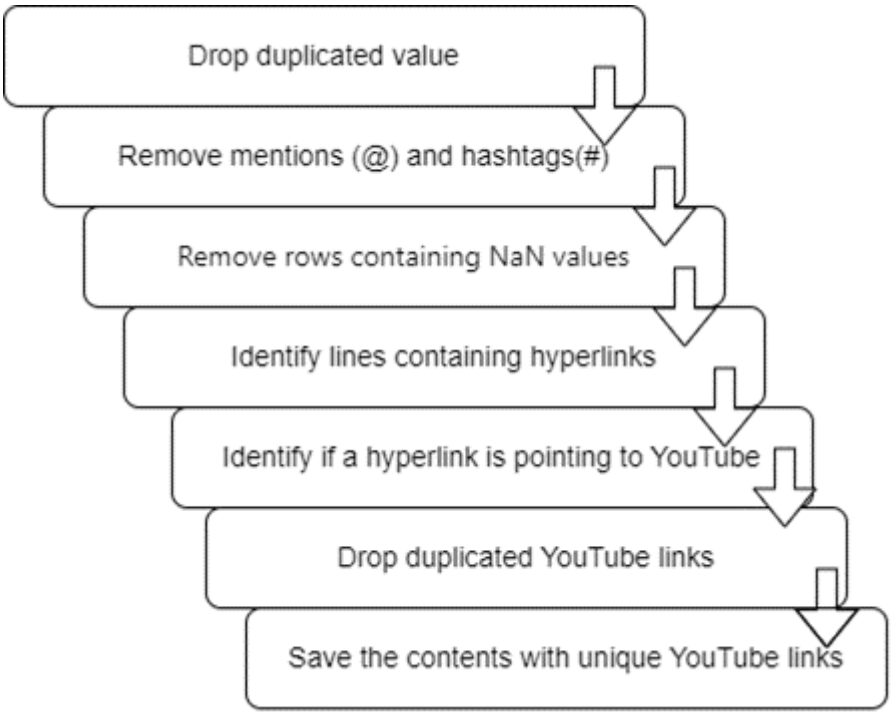


Figure 3.2: Twitter text data cleaning process

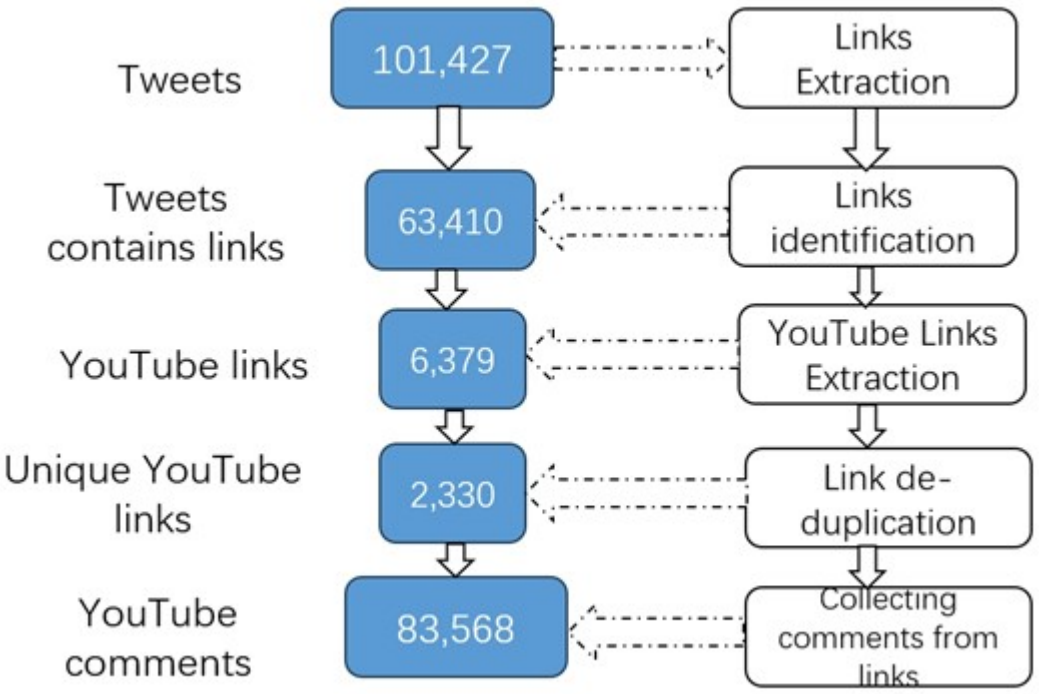


Figure 3.3: The strategy of collecting YouTube comments via Twitter-shared video links

It is crucial to distinguish between the YouTube music video data sourced via Twitter hyperlinks and the T-video dataset, which will be discussed subsequently. The former dataset comprises YouTube music videos systematically aggregated through Twitter hyperlinks. This collection underwent a comprehensive process of crawling and cleaning to ensure the integrity and relevance of the data. This differentiation lays the groundwork for the analysis, emphasising the methodological rigour applied in the compilation and preparation of the YouTube music video dataset derived from Twitter.

3.2.3 YouTube Data Collection and Processing (YouTube video dataset, Y1)

(1) YouTube data collection strategy

The YouTube data collection strategy in this study focuses on music videos, aligning with the research topic. Unlike the Twitter data collection approach, which is constrained by time limitations (e.g., access to tweets from only the past seven days in the free API version), the YouTube API allows access to videos regardless of their upload date, enabling a broader temporal analysis. Therefore, a strict time frame was not applied for YouTube data collection, unlike Twitter. Nevertheless, the process of collecting YouTube data was concluded at the same time as the Twitter data collection.

However, there are more than 40 different categories of channels that are available on YouTube, for example, gaming, entertainment, sports, pets, etc. To ensure that the collected data specifically related to music videos, this study first identified relevant YouTube channels by using 'pop music' as the search keyword, aligning the dataset with the research focus. Instead of relying on a predefined YouTube category, a keyword-based search was conducted to locate relevant channels. To enhance data diversity, ensuring that the collected music videos were not biased toward a single artist, the study excluded official YouTube channels of pop music celebrities. Instead, the six channels with the most subscribers in the pop music category were selected to provide a broader representation of music content.

The data collection of the six popmusic-related YouTube channels is shown in Figure

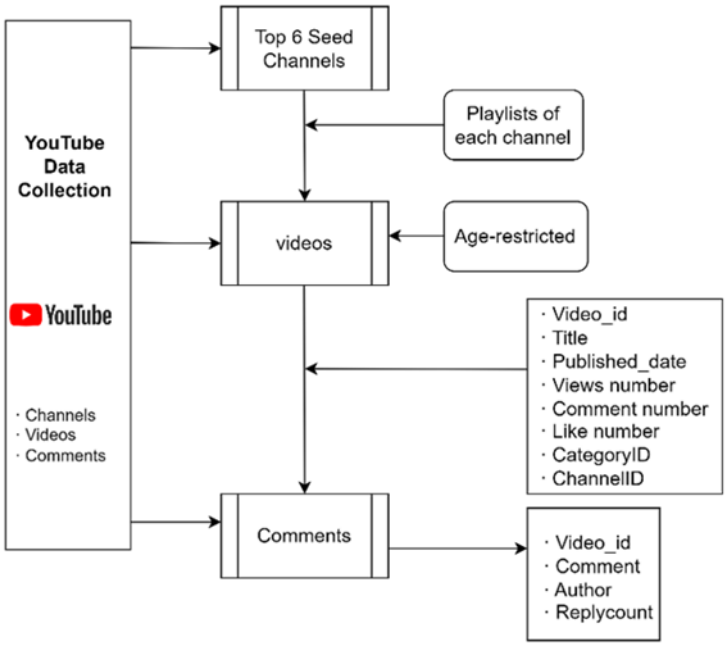


Figure 3.4: YouTube data extraction workflow

3.4. Specifically, according to the playlist ID from each channel, the video layer dataset beneath each channel is captured by utilising each channel's playlist ID, and then, according to the video ID, the comment layer dataset is collected. From Table 3.2, the basic information about these 6 YouTube channels can be learned: The number of subscribers for each of these six YouTube pop music channels suggests their potential audience reach, reflecting their popularity within the platform. At the same time, the study was also able to gather as much data as possible regarding the videos that present pop music topics and as many reviews and comments as feasible through these 6 channels. Table 3.2 presents the basic statistical information of the six selected pop music channels on YouTube.

Channel Name	**Subscribers**(approximate)	**Views**	**Total Videos**	**Channel ID**
Pixl Networks	4,110,000	1,703,122,834	71	UU1iqebKNH36JIdBIjEy8-iQ
Epidemic Pop	1,000,000	292,751,113	1,837	UUzMxEa-lDX2AfgotszScOFg
PopCrush	396,000	196,915,421	613	UUWxt8IN-Uhrpz9RxNV1Of3A
The Pop Song Professor	190,000	21,043,367	676	UUFhmJZnna3LlznyBEooxTDQ
Make Pop Music	172,000	9,745,556	234	UUvARrwO4x0VInVZr4pQRB2A
N&D Music Mashup	68,900	20,663,538	79	UU34k8ID4P-xh4k7OB1njbwQ

Table 3.2: Summary information of the six pop music channels on YouTube

The basic information of the 6 YouTube channels can be seen in Table 3.2. Utilizing the playlists from the aforementioned six channels, this study extracted videos associated with them. The comprehensive YouTube dataset is structured into two layers: videos and video comments. The data collection methodology employed the YouTube API. This process culminated in an initial set of 3,360 unique YouTube video IDs, and based on these IDs, the comments under these videos were further captured and stored. As a consequence, 152,182 corresponding comments were saved.

(2) YouTube data processing strategy

In this part, the preprocessing of the comments layer data includes the cleaning process of the whole comments data (see Figure 3.5). The strategy applied to the YouTube comments level dataset not only from the YouTube data collection strategy but also from the links based on tweets, this study described later. As shown in Figure 3.6, the video layer information includes **video_id**, **comment**, **author**, **replycount**. Notably, the column of **author** describes the username of the person who commented. And the **replycount** is the number of replies to that comment, not the number of replies to this video. The **comment** column describes the comments associated with each video.

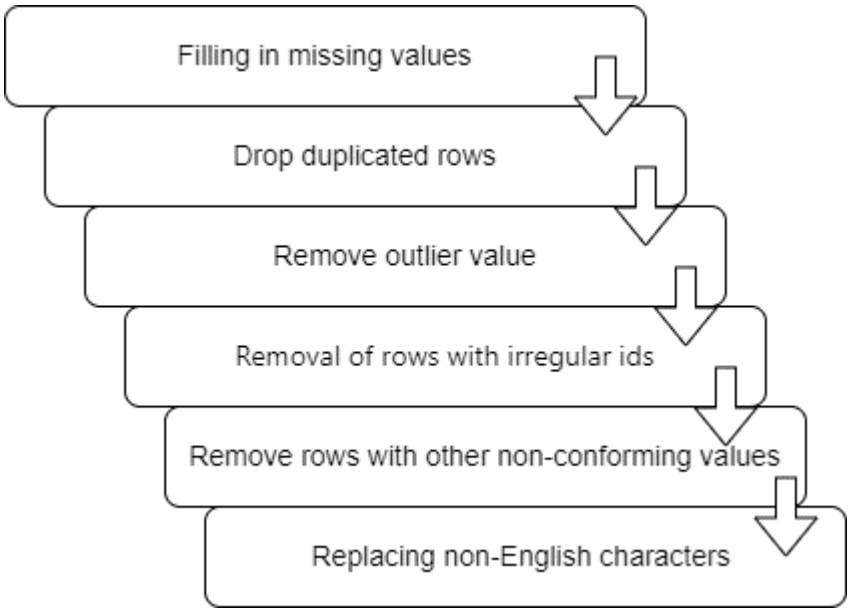


Figure 3.5: YouTube Comments data cleaning process

	video_id	comment	author	reply_count	reply_time
0	136-BtyKfqs	This is honestly pretty sad to me. I've be...	Athanasios Papadopoulos	1	2022-06-28 14:22:03+00:00
1	136-BtyKfqs	Retirement is going back and reliving the nost...	JLAndersonMusic	1	2022-06-28 15:24:02+00:00
2	136-BtyKfqs	Interview Tyler and ask him which song meaning...	DarkEchoes	1	2022-06-28 14:43:40+00:00
3	136-BtyKfqs	Man I remember three years ago in my trench er...	DictatorKenji	0	2022-12-05 00:23:11+00:00

Figure 3.6: Example of a YouTube comment layer

It is imperative to clarify that the YouTube videos initially collected in this study represent the broader dataset obtained directly from YouTube, whereas the Y-video dataset, examined later, consists of a refined subset selected based on overlapping topics with the T-video data. The Y1 (N = 3,360) dataset presented herein is larger than the YouTube music video dataset T1 (N = 2,330) derived from Twitter hyperlinks. However, its broader scope does not inherently align with the thematic focus of the latter. At present, this variance highlights the diversity in content between the two collections, with the larger dataset encompassing a wider range of topics, whereas the Twitter-linked dataset T1 remains more specialised within the realm of music video discussions.

3.2.4 The Global Dataset (T&Y-video dataset)

During the data collection and preprocessing steps, this study constructed two initial datasets: **(1) T1**, a YouTube music video dataset derived from Twitter hyperlinks, and **(2) Y1**, a YouTube music video dataset obtained by crawling six YouTube music channels. To establish a global dataset (T&Y-video) based on common topics, the following steps are performed (see Figure 3.7):

1. Feature Extraction: For each music video in T1 and Y1, natural language processing (NLP) techniques are applied to extract keywords from three descriptive fields: title, tags, and description. Define the set of extracted keywords for each video

v as $K(v)$, where $K(v) = \{k_1, k_2, ..., k_m\}$.

2. Topic Matching: For each video $v_i \in T1$ and each video $v_j \in Y1$, compute the intersection of their keyword sets:

$$I(v_i, v_j) = K(v_i) \cap K(v_j)$$

If $I(v_i, v_j) \neq \emptyset$, the two videos are considered to share a common topic.

3. Dataset Construction: A T&Y-video dataset is created by selecting all videos in T1 and Y1 that have at least one keyword match with a video in the other dataset. A subset of T1 and a subset of Y1 belonging to T&Y-video are T-video and Y-video.

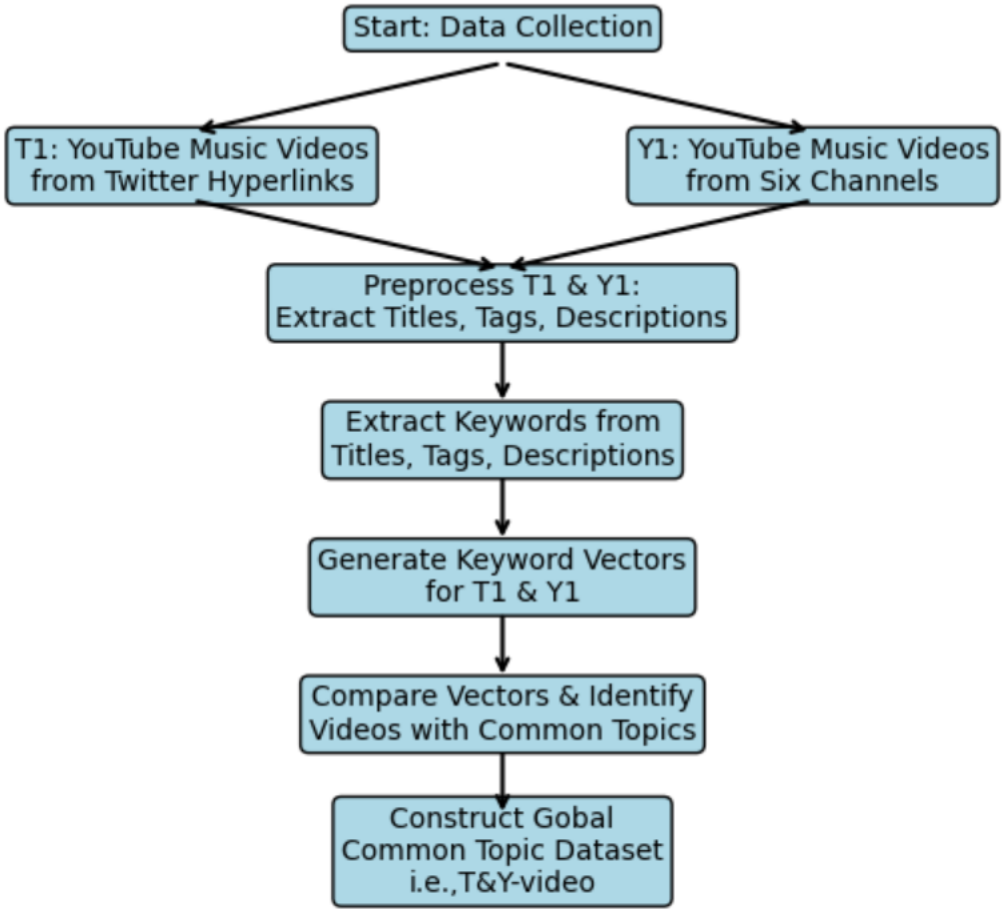


Figure 3.7: The T&Y-video dataset generation process

3.3 Prediction Methods

In this section, the methods that will be used in chapters 5 and 6 will be introduced. In order to fulfil the objectives of the study, ensemble machine learning techniques including Random Forest (RF), Gradient Boosting (GB) and Bagging Regression (BR) were implemented using Python scripting. Machine learning techniques are commonly employed to predict desired outcomes based on input characteristics. The evaluation of the training dataset and the test dataset involves comparing the Root Mean Square Error (RMSE) and Coefficient of Determination (R^2) values for each algorithm. Additionally, the 10-fold cross-validation method is used to provide a more accurate validation of model performance. The final method is determined by combining the results obtained from the aforementioned evaluations.

3.3.1 Preliminary Comparison of Machine Learning Models

In order to determine the most suitable models for predicting user engagement, a preliminary comparison was conducted between several machine learning methods, including linear models (Lasso and Ridge regression) and ensemble models (Bagging). Lasso and Ridge regression were used to examine linear relationships and extract interpretable coefficients. In contrast, Bagging allowed for evaluating non-linear feature importance across the two datasets.

Table 3.3: Feature Influence Comparison Across ML Models

Feature	**Lasso Coef.**	**Ridge Coef.**	**Bagging Imp. (Yvideo)**	**Bagging Imp. (Tvideo)**
Likes	**193,146,120**	**192,577,590**	–	–
Time	22,127,114	17,442,145	–	–
AM_PM_Flag	–14,597,846	–11,824,732	–	–
Day_Num	–2,125,798	–4,885,792	0.0016	–
CategoryID	–3,063,243	–6,012,121	0.0225	0.0024
reply_count	–2,764,514	–2,943,370	0.0046	0.0915
cross_platform	1,570,930	132,940	–	–
Views	–	–	**0.8589**	**0.9061**

To assist in selecting appropriate machine learning models for user engagement prediction, this study tested Lasso, Ridge, and Bagging models on both the YouTube and Twitter datasets. Table 3.3 displays the coefficients (for linear models) and feature

importances (for ensemble models) across these approaches.

As shown, Lasso and Ridge both assign very high coefficients to 'Likes' and 'Time', suggesting their dominant role in linear relationships. However, Bagging models demonstrate that 'Views' carries overwhelming importance in non-linear contexts, especially in both datasets, indicating a strong generalisation. These results support the decision to adopt tree-based ensemble methods in the main experiments.

3.3.2 Bagging Regressor Approach

The Bagging Regressor (BR), proposed by Leo Breiman (Breiman, 1996), is a representative method of ensemble learning that operates in a parallel manner. In this study, BR refers to a bagging-based regression model that aggregates multiple base learners trained on different bootstrap samples to improve predictive accuracy and reduce variance. As a supervised learning technique, BR trains multiple base models independently on different bootstrap samples of the training data. Each bootstrap sample contains M instances randomly drawn with replacement from the original dataset, meaning that some samples may appear more than once while others may be omitted.

In this study, Bagging is implemented using the BaggingRegressor from Scikit-learn, with default settings. Specifically, the base estimator is a DecisionTreeRegressor, which is used to train multiple decision trees on bootstrapped subsets of the data. This method is well-suited for high-dimensional data and complex feature interactions in user behavioural prediction tasks, particularly for modelling user engagement. In this context, the number of Likes on a video is used as the prediction target and serves as a proxy indicator of user engagement.

By combining the outputs of multiple models, bagging reduces model variance and helps prevent overfitting. Its effectiveness stems from the diversity among the individual models trained on different subsets of the data, whose prediction errors can offset each other during aggregation. The ensemble prediction is typically obtained by averaging the outputs of all base regressors.

Furthermore, this approach allows for a more robust estimation process by leveraging complementary data subsets. Figure 3.8 illustrates the flowchart of the Bagging

Regressor (BR) algorithm, adapted from Zou et al. (Zou et al., 2022), where the original "classifier" components have been relabeled as "regressor" to reflect the regression-based application in this study. The diagram outlines the main stages leading to the final prediction of user engagement.

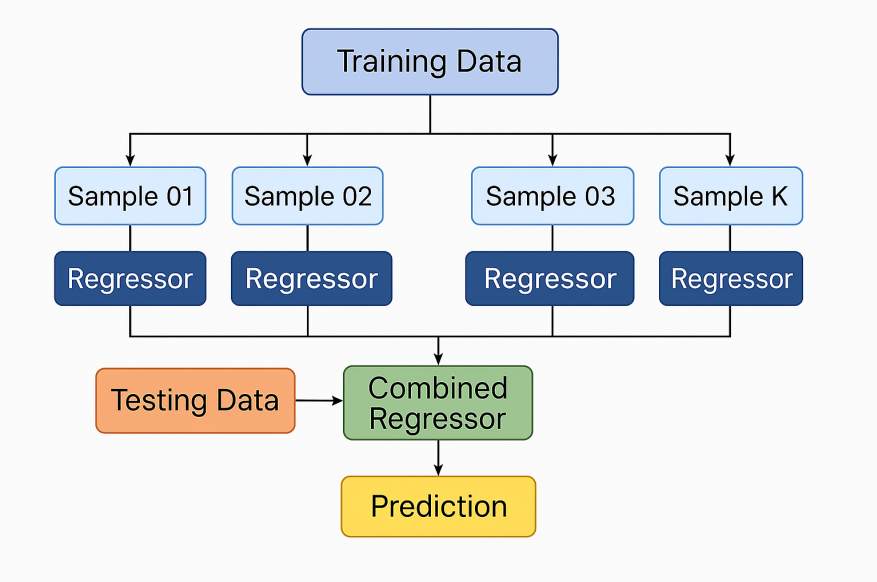


Figure 3.8: Schematic representation of the bagging regressor technique

3.3.3 Bagging Ensemble Approaches

Research scholars have given their insights into the study of the application of random forests and support vector machines in regression prediction problems.

(1) Random Forest

Random Forests (RFs) stand as widely embraced ensemble learning techniques, comprised of multiple decision trees (Breiman, 2001). The core methodology encapsulates a three-step process involving random sampling, random feature selection, and majority voting:

Random Sampling: This step involves selecting n data points from a training dataset of size N with replacement, forming a new training set. The objective is to mitigate over-fitting risks by creating multiple random subsets, thereby enhancing the model's ability to generalise.

Random Feature Selection: For each node in every decision tree, RFs select features. Instead of utilising all features for training, k subsets of features are randomly chosen. Each decision tree then picks an optimal feature attribute as the dividing node. This ensures that each node considers a randomised subset, effectively reducing algorithm variance and bolstering stability.

Majority Voting: The final RFs prediction results from aggregating outcomes from individual decision trees. In regression scenarios, the regression findings from each tree undergo averaging to produce the forest's output. The majority of voting consolidates results from diverse trees, enhancing model precision and comprehensiveness.

The random forest algorithm leverages ensemble learning to amalgamate predictions from numerous decision trees, often outperforming individual trees due to introduced randomness that reduces model variance. Key benefits of random forests include robustness against dataset outliers and minimal need for parameter tuning, with the number of trees being the primary parameter requiring consideration.

The introduction of randomisation mitigates correlation among different decision trees within RF, introducing variability during tree creation. Each tree contributes to the overall output, and when combined, the forest yields a more accurate and resilient outcome. RF has demonstrated high effectiveness in addressing classification and regression problems marked by high dimensionality, non-linearity and illpawedness (Rodriguez-Galiano et al., 2015), commonly encountered in phenomena with numerous dimensions and significant input features. At the same time, a notable advantage of RFs lies in their ability to calculate feature relevance scores, facilitating the assessment of each feature's importance in prediction outcomes (Stulp and Sigaud, 2015).

(2) Bagging-based ensemble algorithms

Ensemble learning may be implemented to enhance the quality of regression results. Ensemble learning is the process of achieving superior predictive performance by combining multiple models that would be possible with any of the individual models alone. Sutton (Sutton, 2005) believes that bagging can be used with tree - based methods to improve the accuracy of the resulting predictions; however, it should be noted that it

can also be used with nontreebased methods such as neural networks.

Random forests are similar to bagging in that they involve constructing several trees using bootstrap samples. However, they distinguish themselves by nourishing each tree with a randomised subset of predictors, hence the name "random." A "forest" of trees is formed by cultivating between 500 and 2,000 trees. Combining bagging and random forest (RF) modelling methods is highly reliable for predictive mapping, especially when utilized together to leverage their individual strengths (Prasad et al., 2006).

Having chosen the bagging algorithm as the primary regression approach, it is worth noting that the bagging algorithm can be integrated with other regression algorithms to enhance its precision and stability, while simultaneously reducing overfitting by lowering result variance. To improve the model's predictive capability, this study initially considered both Random Forest and Support Vector Regression (SVR) as base estimators for the bagging framework. However, only the better-performing method—Random Forest—was retained based on empirical evaluation.

3.3.4 10-Fold Cross-Validation

To reduce the chance due to a single division of training and validation sets, the existing dataset is fully utilised to perform multiple divisions, thereby avoiding the selection of chance hyperparameters and models that do not have the ability to generalise due to ad hoc divisions. This study first presents the results of 10-fold cross-validation by choosing to reduce the chance and improve the generalisation ability of the model through cross-validation.

The models underwent validation using k-fold cross-validation and various statistical approaches. The k-fold methodology is commonly employed to assess the efficacy of a strategy (Raschka and Mirjalili, 2015) in which the associated dataset is randomly partitioned and categorised into ten distinct classes. Figure 3.9 summarizes the concept behind k-fold cross-validation with $k = 10$. The dataset is partitioned into k subsets that are mutually exclusive and of similar size. Each subset is designed to preserve the consistency of the data distribution. This is achieved by employing hierarchical sampling on the dataset. In each iteration, k-1 subsets are combined to form a training

Figure 3.9: 10-fold cross-validation for model evaluation (Raschka and Mirjalili, 2015)

set, while the remaining subset serves as the test set. This process is repeated k times, resulting in k sets of training and testing sets. The k-cross-folding validation method is then applied, where the final result is obtained by averaging the outcomes of the k-training and testing iterations.

In the course of performing a 10-fold cross-validation analysis, it is essential to recognise that while the average root mean squared error (RMSE) provides valuable insights into the overall prediction error across different training and test sets, it may not fully capture the potential presence of overfitting within the model. To thoroughly assess the model's performance, additional analyses were conducted on both the training and test sets, with particular attention to metrics such as RMSE and $\mathbf{R^2}$. Model accuracy improves when the RMSE exhibits lower values, indicating smaller prediction errors, while a higher $\mathbf{R^2}$ value reflects a better fit of the model to the data. The statistical evaluation of the predictive performance was carried out using Equations 3.1 and 3.2.

$$\mathrm{RMSE}(X, h) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(h(x_i) - y_i)^2} \tag{3.1}$$

(worst value = 0, best value = $+\infty$)

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_t hisstudy - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_t hisstudy - \bar{y})^2} \tag{3.2}$$

(worst value = $-\infty$, best value = $+1$) (Chicco et al., 2021)

Where:

•m represents the number of samples, indicating the total number of instances in the dataset.

•$\bar{x}$ represents the mean of the sample.

•x_i stands for the feature vector of the i-th sample, containing multiple feature values.

•y_i is the actual observed value of the i-th sample, which corresponds to the true label or target value.

•$h(x_i\ )$ denotes the prediction made by the model h for the i-th sample. In the context of regression tasks, the model h is a regression model that predicts the output value based on the input features x_i.

•$\hat{y}_i$ is the predicted value of the dependent variable for the i-th sample.

•$\bar{y}_i$ represents the average of the target values (or labels) of the i-th sample.

3.3.5 Experiment Design

(1) The Experiment Design of Predicting Users' Engagement

Both feature extraction and feature selection offer advantages in enhancing learning performance, improving computational efficiency, reducing memory storage, and constructing more effective generalisation models. Consequently, both methods are considered valuable techniques for dimensional reduction.

In scenarios where raw input data lacks features understandable to a specific learning algorithm, feature extraction is typically preferred. However, as feature extraction generates a new set of features, subsequent analysis may become challenging as it compromises the retention of the physical meanings associated with these features. In contrast, feature selection, which preserves some original features, retains the physical meanings of the initial features. This approach provides models with enhanced readability and interpretability. Therefore, feature selection is often favoured in appli-

cations like text mining and genetic analysis. It's important to note that even in cases where feature dimensionality is not exceptionally high, feature extraction/selection remains crucial for improving learning performance, preventing over-fitting, and reducing computational costs.

The video data utilized in this study comprises four components: the number of replies to the video (reply_count), the timing of the video's posting (published_date) and commenting (reply_time), the textual content contained inside the video (title, tags, comments), and the video's category. For this investigation, this study utilised the YouTube API to acquire the Y-video dataset and the T-video dataset, which were collected via hyperlinks in tweets directed to YouTube. These datasets included attributes such as title, hashtag, description, and comments. The video categories are determined by extracting information from their labelled categories. Subsequently, the categories undergo prepossessing.

The curse of dimensionality is a significant concern when using data mining and machine-learning algorithms on high-dimensional data. Sparse data in highdimensional space is a phenomenon that negatively impacts algorithms intended for lowdimensional space (Hastie et al., 2009). Therefore, in this experiment, as shown in Table 3.4, this study adopts different strategies to test the above ensemble model and compare the model performance.

In Table 3.4, this study elaborates on M3 as M3_a, M3_b, and M3_c, each utilising distinct techniques to convert natural language into numerical representations that may be interpreted by machine algorithms. They are, respectively: M3_a uses raw term frequency as input features, constructing a frequency-based representation of the words found in titles, tags, and comments. M3_b employs Word2Vec embeddings, where each word is mapped to a dense vector based on its contextual similarity, and the resulting word vectors are used to represent the text. M3_c relies on TF-IDF (Term Frequency-Inverse Document Frequency) values to represent words, emphasising words that are frequent in a document but rare across the entire corpus. Notably, there is no simple "M3".

No.	Features Selection	Description
M1	M1: num of replies	Only reply_count as inputs
M2	M2: time features	Only time features as inputs, i.e., published_date and reply_time
M3_a	M3_a: words based on term frequency	words from titles, tags, and comments. Word frequency as input
M3_b	M3_b: words based on Word2vec	words from titles, tags, and comments. Word Word2vec matrix as input
M3_c	M3_c: words based on TF-IDF	words from titles, tags, and comments. Word TF-IDF matrix as input
M4	M4: categories	Only categories as inputs, i.e., Music, Education, Entertainment, Gaming etc...
M1+M3_a	number of replies, words based on term Frequency	—
M1+M3_b	number of replies and words based on Word2vec	—
M1+M3_c	number of replies and words based on TF-IDF	—
M1+M2+M3_a	number of replies, time and words based on term frequency	—
M1+M2+M3_b	number of replies, time and words based on Word2vec	—
M1+M2+M3_c	number of replies, time and words based on TF-IDF	—
M1+M2+M4+M3_a	number of replies, time, categories and words based on term frequency	—
M1+M2+M4+M3_b	number of replies, time, categories and words based on Word2vec	—
M1+M2+M4+M3_c	number of replies, time, categories and words based on TF-IDF	—

Table 3.4: Different feature selection strategies as inputs in models

(2) Term frequency

Term Frequency (TF) is the frequency of occurrence of a word in a document, usually divided by the number of occurrences of the word in the document by the total number of words in the document.TF can be expressed by the following mathematical formula:

$$\text{TF(t, d)} = \frac{\text{Number of occurrences of term t in document d}}{\text{Total number of words in document d}} \tag{3.3}$$

Where:

- *TF(t,d)* represents the term frequency of term *t* in document *d*.
- *The number of occurrences of term t in document d* is obtained by counting each occurrence of the term in the document.
- *The total number of words in document d* is the count of all words in the document.

This algorithm 3.3 usually produces a value between 0 and 1, representing the relative frequency. Text mining and information retrieval practitioners frequently apply the idea of *term frequency (TF)* to evaluate a term's importance within a document.

(3) TF-IDF (Term Frequency-Inverse Document Frequency)

The use of word frequency as the sole metric for determining the significance of a word in a document is inappropriate, given that meaningless words may appear frequently in the document and represent the keywords only a small proportion of the time. The idea of *TF-IDF* is to combine *word frequency* and *inverse document frequency* to emphasise lexical items that occur frequently in the current document but less frequently in the whole document set.

It generally consists of two parts, and here this study presents the *IDF* part. In a set of documents, the importance of a lexical item is determined by its *inverse document frequency* or *IDF*. The basic concept of *IDF* is that a word gets a higher weight if it appears less frequently in the collection of documents and can, therefore provide more unique information. In contrast, a word that appears in the majority of the texts is probably a common word with minimal information, giving it a lower weight. The *IDF* formula as below:

$$\text{IDF}(t, D) = \log\left(\frac{\text{Total number of documents in corpus } D}{\text{Number of documents containing term } t + 1} + 1\right) \tag{3.4}$$

Where:

• number of documents in corpus D indicates the total number of documents in the document collection.

• Number of documents containing term t represents the number of documents that contain the term t.

In particular, the number of documents in the document collection divided by the number of documents containing the specific lexical word generates the logarithmic form of the *IDF*. For the size of the whole document collection, the logarithm of this ratio shows how rare the phrase is in relation to the number of documents that contain it. Therefore, a lexical item's importance relative to the corpus as a whole increases with its *IDF*. In general, the *IDF* emphasises words that are less common in the collection of documents, emphasizing their significance within the text.

Hence, the equation for *TF-IDF* can be expressed as:

$$\text{TF-IDF}(t, d, D) = \text{TF}(t, d) \times \text{IDF}(t, D) \tag{3.5}$$

The number of documents in which the term t occurs, the larger this number is, the smaller the *IDF* is, indicating that the term is less important in the whole document set.

(4) Word2Vec

Word2Vec is a familiar algorithm for calculating word vectors, which attracted a lot of attention from industry and academia as soon as it was open-sourced by Google in 2013 (Mikolov et al., 2013). Generally speaking, it is a shallow neural network model that can map each word to a point in a vector space by learning a large amount of textual data and can preserve the semantic and syntactic relationships between words.

As Figure 3.10 shows, Word2Vec is divided into two models: CBOW (Continuous Bag-of-Words) and Skip-gram. The CBOW model predicts the target word from the

contextual word, while the Skip-gram model predicts the contextual word from the target word. Both models are neural network-based language models, where a neural network is trained to learn a vector representation of each word. In this study, the Continuous Bag of Words (CBOW) model is applied by default in the Word2Vec implementation using the Gensim library.

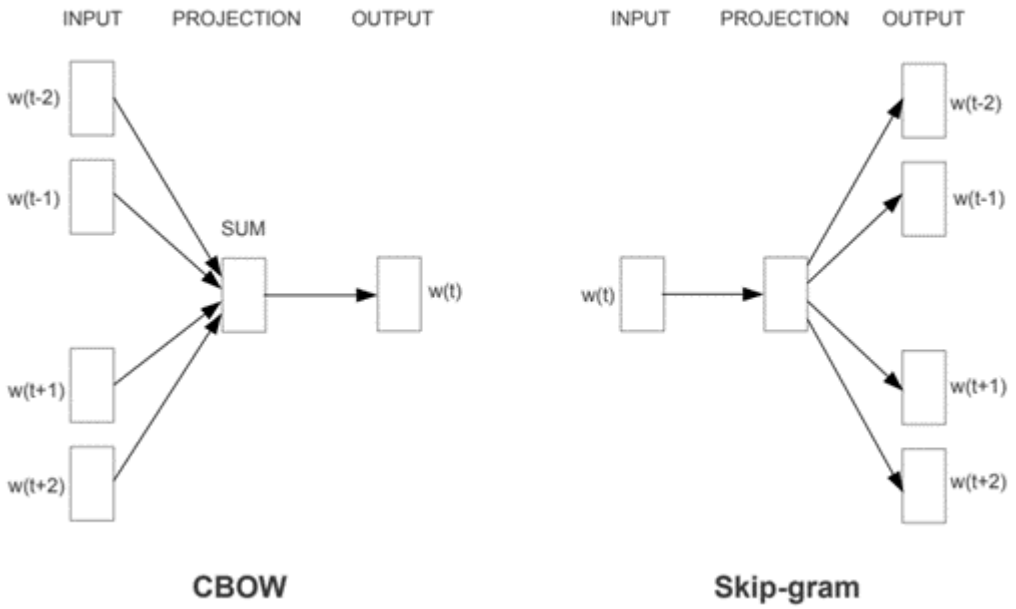


Figure 3.10: The CBOW architecture predicts the current word based on the context, and the skip-gram predicts

Specifically, Word2Vec maps each word to a point in a high-dimensional vector space, and each dimension represents a certain semantic feature of the word. For example, a dimension may represent the "gender" of a word, and a larger value of that dimension for a word means that the word is more "male" oriented, and vice versa for a word that is more "female" oriented. The opposite is true for a word that is more "feminine".

The experimental process will unfold in two distinct phases. Initially, the three algorithms of M3 will undergo a series of validations, concurrently with updates to M1+M3_a, M1+M3_b, and M1+M3_c. Subsequently, the performance of these variants will be evaluated to identify the most effective natural language processing algorithm.

Upon entering the second phase, the good natural language processing algorithms from the first phase will be directly combined with each other. The next steps include presenting the final algorithms from M1 to M1+M2+M3+M4 so that the most effective prediction algorithm can be selected.

3.4 Sentiment Analysis Methods

Based on the data collection processing, preprocessing and clearing discussed above, this study will mainly describe the methods section of Chapter 7, as shown in Figure 3.11.

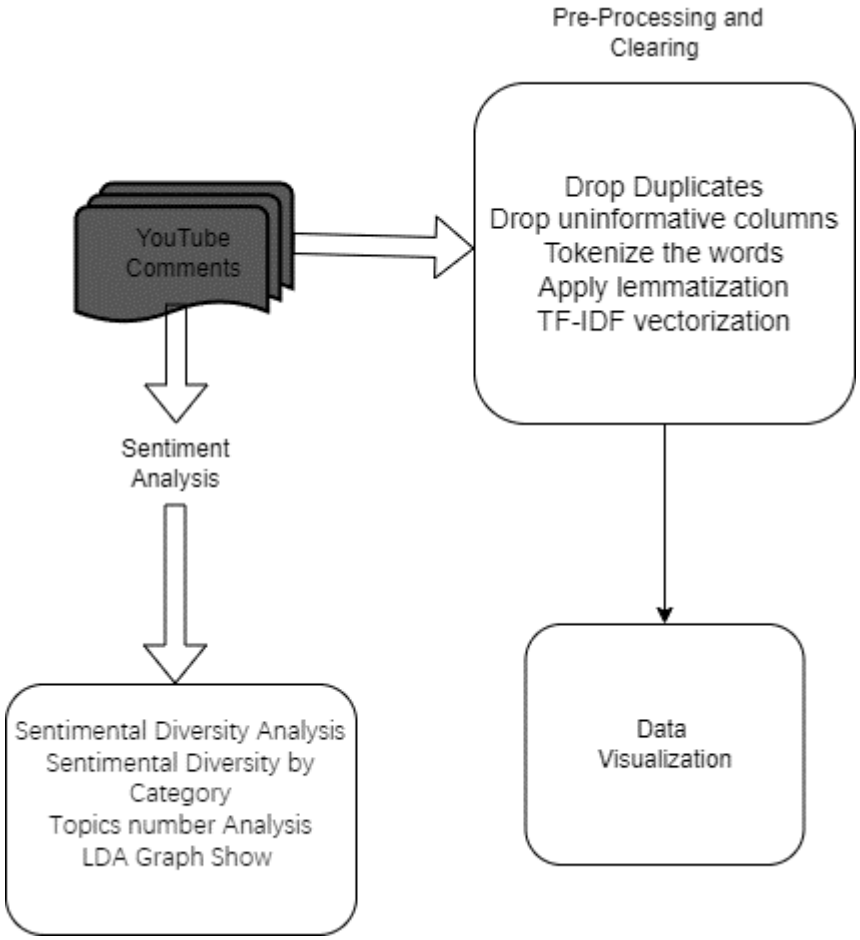


Figure 3.11: Sentiment analysis workflow chart

3.4.1 VADER: A Sentiment Analysis for Social Media

(1) The polarity classification of comments

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool especially suited for sentiment detection in social media and text data (Hutto and Gilbert, 2014). VADER analyses text data to predict the sentiment tendency of a piece of text, be it positive, negative or neutral. VADER builds upon the strengths of traditional sentiment lexicons like LIWC, offering enhancements in several areas. It maintains the advantages of being large in size while remaining easy to inspect, understand, and apply quickly without the need for extensive learning or training. Additionally, it can be easily extended. Much like LIWC (Pennebaker et al., 2001), and unlike some other lexicons or machine learning models, VADER's sentiment lexicon is of gold-standard quality and has been validated by human evaluations. VADER sets itself apart by being more sensitive to the nuances of sentiment expressions in social media contexts, and it also generalises more effectively across various domains. And it is freely available for anyone to download and use.

VADER contains a predefined sentiment lexicon in which each word is associated with a sentiment score (usually between -4 and +4). The words in these lexicons include common emotion words (e.g., 'good', 'bad') and common Internet slang (e.g., 'lol'). VADER uses a set of rules to adjust the sentiment score, taking into account syntactic and semantic factors in the text that may affect the strength of the sentiment. These rules include:

- **Negative words**, such as "not", can diminish or reverse emotion.
- **Intensity words**, such as "very", can amplify the intensity of the emotion.
- **Capitalised words**, such as "AMAZING", can intensify the emotion.
- **Punctuation**: e.g. "!!!!" or "..." that can affect the intensity and direction of the emotion.

In the sentiment analysis, this study expands on the traditional categories of negative, neutral, and positive texts. Drawing inspiration from Elbagir and Yang (Elbagir and Yang, 2019), this study further classifies texts using VADER into five distinct sentiment categories: extreme negative, negative, neutral, positive, and extreme positive

based on their scores. In this study, comments are assigned a positive, negative, or neutral label based on the compound score calculated using VADER (Valence Aware Dictionary and sEntiment Reasoner). This compound score represents the overall sentiment of a comment, normalised to a range between -1 (most negative) and +1 (most positive), and is used to classify sentiment according to predefined thresholds.

To capture a finer-grained understanding of sentiment, each YouTube comment is assigned a polarity score ranging from **-2** (highly negative) to **+2** (highly positive), based on the compound, positive, and negative scores generated by VADER's sentiment analysis. This classification accounts for both the overall polarity direction and the intensity of emotional expression, mapping each comment into one of five sentiment levels:

- Compound score > 0.001 and positive score $> 0.5 \rightarrow$ **Highly Positive** (score = +2);
- Compound score > 0.001 and positive score $\leq 0.5 \rightarrow$ **Positive** (score = +1);
- Compound score between -0.001 and $0.001 \rightarrow$ **Neutral** (score = 0);
- Compound score < -0.001 and negative score $\leq 0.5 \rightarrow$ **Negative** (score = –1);
- Compound score < -0.001 and negative score $> 0.5 \rightarrow$ **Highly Negative** (score = –2).

This refined scoring approach enables a more nuanced interpretation of user attitudes, capturing both the polarity and the emotional intensity expressed in the comments.

(2) The polarity classification of videos

The classification of videos as 'positive' or 'negative' is determined by the aggregate sentiment polarity of their associated comments. Each comment was ranked into one of five categories. As such, each comment is assigned a polarity value ranging from –2 to +2. This study uses the sum of the values of ALL the comments for a given video. If the total is negative, then that is the overall sentiment of the video; if the total is positive, then that video is positive. If the sum of the comment polarity scores equals

zero, the video is classified as neutral. In practice, this study adopts a threshold rule: if the total polarity score of all comments falls within the range of [–N/10, +N/10], where N is the total number of comments on the video, the video is also considered neutral.

Polarity	num of replies_A	num of replies_B	num of replies_C
Highly positive	5	3	2
Positive	4	3	2
Neutral	1	4	3
Negative	6	4	6
4 Highly Negative	0	2	3
Polarity Score	**+8**	**+1**	**-6**
Polarity	**Positive**	**Neutral**	**Negative**

Table 3.5: Illustration of the video polarity classification method

Table 3.5 demonstrates the process of classifying video polarity based on aggregated comment sentiment scores. In this example, three videos, each containing 16 comments, are evaluated by summing the individual comment polarity values (ranging from –2 to +2). To determine whether a video is classified as positive, neutral, or negative, this study applies a threshold-based rule: if the total polarity score lies within ±N/10 (where N is the number of comments), the video is considered neutral. Given that N = 16, the threshold is ±1.6.

- **Video A**: $(5 \times 2) + (4 \times 1) + (1 \times 0) + (6 \times -1) + (0 \times -2) = +8 \rightarrow$ classified as **Positive**.
- **Video B**: $(3 \times 2) + (3 \times 1) + (4 \times 0) + (4 \times -1) + (2 \times -2) = +1 \rightarrow$ classified as **Neutral**.
- **Video C**: $(2 \times 2) + (2 \times 1) + (3 \times 0) + (6 \times -1) + (3 \times -2) = -6 \rightarrow$ classified as **Negative**.

3.4.2 Standard Deviation and Entropy

In sentiment analysis, using Standard Deviation (SD) for comparison is an effective method as it reveals the consistency or diversity of the sentiment of the replies and it

helps us to understand the degree of dispersion in the data. The greater the Standard Deviation, the greater the variability of sentiment scores within that category. This could mean that there is a great deal of variation in how people feel and react to a particular topic or content. If a category has a small standard deviation, it means that sentiment scores are more consistent under that category, and most people feel similarly about the topic. The concept of entropy first originated in physics as a measure of the degree of disorder in a thermodynamic system (Wehrl, 1978). As for information theory, the concept of entropy was first introduced by Shannon (Lin, 1991) to find a way to encode information efficiently/lossless. In this study, entropy is a measure of uncertainty within information theory. The higher the entropy, the more evenly distributed and diverse the sentiment is.

(1) Standard Deviation (SD)

In information theory, Standard Deviation (SD) is used in statistical analyses that deal with information and noise. When it comes to statistical aspects of information, the standard deviation can be used to measure the degree of dispersion in the distribution of an information source's output or signal. When discussing Gaussian noisy channels, the standard deviation of the Gaussian distribution (i.e., the square root of the variance) is commonly used to describe the level of the channel noise. The general formula for standard deviation is:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i hisstudy - \mu)^2} \tag{3.6}$$

where:

- σ denotes the standard deviation.
- x_i is each data point in the data set.
- μ is the mean of the data set.
- N is the number of data points.

In information theory, suppose this study is discussing Gaussian noise channels, where the input signal is X, the output signal is Y, and the noise Z is assumed to obey a Gaussian distribution with zero mean and σ^2 variance:

$$Y = X + Z \tag{3.7}$$

In this scenario, the standard deviation indicates the strength of the noise and is closely related to the channel capacity. For example, for a Gaussian channel with a bandwidth of W and a signal-to-noise ratio of S/N, the channel capacity C can be expressed as:

$$C = W \log_2 \left(1 + \frac{S}{N}\right) \text{ bits/second} \tag{3.8}$$

where S is the power of the signal and N is the noise power. The noise power N is usually related to the noise standard deviation σ, specifically $N \propto \sigma^2$. When this study says N, this study means that there is a positive proportionality between the noise power N and the variance of the noise σ^2, i.e.:

$$N = k \cdot \sigma^2 \tag{3.9}$$

where k is a constant of proportionality. This means that if the variance of the noise σ^2 increases, the noise power N also increases. The larger the variance, the greater the effect of noise in the channel, resulting in a larger noise power.

(2) Entropy

Entropy is a central concept in information theory and is used to measure the uncertainty or amount of information. The concept of entropy was first proposed by Claude Shannon and is known as Shannon Entropy (Lin, 1991). Entropy can be thought of as a measure of the uncertainty of a random variable. The difference between this and the previous (1) is that Gaussian noise and standard deviation focus on the noise in the data transmission, whereas entropy deals with the uncertainty or information content in a random variable. When this study is confronted with a random variable whose outcome is highly uncertain (e.g., there are many possibilities for the outcome and each of them is about the same), then this study has higher entropy for that variable.

For a discrete random variable X(e.g., the output of an information source), the possible values it can take are x_1, x_2, ..., x_n each value occurs with probability $p(x_1)$,

$p(x_2), \ldots, p(x_n)$ then the Shannon entropy $H(X)$ is defined as:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \tag{3.10}$$

Where:

- $H(X)$: denotes the random variable X 's entropy, in units usually of bits.

- $p(x_i)$: denotes the probability that X takes on the value x_i.

- $log_2 p(x_i)$: This is the logarithm of the probability, with a base of 2, since the amount of information is usually measured in bits. If the natural logarithm is used log_e, the unit of entropy is the nat.

The entropy $H(X)$ measures the uncertainty about the possible values of X before this study observes it. If X always takes a fixed value, then the entropy is 0 because there is no uncertainty, and the amount of information is zero. If X takes on a uniformly distributed value with equal probability for all values, then entropy is at its maximum. For instance, in the case of flipping a fair coin, where the probability of each side landing face up is 1/2, the entropy is calculated as:

$$H(X) = -\left(\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right) = 1 \text{ bit} \tag{3.11}$$

This indicates that 1 bit of information is required to describe the outcome of the coin flip.

In information theory, informativeness refers to the minimum amount of information needed to eliminate uncertainty. Entropy can be interpreted as the expected value of the amount of information (i.e., the average amount of information). When the probability of an event occurring is low (i.e., it is rare or unexpected), the occurrence of that event conveys more information. Thus, rare events are highly informative, and the information content of a single event $I(x_i)$ is given by $I(x_i) = -\log_2 p(x_i)$. In contrast, when an event is common or certain, it provides less information because it does not significantly reduce uncertainty. For example, if a coin consistently lands heads, the outcome provides little to no new information, as the result is already known (Equation 3.10).

3.4.3 Topic Extraction

A comment usually consists of several different words that collectively convey the user's feelings or opinions. These words can be analysed by measuring specific similarities of certain distances. Latent Semantic Indexing (LSI) is a simple and practical topic model, which is based on the singular value decomposition (SVD) method to obtain the hidden concepts of the unstructured text (Nan et al., 2010)(Li and Wu, 2010). At first, LSI is introduced to extract topics, but it is a purely mathematical method, and it therefore does not consider the influence of the probability of occurrence between words (Deerwester et al., 1990). LDA (Latent Dirichlet Allocation) provides the topic of each document of the set in the form of a probability distribution, so that after analysing documents to extract their topic distribution, topic clustering or text classification can be performed according to the topic distribution [1]. However, previous research has shown that LDA creates poor analytic results on short texts (Blei et al., 2003), and most comments on YouTube can be said to consist of short texts.

Topic modelling stands as a pivotal technique within the realm of text mining, offering significant capabilities for data mining, uncovering latent patterns, and elucidating the relationships within textual datasets. This approach has fostered extensive research, yielding a multitude of publications that demonstrate its applicability across diverse domains, including software engineering, political science, medicine, and linguistics.

Among the various methodologies employed in topic modelling, Latent Dirichlet Allocation (LDA) is notably prevalent. LDA (Blei et al., 2003) has been the foundation for numerous models due to its effectiveness in identifying and categorising topics within large text corpora. The continuous development of LDA-based models underscores their significance and evolving nature within the field of topic modelling, reflecting ongoing enhancements and adaptations to meet specific research needs in these varied disciplines.

There are 2 main existing methods for predicting the optimal number of topics in LDA: coherence and perplexity. Perplexity is a statistical method for testing the

[1] Introduction to latent semantic analysis, http://lsa3.colorado.edu/papers/LSATutorial.pdf, last accessed 2024/07/09.

efficiency of a model in handling new data that has never been seen before. In LDA, it is used to find the optimal number of topics. In general, this study believes that the lower the complexity value, the higher the accuracy. For a test set consisting of M documents, the ease of confusion (P) is defined as equation 3.12, where $p(w\_d)$ is the probability of observing a word in document d. The probability of observing a word in document d is the total number of words in document d.

$$P = \exp\left\{-\frac{\sum_{d=1}^{M} \log p(w_d)}{\sum_{d=1}^{M} N_d}\right\} \tag{3.12}$$

Coherence, as defined by Röder et al. (Röder et al., 2015), is a measure used to evaluate the semantic relatedness of topics generated by an LDA model. In this study, coherence is computed using the metrics shown in equations 3.13 and 3.14. For example, in a corpus of medical text data, if an LDA model induces a topic that does not align well with the main themes of the data, this topic might be categorized as an outlier. Higher coherence values indicate a greater likelihood of the model achieving higher accuracy in representing related topics.

$$C = \sum_{i<j} \text{score}_{\text{UMass}_{(w_i, w_j)}} \tag{3.13}$$

$$\text{score}_{\text{UMass}_{(w_i, w_j)}} = \log\left(\frac{D(w_i, w_j) + 1}{D(w_i)}\right) \tag{3.14}$$

where $D(w_i)$ denotes the number of documents that contain the word w_i, $D(w_i, w_j)$ represents the number of documents containing both w_i and w_j, and D is the total number of documents in the corpus. These metrics are used in this study to determine the appropriate number of LDA topics for the T-video and Y-video datasets, respectively.

3.5 Summary

This chapter introduced the datasets used in this study and described the data collection, cleaning, and integration procedures in detail. It presented the construction of a cross-platform dataset combining YouTube and Twitter data, along with the ratio-

nale for building a new dataset to capture multimodal features relevant to music video engagement. Key variables—including textual content, metadata, time-based indicators, and platform-specific features—were defined and extracted to enable subsequent modelling.

In addition, the chapter outlined the machine learning methods selected for predicting user engagement, including Random Forest Regression, Gradient Boosting, and BDTR, as well as baseline models such as Lasso and Ridge Regression for feature comparison. The justification for method selection and the construction of feature matrices were also discussed.

This methodological foundation sets the stage for the next chapters, where experimental designs, feature analysis, and model performance evaluation will be presented and discussed.

Chapter 4

Exploratory Data Analysis Based on T-video, Y-video and T&Y-video Datasets

In this chapter, this study will introduce the T&Y-video dataset in detail, which is a full-domain dataset based on the merging of T-video (YouTube videos referenced in Twitter, henceforth referred to as "T-video") and Y-video (YouTube videos taken from the YouTube dataset, henceforth referred to as "Y-video"). This study focuses on the data details of these three datasets, the data crawling strategy and data preprocessing strategy, as well as the various data features that will be used afterwards. The T&Y-video dataset will be used in Chapter 5, while the T-video and Y-video datasets will be used in Chapter 6. The text data (i.e., the replies from both the T-video and Y-video datasets) will be used in Chapter 7 for sentiment analysis.

4.1 Introduction

After data preprocessing in the previous chapter, a total of 2,119 YouTube videos with unique video IDs were identified in the Y-video dataset, accompanied by 20,754 corresponding comments. These were then integrated with the refined T-video dataset, which contains 1,538 unique video IDs and 76,171 comments. The merged dataset thus

comprises two distinct sources, T-video and Y-video, each with its own set of videos and user-generated comments. Both datasets include relevant metadata such as comment timestamps and video-related information, forming a comprehensive foundation for subsequent analysis.

Figure 4.1 illustrates the data composition, which consists of three distinct data sources and three distinct user groups. Specifically, there is a dataset consisting of tweets collected using specific keywords, another dataset including YouTube music videos that are cited in those tweets (referred to as T-video), and a third dataset consisting of search results from YouTube itself based on the same keywords (referred to as Y-video). Evidently, the user community may be classified into three distinct categories. Specifically, those who utilize Twitter, individuals who utilize YouTube, and a collective group of individuals who utilize both Twitter and YouTube. In this context, the term "users" pertains to individuals in the Twitter dataset who actively engage in tweeting and retweeting pertinent content. These users not only consume YouTube music videos but also actively discuss and comment on them. YouTube users, conversely, include the subset of users who exclusively engage in viewing and providing commentary on YouTube channels. The third category encompasses elements from both of the aforementioned categories.

4.2 Description of the T-video and Y-video Datasets within T&Y-video Dataset

Next, in this section, this study describes the information of T-video and Y-video, and the different features within the two datasets.

4.2.1 Basic Descriptive Information of Y-video and T-video

As mentioned before, the T-video dataset has fewer videos and more comments (1,538 videos and 76,171 comments), while the Y-video dataset has more videos but fewer comments (2,119 videos and 40,754 comments). The T&Y-video dataset is the composite of these two datasets (T-video and Y-video).

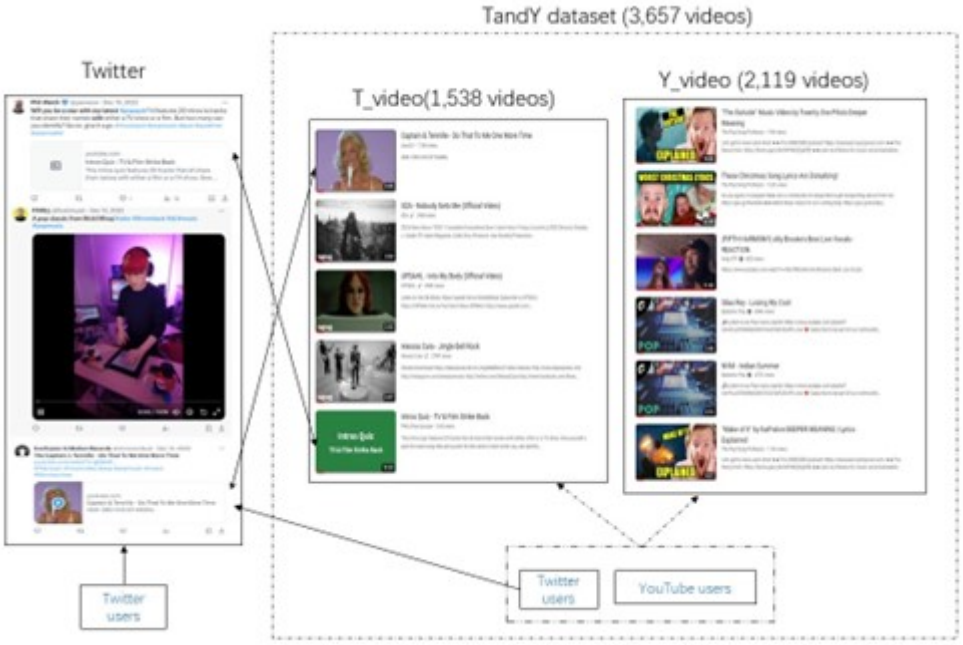


Figure 4.1: The structure of T&Y-video dataset

(1) The basic descriptive information of Y-video dataset

After the above data cleansing and merge calculation for videos with the same video ID, this study ends up with the final Y-video dataset as shown in Table 4.1 below.

Parameter	Views	Likes	reply_count
mean	887,115	7,824	46
std	9,056,872	54,981	126
min	106	2	1
25%	12,752	4065	2
50%	42,172	1,111	6
75%	155,613	2,923	33
max	244,040,100	1,363,134	2,270

Table 4.1: Descriptive of numerical factors associated with Y-video dataset (N= 2,119

(2) The basic descriptive information of T-video

This study utilized the Twitter Streaming API to procure tweets by monitoring the hashtags "#pop music" and "#popmusic" along with the expressions "pop music" and "popmusic". This approach encompasses textual references to pop music, links related to pop music, and those associated with YouTube's URL shortener (youtu.be).

This process yielded a total of 44,163 tweets gathered over six weeks, spanning from December 3, 2022, to January 11, 2023. After preprocessing and the generation process (details see section 3.2.4), resulting in 1,538 unique video IDs and a total of 76,171 comments on these videos. Table 4.2 provides an overview of the top 10 hashtags frequently observed in the dataset of tweets containing YouTube links.

Hashtag	**Freq.**
#music	2,640
#popmusic	2,203
#pop	2,003
#new	991
#PopMusic	818
#gretagrace	695
#singer	671
#song	642
#youtube	612
#newmusic	599

Table 4.2: Top 10 most popular hashtags in Tweets

Table 4.2 shows that in addition to the hashtags used for data mining, there are other often-used tags. These hashtags appear in the data stream because Twitter users tag them simultaneously with those keyword tags.

Parameter	Views	Likes	reply_count
mean	22,809,090	200,373	207
std	148,756,900	1,046,923	361
min	6	0	1
25%	1,328	30	6
50%	65,942	904	69
75%	1,527,505	24,793	217
max	3,415,569,000	19,424,148	3,241

Table 4.3: Descriptive of numerical factors associated with T-video dataset (N=1,538)

A comparison is made between the Y-video dataset (Table 4.1) and the T-video dataset (Table 4.3); it is clear to see that videos in the T-video dataset tend to have much higher view rates than in the Y-video dataset.It is clear that the variation in the number of views is much larger for the T-video dataset than it is the case in the Y-video dataset, with standard deviations of around 149M for T-video as opposed to just over

9M for Y-video. The maximum number of views is over 3B as compared to only 244M for the T-video and Y-video datasets, respectively. Similar patterns in terms of general distributions can also be seen to be the case for the number of Likes, and to a slightly less dramatic extent for the number of replies.

4.2.2 The Numerical Features of the T-video and Y-video Datasets

The numerical data within the T&Y-video dataset focuses on the number of video views, likes, and replies. This study first presents these numerical data for the T-video and Y-video datasets within the T&Y-video dataset to observe any similarities and differences under the same numerical features of the two sub-datasets. Then this study presents the characteristics of the numeric data within the whole T&Y-video dataset in general.

(1) Views, Likes and Replies in T-video and Y-video

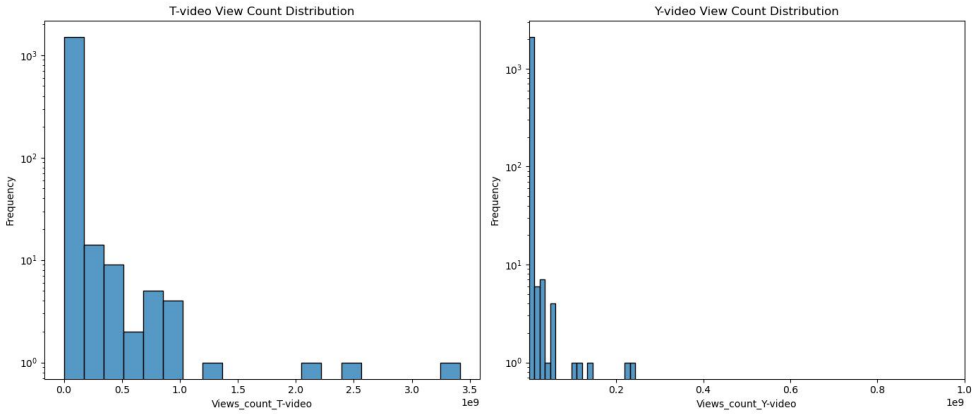
Figure 4.2a illustrates that in T-video, a majority of videos attract a low number of views, while only a few videos achieve high view counts. This suggests a scenario where videos on T-video are primarily viewed by a limited audience, with only a select few gaining significant viewership. This pattern indicates a concentration of viewer interest in a small set of popular videos, signifying diminished interest in the majority.

In contrast, the Y-video chart portrays a more evenly distributed distribution of video views. This suggests a broader spectrum of viewer interests on the Y-video platform, implying that viewers engage with a variety of videos rather than concentrating on a few popular ones.

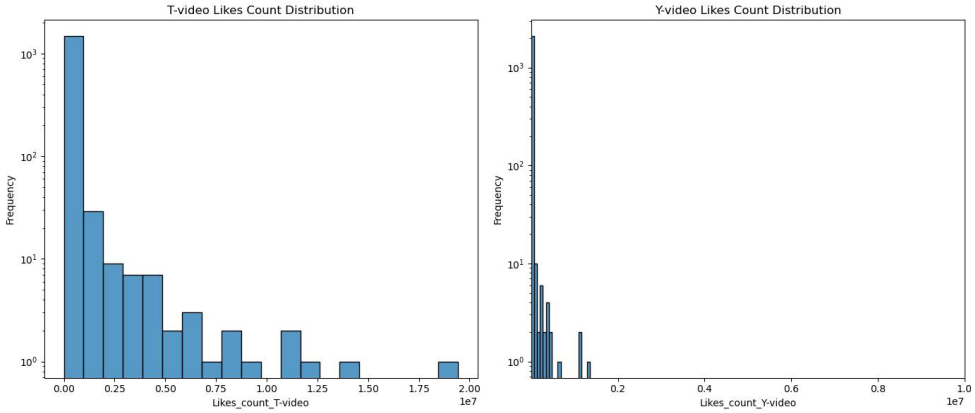
In Figure 4.2b, the T-video data reveals that the majority of videos receive a low number of likes, with only a select few attaining a high like count. This suggests a scenario where videos are primarily appreciated by a limited audience, and only specific videos garner a substantial number of likes. Unlike the left graph, the Y-video data indicates a more evenly distributed pattern in the number of likes for videos. While there are fewer videos with over 2 million likes, these videos exhibit a broader distribution of likes, potentially encompassing some highly popular content.

Figure 4.2c illustrates the distribution of the number of comments in the two

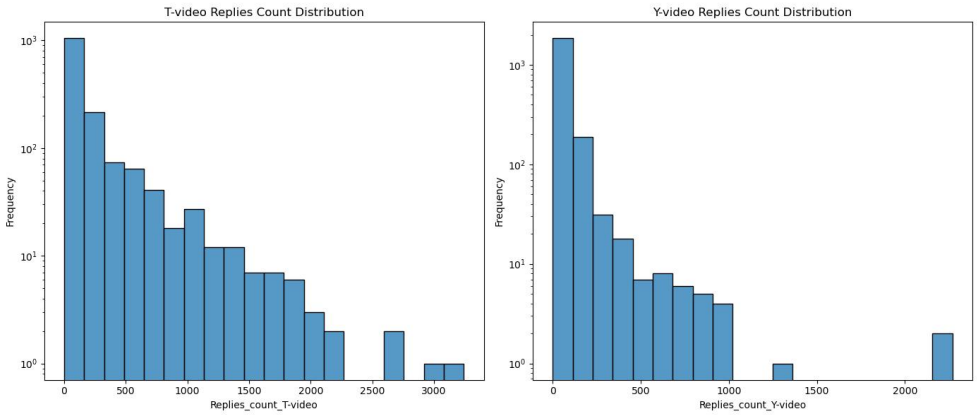
datasets, showing a wide range. The T-video dataset shows a large number of videos with a large number of comments, mainly centred between 0-300. It is worth noting that there are also a significant number of videos with comment counts between 300-500. Interestingly, videos with 500 comments outnumber videos with 300 to 400 comments. In contrast, the Y-video dataset has a much more concentrated distribution of comments between 0 and 200, with a large number of videos having no more than 100 comments. It is worth noting that videos in the middle range of 300 comments do not exist. There are only a few videos with comments between 300 and 500.



(a) Distribution of the number of Views on T-video and Y-video



(b) Distribution of the number of Likes on T-video and Y-video



(c) Distribution of the number of Replies on T-video and Y-video

Figure 4.2: Histogram numerical input factors used from T-video and Y-video datasets

(2) Categories

In the context of YouTube, video categories are essential for organising both channels and content. For instance, categories such as Music and Gaming help group related videos, thereby improving content discoverability and user navigation. Accurate category assignment enhances the visibility of videos and facilitates the search process for target audiences. Given the significance of video categories, this study includes them as an important factor in the analysis.

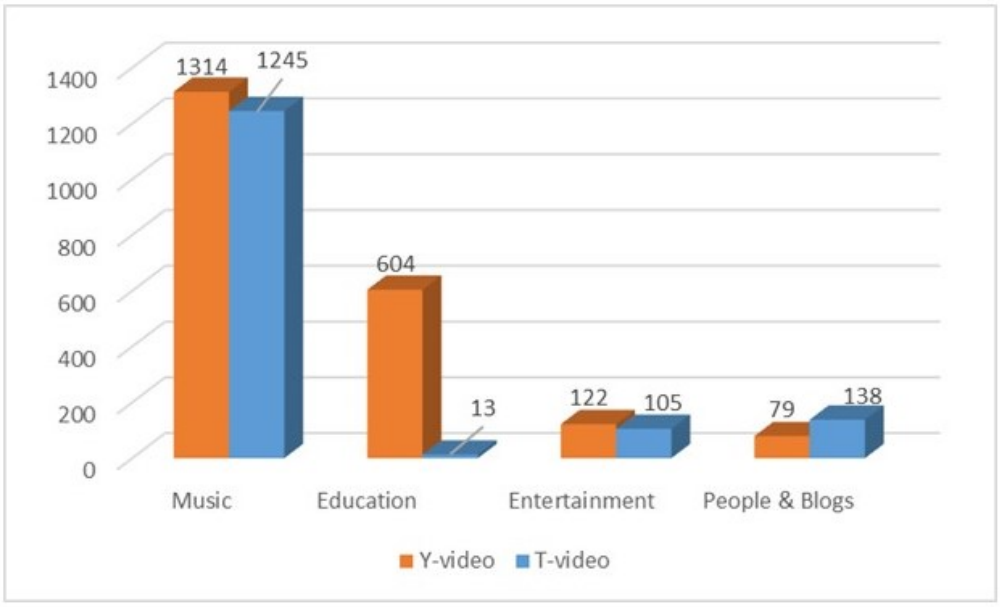


Figure 4.3: The categories distribution of T-video and Y-video.

Chowdhury and Makaroff (Chowdhury and Makaroff, 2013) investigated video attributes across different categories, examining factors such as view count distribution, time-to-peak popularity, post-peak decline rate, and cumulative view count distribution. Their findings revealed distinct popularity patterns for videos in various categories. For instance, videos in news and sports genres experienced rapid initial popularity followed by a swift decline, while those in music and movie genres tended to have prolonged periods of popularity with more consistent viewership rates.

Bärtl (Bärtl, 2018) highlighted the significant disparities among different types of

YouTube videos in terms of channels, upload volume, and view counts. These disparities are evident in the distribution of views, where a small proportion of channels, approximately 3%, accounted for around 85% of the total views. The data suggested that older channels had a higher likelihood of accumulating substantial viewership, while newer channels still had a modest chance of rapid success, especially when making strategic genre selections. Figure 4.3 presents the distribution of video categories in the two datasets. Y-video predominantly features videos from four major categories: Music (1,314), Education (604), Entertainment (122), and People & Blogs (79). The observed diversity of categories in the T-video dataset primarily results from differences in the data collection strategies used for the two datasets, with the top four being Music (1,245), Education (13), Entertainment (105), and People & Blogs (138). This variation in T-video's dataset reflects a broader range of video categories compared to Y-video, indicating greater thematic diversity. Although the number of videos in T-video belonging to less-represented categories is relatively small (a total of 37), they span across eight distinct categories, including Gaming (8), News & Politics (5), Film & Animation (11), Comedy (3), Sports (2), Science & Technology (3), Autos & Vehicles (2), Travel & Events (1), and Howto & Style (2). This suggests that, despite the lower volume, T-video encompasses a wider spectrum of video themes.

Table 4.4 presents the video count for each category in T-video and Y-video across multiple years. While the table appears to show a steady increase in the number of videos, particularly a notable rise after 2017, peaking in 2023-this trend is largely shaped by the differing data collection strategies used for the two datasets. In particular, T-video shows significant growth since 2017 due to its broader collection scope, while Y-video remains relatively limited in volume.

This study also notes that no Y-video data was collected for 2023. According to the Y-video data collection method, data gathering ended on January 11, 2023, and a review of the six channels confirmed that none had uploaded new content during that period.

Across categories, "Music" consistently remains the most represented, while "Education" shows a relatively noticeable increase over time. Other categories, such as

Year	T-video_Categories				Y-video_Categories			
	Music	Edu.	Ent.	People & Blogs	Music	Edu.	Ent.	People & Blogs
2006	3	—	—	—	—	—	—	—
2007	7	—	—	—	—	—	—	—
2008	13	—	2	1	—	—	—	—
2009	60	—	1	—	—	—	—	—
2010	49	2	—	2	—	—	—	—
2011	49	—	1	1	—	—	2	—
2012	40	—	3	—	—	—	10	—
2013	36	1	2	—	—	—	8	—
2014	26	—	2	7	83	—	8	—
2015	33	—	—	7	35	—	1	—
2016	38	—	4	4	118	35	—	4
2017	40	—	3	8	155	83	42	4
2018	51	1	3	5	197	148	47	11
2019	68	1	2	—	378	181	3	32
2020	63	—	6	3	210	86	1	—
2021	85	2	8	10	76	40	—	28
2022	259	2	40	47	62	31	—	—
2023	325	4	28	43	—	—	—	—

Table 4.4: Video Categories Comparison

"Entertainment" and "People & Blogs", demonstrate more stable distributions across years.

(3) The Number of Replies

The number of replies to a video refers to the number of comments posted by users on that video. The replies of the T-video dataset from 2007-12-12 to 2023-4-11 compared to Y-video's (from 2011-11-17 to 2023-01-27). This is one of the most important metrics that provides the following information:

User Engagement: The volume of replies reflects the extent to which viewers are interacting with the video content. More comments usually indicate that viewers are reacting to the video content, increasing user engagement with the content.

Viewer Feedback: Comments allow viewers to express their opinions, thoughts, feelings, and suggestions about the video. The number of responses to a video can be used to measure positive or negative feedback from viewers on the content.

Social Interaction: Comments are one of the main forms of social interaction between users and between users and video creators. A high number of replies may

indicate that the video is more likely to be shared and discussed on social media.

Content Quality: Content in comments can provide a more in-depth assessment of the video's quality and content. Creators can use comments to learn about viewers' needs and preferences so they can improve future content.

Community Building: Through replies and comments, video creators can build a sense of community with their viewers. This is important for long-term viewer loyalty and brand building.

User Interaction Behaviour : The number of video replies can also be used to analyse viewer interaction behaviour. For example, whether viewers have asked questions, interacted with other comments, etc.

Average response time is the average length of time it takes for a viewer to comment or reply to a video after it has been posted. Specifically, it can be calculated in the following way:

$$AverageResponseTime = (TotalResponseTime)/(NumberofResponses) \tag{4.1}$$

$$TotalResponseTime = Thetimeofthelastreply - publishedtimeofavideo \tag{4.2}$$

$$TotalAverageResponse = \frac{1}{n}\sum_{i=1}^{n} \text{AverageResponseTime}_i \tag{4.3}$$

$$\text{Relative reply} = \frac{\text{Total number of video plays per year}}{\text{The sum of the total number of video plays over the years}} \tag{4.4}$$

Where n denotes the number of pings and i denotes the ith video. The average response time can provide the following information:

Engagement and Interaction: Shorter average response times may indicate that the viewership is more interested in the video because they respond faster. This may relate to the video's content or the discussion.

Audience Satisfaction: Long average response times may suggest viewer dissatisfaction or disinterest. This may require a review of the video or an improvement in the way of interacting with the audience.

Effectiveness of Communication: Average response times are also indicative of

the efficacy of communication between creators or community managers and viewers. Rapid responses may increase the efficacy of communication.

Viewer Retention: A low average response time may assist viewers in retaining and maintaining their interest in the video. If a viewer receives a quick response, they may be more likely to return to observing.

Figure 4.4 illustrates the relationship between the reply count and the average response time (in months) for videos in the T-video and Y-video datasets. The x-axis, plotted on a logarithmic scale, represents the reply count, while the y-axis, also on a logarithmic scale, represents the average response time. Each point in the figure corresponds to a video, with the size of the point determined by the video's lifespan (longevity_month), longer lifespans are represented by larger points. Green points denote videos from the T-video dataset, while orange points represent those from the Y-video dataset. Additionally, two horizontal dashed lines indicate the overall average response times for T-video (green dashed line) and Y-video (orange dashed line), respectively.

From the distribution of the points, the green points representing T-video are more concentrated in the bottom-right corner (indicating high reply counts and short response times), suggesting that T-video videos tend to receive a large number of replies quickly. In contrast, the orange points for Y-video are more scattered, with some clustering in the top-left corner (indicating low reply counts and long response times), reflecting lower interaction efficiency for Y-video. The lower position of T-video's green dashed line further confirms its shorter overall response time. In summary, T-video demonstrates better user engagement efficiency compared to Y-video. T-video videos not only receive more replies but also have shorter response times. This is particularly evident for long-lifespan videos (represented by larger points), which perform better in sustaining user engagements. On the other hand, Y-video videos generally receive fewer replies, have longer response times, and exhibit weaker user engagement.

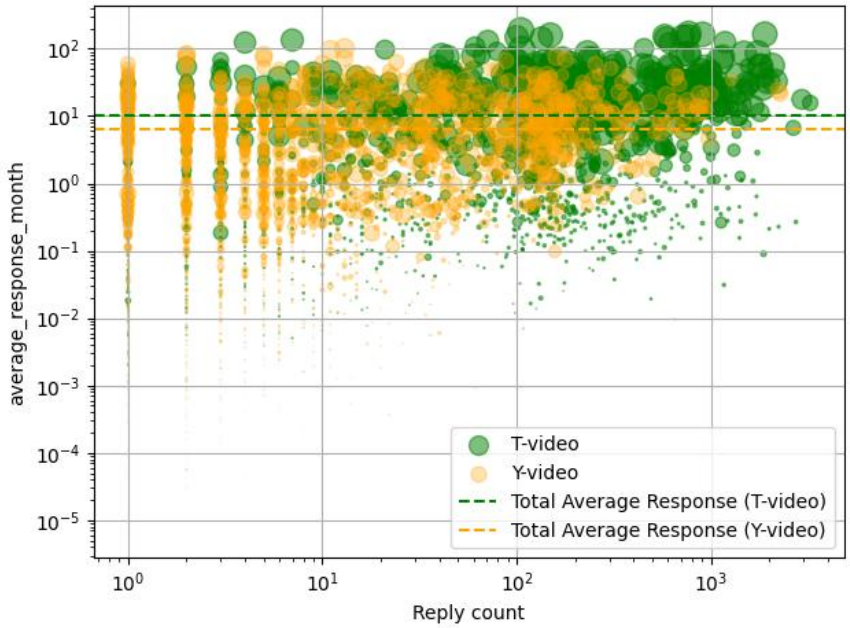


Figure 4.4: Number of replies and average response time from Y-video and T-video datasets

In the visual representation, one of the green dashed lines corresponds to the average response time for the entire T-video dataset, which is recorded as 10.14 months. In contrast, the orange dashed line represents the average response time for the Y-video dataset, which is documented as 6.26 months. This observation suggests that the music videos in the T-video dataset tend to receive user comments over a more extended period compared to those in the Y-video dataset, both in terms of total response time and on a per-video level. As illustrated in Figure 4.4, T-video content appears to sustain user interaction over a longer period, as indicated by its generally higher average monthly response across various reply counts. This pattern may reflect the prolonged visibility and recurring discussions enabled by Twitter sharing dynamics. In contrast, Y-video content tends to attract more immediate, short-lived bursts of attention, consistent with YouTube's recommendation-driven exposure patterns. These results suggest that T-videos may benefit from extended relevance through ongoing engagement, while Y-videos rely more heavily on initial platform-driven discovery.

(4) The Number of Views

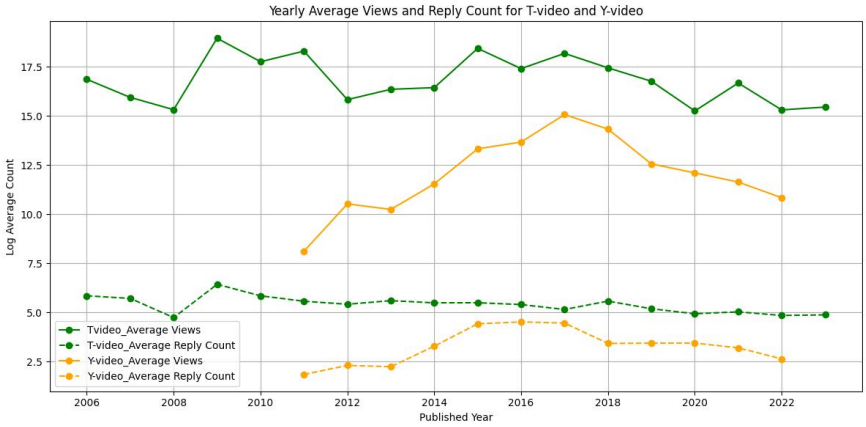


Figure 4.5: Yearly average views and reply count for T-video and Y-video

A video's views represent the count of times users have watched the video, serving as one of the paramount indicators of a video's popularity and impact on the platform. Video views offer valuable insights into the following aspects:

1. Popularity: The number of video views directly mirrors user interest, with higher view counts indicating increased popularity.

2. Content Appeal: Elevated views suggest that the video content resonates with viewers, whether due to its interesting, educational, touching, or unique nature.

3. Influence: The quantity of video views reflects the video's influence and reach. Videos with a broader viewership are more likely to be shared and disseminated on social media platforms.

4. Advertising and Revenue: In the context of ad-supported platforms, video views exhibit a direct correlation with ad revenue. Advertisers tend to favour videos with higher viewership, as it enhances ad exposure and potential returns.

5. User Engagement: Video views serve as a key indicator of user engagement. Videos with a substantial number of views are likely to prompt increased user engagement, including comments, likes, and shares.

6. Platform Performance: For video-sharing platforms, the total number of views becomes a pivotal metric reflecting the overall platform performance. The cumulative views of videos on the platform can significantly influence its reputation and competitive standing.

Figure 4.5 illustrates the average annual views and replies for T-video and Y-video from 2006 to 2023. The data collection process resulted in a more extended duration for YouTube videos promoted on Twitter compared to those exclusively collected on the YouTube channel.

Figure 4.5 shows the yearly average (log-transformed) views and reply counts for T-video and Y-video. T-video exhibits consistently higher average view counts across the entire period, with peaks observed around 2010 and relatively stable reply counts throughout. In contrast, Y-video demonstrates a more pronounced growth trend, reaching a peak in average views around 2017, followed by a gradual decline. Its average reply count is notably lower than that of T-video and declines steadily after 2018.

These trends suggest that T-video content maintains more stable and long-term engagement, possibly driven by ongoing discussions and sharing on external platforms like Twitter. Y-video, however, appears to attract more time-sensitive attention, consistent with YouTube's recommendation-based exposure dynamics. The overall decline in reply counts across both datasets may reflect a broader decrease in active commenting behaviour over time.

As illustrated in Figure 4.6, the portion of views attributed to the T-video dataset (green solid line) remained relatively stable from 2006 to 2023, with minor fluctuations and a moderate peak around 2010. In contrast, the Y-video dataset (orange solid line) experienced a notable rise in view proportions beginning in 2013, peaking sharply around 2017, before declining in subsequent years. This trend may reflect increased content output, promotional activity, or a growing viewer base during that period.

Regarding reply proportions (dotted lines), the T-video dataset exhibits overall consistency, whereas the Y-video dataset shows more volatility. Notably, Y-video's reply share peaked around 2016–2017, suggesting heightened engagement with specific content during that time. The spikes in Y-video may be attributed to a smaller number of highly interactive videos rather than a platform-wide trend. Overall, T-video displays more consistent audience interaction, while Y-video reflects episodic bursts of engagement likely driven by viral or topical content.

Both Fig. 4.5 and Fig. 4.6 depict a discernible "growth and decay" trend in

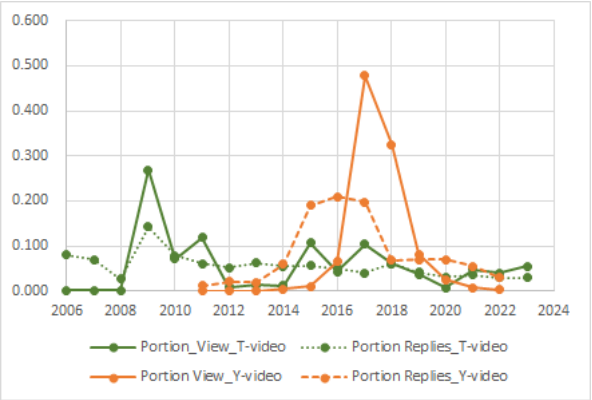


Figure 4.6: The portion of all Views and Replies for each year in video and Y-video

the Y-video data, evident in both the views and responses. In contrast, the T-video data remains relatively consistent during the observed period. Based on the combined analysis of the total number of video categories represented in T-video and Y-video, along with the relative levels of video plays and replies, this study identifies several possible reasons for the observed disparity:

Content Types and Trends: Videos in the Y-video dataset are more likely to originate from larger channels with a substantial following on the YouTube platform. Consequently, they may contain specific content or trends experiencing explosive growth. When videos of a particular type or topic become popular, they may witness a rapid increase in plays and engagements. Conversely, the videos in the T-video dataset are YouTube videos directed from Twitter, potentially losing some users in the redirection process. T-video encompasses multiple categories of videos, while Y-video focuses on only a few. This might explain why T-video's data is more stable, covering a broader content range that is more stable, enduring, and less susceptible to short-term trends, whereas Y-video's data may be influenced by its specific genres.

User Engagement: The Y-video dataset comprises fewer categories of videos capable of gathering users with similar preferences. Some videos may be more controversial or generate more discussion, attracting increased viewer engagement and leading to short-term growth and decay. Videos in the T-video dataset may prioritise engaging viewers over the long term rather than seeking short-term results.

Difference in Timeframe: T-video spans videos from 2006 to 2023, while Y-

video only includes videos from 2011 to 2023. This may explain why T-video's total plays surpass Y-video's, as T-video's videos were released earlier and had more time to attract viewers. It also clarifies why Y-video's fluctuations are more pronounced, with shorter video durations causing relative video plays to decline year by year and drop rapidly after reaching their peak.

Difference in Promotion Strategies: Y-video's videos are on YouTube's platform and haven't been promoted on Twitter, whereas T-video's videos are promoted from Twitter to YouTube. This distinction may contribute to T-video's more consistent statistics, possibly resulting from a steadier audience due to Twitter marketing.

4.2.3 The Textual Features of the T-video and Y-video datasets

In order to obtain the content-related features, specifically the textual and numerical aspects, a series of Python scripts were developed utilising the Natural Language Tool Kit (NLTK). The textual elements were extracted to include the type and quantity of words, as well as the video hashtags utilised in the caption sections of the postings. To decrease the dimensionality of the feature vectors, this study initially employed feature extraction by stemming to combine words with similar meanings. Subsequently, this study focused solely on words that appeared at least 30 times in the word frequency distribution. This selection process yielded a total of 2,276 words. In addition, this study also considered "popular", "music", "song" and "video", which are high-frequency words due to data crawling strategies that may have an impact on the subsequent prediction model. Therefore, this study puts these words into the stopwords list during the extraction of text features. Table 4.5 shows the top 10 most frequent words and their average TF-IDF value in the whole document from T-video and Y-video separately.

The table 4.5 provides a comparison of the top 10 terms in the T-video and Y-video datasets, taking into account their frequency and TF-IDF values. The datasets exhibit a notable presence of shared phrases such as "love," "good," and "great," suggesting the presence of common content themes. Nevertheless, T-video displays elevated frequencies for these phrases, indicating a possible focus on positive attitudes. Terms

Words_T-video	Freq.	TF-IDF_T-video	Words_Y-video	Freq.	TF-IDF_Y-video
love	15,854	0.144	love	9,094	0.0855
good	7,257	0.073	good	4,729	0.0457
great	6,004	0.0665	lyric	4,522	0.0767
beauti	4,033	0.049	sound	3,069	0.0491
hear	3,988	0.0395	mean	3,022	0.0564
amaz	3,912	0.0459	youtub	2,570	0.06
voic	3,858	0.0453	well	2,406	0.0247
year	3,694	0.0378	free	2,393	0.0592
never	3,655	0.0366	voic	2,269	0.0262
well	3,480	0.0363	great	2,192	0.0258

Table 4.5: The top 10 most frequently occurring words in T-video and Y-video

in T-video that are unique, such as "beautiful," "hear," and "amazing," have higher TF-IDF values, which suggests that they are more contextually important. In contrast, Y-video has distinctive words such as "lyric," "sound," and "mean" that have higher TF-IDF values. This indicates their significant relevance in the Y-video dataset, potentially emphasising a concentration on musical and meaningful content.

4.2.4 The Temporal Features of the T-video and Y-video datasets

The temporal features were obtained by extracting and quantifying the frequency of video posting time periods (hour periods, days of the week), commenting time periods (hour periods, days of the week), and distinguishing between morning and afternoon for both video posting and commenting times. Figure 4.7 shows the time proportions of videos posted and replies received. Calculating the number of replies to a video and the number of videos at the time of posting a video uses different methods: no doubt, the number of replies counts the time of replies to each entry in all the datasets; whereas calculating the number of videos posted at the time of posting a video counts videos with the same video ID as a single video, i.e., de-duplicates them to get the unique time of posting of the video, and then calculates the number of videos posted in each different period.

Figure 4.7a presents the distribution of video publishing and reply times across four time periods of the day. A clear distinction is observed in the posting behaviour between the two datasets: in T-video, videos are predominantly published in the afternoon

(33.4%), whereas in Y-video, the majority are released in the evening (41.6%).

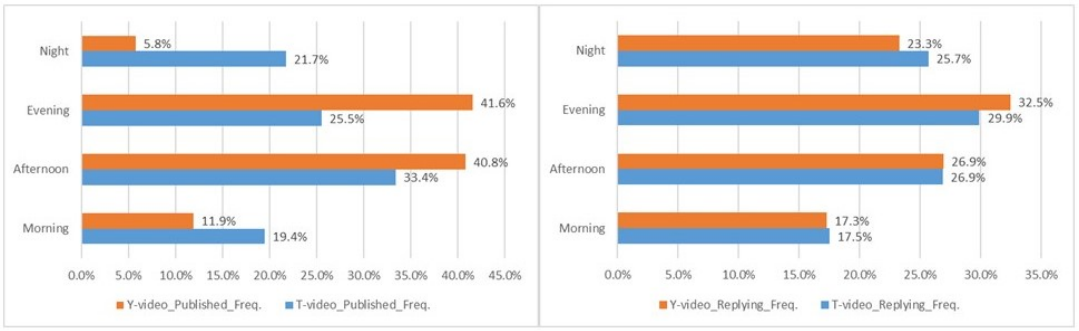
In terms of replying behaviour, both datasets exhibit similar temporal patterns. The evening period receives the highest proportion of replies in both T-video (29.9%) and Y-video (32.5%), followed closely by the afternoon and night. The consistency in replying frequencies suggests that user engagement in terms of comment activity tends to peak during the latter part of the day, regardless of the original posting time.

As illustrated in Figure 4.7b, Friday exhibits the highest video publication frequency in both datasets—24.6% in Y-video and 22.1% in T-video. This notable peak suggests that users are more likely to release new content ahead of the weekend, possibly to maximise visibility and engagement. In terms of replying activity, Friday also shows the highest response rate, with 17.4% in Y-video and 17.0% in T-video, indicating increased user interaction leading into the weekend. These aligned peaks in both posting and replying frequencies suggest that Fridays may represent a strategic period for both content creators and audiences across platforms.

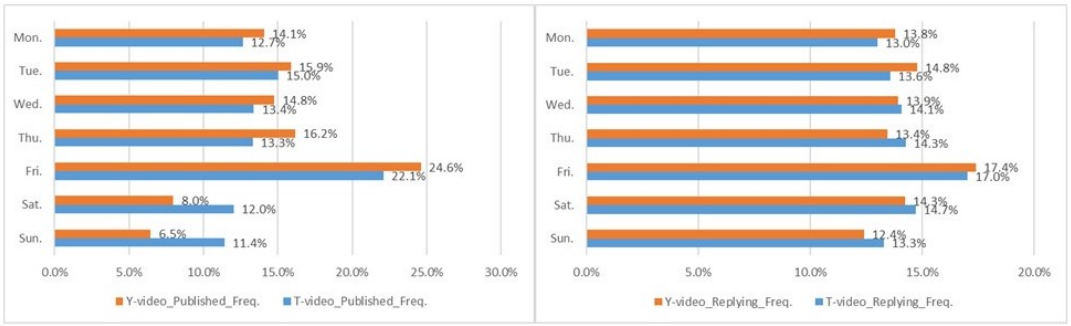
Notes: Morning is from 6:00 a.m. to 12 noon; Afternoon is between 12 noon and 18:00 p.m.; Evening is between 18:00 p.m. to 24:00 a.m.; Night is between 24:00 a.m. to 6:00 a.m.

4.3 Description of the T&Y-video Dataset

In this section, this study concentrates on analysing and presenting the information about the T&Y-video dataset, which is the main data source for the research this study carries out in Chapter 5. The purpose of the analysis is twofold: firstly, to deepen the understanding of the target variable, which is user engagement; and secondly, to gain a deeper understanding of the textual and numerical features within the dataset, to establish a basis for subsequent feature selection and the creation of different feature combinations. This dual approach allows us to refine the predictive model by identifying and exploiting the most relevant variables that influence user engagement.



(a) Proportions of hours for publishing and replying to videos



(b) Proportions of days of the week for publishing and replying to videos

Figure 4.7: Times at which publishing of and replying to take place for the T-video and Y-video datasets

4.3.1 Basic Data Description of Views, Likes and reply_count

As shown in Table 4.6, descriptive information about the T&Y-video dataset has been given. The T&Y-video dataset has 3,657 unique video IDs with 116,925 comments from those videos. For the Views, the data below shows that there are a few videos, but a significant number of plays. The Likes experienced the same situation. The more views a video has, the more likes it has. Videos with a greater number of views are also more likely to elicit more discussion, such as replies.

This study includes 3,657 unique videos in the combined T&Y-video dataset. While both T-video and Y-video show long-tail characteristics—where a small number of videos attract most of the views and likes—their distributional patterns are not identical.

Figure 4.8 illustrates that the distribution of T&Y-video Views and Likes follows a

Parameter	Views	Likes	reply_count
mean	10,106,690	88,803	114
std	97,301,760	686,711	265
min	6	0	1
25%	8,494	230	3
50%	45,227	1,051	12
75%	296,250	4,842	114
max	3,415,569,000	19,424,148	3,241

Table 4.6: The description of Views, Likes and reply_count in T&Y-video dataset (N= 3,657)

pattern resembling a normal distribution, particularly when analysed on a logarithmic scale. This pattern suggests that the majority of video views and likes are centrally concentrated, showing symmetry around the median value. Essentially, this means that most videos receive a moderate number of views and likes, with significantly high or low counts being relatively rare. Employing a logarithmic scale for depicting the distribution of views and likes enables a clearer examination of data that spans a wide range.

Contrasting with the Views and Likes, the distribution of comments per video presents a distinctly different picture, aligning more with a long-tailed distribution. Here, a substantial majority of videos attract very few comments, with counts near zero, whereas a small minority of videos garner a significantly higher number of comments. This pattern indicates that while a few videos achieve above-average engagement in terms of comments, the vast majority fall below this average, with many videos receiving hardly any comments at all. This difference highlights the varied nature of engagement metrics, such as views, likes, and comments, within the video dataset.

4.3.2 The Textual Features and Temporal of the T&Y-video Dataset

The textual features of the T&Y-video dataset

Table 4.7 shows the top 10 most frequent words and their average TF-IDF values in the whole document from the T&Y-video dataset.

As shown in Table 4.7, the high frequency of words such as "love", "good", "hear", "well", and "thank" in the text data, paired with their relatively low TF-IDF values,

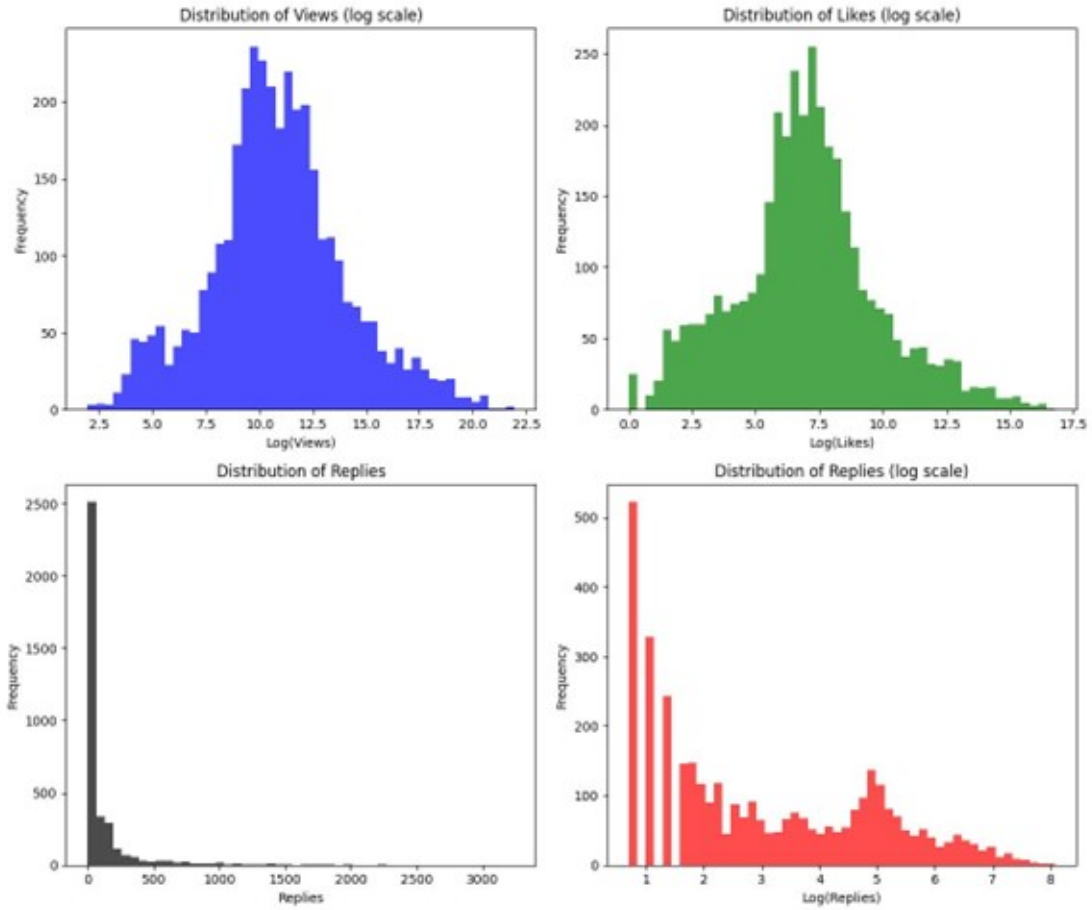


Figure 4.8: Distribution of Views (log scale), Likes (log scale) and Replies (actual and log scale) in the T&Y-video dataset

indicates that they are commonly used throughout the dataset. In contrast, terms like "great", "lyric", and "sound" exhibit moderate frequency but relatively higher TF-IDF values, suggesting they are more specific and contextually distinctive. When comparing the top 10 most frequent words in the T-video and Y-video datasets, there is considerable overlap, highlighting a shared vocabulary of commonly used terms across both datasets.

(2) The Temporal Features

Figure 4.9a reveals a clear pattern regarding the timing of user activities. Replying frequency peaks in the evening, gradually decreasing through the late night into the early hours of the morning, when it reaches its lowest point. In contrast, the frequency

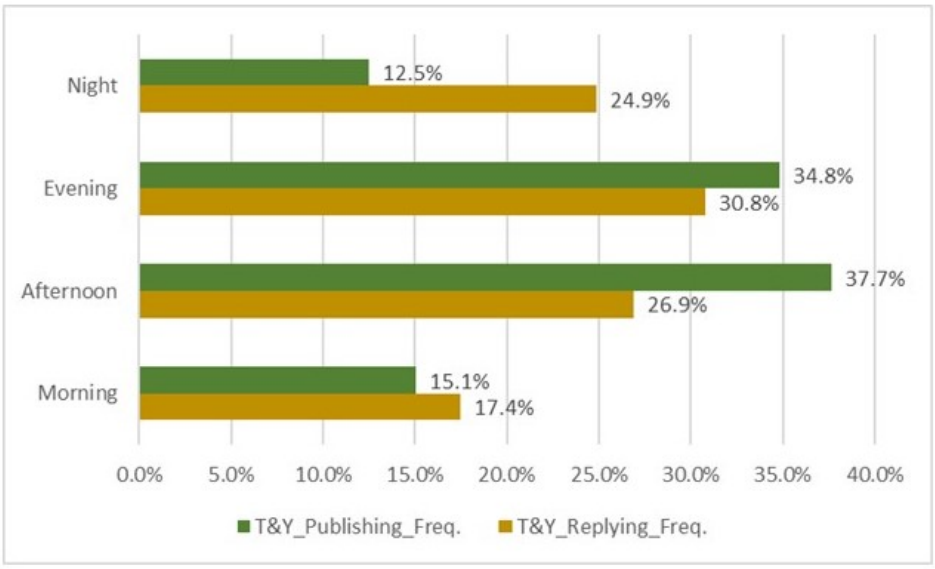
Words_T&Y-video	Freq.	TF-IDF_T&Y-video
love	24,948	0.1037
good	12,015	0.0547
great	8,196	0.0416
lyric	6,393	0.0523
voic	6,162	0.0334
amaz	6,068	0.0318
hear	5,892	0.0287
well	5,886	0.0274
sound	5,528	0.0392
think	5,270	0.0299

Table 4.7: The top 10 most frequently occurring words in T&Y-video dataset

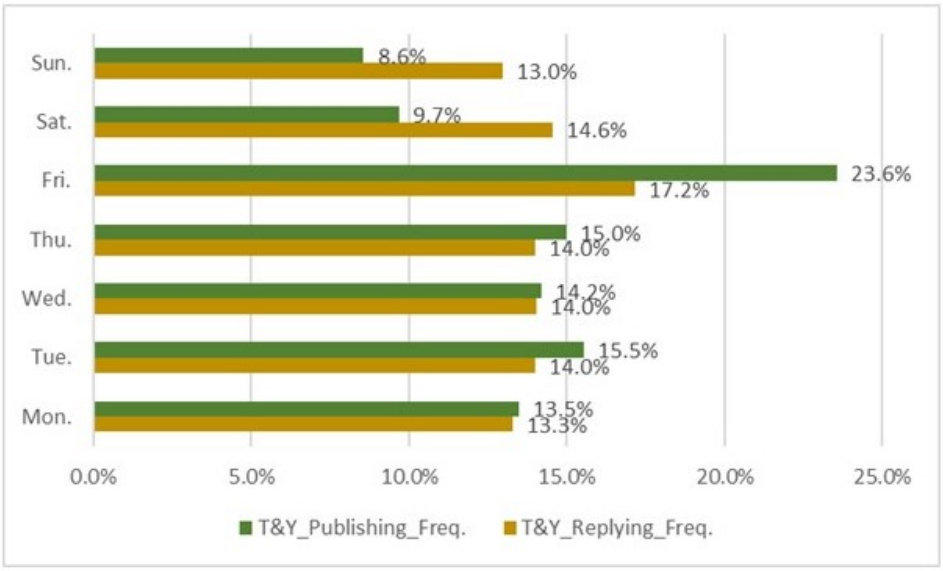
of publishing is highest in the afternoon, before declining during the evening hours. Publishing activity is notably lower in the late night and early morning, with significant variations observed between these periods.

Figure 4.9b shows that reply frequency peaks on Friday (17.2%), followed by a notable drop on Saturday. While reply frequencies across other weekdays remain relatively stable (around 13–15%), Friday stands out as the day of highest user interaction.

Regarding posting frequency, the bar chart shows a generally consistent level above 800 across most weekdays, with a noticeable peak on Fridays and a sharp decline on Saturdays. While the weekly fluctuation is less pronounced than that observed in reply frequency, the alignment of higher posting activity with increased response levels suggests a potential temporal clustering of user engagement, even though a direct correlation is not explicitly established.



(a) Proportions of hours for publishing and replying to videos



(b) Proportions of days of the week for publishing and replying to videos

Figure 4.9: Times at which publishing of and replying to take place for T&Y-video dataset

4.4 Exploratory Data Analysis of Sentiment Analysis

This section presents an overview of the textual data and the polarity data (i.e., Replies content) from the T-video and Y-video datasets utilised in Chapter 7 for sentiment analysis. This study examined the keywords from the titles of the music videos in each dataset to discern thematic differences. Additionally, this study compared the diversity of sentiment tags to highlight variations in how sentiments are expressed under music videos across these platforms. This study also analysed the frequency of positive and negative words in the datasets to understand the focal points of discussions among different user groups.

Figure 4.10 features two-word clouds representing the most frequently occurring words irrespective of polarity in the titles of the music videos from the T-video and Y-video datasets, respectively. In the T-video word cloud, like "lyric", "new", "cover" and "love" indicate a concentration on the content of music videos, while terms such as "night", "girl" and "boy" suggest the thematic content and audience groups for the music videos in T-video. At the same time, the Y-video word cloud includes words like "lyric", "love" and "album", pointing towards an emphasis on lyrical content or emotionally resonant, top-charting songs. The presence of terms such as "meaning", "live," and "one" suggests a focus on the meaning of songs and live performances. As can be seen, although the titles of music videos within T-video and Y-video contain many similar music-related terms ("love", "album", "one"), there are subtle differences in terms of specific themes and content; T-video focuses more on the message of the specific music itself ("lyric", "beat", "new"), whereas Y-video covers a wider range of topics, focusing on the meaning that the music video brings to the viewer, including the level of music production ("production") and the thoughts it brings to the viewer ("meaning", "think", "mean"). These differences help us to better understand the characteristics and focus of the two datasets. These differences may reflect the user characteristics of the respective platforms or the types of videos that are typically popular on each platform.

Figure 4.10: Words clouds showing common words from videos' titles in T-video and Y-video

Figure 4.11 illustrates the proportions of different polarities in the two datasets. It can be seen that the distribution of the proportions of the five polarities (Highly Positive, Positive, Neutral, Negative and Highly Negative) in the two datasets is basically the same, and it is difficult to see major differences. This may indicate that there may be similarities in content types, topics, or user groups between the two datasets. The proportions of each polarity in their respective datasets were not significantly biased.

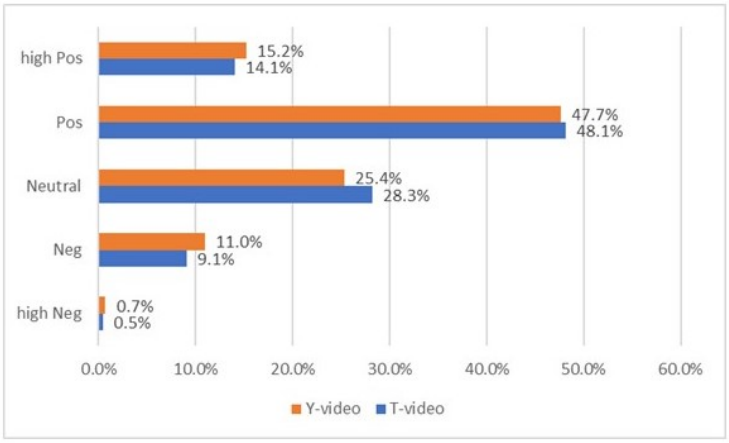


Figure 4.11: The Polarity Proportions in T-video comments and Y-video comments

Although the sentiment polarity distributions are similar, the content specific to each sentiment polarity category may differ. Further analysis of the word clouds in each sentiment category can help us understand the similarities and differences between the two datasets in terms of specific content.

Figure 4.12: Most frequent positive and negative words in T-video and Y-video

When considering positive sentiment (top panel of Figure 4.12), many common words, such as "good", "love", "amazing", "thank", "great", and "well" are found in both word clouds. These words indicate that users in both datasets use some common words in expressing positive emotions, conveying love and appreciation for the video content. The inclusion of the word "thank" in both word clouds suggests that users often express gratitude to video creators or video content when expressing positive emotions. The difference is that in T-video, words such as "voice", "listen", "play", "hear", "look", and "watch" appear, which may indicate that the video content in T-video is more related to the music, lyrics, listening experience, and viewing experience. In Y-video, words such as "people", "life", "thing", "sound", "try", "feel", and "way" appear, which may indicate that the content of the videos in Y-video deals more with life and personal feelings.

Both word clouds contain some identical negative words such as "never", "people", "bad", "mean", "cry", "look", and "hear". These words indicate that users in both datasets used some common words in expressing negative emotions, conveying strong

dissatisfaction and negative feelings towards the video content. The sentiment intensity of the negative words in both word clouds is high, such as "die", "cry", "stop", "hate", and "lost", which convey strong negative emotions from the users. The difference is that in T-video, some words such as "die", "hear", "love", "run" suggest that the negative emotions in this dataset may be more specific to specific emotional experiences and events, such as strong negative emotions, such as loss and escape. In Y-video, some words such as "people", "life", "thing", and "lyric" suggest that the negative emotions in this dataset may be more directed towards relationships, specific issues in life, and lyrical content. Some of the same negative words are found in both datasets, reflecting users' strong dissatisfaction with video content. However, T-video focuses more on negative evaluations of specific events and emotional experiences, whereas Y-video focuses more on negative evaluations of life, relationships, and lyrical content.

4.5 Discussion

4.5.1 Findings

(1) Replies, Views and Likes

The numbers of Replies, Views, and Likes in Y-video indicate significant variations in the dataset, with a small group of videos receiving a large number of views and many others garnering very few, resulting in a long tail. In contrast, T-video's outliers in Views, Replies, and Likes suggest the presence of exceptionally popular videos. Y-video's more even distribution points to a broader range of viewer interests, not confined to a select few popular videos. This observed difference in viewer behaviour between the two platforms provides valuable insights into their viewer dynamics, highlighting the potential factors influencing video viewership patterns.

While Cheng et al. (Cheng et al., 2008) reported that most YouTube videos receive the majority of their views within six days of publication, this study does not track such short-term dynamics. However, Figure 4.6 shows that Y-video engagement is clustered within a narrow range of publication years, indicating a similarly compressed visibility period. In contrast, T-video content demonstrates a more prolonged period of

engagement, with view counts accumulating more gradually over time. This difference suggests that T-video videos may benefit from extended visibility and sharing cycles on Twitter. External referrers, such as search engines and social media, notably Twitter, play a substantial role in directing visitors to YouTube videos (Figueiredo, Almeida, Gonçalves and Benevenuto, 2014). Twitter, in particular, has emerged as a primary source for discovering web videos, leveraging its efficiency in information propagation and follower-follower architecture (Yan et al., 2014).

In contrast, T-video, despite hosting fewer videos, garners a higher quantity of comments. These comments originate not only from the YouTube platform but also from Twitter users. The cross-platform engagement highlights Twitter's effectiveness as a medium for promoting and engaging with audiences, leveraging its significant information propagation efficiency and follower-follower architecture. This analysis provides valuable insights into the dynamics of comment distribution across platforms, shedding light on the role of external factors in influencing viewer engagement.

The T-video data reveals that the majority of videos receive a low number of likes, with only a select few attaining a high like count. This suggests a scenario where videos are primarily appreciated by a limited audience, and only specific videos garner a substantial number of likes. Unlike T-video, the Y-video data indicates a more evenly distributed pattern in the number of likes for videos. While there are fewer videos with over 2 million likes, these videos exhibit a broader distribution of likes, potentially encompassing some highly popular content. This observation underscores the distinct patterns of viewer engagement on T-video and Y-video platforms, where T-video tends to have a concentration of likes on a few videos, while Y-video shows a more uniform distribution across a range of videos. The nuanced differences in liking behaviour shed light on viewer preferences and content appreciation dynamics on the respective platforms.

(2) Categories

T-video demonstrates greater thematic diversity across video categories compared to Y-video. For instance, the T-video dataset includes content from 13 distinct categories, such as Music (1245 videos), Entertainment (105), People & Blogs (138), Film

& Animation (11), and Gaming (8). In contrast, the Y-video dataset is heavily concentrated in just two categories—Music (1314 videos) and Education (604)—with minimal representation in others. The only category where Y-video exceeds T-video significantly is Education, which contains approximately 17 times more videos, primarily due to the selection of educational YouTube channels during data collection.

This difference in category distribution reflects divergent data collection strategies: T-video was compiled using keyword- and hashtag-based searches on Twitter, capturing a wide variety of videos across different channels and topics. In contrast, Y-video was constructed by collecting data directly from six predefined YouTube channels, resulting in more homogeneous content. Consequently, T-video represents a broader and more varied sample of video content.

(3) The Textual Features

By showing the top 10 words within the T-video and Y-video datasets, the study finds that there is a notable overlap in the words discussed by users in the two datasets. However, T-video shows a higher frequency of these phrases, suggesting a possible focus on positive attitudes. Y-video, on the other hand, may place more emphasis on music and meaningful content. By showing the top 10 words within the T-video and Y-video datasets, the study also found that there is no doubt about the existence of common content themes in both datasets. However, T-video shows a higher frequency of these phrases, suggesting a possible focus on positive attitudes. Y-video, on the other hand, may place more emphasis on music and meaningful content.

This study also found that users in both datasets used some common words in expressing negative emotions, conveying strong dissatisfaction and negative feelings towards the video content. However, T-video focuses more on negative evaluations of specific events and emotional experiences, whereas Y-video focuses more on negative evaluations of life, relationships, and lyrical content.

(4) The Temporal Features

The timing of user replies is broadly consistent across both datasets, with the majority of replies concentrated in the evening hours. Beyond this peak period, no substantial variation is observed in reply patterns throughout the day. In contrast, the datasets dif-

fer more noticeably in terms of video upload times: in T-video, uploads predominantly occur in the afternoon, whereas in Y-video, videos are more frequently uploaded in the evening. This contrast suggests that while creators' posting behaviours vary across platforms, user engagement through replies follows a more uniform daily pattern.

From the number of replies and average response time from Y-video and T-video datasets, it can be inferred that the videos hosted on T-video exhibit a greater longevity and continued user engagement over time, as evidenced by the sustained influx of comments. Conversely, Y-video experiences a surge in user engagement shortly after video publication, but subsequently witnesses a decline in comment volume and user willingness to engage with the content as time progresses.

Distinct user behaviour patterns emerge across the T-video and Y-video datasets, particularly in relation to daily engagement cycles. Engagement levels in terms of replies peak during the evening, likely reflecting increased user availability during leisure hours. As the night progresses, both posting and response activities decline, consistent with reduced online activity during typical rest periods. In contrast, posting activity is most frequent during the afternoon, indicating a temporal mismatch between when users tend to upload content and when audiences are most responsive. Recognising these daily behavioural trends is important for optimising content release strategies and enhancing user interaction.

At the same time, although the relationship between posting and response frequencies in the T&Y-video dataset is not directly proportional, there is a discernible overlap where heightened activity in one metric aligns with elevated levels in the other. This observation suggests underlying behavioural patterns in user engagement and interaction dynamics over the course of the week, with both metrics notably intensifying on Fridays.

4.5.2 Limitations

As far as possible, this chapter has examined the various data types in the T&Y-video dataset and transformed them into forms suitable for algorithmic analysis. Due to limitations in both cognitive and computational resources, the study initially prioritised

identifying a broad range of potentially influential features. TF-IDF was employed to quantify the relative importance of words within the textual content. While this method offers a comprehensive overview of language usage, it also generates a high-dimensional feature space.

Future research will focus on dimensionality reduction by selecting a smaller subset of more informative features to enhance model interpretability and predictive performance. This will involve exploring and comparing alternative textual and computational techniques to identify approaches better suited to capturing the nuances of language in social media contexts.

4.6 Summary

In this chapter, the data elements needed for the subsequent chapters 5, 6 and 7 are presented and analysed. These data were initially used to establish an overall understanding of the three databases: T-video, Y-video, and the T&Y-video dataset.

This chapter begins by presenting the overall framework of the entire T&Y-video dataset (N = 3657), which is derived from the merging of both T-video (N = 1538) and Y-video (N = 2119). After that, this chapter presents the results of exploratory analyses on numerical data (e.g., number of likes, views, and replies), temporal data (e.g., time of replies and video posting), and textual data (e.g., video title, tags, and comment content) within the T&Y-video dataset. These analyses also include a comparative examination of the T-video and Y-video subsets, which will be further used in Chapters 5 and 6. Finally, this chapter focuses on the preparation of text-based data, specifically video titles and user comments, for subsequent analysis. These textual elements will be further utilised in Chapter 7 for sentiment classification and related natural language processing tasks.

Chapter 5

Predicting User Engagement based on T&Y-video Dataset

This chapter explores the factors influencing user engagement with YouTube music videos through the application of machine learning models. The analysis integrates both content-based and contextual features, aiming to address two primary research questions:

RQ1: Which machine learning methods demonstrate the strongest performance in predicting user engagement based on textual and numerical content features?

RQ2: What is the impact of time-related features (e.g., publish date, comment date) and content-related features (e.g., comments, titles, tags) on user engagement with music videos?

In this study, user engagement is operationalised as the number of Likes received by each video, which serves as the target variable in the predictive modelling tasks. The objective is to evaluate the predictive value of these features using ensemble models, offering insights for marketers and content creators to enhance engagement strategies. The findings contribute to the broader understanding of user behaviour in social media environments.

5.1 Model Performance Evaluation

Figure 5.1 illustrates the RMSE distributions obtained from 10-fold cross-validation of the Random Forest, Gradient Boosting, and Bagging Regressors. In this study, Bagging is implemented using Scikit-learn's BaggingRegressor with default settings, where the base estimator is a DecisionTreeRegressor. This approach, referred to as Bagged Decision Tree Regression (BDTR) in this study, involves training multiple decision trees on bootstrapped subsets of the data and aggregating their predictions to reduce variance. BDTR is particularly well-suited for high-dimensional datasets and complex feature interactions, which are common in user behaviour prediction tasks. RMSE values are used to evaluate model stability and predictive variance, with narrower boxes in the figure indicating more consistent performance across folds.

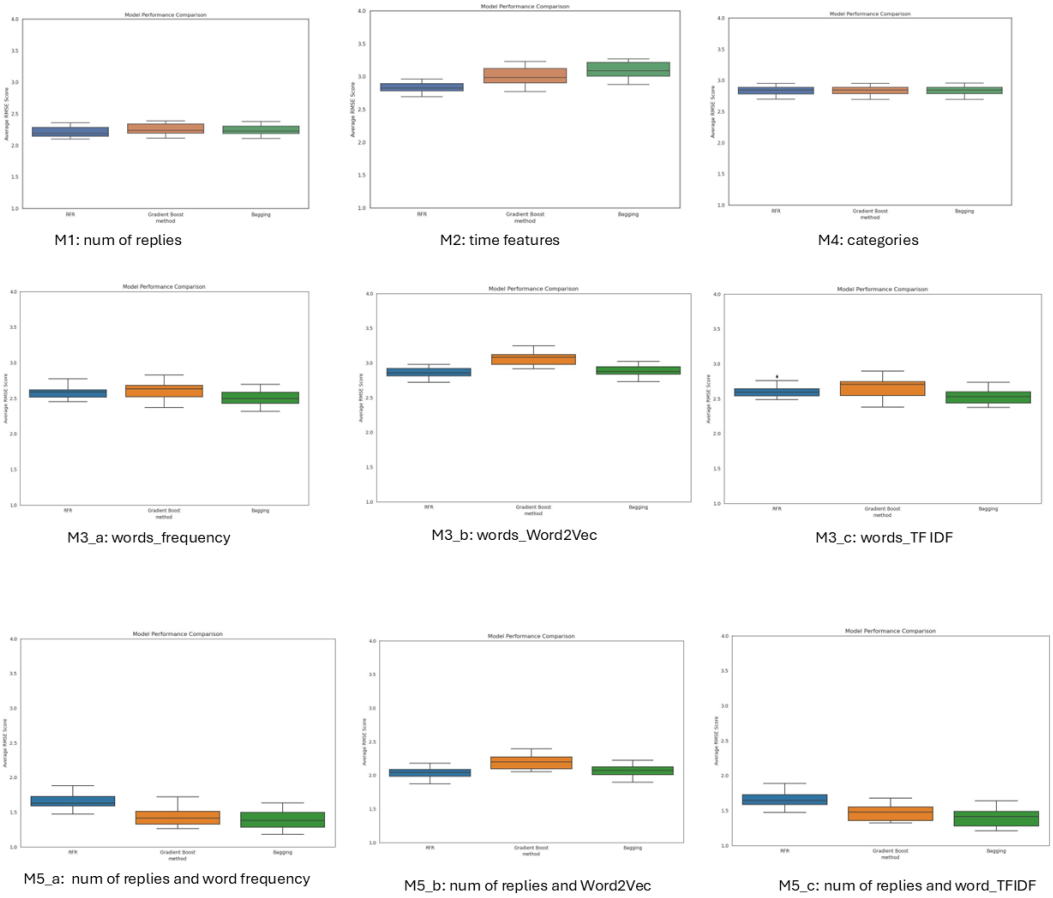


Figure 5.1: Model Performance Comparison with 10-Fold Cross-Validation (Random Forest Regression, Gradient Boost, BDTR)

According to the experimental design outlined in the previous section (see Table 3.2 for details), this study validated three different text feature extraction techniques by applying them to corresponding feature selection strategies.

These techniques are represented by models M3_a, M3_b, and M3_c, each transforming the text data into vector matrices used as input for prediction. Among them, M3_a and M3_c demonstrated similar performance, both outperforming M3_b. When combined with M1 to form M5_a, M5_b, and M5_c, respectively, the overall average RMSE scores were relatively low. The inclusion of an additional feature (i.e., number of replies) further improved prediction performance. Among the three strategies, the TF-IDF approach, used in M3_c, achieved the lowest RMSE, indicating its superior effectiveness in predicting user performance. Supporting evidence is presented in Figure 5.1, with additional studies from Qiu and Yang (Qiu and Yang, 2021), Cahyani and Patasik (Cahyani and Patasik, 2021), and Abubakar et al. (Abubakar et al., 2022) corroborating these findings.

(1) Performance Metrics Validation

In the previous section, we initially evaluated RMSE on the full dataset to preview its performance without data leakage (i.e., ensuring the test data was not used during training). Given the models' generalisation and stability, and no statistically significant differences, we then randomly split the dataset into 80% training and 20% test sets for cross-validation. We compared the RMSE and R^2 metrics on both sets to determine the best-performing model.

Table 5.1 presents the model performance of three different algorithms, Gradient Boost Regression, Bagged Decision Tree Regression (BDTR) and Random Forests Regression- based on the test set and training set across different feature combinations (M1 to M7_c). Evaluation measures such as predictive accuracy and generalisation capacity are employed for assessing model performance. Figure 5.2 illustrates the performance of these algorithms on True vs. Predicted Values based on M1 to M1+M2+M4+M3 (i.e., M7), where the black dashed line of Figure 5.2 represents the y=x diagonal, and the dashed line in the corresponding colour of the algorithm plots the predicted values against the true values in a linear relationship.

Table 5.1: Performance metrics validation results for different models

		Training data		Test data	
Features strategy	Algorithm	R^2	RMSE	R^2	RMSE
M1	Gradient Boost	0.56	1.89	0.54	1.92
	BDTR	0.56	1.90	0.53	1.93
	Random Forest	0.52	1.98	0.56	1.87
M2	Gradient Boost	0.23	2.5	-0.11	2.96
	BDTR	0.27	2.44	-0.17	3.04
	Random Forest	0.05	2.78	-0.01	2.82
M3_a	Gradient Boost	0.56	1.90	0.54	1.92
	BDTR	0.56	1.90	0.53	1.93
	Random Forest	0.52	2	0.56	1.87
M3_b	Gradient Boost	0.97	0.48	-0.2	3.09
	BDTR	0.86	1.07	0	2.82
	Random Forest	0.09	2.73	0	2.81
M3_c	Gradient Boost	0.82	1.2	0.16	2.56
	BDTR	0.89	0.93	0.22	2.48
	Random Forest	0.22	2.52	0.15	2.6
M4	Gradient Boost	0.03	2.82	-0.01	2.83
	BDTR	0.03	2.82	0	2.82
	Random Forest	0.02	2.82	0.01	2.81
M5_a=M1+M3_a	Gradient Boost	0.93	0.75	0.75	1.41
	BDTR	0.97	0.51	0.77	1.36
	Random Forest	0.69	1.58	0.69	1.56
M5_b=M1+M3_b	Gradient Boost	0.99	0.31	0.48	2.04
	BDTR	0.93	0.76	0.54	1.9
	Random Forest	0.55	1.92	0.55	1.89
M5_c=M1+M3_c	Gradient Boost	0.95	0.62	0.75	1.42
	BDTR	0.97	0.51	0.77	1.35
	Random Forest	0.71	1.58	0.7	1.56
M6_c=M1+M2+M3_c	Gradient Boost	0.95	0.62	0.75	1.42
	BDTR	0.97	0.51	0.77	1.35
	Random Forest	0.69	1.58	0.69	1.56
M7_c=M1+M2+M4+M3_c	Gradient Boost	0.96	0.6	0.75	1.41
	BDTR	0.97	0.51	0.78	1.33
	Random Forest	0.7	1.58	0.7	1.53

The following observations can be made from Table 5.1 and Figure 5.2:

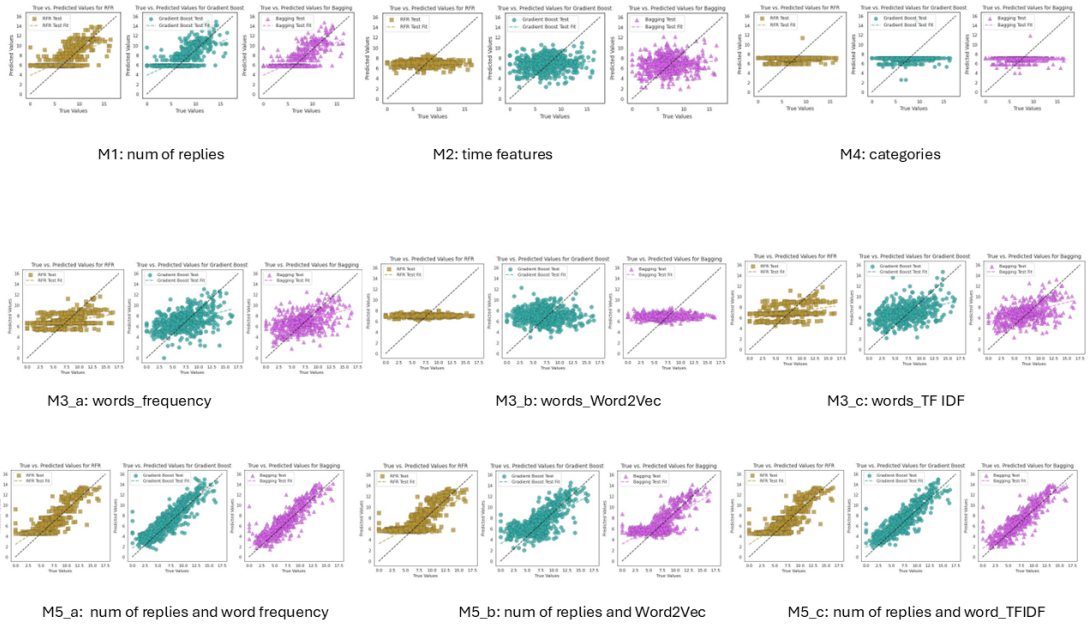


Figure 5.2: True vs. Predicted Values for RFR, GB and BDTR based on M1 to M1+M3_c

Among the feature strategies examined, M2 (temporal features) and M4 (category features) consistently underperform across all models. These strategies yield low or negative R^2 values and relatively high RMSE scores on both training and test sets. Scatter plots further illustrate the lack of predictive alignment, with widely dispersed points and minimal correlation to actual values.

This pattern suggests that time- and category-based metadata alone do not effectively capture the complexity of user engagement. Possible reasons include the general nature of categories (e.g., "Music") and the limited granularity of temporal stamps (e.g., upload or comment times), which may not directly reflect the interactive or emotional dynamics influencing video popularity. As such, M2 and M4 are excluded from further modelling in Chapter 6, where the analysis focuses on more predictive features, notably interaction and content-related variables (M1 and M3).

Gradient Boosting (GB): The evaluation of model performance across different

feature strategies reveals that the inclusion of temporal (M2) and categorical (M4) features contributes little to no predictive value. All three models, Gradient Boosting (GB), Bagging Decision Tree Regression (BDTR), and Random Forest Regression (RFR), consistently yield near-zero or even negative R^2 scores on the test set for these configurations, with GB achieving -0.11 and -0.01 on M2 and M4, respectively. These results indicate that models relying solely on M2 or M4 perform worse than a naïve mean predictor, underscoring the limited utility of time and category features when used in isolation.

When excluding M2 and M4, performance improves significantly across models, particularly for feature combinations involving interaction metrics and textual features (e.g., M1, M1+M3). Under these more informative configurations, GB demonstrates consistently strong predictive power, often yielding R^2 scores close to or above 0.75 on the test set, especially in M5_a and M5_c. However, signs of overfitting are visible in some scatter plots, for example, under M1, where predicted values cluster around y = 5.6, suggesting reduced generalisation.

Bagged Decision Tree Regression (BDTR): BDTR demonstrates solid performance across several informative feature combinations. While it does not yield meaningful results under M2 (time features), its performance improves significantly when applied to feature sets that combine user interaction and textual data. For instance, in the M5_c configuration (reply count + TF-IDF), BR achieves an R^2 of 0.77 on the test set, among its best outcomes, along with a relatively low RMSE. As more relevant features are introduced, BR's predicted values show increasing alignment with the actual values, as reflected in the scatter plots, indicating enhanced fit and reduced variance.

Random Forest Regression (RFR): RFR shows relatively stable performance across most feature configurations, though it generally underperforms compared to GB and BDTR. When using M1 alone, RFR achieves the highest test R^2 of the three models (0.56), suggesting it is slightly more effective at leveraging reply count as a standalone feature. However, this advantage does not hold when incorporating additional textual or contextual features.

For feature combinations such as M5_a (M1 + term frequency), M5_b (M1 + Word2Vec), and M5_c (M1 + TF-IDF), RFR consistently yields lower test R^2 values than both GB and BDTR (e.g., 0.69 in M5_a vs. 0.77 for BDTR, and 0.70 in M5_c vs. 0.77 for BDTR). Similarly, in more comprehensive feature combinations like M6_c and M7_c, RFR continues to lag behind with test R^2 values of 0.69 and 0.70, respectively.

While RFR appears moderately effective when reply count is the dominant feature, it shows limited benefit from the addition of content-based and contextual features. These results suggest that compared to BDTR and GB, RFR has weaker generalisation performance in more complex feature scenarios.

As illustrated in Figure 5.2, all three models, RFR, GB, and BR, display a concentration of predicted values within the range y = [4, 6], particularly under M1 and M5_c. This may be due to:

- **Local Performance Variation:** Models may exhibit reduced sensitivity to input variation within this range, leading to compressed predictions and stacked values along a straight line.
- **Threshold/Saturation Effects:** Certain input characteristics yield similar outputs, suggesting saturation behaviour.
- **Robustness Limitations:** The models may respond uniformly to slightly different inputs, limiting differentiation.

To assess whether the differences in model performance across the compared machine learning algorithms were statistically significant, a Chi-square test was conducted on the R^2 and RMSE values reported in Table 5.1. As shown in Table 5.2, the p-values for both R^2 ($p = 0.486$) and RMSE ($p = 0.529$) exceed the 0.05 significance threshold, indicating that the observed variations are not statistically significant. This suggests that, while performance metrics vary numerically, these differences may be due to sampling variation rather than systematic differences in algorithmic performance.

Table 5.2: Results of significance tests on model performance metrics

Metric	Chi-square (X^2)	p-value	Significant ($p < 0.05$)?
R^2	1.44	0.486	No
RMSE	1.27	0.529	No

Overall, the significance test results (Table 5.2) indicate that ensemble methods, particularly Bagged Decision Tree Regression (BDTR) and Gradient Boosting (GB), outperform Random Forest in leveraging interaction-based features such as reply frequency. These models achieve consistently higher test R^2 scores in feature combinations including M1 (reply count), suggesting stronger capability in modelling user engagement patterns. While textual features—especially term frequency and TF-IDF, provide useful contextual signals, temporal and category-based features contribute relatively little predictive value across models.

The stable performance of BDTR across integrated feature strategies (e.g., M5_c and M7_c) demonstrates its generalisation ability, making it a suitable choice for modelling complex behavioural interactions. In contrast, although Random Forest shows stable results and reasonable generalisation, it appears less effective in capturing higher-order interactions than GB or BDTR. When employed as the base estimator in the BDTR framework, Random Forest benefits from the variance-reducing effect of bagging, leading to improved robustness and predictive accuracy for more complex feature combinations.

(2) Performance BDTR-Random-Forest

Table 5.3 presents performance metrics for models employing Bagging as the ensemble technique, using **Random Forest as the base estimator** (referred to as BDTR-Random-Forest).

On the test set, the BDTR-Random-Forest model demonstrates consistently strong performance under feature combinations that include interaction-based and text-based inputs, particularly M1 (reply count), M3_c (TF-IDF), and their combination in M5_c. For instance, under M5_c, the model achieves an R^2 of 0.76 with an RMSE of 1.35, indicating solid generalisation. In contrast, its predictive ability declines sharply when using temporal (M2) or categorical (M4) features alone, with R^2 scores of -0.10 and

Table 5.3: Model performance results for BDTR-Random-Forest

		Training data		Test data	
Features strategy	Algorithm	R^2	RMSE	R^2	RMSE
M1	BDTR-Random-Forest	0.55	1.92	0.54	1.91
M2	BDTR-Random-Forest	0.25	2.48	-0.10	2.95
M3_c	BDTR-Random-Forest	0.73	1.49	0.21	2.50
M4	BDTR-Random-Forest	0.02	2.82	0.00	2.82
M5_c = M1+M3_c	BDTR-Random-Forest	0.91	0.84	0.76	1.35

0.00, respectively.

Training results show higher R^2 values across all settings, but the performance gap between training and test sets, especially under M2 and M4, suggests potential overfitting and confirms that these features provide limited value in isolation.

Overall, BDTR-Random-Forest is selected as the preferred model due to its consistent test performance, robustness against overfitting, and ability to handle high-dimensional feature spaces. These properties make it well-suited for modelling complex, heterogeneous user engagement behaviours in large-scale social media datasets such as those derived from YouTube.

5.2 Result analysis

In this section, the main goal is to analyse the results of the experiment to answer the research questions.

5.2.1 M1 (reply_count), M2 (time feature) and M4 (categories)

Here this study discusses the case when a single feature is used as input data for user engagement prediction, i.e., M1, M2, and M4.

(1) M1 (reply_count)

Figure 5.2 illustrates that when using only the reply count as the input feature (M1), the predicted values generated by all three models, Random Forest (RFR), Gradient Boosting (GB), and Bagging Regressor (BR), tend to cluster around the true values. However, the points do not fall perfectly along the diagonal line, indicating that none

of the models achieves ideal prediction accuracy. This pattern suggests that while reply count is moderately correlated with user engagement, it is not sufficient on its own to fully explain or predict engagement levels.

As shown in Table 5.1, the training R^2 values for all three models range narrowly from 0.52 to 0.56, suggesting similar explanatory capacity when using reply count alone. On the test set, Random Forest achieves the highest R^2 (0.56), though the differences across the three models are minimal. This indicates that none of the models offers a substantial advantage over the others under the M1 feature strategy, and all exhibit comparable prediction error when relying solely on reply count.

These findings underscore the limitations of relying on a single interaction-based variable. While reply count contributes some predictive value, the modest R^2 scores and similar RMSEs across models highlight that important explanatory information is missing. To address this, the reply count is subsequently combined with textual features in the M5 series of models (i.e., M5_a, M5_b, M5_c), resulting in substantial performance improvements. For example, the Gradient Boosting model in M5_a achieves an R^2 of 0.75 and an RMSE of 1.41 on the test set, compared to an R^2 of 0.54 and RMSE of 1.92 in M1. This improvement demonstrates that integrating interaction metrics with content-based features significantly enhances the model's ability to predict user engagement more accurately.

(2) M2 (time features)

As shown in Table 5.1, the models performed poorly when using only temporal features (M2). The R^2 values on the training set were low (0.23, 0.27, 0.05), and even negative on the test set (-0.11, -0.17, -0.01), indicating that the models failed to capture meaningful patterns and performed worse than simply predicting the mean. RMSE values were relatively high for both training (2.44-2.78) and test sets (2.82-3.04), confirming the models' weak fit and poor generalisation.

These findings are further supported by Figure 5.2, where predicted values deviate considerably from the true values across all three models. The scattered distribution and lack of alignment with the diagonal line reflect the models' limited predictive capacity.

(3) M4 (categories)

In the case of M4, all regression models perform similarly on both the training and test datasets, which generally indicates an absence of significant overfitting or underfitting. However, the R^2 values remain close to zero, implying that the models fail to capture meaningful trends in the data or that categorical features alone do not provide sufficient information to predict the target variable (user engagement).

In summary, the poor performance observed when using time or category features as the sole input for prediction indicates that these features do not provide sufficient information for effective prediction. Therefore, it is necessary to consider incorporating additional relevant features or adopting alternative modelling approaches that are better suited to handling categorical data.

5.2.2 M3 (words)

(1) M3_a (term frequency)

For M3_a, the relatively high R^2 values and low RMSE scores indicate that the model captures some correlation between term frequency and user preferences. The fact that the points on the scatter-plot are mostly clustered around the diagonal but dispersed further confirms that the model has some predictive power using term frequency as a feature, although there is still room for improvement.

(2) M3_b (Word2Vec)

For M3_b, although the gradient boosting model performs well on the training data, it generalises poorly, as indicated by negative R^2 values on the test set. Similarly, the BDTR and random forest models, despite strong performance during training, fail to produce valid predictions on unseen data. This may suggest overfitting to the training set or a mismatch in the feature distributions between training and test sets, preventing effective transfer of learned patterns. Addressing this issue may require revisiting the feature engineering process, adjusting model complexity, or incorporating additional features to enhance generalisation.

(3) M3_c (TF-IDF)

For M3_c, the BDTR performs best on the training set, but this advantage does

not carry over to the test set. Both Gradient Boosting and Random Forest show poor test performance, with Gradient Boosting exhibiting a particularly sharp drop. These results suggest that all three models are affected by overfitting, especially Gradient Boosting. The models may be overly complex or too narrowly fitted to the training data, limiting their ability to generalise to unseen examples.

In conclusion, regardless of the text feature extraction technique, all models suffer from the problem of over-fitting.

5.2.3 Combined Fetures (M1+M3, M1+M2+M3)

Given the weak standalone performance of M2 and M4, their predictive contributions were considered negligible. This section therefore focuses on analyzing M5_a (M1 + M3_a), M5_b (M1 + M3_b), and M5_c (M1 + M3_c).

Compared to M3_b, the performance of M5_b improves across all three models, especially for Gradient Boosting on the test set. This indicates that adding reply_count enhances the generalizability of the word2vec-based model.

On the test set, M5_a (reply count + term frequency) generally outperforms M5_b (reply count + word2vec), suggesting that in this task, term frequency features integrate more effectively with reply count than word2vec. While the improvement of M5_b over M3_b highlights the value of reply_count, word2vec may still require a more complex model structure to be fully effective.

When comparing M5_a, M5_b, and M5_c, the M5_c model (reply count + TF-IDF) demonstrates performance comparable to M5_a and better than M5_b on the test set. Among the three, BDTR in M5_c achieves the lowest test RMSE (1.35). As shown in Figure 5.2, all three regressors (RFR, GB, and BDTR) in the M5_c configuration exhibit tighter alignment between predicted and true values than in M3_b and M3_c, particularly in the BDTR model.

5.2.4 The Importance of Features on User Engagement with Music Videos

This section examines feature importance based on the best-performing model, the BDTR-Random-Forest applied to the T&Y-video dataset. As shown in Table 5.4, the most influential content- and context-related features were identified using the model's internal mechanism, which estimates importance by averaging impurity reduction (e.g., variance reduction) across all base decision trees.

The top 10 features include reply_count and words extracted from video titles, tags, and comments. These textual features were selected using stemming to consolidate semantically similar terms. Among them, reply_count stands out with an importance score of 0.5751, significantly higher than any other feature, highlighting its critical role in predicting user engagement.

Features	Importance_TandY	TF-IDF_TandY	Freq.
reply_count	0.5751	—	—
2010	0.0492	0.0204	833
clifford	0.0321	0.0175	588
pixl	0.0280	0.0024	57
make	0.0259	0.0160	585
professor	0.0130	0.0184	822
stumm	0.0108	0.0175	587
instrument	0.0072	0.0092	741
network	0.0068	0.0027	62
poptast	0.0064	0.0199	770

Table 5.4: Top 10 features based on feature importance scores in the T&Y-Video Dataset

The most influential content-related words can be grouped into three categories:

(1) Title and Tag Related

The word "2010" appears frequently in video titles and tags, typically indicating the year of production. Its TF-IDF score is 0.0204, suggesting that it is important in specific documents but not widespread across the entire dataset.

(2) Channel Owner Related

Several terms originate from the names of prominent YouTube music channels. For

instance, "clifford" and "pixl" refer to the owners of "The Pop Song Professor" and "Pixl Networks" respectively, with TF-IDF scores of 0.0175 and 0.0024. Additional words such as "professor", "stumm", "network" and "poptast" also stem from these or similar channels. Notably, "poptast" is associated with "Epidemic Pop", a popular channel known for high-quality lyric videos, with over one million subscribers. Although the TF-IDF scores for these terms are relatively low (ranging from 0.0024 to 0.0199), their presence reflects the strong branding influence of these music-focused channels within the dataset.

(3) Commonly Used Words

The word "make" ranking as the fifth most important feature, is a general term with broad semantic usage, commonly associated with production, creation, or causality. Its frequent appearance may reflect the recurring language style in video metadata or comments related to content creation.

These observations suggest that user engagement is strongly linked to identifiable creators and well-established music channels. Notably, many of the high-impact terms originate from the Y-video subset, while the T-video portion contributes fewer high-importance words, highlighting an asymmetry in the feature distribution across the combined dataset.

As shown in Table 5.5, nearly all of the top 10 most viewed music videos in the dataset are official releases from the artists' own channels, with the exception of "SHAKIRA ||BZRP Music Sessions "#53". However, this video was uploaded by Bizarrap, a well-known Argentine producer and DJ recognized for his popular "Bzrp Music Sessions" series. These sessions, featuring impromptu performances by guest artists, have attracted millions of views and a large global following.

Table 5.4 further emphasizes the importance of user interaction-particularly the number of replies, in determining a video's popularity. Other high-importance features also point to the influence of prominent YouTube channels. Channels with larger subscriber bases tend to generate more comments shortly after publication, which can help their videos gain priority in YouTube's search and recommendation algorithms. This increased visibility contributes to higher view counts and engagement levels.

video_id	video_title	Views	Likes	Source
2Vv-BfVoq4g	Ed Sheeran - Perfect (Official Music Video)	3,415,569,000	19,424,148	T-video
fLexgOxsZu0	Bruno Mars - The Lazy Song (Official Music Video)	2,527,982,000	13,832,654	T-video
VYOjWnS4cMY	Childish Gambino - This Is America (Official Video)	875,727,100	12,422,504	T-video
YBHQbu5rbdQ	Fifth Harmony - Worth It (Official Video) ft. Kid Ink	2,170,534,000	11,366,890	T-video
CocEMWdc7Ck	SHAKIRA ||BZRP Music Sessions #53	501,340,400	10,834,852	T-video
ZmDBbnmKpqQ	Olivia Rodrigo - drivers license (Official Video)	439,828,300	8,974,255	T-video
h_D3VFfhvs4	Michael Jackson - Smooth Criminal (Official Video)	825,184,500	8,539,399	T-video
HUHC9tYz8ik	Billie Eilish - bury a friend	456,584,400	8,263,927	T-video
izGwDsrQ1eQ	George Michael - Careless Whisper (Official Video)	1,008,453,000	6,895,143	T-video
4fndeDfaWCg	Backstreet Boys - I Want It That Way (Official HD Video)	1,217,145,000	6,696,453	T-video

Table 5.5: Top 10 videos with the highest view counts in T&Y-video dataset

The data supports broader patterns observed in digital content distribution, specifically the "long tail" effect (Anderson, 2006) and the "Pareto principle" (Pareto, 1919). While a small number of highly popular music videos capture the majority of views and attention, a vast number of lesser-known videos collectively account for a significant share of total engagement. This distribution reflects both user preferences and platform dynamics, whereby YouTube's algorithm tends to promote already popular content, further amplifying its reach.

In conclusion, viewer interactions, especially comment activity, combined with the reach of well-established channels, play a critical role in shaping the popularity of music videos. The dominance of a few viral videos coexists with the meaningful cultural and economic contributions of niche content. For individual creators, this highlights the importance of cultivating a subscriber base, encouraging engagement, and developing content strategies that align with recommendation algorithms. Even non-mainstream videos, when targeted effectively, can attract and retain a loyal audience. Building vis-

ibility on the platform requires consistency, content quality, and strategic engagement with viewers over time.

5.3 Discussion and Limitations

5.3.1 Discussion

The purpose of this study is to enhance understanding of the key factors on YouTube that influence the engagement of popular music videos. By applying an ensemble machine learning method, particularly a Bagging Regressor based on the Random Forest algorithm, this research provides insights into model aggregation strategies that can be used to explore the predictive capacity of social media data in explaining user engagement and preferences.

The comparative analysis of traditional and modern ensemble methods (Alsayat, 2022; Ninaus et al., 2019; Sundaramurthy et al., 2020; Wassermann et al., 2019) also incorporates diverse forms of social media data, reflecting the evolving nature of online platforms. In doing so, the study contributes to ongoing efforts to understand user engagement in digitally and socially heterogeneous environments. Notably, it offers a perspective on how user-generated content in music videos contributes to platform dynamics, particularly in the context of personal YouTube channels.

These findings align partially with previous studies (Burgess and Green, 2009; Kietzmann et al., 2011; Liikkanen and Salovaara, 2015), supporting the view that shifts in the social media landscape can affect patterns of engagement with popular music content. They also validate the influence of contextual features, such as time and date, as noted in earlier research (Nathan, 2022; Spasojevic et al., 2015). Although temporal features alone exhibit limited predictive power, their contribution when combined with content-based features (e.g., video titles or descriptions) warrants further exploration, as interaction effects might still exist in more complex feature configurations.

Additionally, the results support the notion that content creators, whether affiliated with official channels or independent, benefit from strategic self-presentation to increase visibility and viewer interaction. The importance of channel names and well-known

YouTubers as features further illustrates the impact of branding on video popularity. Early-stage viewer activity, such as commenting and reposting, also appears to contribute to broader visibility, reinforcing the value of early engagement signals across the dataset. Cross-platform promotion, moreover, plays a role in drawing external audiences to YouTube videos and stimulating comment activity.

From a novelty perspective, this study contributes to the literature (Gatta et al., 2023; Yan et al., 2015) by highlighting emerging social media behaviours. These include multi-platform engagement, content sharing during real-time consumption, linking music to personal performance contexts, and simultaneous interactions with multiple media forms. Such behaviours are particularly pronounced on platforms like Twitter, underscoring the increasingly interconnected nature of digital music experiences.

5.3.2 Limitations

Due to limited knowledge and available resources, those experiments are currently only 3 machine learning algorithms for regression prediction analyses on two datasets, and for future work, other studies hope to explore more possibilities with more regression prediction algorithms.

This study found that a small but non-negligible proportion of the comments were written in non-English languages, based on the results of natural language processing applied to the contextual data. The restricted generalizability of the study's conclusions stems from the dataset's exclusive focus on YouTube music videos from English-speaking nations, making it one of its weaknesses. For data from other social media sites, a similar methodology to this paper might be applied with ease to improve the generalizability of the findings.

Despite those efforts to consider several elements that could impact user engagement, the understanding of this study is constrained by limited knowledge and the limitations of data crawling techniques. As further work, this study plans to improve the quality and coverage of the classification algorithms, especially by improving the way different model features are combined. There exist potential variables that have not been considered, such as data about YouTube channels and demographic character-

istics of YouTubers, among other factors. In forthcoming investigations, other studies intend to incorporate these measures to augment the quantity of noteworthy features within the regression model.

5.4 Summary

The primary aim of this study is to contribute to the existing literature by exploring contemporary approaches for assessing user engagement and preferences related to YouTube music videos. This was achieved through the analysis of data collected directly from YouTube and indirectly via external hyperlinks to YouTube videos shared on the social media platform Twitter.

The study applied three ensemble machine learning methods, Random Forest (RF), Gradient Boosting (GB), and Bagged Decision Tree Regression (BDTR), to predict user engagement. Among these, the Bagged Random Forest (BDTR) model produced the most accurate and consistent results across various feature strategies. Each model was evaluated based on its predictive performance, with the goal of identifying the most effective approach. However, developing optimal machine learning models for this task remains challenging, as model accuracy is influenced by the nature of input variables and the size of the training dataset. Ensemble methods, such as BDTR, typically rely on weak learners (e.g., Decision Trees) to construct sub-models, which are then aggregated to enhance predictive accuracy (Ardabili et al., 2019).

Future research should consider additional factors such as data sparsity, model learning capacity, and alternative evaluation metrics. Incorporating variables like upload timing, response time, and contextual descriptors from video comments may further improve the responsiveness and interpretability of engagement prediction models.

As far as can be determined from the existing literature, this is the first study to use both content- and context-based variables (e.g., views, categories, reply counts, word-based features) across multi-platform sources to explain user engagement with YouTube music videos using an ensemble learning approach. The findings highlight the importance of content creators and channels in influencing user interaction and suggest that distributing music videos across platforms can enhance engagement outcomes. In

particular, variables such as view counts, category type, and timing of video release were shown to be significant predictors of engagement.

Chapter 6

Predicting User Engagement based on Y-videos and T-videos

To better understand the factors influencing user engagement with YouTube music videos across platforms, this chapter extends the analysis to both the T-video and Y-video datasets. Regression models are applied separately to each dataset to identify the best-performing model, followed by a comparison of the key features driving engagement in each context.

This approach enables a more nuanced examination of how specific video characteristics affect audiences across different platforms. By comparing the relative importance of predictors, the study highlights differences in user behaviour between Twitter-linked and YouTube-native content. These insights contribute to a deeper understanding of digital music video engagement and support the strategic development of cross-platform content.

6.1 Research Questions

Social media platforms serve diverse user groups, which may share similarities or exhibit distinct behaviours. Understanding these user segments, whether overlapping or unique, provides valuable insights for YouTubers seeking to promote their content more effectively. This chapter investigates the key factors influencing user engagement

with YouTube music videos using data from both T-video and Y-video datasets.

The analysis is guided by two research questions focused on algorithmic effectiveness and influential features across platforms:

RQ3: Which is the most suitable machine learning method for predicting user engagement based on two datasets, which are different data sources, respectively?

RQ4: What are the factors (textual factors, context factors) that influence user video engagement on each of the two different platforms?

To address RQ3, we apply multiple regression models to both datasets and evaluate their performance to identify the most suitable algorithm for each. For RQ4, we analyse the feature importance generated by the best-performing models, enabling a comparative interpretation of what drives engagement across the two platforms.

6.2 10-Fold Cross-Validation Results

Building on the methodology described in Section 3.3.3, this chapter continues to evaluate model performance using 10-fold cross-validation and R^2 (Coefficient of Determination) as the primary metric. The average R^2 scores across folds are presented using box plots and summary tables to facilitate comparison under different feature selection strategies. As a measure of explained variance, R^2 ranges from $-\infty$ to 1, with higher values indicating better model fit.

Model performance is assessed for Random Forest Regression (RFR), Gradient Boosting (GB), and Bagged Decision Tree Regression (BDTR) across both the T-video and Y-video datasets. Feature strategies include the number of replies (M1), and three textual representations: term frequency (M3_a), Word2Vec (M3_b), and TF-IDF (M3_c), used independently or in combination. Based on the findings in Chapter 5 (see Section 5.2), features M2 (temporal) and M4 (categories) were found to have negligible predictive power and are therefore excluded from the analyses in this chapter.

This section aims to provide a comprehensive overview of model performance under the refined feature strategies, focusing on M1 and M3 variants. A comparative analysis of the T-video and Y-video datasets is conducted to examine how different inputs affect model effectiveness in predicting user preferences.

6.2.1 M1 and M3 as Input Data based on T-video and Y-video

This section analyses the average R^2 scores of single-feature strategies-M1 (number of replies), M3_a (term frequency), M3_b (Word2Vec), and M3_c (TF-IDF), across the T-video and Y-video datasets. These features serve as input to three machine learning models: Random Forest Regression (RFR), Gradient Boosting (GB), and Bagged Decision Tree Regression (BDTR). Detailed definitions of the models can be found in Section 3.4.3.

Table 6.1 presents the average R^2 scores for both datasets under each feature strategy, while Figure 6.1 visualises the corresponding results.

Table 6.1: The average R^2 score with 10-fold cross-validation (Random Forest Regression, Gradient Boost, BDTR) for M1, M3_a, M3_b and M3_c

		Average R^2 Score	
Features strategy	Algorithm	T-video	Y-video
M1	RFR	0.82	0.14
	GB	0.85	0.18
	BDTR	0.81	0.14
M3_a	RFR	0.20	0.45
	GB	0.22	0.45
	BDTR	0.11	0.40
M3_b	RFR	-0.06	-0.04
	GB	-0.09	-0.07
	BDTR	-0.14	-0.14
M3_c	RFR	0.17	0.44
	GB	0.22	0.35
	BDTR	0.09	0.41

As shown in Table 6.1, under the M1 feature strategy, all three models perform well on the T-video dataset. RFR and BDTR yield similar R^2 scores of 0.82 and 0.81, respectively, while GB achieves a slightly higher score of 0.85. In contrast, performance on the Y-video dataset is significantly lower. GB performs slightly better with an R^2 of 0.18, while RFR and BDTR both score 0.14.

Under M3_a (term frequency), model performance declines on the T-video dataset, with R^2 values of 0.20 (RFR), 0.22 (GB), and 0.11 (BDTR). However, results improve

on the Y-video dataset. RFR and BDTR perform comparably with R^2 values around 0.40–0.45, while GB scores slightly lower at 0.40.

For M3_b (Word2Vec), all models perform poorly across both datasets, with negative R^2 values. This indicates a failure to capture meaningful patterns and suggests that this feature strategy may lack predictive relevance in this context.

Under M3_c (TF-IDF), moderate performance is observed on the T-video dataset. RFR and GB yield R^2 scores of 0.17 and 0.22, respectively, while BDTR lags behind at 0.09. In contrast, results improve significantly on the Y-video dataset. RFR and BDTR again perform similarly, with R^2 scores between 0.41 and 0.44, while GB performs slightly worse, scoring 0.35. These results closely mirror those obtained under M3_a (term frequency), indicating that TF-IDF and term frequency features exhibit similar predictive capacities across both datasets.

6.2.2 M1+M3 as Input Data based on T-video and Y-video

This section examines model performance under combined feature strategies, specifically M1 with each variant of M3. Given that M3 includes multiple text analysis methods, the evaluation is organised by M3 subtype to assess the impact of each combination. This approach helps identify the most effective feature configurations for predicting user preferences and engagement with video content.

Table 6.2: The average R^2 score with 10-fold cross-validation (Random Forest Regression, Gradient Boost, BDTR) for M1+M3 input data.

		Average R^2 Score	
Features strategy	Algorithm	T-video	Y-video
M1+M3_a	RFR	0.85	0.58
	GB	0.85	0.51
	BDTR	0.84	0.54
M1+M3_b	RFR	0.84	0.16
	GB	0.85	0.14
	BDTR	0.83	0.09
M1+M3_c	RFR	0.85	0.54
	GB	0.85	0.48
	BDTR	0.83	0.50

Analysing Table 6.2, these model performances on the T-video and Y-video datasets were compared by using combined feature strategies: M1 (number of replies) with each M3 variant (M3_a: term frequency, M3_b: Word2Vec, M3_c: TF-IDF).

Under the M1+M3_a strategy, RFR and GB both achieve the highest R^2 score of 0.85 on the T-video dataset, indicating strong predictive performance. On the Y-video dataset, RFR records a relatively higher R^2 of 0.58, followed by BDTR at 0.54 and GB at 0.51. Overall, all three models exhibit comparable performance on T-video, while RFR and BDTR show a slight advantage over GB on Y-video.

For M1+M3_b, all three models perform well on the T-video dataset (R^2 between 0.83 to and 0.85), largely due to the contribution of the M1 feature (reply count). However, their performance drops sharply on the Y-video dataset, with R^2 values ranging from only 0.09 to 0.16. This indicates that the Word2Vec features (M3_b) contribute little to no predictive power in this context, an observation consistent with findings in Chapter 5, where M3_b failed to improve model performance on either dataset when used alone.

In the M1+M3_c setting, RFR and GB both achieve the highest R^2 score on T-video (0.85), indicating strong predictive accuracy when reply count is combined with TF-IDF features. BDTR also performs well with an R^2 of 0.83. On the Y-video dataset, performance improves relative to using TF-IDF alone, with GB and BDTR achieving R^2 scores of 0.48 and 0.50, respectively, suggesting better adaptability when interaction features are integrated.

RFR consistently performs best across both datasets, particularly on Y-video. Its robust generalisation likely benefits from randomised feature and sample selection, which helps capture dataset-specific patterns. GB performs similarly to RFR on T-video but lags behind on Y-video, especially when used with Word2Vec features, indicating potential sensitivity to complex, high-dimensional text inputs.

Models generally perform similarly on T-video, but diverge significantly on Y-video. This difference suggests that the Y-video dataset may have higher complexity or distinct underlying data characteristics, warranting careful consideration during model selection and tuning. Word2Vec-based features do not provide any meaningful predictive power

in the current modelling setup, as evidenced by consistently poor R^2 scores across both datasets. These results suggest that, within the context of ensemble learning approaches used in this study, Word2Vec may not capture relevant signals for predicting user engagement.

6.3 Performance Metrics Validation Results

Building on the 10-fold cross-validation results, this section further evaluates model performance using an 80/20 train-test split. This approach enables a direct comparison between predicted and actual values, offering additional insight into model accuracy under different feature strategies.

The analysis is organised into three parts for clarity: (1) single-feature input (M1), (2) individual text analysis methods (M3 variants), and (3) combined feature strategies. This structure allows for a focused assessment of how each model responds to increasing data complexity and different input types when predicting user engagement with video content.

6.3.1 Single-feature Input Data based on T-video and Y-video

In this section, the RMSE (Root Mean Square Error) and R^2 (Coefficient of Determination) metrics are presented for the Random Forest (RFR), Gradient Boosting (GB), and BDTR models using single-feature inputs-M1 (number of replies), M3_a (term frequency), M3_b (Word2Vec), and M3_c (TF-IDF), on both training and testing subsets of the T-video and Y-video datasets.

Additionally, we present scatter plots comparing predicted and actual values to evaluate model fit. This provides a clearer view of each model's accuracy under different feature inputs and reveals strengths and limitations in predicting user engagement.

Table 6.3: Performance metrics validation results of T-video and Y-video for RFR, GB and BDTR based on M1 and M3

		Training data				Test data			
		R^2		RMSE		R^2		RMSE	
Features strategy	**Algorithm**	**T-video**	**Y-video**	**T-video**	**Y-video**	**T-video**	**Y-video**	**T-video**	**Y-video**
M1	Random Forest	0.87	0.22	1.4	1.49	0.84	0.19	1.61	1.49
	Gradient Boost	0.92	0.27	1.09	1.44	77	0.13	1.9	1.55
	BDTR	0.92	0.26	1.13	1.44	0.79	0.13	1.83	1.54
M3_a	Random Forest	0.29	0.39	3.29	1.32	0.13	0.3	3.68	1.4
	Gradient Boost	0.89	0.88	1.28	0.59	0.08	0.38	3.86	1.3
	BDTR	0.89	0.93	1.29	0.46	0.16	0.43	3.68	1.25
M3_b	Random Forest	0.19	0.11	3.51	1.59	0	-0.01	4.01	1.67
	Gradient Boost	0.99	0.99	0.11	0.12	-0.19	-0.15	4.38	1.78
	BDTR	0.86	0.86	1.47	0.63	-0.02	-0.05	4.06	1.7
M3_c	Random Forest	0.96	0.93	0.81	0.45	0.09	0.34	3.83	1.34
	Gradient Boost	0.89	0.93	1.3	0.46	0.15	0.46	3.71	1.22
	BDTR	0.31	0.4	3.24	1.3	0.13	0.36	3.74	1.33

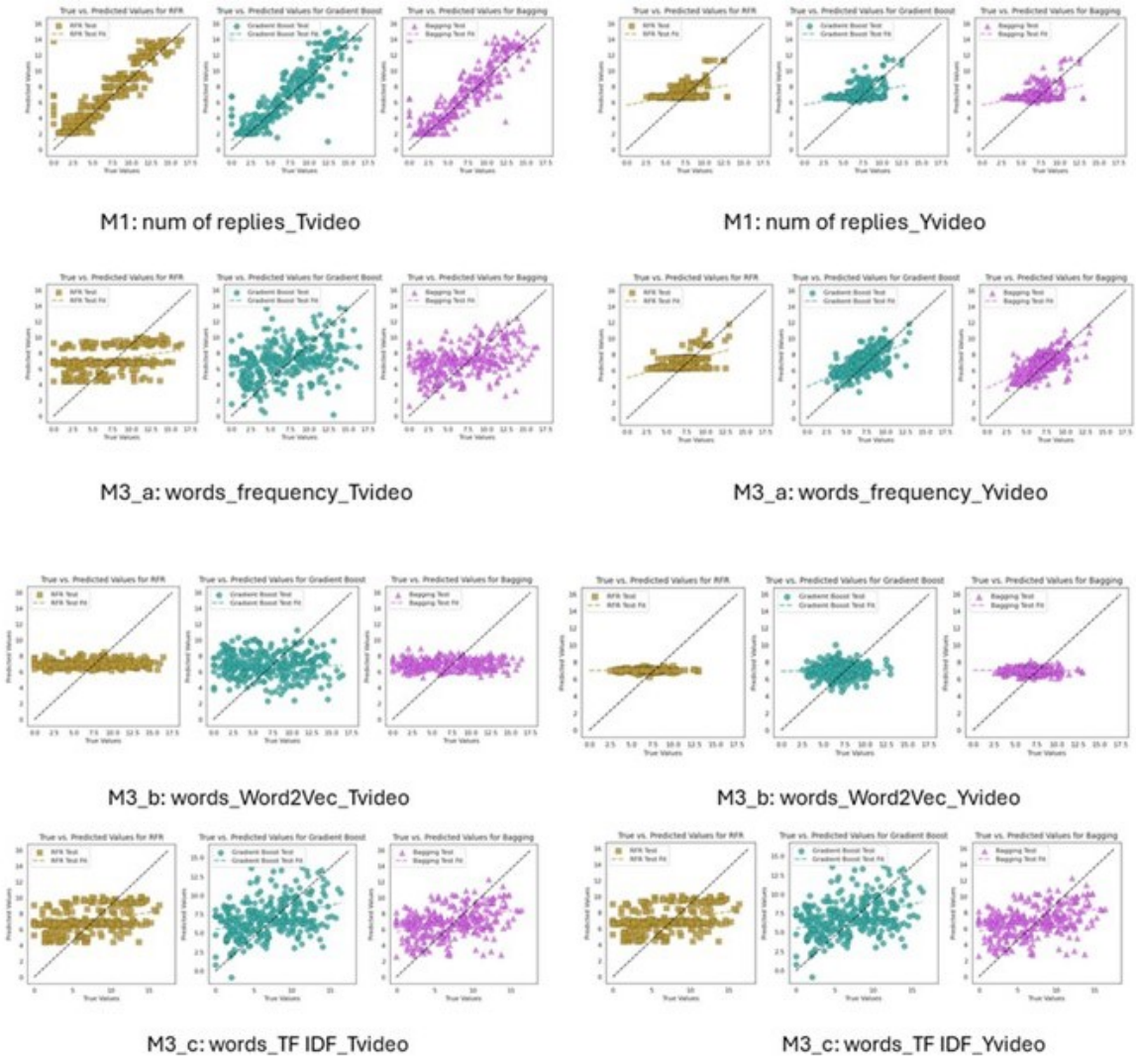


Figure 6.1: True vs. Predicted Values in T-video and Y-video for RFR, GB and BDTR based on M1 and M3

Based on Table 6.3 and Figure 6.1, we evaluate model performance across different feature strategies using both the T-video and Y-video datasets.

M1 (Number of Replies): On the T-video dataset, all three models, RFR, GB, and BDTR, achieve high training R^2 values (close to or above 0.90), indicating excellent fit. RFR performs best on the test set ($R^2 = 0.84$), with GB and BDTR slightly lower, suggesting that RFR generalises more effectively. On the Y-video dataset, despite good training performance, all models show low test R^2 scores (below 0.19), pointing to overfitting and limited generalisation. Among them, RFR again delivers the most stable performance across both sets.

M3_a (Term Frequency): Performance drops significantly for both datasets. On T-video, RFR achieves only 0.29 on the training set, with even lower scores on the test set for all models, suggesting that term frequency features are not well-suited to this task. Y-video shows similarly weak results, with all test R^2 values below 0.4.

M3_b (Word2Vec): This feature strategy performs poorly across both datasets, with R^2 values near zero or negative, indicating that Word2Vec features fail to capture relevant patterns and offer little predictive value.

M3_c (TF-IDF): On the T-video dataset, RFR and GB fit the training data well but show significant performance drops on the test set, indicating overfitting. Y-video results are slightly better but still limited, implying that TF-IDF features do not generalise effectively across platforms.

As shown in Figure 6.1, under the M1 strategy, predicted values closely align with the true values on the T-video dataset, forming tight clusters along the diagonal. On the Y-video dataset, predictions are more dispersed, though Gradient Boosting shows relatively better alignment. In contrast, the M3_a and M3_c strategies, particularly on Y-video, exhibit noticeably higher prediction errors. For M3_b, the predicted values, especially from GB, deviate significantly from the actual values, reflecting the instability of Word2Vec-based features.

Across models, Random Forest Regression (RFR) consistently delivers the most stable predictions on both datasets. GB and BDTR tend to overfit, performing well

on training data but poorly on test data, especially with the T-video dataset. This pattern is reflected in the scatter plots, where only the M1 strategy shows concentrated points along the diagonal, while all M3-based strategies yield scattered and less accurate predictions, particularly for Y-video.

Table 6.4: Significance test results (p-values) for T-video and Y-video R^2 scores across models

Comparison	T-video p-value	Y-video p-value
M1: RFR vs GB	0.041	0.054
M1: RFR vs BDTR	0.032	0.039
M1: GB vs BDTR	0.276	0.198
M3_a: RFR vs GB	0.482	0.503
M3_a: RFR vs BDTR	0.215	0.421
M3_a: GB vs BDTR	0.318	0.372
M3_b: RFR vs GB	0.058	0.062
M3_b: RFR vs BDTR	0.067	0.048
M3_b: GB vs BDTR	0.462	0.417
M3_c: RFR vs GB	0.026	0.031
M3_c: RFR vs BDTR	0.045	0.042
M3_c: GB vs BDTR	0.397	0.389

To assess whether the observed differences in predictive performance across models were statistically significant, independent two-sample t-tests were conducted on the R^2 scores for T-video and Y-video datasets, respectively. The results in Table 6.4 show that for the M1 feature strategy, the R^2 differences between Random Forest Regression (RFR) and Gradient Boosting (GB), as well as between RFR and Bagged Decision Tree Regression (BDTR), are statistically significant for the T-video dataset ($p < 0.05$), with similar significance observed between RFR and BDTR for the Y-video dataset.

In contrast, for feature strategies M3_a and M3_b, no statistically significant differences ($p > 0.05$) were found for most model comparisons, suggesting that these feature sets do not lead to substantial performance variation across algorithms. For M3_c, significant differences were observed between RFR and both GB and BDTR in both datasets ($p < 0.05$), indicating that algorithm choice plays a more critical role when the feature composition changes.

The significance tests were performed separately for T-video and Y-video rather

than directly comparing their results. This is because each dataset represents a different prediction task with distinct underlying distributions, making a direct cross-dataset statistical comparison inappropriate. Instead, running tests within each dataset allows for a fairer assessment of algorithmic differences while controlling for dataset-specific effects.

Overall, these findings confirm that the choice of model significantly impacts predictive performance for certain feature strategies (particularly M1 and M3_c), while in other cases (e.g., M3_a, M3_b), feature set composition appears to be the dominant factor influencing results.

In summary, M1 (number of replies) remains the most effective single-feature strategy for both datasets, although overfitting is observed on Y-video. The M3 strategies, term frequency, Word2Vec, and TF-IDF, generally underperform, underscoring the need for more advanced preprocessing and refined feature selection. These results highlight the importance of aligning feature engineering with dataset characteristics to improve model generalisation and predictive accuracy.

6.3.2 M1+M3 as Input Data based on T-video and Y-video

This section evaluates the performance of Random Forest Regression (RFR), Gradient Boosting (GB), and BDTR models using the M3 feature strategy, which includes term frequency (M3_a), Word2Vec (M3_b), and TF-IDF (M3_c). The analysis considers both RMSE and R^2 across training and testing sets for the T-video and Y-video datasets.

In addition, the predicted versus actual values are compared to assess model accuracy. This analysis offers insights into the predictive effectiveness of different text representation methods for modelling user engagement with video content.

Based on Table 6.5 and Figure 6.2, this section evaluates model performance under the M1+M3 feature combinations, term frequency (M3_a), Word2Vec (M3_b), and TF-IDF (M3_c), on both the T-video and Y-video datasets. Results from both training and testing sets are used to assess each model's ability to predict user engagement with video content.

Table 6.5: Performance metrics validation results of T-video and Y-video for RFR, GB and BDTR based on M1+M3

		Training data				Test data			
		R^2		RMSE		R^2		RMSE	
Features strategy	**Algorithm**	**T-video**	**Y-video**	**T-video**	**Y-video**	**T-video**	**Y-video**	**T-video**	**Y-video**
M1+M3_a	Random Forest	0.89	0.47	1.31	1.23	0.84	0.43	1.59	1.25
	Gradient Boost	0.99	0.91	0.46	0.5	0.82	0.51	1.69	1.16
	BDTR	0.98	0.94	0.55	0.4	0.82	0.57	1.69	1.08
M1+M3_b	Random Forest	0.89	0.29	1.3	1.42	0.84	0.18	1.59	1.51
	Gradient Boost	0.99	1	0.04	0.11	0.81	0.06	1.74	1.61
	BDTR	0.98	0.89	0.57	0.57	0.84	0.15	1.62	1.53
M1+M3_c	Random Forest	0.89	0.89	1.28	1.28	0.84	0.84	1.59	1.59
	Gradient Boost	0.99	0.99	0.37	0.37	0.83	0.83	1.64	1.64
	BDTR	0.98	0.98	0.55	0.55	0.83	0.83	1.65	1.65

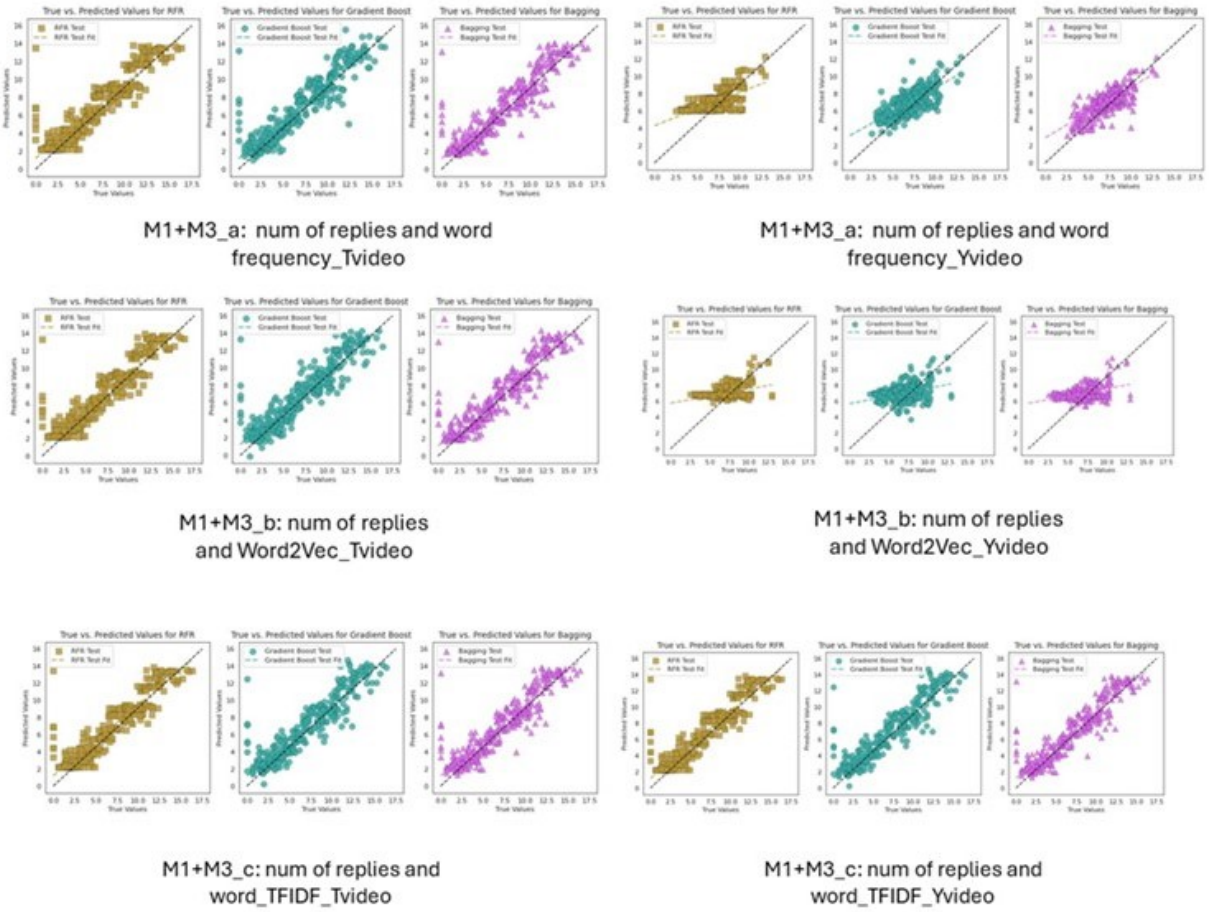


Figure 6.2: True vs. Predicted values in T-video and Y-video for RFR, GB and BDTR based on M1+M3

From Table 6.5 and Figure 6.2, we can state the following:

M1+M3_a (Number of Replies + Term Frequency): On the T-video dataset, GB and BDTR exhibit strong training performance, with R^2 values approaching 1. However, on the test set, their R^2 scores drop to 0.82 with RMSEs around 1.69, suggesting potential overfitting, particularly for GB. Scatterplots show tight alignment between predicted and true values, especially for GB, though this consistency does not fully translate to unseen data. On the Y-video dataset, RFR achieves a test R^2 of 0.43 and an RMSE of 1.25, indicating more stable generalisation. While GB and BDTR yield slightly better R^2 scores, their high training performance ($R^2 \approx 1$) alongside modest test scores further point to overfitting.

M1+M3_b (Number of Replies + Word2Vec): On the T-video dataset, RFR maintains a consistent R^2 of 0.84 across both training and test sets, with GB and BDTR also achieving comparable performance (0.81 and 0.84, respectively). However, these results show no meaningful improvement over using M1 alone, indicating that the inclusion of Word2Vec features adds little or no additional predictive value. Although scatterplots show relatively tight clustering of predicted values, the slight performance drop for GB again suggests overfitting. On the Y-video dataset, all models exhibit poor generalisation: GB performs the worst with an R^2 of just 0.06, and the scatterplots reveal wide dispersion. These results further confirm that Word2Vec features are not informative for this dataset.

M1+M3_c (Number of Replies + TF-IDF): On T-video, RFR and GB reach R^2 scores of 0.84 and 0.83, respectively, with RMSEs of 1.59 and 1.64. BDTR performs similarly. The predicted values align closely with actual ones in the scatterplot, confirming stable performance. On Y-video, all models again achieve an R^2 of 0.84, though RMSEs vary slightly. While predictive accuracy is slightly lower than in T-video, overall generalisation remains strong.

Across feature strategies, Gradient Boosting (GB) consistently achieves high R^2 on training sets but often suffers significant drops on test sets, indicating a tendency toward overfitting. Random Forest Regression (RFR), in contrast, demonstrates supe-

rior generalisation, especially notable in both M1+M3_a and M1+M3_c. BDTR offers stable performance across most configurations, though it tends to underperform on the Y-video dataset with M3_a.

Scatterplots (Figure 6.2) corroborate these findings: while predicted values cluster near the diagonal under M1+M3_c, the dispersion widens under M3_b, particularly for GB on Y-video.

The M1+M3_c combination consistently delivers the most robust and generalizable results across both datasets.

6.4 Models Performance and Feature Impact on User Engagement

In this section, the performance of different machine learning models is examined to assess their effectiveness in predicting user engagement metrics. The analysis particularly focuses on the combination and comparison of ensemble methods such as Random Forest, Gradient Boosting, and BDTR-Random-Forest. By evaluating their predictive accuracy and robustness, this section aims to identify which modeling approach yields the most reliable results under varying feature configurations.

6.4.1 Performance of BDTR-Random-Forest, BDTR and Random Forest

Following the identification of Random Forest as the optimal model for the T-video dataset and BDTR as the preferred choice for Y-video, this section introduces a third approach: the BDTR-Random Forest model. By integrating the mechanisms of both Random Forest and BDTR, this ensemble aims to further enhance predictive performance. For detailed methodology, see Section 3.3.

Table 6.6 presents the performance metrics of BDTR, Random Forest, and the hybrid BDTR-Random Forest model across both datasets. This comparison evaluates whether the combined approach can outperform its individual components, offering a more effective strategy for modelling user engagement with video content.

Table 6.6: Comparison of model performance results for BDTR-Random-Forest, BDTR and Random Forest

		Training data				Test data			
		R^2		RMSE		R^2		RMSE	
Features strategy	**Algorithm**	**T-video**	**Y-video**	**T-video**	**Y-video**	**T-video**	**Y-video**	**T-video**	**Y-video**
M1+M3_c	Random Forest	0.89	0.89	1.28	1.28	0.84	0.84	1.59	1.59
	BDTR	0.98	0.98	0.55	0.55	0.83	0.83	1.65	1.65
	BDTR-Random-Forest	0.95	0.95	0.89	0.89	0.84	0.84	1.6	1.6
M1+M2+M3_c	Random Forest	0.89	0.48	1.29	1.22	0.84	0.43	1.58	1.25
	BDTR	0.98	0.94	0.56	0.42	0.84	0.55	1.6	1.11
	BDTR-Random-Forest 0.95	0.84	0.89	0.69	0.85	0.55	1.58	1.11	
M1+M2+M4+M3_c	Random Forest	0.89	0.48	1.29	1.21	0.84	0.43	1.58	1.25
	BDTR	0.98	0.94	0.56	0.41	0.83	0.55	1.63	1.11
	BDTR-Random-Forest	0.95	0.84	0.89	0.68	0.85	0.55	1.58	1.11

In Table 6.6, all models exhibit strong training performance on the T-video dataset under the M1+M3_c strategy, with BDTR and BDTR-Random-Forest performing notably well. On the test set, Random Forest and BDTR achieve comparable R^2 scores, though Random Forest yields a slightly lower RMSE, indicating marginally better predictive accuracy.

When extended to the M1+M2+M3_c strategy, BDTR achieves the highest R^2 and the lowest RMSE on the test set, suggesting optimal performance. The inclusion of the M2 feature (time) offers limited improvement, indicating it contributes little to predictive value. Adding M4 (category) in the M1+M2+M3_c+M4 strategy maintains BDTR's lead, though overall performance gains remain minimal.

On the Y-video dataset, under M1+M3_c, BDTR and Random Forest achieve identical R^2 scores, but BDTR records a lower RMSE, suggesting slightly better generalisation. It continues to outperform under M1+M2+M3_c and retains this advantage with the inclusion of M4, showing consistent performance across feature combinations.

In summary, BDTR demonstrates the most stable and accurate predictions across both datasets under the full M1+M2+M3_c+M4 strategy. However, the limited improvements from M2 and M4 indicate that time and category features offer marginal added value. These results highlight the model's strength in integrating relevant features without relying on less informative contextual inputs.

6.4.2 The Importance of Features on User Engagement with Music Videos on T-video and Y-video

Following comparative analysis, the relevance of feature combinations in M1+M3_c, M1+M2+M3_c, and M1+M2+M3_c+M4applied using the BDTR model across both datasets, confirms earlier findings: features from M2 (time) and M4 (categories) contribute minimally to predictive performance. As detailed in Chapter 5, the specific attributes within M2 and M4 do not significantly improve accuracy in modelling user preferences.

Based on these insights, Table 6.7 presents the key features identified by the BDTR model under the M1+M3_c strategy. This highlights the predictive strength of combining reply counts (M1) with TF-IDF-based textual features (M3_c), emphasizing the value of these variables in enhancing model accuracy for user engagement prediction.

Feature_M1+M3_c_T-video	Importance	Feature_M1+M3_c_Y-video	Importance
reply_count	0.863	reply_count	0.234
offici	0.004	pixl	0.072
omg	0.003	poptast	0.042
miss	0.003	stumm	0.026
like	0.002	instrument	0.024
mean	0.002	channel	0.018
new	0.002	one	0.018
world	0.002	clifford	0.016
love	0.001	network	0.016
great	0.001	make	0.015

Table 6.7: Top 10 features based on feature importance scores in T-video and Y-video dataset based on M1+M3_c

As shown in Table 6.7, in the T-video dataset, reply_count is the most influential feature, underscoring the importance of user engagement on Twitter. Discussions and replies surrounding video content enhance visibility and viewer interaction on YouTube. Emotional or expressive terms such as omg, miss, like, love, and great also rank highly, suggesting that emotionally resonant language contributes to user interest. The inclusion of the office likely reflects engagement with content from official artist channels, emphasising the influence of credibility and artist recognition on viewer behaviour.

In contrast, reply_count is less impactful in the Y-video dataset, indicating that direct interaction plays a smaller role in engagement on YouTube itself. Instead, features related to channel branding, such as pixl, poptast, and stumm, are more prominent, suggesting user loyalty to particular channels. This highlights a platform-specific dynamic, where engagement in T-video is driven by social interaction and sentiment, while Y-video engagement is shaped by branding and creator identity.

While time and category features contribute little overall, the Category - Music variable shows limited relevance in Y-video, possibly due to differences in content origin. T-video content often originates from official artist accounts, where fan bases drive engagement regardless of metadata. This aligns with the principle of source credibility, where audiences are more likely to trust and interact with content from verified sources.

In contrast, Y-video content frequently stems from individual or semi-professional music channels, where metadata such as category selection plays a more active role in discoverability and promotion, particularly through YouTube's recommendation algorithms.

Finally, reply_count retains value in T-video due to the social nature of Twitter. User decisions to engage with video content may be shaped by visible reactions from others, such as comments or retweets. On YouTube, engagement appears more influenced by creator branding and platform-driven discovery, pointing to longer-term viewer relationships and algorithmic visibility.

In summary, the distinction between T-video and Y-video reflects differing platform dynamics: T-video emphasises immediate social interaction, while Y-video prioritises sustained engagement through creator consistency and channel credibility.

6.5 Discussion and Limitation

6.5.1 Discussion

This chapter deepens the understanding of what drives the popularity of YouTube music videos across platforms. By analysing user preferences and identifying key predictive features in the T-video and Y-video datasets, each representing different sources of

YouTube content, this study assesses the effectiveness of various modelling and feature selection strategies. It further explores ensemble machine learning methods, particularly Bagging Regression based on Random Forest, to evaluate the predictive power of cross-platform user engagement signals.

These findings build upon previous cross-platform research on YouTube content (Gatta et al., 2023; Ginossar et al., 2022; Golovchenko et al., 2020; Wallin, 2021), incorporating both classical and contemporary ensemble approaches. Like these studies, the work integrates social media data across technological and social dimensions, contributing to a broader understanding of user engagement across digital platforms.

Importantly, this study highlights the interaction between user-generated content and platform dynamics in the domain of music videos. It shows how user engagement and dissemination patterns vary between platforms, offering insights into how the socio-technical environment shapes content visibility and reach across YouTube and Twitter.

Those results also echo prior research (Nathan, 2022; Nisa et al., 2021; Spasojevic et al., 2015), affirming the influence of contextual factors, such as timing and self-presentation, on user engagement. We find that YouTube creators increasingly emphasise personal branding and professional presentation, often surpassing official channels in visibility and influence. This trend aligns with YouTube's ethos of "broadcast yourself", where self-curated content gains traction through credibility and consistency.

Conversely, Twitter fosters more immediate, interaction-driven engagement. Comments and retweets significantly influence how videos gain attention, highlighting the importance of conversational context over passive viewership. This contrast is evident in the differing relevance of features across the T-video and Y-video datasets.

In addition, those findings reflect broader trends in social media use, including simultaneous media consumption, music sharing, and real-time commentary, especially on Twitter. While both platforms support vibrant fan communities, the modes of interaction differ: Twitter prioritises brief, public exchanges, while YouTube supports longer-form engagement and more specialised communities, often centred around individual creators or official channels.

Ultimately, platform design strongly shapes user behaviour. Twitter's structure

encourages rapid interaction and broad visibility, while YouTube's recommendation algorithms and content depth foster sustained, interest-based communities. Together, these differences reveal how platform affordances influence not only engagement levels but also the nature of content consumption and social exchange.

6.5.2 Limitation

While Random Forest, Gradient Boosting, and BDTR offer a robust framework for analysing user engagement and preferences, their effectiveness may vary across datasets and contexts. Relying solely on these ensemble models may overlook the complexity and variability of user behaviour across platforms. Future studies could explore alternative approaches such as deep learning methods (e.g., LSTM, Transformers) to capture sequential or semantic nuances in user-generated content, or hybrid models that integrate rule-based and learning-based components to improve interpretability and domain adaptability.

Another limitation lies in the method of linking YouTube video popularity with Twitter sharing activity. We collected tweets using common hashtags and keywords to associate them with YouTube music videos. However, this approach tends to favour high-production-value videos, which are more frequently shared on Twitter. As a result, videos with lower visibility or niche appeal may be underrepresented, potentially biasing the understanding of user engagement. Additionally, differences in the statistical distribution of the T-video and Y-video datasets may further impact the interpretation of model results. In particular, the T-video dataset appears to contain a small number of videos with exceptionally high reply counts and engagement levels. These outliers may disproportionately influence the feature importance rankings, making reply_count appear more predictive in T-video than in Y-video. This skewed distribution introduces a potential bias in the evaluation of feature effectiveness.

This pattern of selective visibility mirrors the platform-specific "content ecologies" described by Segerberg and Bennett (Segerberg and Bennett, 2011), where certain narratives gain prominence due to the affordances of hashtags and platform dynamics. Similarly, the dataset may privilege specific types of content, limiting the generalizabil-

ity of those findings. This highlights the challenge of capturing the full diversity of user engagement across platforms through aggregated social media data.

6.6 Summary

This research aims to enhance both academic and practical understanding of how user engagement and preferences toward YouTube music videos are evaluated across platforms. To this end, it analyses data from two primary sources: YouTube directly and YouTube links shared via Twitter.

The study demonstrates the effectiveness of ensemble machine learning techniques, Random Forest, Gradient Boosting, and BDTR, in predicting user engagement. Results show that Random Forest achieves the highest predictive accuracy on the T-video dataset, as indicated by the highest R^2 and lowest RMSE values. In contrast, BDTR demonstrates superior generalisation performance on the Y-video dataset, outperforming other models in terms of both R^2 and RMSE on the test set. These findings are validated using 10-fold cross-validation, train-test data splitting, and comparisons between predicted and actual values. Among all tested models and feature strategies, the BDTR combined with the M1+M3_c feature set delivers the highest predictive accuracy.

The analysis also highlights distinct factors that influence video popularity across platforms. On YouTube, fostering a strong brand and channel identity through consistent theming, creator appeal, and content innovation is key to sustaining user engagement. In contrast, videos shared via Twitter benefit more from content quality and artist reputation, as engagement on this platform is typically broader and driven by emotional or social response.

Overall, these insights emphasise the importance of tailoring content strategies to platform-specific user behaviours and leveraging appropriate machine learning tools for engagement prediction.

Chapter 7

Heterogeneity of Cross-Platform User Sentiment

7.1 Introduction

Social media platforms like YouTube, Facebook, and Twitter have profoundly transformed the manner in which individuals interact by facilitating the interchange of perspectives regardless of temporal or geographical constraints. Simultaneously, the utilisation of social media carries inherent hazards, including susceptibility to manipulators like social bots, the rapid dissemination of false information, and the creation of "echo chambers", online environments where users are only exposed to information and viewpoints that align with their (Guess et al., 2018). Within such virtual environments, people are only subjected to information and viewpoints that align with their perspectives. When viewers watch music MVs, they are often influenced by fan culture. As a result, they tend to prefer videos published by their favourite YouTubers or stars. Additionally, they are more likely to accept video suggestions from other users who have similar views.

Sentiment analysis (SA) is a research area that seeks to analyse people's feelings or views on entities such as topics, events, individuals, issues, services, goods, organizations, and their characteristics (Liu, 2022). Although the history of natural language processing (NLP) dates back to the 1950s, little attention was paid to people's opinions

and sentiment analysis until 2005. For the past few years, the growth of social media has fueled the development of sentiment analysis (Saberi and Saad, 2017).

Sentiment analysis holds potential utility in predicting the successful outcomes of various events, products, and entities by examining the collective sentiment expressed on social media platforms. By analysing text messages from individual users, valuable insights can be gleaned about their ideas, behaviour, and personality traits. This individual-level analysis is crucial as it forms the basis for aggregating data to reflect broader public opinion. Furthermore, a topic-centric approach allows for the examination of comments related to specific events or entities, providing a comprehensive view of the public's stance, emotions, and attitudes towards these topics. By employing sentiment and opinion mining models, it is possible to aggregate these individual responses to deduce the overall emotional and subjective reactions of the public towards a particular object or event. This methodological approach enables us to move from understanding personal sentiments to making informed predictions about the general public sentiment.

Previous studies have also denoted sentiment analysis as opinion mining, encompassing a diverse array of techniques spanning natural language processing (NLP), information extraction, artificial intelligence, machine learning (ML), data mining (DM), and even psychoanalysis. According to Asghar et al., there exist four distinct categories of comments found on the YouTube platform (Asghar et al., 2015):

(1) Brief Evaluative Remarks: These remarks typically exhibit a pleasant sentiment while lacking intellectual depth.

(2) Advertisements: The purpose of these comments is to promote the products or services of an organisation or corporation.

(3) Adverse critique: The aforementioned remarks pertain to the act of demeaning an individual.

(4) The incoherent and disjointed argument: These comments pertain to religious and political videos, consistently expressing opposition towards the ideas endorsed in the video.

Hence, based on those kinds of social media comments, as the main application of

sentiment analysis, sentiment classification can be divided into three main steps (Chen et al., 2017):

(1) The subjective analysis: The text identifies instances of subjective opinions that communicate personal sentiment, as well as objective statements of facts;

(2) Polarity classification: The objective is to identify individual feelings related to special events;

(3) Sentiment strength detection: Once the sentiment classification process is finished, the analysis proceeds to detect the strength of the sentiment. For example, while videos deemed worthy of rewatching and those regarded as good quality are both characterised by positive emotion, they may exhibit varying degrees of sentiment polarity.

Hence, in order to obtain supplementary sentiment information, it is necessary to conduct a more in-depth examination of the intensity of both positive and negative sentiments in sentiment analysis. The field of sentiment classification, as discussed in previous studies (Pang et al., 2002; Thomas et al., 2006), focuses on the task of automatically assigning opinion values, such as "positive", "negative", or "neutral" to documents or topics. This is achieved through the utilisation of diverse text-oriented and linguistic aspects.

Krouska et al. explored the impact of different preprocessing techniques on the performance of sentiment analysis models applied to Twitter data in 2016. Their study compared various methods such as text cleaning, stopword removal, stemming, and lemmatisation, showing that appropriate preprocessing significantly improves the accuracy of sentiment classification, especially when dealing with noisy and informal text common on social media platforms. The findings suggest that preprocessing not only reduces noise but also enhances the model's ability to accurately identify sentiment. The study highlights the critical role of selecting the right preprocessing techniques in improving the overall performance of sentiment analysis models (Krouska et al., 2016).

As the digital media landscape evolves, understanding how users express emotions on different platforms and how these emotional expressions reveal consistency and differences among users has become a cutting-edge topic of research in the social sciences

and computational communication. This chapter aims to delve into this phenomenon by addressing the following research questions:

RQ5: Does the comparison of sentiment expressed by users' comments on similar topics on Twitter and YouTube reveal consistency or differences in user sentiment across platforms?

RQ6: If heterogeneity exists, is there an association with the categories used by the creators when they are uploading their music videos?

Through a systematic analysis of these issues, this study can not only reveal the complexity of cross-platform user behaviour but also gain a deeper understanding of the evolution of YouTube user sentiment dynamics in the digital age. This is important for developing effective communication strategies, optimising content creation, and enhancing user experience on social media platforms.

7.2 Sentiment Diversity in T-video and Y-video

Results relating to data preprocessing and data exploration have already been discussed in Sections 3.2 and 4.2. This section examines the sentiment trends in the T-video and Y-video datasets after applying VADER Sentiment Analysis.

7.2.1 The Positive and Negative Videos in T-video and Y-video

This section explores the diversity of sentiment in the two datasets, thus illustrating the breadth or consistency of the distribution of sentiment in video comments. In order to compare the sentiments expressed in the two datasets, this study calculated the sentiment scores of each comment using VADER, aggregated the comments by video ID, and further analysed the main concerns and the intensity of the sentiments expressed by the users in the two datasets with different polarity of sentiments (Refer to section 3.4.1 for details).

In the T-video dataset, comprising 1,538 videos, there are 1,370 positive videos and 31 negative ones. Similarly, the Y-video dataset contains 2,119 videos, with 1,714 positive and 146 negative videos. This indicates a significantly higher occurrence of positive

videos in both datasets. Regarding neutral videos and comments, their numbers are relatively low in both datasets. Additionally, neutral content often lacks clear sentiment, making it less informative for sentiment analysis. Therefore, this study focuses on analysing and comparing the two polar sentiments, positive and negative, to derive more meaningful insights. In order to help balance the perspective of the data, particularly when assessing what types of content are more likely to trigger user engagement, a random selection of 10% positive comments was taken in order to control the amount of data being compared, allowing for a more focused and manageable analysis, whilst at the same time providing a clear control that highlights the differences between positive and negative feedback and the characteristics of each. This approach allows us to analyse the characteristics of positive comments in a more focused manner, avoiding the challenge of being overwhelmed by the sheer volume of data. Additionally, this study will include a zoomed-in view of the local data alongside the full data comparison chart, providing deeper insights into the differences between the datasets.

This section is dedicated to exploring the diversity of sentiment in the two datasets, thus illustrating the breadth and consistency of the distribution of sentiments in video comments. In order to compare the sentiments expressed in the two datasets, this study calculated the sentiment scores of each comment using VADER, aggregated the comments by video ID, and further analysed the main concerns and the intensity of the sentiments expressed by the users in the two datasets with different polarities of sentiments.

(1) The Relationship between the Number of Replies and Likes in Positive Videos of T-video and Y-video

To clarify the process depicted in the upper panel of Figure 7.1, we analysed the relationship between the number of replies and the number of likes for positive videos in both the T-video and Y-video datasets. This involved plotting each positive video as a point on a scatter plot, where the x-axis represents the number of likes and the y-axis represents the number of replies. Blue dots correspond to T-video data, and orange dots represent Y-video data. Both T-video's and Y-video's data points are widely distributed in both the number of replies, but particularly in the number of

likes dimension, showing a wide range from less than ten likes (10) to more than 10 million likes (107). A relatively large number of data points lie in areas with low response counts (under 100). This may indicate that even though the comments had a low number of engagements, some of them received high likes due to the quality of their content or audience empathy. Specifically, the figure below in Figure 7.1 shows that although the majority of points are concentrated in the low likes range (i.e., between 102 and 104), there are still a significant number of points in the higher likes level (105 to 106). This reflects the fact that even with lower response counts, certain comments received a higher number of likes due to the appeal or resonance of their content. The left panel also shows that in the lower response count range, the distribution of likes shows greater variation, with some comments attracting a large number of likes, while most receive fewer.

The increase in the number of likes after the number of replies exceeded 100 was not as significant as when the number of responses was low. This could mean that while certain videos trigger more replies to engagements, this does not always translate into a corresponding percentage increase in likes. T-video performs more prominently in the high likes range and is especially denser in the 105 to 106 range, which could indicate that the content or the way user engagements are carried out on T-video's platform are able to elicit stronger positive reactions in the case of a few replies' situations elicit stronger positive reactions. Y-video's dots, while denser in the low number of likes region, also show some distribution in the high number of likes region, suggesting that there is also an ability to generate attention-grabbing content on Y-video.

The correlation coefficients between the number of replies and likes for positive videos are 0.23 in the T-video dataset and 0.51 in the Y-video dataset. These values indicate a weak positive correlation for T-video and a moderate positive correlation for Y-video. This suggests that in the Y-video dataset, videos with more likes are more likely to receive a higher number of replies compared to those in the T-video dataset. This analysis reveals a positive but weak correlation for T-video, suggesting that the influence of replies on likes is minimal, indicating dispersed engagements among users. In contrast, Y-video exhibits a moderate to strong positive correlation, suggesting that

replies significantly influence likes. This stronger correlation indicates that users on Y-video engage more actively and consistently, likely to favour videos with higher engagement levels. The data points on the Y-video are more concentrated, which supports the idea that there is a uniform pattern of engagement. On the other hand, the fewer data points at higher levels of replies and likes on T-video suggest that while some positive videos receive more attention, the overall pattern of engagement is less consistent. In analysing user engagement on YouTube, comments are a key metric, reflecting direct viewer interaction and interest. Therefore, this study considers the number of videos as a primary indicator of user engagement (Benevenuto, Rodrigues, Almeida, Almeida and Ross, 2009). The observed variability in videos and likes suggests that while some positive videos receive significant attention, overall engagement patterns are less consistent.

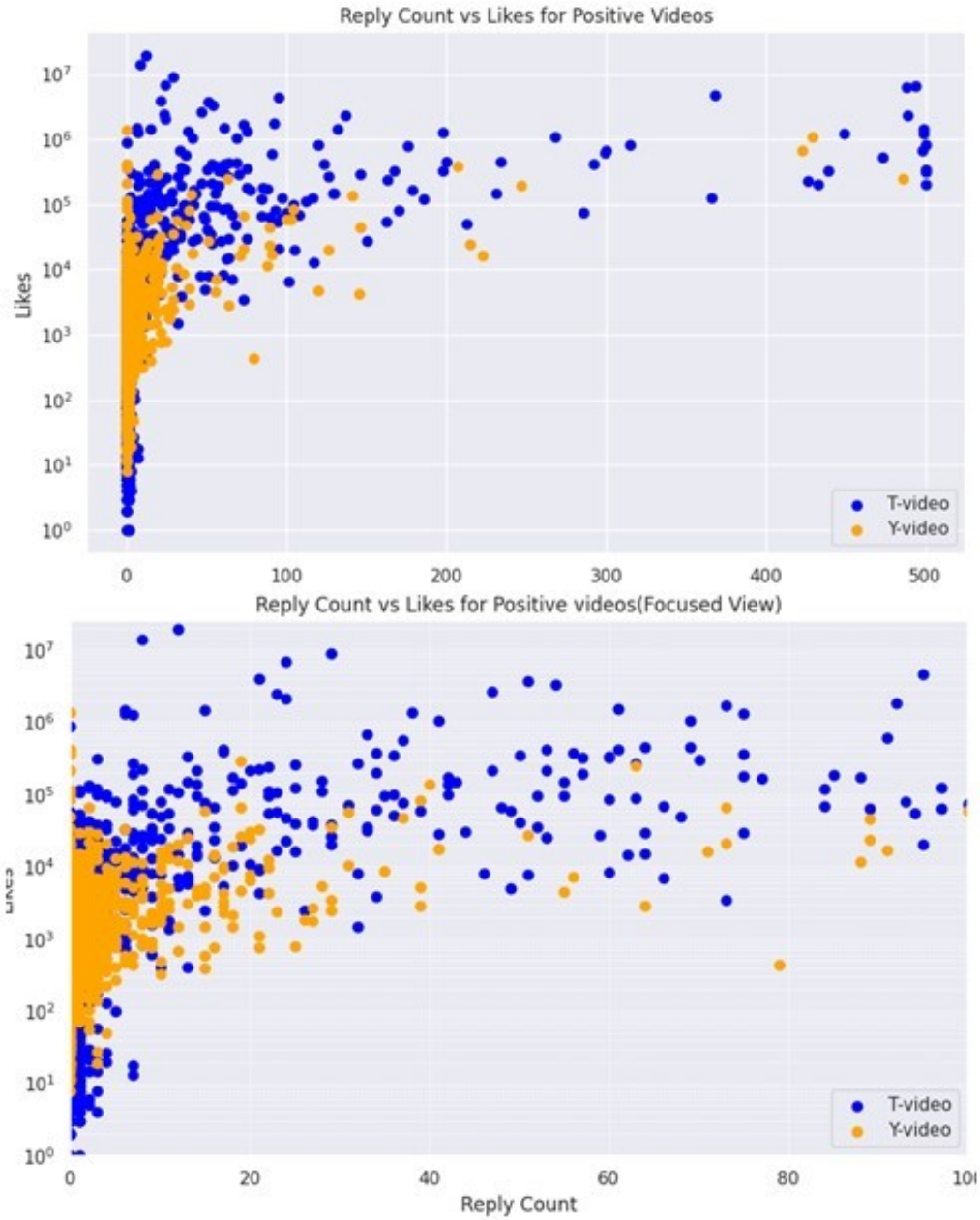


Figure 7.1: The Reply Count vs Likes (on log10 scale) on positive videos: T-video (blue) and Y-video (orange)

(2) The Relationship between the Number of Replies and Likes in Negative Videos of T-video and Y-video

The upper panel of Figure 7.2 compares the number of replies and likes for negative videos in T-video (blue dots) and Y-video (orange dots), with correlation coefficients of 0.11 and -0.02, respectively. The correlation coefficient of 0.11 may indicate that there is a very weak positive correlation between the number of replies and the number of likes. This means that there is a slight tendency for the number of likes to increase as the number of replies increases in the T-video species. The coefficient of -0.02 is close to zero, indicating that there is no significant correlation between the number of replies and the number of likes for Y-video. In other words, the number of replies has no significant effect on the number of likes. These coefficients suggest that users have different dynamics of engagement with negative videos on T-video and Y-video. Y-video shows no significant correlation between replies and likes, suggesting that the two behaviours are independent.

The data is widely distributed across both the number of replies and the number of likes dimensions, especially with the number of likes ranging from 102 to 107. This broad distribution suggests that even negative videos are likely to receive a large number of likes, reflecting the fact that they may touch on points of resonance or topics of widespread interest. The fact that some negative videos received up to millions of likes despite a low number of replies may indicate that they contained strong emotion or sharp criticisms of an issue that elicited empathy from a large number of viewers. It can be observed in this figure that even though the number of replies is between 0 and 100, some videos still receive hundreds of thousands or even millions of likes. This is also supported by the figure below in Figure 7.2, which is a zoomed-in view of the data in this part of the region. As can be seen from the figure, T-video has more dots in the high number of likes area, which may indicate that negative videos on its platform are more likely to receive a higher level of user attention and reaction. Y-video has more densely populated data points in the low to medium number of likes region, which suggests that it has more negative videos but generally fewer likes.

For the few data points with more than 300 responses, the number of likes remains consistently high, suggesting that these videos may relate to important or highly controversial topics. Overall, T-video's negative videos have a denser distribution of points in the low to medium response count range, suggesting that its platform may have strong user activity and engagement. Y-video's negative videos are also denser in the low point count area, but in the very high point count range, its videos are fewer in number, possibly suggesting that its platform varies in its proliferation of negative content or its users' reactions to it.

From Figures 7.1 and 7.2, it can be seen that the number of positive videos in the two datasets is much larger than the number of negative videos in each dataset. That is, the number of positive_T-video is 1,370, the number of positive_Y-video is 1,714; while the number of negative_T-video is 31, the number of negative_Y-video is 146. This study, therefore, randomly chooses 10% of the number of positive videos from each dataset when making a comparison between the positive and the negative videos in the next section.

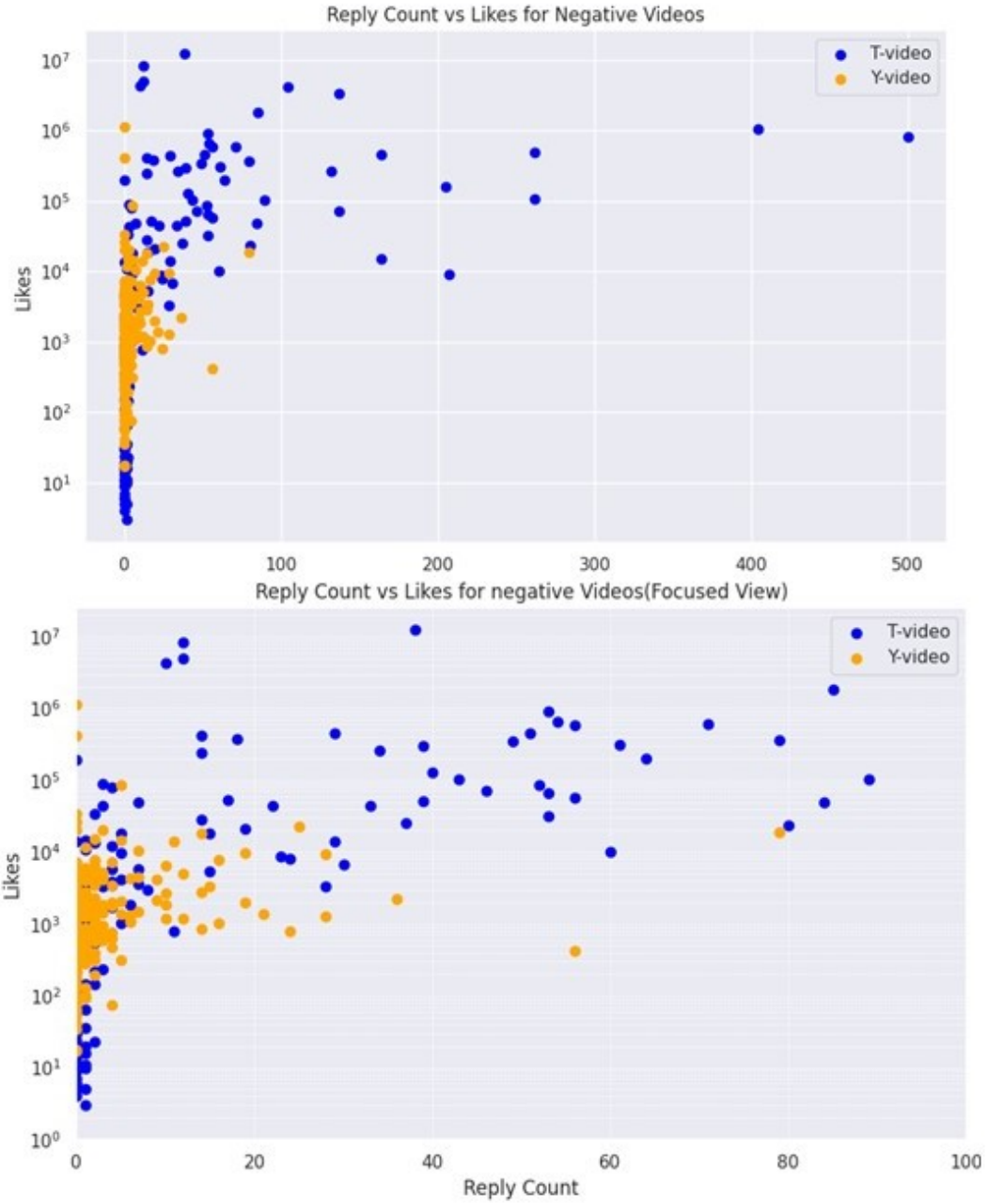


Figure 7.2: The Reply Count vs Likes (on log10 scale) on negative videos: T-video (blue) and Y-video (orange)

(3) The Relationship of Replies and Likes between 10% Positive and All Negative T-video.

Figure 7.3 presents two graphs comparing the number of replies and likes for 10% of randomly selected positive videos and all negative videos on the T-video platform. The top graph shows the overall distribution, while the bottom graph zooms in on videos with fewer than 100 replies for closer examination. The results show that negative videos tend to receive more likes and replies, particularly in the low-reply range. These videos are more concentrated in high-like areas, suggesting that negative content may attract attention due to its association with sensitive or widely discussed topics. In contrast, although positive videos are fewer, they tend to receive relatively higher likes when replies are limited, indicating potentially higher-quality engagement.

Within the range of fewer than 100 replies, both positive and negative videos exhibit a dense distribution of likes, with negative videos notably receiving 105 likes or more. While positive videos generally accumulate fewer likes, some still achieve relatively high like counts (in the 10^3 to 10^4 range), suggesting that even limited positive engagement may resonate strongly with users. Negative videos tend to attract more likes in low-reply regions, possibly reflecting strong audience reactions to controversial or widely relevant content.

The right-hand graph provides a zoomed-in view (0 - 100 replies; 0–200,000 likes), where most T-video entries cluster. In this localised view, fewer blue points (positive videos) are visible compared to red (negative), which may be due to the smaller proportion of sampled positive videos or the lower representation of positive content within low-reply intervals.

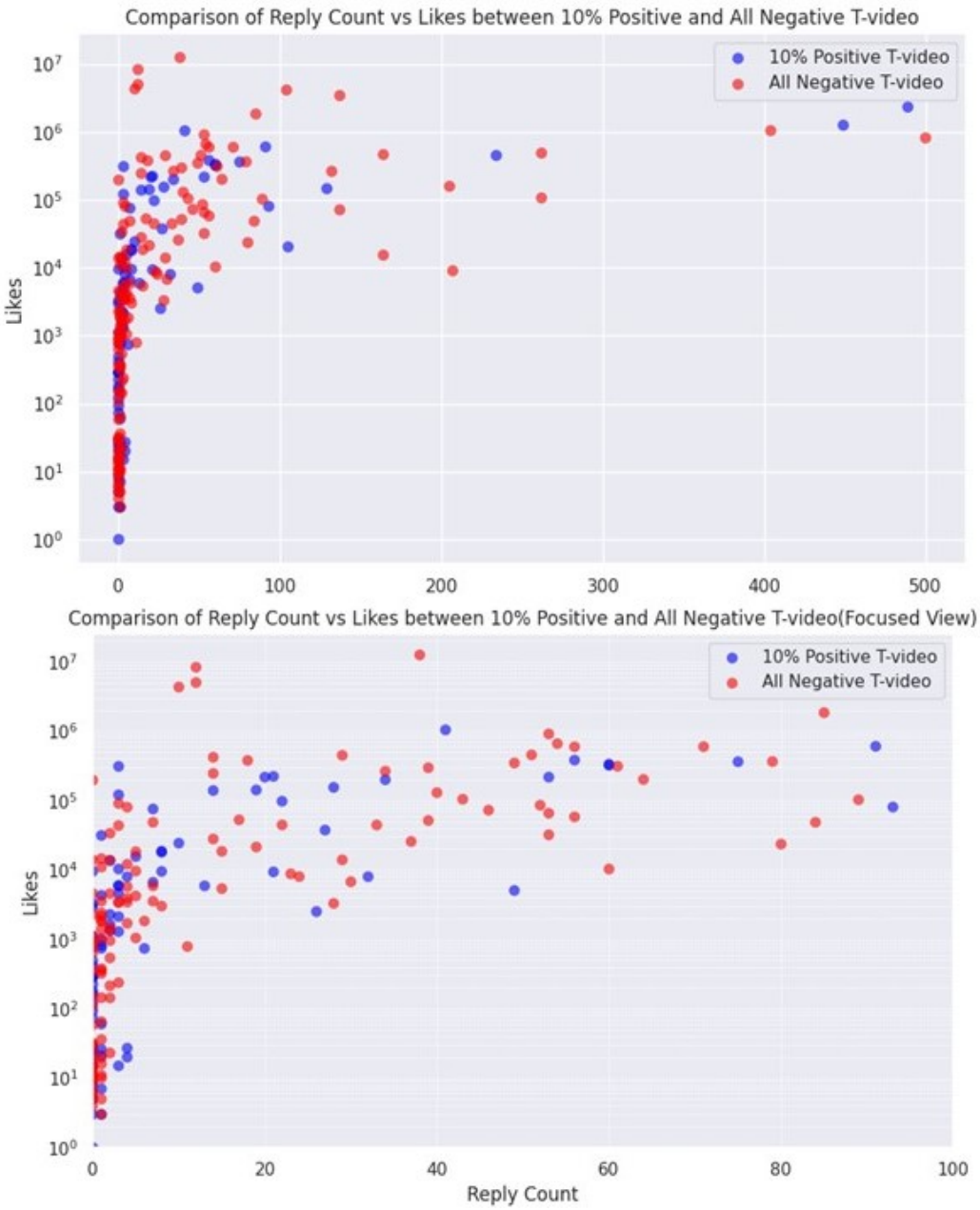


Figure 7.3: Comparison of Reply Count vs Likes between 10% positive videos (blue) and all negative videos (red) of T-video

(4) The Relationship of Replies and Likes between 10% Positive Videos and All Negative Videos in Y-video

The two graphs in Figure 7.4 show the number of replies versus the number of likes between 10% of randomly selected positive videos and all negative videos on the Y-video platform, respectively. The right figure provides a global view, while the left figure focuses on data with response counts below 100. Negative videos generally outperform positive videos in terms of number of likes and replies, especially in the higher range of likes. Negative videos are concentrated in the higher likes range, while positive videos, although fewer in number, also show some concentration in the lower likes range.

Negative videos are concentrated in the lower range of likes (between 103 and 105) in areas with less than 100 replies, but there are also videos that stand out with higher likes, reflecting the fact that negative videos can trigger stronger user responses even with few replies. Positive videos, while generally lower in likes, still had a few videos that reached higher likes, which may indicate that certain positive videos have stronger resonance or appeal.

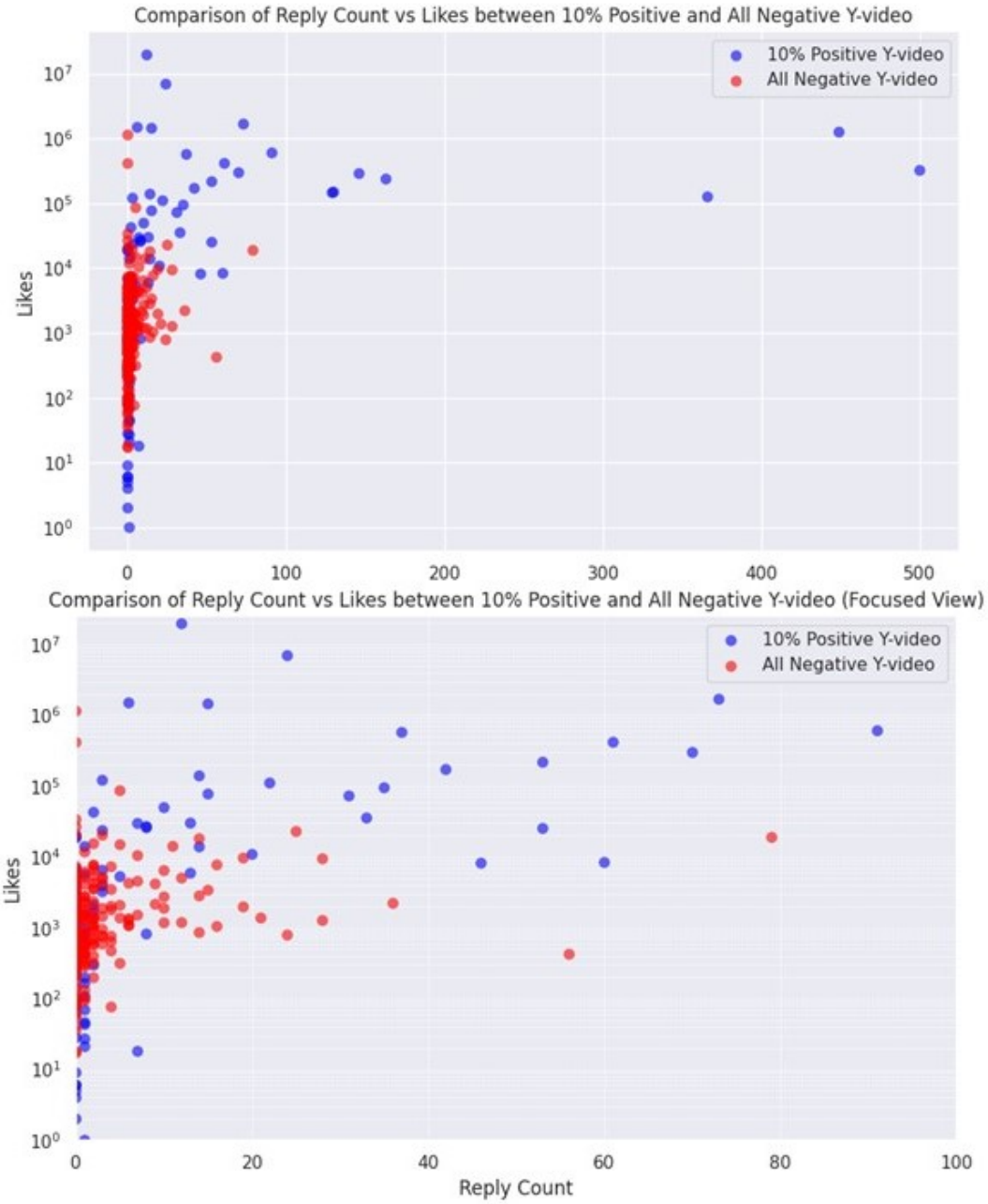


Figure 7.4: Comparison of Reply Count vs Likes between 10% positive videos (blue) and all negative videos (red) of Y-video

7.2.2 The Standard Deviation and Entropy of Sentiment Scores

This section of the study uses three metrics, namely standard deviation and entropy, to analyse the sentiment of the comments under different sentiment videos (positive and negative) within T-video and Y-video for comparison. In sentiment analysis, standard deviation (SD) measures the variability of sentiment responses within the same category, reflecting the degree of consistency in the intensity of the sentiment responses, while entropy measures the complexity of the distribution between the different sentiment categories, reflecting the complexity and diversity of the sentiment responses.

This section presents the distribution of the standard deviation and the entropy of the sentiment scores of the videos within T-video and Y-video. Figures 7.5 and 7.6 illustrate the sentiment diversity in the T-video and Y-video datasets, using Standard Deviation and Entropy as measures, respectively.

(1) The Standard Deviation of T-video and Y-video

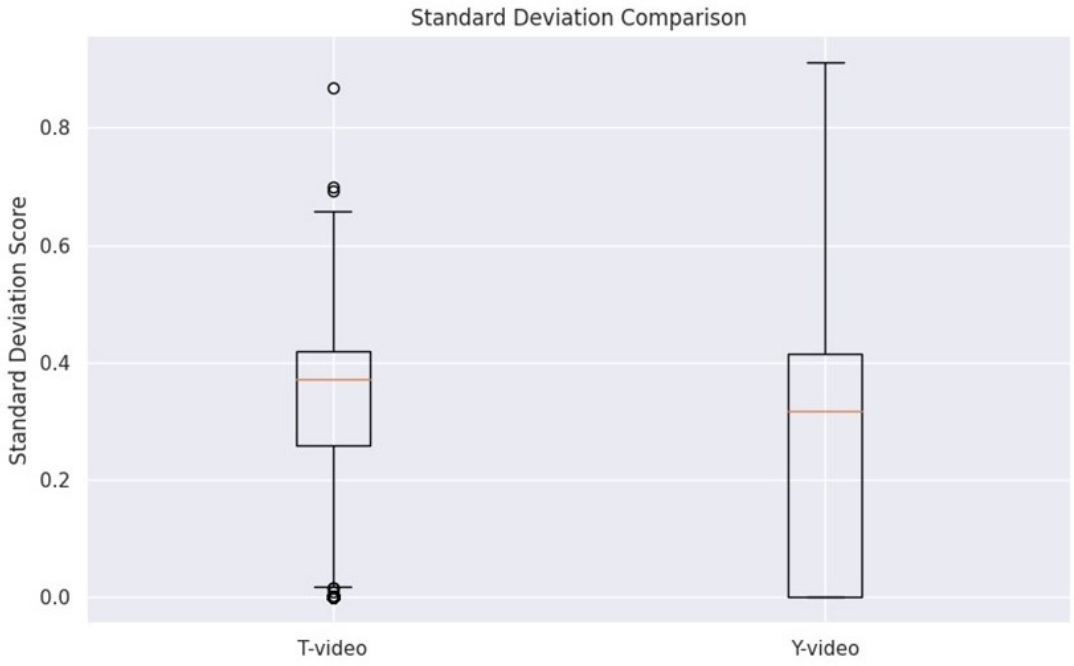


Figure 7.5: Comparison of SD of T-video and Y-video

In examining sentiment scores, a large standard deviation implies more varied and extreme emotions (either very positive or very negative) in the comments, whereas a

small standard deviation indicates more uniform and stable emotional responses. In Figure 7.5, the median for T-video is around 0.37, and the interquartile range (IQR) ranges from 0.25 to 0.42. This box plot illustrates the standard deviation of sentiment for each video. A smaller standard deviation indicates that viewers' emotional responses are more tightly clustered, while a larger one signifies more divergent reactions. The relatively narrow IQR suggests that, for most videos, sentiment variability remains modest, though outliers reveal a few videos with notably broader emotional swings. It is important to note that this consistency in sentiment does not necessarily imply uniformity of video content; rather, it indicates that the emotional responses among viewers for most videos are relatively stable. Further analyses, such as textual or topic modelling, would be required to confirm any thematic uniformity.

The median for the Y-video is around 0.32, slightly lower than that of the T-video dataset. The IQR ranges from 0 to 0.41, with the upper whiskers extending to approximately 0.9 and no apparent outliers. This distribution suggests greater variability in viewer sentiment responses across Y-video content, in contrast to the more clustered responses observed in T-video.

To statistically validate these differences, an independent t-test was conducted ($n_{\mathrm{T}} = 1538$, $n_{\mathrm{Y}} = 2119$), revealing a significant difference in standard deviation scores ($t = 7.65$, $p < 0.001$). The T-video dataset exhibited a higher mean standard deviation (0.314 ± 0.157) than the Y-video dataset (0.266 ± 0.209), confirming the higher emotional variability in the former.

(2) The Entropy of T-video and Y-video

In Figure 7.6, the median entropy of the T-video dataset is higher, the IQR is wider, and the range of entropy distribution is larger, indicating that the distribution of emotions in the T-video dataset is more complex and diversified. The median entropy of the Y-video dataset is lower, the IQR is narrower, and the range of entropy distribution is relatively smaller, indicating that the distribution of emotions in the Y-video dataset is relatively simpler and consistent. There are no outliers in either dataset.

An independent samples t-test further confirmed that the difference in entropy between T-video ($n = 1538$) and Y-video ($n = 2119$) is statistically significant ($t =$

$20.52, p < 0.001$), with T-video showing substantially higher entropy (2.271 ± 1.385) compared to Y-video (1.380 ± 1.226). This statistical evidence supports the observation that the higher entropy and greater variability in the T-video dataset may reflect a greater diversity of video content and more complex emotional responses, while the lower entropy and lower variability in the Y-video dataset may indicate more consistent video content and more uniform emotional responses.

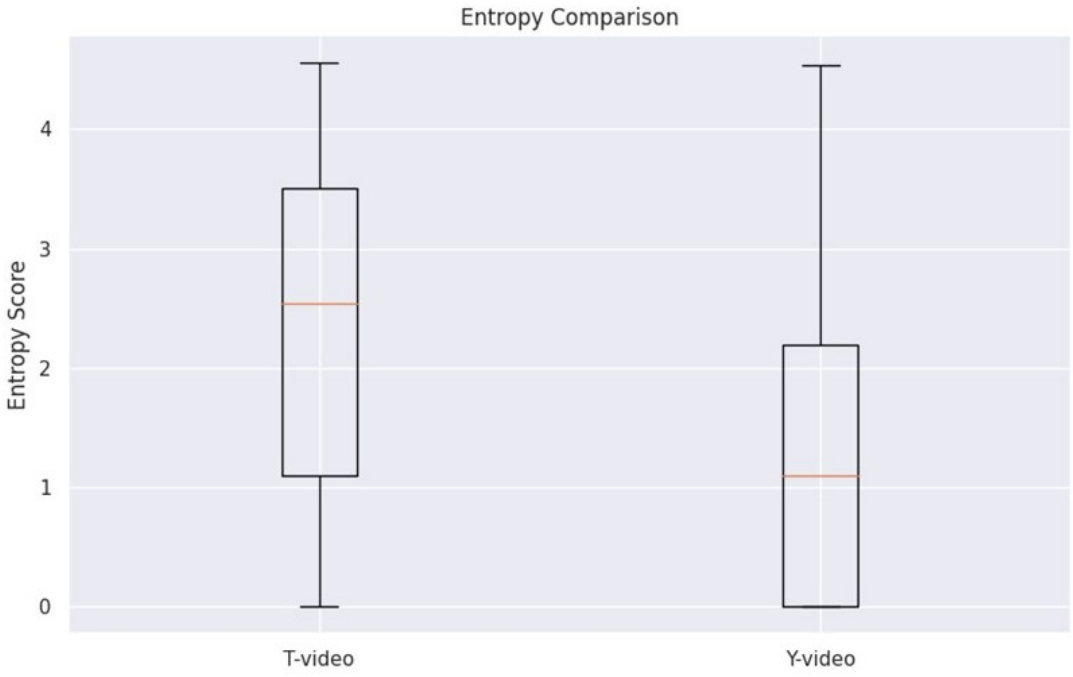


Figure 7.6: Comparison of entropy of T-video and Y-video

In analysing the sentiment scores of the two video datasets, T-video and Y-video, this study used standard deviation (SD) and entropy as metrics. Standard deviation measures the degree of dispersion of the data, i.e., the variability of sentiment scores, while entropy measures the degree of uncertainty or diversity in sentiment categories, i.e., the complexity of sentiment. These two metrics capture different aspects of emotional responses: a dataset can have high SD (indicating large variation in emotional intensity) but low entropy (indicating that sentiment categories are concentrated in a few types), or vice versa.

The standard deviation analysis shows that T-video has a relatively low SD, in-

dicating more consistent affective responses across videos, though a few videos elicit extreme reactions. Y-video, by contrast, shows a higher SD, meaning that emotional intensity varies more widely between videos.

Entropy analysis, however, reveals the opposite trend in complexity: T-video has higher entropy, reflecting a broader and more diverse mix of emotional categories in viewer responses. Y-video has lower entropy, suggesting that despite its higher variability in intensity, the types of emotions expressed are more uniform and concentrated.

In conclusion, T-video exhibits relatively high emotional complexity (high entropy) alongside overall response consistency (low SD), implying that while its videos evoke a wide range of emotional types, the strength of these emotions remains relatively stable for most videos, with only a few extreme outliers. Y-video, on the other hand, shows low emotional complexity (low entropy) but high variability in emotional intensity (high SD), suggesting that while the emotional categories remain simple and uniform, the strength of these responses varies greatly from video to video.

7.3 Sentiment Diversity by Category in T-video and Y-video

In this section, this study presents the sentiment standard deviation, and sentiment entropy value of each video under the main 4 different categories(Music, Entertainment, People& Blog, and Education) in the two datasets, with a focus on comparing and analyzing the differences in the sentiment embodied in music video comments under different categories in the two datasets, as well as analysing how the different categories of videos affect the viewer's affective response. Although the earlier sections did not analyse video categories in detail, we introduce a category-level analysis here to investigate whether different types of music videos elicit distinct emotional responses. Grouping videos into four main categories, Music, Entertainment, People & Blogs, and Education, allows us to compare and contrast sentiment patterns across diverse content. By examining standard deviation and entropy within each category, we can identify whether certain categories tend to evoke more varied or more consistent

emotional reactions, thereby offering additional insights into how video genre influences viewer affect.

7.3.1 The Sentiment Score of Standard Deviation by Categories

Figure 7.7 demonstrates that the T-video and Y-video datasets provide the standard deviation of sentiment scores based on different video categories.

In the Music category, both T-video and Y-video show a median standard deviation of about 0.4, indicating a moderate level of emotional response dispersion. Large box sizes in the plots suggest a broad range of emotional responses across videos on both platforms. However, there are outliers present, indicating that some videos elicit exceptionally strong emotional reactions, although the description seems contradictory about the presence of outliers in Y-video, initially stating their existence and then denying any extreme emotional responses.

For the Education category, the analysis shows more consistency in emotional responses in T-video, with a median standard deviation close to 0.4. The smaller boxes in the T-video indicate that most videos have similar levels of emotional engagement, characterised by low dispersion and no significant outliers, which underscores a uniformity in viewer responses. Conversely, Y-video displays larger bins in this category, suggesting a wider range of emotional reactions among viewers. Despite also having a median standard deviation close to 0.4, the broader range between the upper and lower quartiles points to a higher diversity in emotional responses, though no significant outliers were noted. In summary, T-video tends to show more consistent and less varied emotional responses in the Education category, suggesting that the videos in this category are relatively uniform in how they impact viewers. On the other hand, Y-video exhibits greater variability in emotional responses in the same category, indicating a broader spectrum of viewer reactions to educational content. This diversity could be indicative of varied content styles or differing audience engagement strategies on Y-video compared to T-video.

In the People & Blogs category on T-video, the median standard deviation is about 0.3. The large box size in the plots indicates a high dispersion of emotional responses,

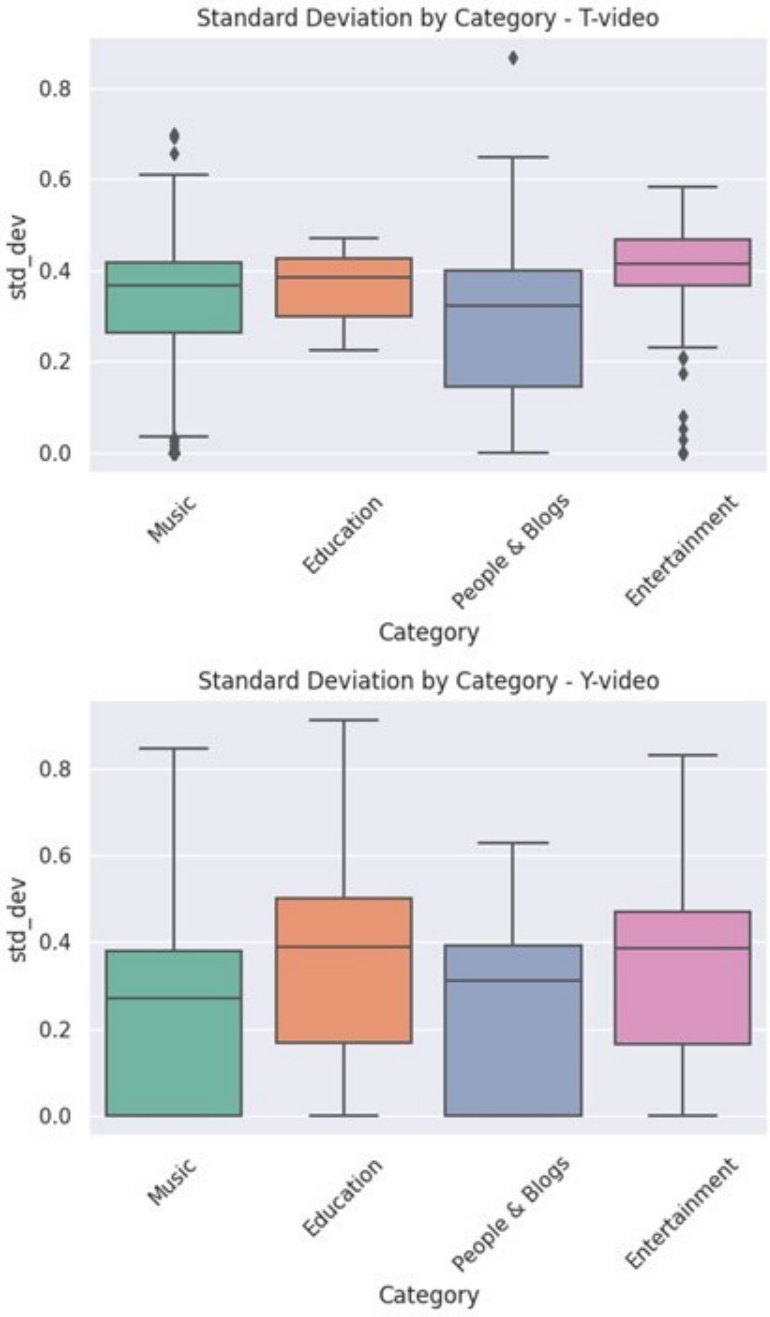


Figure 7.7: The Standard Deviation of sentiment scores by Categories of T-video and Y-video

showing that videos in this category exhibit a broad range of viewer emotions. The significant distance between the upper and lower quartile points to a diverse array of emotional responses. Additionally, the presence of outliers suggests that some videos provoke especially strong emotional reactions. For Y-video in the same category, the median standard deviation is also around 0.3. Despite a similarly large box indicating significant dispersion, the range between the upper and lower quartiles is narrower, which suggests more consistent emotional responses among videos. There are no significant outliers, indicating that emotional responses generally stay within a predictable range.

In the Entertainment category on T-video, the median standard deviation exceeds 0.4. The smaller box size here suggests less dispersion and more consistency in emotional responses across most videos. The smaller range between quartiles underscores the concentration of similar emotional responses. There are some outliers, highlighting that a few videos generate distinctly different emotional reactions compared to the majority. In the Entertainment category on Y-video, the median standard deviation is around 0.4. The larger bins indicate a higher degree of emotional response dispersion, with a broader range between the upper and lower quartiles showing a greater diversity in viewer emotions. The absence of significant outliers indicates that while emotional responses vary, they do not often reach extreme levels. These observations highlight that both platforms exhibit variability in emotional responses; T-video tends to show more uniform emotional reactions in the Entertainment category, while Y-video demonstrates more consistency in the People & Blogs category. Understanding these patterns is crucial for content creators and platform strategies aiming to enhance audience engagement.

7.3.2 Sentiment Entropy Score by Categories

Figure 7.8 illustrates the distribution of category-based sentiment entropy values for T-video and Y-video.

1.The Sentiment Score of Standard Deviation by Categories

Figure 7.8 demonstrates that the T-video and Y-video datasets provide the standard

deviation of sentiment scores based on different video categories.

In the Music category, the median entropy for T-video is approximately 3, with a wide interquartile range. This suggests a high diversity of emotional responses among the videos. The larger box size reflects the broad spectrum of sentiments elicited, while several outliers indicate that a few videos triggered responses markedly different from the rest. In contrast, the Y-video dataset shows a slightly lower median entropy of around 2.5. Although fewer outliers are observed, the box size and interquartile range remain comparable to those of T-video, implying that music videos on both platforms generate diverse emotional reactions, with T-video exhibiting slightly greater variability.

A similar pattern is observed in the Education category. On T-video, the median entropy exceeds 3, accompanied by a large interquartile range. This reflects considerable variation in user sentiment, with most emotional responses falling within a broad yet expected range, evidenced by the absence of notable outliers. On Y-video, although the median entropy is lower at approximately 2, the overall spread of values remains wide. The comparable interquartile range again indicates substantial diversity in emotional responses, albeit with less intensity than that seen on T-video.

Together, these findings demonstrate that educational and music content on both platforms elicits a wide range of emotional engagement. However, T-video consistently exhibits slightly higher median entropy, suggesting that it fosters more varied emotional responses. This highlights platform-specific differences in how users interact with and emotionally react to content, particularly in categories associated with strong personal or thematic resonance.

In the People & Blogs category, the T-video dataset shows a median entropy of approximately 2, with the upper quartile nearing 3 and the lower quartile close to 1. This wide interquartile range, reflected in the relatively large box size, indicates a high degree of variation in emotional responses across videos. In contrast, Y-video presents a slightly lower median entropy of about 1.5, with a somewhat narrower box, suggesting a modestly reduced emotional diversity compared to T-video. Notably, both datasets include outliers with entropy values approaching 4 and close to 0, pointing to videos that elicit highly distinctive emotional reactions.

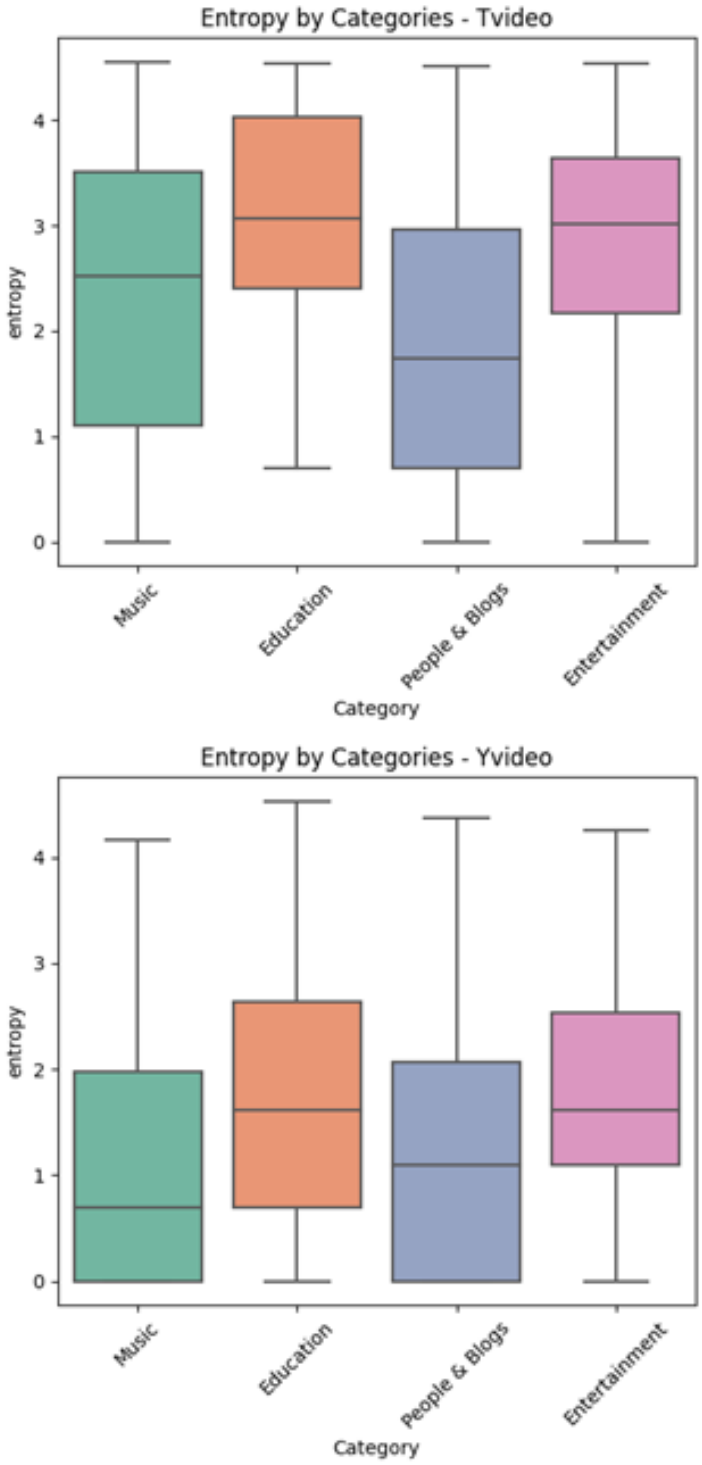


Figure 7.8: The Sentiment Entropy of sentiment score by Categories of T-video and Y-video

In the Entertainment category, T-video records a median entropy of around 3, accompanied by a more concentrated distribution. This indicates consistently high emotional diversity across entertainment videos, though slightly lower than that observed in the Education category and higher than in People & Blogs. For Y-video, the median entropy in this category is about 2, reflecting moderate emotional variation. Compared to other categories within the Y-video dataset, the Entertainment category demonstrates greater emotional diversity than Music and People & Blogs, but slightly less than Education.

Taken together, these findings suggest that both T-video and Y-video exhibit substantial emotional variation across content categories, with notable differences in the extent of diversity. This analysis offers insight into how various content types elicit emotional engagement on each platform and underscores the importance of content categories in shaping user responses.

7.4 Characterising Topics Present in Comments on T-video and Y-video

In this section, the results were presented and analysed, which were obtained during an exploration of the coherence and perplexity scores under differing assumptions around the number of topics present in the T-video and Y-video datasets, using the Latent Dirichlet Allocation (LDA) approach. Based on the principle that the lower the perplexity, the better the model interprets the data, while a high coherence score indicates a greater semantic coherence of the words in the topic, the topic counts will be finalised for both datasets, which prepares us for the next step of topic extraction.

7.4.1 Identification of the Number of Topics

Figures 7.9 and 7.10 illustrate the impact of parameter adjustments on the performance of the topic model for the T-video and Y-video datasets, respectively, focusing on perplexity and coherence scores. For this analysis, this study configured the model with a range of possible topics [Num_Topics = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50] and

model iterations [passes_list = 10, 20, 30, 40, 50, 60, 70, 80], to explore how different settings influence the Latent Dirichlet Allocation (LDA) topic model's effectiveness. This systematic tuning helps in selecting suitable model parameters.

Additionally, the visual analysis facilitated by these figures allows for adjustments in the topic model, enhancing its interpretability and improving the quality of the topics generated for these specific text datasets. Specifically, visualisation helps to identify issues such as overlapping topics, incoherent keywords, or redundant themes. For example, if certain topics have a high level of overlap, this could indicate that the number of topics chosen is too large, and they need to be reduced or merged. Alternatively, if the key terms in a topic do not make coherent sense, the model's parameters can be adjusted (e.g., increasing the number of iterations) to improve the coherence. These insights from the visual analysis directly inform the iterative adjustment of the LDA model, ensuring the topics generated are both distinct and meaningful, ultimately enhancing their interpretability.

Figure 7.9 highlights how the perplexity score in the T-video dataset decreases as the number of topics increases, indicating a better model, with the poorest perplexity score being observed at 5 topics. The perplexity metric only varies slightly across different numbers of iterations. Similarly, coherence scores show a noticeable fluctuation with an increasing number of iterations. However, after the topic count reaches 30 (combined with Figure 7.11), the coherence scores tend to stabilise, indicating that adding more topics beyond this point does not significantly improve the coherence of the model. This configuration seems to best enable the model to capture meaningful topics in the data, suggesting that refining the focus by reducing the number of topics and a sufficient number of iterations can improve the performance of the topic model.

Figure 7.10 depicts the trend in perplexity for Y-video, showing that perplexity generally decreases as the number of topics increases, reaching the lowest level of perplexity at a topic count of 50. This trend suggests that the model explains the data better and reduces the perplexity when more topics are used. In addition, the increase in the number of iterations seems to enhance the reduction in perplexity, especially at 70 or 80 iterations, where the perplexity for the number of topics in the Y-video is

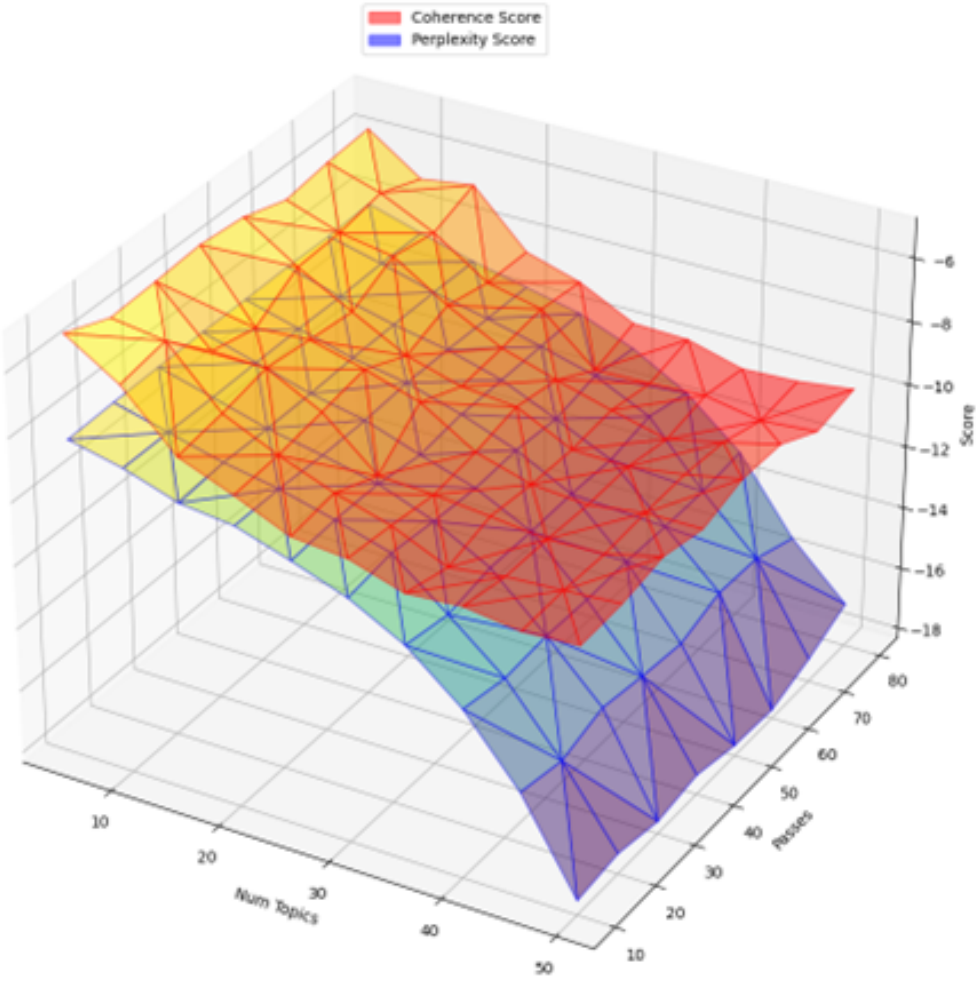


Figure 7.9: The perplexity and coherence score of T-video

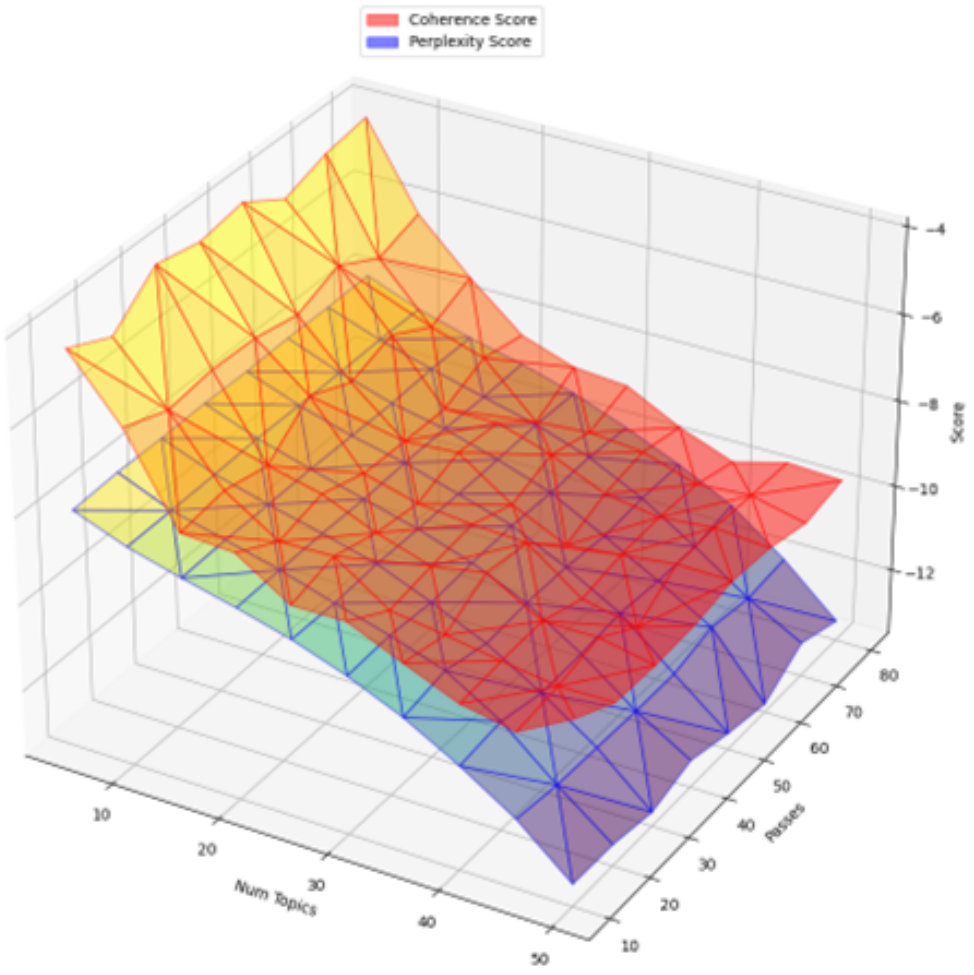


Figure 7.10: The perplexity and coherence score of Y-video

closer to the perplexity mean value for that number of topics. Coherence scores also indicate the best results at a topic count of 5, decreasing as the topic count increases, suggesting that fewer topics were able to maintain stronger correlation and consistency among the topics generated. Although the effect of the number of iterations on the consistency is not significant compared to the perplexity, it can be observed that the coherence score peaks at a topic count of 5 and 70 iterations.

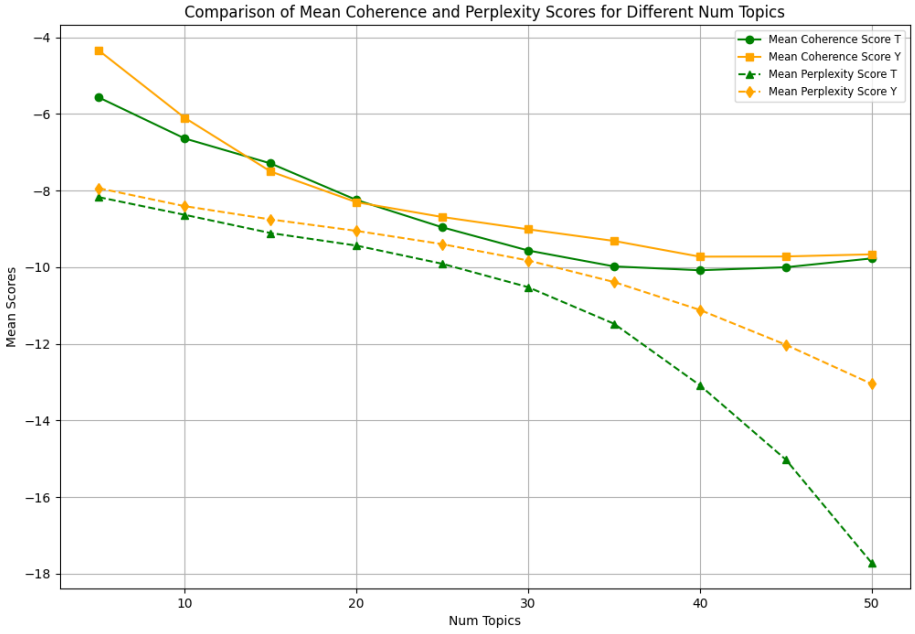


Figure 7.11: Comparison of Mean Coherence and Perplexity Scores for Different Numbers of Topics for both T-video and Y-video datasets

Figure 7.11 shows a comparison of the mean coherence score and perplexity scores for different numbers of topics (Num Topics) for T-video and Y-video. Overall, both the coherence score and perplexity score decrease as the number of topics increases, which may indicate that more topics lead to improved coherence and perplexity of the model.

Specifically, the T-video dataset performs better in terms of perplexity because it decreases more, which usually indicates a decrease in model perplexity and an improvement in model quality. The downward trend of the mean perplexity scores in T-video

(green line) is significantly larger than the downward trend of the mean coherence scores in T-video (green dashed line), especially after topic num = 30, where the gap between the two becomes larger and larger, with perplexity continuing to drop while coherence levels off.

In the analysis of the Y-video and T-video datasets, the model exhibits better performance at a topic count of around 30, indicating that a moderate number of topics is effective in capturing the underlying topic structure in both video datasets. Based on the dual metrics of perplexity and coherence scores, the optimal configuration for both datasets was determined to be 30 topics and 30 iterations.

7.4.2 The Visualisation of LDA Topics

In this section, the layout of the LDA for T-video and Y-video is presented, respectively. In the graphs that follow, a visualisation of the entire LDA topic is presented, with the global topic view on the left and the term bar chart on the right. The graphs provide a concise representation of the relationship between topics and relevant terminology, allowing all key aspects to be easily visualised in a single view.

In the visualisation graphs of LDA topics, each bubble represents a topic (theme), and the size of each bubble indicates the weight or importance of that topic relative to other topics. Specifically, the larger the bubble, the larger the proportion of that topic in the entire text corpus, i.e., the topic covers more documents or words. Thus, the difference in the size of different bubbles reflects the relative distribution of topics in the corpus.

The main components of the LDA visualisation maps, such as the first one shown in Figure 7.12, are: **(1) Left side:** Intertopic Distance Map. This map is generated by Multidimensional Scaling (MDS) and shows the relationship between different topics. Each blue circle in the map represents a topic, and the size of the circle indicates the relative weight or importance of the topic in the overall document set. The distance between the circles indicates the similarity between the topics. The closer the distance, the more similar the content between the two topics; the further the distance, the more different their content. **(2) Right Side:** Histogram of Top-30 Most Relevant Terms.

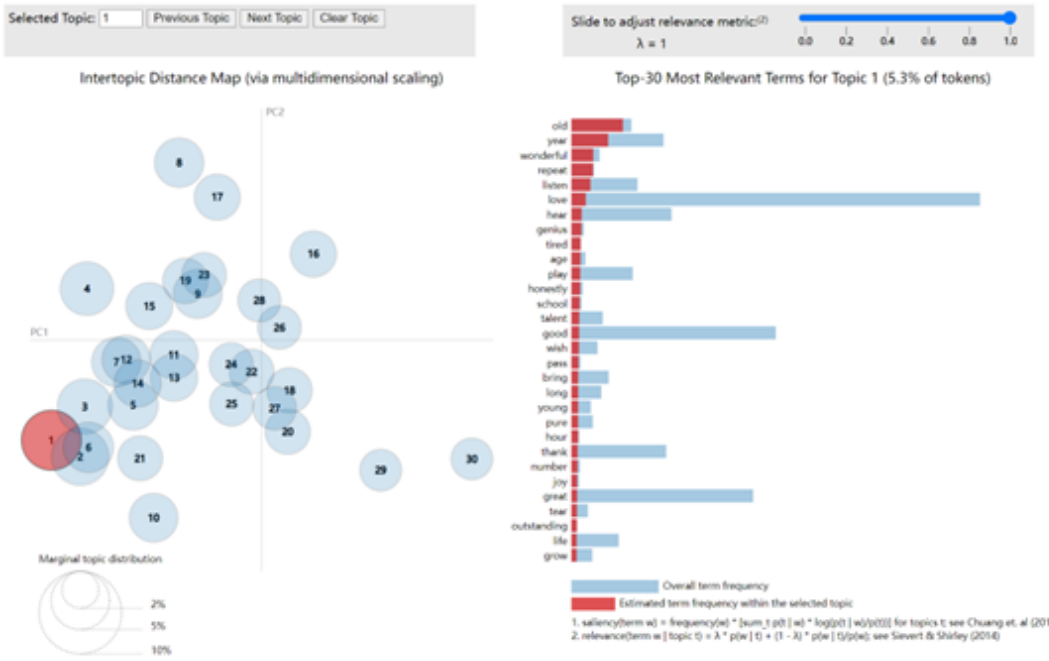


Figure 7.12: The LDA Topics Visualization of T-video (Topic Num = 30, passes = 30)

The bar chart on the right shows the 30 most salient terms for the selected topic (in this case, shown for Topic 1). Each word has two bars, red and blue. The blue bar indicates the number of times the word appears in the entire document set, while the red bar indicates the relative importance of the word in the currently selected topic. **(3) Slider for adjusting relevance metric (λ):** In the upper right corner of the graph, there is a slider for adjusting the value of λ. λ ranges from 0 to 1, with a default value of 1. When λ is 1, the relevance is based on the frequency of the word in the topic only, and thus important words specific to the topic are shown. When λ is 0, the visualisation shows the words in the topic that are most useful in distinguishing it from other topics. This value allows you to see which words are more distinguishable among topics.

Figure 7.12 shows a visualisation of the LDA topic model for the comments linked to the T-video dataset. There are a number of topics in T-video that are clearly independent and do not intersect with any other topic. In particular, Topic 4, Topic 8, Topic 10, Topic 16, Topic 17, Topic 29 and Topic 30. There are also a number of topics

that overlap with each other, often to a significant degree, such as Topics 1, 2, 3 and 6, indicating that there is a high degree of similarity and shared content between these topics. Among them, the almost complete overlaps between Topics 2 and 6, or between Topics 7 and 12, indicate that they contain very similar high-frequency vocabulary and semantics. These topics may express almost the same content, which may require further optimisation of the topic model to merge these highly overlapping topics into one. There are also topics that are connected through third parties. There are topics that do not directly intersect but are linked through other topics (e.g., between Topic 1, Topic 3, and Topic 5). This relationship can be understood in the sense that although there is little lexical similarity between Topics 1 and 5, they both have some overlap with Topic 3, which may play a role in connecting Topic 5 and 7. It can be seen as an indirect similarity, where the topics may have different directions of discussion, but are linked to certain key points.

The phenomenon of topic overlaps or partial crossover that occurs during the selection process of the LDA topic model can be explained by the following aspects:

1. The essential overlap of topic distribution

Lexical sharing: In a real corpus, certain keywords may appear frequently in several different contexts or topics. For example, if the corpus involves music reviews, words such as "love" and "amazing" may appear simultaneously in reviews describing music styles, concert experiences, or specific songs. This leads to keyword overlapping across topics.

Semantic ambiguity: Topic models rely on algorithms to extract statistical patterns from textual data, but these patterns are not always able to perfectly distinguish between all semantically close concepts, especially if the keywords are important in multiple topics.

(2) Model parameters and data characteristics

Number of topics selected: Although the number of topics is selected based on confusion and consistency scores, this number may still be insufficient to fully resolve semantic overlaps between topics. Sometimes, even if the metrics indicate that a certain number is optimal, some topics may not be sufficiently dispersed in real applications

due to the specific nature of the corpus.

Training iterations of the model: The number of iterations and training details (e.g., learning rate, prior distribution) of the LDA model also affect the clarity and separation of the final topics. An insufficient number of iterations may result in the model failing to adequately learn the complex structure in the data.

(3) Data inhomogeneity

Document length and quality: Varying length and detail of documents may cause the model to favour documents that are richer or more consistently worded, thus affecting topic differentiation.

Word frequency and document frequency: High-frequency words may dominate across multiple topics, especially if word frequencies are not adequately adjusted using TF-IDF or other normalisation techniques.

While LDA has known limitations in modelling short-text data due to its assumptions of topic mixture and word co-occurrence, it was used in this study for several reasons. Firstly, prior sentiment analysis using VADER helped filter and cluster comments based on emotional polarity, thus enabling LDA to be applied to thematically meaningful subsets of comments. Secondly, LDA remains one of the most interpretable and widely used methods for unsupervised topic modelling, which makes it suitable for identifying broad themes within emotionally charged user responses. Lastly, its use enables comparison with findings from existing studies that have adopted similar methods in social media analysis.

Topic 4	tune, watch, movie, job, rip, great, summer, sound, love, good
Topic 8	esta_cancion, deserve, lovely, sweet, niall, gem, magical, game, beatle, solo
Topic 10	playlist, real, shit, style, mood, sick, add, todo, film, rap
Topic 16	miss, childhood, man, post, life, beautifully, person, love, gold, thank
Topic 17	damn, queen, hard, twice, abba, seriously, good, harmony, write, youth
Topic 21	proud, girl, world, love, baby, rock, dream, generation, hold, cute
Topic 29	love, wait, lt, guy, bro, absolutely, sooo, god_bless, album, epic
Topic 30	hit, voice, thank, gorgeous, perfection, love, angel, lady, different, record

Table 7.1: The top 10 most weighted keywords in independent topics of T-video

Based on T-video's LDA topic model visualisation (Figure 7.14) and the top 10 most weighted keywords in independent topics (Table 7.1), these keywords represent

the most significant terms after analysing the dataset with the LDA model. The 8 clusters were selected because they demonstrated complete independence from other topics in the LDA visualisation. The visualisation, which plots topics based on their distributional similarity, revealed that while most topics overlap or intersect in the shared topic space, these 8 clusters are positioned as isolated, non-overlapping points. This spatial independence indicates that these clusters represent unique, self-contained topics, with minimal or no semantic overlap with other topics. As such, they were deemed highly interpretable and representative of distinct content, aligning well with the goals of this analysis. For the Y-video dataset, the same presentation method applies. These words are the top 10 most frequently occurring words in each of these 8 separate topics.

As can be seen from the keywords of each topic, most of the independent topics of T-video have positive emotional tendencies, such as 'love', 'proud', 'great', etc. These keywords reflect that viewers generally rate the video content highly. Topics covered include music (e.g. Topic 4, Topic 8, Topic 10), nostalgia and reminiscence (e.g. Topic 16), women and inspiration (e.g. Topic 21), etc., demonstrating the diversity of the content, which touches on different topics such as music, growing up, and inspiration. Keywords in many of the topics (e.g., Topic 8, Topic 29, Topic 30) indicate viewers' passion and support for an artist or a specific piece of work, illustrating that T-video has a high level of fan engagement and that viewers tend to express strong, personal feelings about these works.

Figure 7.13 shows the visualisation results of Y-video's LDA topic model. Topics 1, 2, 3, 4 and 5 are close to each other, which indicates that these topics have more lexical overlaps and may discuss similar content. The bar chart on the right side shows that the high-frequency words in Topics 1, 17, 6, and 9 are obviously located farther away from the other topics, which suggests that the contents of these topics have little overlap with the other topics and belong to relatively independent and special contents. This distribution helps to better understand the characteristics of each topic. The circles for each topic in this graph are not significantly different in size from each other, and many of the topics have some intersections or overlaps, which suggests lexical content

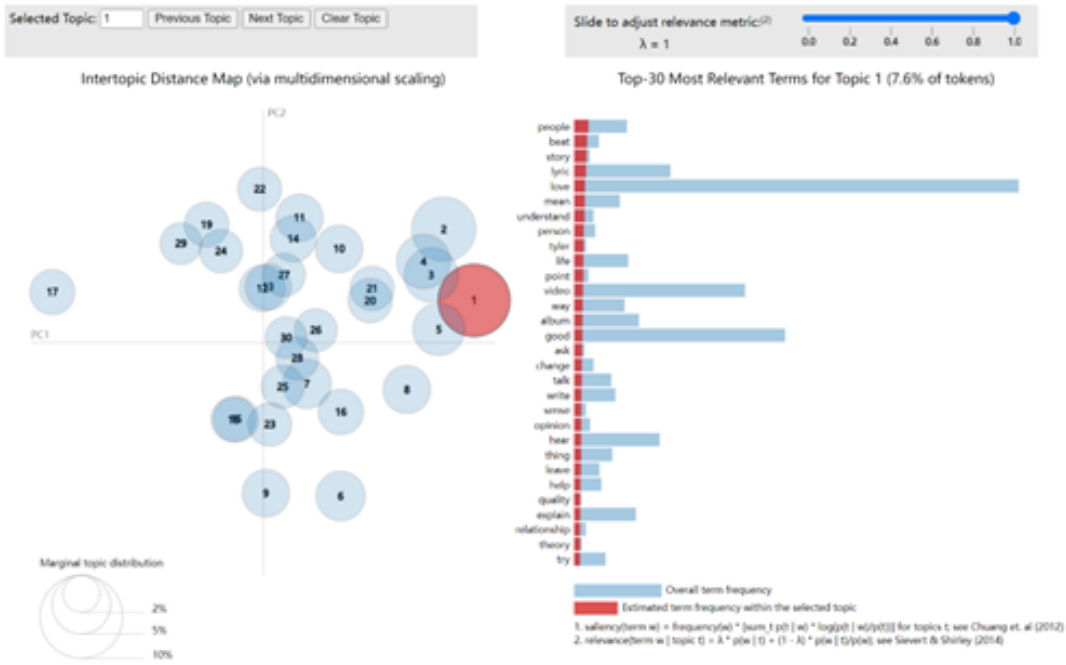


Figure 7.13: he LDA Topics Visualization of Y-video (Topic Num = 30, passes = 30)

overlap among these topics and that the distinctions between the topics are not very clear.

Topic 6	favorite, dude, band, rock, late, walk, begin, summer, mom, album
Topic 8	perfect, cool, hear, incredible, your, bro, fast, kill, type, wish
Topic 9	panic, fall, boy, top, vocalist, disco, friend, generation, good, imagine_dragon
Topic 17	thank, underrate, tutorial, happy, content, help, playlist, great, video, share
Topic 22	amazing, voice, listen, note, ally, copyright, high, list, breakdown, interview

Table 7.2: The top 10 most weighted keywords in independent topics of Y-video

.

Based on Figure 7.13 and the top 10 most representative keywords in each independent topic (Table 7.2), the top 10 keywords in each independent topic are given as the most important words under the topic after analysing the documents according to the LDA model. As can be seen from the keywords, most of the sentiment tendencies in Y-video's independent topics are again positive. For example, in Topic 8 and Topic 17, viewers use words such as 'perfect', 'thank', and 'happy' to express their content highly, suggesting that the content of these videos was generally well-received and appreciated

by viewers. Y-video topics covered music discussions (e.g., Topic 6, Topic 9, Topic 22), tutorial videos (e.g., Topic 17), and generational influences that may relate to musical styles or groups (e.g., Topic 9). The topics reflect the diversity of content, including both music performances and album discussions, as well as tutorial- and help-related content. Keywords such as 'tutorial', 'playlist', and 'help' indicate that viewers are not only interested in entertainment content, but also in educational and learning content, which attracts more participation and engagement from viewers.

There are also some closely related topics in Y-video; for example, topics 12, 13, and 27 are closely intersected in the figure, and these topics contain many comments about performance and emotion, especially discussions of specific performers (e.g., Taylor Swift), dance, and classical performance. There is a high degree of keyword overlap between them, reflecting the fact that these topics collectively explore emotional responses to music and performance, and audience evaluations of these elements. For example, Topic 12 relates to 'dance' and 'legend', whilst Topic 13 and Topic 27 relate to comments on specific characters and the effectiveness of the remix, respectively, which relate to audience emotion and feedback on the performance, making these topics closely linked. These relate to the audience's emotions and feedback on the performance, making these themes closely intertwined.

7.4.3 The LDA Topics Overlap

The visualisations shown so far for the LDA model were based on the number of topics to be 30 and the number of passes to be 30. However, Figures 7.14 and 7.15 indicated that there was a fair degree of overlap between many topics. In this section, further visual comparisons of the datasets T-video and Y-video based on different numbers of topics are performed by setting num_topics = [5, 15] (with the number of passed and remaining at 30)and passes = 30. The main objective is to compare the characteristics of the two datasets in terms of topic distribution, topic differences and coherence based on different numbers of topics.

Combining Figure 7.14, and Tables 7.3 and 7.4, it can be found that when the num-

(a) The LDA Topics Visualization of T-video (Topic Num = 15, passes = 30)

(b) The LDA Topics Visualization of T-video (Topic Num = 5, passes = 30)

Figure 7.14: The LDA Topics Visualization of T-video (Topic Num = [5,15], passes = 30)

Topic	Keywords
Topic 11	que, vibe, deserve, era, por, view, incredible, cancion, proud, saudade
Topic 12	wait, esta_cancion, underrate, iconic, finally, unique, talent, pure, head, word
Topic 13	masterpiece, spotify, damn, rest, peace, jam, increible, light, que, siempre
Topic 14	love, absolutely, brilliant, lovely, voice, fall, guy, god_bless, radio, hear
Topic 15	repeat, video, obsess, bro, dance, lol, niall, abba, genius, disappoint

Table 7.3: Topics and Keywords for `topic_num_T-video = 15`

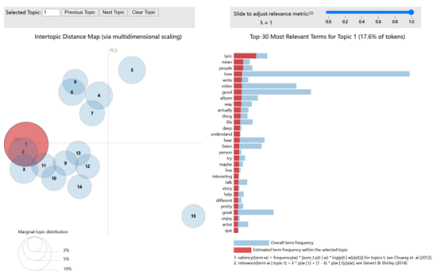
Topic	Keywords
Topic 2	album, favorite, nice, year, hear, listen, love, new, good, play
Topic 3	love, beautiful, voice, awesome, amazing, sound, great, masterpiece, vibe, absolutely
Topic 5	que, old, favourite, era, esta_cancion, por, wait, cancion, gorgeous, nostalgia

Table 7.4: Topics and Keywords for `topic_num_T-video = 5`

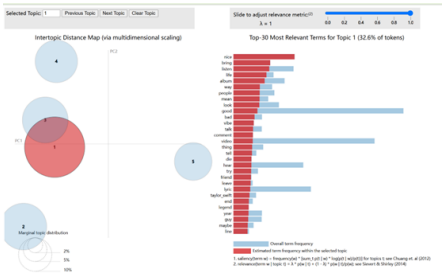
ber of topics is equal to 15, there are some relatively independent topics in the T-video dataset, e.g., Topics 11, 12, 13, 14, and 15, which contain some high-weight keywords, e.g., "queue", "vibe", "masterpiece", etc., indicating that there are significant differences between these topics, which can better distinguish different content types, "vibe", "masterpiece", etc. These topics contain some high-weighted keywords, such as "que", "vibe", "masterpiece", etc., which indicate that there are significant differences between these topics, and they can better distinguish different content types. When the number of topics decreases to 5, the boundaries between topics become blurred. For example, from the keywords in Topic 2 and Topic 3, many keywords (e.g., "love", "hear", "voice")

appeared in multiple topics, suggesting that the overlap of topics has increased, reducing the distinguishability of the model. This implies that while reducing the number of topics can simplify the model, it may also make the different topics less independent.

A similar approach to exploring differing numbers of topics is now illustrated for the comments associated with the Y-video dataset.



(a) The LDA topics visualization of Y-video (Topic Num = 15, passes = 30).



(b) The LDA topics visualization of Y-video (Topic Num = 5, passes = 30)

Figure 7.15: The LDA topics visualization of Y-video (Topic Num = [5,15], passes = 30)

Topic	Keywords
Topic 5	amazing, voice, thank, cool, love, singer, beautiful, lol, good, unique
Topic 15	love, hear, talente, version, sooo, wonderful, incredible, cover, fan, video

Table 7.5: Topics and Keywords for `topic_num_Y-video = 15`

Topic	Keywords
Topic 2	good, amazing, voice, beautiful, sing, vocal, live, singer, cool, dance
Topic 4	video, thank, great, omg, man, finally, sound, explain, hear, work
Topic 5	love, awesome, lol, channel, spotify, light, laugh, vid, early, bro

Table 7.6: Topics and Keywords for `topic_num_Y-video = 5`

Combining Figure 7.15 and Tables 7.5 and 7.6, when the number of topics is high (15), this study can see a relatively large number of independent topics. In the above figure, the topic distance graph shows that some topics are relatively independent and

far away from each other, for example, topic 5. In the case of 15 topics, the top 10 words with the highest weights in each topic show that the topics are more concentrated. For example, the keywords of topic 5, such as 'amazing', 'voice', and 'thank', are all related to music and singing, which indicates that the topics are more specific and the topics are more specific and subdivided. When the number of topics decreases to 5, the most weighted keywords show a tendency to be more generalised. For example, topic 2 contains descriptive words such as 'good', 'amazing', 'voice', etc. These words cover a wider range of topics and lack details, which suggests that when the number of topics decreases, the specificity of the topics decreases, and the content becomes more general.

The distinction between independent topics becomes more evident as the number of topics increases, with the top keywords in each topic highlighting more specific themes. For instance, in the T-video and Y-video datasets, higher topic counts allowed for finer distinctions, such as focusing on specific music genres or emotions. Conversely, as the number of topics decreases, topic boundaries blur, and the content becomes more generalised and overlapping.

7.5 Discussion

7.5.1 Findings

The research in this chapter aims to deepen this study's understanding of YouTube music video user sentiment and discussion topics around videos that were referenced on the different platforms (i.e., YouTube and Twitter). By analysing the sentiment tendencies and topics demonstrated by user comments in the T-video and Y-video datasets, this study aims to highlight the similarities and heterogeneities in user sentiment across different sources and the topics they discuss in their sentiment exchanges.

In the T-video dataset, the quality of the content of users 'comments under both positive and negative videos seems to have an effect on users' likes. In the Y-video dataset, the number of user comments in the positive video dataset has a more significant effect on Likes, indicating that users have a more focused engagement pattern

on the platform and tend to like videos that already have more comments. Users under the Y-video negative video tend to interact more independently, and there is no obvious correlation between liking and replying to behaviours, suggesting that users have a more varied engagement pattern and a relatively lower attitude and reaction to negative comments. They are more varied and have relatively lower attitudes and reactions to negative comments. Specifically, in both T-video and Y-video, negative videos typically generate broader user engagement, while positive videos rely more on the quality of the content to engage users. Negative videos are more topical and interactive on the two platforms, whereas positive video engagements are more focused on individual, contagious comments.

In terms of Entropy and Mode sentiment analysis, T-video's video content is more diverse in style and can trigger a wide range of emotional types, resulting in high emotional complexity and diversity. However, emotional responses to similar videos are more focused and consistent, so most emotional responses are relatively stable, but some videos may trigger extreme emotional fluctuations. The Y-video videos have lower emotional complexity, and viewers' emotional responses to individual videos are usually more consistent with fewer emotional fluctuations, but there is a large variation in emotional responses between videos, suggesting that there is significant variability in emotional responses on the platform.

In terms of different categories, in the Education & Entertainment category, emotional consistency is high, with a relatively uniform style of video content and a concentration of emotional responses, but there is also some content that is capable of provoking extreme emotional responses. In the People & Blogs and Music categories, emotional complexity and diversity are high, with a wide range of emotional responses from viewers, and content that elicits different types of emotional responses, with the potential to appeal to a wide range of viewers. T-video's overall emotional response is neutral and stable, and it is suitable for viewers who are looking for a stable emotional experience.

In the Education and Entertainment category, Y-video shows higher emotional dispersion and greater diversity and variability in emotional responses, suggesting that

its content is rich and varied and capable of triggering a wider range of emotional responses. In the People & Blogs category, emotional responses are more focused, with higher consistency in viewer emotional responses, making it suitable for users who want a consistent viewing experience. Overall, Y-video's emotional responses are more positive and varied, especially in the Entertainment and People & Blog categories, where viewers tend to express positive emotions, which may be related to its content strategy and viewers' emotional preferences.

This study contributes to the broader debate generated by previous cross-platform sentiment analyses of user review content that employ both contemporary and classic sentiment analysis learning methods, such as those of Mejova and Srinivasan (Mejova and Srinivasan, 2012), Shevtsov et al. (Shevtsov et al., 2023), and Bilinski (Bilinski, 2024). Mejova and Srinivasan analyse political discussions on YouTube and Twitter, comparing user behaviour, discourse nature, and content types on these platforms. They find that Twitter features rapid, news-driven updates, while YouTube encourages extended conversations in comment sections, leading to distinct dynamics in political communication on each platform. Shevtsov et al. examine how political content related to the 2020 U.S. elections is discussed on Twitter and YouTube, focusing on platform engagements, particularly through the sharing of YouTube links on Twitter. Sentiment analysis shows that on Twitter, 35.2% of users expressed positive sentiment toward Trump, compared to 28% for Biden. On YouTube, positive sentiment for Trump was 18%, while Biden received 12%. These findings underscore the distinct dynamics in user engagement and sentiment expression on each platform. Bilinski examines how companies leverage various social media platforms to share earnings announcements and their effectiveness in shaping investor responses. The study finds that corporate tweets are particularly impactful when they include essential financial information, mention senior executives (e.g., CEO or CFO), use visual aids, and maintain a moderate tone. These tweets generate stronger investor reactions, especially when retail ownership is high. In contrast, YouTube and Instagram add limited value to corporate communication for influencing stock prices, as Twitter's real-time, information-rich format more effectively engages investors, despite the broader audience reach of other platforms. These studies

synthesise YouTube and Twitter datasets from different technical and social domains, enriching the conversation about user sentiment, thematic relationship modelling, and cross-platform engagement.

Importantly, this study's findings reveal similarities and heterogeneities in review content and emotional tendencies among users of different platforms/datasets in the music video domain. By examining sentiment categorisation in cross-platform user comment content, thematic modelling, and discovering the construction of communities with similar sentiments, this study gains a nuanced understanding of the commenting and dissemination patterns of YouTube videos across platforms. This exploration is intended to enhance this study's grasp of how the current socio-technical environment affects the distribution and sharing of content on YouTube channels, thereby strengthening the strategic development of content that resonates with users across multiple digital environments.

When comparing user emotions expressed on T-video and Y-video, both similarities and differences emerge. T-video comments generally exhibit more consistent and neutral sentiments, while Y-video comments display a broader emotional spectrum, often leaning towards positive. This difference may be attributed to the diverse content and user base of Y-video. The sentiment analysis highlights distinctions in content diversity and emotional engagement between the two platforms. T-video, with its wide variety of content categories, appeals to a broad range of audience interests, which leads to varied levels of emotional engagement. In contrast, Y-video focuses primarily on categories such as "Music" and "Education", which may foster more consistent emotional responses. These targeted content areas on Y-video result in a more balanced sentiment overall, whereas T-video's diversity naturally results in a wider range of emotional reactions. The "People & Blogs" and "Entertainment" categories on Y-video both show higher positive emotional dominance and variability compared to T-video. This could suggest that Y-video's content strategies in these categories are more effective in engaging viewers or that Y-video's audience is more responsive to the content.

Increasing the number of iterations is an important step in improving the performance of the model to ensure that the model is able to fully learn the complex semantic

structures in the data. In this study, by significantly increasing the number of iterations, the performance of the model did not significantly improve on either dataset. This supports each other with the approaches proposed by Tang et al. (Tang et al., 2014) and Hong et al. (Hong and Davison, 2010). The research article by Tang et al. investigated how text length affects LDA performance, pointing out that short texts (in this study, user comments) can make it difficult to accurately estimate topic distributions due to lexical sparsity issues, which affect the model performance. Hong et al.specifically analysed the limitations of short texts (e.g., Twitter data) on the performance of LDA models, and suggested that the problem of short texts can be mitigated by text aggregation or pre-training topic models. Meanwhile, Stevens et al. (Stevens et al., 2012) also argued that when the amount of data is insufficient, it will lead to low consistency among topics, and the increase in the number of iterations will have limited improvement in the model performance when the data is insufficient.

Increasing the number of topics allows the model to capture a wider variety of content, maintaining topic independence and specificity. Conversely, reducing the number of topics blurs topic boundaries and decreases their distinctiveness. To better capture the diversity of content, an appropriate increase in the number of topics can improve the model's accuracy. Differences in user content preferences and engagement patterns between T-video and Y-video result in notable variations in topic distribution and keyword content. These differences likely reflect variations in user demographics, content styles, and engagement habits. T-video users tend to engage in discussions about specific content, while Y-video users focus more on overall sentiment and evaluations.

7.5.2 Limitations

In this study, while employing methods such as perplexity, consistency, and scoring provides a powerful framework for building user review topic models, the effectiveness of these machine learning methods may not be uniformly applicable across different datasets and scenarios. Sometimes, even if the metrics indicate that a certain number is optimal, real-world applications may result in some topics not being sufficiently disaggregated due to the specific nature of the corpus. Lack of homogeneity in data

inhomogeneity (e.g., varying lengths of user comments) may cause the model to favour documents that are richer or more consistently worded, thus affecting topic differentiation. The number of iterations and training details (e.g., learning rates, prior distributions) of the LDA model can also affect the clarity and separation of the final topics. Obviously, the expression of article ideas will be affected not only by the context of words but also by the semantics of words.

In addition, this study encounters limitations in the approach to building online communities based on the same topics and the same sentiment tendencies. By basing this study approach on extracting common themes in music video titles and similar sentiments in each comment, this study aimed to find and establish potential community connections between users on both platforms. However, this approach shows a bias in that the limitations of the way themes are extracted may not fully capture the thematic connections between music videos, suggesting that not all relationships established based on the same themes are based on a strong relational foundation. This highlights the limitations of this study approach as it may not fully capture the strong relationships that build potential social networks.

7.6 Summary

In this chapter, this study explored the dynamics of user sentiment across multiple platforms, focusing on differences in comments relating to YouTube music videos coming from Twitter and YouTube. This study delved into the consistency and variability of sentiments expressed on similar topics, the influence of platform heterogeneity on video categorisation and sentiment expression, and the potential for detecting homogenization in user views.

This study's analysis of entropy and standard deviation as measures of sentiment diversity and dispersion reveals significant insights into the emotional impact of content on T-video and Y-video platforms. T-video shows a pattern of maintaining consistent sentiment responses across most of its content, with high sentiment diversity observed only in a few specific videos. This indicates that T-video may be targeting a specific

audience segment that values consistency and a uniform viewing experience, contributing to a cohesive community with shared emotional responses. In contrast, Y-video displays more diverse emotional responses across its content, which suggests a strategy focused on reaching a broader audience with varied content preferences. This broader emotional range might indicate that Y-video aims to cater to diverse user interests, thus fostering a more dynamic and potentially polarised viewer base. These differences imply that T-video's content strategy might be more effective at creating loyalty and consistent engagement, whereas Y-video's approach could be driving higher engagement by tapping into a wider variety of emotional needs and interests.

Analyses across different categories show that T-video typically exhibits more uniform affective responses, suggesting stability in viewer reactions across various types of content. On the other hand, Y-video demonstrates a wider range of emotional diversity, indicating a varied audience reception that could reflect a more heterogeneous viewer base or diverse content strategies.

Furthermore, an examination of the dominant sentiment scores across different content categories highlights distinct differences between T-video and Y-video. T-video tends to elicit more consistent and neutral sentiment responses, particularly in relation to educational and entertainment content. Conversely, Y-video exhibits greater emotional diversity with generally more positive affective tendencies, especially noticeable in the People & Blogs category.

These findings underscore the unique ways each platform engages its audience and the potential strategies content creators might employ to optimize viewer engagement and satisfaction. This chapter underscored the complex interplay between user sentiment, platform characteristics, and content dynamics. While there are universal themes in how sentiments are expressed across platforms, the nuances that define each platform's unique ecosystem contribute significantly to the shaping of online discourse. Understanding these differences is crucial for content creators, advertisers, and platform developers aiming to foster engaging and positive online environments. Moreover, the ability to detect and understand sentiment homogenization and network formation can aid in anticipating shifts in public opinion and cultural trends, providing valu-

able insights for strategic decision-making in social media management and content curation.

Chapter 8

Conclusion

This Chapter concludes this thesis by examining this study's findings in light of this study's research questions and reiterating the contributions to knowledge. Next, this chapter discusses the limitations of this research and proposes some recommendations for future research. Finally, this chapter addresses the broader significance of YouTube music videos for understanding the behaviours, preferences, and engagement patterns of users on two platforms: YouTube and Twitter.

8.1 Revisiting Questions

8.1.1 Key Factors in YouTube Music Video Engagement

RQ1: Which machine learning method is most suitable for predicting user engagement based on textual and numerical content features?

Chapter 5 describes the use of machine learning algorithms based on the global dataset to predict the predictors of users' engagement with YouTube music videos. The whole experiment was carried out by setting up different feature combination strategies and applying them to the Gradient Boost, BDTR and Random Forest algorithms. Through the performance of different combination strategies on different algorithms, this study found the most suitable algorithms for the full domain dataset and followed extensive experiments, validation, and analysis of the results. Through experiments, validation, and analysis of the results, this study has demonstrated that the BDTR

prediction model using the M1+M3_c feature strategy outperforms all the other models and strategies that have been tested. Therefore, this study chooses to adopt the BDTR model as the underlying framework for the regression prediction analysis.

More specifically, first this study validated the machine learning algorithms based on all the different combination strategies using 10-fold Cross-Validation. Afterwards, considering the generalization ability and stability of the models, as well as the fact that there are no statistically significant differences, this study went on to evaluate and compare the performance attributes of the models on both the test and training datasets in order to further identify the best combination strategies and algorithms based on the global dataset. In order to further explore the accuracy of the Bagging algorithm, different base estimators, i.e., Bagged Decision Tree Regression (BDTR), was used as pooling techniques to compare the model performances among all the combined strategies again. Therefore, BDTR was adopted as an extension of the regression prediction method in this study. The reason for this choice is the intricate and large dimensionality of the dataset. The decision to use a combination of random forests and bagging regressors was made because they have been shown to be effective in mitigating overfitting. The dataset used in this study is from YouTube, a platform that aggregates data from different social media sources with the aim of identifying key attributes that profoundly affect user engagement.

RQ2: What is the impact of time-related features (e.g., release date, comment date) and context-related features (e.g., comment content, title, tags) on user engagement with music videos?

Chapter 5 also answered this question. Using different feature combination strategies, this study eliminates unsuitable elements and combinations of elements for predicting user engagement and finally confirms that the features in datasets consist of "reply_count" and content-related factors, specifically words derived from music video titles, tags, and comments. one of the possible implications of this finding could be that use of certain terms may draw more user attention. In the previous portion of the methodology, this study referred to the application of stemming as a technique for extracting words that share similar meanings.

At the same time, this study's findings reveal differences in the keywords associated with user engagement and contextual relevance across the two platforms. One possible implication of this observation is that the use of certain terms, particularly those related to YouTube channels or channel owners, may draw more user attention in the Y-video dataset, suggesting that YouTubers in Y-video might have a greater visible presence or recognition among viewers compared to those in T-video. Additionally, the fact that the number of comments on music videos in Y-video is significantly lower than in T-video suggests that user preferences may be driven more by the channel's identity or theme rather than individual engagements.

8.1.2 Factors Affecting User Engagement on Two Platforms and the Corresponding Predictive Models

RQ3: Which is the most suitable machine learning method for predicting user engagement based on two datasets, which are different data sources, respectively?

RQ4: What are the factors (textual factors, context factors) that influence user video engagement on each of the two different platforms?

Unlike Chapter 5, Chapter 6 focuses more specifically on the differences between the two datasets, T-video and Y-video, addressing both research questions in detail. While Chapter 5 is based on the combined T&Y-video dataset, which merges the T-video and Y-video datasets to explore their overall patterns, Chapter 6 conducts a more granular analysis by examining them separately. Informed by the experimental results from Chapter 5 (see Section 5.2), some feature strategies are simplified for Chapter 6. Since M2 (time features) and M4 (categories) were found to have little predictive power for the target variable (i.e., the number of Likes), these features are excluded from the analysis of T-video and Y-video in Chapter 6.

Based on the preliminary 10-fold cross-validation and further combined analyses on the training and test sets, the Random Forest model was found to be the preferred model for the T-video dataset as far as a single machine learning algorithm is concerned, whereas the BDTR algorithm outperformed the Y-video dataset based on the respective R^2 and RMSE metrics on the test dataset. These findings highlight the

goodness of fit and prediction accuracy of each model. Overall, for both datasets, the BDTR model demonstrates the most robust performance under the comprehensive M1+M2+M3_c+M4 strategy. However, the marginal improvements offered by the addition of M2 and M4 features suggest that these elements do not significantly enhance the prediction process. This analysis underscores BDTR's effectiveness in leveraging integrated feature strategies without substantial benefit from the inclusion of additional time and category features.

For RQ4, in addition to the significant effect of reply_count as a predictor on both datasets, there are clear differences between Y-video and T-video in terms of the factors that influence the engagement of various users. In the T-video dataset, the influencing factors place greater emphasis on the emotional and descriptive resonance of the content of the tweets, likely reflecting the appeal of the video and the emotional response of the viewer. The word 'office' is often associated with 'official' channels, suggesting that music videos released by an artist's official channels tend to attract a large amount of user attention and engagement, which is indicative of the artist and their official presence appeal. In contrast, the Y-video dataset shows the channel name as a prominent feature, implying that users may be loyal to specific creators or brands on YouTube. Words such as 'pixl', 'poptast' and 'stumm' appear among the significant features, suggesting that channel branding may influence viewer ratings, suggesting that users prefer content based on the channel's identity or themes rather than personal engagements.

These findings reveal different dynamics for engaging subscribers and promoting video content across platforms, with T-video's engagements emphasising direct subscriber engagement and emotional response, while Y-video points to the importance of channel branding and creator influence in engaging viewers.

8.1.3 Sentiment Analysis on T-video and Y-video

RQ5: Does the comparison of sentiment expressed by users' comments on similar topics on Twitter and YouTube reveal consistency or differences in user sentiment across platforms?

RQ6: If heterogeneity exists, is there an association with the categories used by the

creators when they are uploading their music videos?

Chapter 7 performs sentiment analysis through a lexicon-based approach, namely VADER. This study first introduces and explains the advantages of using VADER for sentiment analysis. After that, this study uses topic extraction to extract topics from the users' comment data within T-video and Y-video, and finally, it performs topic visualisation with LDA.

For RQ5, the comparison of sentiment expressed by users' comments on similar topics reveals both consistencies and differences in user sentiment across platforms. On both platforms, positive videos tend to rely on the quality of user comments to drive engagement, suggesting that well-crafted or emotionally resonant comments play a significant role in attracting likes. Negative videos, however, generate broader user engagement because of their topical nature or ability to provoke stronger emotional responses, leading to more discussions and replies.

The differences between platforms are notable in the emotional complexity and engagement patterns of user comments. On T-video, user sentiment exhibits higher emotional complexity and diversity, with content triggering a wider range of emotional responses. However, emotional reactions within individual videos remain relatively stable and consistent, with only a few videos provoking extreme emotional fluctuations. In contrast, the Y-video demonstrates lower emotional complexity, where emotional responses within individual videos are more uniform and predictable. At the same time, the variability of sentiment between videos on Y-video is greater, reflecting more diverse emotional responses to different content. Additionally, user engagement patterns differ: in Y-video, there is a stronger correlation between the number of comments and likes, indicating that users are more likely to engage with videos that already show high engagement levels. On T-video, likes are more influenced by the quality and content of user comments rather than the sheer volume of comments.

For RQ6, the heterogeneity of user perspectives across platforms does lead to significant differences in both the categories of videos chosen by uploaders and the sentiments expressed in user comments. In the Education and Entertainment categories, T-video and Y-video show differing patterns of emotional responses. On T-video, emotional

responses are more consistent and stable, suggesting that videos in these categories maintain a relatively uniform content style that appeals to audiences looking for predictability and stability. In contrast, Y-video displays greater emotional dispersion and diversity in these categories, indicating that its content is more varied and capable of eliciting a wider range of emotional reactions from viewers. This may reflect Y-video's strategy of offering diverse content to cater to broader audience preferences.

In the People & Blogs and Music categories, the emotional responses on both platforms are more complex and diverse, reflecting the ability of these categories to trigger a wide range of emotions among viewers. However, there are notable platform-specific differences. On Y-video, user sentiment tends to be more positive and consistent, especially in the People & Blogs category, where emotional responses show a higher degree of uniformity. This suggests that Y-video's content strategy in this category aligns with viewer preferences for more consistent and predictable emotional experiences. On T-video, emotional responses in these categories show higher variability and complexity, indicating that its content is more diverse and capable of appealing to different audience segments by eliciting a broader range of emotional reactions.

In summary, while both platforms show a reliance on quality content for positive video engagement and broader user engagement for negative videos, T-video exhibits greater emotional diversity and stability within individual videos, whereas Y-video displays a more concentrated and positive emotional engagement across its content. And the heterogeneity in user perspectives across platforms influences the emotional responses within different video categories. T-video's content in categories like People & Blogs and Music tends to elicit a broader range of emotional reactions, while Y-video's content strategy results in more focused and positive emotional engagement, particularly in the People & Blogs and Entertainment categories.

8.2 Research Contributions

8.2.1 Optimising Feature Combinations for User Engagement Prediction

This study contributes to the field of user engagement prediction by systematically exploring the interplay of textual, numerical, and contextual features. Through experimentation with various machine learning models and feature combination strategies, this study identified key features and the most effective algorithms for predicting user engagement with YouTube music videos. This work builds a comprehensive framework to address the challenges posed by the diverse and high-dimensional nature of user-generated data, thereby providing actionable insights for content creators and platform managers.

To achieve this, the study developed and validated multiple feature combination strategies that integrate time-related aspects (e.g., publishing dates and commenting activity) with content-related attributes (e.g., video titles, tags, and user comments). Using advanced machine learning techniques, including Gradient Boost, Bagging, and Random Forest algorithms, this study evaluated the predictive power of these combinations across multiple dimensions. Among the tested models, the BDTR approach, utilising the M1+M3_c (i.e. num of reply+words_TFIDF) feature strategy, emerged as the most robust and effective. This model was specifically chosen for its ability to manage the intricate relationships within the dataset while mitigating overfitting, ensuring strong generalisation capabilities.

Key findings highlight that features like "reply_count" and words derived from music video tags, titles, and comments are crucial in driving user engagement. For instance, incorporating high-value keywords within a video's metadata significantly increases its likelihood of attracting user attention. Moreover, this study reveals distinct differences in user engagement patterns across platforms. In Y-video, keywords associated with channel identity and channel owners demonstrate a stronger influence on user preferences, suggesting that users are more inclined to engage with content based on the creator's reputation or thematic consistency rather than individual comment engage-

ments. In contrast, T-video showcases a larger volume of user comments, implying a higher reliance on direct engagements and user-generated discussions to shape engagement.

By implementing an iterative validation process, including 10-fold cross-validation and performance evaluations on test and training datasets, this study was able to refine feature selection and optimise model stability. Additional exploration using various base estimators for Bagging models (e.g., Random Forest) reinforced the robustness of the Bagged Decision Tree Regression (BDTR), which consistently outperformed other combinations and demonstrated scalability for large datasets. This approach aligns with findings from prior research, such as Tang et al. (Tang et al., 2014) and Hong et al. (Hong and Davison, 2010), which emphasise the importance of text aggregation and feature selection in improving prediction accuracy for short text datasets like user comments.

This study not only identifies the most suitable algorithms and feature strategies for user engagement prediction but also uncovers the nuanced relationships between content features, user behaviours, and platform dynamics. These insights are vital for developing more effective content strategies, enabling creators to design videos that resonate with users and sustain long-term engagement across diverse digital environments.

8.2.2 Advancing User Engagement Prediction Across Platforms

This study highlights a detailed comparison of user engagement prediction and the factors influencing user preferences on two distinct platforms, T-video and Y-video. By leveraging machine learning algorithms and feature combination strategies, the research identifies platform-specific dynamics in engaging viewers and provides insights into predictive modelling for content performance.

These findings provide valuable insights into the nuances of engagement strategies on different platforms. T-video benefits from fostering direct emotional connections and leveraging official branding to attract fans, while Y-video relies on the power of channel identity and creator influence to engage viewers. The distinct user preferences and

engagement patterns between the two platforms illustrate the importance of tailoring content strategies to platform-specific dynamics.

This research not only advances the understanding of cross-platform engagement modelling but also underscores the need for adaptable and feature-driven predictive frameworks that account for the diversity in user behaviour and content engagement. By focusing on the integration of machine learning models with platform-specific insights, this study contributes to the development of targeted strategies for maximising engagement across digital environments.

8.2.3 Revealing Platform-Specific Sentiment Dynamics and Engagement Drivers

This study provides a comprehensive examination of cross-platform emotional dynamics and content strategies that influence user engagement on T-video and Y-video. By conducting sentiment analysis and topic modelling, this research highlights the similarities and differences in user engagement, emotional responses, and the role of content strategies across these platforms.

For user sentiment consistency and differences, the findings reveal shared patterns, such as positive videos relying on high-quality, emotionally resonant comments to drive engagement. Conversely, negative videos evoke broader user engagement by sparking discussions and emotional responses. However, platform-specific differences emerge: On T-video, user sentiment is characterised by greater emotional complexity and diversity, with responses ranging widely across videos but maintaining stability within individual videos. This indicates a platform geared toward fostering varied emotional reactions while ensuring consistency within specific content. On Y-video, emotional responses are more uniform and predictable within individual videos but show greater variability between videos. This suggests a strategy focused on engaging users through targeted content themes and leveraging engagement metrics like comment volume to drive likes and engagement.

For heterogeneity in user perspectives and its impact on video categories, the study underscores distinct platform dynamics: In the Education and Entertainment cate-

gories, T-video exhibits stable and consistent emotional responses, appealing to audiences seeking predictability. In contrast, Y-video showcases greater emotional dispersion, reflecting its diverse content strategy that caters to broader audience preferences. In the People & Blogs and Music categories, both platforms evoke more complex and diverse emotional responses. T-video's content drives a broader range of emotional reactions, indicating its strength in appealing to varied audience segments. Y-video, however, demonstrates more positive and uniform sentiment, particularly in the People & Blogs category, aligning with viewer preferences for consistent emotional experiences.

These findings highlight that while both platforms rely on high-quality content for positive video engagement and topicality for negative videos, their approaches diverge significantly. T-video excels in eliciting diverse emotional reactions, making it well-suited for viewers seeking varied experiences. Y-video, on the other hand, fosters more concentrated and positive emotional engagement, aligning with a content strategy that prioritises consistency and predictability, particularly in categories like People & Blogs and Entertainment. This research contributes to the understanding of cross-platform user engagement by showcasing how emotional dynamics and content strategies influence engagement on different platforms. It underscores the importance of tailoring content approaches to platform-specific user behaviours and preferences, offering actionable insights for content creators and platform strategists aiming to optimise audience engagement.

8.2.4 Revealing Platform-Specific Dynamics in Music Video Engagement

By building and analysing datasets from T-video (Twitter) and Y-video (YouTube), this research uncovers notable differences in platform-specific data structures and engagement patterns under similar topics. These findings contribute to a deeper understanding of how platform dynamics influence music video engagements and highlight the contrasting roles of content diversity and creator-driven engagement in shaping user behaviours.

The T-video dataset, derived from Twitter, demonstrates substantial diversity in its

content. It includes a wide variety of music video genres, spans a significant temporal range, and features a large number of videos with both high view counts and extensive comments. This breadth reflects the platform's open, dynamic nature, which supports the dissemination of official content while fostering user-driven discussions across a wide array of topics. Twitter's conversational framework allows for real-time engagement, making it a hub for engaging with trending and diverse music content.

In contrast, the Y-video dataset, sourced from YouTube, is characterised by its focus on a specific genre, pop music, and is predominantly driven by a small number of highly-followed creators. These music channels exhibit limited diversity in video categories, and their engagement metrics, such as play counts and comments, are heavily dependent on the size of the creators' follower bases. The dataset's temporal scope aligns with the lifecycle of the affiliated channels, suggesting that audience loyalty to specific creators or channels significantly shapes user engagements. This indicates that Y-video's engagement patterns are centred around sustained viewer loyalty and creator branding, rather than the wide-ranging topical appeal observed in T-video.

These differences underline how platform-specific attributes influence user engagement. While Twitter facilitates widespread discussion across varied content, fostering emotional and topical diversity, YouTube's structured ecosystem emphasises niche community engagement driven by creator influence. The findings from these datasets reveal distinct strategies for maximising user engagement on each platform, offering valuable insights for content creators, marketers, and platform developers.

This research thus provides a foundational framework for understanding cross-platform user engagement in the music video domain. It contributes to the broader discourse on how platform characteristics shape audience behaviour and offers a basis for future exploration into optimising cross-platform content strategies to align with diverse user expectations and engagement patterns.

8.3 Research Limitations

8.3.1 Bias in Cross-Platform Engagement Analysis

This study encounters a limitation in its approach to linking YouTube music video popularity with Twitter sharing patterns. By aggregating tweets related to YouTube music videos through common hashtags and phrases, this study aimed to establish a connection between user engagement on both platforms. However, this method revealed a bias towards music videos of higher production quality being more frequently shared on Twitter, suggesting that not all content receives equal visibility or opportunity for engagement. This aspect of content sharing on Twitter, which can act as a form of quality control, may not accurately represent the broader spectrum of user preferences. The phenomenon, akin to the distinct ecologies and sharing patterns identified by Segerberg and Bennett (2011) in their analysis of hashtags used in climate demonstrations, points to the complexity of social media engagements and the potential for certain types of content to be privileged over others. This highlights a limitation in the methodologies of this study, as it may not fully capture the diverse factors influencing user engagement and preferences across platforms. Additionally, the study does not distinguish between engagement from a small group of highly active users versus a broader range of participants, which may impact the interpretation of user engagement trends.

8.3.2 Time Constraints

This study was affected by time constraints, especially during the process of data collection and analysis. This time constraint could lead to several implications. Firstly, as the study must be completed within a limited timeframe, the shorter duration of data collection may not result in a large enough sample to adequately represent the overall picture of the study. For example, certain phenomena may require longer observation periods or repeated data collection over time to capture trends, but time constraints prevented the implementation of such a longitudinal design in this study. This may make the data collected somewhat limited and affect the generalisability and external validity of the findings.

Secondly, time constraints may also have affected the depth of data analysis. Under more intense time pressures, data may not be processed and analysed to cover all potential variables and relationships. In order to complete the data analysis within the given timeframe, certain complex analytical methods may not be applied, and thus, potentially important findings may be overlooked. In addition, due to time constraints, the literature review conducted during the study may not be able to cover all relevant studies, which may affect the construction of the theoretical framework and the in-depth discussion of the results.

8.3.3 Data Limitations

1. Data Integrity

The data in this study is mainly captured through the YouTube API, which has some limitations in terms of data integrity. Firstly, the type and scope of data provided by the YouTube API are limited, and it can only access the data that the platform allows to be made public, such as the number of comments, views, likes, etc. of the video, whereas for more in-depth data related to user behaviour or detailed audience statistics, the API may not be able to provide it. This means that there is a certain amount of missing information in the dataset, preventing a more comprehensive and in-depth understanding of this study. In addition, network connectivity issues during API data crawling or limitations in API limits may result in some of the data not being successfully fetched, which in turn results in the appearance of missing values. In order to deal with these missing values, this study adopts corresponding processing methods, such as missing value interpolation and deletion of incomplete samples, but these methods may also introduce certain biases, which affect the accuracy of the study and the reliability of the conclusions.

Another limitation is that the data crawled by the YouTube API is limited by the policies and privacy settings of the YouTube platform. For some videos or users, their privacy settings may make specific data inaccessible, which may lead to insufficient data in certain categories in the data sample, thus affecting the comprehensiveness of the analysis. For example, some of the interviewed videos were not open for commenting

due to the uploader's settings, which makes the sample of this study insufficiently rich in analysing the content of the comments, which in turn may affect the broad applicability of the conclusions.

2. Data Temporality

The data on the YouTube platform is highly temporal, which also imposes significant limitations on the findings. Since the data for this study was crawled over a specific time period through the YouTube API, the data collected only reflects information such as the popularity of the video and the audience's reaction at that point in time or within that time period, which may not be representative of the long-term trend. For example, certain videos may receive a large number of views and discussions during a certain period of time due to a specific current event hotspot, but this hotness may decline rapidly within a short period of time. Therefore, if the point in time of data collection coincides with the peak of a particular video's popularity, the findings may be biased and difficult to generalise to the long-term popularity of the video or the long-term attitude of viewers towards it.

In addition, the content and user behaviour on YouTube are dynamic and changing, especially when the data is affected by certain external events (e.g., breaking news, policy changes, etc.). This study was only able to capture data for a limited period of time due to time and resource constraints, and thus failed to fully consider the dynamic changes in the data and their impact on the findings. This means that the generalisability of the findings may be limited, especially when considering long-term effects and changes over time, and the applicability of the results may be reduced. In order to minimise the impact of data temporality on the study, this study tried to take time into account when analysing the data and selected data from a number of different time points for comparison, but this measure still has limitations. Future studies can collect data over a longer time span or use real-time monitoring with multiple repeated captures to better understand the characteristics of user behaviour and video heat over time on the YouTube platform, thus improving the stability and generalizability of the study's conclusions.

In summary, this study has some limitations in terms of data integrity and tempo-

rality. Firstly, data integrity is affected by the limitations of the YouTube API itself, as well as privacy settings and possible network and quota issues during data crawling. This leads to the possibility that the data may not be comprehensive enough, thus affecting the accuracy and representativeness of the findings. Secondly, the temporal nature of the data also poses a challenge for this study. As the study relies on YouTube data crawled over a specific time period, the trends reflected in the data may only be representative of a specific point in time, making it difficult to fully apply to long-term, dynamic user behaviour and video performance.

Nonetheless, by being as rigorous as possible in its data processing and analysis methods, this study seeks to provide valuable insights within the existing conditions.

8.4 Future Research

Following the research, this study notes that there are a number of future studies that could be of interest. Due to limited knowledge and available resources, this study's experiments are currently only 3 machine learning algorithms for regression prediction analyses on two datasets, and for future work, this study hopes to seek more possibilities with more regression prediction algorithms.

This study discovered that while not very prominent, a noticeable portion of the comments were written in languages from other nations (like Korean and Latin) when this study applied natural language processing to the context data. This reflects a degree of linguistic diversity within the dataset. However, the study's conclusions are limited in generalizability due to the dataset's exclusive focus on YouTube music videos from English-speaking nations. Applying the methodology of creating the dataset and performing the steps of natural language processing in this study to data from other social media platforms can help increase the generalizability of the findings.

Despite this study's efforts to consider several elements that could impact user engagement, this study's understanding is constrained by limited knowledge and the limitations of data crawling techniques. As further work, the plan is to improve the quality and coverage of the classification algorithms, especially by improving the way different model features are combined. There exist potential variables that have not

been considered, such as data pertaining to YouTube channels and demographic characteristics of YouTubers, among other factors. Future research will consider incorporating these measures to increase the set of predictive features used in the regression model.

In this study, while employing methods such as perplexity and consistency and increasing the number of iterations provides a powerful framework for building user review topic models, the effectiveness of these machine learning methods may not be uniformly applicable across different datasets and scenarios. Sometimes, even if the metrics indicate that a certain number is optimal, real-world applications may result in some topics not being sufficiently decentralised due to the specific nature of the corpus. Data inhomogeneity (varying lengths of user comments) may cause the model to favour documents that are richer or more consistently worded, thus affecting topic differentiation. The number of iterations and training details (e.g., learning rates, prior distributions) of the LDA model can also affect the clarity and separation of the final topics. An insufficient number of iterations may result in the model failing to adequately learn the complex structures in the data. This phenomenon is in line with the approach proposed by Alsaedi et al., (Alsaedi et al., 2016) and Zhu et al., (Zhu et al., 2019), which ignores the influence of article context and the problem of word polysemy. Obviously, the expression of article ideas will be affected not only by the context of words but also by the semantics of words. However, in order to further remedy these limitations, future research could take the following steps:

(1) More comprehensive data collection: Future research could overcome the limited scope of the YouTube API data by using more diverse data crawling tools or incorporating other data sources to obtain more comprehensive and in-depth data.

(2) Increase data time span: Future research can extend the time span of data collection or adopt dynamic tracking to collect data multiple times at different points in time to capture long-term trends in user behaviour and video performance, thereby increasing the generality and stability of conclusions.

(3) Application of multiple analysis methods: In addition to traditional data analysis methods, tools such as time-series analysis or machine learning can be introduced to cope with the temporal nature of the data and better understand the changing

patterns of data at different points in time.

In addition, this study encounters limitations in the approach to building online communities based on the same topics and the same sentiment tendencies. By basing the approach on extracting common themes in music video titles and similar sentiments in each comment, this study aimed to find and establish potential community connections between users on both platforms. However, this approach shows a bias in that the limitations of the way themes are extracted may not fully capture the thematic connections between music videos, suggesting that not all relationships established based on the same themes are based on a strong relational foundation. This highlights the limitations of the approach as it may not fully capture the strong relationships that build potential social networks.

Bibliography

Abisheva, A., Garimella, V. R. K., Garcia, D. and Weber, I. (2014), Who watches (and shares) what on youtube? and when? using twitter to understand youtube viewership, *in* 'Proceedings of the 7th ACM international conference on Web search and data mining', pp. 593–602.

Abu-El-Haija, S., Kothari, N., Lee, J., Natsev, P., Toderici, G., Varadarajan, B. and Vijayanarasimhan, S. (2016), 'Youtube-8m: A large-scale video classification benchmark', *arXiv preprint arXiv:1609.08675* .

Abubakar, H. D., Umar, M. and Bakale, M. A. (2022), 'Sentiment classification: Review of text vectorization methods: Bag of words, tf-idf, word2vec and doc2vec', *SLU Journal of Science and Technology* **4**(1), 27–33.

Agarwal, B., Mittal, N., Bansal, P. and Garg, S. (2015), 'Sentiment analysis using common-sense and context information', *Computational intelligence and neuroscience* **2015**(1), 715730.

Ahmad, A. H., Idris, I., Wong, J. X., Malik, I. S. A., Masri, R. and Alias, S. S. (2020), 'Creating brand awareness through youtube advertisement engagement', *Test Engineering & Management* **83**(4), 7970–7976.

Ahmed, M., Spagna, S., Huici, F. and Niccolini, S. (2013), A peek into the future: Predicting the evolution of popularity in user generated content, *in* 'Proceedings of the sixth ACM international conference on Web search and data mining', pp. 607–616.

Akter, S. and Aziz, M. T. (2016), Sentiment analysis on facebook group using lexicon based approach, *in* '2016 3rd international conference on electrical engineering and information communication technology (ICEEICT)', IEEE, pp. 1–4.

Al-Tamimi, A.-K., Shatnawi, A. and Bani-Issa, E. (2017), Arabic sentiment analysis of youtube comments, *in* '2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)', IEEE, pp. 1–6.

Alsaedi, N., Burnap, P. and Rana, O. (2016), Temporal tf-idf: A high performance approach for event summarization in twitter, *in* '2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)', IEEE, pp. 515–521.

Alsayat, A. (2022), 'Improving sentiment analysis for social media applications using an ensemble deep learning language model', *Arabian Journal for Science and Engineering* **47**(2), 2499–2511.

Anderson, C. (2006), *The long tail: Why the future of business is selling less of more*, Hachette UK.

Ardabili, S., Mosavi, A. and Várkonyi-Kóczy, A. R. (2019), Advances in machine learning modeling reviewing hybrid and ensemble methods, *in* 'International conference on global research and education', Springer, pp. 215–227.

Asghar, M. Z., Ahmad, S., Marwat, A. and Kundi, F. M. (2015), 'Sentiment analysis on youtube: A brief survey', *arXiv preprint arXiv:1511.09142* .

Baccianella, S., Esuli, A., Sebastiani, F. et al. (2010), Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining., *in* 'Lrec', Vol. 10, Valletta, pp. 2200–2204.

Bakshy, E., Rosenn, I., Marlow, C. and Adamic, L. (2012), The role of social networks in information diffusion, *in* 'Proceedings of the 21st international conference on World Wide Web', pp. 519–528.

Barbier, G. and Liu, H. (2011), 'Data mining in social media', *Social network data analytics* pp. 327–352.

Bärtl, M. (2018), 'Youtube channels, uploads and views: A statistical analysis of the past 10 years', *Convergence* **24**(1), 16–32.

Bawden, D. and Robinson, L. (2009), 'The dark side of information: overload, anxiety and other paradoxes and pathologies', *Journal of information science* **35**(2), 180–191.

Belinkov, Y. and Bisk, Y. (2017), 'Synthetic and natural noise both break neural machine translation', *arXiv preprint arXiv:1711.02173* .

Benevenuto, F., Rodrigues, T., Almeida, V., Almeida, J. and Ross, K. (2009), 'Video interactions in online video social networks', *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* **5**(4), 1–25.

Benevenuto, F., Rodrigues, T., Cha, M. and Almeida, V. (2009), Characterizing user behavior in online social networks, *in* 'Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement', pp. 49–62.

Berger, J. and Milkman, K. L. (2012), 'What makes online content viral?', *Journal of marketing research* **49**(2), 192–205.

Bhuiyan, H., Ara, J., Bardhan, R. and Islam, M. R. (2017), Retrieving youtube video by sentiment analysis on user comment, *in* '2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)', IEEE, pp. 474–478.

Bilinski, P. (2024), 'The content of tweets and the usefulness of youtube and instagram in corporate communication', *European Accounting Review* **33**(1), 279–311.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003), 'Latent dirichlet allocation', *Journal of machine Learning research* **3**(Jan), 993–1022.

Bonta, V., Kumaresh, N. and Janardhan, N. (2019), 'A comprehensive study on lexicon based approaches for sentiment analysis', *Asian Journal of Computer Science and Technology* **8**(S2), 1–6.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A. and Van Bavel, J. J. (2017), 'Emotion shapes the diffusion of moralized content in social networks', *Proceedings of the National Academy of Sciences* **114**(28), 7313–7318.

Breiman, L. (1996), 'Bagging predictors', *Machine learning* **24**, 123–140.

Breiman, L. (2001), 'Random forests', *Machine learning* **45**, 5–32.

Burgess, J. (2018), *YouTube: Online video and participatory culture*, John Wiley & Sons.

Burgess, J. and Green, J. (2009), 'Youtube: Digital media and society series', *Cambridge: Polity* .

Burgess, J. and Matamoros-Fernández, A. (2016), 'Mapping sociocultural controversies across digital media platforms: One week of# gamergate on twitter, youtube, and tumblr', *Communication Research and Practice* **2**(1), 79–96.

Cahyani, D. E. and Patasik, I. (2021), 'Performance comparison of tf-idf and word2vec models for emotion text classification', *Bulletin of Electrical Engineering and Informatics* **10**(5), 2780–2788.

Cai, J., Luo, J., Wang, S. and Yang, S. (2018), 'Feature selection in machine learning: A new perspective', *Neurocomputing* **300**, 70–79.

Cambria, E., Schuller, B., Xia, Y. and Havasi, C. (2013), 'New avenues in opinion mining and sentiment analysis', *IEEE Intelligent systems* **28**(2), 15–21.

Cayari, C. (2011), 'The youtube effect: How youtube has provided new ways to consume, create, and share music.', *International journal of education & the arts* **12**(6), n6.

Celma Herrada, Ò. et al. (2009), *Music recommendation and discovery in the long tail*, Universitat Pompeu Fabra.

Cha, M., Kwak, H., Rodriguez, P., Ahn, Y.-Y. and Moon, S. (2007), I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system, *in* 'Proceedings of the 7th ACM SIGCOMM conference on Internet measurement', pp. 1–14.

Chau, C. (2010), 'Youtube as a participatory culture', *New directions for youth development* **2010**(128), 65–74.

Chekima, K. and Alfred, R. (2018), Sentiment analysis of malay social media text, *in* 'Computational Science and Technology: 4th ICCST 2017, Kuala Lumpur, Malaysia, 29–30 November, 2017', Springer, pp. 205–219.

Chen, T. and Guestrin, C. (2016), Xgboost: A scalable tree boosting system, *in* 'Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining', pp. 785–794.

Chen, Y.-L., Chang, C.-L. and Yeh, C.-S. (2017), 'Emotion classification of youtube videos', *Decision Support Systems* **101**, 40–50.

Cheng, X., Dale, C. and Liu, J. (2008), Statistics and social network of youtube videos, *in* '2008 16th Interntional Workshop on Quality of Service', IEEE, pp. 229–238.

Chowdhury, S. A. and Makaroff, D. (2013), Popularity growth patterns of youtube videos-a category-based study, *in* 'International Conference on Web Information Systems and Technologies', Vol. 2, SCITEPRESS, pp. 233–242.

Covington, P., Adams, J. and Sargin, E. (2016), Deep neural networks for youtube recommendations, *in* 'Proceedings of the 10th ACM conference on recommender systems', pp. 191–198.

Crane, R. and Sornette, D. (2008), 'Robust dynamic classes revealed by measuring the response function of a social system', *Proceedings of the National Academy of Sciences* **105**(41), 15649–15653.

Das, B. and Chakraborty, S. (2018), 'An improved text sentiment classification model using tf-idf and next word negation', *arXiv preprint arXiv:1806.06407* .

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. and Harshman, R. (1990), 'Indexing by latent semantic analysis', *Journal of the American society for information science* **41**(6), 391–407.

Dhaoui, C., Webster, C. M. and Tan, L. P. (2017), 'Social media sentiment analysis: lexicon versus machine learning', *Journal of Consumer Marketing* **34**(6), 480–488.

Dietterich, T. G. (2000), Ensemble methods in machine learning, *in* 'International workshop on multiple classifier systems', Springer, pp. 1–15.

Drus, Z. and Khalid, H. (2019), 'Sentiment analysis in social media and its application: Systematic literature review', *Procedia Computer Science* **161**, 707–714.

El Rahman, S. A., AlOtaibi, F. A. and AlShehri, W. A. (2019), Sentiment analysis of twitter data, *in* '2019 international conference on computer and information sciences (ICCIS)', IEEE, pp. 1–4.

Elbagir, S. and Yang, J. (2019), Twitter sentiment analysis using natural language toolkit and vader sentiment, *in* 'Proceedings of the international multiconference of engineers and computer scientists', Vol. 122, sn.

Eppler, M. J. and Mengis, J. (2008), 'The concept of information overload-a review of literature from organization science, accounting, marketing, mis, and related disciplines (2004) the information society: An international journal, 20 (5), 2004, pp. 1–20', *Kommunikationsmanagement im Wandel: Beiträge aus 10 Jahren= mcminstitute* pp. 271–305.

Figueiredo, F., Almeida, J. M., Benevenuto, F. and Gummadi, K. P. (2014), Does content determine information popularity in social media? a case study of youtube videos' content and their popularity, *in* 'Proceedings of the SIGCHI conference on human factors in computing systems', pp. 979–982.

Figueiredo, F., Almeida, J. M., Gonçalves, M. A. and Benevenuto, F. (2014), 'On the dynamics of social media popularity: A youtube case study', *ACM Transactions on Internet Technology (TOIT)* **14**(4), 1–23.

Figueiredo, F., Benevenuto, F. and Almeida, J. M. (2011), The tube over time: characterizing popularity growth of youtube videos, *in* 'Proceedings of the fourth ACM international conference on Web search and data mining', pp. 745–754.

Gareth, J., Daniela, W., Trevor, H. and Robert, T. (2013), *An introduction to statistical learning: with applications in R*, Spinger.

Gatta, V. L., Luceri, L., Fabbri, F. and Ferrara, E. (2023), The interconnected nature of online harm and moderation: investigating the cross-platform spread of harmful content between youtube and twitter, *in* 'Proceedings of the 34th ACM conference on hypertext and social media', pp. 1–10.

Ginossar, T., Cruickshank, I. J., Zheleva, E., Sulskis, J. and Berger-Wolf, T. (2022), 'Cross-platform spread: vaccine-related content, sources, and conspiracy theories in youtube videos shared in early twitter covid-19 conversations', *Human vaccines & immunotherapeutics* **18**(1), 1–13.

Golovchenko, Y., Buntain, C., Eady, G., Brown, M. A. and Tucker, J. A. (2020), 'Cross-platform state propaganda: Russian trolls on twitter and youtube during the 2016 us presidential election', *The International Journal of Press/Politics* **25**(3), 357–389.

Guess, A., Nyhan, B., Lyons, B. and Reifler, J. (2018), 'Avoiding the echo chamber about echo chambers', *Knight Foundation* **2**(1), 1–25.

Gundecha, P. and Liu, H. (2012), 'Mining social media: a brief introduction', *New directions in informatics, optimization, logistics, and production* pp. 1–17.

Guo, P. J., Kim, J. and Rubin, R. (2014), How video production affects student engagement: An empirical study of mooc videos, *in* 'Proceedings of the first ACM conference on Learning@ scale conference', pp. 41–50.

Hassan, A. U., Hussain, J., Hussain, M., Sadiq, M. and Lee, S. (2017), Sentiment analysis of social networking sites (sns) data using machine learning approach for the measurement of depression, *in* '2017 international conference on information and communication technology convergence (ICTC)', IEEE, pp. 138–140.

Hastie, T., Tibshirani, R., Friedman, J. H. and Friedman, J. H. (2009), *The elements of statistical learning: data mining, inference, and prediction*, Vol. 2, Springer.

Haury, A.-C., Gestraud, P. and Vert, J.-P. (2011), 'The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures', *PloS one* **6**(12), e28210.

He, W., Zha, S. and Li, L. (2013), 'Social media competitive analysis and text mining: A case study in the pizza industry', *International journal of information management* **33**(3), 464–472.

Hennig-Thurau, T., Houston, M. B. and Walsh, G. (2007), 'Determinants of motion picture box office and profitability: an interrelationship approach', *Review of Managerial Science* **1**, 65–92.

Ho, T. K. (1998), 'The random subspace method for constructing decision forests', *IEEE transactions on pattern analysis and machine intelligence* **20**(8), 832–844.

Hoi, S. C., Sahoo, D., Lu, J. and Zhao, P. (2021), 'Online learning: A comprehensive survey', *Neurocomputing* **459**, 249–289.

Hong, L. and Davison, B. D. (2010), Empirical study of topic modeling in twitter, *in* 'Proceedings of the first workshop on social media analytics', pp. 80–88.

Hudson, S. and Hudson, R. (2013), 'Engaging with consumers using social media: a case study of music festivals', *International Journal of Event and Festival Management* **4**(3), 206–223.

Hutto, C. and Gilbert, E. (2014), Vader: A parsimonious rule-based model for sentiment analysis of social media text, *in* 'Proceedings of the international AAAI conference on web and social media', Vol. 8, pp. 216–225.

Jannach, D. and Adomavicius, G. (2016), Recommendations with a purpose, *in* 'Proceedings of the 10th ACM conference on recommender systems', pp. 7–10.

Jernigan, D. H. and Rushman, A. E. (2014), 'Measuring youth exposure to alcohol marketing on social networking sites: Challenges and prospects', *Journal of public health policy* **35**(1), 91–104.

Ji, S., Lu, X., Liu, M., Sun, L., Liu, C., Du, B. and Xiong, H. (2023), Community-based dynamic graph learning for popularity prediction, *in* 'Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining', pp. 930–940.

Jiang, H., Wang, W., Wei, Y., Gao, Z., Wang, Y. and Nie, L. (2020), What aspect do you like: Multi-scale time-aware user interest modeling for micro-video recommendation, *in* 'Proceedings of the 28th ACM International conference on Multimedia', pp. 3487–3495.

Joachims, T. (2002), Optimizing search engines using clickthrough data, *in* 'Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining', pp. 133–142.

Jones, W. P. and Teevan, J. (2007), *Personal information management*, Vol. 14, University of Washington Press Seattle, WA.

Jönsson, A. M. and Örnebring, H. (2011), 'User-generated content and the news: Empowerment of citizens or interactive illusion?', *Journalism Practice* **5**(2), 127–144.

Kavoori, A. (2011), *Reading YouTube: The critical viewers guide*, Peter Lang Publishing.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.-Y. (2017), 'Lightgbm: A highly efficient gradient boosting decision tree', *Advances in neural information processing systems* **30**.

Khan, M. L. (2017), 'Social media engagement: What motivates user participation and consumption on youtube?', *Computers in human behavior* **66**, 236–247.

Khan, M. T., Durrani, M., Ali, A., Inayat, I., Khalid, S. and Khan, K. H. (2016), 'Sentiment analysis and the complex natural language', *Complex Adaptive Systems Modeling* **4**, 1–19.

Kietzmann, J. H., Hermkens, K., McCarthy, I. P. and Silvestre, B. S. (2011), 'Social media? get serious! understanding the functional building blocks of social media', *Business horizons* **54**(3), 241–251.

Kim, H.-C., Pang, S., Je, H.-M., Kim, D. and Bang, S. Y. (2002), Pattern classification using support vector machine ensemble, *in* '2002 International Conference on Pattern Recognition', Vol. 2, IEEE, pp. 160–163.

Kim, J. (2012), 'The institutionalization of youtube: From user-generated content to professionally generated content', *Media, culture & society* **34**(1), 53–67.

Krouska, A., Troussas, C. and Virvou, M. (2016), The effect of preprocessing techniques on twitter sentiment analysis, *in* '2016 7th international conference on information, intelligence, systems & applications (IISA)', IEEE, pp. 1–5.

Kümpel, A. S., Karnowski, V. and Keyling, T. (2015), 'News sharing in social media: A review of current research on news sharing users, content, and networks', *Social media+ society* **1**(2), 2056305115610141.

Lange, P. G. (2007), 'Publicly private and privately public: Social networking on youtube', *Journal of computer-mediated communication* **13**(1), 361–380.

Laroche, M., Habibi, M. R. and Richard, M.-O. (2013), 'To be or not to be in social media: How brand loyalty is affected by social media?', *International journal of information management* **33**(1), 76–82.

LeCun, Y., Bengio, Y. and Hinton, G. (2015), 'Deep learning', *nature* **521**(7553), 436–444.

Li, N. and Wu, D. D. (2010), 'Using text mining and sentiment analysis for online forums hotspot detection and forecast', *Decision support systems* **48**(2), 354–368.

Liikkanen, L. A. and Salovaara, A. (2015), 'Music on youtube: User engagement with traditional, user-appropriated and derivative videos', *Computers in human behavior* **50**, 108–124.

Lin, J. (1991), 'Divergence measures based on the shannon entropy', *IEEE Transactions on Information theory* **37**(1), 145–151.

Liu, B. (2022), *Sentiment analysis and opinion mining*, Springer Nature.

Liu, Y. and Yao, X. (1999), 'Ensemble learning via negative correlation', *Neural networks* **12**(10), 1399–1404.

Louppe, G., Wehenkel, L., Sutera, A. and Geurts, P. (2013), 'Understanding variable importances in forests of randomized trees', *Advances in neural information processing systems* **26**.

Mahtab, S. A., Islam, N. and Rahaman, M. M. (2018), Sentiment analysis on bangladesh cricket with support vector machine, *in* '2018 international conference on Bangla speech and language processing (ICBSLP)', IEEE, pp. 1–4.

McAuley, J., Pandey, R. and Leskovec, J. (2015), Inferring networks of substitutable and complementary products, *in* 'Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining', pp. 785–794.

Medhat, W., Hassan, A. and Korashy, H. (2014), 'Sentiment analysis algorithms and applications: A survey', *Ain Shams engineering journal* **5**(4), 1093–1113.

Mejova, Y. and Srinivasan, P. (2012), Political speech in social media streams: Youtube comments and twitter posts, *in* 'Proceedings of the 4th Annual ACM Web Science Conference', pp. 205–208.

Mekouar, S., Zrira, N. and Bouyakhf, E.-H. (2017), Popularity prediction of videos in youtube as case study: A regression analysis study, *in* 'Proceedings of the 2nd international Conference on Big Data, Cloud and Applications', pp. 1–6.

Menard, S. (2002), *Applied logistic regression analysis*, Vol. 106, Sage.

Mikolov, T., Chen, K., Corrado, G. and Dean, J. (2013), 'Efficient estimation of word representations in vector space', *arXiv preprint arXiv:1301.3781* .

Mishra, S. (2019), Bridging models for popularity prediction on social media, *in* 'Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining', pp. 810–811.

Montgomery, D. C., Peck, E. A. and Vining, G. G. (2021), *Introduction to linear regression analysis*, John Wiley & Sons.

Muhammad, A. N., Bukhori, S. and Pandunata, P. (2019), Sentiment analysis of positive and negative of youtube comments using naïve bayes–support vector machine (nbsvm) classifier, *in* '2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE)', IEEE, pp. 199–205.

Mullen, T. and Collier, N. (2004), Sentiment analysis using support vector machines with diverse information sources, *in* 'Proceedings of the 2004 conference on empirical methods in natural language processing', pp. 412–418.

Multisilta, J., Suominen, M. and Östman, S. (2012), 'A platform for mobile social media and video sharing', *International Journal of Arts and Technology* **5**(1), 53–72.

Napoli, P. M. (2011), *Audience evolution: New technologies and the transformation of media audiences*, Columbia University Press.

Natarajan, T., Balakrishnan, J., Balasubramanian, S. A. and Manickavasagam, J. (2014), 'Perception of indian consumers towards social media advertisements in facebook, linkedin, youtube and twitter', *International Journal of Internet Marketing and Advertising* **8**(4), 264–284.

Nathan, E. (2022), 'The best times to post on social media in 2023: An analysis of more than 35 million posts [original research]', `https://coschedule.com/blog/best-times-to-post-on-social-media`. Available from CoSchedule Blog.

Ninaus, M., Greipl, S., Kiili, K., Lindstedt, A., Huber, S., Klein, E., Karnath, H.-O. and Moeller, K. (2019), 'Increased emotional engagement in game-based learning–a machine learning approach on facial emotion detection data', *Computers & Education* **142**, 103641.

Nisa, M. U., Mahmood, D., Ahmed, G., Khan, S., Mohammed, M. A. and Damaševičius, R. (2021), 'Optimizing prediction of youtube video popularity using xgboost', *Electronics* **10**(23), 2962.

North, A. C. and Hargreaves, D. J. (1999), 'Music and adolescent identity', *Music education research* **1**(1), 75–92.

Paltoglou, G. and Thelwall, M. (2012), 'Seeing stars of valence and arousal in blog posts', *IEEE Transactions on Affective Computing* **4**(1), 116–123.

Pang, B., Lee, L. and Vaithyanathan, S. (2002), 'Thumbs up? sentiment classification using machine learning techniques', *arXiv preprint cs/0205070* .

Pang, B., Lee, L. et al. (2008), 'Opinion mining and sentiment analysis', *Foundations and Trends® in information retrieval* **2**(1–2), 1–135.

Pareto, V. (1919), *Manuale di economia politica con una introduzione alla scienza sociale*, Vol. 13, Società editrice libraria.

Park, S. J., Lim, Y. S. and Park, H. W. (2015), 'Comparing twitter and youtube networks in information diffusion: The case of the "occupy wall street" movement', *Technological forecasting and social change* **95**, 208–217.

Peng, C.-Y. J., Lee, K. L. and Ingersoll, G. M. (2002), 'An introduction to logistic regression analysis and reporting', *The journal of educational research* **96**(1), 3–14.

Pennebaker, J. W., Boyd, R. L., Jordan, K. and Blackburn, K. (2015), 'The development and psychometric properties of liwc2015'.

Pennebaker, J. W., Francis, M. E. and Booth, R. J. (2001), 'Linguistic inquiry and word count: Liwc 2001', *Mahway: Lawrence Erlbaum Associates* **71**(2001), 2001.

Postigo, H. (2016), 'The socio-technical architecture of digital labor: Converting play into youtube money', *New media & society* **18**(2), 332–349.

Prasad, A. M., Iverson, L. R. and Liaw, A. (2006), 'Newer classification and regression tree techniques: bagging and random forests for ecological prediction', *Ecosystems* **9**, 181–199.

Qiu, Y. and Yang, B. (2021), Research on micro-blog text presentation model based on word2vec and tf-idf, *in* '2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)', IEEE, pp. 47–51.

Raschka, S. and Mirjalili, V. (2015), 'Python machine learning packt publishing ltd'.

Ratkiewicz, J., Menczer, F., Fortunato, S., Flammini, A. and Vespignani, A. (2010), Traffic in social media ii: Modeling bursty popularity, *in* '2010 IEEE second international conference on social computing', IEEE, pp. 393–400.

Röder, M., Both, A. and Hinneburg, A. (2015), Exploring the space of topic coherence measures, *in* 'Proceedings of the eighth ACM international conference on Web search and data mining', pp. 399–408.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M. and Chica-Rivas, M. (2015), 'Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines', *Ore Geology Reviews* **71**, 804–818.

Roma, P. and Aloini, D. (2019), 'How does brand-related user-generated content differ across social media? evidence reloaded', *Journal of Business Research* **96**, 322–339.

Rothschild, P. C. (2011), 'Social media use in sports and entertainment venues', *International Journal of Event and Festival Management* **2**(2), 139–150.

Rotman, D., Golbeck, J. and Preece, J. (2009), The community is where the rapport is–on sense and structure in the youtube community, *in* 'Proceedings of the fourth international conference on Communities and technologies', pp. 41–50.

Saberi, B. and Saad, S. (2017), 'Sentiment analysis or opinion mining: A review', *Int. J. Adv. Sci. Eng. Inf. Technol* **7**(5), 1660–1666.

Schäfer, T., Sedlmeier, P., Städtler, C. and Huron, D. (2013), 'The psychological functions of music listening', *Frontiers in psychology* **4**, 511.

Segerberg, A. and Bennett, W. L. (2011), 'Social media and the organization of collective action: Using twitter to explore the ecologies of two climate change protests', *The Communication Review* **14**(3), 197–215.

Shen, J. (2024), 'The research of the factors that influence the popularity of youtube videos', *Theoretical and Natural Science* **51**(1), 187–193.

Shevtsov, A., Oikonomidou, M., Antonakaki, D., Pratikakis, P. and Ioannidis, S. (2023), 'What tweets and youtube comments have in common? sentiment and graph analysis on data related to us elections 2020', *Plos one* **18**(1), e0270542.

Siersdorfer, S., Chelaru, S., Nejdl, W. and San Pedro, J. (2010), How useful are your comments? analyzing and predicting youtube comments and comment ratings, *in* 'Proceedings of the 19th international conference on World wide web', pp. 891–900.

Smith, A. N., Fischer, E. and Yongjian, C. (2012), 'How does brand-related user-generated content differ across youtube, facebook, and twitter?', *Journal of interactive marketing* **26**(2), 102–113.

Snelson, C. (2011), 'Youtube across the disciplines: A review of the literature', *MERLOT Journal of Online learning and teaching* .

Spasojevic, N., Li, Z., Rao, A. and Bhattacharyya, P. (2015), When-to-post on social networks, *in* 'Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', pp. 2127–2136.

Stevens, K., Kegelmeyer, P., Andrzejewski, D. and Buttler, D. (2012), Exploring topic coherence over many models and many topics, *in* 'Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning', pp. 952–961.

Stulp, F. and Sigaud, O. (2015), 'Many regression algorithms, one unified model: A review', *Neural Networks* **69**, 60–79.

Sundaramurthy, S., Saravanabhavan, C. and Kshirsagar, P. (2020), Prediction and classification of rheumatoid arthritis using ensemble machine learning approaches, *in* '2020 International Conference on Decision Aid Sciences and Application (DASA)', IEEE, pp. 17–21.

Sutton, C. D. (2005), 'Classification and regression trees, bagging, and boosting', *Handbook of statistics* **24**, 303–329.

Tang, J., Meng, Z., Nguyen, X., Mei, Q. and Zhang, M. (2014), Understanding the limiting factors of topic modeling via posterior contraction analysis, *in* 'International conference on machine learning', PMLR, pp. 190–198.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D. and Kappas, A. (2010), 'Sentiment strength detection in short informal text', *Journal of the American society for information science and technology* **61**(12), 2544–2558.

Thomas, M., Pang, B. and Lee, L. (2006), 'Get out the vote: Determining support or opposition from congressional floor-debate transcripts', *arXiv preprint cs/0607062* .

Thomas, S., Bestman, A., Pitt, H., Deans, E., Randle, M., Stoneham, M. and Daube, M. (2015), 'The marketing of wagering on social media: An analysis of promotional content on youtube, twitter and facebook'.

Thorson, K., Driscoll, K., Ekdale, B., Edgerly, S., Thompson, L. G., Schrock, A., Swartz, L., Vraga, E. K. and Wells, C. (2013), 'Youtube, twitter and the occupy movement: Connecting content and circulation practices', *Information, Communication & Society* **16**(3), 421–451.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **58**(1), 267–288.

Trilling, D., Tolochko, P. and Burscher, B. (2017), 'From newsworthiness to shareworthiness: How to predict news sharing based on article characteristics', *Journalism & mass communication quarterly* **94**(1), 38–60.

Vallet, D., Berkovsky, S., Ardon, S., Mahanti, A. and Kafaar, M. A. (2015), Characterizing and predicting viral-and-popular video content, *in* 'Proceedings of the 24th ACM International on Conference on Information and Knowledge Management', pp. 1591–1600.

Van Dijck, J. (2013), 'Youtube beyond technology and cultural form'.

Vernallis, C. (2013), *Unruly media: YouTube, music video, and the new digital cinema*, Oxford University Press.

Wallin, M. (2021), New media incentives: a cross platform analysis of social media discourse on 4chan, twitter and youtube, Master's thesis.

Wang, A. and Gao, X. (2019), 'Hybrid variable-scale clustering method for social media marketing on user generated instant music video', *Tehnički vjesnik* **26**(3), 771–777.

Wassermann, S., Wehner, N. and Casas, P. (2019), 'Machine learning models for youtube qoe and user engagement prediction in smartphones', *ACM SIGMETRICS Performance Evaluation Review* **46**(3), 155–158.

Wehrl, A. (1978), 'General properties of entropy', *Reviews of Modern Physics* **50**(2), 221.

Westenberg, W. (2016), The influence of youtubers on teenagers: a descriptive research about the role youtubers play in the life of their teenage viewers, Master's thesis, University of Twente.

Wu, S., Rizoiu, M.-A. and Xie, L. (2018), Beyond views: Measuring and predicting engagement in online videos, *in* 'Proceedings of the International AAAI Conference on Web and Social Media', Vol. 12.

Yan, M., Sang, J. and Xu, C. (2014), Mining cross-network association for youtube video promotion, *in* 'Proceedings of the 22nd ACM international conference on Multimedia', pp. 557–566.

Yan, M., Sang, J., Xu, C. and Hossain, M. S. (2015), 'Youtube video promotion by cross-network association:@ britney to advertise gangnam style', *IEEE Transactions on Multimedia* **17**(8), 1248–1261.

Yee, W. G., Yates, A., Liu, S. and Frieder, O. (2009), Are web user comments useful for search?, *in* 'LSDS-IR@ SIGIR'.

Yu, H., Xie, L. and Sanner, S. (2014), Twitter-driven youtube views: Beyond individual influencers, *in* 'Proceedings of the 22nd ACM international conference on Multimedia', pp. 869–872.

Yu, H., Xie, L. and Sanner, S. (2015), The lifecyle of a youtube video: Phases, content and popularity, *in* 'Proceedings of the international AAAI conference on web and social media', Vol. 9, pp. 533–542.

Zhang, S., Yao, L., Sun, A. and Tay, Y. (2019), 'Deep learning based recommender system: A survey and new perspectives', *ACM computing surveys (CSUR)* **52**(1), 1–38.

Zhou, R., Khemmarat, S. and Gao, L. (2010), The impact of youtube recommendation system on video views, *in* 'Proceedings of the 10th ACM SIGCOMM conference on Internet measurement', pp. 404–410.

Zhu, Z., Liang, J., Li, D., Yu, H. and Liu, G. (2019), 'Hot topic detection based on a refined tf-idf algorithm', *IEEE access* **7**, 26996–27007.

Zou, Y., Zheng, C., Alzahrani, A. M., Ahmad, W., Ahmad, A., Mohamed, A. M., Khallaf, R. and Elattar, S. (2022), 'Evaluation of artificial intelligence methods to estimate the compressive strength of geopolymers', *Gels* **8**(5), 271.