

# Classification of Arabic Real and Fake News Based on Arabic Textual Analysis

---

A Thesis Presented to

Department of Computer and Information Sciences

University of Strathclyde

---

in Partial Fulfillment

of the Requirements for the Degree

of Doctor of Philosophy

of Ph.D. of Computer Science (Ph.D.)

---

by

Hanen Tarik Hindi

June 10, 2022

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.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

# Abstract

The ease of communication that has been made possible by chat messaging platforms, and their increased use and ubiquity in society, have motivated purveyors of fake news to create and present their news as legitimate. Though many countries have introduced severe penalties for distributing fake news, monitoring the myriad articles involved has been burdensome. While different organisations have continued their efforts to resolve this problem, many of the solutions rely on verifying the associated metadata for further validation. In the case of text sent through social messaging, these metadata are not always present. Several studies have attempted to identify fake news by analysing the textual content of these pieces, however, there is a dearth of studies on Arabic language sources. This study fills that gap.

This research compiled a machine learning (ML) model that classifies real and fake articles in Arabic based on textual analysis. It is important not only for its development of the classification model but also because of the ability of the model to classify other types of fake news, such as satire and the article's country of origin. This work employed qualitative approaches to create five Arabic datasets that may be used for other research projects in Arabic. Then, through comprehensive textual analysis using Natural Language Processing (NLP) tools, quantitative approaches were used in several supervised ML classifiers.

This research thus puts forward a comprehensive supervised ML classification model that identifies fake news articles that are written in a formal journalistic

genre imitating real news articles. The novelty of this model lies in the fact that it classifies real and fake news articles in Arabic, with fake articles written in a journalistic style, which causes only minor differences between them and real articles. To examine these differences, four textual features were analysed—part of speech (POS), emotion, polarity, and linguistics—that have been successful in identifying fake news in other languages, but have not been fully tested in Arabic fake news. Probing these textual features showed how influential each of them was in identifying fake news in Arabic. With the aid of NLP to extract the textual features combined with ML classifiers, this research compiled a model that reached an accuracy score of 77.2%. Moreover, the model correctly predicted 6 out of 10 articles within the same topic domain, the Hajj, and 17 out of 26 fake articles within another topic domain, COVID-19.

The proposed model achieved promising results and it also successfully classified satire articles, as well as the articles' country of origin within the same topic domain. The research concludes by recommending making use of the contributions provided to conduct this research and to work more on this topic using new methods.

# Contents

<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Problem Statement . . . . .	4
1.2 Motivation . . . . .	6
1.3 Purpose Statement and Research Objectives . . . . .	7
1.4 Research Questions . . . . .	8
1.5 Research Contributions . . . . .	8
1.6 Thesis Organisation . . . . .	9
<b>2 Background and Related Work</b>	<b>11</b>
2.1 Definitions and Classification . . . . .	11
2.1.1 Real News . . . . .	12
2.1.2 Fake News . . . . .	13
2.1.3 Sensationalism and Lies . . . . .	14
2.1.4 Data Mining . . . . .	17
2.1.5 Machine learning . . . . .	17
2.1.6 Textual Analysis . . . . .	19
2.2 Characteristics of Deceptive Text . . . . .	20
2.2.1 Physiological Responses . . . . .	20

## Contents

2.2.2	The Imagination . . . . .	21
2.2.3	Persuasion . . . . .	22
2.3	Deceptive Text Detection Systems . . . . .	22
2.3.1	Explored Textual Features . . . . .	29
2.3.2	Arabic Fake News Detection Systems . . . . .	34
2.4	Research Gap . . . . .	38
2.5	Summary . . . . .	40
<b>3</b>	<b>Characteristics of Modern Standard Arabic (MSA)</b>	<b>41</b>
3.1	The Arabic Root-Pattern System . . . . .	42
3.1.1	Content Words . . . . .	45
3.1.2	Function Words . . . . .	49
3.2	Arabic NLP Challenges . . . . .	53
3.2.1	Orthographic Variations . . . . .	54
3.2.2	Lack of Capitalisation and Punctuation . . . . .	54
3.2.3	Homographs . . . . .	54
3.2.4	Lack of Arabic Lexicons . . . . .	55
3.2.5	Lack of Arabic NLP Tools . . . . .	56
3.3	Tasaheel Tool . . . . .	57
3.3.1	Creation of Wordlists . . . . .	59
3.3.2	Wordlist Revision . . . . .	62
3.3.3	Invoking Wordlists . . . . .	64
3.4	Summary . . . . .	64
<b>4</b>	<b>Methodology</b>	<b>65</b>
4.1	Research Paradigm . . . . .	65
4.2	Research Design . . . . .	67
4.3	Research Method . . . . .	70
4.3.1	Data Collection . . . . .	71

## Contents

4.3.2	Data Processing . . . . .	74
4.3.3	Data Analysis . . . . .	84
4.3.4	Experimental Tools . . . . .	93
4.4	Validity and Reliability . . . . .	95
4.5	Summary . . . . .	95
<b>5</b>	<b>Gold Standard Datasets</b>	<b>96</b>
5.1	Dataset 1: real_fake Dataset . . . . .	96
5.1.1	Real News Collection . . . . .	97
5.1.2	Fake News Generation . . . . .	105
5.2	Dataset 2: Satire Dataset . . . . .	110
5.3	Datasets 3 and 4: News Country-of-Origin Datasets . . . . .	112
5.4	Dataset 5: Covid Dataset . . . . .	113
5.5	Summary . . . . .	116
<b>6</b>	<b>Implementation</b>	<b>117</b>
6.1	Data Preparation . . . . .	117
6.2	Data Pre-processing . . . . .	117
6.3	Textual Feature Extraction . . . . .	121
6.3.1	Extraction of POS Features . . . . .	121
6.3.2	Extraction of Linguistic Features . . . . .	121
6.3.3	Extraction of Emotion and Polarity . . . . .	126
6.4	Datasets Preparation . . . . .	128
6.4.1	Primary Datasets Preparation . . . . .	128
6.4.2	Secondary Datasets Preparation . . . . .	128
6.5	Experimental Setup . . . . .	129
6.5.1	Microsoft Excel . . . . .	129
6.5.2	WEKA . . . . .	130
6.5.3	Orange . . . . .	132

## Contents

6.5.4	Participants for Human Evaluation . . . . .	133
6.5.5	Evaluation Metrics . . . . .	134
6.6	Summary . . . . .	135
<b>7</b>	<b>Experiments and Analysis</b>	<b>136</b>
7.1	Evaluation. Part 1: real_fake Classification . . . . .	136
7.1.1	Experiment 1: Testing Textual Feature Sets Individually .	137
7.1.2	Experiment 2: Testing and Evaluating All of the Features	138
7.1.3	Experiment 3: Testing and Evaluating Combined Features	139
7.1.4	Dataset Lexical Densities . . . . .	141
7.1.5	Important Features . . . . .	143
7.2	Evaluation. Part 2: Experiments Based on the Research Questions	144
7.2.1	Experiment 4: Satire_Nonsatire Classification . . . . .	145
7.2.2	Experiment 5: Country of Origin Classification . . . . .	146
7.2.3	Experiment 6: real_fake Classification Using Unseen (COVID) Dataset . . . . .	148
7.2.4	Experiment 7: Evaluation of Model against Human Perfor- mance . . . . .	150
7.2.5	Experiment 8: Model Evaluation With Stemming . . . . .	152
7.3	Summary . . . . .	154
<b>8</b>	<b>Discussion</b>	<b>155</b>
8.1	Objective 1: Compile an Arabic Fake News Dataset That Includes Real and Fake Articles in Journalistic Writing Style . . . . .	156
8.2	Objective 2: Investigate the Influence of Four Textual Feature Sets on Identifying Fake News in Arabic . . . . .	157
8.2.1	Emotion and Polarity Features . . . . .	157
8.2.2	Linguistic Features . . . . .	159
8.2.3	POS Features . . . . .	163



## Contents

8.3	Objective 3: Develop a Supervised Machine Learning Model That Classifies Real and Fake Articles in Arabic. . . . .	165
8.3.1	Research Question 1: How well does the proposed approach to classification model also classify another type of fake news, such as satire? . . . . .	165
8.3.2	Research Question 2: How well does the proposed approach to classification model be used for another classification task, such as classifying an article’s country of origin? . . .	167
8.3.3	Research Question 3: What is the Performance of the Proposed Model on Unseen Real and Fake Articles Distributed in the Real World? . . . . .	168
8.3.4	Research Question 4: How might the proposed model perform compared to humans in classifying real and fake articles?170	
8.3.5	Research Question 5: What is the effect of stemming articles on the model’s performance? . . . . .	171
8.4	Summary . . . . .	172
<b>9</b>	<b>Conclusion, Limitations, and Future Work</b>	<b>173</b>
9.1	Conclusion . . . . .	173
9.1.1	Arabic Fake News Datasets . . . . .	173
9.1.2	Arabic Wordlists . . . . .	175
9.1.3	Tasaheel Tool . . . . .	176
9.1.4	Model Compilation . . . . .	176
9.2	Limitations . . . . .	177
9.3	Future Work . . . . .	178
	<b>Appendices</b>	<b>179</b>
<b>A</b>	<b>Tasaheel Tool</b>	<b>180</b>
A.1	Part 1. . . . .	181

Contents

A.2 Part 2 . . . . .	185
<b>B Satire_Nonsatire &amp; Hajj_CO and Brexit_Co Lexical Densities</b>	<b>192</b>
B.1 Satire_Nonsatire Lexical Density . . . . .	192
B.2 Hajj_CO and Brexit_Co Lexical Densities . . . . .	194
<b>C Model Evaluation Supplements</b>	<b>198</b>
<b>D Ethical Approval</b>	<b>199</b>
<b>E Publications</b>	<b>210</b>
<b>Bibliography</b>	<b>211</b>
References . . . . .	211

# List of Figures

3.1	A Full Sentence in One Word. . . . .	45
3.2	Tasaheel GUI. . . . .	58
4.1	Details Regarding the Research Paradigms. . . . .	68
4.2	Research Design Steps. . . . .	71
4.3	Research Framework. . . . .	72
4.4	Posit Summary File. . . . .	83
4.5	Tools in Blue and Metrics in Orange, Used For Model Evaluation.	85
4.6	SVM Algorithm. . . . .	87
4.7	Logistic Regression Algorithm. . . . .	88
4.8	Random Forest Algorithm. . . . .	89
5.1	A News Article From the Ministry of Hajj and Umrah in Saudi Arabia Describing the Same Details as the Real Article. . . . .	101
5.2	The NO_RUMORS Platform Displays the Red Graphic as Fake Content and the Green Graphic as Its Content Verification. . . . .	103
5.3	The Annotation Scheme for the real_fake Dataset. . . . .	108
5.4	The Red Side Shows the Fake Article, and the Green Side Shows the original Article. . . . .	114
5.5	Dataset Creation Specifications. . . . .	116
6.1	An Example of a News Article About Brexit. . . . .	118

## List of Figures

6.2	Segmented Text. . . . .	119
6.3	POS Summary Output by Tasaheel. . . . .	122
6.4	StanfordNLP Tagged Text. . . . .	122
6.5	Query Place Result. . . . .	124
6.6	Query output sample. . . . .	125
6.7	Query Place Result . . . . .	126
6.8	Emotion, Polarity, and Linguistic Tagging in Tasaheel. . . . .	127
6.9	Summary of the Pre-Processing and Feature Extraction Tools. . .	127
6.10	Dataset's Worksheets. From Right to Left, The First Column Shows the Articles, Then the Article Class in Gray. The Green Columns Contains the Linguistic Categories, Pink Contains the Polarity, Yellow Contains the Emotions, and Blue Contains the POS. The Last Column Contains the Number of Words in Each Article. . . . .	130
6.11	Dataset in the Arrf File Format. . . . .	131
6.12	Sample of Word Cloud in Orange. . . . .	133
6.13	Dataset Compilation Details. . . . .	135
7.1	Word Cloud of the Unseen Dataset. . . . .	149
8.1	Lexical Densities For All Textual Features. . . . .	158
8.2	Sample of Fake Article Containing Justification, Assurance, Temporal, Spatial, and Illustration Terms. . . . .	161
8.3	Fake Article Contains Negative Polarity in Blue, Oppositions in Red, Then Positive Polarity in Orange. . . . .	162
8.4	Model's Performance With All Textual Features. . . . .	166
A.1	Tasaheel GUI. . . . .	182
A.2	A file Segmented using the Farasa Segmenter. . . . .	182
A.3	A File POS-Tagged by the StanfordNLP Tagger. . . . .	183

## List of Figures

A.4	A File POS Tagged by the Farasa Tagger. . . . .	183
A.5	Emotion, Polarity, and Linguistic Word Tagging Process. . . . .	184
A.6	Summary File With a Summary of All the Tag. . . . .	185
A.7	Tag Identification Example in the Farasa Format. . . . .	187
A.8	Tag Identification Example in StanfordNLP Format. . . . .	187
A.9	Result of the Query. . . . .	189
A.10	Excel Output. . . . .	191
C.1	Sample of real_fake Classification in WEKA. . . . .	198
C.2	Sample of Predicted Classes For test.arff in WEKA. . . . .	199
C.3	Sample of Satire_Nonsatire.arff Testing in WEKA. . . . .	199
C.4	Sample of Important Emotion Features of Covid using Chi-Square in WEKA . . . . .	200
C.5	Sample of Satire_Nonsatire.arff. . . . .	200
C.6	Sample of Hajj_co.xlsx File Showing the Articles, Country Classes 1,2,3, Emotion, Polarity, Linguistics, and POS Features. . . . .	201

# List of Tables

2.1	Fake News Classification by Type. . . . .	15
2.2	Samples of Linguistic Features from the Literature. . . . .	33
2.3	Summary of Datasets Used for Fake News Detection in Arabic. . .	39
3.1	Example of Verb وقى ‘Protect’ . . . . .	42
3.2	Arabic Root Pattern of كتب ‘Wrote’ With Three Vowels, Red Indicates the Short Vowels. . . . .	43
3.3	Influence of Affixes on the Word كتب ‘Wrote’. . . . .	44
3.4	Details of Affixes With The Root علم ‘Knowledge’, Affixes are in Red. . . . .	44
3.5	Compound Nouns. . . . .	46
3.6	Adjective Examples. . . . .	47
3.7	Verb Examples. . . . .	47
3.8	Adverb Examples. . . . .	48

## List of Tables

3.9	Verb and Noun Affiliation. Note: Red Indicates the Prefix, Blue Indicates the Suffix, and Green Indicates the Root. Fem.=Feminine, Masc.=Masculine. . . . .	49
3.10	Examples of Exception and Negation Particles. . . . .	50
3.11	Examples of Conjunctions. . . . .	51
3.12	Pronoun Types and Examples, Red Indicates the Pronoun. . . . .	52
3.13	Affixes as Function Words: Affixes are Colored in Red , Noun or Verb Colored in Green. . . . .	53
3.14	Emotion and Polarity Wordlists. . . . .	61
3.15	Lexicon Resources. . . . .	62
3.16	Linguistic Wordlists With the Number of Concordant Words Each Contains. . . . .	63
4.1	POS Tags Used in this Research, With Examples. . . . .	77
4.2	Emotion Words Used in This Research, With Examples. . . . .	78
4.3	Polarity Words Used in This Research, With Examples. . . . .	78
4.4	Linguistic Features Used in This Research, With Examples. . . . .	80
4.5	The Fake News Confusion Matrix. . . . .	90
5.1	Real News Collection Sources. . . . .	98
5.2	Real News Veracity Checking. . . . .	102
5.3	Real News Excerpts With Number of Articles (Excerpts) From Each News Agency. . . . .	104
5.4	Application of Rubin et al.,’s (2015) Guidelines. . . . .	106
5.5	Article Distribution and Submission Statistics. . . . .	107
5.6	Statistics of the real_fake Dataset. . . . .	109

## List of Tables

5.7	A Sample of Real Articles and Their Corresponding Fake Articles. Note the Red Text Represents the Manipulated Information, Whilst the Green Text Represents the Original Information in the Article. . . . .	109
5.8	Satire_nonsatire Dataset Distribution. . . . .	111
5.9	Example of Satire and Non-Satire Article. . . . .	111
5.10	Country-of-Origin Datasets Distribution. . . . .	113
5.11	Statistics of the Covid Dataset. . . . .	115
5.12	Sample Real and Fake Articles From the Covid Dataset. . . . .	115
6.1	Details of Constructed Datasets. . . . .	118
6.2	POS Features Categories. . . . .	123
6.3	Linguistic Features Categories and Number of Words Matched. . . . .	126
6.4	Emotion and Polarity Categories and Number of Words Matched. . . . .	127
6.5	All Textual Features Extracted. . . . .	129
6.6	Classifiers' Parameters. . . . .	132
7.1	Results of the Evaluation of the Individual Feature Categories. . . . .	137
7.2	Results of Evaluation of All Features in the real_fake Dataset. . . . .	139
7.3	Results of the Evaluation of Combined Features in the real_fake Dataset. . . . .	140
7.4	Lexical Densities of Feature Categories. . . . .	142
7.5	Important Features in real_fake. . . . .	144
7.6	Important Features in Covid. . . . .	144
7.7	Results of Evaluation of All Features in the Satire_Nonsatire Dataset. . . . .	146
7.8	Results of the Evaluation of the Country-of-Origin Dataset Based on the Hajj_CO and Brexit_CO Datasets. . . . .	147
7.9	RF Confusion Matrix of the Unseen Dataset. . . . .	148
7.10	Evaluation of Human Classification of the real_fake Dataset. . . . .	151



List of Tables

7.11 Results For Emotional Features With the Stemmed Dataset. . . .	153
A.1 List of NLP Packages. . . . .	181
B.1 Satire_Nonsatire Lexical Densities. . . . .	193
B.2 Lexical Densities of POS Tags For Hajj_CO and Brexit_CO Datasets.	196

# Acknowledgements

“When My servants ask the concerning Me, I am indeed close (to them): I listen to the prayer of every suppliant when he calleth on Me.” ( Al-Baqarah, 186) ...  
Alhamdu lillahi rabbi alAAalameen....

I dedicate this thesis to my parents, Sana and Tarik , I feel so honored and blessed to have you as my parents, and want to express my gratitude for your care and support over the years. Thank you for instilling me with a strong passion for learning and for doing everything possible to put me on the path to greatness. I will never forget the important values you have passed down to me and the sleepless nights you stayed to pray for my happiness and health. Thank you so much for believing in me and I owe my success to you.

Hatim, Supporting me each day, never letting me lose hope, never stop believing in me, Thank you for backing me throughout.

Fatin, Sameer, and Tarik you are the reason I wake up with a smile everyday, You are nothing less than a blessing from Allah. Heba and Ahmad, I’m thankful for the one constant lifelong relationship I possess – my steadfast relationship with you two.

Dr.George, I want to thank you for all you have taught me. The knowledge and wisdom you have imparted to me have been a great help and support throughout my study. I believe my success is at least in part due to your sincere support and mentorship. Thank you!

Dr.Fatimah and Dr.Hasanain, I cannot thank you enough for your mentorship over the years.

# Chapter 1

## Introduction

The digitalisation of communication, especially social media, has dominated the Arab world, as can be seen by the fact that 79% of people in the region now use social media or direct messaging at least once a day (Radcliffe & Abuhmaid, 2020). In recent years, the widespread use of social media (e.g., Facebook, Twitter) and chat messaging applications has changed how people communicate and it has affected their trust in the veracity of content. In this context, an estimated 55% of people in the Arab world use social media to look for news daily<sup>1</sup> and they share only those stories they believe are worth sharing with their selected network of friends.

However, despite its many advantages, news shared through social media or messaging platforms has its drawbacks. First, its dissemination is decentralised, raising many questions regarding its authenticity. In addition, the ubiquitous presence of bots on these messaging platforms dramatically increases the likelihood that news stories have not been written or vetted by journalists, but are sensationalised, emotionally charged news stories created by sophisticated AI programs that become fake news.

The concept of fake news gained sudden familiarity around the world during the US presidential election in 2016. It also became a common word usually ref-

---

<sup>1</sup><http://www.mideastmedia.org/survey/2017/chapter/social-media/#s225>

## Chapter 1. Introduction

erenced by President Donald Trump to express his disagreement with reporters<sup>2</sup>. Initially, the term was used to reference false information distributed through the news. However, fake news has evolved and become more adept at imitating the journalistic writing style of real news articles, which has given it multiple interpretations throughout the years. The authors in (Sharma et al., 2019) capture a broader scope of the term and define fake news as “a news article or message published and propagated through media, carrying false information regardless the means and motives behind it.” Fake news can thus be defined based on its intent as misinformation or disinformation (Afroz, Brennan, & Greenstadt, 2012). Misinformation has no harmful intent, while disinformation has malicious intent (Ireton & Posetti, 2018). Fake news may also have different names depending on its purpose (means). It may have a humorous purpose, leading it to be identified as “satire”, or it may be “propaganda” when it is a false exaggerated claim to support an institution (Collins, Hoang, Nguyen, & Hwang, 2020). Moreover, fake news can be referred to as “rumours” when it is unsupported information distributed with no actual values (Bondielli & Marcelloni, 2019). The overarching commonality, however, boils down to its incorrect facts and deception (Afroz et al., 2012). In this study, news articles containing any false information about events, numbers, objects, people, etc. are considered fake, while those containing only true information are considered real.

The dissemination of fake news has variously succeeded in disrupting organisations, damaging the reputations of individuals, and threatening democratic elections and other political processes (X. Zhou & Zafarani, 2018). Regardless of whether the news is spread via social media or websites, identifying fake news is the first step in eliminating or reducing its potentially harmful effects on individuals, companies, and governments. According to Newsguard<sup>3</sup>, a service that rates the credibility and transparency of web news content, most information

---

<sup>2</sup><https://edition.cnn.com/2017/10/08/politics/trump-huckabee-fake/index.html>

<sup>3</sup><https://www.zdnet.com/article/coronavirus-misinformation-is-increasing-newsguard-finds/>

## Chapter 1. Introduction

shared across social media about COVID-19 has fake news content. Because of this, the service has recently launched a Coronavirus Misinformation Tracking Centre, which lists misleading websites that spread false information about the virus, cause panic among the population, and undermine government efforts. As an illustrative example, in 90 days, posts on the US Centers for Disease Control and Prevention and the World Health Organization websites received 364,483 social engagements in the form of likes and retweets. However, according to the service, 76 US misinformative sites that published misleading coronavirus information received 52,053,542 social engagements in the same period. Considering the difference between the two levels of engagement, one can see that misinformation can have enormous, detrimental impacts. The Arab world is facing a similar situation (Alqurashi, Hamoui, Alashaikh, Alhindi, & Alanazi, 2021).

Harmful, fake content often affects governments and social communities, who then cite real news sources to debunk the fake stories. However, due to the massive amount of fake news, written by people and machine generated, it is not always possible to verify individual stories by manual fact checking (Nagoudi, Elmadany, Abdul-Mageed, Alhindi, & Cavusoglu, 2020). Indeed, almost 80% of US consumers reported having encountered fake news during the COVID-19 pandemic, highlighting the major extent of this issue<sup>4</sup>. Therefore, automated fake news detection is essential to protect society and instantly detect potential harm.

Since a significant part of fake news is text-based, it is likely that computational linguistics could generate a solution to detect it. One of the most common approaches in computational linguistics is a focus on reader-response cues related to fake news articles. For example, the metadata and comments sections of articles posted on social media have proven to be valuable contextual indicators that the main post comprises fake text (Yanagi, Orihara, Sei, Tahara, & Ohsuga,

---

<sup>4</sup><https://www.statista.com/topics/3251/fake-news/#dossierkeyfigures>

2020). Other features often evident in fake news articles, and therefore revealing characteristics of fake news, include overuse of emotion and an unprofessional tone (Preston, Anderson, Robertson, Shephard, & Huhe, 2021).

## 1.1 Problem Statement

Along with the wide spread of the COVID-19 virus, fake news about the virus also spread rapidly. Indeed, fake news spreads faster than real articles (Vosoughi, Roy, & Aral, 2018). The fake news associated with the virus caused a flurry of research studies to combat this issue. In fact, to our knowledge, prior to COVID-19, there were only three published articles in Arabic that investigated fake news in Arabic, but this number jumped to nine after COVID-19 appeared. Various claims concerning the pandemic, such as its origins as a biological weapon or home remedies such eating garlic to treat and prevent the virus inspired these studies (Alqurashi et al., 2021). Although some of these articles might seem harmless, there were ones that had a dangerous effect on people’s health, such as fake articles claiming that the vaccine was composed of pig parts, which caused many Muslims not to accept the vaccine<sup>5</sup>.

Fake news detection research has quickly proliferated in the English-speaking world (Atodiresei, Tănăselea, & Iftene, 2018; Faustini & Covoos, 2020; Shao, Ciampaglia, Flammini, & Menczer, 2016), Spanish (Posadas-Durán, Gómez-Adorno, Sidorov, & Escobar, 2019), and Brazil (Resende et al., 2019), where researchers have had access to robust datasets for years and have thus been steadily identifying diagnostic features across numerous linguistic parameters that could be used to detect fake news. However, research on this subject in the Arabic language has lagged for several reasons. First, Arabic lacks sufficient resources such as datasets that may be used for such a study. Existing resources, such as Arabic corpora,

---

<sup>5</sup><https://www.dw.com/en/pakistan-conspiracy-theories-hamper-covid-vaccine-drive/a-56853397>

## Chapter 1. Introduction

are limited in scope or not available to the public (Helwe, Dib, Shamas, & Elbasuoni, 2020). Second, Arabic’s rich morphology is ambiguous, leading to a lack of available Arabic Natural Language Processing (NLP) tools. Of the extant tools, three are widely known, Farasa (Abdelali, Darwish, Durrani, & Mubarak, 2016), Camel (Obeid et al., 2020), and Madamira (Pasha et al., 2014), but the latter two are not publicly available. Finally, some governments in the Arab world have banned fake news platforms and placed stiff penalties on fake news production. For example, Saudi Arabia enacted a sentence of three years in prison and a one million Saudi riyal fine as a penalty for fake news distribution, and further fired a Saudi media official over publishing an “incorrect” piece of information<sup>6</sup>. In this way, they repress, if not entirely eradicate, the types of collated sites that English-language researchers have found helpful in creating reliable benchmark datasets.

These limitations have caused a lack of research to combat fake news in Arabic. Research conducted to combat fake news in Arabic focusing on COVID–19 has all been on tweets (Alqurashi et al., 2021; Alsudias & Rayson, 2020). Research groups that have worked on studies to detect deceptive Arabic text relied on analysing text in tweets (Rangel, Rosso, Charfi, & Zaghouani, 2019; Alzanin & Azmi, 2019), online reviews (Saeed, Rady, & Gharib, 2022), and user comments on YouTube (Alkhair, Meftouh, Smaili, & Othman, 2019), which are all short texts with informal writing. The current study focuses on fake news articles written in a formal journalistic style that imitates the real articles without any metadata support. This means that we rely on analysing the fake content’s text only. This is the kind of fake news text usually distributed through social media or messaging platforms. In this context, it becomes more challenging to differentiate between real and fake content, since sources associated with the text, such as publishing source or author, are absent. These limitations, especially the lack

---

<sup>6</sup><https://gulfnews.com/world/gulf/saudi/saudi-media-official-fired-over-fake-news-1.84638099>

of NLP tools that support Arabic, have hindered research efforts to identify fake news based on analysing deceptive markers. For example, while superlatives have been investigated by Kapusta, Hájek, Munk, and Benko (2020) to identify fake news, they were never analysed in Arabic fake news. That is because superlatives in Arabic have a complex morphology that requires specific affixes which cause word ambiguity for NLP tools.

This thesis aims to classify real and fake articles in Arabic. The fake articles are identified by analysing specific deceptive markers. For that, it was essential to build a dataset of Arabic fake news to create a supervised Machine Learning (ML) model, as Arabic fake news datasets are limited. Therefore, this thesis involves harnessing the relatively recent innovative crowdsourcing method to clear the hurdles that have put off other researchers and produce a reliable dataset consisting of real and fake news articles in Arabic as a foundation for this research. Once we have established this subject-specific dataset, we use a textual analysis tool for Arabic to perform various NLP tasks. These features are further used to train the proposed model to classify real and fake content.

## 1.2 Motivation

The motivation behind this thesis is the negative impact fake news has on individuals in specific and on the governments in general. With the lack of helpful resources such as datasets and openly available fake news platforms in the Arab world to create datasets, it is challenging to study the nature of fake news and thus create models that target its distribution. As fake news has the aspect of being deceptive, for that, the approaches to identify deceptive text may help also identify fake news. To study the nature of fake news and thus create a model that identifies it, we rely on studies that targeted deceptive text in terms of their linguistic word use. Throughout the thesis, we present approaches and methods



that identified deceptive text, in an effort to relate them to identify fake news. The thesis focuses on analyzing the linguistic word use of real and fake articles to find linguistic cues that signal the presence of fake text.

### 1.3 Purpose Statement and Research Objectives

The primary goal of this research is to compile a classification model that classifies real and fake articles in Arabic based on textual analysis. The novelty of this research lies in its focus on Arabic fake news articles written in a formal journalistic style and its investigation of deceptive markers to identify fake text. In this context, the research objectives are as follows:

1. Compile an Arabic fake news dataset that includes real and fake articles in journalistic style.
2. Investigate the influence of four textual feature sets on identifying fake news in Arabic:
  - **Parts of speech:** The Part of Speech (POS) reflect the words that make up the sentences. These include verbs, nouns, adjectives, adverbs, prepositions, determiners, particles, conjunctions, pronouns, prepositions, and interjections.
  - **Linguistic markers:** These encode the overall meaning in a statement by linking between ideas or aspects present in the text. They are hedges, assurances, temporal and spatial words, exceptions, negations, illustrations, intensifiers, oppositions, justifications, and superlatives.
  - **Emotion features:** These interpret the overall feeling reflected in the article. They are anger, disgust, sadness, joy, fear, and surprise.
  - **Polarity:** These are indicative of the actions and feelings reflected in the article. They include positive and negative polarities.

3. Develop a supervised ML model that classifies real and fake articles in Arabic.

## 1.4 Research Questions

This study answers the following research questions:

1. **RQ1:** How well does the proposed approach to classification model also classify another type of fake news, such as satire?
2. **RQ2:** How well does the proposed approach to classification model be used for another classification task, such as classifying an article's country of origin?
3. **RQ3:** What is the performance of the proposed model on unseen real and fake articles distributed in the real world?
4. **RQ4:** How might the proposed model perform compared to humans in classifying real and fake articles?
5. **RQ5:** What is the effect of stemming articles on the model's performance?

## 1.5 Research Contributions

The major contributions of this study towards detecting fake news in Arabic are outlined below.

1. A benchmark dataset consisting of 549 real Arabic articles and 549 corresponding fake articles, written in a journalistic style.
2. A dataset consisting of 262 satire and 262 non-satire articles in Arabic. The former were collected from Arabic satirist platforms and the later from legitimate news platforms covering the same topic the satire article covered.

## Chapter 1. Introduction

3. A dataset consisting of 26 fake articles in Arabic about COVID-19 from fact checking platforms. The dataset also includes 26 real articles about COVID from verified news platforms.
4. A dataset consisting of 694 Saudi, 694 Egyptian, and 685 Jordanian articles focusing on the topic of the Hajj.
5. A dataset consisting of 486 Saudi, 481 Egyptian, 485 Jordanian articles focusing on the topic of Brexit.
6. A set of specific Arabic linguistic wordlists that include assurance, negations, oppositions, justification, hedges, intensifiers, illustration, exclusion, spatial, and temporal words.
7. The design of an NLP prototype that supports Arabic text. It carries out major NLP tasks, including segmentation, lemmatisation, and POS tagging. However, its novelty lies in its ability to perform emotion, polarity, and linguistic tagging.
8. A developed ML model that classifies real/fake news in Arabic based on training it on four textual features (POS, emotion, polarity, and linguistics).

## 1.6 Thesis Organisation

This thesis is organised as follows. Chapter 1 introduces the problem statement, motivation, objectives, and the research methodology. Chapter 2 provides background information about news and its types, challenges encountered in fake news detection, and related works in deceptive detection systems. Chapter 3 gives an introduction to the Arabic language, along with challenges in Arabic language processing. An Arabic NLP prototype is presented to solve these challenges. The details of the proposed methodology presented in this study are provided in

Chapter 1. Introduction

Chapter 4. Details of all five datasets compiled in this research are explained in Chapter 5, while Chapter 6 presents the preparation and experimental setup of our work. Chapter 7 presents the experiments conducted to evaluate the model and provides answers to the research questions. A thorough discussion of the research objectives and research questions comprises Chapter 8. Finally, the conclusion and suggestions for future work are presented in Chapter 9.

# Chapter 2

## Background and Related Work

This chapter considers the definitions and classifications of news. It also identifies principal elements in this research. The chapter illustrates the challenges of detecting fake news and provides insights into the characteristics of deceptive text. Additionally, it discusses related works on news detection systems and fake news, with a section on related works in Arabic. The chapter concludes by identifying the research gap.

### 2.1 Definitions and Classification

According to J. Fuller (1996), ‘news is a report of what a news organization has recently learned about matters of some significance or interest to the specific community that the news organization serves.’ Gans (2004) defined news as ‘information which is transmitted from the source to recipients by journalists who are both employees of bureaucratic, commercial organizations and also members of a professional group.’ Although the news itself might not be attractive to everyone, news writers (journalists) strive to portray the content in a manner that attracts a broad audience. Thus, the way news content is handled and presented in the media must further enhance its appeal in order to keep those broad audiences. Many journalists adopt unique writing and presentation styles

## Chapter 2. Background and Related Work

to make the news more appealing to gain more readers. News may be presented in the form of text, videos, or images supported by some explanatory text. As each piece of news is associated with an author and an agency, the news provided in a different media is supervised by the news agency's editorial team, rendering the news reasonably reliable<sup>1</sup>.

Unlike the real news produced by news agencies, which usually use an oversight process before publishing their articles, fake news can be written either by humans (Alqurashi et al., 2021; Elhadad, Li, & Gebali, 2020; Golbeck et al., 2018) or generated by computers (Nagoudi et al., 2020). Fake news can be published anywhere and in many cases, without monitoring<sup>2</sup>. Moreover, according to Torabi Asr and Taboada (2019), fake news can be written by amateurs, students, or professional journalists<sup>3</sup>. The intent of producing fake articles varies and can range from financial gain from advertisements to promoting political propaganda and even just for fun (Aldwairi & Alwahedi, 2018).

The following sections provide further detail about real and fake news.

### 2.1.1 Real News

According to Merriam-Webster, the word real is defined as 'genuine' or 'not artificial.' This is a simple, general definition of real that can be applied to the media to define genuine news content. However, there is an ongoing public debate on what is considered real news, which is at least partly the result of the challenge to authenticate news content without actually witnessing the event oneself. In this context, to generate a sense of credibility among the public, news outlets have, over the years, aimed to emphasise the veracity of their content by illustrating it with images and videos. For example, news articles detailing the medical crisis

---

<sup>1</sup><https://www.preservearticles.com/journalism/what-are-the-role-of-editorial-department-of-a-newspaper/15717>

<sup>2</sup><https://www.bbc.com/future/article/20190528-i-was-a-macedonian-fake-news-writer>

<sup>3</sup><https://www.theguardian.com/world/2018/dec/23/anti-america-bias-der-spiegel-scandal-relotius>

## Chapter 2. Background and Related Work

the COVID-19 pandemic is causing post images or links to videos of packed hospitals with ill patients. However, even with this effort to gain the public’s trust, people tend to believe what fits their perspectives (Preston et al., 2021).

While there is no fool-proof method for verifying news articles, some approaches have been developed to help ensure veracity. For example, Pérez-Rosas, Kleinberg, Lefevre, and Mihalcea (2017) proposed a technique that involves cross-referencing information in articles between several sources to verify the article’s legitimacy. Zhang et al. (2018) proposed the so-called Credibility Coalition, a developing framework that uses indicators as signals to determine whether content is credible. These indicators might be clickbait or headlines, and, in some cases, they may be embedded in the publisher’s metadata in the form of revenue and the presence of ads. Using these indicators may create an accurate assessment of an article’s credibility.

### 2.1.2 Fake News

According to Merriam-Webster, the word fake means ‘not true’ or ‘a worthless imitation passed off as genuine.’ As a verb, to fake is defined as ‘to alter’ or ‘to manipulate.’ However, when associated with news, the term fake news has a much deeper meaning than simple manipulation. In fact, (Lazer et al., 2018, p. 2) described it as ‘fabricated information that mimics news media content in form but not in organizational process or intent. Fake news outlets, in turn, lack the news media’s editorial norms and processes for ensuring the accuracy and credibility of information.’ Generally, fake news can be defined as follows: ‘Fake news is information presented as news, but which contains incorrect information designed to deceive the reader into believing the information is correct’ (Molina, Sundar, Le, & Lee, 2021, p. 181). The salient feature of this definition is the deceptive nature of fake news. The goal of real news is to report what is happening or has happened, even though in some cases these reports might be biased. Fake

## Chapter 2. Background and Related Work

news uses the same style, however, it twists the facts in a way that provokes distrust and factionalism (Bondielli & Marcelloni, 2019). It is important to note that these twisted facts may not be easy to detect, because fake news is crafted to sound legitimate while evoking strong emotions in the reader (Cartwright, Nahar, Weir, Padda, & Frank, 2019). In fact, fake news is created in a manner that leads these articles to be disseminated six times faster than real articles (Fox, 2018). Because of that, this research advances the argument that there is a perceptible difference between real news and fake news and that these differences might serve as markers of deceptive content in the text.

### **Classification of Fake News**

Fake news is generally considered an umbrella term for any false statements that are portrayed as news (Wu, Morstatter, Carley, & Liu, 2019). There have been several attempts to classify types of fake news based on different criteria. Collins et al. (2020) classified fake news into five types, based on its form: clickbait, propaganda, satire, hoaxes, and name theft. However, Fallis (2014) classified it based on intention, as either misinformation or disinformation. Vosoughi, Mohsenvand, and Roy (2017) added rumour to this classification due to its deceptive intention. Table 2.1 below provides definitions for each of these classifications.

### **2.1.3 Sensationalism and Lies**

Fake news producers try to convince readers of their content's legitimacy by closely mimicking the writing styles of legitimate websites and article genres (Tandoc Jr et al., 2018). The fabrication is challenging to verify, because once readers accept the source's legitimacy, their ability to judge credibility levels is suspended, leading them to trust the source without further verification. Another feature of fake news is sensationalism, which tries to hide deception by exploiting emotional material. According to Kilgo, Harlow, García-Perdomo, and Salaverría



Table 2.1: Fake News Classification by Type.

Name	Meaning
<b>Clickbait</b>	Clickbait is an eye-catching headline of a made-up story intended to lure the reader into clicking the link. Clicks generate income for the link owner on a pay-per-click basis and the clickbait stories usually consist of gossip unrelated to the headline (Aldwairi & Alwahedi, 2018).
<b>Propaganda</b>	Propaganda consists of news stories created by a political entity to influence public perceptions. They are typically used to report false information to the public to support a certain view (L. Wang, Wang, De Melo, & Weikum, 2018).
<b>Hoaxes</b>	Hoaxes are intentionally fabricated reports that attempt to deceive the public (Shao et al., 2016).
<b>Satire (irony)</b>	Satire presents stories that may or may not be factually incorrect as news, not to deceive, but rather to expose or ridicule bad or shameful behaviour (Saadany, Mohamed, & Orasan, 2020).
<b>Misinformation</b>	Misinformation refers to unintentionally fabricating information (Fallis, 2014).
<b>Disinformation</b>	Disinformation is news that contains misleading information with misleading intent (Fallis, 2014).
<b>Rumours</b>	Rumours are unverified allegations that start from one or more sources and spread from one node to another over time (Bondielli & Marcelloni, 2019).
<b>Parody</b>	These are articles that contain fabricated information presented in a humorous way (Tandoc Jr, Lim, & Ling, 2018).

(2018), sensationalism highlights emotional or dramatic events to arouse the audience's emotions or attract attention. Though not all fake news articles adopt a sensationalist style, it is worth noting that some writers do use this technique to increase user engagement. A study by Tenenboim and Cohen (2015) found that the use of sensationalism led to more online engagement on a popular website. Along the same lines, fake news also contains lies and is considered a form of manipulation of information. Metts (1989) specified three types of lies. The first is falsification, which includes providing information that is untrue. The second is distortion, which manipulates true information through exaggeration, minimisation, or shuffling of words to delude the reader into misinterpreting the information provided. The third is omission, which withholds relevant informa-

## Chapter 2. Background and Related Work

tion.

To understand the reason people fall victim to the lies in deceptive text in general, a study on deceptive text in phishing emails was conducted by Watters (2009). He studied the psychology of users who fall victim to phishing emails. Though deceptive text in phishing emails is different from text in fake news written in a journalistic style, the deceptive intention in both fake news and phishing emails is similar. They both also utilise electronic media. Thus, the findings of this study offer a useful insight into the nature of deceptive text and some of the signals that might identify it. The study reveals some of the signs that should alert users to potential phishing emails, including the following: (a) spelling mistakes, such as the misspelt name of the user's bank in the subject line; (b) an embedded URL different from the one in the HTML code; and (c) being asked for information that the genuine sender would never require, such as bank login credentials. Although these seemingly obvious discrepancies are useful in detecting phishing emails, many users process emails at too shallow a level. Due to this lack of thorough processing, many people fail to read the content of a phishing email carefully, including comparing the phishing link to the actual link of the service in question, preventing them from distinguishing between phishing email content and more reliable resources.

The lack of processing that many people exhibit in regard to phishing emails is also prevalent in the fake news domain and those who consume it. The findings by Watters (2009), which was a thorough qualitative study that focused on how and why readers believe deceptive text in phishing emails, was supported by Preston et al. (2021). They found that participants recognised fake news items through overly emotive language, a lack of supporting data, an unprofessional subjective tone, and unprofessional graphs and visuals. However, those who accepted the fake news item indicated that the items fit their existing beliefs, related to their personal experience, and the supporting visuals, data, and graphs showed good

support for the item. They also stated that the article's content shed light upon a hidden problem for which they may also have agreed.

No matter how fake news producers fabricate stories, they all try to lure readers into believing their content by using tactics that appear as linguistic markers in the text. To better relate to the problem of fake news in this research, we define important terms related to this research as they are main elements that build our supervised ML model to classify real and fake articles in Arabic.

### **2.1.4 Data Mining**

Large quantities of data are consumed every day. Many projects, in fields such as ML and data mining, have been developed to make use of this data. Although the terms ML and data mining are similar because they both analyse data, and are often used interchangeably, they are not identical. Data mining concerns methods of analysing structured data and often presents approaches that focus on commercial applications. In data mining, several methods such as data clustering and data classification are applied. Extracting patterns from data might be useful to get a better insight into the behaviours of the data, therefore training models on this data makes the model capable of addressing any potential issues. Machine learning encompasses the design and evaluation of algorithms for extracting these patterns from data. It involves experimenting with various statistical and computational techniques to process data in order to describe patterns (Kelleher & Tierney, 2018).

### **2.1.5 Machine learning**

As technologies improve, demand rises to acquire, access, store, and process enormous amounts of data. Earlier, we defined ML as similar to data mining; however, ML, along with artificial intelligence, has been increasingly employed to automate

## Chapter 2. Background and Related Work

human performances (D’Alconzo, Drago, Morichetta, Mellia, & Casas, 2019). Artificial intelligence is the logic that enables computers to perform human actions using their ‘minds.’ Examples include reasoning, planning, prediction, and perceptions. Machine learning, on the other hand, is programming that uses training data to optimise performance criteria.

Machine learning models are constructed through defined parameters, and their actions are performed after being learned. There are two learning algorithms, commonly categorised as supervised and unsupervised. As the name suggests, supervised learning works by mapping input data to output data, through training on labelled values (classes). The technique has been widely used in author alias identification (Layton, Watters, & Dazeley, 2015), fake tweet detection (Patel, Padiya, & Singh, 2022), and fake news detection (Tian, Zhang, & Peng, 2020). In contrast, unsupervised learning works by finding similarities between patterns of data with only input data and no labelled values (Zincir-Heywood, Mellia, & Diao, 2021). Usually, unsupervised learning uses a repetitive method to perform data analysis and draw conclusions based on that. Unsupervised learning also entails neural networks, which are networks that analyse training data models and find small connections between variables. After training, they use the knowledge gained to interpret new data. The algorithms usually need massive amounts of training data, so they are suitable for big data classification tasks (Rouse, 2019).

Supervised learning can be divided into classification and regression. In classification, the supervised learning task trains the model on predefined labels. The model then uses the knowledge gained through training to predict incoming data and produce output data with defined labels, usually as discrete values (for example 0 or 1). Regression has the same learning mechanism; however, the outputs are not discrete but continuous values. For example, if a model is to use the classification method to predict whether a customer will purchase a particular

book based on training data specified by customers—such as age, gender, and education level—the output would predict whether the customer would purchase (1) or not purchase (0). On the other hand, a suitable example for the regression model is a prediction of how much an electricity bill will be based on different parameters such as daily voltage usage, power consumption, and energy demand (Ghorui, 2019).

### 2.1.6 Textual Analysis

Textual analysis is a set of methods used to describe and interpret characteristics of a text by extracting information from textual sources (Loughran & McDonald, 2016). It requires the researcher to closely analyse the textual content of an item rather than the structure of that item. The concept of textual analysis is largely conducted by alluring NLP capabilities. Natural language processing refers to the techniques that enable the researcher to extract information from textual sources to perform data analysis (Pandey & Pandey, 2019). Because of that, the field of NLP has been explored by many researchers that aim to automate the extraction process of useful textual items by designing NLP tools for this service (Abdelali et al., 2016; Obeid et al., 2020; G. R. Weir, 2009).

Textual analysis dates back to 1969, when Walker (1969) investigated its usefulness for extracting rich information and proposed a computer system for performing computational linguistic analysis. More recently, with the rise of various NLP tools and ML classifiers, research on textual analysis has been adopted in accounting (Gandía & Huguet, 2021), stock investments (McGurk, Nowak, & Hall, 2020), and identifying Arabic conspiracy theories on Twitter (Al-Hashedi et al., 2022). Textual analysis has also been used to minimise manual efforts needed to analyse qualitative data. A framework by Brown and Collins (2021) was designed to focus on specific linguistic and artistic elements which is based on textual analysis conceptual theory. Textual analysis is not only used in the

## Chapter 2. Background and Related Work

above-mentioned research domains, but also in cultural studies, mass communication, media studies, philosophy, and sociology, to name a few. It is relevant to these fields specifically because they are related to human behaviours and ways of communicating and coexisting (McKee, 2003). Because textual analysis provides the writer's perspective by analysing their written text, it may also enable the finding of deceptive markers used by the writer (Jeronimo, Marinho, Campelo, Veloso, & da Costa Melo, 2019; Penuela, 2019). For example, with regards to the current research, it is more convenient to focus on certain deceptive markers in a text to predict the content's veracity than to attempt to analyse the whole text to trying to find nuanced differences between it and a non-fake text. In fact, textual analysis with the aid of NLP and ML has been employed in research to detect fake news (G. Weir, Owoeye, Oberacker, & Alshahrani, 2018; Cartwright, Weir, & Frank, 2019; Khanam, Alwasel, Sirafi, & Rashid, 2021).

## 2.2 Characteristics of Deceptive Text

Earlier in this chapter, we showed that fake news may intend to mislead, as in the case of disinformation, or not, as in misinformation. However, the thing all deceptive text has in common is the deceptive nature of their content. The most common way to address deceptive elements in fake news content is to rely on examples that contain explicit falsehoods in their text or use identified deceptive text found in other forms, such as tweets, online reviews, and spam emails.

### 2.2.1 Physiological Responses

Verbal cues have always been a focus of deception detection. One of the cues that deceivers might reveal is the fear of being caught. This can be exploited by looking at the use of assurance words in the deceptive context, which might stem from a feeling of needing to hide something and thus involve an element of fear.

## Chapter 2. Background and Related Work

Early studies on deception detection have investigated some textual features, such as the study by Pennebaker and King (1999). Their findings demonstrated that a deceptive text might have a higher frequency of assurance terms than a standard informative text, as these terms are used to assure the reader of the truth of a statement. In this context, assurance words are generally weak oaths that speakers add to statements in a context where lying is either expected or where the stakes are high. The high stakes produce anxiety, which results in the use of assurance words. In other words, in the context of a law enforcement interrogation, for example, both honest and dishonest subjects would be expected to use assurance words because of the stakes involved in the situation. When assurance terms appear in other environments, such as news stories, this can be a case of ‘leakage,’ resulting from the fear of being caught (Mihalcea & Strapparava, 2009) .

Guilt is another emotion elicited by deceptive acts. The sense of guilt can also create emotional leakage that manifests in more emotional expressiveness on the part of deceivers as compared to truth tellers (Burgoon, Blair, Qin, & Nunamaker, 2003; Hancock, Curry, Goorha, & Woodworth, 2007). In this context, Newman, Pennebaker, Berry, and Richards (2003) argued that it is not just the expression of emotion in general that characterises deceptive speech, but specifically negative emotion. Either way, since news stories are generally meant to be informative without conveying high levels of emotion, the presence of emotional words may be indicative of a fake news story (Torabi Asr & Taboada, 2019).

### **2.2.2 The Imagination**

Deception relies on the imagination rather than on memory (Ott, Choi, Cardie, & Hancock, 2011). Consequently, careful fabricators of fake news attempt to anchor their deceptions in real items or memories that they can easily reference (Berezenko, 2018). However, imaginative language is different from descrip-

tive language, even in written texts. With the aid of computational linguistics, these differences become even more apparent. Rayson, Wilson, and Leech (2002) showed that adverbs were more common in imaginative writing, and comparatives and superlatives were more profuse in informative writing (Rayson et al., 2002).

### **2.2.3 Persuasion**

Simply put, fake and real news have two different objectives. Journalists of real news provide factual information to the best of their ability about an event and place it in a broader context for understanding. Their audience wants to be not only informed about the details of a specific event but also to understand its causes and implications. On the other hand, fake news purveyors use speculation to describe events that feed into the fears and biases of their target audience to persuade them that the fake articles are legitimate and rooted in fact. This subtle, deceptive persuasion appears in the language of fake news articles. In this context, Baptista and Gradim (2020) verified that fake news uses persuasive language to convince the reader of the legitimacy of the text, when encountering any unreliable information in the text.

## **2.3 Deceptive Text Detection Systems**

Many studies have focused on social media and thus rely heavily on paratextual information (emojis, hashtags, and comments), which is often platform specific, for clues about the veracity of the news posting (Alkhair et al., 2019; Manzoor & Singla, 2019; Krishnan & Chen, 2018). The aim of the current study is to conduct an accurate content assessment that is applicable across all platforms and contexts. Previous work on deception detection has been investigated, as this includes many forms of deceptive text such as that found in messages, opinions,



## Chapter 2. Background and Related Work

tweets, and emails. We have reviewed the various techniques employed to detect deceptive text for use in our research. These works are related to our objective as they reflect deceptive text that has been written with the intention of sounding authentic.

Depending on the features studied, most current work on deceptive text detection generally falls under one of two general approaches: the bag-of-words (BoW) approach and the textual features approach. In the BoW approach, the feature vector for each document is computed based on word frequencies. In contrast, the textual features approach relies on analysing textual features from the text to capture the nuanced differences between real and deceptive (fake) writing styles. In this section, we present studies that employ BoW and its related techniques, including term frequency, word embedding and textual features approaches, where textual features are usually categorised into POS, linguistics, emotion, and polarity in the text.

Cartwright, Weir, and Frank (2019) published pioneering work that combated disinformation. Their objective was to develop a method for identifying hostile disinformation endeavours in the cloud. By utilising the International CyberCrime Research Centre’s (ICCRC) Dark Crawler (Zulkarnine, Frank, Monk, Mitchell, & Davies, 2016), Strathclyde’s Posit toolset (G. R. Weir, 2009), and TensorFlow (Abadi et al., 2016), they were able to classify posts as real or fake with an accuracy of 90.12% and 89.5% using Posit and TensorFlow, respectively. They used Dark Crawler, a web crawling software tool that captures web content from the open and the Dark Web, to download 90,605 real news tweets from multiple news webpages. They added these tweets to the dataset and labelled them as ‘real news.’ They also included structured content from online discussion forums and social media platforms. To adequately extract features from the text, the Posit analysis toolset was applied to a set of 5,000 tweets (2,500 real news and 2,500 fake news tweets). Features generated by Posit included values for to-

## Chapter 2. Background and Related Work

tal words, total unique words, number of sentences, noun types, verb types, etc., 27 features in all. The best performance of the Posit classification was 90.12% correct. TensorFlow is a ML system that enables deep learning networks. To build the TensorFlow model, a dataset of 2,709,204 tweets was created by merging multiple datasets. This tool resulted in an accuracy of 89.5%. Its drawback for our work is that it is unknown how well these tools perform with Arabic text.

A further study (Cartwright, Nahar, et al., 2019) extended the techniques of the previous study to develop a toolkit that mobilises artificial intelligence to identify hostile disinformation activities. The researchers analysed a wide sample of social media posts that fall in the ‘fake news’ category and were disseminated by the Russian Internet Research Agency. They used the tools mentioned above—Dark Crawler, Posit, and TensorFlow—as well as SentiStrength (Mei & Frank, 2015) and LibShortText (W. Y. Wang, 2017). SentiStrength is a textual analysis tool which assigns positive or negative values to lexical units in the text. LibShortText is an open-source software package that produces superior results when classifying shorter textual items such as tweets (W. Y. Wang, 2017). SentiStrength yielded an acceptable classification accuracy of 74.26%, while LibShortText generated a classification accuracy of 90.2%.

The increased use of social media and the growing ability of deceivers to pass off their writings as legitimate has led to further research. Innovative approaches, such as deep learning (Keya et al., 2021), have contributed to promising automation models. Though their model achieved high accuracy rates (98.71%) in detecting Bengali fake news, deep learning needs a large amount of data for training, which is not feasible in our study. Turning to supervised ML models, the work by Siagian and Aritsugi (2020) proposed a supervised ML method that detects deceptive and truthful pieces of text. The authors explored a combination of word and character n-grams (continuous sequence of N words), function words, and word syntactic n-grams as features to train the classifier. Through sev-

## Chapter 2. Background and Related Work

eral experiments that included testing and evaluating their model on five different datasets, all of which included deceptive and truthful text from different domains, their results indicated that word and character n-gram features performed well in detecting deceptive text. The same researchers Siagian and Aritsugi (2020) conducted another study that exploited function words as features for classifying deceptive opinions. Their results revealed the effectiveness of this feature in detecting deceptive opinions.

A study by Sharma et al (2021) created a binary classification model that classifies actual versus fake articles. Their dataset included actual and fake articles about Covid from Twitter, Facebook, and Reddit. Using three machine learning classifiers – Decision Tree, Bidirectional Long Short Term Memory, and Support Vector Machine. Bidirectional Long Short Term Memory algorithm achieved the highest accuracy, 76.60%. In the same effort to combat fake news related to Covid19, work by Kanaan et al (2021) was conducted. Their dataset included 8195 fake articles and balanced with 7372 real articles about Covid-19. At first they conducted word vector representations such as TF-IDF and Glove. Then, trained their model using six machine learning algorithms namely, Naive Bayes, XGBoost, Recurrent Neural Network, Long Short Term Memory, Logistic Regression, Support Vector Machine, and Random Forest. Through training, every model achieved accuracies above 90%.

Examining other features associated with the textual content was produced by Aphiwongsophon and Chongstitvatana (2021). They collected 948,373 messages posted on Twitter and extracted twenty two attributes from each message. The attributes included: Id, Name, Is Verified, FollowersCount, Location, etc. Training their dataset using three machine learning algorithms, namely: Naive Bayes, Neural network, and Support Vector Machines; their model achieved high accuracies for all algorithms above 95%. An application-based model was created by Shete et al (2021) to identify news as ‘real’ or ‘fake’. Their application made

## Chapter 2. Background and Related Work

use of Logistic Regression algorithm to identify fake articles from real ones. The TF-IDF vectorizer was used to convert the articles into feature indexes in the matrix for further analysis by the Logistic Regression algorithm. Through this process, their application successfully reached an accuracy percentage of 80 % .

With the rise of online platforms that offer user opinions and reviews, such as Yelp and TripAdvisor, came a rise in fake reviews and studies to detect these reviews. One such study is by Layton, Watters, and Ureche (2013), which used authorship analysis to expose fraudulent hotel reviews. This approach extracts and analyses textual attributes—including the author’s age, gender, and native language—which can be used to identify a particular writer. Based on this technique, it is possible to conclude whether the author of one review is the same as the author of another. Thus, Layton et al. (2013) created a dataset of hotel reviews by choosing several prolific authors and then extracting all their reviews. The dataset was used for training and, with several local n-gram methods, as the buttress for their study. The findings showed that this method effectively determined fraudulent reviews by identifying whether two reviews matched more than 66% of the time.

Twitter is a social networking and microblogging service where users can post and share their thoughts as tweets. Tweets are short messages with a maximum length of 280 characters. The service provides a rich platform for data collection and data analysis using tweets as data. Therefore, many studies have relied on extracting the data available in Twitter to analyse deceptive text. Conceptually in the same vein as the current research, Jardaneh, Abdelhaq, Buzz, and Johnson (2019) proposed a ML method to identify fake Arabic tweets based on the supervised classification model. To filter out non-credible tweets, they extracted two sets of tweet features: content-related and user-related features. Content-related features are related to the tweet’s content, such as the number of retweets, symbols, and words, as well as the sentiment score of tweets. User-related features

## Chapter 2. Background and Related Work

were derived directly from the tweeter’s profile. Some features include average number of hashtags, average tweet length, and average number of URLs. They used a dataset of 1,051 credible and 810 non-credible tweets regarding the Syrian crisis. During the learning process, several ML algorithms, including random forest and logistic regression, were applied. Their system reached an accuracy score of 76% in classifying credible and non-credible tweets.

Mouty and Gazdar (2018) surveyed several studies that targeted the detection of truth in Arabic tweets. Their survey explored the works of Al-Hussaini and Al-Dossari (2016), who built a reputation for detecting deceptive tweets using a sentiment lexicon. The survey also included the research of Floos (2016), who analysed important words in a document, using Term Frequency-Inverse Document Features (TF-IDF). Term Frequency-Inverse Document Features are commonly used to measure how frequently a word appears in the document, which shows its importance. Relying on this technique, Floos (2016) studied the content of tweets to determine whether they were news or rumours.

Deceptive text can also be found in published research papers. Markowitz and Hancock (2014) investigated the Diederik Stapel fraud case. Stapel was a social psychologist who was discovered to have published fraudulent papers throughout his career. Markowitz and Hancock (2014) compiled a database of 24 hoax papers containing more than 170,000 words and another dataset of 25 non-hoax papers with more than 180,000 words. The findings showed that the hoax research papers were written with a prominent level of confidence when describing outcomes and had better linguistic dimensions than the original ones. This indicates a greater level of effort put into hoax work than is evident in original papers.

Deceptive text can also be found in spam emails. A study by Douzi, Al-Shahwan, Lemoudden, and Ouahidi (2020) considered the neural network model paragraph vector-distributed memory (PV-DM) and the TF-IDF approaches to assign dual vector representations to individual messages. This hybrid approach

## Chapter 2. Background and Related Work

gives a compact representation that contains information about the content of an email and its associated features. The study confirmed that the trained classifiers got the best results, with 93% accuracy, outperforming the BoW approach.

The same TF-IDF approach has been adopted by researchers in other fields too, specifically to detect spam SMS texts. M. Gupta, Bakliwal, Agarwal, and Mehndiratta (2018) trained their model using several ML classifiers such as convolutional neural networks (CNN), support vector machines (SVM), and naive Bayes (NB). Their dataset was composed of 1,000 spam and 1,000 ham SMS messages collected through crowdsourcing. The results obtained from their classifiers show the highest accuracy was achieved by the CNN classifier with 99.19%.

Earlier, C. M. Fuller, Biros, and Wilson (2009) showed that humans are incapable of detecting deception and, therefore, the task needs automating. The authors focused on textual and spoken transmissions based on a real-world database of false report statements in law enforcement. Examining a set of linguistic cues such as certainty terms, motion terms, modifiers, etc., their system achieved 74% accuracy in detecting fake statements. These projects inspired the development of tools that automate the detection of deceptive text. For example, VeriPol, a model proposed by Quijano-Sánchez, Liberatore, Camacho-Collados, and Camacho-Collados (2018), was developed to detect false robbery reports based only on text. The tool was developed by combining NLP in features such as unigrams, number of tokens, and number of POS tags (adjectives, negations, prepositions) and ML approaches to create a decision support system that predicts the falsehood of a given police report. More than 1,000 Spanish National Police reports were tested through VeriPol, reaching a success rate of 91% in discriminating between true and false reports. Another tool, called Hoaxy, is a platform that combines an analysis of Twitter online misinformation features with related fact-checking sites to classify fake content. Whenever a statement is flagged in Twitter as fake, the statement is run through a fact-checking platform

to verify its authenticity. Its state as real or fake is then shown to readers on Twitter. However, because Hoaxy relies on fact-checking sharing platforms, there may be 10- to 20-hour delays in the confirmation of misinformation (Shao et al., 2016).

These studies provide greater insight into deceptive text in general and fake news specifically, which provides useful information about the textual features under investigation. In the current research, we emphasise textual features, as the news articles distributed through social platforms and messaging applications provide only textual content and no metadata. Ample evidence supports the effectiveness of these textual features, which we elaborate on below.

### **2.3.1 Explored Textual Features**

A number of researchers (Alsudias & Rayson, 2020; Cartwright, Nahar, et al., 2019) have mobilised textual analysis to counter the deceptive text issue. Specifically, they focused on certain textual categories that might have an impact on the recognition of deceptive text. Here, we focus on textual features that have previously been investigated in research aiming to detect deceptive text.

#### **Part of Speech (POS) Features**

Each word in a sentence has a POS depending on its function in the sentence. When comparing the use of POS in real and fake articles, Kapusta and Obonya (2020) identified certain distinctive features. They found that fake articles used foreign words, adjectives, and nouns significantly more frequently than real articles, however, determiners, prepositions, and verbs were more prevalent in the real articles. Burgoon et al. (2003) found that truthful writers used more nouns, adjectives, prepositions, determiners, and coordinating conjunctions than deceivers, who used more verbs and adverbs.

Because fake news and deceptive reviews both have an aspect of manipulation,

## Chapter 2. Background and Related Work

Li, Ott, Cardie, and Hovy (2014) explored deceptive reviews that were deliberately written to sound authentic to deceive the reader. They constructed a dataset composed of three categories of reviews: truthful customer-generated reviews, Turker-generated (crowdsourced) deceptive reviews, and employee-generated deceptive reviews (as domain experts) about doctors and hotels and restaurants. Their study revealed that customer (truthful) reviews included more nouns, adjectives, prepositions, subordinating conjunctions, and determiners. On the other hand, Turker (deceptive) reviews contained more verbs, adverbs, and pronouns. The deceptive reviews by employees also included more nouns, adjectives, and determiners than the truthful customer reviews while containing fewer verbs and pronouns than the customer reviews. The researchers explained that employees have more knowledge of their domain and are thus able to provide more details and descriptions in their reviews than real customers. As a result, they may provide more detail and unintentionally overlook certain events, resulting in fewer verbs and pronouns. Li et al. (2014) study is contradicted by Kapusta and Obonya (2020), who found an increase of verbs and pronouns in fake news.

Horne and Adali (2017) used complexity, psychology, and stylistic features to classify real, fake, and satirical articles. The stylistic features referred to several POS (i.e., adjectives, verbs, pronouns, and nouns). The scholars also extracted negation, comparison, quantifying, interjections, determiners, nouns, and stop words in the same feature set, and they concentrated on casual, assuring, tentative, and emotional words, in addition to negative and positive words, to determine psychology features. Finally, they calculated the lexical diversity, average word length, and average frequency of words in each document for complexity features. With these feature categories, Horne and Adali were able to classify articles as real, fake, or satirical, and their classifier had an accuracy of 78%. However, satirical articles scored similar to fake articles in their writing characteristics. The authors found that fake articles used fewer analytical words, more



## Chapter 2. Background and Related Work

personal pronouns and adverbs, and fewer nouns. Interestingly, they also found that fake articles used more negative words in general.

### **Linguistic Features**

Pérez-Rosas et al. (2017) discussed the usefulness of linguistic features in detecting fake news. The researchers began by compiling two datasets—one by manipulating extracts from news using crowdsourcing and the other composed of articles about celebrity news. They extracted linguistic features such as pronouns, positive and negative polarities, function words, and basic emotions. Their classifier achieved 74% accuracy for the crowdsourced dataset and 73% for the celebrity dataset. Similar to previous studies, they found that fake articles contained more adverbs, verbs, and punctuation marks than real articles. Their work was novel in that their dataset included manipulated real articles to generate fake articles through crowdsourcing. Thus, their dataset mimicked the real-world dispersal of fake news.

Recent studies have been thorough in detecting the styles of writing used to deceive readers, with the belief that liars may inadvertently reveal their deceptive intentions through their writing. More specifically, Burgoon et al. (2003) compared truthful and deceptive communication to distinguish cues that may help detect deception. They found that deceivers' messages contained a lower quality of language in terms of a limited lexicon and sentence types, as well as a lack of specificity, indicating that some insecurity and hedging may be recognisable in their writings. The associated linguistic features included the number of words, adverbs, and conjunctions. Elaborating on this matter, Newman et al. (2003) compared the language style in true and false stories. They created a dataset composed of participants' true and false statements on several topics, such as abortion and feelings about their friends. The authors extracted linguistic features from the stories, such as positive and negative emotions, words expressing

## Chapter 2. Background and Related Work

causation (justification), tentativeness (hedges), assurance, space and time, exclusive words, and the use of pronouns and verbs. The study found that liars used more negative words, expressed more negative emotions, and used fewer exclusive words and more verbs than those who are truthful. Their computer-based text-analysis program correctly classified liars and truth-tellers 61% of the time in their five independent sample analyses.

L. Zhou, Twitchell, Qin, Burgoon, and Nunamaker (2003) studied the same set of features and added redundancy, objectification, and generalisation. Redundancy refers to the total number of function words divided by the total number of sentences. The diversity of content words was calculated as the total number of different content words divided by the total number of content words. By studying a dataset that contained true and deceptive text created by participants as messages between senders and receivers, the researchers found that deceivers' messages were longer, revealed an element of hedging, and were less complex than messages written by truth tellers.

The wide range of linguistic features investigated in prior research is shown in Table 2.2.

### **Unified Group of Features**

A comprehensive and possibly unified group of feature sets is presented in (Burgoon et al., 2003; Newman et al., 2003; L. Zhou et al., 2004). The effectiveness of groups of features in detecting deceptive text was studied by Gravanis et al. (2019), who used five fake news datasets from Kaggle-EXT, PolitiFact, BuzzFeed, UNB, and McIntire. Combined with ML algorithms, the proposed features obtained up to 95% accuracy in all the datasets used to detect fake news.

Although much research has been done in this field on several languages besides English—such as Indonesian (Adha, 2020), Bengali (Hossain, Rahman, Islam, & Kar, 2020), Brazilian Portuguese (Jeronimo et al., 2019; Resende et al.,

Table 2.2: Samples of Linguistic Features from the Literature.

Linguistic Feature	Source
Assurance	(Sabbeh & Baatwah, 2018)
Negations	(Hancock et al., 2007); (Horne & Adali, 2017); (Newman et al., 2003); (Pérez-Rosas et al., 2017); (Rashkin, Choi, Jang, Volkova, & Choi, 2017)
Intensifiers	(Gröndahl & Asokan, 2019); (Karoui, Zitoune, & Moriceau, 2017); (Pérez-Rosas et al., 2017)
Hedges	(Argamon et al., 2007); (Volkova, Shaffer, Jang, & Hodas, 2017); (Wei, Li, Zhou, & Gong, 2016); (Addawood, Badawy, Lerman, & Ferrara, 2019); (Rashkin et al., 2017)
Justification	(Gravanis, Vakali, Diamantaras, & Karadais, 2019); (Gröndahl & Asokan, 2019); (Hancock et al., 2007); (Jupe, Vrij, Leal, & Nahari, 2018); (Newman et al., 2003); (Pérez-Rosas et al., 2017); (Reis, Correia, Murai, Veloso, & Benvenuto, 2019); (Resende et al., 2019); (L. Zhou, Burgoon, Zhang, & Nunamaker, 2004)
Temporal	(Davis, Varol, Ferrara, Flammini, & Menczer, 2016); (Humpherys, Moffitt, Burns, Burgoon, & Felix, 2011); (Jupe et al., 2018); (Reis et al., 2019); (Resende et al., 2019); (Vrij, Kneller, & Mann, 2000)
Spatial	(Humpherys et al., 2011); (Jupe et al., 2018); (Vrij et al., 2000)
Illustration	(DePaulo et al., 2003); (Feng & Hirst, 2013); (Lagutina et al., 2019); (Li et al., 2014); (Rashkin et al., 2017)
Exceptions	(Ott et al., 2011); (Newman et al., 2003)
Opposition	(Conroy, Rubin, & Chen, 2015); (Hancock et al., 2007); (Karoui et al., 2017); (Vartapetian & Gillam, 2012)
Superlatives	(Rashkin et al., 2017)

2019), Italian (Pierri, Artoni, & Ceri, 2020), and Chinese (Liu, Yu, Wu, Qing, & Peng, 2019) little research on Arabic has been conducted. The current research begins to fill that gap. The prior work on Arabic that has been done is discussed in the next section.

### 2.3.2 Arabic Fake News Detection Systems

Researchers have used various methods to analyse and detect deception in Arabic text. Penuela (2019) used BoW and enriched it with features such as the number of words, number of hashtags, user mentions, and emojis. This provided a solid F score, 77%, in detecting deceptive Arabic tweets. Mohdeb, Laifa, and Naidja (2021) performed TF-IDF to generate useful features for classifying Arabic fake news. Applying this approach to a dataset of COVID-19 related social network posts and news articles, their classifier reached an accuracy of 94%. However, both of these studies are narrow in focus, dealing only with social media posts, and neither of these studies focused on news articles alone. The TF-IDF method proposed in the latter study would mandate the removal of stop words, as they are the most frequently used in documents. In this approach, stop words are retained, because many of them are crucial function words that also fall into linguistic categories and are counted in POS features.

As mentioned earlier in this chapter, satire (irony) is a particular type of fake news. Karoui et al. (2017) built a classifier trained on a dataset containing ironic Arabic tweets to detect this type of fake news. They focused on surface features, such as punctuation and opposition words; sentiment features, such as positive and negative opinion words; shifter features, such as exaggeration and reported speech words; and contextual features, such as personal pronouns. The researchers grouped these features from prior studies in other languages and translated lexicons from these languages into Arabic for their research. As a result, their classifier reached an accuracy of 72.36%, despite the difficulty of processing Arabic social media texts and the absence of tools to deal with translating words in lexicons. However, their study focused on tweets, which is not the text genre we investigated in this study.

A broader perspective that included Arabic news articles has been adopted by Saadany et al. (2020). They conducted exploratory analyses to identify the

## Chapter 2. Background and Related Work

linguistic properties of satirical Arabic content as a type of fake news. Satirical articles were analysed by searching for specific features such as journalistic register (terminology commonly used in journalistic writing), sentiment intensity measures, and subjectivity measures (by analysing pronouns that denote first person inflections). Saadany et al. (2020) claimed that these lexical features were sufficient to classify satirical Arabic fake news accurately. However, the nature of satirical articles is quite different from that of news articles. The satirical cues presented above might be obvious, which may cast doubt on the article's veracity. Detecting fake news articles posing as legitimate by imitating the journalistic style found in real articles is a bigger challenge.

Alzanin and Azmi (2019) analysed methods for identifying online rumours and fake news in tweets by utilising three groups of rumour-identification methods: supervised, unsupervised, and hybrid approaches. Supervised approaches rely on social media indicators to allow both humans and machines to assess credibility. Unsupervised approaches focus on specific characteristics of the texts propagating the rumours, along with the type of accounts that typically display them. Finally, the hybrid approaches combine parts of the other two approaches while emphasising morphological analysis. Though the results of their work were promising, the features relied on the metadata associated with each tweet, such as the tweet date, number of followers, and time. This approach is only useful on platforms associated with metadata and is unreliable when the message is distributed through social messaging platforms, where metadata is not associated with the messages.

In an attempt to differentiate real from fake reviews, Saeed et al. (2022) worked to identify false online reviews in Arabic using four identification methods:

- A rule-based classifier
- Machine learning classifiers

## Chapter 2. Background and Related Work

- A majority voting ensemble classifier
- A stacking ensemble classifier

The authors found that the ensemble approach outperformed other Arabic language fake review detection methods. The ensemble approach combines a rule-based classifier with machine learning techniques, while also using content-based features. It achieved a classification accuracy of 95.25% for the first dataset and 99.98% for the second dataset. However, this approach examined Arabic reviews in the form of short written texts that, in many cases, used informal language. As such texts do not belong to a journalistic genre, it can be hard to transfer the method to news articles.

Concerns about the COVID-19 pandemic and the vaccine have motivated various studies, all of which classified the veracity of Arabic statements concerning the pandemic (Alqurashi et al., 2021; Alsudias & Rayson, 2020; Haouari, Hasanain, Suwaileh, & Elsayed, 2020; Mubarak & Hassan, 2020). However, these studies also dealt with statements made on Twitter and rumours that spread on social media. Tweets cannot exceed 280 characters and rumours are often written informally, so this work is not directly applicable to the current study.

### **Datasets For Short-Text Arabic Fake News**

Alkhair et al. (2019) built a novel Arabic language dataset, which they called the rumours dataset, for fake news analysis based on YouTube user comments. The researchers focused their data collection on over 300 rumours on the deaths of three well-known personalities (who were still alive). After training, their model achieved a 95% accuracy rate in detecting rumours. However, though the model achieved commendable accuracy, their dataset was based on YouTube comments, which meant the texts were short and not written in standard Arabic.

Ali, Mansour, Elsayed, and Al-Ali (2021) introduced AraFacts, the first large Arabic language dataset of naturally occurring claims. The dataset covers about

## Chapter 2. Background and Related Work

6,000 claims made on social media platforms, annotated by professional fact-checkers from five different Arabic fact-checking websites. In their dataset, each claim has a rating selected from four labels (false, partly false, sarcasm, true). In addition, the dataset contains articles, images, and videos that were verified using the fact-checking platforms. However, since the claims were made on social media, they were not written in a formal style. Informal writing tends to lead to grammatical and spelling mistakes that are common in a conversational text, but they are also indicators of deceptive text as previously proven by Watters (2009) and Preston et al. (2021) .

Similarly, Alqurashi et al. (2021) built a dataset that included 8,786 tweets concerning the pandemic. Their model achieved promising results after applying term frequency and word embedding methods. Haouari et al. (2020) built the ArCOVID-19 Rumours dataset that covered 9,400 tweets related to the pandemic. The dataset included the claims made and verified tweets. Several similar pieces of work used to detect deceptive tweets about COVID-19 are seen in (Alsudias & Rayson, 2020; Elhadad et al., 2020; Mubarak & Hassan, 2020). However, as indicated earlier, informal language has distinctive features that may not be suitable for identifying deceptive text.

Khouja (2020) built an Arabic corpus that contained 4,547 true and false claims from news titles. She collected true news titles and based on those, generated fake titles through crowdsourcing. These real and fake title were referred to as claims. Though her corpus contains many claims, the text represented in the titles had relatively few words compared to full news articles. Elhadad et al. (2020) made a similar contribution in compiling fake news datasets in Arabic. He built a 220k dataset composed of misleading tweets about COVID-19.

A recently compiled Arabic fake news dataset by Assaf and Saheb (2021) has been published. The dataset contained fake articles and statements collected

## Chapter 2. Background and Related Work

from fact checking websites such as PalKashif<sup>4</sup>. In details, the dataset contains 323 articles (100 reliable and 233 unreliable news). Though the dataset might be a valuable contribution to research that combats fake news in Arabic, its fake articles were collected from not only news articles also from statements posted on social media platforms. Statements posted on social media would not conform with the journalistic writings obligated in news articles published in news platforms. This fact makes the dataset incompatible with our work, which focuses on fake articles written in a journalistic style.

Table 2.3 summarises Arabic datasets used for fake news detection.

### **Datasets For Long-Text Arabic Fake News**

Nagoudi et al. (2020) proposed a method to build an Arabic fake news dataset based on a machine-manipulated approach. First, they collected 5,187,957 news articles from various news outlets; then they generated false claims by substituting the proper nouns using WordNet to link similar, related words and replacing cardinal numbers in the articles with random numbers. The researchers suggested that this approach would generate a vast number of false articles from the real articles. However, some of these computer-generated articles produced unrealistic statements. For example, an article about a soccer match replaced five goals with 500 goals, thus overdoing the fake statement. This sense of exaggeration in numbers clearly revealed the deceptive language of the text.

## **2.4 Research Gap**

A closer look at the literature reveals several gaps. First, due to government regulations on producing and distributing fake news, there is a lack of Arabic datasets to help research fake news. Most of the datasets studied were prepared

---

<sup>4</sup><http://kashif.ps/>



Table 2.3: Summary of Datasets Used for Fake News Detection in Arabic.

Dataset	Genre	Number of Items
Elhadad et al. (2020)	Tweets	220 k
Khouja (2020)	News titles	4,547
Nagoudi et al. (2020)	Machine generated articles	5,187,957
Alqurashi et al. (2021)	Tweets	8,786
Mohdeb et al. (2021)	Social media posts	635
Alkhair et al. (2019)	YouTube comments	300
Ali et al. (2021)	Social media posts	6000
Assaf and Saheb (2021)	News articles and social media posts	323

from social media platforms, such as YouTube (Alkhair et al., 2019) or Twitter (Alqurashi et al., 2021; Alsudias & Rayson, 2020). However, such platforms may include informal language, often contain short texts, and do not adhere to specific journalistic rules. Thus, despite the success of previous work, they do not provide a dataset that can be used to train a model to classify fake articles written in a journalistic style. These types of fake articles, with which our research is concerned, cause severe damage due to their similarity with real articles. In our research, we rely on textual analysis since when fake articles are distributed through social media or messaging platforms, they may not be associated with any metadata. This requires detailed textual analysis of the article to examine textual markers that may serve as deceptive cues.

The aim of this research is to investigate the impact of textual features in classifying Arabic language fake news and thus construct a model that classifies Arabic language fake news based on these features. One advantage of using a textual features compared to BoW is the significant reduction in data requirements. While a typical set of textual features might contain tens or hundreds of features, it is normal for the BoW approach to assess thousands of features (Al-Ayyoub, Jararweh, Rabab’ah, & Aldwairi, 2017). Moreover, within a BoW model, a researcher must decide which N-grams to work with, which makes the

## Chapter 2. Background and Related Work

model dependent on the researcher's choices (Ullah, Amblee, Kim, & Lee, 2016). Another advantage of using textual features is that compared to deep learning methods, deep learning needs to be trained on huge amounts of input data (Elaraby, Elmogy, & Barakat, 2016). Due to the dearth of Arabic language fake news datasets, we have had to compile our own, which limits data availability. The situation presented challenges because of the Saudi government's criminal penalties on fake news production and distribution, which made it difficult to gather many fake articles in Arabic from fake websites, which are often banned. In fact, during our research, we received a fake news text on WhatsApp, linking to a fake news platform. However, when we clicked on the link, it was blocked immediately, so we were unable to get access to the platform from Saudi Arabia. Finally, NLP tools that were used in the related studies might have been successful in English text, however they could not handle Arabic text.

### **2.5 Summary**

This chapter provided definitions of news and of real and fake news. Deceptive text characteristics were defined, and based on that, related works that aid in detecting deceptive text were explored. We discussed related studies on deceptive text in general and on the Arabic language specifically. The chapter concludes with the research gap, which relates to the lack of Arabic datasets and tools for combating fake news in Arabic.

## Chapter 3

# Characteristics of Modern Standard Arabic (MSA)

Arabic is the official language of 20 Middle Eastern and African countries, including Saudi Arabia, Qatar, Bahrain, Jordan, Egypt, Lebanon, and Morocco. Because it is the language of Islam's holy book, the Quran, the number of Arabic speakers has increased due to the growing number of Muslim converts worldwide and the Islamic faith's tradition of reading the text in the original language (Holes, 2004). Since the inception of Islam in the seventh century CE, Arabs have inhabited many traditionally non-Arabic speaking countries, sometimes adopting non-Arabic loanwords into the Arabic language. For example, the word **أُسْتَاذٌ** 'teacher' is a loan from Persian. Furthermore, in the modern era, globalisation has introduced many terms to the Arabic language, such as the internet. Though this is an English term, Arabs have transliterated it phonetically and added it to their terminology with the same meaning and pronunciation. Nevertheless, Modern Standard Arabic (MSA) has retained the syntax, vocabulary, and phraseology of classical Arabic. Because Arabs generally understand this form of Arabic, many use MSA as the formal medium for virtual, written, and spoken broadcasts. It is also used in schools and televised news broadcasts. In symbolic

Table 3.1: Example of Verb وقى ‘Protect’

Root English	Arabic	Pronunciation	Number of Phonemes	Part of Speech
He protected	وقى	Waqā	3	Past tense Verb
He protects	يقي	Yaqy	3	Present tense verb
Protect	ق	Qi	2 (one for letter ق + vowel)	Imperative verb

terms, MSA is an intrinsic part of daily life for Arabs, regardless of their different dialects. In other words, MSA is the dominant language in everyday life in the Arabic speaking world for business, education, and formal government services, as well as televised broadcasts and news articles (Holes, 2004).

### 3.1 The Arabic Root-Pattern System

Arabic morphology is unique. Every word in the Arabic language has a three-letter root representing the base meaning of the word. From each of these roots, dozens of words can be formed. Specific patterns are applied to the roots, changing the root’s meaning to form a related word. The logic and cohesion of the Arabic language, which is highly systematic, originates from this advanced root-pattern system. In its most basic elements, Arabic is composed of consonant roots that work in tandem with vowel patterns.

More specifically, a root or جذر is a relatively invariable discontinuous bound morpheme represented by two to five phonemes—typically three consonants in a specific order—that has lexical meaning and interlocks with a pattern to form a stem (Ryding, 2005). An example is shown in Table 3.1.

The pattern, as mentioned above, is a limited and, in many cases, discon-

Chapter 3. Characteristics of Modern Standard Arabic (MSA)

Table 3.2: Arabic Root Pattern of كَتَب 'Wrote' With Three Vowels, Red Indicates the Short Vowels.

Root English	Arabic	Pronunciation	Word (With Vowels)	Part of Speech	Meaning
Wrote	كَتَب	Ktb	كَتَب katab	Verb	He wrote
			كُتُب kotub	Noun (Plural)	Books
			كُتِب kutib	Past participle verb	Has written

tinuous morpheme consisting of one or more vowels and slots for root phonemes (radicals). These interlock with a root to form a stem, either alone or combined with one to three derivational affixes, and generally the affixes have grammatical meaning (Ryding, 2005). Put simply, patterns are the fixed moulds of words into which roots can be inserted. Together, the root letters and the patterns in which they are placed form words. Patterns, like suffixes and prefixes, also carry meanings.

Thus, the root-pattern system consists of roots that have a general meaning, with more specific meanings and functions created by the patterns in which the roots are placed. To better understand how roots and patterns work together, one can consider the common root of كَتَب or 'wrote', which forms the basis of Arabic words related to writing or inscriptions. A combination of this root with different patterns forms different words, as seen in the example in Table 3.2, illustrating three patterns of كَتَب 'wrote' with short vowels as patterns.

The root similarities between all three words are readily apparent, even to the untrained eye. While the root كَتَب 'wrote' signifies a word or phrase related to 'writing,' it is clear that three new words are formed when the patterns, as short vowels, are added. Another example of different patterns with affixes added to the same root is shown in Table 3.3.

Chapter 3. Characteristics of Modern Standard Arabic (MSA)

Table 3.3: Influence of Affixes on the Word كُتِبَ ‘Wrote’.

English	Arabic	Pronunciation	Type	Example
Writing	كُتِبَ	ktb	Verb	He wrote
Writer	كاتب	katb	Noun	Writer
Book	كتاب	ketab	Noun (singular)	Book
Writers	كتبه	katabah	Noun (plural)	Writers

Table 3.4: Details of Affixes With The Root عِلْم ‘Knowledge’, Affixes are in Red.

Root English	Arabic	Pronunciation	Word Derivations	Added Affixes	POS	Meaning	
Knowledge	عِلْم	Ilm	عَلِمَ	alima	none	Verb	He knew
			عَالِم	aalim	الألف alif = a	Noun	Scientist
			عَالِم	aleem	الياء ya = ee	Noun (singular)	Someone who knows very much
			عُلَمَاء	ulama'a	الألف و الهمزة a'a	Noun (plural)	Scientists

In short, affixes are clitics added to a word for precision and contextual purposes. Their types depend on their position as prefixes, suffixes, or infixes.

Table 3.4 shows an example of patterns, as affixes, added to the root عِلْم which refers to ‘knowledge.’ Thus, each generated word shares the meaning of the root ‘knowledge.’

Much as English relies on the relationship between consonants and vowels,

Sentence	تَعَلَّمُوها
root	عَلِم = learn
Prefix	ت = you
Suffix	و = plural masculine
Post suffix	ها = it

Figure 3.1: A Full Sentence in One Word.

Arabic relies on the relationship between roots and patterns to form words. Roots also allow Arabic speakers to piece together the meaning of new words based on general concepts. In the examples above, readers could identify the general meaning using the root **كتب** ‘wrote’ or **علم** ‘knowledge’, while using the patterned consonants and vowels to extract the precise definition.

Moreover, the combination of roots and patterns is highly distinctive and may produce the equivalent of a complete sentence in one word. An example of this design can be seen in Figure 3.1, which shows an Arabic word **تعلموها** that is the equivalent of an entire three-word sentence in English: ‘You learn it.’ The prefixes and suffixes that serve as patterns are connected to the root, constructing a logical sentence.

With the importance of the Arabic root-pattern system in mind, the difference in content or function types of words in Arabic is now examined.

### 3.1.1 Content Words

As mentioned earlier, content words have individual meanings, and they can include nouns, verbs, adjectives, adverbs, and interjections.

#### Nouns

Derived from lexical roots, Arabic nouns are formed by placing certain patterns into the root to create different nouns. As in English, Arabic nouns can be

### Chapter 3. Characteristics of Modern Standard Arabic (MSA)

common or proper nouns. Compound nouns are formed in Arabic by combining two independent words to form a syntactic unit. Table 3.5 shows examples of compounds (Massey, 2008).

#### Adjectives

Adjectives are words that describe a noun. They are divided into two groups, depending on their role: attributive and predicative. Attributive adjectives are part of a noun phrase and directly follow a noun to describe it further. In this case, the adjective must agree with the gender and number of the noun. A predicative adjective provides information about the sentence's subject, thus completing the clause. It acts as a predicate in a nominal sentence and agrees with the noun's gender and number.

Arabic adjectives have a comparative or superlative degree. Comparative adjectives, which compare two nouns, usually have the root **أفعل** / 'afa al'. Superlative adjectives are used to indicate the highest degree of comparison. These have the same root **أفعل** / 'afa al' for the masculine form and start with prefix **أ** / 'a a'. In the feminine form, they have the root **فعلى** / 'fu la' and the suffix **ى** / 'a a' (Abu-Chacra, 2007). Table 3.6 lists examples of masculine and feminine superlatives and of attributive and predicative adjectives.

Table 3.5: Compound Nouns.

Noun Pronunciation	Arabic	English	Noun Type	Composition
<b>Abdul Rahman</b>	Proper noun + proper noun	Abdul Rahman	Proper noun	عبد + الرحمن Abdul + Rahman
<b>Humat al-watan</b>	Common plural noun + Common plural noun	Nation protectors	Common noun	حماة + الوطن Humat + alwatan



Chapter 3. Characteristics of Modern Standard Arabic (MSA)

Table 3.6: Adjective Examples.

Masculine Superlative			Feminine Superlative	
English	Arabic	Pronunciation	Arabic	Pronunciation
Biggest	الأكبر	al a 'akbar	الكبرى	al kubra'
Smallest	الأصغر	al a 'asghar	الصغرى	al sughra'
Best	الأفضل	al a 'afdhal	الفضلى	al fudhla'
Attributive Adjective			Predicate Adjective	
Ahmad ate a green apple	أكل أحمد تفاحة خضراء		The house is green	البيت أخضر

Table 3.7: Verb Examples.

Sentence		Verb	Type
English	Arabic		
Ahmad ate the apple.	أكل أحمد التفاحة	Ate	Past
Ahmad is eating the apple.	يأكل أحمد التفاحة	Is eating	Present
Eat your apple!	كُل التفاحة	Eat!	Imperative
Ahmad will eat his apple.	سيأكل أحمد التفاحة	Will eat	Future

### Verbs

As in all languages, verbs in Arabic indicate the action in a sentence. Arabic verbs are formed from a combination of two to five consonants as roots that form the base meaning of the verb. Verbs are categorised, according to their tense, into past, present, and future. There are also imperative and future tense verbs. Though not as commonly used as the other verbs, the latter express actions in the future (Abu-Chacra, 2007). Examples are given in Table 3.7.

## Chapter 3. Characteristics of Modern Standard Arabic (MSA)

### Adverbs

Arabic adverbs are mainly derived from nouns or adjectives. Their main function is to modify any part of speech aside from nouns. The adverb can modify verbs, adjectives, other adverbs, and clauses. It also gives extra information about the word in terms of manner, time, and the frequency of performing a specific action. Examples are in Table 3.8.

### Interjections

These are words, such as **أوه** ‘ooh’ or **يع** ‘yuck’, that express the sentimental state of the speaker.

### Gender, Person, and Number

Gender, person, and number are also important components in Arabic morphology. There are three persons—first, second, and third—with the first person having no gender distinction. In the second person, there are five forms, depending on number and gender: masculine singular, feminine singular, dual (two persons), masculine plural, and feminine plural. Finally, in the third person, there are six verbal distinctions and five pronoun distinctions: singular masculine = **هو** ‘he’, singular feminine = **هي** ‘she’, dual masculine = **هما** ‘they’, dual feminine =

Table 3.8: Adverb Examples.

Adverb			Sentence	
English	Arabic	Type	English	Arabic
Usually	عادةً	Frequency	He usually goes in the morning.	عادة ما يذهب في الصباح.
Soon	قريباً	Time	I will graduate soon.	سأُتخرج قريباً.
Together	معاً	Manner!	We went together to the mall.	ذهبنا معاً الى السوق.

### Chapter 3. Characteristics of Modern Standard Arabic (MSA)

Table 3.9: Verb and Noun Affiliation. Note: Red Indicates the Prefix, Blue Indicates the Suffix, and Green Indicates the Root. Fem.=Feminine, Masc.=Masculine.

Description	Base Form	Fem. Singular	Masc. Singular	Fem. Dual	Masc. Dual	Fem. Plural	Masc. plural
English Arabic Pron. Type	Eat أكل a'kl verb	تأكل Ta'kl	يأكل Ya'kl	تأكلان Ta'kluan	يأكلان ta'kulan	تأكلن Ta'kuln	يأكلوا Ya'kulu
English Arabic Pron. Type	Management إدارة idarah noun	مديرة mudirah	مدير Mudir	مديرتان mudiratan	مديران mudiran	مديرات Mudirat	مدراء Mudara'a

هما ‘they’, plural masculine = هم ‘they’, and plural feminine = هن ‘they’. As a result, there are 13 Arabic person categories, whereas English has only seven (Shamsan & Attayib, 2015). Arabic has three numbers: singular, dual, and plural. Thus, there are distinct pronouns for pairs of people, or animals, whereas in English any number more than one is treated as a plural. Arabic does not consider quantities to be plural until they are three or more. Patterns, such as affixes (prefixes/suffixes) and vowels, can be attached to a verb or noun to specify gender, person, and number. Table 3.9 shows an example of verb and noun inflection.

#### 3.1.2 Function Words

Function words are expressed by ‘particle’ حرف, in the Arabic POS basic structure. Function words do not generally carry meaning by themselves, but are a supportive structure that helps to produce organised and detailed meaning in text. There are a limited number of particles—less than 100—in Arabic. Each particle holds a peculiar meaning and functions according to that meaning when

Table 3.10: Examples of Exception and Negation Particles.

Particle English	Arabic	Pronunciation	Type
Except	إلا	ila	Exception (One particle)
Except	ما عدا	Ma ada	Exception (Two particles)
Except	ليس	lais	Negation (One particle)

added to a word or sentence. Two particles can be combined to express a more definitive meaning for the context; for example, لا سيما which means ‘especially,’ contains two particles ‘la’ and ‘siyama’ and precisely means ‘for that’ Particle types differ based on their function, such as exception and negation particles, as shown in Table 3.10. Exception particles are used to express an object as separate from a particular group. Usually, these are followed by the expectant, a noun. Negation particles are used to negate a statement. Further examples of particles include prepositions, conjunctions, and pronouns.

### Prepositions

Though they are limited in number, prepositions play a vital role in signifying the relationship between one word and another. A preposition may consist of only one letter attached to a noun or a separate word composed of several letters. Each preposition has a linguistic meaning that appears when added before a noun, signifying a location or direction. Prepositions also include derivative prepositions that are a form of a temporal or locational adverb. Some examples include في ‘in’ and على ‘on’.

Table 3.11: Examples of Conjunctions.

Conjunction	Translation	Type	Meaning	Example	Translation
Wa و	And	detached	Addition	أَكَلْتُ التفاحة والموز.	I ate the apple and bananas.
Fa ف	Then	attached	Addition	أَكَلْتُ التفاحة فالموز.	I ate the apple, then the bananas.
A'kin الكن	But	detached	Opposition	ذهبتُ للسوق و لكن السوق مقفل.	I went to the mall, but the mall was closed.

### Conjunctions

Conjunctions are particles that primarily function to connect words or sentences to show a link, such as cause and effect, contradiction, or sequence. There are two types of conjunctions: coordinating and subordinating. Coordinating conjunctions are the type used most in Arabic, as they connect two related words, thoughts, or sentences. Subordinating conjunctions, on the other hand, connect two unequal clauses. When one clause contains a verb, the other clause needs an object, if an object is not present then the statement becomes unequal, hence, subordinating conjunctions are used to link the clauses. Conjunctions can be attached to or detached from a word. Because they are function words, each conjunction has a unique meaning and performs a linguistic function (Bouchentouf, 2013). Table 3.11 presents some conjunctions with examples of their use.

### Pronouns

These are words used to replace a noun. Like conjunctions, Arabic pronouns can be attached or detached. If attached, they are linked to a word in place of the person/thing and agree with the word's number and gender. For example, the

Table 3.12: Pronoun Types and Examples, Red Indicates the Pronoun.

Pronoun	Specification	Type	Example	Translation
أنتِ anti	Feminine singular	Detached	أنتِ التي نجحت. You are the one who passed.	You
أنتَ anta	Masculine singular	Detached	أنتَ الذي نجح. You are the one who passed.	You
ي ya	Feminine singular	Attached	ادرسِي درسك You study your lesson.	You
نا na	Feminine and masculine dual	Attached	درسنا الدرس We studied the lesson.	We

pronoun ي 'ya' is assigned when an imperative verb is directed to a feminine subject, however, the pronoun أ 'aa' is attached when the imperative verb is directed to a masculine subject. Detached pronouns are concrete words used in place of persons and things in a sentence. Similar to attached pronouns, they also agree with the person, number, and gender specifications of the subject and object (Bouchentouf, 2013). Table 3.12 lists some examples. As pronouns perform a grammatical function when added to a sentence, they are also categorised as particles in Arabic.

### Determiners

In addition to the function words above, Arabic also uses determiners, which are classified as definite and indefinite. The prefix al- is definite and used at the beginning of nouns and adjectives. The indefinite determiner is the diacritic mark ô attached to the end of case-marking vowels in nouns and adjectives (Ryding, 2005). For example, 'the dog' is expressed as 'al kalb' الكلب, while 'a dog' is

Table 3.13: Affixes as Function Words: Affixes are Colored in Red , Noun or Verb Colored in Green.

Prefix/ Suffix	Meaning, Role	Example	Translation	Similar Function Word
ل li + verb	To do, justifi- cation	ذهب أحمد ليلعب مع أخوه	Ahmed went to play with	Hence Thus In order to
ك ka + Noun	As, similarity	وجهك كالقمر	Your face is as the moon.	Similar Like
س sa + verb	Will, Future action	سيذهب أحمد إلى البقالة	Ahmad will go to the market.	Shall

كلباً 'klbaan'.

### Affixes as Function Words

When attached to a word, some affixes convey a grammatical meaning. For example, in English, the prefix 'un-' has the same grammatical role as the function word 'not.' Though not concrete, these affixes are considered, in their role, as function words because they give a functional meaning when attached to a word. Table 3.13 shows an example.

## 3.2 Arabic NLP Challenges

The unique morphology of Arabic creates challenges for the Arabic language research community. These challenges have a direct impact on NLP tool processing and overall, on deceptive-news detection systems. Some challenges are explained below.

### 3.2.1 Orthographic Variations

Some Arabic letters share the same letter shape but have different pronunciations when marks such as single dots or double dots—hamza (ء), or mada (◌~)—are placed above or below the letter. Thus, NLP tools must distinguish between the letters based on the position of these marks. However, some MSA texts are lax about adding these marks and the proper marks are sometimes omitted. It is usually up to the reader to determine which word is intended, depending on their familiarity with this practice. For example, the word في meaning ‘in’ is sometimes written without the two dots beneath it as في.

### 3.2.2 Lack of Capitalisation and Punctuation

The absence of capitalization and clear punctuation rules in Arabic make pre-processing difficult. During the automatization process, the machine cannot distinguish between one clause and another, as some Arabic sentences may run the length of an entire paragraph without commas, with coordinators linking the statements together and with the whole section having only one final punctuation mark. Additionally, as proper names in Arabic are not capitalized, their shape is not identifiable. In some cases, a proper noun may be mistaken for a common noun. For example, أيقظتني أحلام could mean ‘I was awakened by **dreams**’ or ‘I was awakened by **Ahlam**’ (a personal name), as أحلام ahlam means ‘dreams’ in Arabic and is also a common girl’s name.

### 3.2.3 Homographs

The current habit of readily discarding the written diacritics of words in MSA text creates homographs. As mentioned in Elkateb et al. (2009), diacritics are essential and considered short vowels used to identify the pronunciation of letters. Inevitably, ambiguity arises when diacritics are misplaced or misused, leaving the



reader to identify the word according to the overall context and making it harder for NLP tools to identify the word accurately. As with any language, when there is misuse of a single diacritic, such as شدة ‘shaddah’, which doubles the consonant, it can cause confusion in multilingual contexts and will mean failure to identify words correctly. For example, the word مثل ‘mathal’, when written without a shaddah on the middle letter might imply the meaning, ‘similar.’ However, when a shaddah is added to the middle letter مَثَل, ‘maththal’, it means ‘acting.’

### 3.2.4 Lack of Arabic Lexicons

Standard Arabic lexicons include Lisan Al-Arab<sup>1</sup> and Al-Mujam Al-Ghani<sup>2</sup>, which have entries for over 300,000 words. These lexicons are widely used in text analysis projects, such as sentiment, subjectivity, and author analyses, as well as identifying the author’s gender (Al-Barhamtoshy, Hemdi, Khamis, & Himdi, 2019; Alsmearat, Al-Ayyoub, Al-Shalabi, & Kanaan, 2017; Mohammad, Salameh, & Kiritchenko, 2016). Although these two lexicons are useful, they do not provide easy access to specific lexical categories. As in dictionaries, all the words are arranged in alphabetical order, with each word defined and given its grammatical use, if provided. The researcher needs to search for their desired words and combine similar words that have similar purposes to form a specific lexicon, which could be burdensome. In fact, some researchers have manually compiled specific Arabic lexicons such as Arabic particle lexicons (Namly, Bouzoubaa, Tahir, & Khamar, 2015), verb lexicons (Loukil, Haddar, & Hamadou, 2010), and sentiment lexicons (Mohammad et al., 2016). A recent lexicon is ArDep: An Arabic Lexicon for Detecting Depression (Alghamdi, Mahmoud, Abraham, Alanazi, & García-Hernández, 2020), which was compiled to recognize the Arabic words and phrases used by people suffering from depression.

<sup>1</sup><http://arabiclexicon.hawramani.com/ibn-manzur-lisan-al-arab/>

<sup>2</sup><https://nujoomapps.com/product/mojam-al-ghani/>

On the other hand, researchers have translated available non-Arabic specific lexicons in other languages such as English and French into Arabic for research use. For example, Karoui et al. (2017) translated an intensifier lexicon in French to Arabic (Karoui, Benamara, Moriceau, Aussenac-Gilles, & Belguith, 2015).

To enhance the use of lexicons, some authors of sentiment lexicons have assigned each word a score to test a system’s ability to predict the sentiment intensity score for a given text. The Multi Perspective Question Answering (MPQA) subjectivity lexicon, for example, contains 2,718 positive, 4,911 negative, and 570 neutral words. Each word was assigned a score between 0 and 1 indicating the intensity, with 1 indicating the maximum score for a positive sentiment and 0 for a negative one. Another example is the emotion lexicon by Strapparava and Mihalcea (2008). This lexicon included the six basic human emotions, according to Ekman (1999), as its emotion categories. It has 748 words for expressing anger, 155 for disgust, 425 for fear, 1,156 for joy, 522 for sadness, and 201 for surprise. It gives fine-grained scores to each word, using a scale from 0 to 100 to indicate intensity in the specific emotion category.

#### 3.2.5 Lack of Arabic NLP Tools

Unfortunately, with most research focused on building sentiment lexicons, other domain lexicons have been neglected. The lack of multiple domain lexicons has caused a lack of NLP tools that support the Arabic language, dampening interest in Arabic research projects. Although there is a large set of tools and programming language libraries for English that support NLP—such as NLTK<sup>3</sup>, StanfordNLP<sup>4</sup> by the Stanford group, TextBlob<sup>5</sup>, Genism<sup>6</sup>, and SpaCy<sup>7</sup>—there are fewer such tools in Arabic. Natural language processing tools that do exist

---

<sup>3</sup><https://www.nltk.org/>

<sup>4</sup><https://nlp.stanford.edu/>

<sup>5</sup><https://textblob.readthedocs.io/en/dev/>

<sup>6</sup><https://radimrehurek.com/gensim/>

<sup>7</sup><https://spacy.io/>

for Arabic are:

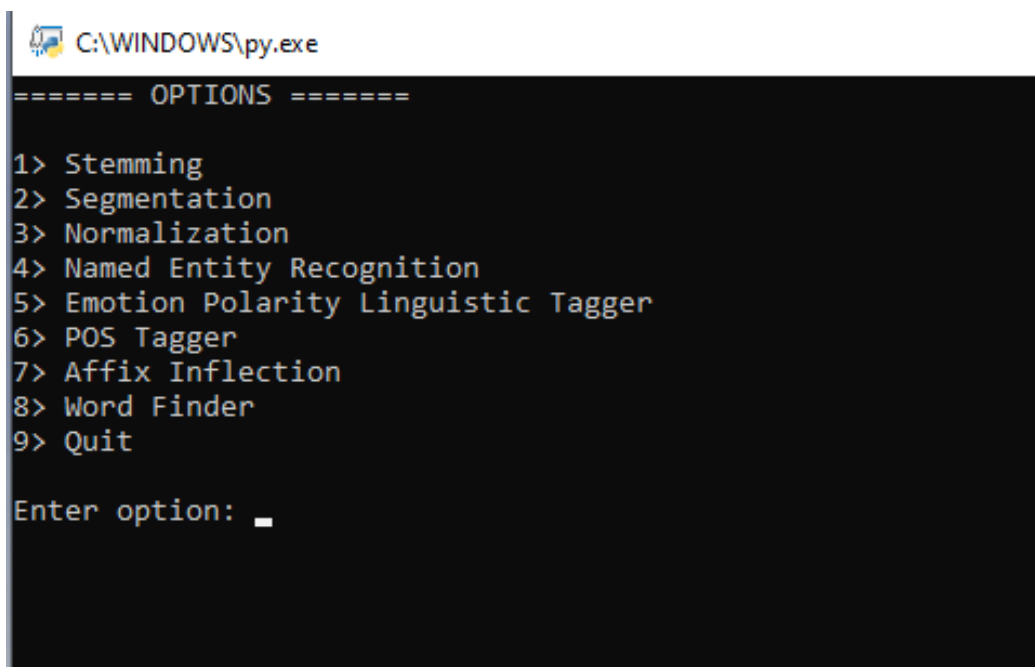
- **Tashaphyne** is an Arabic light stemmer and segmental tool. It provides light stemming, such as removing prefixes/suffixes, and generates segmentation. It uses its own built-in customized prefix and suffix list, which offers precise stemming. Besides stemming and segmentation, it offers normalisation and root extraction.
- **StanfordNLP** is a multilanguage NLP tool used for many languages. For Arabic, it provides parsing, tokenization, sentence splitting, name entity recognition, and POS tagging. It offers utilities through a Python package.
- **Farasa** is an Arabic-specific tool that provides NLP utilities through a collection of Java libraries. The utilities include discretization, segmentation, POS tagging, NER, and parsing.

### 3.3 Tasaheel Tool

For the current research, we have designed a tool to perform various NLP tasks for the Arabic language by combining several Python packages that perform Arabic NLP tasks and upgrading some of their features to include other tasks to support Arabic textual analysis (see Figure 3.2). We designed a two-stage framework to provide a comprehensive solution for Arabic textual research projects that require NLP.

This section describes NLP tasks already available in the packages supported by NLP tools:

- **Stemming:** This was undertaken using Farasa and Tashaphyne.
- **Segmentation:** Due to Arabic's unique morphology, it is necessary to segment text into morphemes to decrease the ambiguity created by the



```
C:\WINDOWS\py.exe
===== OPTIONS =====
1> Stemming
2> Segmentation
3> Normalization
4> Named Entity Recognition
5> Emotion Polarity Linguistic Tagger
6> POS Tagger
7> Affix Inflection
8> Word Finder
9> Quit
Enter option: _
```

Figure 3.2: Tasaheel GUI.

attached affixes. The packages available, such as Tashaphyne and Farasa, were used to mitigate this issue.

- **Normalisation:** It is important in many Arabic research projects dealing with text to perform normalisation that unifies text. Normalisation helps reduce word ambiguity and remove unnecessary noise associated with the text. Here, the Tashaphyne was included. Another set of options was also provided to perform single normalising tasks, such as removing numbers, non-Arabic letters, characters, stop words (user provides the list of stop words), and diacritic marks.
- **POS tagging:** POS taggers were used to assign a POS to each word in a sentence. Further, the user is presented with two POS tagger types: Farasa or Stanford NLP.
- Emotion, polarity, and linguistic tagger offers word tagging of words with emotion polarity or a particular linguistic function.

When creating a tagger function, word resources and extraction methods must be put in place. For example, the natural language tool kit (NLTK) library in Python that supports NLP functions, such as POS tagging, has several corpora that contain words and their POS tags. All the techniques form a method using the position of the word in the sentence to give it its accurate POS tag, where each tag is identified from a corpus; we followed this idea when tagging the textual categories which include emotion, polarity, and linguistics.

Most importantly, we needed to create wordlists for emotion, polarity and linguistics word tagging. In the following section, we detail their construction.

### **3.3.1 Creation of Wordlists**

Not only do languages with unique morphologies such as Arabic lack the relevant corpora available for a high-resource language such as English, but they also lack the basic lexical resources. While this lack of readily available lexical resources created a challenge for this study, it also produced opportunities. Since these lexical resources had to be created from the ground up, they could be crafted to meet the specifications and goals of this research. As previously stated, most Arabic lexicons were either created and annotated manually (Mohammad et al., 2016) or translated from non-Arabic lexicons (Karoui et al., 2017; Saad & Ashour, 2010). The first phase of lexicon creation is quite intense and the second involves the somewhat tedious work of translating words and removing any duplicates that might be produced by translation. Henceforth, the term ‘wordlist’ is used for convenience and to distinguish it from lexicons, which may be associated with scores. In other words, all the words have the same purpose within the feature category.

### Emotion and polarity Wordlists

The emotion and polarity wordlists were created from the words included in previous lexicons. Specifically, to create the emotion wordlist, we extracted the words from Bing Li's English language emotion lexicon<sup>8</sup>. The emotion words fall into the six basic categories of human emotions: غضب 'anger', اشمئزاز 'disgust', خوف 'fear', حزن 'sadness', فرح 'joy', and متفاجئ 'surprise'. Fortunately, the words had previously been translated into Arabic in a study by Saad and Ashour (2010), in their 'Arabic emotion lexicon.' The emotion wordlists categories contains the following six emotion wordlists categories:

- 748 words denoting anger
- 155 words denoting disgust
- 425 words denoting fear
- 1,156 words denoting joy
- 522 words denoting sadness
- 201 words denoting surprise

Similarly, words from the Arabic Sentiment Lexicon created by Mohammad et al. (2016) were extracted to form the polarity wordlists categories. The lexicon included ايجابي 'positive' and سلبي 'negative' words. All the words were included in the positive polarity category comprising 2,006 positive words and the negative polarity with 4,783 negative words.

It is important to note that certain words in the polarity wordlists are unavoidably repeated in the emotion wordlist. This is because emotions involve a broader and larger analysis than sentiment to cover the specific details of the desires, goals and intentions linked to a person's facial expressions. Examples of the polarity and emotion wordlists are provided in Table 3.14.

---

<sup>8</sup><https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>

Table 3.14: Emotion and Polarity Wordlists.

Content Feature	Translation	Example
Anger	anger, exasperation, indignation, resentment	نقمة ، غضب ، سخط ، حنق ، غيظ ، امتعاض
Sadness	worry, crying, tears, tearing	دموع ، الدموع ، بكاء ، همّ ، يدمع ، دمع ، تدمع
Fear	terrible, scary, horrific, terrifying, damned	مفزع ، مرعب ، مروع ، مخيف ، رهيب ، لعين
Joy	good, generous, delicious, beneficial	لذيذ ، كريم ، صالح ، طيب ، فائدة
Surprise	surprise, amazement, confusion	حيرة ، تحير ، مذهول ، مندهش ، حيران
Disgust	disgust, repulsion, loathing	تنافر ، نفور ، مقت ، اشمئزاز ، نفور،تقزز
Positive	wise, free, luxurious, classy	فاخر ، مجاني ، حكيم ، راقى
Negative	dirty, stingy, revolution, mean	بخيل ، لئيم ، وسخ ، ثورة

In total , we have organized six emotion wordlists, anger, fear, sad, surprise, disgust, joy, that are part of the emotion wordlists category and two polarity wordlists, positive and negative , that are part of the polarity wordlists category.

### Linguistic Wordlists

Inspired by the previous work of Saad and Ashour (2010) and Karoui et al. (2017), we heuristically created a wordlist for each linguistic category to be further embedded into Tasaheel. We followed two methods to organise the words to form the linguistic wordlist. First, we formed the intensifier and hedges wordlists by translating the English lexicons available for that category. Intensifiers were translated

Table 3.15: Lexicon Resources.

Lexicon Name	Author	Publisher Country	First Publishing Date
لسان العرب Lisan Al Arab	ابن منظور Ibn Manthur	Tunisia	1290
المعجم الوسيط Al-Mu'jam Al-Waseet	مجمع اللغة العربية بالقاهرة Arabic Language Association in Cairo	Egypt	1960
المعجم الغني Al-Mujam Al-Ghani	عبد الغني أبو العزم Abdul Ghani Abu Al-Azem	Morocco	2016

from the English intensifier lexicon (Wilson, Wiebe, & Hoffmann, 2005), and hedges were translated from the English hedge lexicon (Islam, Xiao, & Mercer, 2020). The words were translated using Google Translate, and duplicate words produced by the translation were removed. Second, as not all the linguistic categories were available in other languages, we created further wordlists by referring to a range of reliable, well-known Arabic lexical resources. The latter was beneficial, as these resources provided words denoting the linguistic categories needed in our work. We organized the words for each linguistic category, relying mainly on the Arabic lexical resources, as shown in Table 3.15. All the words in the wordlists were content words that fit the role of the wordlists for which they were included.

### 3.3.2 Wordlist Revision

Assuming that the emotion and polarity wordlist were credible for use because they had already been used in Arabic research studies (Al-Ayyoub et al., 2017; Al-Barhamtoshy et al., 2019; Karoui et al., 2017; Mohammad et al., 2016; Saad & Ashour, 2010), we revised only the linguistic wordlists. Three female Arabic linguistics scholars from Umm-AlQura University in Makkah, Saudi Arabia, revised



### Chapter 3. Characteristics of Modern Standard Arabic (MSA)

Table 3.16: Linguistic Wordlists With the Number of Concordant Words Each Contains.

Lexical Wordlist	Meaning	Word Example (Translated into English)
Assurance [7]	transitions used to indicate assurance	أن عين نف A'an – a'in – nafs for sure, surely, certainly
Negations [7]	used to dispute the truth of a statement	لا - لن - ل la – lan – lam no, not, never
Intensifiers [14]	to strengthen the meaning of a word	جدا تماما جميع jidan – tamamen - jame very, too, not at all
Hedges [7]	to soften / express hesitation / unassurance	من الممكن يجب احتمال ehtimal- yaji'b- min almomkin maybe/ should/ could/ may
Justification [9]	to show cause / justification	بسبب لذلك من أجل bisabb- lithalik-min ajel because/ to/ for that
Temporal [8]	to show time	البارحه غدا albariha – ghadan yesterday, tomorrow
Spatial [10]	to show space	تحت فوق عند taht – foug- e'nd under, over
Illustration [6]	used to portray	مثال مثال mithal- mathal for example
Exceptions [6]	used to indicate omission	إلا عدى سوى e'la – a'da – siwa except
Opposition [4]	to indicate adversity	لكن إنما lakn – e' nama but /although

the set of linguistic categories. They classified each word as 'approved' or 'not approved' based on its fit with the category. Aggregation was based on voting; at least two scholars had to agree on the word's compatibility with its linguistic category.

As a result, 10 linguistic wordlists category were formed and approved, con-

taining concordant words within the functionality of each category. Table 3.16 describes each linguistic wordlist, along with the number of concordant words it contains, with examples of the words translated into English.

### 3.3.3 Invoking Wordlists

I coded the tool to perform emotion, linguistic, and polarity word tagging and integrated it into Tasaheel. The user can choose which type of textual category they require, and an output file will be produced with all the text files tagged according to the chosen tagger option by word matching. This approach uses the exact string matching method recommended by Al-Sanabani and Al-Hagree (2015), where each word in the list is compared to each word in the text files and a match is displayed in the output file (for more details about Tasaheel, see Appendix A). End users are free to handle and update these customized wordlists separately.

## 3.4 Summary

This chapter provides a comprehensive background of the Arabic language by taking an in-depth look at the language's morphology, content words, and function words. As a result, it reveals challenges in the NLP domain, for which it constructs and details a tool that serves Arabic research.

# Chapter 4

## Methodology

In this chapter, we present the methodology used in this research. The chapter outlines the research principles, procedures, and practices (Kazdin, 1992). It reports the research design, data collection and pre-processing, the selection of textual features, and the tools and measures necessary to train and test the model. A supervised machine learning technique was used in the research to build a model that classifies real and fake Arabic news articles. The proposed model is comprised of multiple stages described in the chapter.

### 4.1 Research Paradigm

A research paradigm is the ‘[...] lens that guides the choice of theory and methods in research’ (Neil, 2007). With this in mind, a methodological approach is typically mandated when conducting research studies. A methodological approach is defined as ‘[...] the approach by which research troubles are solved thoroughly’ (Mishra & Alok, 2017). Thus, finding solutions to problems is one goal of research. The paradigm of a research study to find such a solution consists of a philosophy with specific reasoning. A research philosophy is defined as new and reliable knowledge about the research implemented (Žukauskas, Vveinhardt,

& Andriukaitienė, 2018). In the diverse field of research, several philosophical trends are prevalent, the two most common being positivism and interpretivism. In the first, a researcher is an objective analyst who detaches themselves from personal views and works systematically to produce an objective work that contributes to knowledge. The second trend, interpretivism, depends on interpreting and conducting work in a subjective manner (Mishra & Alok, 2017).

Generally, a research study sketches its aim prior to starting the first step. To reach the aim, researchers should apply a reasoning method that depends on that aim. Common forms of reasoning are inductive and deductive reasoning. Inductive reasoning aims to develop a theory, from the data upward to an output (result), while deductive reasoning intends to test an existing theory, which is a production of prior inductive research. Research designs, then, are based on research reasoning. Research designs depend on the aim of the research and can be descriptive, exploratory, or explanatory. Fundamentally, descriptive research describes a research issue without performing any statistical testing or observative analysis. Explanatory research is performed when testing a relationship between variables, thus generating statistical results and definitive analysis. Exploratory research is based on experiments that help generate new knowledge and, in turn, define findings (Nahar, Al Eroud, Barahoush, & Al-Akhras, 2019).

Ultimately, the conducted research must have a structural framework that enables a researcher to pursue their journey with a systematic approach. Three approaches are commonly used in research: quantitative, qualitative, or mixed methods. The qualitative approach requires descriptive analysis that examines the data. A quantitative approach is concerned with quantitative phenomena, specifically research that handles objects that may be expressed in terms of quantity or countable means. This type of research often requires a more systemic analysis, by conducting experiments and computational techniques, and presenting data in a numeric form, such as percentages, statistics, and formulas (Mishra

& Alok, 2017). The mixed approach is based on both the qualitative and quantitative approaches and is the most common research type.

The research strategy is developed in accordance with the research design, which may be in the form of experiments, case studies, and surveys. To carry out the research, data must be gathered. Data may be collected as a cross-sectional random sample, which means that data from a random sample at a single point in time is collected. Another method of data collection is longitudinal, which collects non-random samples from subjects repeatedly over time in order to study a common trait that connects them. The outcome of these strategies leads to the production of data that must be further analysed.

Data collected may be presented in a quantitative or qualitative form. Accordingly, analysis techniques are employed to make use of this data. For example, quantitative data requires the use of quantitative analysis techniques that rely on calculations defined as frequencies, averages, and correlations, such as descriptive statistics in the form of median, mean, and mode. Quantitative data analysis can also include techniques such as inferential statistics, correlation analysis, and regression analysis. Qualitative data requires the use of qualitative analysis techniques, such as content analysis, thematic analysis, and discourse analysis, to understand words, ideas, and behaviours (Creswell & Creswell, 2017). Figure 4.1 shows details regarding the research paradigms.

## 4.2 Research Design

In the previous section, we explained the research paradigm. However, a crucial pillar of conducting research is the research design, as it allows the researcher to apply the research approaches that are suitable for a successful study. Above, we explained the three common types of research design: descriptive, exploratory, and explanatory. Choosing the appropriate research design is strongly linked to

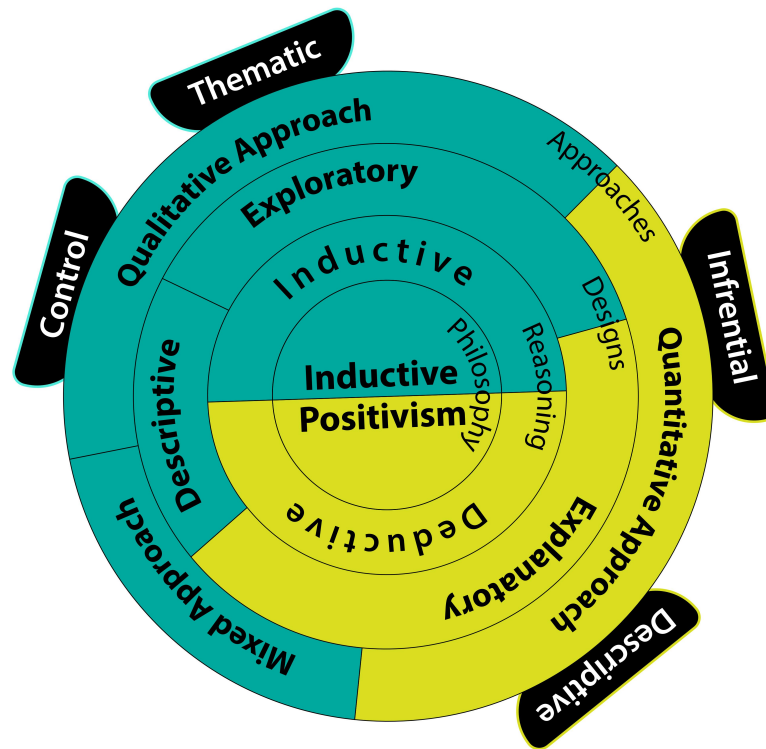


Figure 4.1: Details Regarding the Research Paradigms.

the research aims. The primary objective of this research is to build a supervised machine learning model that classifies real and fake Arabic news articles based on textual analysis. We investigated textual cues in news articles that could point to fake and deceptive text. To answer the research questions presented in Chapter 1, we conducted further work by classifying the following: satire articles, fake articles from real-world cases, and articles from three Arabic speaking countries—Saudi Arabia, Egypt, and Jordan.

Generally, we used the deductive reasoning approach and an exploratory research design, which is based on experiments that help generate new knowledge and define findings. This design was deemed suitable for our research, because studies that build machine learning models based on classification schemes rely heavily on producing a stream of numeric input data (numerical values) that

## Chapter 4. Methodology

are used for model evaluation and analysis (Elhadad et al., 2020; Gröndahl & Asokan, 2019; Rao & Sachdev, 2017). The features extracted from the articles, which are used to further train the model, are converted into numerical data for the computer to analyse. This research design generates reliable and objective data. Further, having the data in numeric form allows it to be processed and analysed in a quantitative analysis approach, which, as explained above, is research concerning objects that may be expressed in terms of quantity or countable means.

Previous studies that have addressed fake news followed a typical automatic data classification scheme (Jeronimo et al., 2019; Tyagi, Pai, Pegado, & Kamath, 2019), where the text was classified as real or fake. They start by collecting data, such as tweets (Helmstetter & Paulheim, 2018), then perform pre-processing steps to compile a clean dataset for training. After that, features are extracted and classified. In most cases, the features are presented as numeric vectors for the computer to comprehend. Finally, a model is tested on these features to enable it to further identify fake articles based on the training it received.

In this respect, a quantitative approach, has been deemed suitable for our research. The construction of datasets that contain news articles, were collected either by crowdsourcing for the primary dataset, or manually for the other datasets. In all cases, the datasets contain data in the form of text. Further, the annotation process addressed in fake news articles depended on the annotator's choices. However, extracting the textual features from the articles generated quantitative data for the computer to work with. This requires quantitative approaches to analyse. Moreover, experiments were performed to examine the applicability of the proposed approach for classification tasks. These experiments rely on highly numeric datasets, all of which are used to gauge objective results.

As shown in Figure 4.2, the research is composed of five main stages, which are part of three phases of the research method. The first phase is data collection,

which includes the stage where data in the form of text articles is collected. Due to the lack of Arabic fake news datasets, we relied on generating fake articles through crowdsourcing. Then, we relied on the annotators' approval for the fake articles. Only articles that met both with their approval were included in the primary dataset. The second phase is the data processing phase, which includes data processing and feature extraction methods and tools. Here, various pre-processing methods, which are crucial when dealing with Arabic language texts, such as normalization and segmentation, were performed. Then, textual features were extracted through NLP tools. The next phase is the data analysis phase, which analyses quantitative results produced in the feature extraction step, utilising using various ML classifiers. Finally, with the exploratory design in mind, we performed experiments, such as testing and evaluating the model's performance. The results of these experiments were analysed through quantitative analysis. This was strictly powered by descriptive and inferential statistics methods to generate new knowledge and assess previous findings on fake news.

### **4.3 Research Method**

The research applies a deductive, based on a positivist research philosophy. Positivist philosophy requires objectiveness in the researcher's manner, with an emphasis on measuring variables and producing numerical data through experiments that lead to new knowledge or support existing knowledge (Tubey, Rotich, & Bengat, 2015). In this section, we detail the three stages of research design mentioned earlier—data collection, data processing, and data analysis. Figure 4.3 illustrates the proposed research framework.





Figure 4.2: Research Design Steps.

### 4.3.1 Data Collection

To train a model in text classification tasks under supervised machine learning, data is an essential component. Due to the lack of available Arabic datasets to train the model, we had to create them through crowdsourcing. This method of data collection was deemed suitable, as we wanted to imitate the production of fake news in the real world. Articles were collected by a cross-sectional probability sampling method. For that, we randomly gathered the articles at one time under the assumption that they represent a sample of bigger data. The real news articles were collected from Arabic news agencies. We collected fake articles upon regulations proposed by Rubin, Chen, and Conroy (2015). To generate fake articles, we relied on crowdsourcing. We assigned annotators to ensure the submitted articles conformed with Rubin et al. (2015) guidelines for fake news creation.

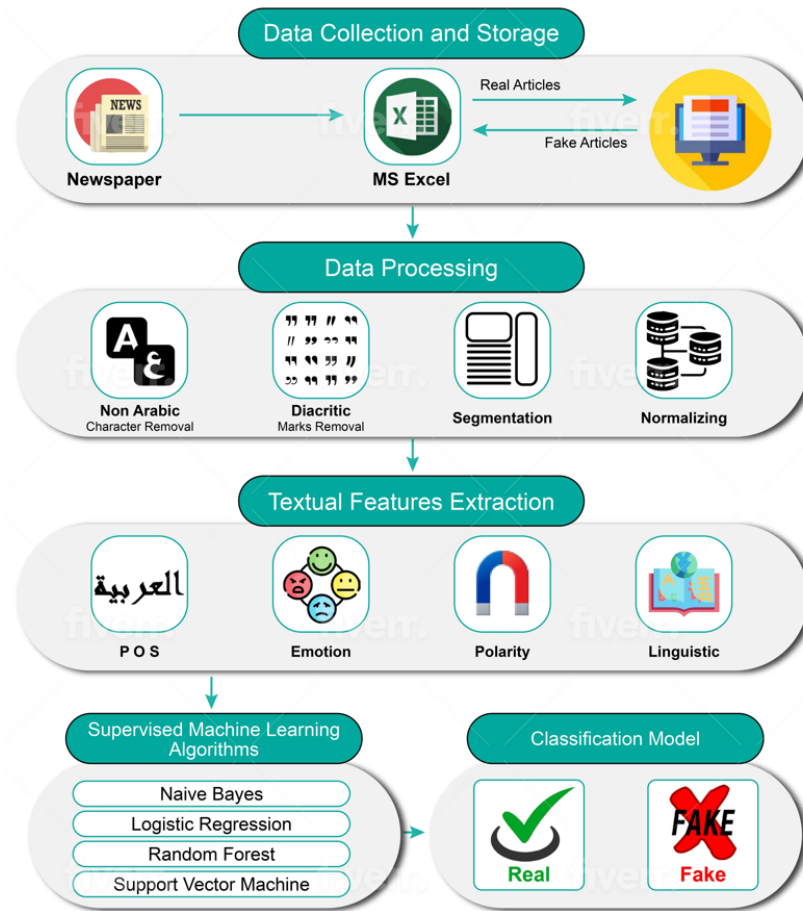


Figure 4.3: Research Framework.

Participants were initially recruited via email from the University of Jeddah. When we did not reach the required number of participants, we posted the task on the Fiverr service platform. The only requirement was for the participants to be Arabic speakers. We did not bar participants from the study based on their age, gender, occupation, or level of education. Rather, we sought diversity among the human participants as well as in the collected fake articles writing style.

In this research, we compiled five datasets. The first dataset included real articles from legitimate news agencies and fake articles generated through crowd-sourcing. The second dataset included satire articles from explicit satire platforms, which were balanced with non-satire articles from non-satire news plat-

forms. The third dataset included fake articles collected from real-world fake articles. The fourth and fifth datasets included news articles about two distinctive topics, the Hajj and Brexit, which were collected from news platforms in three Arabic speaking countries. Through this process, we collected more than 500 fake Arabic articles, more than 400 satire and non-satire articles, and more than 3,000 articles related to the Hajj and Brexit from Arabic news agencies.

### **Verification and Reliability**

To maintain veracity and reliability throughout the data collection, we focused on the data quality. We labelled articles as ‘real’ only after the verification process, such as cross-referencing them through legitimate news agencies and fact-checking websites. Likewise, the fake articles generated were only labelled as ‘fake’, and thus included in the dataset, when they were manipulated by the participants and further approved by the annotators.

The process of constructing fake Arabic articles presented unexpected challenges. First, there was the unavailability of Arabic fake news platforms. In fact, during this study, one fake news platform, Sabk, was blocked in Saudi Arabia within an hour. Second, the number of articles on Arabic fact-checking platforms is limited compared to that on Western fact-checking websites. However, during the COVID-19 pandemic there was slight increase in fake articles, which we managed to include in the Covid dataset.

I encountered different challenges during the process of building the Arabic fake news dataset through crowdsourcing. The first challenge was the lack of interest among potential participants. One reason for this is the stiff penalties Arab governments impose on fake news producers, such as three years jail time, which caused fear in potential participants. To reassure them, we explicitly stated that no personal data would be collected. Also, we explained that the task was purely for research purposes. Appendix C provides the details of the ethics approval

this research. We also encouraged participation in our research by offering a monetary reward to participants. Another challenge was the poor writing of the participants when it came to generating the fake articles. Indeed, around 7% of the submitted fake articles were written in poor, colloquial Arabic. One example used informal words, such as **تاني** ‘second,’ which was written with two dots on the first letter in some articles, while in formal Arabic it the first letter requires three dots **ثاني**. This is a commonly made mistake, as diacritics and dots are often misplaced. To overcome this problem, we set guidelines such as maintaining the journalistic formal writing found in news articles. We also followed the guidelines proposed by Rubin et al. (2015) to generate fake articles according to predefined guidelines that ensure the fake article’s quality and their similarity to fake articles in the real world. Further, the submitted fake articles were annotated by specialised media and journalism focus annotators.

### 4.3.2 Data Processing

#### Pre-Processing

The aim of pre-processing functions is to minimise information loss whilst maintaining maximum data dimensionality. Data pre-processing is defined as cleaning the data from unnecessary and unhelpful information to obtain better results. When dealing with Arabic text, pre-processing is an essential step in data classification. This is due to the unique Arabic morphology, which significantly affects the final results. Arabic diacritics and dots, as discussed in Chapter 3, are commonly misplaced or unmarked, which causes word ambiguity. Another complexity that needs to be addressed is the presence of affixes attached to the words. We rely on exact word matching to extract the textual features (words), and affixes attached to the words can cause mismatching. Therefore, to improve the clarity and word matching of the text, we remove useless information, normalise

the articles in terms of diacritics and clitics, and detach affixes to improve word matching. Various pre-processing steps were performed that handled data in this form. The main pre-processing tasks commonly used were as follows:

- Stemming
- Normalisation
- Segmentation
- Removal of non-Arabic characters
- Removal of diacritic marks

It is important to note that not all NLP tasks give a better impact when applied to text classification models. For example, there is an ongoing debate between studies that reported an enhancement in classification performance when stemming was conducted and others that found lower classification performance with stemming (Al-Anzi & AbuZeina, 2015). We turn now to the output of performing NLP tasks, which generates data that may be further analysed. In fact, for machines to understand the produced data, it must be converted to numeric data, which is used by the machine to distinguish each data from another (Kulkarni & Shivananda, 2019).

### **Textual Features**

Inspired by the fact that an author's writing style may result in language leakage, some of which may signal the presence of deceptive text (Demestichas, Remoundou, & Adamopoulou, 2020), such as hedging words (Islam et al., 2020) and emotion words (Zloteanu, Bull, Krumhuber, & Richardson, 2021) four textual feature sets were composed: POS (10 features), emotional (six features), linguistic (11 features), and polarity (two features). These textual feature sets are composed to analyse the news articles from four different perspectives: POS,

## Chapter 4. Methodology

emotion, polarity, and linguistics. Each of these categories is further explained below.

First, POS features are a word's assigned POS tags that comply with its role in a sentence. These have been effective in detecting fake news (Kapusta & Obonya, 2020; Pérez-Rosas et al., 2017), identifying spammers on Twitter (Alom, Carminati, & Ferrari, 2018; Saeed et al., 2022), and classifying the author's gender in the text (Alsmearat et al., 2017). This textual feature category is capable of producing a comprehensive set of markers to investigate the text, as they are the main blocks used to create statements. However, the analysis was limited to those specific POS that previous studies found useful in deception detection (Kapusta & Obonya, 2020; Horne & Adali, 2017), based on their role, as discussed in Section 2.4.1. The first set of POS features are content words, which include nouns (Bondielli & Marcelloni, 2019; Horne & Adali, 2017; Kapusta et al., 2020; Pérez-Rosas et al., 2017), verbs (Bondielli & Marcelloni, 2019; Kapusta et al., 2020; Pérez-Rosas et al., 2017), adverbs (Adha, 2020; Kapusta & Obonya, 2020), adjectives (Adha, 2020; Seih, Beier, & Pennebaker, 2017), and proper nouns (Horne & Adali, 2017). The second set of POS features is function words, which include the following: conjunctions, prepositions, determiners, pronouns, particles, and interjections. Function words were useful indicators to detect deceptive text in various studies (Chakraborty, Paranjape, Kakarla, & Ganguly, 2016; Demestichas et al., 2020). The set of POS tags is explained using examples in Table 4.1.

Second, emotional features refer to the level of feelings displayed within a given text. Fake news is considered to be lies expressed by writers to readers, and several studies (Sayyed, Sugave, Paygude, & Jazdale, 2021) have associated emotional language use with lies. Early studies suggested that liars tend to cover their lies by embedding emotional language within their writings (Levenson, Ekman, & Friesen, 1990). In previous studies, different emotions have been found to be

Table 4.1: POS Tags Used in this Research, With Examples.

Content Words		
Nouns	Verbs	Adjective
رجل / امرأة / حاج	صعد / نام / صلى	بارد / حار / عال
Man / woman / pilgrims	Climbed/ sleep / prayed	Cold / hot / high
Adverbs	Proper Nouns	
أيضا	أحمد / اليمن / مصر	
Too / also	Ahmad / Yemen / Egypt	
Function Words		
Conjunctions	Prepositions	Pronouns
و / ف / ثم	مع / في	ي / أنت / هي / نحن
And / thus	In / on	We / she / you (singular masculine) / you (singular feminine)
Interjections	Particles	Determiners
أوه	كي ، حيث	ال
Ooh	For that, since	The

related to deceptive text and, in some cases, fake news. More specifically, studies by Jupe et al. (2018) and Hancock et al. (2007) found that liars tend to use more emotional words to hide their lies. Other studies, such as that by Torabi Asr and Taboada (2019), found that emotional language used in news sources may be a good cue for detecting fake news. Meanwhile, some studies found a correlation between the use of highly emotional words and fake news (Baptista & Gradim, 2020; L. Zhou et al., 2004).

Therefore, six essential human emotions—anger, disgust, fear, sadness, joy, and surprise (Levenson et al., 1990)—as described in Table 4.2, were used to analyse the emotional state of each article.

Third, polarity in the deceptive text is not simply being marked by more emotional language, but also by potential positive or negative impacts or actions.

Table 4.2: Emotion Words Used in This Research, With Examples.

	Anger	Sadness	Fear	Disgust	Joy	Surprise
<b>Description</b>	Feeling angry	Feeling sad	Feeling of being scared, frightened	Feeling something is nasty or not right	Feeling happy, joyful	The feeling when something happens unexpectedly
<b>Example (Arabic)</b>	غضب / حنق	بكاء / هم	مروع / مخيف	مقت / نفور	سعيد / فرح	ذهول / مندهش
<b>Example (English)</b>	Angry / exasperation	Worry / cry	Scary / horrific	Disgust / loath	Happy / joyful	Amazement / surprise

Table 4.3: Polarity Words Used in This Research, With Examples.

	Negative	Positive
<b>Description</b>	Words or phrases used to express negative or interrogative context	Words or phrases used to express positive or affirmative context
<b>Example (Arabic)</b>	وسخ / لئيم / بخيل / ثورة	حكيم / مجاني / فاخر / راقى
<b>Example (English)</b>	Dirty / stingy / revolution / mean	Wise / free / luxurious / classy

Hence, polarity includes feelings similar to the emotions described above, plus the associated actions or outcomes of these emotions. For example, an outcome or action of the emotion of joy would be to dance. This can be considered a positive polarity. We use the term ‘polarity’ rather than ‘sentiment’, as the latter is usually linked to scores that measure sentiment, which is not the focus of our research. We rely on the word’s presence only.

By examining these features, researchers have been able to link fake news articles with high polarities (Salgado & Bobba, 2019; Soroka & McAdams, 2015). These features are described in Table 4.3.

Fourth, linguistic features are certain syntactic categories that are too fine-grained to be captured by general POS. Each syntactic unit conforms to a certain linguistic purpose, which is used to build meaningful statements. In recent years, there has been an increasing amount of literature investigating authors’ writing



## Chapter 4. Methodology

styles to identify unique features associated with their writing and to identify certain characteristics (Alsmearat et al., 2017; Alwajeeh, Al-Ayyoub, & Hmeidi, 2014; Burgoon et al., 2003; Gröndahl & Asokan, 2019; Hajja, Yahya, & Yahya, 2019). In this study, the set of linguistic markers investigated, as described in Table 4.4, is as follows: assurance (Sabbbeh & Baatwah, 2018), negations (Hancock et al., 2007; Horne & Adali, 2017; Newman et al., 2003; Pérez-Rosas et al., 2017; Rashkin et al., 2017), justification (Gravanis et al., 2019; Gröndahl & Asokan, 2019; Hancock et al., 2007; Jupe et al., 2018; Newman et al., 2003; Pérez-Rosas et al., 2017; Reis et al., 2019; Resende et al., 2019; L. Zhou et al., 2004), intensifiers (Gröndahl & Asokan, 2019; Karoui et al., 2017; Pérez-Rosas et al., 2017), hedges (Argamon et al., 2007; Volkova et al., 2017; Wei et al., 2016; Addawood et al., 2019; Rashkin et al., 2017), illustrations (DePaulo et al., 2003; Feng & Hirst, 2013; Lagutina et al., 2019; Rashkin et al., 2017; Li et al., 2014), temporal (Vrij et al., 2000; Davis et al., 2016; Humpherys et al., 2011; Reis et al., 2019; Resende et al., 2019; Jupe et al., 2018), spatial (Jupe et al., 2018; Humpherys et al., 2011; Vrij et al., 2000), superlative (Conroy et al., 2015; Hancock et al., 2007; Karoui et al., 2017; Vartapetian & Gillam, 2012), exceptions (Ott et al., 2011; Newman et al., 2003), and oppositions (Conroy et al., 2015; Hancock et al., 2007; Karoui et al., 2017; Vartapetian & Gillam, 2012).

### **Processing Tools**

Successful research with textual datasets is attributed to the fact that NLP techniques allow for numerous methods and styles to classify text. Before detailing the NLP tools used in this research, we elaborate on the creation of emotion, polarity, and linguistic wordlists. The linguistic wordlists were organised and further approved by three academics in the Arabic studies department at Umm–AlQura University in Makkah, Saudi Arabia, to avoid biased word inclusions. Further, the emotion and polarity wordlists were created from available emotion and po-

Chapter 4. Methodology

Table 4.4: Linguistic Features Used in This Research, With Examples.

	<b>Assurance</b>	<b>Negations</b>	<b>Justification</b>
Description	Words used to indicate certainty	Words used to indicate that something is not of the specified	Words that show cause
Example (Arabic)	أن - عين - نفس	لا - لن - لم	سبب - لذلك - من أجل - لام التعليل
Example (English)	for sure, surely, certainly	no, not, never	Because /to /for that
	<b>Intensifiers</b>	<b>Hedges</b>	<b>Illustration</b>
Description	Words that increase intensity of another word	Words that express uncertainty, hesitation	Words used to display or show example
Example (Arabic)	تماما - جميع - جدا	من الممكن - يجب - احتمال	مثال - مثل
Example (English)	very, too, not at all	maybe/should/could/may	for example
	<b>Temporal</b>	<b>Spatial</b>	<b>Superlative</b>
Description	Words that indicate time	Words that indicate place	Words that show the highest degree of comparison
Example (Arabic)	مثال - مثل	تحت - فوق - عند	أفضل - أكبر - أسعد
Example (English)	yesterday, tomorrow	under, over	Best, biggest, happiest
	<b>Exception</b>	<b>Opposition</b>	
Description	Words used to express an omission of something	Words that indicate adversity	
Example (Arabic)	إلا - عدى - سوى	لكن - إنما	
Example (English)	except	However, but	

larity lexicons that were used in various Arabic language projects (Mohammad et al., 2016; Karoui et al., 2017). Chapter 3 provided details of the wordlists.

### **The Tasaheel Tool**

Tasaheel was designed to conduct several NLP functions in Arabic. Some of these functions were provided by packages coded in Python that were available online but scattered in several Arabic research platforms. Our aim was to collect and join several NLP packages that supported Arabic, which provided a comprehensive research utilities tool. Details of these packages are provided in Appendix A.

Pre-processing was performed using NLP functions available in the Tasaheel tool. Specifically, the Farasa package embedded in Tasaheel was used to perform segmentation. Farasa segmenter performance proved to be significantly better in terms of accuracy and speed, compared to MADIMARA and StanfordNLP segmenters, on information retrieval tasks (Al-Yahya, Al-Khalifa, Al-Baity, AlSaeed, & Essam, 2021; Haouari et al., 2020). An additional stemming task was only performed on a copy of the primary dataset to answer research question 5: What is the effect of stemming articles on the model’s performance? For stemming, we used the Arabic light stemmer proposed by Abd, Khan, Thamer, and Hussain (2021), which outperformed five other Arabic stemmers (Khoja, ISRI, Assem, Farasa, and Tashaphyne) in their study. Tasaheel also provided normalisation tasks which were used in this research. Since there were no available utilities to remove diacritic marks and non-Arabic characters, we created a Python code to remove non-Arabic characters, such as commas, URLs, and email signs, and to remove diacritic marks, which were also embedded in Tasaheel.

I also added novel functions, such as emotion, polarity, and linguistic word tagging. This function was based on the foundations of previous related work on building wordlists and word tagging methods, as detailed in Chapter 3. We also made use of the Farasa POS tagger embedded in Tasaheel to tag our datasets. Farasa was chosen for this task due to its high performance in a study by Alluhaibi, Alfraidi, Abdeen, and Yatimi (2021), where it achieved an 86% F-score when tagging 10 text samples, compared to 67% by StanfordNLP and 84%

by Camel. However, Farasa does not explicitly tag proper nouns and interjections; they are tagged as regular nouns. For this reason, we relied on another tool for this task, which is discussed below.

### **The Posit Tool**

Posit was developed by George Weir of the Department of Computer Science at the University of Strathclyde. It is designed to operate under Unix generate quantitative text analysis, engendering frequency data and POS data tagging, whilst accommodating large text corpora. Posit is specifically designed to analyse English language texts, as it uses the Lapos English tagger. However, a new version upgraded its capability by applying POS tagging to the Arabic language using the StanfordNLP POS tagger-3.80 packages. The tag set includes tags of English that comply with Arabic modules and specifically created tags that are applicable to Arabic without considering affixes or inflections.

Amongst its capabilities, Posit generates data that includes the following: values, nouns, verbs, adjectives, total words (tokens), total unique words (types), type/token ratios, number of sentences, average sentence length, number of characters, average word length, noun types, verb types, adjective types, adverb types, preposition types, personal pronoun types, determiner types, possessive pronoun types, interjection types, particle types, prepositions, personal pronouns, determiners, adverbs, adjectives, possessive pronouns, interjections, and particles. There are 27 features in all (G. R. Weir, 2009). Posit also generates a supporting detailed POS summary file for each text file and displays the occurrence of each type of noun (common or proper) and verb (present or past), as well as the occurrence of cardinal numbers, as shown in Figure 4.4. This tool has been successfully used to extract textual data in research focused on detecting fake news (Cartwright, Weir, & Frank, 2019). In this research, Posit was used to tag proper nouns and interjections in the articles, and for searching through tagged

```
{fake_cor/fake2_cor.txt :Input filename
20 :Total words (tokens)
1 :Total unique words (types)
20 :Type/Token Ratio (TTR)
1 :Number of sentences
20 :Average sentence length (ASL)
122 :Number of characters
6.1 :Average word length (AWL)

NUMBER OF POS TYPES
3 :noun_types
2 :verb_types
1 :personal_pronoun_types
1 :adjective_types
0 :preposition_types
0 :possessive_pronoun_types
0 :particle_types
0 :interjection_types
0 :determiner_types
0 :adverb_types

NUMBER OF POS TOKENS
8 :nouns
5 :verbs
2 :adjectives
1 :personal pronouns
0 :prepositions
0 :possessive pronouns
0 :particles
0 :interjections
0 :determiners
0 :adverbs
```

Figure 4.4: Posit Summary File.

adjectives for superlatives.

### Limitations

As stated in the previous section, we relied on Posit and Tasaheel to extract textual features from the articles. Since we relied on exact word matching during the emotion, polarity, and linguistic features extraction, we encountered the problem of being unable to extract an affix that serves a justification purpose. The same

issue was encountered when extracting superlatives. There is a large number of superlatives, which makes collecting them in one wordlist almost impossible. Thus, we made use of their morphology in the form of their affixes. For this task, we designed a code in Python, which was added as a feature in Tasaheel, to search for the desired affix in the POS tagged files produced by Posit, under the Stanford NLP tagger and Farasa. Searching through the POS tagged files for the desired affixes will narrow down the target domain that holds the desired affix.

Another limitation was the issue of homographs. As stated in Section 3.1.2, homographs are an issue when it comes to the removal of diacritics in Arabic text. To overcome this issue, we manually revised all of the results of matched words to ensure that only the target words were included.

### 4.3.3 Data Analysis

Extracting the textual features produced numerical data that was further used for model training and testing in the data classification phase. This involves testing and evaluating the model's performance. Following a quantitative approach, with the aim of exploratory research design, several experiments were conducted, and both descriptive statistics and inferential statistics were proposed. First, experiments were conducted using several ML classifiers with the aid of various tools, as shown in Figure 4.5.

#### Machine Learning Classifiers and Metrics

Classifiers are data mining algorithms used to classify data into predefined classes. From previous related works in data classification, it was found that the Naive Bayes, Random Forest, Support Vector Machines, and Logistic Regressions algorithms performed successfully in general (Al-Barhamtoshy et al., 2019; Elhadad et al., 2020; Penuela, 2019) and in fake news classification in specific (Choudhary, Jha, Saxena, & Singh, 2021; George, Skariah, & Xavier, 2020).

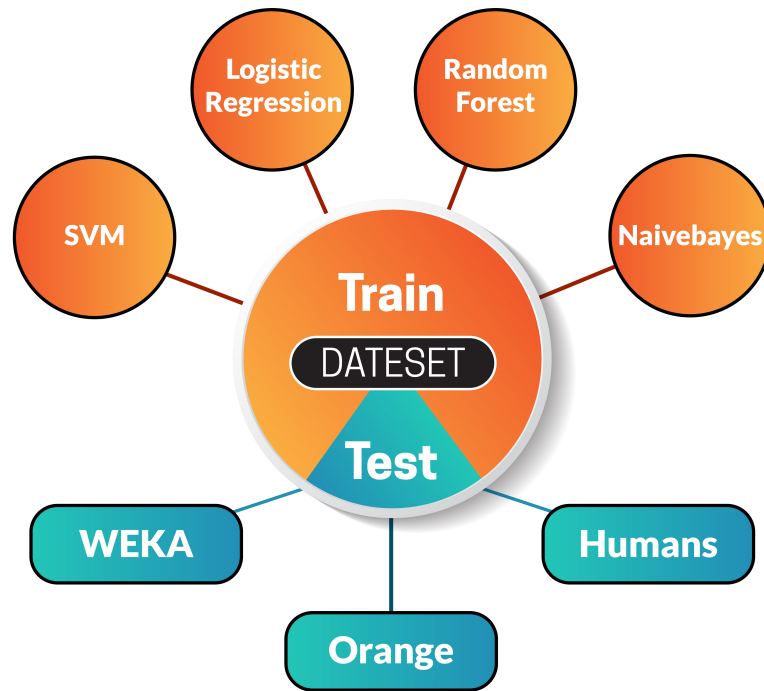


Figure 4.5: Tools in Blue and Metrics in Orange, Used For Model Evaluation.

To avoid any bias predictions from the model or data overfitting, two ML model evaluation methods are commonly used: hold-out or cross-validation. In the first method, the dataset is randomly divided into three subsets: training, validation set, and test set. The training set is used to train and thus build the predictive model, whereas the validation set is a subset of the dataset used to assess the performance of the compiled model in the training phase. This set provides the platform for fine-tuning the model's parameters and, hence selecting the best performing model. The test set is often called the unseen set, as it is a subset of unseen examples of the dataset used to assess the likely future performance of the model (Refaeilzadeh, Thang, & Liu, 2008).

Cross-validation is based on a concept of randomly dividing data into subsets of equal sizes. The model is built n times, each time leaving out one of the subsets of training data and using it as the test set. Usually, the process is repeated until

each group has been used as the test set (Refaeilzadeh et al., 2008).

The proper testing of the models is crucial. Indeed, there is no one-size-fits-all method that applies to all types of models; rather, the testing and evaluation is considered based on the research necessities and both methods are used. Cross-validation is used to find the best model. In this research, we performed the holdout method, which provides a generalised performance of the proposed model and on other unseen data.

### **Support Vector Machine (SVM)**

SVM algorithms are supervised machine learning models that classify input data based on dimensional surfaces, specifically by finding the maximum separating hyperplane between different classes (Vijayan, Bindu, & Parameswaran, 2017). It revolves around the notion of a ‘margin,’ where either side of the hyperplane separates data classes (Kotsiantis, Zaharakis, & Pintelas, 2007). SVM provides data analysis for both regression analysis and classification analysis, and it carries out plotting in the n-dimensional space where the value of each feature is also the value of a certain coordinate. After that, SVM finds an appropriate boundary that maximizes the distance between the closest members of separate classes, as shown in Figure 4.6 (Xie et al., 2017). A strength is its ability to resolve overfitting, especially in a high dimensional space. Due to this, it can model non-linear appropriate boundaries by choosing from many plots available. It has been widely used in text classification projects and proven its applicatory in such projects (George et al., 2020; Shaji, Binu, Nair, & George, 2021) and was thus used in our research.

### **Naïve Bayes**

This simple, probabilistic classifier works by assuming a conditional relation between features of the given data (Vijayan et al., 2017). The classifier is a collec-



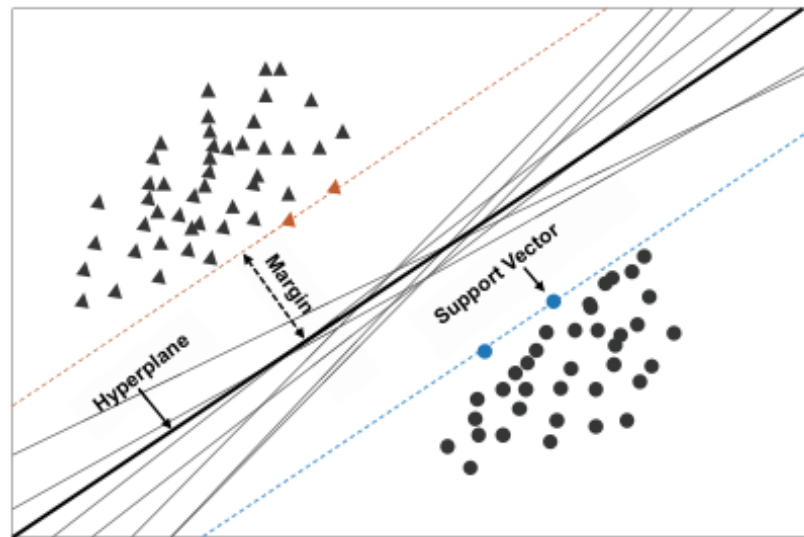


Figure 4.6: SVM Algorithm.

tion of algorithms based on the Bayes Theorem, assuming that all attributes are strictly independent. Thus, it is based on estimating where the model updates its probability table through the training data and predicts new observations by estimating the class probability in the probability table based on its feature values. The classifier's advantages include the fact that it requires little training data which in turn requires less storage space. Moreover, it is naturally robust with regards to missing values, which it ignores in estimating the probabilities, so they have no impact on the final decision. Further, the classifier is considered to be a fast-learning algorithm, which gives faster classification outcomes (Osisanwo et al., 2017). Due to these advantages, Naive Bayes was a deemed suitable choice for the current research.

### Logistic Regression

Logistic Regression: A classifier that finds a relation between features and the probability of a certain outcome (Pramanik, Pal, Mukhopadhyay, & Singh, 2021). It usually states where the boundary between the classes is by using a single

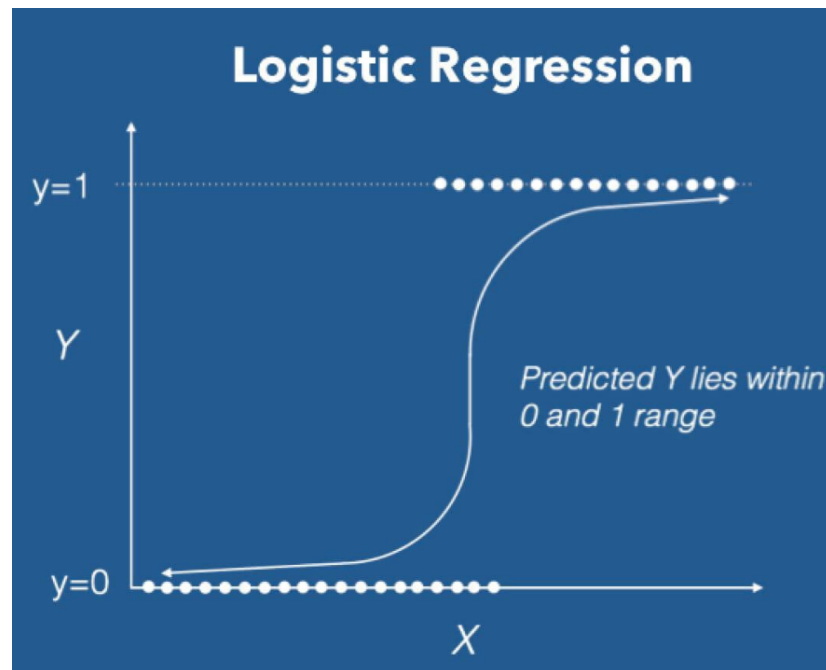


Figure 4.7: Logistic Regression Algorithm.

estimator. The estimator predicts the probabilities depending on the distance from the boundary (Osisanwo et al., 2017). Its main logic is based on logistic function—an S-shaped curve that takes any real valued number and maps it to values ranging between 0 and 1, as shown in Figure 4.7 (Learning, 2021). One of the classifier’s primary advantages is that it is reliable when it comes to solving binary classification problems, since it predicts the probability of input data having only two values (Grover, 2022). Another advantage is that it can be regularised to avoid overfitting. Because of these advantages, this classifier was used in our work.

### Random Forest

This is a meta-estimator classification algorithm that builds a ‘forest’ using decision tree model learning based on bagging techniques (Genuer, Poggi, Tuleau-Malot, & Villa-Vialaneix, 2017), as shown in Figure 4.8 (S. Gupta, 2020). Bag-



Table 4.5: The Fake News Confusion Matrix.

Type	Prediction	
	Real	Fake
Real	True Positive (TP_R)	False Negative (FN_R)
Fake	False Positive (FP_R)	True Negative (TN_R)

assumed that the performance quality of these algorithms is enough to train and test supervised machine learning based models, including those proposed in this research.

The performance evaluation of each classifier is based on average accuracy (A), precision (P), average recall (R), and average F-Measure (F). All measures are computed from the confusion matrix created when performing the classification task. The confusion matrix for fake news detection classifiers is shown in Table 4.5.

The four probabilities are listed below.

1. True Positive (TP\_R): represents the number of real articles correctly classified as real articles.
2. False Negative (FN\_R): represents the number of real articles incorrectly classified as fake articles.
3. True Negative (TN\_R): represents the number of fake articles correctly classified as fake articles.
4. False Positive (FP\_R): represents the number of fake articles incorrectly classified as real articles.

The precision (P), recall (R), and F measures are indicated, respectively, in equations (4.1) – (4.4):

**Precision:** It is calculated as a measure of correctly identified positive cases from all of the predicted positive cases. It is especially useful to inspect the cost of False Positives (FP).

$$Precision(P) = \frac{TPR}{TPR + TNR} \quad (4.1)$$

**Recall:** Recall is calculated as the number of correctly identified positive cases out of all the actual positive cases. It is especially useful for inspecting the cost of False Negatives (FN).

$$Recall(R) = \frac{TPR}{TPR + FNR} \quad (4.2)$$

**F- Measure:** It is the harmonic mean of precision and recall.

$$F - Measure(F) = \frac{2.P.R}{P + R} \quad (4.3)$$

**Accuracy:** It calculates the ratio of sum of true positives and true negatives out of all the predictions.

$$Accuracy(A) = \frac{TPR + TNR}{TPR + FNR + TNR + FNR} \quad (4.4)$$

Descriptive statistics are employed in the form of calculating the lexical densities for classes in each dataset. The intention of this step is to help understand the finer details on the dataset, namely, the differences between word use in real and fake articles. In support of this, several studies have compared word usage in real and fake articles, such as those by Kapusta and Obonya (2020) and Horne and Adali (2017), to gain a better insight into the writing characteristics of deceptive text. In this context, lexical densities are calculated as follows (Yang et al., 2018):

Lexical Density for each category feature in a class (L) =

$$\frac{(\text{total number of occurrence of each feature category in a class}) \times 100}{\text{total number of words in the whole class}} \quad (4.5)$$

Finally, to check the influence of each feature category on our model's perfor-

mance, we apply feature selection methods to identify important features. Feature selection methods are commonly used when the number of features extracted is too high. Reduction or selection of the most discriminating features is useful to enhance the model's performance (Al-Ayyoub et al., 2017). The most common methods include chi-square and principal component analysis (PCA). Principal component analysis is a widely used technique that reduces the dimensions of a feature set by using a linear transformation (Karamizadeh, Abdullah, Manaf, Zamani, & Hooman, 2013). Chi-square investigates the relationships between categorical variables; it calculates the correlation between each feature variable and the target class. Its significance is based on a predefined threshold (usually 0.05). The PCA and chi-square methods have both proven their usefulness for feature reduction in text categorization (Zhai, Song, Liu, Liu, & Zhao, 2018; Taloba, Eisa, & Ismail, 2018).

The formula of the Chi square feature selection is shown in algorithm:

$$X_c^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (4.6)$$

where  $c$  is the degree of freedom (threshold value),  $O$  is the observed value,  $E$  is the expected value, and  $X^2$  is chi-square computed result for the feature.

We chose the chi-square method to demonstrate the most prominent features for the model's performance. The choice is based on the fact that chi-square makes general text classification feature selection not only possible, but also relatively straightforward, simple, and effective (Sanasam, Murthy, & Gonsalves, 2010). Specifically, Ayed, Labidi, and Maraoui (2017) used it for feature selection in Arabic text classification study, and it performed well.

### 4.3.4 Experimental Tools

#### WEKA

Waikato Environment for Knowledge Analysis (WEKA)<sup>1</sup> is a Java-language ML tool developed at the University of Waikato, New Zealand. It is a free data analysis and predictive model software application. For ease of access and visualisation, it provides a graphical user interface. The tool meets demanding real-world applications by carrying out big data tasks, including data processing, classification, clustering, and visualisation. It also supports several algorithms, such as logistic regression, linear regression, Naive Bayes, decision tree, random tree, random forest, decision rule, and neural network. It requires files to be converted into the attribute relation file format (ARRF). WEKA has been widely and successfully used to evaluate several ML models (Adetunji, Oguntoye, Fenwa, & Akande, 2018; Cartwright, Nahar, et al., 2019; G. Weir et al., 2018). It also provides several experiment and evaluation environments, namely percentage split and cross-validation. Further options include testing using the full training set, or testing unseen dataset as a supplied test set. The variety of ML classifiers it provides, along with the specific metrics that enable it to easily optimise parameters, makes this an optimal tool for our research.

#### Orange

Orange<sup>2</sup> is a Python-programmed, component-based data mining software. It provides a range of data visualisation, exploration, pre-processing, and modelling techniques. Similar to WEKA, the software also provides open-source services and has an effective graphical user interface. Uniquely, the software provides drag-and-drop components that perform several data mining tasks. It provides a word cloud visualisation, which may be useful when it comes to analysing the most

---

<sup>1</sup><https://www.cs.waikato.ac.nz/ml/weka/>

<sup>2</sup><https://pypi.org/project/Orange3/>

frequent words in documents. The software mainly fulfils our research demands in terms of this visualisation.

### **Humans**

The proposed model was evaluated by subjects who were not involved in the dataset generation. Subjects were recruited via email from the University of Jeddah's female branch. Appendix C provides the ethics approval. We sent an email through the university's email system to recruit computer science (CS) students to participate in this project, to ensure that the participants were educated and had prior experience using computers. In the end, a total of 10 students agreed to participate.

The subjects were all female students with an average age of 20. Due to the restrictions of Covid-19, such as a ban on meetings of more than 10 people that came into effect a month later, we only conducted this test with a limited number of women and was unable to test it on a larger and more diverse population. To ensure the non-distribution of the generated fake articles, a classroom was organised to meet the participants. Printed handouts of the articles were given to each of the participants. The handouts were composed of 10 random articles: five real articles and five fake ones. We asked the participants to indicate (F) beside the articles they judged as fake and (R) besides the ones they judged as real. We instructed the participants not to write their names or other personal data on the handouts.

The participants' answers were noted in an Excel sheet that contained all 10 articles. Besides each article, We indicated the number of students who correctly predicted the status of the article.



## 4.4 Validity and Reliability

Validity and reliability were ensured throughout this research. First, during data collection, the real articles went through various veracity procedures and were collected following the guidelines of Amjad et al. (2020). Moreover, the fake articles included in the primary dataset were approved by media and journalism specialized annotators, which ensured the dataset’s quality. The other datasets were comprised by collecting articles from platforms that explicitly conformed with the dataset purpose – satire platforms in the case of satire articles and fact-checking platforms in the case of fake articles. Further, we ensured that the country-of-origin datasets, Hajj-Co and Brexit-Co, only included articles about their topics. This was done by coding a Python script that filtered out articles that did not contain the keyword حج ‘Hajj’ for the Hajj-Co dataset, or بريکست ‘Brexit’ for the Brexit-Co dataset.

For the feature extraction phase, we relied on NLP tools that support Arabic, and we only used tools that had been effective in previous studies. Finally, to ensure the validity of our testing and evaluation, we reviewed the datasets, more than once, to ensure that each article was labelled correctly depending on its class. Moreover, the selected features are considered valid, as they are used and cited in the related literature.

## 4.5 Summary

This chapter outlined the fundamentals of research in general and the requirements and basic implementation details of our model’s construction. As discussed, a quantitative approach was considered suitable to our research. The data collection was conducted manually or through crowdsourcing and the quantitative approach for the processing and analysis of said data. The next chapter provides a more detailed account of the data collection.

# Chapter 5

## Gold Standard Datasets

A dataset is needed to train the model to classify real and fake news articles. The objective of this chapter is to explain how a dataset for Arabic fake news was built. We adopted a novel approach in building the Arabic fake news dataset by imitating the production of fake news in reality. The most important aspects of a reliable dataset are veracity and article length, which are discussed in subsequent sections. For this research, five datasets were created: the real\_fake dataset, the COVID-19 fake news dataset, the satire dataset, and the country-of-origin datasets, Hajj\_CO and Brexit\_CO. The first three datasets were used to train and test the model in order to classify the articles as real / fake or satire/ non-satire. The latter two datasets were constructed to train and test the model in order to classify the articles based on the country of origin.

### 5.1 Dataset 1: real\_fake Dataset

This section details the process of collecting the real articles. We present the sources of real news collection, challenges in terms of the veracity and length of the real news collected, and how we overcame these challenges.

### 5.1.1 Real News Collection

To collect a diverse range of articles, real news articles were extracted from Arabic-language news platforms in three predominantly Arabic speaking geographical countries: Saudi Arabia, representing the Arabian Gulf; Egypt, representing North Africa; and Jordan, representing the Mediterranean. We focused on the topic of ‘the Hajj’ حَجّ to collect the real articles. Performed by millions of Muslims worldwide, the Hajj is the annual pilgrimage to the Holy city of Makkah in Saudi Arabia and one of the Five Pillars of Islam. This topic is covered in the news worldwide and is significant in several domains<sup>1</sup>. In addition, the topic is seen in articles on subjects that include politics, economics, and sports, and each has a reasonable amount of information. A Python scrapper searched these news platforms with ‘Hajj’ as the target keyword, spanning the period from October 2016 to December 2019. Through this process, a total of 1,200 news articles were collected. Table 5.1 details each news platform’s name, description, country of publication, and the number of articles collected.

#### Quality of Real News

I removed all duplicate articles, articles where the Hajj was not mentioned, and all blogs that were captured by the scrapper which were mistaken for an article. However, collecting real news articles involved two setbacks: veracity and the content length. These are examined below.

#### Veracity of Real News Articles

The issue of veracity in the selection of news articles and labelling them as ‘real’ has been a matter of dispute in several studies (Horne & Adali, 2017; Ireton & Posetti, 2018). However, unfortunately, the truth remains hard to verify with any degree of certainty. According to Lim (Lim, 2018), even fact-checking websites,

---

<sup>1</sup><https://haj.gov.sa/en>

Table 5.1: Real News Collection Sources.

Newspaper	Description of Newspaper	Country of Publication	Number of Articles
Okaz	Okaz عكاظ is a Saudi Arabian daily Arabic newspaper launched in 1960 and based in Jeddah. The paper, which has several offices in Saudi Arabia, is printed simultaneously in both Riyadh and Jeddah. There is also an online web version. The paper covers several distinct topics and is considered the most popular paper in Saudi Arabia.	Saudi Arabia	200
Masrawy	Masrawy مصرأوي is an Egyptian daily Arabic newspaper launched in 1999 and is based in Cairo, with offices in Giza and Cairo. It covers news from Egypt and the Middle East. It also reports on international topics and is considered the second-most popular newspaper in Egypt.	Egypt	200
AmmonNews	AmmonNews وكالة عمون is a Jordanian daily newspaper launched in 2011 in Amman. It covers news on Jordan, the Middle East, and international topics.	Jordan	200
AlRiyadh	AlRiyadh الرياض is a Saudi newspaper launched in 1965 in Riyadh. It covers a wide range of Saudi topics and international headlines daily. It produces both a printed and an online version of the newspaper.	Saudi Arabia	200
Youm7	Youm7 اليوم السابع is an Egyptian newspaper published online. Launched in 2011 in Cairo, it covers Egyptian and international topics.	Egypt	200
AdDoustur	AdDoustur الدستور is a Jordanian online and printed daily newspaper. Founded in 1975 in Amman, Jordan, the newspaper covers local and international topics.	Jordan	200
<b>Total</b>	1,200		

such as Fact Checker and PolitiFact, have low inter-rater reliability agreement scores between fact-checkers. Thus, Lim (2018) suggested that the accuracy of fact checking of a news statement is almost impossible to verify, even amongst fact-checkers (Lim, 2018). Several approaches have been adopted when verify-

## Chapter 5. Gold Standard Datasets

ing articles: fact-checking, cross-referencing articles, and obtaining articles from legitimate news agencies. In the latter case, the degree of ‘legitimacy’ depends on the reader’s assumption of credibility of news agencies. However, this may be somewhat biased, as there is no specific measure for verifying the legitimacy of news agencies.

In this context, several studies have relied on collecting real news articles from fact-checking websites (Ali et al., 2021; Haouari et al., 2020). Meanwhile, some scholars, such as Pérez-Rosas et al. (2017), cross-referenced all the articles with other resources to ensure their truthfulness. Notably, Amjad et al. (2020) proposed a unique method to verify collected news articles. Their method, which is followed in this study, offers comprehensive veracity measures, as listed below:

1. The article was published by a reliable newspaper.
2. The same news can be found in other news sources that mention the original article.
3. There is supporting metadata, such as pictures, authors, and dates of the article publication.
4. There is a collocation between the title and content of the article; the article has to be read to confirm this.
5. The article’s source is mentioned and is reliable.

I collected the articles from news agencies that have been established for 10 or more years, assuming that their longevity in the industry may add credibility. We also made sure that the articles’ news platforms were supervised under the ministry of media, or equivalent government agencies, in their countries of origin in order to ensure that the news platform was legitimate. Collecting the articles from reliable sources is essential to satisfy points (1) and (5) in the above criteria. Regarding point (2), our efforts to further ensure the veracity of the real articles

collected by the scrapper, beyond cross-referencing fact-checking platforms, involved manually cross-referencing each article with several sources, following the approach outlined by Pérez-Rosas et al. (2017). In order to perform the cross-referencing promptly, we extracted three main points in each article as the key points: who (the main character in the article), what (the main event), and the date of the published article. We matched these key points for each article with archived articles in Google News and on official government websites, such as for example that of the Ministry of Hajj in Saudi Arabia. Even though the matched articles might not be written in exactly the same way, they convey the same information. For example, a collected real article published on 13 December 2019 detailed a meeting between the Saudi Hajj Minister and the United Arab Hajj Minister discussing the Hajj services for United Arab pilgrims. Figure 5.1 shows a matching archived article extracted from the Ministry of Hajj and Umrah in Saudi Arabia with the date of 11 December 2019, which was positively cross checked. Any article that did not have at least one cross-reference was discarded, as it was considered unreliable. The cross-referencing process ensured points (2) and (3), as we also made use of the dates associated with the articles as metadata. Finally, since all of the articles were analysed following this process—which mandated reading the article to confirm the collocation between title and content—point (4) from Amjad et al. (2020) study was also ensured. The mapping process of veracity is explained in Table 5.2.

The culling resulted in 900 verified real articles (300 from each country) from the three countries, Saudi Arabia, Egypt, and Jordan.

As another veracity measure, we verified the articles against fact checking platforms. Unfortunately, there is no unified Arabic fact-checking platform. Instead, nation-dependent, government-monitored outlets check published news or news circulated on social media posts and which are distributed locally. When encountering false news, these fact-checking platforms dynamically add it to their

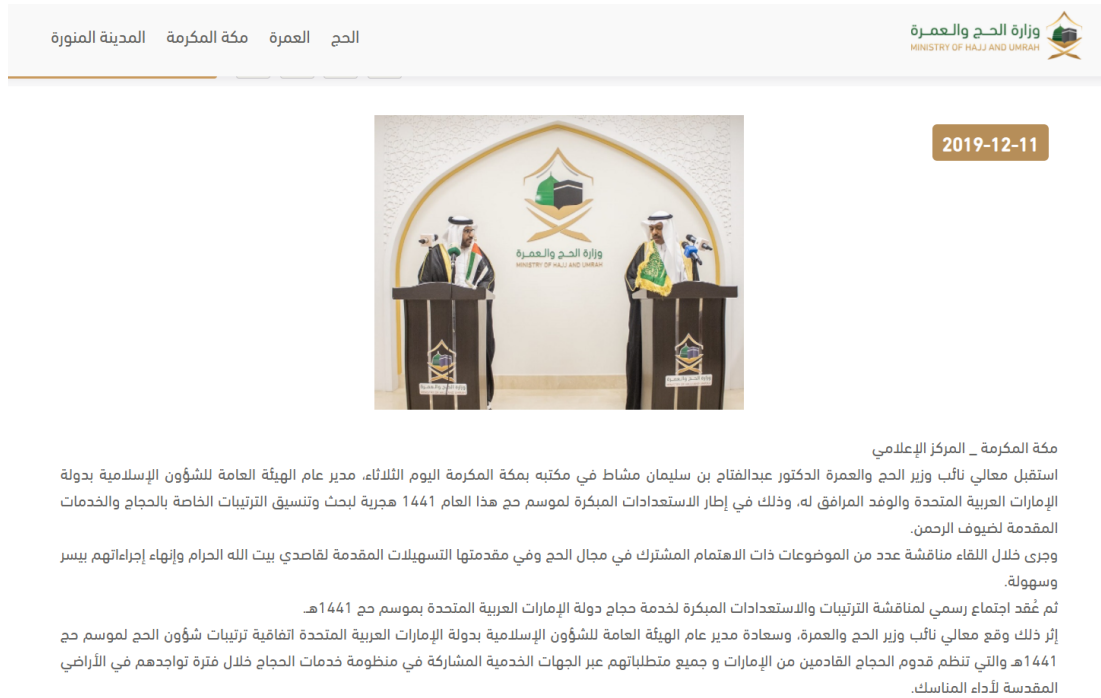


Figure 5.1: A News Article From the Ministry of Hajj and Umrah in Saudi Arabia Describing the Same Details as the Real Article.

extensive database. Usually, they post fake news on their website for public awareness. We used the only four online fact-checking platforms available at the time of research for the article verification process: NO\_RUMORS<sup>2</sup>, Falsoo<sup>3</sup>, AKEED<sup>4</sup>, and Fatabyyano<sup>5</sup>. An example of the platform NO\_RUMORS can be seen in Figure 5.2. These platforms are popular in the Arab region, and Fatabyyano has been certified by the International Fact-Checking Network, which certifies that a website complies with their code of ethics. We ran all the verified 900 real articles through each of the four fact-checking platforms. None of the 900 real articles matched with the fake articles posted in any of the fact checking platforms.

Admittedly, the reliability of labelling articles from news agencies as real is a matter of much debate (Vartapetian & Gillam, 2012; L. Zhou et al., 2004).

<sup>2</sup><http://norumors.net/>

<sup>3</sup>[Falsoo.com](http://Falsoo.com)

<sup>4</sup><https://akeed.jo/ar/>

<sup>5</sup><https://fatabyyano.net/>

Table 5.2: Real News Veracity Checking.

Veracity Property	Action Taken Over Articles	Source
Collecting government-monitored articles	Collection from news agencies that have a license from the Ministry of Media in that country	Amjad Point 1: Reliability
Cross-referencing	Fact-checking websites; check official websites for verification, such as the Ministry of Hajj and official Twitter accounts	Amjad Point 2: Cross-referencing
Metadata support	They were ensured during the process of cross-referencing articles; for example, checking the publication date	Amjad Point 3: Metadata support
Title and Content collocation	They were ensured during cross-referencing	Amjad Point 4: Title Content collocation
Source reliability	Collection from news agencies which have more than 10 years' worth of experience in the industry	Amjad Point 5: Reliability

Nevertheless, we made every effort to keep this study's data as objective as possible.

### Length of Real News Articles

Content length was another challenge in collecting real news articles. This problem does not arise in Tweets or online reviews because these are shorter texts than articles. However, of the 900 collected real articles, 63% were more than seven paragraphs in length and had a wordcount that exceeded 2,000 words. Texts of this length impede the falsification process and cause distraction with unnecessary items when extracting features (Mourão & Robertson, 2019). W. Y. Wang (2017) has developed an approach to address this issue by limiting the real articles to snippets or excerpts of news articles containing important statements,





Figure 5.2: The NO\_RUMORS Platform Displays the Red Graphic as Fake Content and the Green Graphic as Its Content Verification.

which one would want to fact-check. Accordingly, this study collected excerpts from the 900 real news articles, rather than the articles in their entirety. We selected each article’s first and second paragraphs, as news writers tend to give the essential information in these parts. News editors even have a name for this rule of thumb, ‘not burying the lede’—a catch phrase that appears in standard English dictionaries. However, some articles started with quotes or introductory topics before stating the important information. These types of articles were discarded, as it was going to take major effort to manually search for the important information throughout the article. This collection process resulted in 700 news excerpts; we refer to them as ‘articles’ throughout the rest of the research. Details of the number of articles (excerpts), their source newspaper, and their country of

Table 5.3: Real News Excerpts With Number of Articles (Excerpts) From Each News Agency.

Newspaper	Country of Publication	Number of Articles (Excerpts)
Okaz	Saudi Arabia	122
Masrawy	Egypt	110
AmmonNews	Jordan	118
AlRiyadh	Saudi Arabia	122
Youm7	Egypt	110
AdDoustur	Jordan	118
<b>Total</b>		<b>700</b>

origin are provided in Table 5.3. In general, the real articles contained not more than two paragraphs and did not exceed 700 words.

### Fake News Collection Using Crowdsourcing

Previous studies that investigated deceptive Arabic texts have targeted online reviews (Al-Barhamtoshy et al., 2019), YouTube comments (Alkhair et al., 2019), news headlines (Rangel et al., 2019), and Tweets (Mubarak & Hassan, 2020; Rangel et al., 2019; Alzanin & Azmi, 2019). As there are no existing datasets that fit this study, they had to be produced. Inspired by the work of Pérez-Rosas et al. (2017), crowdsourcing was used for the creation of fake news articles for several reasons. First, fake news can be fabricated by amateurs or journalists, because the task requires no specific skills (Torabi Asr & Taboada, 2019). Second, crowdsourcing helps reach a varied and diverse group that will result in rich outcomes with different perspectives. The approach in this thesis is consistent with several studies that have previously relied on crowdsourcing to build datasets of opinion reviews (Li et al., 2014), in addition to opinions on abortion and the death penalty (Mihalcea & Strapparava, 2009). Third, this method involves a lower cost and less effort than manual or technical methods for building datasets (El-Haj, Kruschwitz, & Fox, 2010).

## 5.1.2 Fake News Generation

In this section, we present the guidelines provided to the fake news participants for the generation of fake news articles. These guidelines were taken from Rubin et al. (2015) as the requirements for fake news dataset preparation. We added two instructions to these guidelines to ensure the fake news was written in a proper journalistic style. We also demonstrate the participant selection criteria and the fake articles' annotation scheme.

### Guidelines For Fake News Articles

This study adheres to the 'nine requirements of dataset preparation for fake news' proposed by Rubin et al. (2015) to ensure the reliability of the dataset. We aimed to generate fake articles based on rephrasing and manipulating the content of real news articles, which places the new fake articles into one of UNESCO's seven categories of 'information disorder': manipulated content, which is when authentic information is manipulated for a deceptive purpose (Ireton & Posetti, 2018). Rubin et al. (2015) guidelines were strictly followed, without making any biased decisions, to ensure the reliability of produced fake news.

Table 5.4 shows the guidelines we provided to the participants to ensure Rubin et al. (2015) fake news creation restrictions were followed. In addition to the above guidelines, we explicitly instructed the participants to follow two more guidelines in order to generate a meaningful manipulated version of the original article:

1. Preserve, as much as possible, proper nouns in the form of people or places presented in the original article. Try to avoid adding vague proper nouns.
2. Use different strategies to modify the articles. However, avoid simple negations.

Table 5.4: Application of Rubin et al.,’s (2015) Guidelines.

Rule#	Guideline	Action Taken
(1)	The corpus should include real and fake articles.	50% of the dataset was fake news; 50% of the dataset was real news
(2)	Only text news items should be used.	The collected real articles were only text without any metadata.
(3)	The articles should be obtained from real news from credible news sources.	The collected real articles were obtained from real news from credible news sources.
(4)	The generated fake articles should be the same length as the real articles.	we instructed the participants to keep their article the same word length as the original article, or as close to it as possible.
(5)	The fake articles should be homogeneous in writing style with the real ones; the corpus should be compatible with news genres and topics. For example, if the real article relays political news about a specific event, then the generated fake article about the same event should adopt the same style.	Various participants from different backgrounds were employed. They were instructed to write articles in MSA (the writing style used in newspapers), with good grammar and spelling, and use the journalistic genre writing structure similar to that of a published news article.
(6)	The corpus should be collected within a predefined timeframe.	The real news articles were collected within a three-year timeframe, from 2016 to 2019.
(7)	The fake news articles should be presented similarly to the original articles and with the same purpose, e.g., breaking news.	The participants were requested to manipulate the original articles realistically, which involved changing characters, information, events, and/or numbers, while retaining the purpose of the original article.
(8)	The news articles should be publicly available and easy to access.	Not applicable due to privacy issues and law enforcement in Saudi Arabia.
(9)	The fake articles should be compatible in language and with a similar cultural status to the real ones.	Participants were guided to write articles in MSA following journalistic style like that of the original article.

### Selection of Participants

The advertisements for recruiting fake news producers were carefully worded to ensure participants would be able to adhere to the above guidelines. For example, participants might desire to remain anonymous due to laws, such as the one in

Table 5.5: Article Distribution and Submission Statistics.

Locals	Number of Submitted Articles	Fiverr	Number of Submitted Articles
13	197	26	503

Saudi Arabia, that consider fake news production and dissemination to be crimes. Therefore, potential participants were reassured that no personal information would be asked for. Further, the participants were allowed to write fake news whenever possible, in their free time, and were rewarded a monetary amount for each submitted article. Finally, as some of the participants may have felt that fabricating articles related to a religious topic like the Hajj would violate their moral or religious sensibilities, the consent form included the following statement: ‘This is a research task for textual analysis purposes and does not interfere with anyone’s belief or religion.’

The local university email system at the University of Jeddah was used to invite native Arabic speakers, university students and employees of the University, to participate in this task. Each participant received 20 original and legitimate news articles with the guidelines they needed to follow, as explained in Table 5.5. Real articles were distributed on a first-come-first-serve basis. For example, the first participant to reply to the advertisement received articles 1–20. The second participant was given 21–40, and so forth. Unfortunately, about 43% of the local participants submitted only a subset of the 20 articles requested or cancelled their participation altogether. As a result, in a period of 30 days, participants submitted only about 197 falsified articles. Therefore, the Fiverr<sup>6</sup> platform was used to recruit more participants. Fiverr is a freelance platform that offers connection to freelancers or agencies in various domains such as business, education, and marketing. It allowed us to reach a large and diverse number of participants worldwide and provide these participants with an even greater level

<sup>6</sup>[https://www.fiverr.com/?source=top\\_nav](https://www.fiverr.com/?source=top_nav)

of anonymity than that provided by the university email system.

### Article Annotation Scheme

Two undergraduate female media students and one male Saudi journalist annotated the submitted falsified articles to reduce the likelihood of bias. These annotators were given the full list of submitted articles via email, with the guidelines that were to be followed. They worked individually, and none of the annotators knew each other. As such, their decisions were based on their own analysis. The annotators annotated the submitted falsified articles into three distinct groups: all, partial, and none. Articles that met all of the guidelines received the label ‘all,’ articles that met none of the guidelines received the label ‘none’, and articles that met some of the guidelines received the label ‘partial’. Figure 5.3 describes the annotation scheme. Articles were included in the dataset only when at least two of the three annotators gave it an ‘all’ label.

Articles removed from fake news sets were removed from the original real news dataset as well, to ensure equality in the number of articles in both datasets.

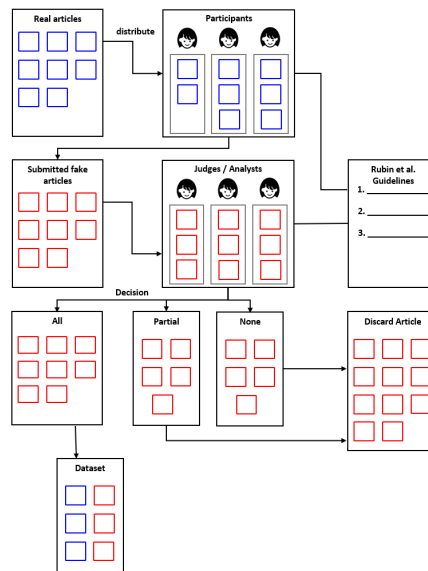


Figure 5.3: The Annotation Scheme for the real\_fake Dataset.

Table 5.6: Statistics of the real\_fake Dataset.

	Real	Fake
No. of articles	549	549
Avg. no. of sentences	3.15	3.38
Avg. no. of characters	606.60	633.13
Avg. sentence length (number of words in sentence)	249.78	303.37

Table 5.7: A Sample of Real Articles and Their Corresponding Fake Articles. Note the Red Text Represents the Manipulated Information, Whilst the Green Text Represents the Original Information in the Article.

Real Article	Fake Article
<p>قالت الصحف السعودية الصادرة الجمعة ان حجاج البيت الحرام هذا العام سيحملون سوار امن الكتروني وذلك بعد الفوضى التي نجمت عن التدافع الدامي في موسم الحج العام الماضي ووضحت صحيفتا اراب نيوز وسعودي غازيت ان اللجوء الى هذه التكنولوجيا من شأنه مساعدة السلطات على معالجة الحجاج والتعرف عليهم في ٢٤ ايلول (سبتمبر) ٢٠١٥ اثناء موسم الحج الاخير اودى حادث تدافع ضخم بحياة ٢٢٩٧ حاجا، بحسب معطيات جمعت من احصائيات حكومات اجنبية. ووجدت هذه الاخير صعوبات في التعرف على الضحايا و بحسب السلطات السعودية فان ٧٦٩ شخصا قتلوا في التدافع المساوي والاشد وقعا في تاريخ الحج.</p>	<p>قالت الصحف السعودية الصادرة السبت ان حجاج البيت الحرام هذا العام سيحملون أجهزة الكترونية صغيرة تلصق على الهواتف المحموله لتحديد أمانهم وذلك بعد الفوضى التي نجمت عن التدافع الدامي في موسم الحج العام الماضي. وأوضحت صحيفتا اراب نيوز وسعودي غازيت ان اللجوء الى هذه التكنولوجيا من شأنها مراقبة السلطات على الحجاج و تتبع أعمالهم وإستخداماتهم الشخصية في الهواتف المحموله و سيكون من السهل عليهم التنظيم والتنسيق بينهم. و في ٢٥ (أكتوبر) ٢٠١٢ اثناء موسم الحج اودى حادث تدافع ضخم بحياة ٣٦٥٤ حاجا، بحسب معطيات من مركز الاحصائيات المحليه. و فجمعت هذه الاخيره بإيجاد صعوبه في التعرف على الضحايا و لكن نأمل أن تنتهي هذه المشاكل بعد استخدام الاجهزة الالكترونية الذكية.</p>
<p>On <b>Friday</b>, the Saudi newspapers <b>published</b> that this year's pilgrims will <b>carry electronic security bracelets</b>, after the chaos that resulted from the bloody stampede during the Hajj season last year. The Arab News and the Saudi Gazette said <b>that resorting to this technology would help the authorities to treat pilgrims better by '[...] getting to know them'</b>. On <b>September 24, 2015</b>, during the last Hajj season, a massive stampede killed 2,297 pilgrims, <b>according to data collected from foreign governments 'statistics'</b>. There were difficulties in identifying the victims. According to the Saudi authorities, <b>769</b> people were killed in the tragic stampede, the most severe in the history of the Hajj.</p>	<p>On <b>Saturday</b>, the Saudi newspapers <b>published</b> that the pilgrims of the Holy House will <b>carry small electronic devices affixed to mobile phones in order to locate them</b> this year, after the chaos that resulted from the bloody stampede during the Hajj season last year. The Arab News and the Saudi Gazette explained that resorting to this technology will <b>allow the authorities to monitor the pilgrims and track their work and personal use on mobile phones, so it will be easy for them to organise and coordinate with each other</b>. Also, on <b>October 25, 2012</b>, during the Hajj season, a huge stampede <b>killed 3,654</b> pilgrims, according to data from <b>the Local Statistics Centre</b>. It was difficult to identify the victims, but the Saudi government confirmed that these problems would end upon using smart electronic devices.</p>

After the annotation process, 549 falsified articles were labelled as ‘fake,’ with the same number of original real news articles labelled as ‘real.’ Statistics of the real\_fake dataset are described in Table 5.6. An example of real and fake articles is demonstrated in Table 5.7. The inter-annotator agreement was measured as a Fleiss’ Kappa of 0.714, indicating a moderate to a substantial agreement beyond chance (Fleiss, Levin, & Paik, 2013).

## 5.2 Dataset 2: Satire Dataset

In order to build a dataset that trains the model to classify satirical and non-satirical articles, satirical articles were extracted from two websites, Alhudood<sup>7</sup> and Dkhlak<sup>8</sup>. These two satirical websites were the only ones openly available online during our research. The objective of creating this dataset was to answer research question 1: How well does the proposed approach to classification model also classify another type of fake news, such as satire? In total, 262 sarcastic Arabic news articles were collected that discussed several topics related to politics, economics, religion, and technology. In order to balance the dataset, non-satirical articles were manually collected from news agencies previously used to collect real articles, as reported in Section 5.1. We tried to collect, as much as possible, the same length of non-satire articles with the length of satire articles. The main focus lay on collecting articles that mentioned the same key dignitaries that the sarcastic articles presented. In some cases, the exact key details could not be found because they were made up. Therefore, articles with similar topics as those discussed by the sarcastic articles were used. For example, if a satirical article discussed ‘the high gas prices in the Gulf countries,’ this was balanced with a real article discussing the same topic. The dataset is called satire\_nonsatire. In total, the dataset includes 524 articles, 262 satire articles and 262 non-satire articles,

---

<sup>7</sup><https://alhudood.net>

<sup>8</sup>[www.dkhlak.net](http://www.dkhlak.net)



Table 5.8: Satire\_nonsatire Dataset Distribution.

	Satire	Non-Satire
No. of articles	262	262
Avg. no. of sentences	4.2	4.7
Avg. no. of characters	452	473
Avg. sentence length (number of words in sentence)	275.1	293.4

Table 5.9: Example of Satire and Non-Satire Article.

Satire	Non-Satire
<p>كشفت مصدرٌ رميَّ رفيعٌ مجازاً ومُخبئٌ على أرض الواقع عن خطة حكوميَّة لتمويل أبحاثٍ طبية تهدف إلى مكلفة علاج مرض ألزهايمر، سعياً منها لمساعدة أميرٍ قدر من المواطنين على نسيان همومهم وواقمهم بواسطة هذا المرض الحميد. وقال المصدر إنَّ نسيان المواطنين حقِّهم في امتلاك حدِّ أدنى من العيش الكريم سيمنح الحكومة فرصة خفض الإنفاق الحكومي على الخدمات العامة والبنى التحتية المتردية، وتركيز الإنفاق الحكومي على ملفَّاتٍ أكثر أهمية، مثل أعضاء الحكومة المبجلين.</p> <p>وأكد المصدر أنَّ وقف علاج المرض سيدعم قدرة المواطنين على عدم تذكر حقوقهم وظروفهم المعيشية البائسة، والتي تدفعهم عادة للقيام بتصرفات طائشة كالمظاهرات والاحتجاجات للمطالبة بها وهو ما سيؤثر سلباً على قمعهم والتكثيف بهم لمساعدتهم على النسيان.</p> <p>وعن الجوانب المدققة التي قد تعود من تفشي ألزهايمر، يوضح المصدر من الممكن أن ينسى المواطنون اسم القائد المفدى، لا سمح الله، أو إنجازاته و مآثره، وهو أثر جانبي بسيط لا يقارن بالفوائد المتحققة، ومن المبل علاج بيت خطابات القائد على التلفاز ورفع صورته في كل مكان وعقد جلسات توعوية للتذكير به.</p>	<p>لا يزال العلماء يلتزمون بالحذر بشأن دواء تم التوصل إليه مؤخراً لعلاج مرض ألزهايمر، وذلك نظراً إلى أن ذلك العقار المسماً أدوكانوماب ما زال في مرحلة التجريب الأولية. ويحفظ كثير من العلماء بشأن تلك الأدوية، التي يسمح باستخدامها ووصفها للمرضى الذين يعانون من فقدان الذاكرة بسبب ألزهايمر، منذ أكثر من عشر سنوات. لكن دراسة جديدة أوضحت أن العقار المذكور فعال، وبإمكانه أن يوقف تدريجياً مشكلة فقدان الذاكرة لدى مرضى ألزهايمر، وقد كانت مسألة تراكم لويحات الأميلويد أحد أهم المسائل التي تستدعيها الدراسات والأدوية التي توصل إليها العلماء لعلاج هذا المرض في السنوات الأخيرة. تعد لويحات الأميلويد تلك نوعاً من أنواع البروتينات التي تنمو بشكل غير طبيعي ثم تتجمع في الأنسجة أو في الأعضاء، وهي لا تفتت بشكل طبيعي مثل البروتينات العادية.</p> <p>ويقول العلماء أن الأجزاء التي تنتج عن تفتت تلك اللويحات قد تكون مسؤولة على موت الخلايا الدماغية، ومن هنا ينظر العلماء إلى أن تلك اللويحات هي المدخل الرئيسي للأدوية الخاصة بمرض ألزهايمر، والتي لا تزال في مرحلة التجريب.</p>
<p>A government plan to fund medical research – which is aimed at combating the treatment of Alzheimer’s disease, in an effort to help as many citizens as possible to forget their worries and reality through this benign disease – has been unveiled by a high-profile official source on the ground. The source said that forgetting citizens’ right to have a minimum decent living will give the government the opportunity to reduce government spending on public services and deteriorating infrastructure, and focus government spending on more important files, such as revered members of the government. The source stressed that stopping the treatment of the disease will support the ability of citizens not to remember their rights and miserable living conditions, which usually pushes them to engage in reckless actions, such as demonstrations and protests, ‘[...] which will remove from us the burden of oppression and abuse to help them forget.’ It is ‘[...] possible for citizens to forget the name of the leader, God forbid, or his achievements and exploits, which is a minor side effect that does not compare to the benefits achieved, and it is easy to cure this by broadcasting the leader’s speeches on television, raising his pictures everywhere, and holding awareness sessions to remind him,’ explains the source.</p>	<p>Scientists are still cautious about a recently developed drug used to treat Alzheimer’s disease, given that the drug, called ‘adocanumab,’ is still considered by many scientists to be in the initial trial phase. Many scientists have been reticent about these drugs, which have been allowed to be used and prescribed to patients with Alzheimer’s amnesia. However, a new study shows that the drug is effective and can gradually stop the problem of memory loss in Alzheimer’s patients. The issue of accumulation of amyloid plaques has been one of the principal issues targeted by studies and drugs that scientists have produced to treat the disease in recent years. These amyloid plaques are a type of protein that grows abnormally and then collects in tissues or organs. They do not break down as naturally as normal proteins. Scientists say that the fragments that result from the fragmentation of these plaques may be responsible for the death of brain cells, and that is why they consider these plaques to be the main entrance to drugs for Alzheimer’s disease, which is still in the experimental stage.</p>

as detailed in Table 5.8. A sample of a satire and non-satire articles is presented in Table 5.9.

### 5.3 Datasets 3 and 4: News Country-of-Origin Datasets

According to Aladhadh, Zhang, and Sanderson (2014), the location of a tweet’s author may influence the reader’s belief in its credibility. To examine the influence of this factor in relation to a real article’s credibility—specifically the country of origin of the published article’s news platform—I created two datasets. The aim was to assess the proposed approach’s feasibility in classification tasks and answer research question 2: How well does the proposed approach to classification model be used for another classification task, such as classifying an article’s country of origin?

Two topics were chosen for the creation of this dataset, the Hajj and Brexit. The topic of the Hajj offers diverse topics in several domains. However, since the Hajj is undertaken in Saudi Arabia, there was concern that the articles gathered from Saudi news agencies would be biased towards this topic. Therefore, the second dataset was created on the topic of Brexit to guarantee that articles were not biased towards one country over another. This has a relatively lower level of importance in Arabic speaking nations. To compile both datasets, news articles from January 2018 to October 2019 were collected by a Python scraper from Okaz, Addastour, and AmmonNews, news agencies. These news agencies represent three Arabic countries, Saudi Arabia, Egypt, and Jordan. For the Hajj dataset, 694 Saudi, 695 Egyptian, and 685 Jordanian news articles were included in a dataset called Hajj\_CO. For the Brexit dataset, 486 Saudi, 481 Egyptian, and 485 Jordanian articles were compiled in the dataset called Brexit\_CO. Table 5.10 shows the distribution of both datasets.

Table 5.10: Country-of-Origin Datasets Distribution.

Hajj_CO Dataset			
	Saudi Arabia	Egypt	Jordan
No. of articles	694	695	685
Avg. no. of sentences	7	7	7
Avg. no. of characters	2134	1809	1480
Avg. sentence length (number of words in sentence)	52	44	35
Brexit_CO Dataset			
No. of articles	486	481	485
Avg. no. of sentences	8	10	8.8
Avg. no. of characters	1163	1995	1404
Avg. sentence length (number of words in sentence)	25	30	32

## 5.4 Dataset 5: Covid Dataset

Additional data was gathered to address the usability of the model’s performance in reality, in dealing with fake news distributed in the real world. In early 2020, with rising concerns about the COVID-19 pandemic, many fake news articles were written and distributed through social media and messaging platforms such as WhatsApp and Facebook. Many of these articles were manipulated from official websites, such as the World Health Organisation, the Ministry of Health in Saudi Arabia, and the Food and Drug Administration in the USA, and distributed through social media and social messaging platforms as genuine.

Following Haouari et al. (2020) approach, a set of fake COVID-19 related articles were manually collected from the popular Arabic fact-checking platform, Fatbyyano. This platform is unique in that it provides fake articles, publication sources, and the verification of the fake articles in the form of real articles from legitimate news agencies. To create the COVID-19 dataset, we carefully chose articles that were written in a journalistic style. This fits our research aim, which is to identify fake articles written in a journalistic manner. We extracted only fake articles published on news or social media platforms that were written in a



Figure 5.4: The Red Side Shows the Fake Article, and the Green Side Shows the original Article.

journalistic style and claimed to come from a legitimate news agency, which in all cases was later confirmed to be fake by Fatbyyano. For example, a fake article was published purporting to come from a legitimate Saudi news agency called ‘Sabq’<sup>9</sup>, however, Fatbyyano determined that it was a non-legitimate website that imitated Sabq’s logo. To give a better perspective, Figure 5.4 shows an example of an article posted on social media referring to a fake website which uses a website address similar to Sabq’s. The red section shows the manipulated fake article, whilst the green section on the other side shows the original article posted on Sabq’s actual website.

Searching in Fatbyyano platform, 26 fake articles and their verified real articles were manually collected. In total, 52 articles, 26 fake and 26 real, about COVID-19 were collected into the COVID dataset. Table 5.11 shows the statistics of the dataset and Table 5.12 shows an example. This dataset forms the unseen data for further testing and evaluating this study’s model in Chapter 7, to answer research

<sup>9</sup>www.sabq.org

Table 5.11: Statistics of the Covid Dataset.

	Real	Fake
No. of articles	26	26
Avg. no. of sentences	2.01	2.89
Avg. no. of characters	143.48	167.56
Avg. sentence length (number of words in sentence)	128.24	134.57

Table 5.12: Sample Real and Fake Articles From the Covid Dataset.

Real Article	Fake Article
<p>تنص على أن أحد مكونات بروتين الحسكة الذي يعلم اللقاح الجسم صنعه يشابه بروتين سينستين-1 (بالإنجليزية: وهو البروتين الذي يساعد على تطور المشيمة التي تحيط بالجنين، مما دفع إلى الاعتقاد بأن الجسم لن يهاجم حركات فيروس كورونا فقط بل سيهاجم البروتينات المكونة للمشيمة ويقوم بإتلافها أيضاً. لا يوجد دليل إلى أن العقم ومشاكل الخصوبة لدى الرجال أو النساء من الآثار الجانبية لأي نوع من لقاحات كورونا التي يتم استخدامها حول العالم، كما أن تقنية الحمض النووي المرسل المستخدمة في إنتاج لقاحات كورونا هي ليست تقنية جديدة وقد تم استخدامها من قبل في إنتاج لقاحات الإنفلونزا</p>	<p>رئيس شركة فايزر السابق لأبحاث الجهاز التنفسي، مايكل ييدون، أن لقاح فايزر الخاص بكورونا يسبب العقم عند النساء. وبينت المنشورات أن اللقاح يحتوي على بروتين سبايك يسمى سينستين-1، الذي لعب دوراً حيوياً في تكوين المشيمة البشرية عند النساء، مما قد يؤدي إلى العقم عند النساء لمدة غير محددة، بحسب ما نقلته عن ييدون. ويحتوي على بروتين سبايك يسمى سينستين-1 وهو حيوي لتكوين المشيمة البشرية عند النساء... مما قد يؤدي إلى العقم عند النساء لمدة غير محددة.</p>
<p>one of the components of the al-Hasakah protein that the vaccine teaches the body to make is similar to the protein Synestin-1 (this is a protein that helps the development of the placenta that surrounds the foetus), which led to the belief that the body will not only attack the stings of the coronavirus but will attack the proteins that make up the placenta and destroy it. There is no evidence that infertility and other fertility problems in men or women are side effects of any type of corona vaccines that are used around the world. The messenger DNA technology used in the production of corona vaccines is not a new technology and has been used before in the production of Influenza Vaccines.</p>	<p>Pfizer's former head of respiratory research, Michael Yidon, said that the Pfizer vaccine for corona causes infertility in women. This may lead to infertility in women for an indefinite period, according to what was quoted from Yedon. The vaccine contains a spike protein called sensitin-1, which is vital for the formation of the human placenta in women, which may lead to infertility in women for an indefinite period.</p>

question 3: What is the performance of the proposed model on an unseen real and fake article distributed in the real world?

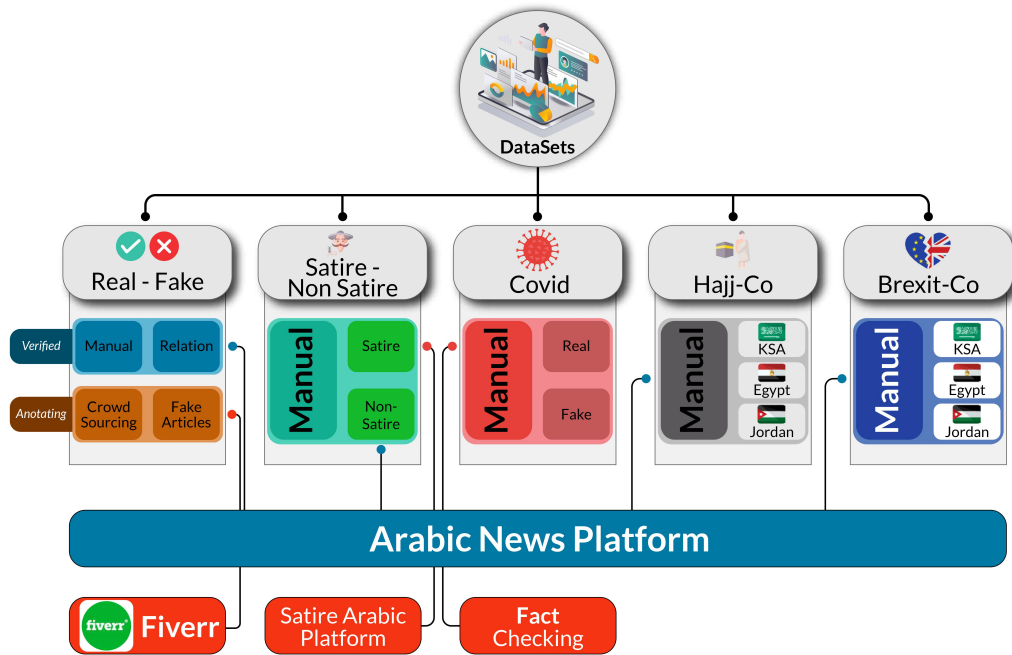


Figure 5.5: Dataset Creation Specifications.

## 5.5 Summary

This section presented the five datasets used in this research. The first dataset was constructed by collecting and verifying real articles and generating a fake version of them using a crowdsourcing, called real\_fake dataset. The second dataset, satire\_nonsatire, was constructed by collecting satirical articles from satirist platforms and balancing them with legitimate news articles on the same topic or about the same dignitaries. The third and fourth datasets, Hajj\_CO and Brexit\_CO, were organized to include more than 2,000 news articles from Saudi Arabia, Egypt, and Jordan about the Hajj and Brexit. The final was created by collecting fake articles distributed on social media about COVID-19 and verified articles on the same subject from legitimate news agencies. This dataset is called COVID. Figure 5.5 shows a summary of all the dataset's compilation process.

# Chapter 6

## Implementation

This chapter contains the core of the current research and discusses the creation of the model.

### 6.1 Data Preparation

One of the main contributions of this research is that it provides five datasets, as detailed in Table 6.1. Each dataset is comprised of articles in text format organized in separate folders. In particular, each text file contained the article's text and was saved as class name and number of article (i.e., real\_22). The real\_fake folder contains 549 real articles and 549 fake articles; the satire\_nonsatire folder contains 262 satire and 262 non-satire articles; the COVID dataset contains 26 real articles and 26 fake articles; Hajj\_CO contains 694 Saudi, 694 Egypt, 685 Jordan articles; and Brexit\_CO contains 486 Saudi, 481 Egypt, 485 Jordan articles. Figure 6.1 shows an example of a Saudi article in the Brexit dataset.

### 6.2 Data Pre-processing

Text pre-processing techniques are generally used to reduce a document's size and increase the processing speed. For text classification tasks, its main purpose

Table 6.1: Details of Constructed Datasets.

Dataset	Description
real_fake	549 real articles; 549 fake articles
Satire_nonsatire	262 satire articles; 262 non-satire articles
COVID	26 fake articles; 26 real articles
Hajj_CO	694 Saudi, 694 Egypt, 685 Jordan
Brexit_CO	486 Saudi, 481 Egypt, 485 Jordan

سجل الاقتصاد البريطاني الثلاثة أرقاما جيدة، مع انخفاض سنوي البطالة وارتفاع الفترة الشريفة، وهي أرقام تحرق عدم اليقين المحيط ببريكست ويتباطأ النمو. وتبدو مؤشرات سوق العمل في المملكة المتحدة بحالة جيدة، مع انخفاض معدل البطالة إلى 6.4% للمرة الأولى منذ 44 عاما في الأشهر الثلاثة حتى شهر نوفمبر، كما أعلن مكتب الإحصاءات الوطني. وهذا المعدل أقل من توقعات الاقتصاديين تحتنت إليهم وكالة بلومبرغ، والذين قالوا إن هذا المعدل سيظل عند مستوى 6.1% سجله في نهاية أكتوبر. وهذه المرة الأولى منذ عام 1975 التي ينخفض فيها معدل البطالة إلى هذا المستوى. ولم يصل تعداد العاطلين في المملكة المتحدة من قبل إلى رقم 32.53 مليون شخص، بارتفاع كبير خلال عام واحد. ووصلت نسبة التوظيف في البلاد إلى 675.8%، الأعلى منذ بدء تجميع هذه الإحصاءات في عام 1971. وسجلت بريطانيا 141 ألف وظيفة صناعية مقارنة بأكثر من 100 ألف وظيفة في بريطانيا. وهو رقم أعلى بكثير مما توقعته وكالة بلومبرغ (87 ألف). وأعتبر هاورد أكر الاقتصادي في مجموعة إي واي البريطانية للتوقعات الاقتصادية، أن سبب هذا الارتفاع في التوظيف هو أن الشركات ترغب في توظيف أكثر عند تمكن من الأشخاص ختية من نقص اليد العاملة الماهرة في بعض القطاعات، خصوصا بسبب تراجع الهجرة من الاتحاد الأوروبي. وكشف مكتب الإحصاءات الوطني في تقريره اليوم أرقاما مشجعة بشأن ارتفاع الرواتب، التي زادت قيمتها بنسبة 3.4% خلال عام واحد، وهي وتيرة لم تعرفها البلاد منذ عام 2008، وأكثر ارتفاعا من التضخم الذي يتباطأ في نشرين الثاني/نوفمبر ليصل إلى نسبة 2.3%. وترفع الرواتب فيما يتراجع ارتفاع الأسعار إلى قرابة 2%، وبذلك، تكسب الأمر قدرة شرائية أقوى، ونتيجة لذلك، ارتفع الدخل الحقيقي بنسبة 1.2 في المئة في شهر نوفمبر.

Figure 6.1: An Example of a News Article About Brexit.

is to reduce data dimensionality, thus reducing the size of text features. Dealing with Arabic necessitates the use of specific pre-processing techniques to reduce any errors. Here, we describe the pre-processing methods applied to the compiled datasets.

## Stemming

This is the process of reducing words to their root (stem). It relies on removing the end of the word, the suffix, to retain the base of the word. In terms of the impact of stemming on Arabic text, the morphological nature of Arabic has led to contradictory results in previous research (Wahbeh, Al-Kabi, Al-Radaideh, Al-Shawakfa, & Alsmadi, 2011). Studies have indicated that stemming might have little-to-no significance in the classifier's performance (Al-Badarneh, Al-Shawakfa, Bani-Ismail, Al-Rababah, & Shatnawi, 2017). Thus, the decision was made to not apply stemming and, instead, use the full form of the word. This was an effort to preserve the author's style as much as possible.



لم يعد ال+مستخدم ب+حاج+ة إلى تحميل ال+أفلام على ال+حاسب من أجل مشاهدة+ت+ها ب+فضل ال+مواقع و+ال+برامج  
التي تسمح ب+مشاهدة ال+أفلام على ال+إنترنت مباشر+ة . لكن ل+ال+وصول إلى ال+أفلام ال+مخصص+ة ل+ال+أطفال  
؛ يحتاج ال+مستخدم إلى إتمام هذه ال+عملي+ة ب+شكل يدوي دون وجود طريق+ة ل+عرض هذه ال+نوع فقط . لذا يمكن  
ل+ال+مستخدم+ين الذي يرغب+ون ب+عرض أفلام ال+أطفال فقط تجرب+ة  
من شركة ديزني ، و+الذي يوفر جميع ال+أفلام ل+مشاهدة+ت+ها على ال+إنترنت . بعد ال+دخول إلى ال+موقع يحتاج ال+مستخدم إلى تسجيل حساب مجاني ل+يتمكن من

Figure 6.2: Segmented Text.

## Segmentation

This process splits a sentence into a set of tokens (words). Its importance comes from converting unstructured text into independent words that can be easily analysed. We used the Farasa segmenter in Tasaheel to perform segmentation on all five datasets. An example of the segmentation is shown in Figure 6.2. The datasets were segmented to remove any affixes attached to the words for better word matching in the features extraction step.

## Normalisation

Several studies that worked with Arabic texts applied normalisation to unify the text for ease of analysis (Al-Badarneh et al., 2017; Alzanin & Azmi, 2019). Most normalisation tasks used to overcome misplaced dots or glitches (ء) in a word were performed by replacing letters as follows:

آ ، إ ، أ → ا

ي → ي

و → و

ه → ه

Because the articles used were written in a journalistic style but were from different news agencies that may have had different writing formats, the dataset

## Chapter 6. Implementation

was normalised to create a unified format. To normalise the datasets, the Tashaphyne<sup>1</sup> normaliser in Tasaheel was applied to all five datasets.

### Stop Words

Stop words were retained, because they constitute an important factor in this analysis. Many stop words act as function words, such as prepositions or conjunctions, which are useful in this work. For example, the stop word ‘except’ **إلا** is a function word used to incorporate an exception into a statement. Removing it would mean losing the statement’s exception clause.

### Filtering Non-Arabic Characters

Several character removals were performed to avoid useless data and remove non-Arabic letters. Further, special characters (**#, %, \$, @, &, \*, ‘,’ ‘;’**) were removed, but full stops were retained because they mark the end of a sentence. Additionally, diacritic marks were removed, though very few were present: the articles were in the journalistic genre, where these marks are in most cases not used. Numbers were retained in the articles, as some generated fake articles had manipulated numbers from the original articles. This is similar to the dataset in Nagoudi et al.’s study, where one of the approaches manipulated numbers in the original article to generate fake articles.

In total, five normalised and segmented datasets that contained articles in text format: `real_fake`, `satire_nonsatire`, `COVID`, `Hajj_Co`, and `Brexit_Co` were prepared for textual features extraction in the next section.

---

<sup>1</sup><https://pypi.org/project/Tashaphyne/>

## 6.3 Textual Feature Extraction

I relied on Tasaheel to extract the textual features from the benchmark datasets `real_fake`, `satire_nonsatire`, `COVID`, `Hajj_CO`, and `Brexit_CO`— as detailed below. Note that all of the datasets contain articles in text format. Assuming that specific words in a sentence indicates the writer’s feelings or thoughts, we extracted words that indicate emotion, polarity, and specific linguistic categories. We also extracted the POS of each word as they could indicate the presence of deceptive text (Newman et al., 2003).

### 6.3.1 Extraction of POS Features

To extract the POS features, Tasaheel was used to tag the datasets using the Farasa POS tagger, as seen in Figure 6.3. However, since Farasa did not generate tags for proper nouns and interjections, we relied on using Posit which generates proper nouns and interjections POS tags according to StanfordNLP tag format. Two folders of each dataset were generated, one tagged using the Farasa tagger, and the other tagged using the StanfordNLP tagger (Figure 6.4).

From each dataset, the nouns, verbs, adverbs, adjectives, prepositions, determiners, pronouns, and conjunctions from the Farasa-tagged folder were noted. Proper nouns and interjections were also noted from each dataset that relied on the generated StanfordNLP-tagged text files. The POS textual feature categories are displayed in Table 6.2.

### 6.3.2 Extraction of Linguistic Features

The linguistic tagger option in Tasaheel allows the tagging of words in a specified linguistic category that denotes a certain grammatical role. The output file produced all matched words, their file destination, and their total number of occurrences. We manually revised the results, and only those that fit the textual



Table 6.2: POS Features Categories.

Content Words		
Nouns	Verbs	Adjective
رجل / امرأة / حاج	صعد / نام / صلى	بارد / حار / عالي
Man / woman / pilgrims	Climbed/ sleep / prayed	Cold / hot / high
Adverbs	Proper Nouns	
أيضا	أحمد / اليمن / مصر	
Too / also	Ahmad / Yemen / Egypt	
Function Words		
Conjunctions	Prepositions	Pronouns
و / ف / ثم	مع / في	ي / أنت / هي / نحن
And / thus	In / on	We / she / you / you
Interjections	Particles	Determiners
أوه	كي / حيث	ال
Ooh	For that, since	The

superlative words in Arabic in one wordlist category. The other issue was the affix ل 'li' representing a justification. Because the affix is attached to a word, exact word matching was not possible. To extract superlatives and the justification ل 'li', we needed to modify the code in Tasaheel to enable searching through the set of POS tags produced by Farasa and StanfordNLP for specific attributes related to these categories. We coded a Python script to search for a designated item by forming a query that fit the item's tag. The code was added as an option in Tasaheel, affix extraction. The POS tagged text files from Tasaheel and Posit in the Farasa and StanfordNLP tag format were input in the Tasaheel affix extraction option.

### Extracting Superlative Category

As stated in chapter 3, superlatives are adjectives that express the highest degree of comparison. Since it is difficult to predefine all superlatives in Arabic and



constant words. In particular, the affix لـ ‘li’ means ‘for that cause’ when attached to a present verb, which produces a justification. To search for this affix, we searched the Farasa tagged files. In Farasa tagged text files, the words and their connected affixes are assigned to detailed POS tags showing the affix type. لـ ‘li’ is tagged as a preposition in this case. Given this, the following query was formed:

```
BASE_ TAG: JJ
Affix: لـ / PREP
```

Figure 6.6 shows the output file with the result of this query for satire\_nonsatire dataset. Where the file indicates that the matching query of this affix is present one time in file “nosat\_3” and one time in file “nosat\_9”. Figure 6.7 demonstrates the place of query match highlighted in the “nosat\_3” file.

The results of both queries were manually reviewed to ensure that they fit the correct linguistic category. Note, using both POS taggers, Farasa and StanfordNLP, to extract justification and superlative categories meant that they supported each other and aided in extracting accurate features. The code for searching within the generated POS tags field for specific queries was added as a new function in Tasaheel as ‘Affix extractor,’ as seen in Appendix A.

To summarise, the linguistic feature categories extracted were as follows: as-

```
nosat_9.txt = 1
],
"لـ+/PREP": [
  "nosat_3.txt = 1",
  "nosat_9.txt = 1"
]
```

Figure 6.6: Query output sample.

ADJ/منفصل +مين/ NOUN+NSUFF-MP/صدر +مين/ NOUN-MS/حصري NOUN-MS/  
 NOUN-MS/ولي/ V/استجاب/ PREP/+ل/ DET+NOUN-MS/سيسي/ ال + PART/أن/ DET+A  
 DET+NOUN-MS/سيسي/ ال + NOUN-MS/محمود/ NOUN-MS/استبعاد/ PREP "/PUN  
 OUN-MS/وقت/ PREP/في/ NOUN-MS/روسيا/ PREP/إلى/ PRON/ه/ NOUN-MS/إيفاد/ C  
 دو/ CONJ/ و/ ADJ+NSUFF-FS/ة+ محلي/ NOUN-MS/إعلام/ NOUN-FP/وسائل/ PREP/في/  
 NOUN+NSUFF-

Figure 6.7: Query Place Result .

Table 6.3: Linguistic Features Categories and Number of Words Matched.

	Assurance	Negations	Justification*
No. of words (All concrete words except*)	7	7	10 (9 concrete + 1 affix)
	Intensifiers	Hedges	Illustration
	14	7	6
	Temporal	Spatial	Superlative
	8	10	
	Exception	Opposition	
	6	4	

assurance, negation, illustration, intensifier, hedges, justification, temporal, spatial, exclusion, superlative, and opposition. Table 6.3 demonstrates each linguistic textual features and the number of words in each feature.

### 6.3.3 Extraction of Emotion and Polarity

To extract the emotion and polarity words from each article in all datasets, the emotion and polarity tagger options in Tasaheel were invoked. For each textual category, words in that category were matched with the text from the articles. The emotional categories extracted were anger, sadness, fear, joy, disgust, and surprise words. Table 6.4 shows the number of words in each category to be matched with the articles. The polarity words extracted were from positive and negative polarity sets. An output file produced words matched, destination files, and their number of occurrences in each file. In each file, words tagged will be displayed with its textual category following, and a summary of the number of the textual categories is provided, as shown in Figure 6.8. Note that the ‘+’



## Chapter 6. Implementation

Table 6.4: Emotion and Polarity Categories and Number of Words Matched.

	Anger	Sadness	Fear	Disgust	Joy	Surprise
	203	214	404	279	337	292
	Negative			Positive		
No. of words	4783			2006		

تتبع فرق التفنيد ال بيبي ل ال هيبة ال عام ال ال ارضاد و حماي ال بيبي ال في منطق مكة ال مكرم تكثيف اصمال ال ميداني ال ال مسح و ال تفنيد على ال منشآت ذات ال تاثير ال بيبي ال محتمل على ال بيبي ال في كل من ال عاصم ال مقدس و محافظة جدة و ذكر مدير ال هيبة ب منطق مكة ال مكرم و وليد ال حجيلي ان ال فرق ضمن خططها ال معتمدة ل حج هذا ال عام تكثف جول ال ها ال تفنيد في كل [intensifiers] من مكة ال مكرم و محافظة جدة و ال طرق ال واصل ال ال مشاعر المقدس ال ال تاكيد من التزام ال منشآت ب ال اشتراطات ال بيبي ال منصوص على ها في ال نظام ال عام ل ال بيبي ال و لوائح ال تنفيذي ال في إطار حرص ال هيبة على صحة و سلامة حجاج بيت الله ال حرام و ال عاملين في ال مابين ان على ال تفنيد تشمل ال منشآت ال صداحي ال منشآت ال اعايش ال و ال تموين و ال منشآت ال صحي ال تابع ال بعث ال ال حج و محطات ال وفود و ال استراحت و جميع [intensifiers] ال منشآت ذات ال [intensifiers] ال تاثير ال بيبي ال محتمل و أكد ان ال فرق ما كل [intensifiers] مزود ب كاف ال تجهيزات ال بتشري ال ال التي تمكن ها من أداء مهام ها على افضل وجه ، و س تعمل على رصد ال مخالفات و اتخاذ ال اجراءات ال نظامي ال حيال ها داخي اصحاب ال منشآت و ال مواطنين و ال مقيمين و حجاج بيت الله ال حرام إلى ال محافظة على بيبي ال حج و ال ابلح عن اي مخالف [negative] ال على ال رقم 988 على مدار ال ساعة .

Count: {'intensifiers': 3, 'negators': 1, 'negative': 1}

Figure 6.8: Emotion, Polarity, and Linguistic Tagging in Tasaheel.

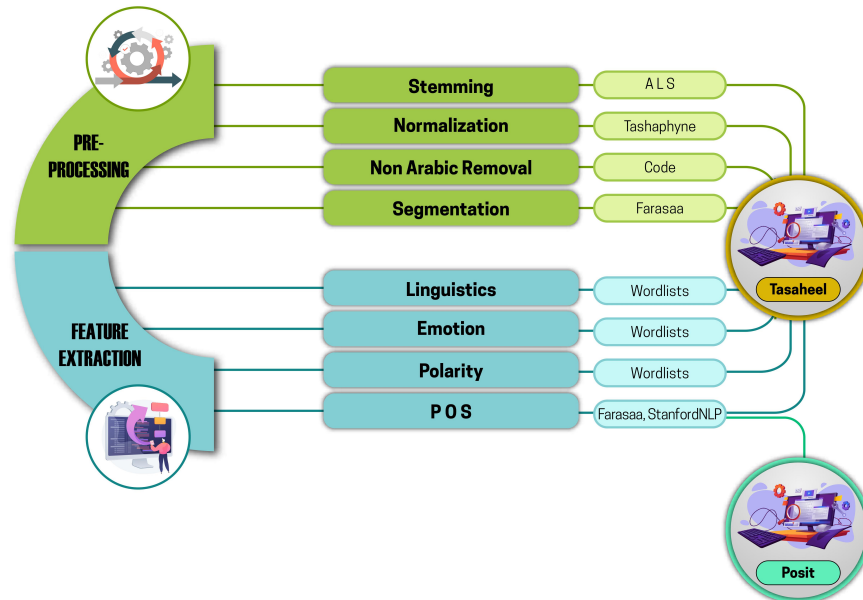


Figure 6.9: Summary of the Pre-Processing and Feature Extraction Tools.

signs included in the text files after segmentation were removed to avoid any mismatching.

Figure 6.9 summarises the pre-processing and feature extraction phases with details of the tasks and tools used.

## 6.4 Datasets Preparation

### 6.4.1 Primary Datasets Preparation

Each dataset includes the quantity of each extracted textual feature from every article in Excel sheets. Note that each article was assigned a specific label, which conforms with its saved name, as previously discussed. Articles were labelled 0 for real and 1 for fake in `real_fake`; 0 for non-satire and 1 for satire in `satire_nonsatire`; 0 for real and 1 for fake in `COVID`; and 1 for Saudi Arabia, 2 for Egypt, and 3 for Jordan in both `Hajj_CO` and `Brexit_CO`. In this way, datasets consisted of the articles' number, class, and number of occurrences of each textual feature. Table 6.5, shows all investigated textual feature categories. In total, the `real_fake` dataset had 544 real and 544 fake articles; the `satire_nonsatire` dataset had 262 satire and 262 non-satire articles; the `COVID` dataset had 26 real and 26 fake articles; the `Hajj_CO` dataset had 694 Saudi, 694 Egyptian, and 685 Jordanian articles; and `Brexit_CO` had 486 Saudi, 481 Egyptian, and 485 Jordanian articles. A total of 30 features were extracted from each article in the five built datasets. The POS feature set had 10 features, the emotional had six, the polarity had two, and the linguistic had 11.

### 6.4.2 Secondary Datasets Preparation

**A.** A small sample of 10 random articles (five real, five fake) from the `real_fake` dataset was grouped in an Excel sheet named `Test.xlsx`, which is used to compare the model's performance with humans in the next chapter. This aims to answer research question 4: How might the proposed model perform compared to humans in classifying real and fake articles?

**B.** We stemmed a copy of the `real_fake` dataset using the ASL stemmer in `Tasaheel` and extracted emotion features. This process produced the `real_fake_stem` file that contained the emotion features of 549 real and 549 fake articles. This was

Table 6.5: All Textual Features Extracted.

<b>POS</b>		
<b>Content Words</b>		
Nouns	Verbs	Adjectives
Adverbs	Proper Nouns	
<b>Function Words</b>		
Conjunctions	Prepositions	Pronouns
Interjections	Particles	Determiners
<b>Emotion</b>		
Anger	Sad	Fear
Joy	Disgust	Surprise
<b>Polarity</b>		
Positive	Negative	
<b>Linguistic</b>		
Assurance	Negations	Illustration
Intensifier	Hedges	Temporal
Spatial	Exclusion	Superlative
Opposition	Justification	

to answer research question 5: What is the effect of stemming articles on the model's performance?

## 6.5 Experimental Setup

In this section, we describe the setup in terms of the software used and details of the parameters set for the ML classifiers. The details of the experiment's environmental setup are outlined below.

### 6.5.1 Microsoft Excel

Microsoft Excel has the necessary computational features, using a grid of cells arranged in numbered rows and letter named columns to organise data. It also offers a wide range of data storage options and fundamental mathematical functions. We used Microsoft Excel 2019 to create five separate spreadsheets for each dataset's feature extraction results. Each spreadsheet was named after its dataset

## Chapter 6. Implementation

NOC	ADJECTI	ADVERB	INTERJE	PARTICI	DETERI	VERBS	PREPOS	NOUNS	joy	disgust	anger	positive	negativ	excepti	time/dé	justifica	order	class	article	
221	0	0	0	1	0	5	4	13	0	0	0	1	0	1	0	0	0	0	0	يُعرض بقصد الأيو حاد الفصيل مستشار خادم الحرمين الشريفين أبو منطقة مكة المكرمة، على خطة قوات من (1)
323	0	0	0	0	1	4	7	30	0	0	1	0	0	1	0	0	2	0	0	يعم الأيو حاد الفصيل لدى استقباله اليوم الإثنين، قائد قوات أمن المنشآت اللواء سعد بن حسن الجدي وبدل
380	0	0	0	0	0	2	4	14	0	0	0	0	0	0	0	0	1	0	0	دم قائد قوات أمن المنشآت أسفة و أنه سيستقر عقد بعض من ورضي العمل والمخاضات التحسين مستوى الجواز
390	3	0	0	0	0	6	7	24	0	0	1	0	2	0	0	0	0	0	0	في نهاية اللقاء وقع الأيو حاد الفصيل على تقرير يري عن بعض التغيير من قوات أمن المنشآت في مهمة الحج (1)
243	0	0	0	0	0	5	3	18	0	0	2	0	0	0	0	0	0	0	0	منحت بصرته وزارة الخارجية الإيرانية، الأحد، لها استمعت القام بالأعمال السعودي لمطالبة ترضي الأبرار (1)
304	0	0	0	0	0	3	4	19	0	0	0	1	0	0	0	0	0	0	0	قلت وكالة الأنباء الإيرانية الرسمية (إرنا)، عن مساعد وزير الخارجية الإيرانية للشؤون العربية والافراسية، حسين نور
243	1	0	0	0	0	5	2	15	0	0	0	0	0	0	0	0	0	0	0	شارع أمير عبدالقهار في أن مسؤول الحج الإيرانيين نطق من ممثل في العهد السعودي والتفكير، أمس، وأكد (1)
138	0	0	0	0	0	5	1	10	0	0	0	0	0	0	0	0	1	0	0	شيع بلجان (إرنا) في أن آخر حصيلة لتحتاج الإيرانيين المغلوبين حتى، السبت، بلغت 25 نكصاء، بينما وصل عدد
156	0	0	0	1	0	3	5	10	0	0	1	0	0	0	0	0	0	0	0	إن العرش الأعلى للثورة الإسلامية بيران، أية الله علي خامنئي، الأربعاء، قد نطق من السلطات السعودية و يوم (1)
327	1	0	0	0	0	2	3	3	23	0	0	0	0	0	0	0	0	0	0	بمساعد وزير الخارجية الأيو حاد الفصيل «صود» إدارة العقوفين السعوديين» خلال العتاد، منبر إلى أن مجلس الوزراء (1)

Figure 6.10: Dataset's Worksheets. From Right to Left, The First Column Shows the Articles, Then the Article Class in Gray. The Green Columns Contains the Linguistic Categories, Pink Contains the Polarity, Yellow Contains the Emotions, and Blue Contains the POS. The Last Column Contains the Number of Words in Each Article.

name and contained the article's text in rows, with each row corresponding to the article number in its original folder. Each column was named after a textual feature category. All tabulated data was then entered into a Microsoft Excel spreadsheet for lexical density calculations in the next chapter. Specifically, each result was noted under its category and corresponded to the article number. For that, each article had a row filled with the results of each textual category. Figure 6.10 shows a sample of one dataset worksheet. The columns include the article's text, class, and its POS, polarity, emotion, and linguistic categories. The last column shows the number of words in the article, which is needed to calculate the lexical density in the next chapter.

As a result, we had a total of five datasets recorded in separate Excel sheets: `real_fake.xlsx`, `satire_nonsatire.xlsx`, `Covid.xlsx`, `Hajj_CO.xlsx`, and `Brexit_CO.xlsx`. To test and evaluate the model in WEKA the results in excel for each dataset must be converted to attribute – relation file format, `arrf`, which will be explained next.

### 6.5.2 WEKA

WEKA is a collection of machine learning algorithms used for data mining tasks. It implements several algorithms to perform classification, clustering, and data

```

@attribute joy numeric
@attribute fear numeric
@attribute sad numeric
@attribute surprise numeric
@attribute proper_NOUNS numeric
@attribute NOUNS numeric
@attribute PREPOSITIONS numeric
@attribute VERBS numeric
@attribute DETERMINERS numeric
@attribute PRONOUNS numeric
@attribute INTERJECTIONS numeric
@attribute ADVERBS numeric
@attribute ADJECTIVES numeric

@data
real0.0,0.0,1.0,0.0,1.0,0.0,2.0,0.0,1.0,1.3,9.4,4.4,3.3,3.2,2.1,1.1,1.1,1.1,1.1,3.4,5.0,0.0,1.0,0.0,41.2,20.5,1.41,221.5,1
real0.2,0.0,1.0,0.0,1.0,0.0,1.1,0.0,0.0,0.6,21.7,10.2,3.6,3.1,1.1,0.3,2.0,30.7,4.1,0.0,0.0,0.0,61.1,61.1,61.3,23.5,3
real0.1,0.0,0.0,1.0,1.0,0.0,0.0,0.0,1.0,2.10,4.4,2.2,2.3,0.3,0.0,2.0,0.14,4.2,0.0,0.0,0.0,0.34,1.34,1.34,180.5,3
real0.0,0.0,0.0,1.0,1.0,1.1,1.0,0.0,0.0,6.14,7.10,2.5,6.6,1.1,0.0,4.3,0.24,7.6,0.0,0.0,0.0,3.63,1.63,1.63,350.5,6
real0.0,0.1,0.0,0.0,0.0,0.0,2.0,0.2,1.1,1.17,3.7,1.4,1.2,1.1,1.1,0.1,1.18,3.5,0.0,0.0,0.0,41.2,20.5,2.20.5,243.5,9
real0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.8,11.4,8.7,3.8,5.0,3.0,0.0,0.0,19.4,3.0,0.0,0.0,0.49,1.49,1.49,304.6,2
real0.0,0.0,0.0,0.0,0.0,0.0,1.0,0.0,0.0,6.9,2.5,5.6,3.0,2.0,0.0,0.0,15.2,5.0,0.0,0.0,0.139,1.39,1.39,243.6,2
real0.1,0.0,0.0,0.0,0.0,0.0,0.0,0.0,1.0,1.10,1.14,3.4,0.0,1.1,0.0,0.0,1.10,1.5,0.0,0.0,0.0,26.3,8.7,1.26,158.6,1
real0.0,0.0,0.0,0.0,0.0,0.0,0.0,1.0,0.0,0.0,5.5,5.3,3.1,5.2,1.1,0.0,0.0,0.10,5.3,0.0,0.1,0.0,0.28,1.28,1.28,156.5,6
real0.0,0.1,0.0,0.0,0.0,0.0,0.0,0.0,0.0,3.20,3.8,6.3,3.1,0.2,0.0,0.2,0.23,3.3,2.0,0.0,0.0,1.53,1.53,2.26,5.3,27.6,2
real0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,6.11,1.7,4.4,6.4,0.0,0.0,0.1,17.1,4.0,0.0,0.0,0.1,40.3,13.3,1.40,224.5,6
real0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,1.0,0.0,3.19,4.8,4.7,3.6,0.0,0.0,0.0,1.77,4.7,0.0,0.0,0.0,56.1,5.1,1.54,331.5,6

```

Figure 6.11: Dataset in the Arrf File Format.

association. After noting the results produced by extracting features from the articles in each dataset, we made another copy of these datasets that were converted to arrf files to be compatible for use in WEKA. Each feature was given in the top of the arrf file, and its value noted in the line corresponding to the article’s label, as shown in Figure 6.11.

As a result, we had a total of five arrf. files: real\_fake. arrf, satire\_nonsatire.arrf, Covid.arrf, Hajj\_Co. arrf, and Brexit\_CO.arrf. Moreover, we had two secondary dataset tests. arrf and real\_fake\_stem.arrf.

WEKA default configurations were set for all NB, SVM, LR, and RF algorithms used in the current research. Details of the parameters can be seen in Table 6.6. The textual feature categories were uploaded in WEKA as the variables for training and testing.

The experimental environment was split by a percentage: 80% training and

Table 6.6: Classifiers' Parameters.

SVM	NB	LR	RF
batchSize 100 kernel linear	batchSize 100	batchSize 100 maxBoostingIterations 500	batchSize 100 bagging with numIterations100 number of trees 100

20% testing for the model's evaluation experiments. Moreover, the supplied test option was used for testing the COVID dataset as unseen testing data. Further, we used the ChiSquaredAttributeEval to evaluate the important features that influence the model's performance.

### 6.5.3 Orange

Orange version 3.31.0 was used in this research. It is an open-source ML and data visualisation software that enables data analysis and maps misclassified data items. Orange provides a word cloud option that displays tokens in the dataset, where it is an excellent widget for displaying the frequency of words. Through BoW features enabled in the tool, words are listed by their frequency. The tool's configuration parameters for the word cloud option performed in this research is based on turning the fake articles into word vectors using BoW n-grams, which is an extension of the BoW approach. An n-gram is a sequence of n tokens (words). In this research, we applied the uni-gram feature to create a word cloud visualisation that generated the frequent words in fake articles. The weight of words are calculated, and the ones with higher weight represent the most frequent ones. These most frequent ones are then displayed, as shown in Figure 6.12 (Orange, 2022).

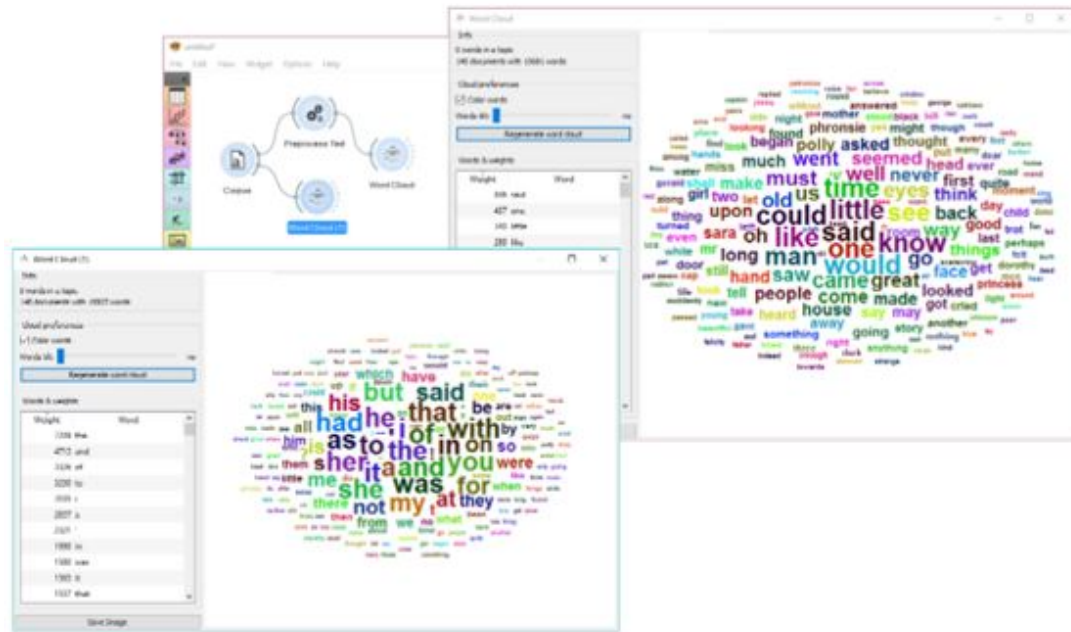


Figure 6.12: Sample of Word Cloud in Orange.

#### 6.5.4 Participants for Human Evaluation

Ten female computer science students participated in the sample testing in a classroom in the University of Jeddah. Handout sheets that comprised 10 articles (5 real and 5 fake), which were randomly extracted from the dataset testing sample for model evaluation, were manually given to the students. We stayed in the classroom and students were not allowed to speak to each other or use their mobile phones during the task. We verbally explained the instructions, as follows: the participants were to note only (R) if they believed the article was real and (F) if they believed the article was fake. No personal data was needed on the form, so it was not collected. The task had no time limit, so it ended when the last participant handed in her sheet.

### 6.5.5 Evaluation Metrics

The performance evaluation of each classifier is based on average accuracy (A), precision (P), average recall (R), and average F-Measure (F).

**Precision:** It is calculated as a measure of correctly identified positive cases from all of the predicted positive cases. It is especially useful to inspect the cost of False Positives (FP).

$$Precision(P) = \frac{TPR}{TPR + TNR} \quad (6.1)$$

**Recall:** It is calculated as a measure of correctly identified positive cases from the actual positive cases. It is especially useful to inspect the cost of False Negatives (FN).

$$Recall(R) = \frac{TPR}{TPR + FNR} \quad (6.2)$$

**F- Measure:** It is the harmonic mean of precision and recall.

$$F - Measure(F) = \frac{2.P.R}{P + R} \quad (6.3)$$

**Accuracy:** It calculates the ratio of sum of true positives and true negatives of all the predictions.

$$Accuracy(A) = \frac{TPR + TNR}{TPR + FNR + TNR + FNR} \quad (6.4)$$

$$\text{Lexical Density for each category feature in a class (L) = } \frac{\text{(total number of occurrence of each feature category in a class )} \times 100}{\text{total number of words in the whole class}}$$



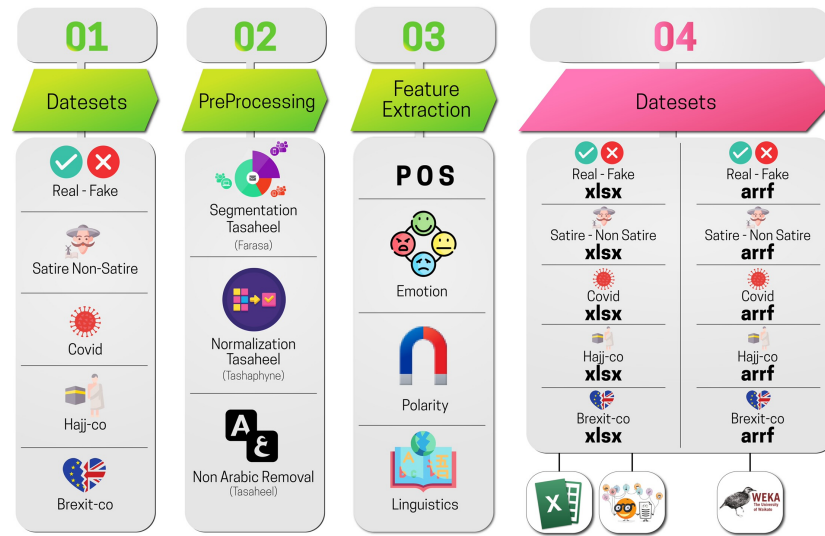


Figure 6.13: Dataset Compilation Details.

The formula of the Chi square feature selection is shown in algorithm:

$$X_c^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (6.5)$$

where  $c$  is the degree of freedom (threshold value),  $O$  is the observed value,  $E$  is the expected value, and  $X_2$  is chi-square computed result for feature. This formula is used to calculate the important features for model's performance.

## 6.6 Summary

In this chapter, we presented the implementation to build the classification model. As detailed in Figure 6.13, a total of five datasets saved in Excel sheets were used for lexical density calculations. We made a copy of these datasets and turned them to arrf files to test and evaluate the proposed model in WEKA.

# Chapter 7

## Experiments and Analysis

This chapter presents the results of the research. It provides details on the experimental procedures carried out and presents the findings and analysis of the supervised machine learning model that classifies real and fake Arabic articles based on textual analysis. This research has two parts. The first part evaluates the model's performance in classifying real and fake articles, whilst the second conducts several experiments to answer the research questions.

### 7.1 Evaluation. Part 1: real\_fake Classification

To evaluate the model's performance, experiments with regards to the textual feature sets were carried out. The WEKA tool was prepared for testing and standard parameters, and configurations for ML classifiers were chosen in WEKA, as discussed in Chapter 6. The `real_fake.arff` is composed of 544 real articles labelled 0 and 544 fake articles labelled 1, with 30 textual features of POS, emotion, polarity and linguistics. Its was randomly split into 80% (874 articles) for training and 20% (215 articles) for testing. All the performance metrics are in percentages.

I present three evaluation scenarios:

Table 7.1: Results of the Evaluation of the Individual Feature Categories.

		POS	Emotions	Polarity(P)	Linguistics(L)
Naive Bayes (NB)	Precision	54.1	56.8	57.4	61.3
	Recall	53.6	54.5	56.0	60.2
	F- Measure	49.4	51.9	54.9	58.7
	Accuracy	53.7	52.1	55.1	60.1
Random Forest (RF)	Precision	63.6	58.4	56.9	64.5
	Recall	63.6	58.4	56.9	64.5
	F- Measure	63.6	58.3	56.8	64.5
	Accuracy	63.6	58.4	56.9	64.5
SVM	Precision	57.3	53.2	49.9	60.0
	Recall	57.7	54.2	48.4	59.4
	F- Measure	56.7	50.2	50.0	58.8
	Accuracy	56.9	51.2	49.0	59.2
Logistic Regression (LR)	Precision	61.9	55.7	57.4	63.1
	Recall	61.9	55.4	57.2	62.9
	F- Measure	61.9	54.9	57.6	62.8
	Accuracy	61.9	54.8	57.8	60.9

- The first experiment evaluated the model’s performance according to each set of textual features individually.
- The second experiment evaluated the model’s performance using all sets of textual features.
- The third experiment evaluated the model’s performance using various combinations of textual feature sets.

### 7.1.1 Experiment 1: Testing Textual Feature Sets Individually

In this experiment, each textual feature set was tested individually, and the model’s performance was evaluated and noted in Table 7.1.

As shown in the table, the four classifiers produced accuracy results ranging between 58.4% and 64.5%. The RF classifier achieved the highest accuracy of

64.5% and 63.6% for the linguistic and POS features, respectively, which might be due to the fact that similar words are found in both POS and linguistic features. For example, the word **أضخم** ‘the biggest’ is an adjective found in the POS features and a superlative in the linguistic features. The NB classifier had the lowest score of 53.7% accuracy for the POS features. The LR classifier’s accuracy decreased by around 0.2 in the accuracy compared to the RF accuracy, for all textual features, except the polarity feature, where it decreased by 0.12. An interesting observation is that emotion and polarity features have relatively similar results in all classifiers, with an accuracy of slightly over 50%. This can be explained by the fact that many words in the emotion features are also found in the polarity features, for example the word **فرح** ‘happy’ is a word that implies the emotion joy and positive polarity.

### **7.1.2 Experiment 2: Testing and Evaluating All of the Features**

In this experiment, all textual feature sets were tested, and the model’s performance was evaluated and noted in Table 7.2.

The results in this table support the findings of the previous experiment with respect to the RF classifier. Random forest performs better than other classifiers. It has an accuracy of 77.3% for precision, 77.2% for recall, 77.3% for F-measure, and 77.2% for accuracy, with all features combined. Although it had a lower accuracy, LR computed an average recall of 69.7%, precision of 69.7%, F-measure of 69.7%, and accuracy of 69.9%. These scores of precision, recall, F-measure, and accuracy have decreased in SVM compared to LR. The NB had the lowest accuracy with 60.6%.

Table 7.2: Results of Evaluation of All Features in the real fake Dataset.

	All Features	
Naive Bayes (NB)	Precision	60.4
	Recall	59.0
	F- Measure	56.8
	Accuracy	60.6
Random Forest (RF)	Precision	77.3
	Recall	77.2
	F- Measure	77.2
	Accuracy	77.2
SVM	Precision	63.7
	Recall	63.4
	F- Measure	63.3
	Accuracy	63.5
Logistic Regression (LR)	Precision	69.7
	Recall	69.7
	F- Measure	69.7
	Accuracy	69.9

### 7.1.3 Experiment 3: Testing and Evaluating Combined Features

At this point, the aim was to determine what features combined might improve the classification results. Combined textual feature sets were tested, and an evaluation of the model's performance is shown in Table 7.3.

The RF outperformed the other three classifiers by producing accuracy scores of above 60% in all textual feature combinations. The highest accuracy scores were achieved by a combination of POS and linguistic features. Logistic regression and RF generated scores of 69.5% and 72.6%, respectively, with the same feature set combination. The lowest accuracy score was achieved by NB where a combination of POS and emotion features achieved 49.1%. Support Vector Machines (SVM) had an average accuracy of 56.3% for POS and emotion, 61.6% for POS and linguistics, 54.5% for POS and polarity, and 62.0% for POS and emotion and linguistics. Moreover, RF produced 67.9% accuracy when polarity

## Chapter 7. Experiments and Analysis

Table 7.3: Results of the Evaluation of Combined Features in the real\_fake Dataset.

		POS+E	POS+L	POS+P	POS+E+L	POS+P+E	E+L+P	E+L	L+P	E+P
Naive Bayes (NB)	Precision	54.2	59.4	58.3	60.1	58.3	62.3	63.0	60.8	58.0
	Recall	53.6	58.1	56.6	58.7	56.6	61.1	61.4	59.9	57.5
	F-Measure	49.1	55.7	52.9	56.4	52.9	59.6	59.7	58.5	57.3
	Accuracy	53.3	59.2	58.2	59.9	56.7	62.1	62.7	60.5	57.9
Random Forest (RF)	Precision	66.4	76.3	61.9	76.3	67.1	75.7	69.6	67.8	61.7
	Recall	66.3	76.2	61.7	76.4	66.9	75.5	69.6	67.8	61.7
	F-Measure	66.3	76.2	61.7	76.3	66.8	75.5	69.6	67.8	61.7
	Accuracy	66.4	76.1	61.8	76.4	66.9	75.6	69.6	67.9	61.7
SVM	Precision	56.8	63.7	55.0	64.2	57.3	61.1	58.7	59.6	52.2
	Recall	57.3	62.4	55.2	62.9	57.4	61.3	58.3	59.8	52.7
	F-Measure	56.3	61.6	54.5	62.0	57.3	61.0	59.0	59.3	50.1
	Accuracy	57.3	63.7	60.1	63.2	57.4	61.1	58.2	59.2	52.1
Logistic Regression (LR)	Precision	62.1	69.6	62.2	68.8	61.9	54.0	62.4	63.6	55.8
	Recall	62.1	69.5	62.2	68.7	61.9	63.9	62.3	63.8	55.7
	F-Measure	62.1	69.5	62.2	68.7	61.9	63.9	62.5	63.7	55.8
	Accuracy	62.1	69.6	62.2	68.7	61.9	59.7	62.1	63.8	55.6

and linguistics were combined. However, when substituting the polarity feature with emotion, the accuracy score increased to 69.6%. Random forest produced 61.7% accuracy when POS was combined with polarity and an increase to 66.3% accuracy when the emotion features were replaced with POS. The same classifier produced 61.7% accuracy when POS was combined with polarity, whereas an increase to 66.3% accuracy was generated when replacing the emotion features with POS.

### Optimal Model

From the previous experiments, we find that the best performing classifier was RF. It achieved an accuracy of 77.2% using all 30 textual features in POS, emotion, polarity, and linguistics to classify real and fake articles in Arabic. Throughout this research, we refer to this model as the ‘optimal model’.

### 7.1.4 Dataset Lexical Densities

The experiments led to an analysis that focused on supervised learning with predefined labels. However, computing lexical densities of each textual feature can provide insights into trends within the data. A comparison of classes' lexical densities, real and fake, for the `real_fake` and `COVID` dataset might provide more meaningful results. We calculated the `COVID` lexical densities as it also includes real and fake articles collected from real world cases and we wanted to quantitatively observe the behaviour of fake articles in both datasets. As a reminder, the `COVID` dataset contained 26 fake articles about COVID-19 collected from fact checking websites and 26 real articles about COVID-19 from legitimate websites. `Real_fake` contained 549 real and 549 fake articles generated through crowdsourcing. The lexical densities of both `real_fake.xlsx` and `COVID.xlsx` were calculated, and their results are noted in Table 7.4.

The comparison between word use in real and fake articles in the `real_fake` and `COVID` datasets shows some slight differences. Specifically, there is a slight increase of .10 in both the nouns and conjunctions in fake articles compared to real ones. A more obvious increase in fake articles is in the number of verbs, adverbs, and particles in the `COVID` dataset compared to the same features for fake articles in the `real_fake` dataset. A noticeable decrease in word use occurs in prepositions, determiners, adjectives, and proper nouns for fake articles in the `COVID` dataset compared to real articles in the same dataset. Moreover, a similar decrease is found in adjectives and proper nouns in the fake articles in the `real_fake` dataset compared to real articles in the same dataset.

The comparison shows a word use increase specifically for sadness, fear, disgust, surprise, negative words, and positive words in fake articles in `real_fake` and `COVID`. However, the number of negative words increased by around .80 in fake articles in the `real_fake` dataset compared to real articles. This was .40 increase in fake articles in the `COVID` dataset. This may be explained by the number

Table 7.4: Lexical Densities of Feature Categories.

<b>Dataset</b>	<b>Real_fake</b>		<b>Covid</b>	
<b>Class</b>	<b>Real</b>	<b>Fake</b>	<b>Real</b>	<b>Fake</b>
Nouns	30.84	31.16	26.29	27.52
Verbs	3.93	4.29	4.04	6.33
Prepositions	8.8	9.63	10.01	9.43
Determiners	18.0	19.03	16.63	14.17
Interjections	0.00	0.01	0.027	0.07
Adverbs	0.125	0.137	0.24	0.31
Adjectives	6.45	6.30	7.23	5.86
Conjunctions	5.34	5.56	4.18	4.68
Proper nouns	10.59	9.64	0.74	0.011
Pronouns	1.61	1.83	2.03	2.64
Particles	10.67	11.78	13.45	15.74
Anger	0.48	0.67	0.08	0.05
Sadness	0.25	0.268	0.02	0.07
Fear	0.21	0.232	0	0.01
Joy	1.40	1.08	0.13	0.13
Disgust	0.03	0.056	0	0.018
Surprise	0.16	0.175	0	0.018
Negative	0.16	0.24	0.7	0.11
Positive	0.31	0.36	0.21	0.31
Assurances	0.43	0.734	0.44	0.577
Negations	0.35	0.49	0.58	0.44
Illustrations	0.04	0.06	0	0.027
Intensifiers	0.12	0.21	0.32	0.34
Hedges	0.39	0.533	0.10	0.18
Justifications	1.20	1.596	0.656	0.41
Temporal	0.65	0.853	0.09	0.05
Spatial	0.50	0.514	0.24	0.19
Exclusive	0.14	0.133	0.262	0.082
Superlatives	0.45	0.364	0.29	0.13
Oppositions	0.08	0.49	0.29	0.11



of fake articles in the `real_fake` dataset, 549, which is more than 10 times the number of fake articles in the COVID dataset, 26. In the same vein, the number of positive words increased by around .10 in fake articles in the COVID dataset compared to a slight increase of .05 in the `real_fake` dataset. Conversely, there was a higher decrease in the number of joy emotion words in fake articles in the `real_fake` dataset, compared to no change in joy words in the COVID dataset.

Analyses of the data demonstrate a general pattern in the amount of words used in fake articles, specifically in assurance, negations, intensifiers, hedges, justification, temporal, and oppositions in `real_fake`. A further increase in the number of assurance words, hedges, and intensifiers was evident in the fake articles in the COVID dataset. However, there was a high decrease in fake articles' negation, justification, temporal, spatial, exclusion, superlatives, and oppositions in the COVID dataset.

### 7.1.5 Important Features

I applied a chi-square test using the `ChiSquaredAttributeEval` option in WEKA to evaluate the important features in both the `real_fake` and the COVID datasets that affected the optimal model's performance. This was to evaluate the influence of each textual feature by statistically computing the chi-square of each textual feature with respect to each class. `Real_fake.arff` and `COVID.arff` were employed for this task. Table 7.5 shows the important features in `real_fake`, while Table 7.6 shows important features in the Covid dataset. Features are arranged in decreasing order of importance

There is excellent agreement between the two datasets in intensifiers, justifications, hedges, negators, superlatives, and oppositions. As these features were found in the top 11 textual features that are influential in classifying real and fake articles, they show their dominance in `real_fake` and COVID. Negative polarity has a higher impact on identifying fake news than positive polarity. Anger is the

Table 7.5: Important Features in real\_fake.

<b>All Features</b>	<b>Intensifier, Conjunctions, Justification, Anger, Determiners, Adverbs, joy, Hedges, Particles, Negators, Superlatives, Oppositions, Assurance</b>
POS	conjunctions, determiners, adverbs, particles, verbs
Polarity	Negative, positive
Emotion	anger, joy, fear, sad, disgust
Linguistics	intensifier, justification, hedges, negators, superlatives, oppositions, assurance, time, place

Table 7.6: Important Features in Covid.

<b>All Features</b>	<b>Proper Nouns, Intensifier, Hedges, Pronouns, Superlatives, Nouns, Adjectives, Surprise, Negators, Oppositions, Justification</b>
POS	proper nouns, nouns, adjectives, prepositions, verbs.
Polarity	Negative, positive
Emotion	surprise, sad, disgust, joy, fear, anger
Linguistics	intensifier, hedges, superlatives, spatial, opposition, exception

highest of the 11 dominant features in the real\_fake dataset while surprise is the highest in COVID. However, it is difficult to draw generalized conclusions based on the results shown here, as there is variance in the dominance of the textual features in both datasets.

## 7.2 Evaluation. Part 2: Experiments Based on the Research Questions

In this section, we conduct experiments to answer the research questions presented in Chapter 1. The following was the experiments scenario:

- The satire\_nonsatire.arff was tested for the model’s performance to classify articles as satire or non-satire.
- To evaluate the model’s performance in classifying the article’s country of

origin, Hajj\_CO.arff and Brexit\_CO.arff were tested.

- The COVID.arff was employed as an unseen test set to evaluate the optimal model's performance with real-world fake articles.
- A sample of 10 articles formed in test.arff from the real\_fake dataset was tested using humans and the optimal model.
- The file real\_fake\_stem.arff was tested to evaluate the optimal model's performance against stemmed files.

### 7.2.1 Experiment 4: Satire\_Nonsatire Classification

In this experiment, we seek to answer Research Question 1: How well does the proposed approach to classification model also classify another type of fake news, such as satire? For that, satire\_nonsatire.arff, which is composed of 262 non-satire articles labelled 0 and 262 satire articles labelled 1, and 30 textual features of POS, emotion, polarity, linguistics was randomly split into 80% (420 articles) for training and 20% (105 articles) for testing in WEKA. The results of this experiment are presented in Table 7.7.

The table reveals the overall accuracy achieved by each classifier is well above 60%. Of the four classifiers, RF has the best performance, with a 73.3% accuracy score. The accuracy score of NB was a close second, with 68.8%. SVM generated 67% for precision, recall, F-measure, and accuracy, while LR produced 63% for precision and recall and 64% for F-measure and accuracy. An average accuracy of 68.8% was generated by NB. Lexical densities for the satire\_nonsatire dataset are provided in Appendix B.

Table 7.7: Results of Evaluation of All Features in the Satire\_Nonsatire Dataset.

	All Features	
Naive Bayes (NB)	Precision	67.4
	Recall	67.5
	F- Measure	68.9
	Accuracy	68.8
Random Forest (RF)	Precision	72.7
	Recall	73.2
	F- Measure	72.6
	Accuracy	73.3
SVM	Precision	67.0
	Recall	67.0
	F- Measure	67.0
	Accuracy	67.1
Logistic Regression (LR)	Precision	63.7
	Recall	63.8
	F- Measure	64.1
	Accuracy	64.2

### 7.2.2 Experiment 5: Country of Origin Classification

The results of this experiment provide an answer to research question 2: How well does the proposed approach to classification model be used for another classification task, such as classifying an article’s country of origin? In this experiment, the Hajj\_CO.arrf and Brexit\_CO.arrf files were employed in WEKA and the model’s performance was evaluated. An 80% training, 20% testing method was used to evaluate the classifiers’ performance in classifying articles based on the country of origin, using the 30 selected features of POS, emotion, polarity, and linguistics. Hajj\_CO contained the results of 30 textual features for 694 articles from Saudi Arabia, 695 articles from Egypt, and 685 articles from Jordan, labelled 1, 2, and 3, respectively. Brexit\_CO contained the results of 486 Saudi Arabian, 481 Egyptian, and 485 Jordanian articles, labelled as 1, 2, and 3, respectively. Classification results for this experiment are presented in Table 7.8.

The RF produced moderate accuracy of 70.3% and 67.9% for Hajj\_CO and

Table 7.8: Results of the Evaluation of the Country-of-Origin Dataset Based on the Hajj\_CO and Brexit\_CO Datasets.

		<b>Hajj_CO</b>	<b>Brexit_CO</b>
Naive Bayes (NB)	Precision	51.2	53.2
	Recall	50.5	54.1
	F- Measure	49.9	51.7
	Accuracy	50.3	52.9
Random Forest (RF)	Precision	70.3	68.0
	Recall	70.4	67.8
	F- Measure	70.4	67.7
	Accuracy	70.3	67.9
SVM	Precision	52.4	52.2
	Recall	46.9	48.2
	F- Measure	45.1	45.0
	Accuracy	48.9	49.1
Logistic Regression (LR)	Precision	65.8	64.8
	Recall	65.8	64.3
	F- Measure	65.8	64.3
	Accuracy	65.8	64.2

Brexit\_CO, respectively, as shown in Table 7.8. SVM and NB produced poor results of less than 53% accuracy for both datasets. SVM had an accuracy score of 48.9% and NB of 50.3% for Hajj\_CO and SVM had an accuracy score of 49.1% and NB of 52.9% for Brexit\_CO. There was a slight increase in accuracy for the LR classifier results for Hajj\_CO (65.8%) compared to Brexit\_CO (64.2%). The proposed model's compilation, based on textual analysis, is thus moderately successful in classifying articles based on country of origin within the same topic domain, using the RF classifier. However, a cross-validation experiment using the Hajj dataset as the training data and Brexit dataset as the testing data revealed poor model performance (11.4% and 33.6%) for precision and recall, respectively. The model did not identify distinct features that could represent the country of origin in different topic domains as well as it did when classifying articles' country of origin within the same topic domain. We further calculated the lexical densities for Hajj\_Co and Brexit\_Co to get a better perspective on the news articles' data

Table 7.9: RF Confusion Matrix of the Unseen Dataset.

Classifier	Real	Fake
Real	14	12
Fake	9	17

pattern, refer to Appendix B.

### 7.2.3 Experiment 6: real\_fake Classification Using Unseen (COVID) Dataset

In this experiment, we seek to answer Research Question 3: What is the performance of the proposed model on unseen real and fake articles distributed in the real world? COVID.arrf, which is composed of 26 real articles labelled 0 and 26 fake articles labelled 1, and 30 textual features of POS, emotion, polarity, linguistics, was uploaded fully as an unseen test dataset under the proposed model compiled in Experiment 1 for real\_fake classification with the RF classifier. We focused on the confusion matrix as this displays the true positive and true negative results as the correctly classified real and fake articles, respectively. The confusion matrix also provides the model’s performance on false negatives and false positives, as these metrics present the incorrectly predicted fake articles as real and real articles as fake, respectively. In this study, false negatives are the most dangerous predictions, as they identify fake articles as real. The classifier’s prediction confusion matrix for this experiment is shown in Table 7.9.

The model classified fake articles correctly more than it correctly classified real articles. From Table 7.9, it is clear that 14 real articles were correctly classified, while the other 12 were mistaken for fake articles. However, in terms of classifying fake articles, the model correctly classified 17 fake articles, which is more than 50% of the dataset. The model identified nine fake articles as real. Although the number is small, it is still around 25%. We further analysed these



## Chapter 7. Experiments and Analysis

by (Liu et al., 2019). They found that fake medical articles contained similar formal terms as those used in real articles to sound legitimate, such as ‘medicine’, ‘food’, and ‘doctor’. However, in the same study, several unrelated words were mentioned in the fake medical articles, such as ‘WeChat’ and ‘kid’. Which is similar to the findings we found of unrelated words such as ‘pig’ and ‘mint’ in the fake COVID articles.

### **Analysis of Real Articles**

I employed the 12 real articles misclassified as fake into the word cloud function in Orange to identify the most frequently occurring words. We found that some of the most frequent words describe the pandemic in a negative way, such as ‘deadly’ and ‘overwhelming’. In fact, eight out of the 12 real articles contained more than four negative words and two anger and sadness emotion words. These articles described the COVID effect on the global economy, the healthcare system being overwhelmed, and death tolls around the world. The highly emotional and negative terms used in these real articles might have confused the model, leading to misclassification of these articles as ‘fake’.

### **7.2.4 Experiment 7: Evaluation of Model against Human Performance**

To measure the relative value of using an automated fake news detection classifier for news articles, and in order to answer Research Question 4—How might the proposed model perform compared to humans in classifying real and fake articles? The model’s performance was compared to unaided human prediction. Test.arrf, a randomly chosen sample of 10 articles (five real, five fake) from the real\_fake dataset, was employed as unseen test data using the optimal model compiled in Experiment 1 for real\_fake classification with RF classifier. The same set of



Table 7.10: Evaluation of Human Classification of the real\_fake Dataset.

Range of Correctly Classified (Out of 10)	Number of Participants
1-3	2
4-5	3
6-7	4
7	Model
8, the highest	1

articles was handed out to 10 participants, with the task to predict the articles as ‘real’ or ‘fake.’ we then compared the number of correctly classified sample articles by the humans and model. Detailed results are provided in Table 7.10.

The proposed model correctly classified seven articles out of 10, which is more than 50%. Manually checking the predictions, we found that the seven correctly classified articles included four true positives (TP) and three true negative (TN) articles. The four TP articles correctly classified by the model and seven of the 10 students were real articles. One contained information about procedures for obtaining Hajj visas, two articles detailed the medical services provided by the Jordanian Hajj Ministry, and one article was an Egyptian article about a meeting between Hajj religious leaders. Overall, the real articles were objective with no identified subjective language.

The three TN were correctly classified by the model and four out of the 10 students, which means the model outperformed six humans in their predictions. These fake articles were contained a sugar coating of intensifiers and justification terms. For example, one article detailed the number of deaths among Egyptian pilgrims by manipulating a high number of deaths, creating fake incidents of fire blazes that were intentionally created at the scene by enemies, and poor services provided by Arab governments for the pilgrims. The fake articles were filled with intensifiers and justification terms to sound reasonable.

On the other hand, the model predicted two false positives (FP), which means two fake articles were classified as real. One of the articles had been manipulated

by changing the numbers, dates, and common nouns relating to the topic. For example, the original articles detailed the number of Egyptian pilgrims coming to the Hajj that year. The articles contained specific descriptions of buses ready to accommodate them and the names the areas they were housed in Makkah. The manipulated fake article changed the number of pilgrims, date of arrival, and details of the bus, such as colour from red to blue and from single-decker to double-decker buses. This article was incorrectly classified as real by six students. The other article incorrectly classified as real by the model was a rather long one, 1,564 words, that detailed the Hajj metro opening. This article was full of journalistic register and thus created an almost perfect imitation of a real news article. The fake article stated, for example: “His Royal Highness King Salaman bin Abdul Aziz AlSaud attended the opening of the Hajj metro, according to Dr. Muneer Bantan, the Hajj Minister of Saudi Arabia. . . .” Words such as “according to” and “the spokesman of” are journalistic registers commonly used to highlight the reliability of a news article and further maximise the credibility effect in readers. The article also detailed the time and place of the event as another measure of credibility for readers. It is not surprising to find that there were eight students that had judged at least one of these two articles as real.

Turning to the one real article that was incorrectly predicted as fake by the model and two of the students, there does not seem to be a specific reason for this error. The article simply discussed a joyful event after Hajj season ended in 2018. It detailed the attendees’ names, place, and some gifts provided in the event.

### **7.2.5 Experiment 8: Model Evaluation With Stemming**

The purpose of this experiment was to answer research Question 5: What is the effect of stemming articles on the model’s performance? To investigate the effect of stemming on the optimal model’s performance, the `real_fake_stem.arrrf`

Table 7.11: Results For Emotional Features With the Stemmed Dataset.

		<b>Emotion (E) Stemmed</b>	<b>Emotion (E) non-Stemmed</b>
Naive Bayes (NB)	Precision	56.3	56.8
	Recall	56.8	54.5
	F- Measure	55.4	51.9
Random Forest (RF)	Precision	56.7	58.4
	Recall	57.2	58.4
	F- Measure	56.0	58.3
SVM	Precision	49.0	53.2
	Recall	48.6	54.2
	F- Measure	43.7	50.2
Logistic Regression (LR)	Precision	57.8	55.7
	Recall	57.1	55.4
	F- Measure	55.9	54.9

file was employed in WEKA. A quick reminder, the `real_fake_stem.arff` contains the results of the emotion textual features set extracted from 549 real and 549 fake articles, all stemmed. The model's performance without stemming in real\_fake classification, Experiment 1, was compared to its performance with stemming with respect to the emotion features only. The results are shown in Table 7.11.

Here, we focus on the F-measure, as it has more precise results for the false negatives and false positives. The RF classifier produced an F-measure of 56% when `real_fake_stem` was tested; however, for `real_fake`, it reached an F-measure of 58.3%. By way of contrast, the SVM classifier reached a lower F-score of 43.7% in stemmed compared to 50.2% in non-stemmed. Overall, each classifier produced low results with slight differences between stemmed and non-stemmed files. Thus, the findings suggest that stemming very slightly affects the model's accuracy, reducing its performance by 2.3 in RF stemmed compared to non-stemmed.

## 7.3 Summary

This chapter discussed various experiments to test and evaluate the model's performance in classifying real and fake articles in Arabic based on four textual feature sets; POS, emotion, polarity, and linguistics. In the first part of the chapter, the four sets of textual features were tested individually, combined, and all together. we found that the RF classifier generated promising results of 77.2% accuracy using all textual features to classify real and fake articles. This part also offered detailed lexical densities for each textual feature in both the real\_fake and COVID datasets. Moreover, important textual features that affected the model's performance were exhibited using chi-square statistical computations. The second part of the chapter discussed experiments conducted to answer the research questions listed in Chapter 1. The experiments included satire and non-satire classification, which had good results with an accuracy of 73.2% with the RF classifier, and classifying articles' country of origin. Textual analysis proved to be unreliable in this classification. This part also included insights into the model's performance on unseen fake articles in the same topic domain, the Hajj, and a different topic, COVID. In both topic domains, the model correctly classified more than 50% of the given articles. Finally, we tested the effect of stemming on the proposed model's performance. We found that stemming has little to no effect on the model's performance. These results add to the rapidly expanding field of Arabic fake news detection.

# Chapter 8

## Discussion

The primary goal of this research was to build a supervised machine learning model that classifies Arabic real and fake news. In the previous chapter, We showed that based on four textual feature sets—POS, emotion, polarity, and linguistics—the proposed model under the RF classifier generated promising results of 77.2% accuracy. The previous chapter also exploited lexical densities and the most important features through chi-square statistics for each textual feature in the real\_fake and COVID datasets. The lexical densities and important features gave a better insight into the nature of data and the behaviour of fake articles in terms of word use. We answered the research questions by conducting experiments to classify satire and non-satire articles, which also had good results with an accuracy of 73.2% with the RF classifier. Moreover, we classified articles based on their country of origin, which showed the feasibility of textual analysis in classification tasks such as this within the same topic domain. The optimal model correctly classified seven out of 10 unseen articles about the Hajj and 17 out of 26 unseen fake articles about COVID-19. In this chapter, we provide a comprehensive discussion of the research objectives and research questions, culminating with the results.

## **8.1 Objective 1: Compile an Arabic Fake News Dataset That Includes Real and Fake Articles in Journalistic Writing Style**

Dataset availability is the primary issue for Arabic classification models for real and fake news. The aim of the current research was to create a model that could accurately classify fake articles written in a journalistic manner which portray themselves as legitimate and thus are more difficult to distinguish. In order to achieve this goal, the dataset needed to include articles written in a journalistic style that imitated legitimate articles. Crowdsourcing aided not only in the recruitment of participants to generate fake articles but it also imitated the production of fake news in real world, as real articles are manipulated by fake news writers to create a mutation of the real article with fake content. This method was chosen as it enabled to recruit many participants with various backgrounds, which offers diverse writing styles of fake articles. We created guidelines for participants to ensure that the generated fake articles followed a journalistic style similar to that in real articles.

There are a number of implications to consider when conducting research using crowdsourcing in a sensitive topic such as fake news. First, participants may be afraid to engage in such a sensitive task, since stiff penalties are assigned by some governments to those who distribute fake news. To assuage this fear, the anonymity of the participants must be ensured. Second, one may expect that the number of participants might not be large enough to contribute to a significant number of fake articles. For this reason, rewards should be considered to recruit and encourage participants. Using service platforms like Fiverr might be beneficial. Third, thorough guidelines must be created for participants to follow. These guidelines may be based on methods developed for previous research. In

this research, we followed the method proposed by Rubin et al. (2015) to generate fake news articles through crowdsourcing and relied on Amjad et al. (2020) method to collect and verify the real articles. Finally, to avoid any bias, we assigned annotators to assess the generated fake articles before including them in the dataset. This ensured the compilation of an Arabic fake news dataset that mirrors the production of fake news in the real world. Since the real news articles were extracted from legitimate news articles, it can be assumed that following these requirements led to generating Arabic fake news articles with a journalistic style similar to that in the real articles. In conclusion, the `real_fake` dataset is the first dataset that includes 549 Arabic real and 549 fake articles written in a journalistic style that may be used for future Arabic research.

## **8.2 Objective 2: Investigate the Influence of Four Textual Feature Sets on Identifying Fake News in Arabic**

Below, we detail each textual feature’s influence in classifying Arabic fake articles in line with the research objective. Figure 8.1 shows the lexical densities of all textual features in both real and fake articles in `real_fake`.

### **8.2.1 Emotion and Polarity Features**

The emotion features reflect the overall feeling expressed in the article. We analysed six emotion features: anger, fear, sadness, disgust, joy, and surprise. Polarity features are indicative of the actions and feelings expressed in the article. We investigated two polarity features: negative and positive. Variant word use was found in the emotion textual feature set for both real and fake articles; a possible explanation for this is that emotions depend on the topic discussed. The

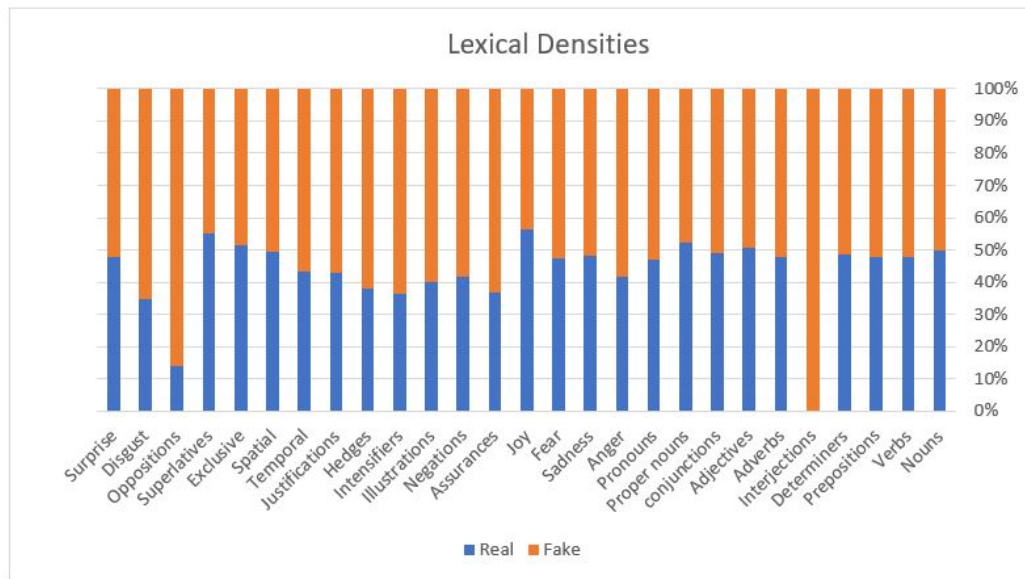


Figure 8.1: Lexical Densities For All Textual Features.

real\_fake dataset contained articles about the Hajj and covered topics such as the pilgrims' death, lack of transportation services, poor catering services, and high prices. These topics raised emotions of anger, fear, and surprise. On the other hand, articles about COVID-19 detailed the pandemic's economic destruction, vaccine fear, and death tolls around the world. This clearly entailed more sadness, fear, and surprise. In fact, we found that anger was the most common emotion feature in the real\_fake dataset, while surprise was the most common in the COVID dataset. The difference in the important emotion features for both datasets implies that no specific emotion feature is indicative of a fake text; rather, the presence of emotional overuse in general implies the probability of fake text. This supports the findings of a recent study by Martel, Pennycook, and Rand (2020) that the use of emotions impacts belief in fake news. Salgado and Bobba (2019) and Soroka and McAdams (2015) have also suggested that news that provokes feelings such as joy or sadness is commonly found to be fake. Deceptive text tends to use a highly emotional tone (Zloteanu et al., 2021).

The same notion that emotional overuse indicates the presence of fake text



could be applied to polarity features. Some studies, such as that by Pérez-Rosas et al. (2017), have shown that positive polarity indicates fake text. However, according to Freelon and Lokot (2020), the overuse of negative polarity might stop readers from believing fake articles. This may be why fake news producers try to use positive polarity terms. However, an opposing study by Rozin and Royzman (2001) has shown the influence of negative polarity in fake text. Rozin and Royzman (2001) study examined the presence of overly used negative polarity in fake text and argued that readers are attracted to such stimuli, leading them to focus on negative topics. In fact, in our research, negative polarity is a dominant feature – more so than positive polarity – in fake articles in `real_fake` and `COVID`. Nevertheless, we cannot generalise this finding to all fake articles since the `real_fake` and `COVID` datasets contained real articles that detailed issues that are by their very nature negative, such as death, pandemic, and negative opinions about Hajj services. The presence of negativity in real articles could have inspired fake news producers to create negative events in their manipulated version of the articles, thus contributing to the overly used negative polarity. In general, emotion or polarity overuse may indicate the presence of deceptive text, which invokes the human’s judgment and model’s prediction of the text as fake.

### 8.2.2 Linguistic Features

These features encode the overall meaning in a statement by linking between ideas or aspects present in the text. In this research, the linguistic features analysed were the following: assurance, negations, illustrations, intensifiers, hedges, justification, temporal, spatial, exclusive, superlatives, and oppositions, 11 features in all.

As stated in Chapter 2, deceivers tend to hide their fear of being caught by overuse of assurance words or persuasive terms to sound credible. We found a similar deceptive pattern when analysing the linguistic terms in fake articles for

real\_fake. Fake news participants supported the statements in their fake articles by using assurance, justification, or both these factors. This is because the fake news producers need to use more effort to sound reasonable in order to balance the inconsistency in their fake articles, which in turn leads to the use of these supporting linguistic features. In fact, assurance and justification features were found to be important features and highly used in fake articles in real\_fake, as seen in Figure 8.1. This supports the results of Addawood et al. (2019) study, in which deceivers were found to use persuasive linguistic cues. Specifically, fake articles in the real\_fake dataset that contained many justification and assurance terms used these words to confirm the manipulated information as well as to provide justification. We found that temporal, spatial, and illustration words were woven into the fake article to support the description fake events in order to make the article appear more credible. An example of this is shown in Figure 8.2, a fake article in real\_fake that contained several justifications in red, temporal phrases in green, spatial in green, illustration in orange, and assurance words in blue:

Another interesting pattern emerged in terms of oppositions and negations. We found there was a relationship between the increase of negations and oppositions, as shown in Figure 8.1. In some fake articles in real\_fake, a statement would include intense use of negative and negations terms that are eased out further by implying opposition terms to introduce a positive polarity. For example, Figure 8.3 shows an article that discusses the negative issue of high prices at the Hajj. The article had negative terms highlighted in blue, negations highlighted in green, and finally eased the article by using an opposition highlighted in red to introduce the positive polarity highlighted in orange. This means that another way fake news producers tend to keep the negative polarity overuse tone down is by using oppositions to create a positive polarity effect in order to balance the negative overuse. We also find that negations and oppositions in the COVID dataset decreased in fake articles compared with real articles.

قال اللواء مصطفى بدير، مساعد وزير الداخلية للشؤون الإدارية والرئيس التنفيذي لبعثة الحج ورئيس بعثة حج القرعة، إنه وصل للأراضي المقدسة الطاهرة 44 ألفا و355 حاجا متضرعا إلى الله سبحانه وتعالى حتى الآن لأداء فريضة الحج

The high executive chief Mustafa Bedair said until now 44 thousand and 355 pilgrims have indeed arrived at the holy land of Makkah to perform the Hajj ritual.

أوضح «بدير» أن 22 ألفا و300 حاج من بعثة القرعة المنظمه وصلوا إلى المملكة العربية السعودية حتى الآن، بواقع 9744 حاجا بمكة المكرمة ..... مثل العام الماضي

He also confirmed the number of pilgrims reached ..... to this day.....similar to last year

حول الأوضاع الصحية للحجاج، أعلن اللواء بدير حزين عن أول حالتين وفاة بين صفوف حجاج البعثة المصرية هذا العام، وأنهما توفيتا إثر إصابتها بجرثومه من أكل غير مطهي ، مشددا على السلطات السعودية بمراقبة المطاعم المقدمه للطعام للحجاج

Considering the health matters of Egyptian pilgrims, he announced the death of two pilgrims this year where they indeed died due to unclean and uncooked food which caused a disease leading to their death. He strictly orders .....

قال ناصر تركي، رئيس لجنة السياحة الدينية بغرفة شركات السياحة، أن لجنة السياحة الدينية تواصل بالتنسيق مع وزارة السياحة جهودها لحل الأزمة، بعد أن أغلقت قنوات الاتصال بكافة الأطراف، مشيرا إلى تجاهل وزير السياحة هشام زعزوع لهذه التحركات وقيامه بالاتصال بوزير الحج السعودي لتحذيره من سرعة حل الأزمة

..... Confirmed that the religious ministry is in contact with the ministry of tourism to solve this matter after the connection have been indeed blocked between the Saudi officials and the Egyptian officials. .... The Egyptian minister of tourism called the Saudi minister to warn him of any consequences and demand fast solutions.

Figure 8.2: Sample of Fake Article Containing Justification, Assurance, Temporal, Spatial, and Illustration Terms.

A third pattern of linguistic usage in fake content was found in the COVID fake articles, where hesitancy in matters related to health issues was expressed using hedge words. An explanation for that may be that fake news writers tend

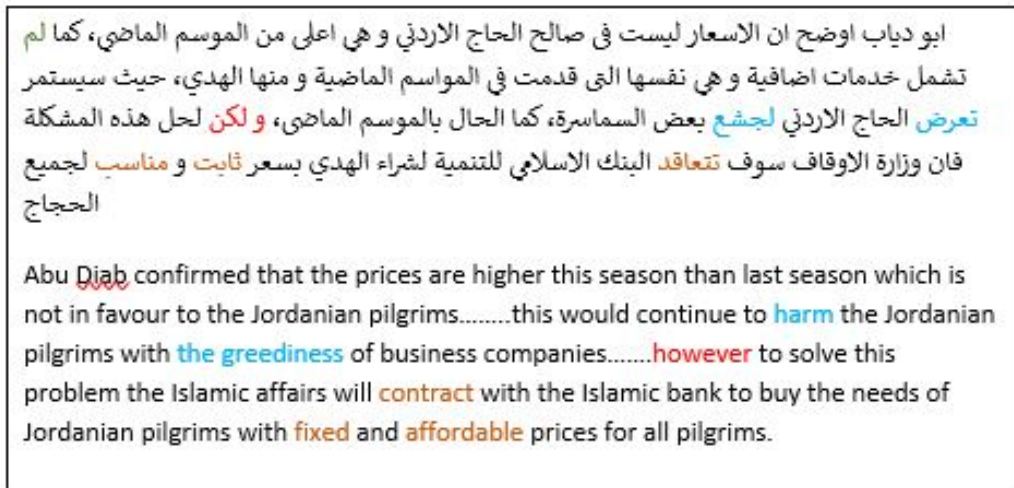


Figure 8.3: Fake Article Contains Negative Polarity in Blue, Oppositions in Red, Then Positive Polarity in Orange.

to be cautious when manipulating information about health matters that could endanger humans. For example, articles that detailed recipes for homemade remedies made from herbs such as mint to combat COVID, used hedging terms such as:

“may be helpful to cure Covid” .... فعالتيه التي قد تساعد في محاربة كورونا

“could be used in the future” .... حيث أنه قد يستخدم في المستقبل كعلاج فعال

Intensifiers were very important features in fake articles in real\_fake and COVID. These findings are in line in with those of Horne and Adali (2017) , Banerjee, Chua, and Kim (2017), Pérez-Rosas et al. (2017), and Sabbeh and Baatwah (2018), who found that fake articles are tailored by exaggeration to attract the reader’s attention. In fact, the fake articles in real\_fake and COVID contained many intensifiers that detailed negative tragic incidents at the Hajj as well as COVID’s deadly effect.

On the other hand, We noted an interesting point that was not expected in superlatives. In contrast to the findings of Dey, Rafi, Parash, Arko, and Chakrabarty (2018) study, where superlatives were frequently used in fake articles when compared to real articles, we found the opposite. The use of superlatives

decreased more than 10% in fake articles in *real\_fake* and COVID compared to real ones. The disparity could be related to the real articles' topics of the Hajj and COVID. Although real news articles are supposed to inform readers about a topic in an objective way, we found that some real articles in *real\_fake* included a high number of superlatives in an attempt to describe the generous services provided at the Hajj. For example, some superlatives described the services provided at the Hajj as *افضل* 'best' and *أجدد* 'newest.' A similar high superlative use was found in real articles in COVID. They detailed the efforts made by governments and health officials to contain the virus by using superlatives such as *أعلى* 'highest' or describing the virus's spreading as *أسرع* 'fastest'. In fact, it is not uncommon for real news writers to use some subjective terms in the form of superlatives to shape the public opinion on a topic (Hamborg, Donnay, & Gipp, 2019).

In general, we found that intensifiers are frequently used in fake articles to attract the readers' attention. However, in order to not expose the intensifier's effect and maintain the implied legitimacy of the fake content, linguistic textual features such as justification, assurance, hedges, negators, oppositions, temporal, and spatial features are woven into the fake content. These textual features help create fake content that sounds legitimate and credible.

### 8.2.3 POS Features

POS features are the POS assigned to each word in the statement, which makes them the building blocks of any article. The POS features analysed in this research were nouns, verbs, adjectives, adverbs, prepositions, particles, conjunctions, determiners, proper nouns, pronouns, and interjections, a total of 11 features.

Content POS such as nouns, verbs, adverbs, and adjectives found in fake articles have been debated in several studies. Pérez-Rosas et al. (2017) find that

verbs and adverbs are more frequently used in fake articles than in real ones, but Kapusta and Obonya (2020) find nouns and adjectives are used a lot in fake articles. Indeed, in the current research, we found that there was an increase in nouns, verbs, and adverbs in fake articles in `real_fake` and `COVID`. We found that adjectives use is reduced in fake articles compared to real ones in both datasets. This is not surprising, because fake articles contain fewer superlatives in our dataset, as discussed above and superlatives are adjectives. Adjective reduction in fake news was also found by Markowitz and Hancock (2014), who found deceptive statements provided far less description than true statements.

Turning to function words, there is an evident increase in prepositions, conjunctions, pronouns, and particles in fake articles in both datasets, as shown in Figure 8.1. Our findings are consistent with those of Posadas-Durán et al. (2019) and Newman et al. (2003), who showed that function words are indicators of people’s feelings, which may expose deceptive writing. A simple reason for the increase in prepositions, conjunctions, pronouns, and particles in fake articles may be that fake writers need to use these function items to create reasonable fabricated events. In fact, conjunctions, particles, and prepositions were found to be dominant textual features in fake articles in both datasets.

Looking deeper into each textual feature set’s influence when tested individually, we find that the RF achieved the highest accuracy, 64.5%, by testing linguistic textual features, followed by POS which reached 63.6% accuracy. However, when emotion and polarity features are combined with POS features, the accuracy reaches a score of 76.6%. Moreover, when all four textual feature sets are combined, they reach an accuracy of 77.2%. Analysing news articles from a linguistic point of view, using linguistic textual features, could offer a better insight into the article’s perspective, which may indicate fake text. Combining that with analysis of the emotional, polarity, and POS of news articles may further assist in identifying fake text.

### **8.3 Objective 3: Develop a Supervised Machine Learning Model That Classifies Real and Fake Articles in Arabic.**

The primary goal of this research was to build a supervised machine learning model that classifies Arabic real and fake news. Fake news distributed through social media or messaging platforms is dangerous as it poses as legitimate while being deceptive. The main issue is that this type of fake content, when distributed through social media or messaging platforms, is not associated with any metadata, such as information about the author or news source. This means that establishing veracity relies only on analysing the text. In our work, we focused on analysing four textual feature sets—POS, emotion, polarity, and linguistics—in news articles to identify deceptive markers that signal the presence of fake text. We compiled a dataset that includes real and fake articles in Arabic, where the fake articles were written in a journalistic manner imitating the real articles. Next, we followed a quantitative approach with the aid of NLP tools to extract four textual feature sets—POS, emotion, polarity, and linguistics—that may identify fake text. In this supervised machine learning technique, we trained our model according to these textual features. The best result was 77.2% accuracy, which was achieved by combining all four textual feature sets, POS, emotion, polarity, and linguistics, for a total of 30 features, as seen in Figure 8.4

#### **8.3.1 Research Question 1: How well does the proposed approach to classification model also classify another type of fake news, such as satire?**

A dataset that comprised 262 satire articles and 262 non-satire articles was created. The satire articles were collected from Arabic satire platforms and the

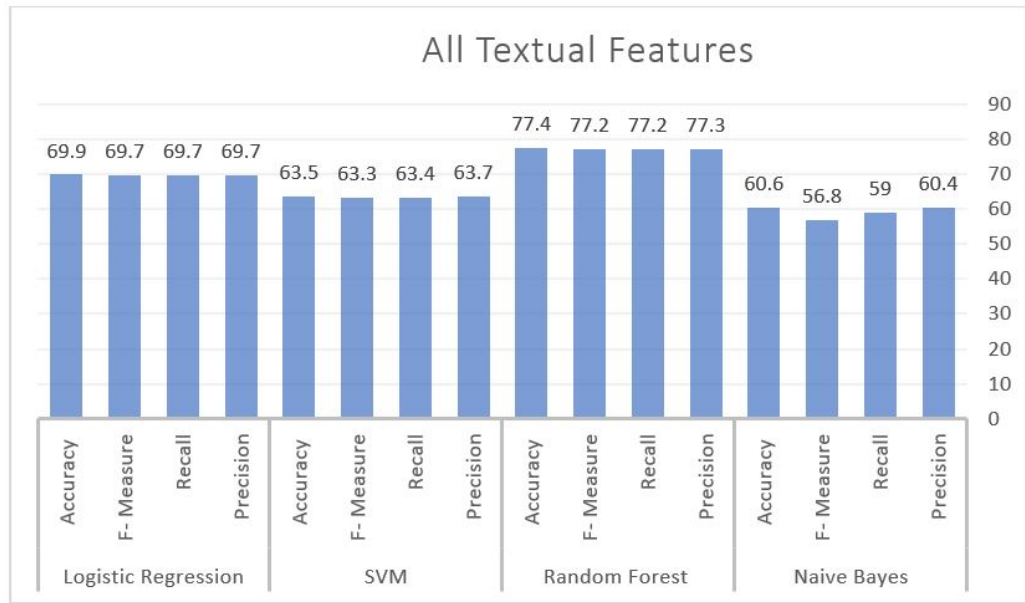


Figure 8.4: Model’s Performance With All Textual Features.

non-satire articles were collected from legitimate news platforms. Each satire article was balanced with a non-satire article that detailed the same topic or key figures introduced in the satire article. We extracted the four textual feature sets—POS, emotion, polarity, and linguistics—from each article. Next, we employed the quantitative results generated from the textual features extraction to compile a model that trains on these data to classify satire and non-satire articles. The proposed model reached a score of 73.4% accuracy with RF, which indicates that the set of features used to classify real and fake articles may be compatible for use to classify satire articles. Moreover, when calculating the lexical densities of the satire\_nonsatire articles (Appendix B), we found similarities in the textual features of intensifiers, superlatives, negations, and oppositions. This suggests that the similarity in the compositions of satire and fake articles, in terms of the deceptive manner, may provide useful insights to compile classification models for other types of deceptive genres.



### 8.3.2 Research Question 2: How well does the proposed approach to classification model be used for another classification task, such as classifying an article’s country of origin?

The objective of this classification task was to ascertain the feasibility of textual analysis in another classification domain: country of origin. For these experiments, we compiled two datasets. The first dataset, Hajj-Co, contained 694 Saudi, 695 Egyptian, and 685 Jordanian news articles about the topic of the Hajj, while the other, Brexit\_CO, contained news articles about Brexit, 486 Saudi, 481 Egyptian, and 485 Jordanian articles. We extracted four textual feature sets—POS, emotion, polarity, and linguistics—for a total of 30 features from all articles in both datasets. The quantitative results were employed to train a model, which performed best with the RF classifier. The RF generated moderate accuracy of 70.4% and 67.7% for Hajj-Co and Brexit-Co, respectively. Thus, these textual features, used to distinguish between real and fake articles, could aid in classifying articles based on their country of origin within the same topic. However, the classification abilities are limited changing the topic domain. All articles were written in formal Arabic and in a journalistic style, as news writers have a recognisable journalistic register, with lexico-grammatical forms commonly used in spoken and written publications (Saadany et al., 2020). For example, words like قال ‘said’ and وضع ‘reported’. Therefore, a native Arabic speaker would immediately realise from these words that the text is part of a news article. However, without unique features that distinguish one country from another, it is challenging for classifiers to link the article to a certain nation. In fact, the model reached a poor 33.2% recall using RF when training the dataset with Hajj articles and testing it with Brexit articles.

At the same time, this research found that these textual features were not

entirely without meaning in the country-of-origin context. Based on the lexical densities of textual features in both country-of-origin datasets, Hajj\_Co and Brexit\_Co (Appendix B), a general understanding of a country’s perspective on a topic might be determined. For example, Hajj articles from Saudi Arabia were more joyful than the articles from the other countries. This maybe due to the fact that Saudi news agencies tend to portray Hajj services provided from the Saudi government to pilgrims in a joyful way. These textual features might provide a better insight into a country’s perspective on a particular topic (Hamborg et al., 2019).

### **8.3.3 Research Question 3: What is the Performance of the Proposed Model on Unseen Real and Fake Articles Distributed in the Real World?**

An experiment that included testing a set of 26 real articles and 26 fake articles about COVID, collected from real-world fake news articles distributed on social messaging platforms, was conducted. The optimal model correctly classified 14 real articles out of 26. However, in terms of classifying fake articles, the model correctly classified 17 fake articles out of 26, which is more than 50% of the dataset. Unfortunately, the model classified nine fake articles as real, which is a cause for concern.

Further analysis in the mistakenly classified fake articles revealed the fake articles included terms that define credibility, similar to those in real ones, such as “according to” and “an insider reported”. Fake articles also contained real world proper nouns such as “Bill Gates”, which also creates a sense of credibility for the reader. These fake articles were well crafted in a journalistic manner with credibility terms inserted, so they are the hardest to differentiate from real articles without prior knowledge. However, fake articles correctly classified by

the model contained deceptive terms. Analysing the textual features proposed helped identify deceptive terms and expose the fake content.

Apart from the models' performances, one intriguing discovery of this experiment was the presence of unrelated words to the topic in fake articles. For example, the words "mint" and "pig" were frequently used in several fake articles, but they have no connection to the medical topic of COVID-19. This was also noted by Cartwright, Nahar, et al. (2019), where one third (399) of tweets from a random sample in their collected fake news dataset ( $N = 1,250$ ) contained what they called "apolitical chatter". These are hashtags unrelated to the subject that do not advance divisive issues. This is very similar to the unrelated terms we found in the COVID fake articles. According to the study, one reason for the apolitical chatter might be that it seems innocuous on the surface, however, it may be used to create difficulties in retrieving tweets linked to the fake producers (a Russian Agency in the study in question). In the case of the current research, the unrelated terms found in the fake articles in COVID might have been used to allay the reader's suspicion by integrating words related to everyday things and seemingly close to the reader's own beliefs. Mint, for example, is a herb considered to have medical remedies by some Arabs, so linking the fake articles to a "known" medical herb such as mint might hide the fake article's deceptive text and focus on the medical aspect of mint. Words such as "pig" were used to support the medical aspect of topics related to the vaccine. The fake articles detailed the vaccine production by talking about pig organs as an ingredient. The presence of these unrelated terms might be a performance by the fake news writer to divert attention from their deceptive text.

Whether these non-related words are referred to as apolitical chatter or simply non-related words, the exact reason of this phenomenon is still not fully understood. However, according to our findings, their presence might be a signal of fake text which could be used in future work to identify fake news. In general,

applying textual analysis on the news articles and training our model on specific textual features to classify real and fake articles in Arabic contributed in the model's ability to correctly classify more than half of first time seen fake articles.

#### **8.3.4 Research Question 4: How might the proposed model perform compared to humans in classifying real and fake articles?**

An experiment that included testing 10 unseen articles about the Hajj (five real and five fake) was performed with the aid of humans and the optimal model. The 10 articles were given to the humans and to the optimal model as an unseen test dataset. In order to ensure that the participants for this experiment were educated, computer science students were chosen. The optimal model performed similarly to the majority of the students, classifying seven out of 10 articles correctly. Specifically, the model correctly classified three fake articles that contained many intensifiers and justification terms to sound reasonable. However, these articles were correctly predicted by only four out of 10 students. On the other hand, two fake articles that were incorrectly predicted as real by the model contained informative text similar to that in real articles. The fake articles were generated by manipulating numbers, dates, and simple adjectives in the text, such as the bus colours from red to blue. These types of fake articles are hard to identify without prior knowledge of the topic. The other article mistakenly predicted as real discussed the metro opening. The details provided were displayed in an objective and informative manner using journalistic registers. This fake article would be hard to identify without attending the event and seeing the details upfront. The argument that well-crafted fake articles that sound legitimate are hard to identify is also supported by the performance of eight out of 10 students that incorrectly predicted at least two of the fake articles written in this manner

as real.

The three fake articles correctly classified by the model and four out of 10 students raised concerns that six students incorrectly classified them as real. We analysed the three fake articles and found a potential reason six out of 10 students believed the fake articles were real. The articles detailed tragic events of Egyptian pilgrims' deaths due to food poisoning, arson incidents from Iranian pilgrims, and a lack of services provided by the Arab governments. Though these articles are fake, the students might have had prior knowledge on the topic or read similar news with regards to such incidents in prior seasons of the Hajj. The fake articles might have been corroborated by students' personal experiences at the Hajj or the topic discussed may have brought to light an unexpressed problem. In fact, this notion is also apparent in a study by Preston et al. (2021), where people believed fake news when it agreed with their personal experiences or fit their beliefs as an important unapparent or unexpressed matter.

These findings reinforce the general belief that the model identifies fake articles based on analysing the textual content and trying to find deceptive markers without emphasis on subjective issues. Yet, people might be influenced by the topics discussed causing subjectivity in their judgments.

### **8.3.5 Research Question 5: What is the effect of stemming articles on the model's performance?**

In this part of the research, we investigated the effect of stemming on the model's performance. To do so, we stemmed a copy of the `real_fake` dataset and extracted emotion textual features. We compared the `real_fake_stem` to the non-stemmed `real_fake` file. The stemmed file `real_fake_stem` reached an F-measure of 56%, while the non-stemmed `real_fake` reached an F-measure of 58.3%, which implies that stemming had little to no effect on the model's performance. In fact, stemming might remove some important affixes used as linguistic categories, such as the

justification affix ُ ‘li’. Stemming is thus not appropriate for this work, and full words are preferred in such a classification task. These results essentially confirm the findings of (Al-Badarneh et al., 2017), whose study demonstrated that stemming might have little or no significance in a classifier’s performance.

## 8.4 Summary

This chapter set out the important theories that underpin textual analysis to classify real and fake articles in Arabic. We found that analysing the article’s linguistic aspects in the form of linguistic textual features might be influential in identifying fake text in fake news articles. A key strength of the research lies in the fact that news articles are analysed in four different aspects: POS, emotion, polarity, and linguistic features, which are useful to identify fake text. We also provided insights into fake news construction that sounds legitimate, concluding that these types of articles are the hardest to differentiate by both the model and by humans, due to their similarity with real articles construction. These insights shed light on the textual content of fake news that may help in constructing models targeted to these implications. We also found similarity in the deceptive nature of satire articles and fake articles, which gave the ability for the optimal model to classify satire and non-satire articles. The optimal model also successfully classified news articles based on their country of origin within the same topic domain following the same textual analysis approach. In general, textual analysis using the four textual feature sets POS, emotion, polarity, and linguistic, followed in the research have proved its feasibility in vital classification tasks.

# Chapter 9

## Conclusion, Limitations, and Future Work

### 9.1 Conclusion

In this research, Arabic news articles were analysed using textual features extracted from the articles' text only. Textual features included POS, emotion, polarity, and linguistics. To extract the textual features, NLP tools were used; however, due to the lack of available NLP tools that support Arabic, a tool had to be constructed that serves the Arabic research community. To construct this, lexical wordlists, which were approved by professionals, were created that can be further used as lexicons for future Arabic research.

This research has made a substantial contribution to research into Arabic and fake news in general, as discussed in the next few sub-sections.

#### 9.1.1 Arabic Fake News Datasets

A key strength of the research lies in the compilation of a dataset that contains fake news in the Arabic journalistic style. Building an Arabic dataset for fake news was not a trivial undertaking, as many Arab governments, such as those

in Saudi Arabia, Egypt, Jordan, Morocco, and Bahrain, prohibit the generation and transmission of fake news, leading to a lack of online fake news platforms to gather samples from. We have overcome this challenge by following the strategy proposed by Rubin et al. (2015) to compile a fake news dataset in Arabic. We followed a crowdsourcing approach. While we faced challenges with this method, such as a lack of participant interest, we offered monetary awards, similar to encouragements made in reality by malicious parties to fake news producers. We also made use of the Fiverr service platform to recruit as many participants as possible. Because of this, the fake articles in the dataset contain diverse writing styles. We ensured the quality of the dataset by applying various veracity measures for real articles, such as that proposed by Amjad et al. (2020), and by manually cross referencing with other news and fact checking platforms. We also ensured the quality of the fake news articles by assigning annotators from the media and journalism domain to annotate them. To our knowledge, this is the first fake news dataset that contains journalistic articles written in Arabic. The recruitment process for participants, the veracity measures carried out for real articles, the annotation process for fake news, and the compilation of real and fake articles in one dataset with research standards offers several valuable contributions to the Arabic research domain.

We also built the `sartire_nonsatire` dataset, which contains 262 satire articles from Arabic satire platforms. We balanced the dataset by collecting 262 non-satire articles on the same topics or figures. In this way, the dataset combined satire and non-satire articles that discussed the same matters and thus provided an opportunity for comparison on these two types of writing. This dataset can be used for further Arabic research on satire detection. Further, we built two topic-specific datasets that included articles from three Arabic countries, Saudi Arabia, Egypt, and Jordan, about the Hajj and Brexit. The `Hajj-Co` dataset contained 694 Saudi, 695 Egyptian, and 685 Jordanian news articles about the topic of the



Hajj. The Brexit\_CO dataset contained news articles about Brexit, 486 Saudi Arabian, 481 Egyptian, and 485 Jordanian articles. These datasets were used to classify the articles' country of origin and can also be used for further Arabic research that necessitates analysing textual specific topic domains.

### 9.1.2 Arabic Wordlists

The textual features investigated in this research were found to be useful in previous non-Arabic research projects that studied deceptive text. We focused on four textual features, POS, emotion, polarity, and linguistics, and tested their influence in identifying fake news in Arabic. To ensure these textual features were properly studied, and because of the lack of available Arabic resources, this process followed several steps. First, we collected and organised 10 linguistic wordlists that included assurance, negation, justification, opposition, exclusion, hedges, intensifier, temporal, spatial, and illustration words. The wordlists were composed from a thorough search of rich Arabic resources—some dating back to 1290. The wordlists were revised and approved by three Arabic scholars. These are the first Arabic linguistic wordlists that offer a variety of fine-grained linguistic textual categories.

Second, we relied on emotion and polarity Arabic lexicons available from Bi Ling lexicons. We extracted the words in each lexicon and organised them in two separate wordlists. The words in the emotion lexicon formed the emotion wordlist and the words in the polarity lexicon formed the polarity wordlist. Because these words came from lexicons widely used in Arabic research, this method ensured that the words included in the wordlists did not need further scholarly approval. The emotion wordlists included anger, sadness, fear, surprise, and disgust. The polarity wordlists included positive and negative. In total, we compiled 18 wordlists that included 10 linguistic categories, six emotion categories, and two polarity categories. Each wordlist can be further used in Arabic research projects.

### 9.1.3 Tasaheel Tool

The tools available to the Arabic research community carry out basic NLP utilities such as stemming, lemmatisation, normalisation, and POS tagging. However, these NLP tools are scattered in several research projects and are not necessarily easily accessible. Some are also not widely known. Because of that, we designed an NLP tool in the Python programming language, named Tasaheel, which offers basic NLP utilities, including novel ones we coded. We collected the available packages freely available online and then combined several of the packages' utilities which offered a variety of functions. For example, we coded the stemming function from the ISRI, Farasa, and Tashaphyne packages. This offers the user several options in one place. Inspired by the POS tagging function and due to the necessity of extracting the emotion, polarity, and linguistic words in the dataset, we also made use of the previously compiled wordlists. We coded a Python script to include a function that offers automatic emotion, polarity, and linguistics tagging. The output of this function is comprehensive, as it offers a summary file with all the tags included for each text file with their number of occurrences. Another option was coded to extract all the summary files into an Excel worksheet for further usage and saving. The tool also provides an option to extract the desired affixes from the text. For that, it offers several novel utilities in the form of emotion, polarity, and linguistic word tagging and affix extraction that will be beneficial for future Arabic projects. Appendix A provides details of the tool.

### 9.1.4 Model Compilation

Throughout this research, qualitative approaches were applied to compile the datasets. For the computer to handle the data in the form of articles and text, we applied quantitative approaches. We extracted the textual features using Tasaheel and another useful tool, Posit. Both tools generated quantitative data on the

textual features investigated. To build a supervised machine learning model that identifies fake news in Arabic, we trained the model on these quantitative data in the form of four textual feature sets: POS, emotion, polarity, and linguistics. Next, we relied on widely used ML classifiers, namely SVM, RF, LR, and NB, to train and test the model. The proposed model achieved an accuracy of 77.2% with the RF classifier. Our model is the first supervised machine learning model that identifies Arabic journalistic written articles as real or fake relying on their textual content only.

## 9.2 Limitations

This work suffers from a number of limitations, notably related to the dataset quantity. The small size of our dataset entailed limited extracted textual features and restricted us from using other approaches such as deep learning, as deep learning requires a large amount of data for training. In the future, we hope to repeat the work on a larger dataset and apply various methods such as deep learning and reinforcement learning.

Furthermore, this work is also limited by its use of NLP tools that detect homographs. Homographs are two words that are written similarly, however, they have different meanings. They are common in Arabic text as diacritic marks are often misplaced, causing accidental homographs. In this research, we manually checked each result produced by the NLP tools to avoid homographs.

A final limitation of this study relates to the type of analysis used for classifying real and fake news articles. We relied on textual analysis by analysing four textual feature sets: POS, emotion, polarity, and linguistics. Though these textual features provided useful training for the model to classify real and fake articles, other non-textual components found in the articles, such as punctuation marks, numbers, and URLs, might aid the classification process.

### 9.3 Future Work

The process of building the supervised machine learning model to classify real and fake Arabic news can be used to build a model that detects other forms of deceptive text, such as spam emails, spam tweets, or fake online reviews. Each of these deceptive texts has its own textual characteristics which may be investigated through textual analysis. We hope to train our model with more samples of fake news and further deploy the model in the cloud for use by governments or to be embedded in chat messaging platforms to alert readers of the presence of fake text. We also want to offer the tool, Tasaheel, to the research community interested in Arabic projects and employ it in the cloud similar to the work of (G. Weir et al., 2018), who deployed their textual analysis tool, Posit, in the cloud for easy use.

Due to the COVID-19 pandemic, experiments on a larger population were restricted, so we were limited to 10 participants to undergo the test. We hope to conduct several tests on larger populations in the future in order to reach a better understanding of fake news in Arabic. The datasets, Tasaheel, the supervised machine learning model, and the insights offered pertaining to fake news in Arabic can be further used to conduct other research projects to combat fake news.

# Appendices

# Appendix A

## Tasaheel Tool

With the rising interest in Arabic research in recent years, it is now possible to find several tools that provide individual NLP tasks or multiple utilities in the form of a comprehensive toolkit. Most offer packages in programming languages such as Python and Java to group classes together to perform certain tasks, thus making it easier for the user to make use of these utilities in their coding scheme. Some tools that provide such packages are Tashaphyne and Information Science Research Institute (ISRI). On the other hand, there are packages that provide multiple tasks in the form of a unified toolkit, such as Farasa and StanfordNLP. A description of each tool is described below:

- Tashaphyne is an Arabic light stemmer and segmental tool. It provides light stemming, such as removing prefixes / suffixes and generates segmentation from that. It relies on using its own built-in customized prefix and suffixes list, which offers more precise stemming. Besides stemming and segmentation, it offers normalization and root extraction.
- ISRI is an Arabic stemming tool without a root dictionary.
- Stanford NLP is a multilanguage NLP tool that can be used for many languages. For Arabic, it provides parsing, tokenization, sentence splitting,

Table A.1: List of NLP Packages.

Package	StanfordNLP	Farasa	Tashahyne	ISRI
Source	StanfordNLP/ CoreNLP	farasa.qcri.org	pypi.org/project/ Tashaphyne	
Language Support	Multilingual	Arabic	Arabic	Arabic
Programming Language	JAVA	JAVA	PYTHON	PYTHON
Stemming	X	X	X	X
Segmentation	X	X	X	
Normalization		X	X	
Name Entity Recognition	X	X		
POS Tagging	X	X		

name entity recognition, and POS tagging. It offers utilities through a Python package.

- Farasa is an Arabic-specific tool that provides NLP utilities through a collection of Java libraries. The utilities include discretization, segmentation, POS tagging, NER, and parsing.

Table A.1 shows each tools' description.

Here, a two-stage framework was proposed to provide a comprehensive solution for Arabic textual research projects that require NLP.

## A.1 Part 1.

This part describes NLP tasks that are already available in the packages supported by NLP tools described in Table A.1. Figure A.1 shows the tool's GUI.

### Stemming

This was undertaken using the following packages—ISRI, Farasa, and Tashaphyne.

## Appendix A. Tasaheel Tool

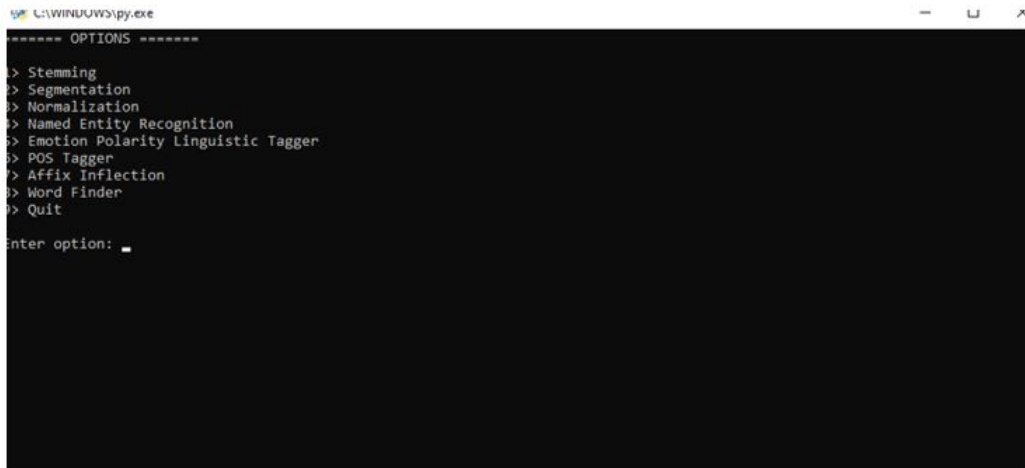


Figure A.1: Tasaheel GUI.

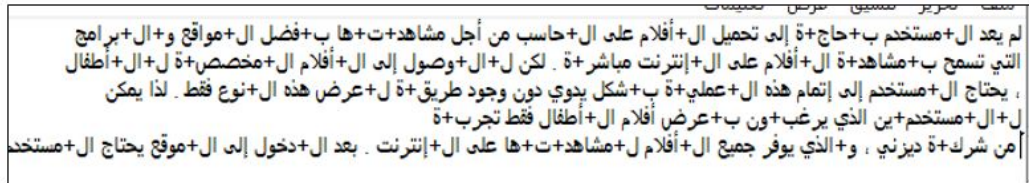


Figure A.2: A file Segmented using the Farasa Segmenter.

### Segmentation

Due to Arabic's unique morphology, it is necessary to segment text into morphemes to decrease the ambiguity in Arabic text that is created from the attached affixes. Here, the packages available, such as Tashaphyne and Farasa, were also used. Figure A.2 shows a segmented file under Farasa.

### Normalization

It is important in many Arabic research projects dealing with text to perform normalization in a manner that unifies text. Normalization helps reduce word ambiguity and remove unnecessary noise associated with the text. Here, the Tashaphyne package was integrated. Another set of options was also provided to perform single normalizing tasks, such as removing numbers, non-Arabic letters,



Appendix A. Tasaheel Tool

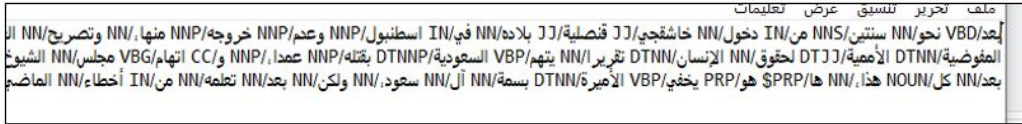


Figure A.3: A File POS-Tagged by the StanfordNLP Tagger.

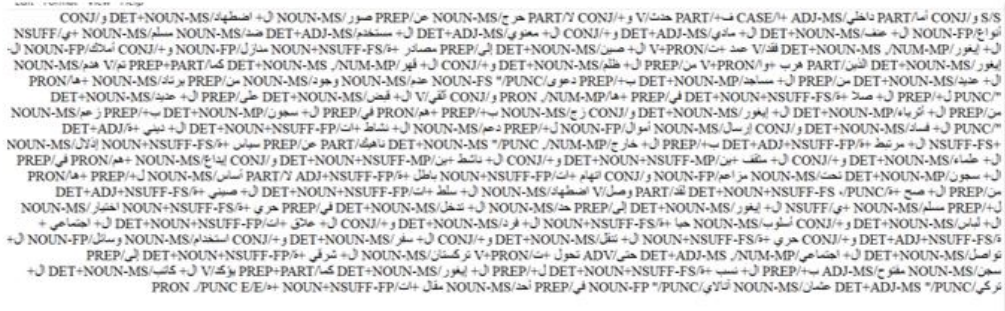


Figure A.4: A File POS Tagged by the Farasa Tagger.

characters, stop words (user provides the list of stop words), and diacritic marks.

**Name Entity Recognition (NER)**

To identify certain figures that might be important during research, NER capabilities were embedded as provided by Farasa and StanfordNLP packages.

**POS Tagging**

To assign a POS to each word in a sentence, POS taggers were used. Further, the user is given two POS tagger types: Farasa or StanfordNLP. Figure A.3 shows a file tagged by the StanfordNLP tagger. Figure A.4 shows file tagged by Farasa tagger.

For each of these tasks, an input folder is provided by the user (which may include an unlimited number of text files), and an output folder named with the option chosen will be generated with the chosen task performed on all the files in that folder. In the figures below, the tools' outputs are shown.

## Appendix A. Tasaheel Tool

حذر [fear] [negative] رئيس مجموعة إيرباص يوم إنترز الخميس [time] من أن شركة الطيران الأوروبية العملاقة قد تتخذ قرارات صارمة جدا بالنسبة لبريطانيا في حال عاينت الاتحاد الأوروبي بدون اتفاق. وذلك في آخر تحديث [fear] من توجه من شركة متعددة الجنسيات.

ووصف إنترز تعامل الحكومة مع بريكتس بالمعيب وقال إن 14 ألف وظيفة في منشآت التصميم وتصنيع الأجنحة التابعة لإيرباص في فيلنكون جنوب غرب انكلترا وفي براونون بشمال ويلز، مهددة.

وحذر إنترز في فيديو نشر على موقع المجموعة إذا حصل بريكتس بدون اتفاق، سيتعين علينا نحن [pronouns] في إيرباص اتخاذ قرارات بحتمل أن تكون صارمة جدا [intensifiers] بالملكة المتحدة.

وقال المدير الإقليمي إن ما [negators] يعمل إلى 110 آلاف وظيفة في بريطانيا تعتمد على مصانع إيرباص البريطانية، التي [linking\_pronouns] تحلق رقم أعمال بلغ حوالي ستة مليارات جنيه (7,8 مليارات دولار) سنويا.

وأضاف أن قطاع الطيران والقضاء بأكمله في البلاد يقف الآن على حافة كارثة.

وحدثت المجموعة التي [linking\_pronouns] كثيرا ما [negators] قالت إن بريكتس قد يعني سحب استثمارات من بريطانيا، عن استيائها الشديد من استراتيجيتها ليريزا ماي الخاصة بريكتس والتي تراوح مكانها.

وقال إنترز من المعيب أنه بعد [time\_place] [place] أكثر من عامين على نتيجة [justification] استفتاء 2016، لا [negators] تزال الشركات غير [exception] فترة على التخطيط بشكل صحيح [positive] للمستقبل.

وأضاف نحن، إلى جانب العديد من زملائنا، قلقينا مرارا بالوضوح لكن [hedge] ما [negators] زلنا لا [negators] نملكه أي فكرة عما يجري هنا.

ووضعت علاقة صانعي الطيران بالاتحاد الأوروبي تحت [time\_place] [place] [time\_place] المحجر مؤخرا، عندما [time\_place] [place] قدمت منظمة التجارة العالمية العام الماضي بأن الشركة تلقت مبالغ [negative] دعم [positive] مخالفة [negative] لتأصل من التكتل، بعدما قالت الولايات المتحدة إنه قدم 22 مليار دولار على شكل مساعدة [positive] حكومية لإطلاق طرازي إي-380 وإيه-350.

وفي وقت لاحق، قالت كاترين بييت، نائبة رئيس شركة إيرباص في المملكة المتحدة لقاء سكاى نيوز إن الحكومة البريطانية طلبت من الشركة توضيح الآثار المحتمل لعدم التوصل إلى اتفاق [positive] (...) ونحن سعداء لتفهم ذلك.

ورفض النواب البريطانيون الأسوع الماضي بأغلبية كبيرة شروط الإسحاب التي [linking\_pronouns] أفلت عليها ماي مع بروكسل، ما [negators] آثار مخاوف جديدة من احتمال [positive] خروج بريطانيا في 29 آذار أيارس بدون اتفاق.

وحدثت مجموعة الضغط القوية بسبب أي المدافعة عن مصانع [positive] الشركات عن الاستياء لأن [justification] ماي لم [negators] تعرض طريقة أكثر وضوحا للنسبي قديما.

وقال إنترز إن السوق العالمي للطيران بنمو بمعدل 5 بالمئة كل [intensifiers] عام لكن [hedge] مستقبلا لا [negators] يعتمد على المملكة المتحدة. سننفي إيرباص على قدر الحداثة وتزدهر مهما [negative] كانت النتيجة.

وتابع السؤال المطروح: هل نرغب المملكة المتحدة في أن تكون جزءا من الحزام المستقبلي هذا؟

من ناحية أخرى، صرح جارك كارني رئيس بنك إنكلترا للمستوى الاقتصادي العالمي في دافوس الخميس [time] أنه لا [negators] تزال هناك [time\_place] [place] [time\_place] أفضاها لوجستية كبيرة تتعين حلها في حال عدم التوصل لاتفاق. وقال الشركات تعمل ما [negators] بوسعها، لكن [hedge] في كثير [intensifiers] من الحالات لا [negators] تستطيع القيام بذلك.

وقالت رئيسة صندوق النقد الدولي كريستين لاغارد أيضا في مقابلة مع إذاعة آر تي إن الفرنسية إن عدم التوصل إلى اتفاق [positive] سيكون بمثابة كارثة يمكن [hedge] أن تؤدي إلى انخفاض قطاعات أعمال كاملة في بريطانيا.

وحدثت ماي بإجراء نقاش وتصويت في 29 كانون الثاني، فيما [time\_place] [place] [time\_place] يدرس النواب لائحة من التغييرات منها إجراء [positive] استفتاء ثان أو إرجاء موعد الخروج المقرر في 29 آذار.

لكن [hedge] كثير [positive] مفاوضات الاتحاد الأوروبي ميشال بارنييه حذر [fear] [negative] الأربعة [time] من أن خطوة [positive] برلمانية لمنع بريكتس بدون اتفاق، ستفشل، ما [negators] تم.

وقال بنو أن هناك [time\_place] [place] [time\_place] هناك [time\_place] [place] [time\_place] أكثرية في مجلس العموم تعارض الخروج بدون اتفاق، ولكن معارضة [negative] الخروج بدون اتفاق [positive] لن [negators] تتبع حصول خروج بدون اتفاق في الموعد الرسمي لتأسيح.

وأضاف لمنع حصول خروج بدون اتفاق، يتعين أن تكون هناك [time\_place] [place] [time\_place] هادئة [intensifiers] موافقة [positive] على حل [positive] آخر.

وحصل إنترز البريطانيون على دعم الإصغاء لجنون مؤيدي بريكتس القائلين إن الشركات المتعددة الجنسيات سننفي يوما في بريطانيا.

Figure A.5: Emotion, Polarity, and Linguistic Word Tagging Process.

### Emotion, Polarity, Linguistic Tagger:

Words may be assigned specific tags that could portray the word from a particular analysis aspect. Here, the emotion, polarity, and linguistic wordlists organized in Appendix B were used and a tagger that assigns emotion, polarity, and linguistic features to words that match those in the wordlist was provided. Further, the following were created: a summary folder that provides each file tagged with appropriate tags, a summary file that contains tag occurrence numbers, and destination files. Here, it should be noted that the user may choose to apply emotion, polarity, or linguistic tagging separately. Each tagger category provides tagging of the files with each category. Figure A.5 shows an example.

Moreover, for both POS tagging and emotion, polarity, and linguistic tagging, two file outputs are produced after this process:

1. Tagged text files
2. A summary file for each tagged document, which displays the number of occurrences of each tag and its ratio, as shown in figure A.6.

## Appendix A. Tasaheel Tool

```
بر التسلق عرض التعليمات
ناتم /NOUN
ين /NSUFF-FD
./PUNC
E/E

Tag data:
S : 1 time
NOUN-MS : 37 time
DET : 34 time
ADJ-MS : 4 time
NOUN : 21 time
NSUFF-FS : 14 time
ADJ : 7 time
NSUFF-FD : 5 time
PREP : 15 time
NUM-MP : 1 time
NOUN-MP : 3 time
NSUFF-FP : 8 time
CONJ : 16 time
PRON : 4 time
PUNC : 8 time
V : 8 time
PART : 7 time
NSUFF-MP : 1 time
E : 1 time

----- List and Tag Information for all_fake_329.txt -----
S/S
ناتم /NOUN-MS
إتام /NOUN-MS
و /CONJ
خطيب /NOUN-MS
ل /DET
```

Figure A.6: Summary File With a Summary of All the Tag.

The ratio was calculated as follows:

$$\text{Ratio} = \frac{\text{tag occurrence in that file}}{\text{number of words in the file}} \quad (\text{A.1})$$

## A.2 Part 2

This part invokes different utilities for textual analysis purposes. This gives a comprehensive approach for the user to analyse the text provided without the hurdle of going through all the text. The analysis utilities are as given below.

### Affix Analyzer

Making use of the affixes produced may provide a more precise analysis of the text. Some affixes are prepositions, pronouns, or conjunctions that perform similar grammatical roles to detached prepositions, pronouns, or conjunctions. To extract affixes, tagging files under Farasa were used, as it provides specific Arabic tags with consideration to Arabic affixes and inflections. Below is an output to

## Appendix A. Tasaheel Tool

the affix extractor. The user is given two options to extract affixes: to extract prefixes or suffixes. Next, two separate prefix / suffix wordlists were used, which allows the tool to look up these affixes to match the input given, here, a folder is provided that includes the following: a summary of each file that contains the affix assigned, the destination files and occurrence of these affixes, an Excel sheet that displays all the affixes and their number of occurrences, and their destination files in an Excel sheet.

As the extraction of affixes was required for this study, this function was added to the Tasaheel tool. Moreover, Igaab and Kareem (2018) stated that affixes play a significant role in changing words' meaning and thus, the grammatical function. As extracting the affixes might be helpful for NLP projects, especially those focusing on textual analysis, this option was added to the tool. To the researcher's knowledge, no work has previously been done to extract affixes in Arabic, although some related work was performed to remove affixes to obtain the roots of the words, and some Arabic NLP tools have employed specific tools for this purpose (Yaseen & Hmeidi, 2014). These tools relied on lemmatization skills that include the removal of the prefixes and suffixes attached to a word (Freihat, Bella, Mubarak, & Giunchiglia, 2018; Al-Shammari & Lin, 2008; Mubarak, 2017). In this context, by identifying the word POS in a sentence, one may create a query based on a syntactic rule that may help identify any affixes attached to it.

Moreover, a unique approach was followed that is similar to the information retrieval method when searching for a query to extract affixes. The tool performs string matching from right to left for prefixes and antefixes (see Algorithm 1). However, string matching is served from left to right when looking for a suffix or postfix (see Algorithm 2). In addition, the tool gives the user the option to search for affixes.

Each tool of Farasa and StanfordNLP, as well as the POS taggers, has a different tag set format. To use the tagged text files produced by different taggers, this

adj / +NSUFF / فعال + NOUN/أخذ PREP/+ب
---

Figure A.7: Tag Identification Example in the Farasa Format.

study's tool was coded to extract affixes in the text files that could be useful in textual analysis. This tool was adjusted to accept queries from the user that contains (affix + BASE\_TAG). The affix is the prefix/antefix or suffix/postfix selected, and the BASE\_TAG is the main word's POS tag. When a query is executed by the user, the tool searches for the chosen query by applying two search methods, depending on the affix. For files tagged under the Farasa tag set format, affixes as prefixes are detected as preposition [PREP] or conjunction [CONJ] tags, whereas suffixes are tagged as [SUFF]. The word's tag may maximally be formed as (BASE\_TAG + suff +affix + affix). An example is displayed in Figure A.7.

However, StanfordNLP follows an approach of tagging each word with its base tag, discarding the affixes attached. The exact word, as in the previous example, is tagged here in the Stanford POS tagger format, as shown in Figure A.8.

The code was modified to include the StanfordNLP POS tag format to extract affixes. However, because affixes are attached, an effort was made to narrow the search domain of affixes to words that start with the same clitics of the affix within the BASE\_TAG matches. Thus, similar to the previous search method,

// /فعالة NN / بأخذ
------------------------

Figure A.8: Tag Identification Example in StanfordNLP Format.

---

**Algorithm 1**

---

```

W = w1, w2, w3.                                ▷ words
T = t1, t2, t3.                                ▷ tags
F(TEXT_FILE) = w1t1, w2t2, w3t3...           ▷ word_tag
input(affix + BASE_TAG)
search(right to left string matching)
for (i=0; i < EOF; i++) do
    if ti= (BASE_TAG) && Wi (first_letter = affix) then
        return Wi
    end if
end for

```

---

**Algorithm 2**

---

```

W = w1, w2, w3.                                ▷ words
T = t1, t2, t3.                                ▷ tags
F(TEXT_FILE) = w1t1, w2t2, w3t3...           ▷ word_tag
input(affix + BASE_TAG)
search(left to right string matching)
for (i=0; i < EOF; i++) do
    if ti= (BASE_TAG) && Wi (first_letter = affix) then
        return Wi
    end if
end for

```

---

the user would input a query made up of (affix+ BASE\_TAG). Moreover, right-to-left string matching and left-to-right string matching were applied to search for prefixes and suffixes.

For example, the affix **س** is considered a preposition denoting ‘future’ when attached to a present-tense verb. For that, the following query was formed:

<p>Affix: <b>س</b></p> <p>BASE_Tag: VBD</p>
---

**Output:** Results of all the words that are tagged with the same BASE\_TAG

```
----- Prefix/Suffix Information for tagged_text2.txt -----
VBP + س:
\ --- سيتئون
```

Figure A.9: Result of the Query.

---

**Algorithm 3** Prefix

---

```

W = w1, w2, w3.                                ▷ words
T = t1, t2, t3.                                ▷ tags
F(TEXT_FILE) = w1t1, w2t2, w3t3...           ▷ word_tag
input(affix\PREP + BASE_TAG) ∨ (affix\CONJ + BASE_TAG)
search(right to left string matching)
for (i=0; i < EOF; i++) do                    ▷ search until end of file
    if ti = ti = (affix\PREP) ∨ (affix\CONJ) then    ▷ if tag = affix
        for (i=0; i < 3; i++) do                ▷ search within three tags
            if ti = (BASE_TAG) then            ▷ is tag = base_tag
                return Wi                        ▷ return the word
            end if
        end for
    end if
end for

```

---

and start with the same clitic letter as the affix.

Figure A.9 shows one of the results of the previous query. At the same time, it shows the affix with its base tag and the word that matches its query.

In all cases, the tool displays all matches that fit the query, their destination file, and their number of occurrences in that file. For example, the figure shows that a query result is in tagged text two and occurred once.

To extract the affixes in the Farasa tag set format, Algorithms 3 and 4 were used:

For example, the affix ك is considered a preposition denoting ‘similarity’. We, therefore, consider the affix constructed as affix / PREP, and the previous query

can be constructed as follows:

Affix: / PREP
BASE_Tag: NOUN

---

**Algorithm 4** Suffix.

---

```

W = w1, w2, w3.                                ▷ words
T = t1, t2, t3.                                ▷ tags
F(TEXT_FILE) = w1t1, w2t2, w3t3...           ▷ word_tag
input(affix\SUFF + BASE_TAG)
search(left to right string matching)
for (i=0; i < EOF; i++) do
  if ti = (BASE_TAG) then
    for i=0; i < 3; i++ do
      if ti = (affix\SUFF) ∨ (affix\SUFF) then
        return Wi
      end if
    end for
  end if
end for

```

---

**Inflection Analyzer**

Similar to the Affix extractor, the tagged files from Farasa were used, and the words/ tags that were tagged with specific inflections in terms of gender and number were extracted. The inflections provided are masculine singular, masculine dual, masculine plural, feminine singular, dual, and plural. Here, a wordlist was created for each of these inflections, embedded in the tool, and the user was given the option to choose the inflection analyses desired. For output, a folder is provided that includes a summary of each file that contains the assigned affix, the destination files and occurrence of these affixes, and an Excel sheet that displays all the affixes and their number of occurrences and their destination files in an Excel sheet.

**Word Matching**

There might be an interest in investigating a certain word use in a group of text files for further implementations. A word matching function is included that gives the option for the user to input a word, and when the match with this word is found in the files, an output file of the word match's file and number of occurrence



## Appendix A. Tasaheel Tool

File	S	PUNC	V	NOUN-MSDET	ADJ-MS	NOUN	NSUFF-FS ADJ	NSUFF-FD PREP	NUM-MP	NOUN-MF	NSUFF-FP CONJ	PRON	E	PART	NSUFF-F			
ksa11.txt	9	38	27	148	131	10	64	51	27	24	58	3	15	14	45	7	9	14
ksa1010.txt	1	5	1	11	17		4	6	6	1	6	4	1	2	4	3	1	1
ksa100100.txt	7	17	26	157	101	10	62	43	30	15	56	7	10	27	29	3	7	10
ksa101101.txt	3	16	8	62	102	9	43	43	30	5	38	2	16	20	31	10	3	5
ksa102102.txt	50	25	80	215	147	25	59	44	25	12	92	17	27	24	50	32	50	77
ksa103103.txt	2	17	5	67	32	5	19	17	8	5	29	4	6	2	13	4	2	5
ksa104104.txt	6	24	18	103	105	14	47	29	26	19	37	4	4	18	41	10	6	12
ksa105105.txt	4	17	14	51	41	4	35	25	16	13	12	4	3	13	17	3	4	2
ksa106106.txt	1	5	3	30	36	10	14	8	5	3	15	1		6	2	5	1	2
ksa107107.txt	2	14	3	39	39	2	20	16	8	6	14		5	5	10		2	
ksa108108.txt	5	14	20	80	97	12	60	52	33	14	47	7	3	22	17	7	5	8
ksa109109.txt	3	12	3	41	33	5	13	14	6	5	11	1	1		6	1	3	
ksa11111.txt	5	29	32	102	110	11	66	53	26	10	64		9	27	40	19	5	25
ksa110110.txt	3	10	5	64	59	13	18	14	11	8	16	4	3	6	15	2	3	2
ksa111111.txt	3	15	24	114	117	24	48	44	24	8	62	1	13	17	40	17	3	14
ksa112112.txt	6	35	8	86	101	10	58	46	26	16	32		13	20	26	1	6	
ksa113113.txt	2	19	11	48	68	10	46	45	23	7	30	7	1	14	13	7	2	4
ksa114114.txt	3	31	14	101	115	7	41	41	27	10	53	1	15	12	40	11	3	7
ksa115115.txt	1	12	3	55	49	14	11	10	5	2	14	1	2	3	10	2	1	2

Figure A.10: Excel Output.

in that file.

### Excel Output

To make it more convenient for the researcher to keep track of the generated data, this option is provided to automatically upload all the generated results from any summary file produced in the previous options, into an Excel sheet that contains the tags as columns and file numbers as rows, with the number of occurrences of each tag in each file, shown in Figure A.10.

One limitation of the program is that it cannot overcome the problem of homographs. Homographs occur when the diacritics are removed. Similar words may be written the same way but have different meanings due to the different positions of the diacritics on the letters. This study's tool resolves research question two, concerning recommendations in building an NLP tool for Arabic textual functions.

# Appendix B

## Satire\_Nonsatire & Hajj\_CO and Brexit\_Co Lexical Densities

### B.1 Satire\_Nonsatire Lexical Density

Here I present the lexical densities of all four textual feature sets of POS+ emotion +linguistic + polarity, Table B.1.

According to Table B.1, I find an increase in most satire articles' textual features categories compared to the non-satire. With attention to nouns, verbs, prepositions, adverbs, proper nouns, and conjunctions in the POS feature sets. This might be due to the fact that satire articles were fused with may "made up" events which involved crafting proper nouns and nouns to aid these made-up events. Along the creation, prepositions and conjunctions were required for an artifice article. This findings is in line with the finds of (Rubin et al., 2015) where shallow syntax (POS) were highly indicative of the presence of satire.

On the other hand, I find that adjectives and determiners were less in satire articles. This might be the fact as the nature of these articles compromised of various topics that were made up, so the focus was on the creation of events than imaginary details using adjectives. Determiners are dominantly used in non

Table B.1: Satire\_Nonsatire Lexical Densities.

Dataset	Satire_nonSatire	
	Non Satire	Satire
Nouns	28.6	29.4
Verbs	5.57	6.65
Prepositions	9.72	9.99
Determiners	16.5	15.5
Interjections	0.01	0.04
Adverbs	0.22	0.29
Adjectives	6.30	5.77
Conjunctions	5.06	5.9
Proper nouns	4.6	5.2
Pronouns	6.10	2.33
Anger	0.08	0.096
Sadness	0.04	0.033
Fear	0.06	0.05
Joy	0.12	0.133
Disgust	0.003	0.015
Surprise	0.02	0.032
Negative	0.93	1.00
Positive	1.145	1.21
Assurances	0.04	0.07
Negations	0.14	0.277
Illustrations	0.06	0.05
Intensifiers	0.03	0.14
Hedges	0.03	0.05
Justifications	0.09	0.04
Temporal	0.08	0.08
Spatial	0.07	0.08
Exclusive	0.018	0.02
Superlatives	0.17	0.22
Oppositions	0.03	0.18

satire articles for referral to key figures with their full title, for example; when referring to princes in the on satire articles they were referred as “Al-Amir”, which means prince. However in some satire articles that referred to the same prince was referred to in a mocking format such as ‘our friend’ صديقنا, which caused the decrease in determiners use.

In like manner, anger, joy, disgust, surprise, and negative were higher used

## Appendix B. Satire\_Nonsatire & Hajj\_CO and Brexit\_Co Lexical Densities

in satire articles compared to positive, fear, and sadness. As anticipated these emotions are geared to portray the humorous mocking nature of the satire articles which may have included anger emotion terms to shift the reader’s perspective toward a hating topic. My experiment is in line with Rubin et al. (2015) that found negative semantic orientations improved their model’s performance to 83%.

As for the linguistic categories, I find that an increase in intensifiers, superlatives, oppositions, negations, and hedges. These sub features are in line with the study of (Karoui et al., 2017) where exaggeration words in terms of intensifiers and opposition words were highlighted as features to identify satire content. Unexpectedly, justification and illustration terms were less used in satire articles. A possible explanation may be that these type of articles are known to be aimed for criticism and call for action with mockery components (Ermida, 2012). This means that no justification terms are needed to persuade the reader into believing the content as legitimate. Similarly, since the articles are based on “made up “events, it is not very necessary to add illustration terms for comparisons. However, uniquely, the “call for action” components in the satire articles would include necessary verbs, intensifiers, and superlatives, which I clearly find in this analysis.

These findings have lead us to conclude textual similarities in both satire and fake news, which implies that the proposed approach might also identify satire articles.

## **B.2 Hajj\_CO and Brexit\_Co Lexical Densities**

Interestingly, the data in Table B.2 shows that articles from KSA used more nouns, adjectives, conjunctions, and pronouns in relation to Hajj than those from the other two countries. This implies that KSA articles detailed Hajj events comprehensively by using many nouns and describing them using adjectives. Further,

conjunctions and pronouns supported the adjective and noun associations. These POS relationships may partly be explained by the fact that Hajj is performed on Saudi land, which in this case, gives the local news agencies better insight into Hajj services such as catering and events such as training. However, the difference is clear in terms of the POS use in comparison to the Brexit dataset, where nouns and adjectives were fairly used as this topic might not directly affect KSA as Hajj does. However, on the other hand, Egyptian and Jordanian articles used more verbs, adverbs, and prepositions in Hajj articles than those from KSA. This finding is expected since articles from both countries published detailed Hajj articles on visa application, services and accommodation procedures, and the financial aspects in their respective countries. In fact, around 48% of Jordanian and Egyptian articles detailed the events that happened in the Hajj season, such as those that happened to their pilgrims during the Hajj journey, such as transportation shortages. As a result, their articles used many verbs to describe the procedures and details that included time and place as well as series of adverbs and prepositions that described time and space.

Another interesting finding was that KSA articles used the highest number of determiners. This can be explained by the fact that these articles introduced important government officials and figures with full titles. For example, to introduce Prince Khaled AlFaisal, the minister of Makkah and Madinah, they referred to him as ‘Saheb Al Sumou Al Malaki Al Amir Khaled AlFaisal AlSaud, amir Makakah Al-Mukarramah wa AlMadina AlMunwarah’. This meant that many ‘Al’ determiners had to be used in KSA articles as compared to other countries.

Based on the results from the table, the following observations can be made. First, KSA articles had more joy and positive words in Hajj than Brexit articles. This is similar to Hamborg et al.’s (Hamborg et al., 2019) statement that countries tend to express the positive side through the news to change the readers’ perceptions of a topic. Since Hajj is performed in KSA, this explains why

Appendix B. Satire\_Nonsatire & Hajj\_CO and Brexit\_Co Lexical Densities

Table B.2: Lexical Densities of POS Tags For Hajj\_CO and Brexit\_CO Datasets.

Dataset	Hajj_CO			Brexit_CO		
	KSA	EGY	JOR	KSA	EGY	JOR
Nouns	30.24	29.83	28.5	29.1	30.2	29.24
Verbs	3.73	4.63	4.81	5.9	6.12	6.00
Prepositions	8.78	9.32	8.72	9.88	10.15	9.36
Determiners	18.1	16.76	15.3	15.26	14.9	15.56
Interjections	0.01	0.01	0.014	0.06	0.05	0.05
Adverbs	0.12	0.20	0.155	0.25	0.26	0.21
Adjectives	6.67	5.86	5.73	6.71	3.32	6.98
Conjunctions	6.42	5.08	5.35	3.49	3.36	3.27
Proper nouns	8.01	7.45	7.28	7.00	6.86	7.09
Pronouns	2.62	1.73	2.06	2.13	1.02	2.27
Anger	0.07	0.17	0.21	0.21	0.28	0.15
Sadness	0.09	0.06	0.10	0.059	0.089	0.057
Fear	0.04	0.02	0.08	0.16	0.20	0.21
Joy	0.52	0.42	0.73	0.43	0.025	0.99
Disgust	0.01	0.02	0.04	0.13	0.071	0.08
Surprise	0.01	0.07	0.04	0.052	0.07	0.033
Negative	0.40	0.49	0.50	0.86	0.99	0.77
Positive	1.08	0.977	0.95	1.37	1.32	1.35
Assurances	0.22	0.22	0.10	0.16	0.09	0.11
Negations	0.25	0.32	0.67	0.22	0.236	0.24
Illustrations	0.14	0.15	0.25	0.043	0.032	0.03
Intensifiers	0.07	0.17	0.17	0.265	0.202	0.17
Hedges	0.11	64	60.0	0.167	0.105	0.11
Justifications	0.03	0.20	0.13	0.267	0.15	0.25
Temporal	0.30	0.65	0.94	0.73	0.825	0.68
Spatial	0.61	0.31	0.52	0.21	0.22	0.32
Exclusive	0.02	0.03	0.02	0.04	0.05	0.07
Superlatives	0.33	0.06	0.12	0.172	0.14	0.09
Oppositions	0.17	0.05	0.09	0.88	0.72	0.48

these articles tended to be more positive. Meanwhile, Egyptian articles had more negative and sadder/fear emotional words in Brexit articles as compared to the other two countries. This can be explained by the fact that 33% of Egyptian articles discussed many of the financial impacts Brexit might have on their economy. Finally, Jordanian articles detailed the most negative emotions toward Hajj, as many of the articles, as stated previously, detailed some shortcomings of Hajj

## Appendix B. Satire\_Nonsatire & Hajj\_CO and Brexit\_Co Lexical Densities

services toward the Jordanian pilgrims. This caused a negative emotional impact.

A closer inspection shows that KSA used more superlatives and intensifiers in the Hajj articles than those on Brexit. This supports the previously stated fact that KSA articles contained more adjectives that were superlatives. Furthermore, KSA articles tended to portray Hajj services using the highest degree of comparison, and thus superlatives and intensifiers were highly used. A similar finding can be seen with both Jordanian and Egyptian articles as they had more prepositions that mentioned time and spatial details; these categories were highly found in these two countries in support of this finding. However, contrary to expectations, these linguistic categories did not significantly have differences the Brexit\_CO dataset. This may be explained by the fact that many articles as possible about Brexit were quoted or sourced from foreign news agencies, resulting in the articles being written and composed similarly without distinctive differences.

# Appendix C

## Model Evaluation Supplements

This appendix demonstrates some of the works done in this research for real\_fake classification or satire\_nonsatire. It also provides sample of arrf file for use in WEKA.

```
-----
Root mean squared error          0.4107
Relative absolute error          53.7267 %
Root relative squared error      81.7488 %
Total Number of Instances       218

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      0.691    0.189    0.825     0.691   0.752     0.498    0.845    0.880    0
      0.811    0.309    0.670     0.811   0.733     0.498    0.845    0.801    1
Weighted Avg.  0.743    0.242    0.757     0.743   0.744     0.498    0.845    0.846

=== Confusion Matrix ===

 a b <-- classified as
85 38 | a = 0
18 77 | b = 1
```

Figure C.1: Sample of real\_fake Classification in WEKA.



```

@relation inf+mis2-weka.filters.unsupervised.attribute.Remove-R2_predicted

@attribute 'prediction margin' numeric
@attribute 'predicted class' {0,1}
@attribute class {0,1}
@attribute justification numeric
@attribute time numeric
@attribute pronouns numeric
@attribute exception numeric

@data
-0.940056,1.7,0.1,0.0,0.0,0.0,0.0,0.1,2,1,0,0,0,0,0,9,22,7,15,4,4,9,1,0,0,0,0,2,1,0,33,7,5,0,0,0,0,0,1,1,68,1,68,2,34,374,5,5
-0.982523,1.7,0.1,0.0,0.0,0.0,0.0,0.0,0,0,0,0,0,0,0,2,10,3,4,5,7,2,1,0,0,0,0,1,0,0,13,3,7,0,0,0,0,0,0,0,6,39,1,39,1,39,228,5,8
-0.997412,1.7,0.0,0.0,0.0,0.0,0.0,0.0,0,2,0,0,0,0,6,16,4,7,2,6,6,0,0,6,0,0,1,5,0,23,4,6,0,0,0,0,0,0,0,54,1,54,3,18,339,6,3
1,0,7,0,1,0,0,3,0,0,1,0,0,1,3,1,0,0,0,0,0,12,23,4,12,5,9,12,0,2,3,2,0,4,4,0,39,4,12,1,0,0,0,0,0,1,84,1,84,2,42,518,6,2
-0.999715,1.7,0.0,0.0,0.1,0,0,0,0,0,2,0,0,0,0,0,3,9,1,2,3,3,3,0,0,2,1,0,1,1,1,13,1,3,0,0,0,0,0,0,0,28,2,14,1,28,165,5,9

```

Figure C.2: Sample of Predicted Classes For test.arrf in WEKA.

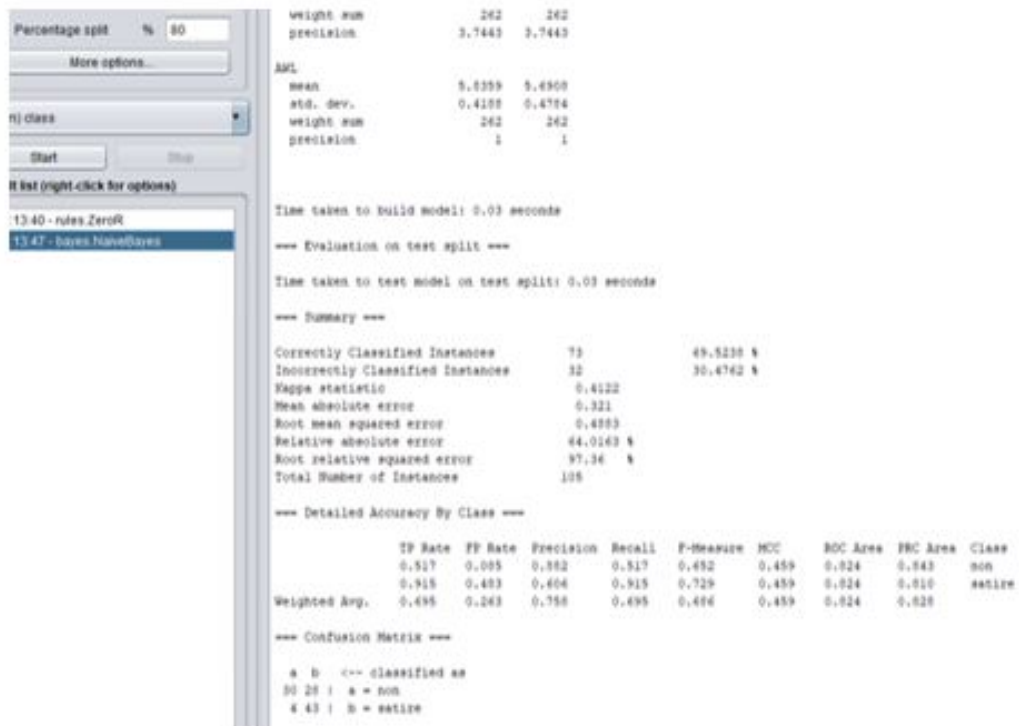


Figure C.3: Sample of Satire\_Nonsatire.arrf Testing in WEKA.

## Appendix D. Ethical Approval



Figure C.4: Sample of Important Emotion Features of Covid using Chi-Square in WEKA .

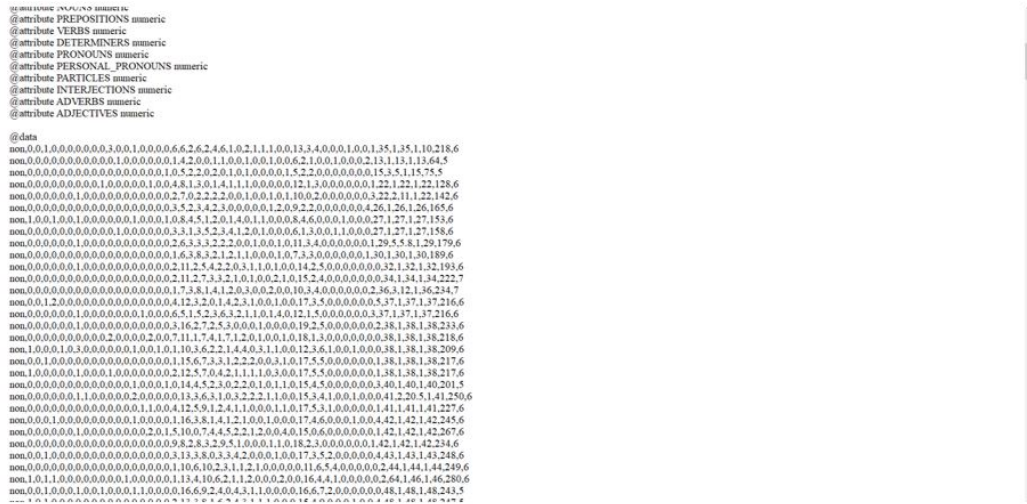


Figure C.5: Sample of Satire\_Nonsatire.arrrf.

# Appendix D

# Ethical Approval

Application ID: 1219

200

Title of research: Classifying Arabic fake news based on Arabic textual analysis

## Appendix D. Ethical Approval

DT/JJ	DTNN	IN	NN-POS	TY-PROPER N	example	Qualifiers	intensifier	negators	time	justif	pronouns	joy	Surprise	Fear	Disgust	Sad	Anger	country	article
10	55	59	26	93	71	0	1	1	0	1	2	0	5	0	0	0	0	0	1
17	17	50	21	129	56	0	0	0	1	1	0	0	6	0	0	0	0	0	1
26	18	52	34	124	54	1	0	2	2	1	0	0	4	0	0	0	0	0	1
19	22	70	33	125	29	1	2	0	0	3	0	0	4	0	0	0	0	0	1
12	27	38	24	109	81	2	0	0	3	1	0	0	7	0	0	0	0	0	1
16	20	44	27	138	34	1	0	0	3	3	0	0	7	0	0	0	0	0	1
15	31	62	24	119	53	0	0	0	0	2	0	1	12	0	0	0	0	0	1
14	20	51	29	109	55	2	0	0	0	3	1	0	2	0	0	0	0	0	1
17	23	44	30	115	64	1	0	2	1	0	1	0	2	0	0	0	0	0	1
14	23	56	41	117	27	3	0	0	6	2	0	6	3	1	0	4	2	0	1
15	23	47	23	145	27	1	1	0	0	0	0	0	4	0	0	0	0	0	1
15	23	46	23	143	27	1	1	0	1	0	0	0	4	0	0	0	0	0	1
24	20	58	35	139	17	1	0	0	9	5	0	2	6	0	0	3	2	4	1
11	40	56	2	72	147	2	0	0	0	1	0	0	0	0	0	0	0	0	1
18	22	50	29	129	43	1	0	0	0	0	0	0	13	1	0	0	0	0	1
18	24	53	29	123	45	0	0	1	2	2	0	0	4	0	0	0	0	0	1
20	23	53	27	129	31	0	0	0	0	2	1	2	8	0	0	0	0	0	1
18	9	51	41	137	25	1	2	0	12	4	0	5	1	0	2	0	3	9	1
18	35	44	19	102	43	1	0	0	0	0	0	5	0	0	0	0	0	0	1
12	33	54	22	93	38	1	0	2	0	3	1	0	3	0	0	0	0	0	1
13	33	52	22	93	37	1	0	2	0	3	0	0	3	0	0	0	0	0	1
13	29	57	15	113	50	2	0	1	3	0	0	1	2	0	0	0	0	0	1
15	45	50	20	123	53	2	0	0	0	2	1	1	5	0	0	0	9	3	1
15	22	52	30	108	46	1	0	0	3	3	2	0	4	0	0	0	0	0	1
19	18	57	23	108	73	0	0	1	0	3	0	0	1	0	0	0	2	0	1
13	34	49	21	110	35	3	1	0	0	3	0	0	0	0	0	0	0	0	1
21	25	52	21	113	30	0	0	1	0	0	0	0	3	0	0	0	0	0	1

Figure C.6: Sample of Hajj\_co.xlsx File Showing the Articles, Country Classes 1,2,3, Emotion, Polarity, Linguistics, and POS Features.

**Summary of research** (short overview of the background and aims of this study):

This study is part of ongoing research that enables the use of a textual analysis approach to classify Arabic fake news articles and classify their country of origin. The study focuses on applying Arabic textual analysis using natural language tools. I will apply supervised machine learning techniques to train my model on certain linguistic markers based on Arabic textual analysis. my model's input will be a collection of real and fake Arabic news articles. Hence, its output will be the system's class prediction (fake or real). my study includes three main stages:

1. **Stage one**—Data collection: Due to the lack of Arabic fake news articles dataset, I will make my own dataset. To this end, participants will be asked to fabricate legitimate Arabic news articles.
2. **Stage two**—Arabic linguistics review: Due to the limitation of Arabic linguistics lexicons, I will make my own linguistics wordlist (simple lexicons). I will create several linguistic wordlists that correspond to certain linguistic purposes. To create the wordlists, I will search online for Arabic wordlists available, and if I cannot find any for a certain linguistic purpose, I will find an English wordlist corresponding to the same category and translate

## Appendix D. Ethical Approval

it to Arabic. Each word in each linguistic wordlist will be matched with the text in the article's content and calculated. This method complies with the textual analysis approach, using natural language processing tools. For example, a wordlist of the 'negators' category includes no, never, not, etc. (the words will be in Arabic, but they were translated here for clarification). This list of words corresponds to the linguistic purpose of 'negators', which gives a nullifying logic. All the wordlists will be matched with the article's content (text); if a match is found, it is calculated. Hence, the number of each linguistic category in each article is calculated. Moreover, I need Arabic linguistic academics to assess my created wordlist.

3. **Stage three**—Test part of the dataset on students to compare their results with my system. I will test a set of given articles from my dataset on participants and my system to assess them. The participants result will then be compared with the system's results.

### **How will participants be recruited?**

1. In stage one, for data collection, I need more than 30 but less than 50 Arabic native speaker participants from both genders to fabricate a given news article. Participants will be invited to participate in this study using the methods listed below.
  - (a) via email: participants will be sent an invitation email first. The email will be sent through the research centre services in Jeddah University. The centre offers an email broadcasting system for researchers to help them reach participants by the student and faculty emails registered in the university. Basically, I will send the invitation email to the research centre, and they will broadcast it via email to all the faculty and students registered in the university. The invitation email will include the inclusion criteria, which are: Arabic native speakers, ages

## Appendix D. Ethical Approval

18–65. There are no exclusion criteria. The email invitation will ask if they would like to participate in this study. If participants reply with agreement, the task will be sent to them to complete and send back to the researcher. Since the task only needs Arabic native speakers at ages between 18 and 65, Jeddah University based in Saudi Arabia will have participants who meet these inclusion criteria. Most of the university faculty and students are native Arabic speakers and are students / faculty members / workers between the ages of 18 and 65. If I do not get the number of participants needed for the study, I will use the following:

- (b) crowdsourcing services such as Amazon Turk, Upwork, and Fiver. A post will be posted in the crowdsourcing website stating the details of the task and email of the researcher for participants to respond. Participants who accept the task posted on the website will be emailed with the task. After completing the task, they will send it back to the researcher.
2. In stage two, for the Arabic linguistic wordlist assessment, 3–5 Arabic linguistics academics participants are needed and will be invited to participate in this study by the following methods.
    - (a) Via email: an invitation email will be sent through the research centre services in Jeddah University. The centre offers an email broadcasting system for researchers to help them reach participants by the student and faculty emails registered in the university. Basically, I will send the invitation email to the research centre, and they will broadcast it via email to all the academics in the Arabic linguistic department. The invitation email will include the inclusion criteria, which are Arabic speakers who have doctoral degrees or a higher academic degree in

## Appendix D. Ethical Approval

linguistics. The email invitation will ask if they would like to participate in this study. If participants reply with agreement, the task will be sent to them to complete and send back to the researcher.

3. In stage three, data testing, I need 30 Arabic native speaker participants, which will be organised through the following methods.

- (a) Via face to face: I will organize with the computer science department chairman in Jeddah University to set aside 10 minutes at the end of a class of one of the faculty members in the department. The class should include not more than 35 students but not less than 30 students. I will organize with the class faculty member to ask the students who want to participate in the survey to stay in class, and those who do not to be dismissed. Since I only need 30 Arabic native speaker students as participants, I am confident that this method will satisfy my needs. They will be given a hard copy of 40 news articles. For articles from 1–20, they should indicate if they perceive it as ‘true’ or ‘not true’. For articles from 20–40, they should indicate which Arabic country they come from, given the options of Saudi Arabia, Jordan, or Egypt.

What will the participants be told about the proposed research study? Either upload or include a copy of the briefing notes issued to participants. In particular, this should include details of yourself, the context of the study and an overview of the data that you plan to collect, your supervisor, and contact details for the Departmental Ethics Committee.

For all three stages of the study:

- Participants will be told that by completing and submitting the task, they are indicating their consent to participate in the study.
- Participants will be informed that their participation is voluntary, and that

## Appendix D. Ethical Approval

there is no harm or risk associated with participation in this study.

- No information will be reported other than to the researchers.
- All responses will be confidential and anonymous.
- Participants have the right to withdraw at any time before submitting their responses. Yet, after submitting their responses, it would be hard to identify the participants' individual responses as no identification information will be collected to identify a particular participant.
- Participants have the right to not answer any question that makes them uncomfortable.
- If participants have any concerns regarding their participation in this study or any other queries, they can contact the researcher or the department's Ethics Committee via email.

Attached is a detailed participant information sheet.

How will consent be demonstrated? Either upload or include here a copy of the consent form/instructions issued to participants. It is particularly important that you make the rights of the participants to freely withdraw from the study at any point (if they begin to feel stressed, for example), nor feel under any pressure or obligation to complete the study, answer any particular question, or undertake any particular task. Their rights regarding associated data collected should also be made explicit.

For all three stages—

1. The first page of the tasks will include consent details, which will make it clear to the participants that:
  - (a) Their participation will be totally anonymous and confidential.

## Appendix D. Ethical Approval

- (b) Their participation is voluntary. Participants have the right to not answer any question that makes them uncomfortable.
- (c) Participants have the right to withdraw at any time before submitting their responses. Yet, after submitting their responses, it would be hard to identify the participants' individual responses as no identification information will be collected to identify a particular participant.
- (d) The questions should not cause any discomfort. Yet, participants have the right to not answer any question that they feel does so.
- (e) Hard copy records will be stored and locked in a cabinet within a locked office, and accessed only by the researcher.
- (f) Electronic data will be stored on the Strathclyde University secure network space in password-protected files.
- (g) If participants have any concerns regarding their participation in this study or any other queries, they could contact the researcher or the department's Ethics Committee via email.

2. The second paper will introduce the question details and the questions.

For stages one and two, data collection and wordlist review, participants who reply to the invitation email and state in the email that they agree to participate (in the email content), only then will the survey be sent to them. As for stage three, data testing, I will meet the student face-to-face. I will announce that students who agree to participate to stay in class and those who do not to be dismissed. Only then will the survey have given to the students in class.

**Attached is details of a consent form.**

What will participants be expected to do? Either upload or include a copy of the instructions issued to participants along with a copy of or link to the survey, interview script or task description you intend to carry out. Please also



## Appendix D. Ethical Approval

confirm (where appropriate) that your supervisor has seen and approved both your planned study, and this associated ethics application.

- Participants who are approached face-to-face will be given a hard copy of the survey, and they will have to complete it and return it to the researcher.
- Participants who are invited via email or through crowdsourcing website. They should complete the questions given to them and then send it back to the researcher.

The survey will be attached and has the tasks for each stage. Note: the survey will be in Arabic so that participants can understand.

What data will be collected and how will it be captured and stored? In particular indicate how adherence to the Data Protection Act and the General Data Protection Regulation (GDPR) will be guaranteed and how participant confidentiality will be handled.

1. No sensitive personal or identifiable information will be collected.
2. Since the collected data will not include any personal or sensitive data, my work does not need to adhere to the provisions of the General Data Protection Regulation (GDPR).
3. The data will be processed and analysed fairly, with limited purposes, and not kept longer than necessary.
4. Hard copy records will be stored in a locked cabinet within a locked office and accessed only by the researcher.
5. electronic data will be stored on the Strathclyde University secure network space in password-protected files.

## Appendix D. Ethical Approval

**How will the data be processed? (e.g., analysed, reported, visualised, integrated with other data, etc.) Please pay particular attention to describing how personal or sensitive data will be handled and how GDPR regulations will be met.**

1. The collected data will be used only for this investigation.
2. The data will be analysed and processed based on the research objectives, using storage and calculation tools such as Microsoft Excel and natural language processing tools such as Python programming language.
3. Access to the data will be suitably secure and restricted to the researcher and the research supervisor.
4. Hard copy records will be stored in a locked cabinet within a locked office and accessed only by the researcher.
5. Electronic data will be stored on the Strathclyde University secure network space in password-protected files.
6. The data will be compared with other sources such as related articles and results from other studies. This comparison will support the reliability of the present study.
7. A report of the findings of this study may be published as a journal article or conference proceeding. The collected data will not include any information that could identify any individual participant.

**How and when will data be disposed of? Either upload a copy of your data management plan or describe how data will be disposed.**

1. Data will not be stored more than necessary and will be disposed of immediately after the conclusion of the researcher's present degree. This will be

## Appendix D. Ethical Approval

done three months after completion of the degree to allow for any minor corrections needed after the viva in August 2021.

2. When data no longer requires to be kept, it will be disposed of appropriately.

This will include:

- (a) Confidential shredding of the collected hard copy surveys.
- (b) Permanent deletion of electronic survey data.

# Appendix E

## Publications

### First Author

- Arabic Fake News Detection based on Textual Analysis, *The Arabian Journal for Science and Engineering*, reference number:AJSE-D-21-02982R3, DOI:10.1007/s13369-021-06449-y
- Arabic News Dataset and Verification System, *IEEE Access*, submitted 20 December 2021.

### Co-author

- Semantic And Sentiment Analysis for Arabic Texts Using Intelligent Model. *Bioscience Biotechnology Research Communications*.(2019). DOI:10.21786/bbrc/12.2/10

## References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., . . . Isard, M. (2016). {TensorFlow}: A system for {Large-Scale} machine learning. In *12th usenix symposium on operating systems design and implementation (osdi 16)* (pp. 265–283).
- Abd, D. H., Khan, W., Thamer, K. A., & Hussain, A. J. (2021). Arabic light stemmer based on isri stemmer. In *International conference on intelligent computing* (pp. 32–45).
- Abdelali, A., Darwish, K., Durrani, N., & Mubarak, H. (2016). Farasa: A fast and furious segmenter for arabic. In *Proceedings of the 2016 conference of the north american chapter of the association for computational linguistics: Demonstrations* (pp. 11–16).
- Abu-Chacra, F. (2007). *Arabic: An essential grammar* (??th ed.). Routledge. Retrieved from <https://www.taylorfrancis.com/books/9781134119189> doi: 10.4324/9780203088814
- Addawood, A., Badawy, A., Lerman, K., & Ferrara, E. (2019). Linguistic cues to deception: Identifying political trolls on social media. In *Proceedings of the international aai conference on web and social media* (Vol. 13, pp. 15–25).
- Adetunji, A., Oguntoye, J., Fenwa, O., & Akande, N. (2018). Web document classification using naïve bayes. *Journal of Advances in Mathematics and Computer Science*, 29(6), 1–11.
- Adha, A. (2020). Linguistic based cues in detecting deception in indonesian

## Appendix E. Publications

- language use. *Argumentum*, 16, 14–30.
- Afroz, S., Brennan, M., & Greenstadt, R. (2012). Detecting hoaxes, frauds, and deception in writing style online. In *2012 IEEE Symposium on Security and Privacy* (pp. 461–475).
- Aladhadh, S., Zhang, X., & Sanderson, M. (2014). Tweet author location impacts on tweet credibility. In *proceedings of the 2014 Australasian Document Computing Symposium* (pp. 73–76).
- Al-Anzi, F. S., & AbuZeina, D. (2015). Stemming impact on arabic text categorization performance: A survey. In *2015 5th International Conference on Information & Communication Technology and Accessibility (ICTA)* (pp. 1–7).
- Al-Ayyoub, M., Jararweh, Y., Rabab'ah, A., & Aldwairi, M. (2017). Feature extraction and selection for arabic tweets authorship authentication. *Journal of Ambient Intelligence and Humanized Computing*, 8(3), 383–393.
- Al-Badarneh, A., Al-Shawakfa, E., Bani-Ismael, B., Al-Rababah, K., & Shatnawi, S. (2017). The impact of indexing approaches on arabic text classification. *Journal of Information Science*, 43(2), 159–173.
- Al-Barhamtoshy, H. M., Hemdi, H. T., Khamis, M. M., & Himdi, T. F. (2019, Jun). Semantic and sentiment analysis for arabic texts using intelligent model. *Bioscience Biotechnology Research Communications*, 12(2), 266–274.
- Aldwairi, M., & Alwahedi, A. (2018). Detecting fake news in social media networks. *Procedia Computer Science*, 141, 215–222.
- Alghamdi, N. S., Mahmoud, H. A. H., Abraham, A., Alanazi, S. A., & García-Hernández, L. (2020). Predicting depression symptoms in an arabic psychological forum. *IEEE Access*, 8, 57317–57334.
- Al-Hashedi, A., Al-Fuhaidi, B., Mohsen, A. M., Ali, Y., Gamal Al-Kaf, H. A., Al-Sorori, W., & Maqtary, N. (2022). Ensemble classifiers for arabic sentiment analysis of social network (twitter data) towards covid-19-related conspiracy

## Appendix E. Publications

- theories. *Applied Computational Intelligence and Soft Computing*, 2022.
- Al-Hussaini, H., & Al-Dossari, H. (2016). A lexicon-based approach to build reputation from social media. *International Research Journal of Electronics and Computer Engineering*, 2(2), 14–19.
- Ali, Z. S., Mansour, W., Elsayed, T., & Al-Ali, A. (2021). Arafacts: the first large arabic dataset of naturally occurring claims. In *Proceedings of the sixth arabic natural language processing workshop* (pp. 231–236).
- Alkhair, M., Meftouh, K., Smaïli, K., & Othman, N. (2019). An arabic corpus of fake news: collection, analysis and classification. In *International conference on arabic language processing* (pp. 292–302).
- Alluhaibi, R., Alfraidi, T., Abdeen, M. A., & Yatimi, A. (2021). A comparative study of arabic part of speech taggers using literary text samples from saudi novels. *Information*, 12(12), 523.
- Alom, Z., Carminati, B., & Ferrari, E. (2018). Detecting spam accounts on twitter. In *2018 ieee/acm international conference on advances in social networks analysis and mining (asonam)* (pp. 1191–1198).
- Alqurashi, S., Hamoui, B., Alashaikh, A., Alhindi, A., & Alanazi, E. (2021). Eating garlic prevents covid-19 infection: Detecting misinformation on the arabic content of twitter. *arXiv preprint arXiv:2101.05626*.
- Al-Sanabani, M., & Al-Hagree, S. (2015). Improved an algorithm for arabic name matching. *Open Transactions on Information Processing*, 2015, 2374–3778.
- Al-Shammari, E., & Lin, J. (2008). A novel arabic lemmatization algorithm. In *Proceedings of the second workshop on analytics for noisy unstructured text data* (pp. 113–118).
- Alsmearat, K., Al-Ayyoub, M., Al-Shalabi, R., & Kanaan, G. (2017). Author gender identification from arabic text. *Journal of Information Security and Applications*, 35, 85-95. doi: <https://doi.org/10.1016/j.jisa.2017.06.003>
- Alsudias, L., & Rayson, P. (2020, July). COVID-19 and Arabic Twitter: How can

## Appendix E. Publications

- Arab world governments and public health organizations learn from social media? In *Proceedings of the 1st workshop on NLP for COVID-19 at ACL 2020*.
- Alwajeeh, A., Al-Ayyoub, M., & Hmeidi, I. (2014). On authorship authentication of arabic articles. In *2014 5th international conference on information and communication systems (icics)* (pp. 1–6).
- Al-Yahya, M., Al-Khalifa, H., Al-Baity, H., AlSaeed, D., & Essam, A. (2021). Arabic fake news detection: Comparative study of neural networks and transformer-based approaches. *Complexity*, 2021.
- Alzanin, S. M., & Azmi, A. M. (2019). Rumor detection in arabic tweets using semi-supervised and unsupervised expectation–maximization. *Knowledge-Based Systems*, 185, 104945.
- Amjad, M., Sidorov, G., Zhila, A., Gómez-Adorno, H., Voronkov, I., & Gelbukh, A. (2020). “bend the truth”: Benchmark dataset for fake news detection in urdu language and its evaluation. *Journal of Intelligent & Fuzzy Systems*, 39(2), 2457–2469.
- Argamon, S., Whitelaw, C., Chase, P., Hota, S. R., Garg, N., & Levitan, S. (2007). Stylistic text classification using functional lexical features. *Journal of the American Society for Information Science and Technology*, 58(6), 802–822.
- Assaf, R., & Saheb, M. (2021). Dataset for arabic fake news. In *2021 ieee 15th international conference on application of information and communication technologies (aict)* (pp. 1–4).
- Atodiresei, C.-S., Tănăselea, A., & Iftene, A. (2018). Identifying fake news and fake users on twitter. *Procedia Computer Science*, 126, 451–461.
- Ayed, R., Labidi, M., & Maraoui, M. (2017). Arabic text classification: New study. In *2017 international conference on engineering & mis (icemis)* (pp. 1–7).



## Appendix E. Publications

- Banerjee, S., Chua, A. Y., & Kim, J.-J. (2017). Don't be deceived: Using linguistic analysis to learn how to discern online review authenticity. *Journal of the Association for Information Science and Technology*, 68(6), 1525–1538.
- Baptista, J. P., & Gradim, A. (2020). Understanding fake news consumption: A review. *Social Sciences*, 9(10), 185.
- Berezenko, V. (2018). Verbal and nonverbal cues to deception in modern english discourse. *Naukovyi visnyk kafedry UNESCO KNLU*, 36, 64–71.
- Bondielli, A., & Marcelloni, F. (2019). A survey on fake news and rumour detection techniques. *Information Sciences*, 497, 38–55.
- Bouchentouf, A. (2013). *Arabic for dummies*. John Wiley & Sons.
- Brown, N., & Collins, J. (2021). Systematic visuo-textual analysis: A framework for analysing visual and textual data. *The Qualitative Report*, 26(4), 1275–1290.
- Burgoon, J. K., Blair, J. P., Qin, T., & Nunamaker, J. F. (2003). Detecting deception through linguistic analysis. In *International conference on intelligence and security informatics* (pp. 91–101).
- Cartwright, B., Nahar, L., Weir, G., Padda, K., & Frank, R. (2019). The weaponization of cloud based social media. *Cloud Computing 2019*, 17.
- Cartwright, B., Weir, G. R., & Frank, R. (2019). Cyberterrorism in the cloud. *Security, Privacy, and Digital Forensics in the Cloud*, 217.
- Chakraborty, A., Paranjape, B., Kakarla, S., & Ganguly, N. (2016). Stop click-bait: Detecting and preventing clickbaits in online news media. In *2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)* (pp. 9–16).
- Choudhary, M., Jha, S., Saxena, D., & Singh, A. K. (2021). A review of fake news detection methods using machine learning. In *2021 2nd international conference for emerging technology (incet)* (pp. 1–5).

## Appendix E. Publications

- Collins, B., Hoang, D. T., Nguyen, N. T., & Hwang, D. (2020). Fake news types and detection models on social media a state-of-the-art survey. In *Asian conference on intelligent information and database systems* (pp. 562–573).
- Conroy, N. K., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. *Proceedings of the association for information science and technology*, 52(1), 1–4.
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). Botnot: A system to evaluate social bots. In *Proceedings of the 25th international conference companion on world wide web* (pp. 273–274).
- Demestichas, K., Remoundou, K., & Adamopoulou, E. (2020). Food for thought: Fighting fake news and online disinformation. *IT Professional*, 22(2), 28–34.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological bulletin*, 129(1), 74.
- Dey, A., Rafi, R. Z., Parash, S. H., Arko, S. K., & Chakrabarty, A. (2018). Fake news pattern recognition using linguistic analysis. In *2018 joint 7th international conference on informatics, electronics & vision (iciev) and 2018 2nd international conference on imaging, vision & pattern recognition (icivpr)* (pp. 305–309).
- Douzi, S., AlShahwan, F. A., Lemoudden, M., & Ouahidi, B. (2020). Hybrid email spam detection model using artificial intelligence. *International Journal of Machine Learning and Computing*, 10(2), 316–322.
- D’Alconzo, A., Drago, I., Morichetta, A., Mellia, M., & Casas, P. (2019). A survey on big data for network traffic monitoring and analysis. *IEEE Transactions on Network and Service Management*, 16(3), 800–813.
- Ekman, P. (1999). Basic emotions. *Handbook of cognition and emotion*, 98(45–

## Appendix E. Publications

60), 16.

- Elaraby, N. M., Elmogy, M., & Barakat, S. (2016). Deep learning: Effective tool for big data analytics. *International Journal of Computer Science Engineering (IJCSE)*, 9.
- Elhadad, M. K., Li, K. F., & Gebali, F. (2020). Detecting misleading information on covid-19. *Ieee Access*, 8, 165201–165215.
- El-Haj, M., Kruschwitz, U., & Fox, C. (2010). Using mechanical turk to create a corpus of arabic summaries.
- Elkateb, S., Black, W., Vossen, P., Farwell, D., Pease, A., & Fellbaum, C. (2009). The challenge of arabic for nlp/mt arabic wordnet and the challenges of arabic. *Citeseer*.
- Ermida, I. (2012). News satire in the press: Linguistic construction of humour inspoof news articles. *Language and humour in the media*, 185.
- Fallis, D. (2014). A functional analysis of disinformation. *IConference 2014 Proceedings*.
- Faustini, P. H. A., & Covoos, T. F. (2020). Fake news detection in multiple platforms and languages. *Expert Systems with Applications*, 158, 113503.
- Feng, V. W., & Hirst, G. (2013). Detecting deceptive opinions with profile compatibility. In *Proceedings of the sixth international joint conference on natural language processing* (pp. 338–346).
- Fleiss, J. L., Levin, B., & Paik, M. C. (2013). *Statistical methods for rates and proportions*. john wiley & sons.
- Floos, A. Y. M. (2016, April). Arabic Rumours Identification By Measuring The Credibility Of Arabic Tweet Content. *International Journal of Knowledge Society Research (IJKSR)*, 7(2), 72-83.
- Fox, M. (2018). Fake news: Lies spread faster on social media than truth does. *NBC News*.
- Freelon, D., & Lokot, T. (2020). Russian twitter disinformation campaigns reach

## Appendix E. Publications

- across the american political spectrum. *Harvard Kennedy School Misinformation Review*, 1(1).
- Freihat, A. A., Bella, G., Mubarak, H., & Giunchiglia, F. (2018). A single-model approach for arabic segmentation, pos tagging, and named entity recognition. In *2018 2nd international conference on natural language and speech processing (icnlsp)* (pp. 1–8).
- Fuller, C. M., Biros, D. P., & Wilson, R. L. (2009). Decision support for determining veracity via linguistic-based cues. *Decision Support Systems*, 46(3), 695–703.
- Fuller, J. (1996). *News values: Ideas for an information age* (Vol. 10). University of Chicago Press.
- Gandía, J. L., & Huguet, D. (2021). Textual analysis and sentiment analysis in accounting: Análisis textual y del sentimiento en contabilidad. *Revista de Contabilidad-Spanish Accounting Review*, 24(2), 168–183.
- Gans, H. J. (2004). *Deciding what's news: A study of cbs evening news, nbc nightly news, newsweek, and time*. Northwestern University Press.
- Genuer, R., Poggi, J.-M., Tuleau-Malot, C., & Villa-Vialaneix, N. (2017). Random forests for big data. *Big Data Research*, 9, 28–46.
- George, J., Skariah, S. M., & Xavier, T. A. (2020). Role of contextual features in fake news detection: a review. In *2020 international conference on innovative trends in information technology (icitiit)* (pp. 1–6).
- Ghorui, k. (2019). *What is big data?* <https://www.geeksforgeeks.org/what-is-big-data/>.
- Golbeck, J., Mauriello, M., Auxier, B., Bhanushali, K. H., Bonk, C., Bouzaghrane, M. A., ... Everett, J. B. (2018). Fake news vs satire: A dataset and analysis. In *Proceedings of the 10th acm conference on web science* (pp. 17–21).
- Gravanis, G., Vakali, A., Diamantaras, K., & Karadais, P. (2019). Behind the

## Appendix E. Publications

- cues: A benchmarking study for fake news detection. *Expert Systems with Applications*, 128, 201–213.
- Gröndahl, T., & Asokan, N. (2019). Text analysis in adversarial settings: Does deception leave a stylistic trace? *ACM Computing Surveys (CSUR)*, 52(3), 1–36.
- Grover, K. (2022). *Advantages and disadvantages of logistic regression*. <https://iq.opengenus.org/advantages-and-disadvantages-of-logistic-regression/>.
- Gupta, M., Bakliwal, A., Agarwal, S., & Mehndiratta, P. (2018). A comparative study of spam sms detection using machine learning classifiers. In *2018 eleventh international conference on contemporary computing (ic3)* (pp. 1–7).
- Gupta, S. (2020, Jun). *Random forest (easily explained)*. <https://medium.com/@gupta020295/random-forest-easily-explained-4b8094feb90>.
- Hajja, M., Yahya, A., & Yahya, A. (2019). Authorship attribution of arabic articles. In *International conference on arabic language processing* (pp. 194–208).
- Hamborg, F., Donnay, K., & Gipp, B. (2019). Automated identification of media bias in news articles: an interdisciplinary literature review. *International Journal on Digital Libraries*, 20(4), 391–415.
- Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2007). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, 45(1), 1–23.
- Haouari, F., Hasanain, M., Suwaileh, R., & Elsayed, T. (2020). Arcov19-rumors: Arabic covid-19 twitter dataset for misinformation detection. *arXiv preprint arXiv:2010.08768*.
- Helmstetter, S., & Paulheim, H. (2018). Weakly supervised learning for fake news detection on twitter. In *2018 ieee/acm international conference on*

## Appendix E. Publications

- advances in social networks analysis and mining (asonam)* (pp. 274–277).
- Helwe, C., Dib, G., Shamas, M., & Elbassuoni, S. (2020). A semi-supervised bert approach for arabic named entity recognition. In *Proceedings of the fifth arabic natural language processing workshop* (pp. 49–57).
- Holes, C. (2004). *Modern arabic: Structures, functions, and varieties*. Georgetown University Press.
- Horne, B., & Adali, S. (2017). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *Proceedings of the international aaai conference on web and social media* (Vol. 11, pp. 759–766).
- Hossain, M. Z., Rahman, M. A., Islam, M. S., & Kar, S. (2020). Banfakenews: A dataset for detecting fake news in bangla. *arXiv preprint arXiv:2004.08789*.
- Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, 50(3), 585–594.
- Igaab, Z. K., & Kareem, I. (2018). Affixation in english and arabic: A contrastive study. *English Language and Literature Studies*, 8(1), 92–103.
- Ireton, C., & Posetti, J. (2018). *Journalism, fake news & disinformation: handbook for journalism education and training*. Unesco Publishing.
- Islam, J., Xiao, L., & Mercer, R. E. (2020). A lexicon-based approach for detecting hedges in informal text. In *Proceedings of the 12th language resources and evaluation conference* (pp. 3109–3113).
- Jardaneh, G., Abdelhaq, H., Buzz, M., & Johnson, D. (2019). Classifying arabic tweets based on credibility using content and user features. In *2019 ieee jordan international joint conference on electrical engineering and information technology (jeeit)* (pp. 596–601).
- Jeronimo, C. L. M., Marinho, L. B., Campelo, C. E., Veloso, A., & da Costa Melo, A. S. (2019). Fake news classification based on subjective language. In

## Appendix E. Publications

- Proceedings of the 21st international conference on information integration and web-based applications & services* (pp. 15–24).
- Jupe, L. M., Vrij, A., Leal, S., & Nahari, G. (2018). Are you for real? exploring language use and unexpected process questions within the detection of identity deception. *Applied Cognitive Psychology*, *32*(5), 622–634.
- Kapusta, J., Hájek, P., Munk, M., & Benko, L. (2020). Comparison of fake and real news based on morphological analysis. *Procedia Computer Science*, *171*, 2285–2293.
- Kapusta, J., & Obonya, J. (2020). Improvement of misleading and fake news classification for fleective languages by morphological group analysis. In *Informatics* (Vol. 7, p. 4).
- Karamizadeh, S., Abdullah, S. M., Manaf, A. A., Zamani, M., & Hooman, A. (2013). An overview of principal component analysis. *Journal of Signal and Information Processing*, *4*.
- Karoui, J., Benamara, F., Moriceau, V., Aussenac-Gilles, N., & Belguith, L. H. (2015). Towards a contextual pragmatic model to detect irony in tweets. In *53rd annual meeting of the association for computational linguistics (acl 2015)* (pp. PP–644).
- Karoui, J., Zitoune, F. B., & Moriceau, V. (2017). Soukhria: Towards an irony detection system for arabic in social media. *Procedia Computer Science*, *117*, 161–168.
- Kazdin, A. (1992). *Methodological issues & strategies in clinical research*. JSTOR.
- Kelleher, J. D., & Tierney, B. (2018). *Data science*. MIT Press.
- Keya, A. J., Afridi, S., Maria, A. S., Pinki, S. S., Ghosh, J., & Mridha, M. F. (2021). Fake news detection based on deep learning. In *2021 international conference on science contemporary technologies (icsct)* (p. 1-6). doi: 10.1109/ICSCT53883.2021.9642565

## Appendix E. Publications

- Khanam, Z., Alwasel, B., Sirafi, H., & Rashid, M. (2021). Fake news detection using machine learning approaches. In *Iop conference series: Materials science and engineering* (Vol. 1099, p. 012040).
- Khouja, J. (2020). Stance prediction and claim verification: An arabic perspective. *arXiv preprint arXiv:2005.10410*.
- Kilgo, D. K., Harlow, S., García-Perdomo, V., & Salaverría, R. (2018). A new sensation? an international exploration of sensationalism and social media recommendations in online news publications. *Journalism*, 19(11), 1497–1516.
- Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1), 3–24.
- Krishnan, S., & Chen, M. (2018). Identifying tweets with fake news. In *2018 IEEE international conference on information reuse and integration (iri)* (pp. 460–464).
- Kulkarni, A., & Shivananda, A. (2019). *Natural language processing recipes: Unlocking text data with machine learning and deep learning using python*. doi: 10.1007/978-1-4842-4267-4
- Lagutina, K., Lagutina, N., Boychuk, E., Vorontsova, I., Shliakhtina, E., Belyaeva, O., ... Demidov, P. (2019). A survey on stylometric text features. In *2019 25th conference of open innovations association (fruct)* (pp. 184–195).
- Layton, R., Watters, P., & Ureche, O. (2013). Identifying faked hotel reviews using authorship analysis. In *2013 fourth cybercrime and trustworthy computing workshop* (pp. 1–6).
- Layton, R., Watters, P. A., & Dazeley, R. (2015). Authorship analysis of aliases: Does topic influence accuracy? *Natural Language Engineering*, 21(4), 497–518.



## Appendix E. Publications

- Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... Rothschild, D. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096.
- Learning, M. (2021). <https://www.machinelearningplus.com/>, journal=Machine Learning Plus.
- Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, *27*(4), 363–384.
- Li, J., Ott, M., Cardie, C., & Hovy, E. (2014). Towards a general rule for identifying deceptive opinion spam. In *Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 1566–1576).
- Lim, C. (2018). Checking how fact-checkers check. *Research & Politics*, *5*(3), 2053168018786848.
- Liu, Y., Yu, K., Wu, X., Qing, L., & Peng, Y. (2019). Analysis and detection of health-related misinformation on chinese social media. *IEEE Access*, *7*, 154480–154489.
- Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, *54*(4), 1187–1230.
- Loukil, N., Haddar, K., & Hamadou, A. B. (2010). A syntactic lexicon for arabic verbs. In *Proceedings of the seventh international conference on language resources and evaluation (lrec'10)*.
- Manzoor, S. I., & Singla, J. (2019). Fake news detection using machine learning approaches: A systematic review. In *2019 3rd international conference on trends in electronics and informatics (icoei)* (pp. 230–234).
- Markowitz, D. M., & Hancock, J. T. (2014). Linguistic traces of a scientific fraud: The case of diderik stapel. *PloS one*, *9*(8), e105937.
- Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes

## Appendix E. Publications

- belief in fake news. *Cognitive research: principles and implications*, 5(1), 1–20.
- Massey, K. (2008). *Intermediate arabic for dummies*. John Wiley & Sons.
- McGurk, Z., Nowak, A., & Hall, J. C. (2020). Stock returns and investor sentiment: textual analysis and social media. *Journal of Economics and Finance*, 44(3), 458–485.
- McKee, A. (2003). *Textual analysis: A beginner's guide*. Sage.
- Mei, J., & Frank, R. (2015). Sentiment crawling: Extremist content collection through a sentiment analysis guided web-crawler. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 1024–1027).
- Metts, S. (1989). An exploratory investigation of deception in close relationships. *Journal of Social and Personal Relationships*, 6(2), 159–179.
- Mihalcea, R., & Strapparava, C. (2009). The lie detector: Explorations in the automatic recognition of deceptive language. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers* (pp. 309–312).
- Mishra, S. B., & Alok, S. (2017). Handbook of research methodology. *Dimensions Of Critical Care Nursing*, 9(1), 60.
- Mohammad, S., Salameh, M., & Kiritchenko, S. (2016, May). Sentiment lexicons for Arabic social media. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 33–37).
- Mohdeb, D., Laifa, M., & Naidja, M. (2021). An arabic corpus for covid-19 related fake news. In *2021 International Conference on Recent Advances in Mathematics and Informatics (ICRAMI)* (pp. 1–5).
- Molina, M. D., Sundar, S. S., Le, T., & Lee, D. (2021). “fake news” is not simply false information: A concept explication and taxonomy of online content. *American Behavioral Scientist*, 65(2), 180–212.
- Mourão, R. R., & Robertson, C. T. (2019). Fake news as discursive integra-

## Appendix E. Publications

- tion: An analysis of sites that publish false, misleading, hyperpartisan and sensational information. *Journalism Studies*, 20(14), 2077–2095.
- Mouty, R., & Gazdar, A. (2018). Survey on steps of truth detection on arabic tweets. In *2018 21st saudi computer society national computer conference (ncc)* (pp. 1–6).
- Mubarak, H. (2017). Build fast and accurate lemmatization for arabic. *arXiv preprint arXiv:1710.06700*.
- Mubarak, H., & Hassan, S. (2020). Arcorona: Analyzing arabic tweets in the early days of coronavirus (covid-19) pandemic. *arXiv preprint arXiv:2012.01462*.
- Nagoudi, E. M. B., Elmadany, A., Abdul-Mageed, M., Alhindi, T., & Cavusoglu, H. (2020). Machine generation and detection of arabic manipulated and fake news. *arXiv preprint arXiv:2011.03092*.
- Nahar, K. M., Al Eroud, A., Barahoush, M., & Al-Akhras, A. (2019). Sap: standard arabic profiling toolset for textual analysis. *International Journal of Machine Learning and Computing*, 9(2), 222–229.
- Namly, D., Bouzoubaa, K., Tahir, Y., & Khamar, H. (2015). Development of arabic particles lexicon using the lmf framework. In *Colloque pour les etudiants chercheurs en traitement automatique du langage naturel et ses applications (cec-tal 2015), sousse tunisia*.
- Neil, J. (2007). *Qualitative versus quantitative research: Key points in a classic debate*.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin*, 29(5), 665–675.
- Obeid, O., Zalmout, N., Khalifa, S., Taji, D., Oudah, M., Alhafni, B., . . . Habash, N. (2020). Camel tools: An open source python toolkit for arabic natural language processing. In *Proceedings of the 12th language resources and evaluation conference* (pp. 7022–7032).

## Appendix E. Publications

- Orange. (2022). *Word cloud*. Retrieved from <https://orangedatamining.com/widget-catalog/text-mining/wordcloud/>
- Osisanwo, F., Akinsola, J., Awodele, O., Hinmikaiye, J., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3), 128–138.
- Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011, June). Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (pp. 309–319).
- Pandey, S., & Pandey, S. K. (2019). Applying natural language processing capabilities in computerized textual analysis to measure organizational culture. *Organizational Research Methods*, 22(3), 765–797.
- Pasha, A., Al-Badrashiny, M., Diab, M., El Kholly, A., Eskander, R., Habash, N., ... Roth, R. (2014). Madamira: A fast, comprehensive tool for morphological analysis and disambiguation of arabic. In *Proceedings of the ninth international conference on language resources and evaluation (lrec'14)* (pp. 1094–1101).
- Patel, M., Padiya, J., & Singh, M. (2022). Fake news detection using machine learning and natural language processing. In *Combating fake news with computational intelligence techniques* (pp. 127–148). Springer.
- Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, 77(6), 1296.
- Penuela, F. J. F.-B. (2019). Deception detection in arabic tweets and news. In *Fire (working notes)* (pp. 122–126).
- Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2017). Automatic detection of fake news. *arXiv preprint arXiv:1708.07104*.

## Appendix E. Publications

- Pierri, F., Artoni, A., & Ceri, S. (2020). Investigating italian disinformation spreading on twitter in the context of 2019 european elections. *PloS one*, *15*(1), e0227821.
- Posadas-Durán, J.-P., Gómez-Adorno, H., Sidorov, G., & Escobar, J. J. M. (2019). Detection of fake news in a new corpus for the spanish language. *Journal of Intelligent & Fuzzy Systems*, *36*(5), 4869–4876.
- Pramanik, P. K. D., Pal, S., Mukhopadhyay, M., & Singh, S. P. (2021). Big data classification: techniques and tools. In A. Khanna, D. Gupta, & N. Dey (Eds.), *Applications of big data in healthcare* (p. 1-43). Academic Press. doi: <https://doi.org/10.1016/B978-0-12-820203-6.00002-3>
- Preston, S., Anderson, A., Robertson, D. J., Shephard, M. P., & Huhe, N. (2021). Detecting fake news on facebook: The role of emotional intelligence. *PloS one*, *16*(3), e0246757.
- Quijano-Sánchez, L., Liberatore, F., Camacho-Collados, J., & Camacho-Collados, M. (2018). Applying automatic text-based detection of deceptive language to police reports: Extracting behavioral patterns from a multi-step classification model to understand how we lie to the police. *Knowledge-Based Systems*, *149*, 155–168.
- Radcliffe, D., & Abuhmaid, H. (2020). Social media in the middle east: 2019 in review. *Available at SSRN 3517916*.
- Rangel, F., Rosso, P., Charfi, A., & Zaghouni, W. (2019). Detecting deceptive tweets in arabic for cyber-security. In *2019 ieee international conference on intelligence and security informatics (isi)* (pp. 86–91).
- Rao, V., & Sachdev, J. (2017). A machine learning approach to classify news articles based on location. In *2017 international conference on intelligent sustainable systems (iciss)* (pp. 863–867).
- Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., & Choi, Y. (2017). Truth of varying shades: Analyzing language in fake news and political fact-checking. In

## Appendix E. Publications

- Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 2931–2937).
- Rayson, P., Wilson, A., & Leech, G. (2002). Grammatical word class variation within the british national corpus sampler. In *New frontiers of corpus research* (pp. 295–306). Brill.
- Refaeilzadeh, P., Thang, L., & Liu, H. (2008). *Cross validation, arizona state university, 2008*.
- Reis, J. C., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. (2019). Supervised learning for fake news detection. *IEEE Intelligent Systems*, 34(2), 76–81.
- Resende, G., Melo, P., CS Reis, J., Vasconcelos, M., Almeida, J. M., & Benevenuto, F. (2019). Analyzing textual (mis) information shared in whatsapp groups. In *Proceedings of the 10th acm conference on web science* (pp. 225–234).
- Rouse, M. (2019). *Aws analytics tools help make sense of big data*.
- Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and social psychology review*, 5(4), 296–320.
- Rubin, V. L., Chen, Y., & Conroy, N. K. (2015). Deception detection for news: three types of fakes. *Proceedings of the Association for Information Science and Technology*, 52(1), 1–4.
- Ryding, K. C. (2005). *A reference grammar of modern standard arabic*. Cambridge university press.
- Saad, M. K., & Ashour, W. M. (2010). Arabic morphological tools for text mining. In *Corpora, 6th archeng international symposiums, eeecs'10 the 6th international symposium on electrical and electronics engineering and computer science* (Vol. 18).
- Saadany, H., Mohamed, E., & Orasan, C. (2020). Fake or real? a study of arabic satirical fake news. *arXiv preprint arXiv:2011.00452*.

## Appendix E. Publications

- Sabbeh, S. F., & Baatwah, S. Y. (2018). Arabic news credibility on twitter: An enhanced model using hybrid features. *journal of theoretical & applied information technology*, 96(8).
- Saeed, R. M., Rady, S., & Gharib, T. F. (2022). An ensemble approach for spam detection in arabic opinion texts. *Journal of King Saud University-Computer and Information Sciences*, 34(1), 1407–1416.
- Salgado, S., & Bobba, G. (2019). News on events and social media: A comparative analysis of facebook users' reactions. *Journalism studies*, 20(15), 2258–2276.
- Sanasam, R., Murthy, H., & Gonsalves, T. (2010, 21 Jun). Feature selection for text classification based on gini coefficient of inequality. In *Proceedings of the fourth international workshop on feature selection in data mining* (Vol. 10, pp. 76–85). Hyderabad, India: PMLR.
- Sayed, H. A., Sugave, S. R., Paygude, S., & Jazdale, B. (2021). Study and analysis of emotion classification on textual data. In *2021 6th international conference on communication and electronics systems (icces)* (pp. 1128–1132).
- Seih, Y.-T., Beier, S., & Pennebaker, J. W. (2017). Development and examination of the linguistic category model in a computerized text analysis method. *Journal of Language and Social Psychology*, 36(3), 343–355.
- Shaji, A., Binu, S., Nair, A. M., & George, J. (2021). Fraud detection in credit card transaction using ann and svm. In *International conference on ubiquitous communications and network computing* (pp. 187–197).
- Shamsan, M. A.-H. A., & Attayib, A.-m. (2015). Inflectional morphology in arabic and english: A contrastive study. *International Journal of English Linguistics*, 5(2), 139.
- Shao, C., Ciampaglia, G. L., Flammini, A., & Menczer, F. (2016). Hoaxy: A platform for tracking online misinformation. In *Proceedings of the 25th*

## Appendix E. Publications

- international conference companion on world wide web* (pp. 745–750).
- Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(3), 1–42.
- Siagian, A. H. A. M., & Aritsugi, M. (2020). Robustness of word and character n-gram combinations in detecting deceptive and truthful opinions. *Journal of Data and Information Quality (JDIQ)*, 12(1), 1–24.
- Soroka, S., & McAdams, S. (2015). News, politics, and negativity. *Political communication*, 32(1), 1–22.
- Strapparava, C., & Mihalcea, R. (2008). Learning to identify emotions in text. In *Proceedings of the 2008 acm symposium on applied computing* (pp. 1556–1560).
- Taloba, A. I., Eisa, D., & Ismail, S. S. (2018). A comparative study on using principle component analysis with different text classifiers. *arXiv preprint arXiv:1807.03283*.
- Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining “fake news” a typology of scholarly definitions. *Digital journalism*, 6(2), 137–153.
- Tenenboim, O., & Cohen, A. A. (2015). What prompts users to click and comment: A longitudinal study of online news. *Journalism*, 16(2), 198–217.
- Tian, L., Zhang, X., & Peng, M. (2020). Fakefinder: twitter fake news detection on mobile. In *Companion proceedings of the web conference 2020* (pp. 79–80).
- Torabi Asr, F., & Taboada, M. (2019). Big data and quality data for fake news and misinformation detection. *Big Data & Society*, 6(1), 2053951719843310.
- Tubey, R., Rotich, J. K., & Bengat, J. K. (2015). Research paradigms: Theory and practice. *Research on humanities and social sciences*, 5, 224–228.
- Tyagi, S., Pai, A., Pegado, J., & Kamath, A. (2019). A proposed model for pre-



## Appendix E. Publications

- venting the spread of misinformation on online social media using machine learning. In *2019 amity international conference on artificial intelligence (aicai)* (pp. 678–683).
- Ullah, R., Amblee, N., Kim, W., & Lee, H. (2016). From valence to emotions: Exploring the distribution of emotions in online product reviews. *Decision Support Systems*, *81*, 41–53.
- Vartapetian, A., & Gillam, L. (2012). 'i don't know where he's not': Does deception research yet offer a basis for deception detectives? In *Proceedings of the workshop on computational approaches to deception detection* (pp. 5–14).
- Vijayan, V. K., Bindu, K., & Parameswaran, L. (2017). A comprehensive study of text classification algorithms. In *2017 international conference on advances in computing, communications and informatics (icacci)* (pp. 1109–1113).
- Volkova, S., Shaffer, K., Jang, J. Y., & Hodas, N. (2017). Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 2: Short papers)* (pp. 647–653).
- Vosoughi, S., Mohsenvand, M. & Roy, D. (2017). Rumor gauge: Predicting the veracity of rumors on twitter. *ACM transactions on knowledge discovery from data (TKDD)*, *11*(4), 1–36.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151.
- Vrij, A., Kneller, W., & Mann, S. (2000). The effect of informing liars about criteria-based content analysis on their ability to deceive cbca-raters. *Legal and criminological psychology*, *5*(1), 57–70.
- Wahbeh, A., Al-Kabi, M., Al-Radaideh, Q., Al-Shawakfa, E., & Alsmadi, I. (2011). The effect of stemming on arabic text classification: an empirical study. *International Journal of Information Retrieval Research (IJIRR)*,

## Appendix E. Publications

1(3), 54–70.

- Walker, D. E. (1969). Computational linguistic techniques in an on-line system for textual analysis. In *International conference on computational linguistics coling 1969: Preprint no. 63*.
- Wang, L., Wang, Y., De Melo, G., & Weikum, G. (2018). Five shades of untruth: Finer-grained classification of fake news. In *2018 ieee/acm international conference on advances in social networks analysis and mining (asonam)* (pp. 593–594).
- Wang, W. Y. (2017, July). “liar, liar pants on fire”: A new benchmark dataset for fake news detection. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 2: Short papers)* (pp. 422–426). Vancouver, Canada: Association for Computational Linguistics.
- Watters, P. A. (2009). Why do users trust the wrong messages? a behavioural model of phishing. In *2009 ecrime researchers summit* (pp. 1–7).
- Wei, J., Li, Y., Zhou, T., & Gong, Z. (2016). Studies on metadiscourse since the 3rd millennium. *Journal of Education and Practice*, 7(9), 194–204.
- Weir, G., Owoeye, K., Oberacker, A., & Alshahrani, H. (2018). Cloud-based textual analysis as a basis for document classification. In *2018 international conference on high performance computing & simulation (hpcs)* (pp. 672–676).
- Weir, G. R. (2009). Corpus profiling with the posit tools. In *Proceedings of the 5th corpus linguistics conference. university of liverpool*.
- Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing* (pp. 347–354).
- Wu, L., Morstatter, F., Carley, K. M., & Liu, H. (2019). Misinformation in social media: definition, manipulation, and detection. *ACM SIGKDD Ex-*

## Appendix E. Publications

- plorations Newsletter*, 21(2), 80–90.
- Xie, J., Su, B., Li, C., Lin, K., Li, H., Hu, Y., & Kong, G. (2017). A review of modeling methods for predicting in-hospital mortality of patients in intensive care unit. *J Emerg Crit Care Med*, 1(8), 1–10.
- Yanagi, Y., Orihara, R., Sei, Y., Tahara, Y., & Ohsuga, A. (2020). Fake news detection with generated comments for news articles. In *2020 IEEE 24th International Conference on Intelligent Engineering Systems (INES)* (pp. 85–90).
- Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., & Yu, P. S. (2018). Ti-cnn: Convolutional neural networks for fake news detection. *arXiv preprint arXiv:1806.00749*.
- Yaseen, Q., & Hmeidi, I. (2014). Extracting the roots of arabic words without removing affixes. *Journal of Information Science*, 40(3), 376–385.
- Zhai, Y., Song, W., Liu, X., Liu, L., & Zhao, X. (2018). A chi-square statistics based feature selection method in text classification. In *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)* (pp. 160–163).
- Zhang, A. X., Ranganathan, A., Metz, S. E., Appling, S., Sehat, C. M., Gilmore, N., ... Robbins, M. (2018). A structured response to misinformation: Defining and annotating credibility indicators in news articles. In *Companion proceedings of the the web conference 2018* (pp. 603–612).
- Zhou, L., Burgoon, J. K., Zhang, D., & Nunamaker, J. F. (2004). Language dominance in interpersonal deception in computer-mediated communication. *Computers in Human Behavior*, 20(3), 381–402.
- Zhou, L., Twitchell, D. P., Qin, T., Burgoon, J. K., & Nunamaker, J. F. (2003). An exploratory study into deception detection in text-based computer-mediated communication. In *36th annual hawaii international conference on system sciences, 2003. proceedings of the* (pp. 10–pp).

## Appendix E. Publications

- Zhou, X., & Zafarani, R. (2018). Fake news: A survey of research, detection methods, and opportunities. *arXiv preprint arXiv:1812.00315*, 2.
- Zincir-Heywood, N., Mellia, M., & Diao, Y. (2021). Overview of artificial intelligence and machine learning. *Communication Networks and Service Management in the Era of Artificial Intelligence and Machine Learning*, 19–32.
- Zloteanu, M., Bull, P., Krumhuber, E. G., & Richardson, D. C. (2021). Veracity judgement, not accuracy: Reconsidering the role of facial expressions, empathy, and emotion recognition training on deception detection. *Quarterly Journal of Experimental Psychology*, 74(5), 910–927.
- Žukauskas, P., Vveinhardt, J., & Andriukaitienė, R. (2018). Philosophy and paradigm of scientific research. *Management culture and corporate social responsibility*, 121.
- Zulkarnine, A. T., Frank, R., Monk, B., Mitchell, J., & Davies, G. (2016). Surfacing collaborated networks in dark web to find illicit and criminal content. In *2016 IEEE Conference on Intelligence and Security Informatics (ISI)* (pp. 109–114).