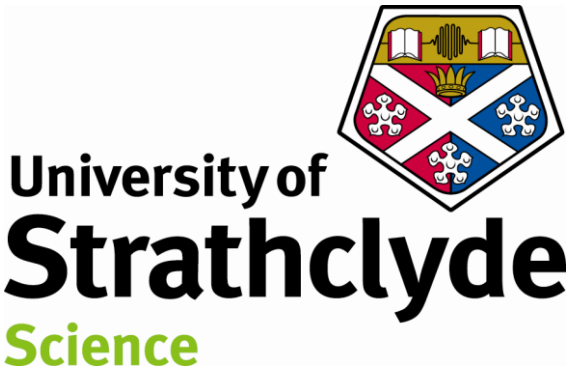
Pragmatic Cost Estimation for Web Applications

Submitted to Department of
Computer and Information Sciences.
University of Strathclyde,
Glasgow.
For the degree of Doctor of Philosophy.

By Sukumar Letchmunan
November 2012



Abstract

Cost estimation for web applications is an interesting and difficult challenge for researchers and industrial practitioners. It is a particularly valuable area of ongoing commercial research. Attaining on accurate cost estimation for web applications is an essential element in being able to provide competitive bids and remaining successful in the market. The development of prediction techniques over thirty years ago has contributed to several different strategies. Unfortunately there is no collective evidence to give substantial advice or guidance for industrial practitioners. Therefore to address this problem, this thesis shows the way by investigating the characteristics of the dataset by combining the literature review and industrial survey findings.

The results of the systematic literature review, industrial survey and an initial investigation, have led to an understanding that dataset characteristics may influence the cost estimation prediction techniques. From this, an investigation was carried out on dataset characteristics. However, in the attempt to structure the characteristics of dataset it was found not to be practical or easy to get a defined structure of dataset characteristics to use as a basis for prediction model selection.

Therefore the thesis develops a pragmatic cost estimation strategy based on collected advice and general sound practice in cost estimation. The strategy is composed of the following five steps: test whether the predictions are better than the means of the dataset; test the predictions using accuracy measures such as MMRE, Pred and MAE knowing their strengths and weaknesses; investigate the prediction models formed to see if they are sensible and reasonable model; perform significance testing on the predictions; and get the effect size to establish preference relations of prediction models. The results from this pragmatic cost estimation strategy give not only advice on several techniques to choose from, but also give reliable results. Practitioners can be more confident about the estimation that is given by following this pragmatic cost estimation strategy.

It can be concluded that the practitioners should focus on the best strategy to apply in cost estimation rather than focusing on the best techniques. Therefore, this pragmatic cost estimation strategy could help researchers and practitioners to get reliable results. The

improvement and replication of this strategy over time will produce much more useful and trusted results.

Acknowledgements

I would like to thank Dr. Marc Roper and Dr. Murray Wood for their great support during the development of this thesis. Without their support and supervision this thesis would never have reached completion. Their constructive comments have helped me a lot since my first year. Thanks a lot.

I must acknowledge the encouragements and support given to me over many years by my parents and sisters. Without their patience and sacrifice during my absence, I would have failed long ago. I would also like to say thanks to Dr Arivalan Ramaiyah, Dr. Teoh Ai Ping and Dr. Sulosanah Sarathi for their support and guidance to push me over the finishing line. To my special brother in Glasgow, Martin Campbell thanks a lot for his advice, company and his support especially during my final year of the thesis. Special thanks to Sorada Pitilertpanya as my best friend all the time during this process.

The research contained within this thesis has been supported by Science University of Malaysia (USM) and the Ministry of Higher Education Malaysia (MOHE). I would like to add special thanks to the staff at the Department of Computer and Information Sciences, University of Strathclyde. I am very grateful for the wonderful and supportive environment which they have provided to me.

Contents

1 Introduction……..1

1.1 Overview 1

1.2 Motivation for Web Application Cost Estimation 2

1.3 Contribution of the Thesis 2

1.4 Research Goals 3

1.5 Research Methodology Outline 4

1.6 Thesis Outline 4

2 Web Application Cost Estimation- State-Of-Art…….6

2.1 Introduction 6

2.2 Challenges to Cost Estimation Techniques 6

2.2.1 Algorithmic Models 9

2.2.2 Expert Judgements 13

2.2.3 Machine Learning 15

2.3 Overview of Some of the Cost Estimation Techniques 18

2.4 Systematic Literature review on Web Applications Cost Estimation 21

2.4.1 Introduction 21

2.4.2 Research Questions 23

2.4.3 Search Strategy 24

2.4.4 Study Selection 25

2.4.5 Threats of Validity 26

2.4.6 Results and Discussion 27

2.4.6.1 Types of Web Application Prediction Techniques 27

2.4.6.2 Superior Web Application Prediction Techniques 28

2.4.6.3 Size Measures for Web Applications 31

2.4.6.4 Prediction Accuracy for Web Applications 32

2.4.6.5 Types of Dataset for Web Applications 33

2.4.6.6 Characteristics of Dataset for Web Applications 33

2.4.7 Recommendations for Practice 35

2.5 Conclusions 36

3 Current Industry Practice on Cost Estimation……39

3.1 Introduction 39

3.2 Methodology 39

3.3 Findings 40

3.4 Conclusions 47

4 Investigating Effort Prediction of Web based Applications using CBR……49

4.1 Introduction 49

4.2 Challenges in CBR 49

4.3 The Dataset 51

4.4 Methodology 52

4.5 Evaluation Criteria 54

4.6 Results and Analysis 54

4.7 Conclusions 71

5 A Pragmatic Cost Estimation Strategy……..73

5.1 Introduction 73

5.2 The Challenges of Cost Estimation 73

5.3 Application Methodology 74

5.3.1 Dataset 75

5.3.2 Estimation Techniques 75

5.3.3 Dataset Characteristics 76

5.3.4 Difficulties in Characterising the Dataset 78

5.4 A Pragmatic Cost Estimation Strategy 84

5.5 Results 91

5.6 Application of Strategy 93

5.7 Summary of the Application of Strategy 114

5.8 Conclusions 121

6 Conclusions and Future Work……..123

6.1 Summary of the Thesis 123

6.2 Thesis Contribution 125

6.3 Lessons Learned 127

6.4 Research Limitations 128

6.5 Future Work 129

6.6 Conclusions 130

References…….131

Appendices

Appendix A……..141

Appendix B……..142

Appendix C……..146

Appendix D……..156

Appendix E……..214

1 Introduction

1.1 Overview

This thesis shows the way in which the choice of prediction techniques can have a serious impact on cost estimation. Although the need for better estimates is clear, there exist a very large number of effort estimation methods and there are few studies which empirically compare all these techniques. To address this problem, this thesis seeks to explore the development of a pragmatic cost estimation strategy to support practitioners rather than argue which techniques are the best.

In order to gather evidence on the relationships of dataset characteristics on web application cost estimation, a systematic literature review has been carried out. The initial aim was to establish which techniques performed best.

The lack of industrial input on web application cost estimation is disturbing. Therefore, this thesis continues by reporting on a survey on industry needs and practice. This process also tries to engage industrial participation in terms of providing the dataset for our future research.

However, this effort failed mainly due to confidentiality concerns. Hence, the International Software Benchmarking Group (ISBSG) Release 10 dataset (ISBSG 2009) was chosen as a vehicle to explore various estimation strategies. Initial investigations using this dataset have been carried out using Case Based Reasoning (CBR) which has proved to be a reasonably effective estimation strategy, although it has not been widely explored in the context of web applications. The main aim of this study was to explore the number of analogies required to provide the best estimate. The results of this investigation were inconclusive but suggest that the effectiveness of CBR is hampered by others factors, including dataset characteristics.

When the systematic literature review and intial experiments results appeared to be inconclusive it was decided to explore the dataset characteristics in order to try and explain the results. However, too many dimensions of dataset characteristics make it hard to structure the dataset characteristics. Therefore it is impractical to study the relationships of the dataset and the techniques.

Following on from this the thesis proposes a pragmatic cost estimation strategy. The steps may be summarised as: 1) Test against means, 2) Test using MMRE, Pred and MAE, 3) Investigate the models, 4) Perform significance testing and 5) Test effect size. The strategy is to follows the series of steps during the analysis while it is sensible thing to do or otherwise stop from continuing and give advice to practitioners about what should they do.

1.2 Motivation for Web Application Cost Estimation

Web application cost estimation provides several benefits to developers: resource estimation, project planning and an estimate of a competitive price to remain competitive in the market. Yet despite several approaches that have been proposed in the literature to address this issue, the problem of finding the right techniques to predict the cost of web applications remain unsolved. When comparing different prediction techniques, the researchers keep on suggesting new techniques to solve the problem. Finding appropriate techniques to estimate web applications, having confidence in their predictions and assessing their performance are all significant issues hampering the applicability of web application cost estimation. Therefore, it is an interesting area to explore further with the aim of benefitting industry practitioners and researchers.

1.3 Contribution of the Thesis

The work presented in this thesis makes the following contributions to the area of web application cost estimation:

An investigation of the literature to gather the evidence on web application cost estimation carried out by a systematic literature review. This can provide guidance to practitioners on web application cost estimation techniques based on the empirical evidence that is collected, and for researchers to identify any areas requiring further study.

An investigation of the issues that confound web application cost estimation prediction techniques in industry.

An investigation on dataset characteristics concluded that it was not practical to accurately analyse the characteristics of the dataset.

Shepperd and MacDonell's validation framework (Shepperd and MacDonell, 2012) and other pieces of advice are applied to create a pragmatic cost estimation strategy which can help practitioners to identify not only a technique, but a "basket of techniques" which they can compare. The strategy also gives advice as to what should be done with the data if it fails to give quality results.

***Key findings*:**

Based on the systematic literature review, it was found that there was a lack of consistency in the findings, there are challenges in establishing the influence of data characteristics, and the role of accuracy measures (PRED, MMRE and MAE) varies between studies.

Based on CBR study it was found that no reliable guidance could be given regarding the number of analogies that should be employed in making a prediction. In addition, the results also do not give any confidence that increasing the size of the dataset results in more accurate predictions. It was also found that outliers could possibly effect the predictions.

The pragmatic cost estimation strategy, does not force practitioners to adopt a particular strategy up front, but allows them to explore the results generated from whatever prediction techniques they care to use, helps them to decide whether or not to use these techniques, and supports them in determining the quality of the results that have been obtained.

1.4 Research Goals

The research goals of this thesis are identified below:

- To determine the state of the art of web based cost estimation by collecting the available evidence from the literature in the form of a systematic literature review.

- To establish current industry practice in terms of cost estimation.

- To provide or create a strategy based on existing advice in the literature which can help practitioners to identify which estimation techniques should be applied in which circumstances.

1.5 Research Methodology Outline

The thesis reviews and synthesizes the available published evidence by carrying out a systematic literature review. It does this by following the high level steps that introduced by Kitchenham (2004).

The thesis investigates current cost estimation practices by carrying out a survey of companies involved in web applications development around Glasgow and Edinburgh.

The thesis investigates the optimal number of analogies to employ when making an estimate using case-based reasoning (CBR) - one of the more popular techniques identified from the literature review. The thesis replicates the approach adopted in previous studies (e.g Kadoda et al. 2000) for this investigation and uses a publicly available web-application dataset from which various different-sized subsets are created in order to take into account the potential impact of larger or smaller pools of data.

The thesis develops a pragmatic cost estimation strategy based on Shepperd and MacDonell's validation framework (Shepperd and MacDonell, 2012) and other pieces of advice such as looking at the value of the accuracy measures employed and investigating the models that produced by the estimators. The thesis demonstrates the approach with examples drawn from the publicly available Desharnais dataset.

1.6 Thesis Outline

The remainder of the thesis is structured in the following way:

Chapter 2: Web Application Cost Estimation – State of Art

The thesis begins with a traditional literature review and then follows with a systematic literature review. The literature review gathers information on web application cost estimation. A systematic literature review is presented in this chapter, discussing the motivation behind the review, the empirical evidence on different cost estimation techniques, and motivating the inspection of the key characteristics of literature dataset. Evidence of existing relationships between dataset characteristics and the techniques involved is also collected in this study.

Chapter 3: Current Industry Practice on Cost Estimation

The survey collects input on the practice that is used in industry such as the method, the data characteristics, the cost estimation practice pattern, and the possibilities of future help in data contribution.

Chapter 4: Investigating Effort Prediction of Web Based Applications Using CBR

The thesis continues by investigating CBR on the ISBSG dataset. It presents the related work on web application cost estimation and a number of challenges to the application of CBR. The main aim of the study is to investigate the optimal number of analogies to employ when making an estimate.

Chapter 5: A Pragmatic Cost Estimation Strategy

A strategy has been demonstrated by using Shepperd and MacDonell's validation framework and other pieces of advice that are found in literature. It also argues that trying to base an estimate on data characteristics is very hard. Therefore, five-stage strategy has been introduced to support practitioners in giving the best estimates they can.

Chapter 6: Conclusions and Future Work

The final chapter of the thesis contains a summary of the work presented and discusses what lessons can be learned for cost estimation strategy, as well as areas of future work including the verification of the cost estimation strategy using different dataset. The conclusion of this thesis is that rather than trying to argue that one technique is best, it draws on results that demonstrate that there is not a universally good approach as cost estimation methods vary according to different contexts. A pragmatic cost estimation strategy is timely appropriate to help practitioners produce the best estimates.

2 Web Application Cost Estimation- State-Of-Art

2.1 Introduction

In recent years cost estimation tools for web applications and web hypermedia systems have seen an increased demand for investigation and further development. Researchers working in the field of cost estimation for web applications and hypermedia are facing greater challenges in order to produce a measurement tool that accurately deals with the developmental effort and scheduling of a designated project. Over the last decade, there has been limited research in this area as this category is considered new and some of the metrics and models proposed in this research area are still being enhanced along with the progression of web development technologies. At present, companies developing Web-based systems face the problems and challenges of estimating the required development effort in a fixed time frame. This problem does not have a standard solution yet (Dahwan et al. 2007). The early sections of this chapter will discuss the state-of-art of web application cost estimation based on current literature. However it's hard to objectively compare the available evidence or establish the research questions based on traditional review. Therefore the second part of this chapter covers a systematic literature review on web application cost estimation studies.

2.2 Challenges to Cost Estimation Techniques

Before exploring the research literature, it is advisable to distinguish between the challenges of estimation faced by traditional approaches and web-based approaches. These differences may enable users and future researchers to determine the metrics that can be adapted from traditional approaches. These metrics need to be produced specifically for the development of web-based applications. Some of the characteristics of traditional versus web development approach have been contrasted in Table 2.1, meanwhile Table 2.2 shows the particular challenges faced by web- based estimation (Reifer 2000).

Characteristics	Traditional Developments	Web Developments
Primary objective	Build quality software products at minimum cost	Build quality products to market as quickly as possible
Typical Project size	Medium to large (Hundreds of team members)	Small (3-5 team members)
Typical timeline	12-18 months	3-6 months
Development approach employed	Classical, requirements-based, phased and/or incremental delivery, use cases, documentation-driven	Rapid application development, gluing building blocks together, prototyping, Rational Unified Process, MBASE
Primary engineering technologies used	Object oriented methods, generators, modern programming languages (C++), CASE tools, etc.	Component-based methods, 4th and 5th generation languages(html, Java, etc.) visualization (motion, animation), etc.
Process employed	CMM-based	Ad hoc
Products developed	Code-based systems, mostly new, some reusable, many external interfaces, often complex	Object-based systems, many reusable components (shopping carts, etc.), few external interfaces, relatively simple
People involved	Professional software engineers with lots of experience	Graphic designers, less experienced software engineers
Estimating technologies used	SLOC or function point-based models, WBS approach for small projects	Wing it

Table 2.1: Characteristics of Traditional Versus Web Development Projects (from (Reifer 2000))

From Table 2.1 and 2.2, it can be concluded that the main challenges between traditional and web development estimation are the duration that is required to develop them. Web developments are in fast mode compared to traditional software. Therefore more challenges have to be overcome to produce quality products in a short period of time. Size measure, which has been used in traditional software, is no longer applicable for web development which requires template and web based objects compared to function points and lines of code.

The data in Table 2.1 indicates that a typical project involves -"hundreds of team members" for traditional developments. This could be an over-exaggeration as there are many traditional developments which involve much smaller teams. While very large teams undoubtedly exist, they are not necessarily typical. The data in Table 2.1 also suggests that all web development is simple. However this is not the always the case for more recent web developments which may often be very ambitious projects (sites such as Facebook being a prime example). More challenges in web developments which make it much more complex

for example large diversity of systems, language and technologies that may be combined to build a website.

	Traditional Approach	Web-based Challenges
Estimation process	Most use analogy supplemented by lessons gleaned from past experience	Job costing done ad-hoc based on inputs from developers (often too optimistic)
Size Estimation	Systems are built to requirements, SLOC or function points are used	Applications are built using templates and a variety of web-based objects (html, applets, component, building blocks). No agreement on the size measure for web applications has yet been reached within the community.
Effort Estimation	Effort is estimated via regression formulas modified by cost drivers (plot project data developed relationships between variables	Effort is estimated by breaking the job down into tasks and identifying what is needed to do the work. Little history is available.
Schedule Estimation	Schedule is estimated using a cube root relationship with effort	Schedule is based on analogy. Models typically estimate schedules high because cube root relationship doesn't hold.
Quality Estimation	Quality is measurable from internal metrics like defect rates and system properties	Quality is hard to measure. New metrics are needed to assess multimedia quality.
Model Calibration	Measurements from past projects are used to calibrate models to improve accuracy	Measurement from past projects are used to identify folklore (too few to be used yet)
"What if " Analysis	Estimating models are used to perform quantitative "what if" and risk analysis. They are used to compute ROI and cost/benefits	Most "what if" and risk analysis is qualitative because models do not exist. ROI and cost/benefits for e-commerce applications remains an open challenge

Table 2.2: Web-Based Estimating Challenges (from (Reifer 2000))

It is reported in Table 2.2 that "little history is available" for effort estimation in web development. However, this ignores the fact that web projects often have a short lifecycle which means that in the time since the study was published (2000), many organisations will have had the opportunity to amass a substantial amount of historical data.

In Table 2.2 under "Quality Estimation" it has been reported that quality is hard to measure. However, there is no reason why quality characteristics such as those identified in the standard ISO-9126 are not applicable in this context. Usability, Efficiency, Maintainability,

Functionality, Reliability and even Portability (in terms of use on different web-browsers) are all as relevant to web-based systems as to "traditional" systems.

A common mistake is the failure to differentiate between **web hypermedia** systems and **web applications**. Web applications are technically distinct from web hypermedia systems in terms of how they are implemented. Within this study, a web application is defined as a software application that uses web sites as a front end for broad and remote access. The back end provides full user functionality so that the user can affect the status of the business logic on the web server. Hypermedia systems are much easier to predict in terms of effort than web applications because web hypermedia systems are much smaller in terms of the expended development effort and are a non-conventional application characterised by authoring information using nodes (chunks of information), links (relations between nodes), anchors, assess structure (for navigation) and delivery of this structure over the web (Mendes et al. 2002d). Consequently, web applications are much more complex to develop and manage than web hypermedia systems. The focus or interest for this study and the literature is on the cost estimation of web applications.

Cost estimation is frequently referred to as effort estimation. However the two are not strictly synonymous as there other project costs to consider as well as effort such as licensing, travel, hosting, training, etc.. This said, the costs of the personnel employed on a project – the effort- is often the dominant cost, which is why the terms are often used interchangeably. Effort estimation can be obtained via parametric equations, previous experience or expert judgement. In general, these are grouped in three different techniques. There are algorithmic models, expert judgement and machine learning. The previous related work in each of these groups is presented in the remainder of this section.

2.2.1 Algorithmic Models

Algorithmic models predict estimates of effort using parametric equations. The models employed are typically derived from statistical data analysis. Most prediction systems currently use these techniques because they are relatively simple and easy to use. Examples are ordinary least-squares regression (OLS) (Briand et al. 2002), the Constructive Cost Model (COCOMO), and Classification and Regression Trees (CART) (Boehm 1981).

The most famous algorithmic approach model is the Constructive Cost Model (COCOMO), introduced by Barry Boehm in 1981. This considers three types of software projects: organic, semi-detached and embedded. Organic systems are essentially data processing systems, while embedded correspond to real time systems. Semi-detached systems combine elements of both. This model is also known as the Basic COCOMO model. The major problem with this model is that it is driven by a variant of lines of code (LOC). Since LOC are only available at the end of the project, we need to make a prediction at the start of a project for estimation, which can be just as hard as estimating the effort. This weakness is overcome by COCOMO II, as this model estimates project size from the specification (Boehm et al. 2000).

As mentioned earlier, COCOMO II was enhanced from the earlier version of COCOMO in which new cost drivers were introduced to provide better estimation accuracy. This model can be used in the area of software development, budget decisions, product trade-off, IT capital planning, management decisions etc. COCOMO II has 29 cost drivers to be considered in software cost estimation, compared to only 15 (cost drivers) in the earlier version and uses Function Points Analysis as its sizing measurement. The cost factors can be divided broadly into five main groups; namely (i) scale factors; (ii) product factors; (iii) project factors; (iv) platform factors; and (v) personnel factors. Each and every factor described has its own weighting value, which is calculated accumulatively, in order to produce a software cost estimate.

As an alternative to measuring project size without the need of LOC, Albrecht (1985) devised a method of estimating effort by measuring the functionality of a system as opposed to size, namely function points (FP) (Briand et al. 1999). FPs take into account the number of interfaces, files and queries in a specification which are then weighted according to their complexity; either simple, average or complex. The sum of these complexity weighted attributes is the Unadjusted FP count (UFC). This is then multiplied by a Technical Complexity Factor (TCF) which is composed of a variety of technical and project factors. The result is an Adjusted FP Count (AFC). The big advantage of FPs over LOC is that they are available before development commences. However despite this, Kichenham et al. (1995) criticized FPs. The major disadvantages mentioned by them were the difficulty in comparing the FP count at the start and the end of the project. The difficulty with this is not so much in automating the counting, but more in other aspects such as unit definitions, measurement

instruments and measurement protocol. All these definitions ensure that the model is correct and in a suitable form for FP counting.

Reifer introduced a sizing metric for web applications known as Web Objects, which is an extension of function points (Reifer 2000). Web Objects consist of all the elements in Function Points such as (i) internal logical files; (ii) external interface file; (iii) external input; (iv) external output; and (v) external inquiries, together with an additional four web related components namely (i) multimedia files; (ii) web building blocks; (iii) scripts; and (iv) links. Each Web Object components needs to be counted and categorised in terms of complexity as low, average or high. The results of comparisons between Function Points and Web Objects for sizing measures show that more accurate effort estimation is obtained from the model using Web Objects. In this paper Reifer strongly suggests that sizing metrics are one of the greatest concerns in estimating the cost of web applications, as the size of a web application becomes the fundamental cost driver. The new size metrics proposed by Reifer are said to accurately estimate the work involved in web based application projects that cannot be accurately estimated using source lines of code (SLOC) or function points (FPs) alone. Web Objects enable the computation of size by considering the elements involved in web application development.

Reifer developed a worksheet known as the Web Object Calculation Worksheet in which he listed Web Object predictors according to their respective complexity weight such as low, average or high. The worksheet and the size metrics for size measurement became the first step in developing a model (WebMo) that accurately estimates the cost and optimal schedule for web development. Reifer developed the WebMo model using expert judgments and data from 46 projects using regression analysis. The WebMo model was also developed using nine cost factors and fixed power laws to estimate the effort accurately. On top of that, the duration was calculated based on a square-root relationship with effort based upon built-in scaling rules. An analysis of Web Objects by Reifer shows that these sizing metrics have many advantages in estimating the development cost for web applications compared to traditional source lines of code (SLOC) and function points. Reifer developed counting conventions and validated that web objects have better predictive accuracy than traditional function points by using counting conventions.

Ruhe, Jeffery and Wieczorek continued this research and focused on the use of Web Objects to estimate the development effort for web applications (Ruhe et al. 2003). In this paper, they investigated the applicability of Web Objects as size measurement metrics compared with traditional function points. Their results, based on web applications in the context of an industrial dataset, show that estimation derived using Web Objects significantly outperformed models using Function Points. This confirmed the earlier study, which indicated that Function Points (FPs) were inappropriate for productivity estimation of framework based web development, as they did not take into account the reuse of components (Morisio et al. 1999).

Rollo introduced a different sizing measurement known as Full Function Points (FFPs), but which has not been subjected to full empirical evaluation. FFP is a functional measure based on standard FP techniques (Rollo 2006). The FFP transactional functions types are identified at the sub process level, instead of the process level as is done with traditional FP. It can thus be said that FFP takes into account a finer level of granularity, (the sub process level), while FP only considers the process level. In his study he claims without any empirical results that FFP's are the most flexible method for counting the functional size of web applications.

A new web application cost estimation model was introduced by Mangia and Paiano, known as Metric Model for Web Application (MMWA) (Mangia and Paiano 2003). MMWA metrics represent a solution to the problems of estimating the development cost and size by taking into account all complexity factors in the development of a web based application. The advantage of this model, unlike the early version of COCOMO, is that it is independent as the model allows decisions concerning the development of web applications to be taken using a view without concerning the logic of the programmer. MMWA is sub-divided into four sub-models identified as (i) Functional Sizing Model; (ii) Navigational Structures Sizing Model; (iii) Publishing Sizing Model; and (iv) Multimedia Sizing Model. Each of these models is related to a particular factor of complexity in web applications. According to Mangia and Paiano, "… each module is categorized by (i) component, the set of information necessary for the correct implementation; (ii) tools, the coding and formalization diagrams of this information; and (iii) counting rules, the identification of cost indicators and techniques to produce final measurements".

The functional sizing model is used to identify all the main functions required by the application. The main concern in this model is the measurement which is based on the size of data with regards to each function and a series of weights associated with it. Meanwhile, the navigational structure-sizing model is used to measure the development effort for the navigational or browsing structures. Each sub model provides a measure of complexity required for the development of each specific component such as the macro-functions required by the user, input data, output data and series of information units. The sub models will produce results by using a standard measurement unit known as Unadjusted Web Complexity Point (UWCP).

A tool has been developed to enable the estimator to codify, using diagrams, the information gathered, preparing it for the estimate of its relative complexity and expressing this as UWCP (Mangia and Piano 2003). Additional to these sub models, a calibration phase is also used, which takes into account factors concerning project environment and the type of applications. From extensive testing and analysis it has been shown that at the early stage MMWA produces results which are accurate in estimating the development effort of web-based applications. However, this sizing measure has not gained any popularity or continuity from other researchers in a web applications development context as W2000 is used as the design framework. This W2000 design framework uses a consolidated methodology or systematic approach to design web applications. By using this framework, it is hard to collect the data of previous projects and is therefore not relevant for web application development estimation.
The trend discussed in the literature is mainly focused on Web Objects and Function Point Analysis as sizing measurements. However, some recent research has been conducted such as case based reasoning (Mendes et al. 2002b), artificial neural networks (ANN) (Idris et al. 2008) and genetic algorithms (Burgess 2007). All of these fall in to machine learning approaches providing the basis for development effort estimation models in contrast to algorithmic models. However, in the next section expert judgement will be covered first before discussing machine learning approaches in detail.

2.2.2 Expert Judgement

Expert judgement involves making predictions based on the skill and experience of one or more experts. This method is not highly regarded amongst the research community as it is considered to be subject to bias and political pressure and also highly dependent on the

caliber and experience of the expert. Hammond (1987) claims that experience has been shown to be unrelated to the empirical accuracy of expert judgment. It has also been described as guessing (Kitchenham 1991).

Ruhe, Jeffery and Wieczorek (2003) proposed an approach to web based cost estimation by investigating the application of COBRA (Cost Estimation, Benchmarking, and Risk Assessment). COBRA consists of two main components identified as (i) the causal model, consisting of major factors influencing the cost of projects within local environments, also known as cost drivers; and (ii) data from previous projects. The second component is equally as important as the first one because it determines the relationship between cost overheads and costs such as qualitative and quantitative cost factors. This qualitative information is quantified using expert opinion. The quantification is the percentage of cost overhead above that of a nominal project. The development effort of COBRA can be simplified as (i) determining the data source in terms of functionality and categorization schema; (ii) distinguishing the type of web development - web hypermedia or web application; (iii) determining the cost factors by using either traditional cost factors or performing personal interviews with experts; (iv) developing the causal model; and (v) refining the qualification of the relationship with the causal model by performing personal interviews with experts to obtain the value for multipliers.

The accuracy of COBRA is validated by using three different techniques namely Web-CORBA, Ordinary Least Square Regression and Allete Systems Informal Method (Ruhe et al. 2003). The estimates are compared with actual effort by calculating the magnitude of relative error (MRE) and prediction level, Pred. COBRA is said to be one of the most accurate models for estimating the development cost of web applications from the testing conducted in the research paper. The latest version of COBRA, modified solely to estimate web development effort identifies, some serious issues that need to be considered when developing this model, especially those associated with size measurements. In the earlier version, CORBA used line of code (LOC) as its size measurement, which was not appropriate for web development estimation. This is because LOC are only available at the end of the project and it is quite difficult to predict the LOC of web applications at the start of development. Furthermore, web applications typically involve a mixture of languages and applications that can make the LOC cost estimation even harder. In conclusion, certain issues arise from the development of COBRA such as (i) accurate cost factor definition; (ii) early

size measurement for web applications; and (iii) additional characteristics to determine the cost estimation accurately.

2.2.3 Machine Learning

Machine learning is based on computational intelligence techniques such as artificial neural networks, genetic algorithms, and analogy approaches such as case based reasoning. They have been developed to avoid the shortcomings of the above mentioned techniques. The major focus of machine learning is to automatically learn to recognise complex patterns that exist and make intelligent decisions based on the data.

Genetic Algorithms are a type of evolutionary computation technique. This technique provides a general structure for solving problems, which mimic the biological paradigm of the "survival of the fittest" (Back et al. 1997). Research carried out by Dolado (2001) shows promising results for GA based estimation systems on a single input variable. In Dolado's research, both standard regression analysis and GA were applied and compared on several data sets. However, regardless of the method, the basic size-effort relationship did not show satisfactory results, from a predictive point of view, across all data sets. Burgess et al. (2007) extended this idea into richer models requiring larger populations and much longer learning lifetimes. They also investigated the potential for the use of genetic programming (GP) methods to build software cost prediction systems and compare preliminary results against other previously researched approaches. Despite the evidence that GP has the potential to be a valid additional tool for software effort estimation, they concluded that the set up and running effort was high and the interpretation was difficult.

A **neural network** (NN) is a computer system that simulates the learning process of the human brain. NN are massively parallel systems inspired by the architecture of biological neural networks, comprising simple interconnected units (artificial neurons). Neurons compute a weighted sum of their input and generate an output if the sum exceeds a certain threshold. This output then becomes an excitatory (positive) or inhibitory (negative) input to other neurons in the network. The process continues until one or more outputs is generated (Mair et al. 2000). NNs are used widely in many industrial areas, including software effort estimation. The applicability of NNs to software effort estimation has been extensively

studied (Mair et al. 2000 and Aggrawal 2005). Srinivasan and Fisher (1995) pointed out that the performance of neural network approaches was very sensitive to configuration choices, such as the number of hidden units, the stopping criteria, and the initial weight settings. The appropriate settings of these choices can only be determined empirically. Lee et al. (1998) extended this research with a new approach that integrated neural network methods with cluster analysis to improve both training efficiency and network performance. A failing of neural networks is that they operate as 'black boxes' and provide the user with no information about how output is reached (Gray et al. 1997). In fact, the ability to generate explanations is important in order to gain user acceptance of artificial intelligence techniques. Another problem with neural networks is catastrophic forgetting, where training on new data causes the network to lose existing knowledge. However, given the relatively small sizes of software metric data sets this is unlikely to be problematic.

Fuzzy Logic is firmly grounded in terms of its theoretical foundations and applications in the various fields in which it is being used, such as robotics, medicine and image processing. Fuzzy Logic systems have only been used in a few publications for software development models (Kumar et al. 1994). A fuzzy system is a mapping between linguistic terms, such as 'very small' attached to variables. Thus, an input into a fuzzy system can be either numerical or linguistic, with the same applying to the output. The most obvious strength of fuzzy systems is that by using linguistic mapping, a highly intuitive model can be created that anyone, even without any training, can understand and, if necessary, criticise. On the negative side fuzzy systems suffer from some limitations, including the difficulty of specifying a system with very high accuracy while maintaining a degree of meaningfulness, "Generally more accuracy requires more rules, with a greater numbers of rules leading to more complex and less interpretable systems" (Gray et al. 1997).

Analogy involves the comparison of one or more completed projects with the details of a new project to predict cost and duration. The main issue with this technique is the requirement for a data set with which to compare to the new project. The analogy approach that has been explored most for the problem of cost prediction is that of Case Based Reasoning, in part because the idea of formalising the process of predicting by analogy is attractive (Shepperd 2007). A study of nearly 600 organizations reported that analogy is the most widely used estimation method in the software industry (Heemstra 1992). This is most likely to be because users may be more willing to accept a solution from a form of reasoning

which is similar to human problem solving (Shepperd 1997). Analogy based estimation has also been extensively studied and applied, due to its conceptual simplicity and empirical competitiveness. In reality there is no single best software cost estimation model, but CBR is rated among the best methods in a variety of circumstances (Ruhe et al. 2003).

Web Objects proposed by Reifer (2000) have been further analyzed and tested by some other researchers using Analogy techniques. Mendes, Mosley and Counsell (2003a), conducted a survey to identify early size measures for web cost estimation and to compare the prediction accuracy of web company-specific data with the data from a multi-organizational database. In their survey, online quotes for Web development projects based on gathered size measures using Web forms were collected. The measures used were organized into five categories, namely (i) Web application static measures; (ii) Web application dynamic measures; (iii) Web project measures; (iv) Web company measures; and (v) Web interface design measures. The survey on the size identified three attributes : (i) length - physical size of web applications, (ii) functionality - functions by web application to user; and (iii) complexity - complexity of the application. From the survey, it was found that there were two dominant factors which were identified as the total number of Web pages and features/functionality, which influence the web cost estimation.

Case Based Reasoning (CBR) works by comparing the new project, for which an estimate is required, to those similar finished projects with known efforts. The known efforts are then used to produce the prediction of the effort for a new project based on attributes similar to the finished projects. Applying CBR takes into consideration several parameters such as feature subset selection, similarity measure, scaling, number of analogies, analogy adaptation and adaptation rules.

Mendes et al. (2003d) investigated the use of adaptation rules to improve web cost estimation. In this study the research employed two types of adaptation rules, adaptation without weights and adaptation using weights. According to their research, adaptation rules are used to reflect feature differences between the new problem and the retrieved cases. Adaptation rules are used to adapt the estimated effort, so that it will reflect the characteristics of the target project more closely. According to Walkerden and Jeffery (1999), once the most similar finished project in the case base has been retrieved, its effort value is adjusted to reflect the effort of the new project. The type of adaptation rules, methods to

derive adaptation rules, and formula derived from the study are explained in detail in the research paper. The types of adaptation rules employed are as follows: adaptation without weight, adaptation with different weights that indicate the strength of the relationship between size measure and effort, and weighted Euclidean distance. The comparison of adaptation techniques was made using the jack-knife method, also known as cross validation.

In Mendes et al's research, two types of data sets were used; web project data from the same company and web project data from many organizations. The purpose of using two different datasets was to compare the accuracy level produced by these two datasets based on the adaptation rules. The final results obtained from the research indicate results obtained without applying adaptation rules and demonstrate significant outcomes compared to their counterpart. On top of that, the results also demonstrated that for datasets based on the same company, the adaptation rules without weights gave the best predictions for less 'messy' datasets. Meanwhile, for datasets obtained from multiple organizations, better results were produced when there were no adaptation rules applied. Although there is no clear definition of messy dataset here, the author classified that the dataset from the same company was less messy compared to the dataset from a different company. The research also proposed additional elements to adaptation rules in order to predict the cost of web applications accurately, known as Feature Subset Selection (FSS). FSS involves determining the optimum subset of features that give the most accurate estimation (Mendes et al. 2002). Some CBR tools offer functionality to support FSS such as the Angel tool (Shepperd et al. 1997), while CBR-works (Schulz 1999) does not offer this functionality.

Gray and MacDonell (1997) compared least square regression, robust regression, neural networks, fuzzy systems, hybrid neuro-fuzzy systems, rule-based systems, case based reasoning, and classification and decision trees. They concluded that among the nine different predictive models of software metrics compared, the CBR approach was worth further study due to its encouraging results.

2.3 Overview Some of the Cost Estimation Techniques

In next section an overview of some of the techniques that will be used throughout this research.

Linear Regression

Linear regression or Stepwise regression (SWR) is a statistical technique whereby a prediction model (Equation) is built to represent the relationship between independent and dependent variable. At each stage this technique builds the model by adding the independent variable with highest association to the dependent variable, taking into account all variables currently in the model. It aims to find the set of independent variables that best explains the variation in the dependent variable. The goal of regression is to find the function *f*(x) that best models the data. In linear regression, this is done by finding the line that minimizes the sum squares error on the data.

Support Vector Regression

Support Vector Regression (SVR) is a regression technique based on Support Vector Machines (SVM), a very effective machine learning approach (Corazza et al 2011).

SVM is used for binary classification; it looks for the hyperplane which separates the elements of the two considered classes with the largest margin. In the parlance of SVM literature, a predictor variable is called an *attribute*, and a transformed attribute that is used to define the hyperplane is called a *feature*. The task of choosing the most suitable representation is known as *feature selection*. A set of features that describes one case (i.e., a row of predictor values) is called a *vector*. So the goal of SVM modeling is to find the optimal hyperplane that separates clusters of vector in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other size of the plane. The vectors near the hyperplane are the *support vectors* (Dtreg 2011).

The distance between the dashed lines is called the *margin*. The vectors (points) that constrain the width of the margin are the *support vectors*. Rather than fitting nonlinear curves to the data, SVM handles this by using a *kernel function* to map the data into a different space where a hyperplane can be used to do the separation. Usually data in the space are non-linear, and thus kernel functions can be considered to map a problem in a feature space where the target function consists of a line (Dtreg 2011). In effort estimation, the input space consists of the attribute quantifying the cost drivers for software projects and the target function is an effort estimate. In this investigation, the kernel functions that will be analyzed are linear:

SVR, Polynominal: SVR-Poly and Radial Basis Function (RBF). These kernels are chosen for further investigation as it been reported in systematic literature review that these kernels are widely used and implemented in WEKA.

Classification and regression trees

The objective of CART models is to build a binary tree by recursively partitioning the predictor space into subsets where the distribution of the response variable is successively more homogeneous. The partition is determined by splitting rules associated with each of the internal nodes. Each observation is assigned to a unique leaf node, the conditional distribution of the response variable is determined. The best splitting for each node is searched based on a purity function calculated from the data. The data is considered to be pure when it contains data samples from only one class (Fewster et al 2001). Trees used for problems with numerical features are often called regression trees and trees used for problem with categorical features are often called as classification trees (Mendes 2008).

CART uses backward pruning algorithms. This means that they will grow a tree until it is not possible to grow it any further and thus the only stopping rule is when there are only 2 instances left in a node. This will lead to a very large tree that over fit the data. In that case, we use Reduced Error Pruned trees (REPTrees) which are simulated by WEKA tools that use the concept of pruning to build smaller tree models that perform better on new data. The idea is to remove leaves that have a high error rate. There are two methods of pruning that are used in CART algorithm. The first is to use an independent testing sample, usually made by holding back a proportion of the data in reserve and building the model with the remaining data. The testing data is then used to estimate the error rate for each node. Working back from the leaves upwards, each nodes error rate is compared with the weighted average of the error rates of all the leaf nodes in its subtree. If the error rate of the node is lower, the whole subtree is removed and the node in question is changed to be a leaf node. If it is not lower, then the subtree is left intact and the node above will be examined next. This continues until the root node is reached. In this way, the tree model is reduced in size up to a point where further reduction will not yield lower error rates. The method described above is referred to as reduced-error pruning. Its main disadvantage is that some of the data is being held back for pruning and that this data cannot be used in helping to build a better tree model. This can be a serious problem when dealing with small datasets (Weka 2011b).

Case Based Reasoning

Case based Reasoning (CBR) is a branch of Artificial Intelligence where knowledge of a similar past cases is used to solve new cases (Shepperd and Kadoda 2001). Herein completed projects form the case base. The new project is referred as the target case which use the all the features and estimate the effort. There are some issues on this method such as similarity function, the number of analogies to select the similar projects to consider for estimation and the analogy adaptation strategy for generating the estimation. To understand these better, more details are discussed in next chapter. An initial investigation has been carried out using this technique on ISBSG dataset which motivates to study the impact of dataset characteristics on prediction techniques (Letchmunan et al 2010). The similarity measure used in this study is the Euclidean distance and effort estimates were obtained using the effort for the most similar project in the case base (CBR1), and the average of the two (CBR2) and three (CBR3) most similar projects.

2.4 Systematic Literature Review on Web Application Cost Estimation

2.4.1 Introduction

Traditional literature review shows that cost estimation is an intrestering area to explore. However based on traditional literature it's hard to systematically review the available evidence. Therefore this initial traditional literature review motivated the need for a systematic review on web application cost estimation studies. The aim was to systematically review and report the available evidence in the current literature to support the proposed research questions.

Based on the traditional review it is found that there are a wide range of existing approaches, so one natural line of enquiry is to try and determine which of these work best. Another point that arises in the previous sections is that some authors mention that certain techniques perform better with messy data, and so examining the impact of the dataset is another line of enquiry. These potential areas of investigation are formulated into more precise research questions as part of systematic literature review.

This process was initiated by an initial pilot study on a subset of papers to test the viability of the proposed research questions – did the papers include the necessary data to answer the proposed questions and how feasible was the proposed analysis? Based on this, the plan was refined and a full, thorough systematic literature review (SLR) of the web application literature was performed.

A systematic literature review is defined as identifying, evaluating and interpreting all available research relevant to a particular research question, or topic area, or a phenomenon of interest (Kitchenham 2004). The rationale for performing such a review can include: identifying the existing evidence regarding the use of a particular technology, to identify gaps in the existing research or to provide a context for properly placing new research activities (Riaz et al. 2009). Most of the literature on conducting systematic reviews suggests three phases: planning the review, conducting the review and reporting the review. Here it is proposed to use a refinement of these high level steps (Kitchenham 2004):

1. Define the research question.
2. Identify a few relevant studies and perform a pilot study.
3. Run searches on all relevant databases (IEEE, ACM, Google scholar, CiteSeer).
4. Document the search strategy.
5. Appraisal and selection of studies.
6. Analysing and presenting the results.
7. Discuss generalisability of conclusions and limitations of the review.
8. Make recommendations for practice.

The overall objective of the planned systematic review is to analyse and summarise the results to date on web application cost estimation and to identify needs and opportunities for future research in this area.

The remainder of this chapter is structured as follows. In Section 2.4.2, the proposed research questions for the systematic review are presented. Section 2.4.3 documents the search strategy that was used. This is followed by the search selection in Section 2.4.4. In Section 2.4.5, potential threats to validity are presented, followed by the results and discussion of each research question in Section 2.4.6. Section 2.4.7 provides the recommendation for practice. The conclusions are presented in Section 2.4.8.

2.4.2 Research Questions

In order to understand the state of the art of web application cost estimation in existing empirical research, it is important to find the right research questions. The main driver to perform this systematic review was to identify the trends and factors that impact on web application cost estimation. The research questions were identified and structured with the help of the Population, Intervention, Outcome, Context (PIOC) criteria (Kitchenham 2004).

Population	Web applications
Intervention	Methods/techniques for cost estimation
Outcome	Accuracy of cost estimation methods/techniques, successful cost estimation methods/techniques
Context	Encompass academia as well as software industry. All types of empirical studies including observation, interview, questionnaires, experiments, and case studies.

Table 2.3: Research question criteria

As a result, the research questions to be addressed in this systematic review were identified as follows:

Q1: What empirical evidence currently exists to support the effectiveness of the different cost estimation techniques for web applications?

Q1a: What techniques have been reported to estimate cost for web applications?

Q1b: What estimation techniques are reported to be superior for web applications based on what empirical evidence?

Q1c: What size measures have been used for measuring the accuracy of the estimation techniques for web applications?

Q1d: Which prediction accuracy methods have been used for web applications?

Q2: What are the key characteristics of the study datasets used in the study? Do these characteristics appear to affect the results?

Q2a: Which type of dataset has been used for this research (student/professional)?

Q2b: What are the main characteristics of datasets for web applications? Do they affect the results?

Most systematic reviews follow the process of planning, realization and reporting activities, each of which themselves consists of several steps. Although the plan is to follow the systematic review according to the procedure that is described by Kitchenham, it is proposed to start with a pilot study to investigate the appropriateness of the research questions and to explore the feasibility of gathering and analysing data which helps answer these questions based on existing empirical evidence. This process was initiated by an initial pilot study on a subset of papers to test the viability of the proposed research questions (Kitchenham, 2004) – did the papers include the necessary data to answer the proposed questions and how feasible was the proposed analysis? Based on this, the plan was refined and a full, thorough SLR of the web application literature was performed.

2.4.3 Search Strategy

In a systematic review, a well-planned search strategy is very important so that every relevant piece of work can be found in the search results. Therefore, an extensive search for research papers was conducted to try to answer the proposed research questions. The search terms used in this systematic review were developed using the following steps (Kitchenham et al. 2007):

1. Derive major search terms from the research questions by identifying Population, Intervention, Outcome and Context.
2. Identify keywords in the relevant papers.
3. Identify alternative spellings and synonyms for search terms with the help of a thesaurus.
4. Use Boolean OR to construct search strings from the search terms with similar meanings.

5. Use Boolean AND to concatenate the search terms and restrict the research.

The resulting search string was as follows:

(web or hypermedia) AND
(systems OR application OR method OR process OR system OR technique OR methodology OR procedure) AND
(cost OR effort OR development) AND
(estimation OR prediction OR assessment)

The distinction between hypermedia and web application is not that clear in literature. Therefore the search criteria included them to ensure that any papers confusing the terms were included. However, the studies will be dropped in the study selection stage if it is not related to web applications.

The search strategy contained the following decisions:

Searched databases: IEEE Xplore, Spinger Link, Science Direct, ACM digital Library.
Search items: Journal articles, workshops papers and conference papers..
Search applied on: Full text - to avoid exclusion of papers that do not include the keywords in the title or abstract, but are still relevant to the review.
Publication period: Since 1999.

This search was limited until July 2010 as that is the time this literature performed. Hence, any paper published after July 2010 is not included.

2.4.4 Study Selection

The search strategy resulted in 132 candidate papers. In the next stage all the irrelevant studies were excluded by reading the abstract. This process left 47 candidate papers. Further reading the full text and critically appraising the empirical work left 30 candidate papers. The complete list of these, along with their summaries, can be found in Appendix C

The following inclusion and exclusion criteria[1] were followed throughout this process, which was refined during the pilot studies.

Inclusion criteria:

- Estimation using web application dataset.
- Estimate web application effort.
- Size measure in web applications.
- Prediction techniques for web applications cost estimation.
- For duplicate publications of the same study, only the most complete and most recent was included.

Exclusion criteria:

- Web hypermedia
- Propose metrics models
- Literature on web size metrics
- Cost on writing code on OO framework
- Web application design techniques or Requirement methodology

All the studies that were identified in the primary study were read in depth in order to extract the data needed to answer the research questions. All the information that was extracted from the studies was inserted in a table form, which was designed during the pilot studies. All the information that was extracted is highlighted in different colours according to the different research questions. This approach helps the researcher to locate and validate the extracted information for future reference. For the reported SLR, the data that was extracted is presented in Appendix C and synthesised here when answering the research questions.

2.4.5 Threats to Validity

This section discusses the possible threats to the validity of the proposed review. These should be taken into account while interpreting or using the reported findings.

[1] Inclusion and exclusion criteria define the studies in the review and thus what the search strategy is attempting to locate. The inclusion criteria specify which studies are to be included in the review. Logically, those to be excluded from the review are listed in the exclusion criteria.

Studies that do not mention “Web” or “Hypermedia” in the title of article have not been included in the primary study set. Thus, it is possible that the search procedure has missed a number of studies that are concerned with web applications, but where it is not mentioned in the research title.

Studies that lack scientific rigor are excluded. Much of the research work reported by industry practitioners may fall into this category. It would have been beneficial to compare evidence from the practitioner community with the research or academic community but lack of empirical detail has made this difficult or impossible.

It is possible that the detailed dataset descriptions for published papers exist in non-published but referenced ‘grey literature’ such as technical reports. There is a danger that the systematic literature review fails to identify such papers and their important data.

2.4.6 Results and discussion

The findings and analysis of the data extracted from the reviewed papers in order to answer the research questions are presented in this section. The summary of the evidence for each research question is presented in Appendix C.

2.4.6.1 Types of Web Application Prediction Techniques (RQ1a)

Nine techniques were identified in the literature to estimate the web application effort. They are as follows:

- Case-Based Reasoning(CBR) or (analogy based estimation)
- Ordinary least squares regression (OLS)
- Linear Regression (LR)
- Stepwise regression (SW)
- Classification and Regression Trees (CART)
- Expert based estimate (ES)
- Bayesian Network (BN)
- Fuzzy radial basis function neural network (FRBFN)
- Support vector regression (SVR)

In this study CBR and analogy based estimation were classified as the same techniques. Among the above listed web application prediction techniques, the most popular techniques that were used by researchers was CBR. However, regression models which were separated into different categories were not far behind and accounted for 17 studies. Figure 4.1 shows the number of studies using the different techniques.

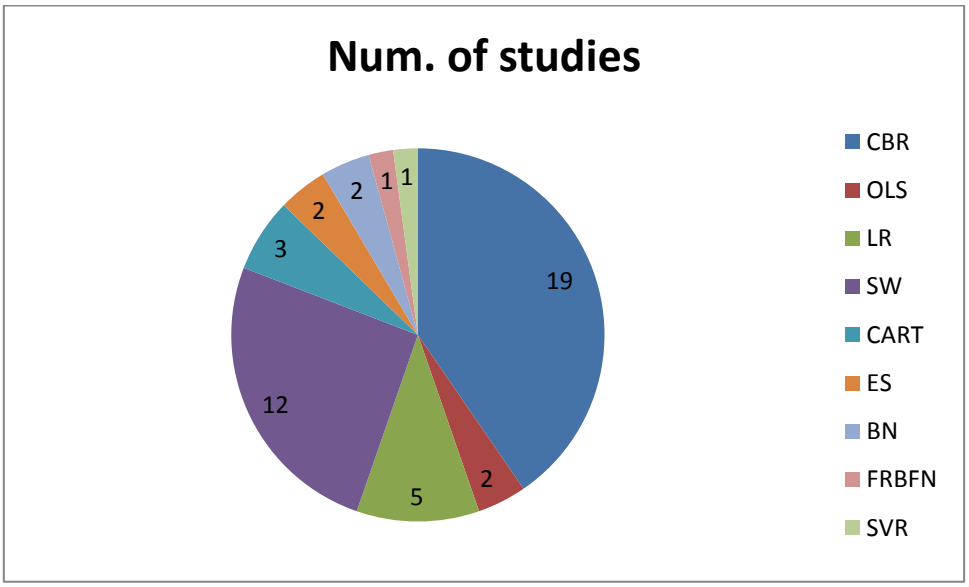


Fig. 2.1 Number of studies using the different prediction techniques

It was common in the early years for researchers to compare the results from CBR and different types of regression techniques. Recently, more research on network types of prediction techniques has been applied. Most recently the techniques that have been reported are a new generation of machine learning algorithms called Support Vector Regression (SVR) (Corazza et. al 2009).

2.4.6.2 Superior Web Application Prediction Techniques (RQ1b)

In the area of cost estimation, most researchers and practitioners would like to know which are the best prediction techniques. Due to the fact that, this question is not easy to answer, this section tries to gather evidence in the literature to give researchers and practitioners some guidelines on which techniques to choose for web applications.

Some existing works reported the best prediction techniques in their papers. The evidence is summarised in Appendix C in the column “Best Techniques”. Different features have been compared to select the prediction techniques. The best prediction techniques and the features that have been compared are shown in Table 4.2. Some of the papers mention which techniques performed better using certain size measures and type of dataset.

Overall there are mixed results in terms of the best prediction techniques in the literature. Different settings that have been analysed by researchers provide different results which make it hard to conclude which are the best prediction techniques. However, an interesting finding of this research question is that no research has been carried out into the details of the characteristics of the dataset that are used. Researchers such as Mendes et al. (2004, 2007, 2008b) and Kitchenham et al. (2004) tested different dataset groups such as a single company or cross company dataset.

Mendes et al. (2004) identified several factors which could explain this such as the small size of the datasets and the presence of outliers. They also reported difficulties in obtaining industrial data, although companies see the benefit of contributing data on their projects. They also pointed out that CBR estimation is considered to be a good technique in estimating the effort of web based applications within the same organization as the data and cost factors are approximately the same for each and every development effort. However, in a previous chapter using the ISBSG dataset which consisted of data from different companies, the results that were produced by CBR were disappointing. None of the averages of MMRE results of this study was anywhere near the 25% value – in fact values below 100% were rare Therefore several factors could have contributed to these kinds of research results, such as the small dataset size, the presence of outliers, and data obtained without rigorous quality assurance procedures.

Study	Best Techniques	Features compared
Ruhe et al. 2003	OLS- Web Objects*	Web Objects vs Function Points
Costagliola et al. 2006	RT and CBR – Length Measure* SW- Functional Measure*	Length Measure vs Functional Measure
Mendes et al. 2002	CBR with different settings	Size measure, Euclidean distance and analogies
Ruhe et al. 2003b	Web Cobra*	OLS vs ES vs Web Cobra
Mendes et al. 2001a	Linear Regression	LR vs SW
Mendes et al. 2001b	CBR	LR vs SW vs CBR
Mendes et al. 2002	No single technique	LR vs SW vs CBR
Mendes et al. 2003	SW	SW vs CART vs CBR
Mendes et al. 2002c	SW	SW vs CBR
Watson et al. 2002	CBR-Weighted Euclidean distance	Different adaptation Settings
Mendes et al. 2002d	LR and SW	LR vs SW and CBR
Mendes et al. 2003a	CBR- Company Specific dataset#	CBR vs SW
Mendes et al. 2003b	CBR- adaptation rules	
Kitchenham et al. 2004	Within company models#	Cross company dataset vs Within company dataset
Mendes et al. 2004	SW- Within Company# CBR- Cross company#	SW vs CBR with dataset setting
Sergio et al. 2007	SW- Length Measure* CBR-Tukutuku Measure*	SW vs CBR with different size measure
Mendes et al. 2007b	None superior	SW vs CBR vs CART
Mendes et al. 2007a	Single Company#	SW vs CBR
Mendes 2007	BN	BN
Mendes 2008	BN Hybrid Model	SW vs CBR vs BN
Idris et al. 2008	FRBFN- C Means	Fuzzy C- means vs FRBFN using hard C-means
Mendes et al. 2008b	Single company datasets#	SW vs CBR
Corazza et al. 2009	SVR	SVR vs SW vs CBR vs BN

***Prediction model that appear the best techniques using size measures**

#Prediction model that appear the best techniques using types of data

Table 2.4: Evidence of Best techniques

From the results in Table 2.4, it is found that although CBR appears as the "Best Technique" 7 times, it only involves 3 distinct researchers (first named) i.e. Mendes, Watson and Sergio. But both Watson and Sergio also include Mendes on the author list which suggests that much of the research could be emanating from the same group. One reason for this could be the availability of the Tukutukutu dataset, a relatively large dataset of web application cost data initiated by Emilia Mendes at the end of 2002. However, the Tukutukutu dataset is not publicly available, being only accessible by companies that make a contribution to the data or to collaborators of Mendes. These factors make it difficult to speculate about the performance of CBR on other datasets, and also hard, if not impossible, to validate the results due to the unavailability of the data.

2.4.6.3 Size Measures for Web Applications (RQ1c)

In software development, function points have been widely used as size measures. However, the framework of web applications challenges the use of function points which take into account the number of interfaces, files and queries in a specification. As a result, different researchers have tried to solve this problem by introducing different types of size measures for web applications.

In Chapter 2 in the traditional literature review the introduction of web objects was covered in detail. Web objects were introduced by Reifer as an alternative for function points to solve the problem of web application features. Ruhe et al. (2003) pointed out that web objects were much more preferable for web applications compared to function points. The results of their empirical analysis also revealed that models based on web objects showed significantly better prediction accuracy. Based on an email reply from Reifer it was noted that the research on web objects had not continued because of limited funds.

Mendes actively introduced and researched different types of measures such as size metrics, reusability metrics, complexity metrics, effort metrics and confounding metrics. However, most of these studies focused on web hypermedia. Measures that have been used such as page counts and media counts might not be practical for industrial practice.

Costagliola et al. (2004) studied size measures and first introduced COSMIC-FPP as an alternative size measure for web applications. They also studied length measures (e.g. number of pages, number of media, number of clients and server side scripts) and functional measures (e.g. external input, external output, external queries, etc…) using both stepwise

linear regression and case based reasoning (Costagliola et al. 2006). Their empirical results revealed that length measures provided better estimates when using CBR and functional measures provided better results when using stepwise regression. However, their results have not shown any significant differences in terms of the estimation between them.

In the latest studies on the comparison of size measures Sergio et al. (2007) compared all the existing size measures which were reported in the literature. Four sets of size measures were identified such as web objects, length measures, functional measures and Tukutuku measures (Mendes et al. 2003c). From this investigation, based on industrial datasets, the empirical results showed that all the measures gave good predictions in terms of prediction accuracy measures for both SWR and CBR. Moreover, using SWR, length measures and web objects yielded significantly better results than functional measures, but presented similar results to the Tukutuku measures. Meanwhile for CBR, the results did not show any significant differences amongst the four sets of size measures.

Although a large number of size measures have been introduced and researched, it appears that there is no standard size measure which can be reliably used for web applications.

2.4.6.4 Prediction Accuracy for Web Applications (RQ1d)

Prediction accuracy can be measured using various metrics. In terms of effort estimation several metrics measure the accuracy from different aspects. Therefore to answer the research questions (RQ1d), it is important to report the metrics that have been used by researchers to date.

It was found that MMRE (Mean Magnitude of Relative Error), Pred (25) (Percentage of prediction that is within 25% of the actual value) and MdMRE (Median Magnitude of Relative Error) were the three most popular accuracy metrics. In some studies boxplots of absolute residuals (actual effort- estimate) and boxplots of z (estimated effort / actual effort) have also been reported.

Kitchenham et al. (2001) criticised MMRE on the basis that it is essentially a measure of the spread of z (z=estimate/actual) rather than accuracy and suggested that boxplots of residuals and boxplots of z were better alternatives or a complement to summary statistics. As a result, more researchers are now including these boxplots with their results.

2.4.6.5 Types of Dataset for Web Applications (RQ2a)

The types of dataset that have been used in the literature are either student or industrial datasets. Therefore RQ2a gathers information in term of the types of dataset that have been used and the availability of industrial datasets.

There is an equal split between student and industrial datasets that have been used in the literature. The trend also shows that more research on industrial datasets has been carried out recently compared to student datasets. Out of 15 studies that use industrial datasets, 7 are from studies that were reported after 2006. The reason for this is the collection of the Tukutuku database. Tukutuku means "Web" in Maori, which is the native language of New Zealand (http://www.metriq.biz/tukutuku/). This database consists of 150 data sets from a combination of web hypermedia and web applications. However, this dataset is not available for other public researchers to use because of the confidentiality of the dataset.

2.4.6.6 Characteristics of Dataset for Web Applications (RQ2b)

The main idea of the research question (RQ2b) is to gather information on the characteristics of the datasets that have been used in the literature. The main supporting idea to look at in this question is to see the definition of the different terms that have been used to describe the dataset in the literature and also to consider the effect of using this kind of dataset. This is because in earlier studies it was found that Shepperd and Kadoda suggested that data set characteristics could have a strong influence on the choice of techniques to employ to obtain effort estimates (Shepperd 2001). As a result in this section, evidence from the literature that describes the characteristics of the dataset using terms such as "messy", "less-messy", "smaller" and "homogenous" will be presented. This is followed by looking at whether the use of these dataset characteristics has had any effect on the techniques as has been suggested by Shepperd.

The general definition of "messy" means untidy and in a disordered condition. However, from the literature it was found that "messy" datasets were defined as a "discontinuous cost" function, where there were no linear or log-linear relationships between size and effort "(Mendes 2003). Meanwhile in most of Mendes's studies the "less messy" data set definition refers to: a small number of outliers, a small amount of colinearity, strong relationships

between predictors (independent) and response (dependent) variables. However, there is no indication of how small the dataset is. Collinearity represents the number of statistically significant correlations with other independent variables out of the total number of independent variables (Shepperd and Kadoda 2001). Mendes (2003) also claims that by choosing web projects from a single company, they hoped to obtain a dataset which would be "less messy".

Mendes also claims that using regression analysis with "less messy" datasets will give the best estimation accuracy and CBR should be used with more "messy" datasets to obtain more accurate effort estimates (Mendes et al. 2004 and 2007). An interesting result reported in another study (Mendes 2003), was that adaptation rules improve prediction accuracy if used on datasets which are "less messy", while predictions obtained on very "messy" datasets do not improve by using the adaptation rules. Similar to trends in further studies (Mendes et al 2004), CBR was also reported to be better for prediction across large heterogeneous datasets, but regression was better for within company predictions. The results in Mendes (2003) confirm previous work where, for normal datasets with co-linearity, stepwise regression had a better prediction accuracy more often than CBR or Classification and Regression Trees (CART). There are no definitions of heterogeneous datasets in the literature; however, it is believed that this is a similar result to that obtained in Mendes et al. (2004 , 2007) which again refers to "messy" datasets. Meanwhile Mosley et al. (2003) claimed data may be more "homogenous" when the range of the data is smaller than other datasets. The other study that mentioned "homogeneity" of datasets was in the related work of Lokan et al. (2008), which argued that it was better to train models using only homogeneous data rather than all the data available. There is also considerable interest in seeing how much data is sufficient for a company to perform useful estimations. Costagliola (2006), mentioned that their study was composed of 15 projects which is quite a small number from a statistical point of view. However, in other studies Kitchenham et al. (2004), web cost estimation demonstrated good prediction accuracy using 12 projects from their own data and agreed with Shepperd et al. (1997) that in some circumstances such as a stable development process, and depending on the number of variables included in a model, a dataset of 12 or 13 was sufficient.

Kitchenham et al. (2004) reported that a within company regression model was significantly better than a cross company model. One possible reason for the better performance of the within company dataset compared to the cross company one may be related to the likely of

the single company dataset. In their systematic review of single company and cross company studies they found that all studies where single company predictions were significantly better than cross company predictions a smaller number of projects were employed than in the cross company model. In addition, such datasets were characterized by smaller maximum effort.

Overall it can be concluded that most studies refer to "messy" data as having discontinuous cost function, where there are no linear or log-linear relationships between size and effort (Mendes 2002b). Data from cross-company datasets was also described as messy or heterogeneous. From the literature it was found that CBR was most favourable for this type of dataset. Meanwhile, less-messy datasets are related to single or within company data and are also referred to as homogeneous datasets. It has been suggested regression analysis will give a better estimation using this kind of dataset. This supports the claim of Shepperd and Kadoda (2001) that data set characteristics could have a strong influence on the choice of techniques to employ to obtain effort estimates.

2.4.7 Recommendations for Practice

This Systematic Literature Review has found that stepwise regression (SW) and case based reasoning (CBR) are the most common techniques that have been used and studied. However, recent research on machine learning techniques such as Support Vector Regression (SVR) is increasing and shows the potential to be chosen as prediction techniques. Despite the large number of empirical studies in this area, inconsistent results have been reported and it is hard for practitioners to use any of the findings as their guideline. Therefore, researchers are encouraged to conduct a standard operating procedure to do the research on cost estimation. There is no point in introducing a new technique for cost estimation. More research on existing techniques in a systematic manner could produce better guidelines for practitioners in industry.

The question as to which techniques are the best or superior should be eliminated from the researchers mind. As suggested by Shepperd and Kadoda (2001), the focus should be on which techniques are best suited in which circumstance. Different features and different characteristics of the dataset will not provide any consistent conclusions. Therefore, a guideline on choosing prediction techniques should be based on dataset characteristics.

This review has found that several size measures have been introduced and that there are no standard size measures yet. Therefore the research community in this area should agree that certain size measures should be used by all researchers and practitioners. Indeed, without a uniform size measure for comparison, confidence in prediction results will not be gained.

The most interesting finding of this review concerns is the datasets that have been used by researchers. In the early years of cost estimation research, most researchers used student datasets. The reason for lack of industrial datasets was confidentiality of the dataset. Therefore, this review suggests that there should be a standard on how these datasets can be captured without revealing confidential information. This will encourage replication studies to be carried out and thereby increase the contribution to a body of knowledge.

The review has found different data characteristics mentioned in the literature; however, there has been no empirical investigation of the dataset characteristics which could influence prediction techniques. Therefore, a framework of dataset characteristics which influence the prediction techniques should be introduced as a guideline for practitioners. The repetitions of such studies could also improve such frameworks.

2.5 Conclusions

In the earlier part of this chapter some challenges of web application cost estimation were revealed. Cost estimation techniques were put into three different groups in this chapter: algorithmic models, expert judgement and machine learning. There are a variety of new techniques that have been proposed for web application cost estimation. However, there are no clear conclusions as to which techniques should be used. Therefore the thesis continues with a systematic literature review to report the available evidence. This systematic literature review investigated web application cost estimation. An extensive literature review searched for relevant studies published in the period 1999-2010, finally identifying 30 primary studies that were used to try to answer the research questions (RQs) which were mentioned in this review. The principal findings of this review are summarized as follows:

(RQ1a) The techniques that have been reported to estimate the cost of web applications are Case-Based Reasoning (CBR) or (analogy based estimation), Ordinary Least Squares

Regression (OLS), Linear Regression (LR), Stepwise Regression (SW), Classification and Regression Trees (CART), Expert Based Estimate (ES), Bayesian Network (BN), Fuzzy Radial Basis Function Neural Network (FRBFN) and Support Vector Regression (SVR). Among them CBR and SW are the most frequently used. Recently, however, SVR has become the most popular among them.

(RQ1b) There is not one estimation technique that can be proven to be superior. Different techniques with different features give mixed results in terms of prediction techniques.

(RQ1c) There is not yet any standard software size measure which can be used to measure the accuracy of web application prediction techniques.

(RQ1d) There is consistency in terms of the prediction accuracy that has been used. The most popular are MMRE, MdMRE and Pred(25). Since 2001, most studies also included boxplots of z and residuals.

(RQ2a) In the early years most studies used student datasets due to a lack of industrial datasets. However, since the Tukutuku database was created more studies have been published using this dataset. Unfortunately, the studies that have used this dataset are from the same group of researchers as this dataset is restricted due to confidentiality.

(RQ2b) In terms of data characteristics, most research has focused on single versus cross company datasets. However, the other characteristics of the dataset are only mentioned as a possible reason for the outcome of the prediction techniques. A summary of the main characteristics of the dataset for web applications is:

Messy/Non Messy	**Characteristics**	**Prediction Techniques**
Messy	Discontinues cost function No linear or log-linear relationships between size and effort Cross company Large heterogeneous	CBR
Non Messy	Small number of outliers Small number of collinearity	Regression

	Strong relationships between independent and dependent variables Single company Homogeneous	

These findings show that there may not be much to be gained from looking at which techniques are the best. It may be more fruitful to look at which dataset characteristics will suit which techniques the best. This systematic litearature review has explored the state of the art of web-based cost estimation research, but has not been able to consider what the industry practice is. To complement this review, and find out how practitioners are dealing with this problem, the following chapter will first investigate the current industry practice on cost estimation.

3 Current Industry Practice on Cost Estimation

3.1 Introduction

The literature (Shepperd et al. 1997) (Fewster et al. 2001) (Mendes et al. 2007b) contains some excellent work for constructing and understanding web applications and the methods of cost estimation to apply. However, much of this research did not study real practice in industry, but instead were mostly based on web applications developed by students, which may affect the external validity of their results. Therefore, a survey of web application cost estimation practice in industry will reveal current practice and help to gather data for the future development of best estimation practices.

The main research objective was to identify the current cost estimation practice in the small software development industry. The survey was carried out in Scotland, mainly in Glasgow and Edinburgh, primarily as a convenience sample but also because one of the intentions was to try and build stronger links with companies which interested in the topic and purse these by visiting them. This survey helps to identify the differences between cost estimation techniques in the literature and those actually used in practice, and also discover the type and nature of web applications being developed.

3.2 Methodology

An online survey of cost estimation practice in Scotland, mainly in Glasgow and Edinburgh, was conducted. Contact information was obtained for 160 web development companies in Scotland using www.yell.com.uk. Several techniques were used to approach them such as email, letters (see Appendix A) and also telephone calls. Although all the companies had been sent a letter of invitation to the survey, the researcher still contacted them by mail and telephone to make sure they had received the letter and understood the survey process. Some help was also given to some companies via a phone call in terms of how they can participate in this survey. The survey instrument was an online questionnaire with 19 questions (see Appendix B for the full questionnaire). The decision was made to design the initial questionnaire with a limited number of questions in order to engage participation from as many companies as possible (Mendes et al. 2003a).

Respondents of the questionnaire had the email address and phone number of the researcher so that they could make contact should they have difficulties understanding the questionnaire or completing the survey. However, no such questions arose from the respondents.

This exercise yielded replies from 16 companies, giving a 10% response rate. Of these 16 companies, 10 companies indicated that they were willing to participate in a further interview. The primary reasons for low-participation were that they were either too busy or not interested. One of the aims of the further interviews was to open up discussions on the topic of obtaining data from interested organisations. However this further interview failed to achieve its objective in obtaining the industry data. All the companies were reluctant to contribute the data for the reasons of confidentiality and consequently no follow-up interview conducted.

3.3 Findings

Organisation Size

The responses indicated that the majority of these companies were small, consisting of fewer than 5 employees (see Figure 3.1). All responding companies had less than 10 employees. Four out of sixteen companies were owned by an individual and there were no other staff in these companies.

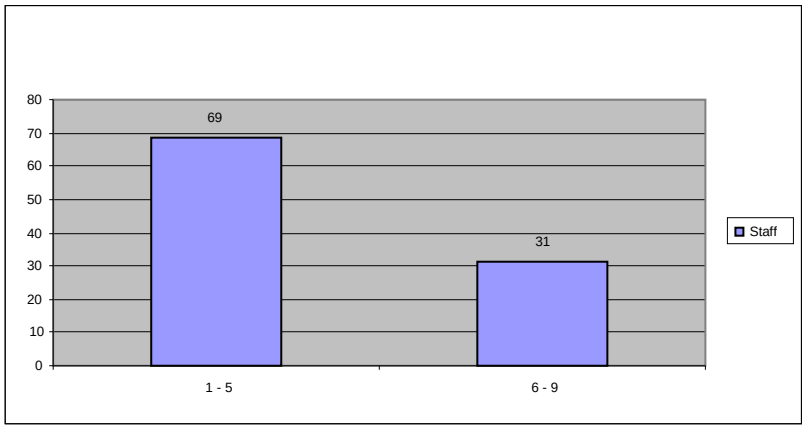


Figure 3.1: Organization Size (%)

Type of Web Application Developed

Figure 3.2 shows the range of web applications developed. Most of the companies developed Customer Management Solutions. The second highest proportion was E-commerce (27%),

followed by web sites (13%), and billing solutions (7%). Overall web applications which comprise billing solutions, customer management solutions and e-commerce dominate the type of web applications developed. Web sites which fall in web hypermedia are the least in terms of the proportion that have been developed.

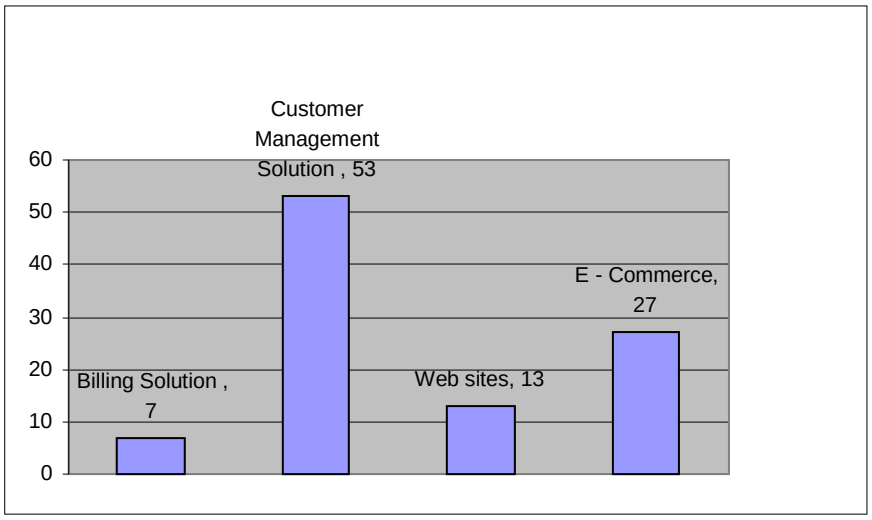


Figure 3.2: Type of Web Application Developed (%)

Typical Size of Web Application Developed

Concerning the typical size of web application developed (see figure 3.3), 56% fell in the range of 3 to 5 person months, 38% in the range of 1 to 2 person months and lastly only 6% fell in the range 6 to 7 person months.

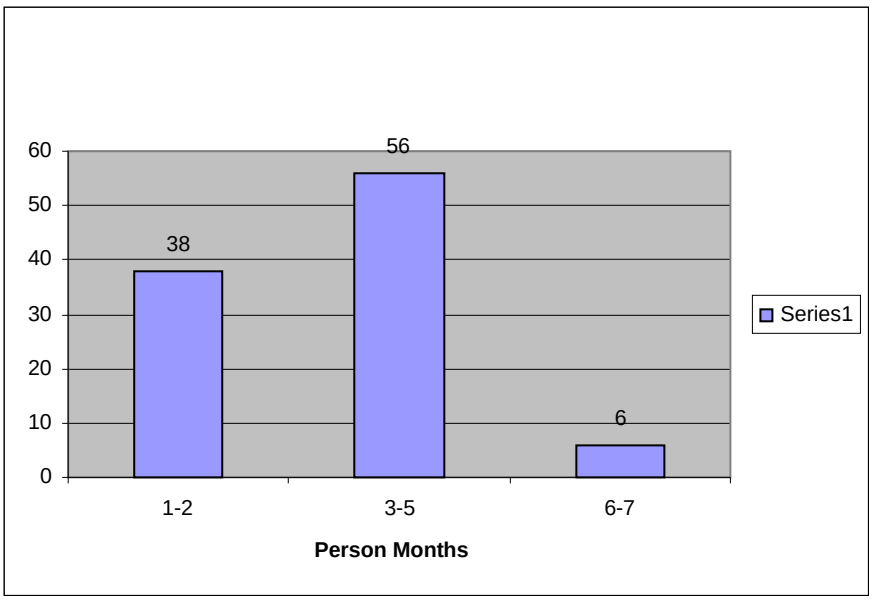


Figure 3.3: Typical Size Of Web Application Developed (%)

Tools

When asked how these companies carried out estimation, it was discovered that most of the companies (87%) that answered this question did not use any tools (see Figure 3.4). Only two companies used tools for their estimation, one stated that they used a home grown application, and another used a time sheet and calculator. These companies might be lacking awareness about existing estimation tools. Respondents also commented that they did not know how other companies were carrying out their estimates, but were keen to find out.

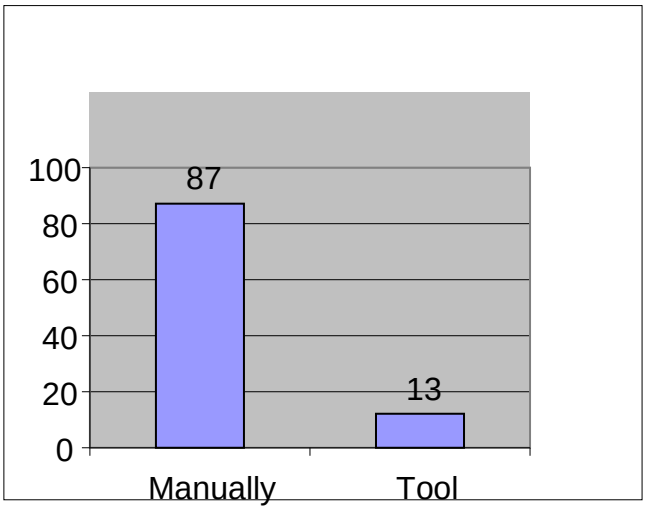


Figure 3.4: How do you carry out estimation?

Awareness of Cost Estimation Method

Interestingly, when respondents were asked which type of cost estimation methods they were aware of, a majority of the survey respondents (63%) answered that they had awareness of Expert Judgement methods; followed by Top Down estimation[2] (50%), Bottom Up estimation[3] (38%) and Estimation by Analogy (also 38%) (see Figure 2.5).

[2] Top-down estimation is applied to get an overall estimate for the project, usually early in the project life cycle. Functions point could provide basis for a top-down estimate

[3] Bottom-up estimation is used to estimate effort at the task level. It can be used by the project manager to monitor progress by task for a special stage of work. However it is not capable of providing overall project estimates for all aspects of the projects life cycle.

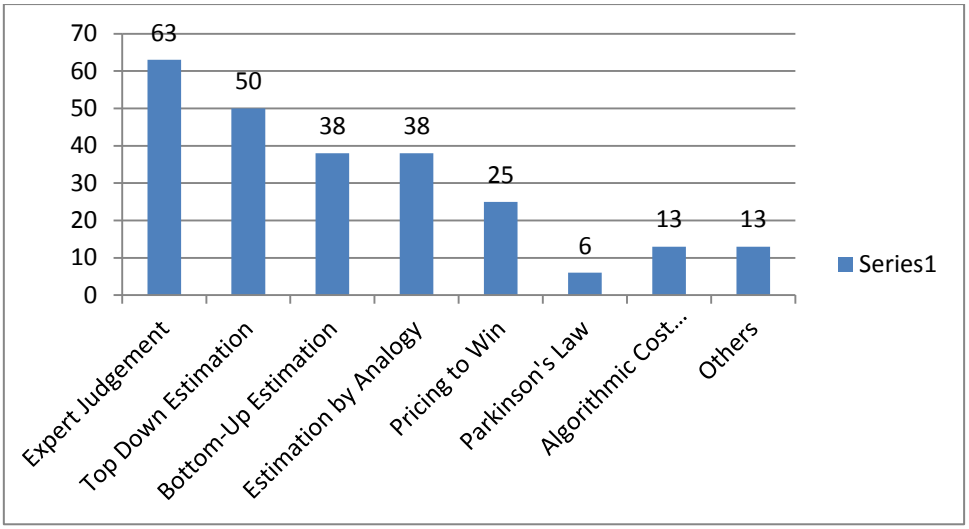


Figure 3.5: Awareness of Cost estimation Methods (%)

Cost Estimation Methods Currently Used

In the survey, respondents were questioned about the cost estimation methods that they currently used. A list drawn from the literature was presented. Although not every method or technique was listed, the list was considered adequate to get a picture of the general types of cost estimation methods being employed. From Figure 3.6, Expert Judgement was the most widely used. The next most widely used methods were Bottom-up estimation, Estimation by analogy and Top-down estimation.

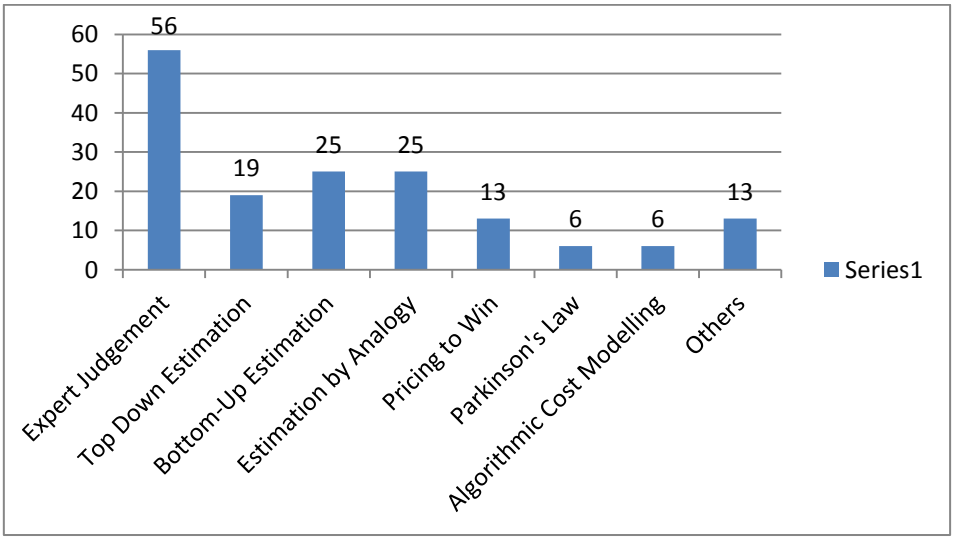


Figure 3.6: Cost Estimation Methods Currently Used (%)

Approach to the Estimation Stage

The next question was on what lifecycle stage the estimation was carried out. This uncovered some interesting findings (see Figure 3.7). Most of the companies carried out the estimation

at the initial or proposal stage, although about one in four respondents estimated the cost after the basic requirements had been captured. In Figure 3.7, the survey does not explain further the distinction between the different estimation stages, leaving this open to interpretation. These are meant to represent two early, but separate, stages in a project's lifetime. However, it is understandable that some degree to requirements analysis must take place in order to formulate an initial brief/proposal, and so the distinction between the phases could be somewhat blurred. This said, there were no queries or comments from questionnaire participants regarding these terms.

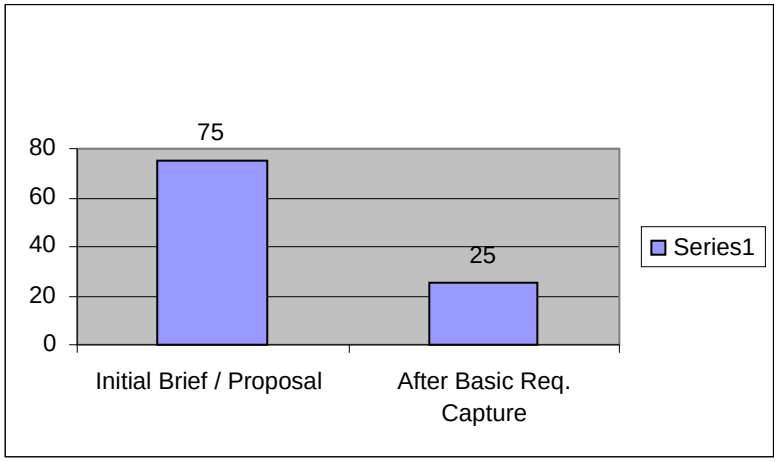


Figure 3.7: At what stage of the lifecycle do you carry out estimation? (%)

Cost Driver for the Estimation

The survey asked respondents to identify the cost drivers that were being used for web application cost estimation. A summary of their responses is shown in Figure 3.8. The majority of the respondents used duration (75%) and total effort (63%) as the cost driver in the estimation.

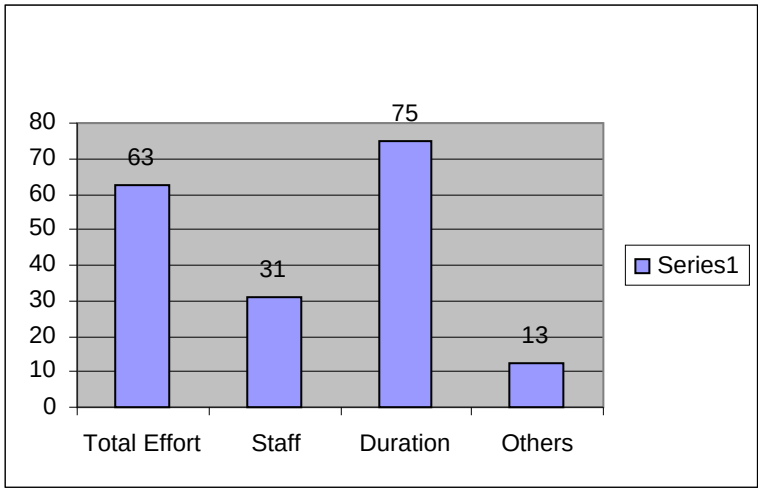


Figure 3.8: What was being estimated? (%)

Purpose of the Estimation

A majority of the respondents reported that the purpose of estimation was to win the bid (81%). Several respondents also reported using the estimation to determine person effort (50%) and budget approval (38%) (see Figure 3.9).

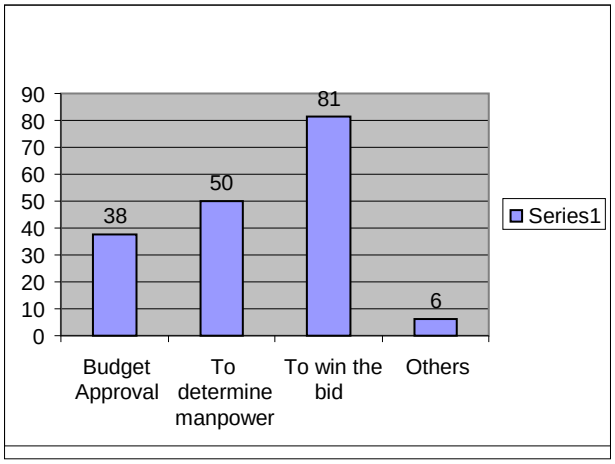


Figure 3.9: Purpose of estimation (%)

Person in Charge for Cost Estimation

The responses indicated that most of the companies' directors (75%) were the ones who determined the cost of the web application (see Figure 3.10). While this may be different for larger organisations, the respondents in this survey were all from small software firms which had fewer than 10 employees.

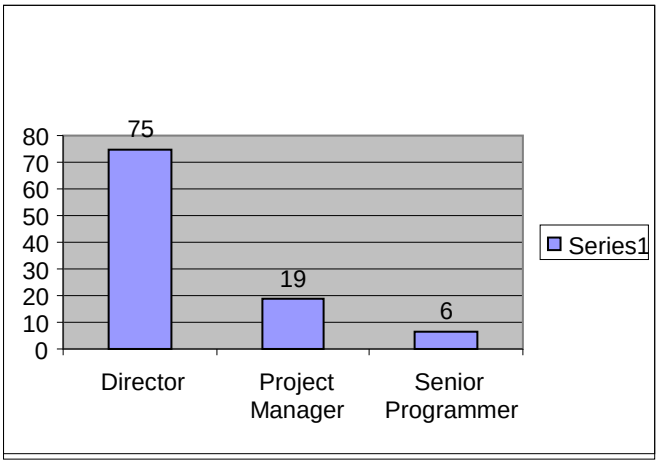


Figure 2.10: Person in charge for the cost estimation (%)

Experience in Web Applications Cost Estimation

Interestingly, 38% of the survey respondents had more than 8 year's experience (see figure 2.11), followed by 1 to 3 years (31%), and 4 to 7 years (25%). However, because the survey did not ask the respondents to give the number of web applications they had estimated in a month or a year, it is difficult to discern their true level of expertise in the area of web application cost estimation.

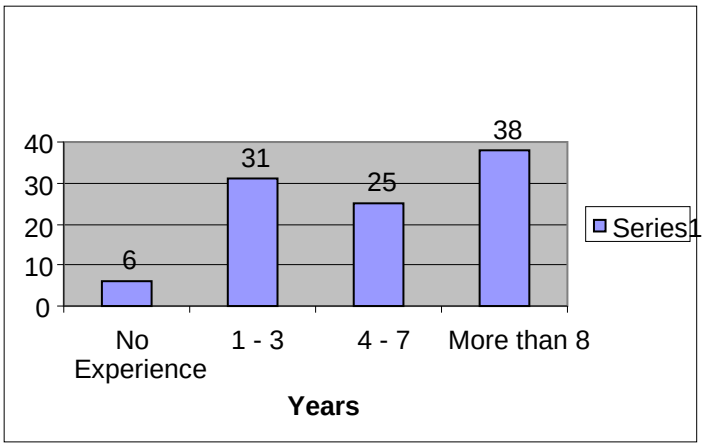


Figure 3.11: Experience in Web Application Cost Estimation (%)

Reasons for Inaccuracies

Finally, the reasons for estimation inaccuracies by the respondents are shown in Figure 3.12. Approximately 50% of the respondents reported that insufficient requirement analysis was the cause of the inaccuracies of web application cost estimation. Other reasons for inaccuracies were under-costing (19%), learning curve (19%) and early estimation (12%).

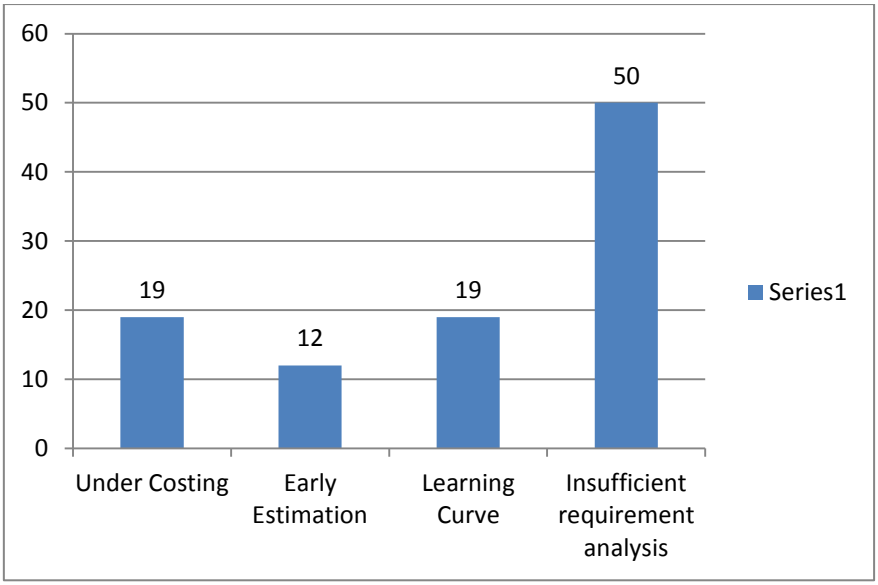


Figure 3.12: Reasons for inaccuracies (%)

3.4 Conclusions

The analysis showed that all the companies which participated were from small software organisations. In most of these companies, the decision on cost estimation was made by the owner or the director. The highest percentage of the studied respondents were aware of and had used expert judgement as their method of web application cost estimation. Almost all of them did not use any tool for this estimation. The majority of the respondents were experienced in web application cost estimation methods and their main purpose for estimating was to win the bid.

Further research efforts in this area are timely. It has been several years since researchers proposed different techniques for web application cost estimation. The fact that these methods have slowly trickled down to practitioners is disappointing for academics - at some level theory should inform practise. Interpreting why this has occurred is not easy. Although Reifer's study of web objects to support estimation (Reifer 2000) is a solid reference point for researchers in this field, it appears not to be used in practise. Lack of data from industry on the features or attributes that are needed for web object measurement made the use of web objects less practical.

The systematic literature review and industry survey motivate the need for further investigation into CBR. Based on the industry survey, estimation by analogy (which is how CBR operates) was by far the most popular technique that did not rely entirely on judgement

or expertise, which influenced the decision to investigate it further. CBR has also been chosen for further investigation as it is the most favourable among nine predictive models that were tested by Gray and MacDonell (1997). However, this investigation was only on the software dataset. As a results, an investigation into effort prediction of web application datasets using CBR will be presented in the next chapter.

4 Investigating Effort Prediction of Web based Applications Using CBR

4.1 Introduction

The systematic literature review and survey in the area of web application cost estimation motivates the need for further experimentation on one of the techniques used on web applications datasets. CBR has been chosen for further investigation as this is the technique that is sparking popular interest amongst academic researchers as well as from industrial practitioners. Furthermore, CBR was also reported to be better for prediction across large heterogeneous datasets.

There are several issues concerning CBR which will be covered in detail in the next section. However, the main purpose of the study is to investigate the optimal number of analogies (i.e. how many of the most similar cases should be taken into account) to employ when making an estimate, in addition to noting any other issues that arise when using CBR on web application data.

4.2 Challenges in CBR

Although CBR has been explored the most in the context of cost estimation, there are still a number of challenges regarding the effective application of CBR, some of which are general to a domain and others which may only be relevant to a particular dataset. The problems that most researchers encounter in applying CBR fall into the following categories (Shepperd et al. 2001):

(i) Feature Subset Selection

There are many features in the dataset but not all of them are necessarily relevant for predicting the project effort. They might be redundant or contain error data.

(ii) Scaling

Scaling or standardization involves the transformation of attribute values according to a defined rule such that all attributes are measured using the same unit. Angel (Shepperd et al. 1997) for example assigns zero to the minimum observed value and one to maximum observed value.

(iii) Similarity Measure

A distance measure in CBR is the degree of similarity between two projects in terms of their effort drivers. Euclidean distance is the most commonly used to solve this problem. Similarity measures for categorical data typically employ a value of 1 to represent a match and 0 otherwise.

(iv) How Many Analogies To Use

The number of analogies refers to the number of most similar cases that will be used to generate the estimate. Most of the previous work employs 1, 2 and 3 analogies, but there is no clear rule on how many analogies are to be used (Kadoda et al. 2000) (Mendes et al. 2003b) (Mendes et al. 2002b).

(v) Analogy Adaptation

Analogy adaptation concerns how to generate the estimate once the analogies are retrieved. Different approaches include using the mean of analogies or the nearest neighbour.

Several papers have investigated this last aspect in detail (Kadoda et al. 2000) (Shepperd et al. 1997), focusing on dataset size as one of the major factors concerning the accuracy of analogy based methods by analyzing the trends in estimation accuracy as the datasets grow. Although the work of Kadoda et al. confirmed that analogy based estimation achieves better results by employing larger training sets, Shepperd and Schofield claim that accuracy in analogy based estimation does not always increase within the number of projects or datasets – showing instead that it can be affected greatly by the introduction of outlying projects (Shepperd et al. 1997).

Therefore among all the challenges this study will focus on these two questions:

1. Does accuracy improve as the number of analogies increases?
2. Does accuracy improve as the number of projects increases?

Much of the work that has focused on these questions uses public datasets, many of which are old and do not employ web application data. Therefore it may be fruitful to investigate these questions by using a web application dataset.

4.3 The Dataset

The investigations in this chapter are all based upon the International Software Benchmarking Group (ISBSG) Release 10 dataset (ISBSG 2009). The data in the ISBSG repository comes from over twenty-five countries, with 60% of the projects being less than 7 years old. Software practitioners voluntarily submitted the projects in the ISBSG data set which was collected by questionnaire. The ISBSG collection pays much attention to the quality of the gathered data. There are special data validation forms and the project managers were asked to report the confidence they have in the information provided (Angelis et al. 2001). A specific field was used containing a rating code of A, B, or C applied to the project data by the ISBSG quality reviewers to denote the following:

A= The submission satisfies all the criteria for seemingly sound data.

B= The submission appears fundamentally sound but there is some evidence to question some of the supplied data.

C= The submission has some fundamental shortcomings in the data.

As the ISBSG point out, in any statistical analysis only projects with an A or B rating should be used. Of the 4,106 project summaries in the repository, 422 were related to web applications, and it is this subset which is the subject of this study. Therefore the data used in this study were web applications dataset with an A or B rating.

The dataset covered a wide range of applications, development techniques and tools, languages and platforms. Of the total of 109 features that may potentially appear in the ISBSG dataset, just 9 were selected which were considered relevant to this work, or which could potentially have an impact on effort or could consider the rest of attributes are

irrelevant for effort estimation studies. Other information that excluded such as variables that describing software quality – number of defects per projects categorized by their severity, i.e. extreme, major and minor defects. Table 4.1 lists the features used in this study.

Name	Description
Case Name	*Index*
CountApproach	*Counting approach that has been used such as IFPUG, LOC*
WorkEffort	*Summary of work effort in hours*
DevType	*Development Type*
AppType	*Application Type*
PriProgLang	*Primary Programming Language*
Database	*Database system*
FunctionalSize	*Functional Size*
AdjustedFP	*Adjusted Function Points Count*

Table 4.1: Description of selected features

4.4 Methodology

The main aim of this study is to investigate the impact of the number of analogies on the accuracy of estimates obtained through case-based reasoning. Consequently, the large dataset needed to be broken down into smaller subsets in order to provide more opportunities to experiment using different numbers of analogies, and also to mimic more closely the data set size likely to be available in an industrial context. The 422 web application records in the ISBSG dataset were divided into 3 groups, each consisting of 67 unique records (cases). Care was also taken not to include any cases that were incomplete.

Only 9 out of 109 features been selected for our studies as the rest of attributes are irrelevant for effort estimation studies. No investigations were carried on correlations between any of these characteristics with the summary work effort. The ISBSG dataset clearly separate these 9 features as relevant for effort estimation, therefore the thesis selects these 9 features for further investigation. Although function points and adjusted function appear to be closely

related but in this thesis both of these are retain as they are frequently used for cost estimation studies in literature.

Similarly to previous studies (e.g. Kadoda et al. 2000), in order to explore the impact of the number of cases, these three datasets were further subdivided (randomly again) to populate smaller datasets consisting of 17, 33, and 49 records. The reasons for the dividing the data in such way are to investigate the impact of increasing size of dataset. This exercise yielded a total of twelve data sets: three initial groups (labelled G1, G2 and G3) each containing 67 cases, each randomly subdivided into groups of 17, 33, 49 and labelled G1-Ran1-17, G1-Ran1-33, G1-Ran1-49, G1-Ran1-67, G2-Ran1-17, G2-Ran1-33, … G3-Ran1-67. Each of groups has 67 unique records. For each group there is overlap between the subsets.

For example:

G1-17 ⊂ G1-33 ⊂ G1-49 ⊂ G1-67

This procedure was then repeated a further two times to guard against any freak results introduced by the randomising the process (De Almeida et al 1998) producing a second (G1-Ran2-17, G1-Ran2-33, … G3-Ran2-67) and third (G1-Ran3-17, G1-Ran3-33, … G3-Ran3-67) – thirty-six data sets in all[4]. There is no-overlap between G1, G2 and G3.
The CBR tool Angel (Shepperd et al. 1997) was used for this experiment to determine the prediction value of the effort using the jack-knife method (also known as leave one out cross-validation). This procedure was the same as that adopted by others, including (Mendes et al. 2002b), and followed the procedure outlined below. This was applied to all 36 datasets.

For each case in the data set:
Discard the effort data for that case (marked as "unconfirmed" - in order to simulate a new project)
Using from 1 to 7 analogies:
Use the remaining cases to estimate the effort for the unconfirmed case
Restore the original effort value for the unconfirmed case and return it to the dataset

In Angel tool similarity is defined as Euclidean distance in n-dimensional space where n is the number of project features. Each dimension is standardized so all dimensions have equal

[4] Note that GNRanM-67 will be identical for all values of M, but are included in the results for the purposes of comparison.

weight. The notion of distance gives an indication of the degree of similarity (Shepperd et al. 1997). The Euclidean similarity is based on the Euclidean distance between two projects:

$$SIM(C_1, C_2, P) = \frac{1}{\sqrt{\sum_{1 \in P} Feature\_dissimilarity(C_{1j}, C_{2j})}}$$

where P is the set of n features, C_1 and C_2 are cases and

$$Feature\_dissimilarity(C_{1j}, C_{2j}) \begin{cases} (C_{1j} - C_{2j})^2 \\ 0 \\ 1 \end{cases}$$

where 1) the features are numeric, 2) if the features are categorical and $C_{1j} = C_{2j}$, or 3) where the features are categorical and , $C_{1j} \neq C_{2j,,}$ respectively (Shepperd et al. 1997).

4.5 Evaluation Criteria

There are several criteria to evaluate the predictions of a model (Conte et al. 1986). To gauge the accuracy of each estimated effort value, two values are calculated for each predictive models used for each dataset: the Mean Magnitude of Relative Error (MMRE) and the Prediction at level n (Pred(n)) (Pickard et al. 1999). MMRE is calculated to indicate the relative amount by which the predictions over or underestimate the real value, and Pred(25) to indicate how many of the predictions lie within 25% of the real values. Conte et al. (1986) suggest that MMRE <=25% and PRED(25) >= 75% as a criterion for acceptable model performance. MMRE, is an average of the magnitude of relative error (MRE) where MRE is calculated as |actual – estimate|/ actual. PRED(25) is calculated based on the percentage of projects that have an MRE value of <= 0.25.

4.6 Results and Analysis

In this section, graphs are used to illustrate the results. The study reports analogies (k) up to 7 on the x-axis and the value of MMRE on the y-axis. To study and illustrate the results better, all the graphs are scaled to a standard size in axis x and y. Axis- y represents the MMRE values in term of percentage. 1 in axis-y represents 100%.

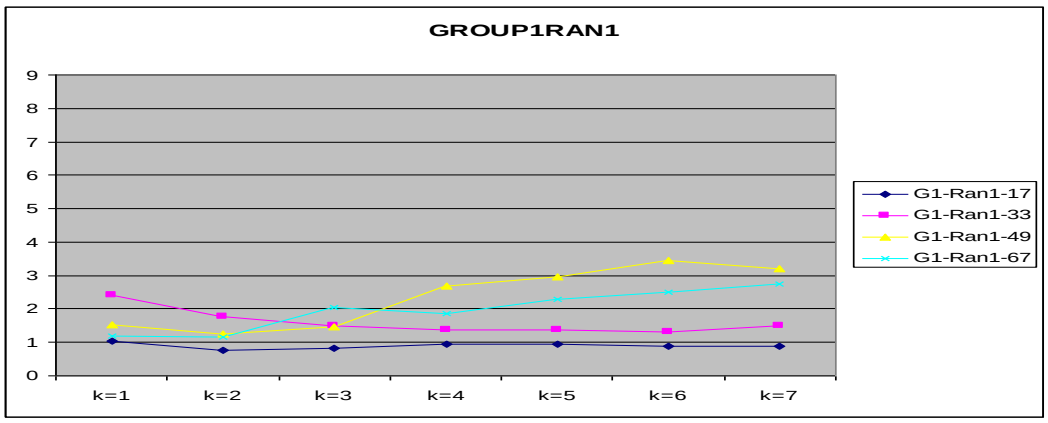


Figure 4.1: Result of MMRE vs Analogies on Group1Ran1

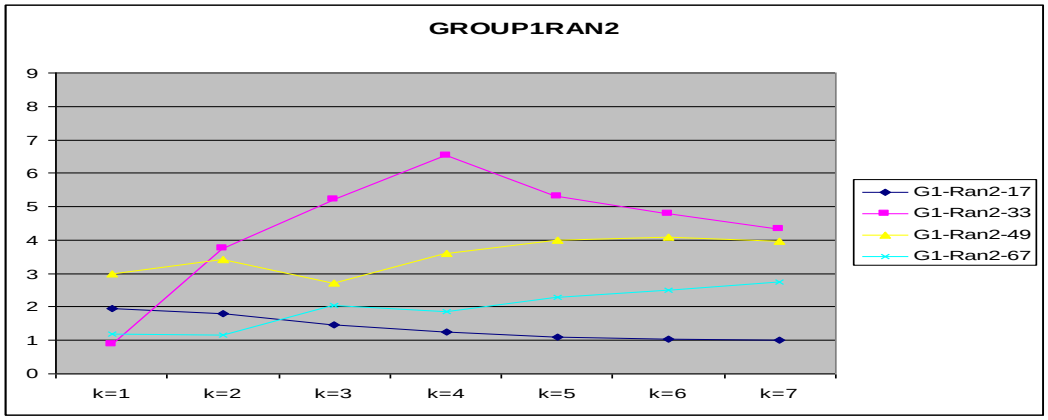


Figure 4.2: Result of MMRE vs Analogies on Group1Ran2

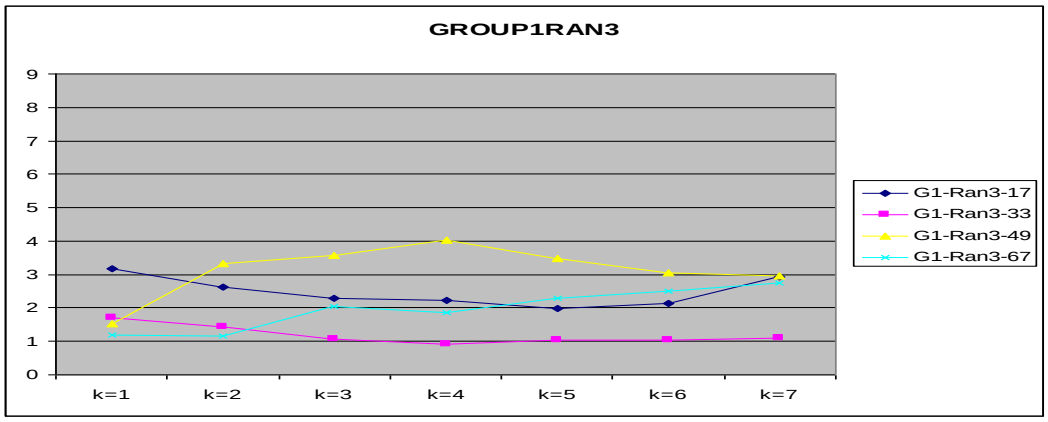


Figure 4.3: Result of MMRE vs Analogies on Group1Ran3

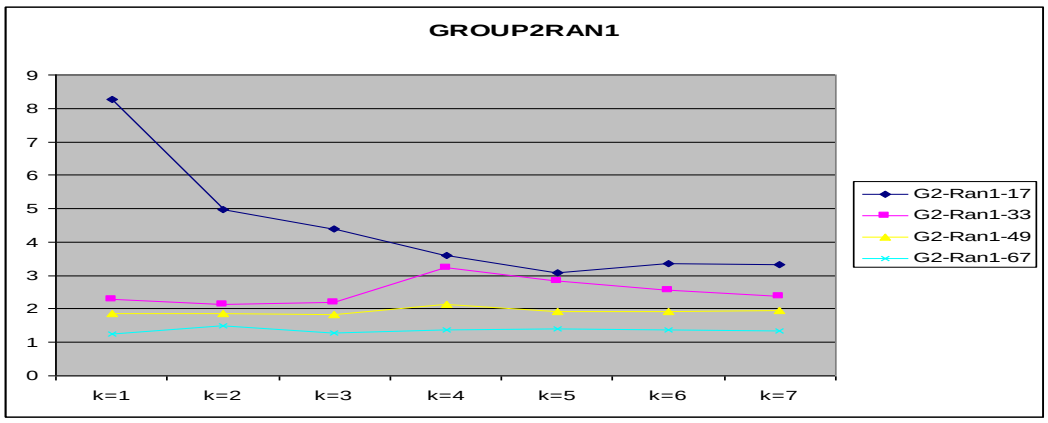


Figure 4.4: Result of MMRE vs Analogies on Group2Ran1

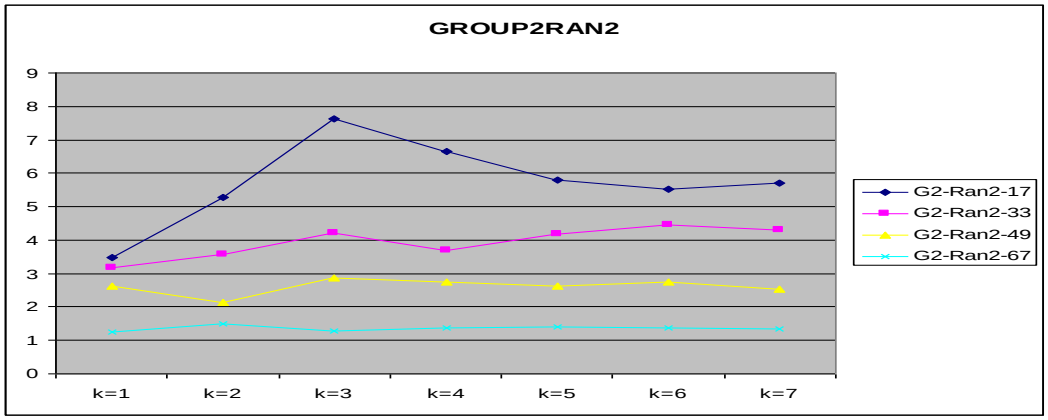


Figure 4.5: Result of MMRE vs Analogies on Group2Ran2

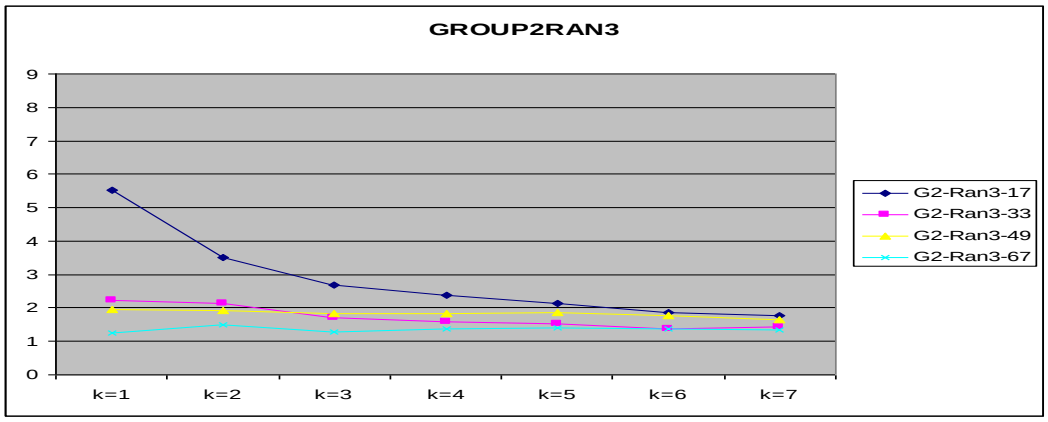


Figure 4.6: Result of MMRE vs Analogies on Group2Ran3

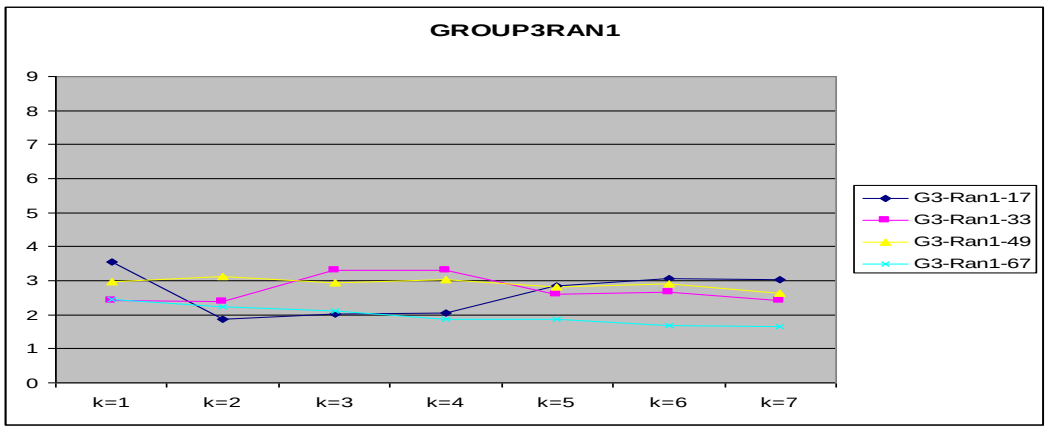


Figure 4.7: Result of MMRE vs Analogies on Group3Ran1

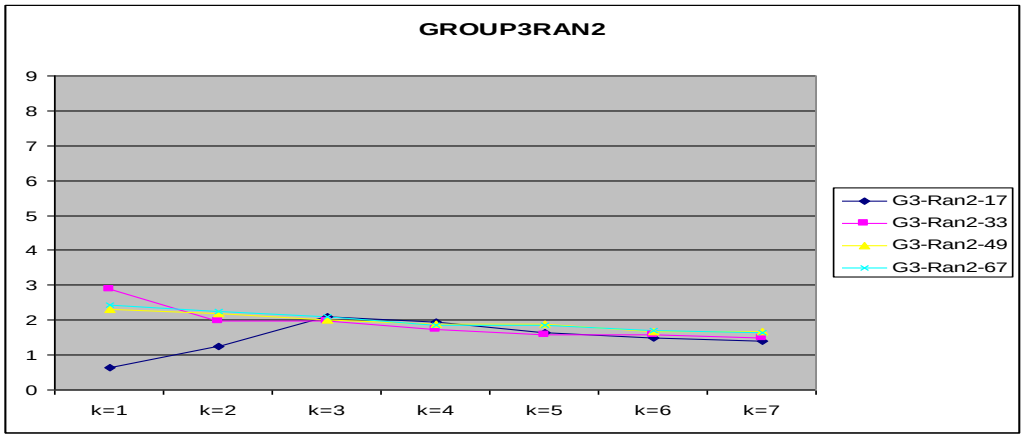


Figure 4.8: Result of MMRE vs Analogies on Group3Ran2

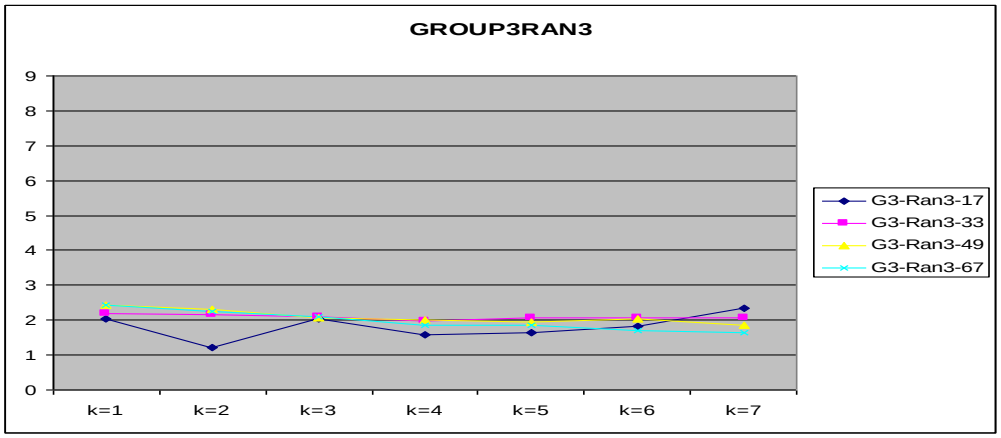


Figure 4.9: Result of MMRE vs Analogies on Group3Ran3

There are two immediately notable results concerning the MMRE values. Firstly, none of the averages is anywhere near the 25% value – in fact values below 100% are rare. Secondly, the graphs typically do not display any common trends. In some cases there is a general lowering

of the MMRE values as k (the number of analogies) increases (for example Group2Ran3, which shows a gradual convergence as k gets larger), whilst other cases show completely the opposite trend, and others still display sudden peaks or troughs. The remainder of this section will attempt to provide an explanation for some of these more pronounced patterns by considering some particular questions.

4.6.1 What is the reason for the peak in the results for G1Ran2-33 in Figure 3.2?

As can be seen in Figure 4.2, the results for this set show a very different pattern compared to G1Ran1-33 (Figure 4.1) and G1Ran3-33 (Figure 4.3) (drawn from the same set of 67 cases) and even for other configurations of the Group1 data (a similar shape can be observed in G1Ran3-49 (Figure 4.3), but the peak value is considerably lower). Also, it is unusual that the MMRE value starts off as one of the lowest for k=1 and climbs to one of the highest for k=4. To investigate this result in more detail it is necessary to look more closely at the dataset (up to k=4 for space reasons), shown in Table 4.2.

As can be seen for k=1, the most frequently predicted effort value is 47. This can be examined in more detail by looking at two different cases (those named 13700 and 10566) which have very different values of actual effort (352 and 8580 respectively) but which show the same predicted effort value of 47 when k=1.

Each entry in the dataset conforms to the following format:

> Case Name, Count Approach, Summary Work Effort, Development Type, Application Type, Primary Programming Language, First Database System, Functional Size, Adjusted Functional Points.

Summary work effort is the field that is left blank in Angel tools. Once all other features and number of analogies are inserted, the tools will provide predicted effort values.

For case name 13700 which holds the following data:

- 13700, IFPUG, 352, Enhancement, Process Control, ASP, SQL SERVER, 133, 133

the nearest calculated data points are:

- Rank 1, Distance: 0.654

  13319, IFPUG, 47, New Development, other: Sales contact management, ASP, ORACLE, 113, 113
- Rank 2, Distance: 0.713

  13127, IFPUG, 7496, New Development, Workflow support & management, ASP, SQL Server7, 786, 786
- Rank 3, Distance: 0.755

  15603, IFPUG, 756, Enhancement, Financial application area, Java, Interactive, 124, 124
- Rank 4, Distance: 0.755

  15008, IFPUG, 626, New Development, Financial application area, Java, Interactive, 116, 116

G1-Ran2-33		k=1	k=2	k=3	k=4
Case Name	**Actual Effort**	**Pred. Effort**	**Pred. Effort**	**Pred. Effort**	**Pred. Effort**
15720	*934*	*2240*	*1688*	*1225*	*988*
15008	*626*	*47*	*351*	*486*	*444*
13034	*4295*	*352*	*1621*	*3941*	*4361*
14779	*2891*	*2240*	*3267*	*5038*	*5184*
11100	*2240*	*2891*	*3593*	*5255*	*5346*
11648	*1056*	*1136*	*936*	*1935*	*1539*
10180	*2340*	*352*	*2986*	*4851*	*4712*
15440	*301*	*1136*	*707*	*782*	*691*
13127	*7496*	*352*	*199*	*1325*	*3836*
11283	*410*	*543*	*480*	*462*	*426*
15444	*2504*	*9231*	*10301*	*9366*	*7918*
14260	*3576*	*11372*	*9976*	*6666*	*6873*
15137	*543*	*410*	*364*	*382*	*393*
10358	*737*	*352*	*2143*	*1780*	*1619*
10427	*11372*	*8580*	*4313*	*5374*	*4924*
12078	*54*	*36*	*30*	*331*	*532*
11421	*36*	*24*	*39*	*337*	*537*
11132	*278*	*301*	*718*	*790*	*697*
13369	*418*	*425*	*417*	*371*	*414*
13700	*352*	*47*	*3771*	*2766*	*2231*
15603	*756*	*626*	*489*	*341*	*420*
14487	*3116*	*2240*	*1259*	*978*	*840*
13319	*47*	*352*	*4466*	*5476*	*6950*
12408	*425*	*418*	*414*	*368*	*412*
11718	*3934*	*47*	*4313*	*6666*	*5184*
13744	*1136*	*1056*	*704*	*715*	*1519*
13896	*1136*	*47*	*4313*	*6666*	*5075*
15468	*9231*	*11372*	*9976*	*7485*	*7488*
14911	*319*	*47*	*351*	*415*	*467*
11730	*5621*	*8580*	*5460*	*5071*	*4526*
13254	*24*	*36*	*45*	*341*	*540*
10566	*8580*	*47*	*5709*	*5680*	*4845*
11809	*655*	*47*	*336*	*330*	*437*

Table 4.2: Predicted effort for G1-Ran2-33

For case name 10566 which holds the following data:

- 10566, IFPUG, 8580, New Development, Financial transaction process/accounting, SQL, Oracle, 359, 359

The nearest calculated data points are:

- Rank 1, Distance: 0.663
  13319, IFPUG, 47, New Development, other: Sales contact management, ASP, ORACLE, 113, 113
- Rank2, Distance: 0.689
  10427, IFPUG, 11372, New Development, Financial transaction process/accounting, SQL, ORACLE, 859, 859
- Rank 3, Distance: 0.755
  11730, IFPUG, 5621, Enhancement, Document management; Financial transaction process/accounting Image video or sound processing, COBOL, IDMS-DB, 344, 344
- Rank 4, Distance: 0.756
  10180, IFPUG, 2340, New Development, Financial transaction process/accounting, Visual Basic, SQL-Server, 309, 309

In this example it appears that the similarity measure used in the Angel tool is having an effect on prediction. Consideration of case 10566 suggests that the best fit (and highest rank) should be case 10427 as it has several of the categorical fields in common (Development Type, Application Type, Primary Programming Language, and First Database System). However, it is pushed into second place as the distance measure appears to be dominated by the numeric fields (categorical fields are given the value 1 if they match and 0 if not), and consequently case 13319, whose numeric function point values are closer to case 10566 than case 10427, is ranked higher even though it has fewer categorical fields in common. This is quite a frequent occurrence – not just in this case but throughout the entire dataset. In many cases this will result in a less appropriate case appearing as the first ranked match which may go some way towards accounting for the relatively poor MMRE values.

As the value of k increases, then so does the MMRE – quite dramatically – resulting in an MMRE of 6.538 when k = 4. This average is skewed by some extremely high MRE values – as high as 146 in some cases. Case 13319 is an example of this:

- 13319, IFPUG, 47, New Development, other: Sales contact management, ASP, ORACLE, 113, 113

The nearest cases for 13319 are:

- Rank 1, Distance: 0.654
  13700, IFPUG, 352, Enhancement, Process Control, ASP, SQL SERVER, 133, 133
- Rank 2, Distance: 0.663
  10566, IFPUG, 8580, New Development, Financial transaction process/accounting, SQL, Oracle, 359, 359
- Rank 3, Distance: 0.716
  13127, IFPUG, 7496, New Development, Workflow support & management, ASP, SQL Server7, 786, 786
- Rank 4, Distance: 0.730
  10427, IFPUG, 11372, New Development, Financial transaction process/accounting, SQL, ORACLE, 859, 859

Clearly, the effort associated with all these closely ranked cases is some way off the target value (47), but that associated with the second, third, and particularly fourth cases are substantially different. So as k increases the MRE gets significantly larger: 115(refer to footnote[5]) for k = 3 and 146(refer to footnote[6]) when k = 4. Admittedly, this data point is the only one that has an MRE value of more than 100; the rest of the cases result in values less than 8, and the majority of them are less than 1. Nevertheless, this is the main reason that the MMRE is so large. It is a poignant illustration of the impact that outliers, or even the lack of close matches in the dataset, can have on the accuracy of effort predictions. Furthermore, it also demonstrates the rather unpredictable effect of increasing the number of analogies.

4.6.2 Why does G1-Ran3-33 display such a different trend compared to G1-Ran2-33?

In contrast to G1-Ran2-33, G1-Ran3-33 has a very different trend of MMRE values, showing a slight downward trend until k = 4 and a very slight increase thereafter. There are no peaks or extreme values as in the case of G1-Ran2-33, and the MMRE values range between 1.716 and 0.909. In some ways this is curious as the pattern of data in the two sets is apparently dissimilar as can be seen by the summary Table 4.3:

[5] The mean of the predicted effort is (352+8580+7496)/3 = 5476 and the MRE is Abs(47 – 5476)/47=115
[6] Abs(47 – (352+8580+7496+11372)/4)/47

Dataset	Mean	Median	Min	Max	Skewness
G1-Ran2-33	*2346*	*934*	*24*	*11372*	*1.706*
G1-Ran3-33	*2605*	*1136*	*24*	*11372*	*1.462*

Table 4.3: Statistics description of Particular Group Dataset

Both have the same minimum and maximum values, so why does G1-Ran3-33 not display any of the extreme values of G1-Ran2-33? From tables 4.4 and 4.5 it can be seen that the MRE for the predicted effort based on one analogy is better for G1-Ran2-33 than for G1-Ran3-33. This is caused largely by the poor initial matches for G1-Ran3-33, in addition to the frequent predicted effort of 47 for G1-Ran2-33 – often a very poor match but still yielding a MRE value of less than 1 (one of the weaknesses of the MRE calculation).

Case no.	Actual effort	Predicted effort	MRE
13319	*47*	*352*	*6.489362*
15440	*301*	*1136*	*2.774086*
15444	*2504*	*9231*	*2.686502*
14260	*3576*	*11372*	*2.180089*
15720	*934*	*2240*	*1.398287*
10566	*8580*	*47*	*0.994522*
11718	*3934*	*47*	*0.988053*
13896	*1136*	*47*	*0.958627*
13127	*7496*	*352*	*0.953042*
11809	*655*	*47*	*0.928244*

Table 4.4: Top 10 MRE values for G1-Ran2-33 (k=1)

Case no.	Actual effort	Predicted effort	MRE
13700	*352*	*7496*	*20.29545*
12573	*1671*	*11372*	*5.805506*
10173	*118*	*578*	*3.898305*
10178	*2503*	*11372*	*3.543348*
15940	*66*	*210*	*2.181818*
14260	*3576*	*11372*	*2.180089*
15675	*2762*	*8580*	*2.106445*
14194	*210*	*578*	*1.752381*
13254	*24*	*66*	*1.75*
14485	*484*	*1136*	*1.347107*

Table 4.5: Top 10 MRE values for G1-Ran3-33 (k=1)

In contrast, when four analogies are used the position is reversed and the top MRE values for G1-Ran2-33 are much higher (the value of 146 has already been illustrated) than those for G1-Ran3-33. These values are summarised in Table 4.6 and Table 4.7

Case no.	Actual effort	Predicted effort	MRE
13319	*47*	*6950*	*146.8723*
13254	*24*	*540*	*21.5*
11421	*36*	*537*	*13.91667*
12078	*54*	*532*	*8.851852*
13700	*352*	*2231*	*5.338068*
13896	*1136*	*5075*	*3.46743*
15444	*2504*	*7918*	*2.162141*
11132	*278*	*697*	*1.507194*
11100	*2240*	*5346*	*1.386607*
15440	*301*	*691*	*1.295681*

Table 4.6: Top 10 MRE values for G1-Ran2-33 (k=4)

Case no.	Actual effort	Predicted effort	MRE
13700	*352*	*3388*	*8.625*
12573	*1671*	*7921*	*3.740275*
13254	*24*	*112*	*3.666667*
10178	*2503*	*6911*	*1.761087*
14485	*484*	*1078*	*1.227273*
15675	*2762*	*5709*	*1.06698*
12078	*54*	*104*	*0.925926*
10173	*118*	*227*	*0.923729*
14260	*3576*	*6705*	*0.875*
10802	*578*	*112*	*0.806228*

Table 4.7: Top 10 MRE values for G1-Ran3-33 (k=4)

Although the worst case for G1-Ran3-33 produces a very high MRE value (8.625), this is substantially lower than the value of 146 which is primarily responsible for the overall high MMRE for G1-Ran2-33. Looking at this worst case in more detail it can be seen that the predicted effort values get closer to the actual effort (having started off some considerable distance away), which reduces the MRE. This is in contrast with the case of 13319 in G1-Ran2-33 where the values deviate even further as more analogies are brought into play.

- 13700, IFPUG, 352, Enhancement, Process Control, ASP, SQL SERVER, 133, 133

The nearest data points for 13700 are:

- Rank 1, Distance: 0.693
  13127, IFPUG, 7496, New Development, Workflow Support & Management, ASP, SQL Server7, 786, 786
- Rank 2, Distance: 0.707
  13981, IFPUG, 4648, New Development, Other: Sales Promotion Tool, Visual Basic, SQL SERVER, 895, 895
- Rank 3, Distance: 0.755
  15603, IFPUG, 756, Enhancement, Financial Application Area, Java, Interactive, 124, 124
- Rank 4, Distance: 0.755

11809, IFPUG, 655, Enhancement, Financial Application Area, Java, Interactive, 113, 113

From this it could be argued that projects distribution in the dataset is important: rather obviously, a case base that does not contain projects that are remotely close to those for which predictions are being made is unlikely to produce accurate results. This point is illustrated by group G2-Ran3. The trend for all subcategories in this group is the same: initially disparate values for k=1 quickly converge to a much smaller range as k increases. The MMRE values are still too high for this to be considered a "good" prediction, but the pattern of the graph follows the shape that might intuitively be expected. The reason for this is that the group (and subgroups) consists of data which is spread evenly from the lowest to the highest value. All groups have the same maximum (21700) but also contain other large values (19306, 14992 and 11165), which tend to be chosen as close matches to each other and result in relatively good estimates, or at least not very poor ones.

This appears to confirm the observations of Kadoda et al. 2001 and Shepperd et al. 2001, that there is likely to be a strong interaction between the accuracy of a given prediction system and the underlying characteristics of the dataset it is applied to. However, looking at the graphs of the results, it does not appear that increasing the size of the dataset improves the accuracy of the prediction – larger datasets appear to display similarly erratic results to the smaller ones. This interaction between the dataset and the predictions can be clearly observed in the graphs below which group the results by different sized datasets.

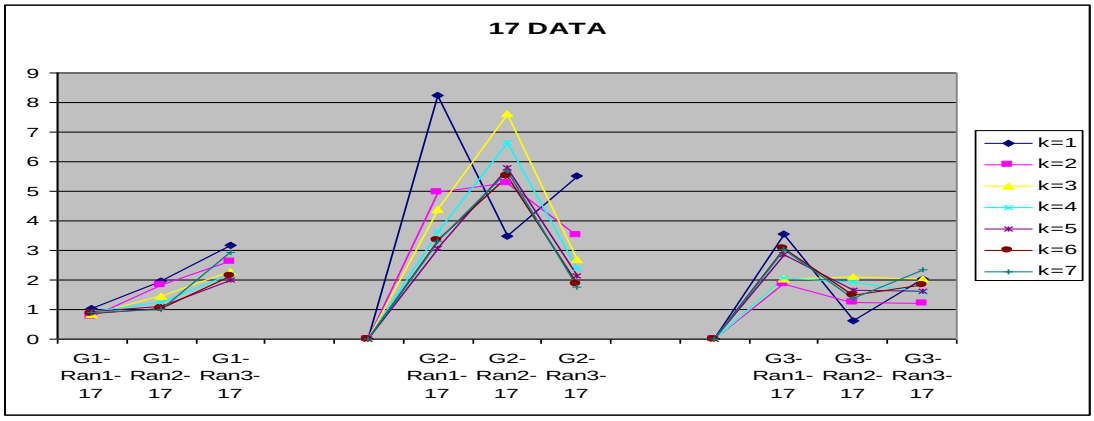


Figure 4.10: Result of MMRE vs Groups for 17 data

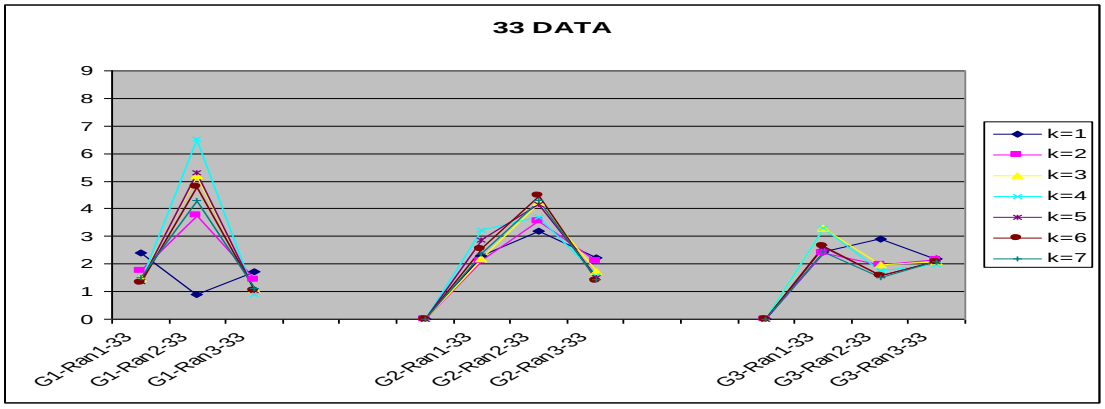


Figure 4.11: Result of MMRE vs Groups for 49 data

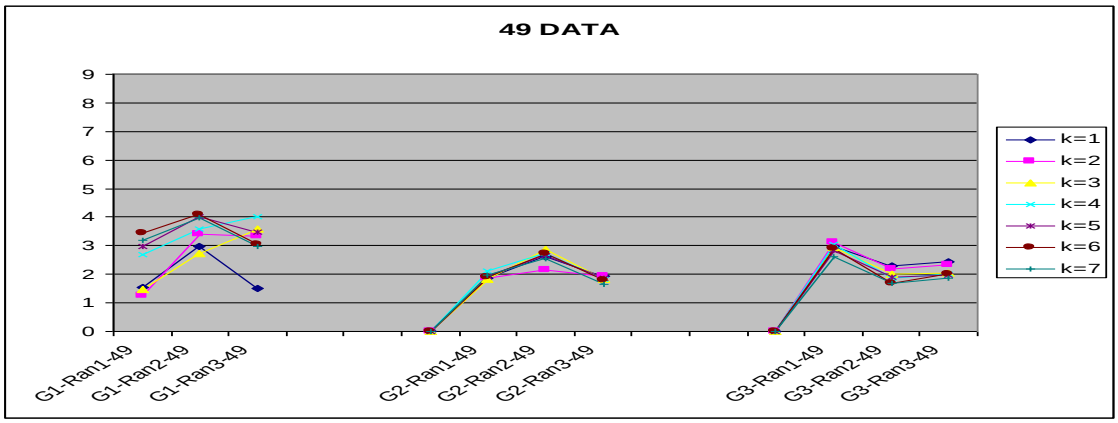


Figure 4.12: Result of MMRE vs Groups for 49 data

4.6.3 Why is the MMRE for k=1 for G2Ran2-17 so high?

When using only one analogy there is obviously no opportunity to average the results and so the difference in the value of effort could affect it. In G2Ran2-17 there are two big values in this group (14992 and 11165) and the next value is 5018, followed by 3303 and below. The presence of these high values could skew the effort predictions. We can investigate this further by looking at the results of the data set in both groups (see table 4.8).

G2Ran2-17		k=1	k=2	k=3	k=4
Case no.	**Actual Effort**	**Predicted Effort**	**Predicted Effort**	**Predicted Effort**	**Predicted Effort**
16023	*525*	*1712*	*1556*	*2138*	*1863*
16076	*105*	*51*	*156*	*1341*	*1434*
16612	*465*	*1037*	*2170*	*2380*	*2135*
17461	*1400*	*3303*	*1914*	*1621*	*1644*
17614	*3303*	*1400*	*1218*	*987*	*856*
18030	*2800*	*1009*	*737*	*751*	*823*
18398	*14992*	*11165*	*5973*	*4318*	*3938*
18705	*1009*	*1712*	*1246*	*1764*	*1439*
19107	*1712*	*1009*	*767*	*978*	*1661*
19673	*3712*	*262*	*183*	*693*	*532*
20145	*51*	*105*	*183*	*1359*	*1447*
20426	*147*	*5018*	*10005*	*10391*	*7989*
20896	*5018*	*147*	*7569*	*5979*	*7262*
21180	*781*	*1009*	*1904*	*4991*	*3859*
21550	*11165*	*781*	*895*	*1530*	*1263*
22177	*1037*	*465*	*1884*	*1722*	*1423*
22409	*262*	*3712*	*1908*	*1289*	*1395*

Table 4.8: Predicted effort for G2-Ran2-17 (up to k=4)

The data points that have the greatest impact on the MMRE are 20426 and 22409, which are considered in more detail below.

- 20426, COSMIC-FFP, 147, New Development, Transaction/Production System, Visual Basic, SQL Server7, 751, 751

The nearest data points for 20426 are:

- Rank 1, Distance: 0.755
  20896, COSMIC-FFP, 5018, New Development, Document management, ASP, SQL SERVER, 762, 762
- Rank 2, Distance: 0.846
  18398, IFPUG, 14992, New Development, Customer Billing/Relationship Management, HTML, ORACLE, 694, 694
- Rank 3, Distance: 0.902
  21550, IFPUG, 11165, New Development, Document mngnt; Financial trans process/acc; Image video or sound processing, Visual Basic, SQL SERVER, 307, 307
- Rank 4, Distance: 0.921
  21180, IFPUG, 781, New Development, Trading, Visual Basic, Oracle 8i, 235, 235

When k=1 for this data point the MRE is 33.13, which is the highest in this group. While when k=2, the MRE is 67.06 and the second highest MRE for this group is only 6.28. This again illustrates the impact of the numeric values (the final two size estimates) in the distance calculation.

The second data point also illustrates this issue but raises another interesting question:

- 22409, IFPUG, 262, Enhancement, Financial application area, Java, Interactive, 46, 46

The nearest data points for 22409 are:

- Rank 1, Distance: 0.755
  19673, IFPUG, 3712, New Development, Catalogue/register of things or events; Document management; Online analysis and reporting; Workflow support & management, Java, ORACLE, 51, 51
- Rank 2, Distance: 0.756
  16076, IFPUG, 105, Enhancement, Financial application area, Java, Interactive, 19, 19
- Rank 3, Distance: 0.756
  20145, IFPUG, 51, Enhancement, Financial application area, Java, Interactive, 9, 9
- Rank 4, Distance: 0.756
  19107, IFPUG, 1712, Enhancement, Relatively complex application, 4GL, Interactive, 89, 89

Again this leads to similarly high values for the MRE but illustrates another issue with the data. In all cases the size calculations are relatively low numbers of function points (46, 51, 19, 9, 89), but the effort values vary disproportionately (262, 3712, 105, 51, 1712) except where there is a close categorical match where the effort is almost consistently 5.5 times the size. This may be coincidence or may also indicate data which comes from the same company or even the same team. Unfortunately, such information is not available in the data set for reasons of privacy, even though it is potentially useful in finding matching cases.

4.6.4 Questions arising from the Pred(25) results.

As mentioned earlier, only a subset of the PRED(25) results are included for reasons of space (see Figure 4.13), even though they are considered a more preferable mechanism to MMRE for assessing the accuracy of prediction mechanisms given the weaknesses associated with MMRE (Foss et al. 2003). The PRED(25) results display similar characteristics to the MMRE results: no general trends regarding the accuracy of the estimate and the number of analogies, and a clear indication of the impact of the underlying data set. The y-axis for these figures is based on how many predictions lie within 25% of the real values. 0.4 in the y axis means 40% of data prediction that lie within this 25%. However according to Conte et al. (1986), Pred(25) should be more than 75%, which is 0.75 for the graph that is represented here. None of the results below show any of the Pred(25) results that match to the Conte criterion as an acceptable model.

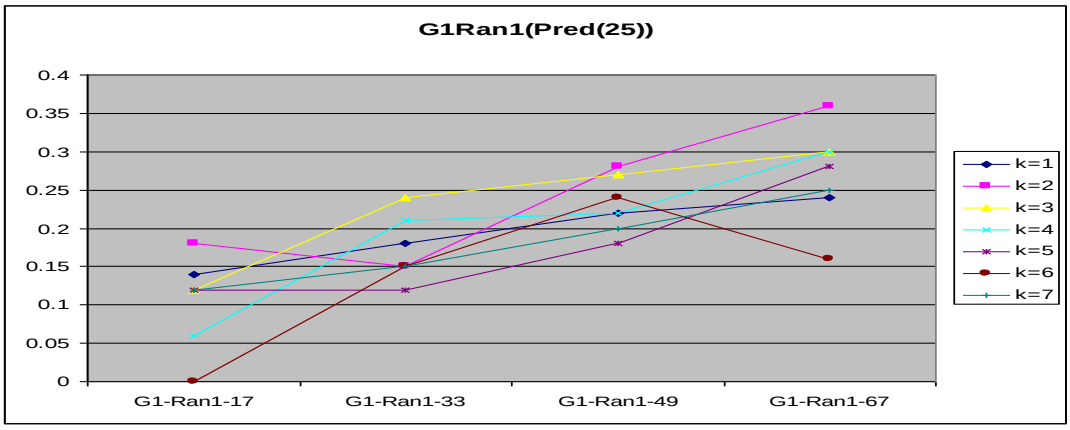


Figure 4.13: The Pred (25) results on Group1Ran1

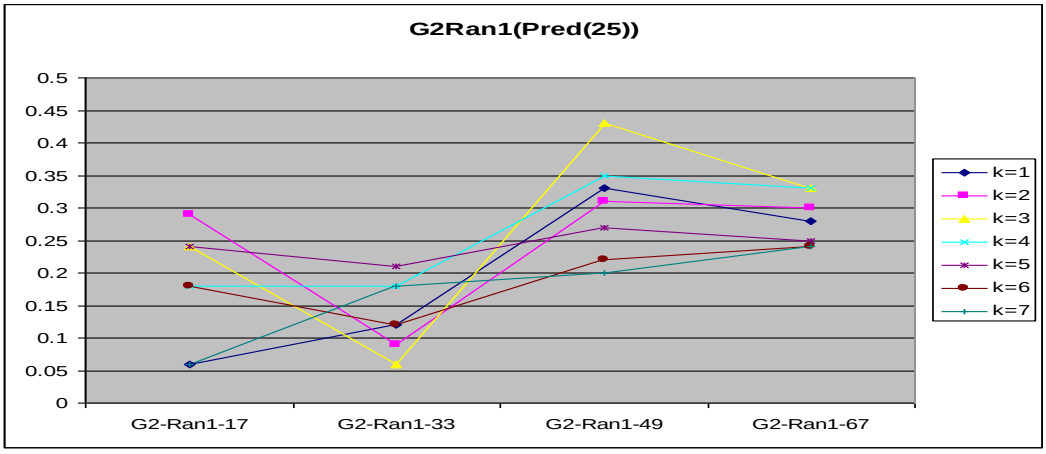


Figure 4.14: The Pred (25) results on Group2Ran1

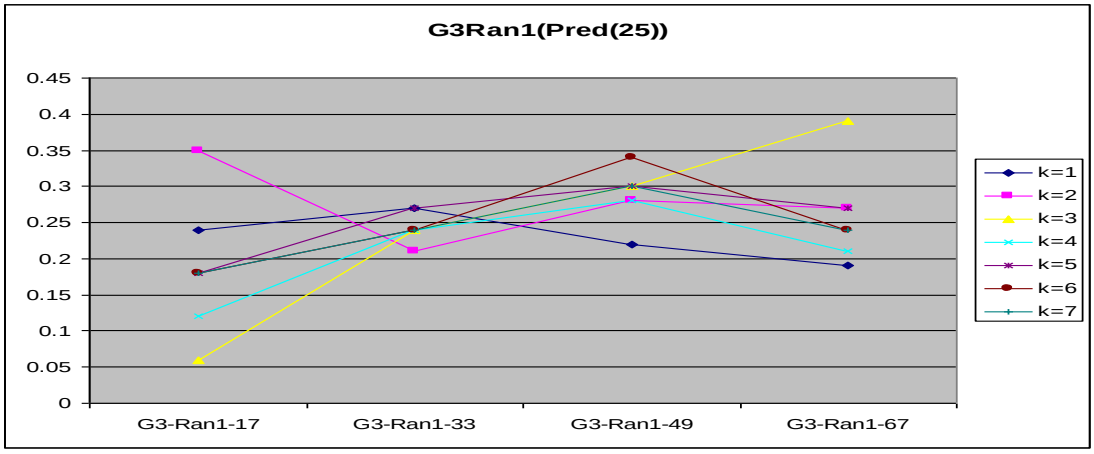


Figure 4.15: The Pred (25) results on Group3Ran1

4.7 Conclusions

The main finding of this investigation is that no reliable guidance can be given regarding the number of analogies that should be employed in making a prediction. In some cases there is a tendency for the data to converge as k increases, whilst in others it diverges. Most of the graphs seem to suggest that the data has a big influence in the calculation of the MMRE and also the PRED(25) values.

In addition, the results do not give any confidence that increasing the size of the dataset results in more accurate predictions. In some cases the smallest set (17 cases) is the least accurate, but in others it is the most! The larger datasets (with 33, 49 and 67 values) tend to

gravitate towards each other more and display less volatility, but their relationship to each other is not always predictable.

It was also found that outliers in the form of large or small values could possibly affect these predictions. Related to this is the distribution of data within the dataset – understandably, those with a more even spread of data tended to produce lower MMRE values. The quality of the data set seems to plays a major role in the precision of the prediction.

Another important result of this study is the relationship between the features used and the distance calculation. In this study only 8 features were employed, and only 2 of these were numeric - Functional Size, Adjusted Functional Points (Effort is also numeric but is not employed in the distance measure as it is the value which is being predicted) and the rest is categorical. Again the characteristics of the dataset could influence prediction accuracy because categorical data contributes either 1 or 0 to the distance calculation depending on whether there is a match or not. As a consequence the numeric values tend to dominate the distance calculation, resulting in cases which are arguably slightly poorer matches being ranked higher than apparently better ones.

This investigation motivates the need for further exploration in terms of dataset characteristics. Therefore, the next chapter describes an investigation that was initially aimed at characterising the dataset with a view to mapping dataset characteristics to technique to provide a way forward for practitioners.

5 A Pragmatic Cost Estimation Strategy

5.1 Introduction

Through a systematic literature review, industrial survey, and experimentation it was found that there was still a lack of information about how to carry out cost estimation. Although there are a lot of techniques available, the literature appears to be unhelpful for practitioners to apply a strategy upon which they can rely. Most of the literature still focuses on the best techniques rather than which is the best strategy to apply.

The aim of this chapter is to demonstrate a strategy that can be applied by those charged with the task of cost estimation within organisations. Rather than trying to argue that one strategy is best, it draws on results that demonstrate that there is not a universally good approach and that cost estimation methods vary according to context, data characteristics etc.. Although the results of the previous chapter suggest that the effectiveness of prediction techniques was hampered by several factors including the characteristics of underlying dataset, this chapter argues that trying to base an estimate on data characteristics is very hard. This due to the fact that there are so many dimensions to consider, particularly for those who are not highly skilled in data analysis.

The basics of the approach are to apply the advice contained within Shepperd and MacDonell's validation framework (Shepperd and MacDonell, 2012) and other sound pieces of advice such as looking at the value of Pred and MMRE and investigating the models produced by the estimators. The approach follows a series of steps and continues while the analysis is still a sensible thing to do; otherwise stopping, and again giving advice as to what the practitioner should do. The chapter illustrates this approach with examples drawn from the Desharnais dataset.

5.2 The Challenges of Cost Estimation

Accurate cost estimations are crucial for better project planning, monitoring and control. In industry the stress of getting better estimates is usually high in demand. Over the last three decades a variety of estimation techniques have been developed and investigated to provide improved estimates. Despite intense research, given the diversity of estimation techniques it is difficult to assess which techniques will be the best in any given circumstance.

There are several challenges in finding the best techniques for cost estimation such as lack of knowledge of prediction techniques, various accuracy measures, complexity of integrated solutions, data characteristics etc.. Some of the challenges discussed further:

Lack of knowledge of prediction techniques Most researchers are familiar with only one or two techniques, not all the techniques that exist in the market. Furthermore, there are many prediction techniques which offer no clear guidance as to which should be chosen.

Various accuracy measures There are several accuracy measures in literature for assessing the prediction techniques. However, the accuracy measures that are reported may give invalid results or favour to certain types of data. In order to assess technique appropriateness, practitioners should understand the strengths and weaknesses of the accuracy measures and use them to guide the results.

Data Characteristics Based on literature, (Mendes et al. (2003), Shepperd (2001), Kitchenham et al. (2004) and Lokan et al. (2008)) it was reported that the characteristics of the dataset could be the factor as there was an indication that when one technique predicts poorly, one or more of the other tend to perform significantly better. However there is no strong recommendation in the literature which techniques should be chosen in which circumstance. The difficulty in characterising the data could be a factor why there is no strong recommendation for practitioners. The following sections will demonstrate in more detail why it is very difficult to characterise a dataset.

5.3 Application Methodology

Previous experiment results and Shepperd (2001) show that data set characteristics could have a strong influence on the choice of the techniques to be employed in obtaining effort estimates. In this study Desharnais dataset was selected to identify the characteristics of the dataset. Initial investigation of this study showed that there was a lack of consistency in the characteristics of each dataset attribute. This lack of consistency in characteristics made it difficult to define the type of dataset for future analysis. Therefore, this study provides an essential platform to develop cost estimation strategies to support future research in providing reliable cost estimation results.

5.3.1 Dataset

The investigations in this chapter are all based upon the Desharnais dataset (Desharnais 2011). The Desharnais dataset is a publicly available dataset in Promise repository. The dataset comprised 81 software projects derived from a Canadian software house. Although types of projects that used in Desharnais is not exclusively web-based, this is not relevant for the purposes of this study as the approach is relevant for any effort-related dataset.

The dataset comprised 12 (one dependent and eleven independent) features summarised in the table below. Four out of 81 projects contained missing values and hence were excluded from further investigation. Table 5.1 lists the features that were used in this study.

Name	Description
Project Name	*Numeric identifier*
Effort	*Measured in hours*
ExpEquip	*Team experience in years*
ExpProjMan	*Project Managers experience in years*
Trans	*Number of transactions processed*
Entities	*Number of entities*
PointsAdjust	*Adjusted Function Points*
Envergure	*Scale*
Length	*Actual project schedule in months*
Language	*Programming language used*
PointsNonAdjust	*Unadjusted Function Points*
YearFin	*Year of completion*

Table 5.1: Desharnais dataset description

This experiment focuses on 10 out of 12 attributes in the Desharnais dataset that could potentially affect the prediction values. The attributes are Team experience, ProjectManagerExperience, Length, Effort, Transactions, Entities, PointsAdjust, Envergure, PointsNonAdjust and Language. The remaining attributes were discarded (Project name and Year of completion), for the following reasons: Project Name (number of case or index) and Year of completion (irrelevant).

5.3.2 Estimation Techniques

The estimation techniques that were studied in this experiment were based on the popular techniques that were discovered in the systematic literature review. The techniques were:

Linear Regression (LR)

Radial Basis Function Network (RBFN)

Support Vector Regression- Poly (SVRP)

Reduced Error Pruned Trees (RepTrees)

Case Based Reasoning with 1-analogy (CBR-1)

Case Based Reasoning with 2-analogy (CBR-2)

Case Based Reasoning with 3-analogy (CBR-3)

Bayesian Networks was excluded as this technique had been repeatedly reported as giving the worst result when compared to simpler models (Mendes and Mosley 2008). An overview of the techniques explored may be found in section 2.3.

5.3.3 Dataset Characteristics

Based on systematic literature review, it was found that different dataset characteristics were reported to influence the result of cost estimation; however, studies of dataset characteristics in this area by previous researchers were incomplete ((Shepperd 2001), (Mendes 2003), (Mendes et al. 2007) and (Lokan et al. 2008)). Therefore, this study gathered the evidence of characteristics that been mentioned in systematic literature review and then defined them. These characteristics were identified as part of the answer to question RQ2b in the previous chapter.

> RQ2b : What are the main characteristics of datasets for web applications? Do they affect the results?

Normal means the dataset were normally distributed. In order to identify a *normal* group, a histogram may be used as they give a good indication of the dataset distribution.

Kurtosis: “A measure of the "peakedness" or "flatness" of a distribution. A kurtosis value near zero indicates a shape close to normal. A negative value indicates a distribution which is more peaked than normal, and a positive kurtosis indicates a shape flatter than normal. An extreme positive kurtosis indicates a distribution where more of the values are located on the edges of the distribution rather than around the mean. A kurtosis value of +/-1 is considered very good for most psychometric uses, but +/-2 is also usually acceptable. “(SPSS 2011)

The degree of kurtosis is defined as:

$$\frac{\sum_{i=1}^{N}\left(Y-\bar{Y}\right)^4}{\left(N-1\right)s^4}-3$$

where $\bar{Y}$ is the mean, s is the standard deviation, and N is the number of data points.

If the kurtosis statistic sign is positive, we can assume that the distribution is leptokurtic (too tall). Alternatively, if the kurtosis statistic is negative, we know that the distribution of data is platykurtic (too flat).

Skewness: "The extent to which a distribution of values deviates from the symmetry around the mean. A value of zero means the distribution is symmetric, while a positive skewness indicates a greater number of smaller values, and a negative value indicates a greater number of larger values. Values for acceptability for psychometric purposes (+/-1 to +/-2) are the same as with kurtosis." (SPSS 2011)

The degree of skewness is defined for $Y_1, Y_2, \ldots Y_N$ as : $\left(\frac{\sum_{i=1}^{N}(Y-\bar{Y})^3}{(N-1)s^3}\right)$

where $\bar{Y}$ is the mean, s is the standard deviation, and N is the number of data points.

Outlier: An outlier may be defined as an outlying observation which appears to deviate markedly from other members of the sample in which it occurs **(**Grubbs 1969). It has been shown that the presence of a single outlier can greatly alter the results (Altman 1991) and (James 1993). In this investigation the R-project (R 2011) is used to determine the exact values that can be labeled as outliers besides the number of outliers that is reported in each distribution. In R, outliers are defined as "finds values with the largest difference between it and the sample mean"

Collinearity/ multicollinearity: The number of variables that exhibit significant correlations with other independent variables out of the total number of independent variables (Shepperd and Kadoda 2001).

Heterogeneous/ homogeneous: Homogeneous relates to the same set of distribution or is uniform in character composition. Lokan and Mendes (2009) studied the issue of homogeneity by using chronological splitting to compare cross company and single company effort predictions. Their results showed that single company predictions significantly outperformed cross company ones. They also concluded that single company projects are more homogeneous than cross company projects.

This study followed Shepperd and Schofield in terms of the minimum number required for cost estimation. According to their experience, data sets of 10-12 projects already provided a stable basis for estimation (Shepperd and Schofield 1997). In that case, this study investigated the characteristics of "small" when the dataset groups contained 15 numbers of data and classified them as "big" when the groups contained 50 numbers of data.

The study divided the Desharnais dataset into 14 subsets of data. Based on the approach employed by Shepperd and Schofield (1997), the data has been divided to small (15) and big (50) groups giving 7 small datasets and 7 big datasets.

Dataset-(odd) denotes the small sets and dataset-(even) the big ones. There are overlap between small dataset and the big group of dataset. This approach can provide a better picture into the applicability of the techniques as well as more confidence in the results.

Studies on collinearity/ multicollinearity and heterogeneous/ homogenous are to be left for future studies . The existing dataset characteristics are complex enough as they stand.

5.3.4 Difficulties in characterising the dataset

Based on the results in the previous chapter, the thesis attempts to investigate which dataset characteristics are favourable for certain techniques. The characteristics that were identified were used to generalise each dataset group. However, these attempts failed due to the difficulties in characterising the dataset. This section will demonstrate why it is difficult to characterise a dataset.

The skewness, kurtosis values and number of outliers of each attribute in the subset of each dataset generated are reported in Table 5.2. The values highlighted in red indicate that they have extreme skewness and kurtosis values.

	Dataset-1			Dataset-2		
	Skewness	Kurtosis	#Outliers	Skewness	Kurtosis	#Outliers
TeamExp	-0.381	-0.764	0	0.094	-1.291	0
ManagerEXp	-0.147	-1.546	0	0.143	-0.353	0
Length	1.067	1.009	0	1.791	5.486	1
Effort	-0.027	0.464	0	0.311	-0.383	0
Transactions	0.801	0.366	0	2.25	6.806	3
Entities	1.566	2.161	2	1.503	2.323	4
PointsAdjust	0.422	-0.721	0	1.469	3.596	2
Envergure	-0.067	-0.455	0	-0.361	-0.413	0
PointsNonAdjust	0.684	-0.283	0	1.300	2.311	1
	Dataset-3			**Dataset-4**		
	Skewness	Kurtosis	#Outliers	Skewness	Kurtosis	#Outliers
TeamExp	-0.382	-0.764	0	0.094	-1.291	0
ManagerEXp	-0.147	-1.546	0	0.143	-0.353	0
Length	2.282	6.492	1	2.109	5.811	2
Effort	3.086	10.995	1	4.822	29.434	1
Transactions	2.015	4.651	1	2.815	9.679	4
Entities	1.441	1.431	2	1.432	1.923	4
PointsAdjust	1.992	5.078	1	2.499	8.896	3
Envergure	-0.069	-0.029	0	-0.390	-0.449	0
PointsNonAdjust	1.889	4.281	1	2.399	8.316	2
	Dataset-5			**Dataset-6**		
	Skewness	Kurtosis	#Outliers	Skewness	Kurtosis	#Outliers
TeamExp	-0.087	-1.477	0	0.922	-1.322	0
ManagerEXp	1.045	1.739	0	0.277	0.257	0
Length	-0.038	-0.752	0	-0.026	7.491	2
Effort	0.168	-1.478	0	-0.272	-1.037	0
Transactions	0.957	0.330	1	0.682	18.022	1
Entities	2.491	7.463	1	3.254	3.307	5
PointsAdjust	1.218	0.767	2	2.059	9.305	1
Envergure	0.211	-0.906	0	1.855	-0.728	0
PointsNonAdjust	1.642	2.333	2	1.871	5.997	2
	Dataset-7			**Dataset-8**		
	Skewness	Kurtosis	#Outliers	Skewness	Kurtosis	#Outliers
TeamExp	-0.767	0.367	0	0.066	-1.262	0
ManagerEXp	-0.593	-1.189	0	0.088	-0.280	0
Length	0.634	0.132	0	1.659	5.090	1
Effort	1.130	1.225	2	1.751	4.334	4
Transactions	0.665	0.159	0	1.828	4.697	3
Entities	1.313	1.088	2	1.424	2.221	4
PointsAdjust	0.485	-0.752	0	1.183	2.053	3
Envergure	-0.532	0.096	0	-0.568	-0.219	0
PointsNonAdjust	0.653	-0.564	0	1.001	0.947	1

Table 5.2.1: Skewness, kurtosis and outlier values for each attribute in Dataset-(1-8)

	Dataset-9			Dataset-10		
	Skewness	Kurtosis	#Outliers	Skewness	Kurtosis	#Outliers
TeamExp	0.122	-1.548	0	0.520	-1.117	0
ManagerEXp	0.460	-1.642	0	0.644	0.767	0
Length	0.842	0.098	0	2.045	5.437	3
Effort	0.910	0.740	1	1.605	2.309	3
Transactions	-0.401	-0.469	0	2.004	5.762	2
Entities	0.036	-1.457	0	1.237	1.207	1
PointsAdjust	-0.373	-0.983	0	1.014	1.275	1
Envergure	0.224	-0.548	0	0.041	-0.339	0
PointsNonAdjust	-0.421	-0.981	0	0.759	-0.041	1
	Dataset-11			**Dataset-12**		
	Skewness	Kurtosis	#Outliers	Skewness	Kurtosis	#Outliers
TeamExp	0.350	-1.255	0	0.444	-1.224	0
ManagerEXp	0.217	-1.823	0	0.604	0.632	0
Length	1.663	2.546	2	1.990	4.274	4
Effort	2.081	3.946	2	1.915	3.705	4
Transactions	-0.093	-0.241	0	2.704	9.369	3
Entities	2.167	4.779	2	1.135	0.832	1
PointsAdjust	0.886	0.539	0	2.026	6.376	2
Envergure	1.151	1.855	1	0.041	-0.339	0
PointsNonAdjust	1.240	2.170	1	1.833	5.460	2
	Dataset-13			**Dataset-14**		
	Skewness	Kurtosis	#Outliers	Skewness	Kurtosis	#Outliers
TeamExp	0.206	-1.261	0	-0.086	-1.285	0
ManagerEXp	-0.173	-1.213	0	0.170	-0.140	0
Length	1.772	3.891	1	1.233	2.512	1
Effort	0.564	-0.692	0	1.158	1.285	1
Transactions	1.711	3.843	1	0.969	0.416	0
Entities	2.144	5.318	2	0.990	0.454	2
PointsAdjust	1.718	4.253	1	0.633	0.005	0
Envergure	-0.781	0.082	0	-0.467	0.294	2
PointsNonAdjust	1.650	3.996	1	0.700	-0.067	0

Table 5.2.2: Skewness, kurtosis and outlier values for each attribute in Dataset-(9-14)

Based on descriptive analysis of each dataset in Table 5.2.1 and Table 5.2.2, it was found that each attribute demonstrated a totally unique relationship among itself. The analysis on Dataset-10, shows that the kurtosis values of transactions indicates a values of 5.762. On the other hand the kurtosis value for PointsAdjust was only 1.207. The other descriptive analysis found that in Dataset-2, while the effort attributes indicated acceptable values for skewness, kurtosis and no outliers, the other attributes seemed to have extreme skewness, extreme kurtosis and a high number of outliers. Due to these different dimensions of extreme values, the characteristics were unable to be structured in this study.

From the observation on the groups of our subset dataset, the cleanest dataset was **Dataset-9**. Although this subset represents the original dataset which is skewed, the skewness and kurtosis value were (+/-1 to +/-2) with only one outlier. Meanwhile, the messiest dataset were **Dataset-4**. This is because this subset demonstrated the highest skewness and kurtosis for more than one attribute.

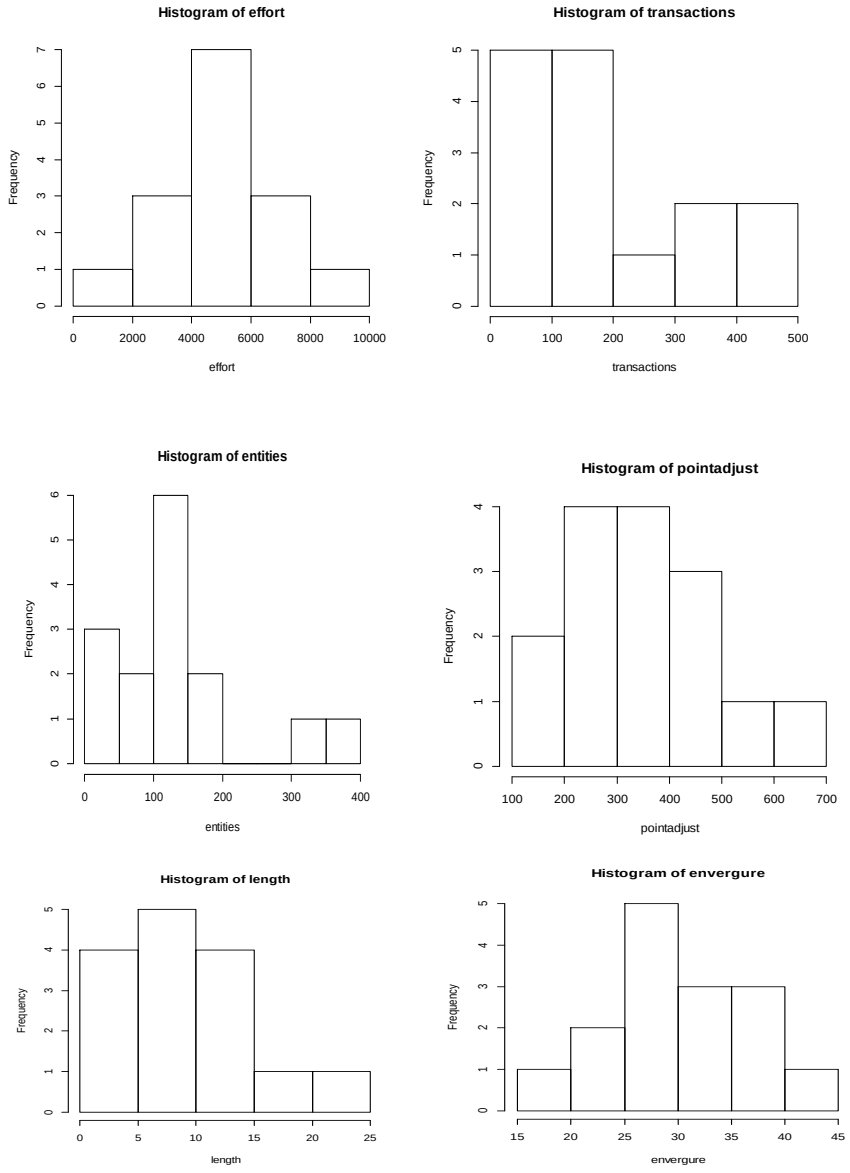


Figure 5.1: Histogram of some of attributes in *Dataset-1*

Figure 5.1 shows some of attributes in Dataset-1 where there are different dimensions. Although the effort attributes seem to be distributed normally but the rest show a different pattern. Based on the histogram it is clearly show that each attribute seems to show different

characteristic. Further investigation reveals the relationships between the attribute and the effort in Dataset-1, Dataset-4 and Dataset-9.

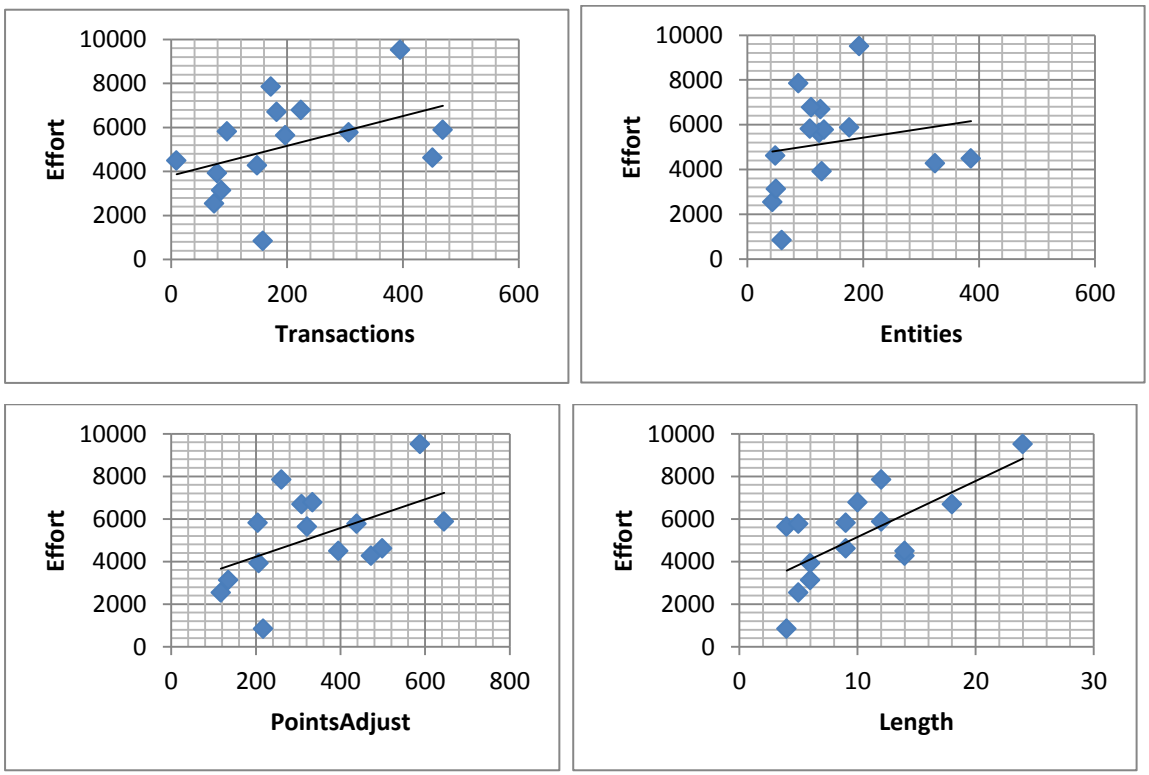


Figure 5.2: Graph of some of the attributes against effort in *Dataset-1*

Based on Figure 5.2, the graphs shows that when the dependent attribute (effort) is normal in the histogram, the independent data seems to tend towards linear relationships with some of the attributes but it is by no means clear. A similar pattern appears in Dataset-9, which been classified as a clean dataset.

However, for a messy dataset, which is illustrated in Figure 5.4, the situation is even worse and the relationships are even more clustered. In addition some of data points are far away from this cluster, which appears to be the outliers

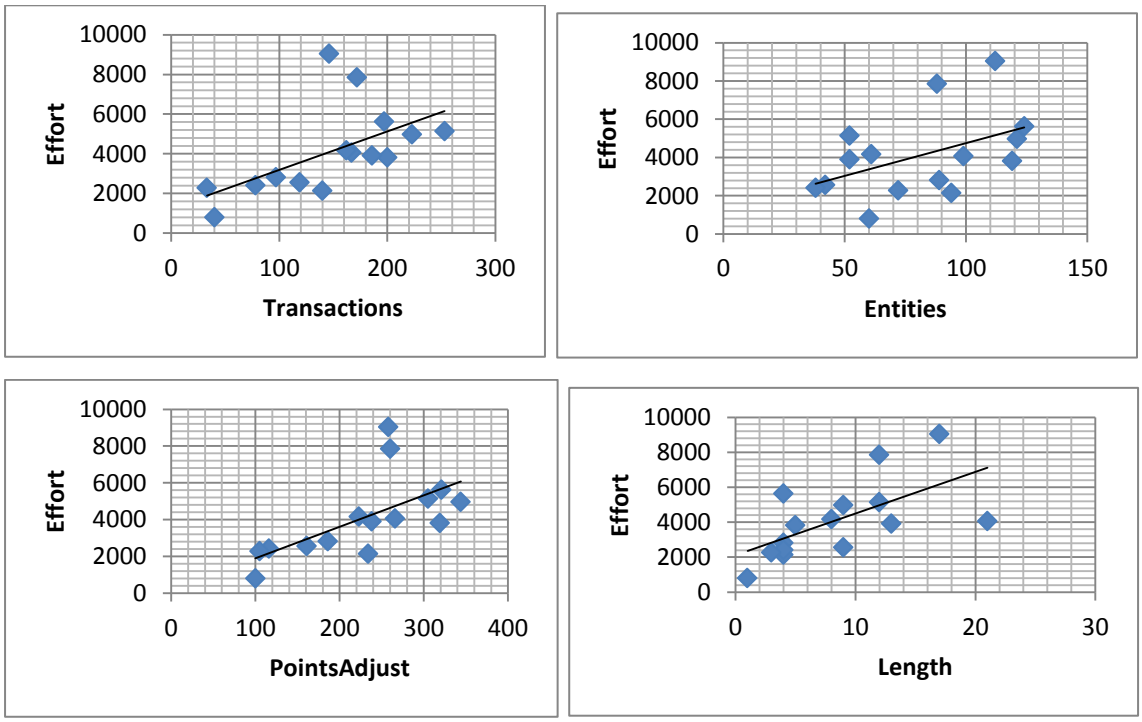


Figure 5.3: Graph of some of the attributes against effort in *Dataset-9*

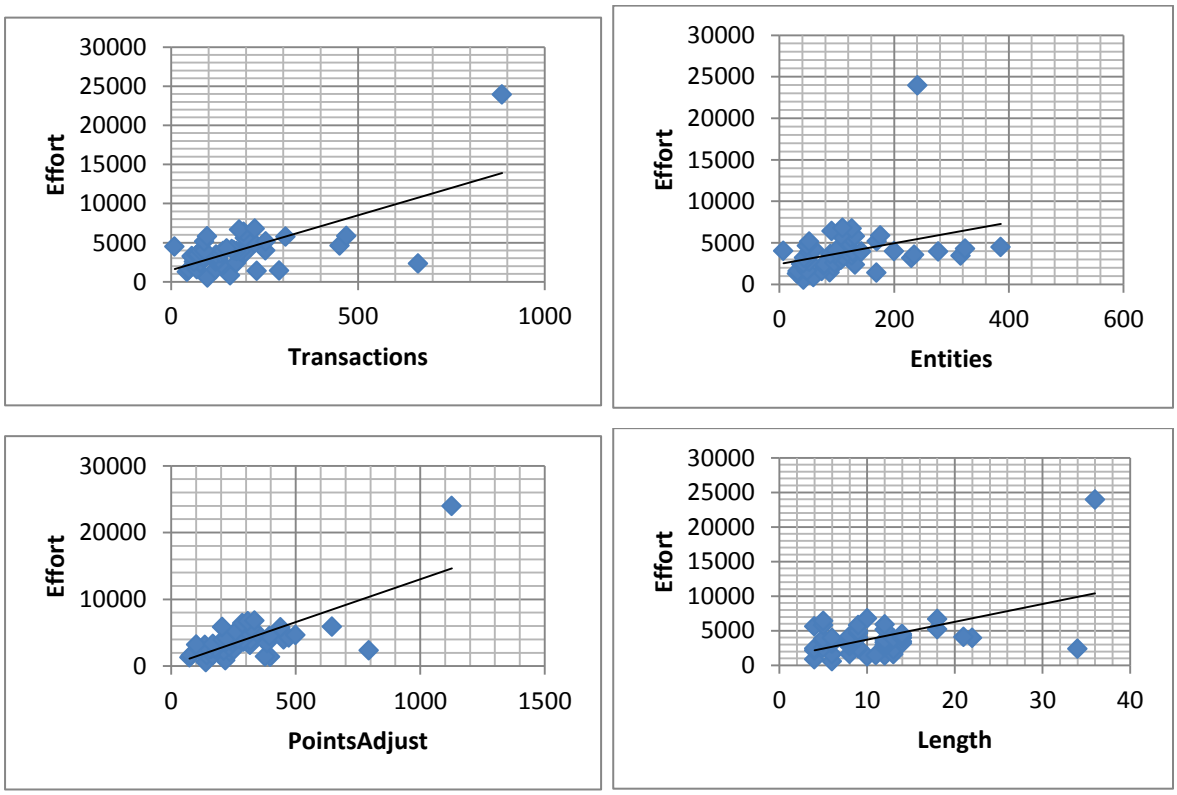


Figure 5.4: Graph of some of the attributes against effort in *Dataset-4*

It can be concluded that it is hard to structure the characteristics of a dataset as they have many attributes and there are no simple relationships between each element- as some

increase, other decrease etc. It is also difficult and impractical to explore all the possible combinations of attributes. Therefore, to address the difficulty of structuring dataset characteristics, a pragmatic strategy derived from Shepperd's validation framework is proposed.

5.4 A **Pragmatic Cost Estimation Strategy**

Based on previous discussions confirming the difficulties in structuring the characteristics of a dataset, a pragmatic cost estimation strategy was proposed. The cost estimation strategies in this study were identified and analysed using different strategies or advice from various reputable researchers. Analyses were performed based solely on literature input. The outcome of the analysis showed that the proposed strategies in this study would provide reliable results that could assist practitioners and researchers in cost estimation.

As practitioners and researchers in cost estimation need a way to progress with their data, the advice is to run a variety of estimation models on their dataset and then subject it to analysis based upon:

- ***Shepperd and MacDonell's Validation framework***

  Shepperd and MacDonell (2012) proposed a validation framework which was believed to provide a basis for the rigorous appraisal of results. However in this thesis, this validation framework could be used as part of a cost estimation strategy.

  By establishing this validation framework, researchers need to answer three fundamental questions:

  1. Does the prediction system outperform a baseline of random guessing?
  2. Is the difference of prediction techniques statistically significant?
  3. Is the effect size large enough to justify two prediction techniques relations in practice?

- ***Functional form of the model***

  Myrtveit and Stenstrud (2012) proposed that the evaluation procedure must include a theoretical justification of the functional form of the prediction model and that this ought to be a major evaluation and selection criterion.

The functional form of the model must as a minimum be *theoretically justifiable*, preferable and *testable*. To be theoretically justifiable the minimum criterion is that it is monotonous, meaning that the predicted effort must increase with software size.

- ***MMRE, Pred and MAE***

There are several criteria to evaluate model predictions (Conte et al. 1986). To gauge the accuracy of each estimated effort value, two values are calculated for each predictive model used for each dataset: the Mean Absolute Error[7] and Mean Magnitude of Relative Error[8] (MMRE) (Pickard et al. 1999).

Mean Magnitude of the Relative Error (MMRE) is calculated to indicate the relative amount by which the predictions over or underestimate the real value. The other advantage of MMRE besides it being commonly used in literature is it is independent of units. Independence of units means that it does not matter whether the effort is reported in workhours or workmonths. An MMRE will be, say 10% whatever unit is used (Foss et al, 2002).

Pred is a measure of what proportion of predicted values which have MRE are less than or equal to a specified value. In this thesis, Pred(25) was chosen as they are commonly used in most software effort prediction literature.

Pred(25) = (number of data where MRE<= 0.25)/ total amount of data

For example, Pred(25) = 50% means that half of the estimates are within 25 percent of the actual. Note that this is inverse to MMRE, where high Pred values are desirable.

Based on literature (Shepperd and Kadoda 2001) (Kitchenham et al. 2001) (Foss et al. 2003), it is understood in this study that MMRE and Pred have their limitations. Although these accuracy measures have limitations, in this

[7] The average absolute error for each dataset, where the absolute error is defined as |actual – estimate|.
[8] The average MRE for each dataset, where the MRE is defined as |actual – estimate|/actual.

study they are not left over but incorporate all these accuracy measures including the MAE to provide quality results. By understanding the limitations and strengths of this accuracy measure, it will help practitioners understand the quality of their results.

Although MMRE is a meaningful summary statistic and is necessary, these accuracy measure optimized by choosing a model that underestimates (Foss et al. 2003). It is possible, therefore, that the MMRE will favour a prediction technique that underestimates. However, the advantage of this accuracy measure as a means of assessing competing models.

Meanwhile Pred(25) is simply the percentage of estimates that are within 25% of the actual value. Therefore, Pred(25) is insensitive to the degree of estimates inaccuracy outside the specified measure. For example, a Pred(25) measure will not distinguish between a prediction system whose predictions deviate by 26% and one for whose predictions deviate by 260% (Kitchenham et. al, 2001).

The reasons for including these accuracy measures in our strategy are:

1) They are the most commonly and widely used accuracy measures in Software engineering

2) MMRE will help to measure the spread of the accuracy results in terms of variable z where z= abs(estimate-actual)/actual

3) Pred will help to measure the kurtosis of the accuracy results in terms of variable z. The percentage of Pred(25) will help practitioners to understand how much data has achieved 25% of actual value.

4) MAE will help the researcher to understand the distribution of MRE.

Conte et al. suggest that MMRE <=25% and PRED(25) >= 75% as a criterion for acceptable model performance. However, this minimum acceptable model

is hardly ever achieved in most literature. Therefore in this study, it is suggested that each experiment should have its own cut-off point. These cut-off points could be determined by project managers or researchers based on the best cut-off point for their estimation results. The strategy is to only use the techniques which beat the cut-off point. In this thesis the cut-off points are set to 50 as they represent at least half of the distribution which gives the best estimation according to the distribution of MRE values. Therefore the strategy will look at the best MAE results among all the competing techniques and MMRE <= 50% or Pred(25) >=50%. The use of all these accuracy measures is to have a better understanding of the quality of results between different prediction techniques.

Many recent papers are still basing their conclusions solely on the value of MMRE and Pred(25) without any statistical test (Aroba et al. 2008) (Huang et al. 2006) (Kumar et al. 2008). As a result, this study will also propose incorporating statistical tests as part of a pragmatic cost estimation strategy.

Based on the advice above the steps of this pragmatic cost estimation strategy are:

1. Test against means

The first step of the strategy is to test against means. Mean dataset values will be used as a baseline of random guessing. The reason for the mean to be chosen as an appropriate substitute for random guessing is that it is a fairly simple benchmark or baseline with which to compare the performance of other competing prediction techniques, as prediction system should outperform random guessing. Random guessing is simply to assign y-value of another case to the target case. If the prediction systems fail to do so it means it is actually not predicting in any meaningful sense. A standardised accuracy measure (SA) for prediction techniques was suggested by Sheppered and MacDonell (2012).

SA= (1-MARpi/MARpo)*100

where MARpo is the sample mean as an estimator and MARpi is the mean absolute residual of the model estimator.

The interpretation of SA is that the ratio represents how much better it is as a predictive model (pi) than the mean or random guessing (po). A value close to zero is discouraging and a negative value would be worrisome. The positive sign of SA means the predictive models are better than mean or random guessing. Meanwhile the negative sign is shows how bad the predictive models are against the mean as an estimator.

2. Test using MMRE, Pred and MAE.

The next step is to test using MMRE, Pred and MAE. Although there are some limitations using this accuracy measure in literature which was discussed in earlier sections, the strategy still adopts this measurement as part of a pragmatic strategy as they are referring to estimators of a function of the parameters related to the distribution of MRE values. The strategy can be used by applying a cut-off point. The cut-off points for this experiment are 50% for MMRE and Pred(25). The strategy is to only use the techniques which beat the cut-off point. MAE will be used in comparing the prediction techniques in each dataset group (see step 4).

3. Investigate the models.

To increase the confidence in the prediction model, strategies are required to investigate the models. The model should make sense, is valid and theoretically justifiable and preferably testable.

4. Perform significance testing

Once the model is understandable and sensible the strategies require significant testing. Statistically significant obtain in order to verify if the differences observed using summary accuracy measures were actual or due to chance, and also to check whether the absolute residuals for the techniques used came from the same populations (Corazza et al. 2011). The strategy will lead the researchers or practitioners to interpret their results with more confidence i.e. that the results produced were not due to chance.

5. ***Test effect size***.

This step is performed as the approach to null hypothesis testing has been criticised. The p values by statistical testing may not be informative if the sample size is large as even a small difference will be recorded as being significant. Statistical testing is important to draw firm conclusions on confidence levels; however, statistical significance does not present the effect size. Statistical significance only tells the researcher how likely it is an observed finding could have occurred by chance. One way to overcome this is to report the effect size. The effect size will give confidence in how meaningful the results are and not only based on how significant the results are. To address meaningfulness, researchers can report and interpret an effect size estimate.

Effect size is a name given to a group of statistics that measure the magnitude of a treatment effect. In many cases, effect size is a better measure of research outcomes than the significance level. This is because with large samples, one can observe statistically significant group differences even when only a tiny effect is present. Unlike significance tests, effect size indices are independent of sample size.

Effect-size estimates are metrics designed specifically to characterize results in more functional and meaningful ways by discussing the *magnitude* of an effect in addition to estimates of probability. The most commonly used effect size estimate is Cohen's *d* (Rosenthal, R. 1994). Cohen's *d* is computed by dividing the mean difference between groups by the pooled standard deviation. The effect size is just the standardised mean difference between the two groups. In other words for this research:

Effect Size = (Mean of absolute residual of Prediction Techniques – Mean of Values of sample) / Standard Deviation of mean as an estimator

To interpret the effect size in this thesis it closely follows the categories introduced by Cohen and used by Shepperd and MacDonell (2012) where small (≈0.2), medium (≈0.5) and large (≈0.8).

Therefore, the strategy which will accept the result is not a chance outcome if they are statistically significant and a medium or large effect size. However, if the results have only a

small effect size, then the results are significant but not interesting or not worth bothering about.

Based on this pragmatic cost estimation strategy, practitioners can choose to select any prediction models that are available to them. Furthermore, by using this pragmatic cost estimation strategy, practitioners can understand better what prediction techniques are actually estimating and have confidence in the results that are being produced by the prediction techniques.

5.5 Results

	LinearRegression			RBF Network			SVR-Poly			REPTrees			CBR								
													k=1			k=2			k=3		
	MAE	MMRE	Pred	MAE	MMRE	Pred	MAE	MMRE	Pred	MAE	MMRE	Pred	MAE	MMRE	Pred	MAE	MMRE	Pred	MAE	MMRE	Pred
Dataset-1	2160.8	64.0	0.26	1602.3	60.0	0.66	2087.7	49.0	0.26	1528.2	53.0	0.53	1795.2	46.0	0.40	1619.5	47.0	0.40	1652.9	53.0	0.40
Dataset-2	1048.6	36.0	0.50	1304.9	56.0	0.44	1267.5	44.0	0.44	1310.6	51.0	0.42	1247.8	36.0	0.40	1273.1	43.0	0.38	1061.5	38.0	0.52
Dataset-3	3152.3	60.0	0.13	2730.1	70.0	0.66	3285.7	73.0	0.26	2847.8	65.0	0.40	2835.4	56.0	0.40	2577.8	51.0	0.46	2614.7	53.0	0.40
Dataset-4	1658.1	51.0	0.42	1739.8	59.0	0.46	1335.9	39.0	0.38	1785.6	67.0	0.44	1429.9	33.0	0.46	1241.7	32.0	0.62	1337.5	39.0	0.58
Dataset-5	421.3	45.0	0.33	321.6	34.0	0.40	432.7	41.0	0.27	408.9	44.0	0.27	332.2	35.0	0.53	339.4	38.0	0.40	408.3	45.0	0.33
Dataset-6	807.4	45.0	0.46	862.1	57.0	0.42	823.8	41.0	0.42	791.8	51.0	0.50	787.4	40.0	0.50	738.8	42.0	0.46	633.1	40.0	0.60
Dataset-7	4100.0	101.0	0.26	2986.8	82.0	0.40	3191.2	80.0	0.33	3200.6	91.0	0.40	4604.1	93.0	0.20	3810.1	82.0	0.26	3481.9	77.0	0.26
Dataset-8	1834.8	51.0	0.48	2254.0	77.0	0.26	1563.1	39.0	0.46	1811.0	63.0	0.50	1960.2	44.0	0.36	1774.8	49.0	0.34	1915.4	54.0	0.44
Dataset-9	2088.2	49.0	0.53	1725.0	55.0	0.46	1896.3	52.0	0.33	1661.8	46.0	0.46	1968.9	66.0	0.40	1396.0	49.0	0.53	1529.2	54.0	0.46
Dataset-10	2542.0	70.0	0.34	2523.7	76.0	0.36	2400.2	65.0	0.26	2527.7	76.0	0.28	2554.5	59.0	0.32	2225.9	56.0	0.34	2032.7	50.0	0.34
Dataset-11	1556.3	45.0	0.46	3796.8	127.0	0.33	2088.1	53.0	0.20	4334.3	127.0	0.06	2527.0	67.0	0.33	2245.1	58.0	0.33	2670.4	66.0	0.26
Dataset-12	2867.8	72.0	0.22	3030.0	87.0	0.28	2307.1	67.0	0.30	3502.4	113.0	0.22	2258.8	53.0	0.36	2200.6	47.0	0.38	2187.5	51.0	0.36
Dataset-13	2488.3	131.0	0.26	1876.2	117.0	0.40	1978.9	91.0	0.13	1948.4	92.0	0.26	2526.3	80.0	0.20	2070.1	94.0	0.13	1702.6	71.0	0.26
Dataset-14	2676.3	64.0	0.28	2983.0	92.0	0.26	2417.5	48.0	0.28	2944.4	106.0	0.28	2812.9	46.0	0.40	2834..5	57.0	0.34	2824.6	74.0	0.32

Table 5.3: MAE, MMRE and Pred(25) result

*Highlighted values indicate the best results for MAE/MMRE/PRED

The results based on different characteristic groups of dataset against different types of techniques in terms of MAE, MMRE and Pred results are presented in Table 5.3. The best results in Table 5.3 are highlighted; however, to show better picture of this Table 5.4 is shown. Here the best results are included more than once in some cases when the results are too close to choose the best results. It was selected based on the next best results which did not exceed more than a 5% change.

Best Result

Group	MAE	MMRE	Pred
Dataset-1	REPTrees	CBR1	RBFN
	RBFN	CBR2	
Dataset-2	LR	LR	CBR3
	CBR3	CBR3	LR
Dataset-3	CBR2	CBR2	RBFN
	CBR3	CBR3	
Dataset-4	CBR2	CBR2	CBR2
Dataset-5	RBFN	RBFN	CBR1
	CBR1	CBR1	
Dataset-6	CBR3	CBR3	CBR3
		CBR1	
Dataset-7	RBFN	CBR3	RBFN
			REPTrees
Dataset-8	SVR-P	SVR-P	LR
			SVR-P
Dataset-9	CBR2	REPTrees	CBR2
		CBR2	LR
Dataset-10	CBR3	CBR3	RBFN
			CBR3
Dataset-11	LR	LR	LR
Dataset-12	CBR3	CBR2	CBR2
	CBR2	CBR3	CBR3
Dataset-13	CBR3	CBR3	RBFN
Dataset-14	SVR-P	CBR1	CBR1
		SVR-P	

Table 5.4: Best MAE, MMRE and Pred(25) results

It can be observed that there are different techniques which appear to be the best across different accuracy measures. There are three datasets out of 14 that achieved

the best results across MAE, MMRE and Pred which are Dataset-4, Dataset-6 and Dataset-11.

5.6 Application of Strategy

In this section the strategy that was introduced earlier will be applied here to further analyse the results. The steps in pragmatic cost estimation strategy are:

1. Test against means
2. Test using MMRE, Pred and MAE.
3. Investigate the models.
4. Perform significance testing
5. Test effect size.

The strategy follows a series of steps during the analysis while it is still a sensible thing to do. Otherwise stop from continuing and give advice to practitioners about what should they do.

1. Test against means

Here in this step, the prediction will test against means. This step will be used to examine how much better the predictive models are than random guessing, which also works as a notion of some fundamental baseline or benchmark. In this investigation, the mean of the datasets is used as the random guess.The strategy is to choose the SA which gives a positive sign where they are better than mean or random guessing.

	LR	RBFN	SVR-P	RepTrees	CBR-1	CBR-2	CBR-3
Dataset-1	-30.0	3.57	-25.6	8.03	-8.03	2.54	0.53
Dataset-2	17.49	-2.68	0.26	-3.12	1.81	-0.17	16.47
Dataset-3	-14.32	0.99	-19.1	-3.27	-2.82	6.51	5.17
Dataset-4	-1.37	-6.37	18.32	-9.17	12.57	24.1	18.2
Dataset-5	-8.87	16.8	-11.8	-5.64	13.9	12.3	-5.50
Dataset-6	10.95	4.91	9.13	12.66	13.15	18.51	30.16
Dataset-7	-51.91	-10.66	-18.23	-18.58	-70.58	-41.16	-29.0
Dataset-8	12.19	-7.86	25.19	13.33	6.19	15.06	8.33
Dataset-9	-28.78	-6.38	-16.94	-2.48	-21.46	13.90	5.69
Dataset-10	21.4	21.96	25.77	21.83	21.01	31.16	37.14
Dataset-11	55.71	-8.04	40.58	-23.33	28.09	36.11	24.01
Dataset-12	20.88	16.41	36.35	3.37	37.68	39.28	39.65
Dataset-13	-11.38	16.01	11.41	12.78	-13.08	7.33	23.78
Dataset-14	17.28	7.80	25.28	8.99	13.06	12.39	12.69

Table 5.5: Standardised accuracy measures (SA) results

Table 5.5 shows mixed results for different predictive models. In *Dataset-7* there is not a single technique which performs better than random guessing, while there are several dataset groups (Dataset-6, Dataset-10, Dataset-12 and Dataset-14) where all the techniques perform better than random guessing. In Dataset-8, RBFN is the only technique that performs worse than random guessing.

Although the results in Table 5.5 show that not all of them are actually predicting, since there are yields which have negative signs, there also certain predictive models which yield considerably better (1-56%) accuracy levels than random guessing.

If none of the techniques are better than the mean then there is very little point in proceeding any further. However, before giving any advice as part of a pragmatic strategy further investigation was carried out on the results of each prediction technique in Dataset-7 which showed none of the techniques were better than means. It was observed that the smallest data which was 847 appears to give by far the worst result in the entire prediction model. Table 5.6 shows the top 3 MRE across all predictions.

LR		RBFN		SVRP		RepTrees	
Effort	**MRE**	**Effort**	**MRE**	**Effort**	**MRE**	**Effort**	**MRE**
847	7.37	847	6.88	847	5.99	847	7.37
5880	1.23	3136	1.07	4494	0.83	3136	1.21
9520	1.15	4494	0.72	4277	0.74	4277	1.07

Table 5.6.1: Top 3 MRE results for (LR, RBFN, SVRP and RepTrees) techniques for *Dataset-7*

CBR1		CBR2		CBR3	
Effort	**MRE**	**Effort**	**MRE**	**Effort**	**MRE**
847	3.64	847	4.75	847	4.07
4494	2.33	5775	1.55	5775	1.03
5880	1.55	4494	1.1	3927	0.99

Table 5.6.2: Top 3 MRE results for (CBR1, CBR2 and CBR3) techniques for *Dataset-7*

Based on the results it can be shown that the actual effort of 847 demonstrated extremely high MRE across all predictions. Although the outliers were two different values, in this case the smallest value and the next smallest value had huge difference between them. Therefore it is not only the outliers which affect prediction techniques,

but the overall distribution of the dataset which is important in suggesting whether the dataset group is suitable for cost estimation.

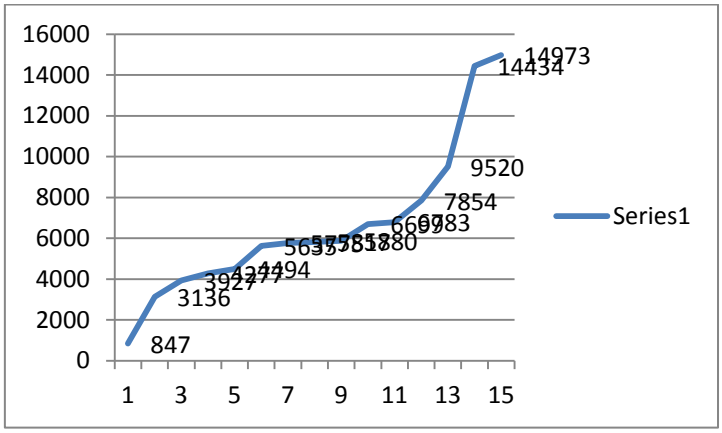


Figure 5.5: Actual effort for *Dataset-7*

In most prediction techniques this small actual effort could influence prediction techniques when there is no other data which is in the same range of actual effort. Based on Figure 5.5, the next actual effort after 847 is 3136. In this case even if the prediction techniques give the best prediction, it still gives a huge MRE.

For example:

Let's say actual effort 847 and predicted effort 3136.

MRE for actual effort 847= (3136-847)/847 = 2.70

Let's say actual effort 14434 and predicted effort 9520.

MRE for actual effort 9520= (14434-9520)/9520 = 0.51

It is clear that the data which has actual effort 847 gives huge MRE by defaults in all the prediction techniques. This will give relatively huge results for the MAE. The example of using outlier data which has actual effort 14434 shows that even outliers are not be able to predict worse by using the next nearest data. This could be the reason why there is not even one technique which performs better than random guessing in Dataset-7. As a piece of pragmatic advice if none of the techniques are better than the mean then the strategy should be:

1. Use Case selection techniques which identify and remove redundant and noisy projects.

2. Employ data analysts to improve the dataset by using data analysis techniques such as principle component analysis.

3. Clean the dataset using human experts based on the similarity of data before the use of prediction tools.

2. Test using MMRE, Pred(25) and MAE results

The red highlighted results in Table 5.7 indicate the MMRE results which have an MMRE of less than 50%, meanwhile in Table 5.8 the red highlighted results show Pred results which have Pred(25) more than 50%.

	LR	RBFN	SVRP	RepTrees	CBR-1	CBR-2	CBR-3
	MMRE	MMRE	MMRE	MMRE	MMRE	MMRE	MMRE
Dataset-1	64.00	60.00	49.00	53.00	46.00	47.00	53.00
Dataset-2	36.00	56.00	44.00	51.00	36.00	43.00	38.00
Dataset-3	60.00	70.00	73.00	65.00	56.00	51.00	53.00
Dataset-4	51.00	59.00	39.00	67.00	33.00	32.00	39.00
Dataset-5	45.00	34.00	41.00	44.00	35.00	38.00	45.00
Dataset-6	45.00	57.00	41.00	51.00	40.00	42.00	40.00
Dataset-7	101.00	82.00	80.00	91.00	93.00	82.00	77.00
Dataset-8	51.00	77.00	39.00	63.00	44.00	49.00	54.00
Dataset-9	49.00	55.00	52.00	46.00	66.00	49.00	54.00
Dataset-10	70.00	76.00	65.00	76.00	59.00	56.00	50.00
Dataset-11	45.00	127.00	53.00	127.00	67.00	58.00	66.00
Dataset-12	72.00	87.00	67.00	113.00	53.00	47.00	51.00
Dataset-13	131.00	117.00	91.00	92.00	80.00	94.00	71.00
Dataset-14	64.00	92.00	48.00	106.00	46.00	57.00	74.00

Table 5.7: MMRE result

	LR	RBFN	SVRP	RepTrees	CBR-1	CBR-2	CBR-3
	Pred	Pred	Pred	Pred	Pred	Pred	Pred
Dataset-1	0.26	0.66	0.26	0.53	0.40	0.40	0.40
Dataset-2	0.50	0.44	0.44	0.42	0.40	0.38	0.52
Dataset-3	0.13	0.66	0.26	0.40	0.40	0.46	0.40
Dataset-4	0.42	0.46	0.38	0.44	0.46	0.62	0.58
Dataset-5	0.33	0.40	0.27	0.27	0.53	0.40	0.33
Dataset-6	0.46	0.42	0.42	0.50	0.50	0.46	0.60
Dataset-7	0.26	0.40	0.33	0.40	0.20	0.26	0.26
Dataset-8	0.48	0.26	0.46	0.50	0.36	0.34	0.44
Dataset-9	0.53	0.46	0.33	0.46	0.40	0.53	0.46
Dataset-10	0.34	0.36	0.26	0.28	0.32	0.34	0.34
Dataset-11	0.46	0.33	0.20	0.06	0.33	0.33	0.26
Dataset-12	0.22	0.28	0.30	0.22	0.36	0.38	0.36
Dataset-13	0.26	0.40	0.13	0.26	0.20	0.13	0.26
Dataset-14	0.28	0.26	0.28	0.28	0.40	0.34	0.32

Table 5.8: Pred(25) result

Based on the results in Table 5.7 and Table 5.8, the techniques which are better than means and have less than 50% of MMRE or 50% higher for Pred(25) are presented in Table 5.9. The highlighted technique in red shows the best MAE results.

Better than Means, MMRE(<50%) or Pred(25) > 50%							
Dataset-1		RBFN		RepTrees		CBR-2	CBR-3
Dataset-2	LR		SVRP		CBR-1		CBR-3
Dataset-3		RBFN					
Dataset-4			SVRP		CBR-1	CBR-2	CBR-3
Dataset-5		RBFN			CBR-1	CBR-2	
Dataset-6	LR		SVRP	RepTrees	CBR-1	CBR-2	CBR-3
Dataset-7							
Dataset-8			SVRP	RepTrees	CBR-1	CBR-2	
Dataset-9						CBR-2	
Dataset-10							CBR-3
Dataset-11	LR						
Dataset-12						CBR-2	
Dataset-13							
Dataset-14			SVRP		CBR-1		

Table 5.9: Techniques which give results better than Means and MMRE(<50%) or Pred(25) (>50%)

Table 5.9 shows "a basket of techniques" which have survived after two steps of the pragmatic cost estimation strategy. All these techniques are better than the mean and have MMRE<50% or Pred(25) >50%. The ones highlighted in red shows the techniques which also appear to give the best MAE. After Step 2 in the pragmatic cost estimation strategy, it was found that SVRP and CBR were more favourable in most datasets.

As a piece of pragmatic advice if none of the techniques are better than the mean and MMRE <50% or Pred(25) >50 then the strategy should be:

1. Use all the advice that has been given earlier which is to use case selection techniques, employ data analysts or clean the dataset using human experts.

2. To get better MMRE and Pred(25) eliminate the data which is causing the problem as part of the cleaning strategy.

3. Investigate the models

The remaining techniques which are better than means and MMRE < 50% or Pred(25) >50% will be investigated further in terms of their model. As mentioned earlier this strategy will investigate the models to check whether they are theoretically justifiable and most importantly whether the models make sense. This criteria needs to be fulfilled before it is used as an acceptable prediction model. The models that will be investigated further based on fulfilment of strategies on step 2 are:

Linear Regression – Dataset-2 or Dataset-11

RBFN – Dataset-5

SVRP- Dataset-8

CBR- Dataset-4

RepTrees – Dataset-1

This dataset yields better results than mean, best MAE, MMRE<50% or Pred>50%. Therefore those dataset models will be investigated further in this step.

Linear Regression

Dataset-11:

Effort = 189.5066 * Length -30.282 * Transactions + 37.02 * PointsNonAdjust -1001.1841

The Linear Regression model incorporates most important attributes which gives a sensible model. The model appears to be good because it grows with length and PointNonAdjust; however it is still not that good model as it gets smaller as Transactions grow.

RBFN model

Dataset-5:

Effort = -205.1876 * pCluster_0_0 + 205.1879 * pCluster_0_1 + 1101.4097

Based on the RBFN algorithm to construct the model, the classifiers conduct a cluster analysis on the training data set and allocate one hidden unit for each cluster (Oyang et al 2005). The algorithm differs by the clustering algorithm employed and how the parameters of the RBF network are set. Based on this definition and the model that is generated it appears that the RBFN model is hard to understand, especially when the model incorporates hidden units and has different cluster algorithms (a similar point was raised in the literature review about the difficulty on understanding the models created by techniques such as neural nets). However, based on our criteria the model should be theoretically justifiable. Therefore, this model failed to be considered as an acceptable model as it could not be explained to the end user and it was not an understandable or sensible model.

SVRP model

Dataset-8:

weights (not support vectors):

- 0.0141 * (normalized) TeamExp
+ 0.1057 * (normalized) ManagerEXp
+ 0.1151 * (normalized) Length
+ 0.0644 * (normalized) Transactions
+ 0.058 * (normalized) Entities
+ 0.0889 * (normalized) PointsAdjust
+ 0.0617 * (normalized) Envergure
+ 0.2024 * (normalized) PointsNonAdjust
- 0.1513 * (normalized) Language
+ 0.0484

Based on the model that is generated by WEKA using SVRP, it can be shown that the model only shows different weights on the entire attribute. The entire attribute was considered when building the model. However, discovering the model is quite tricky and involves the systematic investigation of the impact of every attribute. Again this raises more general point about the use of such "black box" techniques – they may produce good results but can they be trusted if you don't know the reason behind them. SVRP models which are considered as black-box models only explain the weights of the attribute and do not provide any other details pertaining to:-

a) Attribute relationships

b) Influence of attribute on the effort

c) Relevance of attribute

Based on the definition of black- box techniques practitioners should understand when they choose these techniques that they ignore the internal mechanism of the technique and focus solely on the outputs generated in response to selected datasets. Therefore, if practitioners opt to choose this technique then verification and validation of the output is necessary. This will help practitioners to verify whether the model has been built correctly, and validate whether the correct model has been built.

Based on our definition of acceptable model, these techniques seem hard to understand. It is also hard to explain to the end user the reasons behind the results.

Case Based Reasoning

A distance measure in CBR is the degree of similarity between two projects in terms of their effort drivers. Euclidean distance is most commonly used to solve this problem. However, in the Angel tool there is no other model to investigate further. The only information is the distance measures. Based on the distance measure for this dataset it was found that the prediction was based on the nearest data based on the distance between them. Although it is sensible to obtain the cases that show the measurement there is high computational expense to generate them. Having said that, however this technique is much easier to understand compared to the SVRP. In addition, it is sensible as the distance measures which is used, is based on historic

data. However, the rule of thumb for this step is to have theoretically justifiable models. Although the model does not appear in CBR the distance measure for each case shows which data has been used to obtain the prediction results. Therefore, this technique still can be considered as giving sensible predictions based on the distance measure.

RepTrees Model

Based on model investigation it was found RepTrees produces a model which uses only one node in most cases.

Model of REPTrees for *Dataset-1*:

REPTree

==========

: 5187.47 (10/3250995.65) [5/6585200.65]

Size of the tree : 1

Based on the REPTrees model it was found that for this dataset a tree can be built with just one node- the value of 5187.47, which means it is going to guess this every single time. However in our experiment an n-fold cross validation was used. The number of n depends on the size of the dataset. Slightly different results will be obtained as it builds a new model each time (i.e. creates a tree using 14 data items, then uses it to predict the 15^{th}).

Further investigation into the model that was created by REPTrees shows that the node value of 5187.47 referred to the mean of actual effort in *Dataset-1*. This is actually a default value that was given by WEKA for REPTrees which failed to generalise decision trees.

REPTrees clearly failed to build a model for this dataset. The reason for this could be the appropriateness of the dataset that was used. The reason REPTrees could not generate a true model for this experiment was because the dataset consisted of only numerical attributes. By definition REPTrees is much more appropriate for datasets which have categorical features. The other reason for REPTrees failed to build a model is the dataset that was used in this study was small. In a small dataset

REPTrees should be avoided. REPTrees requires a large dataset as the data has to split into three (training, validation and testing) data. Although one disadvantage of REPTrees is that it needs a large amount of data, the advantage of this results in more accurate classification trees (Quinlan, 1987). Therefore, when the dataset failed to generate a tree using REPTrees in WEKA, it predicted one node which was the mean of distribution. By default, when using cross validation this happens to give better results than the mean.

As part of the pragmatic advice for this step:

1. Practitioners could use linear regression if the model seems sensible for individual cases.

2. Practitioners may also use RBFN and SVRP only if they understand what they are actually estimating.

3. Practitioners should refrain from using RepTrees for numerical dataset as it failed to generalise a prediction model.

4. Perform Significance Testing

The next step in pragmatic cost estimation strategy is to perform significance testing to see if any of the techniques are better than others. To address this, each dataset group was tested across all prediction techniques. In particular the following null hypothesis was formulated:

Hn0: All prediction techniques provide estimates that are not significantly different.

While the null hypotheses can be rejected with relatively high confidence, it is possible to formulate an alternative hypothesis:

Ha0: Among the prediction techniques there is one that provides estimates which are (significantly) better than others.

Thus, specific null hypotheses formulated according to the group:

HnX: All prediction techniques in (***Dataset-X***) provide estimates that are not significantly different.

X is the index of the dataset group

The related alternative hypotheses will be shown individually in Table 5.10. The R-Project was used to generate the results for statistical testing using Wilcoxon Two Sample t-test. The attribute used for this test was absolute residuals against each technique. Paired absolute residuals are used, since they are less vulnerable to bias than the magnitude of relative error (Shepperd and Kadoda, 2001).

Dataset-1	**RBF Network**	**SVR-Poly**	**REPTrees**	**CBR**		
				k=1	**k=2**	**k=3**
Linear Regression	0.1607	0.8702	0.1261	0.4185	0.2328	0.2997
RBF Network	-	0.116	1	0.34	0.713	0.6482
SVR-Poly	-	-	0.09753	0.5068	0.2854	0.2997
REPTrees	-	-	-	0.3614	0.5949	0.6783
CBR-1	-	-	-	-	0.5614	0.6481
CBR-2	-	-	-	-	-	0.9669

Hn1: All prediction techniques in (***Dataset-1***) provide estimates that are not significantly different.

Dataset-2	**RBF Network**	**SVR-Poly**	**REPTrees**	**CBR**		
				k=1	**k=2**	**k=3**
Linear Regression	0.3683	0.2034	0.1669	0.6124	0.1822	0.9972
RBF Network	-	0.7174	0.7071	0.7123	0.7538	0.4586
SVR-Poly	-	-	0.9204	0.6221	0.9698	0.1392
REPTrees	-	-	-	0.4713	0.904	0.1546
CBR-1	-	-	-	-	0.6075	0.6003
CBR-2	-	-	-	-	-	0.1777

Ha2: All prediction techniques in (***Dataset-2***) provide estimates that are not significantly different.

Dataset-3	**RBF Network**	**SVR-Poly**	**REPTrees**	**CBR**		
				k=1	**k=2**	**k=3**
Linear Regression	0.08143	1	0.4363	0.34	0.2017	0.2169
RBF Network	-	0.1485	0.6529	0.4067	0.6529	0.461
SVR-Poly	-	-	0.2169	0.3194	0.137	0.1873
REPTrees	-	-	-	0.8357	0.7437	0.8381
CBR-1	-	-	-	-	0.5612	0.6186
CBR-2	-	-	-	-	-	0.9674

Hn3: All prediction techniques in (***Dataset-3***) provide estimates that are not significantly different.

Dataset-4	**RBF Network**	**SVR-Poly**	**REPTrees**	**CBR**		
				k=1	**k=2**	**k=3**
Linear Regression	0.7123	0.2995	0.4421	0.1439	0.03231	0.1777
RBF Network	-	0.1506	0.7174	0.125	0.01998	0.09525
SVR-Poly	-	-	0.1016	0.8659	0.3946	0.9917
REPTrees	-	-	-	0.04983	0.009527	0.04672
CBR-1	-	-	-	-	0.4819	0.7774
CBR-2	-	-	-	-	-	0.3142

Hn4: Among the prediction techniques in (***Dataset-4***) there is one that provides estimates (significantly) better than others

Dataset-5	**RBF Network**	**SVR-Poly**	**REPTrees**	**CBR**		
				k=1	**k=2**	**k=3**
Linear Regression	0.3669	0.8063	0.6041	0.2283	0.1873	0.6827
RBF Network	-	0.1607	0.1261	1	0.5125	0.1736
SVR-Poly	-	-	0.8381	0.1978	0.4864	0.6827
REPTrees	-	-	-	0.1294	0.2496	0.9025
CBR-1	-	-	-	-	0.5196	0.2447
CBR-2	-	-	-	-	-	0.2496

Ha5: All prediction techniques in (***Dataset-5***) provide estimates that are not significantly different.

Dataset-6	**RBF Network**	**SVR-Poly**	**REPTrees**	**CBR**		
				k=1	**k=2**	**k=3**
Linear Regression	0.4971	0.8659	0.8931	0.8442	0.8496	0.229
RBF Network	-	0.5625	0.361	0.3794	0.2398	0.04672
SVR-Poly	-	-	0.8388	0.702	0.7643	0.1338
REPTrees	-	-	-	0.8442	0.8767	0.2398
CBR-1	-	-	-	-	0.9313	0.4628
CBR-2	-	-	-	-	-	0.2868

Hn6: Among the prediction techniques in (***Dataset-6***) there is one that provides estimates (significantly) better than other.

Dataset-7	**RBF Network**	**SVR-Poly**	**REPTrees**	**CBR**		
				k=1	**k=2**	**k=3**
Linear Regression	0.4124	0.6041	0.4678	0.6481	0.7748	0.9025
RBF Network	-	0.6236	0.8063	0.1843	0.5949	0.2671
SVR-Poly	-	-	0.7748	0.4552	0.8063	0.713
REPTrees	-	-	-	0.2133	0.713	0.3892
CBR-1	-	-	-	-	0.3398	0.7089
CBR-2	-	-	-	-	-	1

Hn7: All prediction techniques in (***Dataset-7***) provide estimates that are not significantly different.

Dataset-8	**RBF Network**	**SVR-Poly**	**REPTrees**	**CBR**		
				k=1	**k=2**	**k=3**
Linear Regression	0.05487	0.4928	0.6867	0.9176	0.551	0.4023
RBF Network	-	0.006814	0.05065	0.1392	0.1251	0.1766
SVR-Poly	-	-	0.2806	0.5327	0.1121	0.09322
REPTrees	-	-	-	0.904	0.7827	0.6172
CBR-1	-	-	-	-	0.7748	0.5742
CBR-2	-	-	-	-	-	0.8469

Ha8: Among the prediction techniques in (***Dataset-8***) there is one that provides estimates (significantly) better than other.

Dataset-9	**RBF Network**	**SVR-Poly**	**REPTrees**	**CBR**		
				k=1	**k=2**	**k=3**
Linear Regression	0.7437	0.713	0.8063	0.4805	0.9025	0.9349
RBF Network	-	1	0.4124	0.8682	0.6236	0.7437
SVR-Poly	-	-	0.3453	0.9009	0.3453	0.461
REPTrees	-	-	-	0.3835	0.8381	0.7437
CBR-1	-	-	-	-	0.4066	0.5336
CBR-2	-	-	-	-	-	0.8063

Hn9: All prediction techniques in (***Dataset-9***) provide estimates that are not significantly different.

Dataset-10	**RBF Network**	**SVR-Poly**	**REPTrees**	**CBR**		
				k=1	**k=2**	**k=3**
Linear Regression	0.5695	0.7433	0.4023	0.9313	0.5556	0.4928
RBF Network	-	0.6766	0.9972	0.6516	0.1868	0.1411
SVR-Poly	-	-	0.5037	0.7907	0.3092	0.361
REPTrees	-	-	-	0.5979	0.1144	0.1037
CBR-1	-	-	-	-	0.438	0.438
CBR-2	-	-	-	-	-	0.978

Hn10: All prediction techniques in (***Dataset-10***) provide estimates that are not significantly different.

Dataset-11	**RBF Network**	**SVR-Poly**	**REPTrees**	**CBR**		
				k=1	**k=2**	**k=3**
Linear Regression	0.1873	0.1261	0.0128	0.042	0.4124	0.1064
RBF Network	-	0.6529	0.6236	0.8356	0.3046	0.8381
SVR-Poly	-	-	0.05553	0.3612	0.9674	0.7748
REPTrees	-	-	-	0.3835	0.0742	0.1607
CBR-1	-	-	-	-	0.5895	0.6781
CBR-2	-	-	-	-	-	0.8381

Ha11: Among the prediction techniques in (***Dataset-11***) there is one that provides estimates (significantly) better than others.

Dataset-12	**RBF Network**	**SVR-Poly**	**REPTrees**	**CBR**		
				k=1	**k=2**	**k=3**
Linear Regression	0.7538	0.3293	0.3362	0.3502	0.1217	0.08295
RBF Network	-	0.2134	0.5556	0.1868	0.0383	0.05065
SVR-Poly	-	-	0.04095	0.8876	0.372	0.5695
REPTrees	-	-	-	0.04303	0.006743	0.006673
CBR-1	-	-	-	-	0.4993	0.4841
CBR-2	-	-	-	-	-	0.9533

Hn12: Among the prediction techniques in (***Dataset-12***) there are more than one that provides estimates (significantly) better than others.

Dataset-13	**RBF Network**	**SVR-Poly**	**REPTrees**	**CBR**		
				k=1	**k=2**	**k=3**
Linear Regression	0.8381	0.6529	0.8381	0.2453	0.5668	0.7748
RBF Network	-	0.7748	1	0.2453	0.6236	0.6236
SVR-Poly	-	-	0.7748	0.2288	0.8381	0.5393
REPTrees	-	-	-	0.1842	0.6236	0.5393
CBR-1	-	-	-	-	0.4805	0.2132
CBR-2	-	-	-	-	-	0.3453

Hn13: All prediction techniques in (***Dataset-13***) provide estimates that are not significantly different.

Dataset-14	**RBF Network**	**SVR-Poly**	**REPTrees**	**CBR**		
				k=1	**k=2**	**k=3**
Linear Regression	0.8931	0.5237	0.9972	0.5463	0.5837	0.8067
RBF Network	-	0.4061	0.9533	0.3945	0.551	0.7277
SVR-Poly	-	-	0.4713	0.7277	0.9808	0.7071
REPTrees	-	-	-	0.5556	0.5327	0.7695
CBR-1	-	-	-	-	0.7801	0.4586
CBR-2	-	-	-	-	-	0.8067

Hn14: All prediction techniques in (***Dataset-14***) provide estimates that are not significantly different.

Table 5.10: Comparison of the p-values on absolute residuals using Wilcoxon two sample t-test

The statistical result that is reported in Table 5.10 led us to the following observations:

1. CBR-3 gives an estimate that is statistically significantly better than RBFN in *Dataset-6* (Ha6).

2. SVR-P gives an estimate that is statistically significantly better than RBFN in *Dataset-8* (Ha8).

3. LR gives an estimate that is statistically significantly better than CBR-1 in *Dataset-11* (Ha11).

4. CBR-2 and CBR-3 gives an estimate that is statistically significantly better than RBFN in *Dataset-12* (Ha12).

5. RepTrees which were discarded in an earlier step were not considered for this strategy although they showed statistically significant results.

The results based on statistically significant results eliminated most of the techniques which did not provide any statistical significance. Therefore, the identified techniques which are better than one or more techniques are Dataset-6 (CBR-3 was better than RBFN), Dataset-8 (SVRP is better than RBFN), Dataset-11 (LR was better than CBR-1) and Dataset-12 (CBR-2 and CBR-3 were better than RBFN). For the rest of dataset any approach is as valid as any other.

Dataset-6, Dataset-8, Dataset-11, and Dataset-12 remain to give preference relations. However based on standard validation frameworks, to get the preference relations between different predictive models, the effect size should be studied. Therefore, the next step will pay attention to effect size.

5. *Test effect size*

This step is the final step of pragmatic cost estimation strategy. As mentioned earlier the strategy is to get effect size and give preference relations. The reason for the use of effect size in practise is because the statistically significant test could be influenced by size. Even small differences may lead to give statistically significant results if the sample size is large. Therefore, effect size will give a better indication regardless of sample size. The effect size in this thesis follows closely the categories that were introduced by Cohen and used by Shepperd and MacDonell (2012), where small (≈0.2), medium (≈0.5) and large (≈0.8). The strategy will eliminate techniques which were statistically significant but had a small effect size as they were not important as mentioned in literature.

Example of this calculation shown below for *Dataset-1* :

	Dataset-1	**Mean Estimator**		**SVR-P**		**RepTrees**	
Size	**Effort**	**Predicted**	**|Pred-Act|**	**Predicted**	**|Pred-Act|**	**Predicted**	**|Pred- Act|**
217	847	5187	4340	1984	1137	4033.1	3186.1
117	2548	5187	2639	2031.4	516.6	5376	2828
135	3136	5187	2051	5403.8	2267.8	5334	2198
207	3927	5187	1260	5991.7	2064.7	5277.5	1350.5
472	4277	5187	910	7899	3622	5252.5	975.5
395	4494	5187	693	2573	1921	3425.3	1068.7
499	4620	5187	567	9047.2	4427.2	5228	608
321	5635	5187	448	3428.4	2206.6	5155.5	479.5
438	5775	5187	588	4152.5	1622.5	5145.5	629.5
204	5817	5187	630	4418.4	1398.6	4232.6	1584.4
645	5880	5187	693	8043.5	2163.5	5138	742
308	6699	5187	1512	7519.2	820.2	6368.8	330.2
334	6783	5187	1596	3725.5	3057.5	6569	214
260	7854	5187	2667	5164.2	2689.8	5766	2088
588	9520	5187	4333	8119.1	1400.9	4878	4642
Average	5187						
MAE			1662		2087.7		1528.2
Std Dev.			1308				
Effect Size					0.32		-0.10

The formulae used for the effect size is:

Effect Size = (Mean of absolute residual of Prediction Techniques – Mean of Values of sample) / Standard Deviation of mean as an estimator

Effect Size for SVR-P = (2087.7 -1662)/ 1308

= 0.32

Effect Size for RepTrees = (1528.2 -1662)/ 1308

= -0.10

Only data sets that have prediction techniques which are statistically significant better than others should be further investigated in term of effect size.

	LR	RBFN	SVR-P	RepTrees	CBR-1	CBR-2	CBR-3
Dataset-6	-0.16	-0.07	-0.13	-0.18	-0.19	-0.27	-0.45
Dataset-8	-0.12	0.07	-0.23	-0.12	-0.05	-0.14	-0.07
Dataset-11	-0.52	0.07	-0.38	0.21	-0.26	-0.33	-0.22
Dataset-12	-0.22	-0.17	-0.38	-0.03	-0.39	-0.41	-0.41

Table 5.11: Effect Size result

Table 5.11 show the results of effect size where the practical effect size is small in most cases. However these effect size results will be useful if the prediction is better than guessing and is also statistically significant. A combination of statistical significance (p-value < 0.05) and large effect size can generate a set of preference relations in which one can be confident that the relation is not a chance outcome. Based on Table 5.11 only LR in Dataset -11 achieved medium effect size; the rest of effect sizes were small which could be considered uninteresting.

Having explored the means, MMRE and Pred, models and significance, preference relations can be generated.

Preference relations between two prediction techniques can be established such as P1 $\prec$ P2. The preference relations may be read as P2 is preferred to P1 or P1 is less

preferable than P2. If the effect size is not large enough, or in this thesis if the effect size is less than medium, then an indifference relation with a non-strict order will be generated, thus P1≼P2 denotes that P2 is not worse than P1.

This results of effect size and preference relations that were generated based on this step will be explained further in the next section accordingly as a summary of the application of the strategy.

5.7 Summary of The Application of The Strategy

The thesis proposed a pragmatic cost estimation strategy by incorporating the information that was gathered from literature. The proposed pragmatic cost estimation strategy consist of 5 steps: 1) Test against means, 2) Test using MMRE, Pred and MAE, 3) Investigate the models 4) Perform significance testing and 5) Test effect size. The strategy follows a series of steps during the analysis and continues while the strategy is still a sensible thing to do or otherwise stop and give advice. This section will show summary results based on the application of the strategy on each dataset group.

Dataset	Dataset-1
Prediction better than mean?	RBFN, CBR-2, CBR-3, Reptrees
MMRE(<50%) Or Pred(25) (>50%) or MAE	MMRE? CBR-2 Pred(25)? RBFN MAE? RepTrees
Acceptable model?	Only for CBR-2
Significance testing?	No Significance
Pragmatic Advice: Can use CBR-2 based on prediction results better than mean and MMRE (<50%). CBR-3 also can be optional as they are slightly higher (53%) than our cut-off points, RBFN and RepTrees drop as the model is not theoretically justifiable. However, these results failed to give any statistically significant results. To get a better prediction, it is suggested that data cleaning is required by using human experts. The other option is to increase the amount of data in the dataset.	

Dataset	Dataset-2
Prediction better than mean?	LR, SVRP, CBR-1, CBR-3
MMRE(<50%) Or Pred(25) (>50%) or MAE	MMRE? LR, SVRP, CBR-1, CBR-3 Pred(25)? LR, CBR-3 MAE? LR, CBR-3
Acceptable model?	Only for LR ,CBR-1 , CBR-3
Significance testing?	No Significance
Pragmatic Advice: Can use LR, CBR-1, and CBR-3 as they satisfy MMRE, Pred and MAE results. Based on the model on LR it is theoretically justifiable as the model is based on PointAdjust and language. The other option is CBR-1 and CBR-3. If the practitioners trust SVRP tools then SVRP can also be the other option for this dataset. Although there are no statistically significant results achieved here the results based on MMRE, Pred(25) and MAE shows that LR and CBR-3 could serve this dataset well enough as these techniques satisfy all accuracy measures.	

Dataset	Dataset-3
Prediction better than mean?	RBFN, CBR-2 ,CBR-3
MMRE(<50%) Or Pred(25) (>50%) or MAE	MMRE? None but CBR-2 (51%) Pred(25)? RBFN MAE? CBR-2
Acceptable model?	Only for CBR-2
Significance testing?	No Significance
Pragmatic Advice: Based on only MAE results CBR-2 could be used for prediction in this dataset. This results are considered really weak therefore the advice for this dataset is that data cleaning is needed using human experts as there is surely no point in forwarding this dataset for prediction as none of them beat our cut-out point for MMRE. RBFN gives a better prediction only in Pred results; however, the model is not understandable. Therefore, data cleaning is required before cost estimation tasks begin for this dataset.	

Dataset	Dataset-4
Prediction better than mean?	SVRP, CBR-1, CBR-2, CBR-3
MMRE(<50%) Or Pred(25) (>50%) or MAE	MMRE? SVRP, CBR-1,CBR-2, CBR-3 Pred(25)? CBR-2, CBR-3 MAE? CBR-2, CBR-3
Acceptable model?	Only for CBR-1, CBR-2, CBR-3
Significance testing?	No Significance
Pragmatic Advice: Can use CBR as they perform well in terms of MMRE, Pred and MAE. However, if the practitioners understand and trust the estimation provided by SVRP tools, then it is optional to use them.	

Dataset	Dataset-5
Prediction better than mean?	RBFN, CBR-1, CBR-2
MMRE(<50%) Or Pred(25) (>50%) or MAE	MMRE? RBFN, CBR-1, CBR-2 Pred(25)? CBR-1 MAE? RBFN, CBR-1
Acceptable model?	Only for CBR-1
Significance testing?	No Significance
Pragmatic Advice: Can use CBR-1 as they perform well in terms of MMRE, Pred and MAE. Although RBFN gives the best MAE the model is not understandable. Therefore CBR-1 is the best alternative for this dataset which has nearly the same results as RBFN in terms of MAE.	

Dataset	Dataset-6
Prediction better than mean?	LR, RBFN, SVRP, RepTrees, CBR-1,CBR-2, CBR-3
MMRE(<50%) Or Pred(25) (>50%) or MAE	MMRE? LR, SVRP, CBR-1, CBR-2, CBR-3 Pred(25)? RepTrees, CBR-1, CBR-3 MAE? CBR-3
Acceptable model?	LR, CBR-1, CBR-3
Significance testing?	RBFN vs CBR-3 (p-value = 0.04672)
Effect Size?	CBR-3 (-0.45) RBFN⩽ CBR-3
Pragmatic Advice: CBR-3 would be the best option to be used for this dataset based on MMRE, Pred(25), MAE and has a statistical significance better than RBFN. In this dataset the effect size is only -0.45 which is not large enough. Therefore the preference relations will be RBFN⩽ CBR-3, which means CBR-3 is not worse than RBFN. The other option is to use LR or SVRP which gives good results in MMRE.	

Dataset	Dataset-7
Prediction better than mean?	Not even one better than mean
Pragmatic Advice: The prediction is not better than the mean which makes the process of the continuation of estimation meaningless. Therefore the advice is data cleaning is needed for this dataset using human experts before the start of the cost estimation task. Besides cleaning the outliers, some kind of case selection should be applied to give a better prediction for this dataset. The increase in the amount of data and selective data could increase accuracy measure.	

Dataset	Dataset-8
Prediction better than mean?	LR, SVRP, RepTrees, CBR-1,CBR-2, CBR-3
MMRE(<50%) Or Pred(25) (>50%) or MAE	MMRE? SVRP, CBR-1,CBR-3 Pred(25)? RepTrees MAE? SVRP
Acceptable model?	CBR-1, CBR-3
Significance testing?	RBFN vs SVRP (p-value = 0.006814)
Effect Size?	SVRP (-0.23) RBFN≼ SVRP
Pragmatic Advice: For this dataset although the SVRP model stands as a black box technique, but it seem to give significance results better than RBFN. The advice is that if the practitioners trust this model then this technique could be used to evaluate the prediction. Based on the effect size for SVRP on this dataset (-0.23) the preference relations will be RBFN≼ SVRP which means SVRP is not worse than RBFN. The other option to choose from is CBR-1 and CBR-3 which satisfy our cut-off point in terms of MMRE.	

Dataset	Dataset-9
Prediction better than mean?	CBR-2, CBR-3
MMRE(<50%) Or Pred(25) (>50%) or MAE	MMRE? CBR-2 Pred(25)? CBR-2 MAE? CBR-2
Acceptable model?	CBR-2
Significance testing?	No Significance
Pragmatic Advice: Although there are no statically significant results on this dataset the researcher could use CBR-2 as they perform well in terms of MMRE, Pred and MAE. This dataset also needs further cleaning for better prediction results; as based on first step only CBR-2 and CBR-3 is better than the mean. The dataset has only a small amount of data which could be the reason why the other prediction results failed to give better predictions. Therefore, the advice is either to use CBR-2 or for a better prediction use human experts to clean the dataset before the start of cost estimation tasks.	

Dataset	Dataset-10
Prediction better than mean?	LR, RBFN, SVRP, RepTrees, CBR-1,CBR-2, CBR-3
MMRE(<50%) Or Pred(25) (>50%) or MAE	MMRE? CBR-3 Pred(25)? None MAE? CBR-3
Acceptable model?	Only for CBR-3
Significance testing?	No Significance
Pragmatic Advice: Although there are no statically significant results in this dataset, the researcher could use CBR-3 as it performs well in terms of MMRE, and MAE. The other option is to use human experts to investigate the dataset and find out why the data seems to give a prediction better than mean for all the models but fails early in our steps of pragmatic cost estimation strategy. Human experts in data analysis could help to eliminate irrelevant data which causes the problem for this dataset.	

Dataset	Dataset-11
Prediction better than mean?	LR, SVRP, CBR-1, CBR-2, CBR-3
MMRE(<50%) Or Pred(25) (>50%) or MAE	MMRE? LR Pred(25)? LR (46%) MAE? LR
Acceptable model?	LR
Significance testing?	LR vs CBR-1 (p-value = 0.042)
Effect Size?	LR (-0.52) CBR-1≺ LR
Pragmatic Advice: In this dataset it is clear that LR is a preferable technique. LR achieves better results than mean, MMRE and MAE. It also achieves slightly lower based on our cut-off point in terms of Pred(25) which is 46%. It is also proven that LR has a statistical significance better than CBR-1 with p-values < 0.05. The effect size for LR on this dataset is -0.52 which gives preference relations as CBR-1≺ LR which means LR is preferred compared to CBR-1 or CBR-1 which is less preferable than LR.	

Dataset	Dataset-12
Prediction better than mean?	LR, RBFN, SVRP, RepTrees, CBR-1,CBR-2, CBR-3
MMRE(<50%) Or Pred(25) (>50%) or MAE	MMRE? CBR-2, CBR-3 Pred(25)? None MAE? CBR-2, CBR-3
Acceptable model?	Only for CBR-2, CBR-3
Significance testing?	RBFN vs CBR-2 (p-value = 0.0383)
Effect Size?	CBR-2 (-0.41) RBFN≼ CBR-2
Pragmatic Advice: For this dataset CBR-2 is the best possible technique which can be used for cost estimation. The results of this pragmatic cost estimation strategy shows that CBR-2 is statistically significanctly is better than RBFN with p-values < 0.05. Based on the effect size for CBR-2 in this dataset (-0.41) the preference relations will be RBFN≼ CBR-2, which means CBR-2 is not worse than RBFN. The other option to choose from is CBR-3 which satisfies our cut-off point in terms of MMRE and also gives best results in terms of MAE.	

Dataset	Dataset-13
Prediction better than mean?	RBFN, SVRP, RepTrees, CBR-2, CBR-3
MMRE(<50%) Or Pred(25) (>50%) or MAE	MMRE? None Pred(25)?None MAE? CBR-3
Pragmatic Advice: This data set requires data cleaning before the start of the cost estimation task as the dataset did not even beat our cut-off point for MMRE and Pred(25). The dataset also has a high MMRE percentage, which shows that the dataset has a serious problem to identify. Human experts in data analysis and better case selection will help this dataset for future prediction. Otherwise the best option is to use CBR-3 based on best MAE results achieved.	

Dataset	Dataset-14
Prediction better than mean?	LR, RBFN, SVRP, RepTrees, CBR-1,CBR-2, CBR-3
MMRE(<50%) Or Pred(25) (>50%) or MAE	MMRE? SVRP, CBR-1 Pred(25)? None MAE? SVRP
Acceptable model?	Only for CBR-1
Significance testing?	No Significance
Pragmatic Advice: Can use CBR-1 for this dataset; however, if the practitioner trusts the results provided by SVRP then it could be the other option. The best option for this dataset is to go through data cleaning as the results still lack confidence and achieve weak results in terms of MMRE and Pred(25). To get a better accuracy prediction which is statistically significant the dataset should be carefully selected using the case selection process.	

Table 5.12: Pragmatic Cost estimation Strategy result

The overall results based on steps that been used as a pragmatic cost estimation strategy are presented in Table 5.12. The results can be viewed as being positive for the continued use of pragmatic cost estimation strategy.

The general recommendation for practitioners using this dataset:

1. Should not use Reptrees as they are not appropriate for numerical and small datasets.

2. RBFN could be dropped from the early stage of the estimation process as the model is hard to understand and justify for the non-expert.

3. Although SVRP results seem hard to trace the model since they produce good results, it is hard to leave out these prediction results. Therefore the advice is if practitioners trust the tools then they can use them. However, in most cases it is only an optional technique.

4. The strategy suggests that not only one technique but a basket of techniques available to be chosen. It also shows the different quality level of the results based on information gathered in a pragmatic cost estimation strategy. Whether they are qualified from just having the best MMRE, Pred and MAE results or whether the model is justifiable, and then has statistical significance and effect size.

5. Based on our results, although there are some results which do not show any statistically significant results, the strategy still can advise the practitioners on some of the techniques which achieve the best results in terms of MMRE, Pred and MAE. However, if higher quality results are required then the best advice is to do some data cleaning using human experts or employ data analysts to improve the dataset.

6. Overall for this dataset, LR, CBR-1, CBR-2, CBR-3 and SVRP could be used as prediction techniques.

5.8 Conclusions

Although there is a lot of debate in the cost estimation world in terms of the best predictive systems, it is not easy to find any conclusion until there is a standard operating procedure on cost estimation. Rather than argue which techniques are the best, the thesis applies the advice contained within Shepperd's validation framework and other sound pieces of advice such as looking at the value of Pred and MMRE and investigating the models produced by the estimators. This will help practitioners to better understand which techniques should be used and which techniques should never be used at all in certain circumstance.

It is also important that the effect size and preference relations should be reported in the literature to give other researchers better ideas and guidelines on the results reported. This pragmatic cost estimation strategy could be an initial guideline for other researchers to perform cost estimation tasks. By using this approach, which follows a series of steps and continues while the analysis is still sensible, it will help practitioners to better understand what they are actually estimating.

By using this pragmatic cost estimation strategy, practitioners will have more techniques to choose from rather than focusing on one technique. This will solve the problem of looking or arguing as to which is the best technique. This strategy gives practitioner's confidence in the results of their prediction tools. They also will understand better what the tools are actually estimating rather than trust the prediction results blindly.

This pragmatic cost estimation strategy helps even the new practitioners in cost estimation industry to understand the estimation accuracy of the tools that they choose. The strategy also helps to determine how well the prediction tool that they choose performs against the other existing tools. However highly skilled on data analysis is required for data cleaning if the dataset fails in the early stage of pragmatic cost estimation strategy. The implementation of this strategy will solve the problem of constantly searching for the best techniques (a fruitful task given the complexities of the data involved in cost estimation) and help practitioners to have better understanding and confidence in the technique that they are using.

6 Conclusions and Future Work

6.1 Summary of the Thesis

The thesis begins with a traditional literature review and then follows with a systematic literature review. The literature review gathers background information on web application cost estimation,- identifying a variety of new techniques that have been proposed for web application cost estimation, but finding that there are no clear conclusions as to which techniques should be used. To explore this further, a systematic literature review is presented, which explores the empirical evidence in support of different cost estimation techniques.

The review searched for relevant studies published in the period 1999-2010, finally identifying 30 primary studies.

It was found that a variety of regression and machine learning based technique have been reported to estimate the cost of web applications with Case-Based Reasoning (CBR) and Stepwise Regression (SW) being the most frequently used but Support Vector Regression (SVR) recently becoming a popular approach.

It was also found that there is not one estimation technique that can be proven to be superior. Different techniques with different features give mixed results in terms of prediction techniques. Another point to be notice are there is not yet any standard software size measure which can be used to measure the accuracy of web application prediction techniques.

However there is consistency in terms of the prediction accuracy that has been used. The most popular are MMRE, MdMRE and Pred(25). Since 2001, most studies also included boxplots of z and residuals.

In the early years most studies used student datasets due to a lack of industrial datasets. However, since the Tukutuku database was created more studies have been published using this dataset, mainly by the same group of researchers (unfortunately, this dataset is restricted due to confidentiality).

In terms of data characteristics, most research has focused on single versus cross company datasets. However, the other characteristics of the dataset are also mentioned as a possible reason for the outcome of the prediction techniques. These findings show that rather than trying to establish which technique works best overall, a more fruitful approach may be to explore the relationships between technique and dataset in order to identify which technique to use in which circumstances.

The thesis continues by doing survey to investigate the current industry practice on cost estimation. The survey collects input on the practice that is used in industry such as the method, the data characteristics, the cost estimation practice pattern, and the possibilities of future help in data contribution.

The analysis showed that all the companies which participated were from small software organisations. In most of these companies, the decision on cost estimation was made by the owner or the director. The highest percentage of the studied respondents were aware of, and had used, expert judgement as their method of web application cost estimation. Almost all of them did not use any tool for this.

Motovated by the findings of the systematic literature review, the thesis continues by investigating CBR on the ISBSG dataset with the aim of investigating the impact of dataset and number of analogies. It presents the related work on web application cost estimation and a number of challenges to the application of CBR. The main finding of this investigation is that no reliable guidance can be given regarding the number of analogies that should be employed in making a prediction. In some cases there is a tendency for the data to converge as k increases, whilst in others it diverges. Most of the graphs seem to suggest that the data has a big influence in the calculation of the MMRE and also the PRED(25) values.

In addition, the results do not give any confidence that increasing the size of the dataset results in more accurate predictions. It was also found that outliers in the form of large or small values could possibly affect these predictions. The quality of the data set seems to plays a major role in the precision of the prediction.

Given the findings regarding the quality of the data, the thesis describes an investigation into characterising the dataset with aim of associating datasets with particular characteristics onto prediction techniques. It was found that it was hard to characterise the dataset structure, therefore the thesis suggests as an alternative a pragmatic cost estimation strategy. This five-stage strategy which has been introduced to support practitioners in giving the best estimates they can by identifying which techniques should be used and which techniques should never be used at all in certain circumstances.

Pulling together various diverse sources of sound of advice, the proposed pragmatic cost estimation strategy consist of 5 steps: 1) Test against means, 2) Test using MMRE, Pred and MAE, 3) Investigate the models 4) Perform significance testing and 5) Test effect size. The strategy follows this series of steps during the analysis and continues while the strategy is still a sensible thing to do or otherwise stops and gives advice.

By being non-prescriptive about the technique used, and allowing the practitioners to use whether tools they have at their disposal (along with a small amount of historical data), this pragmatic cost estimation strategy helps even the new practitioners in cost estimation industry to understand the estimation accuracy of the tools that they choose. The strategy also helps to determine how well the prediction tool that they choose performs against the other existing tools. However highly skilled on data analysis is required for data cleaning if the dataset fails in the early stage of pragmatic cost estimation strategy. The aim of this strategy is to stop the practitioner constantly searching for the best techniques and help them to have better understanding and confidence in the technique that they are using.

6.2 Thesis Contributions

This thesis makes several contributions to the web application cost estimation area. This study has investigated the literature in a systematic manner by conducting a systematic literature review. The aim of this was to systematically review and report the available evidence in current literature to support the proposed research questions. Therefore, in this thesis, a summary of the results to date on web application cost

estimation has analysed and identified the needs and opportunities for future research in this area. Although different techniques have been proposed, there is a lack of information about which techniques should be used in which circumstance.

Additional findings from the review include the possibility of exploring data characteristics as a mechanism for improving the accuracy of prediction, and the lack of a widely accepted web size measure.

Through a survey of practitioners it has identified the pattern of estimation that is involved in industry. The identification of this pattern helps to drive the direction of academic research by informing the pattern that is involved in industry. This helps the future researcher understand the needs of the cost estimation industry. Finally, the researcher can make progress in this research and help industry practitioners to use the findings that are worthwhile for them.

The thesis has also investigated in details on dataset characteristics as the initial work on the ISBSG dataset (Letchmunan et al. 2010) and found that the effectiveness of prediction techniques was hampered by several factors including the characteristics of underlying dataset. The novel findings on this study are that no reliable guidance can be given regarding the number of analogies that should be employed in making a prediction. In addition, the results also do not give any confidence that increasing the size of the dataset results in more accurate predictions. It was also found that outliers could possibly effect the predictions.

It has been found that to generalise or structure dataset characteristics cannot be an easy task, especially for non-highly skilled practitioners. Therefore in this thesis, Shepperd and MacDonell's validation framework and other pieces of advice were used to build and demonstrate a pragmatic cost estimation strategy.

The contribution of this thesis from the initial pragmatic cost estimation strategy will support practitioners in giving several best prediction techniques when carrying out a cost estimation task in their organization. The key contribution of this thesis is the demonstration of how pragmatic strategies can be used in practice during (web-based)

cost estimation. This approach provides good a insight for practitioners to choose the most reliable and trusted prediction techniques.

This pragmatic strategy has demonstrated its usefulness during the evaluation but it may be possible to make alterations for this to be improved in future studies. The other steps which should be included in the pragmatic cost estimation strategy remain an open question for future studies to address.

6.3 Lessons Learned

Several valuable lessons have been learnt which should benefit future research. The survey on cost estimation industry practice has revealed that there is a need to do some academic research in the area of cost estimation which is practical and cost effective for use in industry. Although there is extensive research on this, there is no standard advice for industry practitioners. Most researchers investigate the benefit of each technique without looking at which circumstance is best suited for which prediction techniques. To collect all the evidence on the literature, a systematic literature review is proposed by this thesis. The result of this, although potentially disappointing, is actually quite helpful because it provides a direct suggestion as to why the need for this investigation is timely and necessary.

Rather than choose the best techniques, it is always better to have several options of prediction techniques that best suit after a series cost estimation strategies. In that case, in this thesis Shepperd and MacDonell's validation framework and several pieces of advice have been used. Therefore the use of this pragmatic cost estimation strategy may be beneficial in the meantime. These solve the problem of deciding which prediction techniques are the best; however, more research is required to determine what the best strategy which has been suggested in this thesis.

6.4 Research Limitations

The thesis has a number of limitations identified below:

Literature review

A systematic literature review is regarded as the best approach to review and report the available evidence. However in a systematic literature review it's rare to find any industry approaches or findings. Research using industry dataset, tools and practitioners should be reported more often. Researchers who have industry datasets could make them publicly available by excluding the confidentiality data. This will provide the researchers in this area to research more on this area, increase the opportunity to replicate studies, and help the industry to obtain better prediction techniques.

CBR using ISBSG dataset

There are several challenges that are reported in CBR. More research to solve the challenges in CBR should be carried out as this technique give prediction based on past history dataset.

The results are based on distance measures. However the validity of the results seems to give less priority for categorical data. The reason for this is that it uses only values of 0 (identical) or 1 (different). The tool does not provide the model that has been generated.

Pragmatic cost estimation strategy

Although the thesis has developed and demonstrated a pragmatic cost estimation strategy using 14 datasets, replication of this strategy using different datasets in different experiments will give better insights into the strategy. For example, the strategy used a dataset which has only numerical features. Future research should use a dataset that has categorical features to explore the potential impact of these. In our strategy there is no advice on data cleaning before the start of cost estimation (as this

is regarded as being outside the scope of this thesis). Future refinement should look on this as part of strategy.

These limitations create a number of opportunities for future research which will be discussed further in the next section.

6.5 Future Work

From the work carried out in this thesis, there are several issues that require further investigation.

There is a need to automate the Angel tool

Among all the techniques that were used, Case Based Reasoning (CBR) showed some potential in certain circumstance. However this technique was found to be the most difficult and more time was required to attain the final result due to the difficulties in running the tool automatically in comparison to WEKA. Further attention is required to improve the Angel tool so that it is more automated.

Application of pragmatic cost estimation strategies on industry dataset

This thesis drew its findings from an academic viewpoint, working in an experimental situation. An alternative setting is to use an actual industrial dataset which could result in more accurate findings. One of the steps to evaluate the application of pragmatic strategies presented in this thesis is to use them in an industrial environment

Refinement and improvement of pragmatic cost estimation strategies

Pragmatic cost estimation strategies that have been developed in this thesis need refinement and improvement. A future evaluation on an industrial dataset will give better insight and further refinement of these strategies.

There is a need for pragmatic advice on data cleaning

Based on the pragmatic cost estimation strategies, it was found that data cleaning could influence the overall results of prediction techniques, Therefore, as a result of this thesis there is a need for pragmatic advice on data cleaning before the start of the application of pragmatic cost estimation strategies.

The potential for further research on the relationships between data characteristics and prediction techniques

Due to the fact that the relationships between data characteristics and prediction techniques are currently unknown, there is a potential for further research to explore the possible relationships between data characteristics and prediction techniques.

Standard reporting and the documention of pragmatic cost estimation strategy results

This thesis has also demonstrated how important standard reporting and documenting are for future work. Therefore, the cost estimation community should agree on a standard method to report and document the results so that other researchers can validate and improve their work. Only through this, can better strategies be created to solve the problem of which techniques to use in order to support practitioners in giving the best estimate possible.

6.6 Conclusions

The thesis has shown the way in which the cost estimation area should have a pragmatic cost estimation strategy. This suggests to practitioners and researchers that rather than focusing on which the best prediction techniques are, they should look at which are the best techniques for use in cost estimation strategy. Using the strategy it would be much easier to inspect and study the details of the results. The pragmatic cost estimation strategy that has been produced in this thesis is just a start of a new era of selecting better prediction techniques. However, this needs further analysis and possible refinement using a different industrial dataset.

In conclusion, researchers should agree and adopt a standard operating procedure to report and document their results. This will help industrial practitioners to use their findings and enrich the knowledge about which is the best prediction in which circumstances rather than producing new techniques.

References

Aamodt, A. and Plaza, E., 1994. Case based reasoning: Foundational issues, methodology variations, and system approaches, *AI Communications* 7, pp. 39-54

Albrecht, A.J., 1985. Function Points help managers assess application, maintenance values, *Computerworld Special Report on Software Productivity, CW Communications*, pp. SR20 - SR21.

Altman. 1991. Practical statistics for medical research. London: Chapman and Hall.

Aggarwal, K.K., Singh, Yogesh., Chandra, Pravin., and Puri, Manimala 2005. Bayesian regularization in a neural network model to estimate lines of code using function points, Journal of Computer Sciences 1(4), 505-509

Angelis. L., I. Stamelo and M. Morisio. 2001. Building A Software Cost Estimation Model Based On Categorical Data, Software Metrics, IEEE International Symposium on, p. 4, Seventh International Software Metrics Symposium (METRICS'01)

Aroba, J., Cuadrado-Gallego, J. J., Sicilia, M. Á., Ramos, I., and García-Barriocanal, E. 2008. Segmented software cost estimation models based on fuzzy clustering. *Journal of Systems and Software, 81*(11), 1944-1950.

Back, T., U. Hammel and H. Schwefel. 1997. Evolutionary Computation: Comments on the History and Current State, IEEE Transactions on Evolutionary Computation, pp. 3–17

Baresi, L., Morasca, and S., Paolini, P. 2002. An empirical study on the design effort for Web applications. In: Proceedings of WISE 2002, pp. 345–354

Baresi, L., Morasca, S. 2007. Three Empirical Studies on Estimating the Design Effort of Web Applications. Transactions on Software Engineering and Methodology 16(4)

Baresi, L., Sandro Morasca and Paolo Paolini. 2003. Estimating the Design Effort of Web Applications, Proceedings of the 9th International Symposium on Software Metrics, p.62

Briand, L. C., Khaled El Emam, Dagmar Surmann, Isabella Wieczorek, and Katrina D. Maxwell. 1999. An assessment and comparison of common software cost estimation modeling techniques. In *Proceedings of the 21st international conference on Software engineering* (ICSE '99). ACM, New York, NY, USA

Briand, L. C., I. Wieczorek. 2002. Software Resource Estimation. *Encyclopedia of Software Engineering.* Volume 2, P-Z (2nd ed., 2002), Marciniak, John J. (ed.) New York: John Wiley & Sons, pp. 1160-1196.

Boehm, B. 1981. Software Engineering Economics, Prentice – Hall

Boehm, B. W., C. Abts, A.W. Brown, S. Chulani, B.K. Clark, W. Horowitz, R. Madachy, D. Reifer, B. Steece. 2000. *Software Cost Estimation with COCOMO 11*. Prenctice Hall, NJ

Burgess, Colin J., and Martin Lefley. 2007. Can genetic programming improve software effort estimation? A comparative evaluation, Information and Software Technology, 43 (14), 863-867

Conte, S.D, H.E.Dunsmore, and V.Y.Shen. 1986. Software Enginnering Metrics and Models.

Corazza, A., S. Di Martino, F. Ferrucci, C. Gravino, E. Mendes. 2009. Applying support vector regression for web effort estimation using a cross-company dataset," Empirical Software Engineering and Measurement, International Symposium on, pp. 191-202, 2009 3rd International Symposium on Empirical Software Engineering and Measurement

Corazza, Anna, Sergio Di Martino, Filomena Ferrucci, Carmine Gravino, and Emilia Mendes. 2011. Investigating the use of Support Vector Regression for web effort estimation. Empirical Softw. Engg. 16, 2, 211-243

Costagliola, G., Di Martino, S., Ferrucci, F., Gravino, C., Tortora, G., Vitiello, G. 2006. Effort estimation modeling techniques: a case study for web applications. In: ICWE'06. Procs. Intl. Conference on Web Engineering, pp. 9–16

Costagliola, G., Ferrucci, F., Gravino, C., Tortora, G., and Vitiello, G. 2004. A COSMIC-FFP based method to estimate web application development effort. In Web Engineering---4th International Conference, ICWE 2004 (Munich, Germany, July 26--30). Lecture Notes in Computer Science, vol. 3140. Springer-Verlag, Berlin, Germany, 161--165

Craig Standing. 2002. Methodologies for developing Web applications, Information and Software Technology, Volume 44, Issue 3, Pages 151-159.

Dhawan, S and Kumar, R 2007. Web Metrics for evaluating Effort and design of Hyperdocuments. Association for Computing Machinery New Zealand Bulletin: 14-26 *Shanghai China:* 331-340

Desharnais Dataset. http://tunedit.org/repo/PROMISE/EffortPrediction/desharnais.arff. (Accessed on 28 May 2011)

De Almeida, M.A., Lounis, H.,, and Melo, W.L. 1998. An Investigation on the Use of Machine Learned Models for Estimating Correction Costs. *Proc. of the 30th Int'l Conf.Software Eng*, 473-476

Dteg. 2011. SVM- Support Vector Machine. http://www.dtreg.com/svm.htm (Accessed on 28 May 2011)

Dolado, J.J., 2001. On the problem of the software cost function, Information and Software Technology, Volume 43, Issue 1, 1 January 2001, Pages 61-72

Fewster Rachel and Emilia Mendes. 2001. Measurement, Prediction and Risk Analysis for Web Applications. Software Metrics, IEEE International Symposium on, p. 338, Seventh International Software Metrics Symposium (METRICS'01)

Fewster, R., Mendes, E., 2001. Measurement, Prediction and Risk Analysis for Web Applications. In: Proceedings of IEEE Metrics Symposium, pp. 338–348. IEEE Computer Society Press, Los Alamitos

Foss T, Myrtveit I, and Stensrud E. 2001. A Comparison of LAD and OLS Regression for Effort Prediction of Software Projects. Proc. 12th European Software Control and Metrics Conf., pp. 9-15.

Foss T, Stensrud E, Kitchenham B, Myrtveit I. 2003. A Simulation Study of the Model Evaluation Criterion MMRE. IEEE Transactions On page(s): 985 - 995 , Volume: 29 Issue: 11, Nov. 2003

GraphPad. 2011
http://www.graphpad.com/articles/interpret/Analyzing_two_groups/choos_anal_comp_two.htm (Accessed on 8 August 2011)

Grubbs, F. E. 1969. Procedures for detecting outlying observations in samples. http://en.wikipedia.org/wiki/Outliers. (Accessed on 28 May 2011)

Gray, A., and MacDonell, S., 1997. A comparison of techniques for developing predictive models of software metrics, Information and Software Technology, 39 (6), 425- 437

Hammond, K.R, Hamm R.M., Grassia J. and Pearson T., 1987. Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment, IEEE Transactions on Systems, Man, and Cybernetics, 17 (5), pp. 753–770

Heemstra, F.J., 1992. Software cost estimation. Information and software Technology 34 (10) pp. 627-639.

Huang, S. J., & Chiu, N. H. 2006. Optimization of analogy weights by genetic algorithm for software effort estimation. *Information and Software Technology*, *48*(11), 1034-1045.

Idris Ali, Abdelali Zakrani, Mohamed Elkoutbi and Alain Abran, 2008. Fuzzy Radical Basis function Neural Networks for Web Applications Cost Estimation. *IEEE 2008*: 576-580

ISBSG, Dataset, 2009. International Software Benchmark and Standard Group, www.isbsg.org., April 27, 2009

James N.Miller. 1993. Tutorial Review - Outliers in experimental data and their treatment. Analyst, 118, p. 455-461

Jorgensen, M. 2004. A Review of Studies on Expert Estimation of Software Development Effort, *Journal of Systems and Software* 70(1-2) : 37-60

Jorgensen, M.; Shepperd, M., 2007. A Systematic Review of Software Development Cost Estimation Studies. *IEEE Transactions on Software Engineering* , vol.33, no.1, pp.33-53

Kadoda, Gada, Michelle Cartwright, Liguang Chen, and Martin Shepperd. 2000. Experiences Using Case-Based Reasoning to Predict Software Project Effort. Proceedings of EASE 2000, Keele, UK.

Kadoda, Gada, Michelle Cartwright, and Martin Shepperd. 2001. Issues on the effective use of CBR Technology for software project prediction. In *Case-Based Reasoning Research and Development*, LNCS, Springer, 276-290

Kitchenham, B. 1991. Making process predictions. In Fenton N. Software Metrics – A Rigorous Approach, Chapman and Hall.

Kitchenham, B.; Pfleeger, S.L.; Fenton, N. 1995. Towards a framework for software measurement validation. *IEEE Transactions on Software Engineering* , vol.21, no.12, pp.929-944

Kitchenham BA, MacDonell SG, Pickard LM, and Shepperd MJ. 2001. What Accuracy statistics Really Measure. IEEE Proc. Software, vol. 148, no. 3, pp. 81-85.

Kitchenham, B.A., and Mendes, E. 2004. A Comparison of Cross-company and Single-company Effort Estimation Models for Web Applications. In: Proceedings EASE 2004, pp. 47–55

Kitchenham, B.A., and Mendes, E. and Guilherme H. Travassos, Cross versus Within-Company Cost Estimation Studies: A Systematic Review, IEEE Transactions on Software Engineering, pp. 316-329, May, 2007

Kumar, S., B.A. Krishna and P.S. Satsangi., 1994. Fuzzy systems and neural network in software engineering project management. J.Applied Intelligence, 4, pp. 31-52

Kumar,V., K. Ravi, V. Carr, and Raj Kiran, N. 2008. Software development cost estimation using wavelet neural networks. *Journal of Systems and Software, 81*(11), 1853-1867.

Lee, A., Cheng, C. H., & Balakrishnan, J. 1998. Software development cost estimation: integrating neural network with cluster analysis. *Information & Management*, *34*(1), 1-9.

Letchmunan, Sukumar, Marc Roper and Murray Wood. 2010. Investigating effort prediction of web-based applications using CBR on the ISBSG dataset. Proc. 14th

International Conference on Evaluation and Assessment in Software Engineering (EASE)

Lokan, C. and E. Mendes. 2008. Investigating the use of chronological splitting to compare software cross- company and single-company effort predictions. In *Proceedings of the 12th Conference on Evaluation & Assessment in Software Engineering (EASE 2008)*, pages 151-160.

Lokan, Chris and Emilia Mendes. 2009. Using chronological splitting to compare cross- and single-company effort models: further investigation. In Proceedings of the Thirty-Second Australasian Conference on Computer Science - Volume 91 (ACSC '09), Bernard Mans (Ed.), Vol. 91. Australian Computer Society, Inc., Darlinghurst, Australia, Australia, 47-54.

Mair Carolyn, Gada Kadoda, Martin Lefley, Keith Phalp, Chris Schofield, Martin Shepperd, Steve Webster. 2000. An investigation of machine learning based prediction, Journal of Systems and Software 53 (1), 23-29

Mangia, L., Paiano, R. 2003. MMWA: A Software Sizing Model for Web Applications. In: Proc. Fourth International Conference on Web Information Systems Engineering, pp. 53–63

Mendes Emilia. 2007. The Use of Bayesian Networks for Web Effort Estimation, Lecture Notes in Computer Science, p.90-104, Vol. 4607/2007

Mendes Emilia., Chris Lokan , Robert Harrison and Chris Triggs. 2005. A Replicated Comparison of Cross-Company and Within-Company Effort Estimation Models Using the ISBSG Database, Proceedings of the 11th IEEE International Software Metrics Symposium (METRICS'05), p.36

Mendes Emilia., S. Counsell, N. Mosley. 2005b. Towards a Taxonomy of Hypermedia and Web Application Size Metrics. In Proceedings of International Conference of Web Engineering (ICWE 2005), pp. 110--123, 2005

Mendes, E. and S. Counsell and N. Mosley. 2001. Towards the Prediction of Development Effort for Hypermedia Applications, HT'01, 249–258.

Mendes, E., 2000. Investigating Metrics for a Development Effort Prediction Model of Web Applications. *Proceedings of the 2000 Australian Software Engineering Conference*, Page(s): 31-41.

Mendes, E., C. Pollino and N. Mosley. 2009. Building an expert-based Web effort estimation model using Bayesian networks. Proceedings of the EASE Conference (2009), pp. 1–10.

Mendes, E., Di Martino, S., Ferrucci, F., Gravino, C. 2007a. Effort Estimation: How Valuable is it for a Web Company to use a Cross-company Data Set, Compared to Using Its Own Single Company Data Set? In: Proceedings of WWW 2007

Mendes, E., Ian Watson , Chris Triggs , Nile Mosley, and Steve Counsell. 2003. A Comparative Study of Cost Estimation Models for Web Hypermedia Applications, Empirical Software Engineering, v.8 n.2, p.163-196

Mendes, E., Martino, S.D., Ferrucci, F., Gravino, C. 2008. Cross-company vs. single-company web effort models using the Tukutuku database: An extended study. Journal of System & Software 81(5), 673–690

Mendes, E., Mosley, and N., Counsell, S. 2003a. Early Web Size Measures and Effort Prediction for Web Costimation. In: Proceedings of the IEEE Metrics Symposium, pp. 18–29

Mendes, E., Mosley, N., and Counsell, S. 2002d. Web hypermedia cost estimation: further assessment and comparison off cost estimation modelling techniques. New Review of Hypermedia and Multimedia, 199 – 229

Mendes, E., Mosley, N., and Counsell, S. 2003c. Investigating Early Web Size Measures for Web Cost Estimation, EASE 2003

Mendes, E., Mosley, N., and Counsell, S. 2003d. Do adaptation rules improve web cost estimation?. In *Proceedings of the Fourteenth ACM Conference on Hypertext and Hypermedia* (Nottingham, UK, August 26 - 30, 2003). HYPERTEXT '03. ACM, New York, NY, 173-183.

Mendes, E., Mosley, N., and Watson, I. 2002b. A Comparison of Case-Based reasoning Approaches to Web Hypermedia Project Cost Estimation. In: Proc. WWW'02

Mendes, E., Mosley, N., Counsell, S. 2001. Web metrics - Metrics for estimating effort to design and author Web applications. IEEE MultiMedia, 50–57

Mendes, E., Mosley, N., Counsell, S. 2003b. A Replicated Assessment of the Use of Adaptation Rules to Improve Web Cost Estimation. In: Proc. ISESE, pp. 100–109

Mendes, E., Mosley, N., Counsell, S. 2005. Investigating Web Size Metrics for Early Web Cost Estimation. Journal of Systems and Software 77(2), 157–172

Mendes, E., Mosley, N., Counsell, S. 2005. Web Effort Estimation. In: Mendes, E., Mosley, N. (eds.) Web Engineering, pp. 29–73. Springer, Heidelberg

Mendes, E., Nile Mosley , Steve Counsell. 2002. The Application of Case-Based Reasoning to Early Web Project Cost Estimation, Proceedings of the 26th International Computer Software and Applications Conference on Prolonging Software Life: Development and Redevelopment, p.393-398

Mendes, E., Nile Mosley. 2002c. Further Investigation into the Use of CBR and Stepwise Regression to Predict Development Effort for Web Hypermedia Applications, Proceedings of the 2002 International Symposium on Empirical Software Engineering, p.79

Mendes, E., S. Counsell and N. Mosley. 2001. Measurement and effort prediction for Web applications. , Springer-Verlag, Berlin, Germany

Mendes, E., S. Counsell and N. Mosley. 2002a. Comparison of Web Size Measures for Predicting Web Design and Authoring Effort, IEE Proceedings-Software 149(3), pp. 86—92

Mendes, E., Sergio, M., Filomena, F. and CarmineAn, G., 2007b. A Replicated Study Comparing Web Effort Estimation Techniques, Web Information Systems Engineering, WISE 2007, 423-435

Mendes, E.; Counsell, S. 2000. Web development effort estimation using analogy. *Software Engineering Conference, 2000. Proceedings. 2000 Australian*, vol., no., pp.203-212

Mendes, E.; Mosley, N. 2008b. Bayesian Network Models for Web Effort Prediction: A Comparative Study. *IEEE Transactions on Software Engineering*, vol.34, no.6, pp.723-737

Mendes, Emilia and Barbara Kitchenham. 2004. Further Comparison of Cross-Company and Within-Company Effort Estimation Models for Web Applications, Proceedings of the Software Metrics, 10th International Symposium on (METRICS'04), p.348-357

Mendes, Emilia and Nile Mosley. 2008. Bayesian Network Models for Web Effort Prediction: A Comparative Study. *IEEE Transactions on Software Engineering*, vol.34, no.6, pp.723-737

Mendes., Emilia. 2008a. The Use of Bayesian Networks for Web Effort Estimation: Further Investigation, Proceedings of the 2008 Eighth International Conference on Web Engineering, p.203-216

Morisio, Maurizio., Ioannis Stamelos , Vasilis Spahos and Daniele Romano. 1999. Measuring Functionality and Productivity in Web-Based Applications: A Case Study, Proceedings of the 6th International Symposium on Software Metrics, p.111

Myrtveit, I., and Stensrud, E. 2012. Validity and reliability of evaluation procedures in comparative studies of effort prediction models. *Empirical Software Engineering*, 1-11.

Oyang, Yen-Jen, et al. 2005. Data classification with radial basis function networks based on a novel kernel density estimation algorithm, *Neural Networks, IEEE Transactions on* 16.1 p. 225-236.

Perry Dewayne E., Adam A. Porter, and Lawrence G. Votta. 2000. Empirical studies of software engineering: a roadmap. In *Proceedings of the Conference on The Future of Software Engineering* (ICSE '00). ACM, New York, NY, USA

Pickard Lesley, Barbara Kitchenham and Susan Linkman. 1999. An Investigation of Analysis Techniques for Software Datasets. Software Metrics, IEEE International

Symposium on, p. 130, Sixth International Software Metrics Symposium (METRICS'99)

Quinlan, J. R. 1987. Generating production rules from decision trees. In *Proceedings of the Tenth International Joint conference on Artificial intelligence* (Vol. 30107, pp. 304-307).

R. 2011. R Project. http://www.r-project.org/(Accessed on 28 May 2011)

Riaz, M., Mendes, E., Tempero, E., 2009. A systematic review of software maintainability prediction and metrics, Empirical Software Engineering and Measurement, 2009. ESEM 2009. pp. 367 - 377

Reifer, D.J., 2000. Web development: estimating quick-to-market software. *Software, IEEE* , vol.17, no.6, pp.57-64, Nov/Dec 2000

Ruhe Melanie , Ross Jeffery , Isabella Wieczorek. 2003. Cost estimation for web applications, Proceedings of the 25th International Conference on Software Engineering, 285-294

Ruhe Melanie , Ross Jeffery , Isabella Wieczorek. 2003. Using Web Objects for Estimating Software Development Effort for Web Applications, Proceedings of the 9th International Symposium on Software Metrics, p.30

Rollo, T. 2006. Functional size measurement and COCOMO—a synergistic approach. In *Proceedings of Software Measurement European Forum* (SMEF), Rome, Italy, 259–267.

Rosenthal, R. (1994). Parametric measures of effect size. The handbook of research synthesis, 231-244.

Schofield, C. 1998. An empirical investigation into software estimation by analogy, PhD thesis, Dept. of Computing, Bournemouth Univ., UK,

Schulz, S. 1999. CBR-Works: A state-of-the art shell for case-based application building. In Melis, E., ed., *Proceedings of GWCBR'99, Wiirzburg, Germany*, pp. 166-175.

Sergio Di Martino, F. Ferrucci, C. Gravino. 2009. An empirical Study on the use of Web-COBRA and Web Objects to estimate web application development effort, ICWE 2009

Sergio Di Martino, Filomena Ferrucci, Carmine Gravino and Emilia Mendes. 2007. Comparing Size Measures for Predicting Web Application Development Effort: A Case Study. Empirical Software Engineering and Measurement, International Symposium on, pp. 324-333, First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007)

Shepperd, Martin and Gada Kadoda. 2001. Using Simulation to Evaluate Prediction Techniques. Software Metrics, IEEE International Symposium on, p. 349, Seventh International Software Metrics Symposium (METRICS'01)

Shepperd, Martin and Schofield, Chris. 1997. Estimating software project effort using analogies. *IEEE Transactions on Software Engineering*, vol.23, no.11, pp.736-743

Shepperd, Martin. 2007. Software project economics: a roadmap. *Future of Software Engineering (FOSE'07)*, 304-315

Shepperd, M., and MacDonell, S. 2012. Evaluating prediction systems in software project estimation. *Information and Software Technology*.

Silvia Abrahao , Geert Poels , and Oscar Pastor. 2004. Evaluating a Functional Size Measurement Method for Web Applications: An Empirical Analysis, Proceedings of the Software Metrics, 10th International Symposium on (METRICS'04), p.358-369

Simpson, J. R. and Montgomery, D. C..1998. A robust regression technique using compound estimation. Naval Research Logistics (NRL), 45: 125-139.

SPSS. 2011. SPSS: Descriptive Statistics. http://psychology.illinoisstate.edu/psy138/resources/spss/spss3.html. (Accessed on 28 May 2011)

Srinivasan, K.; Fisher, D., 1995. Machine learning approaches to estimating software development effort, *Software Engineering, IEEE Transactions on* , vol.21, no.2, pp.126-137.

Stamelos I, Angelis L, Morisio M, Sakellaris E, Bleris GL. 2003. Estimating the development cost of custom software, Information & Management, Volume 40, Issue 8, September 2003, Pages 729-741

Stensrud,E., and Myrveit, I. 1998. Human performance estimating with analogy and regression models: an empirical validation. In proceedings Fifth International Software Metrics Symposium (Metrics' 98). IEEE Computer Society Press, Los Alamitos, CA, pp.205-213

Ochodek, M., Nawrocki, J., Kwarciak, K., 2011. Simplifying effort estimation based on Use Case Points. Information and Software Technology 53, pp 200-213

Walkerden, Fiona and Ross Jeffery. 1999. An Empirical Study of Analogy-based Software Effort Estimation. Empirical Software Engineering. 4, 2 (June 1999), 135-158

Watson, I., Mendes, E., Mosley, N., and Counsell, S. 2002. Using CBR to Estimate Development Effort for Web Hypermedia Applications. In Proc. of the Fifteenth Annual Conference of the International Florida Artificial Intelligence Research Society. Menlo Park, CA, AAAI Press (2002) 132–136

Weka. 2011a . http://www.cs.waikato.ac.nz/ml/weka/(Accessed on 28 May 2011)

Weka. 2011b http://wekadocs.com/node/2(Accessed on 28 May 2011)

Appendices

Appendix A ……141

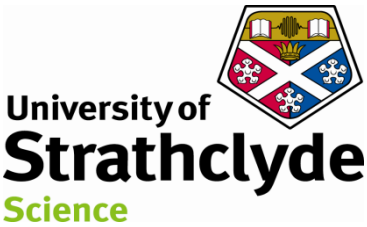
Appendix B…….142

Appendix C…….146

Appendix D…….156

Appendix E…….214

Appendix A



Study Title: Web Application Cost Estimation Best Practice in Scotland
Investigator: Dr Marc Roper and Sukumar Letchmunan

Dear:

I am a PhD research student at University of Strathclyde studying cost estimation techniques for web applications development. As part of this research I am gathering data on the best practice in industry on web application cost estimation, and would value your participation in the study. The aim of this study is to investigate the various approaches to cost estimation for web applications development with a view to comparing and evaluating them.

Every company that participates in this study will receive a copy of the final report which will compare the cost estimation strategies employed across the companies in Scotland. All the volunteered data remains strictly confidential and no references to individuals participating in this nor their companies are disclosed publicly. Please visit the link below to answer this questionnaire:
http://devweb2007.cis.strath.ac.uk/~sukumar/survey/survey.html

I realise that your time is extremely valuable, but if you were able to assist us in completing this questionnaire it would be much appreciated. If you have any questions at any time about the study or the procedures, you may contact me by phone at **07527497286** or by e-mail at sukumar@cis.strath.ac.uk

If you are not in the right position to answer this questionnaire, could you recommend or pass this mail to any other main person in the company that did the decision on cost estimations.

Thank you for your assistance.

Sincerely,

L.Sukumar
Research student
Department of Computer and Information Sciences,
University of Strathclyde.
Tel: 0141 548 3592 Fax: 0141 548 4523

Appendix B

Web Application Cost Estimation Best Practice in Scotland

Welcome.

Thank You for taking part in this project.

The aim of this study is to investigate the various approaches to cost estimation for web applications development with a view to comparing and evaluating them. All the volunteered data remains strictly confidential and no references to individuals participating in this nor their companies are disclosed publicly. Your participation will require approximately 30 minutes to complete the survey.

Name

Position

Size of Company staff (approximately)

Contact No.

Email Address

1. When was your company established?

2. Normally how many people work in the design and development of web applications?

3. What type of web applications does your company mainly develop?

4. What is the typical size of the web application development project?

   person months

5. Which type of cost estimation methods are you aware of?

   - ☐ Algorithmic Cost Modelling
   - ☐ Expert Judgement.
   - ☐ Estimation by analogy
   - ☐ Parkinsonâ€™s Law
   - ☐ Pricing To Win
   - ☐ Top Down Estimation
   - ☐ Bottom-up estimation
   - ☐ Others:

6. Which type of cost estimation methods are used in your company?

   - ☐ Algorithmic Cost Modelling
   - ☐ Expert Judgement.
   - ☐ Estimation by analogy
   - ☐ Parkinsonâ€™s Law
   - ☐ Pricing To Win
   - ☐ Top Down Estimation
   - ☐ Bottom-up estimation
   - ☐ Others:

7. How do you carry out estimation?

   ○ Manually ○ Tool (Please specify):

8. At what stage of the lifecycle do you carry out the estimation?

9. What was being estimated?

   - ☐ Total effort
   - ☐ Staff.

- [ ] Duration
- [ ] Others (Please specify): _____

10. What was the purpose of the estimation?

    - [ ] Budget Approval
    - [ ] To determine manpower
    - [ ] To win the bid
    - [ ] Others (Please specify): _____

11. How many and what was the position of the people involved in the cost estimation process?

    _____

12. Did the people involved in the estimation process have any previous experience on web applications cost estimation? If yes what was their experience?

    _____

13. What was the level of accuracy were you trying to achieve?

    _____

14. If the cost estimation was not accurate what is the cause of inaccuracies?

    _____

15. How would you think the accuracy of web application cost estimations could be improved?

    _____

16. Was there any training on cost estimation provided by your company?

    ○ Yes ○ No

    If your answer is Yes, please continue to question 17. If No, proceed to question 18.

17. Did you think the training did or would improve your ability to estimate accurately?

Yes No

18. Did you save any estimation data from previous projects?

Yes No

19. Would you be willing to contribute some more of your time for a follow-up interview?

Yes No

Submit Reset

Appendix C

Study	Types (Web Hypermedia/ Web applications)	Size Measures	Prediction Techniques	Prediction Measures	Types of dataset	Characteristics	Best Techniques
(Ruhe et. al 2003)	Web applications	Web objects (WebMo) Vs Function points	Ordinary least squares regression Allete systems informal model (Expert System)	Magnitude relative error (MRE) Pred(25) Boxplots	Industrial	Industrial Australian web development company (12 dataset) 9 New developments 1 enhancement 2 redevelopments	OLS - Web Objects
(Mendes et. al 2001)	Web Hypermedia	Node Count, Media Count, Reused media count, Total node allocation, Total Media allocation, total reused media allocation, connectivity, connectivity density, total node complexity, cyclomatic complexity,	Case based Reasoning	UNKNOWN	Students	43 Computer Science Students from University of Auckland	Case based Reasoning

(Mangia et. al 2003)	Do not cover our Research Question Propose a metrics models for web applications						
(Lokan et. al 2008)	Do not used web project dataset which consist in ISBSG dataset ISBSG dataset (remove web project)						
(Costagliola et. al 2006)	Web applications	Length measures (number of web pages, new web pages, scripts, link, references) Functional Measures (Fp+web objects)	Linear Regression(LR) Regression tree(RT) Stepwise regression(SW) Analogy-based estimation(ABE) Combination of RT and LR Combination of RT and ABE	MMRE MdMRE Pred(0.25) Boxplots	Industrial	Italian software company (15 web projects)	LM – RT and ABE FM - SW
(Mendes et. al 2005)	Web applications Do not cover our Research question. Literature of web size metrics reported.						
Mendes et. al 2002	Web Hypermedia	Requirement and design measures(Use case count,	Case based Reasoning	MMRE MdMRE Pred(25) Boxplots of residuals	Students	Computer Science Students from University of Auckland	Requirement and design measures the unweighted Euclidean distance using the mean of the closest two

		entity count, attribute count, node count, anchor count, design effort) Application Measures (Page count, media count, program length, connectivity density, total page complexity, total effort)					analogies gave the best prediction: Application measures: Weighted Euclidean Distance, three analogies gave the best predictions.
(Ruhe et. al 2003)	Web applications	Web Objects	Ordinary least Squares(OLS) regression Expert based estimates Web-COBRA	MRE MMRE Pred(25) Boxplots of residual	Industrial	Australian Company (Allette Systems) 12 datasets 9 New developments 1 enhancement 2 redevelopments	Web –Cobra
(Morisio et. al 1999)	Do not cover our Research question Using Web applications but looking on cost of writing code on OO framework						
(Mendes et. al 2000)	Web hypermedia	Number of documents, reused documents,	Estimation by analogy	MMRE PRED(25)	Srudents	Two datasets from 76 student	Estimation by analogy

		links, compactness, stratum, structure of application					
(Mendes 2000)	Not relevant to our studies as it evaluate on independent variables of hypermedia.						
(Mendes et. al 2001a)	Web Hypermedia	Length size, reusability, complexity, size	Linear regression Stepwise regression	MMRE Box plots	Students	43 Computer Science students.	Linear regression
(Mendes et. al 2001b)	Web Hypermedia	Compactness, stratum, reused docs, connectivity, structure	Linear Regression Stepwise Multiple regression Estimation by Analogy	MMRE MdMRE	Students	76 Computer Science students.	Estimation by analogy
(Fewster et.al 2001)	Web hypermedia	Size, Reusability, complexity	Generalised linear model	Boxplot of residual	Students	43 Computer science students	Not reported
(Craig 2002)	Do not cover our research questions – requirement of methodologies						
(Mendes et. al 2002)	Web Hypermedia	Length, complexity, functionality	Linear, Stepwise regression	Boxplot of residual	Students	43 Computer science students	No single technique

(Mendes et al 2002b)	Web Hypermedia	Page Count, Media Count, Reused media count, Total node allocation, Total Media allocation, total reused media allocation, connectivity, connectivity density, total node complexity, cyclomatic complexity,	Case based reasoning	MMRE MdMRE Pred(25)	Students	Computer science students	Case based reasoning
(Mendes et. al 2003)	Web Hypermedia	Page count, Media Count, Program Count, Connectivity density, total page complexity, reused media count, reused program count	Stepwise Regression Regression Trees CBR	MMRE MdMRE Pred(25) Boxplots	Students	37 web hypermedia projects developed by MSc Students University of Auckland.	Stepwise Regression
(Mendes et. al 2002c)	Web Hypermedia	Page count, Media Count, Program Count, Connectivity density, total page	Stepwise CBR	MMRE MdMRE Pred(25) Boxplots of residuals	Students	37 web hypermedia projects developed by MSc Students University of Auckland.	Stepwise Regression

		complexity, reused media count, reused program count					
(Watson et. al 2002)	Web Hypermedia	Page count, Media Count, Program Count, Connectivity density, total page complexity, reused media count, reused program count	CBR	MMRE MdMRE Pred(25) Boxplots of residuals	Students	37 web hypermedia projects developed by MSc Students University of Auckland.	CBR -weighted Euclidean distance
(Mendes et. al 2002d)	Web hypermedia	Page count, Media Count, Program Count, Connectivity density, total page complexity, reused media count, reused program count	Linear Stepwise regression CBR	MMRE MdMRE Pred(25) Boxplots of residuals	Students	37 web hypermedia projects developed by MSc Students University of Auckland.	Stepwise and multiple linear regression.
(Baresi et. al 2002)	Web Application- Not Related for our studies as its on empirical study on web application design techniques.						
(Baresi et. al 2003)	Web Application Not Related for our						

	studies as its study on web application design techniques.						
(Mendes et. al 2003a)	Web Application	total number of web pages features/functionality	Case based reasoning Stepwise Regression	MMRE Pred(25)	Industrial	133 Tukutuku Datasets	Case based reasoning Company- specific datasets. .
(Mendes et. al 2003b)	Web applications		Case based reasoning	MMRE Pred(25)	Industrial	Tukutuku datasets Dataset 1: 12 web projects developed by single company Dataset2: 37 web projects developed by several companies. Dataset 1: 20 Features, 0 categorical features,80% outliers, 90% colinearty Dataset 2: 20 Features, 0 categorical features,30% outliers, 30% colinearty	CBR using adaptation rules
(Mendes et. al 2003c)	Web applications - literature on web size measure						
(Mendes et. al 2005)							Same as (Mendes et. al 2003c)
(Mendes et. al 2003d)	Web hypermedia	Page count, Media Count, Program Count,	Case based reasoning	MMRE Pred(25)	Students	37 web hypermedia projects developed by MSc Students University of Auckland.	

		Connectivity density, total page complexity, reused media count, reused program count					
(Silvia et. al 2004)	Web applications Not related to our studies - study on OOmFPweb in term of conceptual modelling, and evaluates the efficacy						.
(Kitchenham et. al 2004)	Web Hypermedia		MMRE Median MRE	Pred(25) Median of absolute residuals Boxplots	Industrial	53 web projects from Tukutuku database. Each web project provides 40 variables	Within company models
(Mendes et al 2004)	Web Hypermedia or Web software application Both represent as Web Projects		Forward stepwise regression Case based reasoning	MMRE Median MRE Pred(25) Median of absolute residuals Boxplots	Industrial	67 web projects Tukutuku database	SW- Within Company CBR – Cross company
(Mendes et. al 2005a)	This paper can be exclude as its on software projects						
(Costagliola et. al 2004)	Web application	Cosmic Full Function Point	MMRE Pred(25)		Students	32 web projects Undergraduate student's dataset.	A statistical analysis has been performed to confirm that COMIC-FFP can be used to predict development effort of web based systems.
(Mendes et. al	Hypermedia						Literature on Size metrics

2005b)	Literature on Size metrics						
(Mendes et. al 2005c)	A chapter on Web effort estimation						
(Sergio et. al 2007)	Web applications	Web objects Tukutuku measures Length measures Functional measures	Forward stepwise regression Case based reasoning	MMRE MdMRE Pred(25) Boxplots of residuals	Industrial	15 web applications (Italian software company) 25 variables	LM – SWR TM- CBR
(Mendes et. al 2007)	Web applications		Forward stepwise regression Case based reasoning Classification & Regression Trees (CART)	MMRE MdMRE Pred(25)	Industrial	150 web projects from Tukutuku database. 25 variables.	None of them superior
(Mendes et. al 2007b)	Web applications		Forward stepwise regression Case based reasoning	MMRE MdMRE Pred(25) Boxplots of residuals	Industrial	83 web projects from Tukutuku database. 25 variables. Single company – 15 projects Cross company – 68 projects	Single company model
(Mendes 2007)	Web applications		Bayesian network	MMRE MdMRE Pred(25) Boxplots of residuals	Industrial	150 web projects from Tukutuku database 25 variables	Bayesian network - mean and median effort.
(Mendes 2008)	Web application		Manual stepwise regression	MMRE MdMRE			This paper extends the work presented in [(Mendes 2007)

			Case based reasoning Bayesian network	Pred(25) Mean Median Boxplots of residuals			Hybrid BN model .
(Mendes et. al 2008)	Web applications						Results seem same as presented in (Mendes 2008)
(Mendes et. al 2009)	Web applications		Bayesian network		Industrial	Tukutuku dataset	Bayesian Network used to construct an expert based web effort model.
(Idris et. al 2008)	Web hypermedia		Fuzzy Radial Basis Function Neural networks (FRBFN)	MMRE Pred(25)	Industrial	53 web hypermedia fromTukutuku dataset 9 numerical attributes.	The results show that an RBFN using fuzzy C-means performs better than RBFN using hard C-means.
(Baresi et al)	Web applications Estimating the design effort. Not related for our studies.						
(Mendes et. al 2008b)	Web applications		Forward stepwise regression Case based reasoning	MMRE MdMRE Pred(25)	Industrial	83 web projects of Tukutuku database 25 variables	Single company datasets
(Corazza et. al 2009)	Web applications		Support Vector Regression(SVR) Manual Stepwise Regression Case based reasoning Bayesian networks	MMRE MdMRE Pred(25) Boxplots of residuals	Industrial	130 projects randomly selected from Tukutuku database.	Support Vector Regression(SVR)

Appendix D- Raw Data (ISBSG)

G1-Ran1-13

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
10178,COSMIC-FFP,2503,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;Workflow support & management;Process Control;,Java,Yes;,826,826
15720,IFPUG,934,Enhancement,Financial transaction process/accounting,PL/I,DB2,44,44
11252,IFPUG,3107,Enhancement,relatively complex application,4GL,Interactive,117,117
15008,IFPUG,626,New Development,Financial application area,Java,Interactive,116,116
13034,IFPUG,4295,New Development,other: sales promotion tool,Visual Basic,SQL-Server,422,422
14769,IFPUG,7368,Enhancement,Financial transaction process/accounting,Java,DB2;,1753,1753
13742,IFPUG,12564,New Development,Trading,HTML,ORACLE,1588,1588
14779,IFPUG,2891,New Development,Management or performance reporting;,Java,DB2;,430,430
15130,IFPUG,4045,New Development,Financial transaction process/accounting,Java,Oracle 8i8,435,435
11728,IFPUG,6944,New Development,other: Sales contact management,Visual Basic,SQL SERVER,424,424
11100,IFPUG,2240,Enhancement,Financial transaction process/accounting,Java,DB2,539,539
11009,IFPUG,2800,Enhancement,Financial transaction process/accounting,Java,UDB,124,124
10180,IFPUG,2340,New Development,Financial transaction process/accounting,Visual Basic,SQL-Server,309,309
15675,IFPUG,2762,New Development,Financial transaction process/accounting,Java,Oracle 8i8,297,297
11648,IFPUG,1056,Enhancement,Financial application area,Java,Interactive,165,165
15440,IFPUG,301,Enhancement,Financial application area,Java,Interactive,51,51
14578,IFPUG,470,Enhancement,Financial application area,Java,Interactive,77,77

G1-Ran1-33

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
10178,COSMIC-FFP,2503,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;Workflow support & management;Process Control;,Java,Yes;,826,826
15720,IFPUG,934,Enhancement,Financial transaction process/accounting;,PL/I,DB2;,44,44
11252,IFPUG,3107,Enhancement,relatively complex application,4GL,Interactive,117,117
15008,IFPUG,626,New Development,Financial application area;,Java,Interactive;,116,116
13034,IFPUG,4295,New Development,other: sales promotion tool,Visual Basic,SQL-Server,422,422
14769,IFPUG,7368,Enhancement,Financial transaction process/accounting;,Java,DB2;,1753,1753
13742,IFPUG,12564,New Development,Trading,HTML,ORACLE,1588,1588
14779,IFPUG,2891,New Development,Management or performance reporting;,Java,DB2;,430,430
15130,IFPUG,4045,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,435,435
11728,IFPUG,6944,New Development,other: Sales contact management,Visual Basic,SQL SERVER,424,424
11100,IFPUG,2240,Enhancement,Financial transaction process/accounting,Java,DB2,539,539
11648,IFPUG,1056,Enhancement,Financial application area,Java,Interactive,165,165
11009,IFPUG,2800,Enhancement,Financial transaction process/accounting,Java,UDB,124,124
10180,IFPUG,2340,New Development,Financial transaction process/accounting,Visual Basic,SQL-Server,309,309
15675,IFPUG,2762,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,297,297
15440,IFPUG,301,Enhancement,Financial application area;,Java,Interactive;,51,51
14578,IFPUG,470,Enhancement,Financial application area;,Java,Interactive;,77,77
13127,IFPUG,7496,New Development,Workflow support & management,ASP,SQL Server7,786,786
15598,IFPUG,1028,Enhancement,Web-based Application;,NET,SQL Server;,347,347
11283,IFPUG,410,Enhancement,Financial application area,Java,Interactive,82,82
11436,IFPUG,9058,New Development,other: Sales contact management,C,ORACLE,599,599
15444,IFPUG,2504,New Development,Financial transaction process/accounting;,Visual Basic,SQL Server 2000;,1236,1236
15123,IFPUG,5200,New Development,Financial transaction process/accounting;,Java,DB2;,775,775
15692,IFPUG,11949,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,1285,1285
14260,IFPUG,3576,New Development,Financial transaction process/accounting,Java,Oracle 8i8,778,778
15137,IFPUG,543,New Development,Financial application area;,Java,Interactive;,92,92

14908,IFPUG,2580,New Development,Web-based Application;,Java,Oracle;,264,264
10358,IFPUG,737,New Development,Web-based Application,Visual Basic,SQL Server,191,191
11149,IFPUG,1472,Enhancement,Financial transaction process/accounting,C,ORACLE,407,407
11149,IFPUG,1472,Enhancement,Financial transaction process/accounting,C,ORACLE,407,407
12178,IFPUG,679,New Development,Financial transaction process/accounting,ASP,MSDE SQL Server 2000,302,302
12078,IFPUG,54,Enhancement,Financial application area,Java,Interactive,17,17
12573,IFPUG,1671,New Development,other: Management system,PL/SQL,ORACLE,1216,1216

G1-Ran1-49

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
10178,COSMIC-FFP,2503,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;Workflow support & management;Process Control;,Java,Yes;,826,826
15720,IFPUG,934,Enhancement,Financial transaction process/accounting;,PL/I,DB2;,44,44
11252,IFPUG,3107,Enhancement,relatively complex application,4GL,Interactive,117,117
15008,IFPUG,626,New Development,Financial application area;,Java,Interactive;,116,116
13034,IFPUG,4295,New Development,other: sales promotion tool,Visual Basic,SQL-Server,422,422
14769,IFPUG,7368,Enhancement,Financial transaction process/accounting;,Java,DB2;,1753,1753
13742,IFPUG,12564,New Development,Trading,HTML,ORACLE,1588,1588
14779,IFPUG,2891,New Development,Management or performance reporting;,Java,DB2;,430,430
15130,IFPUG,4045,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,435,435
11728,IFPUG,6944,New Development,other: Sales contact management,Visual Basic,SQL SERVER,424,424
11100,IFPUG,2240,Enhancement,Financial transaction process/accounting,Java,DB2,539,539
11648,IFPUG,1056,Enhancement,Financial application area,Java,Interactive,165,165
11009,IFPUG,2800,Enhancement,Financial transaction process/accounting,Java,UDB,124,124
10180,IFPUG,2340,New Development,Financial transaction process/accounting,Visual Basic,SQL-Server,309,309
15675,IFPUG,2762,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,297,297
15440,IFPUG,301,Enhancement,Financial application area;,Java,Interactive;,51,51
14578,IFPUG,470,Enhancement,Financial application area;,Java,Interactive;,77,77
13127,IFPUG,7496,New Development,Workflow support & management,ASP,SQL Server7,786,786

15598,IFPUG,1028,Enhancement,Web-based Application;,NET,SQL Server;,347,347
11283,IFPUG,410,Enhancement,Financial application area,Java,Interactive,82,82
11436,IFPUG,9058,New Development,other: Sales contact management,C,ORACLE,599,599
15444,IFPUG,2504,New Development,Financial transaction process/accounting;,Visual Basic,SQL Server 2000;,1236,1236
15123,IFPUG,5200,New Development,Financial transaction process/accounting;,Java,DB2;,775,775
15692,IFPUG,11949,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,1285,1285
14260,IFPUG,3576,New Development,Financial transaction process/accounting,Java,Oracle 8i8,778,778
15137,IFPUG,543,New Development,Financial application area;,Java,Interactive;,92,92
14908,IFPUG,2580,New Development,Web-based Application;,Java,Oracle;,264,264
10358,IFPUG,737,New Development,Web-based Application,Visual Basic,SQL Server,191,191
11149,IFPUG,1472,Enhancement,Financial transaction process/accounting,C,ORACLE,407,407
10427,IFPUG,11372,New Development,Financial transaction process/accounting,SQL,ORACLE,859,859
12178,IFPUG,679,New Development,Financial transaction process/accounting,ASP,MSDE SQL Server 2000,302,302
12078,IFPUG,54,Enhancement,Financial application area,Java,Interactive,17,17
12573,IFPUG,1671,New Development,other: Management system,PL/SQL,ORACLE,1216,1216
11421,IFPUG,36,Enhancement,Financial application area,Java,Interactive,6,6
14372,IFPUG,3857,New Development,Financial transaction process/accounting,Visual C++,NCR;TeraData,1521,1521
14372,IFPUG,3857,New Development,Financial transaction process/accounting,Visual C++,NCR;TeraData,1521,1521
12175,IFPUG,1214,Enhancement,Financial application area,Java,Interactive,176,176
13369,IFPUG,418,New Development,Financial application area,Java,Interactive,72,72
13700,IFPUG,352,Enhancement,Process Control,ASP,SQL SERVER,133,133
10261,IFPUG,946,New Development,Financial application area,Java,Interactive,166,166
15940,IFPUG,66,Enhancement,Financial application area;,Java,Interactive;,15,15
15603,IFPUG,756,Enhancement,Financial application area;,Java,Interactive;,124,124
14487,NESMA,3116,Enhancement,Financial transaction process/accounting,Java,DB2,62,62
12057,IFPUG,907,Enhancement,Financial application area,Java,Interactive,168,168
13319,IFPUG,47,New Development,other: Sales contact management,ASP,ORACLE,113,113
13981,IFPUG,4648,New Development,other: sales promotion tool,Visual Basic,SQL SERVER,895,895
12408,IFPUG,425,Enhancement,Financial application area,Java,Interactive,72,72
14194,IFPUG,210,Enhancement,relatively complex application,4GL,Interactive,15,15

11718,IFPUG,3934,New Development,Transaction/Production System,Java,ORACLE V8i,194,194

G1-Ran1-67

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
10178,COSMIC-FFP,2503,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;Workflow support & management;Process Control;,Java,Yes;,826,826
15720,IFPUG,934,Enhancement,Financial transaction process/accounting;,PL/I,DB2;,44,44
11252,IFPUG,3107,Enhancement,relatively complex application,4GL,Interactive,117,117
15008,IFPUG,626,New Development,Financial application area;,Java,Interactive;,116,116
13034,IFPUG,4295,New Development,other: sales promotion tool,Visual Basic,SQL-Server,422,422
14769,IFPUG,7368,Enhancement,Financial transaction process/accounting;,Java,DB2;,1753,1753
13742,IFPUG,12564,New Development,Trading,HTML,ORACLE,1588,1588
14779,IFPUG,2891,New Development,Management or performance reporting;,Java,DB2;,430,430
15130,IFPUG,4045,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,435,435
11728,IFPUG,6944,New Development,other: Sales contact management,Visual Basic,SQL SERVER,424,424
11100,IFPUG,2240,Enhancement,Financial transaction process/accounting,Java,DB2,539,539
11648,IFPUG,1056,Enhancement,Financial application area,Java,Interactive,165,165
11009,IFPUG,2800,Enhancement,Financial transaction process/accounting,Java,UDB,124,124
10180,IFPUG,2340,New Development,Financial transaction process/accounting,Visual Basic,SQL-Server,309,309
15675,IFPUG,2762,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,297,297
15440,IFPUG,301,Enhancement,Financial application area;,Java,Interactive;,51,51
14578,IFPUG,470,Enhancement,Financial application area;,Java,Interactive;,77,77
13127,IFPUG,7496,New Development,Workflow support & management,ASP,SQL Server7,786,786
15598,IFPUG,1028,Enhancement,Web-based Application;,NET,SQL Server;,347,347
11283,IFPUG,410,Enhancement,Financial application area,Java,Interactive,82,82
11436,IFPUG,9058,New Development,other: Sales contact management,C,ORACLE,599,599
15444,IFPUG,2504,New Development,Financial transaction process/accounting;,Visual Basic,SQL Server 2000;,1236,1236
15123,IFPUG,5200,New Development,Financial transaction process/accounting;,Java,DB2;,775,775
15692,IFPUG,11949,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,1285,1285

14260,IFPUG,3576,New Development,Financial transaction process/accounting,Java,Oracle 8i8,778,778
15137,IFPUG,543,New Development,Financial application area;,Java,Interactive;,92,92
14908,IFPUG,2580,New Development,Web-based Application;,Java,Oracle;,264,264
10358,IFPUG,737,New Development,Web-based Application,Visual Basic,SQL Server,191,191
11149,IFPUG,1472,Enhancement,Financial transaction process/accounting,C,ORACLE,407,407
10427,IFPUG,11372,New Development,Financial transaction process/accounting,SQL,ORACLE,859,859
12178,IFPUG,679,New Development,Financial transaction process/accounting,ASP,MSDE SQL Server 2000,302,302
12078,IFPUG,54,Enhancement,Financial application area,Java,Interactive,17,17
12573,IFPUG,1671,New Development,other: Management system,PL/SQL,ORACLE,1216,1216
11421,IFPUG,36,Enhancement,Financial application area,Java,Interactive,6,6
14372,IFPUG,3857,New Development,Financial transaction process/accounting,Visual C++,NCR;TeraData,1521,1521
11132,IFPUG,278,Enhancement,Financial application area,Java,Interactive,58,58
12175,IFPUG,1214,Enhancement,Financial application area,Java,Interactive,176,176
13369,IFPUG,418,New Development,Financial application area,Java,Interactive,72,72
13700,IFPUG,352,Enhancement,Process Control,ASP,SQL SERVER,133,133
10261,IFPUG,946,New Development,Financial application area,Java,Interactive,166,166
15940,IFPUG,66,Enhancement,Financial application area;,Java,Interactive;,15,15
15603,IFPUG,756,Enhancement,Financial application area;,Java,Interactive;,124,124
14487,NESMA,3116,Enhancement,Financial transaction process/accounting,Java,DB2,62,62
12057,IFPUG,907,Enhancement,Financial application area,Java,Interactive,168,168
13319,IFPUG,47,New Development,other: Sales contact management,ASP,ORACLE,113,113
13981,IFPUG,4648,New Development,other: sales promotion tool,Visual Basic,SQL SERVER,895,895
12408,IFPUG,425,Enhancement,Financial application area,Java,Interactive,72,72
14194,IFPUG,210,Enhancement,relatively complex application,4GL,Interactive,15,15
11718,IFPUG,3934,New Development,Transaction/Production System,Java,ORACLE V8i,194,194
14485,IFPUG,484,New Development,Financial application area,Java,Interactive,78,78
12667,IFPUG,6600,Enhancement,Financial transaction process/accounting,C++,HIRDB,1307,1307
13744,IFPUG,1136,Enhancement,Financial application area,Java,Interactive,160,160
13896,IFPUG,1136,Enhancement,Document management,PL/SQL,Oracle 8,50,50
11160,IFPUG,620,Enhancement,Financial application area,Java,Interactive,94,94

15570,IFPUG,7602,New Development,Workflow support & management;,C#,SQL SERVER;,202,202
15468,IFPUG,9231,New Development,Not specified;,C++,ORACLE;,1171,1171
10802,IFPUG,578,Enhancement,relatively complex application,4GL,Interactive,31,31
14911,IFPUG,319,New Development,Financial application area;,Java,Interactive;,103,103
13718,IFPUG,4750,New Development,other: production management system,Visual Basic,Oracle 8,528,528
11730,IFPUG,5621,Enhancement,Document management;Financial transaction process/accounting Image video or sound processing,COBOL,IDMS-DB,344,344
10173,IFPUG,118,Enchancement,Financial application area;,Java,Interactive,25,25
13375,IFPUG,162,Enhancement,Financial application area,Java,Interactive,28,28
13254,IFPUG,24,Enhancement,Financial application area,Java,Interactive,4,4
11873,IFPUG,1846,Enhancement,relatively complex application,3GL,Interactive,202,202
10566,IFPUG,8580,New Development,Financial transaction process/accounting,SQL,ORACLE,359,359
14560,IFPUG,128,Enhancement,Financial application area;,Java,Interactive;,22,22
11809,IFPUG,655,Enhancement,Financial application area,Java,Interactive,113,113

G1-Ran2-17

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
15720,IFPUG,934,Enhancement,Financial transaction process/accounting;,PL/I,DB2;,44,44
15008,IFPUG,626,New Development,Financial application area;,Java,Interactive;,116,116
14779,IFPUG,2891,New Development,Management or performance reporting;,Java,DB2;,430,430
10180,IFPUG,2340,New Development,Financial transaction process/accounting,Visual Basic,SQL-Server,309,309
15440,IFPUG,301,Enhancement,Financial application area;,Java,Interactive;,51,51
11283,IFPUG,410,Enhancement,Financial application area,Java,Interactive,82,82
14260,IFPUG,3576,New Development,Financial transaction process/accounting,Java,Oracle 8i8,778,778
10358,IFPUG,737,New Development,Web-based Application,Visual Basic,SQL Server,191,191
12078,IFPUG,54,Enhancement,Financial application area,Java,Interactive,17,17
11132,IFPUG,278,Enhancement,Financial application area,Java,Interactive,58,58
15603,IFPUG,756,Enhancement,Financial application area;,Java,Interactive;,124,124
14487,NESMA,3116,Enhancement,Financial transaction process/accounting,Java,DB2,62,62

13896,IFPUG,1136,Enhancement,Document management,PL/SQL,Oracle 8,50,50
15468,IFPUG,9231,New Development,Not specified;,C++,ORACLE;,1171,1171
11730,IFPUG,5621,Enhancement,Document management;Financial transaction process/accounting Image video or sound processing,COBOL,IDMS-DB,344,344
10566,IFPUG,8580,New Development,Financial transaction process/accounting,SQL,ORACLE,359,359
11809,IFPUG,655,Enhancement,Financial application area,Java,Interactive,113,113

G1-Ran2-33

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
15720,IFPUG,934,Enhancement,Financial transaction process/accounting;,PL/I,DB2;,44,44
15008,IFPUG,626,New Development,Financial application area;,Java,Interactive;,116,116
13034,IFPUG,4295,New Development,other: sales promotion tool,Visual Basic,SQL-Server,422,422
14779,IFPUG,2891,New Development,Management or performance reporting;,Java,DB2;,430,430
11100,IFPUG,2240,Enhancement,Financial transaction process/accounting,Java,DB2,539,539
11648,IFPUG,1056,Enhancement,Financial application area,Java,Interactive,165,165
10180,IFPUG,2340,New Development,Financial transaction process/accounting,Visual Basic,SQL-Server,309,309
15440,IFPUG,301,Enhancement,Financial application area;,Java,Interactive;,51,51
13127,IFPUG,7496,New Development,Workflow support & management,ASP,SQL Server7,786,786
11283,IFPUG,410,Enhancement,Financial application area,Java,Interactive,82,82
15444,IFPUG,2504,New Development,Financial transaction process/accounting;,Visual Basic,SQL Server 2000;,1236,1236
14260,IFPUG,3576,New Development,Financial transaction process/accounting,Java,Oracle 8i8,778,778
15137,IFPUG,543,New Development,Financial application area;,Java,Interactive;,92,92
10358,IFPUG,737,New Development,Web-based Application,Visual Basic,SQL Server,191,191
10427,IFPUG,11372,New Development,Financial transaction process/accounting,SQL,ORACLE,859,859
12078,IFPUG,54,Enhancement,Financial application area,Java,Interactive,17,17
11421,IFPUG,36,Enhancement,Financial application area,Java,Interactive,6,6
11132,IFPUG,278,Enhancement,Financial application area,Java,Interactive,58,58
13369,IFPUG,418,New Development,Financial application area,Java,Interactive,72,72
13700,IFPUG,352,Enhancement,Process Control,ASP,SQL SERVER,133,133

15603,IFPUG,756,Enhancement,Financial application area;,Java,Interactive;,124,124
14487,NESMA,3116,Enhancement,Financial transaction process/accounting,Java,DB2,62,62
13319,IFPUG,47,New Development,other: Sales contact management,ASP,ORACLE,113,113
12408,IFPUG,425,Enhancement,Financial application area,Java,Interactive,72,72
11718,IFPUG,3934,New Development,Transaction/Production System,Java,ORACLE V8i,194,194
13744,IFPUG,1136,Enhancement,Financial application area,Java,Interactive,160,160
13896,IFPUG,1136,Enhancement,Document management,PL/SQL,Oracle 8,50,50
15468,IFPUG,9231,New Development,Not specified;,C++,ORACLE;,1171,1171
14911,IFPUG,319,New Development,Financial application area;,Java,Interactive;,103,103
11730,IFPUG,5621,Enhancement,Document management;Financial transaction process/accounting Image video or sound processing,COBOL,IDMS-DB,344,344
13254,IFPUG,24,Enhancement,Financial application area,Java,Interactive,4,4
10566,IFPUG,8580,New Development,Financial transaction process/accounting,SQL,ORACLE,359,359
11809,IFPUG,655,Enhancement,Financial application area,Java,Interactive,113,113

G1-Ran2-49

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
10178,COSMIC-FFP,2503,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;Workflow support & management;Process Control;,Java,Yes;,826,826
15720,IFPUG,934,Enhancement,Financial transaction process/accounting;,PL/I,DB2;,44,44
15008,IFPUG,626,New Development,Financial application area;,Java,Interactive;,116,116
13034,IFPUG,4295,New Development,other: sales promotion tool,Visual Basic,SQL-Server,422,422
13742,IFPUG,12564,New Development,Trading,HTML,ORACLE,1588,1588
14779,IFPUG,2891,New Development,Management or performance reporting;,Java,DB2;,430,430
11728,IFPUG,6944,New Development,other: Sales contact management,Visual Basic,SQL SERVER,424,424
11100,IFPUG,2240,Enhancement,Financial transaction process/accounting,Java,DB2,539,539
11648,IFPUG,1056,Enhancement,Financial application area,Java,Interactive,165,165
10180,IFPUG,2340,New Development,Financial transaction process/accounting,Visual Basic,SQL-Server,309,309
15675,IFPUG,2762,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,297,297

15440,IFPUG,301,Enhancement,Financial application area;,Java,Interactive;,51,51
13127,IFPUG,7496,New Development,Workflow support & management,ASP,SQL Server7,786,786
15598,IFPUG,1028,Enhancement,Web-based Application;,NET,SQL Server;,347,347
11283,IFPUG,410,Enhancement,Financial application area,Java,Interactive,82,82
15444,IFPUG,2504,New Development,Financial transaction process/accounting;,Visual Basic,SQL Server 2000;,1236,1236
15123,IFPUG,5200,New Development,Financial transaction process/accounting;,Java,DB2;,775,775
14260,IFPUG,3576,New Development,Financial transaction process/accounting,Java,Oracle 8i8,778,778
15137,IFPUG,543,New Development,Financial application area;,Java,Interactive;,92,92
14908,IFPUG,2580,New Development,Web-based Application;,Java,Oracle;,264,264
10358,IFPUG,737,New Development,Web-based Application,Visual Basic,SQL Server,191,191
10427,IFPUG,11372,New Development,Financial transaction process/accounting,SQL,ORACLE,859,859
12178,IFPUG,679,New Development,Financial transaction process/accounting,ASP,MSDE SQL Server 2000,302,302
12078,IFPUG,54,Enhancement,Financial application area,Java,Interactive,17,17
11421,IFPUG,36,Enhancement,Financial application area,Java,Interactive,6,6
14372,IFPUG,3857,New Development,Financial transaction process/accounting,Visual C++,NCR;TeraData,1521,1521
11132,IFPUG,278,Enhancement,Financial application area,Java,Interactive,58,58
13369,IFPUG,418,New Development,Financial application area,Java,Interactive,72,72
13700,IFPUG,352,Enhancement,Process Control,ASP,SQL SERVER,133,133
10261,IFPUG,946,New Development,Financial application area,Java,Interactive,166,166
15603,IFPUG,756,Enhancement,Financial application area;,Java,Interactive;,124,124
14487,NESMA,3116,Enhancement,Financial transaction process/accounting,Java,DB2,62,62
13319,IFPUG,47,New Development,other: Sales contact management,ASP,ORACLE,113,113
13981,IFPUG,4648,New Development,other: sales promotion tool,Visual Basic,SQL SERVER,895,895
12408,IFPUG,425,Enhancement,Financial application area,Java,Interactive,72,72
11718,IFPUG,3934,New Development,Transaction/Production System,Java,ORACLE V8i,194,194
14485,IFPUG,484,New Development,Financial application area,Java,Interactive,78,78
13744,IFPUG,1136,Enhancement,Financial application area,Java,Interactive,160,160
13896,IFPUG,1136,Enhancement,Document management,PL/SQL,Oracle 8,50,50
15570,IFPUG,7602,New Development,Workflow support & management;,C#,SQL SERVER;,202,202
15468,IFPUG,9231,New Development,Not specified;,C++,ORACLE;,1171,1171

14911,IFPUG,319,New Development,Financial application area;,Java,Interactive;,103,103
13718,IFPUG,4750,New Development,other: production management system,Visual Basic,Oracle 8,528,528
11730,IFPUG,5621,Enhancement,Document management;Financial transaction process/accounting Image video or sound processing,COBOL,IDMS-DB,344,344
13375,IFPUG,162,Enhancement,Financial application area,Java,Interactive,28,28
13254,IFPUG,24,Enhancement,Financial application area,Java,Interactive,4,4
10566,IFPUG,8580,New Development,Financial transaction process/accounting,SQL,ORACLE,359,359
14560,IFPUG,128,Enhancement,Financial application area;,Java,Interactive;,22,22
11809,IFPUG,655,Enhancement,Financial application area,Java,Interactive,113,113

G1-Ran2-67

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
10178,COSMIC-FFP,2503,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;Workflow support & management;Process Control;,Java,Yes;,826,826
15720,IFPUG,934,Enhancement,Financial transaction process/accounting;,PL/I,DB2;,44,44
11252,IFPUG,3107,Enhancement,relatively complex application,4GL,Interactive,117,117
15008,IFPUG,626,New Development,Financial application area;,Java,Interactive;,116,116
13034,IFPUG,4295,New Development,other: sales promotion tool,Visual Basic,SQL-Server,422,422
14769,IFPUG,7368,Enhancement,Financial transaction process/accounting;,Java,DB2;,1753,1753
13742,IFPUG,12564,New Development,Trading,HTML,ORACLE,1588,1588
14779,IFPUG,2891,New Development,Management or performance reporting;,Java,DB2;,430,430
15130,IFPUG,4045,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,435,435
11728,IFPUG,6944,New Development,other: Sales contact management,Visual Basic,SQL SERVER,424,424
11100,IFPUG,2240,Enhancement,Financial transaction process/accounting,Java,DB2,539,539
11648,IFPUG,1056,Enhancement,Financial application area,Java,Interactive,165,165
11009,IFPUG,2800,Enhancement,Financial transaction process/accounting,Java,UDB,124,124
10180,IFPUG,2340,New Development,Financial transaction process/accounting,Visual Basic,SQL-Server,309,309
15675,IFPUG,2762,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,297,297
15440,IFPUG,301,Enhancement,Financial application area;,Java,Interactive;,51,51

14578,IFPUG,470,Enhancement,Financial application area;,Java,Interactive;,77,77
13127,IFPUG,7496,New Development,Workflow support & management,ASP,SQL Server7,786,786
15598,IFPUG,1028,Enhancement,Web-based Application;,NET,SQL Server;,347,347
11283,IFPUG,410,Enhancement,Financial application area,Java,Interactive,82,82
11436,IFPUG,9058,New Development,other: Sales contact management,C,ORACLE,599,599
15444,IFPUG,2504,New Development,Financial transaction process/accounting;,Visual Basic,SQL Server 2000;,1236,1236
15123,IFPUG,5200,New Development,Financial transaction process/accounting;,Java,DB2;,775,775
15692,IFPUG,11949,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,1285,1285
14260,IFPUG,3576,New Development,Financial transaction process/accounting,Java,Oracle 8i8,778,778
15137,IFPUG,543,New Development,Financial application area;,Java,Interactive;,92,92
14908,IFPUG,2580,New Development,Web-based Application;,Java,Oracle;,264,264
10358,IFPUG,737,New Development,Web-based Application,Visual Basic,SQL Server,191,191
11149,IFPUG,1472,Enhancement,Financial transaction process/accounting,C,ORACLE,407,407
10427,IFPUG,11372,New Development,Financial transaction process/accounting,SQL,ORACLE,859,859
12178,IFPUG,679,New Development,Financial transaction process/accounting,ASP,MSDE SQL Server 2000,302,302
12078,IFPUG,54,Enhancement,Financial application area,Java,Interactive,17,17
12573,IFPUG,1671,New Development,other: Management system,PL/SQL,ORACLE,1216,1216
11421,IFPUG,36,Enhancement,Financial application area,Java,Interactive,6,6
14372,IFPUG,3857,New Development,Financial transaction process/accounting,Visual C++,NCR;TeraData,1521,1521
11132,IFPUG,278,Enhancement,Financial application area,Java,Interactive,58,58
12175,IFPUG,1214,Enhancement,Financial application area,Java,Interactive,176,176
13369,IFPUG,418,New Development,Financial application area,Java,Interactive,72,72
13700,IFPUG,352,Enhancement,Process Control,ASP,SQL SERVER,133,133
10261,IFPUG,946,New Development,Financial application area,Java,Interactive,166,166
15940,IFPUG,66,Enhancement,Financial application area;,Java,Interactive;,15,15
15603,IFPUG,756,Enhancement,Financial application area;,Java,Interactive;,124,124
14487,NESMA,3116,Enhancement,Financial transaction process/accounting,Java,DB2,62,62
12057,IFPUG,907,Enhancement,Financial application area,Java,Interactive,168,168#
13319,IFPUG,47,New Development,other: Sales contact management,ASP,ORACLE,113,113
13981,IFPUG,4648,New Development,other: sales promotion tool,Visual Basic,SQL SERVER,895,895

12408,IFPUG,425,Enhancement,Financial application area,Java,Interactive,72,72
14194,IFPUG,210,Enhancement,relatively complex application,4GL,Interactive,15,15
11718,IFPUG,3934,New Development,Transaction/Production System,Java,ORACLE V8i,194,194
14485,IFPUG,484,New Development,Financial application area,Java,Interactive,78,78
12667,IFPUG,6600,Enhancement,Financial transaction process/accounting,C++,HIRDB,1307,1307
13744,IFPUG,1136,Enhancement,Financial application area,Java,Interactive,160,160
13896,IFPUG,1136,Enhancement,Document management,PL/SQL,Oracle 8,50,50
11160,IFPUG,620,Enhancement,Financial application area,Java,Interactive,94,94
15570,IFPUG,7602,New Development,Workflow support & management;,C#,SQL SERVER;,202,202
15468,IFPUG,9231,New Development,Not specified;,C++,ORACLE;,1171,1171
10802,IFPUG,578,Enhancement,relatively complex application,4GL,Interactive,31,31
14911,IFPUG,319,New Development,Financial application area;,Java,Interactive;,103,103
13718,IFPUG,4750,New Development,other: production management system,Visual Basic,Oracle 8,528,528
11730,IFPUG,5621,Enhancement,Document management;Financial transaction process/accounting Image video or sound processing,COBOL,IDMS-DB,344,344
10173,IFPUG,118,Enchancement,Financial application area;,Java,Interactive,25,25
13375,IFPUG,162,Enhancement,Financial application area,Java,Interactive,28,28
13254,IFPUG,24,Enhancement,Financial application area,Java,Interactive,4,4
11873,IFPUG,1846,Enhancement,relatively complex application,3GL,Interactive,202,202
10566,IFPUG,8580,New Development,Financial transaction process/accounting,SQL,ORACLE,359,359
14560,IFPUG,128,Enhancement,Financial application area;,Java,Interactive;,22,22
11809,IFPUG,655,Enhancement,Financial application area,Java,Interactive,113,113

G1-Ran3-17

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
10178,COSMIC-FFP,2503,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;Workflow support & management;Process Control;,Java,Yes;,826,826
15130,IFPUG,4045,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,435,435
11648,IFPUG,1056,Enhancement,Financial application area,Java,Interactive,165,165

15675,IFPUG,2762,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,297,297
13127,IFPUG,7496,New Development,Workflow support & management,ASP,SQL Server7,786,786
15444,IFPUG,2504,New Development,Financial transaction process/accounting;,Visual Basic,SQL Server 2000;,1236,1236
14260,IFPUG,3576,New Development,Financial transaction process/accounting,Java,Oracle 8i8,778,778
15137,IFPUG,543,New Development,Financial application area;,Java,Interactive;,92,92
12573,IFPUG,1671,New Development,other: Management system,PL/SQL,ORACLE,1216,1216
14372,IFPUG,3857,New Development,Financial transaction process/accounting,Visual C++,NCR;TeraData,1521,1521
13700,IFPUG,352,Enhancement,Process Control,ASP,SQL SERVER,133,133
13981,IFPUG,4648,New Development,other: sales promotion tool,Visual Basic,SQL SERVER,895,895
14485,IFPUG,484,New Development,Financial application area,Java,Interactive,78,78
10802,IFPUG,578,Enhancement,relatively complex application,4GL,Interactive,31,31
13718,IFPUG,4750,New Development,other: production management system,Visual Basic,Oracle 8,528,528
13254,IFPUG,24,Enhancement,Financial application area,Java,Interactive,4,4
11809,IFPUG,655,Enhancement,Financial application area,Java,Interactive,113,113

G1-Ran3-33

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
10178,COSMIC-FFP,2503,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;Workflow support & management;Process Control;,Java,Yes;,826,826
13034,IFPUG,4295,New Development,other: sales promotion tool,Visual Basic,SQL-Server,422,422
15130,IFPUG,4045,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,435,435
11648,IFPUG,1056,Enhancement,Financial application area,Java,Interactive,165,165
15675,IFPUG,2762,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,297,297
15440,IFPUG,301,Enhancement,Financial application area;,Java,Interactive;,51,51
13127,IFPUG,7496,New Development,Workflow support & management,ASP,SQL Server7,786,786
11283,IFPUG,410,Enhancement,Financial application area,Java,Interactive,82,82
15444,IFPUG,2504,New Development,Financial transaction process/accounting;,Visual Basic,SQL Server 2000;,1236,1236
15123,IFPUG,5200,New Development,Financial transaction process/accounting;,Java,DB2;,775,775

14260,IFPUG,3576,New Development,Financial transaction process/accounting,Java,Oracle 8i8,778,778
15137,IFPUG,543,New Development,Financial application area;,Java,Interactive;,92,92
10427,IFPUG,11372,New Development,Financial transaction process/accounting,SQL,ORACLE,859,859
12078,IFPUG,54,Enhancement,Financial application area,Java,Interactive,17,17
12573,IFPUG,1671,New Development,other: Management system,PL/SQL,ORACLE,1216,1216
14372,IFPUG,3857,New Development,Financial transaction process/accounting,Visual C++,NCR;TeraData,1521,1521
12175,IFPUG,1214,Enhancement,Financial application area,Java,Interactive,176,176
13700,IFPUG,352,Enhancement,Process Control,ASP,SQL SERVER,133,133
15940,IFPUG,66,Enhancement,Financial application area;,Java,Interactive;,15,15
15603,IFPUG,756,Enhancement,Financial application area;,Java,Interactive;,124,124
12057,IFPUG,907,Enhancement,Financial application area,Java,Interactive,168,168
13981,IFPUG,4648,New Development,other: sales promotion tool,Visual Basic,SQL SERVER,895,895
14194,IFPUG,210,Enhancement,relatively complex application,4GL,Interactive,15,15
14485,IFPUG,484,New Development,Financial application area,Java,Interactive,78,78
13744,IFPUG,1136,Enhancement,Financial application area,Java,Interactive,160,160
11160,IFPUG,620,Enhancement,Financial application area,Java,Interactive,94,94
15468,IFPUG,9231,New Development,Not specified;,C++,ORACLE;,1171,1171
10802,IFPUG,578,Enhancement,relatively complex application,4GL,Interactive,31,31
13718,IFPUG,4750,New Development,other: production management system,Visual Basic,Oracle 8,528,528
10173,IFPUG,118,Enchancement,Financial application area;,Java,Interactive,25,25
13254,IFPUG,24,Enhancement,Financial application area,Java,Interactive,4,4
10566,IFPUG,8580,New Development,Financial transaction process/accounting,SQL,ORACLE,359,359
11809,IFPUG,655,Enhancement,Financial application area,Java,Interactive,113,113

G1-Ran3-49

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
10178,COSMIC-FFP,2503,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;Workflow support & management;Process Control;,Java,Yes;,826,826
11252,IFPUG,3107,Enhancement,relatively complex application,4GL,Interactive,117,117

13034,IFPUG,4295,New Development,other: sales promotion tool,Visual Basic,SQL-Server,422,422
13742,IFPUG,12564,New Development,Trading,HTML,ORACLE,1588,1588
15130,IFPUG,4045,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,435,435
11728,IFPUG,6944,New Development,other: Sales contact management,Visual Basic,SQL SERVER,424,424
11648,IFPUG,1056,Enhancement,Financial application area,Java,Interactive,165,165
11009,IFPUG,2800,Enhancement,Financial transaction process/accounting,Java,UDB,124,124
15675,IFPUG,2762,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,297,297
15440,IFPUG,301,Enhancement,Financial application area;,Java,Interactive;,51,51
13127,IFPUG,7496,New Development,Workflow support & management,ASP,SQL Server7,786,786
11283,IFPUG,410,Enhancement,Financial application area,Java,Interactive,82,82
15444,IFPUG,2504,New Development,Financial transaction process/accounting;,Visual Basic,SQL Server 2000;,1236,1236
15123,IFPUG,5200,New Development,Financial transaction process/accounting;,Java,DB2;,775,775
14260,IFPUG,3576,New Development,Financial transaction process/accounting,Java,Oracle 8i8,778,778
15137,IFPUG,543,New Development,Financial application area;,Java,Interactive;,92,92
10427,IFPUG,11372,New Development,Financial transaction process/accounting,SQL,ORACLE,859,859
12078,IFPUG,54,Enhancement,Financial application area,Java,Interactive,17,17
12573,IFPUG,1671,New Development,other: Management system,PL/SQL,ORACLE,1216,1216
10358,IFPUG,737,New Development,Web-based Application,Visual Basic,SQL Server,191,191
14372,IFPUG,3857,New Development,Financial transaction process/accounting,Visual C++,NCR;TeraData,1521,1521
12175,IFPUG,1214,Enhancement,Financial application area,Java,Interactive,176,176
13700,IFPUG,352,Enhancement,Process Control,ASP,SQL SERVER,133,133
10261,IFPUG,946,New Development,Financial application area,Java,Interactive,166,166
15940,IFPUG,66,Enhancement,Financial application area;,Java,Interactive;,15,15
15603,IFPUG,756,Enhancement,Financial application area;,Java,Interactive;,124,124
14487,NESMA,3116,Enhancement,Financial transaction process/accounting,Java,DB2,62,62
12057,IFPUG,907,Enhancement,Financial application area,Java,Interactive,168,168
13319,IFPUG,47,New Development,other: Sales contact management,ASP,ORACLE,113,113
13981,IFPUG,4648,New Development,other: sales promotion tool,Visual Basic,SQL SERVER,895,895
12408,IFPUG,425,Enhancement,Financial application area,Java,Interactive,72,72
14194,IFPUG,210,Enhancement,relatively complex application,4GL,Interactive,15,15

11718,IFPUG,3934,New Development,Transaction/Production System,Java,ORACLE V8i,194,194
14485,IFPUG,484,New Development,Financial application area,Java,Interactive,78,78
12667,IFPUG,6600,Enhancement,Financial transaction process/accounting,C++,HIRDB,1307,1307
13744,IFPUG,1136,Enhancement,Financial application area,Java,Interactive,160,160
13896,IFPUG,1136,Enhancement,Document management,PL/SQL,Oracle 8,50,50
11160,IFPUG,620,Enhancement,Financial application area,Java,Interactive,94,94
15468,IFPUG,9231,New Development,Not specified;,C++,ORACLE;,1171,1171
10802,IFPUG,578,Enhancement,relatively complex application,4GL,Interactive,31,31
13718,IFPUG,4750,New Development,other: production management system,Visual Basic,Oracle 8,528,528
11730,IFPUG,5621,Enhancement,Document management;Financial transaction process/accounting Image video or sound processing,COBOL,IDMS-DB,344,344
10173,IFPUG,118,Enchancement,Financial application area;,Java,Interactive,25,25
13375,IFPUG,162,Enhancement,Financial application area,Java,Interactive,28,28
13254,IFPUG,24,Enhancement,Financial application area,Java,Interactive,4,4
11873,IFPUG,1846,Enhancement,relatively complex application,3GL,Interactive,202,202
10566,IFPUG,8580,New Development,Financial transaction process/accounting,SQL,ORACLE,359,359
14560,IFPUG,128,Enhancement,Financial application area;,Java,Interactive;,22,22
11809,IFPUG,655,Enhancement,Financial application area,Java,Interactive,113,113

G1-Ran3-67

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
10178,COSMIC-FFP,2503,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;Workflow support & management;Process Control;,Java,Yes;,826,826
15720,IFPUG,934,Enhancement,Financial transaction process/accounting;,PL/I,DB2;,44,44
11252,IFPUG,3107,Enhancement,relatively complex application,4GL,Interactive,117,117
15008,IFPUG,626,New Development,Financial application area;,Java,Interactive;,116,116
13034,IFPUG,4295,New Development,other: sales promotion tool,Visual Basic,SQL-Server,422,422
14769,IFPUG,7368,Enhancement,Financial transaction process/accounting;,Java,DB2;,1753,1753
13742,IFPUG,12564,New Development,Trading,HTML,ORACLE,1588,1588

14779,IFPUG,2891,New Development,Management or performance reporting;,Java,DB2;,430,430
15130,IFPUG,4045,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,435,435
11728,IFPUG,6944,New Development,other: Sales contact management,Visual Basic,SQL SERVER,424,424
11100,IFPUG,2240,Enhancement,Financial transaction process/accounting,Java,DB2,539,539
11009,IFPUG,2800,Enhancement,Financial transaction process/accounting,Java,UDB,124,124
11648,IFPUG,1056,Enhancement,Financial application area,Java,Interactive,165,165
10180,IFPUG,2340,New Development,Financial transaction process/accounting,Visual Basic,SQL-Server,309,309
15675,IFPUG,2762,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,297,297
15440,IFPUG,301,Enhancement,Financial application area;,Java,Interactive;,51,51
14578,IFPUG,470,Enhancement,Financial application area;,Java,Interactive;,77,77
13127,IFPUG,7496,New Development,Workflow support & management,ASP,SQL Server7,786,786
15598,IFPUG,1028,Enhancement,Web-based Application;,NET,SQL Server;,347,347
11283,IFPUG,410,Enhancement,Financial application area,Java,Interactive,82,82
11436,IFPUG,9058,New Development,other: Sales contact management,C,ORACLE,599,599
15444,IFPUG,2504,New Development,Financial transaction process/accounting;,Visual Basic,SQL Server 2000;,1236,1236
15123,IFPUG,5200,New Development,Financial transaction process/accounting;,Java,DB2;,775,775
15692,IFPUG,11949,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,1285,1285
14260,IFPUG,3576,New Development,Financial transaction process/accounting,Java,Oracle 8i8,778,778
15137,IFPUG,543,New Development,Financial application area;,Java,Interactive;,92,92
14908,IFPUG,2580,New Development,Web-based Application;,Java,Oracle;,264,264
10358,IFPUG,737,New Development,Web-based Application,Visual Basic,SQL Server,191,191
11149,IFPUG,1472,Enhancement,Financial transaction process/accounting,C,ORACLE,407,407
10427,IFPUG,11372,New Development,Financial transaction process/accounting,SQL,ORACLE,859,859
12178,IFPUG,679,New Development,Financial transaction process/accounting,ASP,MSDE SQL Server 2000,302,302
12078,IFPUG,54,Enhancement,Financial application area,Java,Interactive,17,17
12573,IFPUG,1671,New Development,other: Management system,PL/SQL,ORACLE,1216,1216
11421,IFPUG,36,Enhancement,Financial application area,Java,Interactive,6,6
14372,IFPUG,3857,New Development,Financial transaction process/accounting,Visual C++,NCR;TeraData,1521,1521
11132,IFPUG,278,Enhancement,Financial application area,Java,Interactive,58,58
12175,IFPUG,1214,Enhancement,Financial application area,Java,Interactive,176,176

13369,IFPUG,418,New Development,Financial application area,Java,Interactive,72,72
13700,IFPUG,352,Enhancement,Process Control,ASP,SQL SERVER,133,133
10261,IFPUG,946,New Development,Financial application area,Java,Interactive,166,166
15940,IFPUG,66,Enhancement,Financial application area;,Java,Interactive;,15,15
15603,IFPUG,756,Enhancement,Financial application area;,Java,Interactive;,124,124
14487,NESMA,3116,Enhancement,Financial transaction process/accounting,Java,DB2,62,62
12057,IFPUG,907,Enhancement,Financial application area,Java,Interactive,168,168
13319,IFPUG,47,New Development,other: Sales contact management,ASP,ORACLE,113,113
13981,IFPUG,4648,New Development,other: sales promotion tool,Visual Basic,SQL SERVER,895,895
12408,IFPUG,425,Enhancement,Financial application area,Java,Interactive,72,72
14194,IFPUG,210,Enhancement,relatively complex application,4GL,Interactive,15,15
11718,IFPUG,3934,New Development,Transaction/Production System,Java,ORACLE V8i,194,194
14485,IFPUG,484,New Development,Financial application area,Java,Interactive,78,78
12667,IFPUG,6600,Enhancement,Financial transaction process/accounting,C++,HIRDB,1307,1307
13744,IFPUG,1136,Enhancement,Financial application area,Java,Interactive,160,160
13896,IFPUG,1136,Enhancement,Document management,PL/SQL,Oracle 8,50,50
11160,IFPUG,620,Enhancement,Financial application area,Java,Interactive,94,94
15570,IFPUG,7602,New Development,Workflow support & management;,C#,SQL SERVER;,202,202
15468,IFPUG,9231,New Development,Not specified;,C++,ORACLE;,1171,1171
10802,IFPUG,578,Enhancement,relatively complex application,4GL,Interactive,31,31
14911,IFPUG,319,New Development,Financial application area;,Java,Interactive;,103,103
13718,IFPUG,4750,New Development,other: production management system,Visual Basic,Oracle 8,528,528
11730,IFPUG,5621,Enhancement,Document management;Financial transaction process/accounting Image video or sound processing,COBOL,IDMS-DB,344,344
10173,IFPUG,118,Enchancement,Financial application area;,Java,Interactive,25,25
13375,IFPUG,162,Enhancement,Financial application area,Java,Interactive,28,28
13254,IFPUG,24,Enhancement,Financial application area,Java,Interactive,4,4
11873,IFPUG,1846,Enhancement,relatively complex application,3GL,Interactive,202,202
10566,IFPUG,8580,New Development,Financial transaction process/accounting,SQL,ORACLE,359,359
14560,IFPUG,128,Enhancement,Financial application area;,Java,Interactive;,22,22

11809,IFPUG,655,Enhancement,Financial application area,Java,Interactive,113,113

G2-Ran1-17

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
16023,IFPUG,525,Enhancement,Financial application area;,Java,Interactive;,101,101
16917,IFPUG,21700,Enhancement,Document mngt;Fin trans process/acc;Image video or sound processing ,COBOL , IDMS-DB,500,500
17227,IFPUG,953,Enhancement,Financial application area;,Java,Interactive;,127,127
17855,IFPUG,252,Enhancement,Financial application area;,Java,Interactive;,42,42
18019,IFPUG,155,Enhancement,Financial transaction process/accounting;,C,ORACLE;,19,19
18398,IFPUG,14992,New Development,Customer billing/relationship management;,HTML,ORACLE;,694,694
18452,IFPUG,750,New Development,Financial application area;,Java,Interactive;,121,121
18956,IFPUG,9456,New Development,Process Control;,Java,Oracle 8i;,931,1182
19659,IFPUG,354,Enhancement,Financial application area;,Java,Interactive;,61,61
19829,IFPUG,2676,Enhancement,Financial transaction process/accounting;,Java,Interactive;,380,353
19990,IFPUG,742,New Development,other: personnel system ;,Java,DB2;,246,246
20409,IFPUG,2803,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,18,18
20591,IFPUG,908,Enhancement,Financial application area;,Java,Interactive;,89,89
21114,IFPUG,175,Enhancement,Financial application area;,Java,Interactive;,33,33
21351,IFPUG,1571,Enhancement,relatively complex application;,4GL,Interactive;,82,82
22168,IFPUG,730,Enhancement,Financial transaction process/accounting;Online analysis and reporting;Space management of schools;,Active Server Pages 2.0 etc,Microsoft SQL Server 2000;,848,822

G2-Ran1-33

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
16023,IFPUG,525,Enhancement,Financial application area;,Java,Interactive;,101,101
16612,IFPUG,465,Enhancement,Financial transaction process/accounting;,PL/SQL,Oracle 8;,162,162
16917,IFPUG,21700,Enhancement,Document mngt;Fin trans process/acc;Image video or sound processing ,COBOL , IDMS-DB,500,500
17227,IFPUG,953,Enhancement,Financial application area;,Java,Interactive;,127,127

17614,IFPUG,3303,Enhancement,Catalogue/register of things or events;Customer billing/relationship management;,COOL:GEN,DB2;,128,128
17855,IFPUG,252,Enhancement,Financial application area;,Java,Interactive;,42,42
18019,IFPUG,155,Enhancement,Financial transaction process/accounting;,C,ORACLE;,19,19
18047,IFPUG,331,New Development,Financial application area;,Java,Interactive;,72,72
18398,IFPUG,14992,New Development,Customer billing/relationship management;,HTML,ORACLE;,694,694
18452,IFPUG,750,New Development,Financial application area;,Java,Interactive;,121,121
18590,IFPUG,61,Enhancement,Financial application area;,Java,Interactive;,11,11
18705,IFPUG,1009,Enhancement,relatively complex application;,4GL,Interactive;,227,227
18956,IFPUG,9456,New Development,Process Control;,Java,Oracle 8i;,931,1182
19107,IFPUG,1712,Enhancement,relatively complex application;,4GL,Interactive;,89,89
19659,IFPUG,354,Enhancement,Financial application area;,Java,Interactive;,61,61
19757,IFPUG,1536,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,127,127
19802,IFPUG,3358,New Development,Web-based Application;,Java,SQL Server;,200,200
19829,IFPUG,2676,Enhancement,Financial transaction process/accounting;,Java,Interactive;,380,353
19990,IFPUG,742,New Development,other: personnel system ;,Java,DB2;,246,246
20104,IFPUG,2000,New Development,other: Sales contact management;,Java,Oracle 8;,473,473
20145,IFPUG,51,Enhancement,Financial application area;,Java,Interactive;,9,9
20409,IFPUG,2803,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,18,18
20487,IFPUG,429,Enhancement,relatively complex application;,4GL,Interactive;,11,11
20591,IFPUG,908,Enhancement,Financial application area;,Java,Interactive;,89,89
21114,IFPUG,175,Enhancement,Financial application area;,Java,Interactive;,33,33
21191,IFPUG,312,New Development,tools or system;,Coldfusion,SQL SERVER;,96,96
21351,IFPUG,1571,Enhancement,relatively complex application;,4GL,Interactive;,82,82
21528,IFPUG,291,Enhancement,relatively complex application;,4GL,Interactive;,22,22
21596,IFPUG,29,Enhancement,Financial application area;,Java,Interactive;,6,6
21816,IFPUG,940,Enhancement,Financial application area;,Java,Interactive;,162,162
22168,IFPUG,730,Enhancement,Financial transaction process/accounting;Online analysis and reporting;Space management of schools;,Active Server Pages 2.0 etc,Microsoft SQL Server 2000;,848,822
22403,IFPUG,480,Enhancement,relatively complex application;,3GL,Interactive;,32,32

22476,IFPUG,2592,Enhancement,Financial transaction process/accounting;,C,SYBASE;,115,115

G2-Ran1-49

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
16023,IFPUG,525,Enhancement,Financial application area;,Java,Interactive;,101,101
16332,IFPUG,2100,New Development,Financial transaction process/accounting;,ASP,SQL SERVER;,257,257
16612,IFPUG,465,Enhancement,Financial transaction process/accounting;,PL/SQL,Oracle 8;,162,162
16917,IFPUG,21700,Enhancement,Document mngt;Fin trans process/acc;Image video or sound processing ,COBOL , IDMS-DB,500,500
17227,IFPUG,953,Enhancement,Financial application area;,Java,Interactive;,127,127
17614,IFPUG,3303,Enhancement,Catalogue/register of things or events;Customer billing/relationship management;,COOL:GEN,DB2;,128,128
17855,IFPUG,252,Enhancement,Financial application area;,Java,Interactive;,42,42
18019,IFPUG,155,Enhancement,Financial transaction process/accounting;,C,ORACLE;,19,19
18047,IFPUG,331,New Development,Financial application area;,Java,Interactive;,72,72
18398,IFPUG,14992,New Development,Customer billing/relationship management;,HTML,ORACLE;,694,694
18452,IFPUG,750,New Development,Financial application area;,Java,Interactive;,121,121
18590,IFPUG,61,Enhancement,Financial application area;,Java,Interactive;,11,11
18705,IFPUG,1009,Enhancement,relatively complex application;,4GL,Interactive;,227,227
18956,IFPUG,9456,New Development,Process Control;,Java,Oracle 8i;,931,1182
19107,IFPUG,1712,Enhancement,relatively complex application;,4GL,Interactive;,89,89
19659,IFPUG,354,Enhancement,Financial application area;,Java,Interactive;,61,61
19757,IFPUG,1536,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,127,127
19802,IFPUG,3358,New Development,Web-based Application;,Java,SQL Server;,200,200
19829,IFPUG,2676,Enhancement,Financial transaction process/accounting;,Java,Interactive;,380,353
19990,IFPUG,742,New Development,other: personnel system ;,Java,DB2;,246,246
19997,IFPUG,419,Enhancement,Financial application area;,Java,Interactive;,91,91
20104,IFPUG,2000,New Development,other: Sales contact management;,Java,Oracle 8;,473,473
20117,IFPUG,3982,New Development,Financial transaction process/accounting;,HTML,HiRDB;,190,190
20145,IFPUG,51,Enhancement,Financial application area;,Java,Interactive;,9,9

20385,IFPUG,4184,New Development,Catalogue/register of things or events;Document management;,Java,ORACLE;,260,260
20409,IFPUG,2803,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,18,18
20426,COSMIC-FFP,147,New Development,Transaction/Production System;,Visual Basic,SQL Server7 ;,751,751
20487,IFPUG,429,Enhancement,relatively complex application;,4GL,Interactive;,11,11
20558,IFPUG,311,Enhancement,Financial application area;,Java,Interactive;,37,37
20591,IFPUG,908,Enhancement,Financial application area;,Java,Interactive;,89,89
20730,IFPUG,1830,New Development,Trading;,Java,Yes;,94,94
20896,COSMIC-FFP,5018,New Development,Document management;,ASP,SQL SERVER;,762,762
21114,IFPUG,175,Enhancement,Financial application area;,Java,Interactive;,33,33
21180,IFPUG,781,New Development,Trading;,Visual Basic,Oracle 8i;,235,235
21191,IFPUG,312,New Development,tools or system;,Coldfusion,SQL SERVER;,96,96
21254,IFPUG,347,Enhancement,Financial application area;,Java,Interactive;,55,55
21351,IFPUG,1571,Enhancement,relatively complex application;,4GL,Interactive;,82,82
21414,IFPUG,2274,Enhancement,Workflow support & management;,C#,DB2;,414,414
21528,IFPUG,291,Enhancement,relatively complex application;,4GL,Interactive;,22,22
21550,IFPUG,11165,New Development,Document mngnt;Financal trans process/acc;Image video or sound processing, Visual Basic, SQL SERVER,307,307
21596,IFPUG,29,Enhancement,Financial application area;,Java,Interactive;,6,6
21609,IFPUG,355,Enhancement,relatively complex application;,4GL,Interactive;,12,12
21816,IFPUG,940,Enhancement,Financial application area;,Java,Interactive;,162,162
21857,IFPUG,626,New Development,other: Sales contact management;,ASP,Oracle 8i;,242,242
22168,IFPUG,730,Enhancement,Financial transaction process/accounting;Online analysis and reporting;Space management of schools;,Active Server Pages 2.0 etc,Microsoft SQL Server 2000;,848,822
22177,IFPUG,1037,New Development,Financial application area;,Java,Interactive;,146,146
22403,IFPUG,480,Enhancement,relatively complex application;,3GL,Interactive;,32,32
22409,IFPUG,262,Enhancement,Financial application area;,Java,Interactive;,46,46
22476,IFPUG,2592,Enhancement,Financial transaction process/accounting;,C,SYBASE;,115,115

G2-Ran1-67

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
16023,IFPUG,525,Enhancement,Financial application area;,Java,Interactive;,101,101
16076,IFPUG,105,Enhancement,Financial application area;,Java,Interactive;,19,19
16332,IFPUG,2100,New Development,Financial transaction process/accounting;,ASP,SQL SERVER;,257,257
16575,IFPUG,1090,Enhancement,Trading;Electronic Data Interchange;,ASP,MS-SQL;,109,109
16612,IFPUG,465,Enhancement,Financial transaction process/accounting;,PL/SQL,Oracle 8;,162,162
16886,IFPUG,440,New Development,Financial application area;,Java,Interactive;,88,88
16917,IFPUG,21700,Enhancement,Document mngt;Fin trans process/acc;Image video or sound processing ,COBOL , IDMS-DB,500,500
17057,IFPUG,664,New Development,Online analysis and reporting;Workflow support & management;,Java,ORACLE;,51,51
17227,IFPUG,953,Enhancement,Financial application area;,Java,Interactive;,127,127
17461,IFPUG,1400,New Development,Catalogue/register of things or events;Document management;,Java,ORACLE;,120,120
17614,IFPUG,3303,Enhancement,Catalogue/register of things or events;Customer billing/relationship management;,COOL:GEN,DB2;,128,128
17739,IFPUG,442,New Development,Financial application area;,Java,Interactive;,68,68
17855,IFPUG,252,Enhancement,Financial application area;,Java,Interactive;,42,42
17989,IFPUG,1824,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,107,107
18019,IFPUG,155,Enhancement,Financial transaction process/accounting;,C,ORACLE;,19,19
18030,IFPUG,2800,Enhancement,Electronic Data Interchange;,ASP,DB2;SQL;,196,196
18047,IFPUG,331,New Development,Financial application area;,Java,Interactive;,72,72
18311,IFPUG,1354,New Development,Financial application area;,Java,Interactive;,188,188
18398,IFPUG,14992,New Development,Customer billing/relationship management;,HTML,ORACLE;,694,694
18444,IFPUG,19306,Enhancement,Financial transaction process/accounting;,COBOL,IDMS-DB;,393,393
18452,IFPUG,750,New Development,Financial application area;,Java,Interactive;,121,121
18537,IFPUG,7063,New Development,Financial transaction process/accounting;,C,DB2 UDB WorkgroupServer;,522,522
18590,IFPUG,61,Enhancement,Financial application area;,Java,Interactive;,11,11
18686,IFPUG,1588,New Development,Web-based application;,SQL,ORACLE;,340,367
18705,IFPUG,1009,Enhancement,relatively complex application;,4GL,Interactive;,227,227
18737,IFPUG,711,Enhancement,Financial application area;,Java,Interactive;,145,145

18956,IFPUG,9456,New Development,Process Control;,Java,Oracle 8i;,931,1182
19062,IFPUG,2595,Enhancement,relatively complex application;,4GL,Interactive;,150,150
19107,IFPUG,1712,Enhancement,relatively complex application;,4GL,Interactive;,89,89
19278,IFPUG,419,Enhancement,Financial application area;,Java,Interactive;,59,59
19659,IFPUG,354,Enhancement,Financial application area;,Java,Interactive;,61,61
19673,IFPUG,3712,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;Workflow support & management;,Java,ORACLE;,51,51
19757,IFPUG,1536,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,127,127
19798,IFPUG,381,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,46,46
19802,IFPUG,3358,New Development,Web-based Application;,Java,SQL Server;,200,200
19829,IFPUG,2676,Enhancement,Financial transaction process/accounting;,Java,Interactive;,380,353
19990,IFPUG,742,New Development,other: personnel system ;,Java,DB2;,246,246
19997,IFPUG,419,Enhancement,Financial application area;,Java,Interactive;,91,91
20104,IFPUG,2000,New Development,other: Sales contact management;,Java,Oracle 8;,473,473
20117,IFPUG,3982,New Development,Financial transaction process/accounting;,HTML,HiRDB;,190,190
20145,IFPUG,51,Enhancement,Financial application area;,Java,Interactive;,9,9
20385,IFPUG,4184,New Development,Catalogue/register of things or events;Document management;,Java,ORACLE;,260,260
20409,IFPUG,2803,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,18,18
20426,COSMIC-FFP,147,New Development,Transaction/Production System;,Visual Basic,SQL Server7 ;,751,751
20487,IFPUG,429,Enhancement,relatively complex application;,4GL,Interactive;,11,11
20558,IFPUG,311,Enhancement,Financial application area;,Java,Interactive;,37,37
20591,IFPUG,908,Enhancement,Financial application area;,Java,Interactive;,89,89
20730,IFPUG,1830,New Development,Trading;,Java,Yes;,94,94
20896,COSMIC-FFP,5018,New Development,Document management;,ASP,SQL SERVER;,762,762
20900,IFPUG,184,Enhancement,tools or system;,Java,Solid;,98,98
21114,IFPUG,175,Enhancement,Financial application area;,Java,Interactive;,33,33
21180,IFPUG,781,New Development,Trading;,Visual Basic,Oracle 8i;,235,235
21191,IFPUG,312,New Development,tools or system;,Coldfusion,SQL SERVER;,96,96
21254,IFPUG,347,Enhancement,Financial application area;,Java,Interactive;,55,55

21351,IFPUG,1571,Enhancement,relatively complex application;,4GL,Interactive;,82,82
21414,IFPUG,2274,Enhancement,Workflow support & management;,C#,DB2;,414,414
21528,IFPUG,291,Enhancement,relatively complex application;,4GL,Interactive;,22,22
21550,IFPUG,11165,New Development,Document mngnt;Financal trans process/acc;Image video or sound processing, Visual Basic, SQL SERVER,307,307
21596,IFPUG,29,Enhancement,Financial application area;,Java,Interactive;,6,6
21609,IFPUG,355,Enhancement,relatively complex application;,4GL,Interactive;,12,12
21816,IFPUG,940,Enhancement,Financial application area;,Java,Interactive;,162,162
21857,IFPUG,626,New Development,other: Sales contact management;,ASP,Oracle 8i;,242,242
22168,IFPUG,730,Enhancement,Financial transaction process/accounting;Online analysis and reporting;Space management of schools;,Active Server Pages 2.0 etc,Microsoft SQL Server 2000;,848,822
22177,IFPUG,1037,New Development,Financial application area;,Java,Interactive;,146,146
22403,IFPUG,480,Enhancement,relatively complex application;,3GL,Interactive;,32,32
22409,IFPUG,262,Enhancement,Financial application area;,Java,Interactive;,46,46
22476,IFPUG,2592,Enhancement,Financial transaction process/accounting;,C,SYBASE;,115,115

G2-Ran2-17

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
16023,IFPUG,525,Enhancement,Financial application area;,Java,Interactive;,101,101
16076,IFPUG,105,Enhancement,Financial application area;,Java,Interactive;,19,19
16612,IFPUG,465,Enhancement,Financial transaction process/accounting;,PL/SQL,Oracle 8;,162,162
17461,IFPUG,1400,New Development,Catalogue/register of things or events;Document management;,Java,ORACLE;,120,120
17614,IFPUG,3303,Enhancement,Catalogue/register of things or events;Customer billing/relationship management;,COOL:GEN,DB2;,128,128
18030,IFPUG,2800,Enhancement,Electronic Data Interchange;,ASP,DB2;SQL;,196,196
18398,IFPUG,14992,New Development,Customer billing/relationship management;,HTML,ORACLE;,694,694
18705,IFPUG,1009,Enhancement,relatively complex application;,4GL,Interactive;,227,227
19107,IFPUG,1712,Enhancement,relatively complex application;,4GL,Interactive;,89,89
19673,IFPUG,3712,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;Workflow support & management;,Java,ORACLE;,51,51

20145,IFPUG,51,Enhancement,Financial application area;,Java,Interactive;,9,9
20426,COSMIC-FFP,147,New Development,Transaction/Production System;,Visual Basic,SQL Server7 ;,751,751
20896,COSMIC-FFP,5018,New Development,Document management;,ASP,SQL SERVER;,762,762
21180,IFPUG,781,New Development,Trading;,Visual Basic,Oracle 8i;,235,235
21550,IFPUG,11165,New Development,Document mngnt;Financal trans process/acc;Image video or sound processing, Visual Basic, SQL SERVER,307,307
22177,IFPUG,1037,New Development,Financial application area;,Java,Interactive;,146,146
22409,IFPUG,262,Enhancement,Financial application area;,Java,Interactive;,46,46

G2-Ran2-33

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
16023,IFPUG,525,Enhancement,Financial application area;,Java,Interactive;,101,101
16076,IFPUG,105,Enhancement,Financial application area;,Java,Interactive;,19,19
16612,IFPUG,465,Enhancement,Financial transaction process/accounting;,PL/SQL,Oracle 8;,162,162
16917,IFPUG,21700,Enhancement,Document mngt;Fin trans process/acc;Image video or sound processing ,COBOL , IDMS-DB,500,500
17461,IFPUG,1400,New Development,Catalogue/register of things or events;Document management;,Java,ORACLE;,120,120
17614,IFPUG,3303,Enhancement,Catalogue/register of things or events;Customer billing/relationship management;,COOL:GEN,DB2;,128,128
17989,IFPUG,1824,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,107,107
18030,IFPUG,2800,Enhancement,Electronic Data Interchange;,ASP,DB2;SQL;,196,196
18398,IFPUG,14992,New Development,Customer billing/relationship management;,HTML,ORACLE;,694,694
18444,IFPUG,19306,Enhancement,Financial transaction process/accounting;,COBOL,IDMS-DB;,393,393
18590,IFPUG,61,Enhancement,Financial application area;,Java,Interactive;,11,11
18705,IFPUG,1009,Enhancement,relatively complex application;,4GL,Interactive;,227,227
19062,IFPUG,2595,Enhancement,relatively complex application;,4GL,Interactive;,150,150
19107,IFPUG,1712,Enhancement,relatively complex application;,4GL,Interactive;,89,89
19673,IFPUG,3712,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;Workflow support & management;,Java,ORACLE;,51,51
19829,IFPUG,2676,Enhancement,Financial transaction process/accounting;,Java,Interactive;,380,353
19997,IFPUG,419,Enhancement,Financial application area;,Java,Interactive;,91,91

20145,IFPUG,51,Enhancement,Financial application area;,Java,Interactive;,9,9
20385,IFPUG,4184,New Development,Catalogue/register of things or events;Document management;,Java,ORACLE;,260,260
20426,COSMIC-FFP,147,New Development,Transaction/Production System;,Visual Basic,SQL Server7 ;,751,751
20487,IFPUG,429,Enhancement,relatively complex application;,4GL,Interactive;,11,11
20591,IFPUG,908,Enhancement,Financial application area;,Java,Interactive;,89,89
20896,COSMIC-FFP,5018,New Development,Document management;,ASP,SQL SERVER;,762,762
21180,IFPUG,781,New Development,Trading;,Visual Basic,Oracle 8i;,235,235
21254,IFPUG,347,Enhancement,Financial application area;,Java,Interactive;,55,55
21414,IFPUG,2274,Enhancement,Workflow support & management;,C#,DB2;,414,414
21550,IFPUG,11165,New Development,Document mngnt;Financal trans process/acc;Image video or sound processing, Visual Basic, SQL SERVER,307,307
21816,IFPUG,940,Enhancement,Financial application area;,Java,Interactive;,162,162
21857,IFPUG,626,New Development,other: Sales contact management;,ASP,Oracle 8i;,242,242
22177,IFPUG,1037,New Development,Financial application area;,Java,Interactive;,146,146
22403,IFPUG,480,Enhancement,relatively complex application;,3GL,Interactive;,32,32
22409,IFPUG,262,Enhancement,Financial application area;,Java,Interactive;,46,46
22476,IFPUG,2592,Enhancement,Financial transaction process/accounting;,C,SYBASE;,115,115

G2-Ran2-49

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
16023,IFPUG,525,Enhancement,Financial application area;,Java,Interactive;,101,101
16076,IFPUG,105,Enhancement,Financial application area;,Java,Interactive;,19,19
16575,IFPUG,1090,Enhancement,Trading;Electronic Data Interchange;,ASP,MS-SQL;,109,109
16612,IFPUG,465,Enhancement,Financial transaction process/accounting;,PL/SQL,Oracle 8;,162,162
16917,IFPUG,21700,Enhancement,Document mngt;Fin trans process/acc;Image video or sound processing ,COBOL , IDMS-DB,500,500
17057,IFPUG,664,New Development,Online analysis and reporting;Workflow support & management;,Java,ORACLE;,51,51
17461,IFPUG,1400,New Development,Catalogue/register of things or events;Document management;,Java,ORACLE;,120,120
17614,IFPUG,3303,Enhancement,Catalogue/register of things or events;Customer billing/relationship management;,COOL:GEN,DB2;,128,128
17855,IFPUG,252,Enhancement,Financial application area;,Java,Interactive;,42,42

17989,IFPUG,1824,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,107,107
18030,IFPUG,2800,Enhancement,Electronic Data Interchange;,ASP,DB2;SQL;,196,196
18047,IFPUG,331,New Development,Financial application area;,Java,Interactive;,72,72
18398,IFPUG,14992,New Development,Customer billing/relationship management;,HTML,ORACLE;,694,694
18444,IFPUG,19306,Enhancement,Financial transaction process/accounting;,COBOL,IDMS-DB;,393,393
18537,IFPUG,7063,New Development,Financial transaction process/accounting;,C,DB2 UDB WorkgroupServer;,522,522
18590,IFPUG,61,Enhancement,Financial application area;,Java,Interactive;,11,11
18705,IFPUG,1009,Enhancement,relatively complex application;,4GL,Interactive;,227,227
18737,IFPUG,711,Enhancement,Financial application area;,Java,Interactive;,145,145
19062,IFPUG,2595,Enhancement,relatively complex application;,4GL,Interactive;,150,150
19107,IFPUG,1712,Enhancement,relatively complex application;,4GL,Interactive;,89,89
19659,IFPUG,354,Enhancement,Financial application area;,Java,Interactive;,61,61
19673,IFPUG,3712,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;Workflow support & management;,Java,ORACLE;,51,51
19802,IFPUG,3358,New Development,Web-based Application;,Java,SQL Server;,200,200
19829,IFPUG,2676,Enhancement,Financial transaction process/accounting;,Java,Interactive;,380,353
19997,IFPUG,419,Enhancement,Financial application area;,Java,Interactive;,91,91
20104,IFPUG,2000,New Development,other: Sales contact management;,Java,Oracle 8;,473,473
20145,IFPUG,51,Enhancement,Financial application area;,Java,Interactive;,9,9
20385,IFPUG,4184,New Development,Catalogue/register of things or events;Document management;,Java,ORACLE;,260,260
20409,IFPUG,2803,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,18,18
20426,COSMIC-FFP,147,New Development,Transaction/Production System;,Visual Basic,SQL Server7 ;,751,751
20487,IFPUG,429,Enhancement,relatively complex application;,4GL,Interactive;,11,11
20558,IFPUG,311,Enhancement,Financial application area;,Java,Interactive;,37,37
20591,IFPUG,908,Enhancement,Financial application area;,Java,Interactive;,89,89
20896,COSMIC-FFP,5018,New Development,Document management;,ASP,SQL SERVER;,762,762
20900,IFPUG,184,Enhancement,tools or system;,Java,Solid;,98,98
21180,IFPUG,781,New Development,Trading;,Visual Basic,Oracle 8i;,235,235
21254,IFPUG,347,Enhancement,Financial application area;,Java,Interactive;,55,55

21351,IFPUG,1571,Enhancement,relatively complex application;,4GL,Interactive;,82,82
21414,IFPUG,2274,Enhancement,Workflow support & management;,C#,DB2;,414,414
21528,IFPUG,291,Enhancement,relatively complex application;,4GL,Interactive;,22,22
21550,IFPUG,11165,New Development,Document mngnt;Financal trans process/acc;Image video or sound processing, Visual Basic, SQL SERVER,307,307
21609,IFPUG,355,Enhancement,relatively complex application;,4GL,Interactive;,12,12
21816,IFPUG,940,Enhancement,Financial application area;,Java,Interactive;,162,162
21857,IFPUG,626,New Development,other: Sales contact management;,ASP,Oracle 8i;,242,242
22168,IFPUG,730,Enhancement,Financial transaction process/accounting;Online analysis and reporting;Space management of schools;,Active Server Pages 2.0 etc,Microsoft SQL Server 2000;,848,822
22177,IFPUG,1037,New Development,Financial application area;,Java,Interactive;,146,146
22403,IFPUG,480,Enhancement,relatively complex application;,3GL,Interactive;,32,32
22409,IFPUG,262,Enhancement,Financial application area;,Java,Interactive;,46,46
22476,IFPUG,2592,Enhancement,Financial transaction process/accounting;,C,SYBASE;,115,115

G2-Ran2-67

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
16023,IFPUG,525,Enhancement,Financial application area;,Java,Interactive;,101,101
16076,IFPUG,105,Enhancement,Financial application area;,Java,Interactive;,19,19
16332,IFPUG,2100,New Development,Financial transaction process/accounting;,ASP,SQL SERVER;,257,257
16575,IFPUG,1090,Enhancement,Trading;Electronic Data Interchange;,ASP,MS-SQL;,109,109
16612,IFPUG,465,Enhancement,Financial transaction process/accounting;,PL/SQL,Oracle 8;,162,162
16886,IFPUG,440,New Development,Financial application area;,Java,Interactive;,88,88
16917,IFPUG,21700,Enhancement,Document mngt;Fin trans process/acc;Image video or sound processing ,COBOL , IDMS-DB,500,500
17057,IFPUG,664,New Development,Online analysis and reporting;Workflow support & management;,Java,ORACLE;,51,51
17227,IFPUG,953,Enhancement,Financial application area;,Java,Interactive;,127,127
17461,IFPUG,1400,New Development,Catalogue/register of things or events;Document management;,Java,ORACLE;,120,120
17614,IFPUG,3303,Enhancement,Catalogue/register of things or events;Customer billing/relationship management;,COOL:GEN,DB2;,128,128
17739,IFPUG,442,New Development,Financial application area;,Java,Interactive;,68,68

17855,IFPUG,252,Enhancement,Financial application area;,Java,Interactive;,42,42
17989,IFPUG,1824,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,107,107
18019,IFPUG,155,Enhancement,Financial transaction process/accounting;,C,ORACLE;,19,19
18030,IFPUG,2800,Enhancement,Electronic Data Interchange;,ASP,DB2;SQL;,196,196
18047,IFPUG,331,New Development,Financial application area;,Java,Interactive;,72,72
18311,IFPUG,1354,New Development,Financial application area;,Java,Interactive;,188,188
18398,IFPUG,14992,New Development,Customer billing/relationship management;,HTML,ORACLE;,694,694
18444,IFPUG,19306,Enhancement,Financial transaction process/accounting;,COBOL,IDMS-DB;,393,393
18452,IFPUG,750,New Development,Financial application area;,Java,Interactive;,121,121
18537,IFPUG,7063,New Development,Financial transaction process/accounting;,C,DB2 UDB WorkgroupServer;,522,522
18590,IFPUG,61,Enhancement,Financial application area;,Java,Interactive;,11,11
18686,IFPUG,1588,New Development,Web-based application;,SQL,ORACLE;,340,367
18705,IFPUG,1009,Enhancement,relatively complex application;,4GL,Interactive;,227,227
18737,IFPUG,711,Enhancement,Financial application area;,Java,Interactive;,145,145
18956,IFPUG,9456,New Development,Process Control;,Java,Oracle 8i;,931,1182
19062,IFPUG,2595,Enhancement,relatively complex application;,4GL,Interactive;,150,150
19107,IFPUG,1712,Enhancement,relatively complex application;,4GL,Interactive;,89,89
19278,IFPUG,419,Enhancement,Financial application area;,Java,Interactive;,59,59
19659,IFPUG,354,Enhancement,Financial application area;,Java,Interactive;,61,61
19673,IFPUG,3712,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;Workflow support & management;,Java,ORACLE;,51,51
19757,IFPUG,1536,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,127,127
19798,IFPUG,381,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,46,46
19802,IFPUG,3358,New Development,Web-based Application;,Java,SQL Server;,200,200
19829,IFPUG,2676,Enhancement,Financial transaction process/accounting;,Java,Interactive;,380,353
19990,IFPUG,742,New Development,other: personnel system ;,Java,DB2;,246,246
19997,IFPUG,419,Enhancement,Financial application area;,Java,Interactive;,91,91
20104,IFPUG,2000,New Development,other: Sales contact management;,Java,Oracle 8;,473,473

20117,IFPUG,3982,New Development,Financial transaction process/accounting;,HTML,HiRDB;,190,190
20145,IFPUG,51,Enhancement,Financial application area;,Java,Interactive;,9,9
20385,IFPUG,4184,New Development,Catalogue/register of things or events;Document management;,Java,ORACLE;,260,260
20409,IFPUG,2803,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,18,18
20426,COSMIC-FFP,147,New Development,Transaction/Production System;,Visual Basic,SQL Server7 ;,751,751
20487,IFPUG,429,Enhancement,relatively complex application;,4GL,Interactive;,11,11
20558,IFPUG,311,Enhancement,Financial application area;,Java,Interactive;,37,37
20591,IFPUG,908,Enhancement,Financial application area;,Java,Interactive;,89,89
20730,IFPUG,1830,New Development,Trading;,Java,Yes;,94,94
20896,COSMIC-FFP,5018,New Development,Document management;,ASP,SQL SERVER;,762,762
20900,IFPUG,184,Enhancement,tools or system;,Java,Solid;,98,98
21114,IFPUG,175,Enhancement,Financial application area;,Java,Interactive;,33,33
21180,IFPUG,781,New Development,Trading;,Visual Basic,Oracle 8i;,235,235
21191,IFPUG,312,New Development,tools or system;,Coldfusion,SQL SERVER;,96,96
21254,IFPUG,347,Enhancement,Financial application area;,Java,Interactive;,55,55
21351,IFPUG,1571,Enhancement,relatively complex application;,4GL,Interactive;,82,82
21414,IFPUG,2274,Enhancement,Workflow support & management;,C#,DB2;,414,414
21528,IFPUG,291,Enhancement,relatively complex application;,4GL,Interactive;,22,22
21550,IFPUG,11165,New Development,Document mngnt;Financal trans process/acc;Image video or sound processing, Visual Basic, SQL SERVER,307,307
21596,IFPUG,29,Enhancement,Financial application area;,Java,Interactive;,6,6
21609,IFPUG,355,Enhancement,relatively complex application;,4GL,Interactive;,12,12
21816,IFPUG,940,Enhancement,Financial application area;,Java,Interactive;,162,162
21857,IFPUG,626,New Development,other: Sales contact management;,ASP,Oracle 8i;,242,242
22168,IFPUG,730,Enhancement,Financial transaction process/accounting;Online analysis and reporting;Space management of schools;,Active Server Pages 2.0 etc,Microsoft SQL Server 2000;,848,822
22177,IFPUG,1037,New Development,Financial application area;,Java,Interactive;,146,146
22403,IFPUG,480,Enhancement,relatively complex application;,3GL,Interactive;,32,32
22409,IFPUG,262,Enhancement,Financial application area;,Java,Interactive;,46,46
22476,IFPUG,2592,Enhancement,Financial transaction process/accounting;,C,SYBASE;,115,115

G2-Ran3-17

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
16076,IFPUG,105,Enhancement,Financial application area;,Java,Interactive;,19,19
16332,IFPUG,2100,New Development,Financial transaction process/accounting;,ASP,SQL SERVER;,257,257
16886,IFPUG,440,New Development,Financial application area;,Java,Interactive;,88,88
16917,IFPUG,21700,Enhancement,Document mngt;Fin trans process/acc;Image video or sound processing ,COBOL , IDMS-DB,500,500
17614,IFPUG,3303,Enhancement,Catalogue/register of things or events;Customer billing/relationship management;,COOL:GEN,DB2;,128,128
17855,IFPUG,252,Enhancement,Financial application area;,Java,Interactive;,42,42
18590,IFPUG,61,Enhancement,Financial application area;,Java,Interactive;,11,11
18956,IFPUG,9456,New Development,Process Control;,Java,Oracle 8i;,931,1182
19107,IFPUG,1712,Enhancement,relatively complex application;,4GL,Interactive;,89,89
19278,IFPUG,419,Enhancement,Financial application area;,Java,Interactive;,59,59
20104,IFPUG,2000,New Development,other: Sales contact management;,Java,Oracle 8;,473,473
20409,IFPUG,2803,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,18,18
20558,IFPUG,311,Enhancement,Financial application area;,Java,Interactive;,37,37
21180,IFPUG,781,New Development,Trading;,Visual Basic,Oracle 8i;,235,235
21550,IFPUG,11165,New Development,Document mngnt;Financal trans process/acc;Image video or sound processing, Visual Basic, SQL SERVER,307,307
22177,IFPUG,1037,New Development,Financial application area;,Java,Interactive;,146,146
22409,IFPUG,262,Enhancement,Financial application area;,Java,Interactive;,46,46

G2-Ran3-33

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
16076,IFPUG,105,Enhancement,Financial application area;,Java,Interactive;,19,19
16332,IFPUG,2100,New Development,Financial transaction process/accounting;,ASP,SQL SERVER;,257,257
16886,IFPUG,440,New Development,Financial application area;,Java,Interactive;,88,88
16917,IFPUG,21700,Enhancement,Document mngt;Fin trans process/acc;Image video or sound processing ,COBOL , IDMS-DB,500,500
17227,IFPUG,953,Enhancement,Financial application area;,Java,Interactive;,127,127

17614,IFPUG,3303,Enhancement,Catalogue/register of things or events;Customer billing/relationship management;,COOL:GEN,DB2;,128,128
17855,IFPUG,252,Enhancement,Financial application area;,Java,Interactive;,42,42
17989,IFPUG,1824,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,107,107
18047,IFPUG,331,New Development,Financial application area;,Java,Interactive;,72,72
18311,IFPUG,1354,New Development,Financial application area;,Java,Interactive;,188,188
18398,IFPUG,14992,New Development,Customer billing/relationship management;,HTML,ORACLE;,694,694
18537,IFPUG,7063,New Development,Financial transaction process/accounting;,C,DB2 UDB WorkgroupServer;,522,522
18590,IFPUG,61,Enhancement,Financial application area;,Java,Interactive;,11,11
18705,IFPUG,1009,Enhancement,relatively complex application;,4GL,Interactive;,227,227
18956,IFPUG,9456,New Development,Process Control;,Java,Oracle 8i;,931,1182
19107,IFPUG,1712,Enhancement,relatively complex application;,4GL,Interactive;,89,89
19278,IFPUG,419,Enhancement,Financial application area;,Java,Interactive;,59,59
19757,IFPUG,1536,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,127,127
19802,IFPUG,3358,New Development,Web-based Application;,Java,SQL Server;,200,200
19990,IFPUG,742,New Development,other: personnel system ;,Java,DB2;,246,246
20104,IFPUG,2000,New Development,other: Sales contact management;,Java,Oracle 8;,473,473
20117,IFPUG,3982,New Development,Financial transaction process/accounting;,HTML,HiRDB;,190,190
20409,IFPUG,2803,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,18,18
20487,IFPUG,429,Enhancement,relatively complex application;,4GL,Interactive;,11,11
20558,IFPUG,311,Enhancement,Financial application area;,Java,Interactive;,37,37
20896,COSMIC-FFP,5018,New Development,Document management;,ASP,SQL SERVER;,762,762
21180,IFPUG,781,New Development,Trading;,Visual Basic,Oracle 8i;,235,235
21191,IFPUG,312,New Development,tools or system;,Coldfusion,SQL SERVER;,96,96
21351,IFPUG,1571,Enhancement,relatively complex application;,4GL,Interactive;,82,82
21550,IFPUG,11165,New Development,Document mngnt;Financal trans process/acc;Image video or sound processing, Visual Basic, SQL SERVER,307,307
21816,IFPUG,940,Enhancement,Financial application area;,Java,Interactive;,162,162
22177,IFPUG,1037,New Development,Financial application area;,Java,Interactive;,146,146

22409,IFPUG,262,Enhancement,Financial application area;,Java,Interactive;,46,46

G2-Ran3-49

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
16023,IFPUG,525,Enhancement,Financial application area;,Java,Interactive;,101,101
16076,IFPUG,105,Enhancement,Financial application area;,Java,Interactive;,19,19
16332,IFPUG,2100,New Development,Financial transaction process/accounting;,ASP,SQL SERVER;,257,257
16612,IFPUG,465,Enhancement,Financial transaction process/accounting;,PL/SQL,Oracle 8;,162,162
16886,IFPUG,440,New Development,Financial application area;,Java,Interactive;,88,88
16917,IFPUG,21700,Enhancement,Document mngt;Fin trans process/acc;Image video or sound processing ,COBOL , IDMS-DB,500,500
17227,IFPUG,953,Enhancement,Financial application area;,Java,Interactive;,127,127
17461,IFPUG,1400,New Development,Catalogue/register of things or events;Document management;,Java,ORACLE;,120,120
17614,IFPUG,3303,Enhancement,Catalogue/register of things or events;Customer billing/relationship management;,COOL:GEN,DB2;,128,128
17855,IFPUG,252,Enhancement,Financial application area;,Java,Interactive;,42,42
17989,IFPUG,1824,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,107,107
18019,IFPUG,155,Enhancement,Financial transaction process/accounting;,C,ORACLE;,19,19
18047,IFPUG,331,New Development,Financial application area;,Java,Interactive;,72,72
18311,IFPUG,1354,New Development,Financial application area;,Java,Interactive;,188,188
18398,IFPUG,14992,New Development,Customer billing/relationship management;,HTML,ORACLE;,694,694
18452,IFPUG,750,New Development,Financial application area;,Java,Interactive;,121,121
18537,IFPUG,7063,New Development,Financial transaction process/accounting;,C,DB2 UDB WorkgroupServer;,522,522
18590,IFPUG,61,Enhancement,Financial application area;,Java,Interactive;,11,11
18705,IFPUG,1009,Enhancement,relatively complex application;,4GL,Interactive;,227,227
18737,IFPUG,711,Enhancement,Financial application area;,Java,Interactive;,145,145
18956,IFPUG,9456,New Development,Process Control;,Java,Oracle 8i;,931,1182
19107,IFPUG,1712,Enhancement,relatively complex application;,4GL,Interactive;,89,89
19278,IFPUG,419,Enhancement,Financial application area;,Java,Interactive;,59,59

19673,IFPUG,3712,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;Workflow support & management;,Java,ORACLE;,51,51
19757,IFPUG,1536,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,127,127
19802,IFPUG,3358,New Development,Web-based Application;,Java,SQL Server;,200,200
19829,IFPUG,2676,Enhancement,Financial transaction process/accounting;,Java,Interactive;,380,353
19990,IFPUG,742,New Development,other: personnel system ;,Java,DB2;,246,246
20104,IFPUG,2000,New Development,other: Sales contact management;,Java,Oracle 8;,473,473
20117,IFPUG,3982,New Development,Financial transaction process/accounting;,HTML,HiRDB;,190,190
20385,IFPUG,4184,New Development,Catalogue/register of things or events;Document management;,Java,ORACLE;,260,260
20409,IFPUG,2803,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,18,18
20487,IFPUG,429,Enhancement,relatively complex application;,4GL,Interactive;,11,11
20558,IFPUG,311,Enhancement,Financial application area;,Java,Interactive;,37,37
20730,IFPUG,1830,New Development,Trading;,Java,Yes;,94,94
20896,COSMIC-FFP,5018,New Development,Document management;,ASP,SQL SERVER;,762,762
20900,IFPUG,184,Enhancement,tools or system;,Java,Solid;,98,98
21180,IFPUG,781,New Development,Trading;,Visual Basic,Oracle 8i;,235,235
21191,IFPUG,312,New Development,tools or system;,Coldfusion,SQL SERVER;,96,96
21254,IFPUG,347,Enhancement,Financial application area;,Java,Interactive;,55,55
21351,IFPUG,1571,Enhancement,relatively complex application;,4GL,Interactive;,82,82
21528,IFPUG,291,Enhancement,relatively complex application;,4GL,Interactive;,22,22
21550,IFPUG,11165,New Development,Document mngnt;Financal trans process/acc;Image video or sound processing, Visual Basic, SQL SERVER,307,307
21609,IFPUG,355,Enhancement,relatively complex application;,4GL,Interactive;,12,12
21816,IFPUG,940,Enhancement,Financial application area;,Java,Interactive;,162,162
21857,IFPUG,626,New Development,other: Sales contact management;,ASP,Oracle 8i;,242,242
22177,IFPUG,1037,New Development,Financial application area;,Java,Interactive;,146,146
22403,IFPUG,480,Enhancement,relatively complex application;,3GL,Interactive;,32,32
22409,IFPUG,262,Enhancement,Financial application area;,Java,Interactive;,46,46

G2-Ran3-67

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
16023,IFPUG,525,Enhancement,Financial application area;,Java,Interactive;,101,101
16076,IFPUG,105,Enhancement,Financial application area;,Java,Interactive;,19,19
16332,IFPUG,2100,New Development,Financial transaction process/accounting;,ASP,SQL SERVER;,257,257
16575,IFPUG,1090,Enhancement,Trading;Electronic Data Interchange;,ASP,MS-SQL;,109,109
16612,IFPUG,465,Enhancement,Financial transaction process/accounting;,PL/SQL,Oracle 8;,162,162
16886,IFPUG,440,New Development,Financial application area;,Java,Interactive;,88,88
16917,IFPUG,21700,Enhancement,Document mngt;Fin trans process/acc;Image video or sound processing ,COBOL , IDMS-DB,500,500
17057,IFPUG,664,New Development,Online analysis and reporting;Workflow support & management;,Java,ORACLE;,51,51
17227,IFPUG,953,Enhancement,Financial application area;,Java,Interactive;,127,127
17461,IFPUG,1400,New Development,Catalogue/register of things or events;Document management;,Java,ORACLE;,120,120
17614,IFPUG,3303,Enhancement,Catalogue/register of things or events;Customer billing/relationship management;,COOL:GEN,DB2;,128,128
17739,IFPUG,442,New Development,Financial application area;,Java,Interactive;,68,68
17855,IFPUG,252,Enhancement,Financial application area;,Java,Interactive;,42,42
17989,IFPUG,1824,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,107,107
18019,IFPUG,155,Enhancement,Financial transaction process/accounting;,C,ORACLE;,19,19
18030,IFPUG,2800,Enhancement,Electronic Data Interchange;,ASP,DB2;SQL;,196,196
18047,IFPUG,331,New Development,Financial application area;,Java,Interactive;,72,72
18311,IFPUG,1354,New Development,Financial application area;,Java,Interactive;,188,188
18398,IFPUG,14992,New Development,Customer billing/relationship management;,HTML,ORACLE;,694,694
18444,IFPUG,19306,Enhancement,Financial transaction process/accounting;,COBOL,IDMS-DB;,393,393
18452,IFPUG,750,New Development,Financial application area;,Java,Interactive;,121,121
18537,IFPUG,7063,New Development,Financial transaction process/accounting;,C,DB2 UDB WorkgroupServer;,522,522
18590,IFPUG,61,Enhancement,Financial application area;,Java,Interactive;,11,11
18686,IFPUG,1588,New Development,Web-based application;,SQL,ORACLE;,340,367
18705,IFPUG,1009,Enhancement,relatively complex application;,4GL,Interactive;,227,227
18737,IFPUG,711,Enhancement,Financial application area;,Java,Interactive;,145,145

18956,IFPUG,9456,New Development,Process Control;,Java,Oracle 8i;,931,1182
19062,IFPUG,2595,Enhancement,relatively complex application;,4GL,Interactive;,150,150
19107,IFPUG,1712,Enhancement,relatively complex application;,4GL,Interactive;,89,89
19278,IFPUG,419,Enhancement,Financial application area;,Java,Interactive;,59,59
19659,IFPUG,354,Enhancement,Financial application area;,Java,Interactive;,61,61
19673,IFPUG,3712,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;Workflow support & management;,Java,ORACLE;,51,51
19757,IFPUG,1536,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,127,127
19798,IFPUG,381,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,46,46
19802,IFPUG,3358,New Development,Web-based Application;,Java,SQL Server;,200,200
19829,IFPUG,2676,Enhancement,Financial transaction process/accounting;,Java,Interactive;,380,353
19990,IFPUG,742,New Development,other: personnel system ;,Java,DB2;,246,246
19997,IFPUG,419,Enhancement,Financial application area;,Java,Interactive;,91,91
20104,IFPUG,2000,New Development,other: Sales contact management;,Java,Oracle 8;,473,473
20117,IFPUG,3982,New Development,Financial transaction process/accounting;,HTML,HiRDB;,190,190
20145,IFPUG,51,Enhancement,Financial application area;,Java,Interactive;,9,9
20385,IFPUG,4184,New Development,Catalogue/register of things or events;Document management;,Java,ORACLE;,260,260
20409,IFPUG,2803,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,18,18
20426,COSMIC-FFP,147,New Development,Transaction/Production System;,Visual Basic,SQL Server7 ;,751,751
20487,IFPUG,429,Enhancement,relatively complex application;,4GL,Interactive;,11,11
20558,IFPUG,311,Enhancement,Financial application area;,Java,Interactive;,37,37
20591,IFPUG,908,Enhancement,Financial application area;,Java,Interactive;,89,89
20730,IFPUG,1830,New Development,Trading;,Java,Yes;,94,94
20896,COSMIC-FFP,5018,New Development,Document management;,ASP,SQL SERVER;,762,762
20900,IFPUG,184,Enhancement,tools or system;,Java,Solid;,98,98
21114,IFPUG,175,Enhancement,Financial application area;,Java,Interactive;,33,33
21180,IFPUG,781,New Development,Trading;,Visual Basic,Oracle 8i;,235,235
21191,IFPUG,312,New Development,tools or system;,Coldfusion,SQL SERVER;,96,96
21254,IFPUG,347,Enhancement,Financial application area;,Java,Interactive;,55,55

21351,IFPUG,1571,Enhancement,relatively complex application;,4GL,Interactive;,82,82
21414,IFPUG,2274,Enhancement,Workflow support & management;,C#,DB2;,414,414
21528,IFPUG,291,Enhancement,relatively complex application;,4GL,Interactive;,22,22
21550,IFPUG,11165,New Development,Document mngnt;Financal trans process/acc;Image video or sound processing, Visual Basic, SQL SERVER,307,307
21596,IFPUG,29,Enhancement,Financial application area;,Java,Interactive;,6,6
21609,IFPUG,355,Enhancement,relatively complex application;,4GL,Interactive;,12,12
21816,IFPUG,940,Enhancement,Financial application area;,Java,Interactive;,162,162
21857,IFPUG,626,New Development,other: Sales contact management;,ASP,Oracle 8i;,242,242
22168,IFPUG,730,Enhancement,Financial transaction process/accounting;Online analysis and reporting;Space management of schools;,Active Server Pages 2.0 etc,Microsoft SQL Server 2000;,848,822
22177,IFPUG,1037,New Development,Financial application area;,Java,Interactive;,146,146
22403,IFPUG,480,Enhancement,relatively complex application;,3GL,Interactive;,32,32
22409,IFPUG,262,Enhancement,Financial application area;,Java,Interactive;,46,46
22476,IFPUG,2592,Enhancement,Financial transaction process/accounting;,C,SYBASE;,115,115

G3-Ran1-17

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
22589,IFPUG,187,New Development,Process Control;,Java,Oracle 9i;,41,41
22869,IFPUG,3566,Enhancement,Device or interface driver;Financial transaction process/accounting;Process Control;,NET,DB2;,211,211
23146,IFPUG,3112,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,49,49
23266,IFPUG,4656,Enhancement,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,621,621
23925,IFPUG,5086,Enhancement,relatively complex application;,4GL,Interactive;,421,421
24043,IFPUG,1442,New Development,Financial application area;,Java,Interactive;,267,267
25180,IFPUG,7760,New Development,Web-based application;,SQL,ORACLE;,927,908
25552,IFPUG,232,Enhancement,relatively complex application;,4GL,Interactive;,7,7
25690,IFPUG,1248,New Development,Catalogue/register of things or events;Online analysis and reporting;,Java,ORACLE;,230,230
25841,IFPUG,5885,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,633,633

26294,IFPUG,17120,New Development,Business enabling service;,C,DB2;,433,524
26382,COSMIC-FFP,7111,New Development,Geographic or spatial information system;,C++,Yes;,106,106
26733,IFPUG,4880,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,649,649
26755,IFPUG,3998,Enhancement,Network Management;,ASP,ORACLE;,1157,1157
27560,IFPUG,3887,Re-development,Operating system or software utility;Other;,C#,Sql Server 2000;,199,221
28046,IFPUG,120,Enhancement,Web Content & Middleware;,C++,JAVA;,30,34
28161,IFPUG,1850,Enhancement,Telecom Data Circuits and Revenue;,C++,ORACLE;,200,232

G3-Ran1-33

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
22508,IFPUG,11052,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,149,149
22589,IFPUG,187,New Development,Process Control;,Java,Oracle 9i;,41,41
22723,IFPUG,5084,Re-Development,Web-based Application;,ASP,SQL Server;,700,700
22869,IFPUG,3566,Enhancement,Device or interface driver;Financial transaction process/accounting;Process Control;,NET,DB2;,211,211
22899,IFPUG,109,Enhancement,Process Control;,Java,Solid;,34,34
23146,IFPUG,3112,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,49,49
23241,IFPUG,144,Enhancement,tools or system;,Java,Oracle 9i;,95,95
23266,IFPUG,4656,Enhancement,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,621,621
23565,IFPUG,1652,Enhancement,Financial transaction process/accounting;,Java,Interactive;,87,87
23925,IFPUG,5086,Enhancement,relatively complex application;,4GL,Interactive;,421,421
24043,IFPUG,1442,New Development,Financial application area;,Java,Interactive;,267,267
24218,IFPUG,2321,New Development,Financial application area;,Java,Interactive;,283,283
24569,IFPUG,83,New Development,Web-based Application;,Visual Basic,SQL Server;,19,19
25180,IFPUG,7760,New Development,Web-based application;,SQL,ORACLE;,927,908
25552,IFPUG,232,Enhancement,relatively complex application;,4GL,Interactive;,7,7
25641,IFPUG,1718,New Development,other: production management system;,SQL,SQL Server7;,776,776
25690,IFPUG,1248,New Development,Catalogue/register of things or events;Online analysis and reporting;,Java,ORACLE;,230,230

25725,IFPUG,17400,Enhancement,Financial transaction process/accounting;,C++,HIRDB;,2099,2099
25841,IFPUG,5885,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,633,633
26019,IFPUG,2640,New Development,Catalogue/register of things or events;Document management;Workflow support & management;,Java,ORACLE;,234,234
26201,IFPUG,3312,New Development,Financial application area;,Java,Interactive;,364,364
26294,IFPUG,17120,New Development,Business enabling service;,C,DB2;,433,524
26331,IFPUG,12000,New Development,Web-based Application;,Java,Sybase;,1800,1800
26382,COSMIC-FFP,7111,New Development,Geographic or spatial information system;,C++,Yes;,106,106
26417,IFPUG,2450,New Development,Financial transaction process/accounting;,ASP,SQL Server2000;,360,360
26733,IFPUG,4880,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,649,649
26755,IFPUG,3998,Enhancement,Network Management;,ASP,ORACLE;,1157,1157
27123,COSMIC-FFP,888,New Development,Customer billing/relationship management;,PHP,MySql;,234,234
27560,IFPUG,3887,Re-development,Operating system or software utility;Other;,C#,Sql Server 2000;,199,221
27824,IFPUG,1163,New Development,other: DB Serch system;,PHP,SQL SERVER;,160,160
28046,IFPUG,120,Enhancement,Web Content & Middleware;,C++,JAVA;,30,34
28127,IFPUG,912,Enhancement,Financial application area;,Java,Interactive;,172,172
28161,IFPUG,1850,Enhancement,Telecom Data Circuits and Revenue;,C++,ORACLE;,200,232

G3-Ran1-49

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
22508,IFPUG,11052,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,149,149
22561,IFPUG,18314,New Development,Financial transaction process/accounting;,SQL,ORACLE;,2245,2245
22589,IFPUG,187,New Development,Process Control;,Java,Oracle 9i;,41,41
22705,IFPUG,1562,New Development,Financial transaction process/accounting;,Java,Oracle 8i;,218,218
22723,IFPUG,5084,Re-Development,Web-based Application;,ASP,SQL Server;,700,700
22869,IFPUG,3566,Enhancement,Device or interface driver;Financial transaction process/accounting;Process Control;,NET,DB2;,211,211

22899,IFPUG,109,Enhancement,Process Control;,Java,Solid;,34,34
23146,IFPUG,3112,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,49,49
23241,IFPUG,144,Enhancement,tools or system;,Java,Oracle 9i;,95,95
23266,IFPUG,4656,Enhancement,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,621,621
23565,IFPUG,1652,Enhancement,Financial transaction process/accounting;,Java,Interactive;,87,87
23925,IFPUG,5086,Enhancement,relatively complex application;,4GL,Interactive;,421,421
24043,IFPUG,1442,New Development,Financial application area;,Java,Interactive;,267,267
24218,IFPUG,2321,New Development,Financial application area;,Java,Interactive;,283,283
24291,IFPUG,557,Enhancement,Financial application area;,Java,Interactive;,96,96
24569,IFPUG,83,New Development,Web-based Application;,Visual Basic,SQL Server;,19,19
24677,IFPUG,2624,Enhancement,relatively complex application;,4GL,Interactive;,201,201
25180,IFPUG,7760,New Development,Web-based application;,SQL,ORACLE;,927,908
25310,IFPUG,1766,New Development,Financial application area;,Java,Interactive;,256,256
25552,IFPUG,232,Enhancement,relatively complex application;,4GL,Interactive;,7,7
25620,IFPUG,20096,Re-development,Customer billing/relationship management;Other;,C#,Sql Server 2000;,1127,1341
25641,IFPUG,1718,New Development,other: production management system;,SQL,SQL Server7;,776,776
25666,IFPUG,449,Enhancement,Financial application area;,Java,Interactive;,68,68
25690,IFPUG,1248,New Development,Catalogue/register of things or events;Online analysis and reporting;,Java,ORACLE;,230,230
25704,IFPUG,2464,New Development,Financial application area;,Java,Interactive;,202,202
25725,IFPUG,17400,Enhancement,Financial transaction process/accounting;,C++,HIRDB;,2099,2099
25741,IFPUG,2088,Enhancement,Equipment Management;,SQL,ORACLE;,109,134
25841,IFPUG,5885,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,633,633
25988,IFPUG,11752,New Development,other: mission-critical system;,Java,ORACLE;,2091,2091
26019,IFPUG,2640,New Development,Catalogue/register of things or events;Document management;Workflow support & management;,Java,ORACLE;,234,234
26093,COSMIC-FFP,3187,New Development,Online System for University fraternities;,ASP,MS SQLServer2000;,655,655
26201,IFPUG,3312,New Development,Financial application area;,Java,Interactive;,364,364
26251,IFPUG,112,Enhancement,Financial application area;,Java,Interactive;,31,31
26294,IFPUG,17120,New Development,Business enabling service;,C,DB2;,433,524

26317,IFPUG,1216,Enhancement,Online analysis and reporting;,Java,ORACLE;,156,156
26331,IFPUG,12000,New Development,Web-based Application;,Java,Sybase;,1800,1800
26382,COSMIC-FFP,7111,New Development,Geographic or spatial information system;,C++,Yes;,106,106
26417,IFPUG,2450,New Development,Financial transaction process/accounting;,ASP,SQL Server2000;,360,360
26733,IFPUG,4880,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,649,649
26755,IFPUG,3998,Enhancement,Network Management;,ASP,ORACLE;,1157,1157
27123,COSMIC-FFP,888,New Development,Customer billing/relationship management;,PHP,MySql;,234,234
27553,COSMIC-FFP,19306,Enhancement,Document management;Financial transaction process/accounting;Image, video or sound processing;",COBOL,398,398
27560,IFPUG,3887,Re-development,Operating system or software utility;Other;,C#,Sql Server 2000;,199,221
27732,IFPUG,2218,New Development,other: Sales contact management;,SQL,SQL SERVER;,609,609
27824,IFPUG,1163,New Development,other: DB Serch system;,PHP,SQL SERVER;,160,160
27941,IFPUG,3068,New Development,Management or performance reporting;Online analysis and reporting;,Datastage,DB2;,349,349
28046,IFPUG,120,Enhancement,Web Content & Middleware;,C++,JAVA;,30,34
28127,IFPUG,912,Enhancement,Financial application area;,Java,Interactive;,172,172
28161,IFPUG,1850,Enhancement,Telecom Data Circuits and Revenue;,C++,ORACLE;,200,232

G3-Ran1-67

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
22508,IFPUG,11052,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,149,149
22524,IFPUG,1139,Enhancement,Financial application area;,Java,Interactive;,156,156
22561,IFPUG,18314,New Development,Financial transaction process/accounting;,SQL,ORACLE;,2245,2245
22570,IFPUG,110,Enhancement,Financial application area;,Java,Interactive;,12,12
22589,IFPUG,187,New Development,Process Control;,Java,Oracle 9i;,41,41
22692,COSMIC-FFP,21600,New Development,Logistic or supply planning & control;,Java,internal tools DB;,115,115
22705,IFPUG,1562,New Development,Financial transaction process/accounting;,Java,Oracle 8i;,218,218
22712,IFPUG,172,New Development,Web-based Application;,Visual Basic,SQL Server;,46,46
22723,IFPUG,5084,Re-Development,Web-based Application;,ASP,SQL Server;,700,700

22869,IFPUG,3566,Enhancement,Device or interface driver;Financial transaction process/accounting;Process Control;,NET,DB2;,211,211
22886,IFPUG,902,Enhancement,relatively complex application;,4GL,Interactive;,110,110
22899,IFPUG,109,Enhancement,Process Control;,Java,Solid;,34,34
23123,IFPUG,1047,New Development,Financial application area;,Java,Interactive;,154,154
23146,IFPUG,3112,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,49,49
23168,IFPUG,838,Enhancement,Financial application area;,Java,Interactive;,133,133
23241,IFPUG,144,Enhancement,tools or system;,Java,Oracle 9i;,95,95
23266,IFPUG,4656,Enhancement,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,621,621
23518,IFPUG,626,Enhancement,relatively complex application;,4GL,Interactive;,25,25
23565,IFPUG,1652,Enhancement,Financial transaction process/accounting;,Java,Interactive;,87,87
23791,IFPUG,48,New Development,other: Sales contact management;,ASP,ORACLE;,62,62
23925,IFPUG,5086,Enhancement,relatively complex application;,4GL,Interactive;,421,421
24017,IFPUG,13728,New Development,Financial transaction process/accounting;,SQL,ORACLE;,1956,1956
24043,IFPUG,1442,New Development,Financial application area;,Java,Interactive;,267,267
24200,IFPUG,5841,Re-development,Geographic or spatial information system;Online analysis and reporting;,C#,SQL;,354,354
24218,IFPUG,2321,New Development,Financial application area;,Java,Interactive;,283,283
24275,IFPUG,5226,New Development,other: production management system;,Java,ORACLE;,578,578
24291,IFPUG,557,Enhancement,Financial application area;,Java,Interactive;,96,96
24483,COSMIC-FFP,8772,Re-development,Workflow support & management;,J2EE,Oracle8i;,465,465
24569,IFPUG,83,New Development,Web-based Application;,Visual Basic,SQL Server;,19,19
24654,IFPUG,3600,New Development,Web-based application;,SQL,ORACLE;,408,408
24677,IFPUG,2624,Enhancement,relatively complex application;,4GL,Interactive;,201,201
25178,IFPUG,1393,New Development,Financial application area;,Java,Interactive;,172,172
25180,IFPUG,7760,New Development,Web-based application;,SQL,ORACLE;,927,908
25287,IFPUG,1186,New Development,Management Information System;,Visual Basic,ORACLE;,128,132
25310,IFPUG,1766,New Development,Financial application area;,Java,Interactive;,256,256
25550,IFPUG,5714,Enhancement,Web-based Application;,Java,Others;,580,580
25552,IFPUG,232,Enhancement,relatively complex application;,4GL,Interactive;,7,7
25620,IFPUG,20096,Re-development,Customer billing/relationship management;Other;,C#,Sql Server 2000;,1127,1341

25641,IFPUG,1718,New Development,other: production management system;,SQL,SQL Server7;,776,776
25666,IFPUG,449,Enhancement,Financial application area;,Java,Interactive;,68,68
25690,IFPUG,1248,New Development,Catalogue/register of things or events;Online analysis and reporting;,Java,ORACLE;,230,230
25704,IFPUG,2464,New Development,Financial application area;,Java,Interactive;,202,202
25725,IFPUG,17400,Enhancement,Financial transaction process/accounting;,C++,HIRDB;,2099,2099
25741,IFPUG,2088,Enhancement,Equipment Management;,SQL,ORACLE;,109,134
25841,IFPUG,5885,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,633,633
25988,IFPUG,11752,New Development,other: mission-critical system;,Java,ORACLE;,2091,2091
26019,IFPUG,2640,New Development,Catalogue/register of things or events;Document management;Workflow support & management;,Java,ORACLE;,234,234
26093,COSMIC-FFP,3187,New Development,Online System for University fraternities;,ASP,MS SQLServer2000;,655,655
26201,IFPUG,3312,New Development,Financial application area;,Java,Interactive;,364,364
26251,IFPUG,112,Enhancement,Financial application area;,Java,Interactive;,31,31
26294,IFPUG,17120,New Development,Business enabling service;,C,DB2;,433,524
26317,IFPUG,1216,Enhancement,Online analysis and reporting;,Java,ORACLE;,156,156
26331,IFPUG,12000,New Development,Web-based Application;,Java,Sybase;,1800,1800
26382,COSMIC-FFP,7111,New Development,Geographic or spatial information system;,C++,Yes;,106,106
26417,IFPUG,2450,New Development,Financial transaction process/accounting;,ASP,SQL Server2000;,360,360
26505,IFPUG,2153,New Development,Financial application area;,Java,Interactive;,234,234
26733,IFPUG,4880,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,649,649
26755,IFPUG,3998,Enhancement,Network Management;,ASP,ORACLE;,1157,1157
27123,COSMIC-FFP,888,New Development,Customer billing/relationship management;,PHP,MySql;,234,234
27553,COSMIC-FFP,19306,Enhancement,Document management;Financial transaction process/accounting;Image, video or sound processing;",COBOL,398,398
27560,IFPUG,3887,Re-development,Operating system or software utility;Other;,C#,Sql Server 2000;,199,221
27732,IFPUG,2218,New Development,other: Sales contact management;,SQL,SQL SERVER;,609,609
27824,IFPUG,1163,New Development,other: DB Serch system;,PHP,SQL SERVER;,160,160
27941,IFPUG,3068,New Development,Management or performance reporting;Online analysis and reporting;,Datastage,DB2;,349,349
28046,IFPUG,120,Enhancement,Web Content & Middleware;,C++,JAVA;,30,34

28127,IFPUG,912,Enhancement,Financial application area;,Java,Interactive;,172,172
28161,IFPUG,1850,Enhancement,Telecom Data Circuits and Revenue;,C++,ORACLE;,200,232

G3-Ran2-17

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
22508,IFPUG,11052,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,149,149
22524,IFPUG,1139,Enhancement,Financial application area;,Java,Interactive;,156,156
22705,IFPUG,1562,New Development,Financial transaction process/accounting;,Java,Oracle 8i;,218,218
22886,IFPUG,902,Enhancement,relatively complex application;,4GL,Interactive;,110,110
23565,IFPUG,1652,Enhancement,Financial transaction process/accounting;,Java,Interactive;,87,87
24043,IFPUG,1442,New Development,Financial application area;,Java,Interactive;,267,267
25178,IFPUG,1393,New Development,Financial application area;,Java,Interactive;,172,172
25552,IFPUG,232,Enhancement,relatively complex application;,4GL,Interactive;,7,7
25620,IFPUG,20096,Re-development,Customer billing/relationship management;Other;,C#,Sql Server 2000;,1127,1341
25725,IFPUG,17400,Enhancement,Financial transaction process/accounting;,C++,HIRDB;,2099,2099
26093,COSMIC-FFP,3187,New Development,Online System for University fraternities;,ASP,MS SQLServer2000;,655,655
26317,IFPUG,1216,Enhancement,Online analysis and reporting;,Java,ORACLE;,156,156
26417,IFPUG,2450,New Development,Financial transaction process/accounting;,ASP,SQL Server2000;,360,360
27123,COSMIC-FFP,888,New Development,Customer billing/relationship management;,PHP,MySql;,234,234
27732,IFPUG,2218,New Development,other: Sales contact management;,SQL,SQL SERVER;,609,609
28127,IFPUG,912,Enhancement,Financial application area;,Java,Interactive;,172,172
28161,IFPUG,1850,Enhancement,Telecom Data Circuits and Revenue;,C++,ORACLE;,200,232

G3-Ran2-33

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
22508,IFPUG,11052,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,149,149
22524,IFPUG,1139,Enhancement,Financial application area;,Java,Interactive;,156,156
22589,IFPUG,187,New Development,Process Control;,Java,Oracle 9i;,41,41

22705,IFPUG,1562,New Development,Financial transaction process/accounting;,Java,Oracle 8i;,218,218
22869,IFPUG,3566,Enhancement,Device or interface driver;Financial transaction process/accounting;Process Control;,NET,DB2;,211,211
22886,IFPUG,902,Enhancement,relatively complex application;,4GL,Interactive;,110,110
23146,IFPUG,3112,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,49,49
23241,IFPUG,144,Enhancement,tools or system;,Java,Oracle 9i;,95,95
23565,IFPUG,1652,Enhancement,Financial transaction process/accounting;,Java,Interactive;,87,87
23791,IFPUG,48,New Development,other: Sales contact management;,ASP,ORACLE;,62,62
24043,IFPUG,1442,New Development,Financial application area;,Java,Interactive;,267,267
24218,IFPUG,2321,New Development,Financial application area;,Java,Interactive;,283,283
24483,COSMIC-FFP,8772,Re-development,Workflow support & management;,J2EE,Oracle8i;,465,465
24569,IFPUG,83,New Development,Web-based Application;,Visual Basic,SQL Server;,19,19
25178,IFPUG,1393,New Development,Financial application area;,Java,Interactive;,172,172
25287,IFPUG,1186,New Development,Management Information System;,Visual Basic,ORACLE;,128,132
25552,IFPUG,232,Enhancement,relatively complex application;,4GL,Interactive;,7,7
25620,IFPUG,20096,Re-development,Customer billing/relationship management;Other;,C#,Sql Server 2000;,1127,1341
25725,IFPUG,17400,Enhancement,Financial transaction process/accounting;,C++,HIRDB;,2099,2099
25841,IFPUG,5885,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,633,633
26093,COSMIC-FFP,3187,New Development,Online System for University fraternities;,ASP,MS SQLServer2000;,655,655
26201,IFPUG,3312,New Development,Financial application area;,Java,Interactive;,364,364
26294,IFPUG,17120,New Development,Business enabling service;,C,DB2;,433,524
26317,IFPUG,1216,Enhancement,Online analysis and reporting;,Java,ORACLE;,156,156
26417,IFPUG,2450,New Development,Financial transaction process/accounting;,ASP,SQL Server2000;,360,360
26505,IFPUG,2153,New Development,Financial application area;,Java,Interactive;,234,234
26755,IFPUG,3998,Enhancement,Network Management;,ASP,ORACLE;,1157,1157
27123,COSMIC-FFP,888,New Development,Customer billing/relationship management;,PHP,MySql;,234,234
27560,IFPUG,3887,Re-development,Operating system or software utility;Other;,C#,Sql Server 2000;,199,221
27732,IFPUG,2218,New Development,other: Sales contact management;,SQL,SQL SERVER;,609,609
27941,IFPUG,3068,New Development,Management or performance reporting;Online analysis and reporting;,Datastage,DB2;,349,349
28127,IFPUG,912,Enhancement,Financial application area;,Java,Interactive;,172,172
28161,IFPUG,1850,Enhancement,Telecom Data Circuits and Revenue;,C++,ORACLE;,200,232

G3-Ran2-49

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
22508,IFPUG,11052,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,149,149
22524,IFPUG,1139,Enhancement,Financial application area;,Java,Interactive;,156,156
22570,IFPUG,110,Enhancement,Financial application area;,Java,Interactive;,12,12
22589,IFPUG,187,New Development,Process Control;,Java,Oracle 9i;,41,41
22705,IFPUG,1562,New Development,Financial transaction process/accounting;,Java,Oracle 8i;,218,218
22712,IFPUG,172,New Development,Web-based Application;,Visual Basic,SQL Server;,46,46
22869,IFPUG,3566,Enhancement,Device or interface driver;Financial transaction process/accounting;Process Control;,NET,DB2;,211,211
22886,IFPUG,902,Enhancement,relatively complex application;,4GL,Interactive;,110,110
23123,IFPUG,1047,New Development,Financial application area;,Java,Interactive;,154,154
23146,IFPUG,3112,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,49,49
23241,IFPUG,144,Enhancement,tools or system;,Java,Oracle 9i;,95,95
23266,IFPUG,4656,Enhancement,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,621,621
23565,IFPUG,1652,Enhancement,Financial transaction process/accounting;,Java,Interactive;,87,87
23791,IFPUG,48,New Development,other: Sales contact management;,ASP,ORACLE;,62,62
24017,IFPUG,13728,New Development,Financial transaction process/accounting;,SQL,ORACLE;,1956,1956
24043,IFPUG,1442,New Development,Financial application area;,Java,Interactive;,267,267
24218,IFPUG,2321,New Development,Financial application area;,Java,Interactive;,283,283
24275,IFPUG,5226,New Development,other: production management system;,Java,ORACLE;,578,578
24483,COSMIC-FFP,8772,Re-development,Workflow support & management;,J2EE,Oracle8i;,465,465
24569,IFPUG,83,New Development,Web-based Application;,Visual Basic,SQL Server;,19,19
24677,IFPUG,2624,Enhancement,relatively complex application;,4GL,Interactive;,201,201
25178,IFPUG,1393,New Development,Financial application area;,Java,Interactive;,172,172
25287,IFPUG,1186,New Development,Management Information System;,Visual Basic,ORACLE;,128,132
25310,IFPUG,1766,New Development,Financial application area;,Java,Interactive;,256,256
25552,IFPUG,232,Enhancement,relatively complex application;,4GL,Interactive;,7,7

25620,IFPUG,20096,Re-development,Customer billing/relationship management;Other;,C#,Sql Server 2000;,1127,1341
25666,IFPUG,449,Enhancement,Financial application area;,Java,Interactive;,68,68
25725,IFPUG,17400,Enhancement,Financial transaction process/accounting;,C++,HIRDB;,2099,2099
25841,IFPUG,5885,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,633,633
25988,IFPUG,11752,New Development,other: mission-critical system;,Java,ORACLE;,2091,2091
26093,COSMIC-FFP,3187,New Development,Online System for University fraternities;,ASP,MS SQLServer2000;,655,655
26201,IFPUG,3312,New Development,Financial application area;,Java,Interactive;,364,364
26251,IFPUG,112,Enhancement,Financial application area;,Java,Interactive;,31,31
26294,IFPUG,17120,New Development,Business enabling service;,C,DB2;,433,524
26317,IFPUG,1216,Enhancement,Online analysis and reporting;,Java,ORACLE;,156,156
26382,COSMIC-FFP,7111,New Development,Geographic or spatial information system;,C++,Yes;,106,106
26417,IFPUG,2450,New Development,Financial transaction process/accounting;,ASP,SQL Server2000;,360,360
26505,IFPUG,2153,New Development,Financial application area;,Java,Interactive;,234,234
26733,IFPUG,4880,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,649,649
26755,IFPUG,3998,Enhancement,Network Management;,ASP,ORACLE;,1157,1157
27123,COSMIC-FFP,888,New Development,Customer billing/relationship management;,PHP,MySql;,234,234
27553,COSMIC-FFP,19306,Enhancement,Document management;Financial transaction process/accounting;Image, video or sound processing;",COBOL,398,398
27560,IFPUG,3887,Re-development,Operating system or software utility;Other;,C#,Sql Server 2000;,199,221
27732,IFPUG,2218,New Development,other: Sales contact management;,SQL,SQL SERVER;,609,609
27824,IFPUG,1163,New Development,other: DB Serch system;,PHP,SQL SERVER;,160,160
27941,IFPUG,3068,New Development,Management or performance reporting;Online analysis and reporting;,Datastage,DB2;,349,349
28046,IFPUG,120,Enhancement,Web Content & Middleware;,C++,JAVA;,30,34
28127,IFPUG,912,Enhancement,Financial application area;,Java,Interactive;,172,172
28161,IFPUG,1850,Enhancement,Telecom Data Circuits and Revenue;,C++,ORACLE;,200,232

G3-Ran2-67

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points

22508,IFPUG,11052,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,149,149
22524,IFPUG,1139,Enhancement,Financial application area;,Java,Interactive;,156,156
22561,IFPUG,18314,New Development,Financial transaction process/accounting;,SQL,ORACLE;,2245,2245
22570,IFPUG,110,Enhancement,Financial application area;,Java,Interactive;,12,12
22589,IFPUG,187,New Development,Process Control;,Java,Oracle 9i;,41,41
22692,COSMIC-FFP,21600,New Development,Logistic or supply planning & control;,Java,internal tools DB;,115,115
22705,IFPUG,1562,New Development,Financial transaction process/accounting;,Java,Oracle 8i;,218,218
22712,IFPUG,172,New Development,Web-based Application;,Visual Basic,SQL Server;,46,46
22723,IFPUG,5084,Re-Development,Web-based Application;,ASP,SQL Server;,700,700
22869,IFPUG,3566,Enhancement,Device or interface driver;Financial transaction process/accounting;Process Control;,NET,DB2;,211,211
22886,IFPUG,902,Enhancement,relatively complex application;,4GL,Interactive;,110,110
22899,IFPUG,109,Enhancement,Process Control;,Java,Solid;,34,34
23123,IFPUG,1047,New Development,Financial application area;,Java,Interactive;,154,154
23146,IFPUG,3112,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,49,49
23168,IFPUG,838,Enhancement,Financial application area;,Java,Interactive;,133,133
23241,IFPUG,144,Enhancement,tools or system;,Java,Oracle 9i;,95,95
23266,IFPUG,4656,Enhancement,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,621,621
23518,IFPUG,626,Enhancement,relatively complex application;,4GL,Interactive;,25,25
23565,IFPUG,1652,Enhancement,Financial transaction process/accounting;,Java,Interactive;,87,87
23791,IFPUG,48,New Development,other: Sales contact management;,ASP,ORACLE;,62,62
23925,IFPUG,5086,Enhancement,relatively complex application;,4GL,Interactive;,421,421
24017,IFPUG,13728,New Development,Financial transaction process/accounting;,SQL,ORACLE;,1956,1956
24043,IFPUG,1442,New Development,Financial application area;,Java,Interactive;,267,267
24200,IFPUG,5841,Re-development,Geographic or spatial information system;Online analysis and reporting;,C#,SQL;,354,354
24218,IFPUG,2321,New Development,Financial application area;,Java,Interactive;,283,283
24275,IFPUG,5226,New Development,other: production management system;,Java,ORACLE;,578,578
24291,IFPUG,557,Enhancement,Financial application area;,Java,Interactive;,96,96
24483,COSMIC-FFP,8772,Re-development,Workflow support & management;,J2EE,Oracle8i;,465,465
24569,IFPUG,83,New Development,Web-based Application;,Visual Basic,SQL Server;,19,19

24654,IFPUG,3600,New Development,Web-based application;,SQL,ORACLE;,408,408
24677,IFPUG,2624,Enhancement,relatively complex application;,4GL,Interactive;,201,201
25178,IFPUG,1393,New Development,Financial application area;,Java,Interactive;,172,172
25180,IFPUG,7760,New Development,Web-based application;,SQL,ORACLE;,927,908
25287,IFPUG,1186,New Development,Management Information System;,Visual Basic,ORACLE;,128,132
25310,IFPUG,1766,New Development,Financial application area;,Java,Interactive;,256,256
25550,IFPUG,5714,Enhancement,Web-based Application;,Java,Others;,580,580
25552,IFPUG,232,Enhancement,relatively complex application;,4GL,Interactive;,7,7
25620,IFPUG,20096,Re-development,Customer billing/relationship management;Other;,C#,Sql Server 2000;,1127,1341
25641,IFPUG,1718,New Development,other: production management system;,SQL,SQL Server7;,776,776
25666,IFPUG,449,Enhancement,Financial application area;,Java,Interactive;,68,68
25690,IFPUG,1248,New Development,Catalogue/register of things or events;Online analysis and reporting;,Java,ORACLE;,230,230
25704,IFPUG,2464,New Development,Financial application area;,Java,Interactive;,202,202
25725,IFPUG,17400,Enhancement,Financial transaction process/accounting;,C++,HIRDB;,2099,2099
25741,IFPUG,2088,Enhancement,Equipment Management;,SQL,ORACLE;,109,134
25841,IFPUG,5885,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,633,633
25988,IFPUG,11752,New Development,other: mission-critical system;,Java,ORACLE;,2091,2091
26019,IFPUG,2640,New Development,Catalogue/register of things or events;Document management;Workflow support & management;,Java,ORACLE;,234,234
26093,COSMIC-FFP,3187,New Development,Online System for University fraternities;,ASP,MS SQLServer2000;,655,655
26201,IFPUG,3312,New Development,Financial application area;,Java,Interactive;,364,364
26251,IFPUG,112,Enhancement,Financial application area;,Java,Interactive;,31,31
26294,IFPUG,17120,New Development,Business enabling service;,C,DB2;,433,524
26317,IFPUG,1216,Enhancement,Online analysis and reporting;,Java,ORACLE;,156,156
26331,IFPUG,12000,New Development,Web-based Application;,Java,Sybase;,1800,1800
26382,COSMIC-FFP,7111,New Development,Geographic or spatial information system;,C++,Yes;,106,106
26417,IFPUG,2450,New Development,Financial transaction process/accounting;,ASP,SQL Server2000;,360,360
26505,IFPUG,2153,New Development,Financial application area;,Java,Interactive;,234,234
26733,IFPUG,4880,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,649,649

26755,IFPUG,3998,Enhancement,Network Management;,ASP,ORACLE;,1157,1157
27123,COSMIC-FFP,888,New Development,Customer billing/relationship management;,PHP,MySql;,234,234
27553,COSMIC-FFP,19306,Enhancement,Document management;Financial transaction process/accounting;Image, video or sound processing;",COBOL,398,398
27560,IFPUG,3887,Re-development,Operating system or software utility;Other;,C#,Sql Server 2000;,199,221
27732,IFPUG,2218,New Development,other: Sales contact management;,SQL,SQL SERVER;,609,609
27824,IFPUG,1163,New Development,other: DB Serch system;,PHP,SQL SERVER;,160,160
27941,IFPUG,3068,New Development,Management or performance reporting;Online analysis and reporting;,Datastage,DB2;,349,349
28046,IFPUG,120,Enhancement,Web Content & Middleware;,C++,JAVA;,30,34
28127,IFPUG,912,Enhancement,Financial application area;,Java,Interactive;,172,172
28161,IFPUG,1850,Enhancement,Telecom Data Circuits and Revenue;,C++,ORACLE;,200,232

G3-Ran3-17

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
22524,IFPUG,1139,Enhancement,Financial application area;,Java,Interactive;,156,156
22570,IFPUG,110,Enhancement,Financial application area;,Java,Interactive;,12,12
22712,IFPUG,172,New Development,Web-based Application;,Visual Basic,SQL Server;,46,46
22899,IFPUG,109,Enhancement,Process Control;,Java,Solid;,34,34
23146,IFPUG,3112,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,49,49
23266,IFPUG,4656,Enhancement,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,621,621
24017,IFPUG,13728,New Development,Financial transaction process/accounting;,SQL,ORACLE;,1956,1956
24218,IFPUG,2321,New Development,Financial application area;,Java,Interactive;,283,283
24275,IFPUG,5226,New Development,other: production management system;,Java,ORACLE;,578,578
24677,IFPUG,2624,Enhancement,relatively complex application;,4GL,Interactive;,201,201
25310,IFPUG,1766,New Development,Financial application area;,Java,Interactive;,256,256
25552,IFPUG,232,Enhancement,relatively complex application;,4GL,Interactive;,7,7
25641,IFPUG,1718,New Development,other: production management system;,SQL,SQL Server7;,776,776
25988,IFPUG,11752,New Development,other: mission-critical system;,Java,ORACLE;,2091,2091

26251,IFPUG,112,Enhancement,Financial application area;,Java,Interactive;,31,31
26733,IFPUG,4880,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,649,649
27941,IFPUG,3068,New Development,Management or performance reporting;Online analysis and reporting;,Datastage,DB2;,349,349

G3-Ran3-33

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
22508,IFPUG,11052,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,149,149
22524,IFPUG,1139,Enhancement,Financial application area;,Java,Interactive;,156,156
22570,IFPUG,110,Enhancement,Financial application area;,Java,Interactive;,12,12
22692,COSMIC-FFP,21600,New Development,Logistic or supply planning & control;,Java,internal tools DB;,115,115
22712,IFPUG,172,New Development,Web-based Application;,Visual Basic,SQL Server;,46,46
22723,IFPUG,5084,Re-Development,Web-based Application;,ASP,SQL Server;,700,700
22899,IFPUG,109,Enhancement,Process Control;,Java,Solid;,34,34
23146,IFPUG,3112,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,49,49
23241,IFPUG,144,Enhancement,tools or system;,Java,Oracle 9i;,95,95
23266,IFPUG,4656,Enhancement,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,621,621
23791,IFPUG,48,New Development,other: Sales contact management;,ASP,ORACLE;,62,62
23925,IFPUG,5086,Enhancement,relatively complex application;,4GL,Interactive;,421,421
24017,IFPUG,13728,New Development,Financial transaction process/accounting;,SQL,ORACLE;,1956,1956
24218,IFPUG,2321,New Development,Financial application area;,Java,Interactive;,283,283
24275,IFPUG,5226,New Development,other: production management system;,Java,ORACLE;,578,578
24483,COSMIC-FFP,8772,Re-development,Workflow support & management;,J2EE,Oracle8i;,465,465
24677,IFPUG,2624,Enhancement,relatively complex application;,4GL,Interactive;,201,201
25180,IFPUG,7760,New Development,Web-based application;,SQL,ORACLE;,927,908
25310,IFPUG,1766,New Development,Financial application area;,Java,Interactive;,256,256
25552,IFPUG,232,Enhancement,relatively complex application;,4GL,Interactive;,7,7
25641,IFPUG,1718,New Development,other: production management system;,SQL,SQL Server7;,776,776

25725,IFPUG,17400,Enhancement,Financial transaction process/accounting;,C++,HIRDB;,2099,2099
25741,IFPUG,2088,Enhancement,Equipment Management;,SQL,ORACLE;,109,134
25988,IFPUG,11752,New Development,other: mission-critical system;,Java,ORACLE;,2091,2091
26251,IFPUG,112,Enhancement,Financial application area;,Java,Interactive;,31,31
26331,IFPUG,12000,New Development,Web-based Application;,Java,Sybase;,1800,1800
26505,IFPUG,2153,New Development,Financial application area;,Java,Interactive;,234,234
26733,IFPUG,4880,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,649,649
27560,IFPUG,3887,Re-development,Operating system or software utility;Other;,C#,Sql Server 2000;,199,221
27824,IFPUG,1163,New Development,other: DB Serch system;,PHP,SQL SERVER;,160,160
27941,IFPUG,3068,New Development,Management or performance reporting;Online analysis and reporting;,Datastage,DB2;,349,349
28127,IFPUG,912,Enhancement,Financial application area;,Java,Interactive;,172,172
28161,IFPUG,1850,Enhancement,Telecom Data Circuits and Revenue;,C++,ORACLE;,200,232

G3-Ran3-49

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
22508,IFPUG,11052,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,149,149
22524,IFPUG,1139,Enhancement,Financial application area;,Java,Interactive;,156,156
22570,IFPUG,110,Enhancement,Financial application area;,Java,Interactive;,12,12
22589,IFPUG,187,New Development,Process Control;,Java,Oracle 9i;,41,41
22692,COSMIC-FFP,21600,New Development,Logistic or supply planning & control;,Java,internal tools DB;,115,115
22712,IFPUG,172,New Development,Web-based Application;,Visual Basic,SQL Server;,46,46
22723,IFPUG,5084,Re-Development,Web-based Application;,ASP,SQL Server;,700,700
22899,IFPUG,109,Enhancement,Process Control;,Java,Solid;,34,34
23123,IFPUG,1047,New Development,Financial application area;,Java,Interactive;,154,154
23146,IFPUG,3112,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,49,49
23241,IFPUG,144,Enhancement,tools or system;,Java,Oracle 9i;,95,95
23266,IFPUG,4656,Enhancement,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,621,621

23565,IFPUG,1652,Enhancement,Financial transaction process/accounting;,Java,Interactive;,87,87
23791,IFPUG,48,New Development,other: Sales contact management;,ASP,ORACLE;,62,62
23925,IFPUG,5086,Enhancement,relatively complex application;,4GL,Interactive;,421,421
24017,IFPUG,13728,New Development,Financial transaction process/accounting;,SQL,ORACLE;,1956,1956
24200,IFPUG,5841,Re-development,Geographic or spatial information system;Online analysis and reporting;,C#,SQL;,354,354
24218,IFPUG,2321,New Development,Financial application area;,Java,Interactive;,283,283
24275,IFPUG,5226,New Development,other: production management system;,Java,ORACLE;,578,578
24483,COSMIC-FFP,8772,Re-development,Workflow support & management;,J2EE,Oracle8i;,465,465
24569,IFPUG,83,New Development,Web-based Application;,Visual Basic,SQL Server;,19,19
24677,IFPUG,2624,Enhancement,relatively complex application;,4GL,Interactive;,201,201
25180,IFPUG,7760,New Development,Web-based application;,SQL,ORACLE;,927,908
25310,IFPUG,1766,New Development,Financial application area;,Java,Interactive;,256,256
25550,IFPUG,5714,Enhancement,Web-based Application;,Java,Others;,580,580
25552,IFPUG,232,Enhancement,relatively complex application;,4GL,Interactive;,7,7
25641,IFPUG,1718,New Development,other: production management system;,SQL,SQL Server7;,776,776
25666,IFPUG,449,Enhancement,Financial application area;,Java,Interactive;,68,68
25704,IFPUG,2464,New Development,Financial application area;,Java,Interactive;,202,202
25725,IFPUG,17400,Enhancement,Financial transaction process/accounting;,C++,HIRDB;,2099,2099
25741,IFPUG,2088,Enhancement,Equipment Management;,SQL,ORACLE;,109,134
25988,IFPUG,11752,New Development,other: mission-critical system;,Java,ORACLE;,2091,2091
26019,IFPUG,2640,New Development,Catalogue/register of things or events;Document management;Workflow support & management;,Java,ORACLE;,234,234
26201,IFPUG,3312,New Development,Financial application area;,Java,Interactive;,364,364
26251,IFPUG,112,Enhancement,Financial application area;,Java,Interactive;,31,31
26317,IFPUG,1216,Enhancement,Online analysis and reporting;,Java,ORACLE;,156,156
26331,IFPUG,12000,New Development,Web-based Application;,Java,Sybase;,1800,1800
26382,COSMIC-FFP,7111,New Development,Geographic or spatial information system;,C++,Yes;,106,106
26505,IFPUG,2153,New Development,Financial application area;,Java,Interactive;,234,234
26733,IFPUG,4880,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,649,649

27123,COSMIC-FFP,888,New Development,Customer billing/relationship management;,PHP,MySql;,234,234
27553,COSMIC-FFP,19306,Enhancement,Document management;Financial transaction process/accounting;Image, video or sound processing;",COBOL,398,398
27560,IFPUG,3887,Re-development,Operating system or software utility;Other;,C#,Sql Server 2000;,199,221
27732,IFPUG,2218,New Development,other: Sales contact management;,SQL,SQL SERVER;,609,609
27824,IFPUG,1163,New Development,other: DB Serch system;,PHP,SQL SERVER;,160,160
27941,IFPUG,3068,New Development,Management or performance reporting;Online analysis and reporting;,Datastage,DB2;,349,349
28046,IFPUG,120,Enhancement,Web Content & Middleware;,C++,JAVA;,30,34
28127,IFPUG,912,Enhancement,Financial application area;,Java,Interactive;,172,172
28161,IFPUG,1850,Enhancement,Telecom Data Circuits and Revenue;,C++,ORACLE;,200,232

G3-Ran3-67

Case Name,Count Approach,Summary work Effort ,Development Type,Application Type,Primary Programming Language ,1st Database System,Functional Size,Adjusted Functional Points
22508,IFPUG,11052,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,149,149
22524,IFPUG,1139,Enhancement,Financial application area;,Java,Interactive;,156,156
22561,IFPUG,18314,New Development,Financial transaction process/accounting;,SQL,ORACLE;,2245,2245
22570,IFPUG,110,Enhancement,Financial application area;,Java,Interactive;,12,12
22589,IFPUG,187,New Development,Process Control;,Java,Oracle 9i;,41,41
22692,COSMIC-FFP,21600,New Development,Logistic or supply planning & control;,Java,internal tools DB;,115,115
22705,IFPUG,1562,New Development,Financial transaction process/accounting;,Java,Oracle 8i;,218,218
22712,IFPUG,172,New Development,Web-based Application;,Visual Basic,SQL Server;,46,46
22723,IFPUG,5084,Re-Development,Web-based Application;,ASP,SQL Server;,700,700
22869,IFPUG,3566,Enhancement,Device or interface driver;Financial transaction process/accounting;Process Control;,NET,DB2;,211,211
22886,IFPUG,902,Enhancement,relatively complex application;,4GL,Interactive;,110,110
22899,IFPUG,109,Enhancement,Process Control;,Java,Solid;,34,34
23123,IFPUG,1047,New Development,Financial application area;,Java,Interactive;,154,154
23146,IFPUG,3112,Enhancement,Financial transaction process/accounting;,Java,DB2 UDB;,49,49
23168,IFPUG,838,Enhancement,Financial application area;,Java,Interactive;,133,133
23241,IFPUG,144,Enhancement,tools or system;,Java,Oracle 9i;,95,95

23266,IFPUG,4656,Enhancement,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,621,621
23518,IFPUG,626,Enhancement,relatively complex application;,4GL,Interactive;,25,25
23565,IFPUG,1652,Enhancement,Financial transaction process/accounting;,Java,Interactive;,87,87
23791,IFPUG,48,New Development,other: Sales contact management;,ASP,ORACLE;,62,62
23925,IFPUG,5086,Enhancement,relatively complex application;,4GL,Interactive;,421,421
24017,IFPUG,13728,New Development,Financial transaction process/accounting;,SQL,ORACLE;,1956,1956
24043,IFPUG,1442,New Development,Financial application area;,Java,Interactive;,267,267
24200,IFPUG,5841,Re-development,Geographic or spatial information system;Online analysis and reporting;,C#,SQL;,354,354
24218,IFPUG,2321,New Development,Financial application area;,Java,Interactive;,283,283
24275,IFPUG,5226,New Development,other: production management system;,Java,ORACLE;,578,578
24291,IFPUG,557,Enhancement,Financial application area;,Java,Interactive;,96,96
24483,COSMIC-FFP,8772,Re-development,Workflow support & management;,J2EE,Oracle8i;,465,465
24569,IFPUG,83,New Development,Web-based Application;,Visual Basic,SQL Server;,19,19
24654,IFPUG,3600,New Development,Web-based application;,SQL,ORACLE;,408,408
24677,IFPUG,2624,Enhancement,relatively complex application;,4GL,Interactive;,201,201
25178,IFPUG,1393,New Development,Financial application area;,Java,Interactive;,172,172
25180,IFPUG,7760,New Development,Web-based application;,SQL,ORACLE;,927,908
25287,IFPUG,1186,New Development,Management Information System;,Visual Basic,ORACLE;,128,132
25310,IFPUG,1766,New Development,Financial application area;,Java,Interactive;,256,256
25550,IFPUG,5714,Enhancement,Web-based Application;,Java,Others;,580,580
25552,IFPUG,232,Enhancement,relatively complex application;,4GL,Interactive;,7,7
25620,IFPUG,20096,Re-development,Customer billing/relationship management;Other;,C#,Sql Server 2000;,1127,1341
25641,IFPUG,1718,New Development,other: production management system;,SQL,SQL Server7;,776,776
25666,IFPUG,449,Enhancement,Financial application area;,Java,Interactive;,68,68
25690,IFPUG,1248,New Development,Catalogue/register of things or events;Online analysis and reporting;,Java,ORACLE;,230,230
25704,IFPUG,2464,New Development,Financial application area;,Java,Interactive;,202,202
25725,IFPUG,17400,Enhancement,Financial transaction process/accounting;,C++,HIRDB;,2099,2099
25741,IFPUG,2088,Enhancement,Equipment Management;,SQL,ORACLE;,109,134
25841,IFPUG,5885,New Development,Financial transaction process/accounting;,Java,Oracle 8i8,633,633

25988,IFPUG,11752,New Development,other: mission-critical system;,Java,ORACLE;,2091,2091
26019,IFPUG,2640,New Development,Catalogue/register of things or events;Document management;Workflow support & management;,Java,ORACLE;,234,234
26093,COSMIC-FFP,3187,New Development,Online System for University fraternities;,ASP,MS SQLServer2000;,655,655
26201,IFPUG,3312,New Development,Financial application area;,Java,Interactive;,364,364
26251,IFPUG,112,Enhancement,Financial application area;,Java,Interactive;,31,31
26294,IFPUG,17120,New Development,Business enabling service;,C,DB2;,433,524
26317,IFPUG,1216,Enhancement,Online analysis and reporting;,Java,ORACLE;,156,156
26331,IFPUG,12000,New Development,Web-based Application;,Java,Sybase;,1800,1800
26382,COSMIC-FFP,7111,New Development,Geographic or spatial information system;,C++,Yes;,106,106
26417,IFPUG,2450,New Development,Financial transaction process/accounting;,ASP,SQL Server2000;,360,360
26505,IFPUG,2153,New Development,Financial application area;,Java,Interactive;,234,234
26733,IFPUG,4880,New Development,Catalogue/register of things or events;Document management;Online analysis and reporting;,Java,ORACLE;,649,649
26755,IFPUG,3998,Enhancement,Network Management;,ASP,ORACLE;,1157,1157
27123,COSMIC-FFP,888,New Development,Customer billing/relationship management;,PHP,MySql;,234,234
27553,COSMIC-FFP,19306,Enhancement,Document management;Financial transaction process/accounting;Image, video or sound processing;",COBOL,398,398
27560,IFPUG,3887,Re-development,Operating system or software utility;Other;,C#,Sql Server 2000;,199,221
27732,IFPUG,2218,New Development,other: Sales contact management;,SQL,SQL SERVER;,609,609
27824,IFPUG,1163,New Development,other: DB Serch system;,PHP,SQL SERVER;,160,160
27941,IFPUG,3068,New Development,Management or performance reporting;Online analysis and reporting;,Datastage,DB2;,349,349
28046,IFPUG,120,Enhancement,Web Content & Middleware;,C++,JAVA;,30,34
28127,IFPUG,912,Enhancement,Financial application area;,Java,Interactive;,172,172
28161,IFPUG,1850,Enhancement,Telecom Data Circuits and Revenue;,C++,ORACLE;,200,232

Appendix E- Desharnais Raw Dataset

Dataset-1

TeamExp	ManagerEXp	Length	Effort	Transactions	Entities	PointsAdjust	Envergure	PointsNonAdjust	Language
1	4	4	847	158	59	217	18	180	3
1	2	5	2548	74	43	117	25	105	2
3	3	6	3136	86	49	135	32	131	1
4	3	6	3927	79	128	207	27	190	1
3	1	14	4277	148	324	472	39	491	1
3	4	14	4494	9	386	395	21	340	2
1	1	9	4620	451	48	499	28	464	1
0	0	4	5635	197	124	321	33	315	1
2	4	5	5775	306	132	438	37	447	1
3	3	9	5817	96	108	204	29	192	1
4	3	12	5880	469	176	645	43	697	3
2	1	18	6699	182	126	308	35	308	1
2	1	10	6783	224	110	334	28	311	2
3	1	12	7854	172	88	260	30	247	1
4	4	24	9520	395	193	588	40	617	1

Dataset-2

TeamExp	ManagerEXp	Length	Effort	Transactions	Entities	PointsAdjust	Envergure	PointsNonAdjust	Language
0	4	6	546	97	42	139	6	99	3
1	4	4	847	158	59	217	18	180	3
1	3	10	1267	42	31	73	27	67	2
4	4	12	1400	229	169	398	39	414	3
2	4	11	1435	289	88	377	28	351	3
1	2	6	1575	47	32	79	14	62	2
4	7	13	1603	69	74	143	14	113	1
3	2	8	1617	119	48	167	26	152	2
0	0	4	2149	140	94	234	24	208	1
2	3	13	2275	134	77	211	13	165	2
2	4	34	2352	661	132	793	23	698	3
3	4	4	2422	78	38	116	24	103	1
4	4	9	2429	174	78	252	41	267	3
1	1	5	2520	78	99	177	14	140	1
2	1	9	2569	119	42	161	25	145	2
1	1	12	2583	61	96	157	18	130	1
1	4	8	2723	124	52	176	14	139	2
1	1	12	2926	126	107	233	23	205	2
3	3	6	3136	86	49	135	32	131	1
4	1	14	3164	86	230	316	33	310	1
4	3	8	3192	57	43	100	43	108	1
1	1	12	3276	55	112	167	12	129	2
4	4	14	3437	68	316	384	20	326	2
3	3	5	3472	120	126	246	15	197	2

2	0	6	3542	71	235	306	37	312	1
1	3	8	3626	194	97	291	35	291	2
1	3	12	3647	132	89	221	5	155	2
0	0	5	3829	200	119	319	30	303	1
1	2	13	3913	186	52	238	25	214	1
4	3	6	3927	79	128	207	27	190	1
0	1	22	3941	139	143	282	22	245	2
4	1	14	3948	175	277	452	37	461	1
1	4	8	3983	89	200	289	33	283	1
2	2	9	4004	252	7	259	28	241	1
4	1	21	4067	167	99	266	24	237	1
3	4	8	4172	162	61	223	32	216	1
3	1	14	4277	148	324	472	39	491	1
3	4	14	4494	9	386	395	21	340	2
1	1	9	4620	451	48	499	28	464	1
4	4	9	4977	223	121	344	28	320	1
1	4	12	5152	253	52	305	34	302	1
2	4	18	5180	88	170	258	34	255	1
0	0	4	5635	197	124	321	33	315	1
2	4	5	5775	306	132	438	37	447	1
3	3	9	5817	96	108	204	29	192	1
4	3	12	5880	469	176	645	43	697	3
1	1	5	6405	194	91	285	35	285	1
2	1	18	6699	182	126	308	35	308	1
2	1	10	6783	224	110	334	28	311	2
4	4	16	7252	116	170	286	27	263	1

Dataset-3

TeamExp	ManagerEXp	Length	Effort	Transactions	Entities	PointsAdjust	Envergure	PointsNonAdjust	Language
1	4	4	847	158	59	217	18	180	3
1	2	5	2548	74	43	117	25	105	2
3	3	6	3136	86	49	135	32	131	1
4	3	6	3927	79	128	207	27	190	1
3	1	14	4277	148	324	472	39	491	1
3	4	14	4494	9	386	395	21	340	2
1	1	9	4620	451	48	499	28	464	1
0	0	4	5635	197	124	321	33	315	1
2	4	5	5775	306	132	438	37	447	1
3	3	9	5817	96	108	204	29	192	1
4	3	12	5880	469	176	645	43	697	3
2	1	18	6699	182	126	308	35	308	1
2	1	10	6783	224	110	334	28	311	2
3	1	12	7854	172	88	260	30	247	1
4	4	36	23940	886	241	1127	34	1116	1

Dataset-4

TeamExp	ManagerEXp	Length	Effort	Transactions	Entities	PointsAdjust	Envergure	PointsNonAdjust	Language
0	4	6	546	97	42	139	6	99	3
1	4	4	847	158	59	217	18	180	3
1	3	10	1267	42	31	73	27	67	2
4	4	12	1400	229	169	398	39	414	3
2	4	11	1435	289	88	377	28	351	3
1	2	6	1575	47	32	79	14	62	2
4	7	13	1603	69	74	143	14	113	1
3	2	8	1617	119	48	167	26	152	2
0	0	4	2149	140	94	234	24	208	1
2	3	13	2275	134	77	211	13	165	2
2	4	34	2352	661	132	793	23	698	3
3	4	4	2422	78	38	116	24	103	1
4	4	9	2429	174	78	252	41	267	3
1	1	5	2520	78	99	177	14	140	1
2	1	9	2569	119	42	161	25	145	2
1	1	12	2583	61	96	157	18	130	1
1	4	8	2723	124	52	176	14	139	2
1	1	12	2926	126	107	233	23	205	2
3	3	6	3136	86	49	135	32	131	1
4	1	14	3164	86	230	316	33	310	1
4	3	8	3192	57	43	100	43	108	1
1	1	12	3276	55	112	167	12	129	2
4	4	14	3437	68	316	384	20	326	2
3	3	5	3472	120	126	246	15	197	2

2	0	6	3542	71	235	306	37	312	1
1	3	8	3626	194	97	291	35	291	2
1	3	12	3647	132	89	221	5	155	2
0	0	5	3829	200	119	319	30	303	1
1	2	13	3913	186	52	238	25	214	1
4	3	6	3927	79	128	207	27	190	1
0	1	22	3941	139	143	282	22	245	2
4	1	14	3948	175	277	452	37	461	1
1	4	8	3983	89	200	289	33	283	1
2	2	9	4004	252	7	259	28	241	1
4	1	21	4067	167	99	266	24	237	1
3	4	8	4172	162	61	223	32	216	1
3	1	14	4277	148	324	472	39	491	1
3	4	14	4494	9	386	395	21	340	2
1	1	9	4620	451	48	499	28	464	1
4	4	9	4977	223	121	344	28	320	1
1	4	12	5152	253	52	305	34	302	1
2	4	18	5180	88	170	258	34	255	1
0	0	4	5635	197	124	321	33	315	1
2	4	5	5775	306	132	438	37	447	1
3	3	9	5817	96	108	204	29	192	1
4	3	12	5880	469	176	645	43	697	3
1	1	5	6405	194	91	285	35	285	1
2	1	18	6699	182	126	308	35	308	1
2	1	10	6783	224	110	334	28	311	2
4	4	36	23940	886	241	1127	34	1116	1

Dataset-5

TeamExp	ManagerEXp	Length	Effort	Transactions	Entities	PointsAdjust	Envergure	PointsNonAdjust	Language
0	4	6	546	97	42	139	6	99	3
0	2	6	595	213	73	286	6	203	3
2	2	3	651	126	49	175	38	180	3
1	1	9	710	145	38	183	27	168	3
4	4	1	805	40	60	100	18	83	1
4	2	5	840	58	34	92	29	86	1
1	4	4	847	158	59	217	18	180	3
3	4	10	1155	101	57	158	9	117	2
1	3	10	1267	42	31	73	27	67	2
4	4	12	1400	229	169	398	39	414	3
2	4	11	1435	289	88	377	28	351	3
1	2	6	1575	47	32	79	14	62	2
4	7	13	1603	69	74	143	14	113	1
3	2	8	1617	119	48	167	26	152	2
3	2	6	1876	101	45	146	15	117	2

Dataset-6

TeamExp	ManagerEXp	Length	Effort	Transactions	Entities	PointsAdjust	Envergure	PointsNonAdjust	Language
0	4	6	546	97	42	139	6	99	3
0	2	6	595	213	73	286	6	203	3
2	2	3	651	126	49	175	38	180	3
1	1	9	710	145	38	183	27	168	3
4	4	1	805	40	60	100	18	83	1
4	2	5	840	58	34	92	29	86	1
1	4	4	847	158	59	217	18	180	3
3	4	10	1155	101	57	158	9	117	2
1	3	10	1267	42	31	73	27	67	2
4	4	12	1400	229	169	398	39	414	3
2	4	11	1435	289	88	377	28	351	3
1	2	6	1575	47	32	79	14	62	2
4	7	13	1603	69	74	143	14	113	1
3	2	8	1617	119	48	167	26	152	2
3	2	6	1876	101	45	146	15	117	2
0	0	4	2149	140	94	234	24	208	1
1	1	10	2174	64	54	118	25	106	1
2	3	13	2275	134	77	211	13	165	2
1	1	3	2282	33	72	105	19	88	1
2	3	8	2331	106	39	145	6	103	1
2	4	34	2352	661	132	793	23	698	3
3	4	4	2422	78	38	116	24	103	1
4	4	9	2429	174	78	252	41	267	3
1	1	5	2520	78	99	177	14	140	1

1	2	5	2548	74	43	117	25	105	2
2	1	9	2569	119	42	161	25	145	2
1	1	12	2583	61	96	157	18	130	1
1	4	8	2723	124	52	176	14	139	2
4	3	12	2800	227	73	300	34	297	1
0	0	4	2821	97	89	186	38	192	1
1	1	12	2926	126	107	233	23	205	2
2	3	7	2989	116	72	188	18	156	1
3	3	6	3136	86	49	135	32	131	1
4	1	14	3164	86	230	316	33	310	1
4	3	8	3192	57	43	100	43	108	1
1	1	12	3276	55	112	167	12	129	2
4	4	14	3437	68	316	384	20	326	2
3	3	5	3472	120	126	246	15	197	2
2	0	6	3542	71	235	306	37	312	1
1	3	8	3626	194	97	291	35	291	2
1	3	12	3647	132	89	221	5	155	2
0	0	5	3829	200	119	319	30	303	1
1	2	13	3913	186	52	238	25	214	1
4	3	6	3927	79	128	207	27	190	1
0	1	22	3941	139	143	282	22	245	2
4	1	14	3948	175	277	452	37	461	1
1	4	8	3983	89	200	289	33	283	1
2	2	9	4004	252	7	259	28	241	1
4	1	21	4067	167	99	266	24	237	1
3	4	8	4172	162	61	223	32	216	1

Dataset-7

TeamExp	ManagerEXp	Length	Effort	Transactions	Entities	PointsAdjust	Envergure	PointsNonAdjust	Language
1	4	4	847	158	59	217	18	180	3
4	4	12	14973	318	269	587	34	581	2
3	3	6	3136	86	49	135	32	131	1
4	3	6	3927	79	128	207	27	190	1
3	1	14	4277	148	324	472	39	491	1
3	4	14	4494	9	386	395	21	340	2
2	3	17	14434	221	121	342	35	342	1
0	0	4	5635	197	124	321	33	315	1
2	4	5	5775	306	132	438	37	447	1
3	3	9	5817	96	108	204	29	192	1
4	3	12	5880	469	176	645	43	697	3
2	1	18	6699	182	126	308	35	308	1
2	1	10	6783	224	110	334	28	311	2
3	1	12	7854	172	88	260	30	247	1
4	4	24	9520	395	193	588	40	617	1

Dataset-8

TeamExp	ManagerEXp	Length	Effort	Transactions	Entities	PointsAdjust	Envergure	PointsNonAdjust	Language
0	4	6	546	97	42	139	6	99	3
1	4	4	847	158	59	217	18	180	3
1	3	10	1267	42	31	73	27	67	2
4	4	12	1400	229	169	398	39	414	3
2	4	11	1435	289	88	377	28	351	3
1	2	6	1575	47	32	79	14	62	2
4	7	13	1603	69	74	143	14	113	1
3	2	8	1617	119	48	167	26	152	2
0	0	4	2149	140	94	234	24	208	1
2	3	13	2275	134	77	211	13	165	2
2	4	34	2352	661	132	793	23	698	3
3	4	4	2422	78	38	116	24	103	1
4	4	9	2429	174	78	252	41	267	3
1	1	5	2520	78	99	177	14	140	1
2	1	9	2569	119	42	161	25	145	2
1	1	12	2583	61	96	157	18	130	1
1	4	8	2723	124	52	176	14	139	2
1	1	12	2926	126	107	233	23	205	2
3	3	6	3136	86	49	135	32	131	1
4	1	14	3164	86	230	316	33	310	1
4	3	8	3192	57	43	100	43	108	1
4	1	20	10577	304	78	382	39	397	1
2	4	15	11361	323	184	507	35	507	2
3	3	5	3472	120	126	246	15	197	2

2	0	6	3542	71	235	306	37	312	1
1	3	8	3626	194	97	291	35	291	2
1	3	12	3647	132	89	221	5	155	2
0	0	5	3829	200	119	319	30	303	1
1	2	13	3913	186	52	238	25	214	1
4	3	6	3927	79	128	207	27	190	1
4	4	12	14973	318	269	587	34	581	2
4	1	14	3948	175	277	452	37	461	1
1	4	8	3983	89	200	289	33	283	1
2	2	9	4004	252	7	259	28	241	1
4	1	21	4067	167	99	266	24	237	1
2	3	17	14434	221	121	342	35	342	1
3	1	14	4277	148	324	472	39	491	1
3	4	14	4494	9	386	395	21	340	2
1	1	9	4620	451	48	499	28	464	1
4	4	9	4977	223	121	344	28	320	1
1	4	12	5152	253	52	305	34	302	1
2	4	18	5180	88	170	258	34	255	1
0	0	4	5635	197	124	321	33	315	1
2	4	5	5775	306	132	438	37	447	1
3	3	9	5817	96	108	204	29	192	1
4	3	12	5880	469	176	645	43	697	3
1	1	5	6405	194	91	285	35	285	1
2	1	18	6699	182	126	308	35	308	1
2	1	10	6783	224	110	334	28	311	2
4	4	16	7252	116	170	286	27	263	1

Dataset-9

TeamExp	ManagerEXp	Length	Effort	Transactions	Entities	PointsAdjust	Envergure	PointsNonAdjust	Language
1	4	12	5152	253	52	305	34	302	1
0	0	4	5635	197	124	321	33	315	1
4	4	1	805	40	60	100	18	83	1
0	0	5	3829	200	119	319	30	303	1
0	0	4	2149	140	94	234	24	208	1
0	0	4	2821	97	89	186	38	192	1
2	1	9	2569	119	42	161	25	145	2
1	2	13	3913	186	52	238	25	214	1
3	1	12	7854	172	88	260	30	247	1
3	4	4	2422	78	38	116	24	103	1
4	1	21	4067	167	99	266	24	237	1
2	1	17	9051	146	112	258	40	271	1
1	1	3	2282	33	72	105	19	88	1
3	4	8	4172	162	61	223	32	216	1
4	4	9	4977	223	121	344	28	320	1

Dataset-10

TeamExp	ManagerEXp	Length	Effort	Transactions	Entities	PointsAdjust	Envergure	PointsNonAdjust	Language
4	4	12	14973	318	269	587	34	581	2
2	4	18	5180	88	170	258	34	255	1
2	4	5	5775	306	132	438	37	447	1
4	1	20	10577	304	78	382	39	397	1
1	4	8	3983	89	200	289	33	283	1
4	1	14	3164	86	230	316	33	310	1
2	0	6	3542	71	235	306	37	312	1
3	1	14	4277	148	324	472	39	491	1
4	4	16	7252	116	170	286	27	263	1
4	1	14	3948	175	277	452	37	461	1
4	3	6	3927	79	128	207	27	190	1
1	1	9	710	145	38	183	27	168	3
4	4	9	2429	174	78	252	41	267	3
1	1	5	6405	194	91	285	35	285	1
2	2	3	651	126	49	175	38	180	3
1	3	17	9135	137	119	256	34	253	2
2	4	11	1435	289	88	377	28	351	3
1	1	8	5922	260	144	404	24	360	1
1	4	4	847	158	59	217	18	180	3
3	3	16	8050	302	145	447	52	523	2
1	1	9	4620	451	48	499	28	464	1
2	4	34	2352	661	132	793	23	698	3
1	1	10	2174	64	54	118	25	106	1
1	4	39	19894	284	230	514	50	591	1

2	1	18	6699	182	126	308	35	308	1
2	3	27	14987	173	332	505	19	424	1
2	2	9	4004	252	7	259	28	241	1
4	3	11	12824	131	180	311	51	361	1
2	3	8	2331	106	39	145	6	103	1
3	3	9	5817	96	108	204	29	192	1
2	3	7	2989	116	72	188	18	156	1
3	3	6	3136	86	49	135	32	131	1
2	3	17	14434	221	121	342	35	342	1
1	1	12	2583	61	96	157	18	130	1
1	3	12	3647	132	89	221	5	155	2
3	7	13	8232	45	387	432	16	350	2
1	1	12	3276	55	112	167	12	129	2
1	4	8	2723	124	52	176	14	139	2
3	3	5	3472	120	126	246	15	197	2
1	2	6	1575	47	32	79	14	62	2
1	1	12	2926	126	107	233	23	205	2
3	2	6	1876	101	45	146	15	117	2
1	1	5	2520	78	99	177	14	140	1
4	7	13	1603	69	74	143	14	113	1
1	3	8	3626	194	97	291	35	291	2
2	1	10	6783	224	110	334	28	311	2
2	4	15	11361	323	184	507	35	507	2
1	3	10	1267	42	31	73	27	67	2
1	2	5	2548	74	43	117	25	105	2
3	4	10	1155	101	57	158	9	117	2

Dataset-11

TeamExp	ManagerEXp	Length	Effort	Transactions	Entities	PointsAdjust	Envergure	PointsNonAdjust	Language
1	4	12	5152	253	52	305	34	302	1
0	0	4	5635	197	124	321	33	315	1
4	4	1	805	40	60	100	18	83	1
0	0	5	3829	200	119	319	30	303	1
0	0	4	2149	140	94	234	24	208	1
0	0	4	2821	97	89	186	38	192	1
2	1	9	2569	119	42	161	25	145	2
1	2	13	3913	186	52	238	25	214	1
3	1	12	7854	172	88	260	30	247	1
3	4	4	2422	78	38	116	24	103	1
4	1	21	4067	167	99	266	24	237	1
2	3	27	14987	173	332	505	19	424	1
1	1	3	2282	33	72	105	19	88	1
3	4	8	4172	162	61	223	32	216	1
1	4	39	19894	284	230	514	50	591	1

Dataset-12

TeamExp	ManagerEXp	Length	Effort	Transactions	Entities	PointsAdjust	Envergure	PointsNonAdjust	Language
4	4	12	14973	318	269	587	34	581	2
2	4	18	5180	88	170	258	34	255	1
2	4	5	5775	306	132	438	37	447	1
4	1	20	10577	304	78	382	39	397	1
1	4	8	3983	89	200	289	33	283	1
4	1	14	3164	86	230	316	33	310	1
2	0	6	3542	71	235	306	37	312	1
3	1	14	4277	148	324	472	39	491	1
4	4	16	7252	116	170	286	27	263	1
4	1	14	3948	175	277	452	37	461	1
4	3	6	3927	79	128	207	27	190	1
1	1	9	710	145	38	183	27	168	3
4	4	9	2429	174	78	252	41	267	3
1	1	5	6405	194	91	285	35	285	1
2	2	3	651	126	49	175	38	180	3
4	4	36	23940	886	241	1127	34	1116	1
2	4	11	1435	289	88	377	28	351	3
1	1	8	5922	260	144	404	24	360	1
1	4	4	847	158	59	217	18	180	3
3	3	16	8050	302	145	447	52	523	2
1	1	9	4620	451	48	499	28	464	1
2	4	34	2352	661	132	793	23	698	3
1	1	10	2174	64	54	118	25	106	1
1	4	39	19894	284	230	514	50	591	1

2	1	18	6699	182	126	308	35	308	1
2	3	27	14987	173	332	505	19	424	1
2	2	9	4004	252	7	259	28	241	1
4	3	11	12824	131	180	311	51	361	1
2	3	8	2331	106	39	145	6	103	1
3	3	9	5817	96	108	204	29	192	1
2	3	7	2989	116	72	188	18	156	1
3	3	6	3136	86	49	135	32	131	1
2	3	17	14434	221	121	342	35	342	1
1	1	12	2583	61	96	157	18	130	1
1	3	12	3647	132	89	221	5	155	2
3	7	13	8232	45	387	432	16	350	2
1	1	12	3276	55	112	167	12	129	2
1	4	8	2723	124	52	176	14	139	2
3	3	5	3472	120	126	246	15	197	2
1	2	6	1575	47	32	79	14	62	2
1	1	12	2926	126	107	233	23	205	2
3	2	6	1876	101	45	146	15	117	2
1	1	5	2520	78	99	177	14	140	1
4	7	13	1603	69	74	143	14	113	1
1	3	8	3626	194	97	291	35	291	2
2	1	10	6783	224	110	334	28	311	2
2	4	15	11361	323	184	507	35	507	2
1	3	10	1267	42	31	73	27	67	2
1	2	5	2548	74	43	117	25	105	2
3	4	10	1155	101	57	158	9	117	2

Dataset-13

TeamExp	ManagerEXp	Length	Effort	Transactions	Entities	PointsAdjust	Envergure	PointsNonAdjust	Language
0	4	6	546	97	42	139	6	99	3
0	2	6	595	213	73	286	6	203	3
1	1	9	710	145	38	183	27	168	3
2	4	11	1435	289	88	377	28	351	3
1	2	6	1575	47	32	79	14	62	2
0	0	4	2821	97	89	186	38	192	1
1	1	12	2926	126	107	233	23	205	2
2	3	7	2989	116	72	188	18	156	1
3	3	6	3136	86	49	135	32	131	1
3	4	14	4494	9	386	395	21	340	2
0	0	4	5635	197	124	321	33	315	1
3	3	9	5817	96	108	204	29	192	1
2	1	10	6783	224	110	334	28	311	2
4	4	16	7252	116	170	286	27	263	1
4	5	26	9100	482	227	709	26	645	2

Dataset-14

TeamExp	ManagerEXp	Length	Effort	Transactions	Entities	PointsAdjust	Envergure	PointsNonAdjust	Language
0	4	6	546	97	42	139	6	99	3
0	2	6	595	213	73	286	6	203	3
2	2	3	651	126	49	175	38	180	3
1	1	9	710	145	38	183	27	168	3
4	4	1	805	40	60	100	18	83	1
3	3	6	3136	86	49	135	32	131	1
4	1	14	3164	86	230	316	33	310	1
4	3	8	3192	57	43	100	43	108	1
1	1	12	3276	55	112	167	12	129	2
4	4	14	3437	68	316	384	20	326	2
3	3	5	3472	120	126	246	15	197	2
2	0	6	3542	71	235	306	37	312	1
1	3	8	3626	194	97	291	35	291	2
1	3	12	3647	132	89	221	5	155	2
0	0	5	3829	200	119	319	30	303	1
1	2	13	3913	186	52	238	25	214	1
4	3	6	3927	79	128	207	27	190	1
0	1	22	3941	139	143	282	22	245	2
4	1	14	3948	175	277	452	37	461	1
1	4	8	3983	89	200	289	33	283	1
3	4	14	4494	9	386	395	21	340	2
1	1	9	4620	451	48	499	28	464	1
4	4	9	4977	223	121	344	28	320	1
1	4	12	5152	253	52	305	34	302	1

2	4	18	5180	88	170	258	34	255	1
0	0	4	5635	197	124	321	33	315	1
2	4	5	5775	306	132	438	37	447	1
3	3	9	5817	96	108	204	29	192	1
4	3	12	5880	469	176	645	43	697	3
1	1	8	5922	260	144	404	24	360	1
1	1	5	6405	194	91	285	35	285	1
2	1	18	6699	182	126	308	35	308	1
2	1	10	6783	224	110	334	28	311	2
4	4	16	7252	116	170	286	27	263	1
3	1	12	7854	172	88	260	30	247	1
3	3	16	8050	302	145	447	52	523	2
3	7	13	8232	45	387	432	16	350	2
2	1	17	9051	146	112	258	40	271	1
4	5	26	9100	482	227	709	26	645	2
1	3	17	9135	137	119	256	34	253	2
4	4	24	9520	395	193	588	40	617	1
4	1	20	10577	304	78	382	39	397	1
2	4	15	11361	323	184	507	35	507	2
4	3	11	12824	131	180	311	51	361	1
2	3	24	13860	473	182	655	40	688	2
2	3	17	14434	221	121	342	35	342	1
4	4	12	14973	318	269	587	34	581	2
2	3	27	14987	173	332	505	19	424	1
1	4	39	19894	284	230	514	50	591	1
3	1	14	4277	148	324	472	39	491	1