Estimating Data Mining Techniques and Bayesian Mathematical Analysis Models a study

Gyanvi Verma
PhD Scholar, Computer Science, Email Id: - higyanvi@gmail.com

Dr. Manish Varshney
MUIT University, Lucknow (U.P.), Email Id: - itsmanishvarshney@gmail.com

Abstract

Software is a set of programs written in a programming language which does a specific and designated task (Somerville 2010). Computer software eventually was invented in 1940's. Computers needed a program to instruct it to do a specific task. For the same assembly language was developed which consists of small instructions written in codes. In 1950 a high level programming language which consists of English like sentences called FORTRAN was developed. FORTRAN stands for Formula Translation and it could perform scientific computations. 1960's saw the development of more programming languages like COBOL, Algol and PL/1. Along with it, operating systems to operate the computers like DOS and VMS. During development of these software the need for a proper method for software development was recognized. The first conference on software engineering was held in 1969. With the onset of a new stream of science called software engineering, many more software related to databases, structured programming languages and operating systems came into existence.

Keyword: - Instructions, Scientific, Development, Existence.

DOI Number: 10.14704/nq.2022.20.9.NQ44499 **Neuro Quantology 2022; 20(9):4371-4374**

Introduction

The question is why is software engineering needed? It was because software were getting more complex and large and were in demand in every field of science and engineering. It was seen that software projects needed a proper method to deal with time, efforts and cost estimation. A systematic approach was therefore needed for software development. Software engineering was that stream of science which dealt with specifying how to handle the software in every phase. The basic phases identified initially were analysis, design, implement and test/maintain (Pressman, R.S. 2005). Few models like waterfall model which suggested completing each phase and then going to next phase was very popular for many years. However, practically it was not very successful.

Practitioners came up with spiral model where each phase was iterated many times. Like while designing, it was required to go back to analysis check the requirements again and make amendments in design. Similarly while implementing, having a iterative approach of going back to analyzing requirements and design phase, ascertained that implementation is going in right directions. Software Engineering underwent many changes. Earlier testing was given minimum attention and in last few years testing became the most important phase.

Software Project Management is the intriguing blend of technical and human component. It has taken a sharp inclination towards 'how' the project has to be done rather than 'what'(Rose 2005). Software project management deals with planning and leading software projects. It can be said that were as software engineering is the science of planning and implementing the software projects, software project management is the art for monitoring and controlling the software project.

Review of Literature

There are many software development life cycle models which follow these steps in different ways. Mentioning here a few, Waterfall model has all phases cascaded to each other in which progress is seen as flowing steadily downwards. The next phase is started only after the defined sets of goals are achieved for previous phase is completed. In Iterative model, the phases are overlapped or iterated with each other. In this way enhancement of requirement can be done while designing and implementing. The spiral Model developed by Boehm was very successful. It is a risk-driven software development model. It keeps the various risks involved in the projects and adopts elements of incremental or iterative model.

Human component has become a very vital factor in project management, since it is the human who transforms process into product and it is their mind that gives shape to the product engineered (Lipke et al. 1992).

Companies aim to produce user friendly, reliable, innovative and maintainable software. Companies look into every aspect for software quality during development lifecycle to have a competitive edge over other similar products. Companies have to struggle in terms of cost effectiveness and time constraints. The project needs to be completed within time and cost limits. Software process is well defined and quantified with help of metrics starting from requirement phase to testing phase (Suma V. et al., 2008). Methods and tools are well equipped before and during software development. Software Engineering focuses upon generation of high quality product through well defined process. Software companies work meticulously for developing high quality software within scheduled time and estimated cost and resources. A systematic approach towards development of high quality software is followed due to the increased competitiveness in today's business world, technological advances, and increase in hardware complexity and frequently changing business requirements.

Data Mining

Data mining is a promising field for knowledge discovery. Data mining is the process of extracting knowledge from large or small data set(Han et al. 2006). It uses a combination of an explicit knowledge base, sophisticated analytical skills, domain knowledge to uncover hidden trends and patterns. These trends and patterns can be extracted on by using various data mining algorithms. The algorithm first analyzes a large set of data and finds specific patterns and displays it in a specific form as per the technique chosen..

Research Objectives

The objectives thus framed are as follows

Objective of this research work is to use a data mining framework for enhancing software development process by deploying right people based on historical data. The historical data taken from previous similar projects will reveal the capability patterns required in project personnel for the current project. Following five specific objectives are decided and achieved:

Objective – 1: To implement a data mining methodology for studying the patterns of performance of project personnel in a software company. This includes the study of the nature of data mining techniques used along with data specifications and empirical results obtained by the use of these techniques.

To achieve this objective existing data mining techniques proposed and implemented by various researchers are comprehensively studied and analyzed. A review is conducted which includes the analytical and empirical results of various data mining techniques. Studies helped in understanding the applicability, advantages and shortcomings of data mining methods for classification of project personnel based on their technical attributes. This also provided path for further research work.

Objective – 2: Human aspect needs deeper investigation before and during the project development since work behavior of project personnel has a direct impact on quality of the product. Software is a human intrinsic product and it is very crucial to look into capability factors of project personnel for fulfilling quality objectives. Thus, objective 2 is to analyze various technical attributes affecting the performance. This includes considering a large set of attributes and then

reducing the set using data mining techniques to bring out the dominant factors.

Objective – 3: Assessment of extensively used data mining techniques in knowledge discovery. To achieve this objective comparison was made by results of decision trees and bayesian mathematical model.

Objective – 4: To identify the impact of human aspect on turnover rates. This includes analyzing the root cause for performance gaps. To accomplish this objective, empirical study was carried out on various projects across five sampled software companies and a relationship was drawn between human aspect and profit.

Objective – 5: To suggest an integrated human capability and company work culture model for improving the quality of software developed.

Research Methodology

In order to achieve the above stated objectives, the instruments used in the study are:

(i) Literature survey

(ii) Conceptual framework development

(iii) Development of Computer Algorithms/ Programs

(iv) Interpretation of Results

(v) Publication of technical papers

(vi) Improvement by the way of feedback of survey and technical publications, models studied in the field.

(vii) Compilation of the final results, the theories developed/ evolved & final preparation of thesis.

Scope of Research Work

This research work focuses on the design and applicability of data mining framework for predicting the performance of project personnel in a software company.

Software industry spans over a wide domain and every domain requires a different skill set or expertise. Therefore its scope had be constrained. The constraint of this study is that it has dealt with short term projects of duration two to three years with windows operating system and C/C++ as programming language for web based applications. Secondly the data was collected only from software companies in Bangalore. Project personnel information is collected from those projects which are developed in similar domain using common technology and programming language

Data was collected from multiple sources like human resource department, project lead and project personnel. Various modes of data collection were used like brainstorming, discussions, interviews, and argumentation. The project leaders provided goals and experts in human resources suggested selection criteria.

Data involving several attributes was thus collected and used to predict the performance class. Attributes involving personal, educational and work related details were collected. However, this study was limited to only technical attributes and web based projects which are more of service than product. These projects need innovations for sustaining in the web world.

Significance of Research Work

The efficiency of the system rests on the quality of the person or project members. Without good person even the best of the system is bound to fail. With good project members even the defects of the project can be largely overcome. This study thus explores potential of a person in terms of ability, aptitude, personality, intelligence and interests which influence his performance and thereby is related to the quality of project.

This study gives an understanding of human aspect of software engineering specially the selection and development of project personnel along with applicability of data mining techniques for achieving the same.

Following are the specific significant benefits of this research work:

First, the proposed research provides a theoretical framework to the concept of data mining; applicability of data mining methods in predicting the performance of the employee at the very beginning of the project based on his technical attributes. This is helpful for designing appropriate research studies, to understand the objectives and make decisions based on these objectives for focusing on human aspect and eliminating one of the biggest software risk.

Second, it gives the root cause analysis for low performance of project personnel by identifying them through data mining techniques so that companies can fill the performance gaps. Apart from software process, this study will draw a relationship between managerial issues like work culture, pay, providing job satisfaction and human resource activities on quality of the product.

Third, this research work measures the efficiency of extensively used data mining techniques by comparing them on the basis of accuracy parameters. This is helpful in selecting best data mining techniques for the study.

Fourthly, based on the results, companies can derive rules for performance analysis. This is helpful for enhancing the quality of the software being developed which provides a competitive edge over other companies.

Fifth, this research work presents an integrated capability work culture model. This approach of integrating staff capability with satisfying work conditions has shown positive impact on software development process. This approach takes care of human component related to project personnel performance which is a major entity of the human aspect of software engineering. This integrated model is an empirical and experiential based development process that ensures project success.

These results are going to be of great help to both researchers and companies in enhancing the quality of the product developed and increasing the profitability of the company.

Research Questions

The research focuses on human aspect and it was based on the following research questions.

- Whether the company's strategy and mission constituted employee as competitive advantage?
- Whether the company followed any parametric method for selection and recruitment?
- Whether the managerial practices related to selection, development and retaining employee was a joint effort of management, human resources and project team or just a responsibility of human resources?
- Whether the company's' employees rated company as giving a good work condition and job satisfaction and does it have impact on performance ?

Conclusion

In today's global world, one of the issues that every sector must confront is the development of customer-satisfying software. As a result, the software industry is concentrating on human-based software since it relies on the human intellect to create new systems in order to compete with other businesses. The right project crew is at the foundation of the personality attribute (Tamara Sumner, 1995).

To have a good pool of talent, various firms use different tactics. The data mining process has obviously exceeded the project team's necessary points. It is critical that all businesses comprehend the vacancies, do a preliminary investigation of the causes of these vacancies, and implement plans to address these needs. Such tools that help to create and maintain a positive talent environment will boost the organization's productivity. It is now feasible to undertake a performance parametric analysis and look in depth at those elements that can increase employee performance and so contribute to the company's growth using data mining techniques.

References

- Sommerville.I. (2010). Software Engineering, 9th Edition, Pearson.
- Lock, Dennis. (2007). Project Management (9th Edition), Aldershot, Gower. Schwalbe, Kathy (2007) IT Project Management (5th edition) , London.
- Hughes Bob (2004), Project Management for IT-related projects, London, The British Computer society.
- Hughes Bob, (2002). Practical Software Measurements, London, Mc-graw hill. Boehm Barry W. (1989). Software Risk management, IEEE, Computer Society press
- Arnold, John, Cary Cooper, Ivan Robertson, (2004). Work Psychology: Understanding Human Behavior in the Work place (4th edition) , London, FT prentice Hall.