ISSN–2277- 8721

EIIRJ Electronic International Interdisciplinary Research Journal

Volume–XII, Issue– I (a) Jan – Feb 2023

OPEN ACCESS Original Research Article

ROLE OF MACHINE LEARNING TECHNIQUES IN COST PREDICTION OF AGILE PROJECTS

Dr. Manju Vyas

Associate Professor,

Jaipur Engineering College & Research Centre,

Jaipur (Raj.)

Abstract

In current scenario of software industry culture, an important and crucial task under project management is accurate estimation of practical measures like cost and effort which subsequently results in successful project completion. Many researchers have analysed and proposed various techniques in the estimation for software projects using conventional frameworks like waterfall, incremental etc. In recent years as there are technological advancements and there is a requirement of adaptation to technological changes, hence agile development methodology has attracted the interest of many researchers and software developers in software companies. Various researchers have proposed several techniques including opinion based, algorithm based and machine learning based techniques for effort and cost estimation of software projects. The proposed work in this paper deals with the study and analysis of the most popular techniques used in every category of estimation practices.

In current scenario of software industry culture, an important and crucial task under project management is accurate estimation of practical measures like cost and effort which subsequently results in successful project completion. Many researchers have analysed and proposed various techniques in the estimation for software projects using conventional frameworks like waterfall, incremental etc. In recent years as there are technological advancements and there is a requirement of adaptation to technological changes, hence agile development methodology has attracted the interest of many researchers and software developers in software companies. Various researchers have proposed several techniques including opinion based, algorithm based and machine learning based techniques for effort and cost estimation of software projects. The proposed work in this study deals with the study and analysis of the most popular techniques used in every category of estimation practices used in agile development. This paper describes a review of the research work conducted in the effort estimation of non-agile and agile projects in the previous years, which consists of different techniques and approaches. The first section describes an introduction of the relevance of estimation in project management, the second section covers a study of all those researches which have analysed the use of various machine learning based techniques for estimation of software projects and in the third section a comparative analysis is done based on the studied literature in terms of

ISSN–2277- 8721

EIIRJ Electronic International Interdisciplinary Research Journal

Volume–XII, Issue– I (a) Jan – Feb 2023

OPEN ACCESS Original Research Article

various factors like techniques used, estimation accuracy, and their applicability in various scenarios as part of summary of the chapter.

Introduction:

Effort Estimation of Software Projects:

Software project management is a key area in the field of computer science as software now-a-days impacts every area related to human life. Managing software means the process for development and the maintenance of software must be completely controlled using various pre-defined set of rules. As the software development process has to follow various parameters and a well-defined life cycle to ultimately deliver all the requirements gathered from the customers hence it has become quite time consuming and expensive process. It is also an evident fact that failure in software is caused mainly due to faulty practices used in project management. Using the right and optimised practices for software management helps both client as well as developers. Because of all the factors the need for highly reliable software is increasing. The reliability of software is mainly dependent on two factors: the selection of proper model for development and the estimation of various parameters. During the last few decades, the former area has been a research interest for many researchers resulting in development of many reliability models. Hence, currently parameter estimation is considered to be a primary activity in software reliability prediction and broadly the most important aspect of software project management. Software reliability models only become useful if they provide a correct and optimal estimation of various parameters.

As the agile process model allows the dynamic changes in the user requirements during the development cycle so this is the preferred process model these days. Initially, the estimation of size, cost, time and effort was done using an expert's opinion or using historical data like Planning Poker and Case Based Reasoning. But the dependence of these techniques on the historical data and in absence of previous data and expert's opinion these techniques were not useful. (Aljahdali, 2015)

In a nutshell, it is observed from all the above mentioned factors that lot of research work is required in the field of parameter estimation of projects specifically developed using Agile methodology. This paper indicates a comparative analysis of some researches where results were indicated by applying various techniques of Machine Learning for effort and cost estimation of the projects over a published dataset

Related Work - Machine Learning Techniques for Estimation:

(Wen *et al.*, 2012) published a SLR consisting of empirical studies on various models of machine learning published in 1991-2010. The paper analysed four key areas which are: Machine Learning technique used, estimation accuracy, analysis of various models and estimation context.

ISSN–2277- 8721

EIIRJ Electronic International Interdisciplinary Research Journal

Volume–XII, Issue– I (a) Jan – Feb 2023

OPEN ACCESS Original Research Article

Around 80 primary studies were explored and concluded that total eight Machine Learning techniques are generally applied in Software effort estimation and proved that the models based on Soft Computing gives better predication accuracy as compared to non ML models. Further, it mentioned that the accuracy of estimation varies when same ML model is applied over different datasets or different experimental designs. The SLR was conducted by developing a review protocol which included six stages: defining research questions, designing a strategy for searching, selecting the appropriate and relevant studies, assessing the quality of publications, data extractions and data synthesis. The following fig. 2.1 outlines these stage

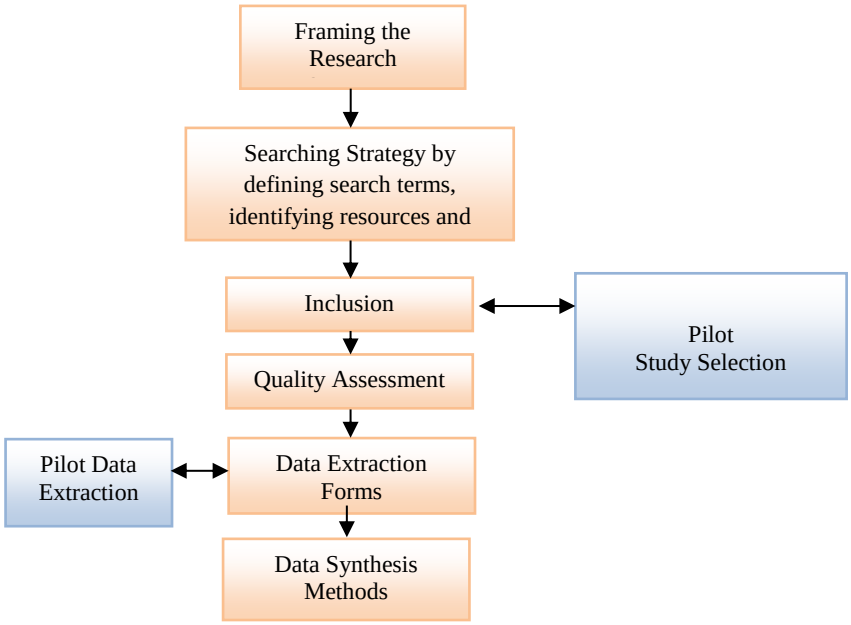

Figure 2.1: Stages of Conducting Systematic Literature Review

The review concluded that very less research has been done on the applications of Bayesian Network, SVR, Genetic Algorithms in the domain of software development effort estimation, hence these areas can be explored by the researchers. Further, it mentioned that the data sets available in this domain are very small in size, hence it suggested that researchers must share their proprietary datasets after removing the confidential information.

(Ziauddin *et al*., 2012) developed a mathematical model which is used to predict effort. The model was calibrated on a 21 row dataset corresponding to 21 projects.The values of size and complexity varies from 1 to 5. A very small and least complex story is indicated by the size value of 1 and a complexity value of 1. This type of story is direct requirement which does not require any research and which has very less number of variables. Thus it takes only few hours. These stories have localized effects. While a very large and an extremely complex story has size value 5 and complexity value 5. It basically requires an expert skill set. Such kind of stories have great impact on other stories.

ISSN–2277- 8721

EIIRJ Electronic International Interdisciplinary Research Journal

Volume–XII, Issue– I (a) Jan – Feb 2023

OPEN ACCESS Original Research Article

Thus the Size*Complexity value determines the value for each user story. The total number of story points are the total user stories of the project which is calculated by the sum of all values of user stories The calibration of velocity is done using factors like Friction and dynamic forces which calculates Deceleration. The effort of the project is then calculated

$$T = \frac{\sum_{i-1}^{n} ES_i}{C_i^D} \times {}^{1}/_{WD^{months}} \qquad 2.1$$

Then the accuracy was checked using the evaluation measures MMRE, PRED. The MMRE observed was 7.19% and Prediction accuracy observed was 57.14%.

(Satapathy *et al*., 2014) suggested Support Vector Machine based technique for increasing the effort prediction accuracy.. A comparison was made for these SVR-kernel results. The function of SVR is based on computation of linear regression. The effort is calculated using story point approach and different kernels of SVR are used for optimization of the results. The proposed approach used dataset given in Zia's work for evaluating the effort and validating the results. The following fig. 2.2 indicates the steps proposed for effort estimation using various SVR kernels.

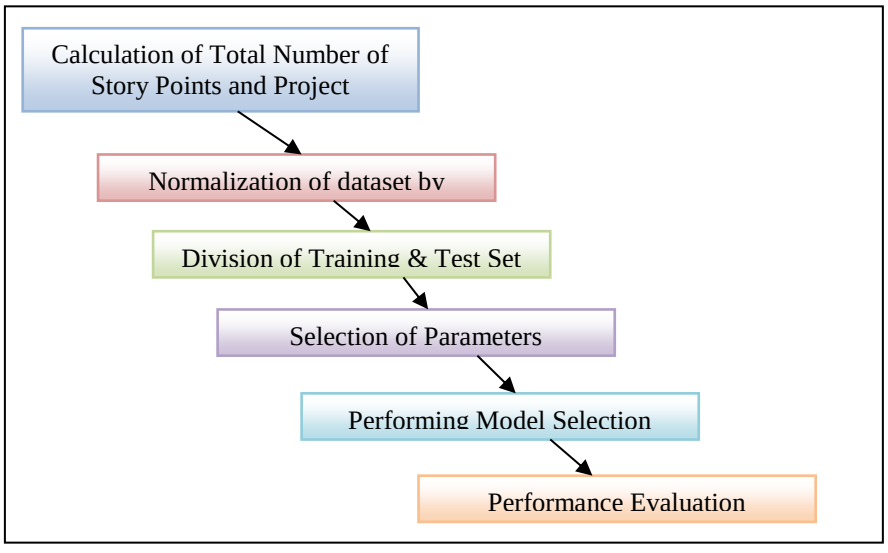

Figure 2.2: Effort Estimation Steps

(López-Martín, 2015) discussed the issues which comes when soft computing models are applied for prediction of the development effort. The accuracy of prediction was compared for various models based on neural network and statistical regression. The study used a published dataset containing function point as predictor variable depicting the size of project and actual effort. The results in the paper indicated that the accuracy of estimation for neural network models were better than that of regression based mathematical model.

(Khuat, *et al*, 2018) introduced a novel formula based on number of story points and team velocity. They applied an ensemble of Artificial Colony Bee (ABC) along with Particle Swarm optimization (PSO) to estimate the cost and effort parameters of

ISSN–2277- 8721

EIIRJ Electronic International Interdisciplinary Research Journal

Volume–XII, Issue– I (a) Jan – Feb 2023

OPEN ACCESS Original Research Article

projects developed using agile. Using experimental evaluation they compared the results of proposed algorithm with existing literature and proved that the proposed hybridization gave more accurate results.

(Zakrani, *et al.*, 2018) proposes a new and improved model for software project's effort estimation. The model applies support vector regression along with optimisation using Grid search. For applying Grid search, the parameters like type, kernel function, complexity values, kernel parameter and Epsilon values were tuned and then 5 fold cross validation was used. It also demonstrates that the result obtained on this paper outperforms the methods suggested in recent relevant literature in terms of various performance evaluation metrics.

(Malgonde, *et al.*, 2019) uses an ensemble based approach for effort estimation of software projects and performs experimental evaluation to evaluate the prediction accuracy of proposed approach and compare it with other published researches. An optimization model is then used to test the model and optimize the sprint planning. The model used five input variables which are independent: priority of particular the user story, number of subtask for that particular story, its size, its sprint duration and the total experience of the programmer. The target variable is effort. The experimental evaluation first calculated RMSE values for various algorithms like SVM, ANN, Ridge Regression etc. and then feature selection was applied to check the impact. For few algorithms the prediction accuracy improved like SVM, ANN while for decision tree, linear regression etc. it decreased. Then the weights were normalized to propose an ensemble technique and the results of proposed technique were compared with other popular ensemble technique.

(Arora *et al.*, 2020) in their systematic literature review studied various literature available related to machine learning estimation approached used in Scrum projects. The work farmed certain research questions, and after selecting various relevant studies related to each research question, summarized their observations and discussed the research gaps and scope for the researchers. The questions framed were which ML model is usually used for estimation of software projects developed using SCRUM, do ML models outperform non ML models, what is the estimation accuracy of ML models, what is the role of meta heuristic algorithms, are ensemble techniques better, what are the factors affecting estimation. The observations included that a wide variety of algorithms based on machine learning techniques are used for effort estimation of SCRUM projects like Bayesian Networks, Support Vector Machines, PSO, Neural Networks and many others and further it was quoted that the expert estimation models often suffers from individual bias, hence normally ML models have better prediction accuracy as compared to other non ML models. It was further concluded that after comparison of all the studies, the firework optimized neural network technique gave minimum error in prediction.

(Gultekin *et al*., 2020) used various ML algorithms to estimate the story point value. As compared to other published works, here it was proposed that the story point values, effort values and cost predictions must be done at each iteration so that deviation observed in the previous iteration may be taken as input for better accurate prediction in the next

ISSN–2277- 8721

EIIRJ Electronic International Interdisciplinary Research Journal

Volume–XII, Issue– I (a) Jan – Feb 2023

OPEN ACCESS Original Research Article

iteration. Various algorithms like SVR, Gradient Boosting, Random Forest etc. are applied to test the accuracy in prediction on various datasets. As the project phases are considered, hence the major advantage of the proposed work is the compensation of risk emerged in one phase into another phase. The results obtained were further compared with the already published results used in story point estimations.

Comparative Study & Summary of Observations:

Based on the studied literature, a comparative analysis is provided. Table 3.1 indicates all the studies using expert judgement, algorithm based techniques and machine learning based techniques for effort estimation of software projects developed using traditional or agile framework. The comparison is based on the estimation approach used, dependent and independent variable used, accuracy metrics and the characteristics of the dataset used. The findings of the study are provided in the last column. The findings shows the performance of the model in terms of the model which it has outperformed and the factors over which it has proved beneficial.

Table 2.3: Comparative Analysis of the Estimation Techniques Studied

Reference	Estimation Technique	Target Variable	Predictor Variables	Accuracy Metric	Dataset Used	Findings
Abrahamsson *et al.*, 2007	Regression, Neural Network	Development Effort	Weighted methods, object points	For global models: MRE(SR)=84% MRE(RBF)=63% For incremental models: MRE(SR)=57% MRE(RBF)=37%	2 commercial projects	The effort estimation by incremental models outperform the global models
Abrahamsson *et al.*, 2011	Expert Judgement, Regression, Neural Networks,	Effort	Character count, Keyword priority	---	1337 stories (2 projects)	The structured user stories improves the effort prediction
Ziauddin *et al.*, 2012	Statistical Model	Effort	No. of Story Points, Velocity	MMRE = 7.19%, PRED= 57.14%	21 Project Dataset	The experimental evaluation proved that model gave better prediction accuracy as compared to expert models.
Satapathy *et al.*, 2014	SVR linear kernel, polynomial kernel, RBF kernel, sigmoid Kernel	Effort	No of story points, team velocity	MMRE (linear) =0.1492 MMRE(polynomial)= 0.4350 MMRE(RBF)=0.0747 PRED(RBF)=95.90% MMRE(sigmoid)=0.1929	21 projects	The results suggested for the comparison of various SVR kernels methods and showed that SVR RBF kernel gave higher accuracy and less MMRE values as compared to other kernel.

ISSN–2277- 8721

EIIRJ Electronic International Interdisciplinary Research Journal

Volume–XII, Issue– I (a) Jan – Feb 2023

OPEN ACCESS Original Research Article

López-Martín, 2015	SLR, MLP, GRNN, RBFNN	Effort	Adjusted Function Points	MAR(SLR)=0.45 MAR(MLP)=0.32 MAR(GRNN)=0.3 MAR(RBFNN)=0.31	Datasets containing new and enhanced in enterprise environment	The results showed that RBFNN can be used for effort estimation of new projects developed using 3GL and training RBFNN is faster.
Khuat *et al.*, 2016	Cascade correlation neural network (CCNN)	Effort	Velocity, Number of story points	MMRE—0.1486	21 projects	It outperformed the models GRNN, PNN, GMDHPN
Zakrani *et al.*, 2018	SVR-RBF-GS	Effort	Number of story points, Velocity	MMRE-0.1640	21 projects	It outperformed the models SVM-RBF
Malgonde, *et al.*, 2019	Ensemble	Effort per story	Priority, no of subtasks, size, developed experience, sprint	MAE=8.167, RMSE=9.243	503 stories, 24 projects	Ensemble based technique improves prediction accuracy
Gultekin *et al.*,2020	Multi Layer Perceptron	Effort per story	No of issues as start, initial velocity, no of issues added, added velocity, to-do velocity	PRED=92.52% MAE=7.48	5 projects with 3233 issues	Machine learning technique based models performed better as compared to non-ML models.

References:

Abrahamsson, P., Moser, R., Pedrycz, W., Sillitti, A., & Succi, G. (2007, September). Effort prediction in iterative software development processes--Incremental versus global prediction models. In First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007) (pp. 344-353). IEEE.

Arora, M., Verma, S., & Chopra, S. (2020). A systematic literature review of machine learning estimation approaches in scrum projects. Cognitive Informatics and Soft Computing: Proceeding of CISC 2019, 573-586.

Dragicevic, S., Celar, S., & Turic, M. (2017). Bayesian network model for task effort estimation in agile software development. Journal of systems and software, 127, 109-119.

Gultekin, M., & Kalipsiz, O. (2020). Story point-based effort estimation model with machine learning techniques. International Journal of Software Engineering and Knowledge Engineering, 30(01), 43-66.

Jorgensen, M., & Shepperd, M. (2006). A systematic review of software development

ISSN–2277- 8721

EIIRJ Electronic International Interdisciplinary Research Journal

Volume–XII, Issue– I (a) Jan – Feb 2023

OPEN ACCESS Original Research Article

cost estimation studies. IEEE Transactions on software engineering, 33(1), 33-53.

Khuat, T. T., & Le, M. H. (2018). A novel hybrid ABC-PSO algorithm for effort estimation of software projects using agile methodologies. Journal of Intelligent Systems, 27(3), 489-506.

López-Martín, C., 2015. Predictive accuracy comparison between neural networks and statistical regression for development effort of software projects. Applied Soft Computing, 27, pp.434-449.

Malgonde, O. and Chari, K., 2019. An ensemble-based model for predicting agile software development effort. Empirical Software Engineering, 24, pp.1017-1055.

Satapathy, S. M., Acharya, B. P., & Rath, S. K. (2014). Class point approach for software effort estimation using stochastic gradient boosting technique. ACM SIGSOFT Software Engineering Notes, 39(3), 1-6.

Wen, J., Li, S., Lin, Z., Hu, Y., & Huang, C. (2012). Systematic literature review of machine learning based software development effort estimation models. Information and Software Technology, 54(1), 41-59.

Ziauddin, K. Z. K., Tipu, S. K., & Zia, S. (2012). An Intelligent Software Effort Estimation System. Journal of Expert Systems (JES), 1(4), 91-98.

Zakrani, A., Najm, A., & Marzak, A. (2018, October). Support vector regression based on grid-search method for agile software effort prediction. In 2018 IEEE 5th International Congress on Information Science and Technology (CiSt) (pp. 1-6). IEEE.

Cite This Article:

Vyas M., (2023). *Role of Machine Learning Techniques in Cost Prediction of Agile Projects* **,** ***Electronic International Interdisciplinary Research Journal****, XII, Issue – I (a), Jan-Feb, 2023, 128-135*