IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

Estimation of Effort Methods in Development of Software using Machine Learning Techniques

Monika Tiwari[1] and Dr. Kalpana Sharma[2]
Research Scholar, Department of CS, Bhagwant University, Ajmer, Rajasthan, India[1]
Assistant Prof., Department of CS, Bhagwant University, Ajmer, Rajasthan, India[2]

Abstract: *Software effort estimation accuracy is a critical factor for effective planning, control and delivering a successful software project within budget and time. Both overestimation and underestimation are major challenges to future software development, hence the continued need for accuracy in software effort estimation. The main objective of this study is to help researchers to know which machine learning techniques predict promising effort estimation accuracy in software development. In this paper, the performance of the machine learning ensemble technique is examined with a single technique based on two of the most commonly used accuracy evaluation metrics.*

Keywords: Accuracy, Metrics, Control, Software & Estimation etc

I. INTRODUCTION

Software size estimation is an important feature in determining the effort required to develop a software product. It is the method of estimating the most practical measure of effort (expressed as personal hours or capital) required to create or maintain development operations in the light of insufficient, questionable, and controversial data. Software effort estimation (SEE) is the process of predicting the most sensible use of effort required to develop or maintain software. SEE is the activity of estimating the total effort required to complete a software project [2]. Effectively estimating the effort required to develop a software product is of fundamental importance to maintaining competitiveness in the marketplace. Both under- and over-estimation indicate undesirable outcomes for the organizations. Underestimating can leadto huge overruns in budget and schedule, which can result in the cancellation of projects; In such a situation, all the hard work so far went in vain. Over-estimated projects cannot be subsidised; As a result, organizational efficiencies are getting hurt. The process of effort estimation needs to be optimized as proper estimation is required from both developer side as well as client side. On the developer side, estimates help plan development and monitor progress. On the client side, they are used to negotiate contracts, set completion dates, prototype release dates, etc. However, as indicated in research work reported by the Brazilian Ministry of Science and Technology-MCT, only 29% of organizations meet the size. estimation and 45.7% received software estimation. Therefore the research work on predictive estimation of proposed software has attracted the attention of many practitioners and theorists.

In 2013, the Standish Group Chaos Manifesto [3] stated that 43% of IT projects were delivered late, over budget, and/or with less than the required features and functions. This indicates that the role of project management is becoming increasingly recognized as a more important aspect of sustainability [4,5].

For accurate evaluation of web applications, datasets from previous web development projects are aggregated from the ISBSG [10] dataset. Similarly, in the case of agile projects, the Story Point Approach (SPA) is used to measure the effort required to implement a user story. Project velocity is obtained by adding up the estimates for user stories that are finished during the iteration (story point iteration). The efficiency of the models obtained using CPA, UCP, WEB and SPA can be improved by employing some intelligent techniques. The proposed research study considered the application of various Machine Learning (ML) techniques such as CPA, UCP, Decision Tree (DT), Stochastic Gradient Boosting (SGB), Random Forest (RF) and Support Vector Regression (SVR) kernel methods on the web. Went. Gone and SPA dataset to improve its prediction accuracy.

These datasets are selected on the basis of their content and its relevance so that inference process can be employed on those datasets. Class point dataset [14], UCP dataset is collected from 3 different sources which include datasets

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

from industries and some are available for academic research purpose. The entire web dataset is collected from the ISBSG repository and the SPA dataset [9]. A detailed description of these datasets is presented in the supporting chapters. The results of different models obtained after applying machine learning techniques are compared with each other as well as with the results availablein the literature to assess their performance.

Machine learning techniques used

The following machine learning techniques are applied to various datasets considered to calculate the effort of a software product. The decision to choose machine learning techniques for the purpose of implementation in the proposed research has been made on the basis of previous research study done in literature survey [12-15]. Many researchers have previously applied some of the following machine learning techniques for their research purposes. But none of these techniques were previously applied to inference using CP, UCP, WEB and SP datasets. Each proposed contribution also describes a detailed representation of the results obtained using these techniques for their respective datasets. Each contribution also shows a detailed comparison of these techniques with earlier results obtained from the literature to reach their performance.

II. LITERATURE REVIEW

Shweta. KR etal. (2022), machine learning is the latest buzzword playing an important role in various fields of medicine, research, and an industrial application. It is difficult to weigh the real values or valueof software. The best way to estimate software development cost, effort, size and time is based on past experience in software development. To measure the standard cost of software, as a unit of software value, machine-learning algorithms are used to enhance the level of end user satisfaction through accurateand quick calculation of software cost and effort estimation. In this research work, an innovative cost estimation for software project management was developed using an improved artificial neural network model. Two publicly available datasets are compared with different machine learning algorithms and the results show that the proposed model has high accuracy and low error rate in predicting the first stage of cost and effort evaluation.

Gautman et al (2021) stated that recently, there has been a growing frustration of programming projectdue to lack of system and financial planning limitations [2]. Deren et al (2020) applied expense evaluationto board development using an ANN model [3]. Fengwei Ning et al (2020) proposed a three-dimensional CNN for feasibility cost estimation [4]. Eric Mattel et al (2019) recommended that quotes allow project directors to assess the deliverability of activities and control costs effectively [5]. Mahmood et al (2019) builds a product cost evaluation model using an AI approach [6]. Michael et al (2018) applied neural convolution computation to cost evolution [7]. Przemys et al (2017) proposed various AI calculations for exertion and time evaluation [8]. TMS Elhag et al (1998) proposed ANNs for the development of programming projects [9]. Richa Yadav et al. (2016) examine that the achievement of a venture undertaking is further characterized by the developed amount and cost valuation strategy that deals with the ideal utilization of assets [10]. Murat Gunaydin et al (2004) investigate the utility of neural organization systems to beat cost assessment issues in the early stages of building configuration processes[11].

Ensemble Techniques and Accuracy Evaluation Metrics

The idea behind using EEE in SEE is that each estimation technique has its merits and drawbacks by integrating the techniques through EEE. We can remove the deficiencies, leading to more accurate estimates that can be obtained from a single model [12, 13].

There are two types of EEE methods [14]:

- Homogeneous EEE: Used to refer to an ensemble that aggregates a base model that contains no less than two different combinations of learning ensembles.
- Heterogeneous EEE: Used to refer to an ensemble that includes two diverse base models or more.

III. RESULT ANALYSIS

It should be noted that for the MMRE evaluation measure, a smaller value is better and for PRED (25) a larger value is better. The mean value of MMRE = 0.17 was smaller than the mean value of MMRE = 0.30 for a single technique on the NPD dataset. The Ensemble technique showed a 13% improvement compared to the Solo technique,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

indicating that the Ensemble technique was better in accuracy than the Solo technique in terms of the mean value of MMRE. Similarly, the mean value of MMRE = 0.21 for the ensemble technique was smaller than the mean value of MMRE = 0.31 for the single technique on the PD dataset and showed a 10% improvement over the single technique. In addition, the mean values of Ensemble_MMRE_PD (0.21) and Ensemble_MMRE_NPD (0.17) were smaller than the mean values of Solo_MMRE_NPD (0.30) and Solo_MMRE_PD (0.31), respectively. The results for the mean value of the MMRE evaluation measure indicated that the ensemble technique was better in accuracy on both the PD and NPD datasets than the single technique.

The mean value of an ensemble using MMRE is 0.16, smaller than the single MMRE = 0.36 on the PD dataset, indicating a 20% improvement over the single technique. The mean value of MMRE = 0.15 was smaller than MMRE = 0.30 for the Ensemble technique, showing a 15% improved accuracy over the single technique for the Ensemble technique on the NPD dataset. The results for the mean value of the MMRE evaluation measure indicated that the ensemble technique was better in accuracy on both the PD and NPD datasets than the single technique. The minimum value of MMRE = 0.10 and maximum of 0.47 obtained by an ensemble technique compared to the single technique (min = 0.12, max = 0.52) means better estimation with an ensemble technique and a similar combination of ensemble rules produces accuracy. The graphical representation of MMRE for an ensemble and single techniques on PD and NPD datasets is shown in Figure 1.1.

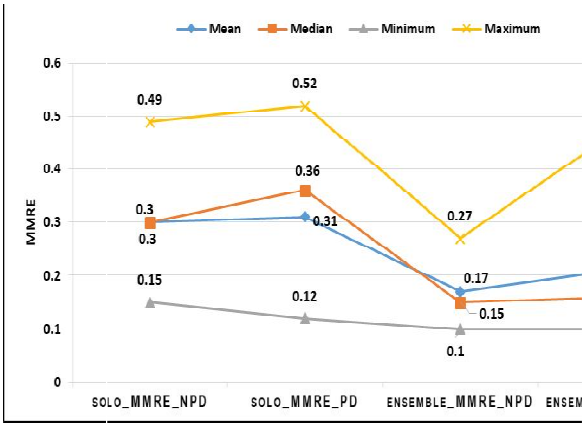

Figure 1.1: MMRE accuracy of an ensemble and single techniques on PD and NPD datasets

An integrated technique demonstrated 15.73% improvement over a single technique. The maximum valueachieved by the ensemble was 88.89% compared to a single value of 88.83% on the PD dataset. However,the maximum accuracy achieved by the single technique (max = 95.32%) on the NPD dataset was better than that of the ensemble technique (max = 84.00%). It should also be noted that the minimum values

obtained by the aggregated technique, 37.50% and 69.37%, respectively, on both the PD and NPD datasets were better in accuracy than the single technique. The graphical representation of PRED(25) foran ensemble and single techniques on PD and NPD datasets is shown in Figure 1.2.

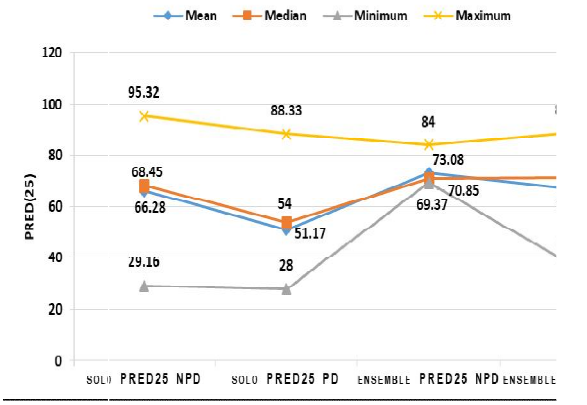

Figure 1.2: PRED(25) accuracy of an ensemble and single techniques on PD and NPD datasets

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

IV. CONCLUSION AND FUTURE WORK

We specifically focused on single techniques for machine learning-based ensemble and effort estimation. Our contribution to this research work is twofold. First, we explore the state of the art in the field of effortestimation using aggregation and singleton techniques. We obtained the primary study by following well- established systematic literature review (SLR) protocols prescribed for the software engineering domain. Second, we compared and evaluated both techniques by applying commonly used accuracy performance metrics (MMRE and PRED (25)) on PD and NPD datasets. Consolidation techniques outperformed single techniques when evaluated under MMRE and PRED (25) evaluation metrics. The ensemble technique was more accurate on both the PD and NPD datasets than the Solo technique. This is because every single estimation technique has merits and demerits that somehow lead to incorrect estimation results. Human expert inference can maximize the use of context-specific knowledge that cannot be accounted for by predictive algorithms, especially when the development team transitions to work with new emerging technologies or work in new application domains. Is. Improving the effort estimation accuracy prediction of the software development effort using a combination of estimation techniques, including expert estimation to build coherence models or frameworks, would be another research direction. In the future, we aim to utilize the knowledge of this investigation and will propose an aggregated model to improve the estimation accuracy prediction of software development efforts involved with algorithmic, expert inference and machine learning techniques

REFERENCES

[1]. Shweta.KR etal. (2022),” IMPROVED ARTIFICIAL NEURAL NETWORK MODEL FOR
[2]. SOFTWARE PROJECT COST ESTIMATION IN EARLIER STAGE”, International Journal of
[3]. Mechanical Engineering, ISSN: 0974-5823 Vol. 7 No. 1 January, 2022.
[4]. Gouthaman P, Suresh Sankaranarayanan, Prediction of Risk Percentage in Software Projects by Training Machine Learning Classifiers”, computers and electrical engineering (2021)
[5]. DareenRyied Al-Tawal, MazenArafah. GhalebJalilSweis, “A model utilizing the artificial neural network in cost estimation of construction projects in Jordan”, ECAM (2020)
[6]. Fangwei Ning, Yan Shia, MaolinCai, Weiqing Xu, Xianzhi Zhang, “Manufacturing cost estimation based on the machining process and deep learning method”, journal of manufacturing systems (2020)
[7]. Erik Matel, FaridaddinVahdatikhaki, SiavashHosseinyalamdary, Thijs Evers,HansVoordijk, “An artificial neural network approach for cost estimation of engineering services”, International journal of construction management (2019)
[8]. Mahmood Mohd Al Asheeri, Mustafa Hammad, “Machine Learning Models for Software Cost Estimation” International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT) IEEE (2019)
[9]. Michal Juszczyk ,AgnieszkaLesniak, Krzysztof Zima, “ANN Based Approach for Estimation of Construction Costs of Sports Fields”, wiley (2018)
[10]. Przemys_awPospieszny ,BeataCzarnacka-Chrobot , Andrzej Kobylin_ski, “An effective approach for software project effort and duration estimation with machine learning algorithms,The Journal of Systems & Software (2017)
[11]. T.M.S. Elhag, A.H. Boussabaine, “An Artificial Neural System for Cost Estimation of Construction Projects”, In: Hughes, W (Ed.), 14th Annual ARCOM Conference, 9-11 September Vol. 1, 219-26. (1998)
[12]. Richa Yadav, Monica Vyas, Vivekanada Vyas, Sanket Agrawal, “Cost Estimation Model (Cem) for Residential Building using Artificial Neural Network”, (IJERT) Vol.5(1) (2016)
[13]. H. Murat Gunaydin, S. ZeynepDogan, “A neural network approach for early cost estimation of structural systems of buildings”, International journal of project management (2004)
[14]. Idri, A, Amazal, F.A., Abran, A., (2015). “Analogy-based software development effort estimation: a systematic mapping and review.” Inf. Softw. Technol. 58, 206–230. doi: 10.1016/j.infsof.2014.07.013 .
[15]. Wen, J., Li, S., Lin, Z., Hu, Y., Huang, C., (2015). “Systematic literature review of machine learning based software development effort estimation models.” Inf. Softw. Technol. 54, 41–59. doi: 10.1016/j.infsof.2011.09.002 .

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

[16]. Elish, M.O., Helmy, T., Hussain, M.I., 2013. "Empirical study of homogeneous and heterogeneous ensemble models for software development effort estimation." Math. Probl. Eng 2013. doi: 10.1155/2013/312067.