International Journal of Global Tech Management (IJGTM)
Volume 1, Issue 2, July-December, 2024
Available online at: https://pgrpublication.com/index.php/ijgtm

Leveraging Historical Data from Project Control Systems for Accurate Cost Estimation: A Machine Learning Perspective

Vivek Singh[1], Neha Yadav[2]

[1]Software Test Engineer - Sr. Consultant, VISA, USA
[2]Independent Researcher, USA

Article history: Received: 15 August 2024, Accepted: 27 August 2024, Published online: 09 September 2024.

ABSTRACT

In this paper, historical data from a project control software is utilised for analysis of the possibility that machine learning models can improve the precision of cost estimation. Which individual and ensemble techniques are appropriate for predicting project cost as well as effort in software development is determined through this research, comparing individual and ensemble strategies.

Case studies and extensive reviews in the literature provide preliminary results showing that ensemble strategies significantly raise the precision of prediction. This advance provides more better resource allocation, planning of finance, and overall success for the project. This research study underlines that one needs to integrate advanced models of machine learning into project management procedures in order to improve on decisions and reduce risk.

Keywords-Machine Learning, Cost Estimation, Project Management, Ensemble Techniques, Software Effort Estimation, Predictive Analytics, Resource Allocation, Financial Planning.

INTRODUCTION

Precise cost estimation is crucial for effective project management, as its determination directly affects all parameters of compliance with the budget and project schedules. The paper will be focused on how some aspects of machine learning methods can be applied to improving the accuracy of cost estimations by using past data that can be extracted from a project control software.

For these purposes, the most efficient techniques for project cost and software effort prediction will be compared - solo techniques and ensemble techniques. The results highlight the impact that highly developed models of machine learning can have on decision-making and the allocation of resources, making a project more effective overall by beginning with readings methodically in literature and case studies.

LITERATURE REVIEW

Evaluating Machine Learning Techniques for Software Effort Estimation Accuracy

According to Mahmood et al.,2022., Where successful projects necessitate a need for accurate estimates of software effort, there is a need to keep projects within budget and to meet timelines. Improving the techniques of estimation can be paramount in that overestimation and underestimation of effort can provide grave problems.

It is within this view that this study compares ensemble methods with solo procedures with regard to their capability for assessing how different ML techniques will predict the accuracy of software effort (Mahmood et al., 2022).

Conducted a panoramic literature review of 28 papers. Of the 14 papers using solo and 14 papers using ensemble approaches to machine learning, we aimed to establish which of these best produced the highest estimation accuracy in accordance with the accepted metrics of Mean Magnitude of Relative Error (MMRE) and PRED (25).

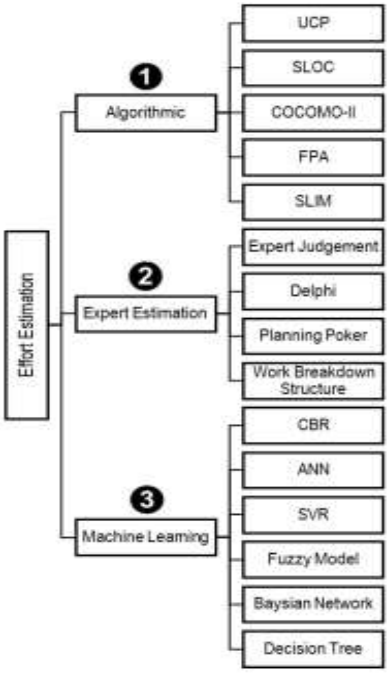

Figure 1: Categorization of effort estimation models

(**Source:** Mahmood et al.,2022)

These measures help in determining how closely projections reflect actual results. Finding relevant studies, conducting quality checks to ensure that the results deduced are valid, and gathering data for an analysis were all part of the review process. The general results of the findings indicated that ensemble methods tend to outperform solo methods since it combines various algorithms for making predictions.

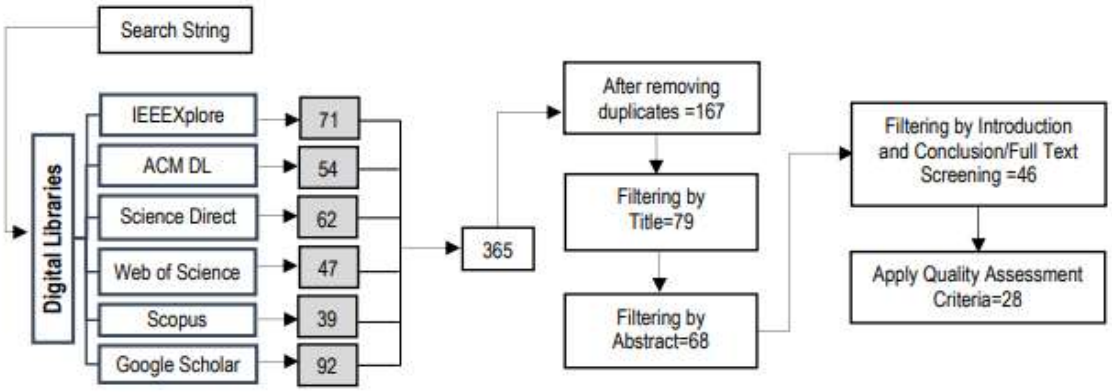

Figure 2: Search Strategy

(**Source:** Mahmood et al.,2022)

This is very essential as ensemble methods combine the best characteristics of several models into getting estimates that turn out to be accurate(Elmousalami, 2020). The results of the study present an increasing trend of using ensemble techniques with higher accuracy generally offered by them in software effort estimation. Such an understanding benefits practitioners and researchers in the domain of software development to choose the best estimating techniques. Findings from the study can guide future research and improve project planning procedures, which ultimately lead to an increased success rate of software projects with development in the software industry. However, groups may improve resource management significantly and control the risks that correspond to the underestimation of effort through applying such complex methods.

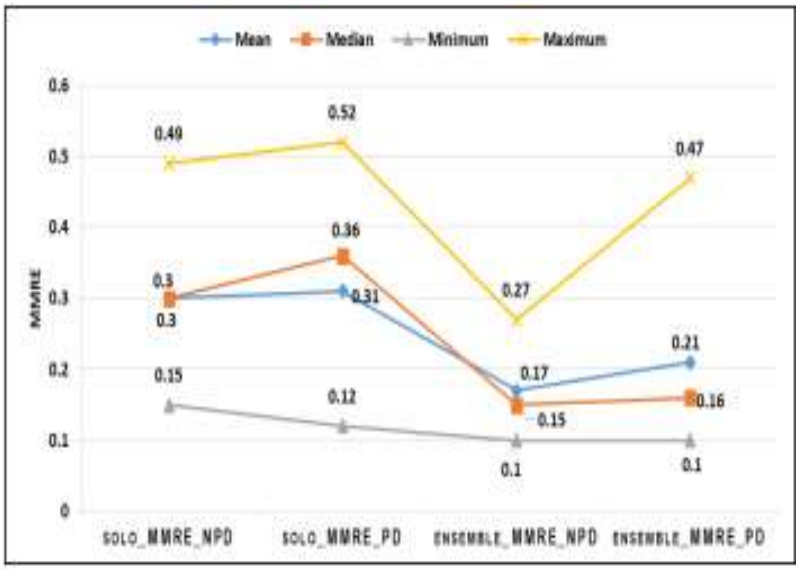

Figure 3: MMRE accuracy of an ensemble and solo techniques on PD and NPD datasets

(**Source:** Mahmood et al.,2022)

Evaluating AI Techniques for Accurate Project Cost Prediction

According to Elmousalami., 2020., during the design stage, for project managers and strategic decision-makers, it is very critical to approximate project costs. The inadequacy of traditional approaches, like statistical and probabilistic algorithms, is often rooted in insufficient and inconsistent data samples. In this work, an attempt shall be made to identify and discuss machine learning and AI approaches-a number of which encompass several models to increase the accuracy of these cost forecasts(Awada et al., 2021). The paper examines 20 AI techniques, which involve genetic fuzzy models, fuzzy logic, artificial neural networks, multiple regression analysis, case-based reasoning, hybrid approaches, and ensemble techniques like random forest and XGBoost.

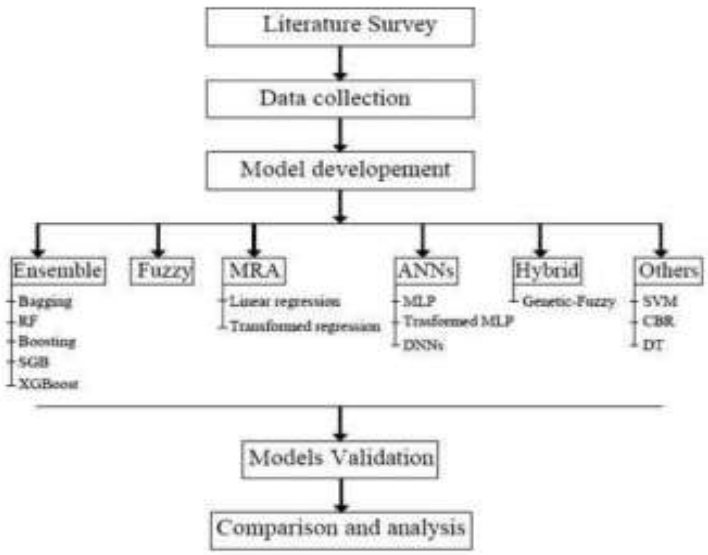

Figure 4: Research methodology

(**Source:** Elmousalami., 2020)

The paper reports the applicability of the models to estimate the costs in field canal improvement projects as a real-world case study. The results show that among the 20 approaches that have been tested, XGBoost emerges to be the one with superior performance in achieving an adjusted R^2 of 0.929 and mean absolute percentage error of 9.091%. These data propose that XGBoost appropriately comprehends cost prediction difficulties.

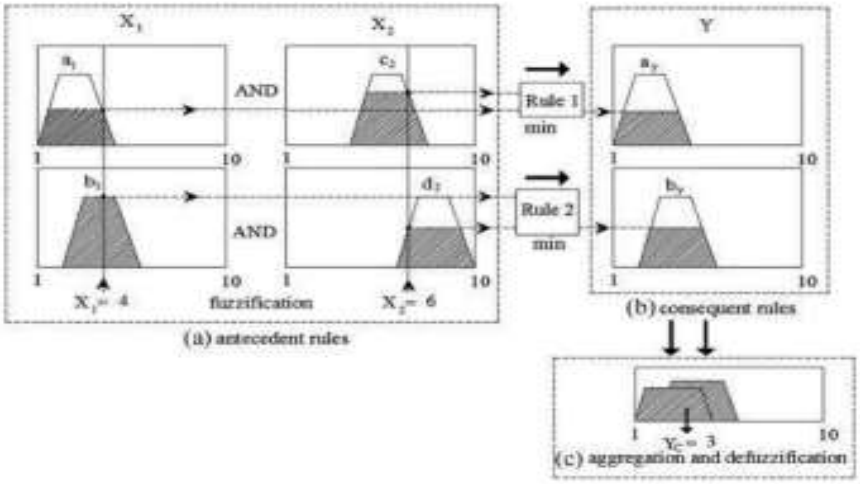

Figure 5: Fuzzy rules firing: (a) antecedent rules; (b) consequent rules; (c) aggregations; and fuzzifications.

(**Source:** Elmousalami., 2020)

This model can be used in practice because it holds strengths in the management of outliers, handling missing information, and nonlinear adaptation(Tahir et al., 2023). In addition to the above, the paper deals with relevant issues concerning model interpretation and uncertainty in order to ensure that users are able to understand and trust predictions produced by these AI models. This work is a significant contribution to this field with the availability of a dataset for FCIPs. Therefore, further analysis and validation of other models will be possible through other researchers working in the area. It enhances the general dependability of budgeting for project management but, at the same time guides practitioners on choosing the right AI algorithms for cost prediction.

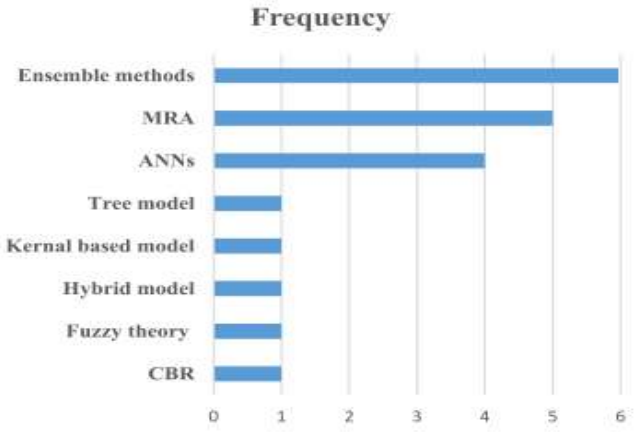

Figure 6: Frequency of each model category.

(**Source:** Elmousalami., 2020)

METHODS

Data collection and data processing

This system has deep fulfilment of the main need for data gathering and processing in order to provide consistent results for machine learning-the source of data forming the basis of historical records from project control software that captures the details on resource use, budget prediction, actual expenditures, and timeframes, which rely on queries and APIs for regular data extraction for proper coverage of a wide range of projects (Hong et al., 2020). The cleaning process begins during the processing stage; it encompasses data cleaning to remove and delete duplicates, fill in missing information, and address inconsistencies. The next step is feature engineering, where appropriate predictors are identified and the data is transformed through normalization and encoding for their use in machine learning. A way to construct and validate the model during the final stages is by splitting the dataset into training and testing sets. This process ensures that data quality.

Figure 7: Data collection and data processing

(**Source:** https://www.pecan.ai/)

Designing of Machine Learning Models

Different critical components are involved in the model design process in improving the cost estimation accuracy with machine learning. This means that the number of algorithms to be considered depends on the complexity of the dataset itself. This study involves several algorithms like neural networks, ensemble methods such as Random Forest and XGBoost, decision trees, regression approaches, and finally neural networks (Akinosho et al., 2020). All of these depend on some various connections in the feature in the data and nonlinear patterns. Another step that will prevent overfitting is feature selection, which makes use of techniques like tree-based significance or recursive feature elimination (Zhang et al., 2021). Hyperparameter adjustment promotes optimization of the model's performance, which consequently minimizes the error of its predictions. Last but not least, in the pursuit of effective cost estimating solutions, the model is trained and validated with metrics like Mean Absolute Error and R^2.

Implementation and Deployment

The first step in the implementation of the study on improving the accuracy of cost estimation through machine learning techniques would be to integrate the models with the already existing project control software using APIs, which enable smooth data flow. When the models are thoroughly tested-that is assured performance-they go into live production in a cloud-based environment to handle real-time project data. For this reason, project managers can then use user training in an effort to become better forecasters and apply insight gained to the decision-making process (Khalil et al., 2022). It uses a monitoring system that allows it to adapt based on suggestions. Regular updates and maintenance prevent loss of accuracy in models so that greater confidence and effective financial planning and management of projects can be achieved.

RESULT

Predictive Analytics in Sales and Demand

Predictive analytics on sales and demand details provide useful information regarding how things may play out in the future. Using past project data, the machine learning model will be better positioned to predict future demand/sales fluctuations (Giray, 2021). This tends to help in the least probable overestimation or underestimation of costs since the project manager is better positioned to make informed decisions in terms of budget and resource allocation. The entire process of cost estimation would become more effective through incorporation of predictive analytics, thereby making financial planning more market-oriented and eventually successful projects.

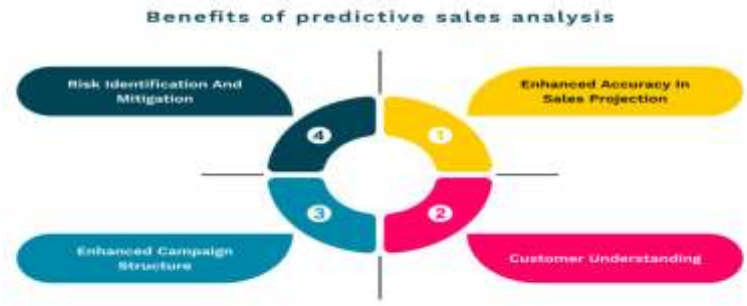

Figure 8: Predictive Analytics in Sales and Demand

(**Source:** https://nektar.ai/)

Innovation Strategies for Inventory Management and Replenishment

Innovative replenishment and inventory management systems incorporate the use of machine learning to optimize levels of stocks and reduce cost. Predictive analytics will enable businesses to estimate demand more accurately, thereby allowing for timely replenishment and minimizing surplus inventory(Hutchinson et al., 2021). The complexity in the process of inventory management can be reduced by using machine learning algorithms that evaluate past data and find trends for improving order quantities. This proactive approach leads to the enhancement of financial results and a better performance rate of projects due to lower holding costs and much more accurate cost projections.

Redesigning the Lines of Logistics and Supply

Re-evaluating supply and logistics chains emphasizes the requirement to leverage machine intelligence in a pursuit to improve efficiency. Organizations can become more responsive, reduce lead times, and make their supply chains efficient by implementing data-driven insights. The machines can explore past data to get proper estimation of the demand to manage it thereby minimizing disruption and influencing the inventory decision made. The strategic redesign would help improve logistical effectiveness, but it also allows for more accurate cost projections, hence better resource allocation and budget management-all of which contribute to project success and financial stability.

DISCUSSION

The conversation discusses how machine learning transforms the accuracy of cost estimates in project management. Models in machine learning could therefore find trends and produce a more reliable cost estimate by fitting past data gotten from the use of project control software. This reduces over-estimation and under-estimation chances. Predictive analytics does enable deeper decision-making insights and proactive reorientations of strategic action and resources on the part of the managers(Filippetto et al., 2021). The study also makes light of how important it is to continuously refine the model and quality of data in order to maintain accuracy. Taking everything together, these advancements underpin that machine learning can become quite important for improving project performance and finance management in not static settings.

Future Directions

Future revised models will include real-time analytics with more diverse input data that could provide increased potential in cost predictability. Indeed, taking factors from external sources into account, such as resource availability, economic data, or any shifts in the market, can allow for model forecasts to be more accurate (Lin et al., 2021). Further study of advanced techniques in machine learning, including reinforcement learning and deep learning, can increase power and flexibility even further. With the user-friendly interfaces, the project managers will be more likely to adopt them, and therefore, the insights should be useful. Finally, the creation of ongoing feedback loops for model training will support accurate maintenance over time, allowing companies to indeed succeed in adjusting to changing project landscapes.

CONCLUSION

In conclusion, Historical data supports ensemble machine learning algorithms, like XGBoost, which increase a project cost estimate with higher accuracy. All the advantages are related to better resource management and financial planning as well as decreasing project risk. Future work should be able to combine real-time data from this scenario with even more complex machine learning models than used here, like reinforcement learning. Indeed, more research into advanced machine learning techniques will continue to further enrich procedures in estimations with improvement sophistications to equip professionals with relevant resources for efficiently handling complicated conditions for projects.

REFERENCE LIST

Journal

[1]. Mahmood, Y., Kama, N., Azmi, A., Khan, A.S. and Ali, M., 2022. Software effort estimation accuracy prediction of machine learning techniques: A systematic performance evaluation. Software: Practice and experience, 52(1), pp.39-65.
[2]. Elmousalami, H.H., 2020. Comparison of artificial intelligence techniques for project conceptual cost prediction: A case study and comparative analysis. IEEE Transactions on Engineering Management, 68(1), pp.183-196.
[3]. Awada, M., Srour, F.J. and Srour, I.M., 2021. Data-driven machine learning approach to integrate field submittals in project scheduling. Journal of Management in Engineering, 37(1), p.04020104.
[4]. Preet Khandelwal, Surya Prakash Ahirwar, Amit Bhardwaj, Image Processing Based Quality Analyzer and Controller, International Journal of Enhanced Research in Science Technology & Engineering, Volume 2, Issue 7, 2013.

[5]. Bhardwaj, Amit. "Literature Review of Economic Load Dispatch Problem in Electrical Power System using Modern Soft Computing," International Conference on Advance Studies in Engineering and Sciences, (ICASES-17), ISBN: 978-93-86171-83-2, SSSUTMS, Bhopal, December 2017.
[6]. Tahir, T., Gencel, Ç., Rasool, G., Tariq, U., Rasheed, J., Yeo, S.F. and Cevik, T., 2023. Early software defects density prediction: training the international software benchmarking cross projects data using supervised learning. IEEE Access.
[7]. Hong, T., Wang, Z., Luo, X. and Zhang, W., 2020. State-of-the-art on research and applications of machine learning in the building life cycle. Energy and Buildings, 212, p.109831.
[8]. Vivek Singh, Neha Yadav. (2023). Optimizing Resource Allocation in Containerized Environments with AI-driven Performance Engineering. International Journal of Research Radicals in Multidisciplinary Fields, ISSN: 2960-043X, 2(2), 58–69. Retrieved from https://www.researchradicals.com/index.php/rr/article/view/83
[9]. Vivek Singh, Neha Yadav,"Deep Learning Techniques for Predicting System Performance Degradation and Proactive Mitigation" (2024). International Journal of Open Publication and Exploration, ISSN: 3006-2853, 12(1), 14-21. https://ijope.com/index.php/home/article/view/136
[10]. Zhang, L., Wen, J., Li, Y., Chen, J., Ye, Y., Fu, Y. and Livingood, W., 2021. A review of machine learning in building load prediction. Applied Energy, 285, p.116452.
[11]. Akinosho, T.D., Oyedele, L.O., Bilal, M., Ajayi, A.O., Delgado, M.D., Akinade, O.O. and Ahmed, A.A., 2020. Deep learning in the construction industry: A review of present status and future innovations. Journal of Building Engineering, 32, p.101827.
[12]. Patel, M., Parikh, H., & Dave, G. (2023). Chitosan flakes-mediated diatom harvesting from natural water sources. Water Science & Technology, 87(7), 1732-1746.
[13]. Parikh, H., Patel, M., Patel, H., & Dave, G. (2023). Assessing diatom distribution in Cambay Basin, Western Arabian Sea: impacts of oil spillage and chemical variables. Environmental Monitoring and Assessment, 195(8), 993
[14]. Khalil, M., McGough, A.S., Pourmirza, Z., Pazhoohesh, M. and Walker, S., 2022. Machine Learning, Deep Learning and Statistical Analysis for forecasting building energy consumption—A systematic review. Engineering Applications of Artificial Intelligence, 115, p.105287.
[15]. Giray, G., 2021. A software engineering perspective on engineering machine learning systems: State of the art and challenges. Journal of Systems and Software, 180, p.111031.
[16]. Navpreet Singh Tung, Amit Bhardwaj, Ashutosh Bhadoria, Kiranpreet Kaur, Simmi Bhadauria, Dynamic programming model based on cost minimization algorithms for thermal generating units, International Journal of Enhanced Research in Science Technology & Engineering, Volume 1, Issue 3, ISSN: 2319-7463, 2012.
[17]. Er Amit Bhardwaj, Amardeep Singh Virdi, RK Sharma, Installation of Automatically Controlled Compensation Banks, International Journal of Enhanced Research in Science Technology & Engineering, 2013.
[18]. Hutchinson, B., Smart, A., Hanna, A., Denton, E., Greer, C., Kjartansson, O., Barnes, P. and Mitchell, M., 2021, March. Towards accountability for machine learning datasets: Practices from software engineering and infrastructure. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 560-575).
[19]. Filippetto, A.S., Lima, R. and Barbosa, J.L.V., 2021. A risk prediction model for software project management based on similarity analysis of context histories. Information and Software Technology, 131, p.106497.
[20]. Lin, S.S., Shen, S.L., Zhou, A. and Xu, Y.S., 2021. Risk assessment and management of excavation system based on fuzzy set theory and machine learning methods. Automation in Construction, 122, p.103490.