

The Nexus between the Machine Learning Techniques and Software Project Estimation

Md. Shelim Miah¹, Mahesh Babu Pasupuleti², Harshini Priya Adusumalli³

¹Assistant Professor, Department of Business Administration, Asian University of Bangladesh, Dhaka, **BANGLADESH**

²Data Analyst, Department of IT, iMinds Technology Systems, Inc., Pittsburgh, PA 15243, **USA**

³Software Developer, CGI, 611 William Penn Pl # 1200, Pittsburgh, PA, **USA**

ABSTRACT

Machine Learning is an application of artificial intelligence that allows computers to learn and develop without explicit programming. In other words, the goal of ML is to let computers learn on their own without human involvement and then alter their activities. ML also allows huge data processing. Project management planning and evaluation are vital in project execution. Project management is difficult without a realistic and logical plan. We give a complete overview of works on Machine Learning in Software Project Management. The first category contains software project management research articles. The third category includes research on the phases and tests that are the parameters used in machine-learning management and the final classes of the results from the study, contribution of studies in production, and promotion of machine-learning project prediction. Our contribution also provides a broader viewpoint and context for future project risk management efforts. In conclusion, machine learning is more successful in reducing project failure probabilities, increasing output ratio for growth, and facilitating analysis on software fault prediction based on accuracy.

Key Words: Software Project, Software Estimation, Software Project Management, Machine Learning Technique

INTRODUCTION

When project requirements are evaluated, it can result in increased project costs and time expenditures, unfulfilled goals, or even project cancellations. This is a natural and unwelcome project danger that can have detrimental impacts on the reliability of software projects. While a software development project is underway, the need to change specifications (in terms of multiple expansion, elimination, and modification) is one of the most significant elements causing problems for the project (Pasupuleti, 2018). It comprises an explanatory examination of the fundamentals of software project assessment and computer training technology. This article is divided into two parts. This page defines the approach, including the source of information, the prerequisites for research eligibility, the Systematic Literature Review (SLR), and the implications of search results from published articles on the outcome of the study. As well as this, it identifies the research questions (RQs) for this study, with threats to validity pointing out significant hurdles to the effectiveness of SLRs. Using an ML technique, the questions of any object from three website papers were divided into four classes, which corresponded to the literary taxonomy on software project management used in this study. A modern approach to risk management in software projects was discussed in the study, as were the sources of inspiration, obstacles, and recommendations in that area of research.

When it comes to good Software Project Management, the ability to accurately forecast software development effort and duration is essential (SPM). Accuracy and dependability of prediction mechanisms are also critical considerations. Being able to make accurate effort estimates, especially during the early stages of a software project, can help to greatly reduce the high risks that are incurred throughout the development of a software product (Pasupuleti, 2020). Unfortunately, most of the existing estimation approaches are frequently inaccurate, resulting in the majority of projects experiencing significant effort overruns. It has been discovered, however, that software project estimating based on machine learning algorithms can deliver more accurate work estimation results.

THREATS TO VALIDITY

A number of other studies (Adusumalli & Pasupuleti, 2017; Rahman et al., 2019; Pasupuleti et al., 2019) have identified substantial difficulties to the effectiveness of SLRs, as well as emerging trends in the use of machine learning algorithms, benchmark data-sets, validation methodologies, and size measures for software effort estimation. Four alternative tactics were employed in order to reduce the hazards posed by these TTVs from a strategic standpoint.

First and foremost, construct validity was established by executing a manual and automated sentence search in order to reduce the estimated SPM data from the data collection process to a bare minimum. Aside from that, the papers chosen for this study analyze the SPM by thoroughly studying the shortened search results. Second, internal validity was investigated using the methods described by Pasupuleti & Amin, 2018; Pasupuleti, 2017; Fadziso et al., 2018). These methods were employed to resolve the internal validity problem. Additionally, in order to prevent biases during the exhaustive search for journal articles, a technique that combines two phases of the search was utilized for a comprehensive selection approach, which was designed to be as extensive as possible. All articles of interest were taken from databases that had previously been utilized for relevant research (Pasupuleti, 2016; Pasupuleti & Adusumalli, 2018), and they were then subjected to rigorous selection methods.

Third, external validity was addressed by merging SPM investigations throughout a ten-year period, which resulted in results that were more generalizable than the original findings. There is a parallel relationship between the cumulative collection of papers and the number of papers that are currently available, which suggests that this SLR can maintain a generalized report that is consistent with the external validity criteria of the research while also remaining cost-effective. Finally, the conclusion validity was handled using SLR methods and guidelines that have been applied by researchers from reputable publications, such as (Maddinget al., 2020), which makes it possible to reproduce the research chronologies of this SLR with quantifiable and identical results, thereby demonstrating the validity of the conclusion.

REVIEW AND SURVEY ARTICLES

The study and research documents provided an overview of the most recent perceptions of machine learning technologies in SPM creation and evaluation, as well as the implementation of machine learning algorithms. This section explains and makes use of the alternative ways. These articles have been organized into distinct subjects and implementations; the works described below have been organized into a specific logical model group.

Author Donepudi et al. (2020b) employed a financial return on investment (ROI) network infrastructure evaluation to determine the viability of his or her network infrastructure

investment. In most academic institutions, purchasing or selling inventories is difficult because the commodity does not provide a "sale" advantage. This is especially true when employing a return on investment (ROI) financial theory. The study Adusumalli (2018) conducts a thorough examination of contemporary program maintenance studies. The findings of the study revealed that the use of ML algorithms in maintenance forecasting has increased significantly since 2005. The difficulties were classified into sectors of competence based on the Project Management Body of Knowledge (PMBOK), and they were also examined (Khan et al., 2020). Artificial intelligence tools and hurdles in agile project management were used to investigate the issues. According to the contribution, modern project management models and information technology techniques that integrate machine learning-based methods with the treatment of inaccuracy, vagueness, or ambiguity will be required in the near future, with crucial performance indicators being correlated with critical areas of expertise.

This section discusses the various alternative ways that have been employed. These articles were sorted into several categories based on their subjects and applications. This category contains three different articles.

The initial model, which was developed for product lifecycle management, was a success (PLM). PLM was presented by the authors of Adusumalli (2017b). The approach creates a layer of functionality that will allow for the next iteration of PLM to be built on top of an existing PLM network. The Ford Powertrain case study will be used to demonstrate how this new PLM will be incorporated into a digital plant automation ecosystem.

The second model, which was tested on a recommendation algorithm, was successful. Donepudi et al. (2020c) proposes a novel software algorithm for solving problems. First, implement a bug-based feature and a special screening mechanism to validate the applicant fixer. Then, using a variety of comments and promises, construct a network of multi-developer commits and rank them in order to select the most acceptable bug fixer. The outcome indicates that the solution was successful in completing the error triage process.

The third model is the Logic Model. Using a real-world case study in a distributed domain, two articles (Adusumalli, 2017a; Adusumalli, 2019) applied agile testing to a selected team and compared their results with those of three other groups in order to determine the impact of including the customer as part of the testing process in order to overcome distributed development difficulties. However, the group that used agile testing was able to verify more than 99 percent of all requests that were submitted during the testing process, which was a significant difference that contributed to the overall productivity of the project.

DISCUSSION

The benefits of utilizing the ML platform for project management are clear and convincing to all parties involved. This section highlights some of the advantages in literature, which are organized into categories based on their unique characteristics. The appropriate sources are given in order to facilitate a more in-depth discussion (see Figure 1).

Prediction Cost Evaluation Model Advantages

The method was utilized to improve the fault-predicting model's performance. The predictive fault model is tested using cost estimation. Preliminary findings: Compared to previous methodologies, the proposed device source metrics are beneficial in software projects with percentages of error classes below the acceptable threshold value.

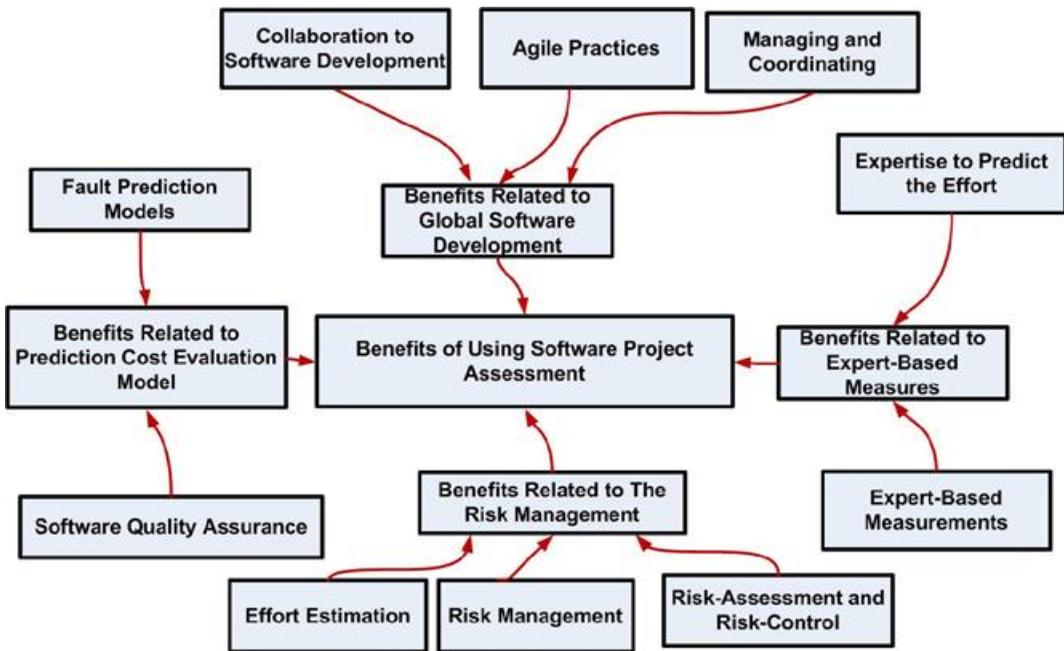


Figure 1: The Advantages of Using a Software Project Assessment

Benefits of Risk Management

Software planning tasks can be separated into two categories: engagement appraisal and risk reduction. Software effort estimation is influenced by various factors, and software risk management needs early discovery, treatment, and eradication. Efforts are determined by software growth commitment. The key activity in the project planning phase is risk assessment. However, traditional risk-reduction strategies rely on human judgment and expertise, and risk assessments are redundant and costly for software projects.

Global Software Development Benefits (GSD)

The rising use of GSD to reduce manufacturing costs and access a wide range of professional analysis is another significant market advancement. GSD initiatives are growing more common, but they are difficult. These include issues with participant connectivity, developing acceptable community contacts, cultural differences, and managing and organizing work in implementation projects. In short, distributed software development requires significant teamwork. Some agile methodologies are difficult to adopt because they rely on face-to-face communication, which is difficult in GSD environments.

Expert-Based Measures Benefits

The predictor uses experience to foretell the work of experts. The estimator's skill depends on the issue and his prior experience. The centered model would gain greatly if the expert measurements were deleted. The position of expert behavior is not to be ignored. The evaluation model research poses numerous limitations in collecting the prediction measures for possible work. The lack of particular methods to measure behavior, consistency, and adherence to a system or measuring accuracy may be a disadvantage.

RECOMMENDATIONS FOR EXPERT BASED MEASURES

These skills are based on project groups' understanding in eleven areas, including pricing, time, distance, productivity, and resource management (often contributes to over-optimistic estimates). That which creates and gathers the knowledge required to perform a project. Micro and macro knowledge levels were built.

Stakeholder profiles were created for each market topic based on the expertise of each stakeholder. Consequently, rather than seeing each effort as equal, investigate answers in a larger spectrum of initiatives. Rahman et al. (2020) defines approaches based on the content of objects assigned to each topic. The distribution of tiny series values for conceptually comparable items is required in some cases. Technically, shipping and monitoring, for example, were unrelated. By forming communities of like stakeholders, cooperation filtering can assess whether the stakeholder will inform on the targeted issues.

Using the optimum software cycle model for the given project based on available data science proof is recommended early in the implementation process. The project management framework challenges the community with the protocol structure advice. Second, it compares alternative classification and selection algorithms, with a proposal model based on historical software creation statistics to estimate a modern software project process model. This information helps project managers choose the appropriate software system for the present project at a critical moment in the creation phase, examines the reciprocal effect between phase models and various forms of factor project, and utilizes the best strategy. An automatic recommendation framework based on the ML software process model helps project managers determine which software process model to utilize in the early stages of a new project.

Personalized suggestions provide an interface for data research teams to efficiently engage in ML projects. One-second responses to queries and new knowledge deliver millions of applicants personalized feedback. So it operates on Antelope's open-source implementation, a flexible information engineering platform. It may also employ a variety of ML tools and technology developed here, some of which only connect with standard data management systems.

RECOMMENDATIONS FOR MANAGEMENT SOFTWARE PROCESS

Because Agile prioritizes features, it reduces the impact of thin consistency. Compared to smaller projects, larger projects have output issues, even with fair overall system requirements. This could apply to agile civilizations vs non-agile crops causing inter-level conflict. The Agile approach may better identify a disagreement to deliver a short-term fix with short-term improvements and a long-term value conflict resolution plan. Cultures are now merging and intertwining, as are religious systems. Agile Software Value Conflicts Agile systems self-organize. There is evidence that an organization is important in the development of agility. If an agile project does not alter or adjust its project environment, the literature shows minimal support for doing so. No advice is given on how to monitor and manage disagreements. By Donepudi et al. (2020a), this technique recognizes issues and remedies. My experience is that of an Agile software development team that works without a project and may not have a meaningful impact. People and institutions' ideals are part of the environment, thus agile notions must be examined.

Automated estimation Agile uses the newest ML techniques to apply automatic card estimate to historical anticipated human data. "Auto Estimate" improves the most used manual poker preparation method in agile organizations. Maximize estimation precision by limiting the

impact of inaccurate estimates; show that auto estimation boosts poker preparation later in the project; and quantify the benefit of producing well designed tale cards.

Intuitive project control is tied to mathematical and linguistic data management. It also looks into new ways of managing businesses and open access applications for computer intelligence in recent decades. It also evaluates niche areas with high thematic applicability. By the key metrics of success, the input relates to the predicted demand for current project management and IT resources that include ML-based frameworks and information imprecision care, vagueness, or ambiguity. Modern learning assessment libraries and open-source project management frameworks open up a new area of study related to technical convergence of IT resources.

In iterative software development, SRP is an essential phase. SRP contains a delivery, like compounding, of soft restrictions, such time, resources, price, or money. The SRP-Plugin shows that applications used with the common application help boost development productivity. The plugin has a comprehensive visual space ecosystem with sophisticated release preparation features, boosting launch readiness, viability, and stakeholder collaboration. SRP-Plugin enriches the Visual Lab with a robust, detailed, and systematic approach. However, preparing release schedules is only the first step in bigger principles that guide strategic release planning decisions.

RECOMMENDATIONS FOR SOFTWARE FAULT PREDICTION MODELS

Active SPM must anticipate software development effort and duration. Prediction accuracy and reliability are also crucial. To anticipate software duration, ML techniques were deployed. These methods require a forecasting component that predicts the possible project commitment and period based on current/past data. Despite this, many ML algorithms go untested. Various ML techniques are employed to build machine models based on project details. The GRNN model predicts tech workers' efficiency. GRNN could be used to estimate programmers' productivity for new and updated lines of code. When predicting the productivity of individual software practitioners, GRNN outperforms statistical regression when using experiences as independent factors. Prediction employing effective data mining techniques measures project success/failure phase-by-phase instead of the standard assessment style of the entire project aspect. Data mining gathers data from numerous initiatives across multiple computing areas to cascade clustering and grouping processes. They also suggested that failed projects were becoming more successful. Foreseeing the severity of a program error was introduced by Nagwani and Bhansali. In order to establish a software bug repository severity cluster, a problem fixing length was used. The suggested approach is implemented using open-source code from MySQL's open-source issue repository.

The data mining categorization approach predicts cost slippage using the baseline budget and timeline for an ICT project. There are three types of falls: natural, medium, and high. The purpose is to explain how a data mining classification model is constructed to estimate cost losses. The suggested model splits a project into three categories based on input (for example, initial budget and timetable) (regular, medium, and large).

An experiment in computational biology with a pattern finding strategy is implemented. The technology reveals inherent connection tendencies, allowing business practitioners to gain useful knowledge and boost decision-making trust. The data sets evaluated produced statistically significant patterns with good grades. In addition, it displays the impact of various budgeting approaches. The first study employs unique patterns mining to identify

incorrect software in software engineering. A positive ranking performance and useful knowledge for decision-makers were found by the research.

Fault-finding models developing a model of defect prediction for a large industrial software firm is beneficial. Even though 4% of the program is flawed, the purpose is to incorporate a malfunctioning factor into a big program project's continuous development process and lessen the evaluation initiative. The latest strategy comes from experience.

CONCLUSIONS

An extensive study in software project management on ML approaches was concluded. Jobs have spread steadily over time. Automatic effort estimation relies on ANN, Fuzzy Logic, Genetic, and Regression Algorithms. Precise effort calculation is a major software development practice. Time and difficulties influenced the software. Various ML works in software project management have basic themes. These ventures are divided into four categories: reviews and surveys of software project management; case studies of software project management methods; experimental publications used in the management of ML, a type of structure or architectural model; and an analysis of a project, form or structure. These articles can help software project managers describe and explain the hazards, benefits, and recommendations of ML techniques. However, due to the sheer number of ML methods, many remain unstudied. Based on the literature, the motivations for utilizing automated SPM, project preparation assessment issues, and ML engineering technologies are explored. Although SPM literature discusses project performance and loss, there is longstanding debate over whether project progress should be measured. These instructions will answer the challenges facing a software project in ML approaches and open up work opportunities in this industry. The calculation of effort using ML techniques for risk assessment is still being researched. To reduce the problem, create districts with the same stakeholders and anticipate whether a stakeholder is aware of the issue. This review offered tentative answers to important problems on ML-based Software Project Management Estimation.

REFERENCES

- Adusumalli, H. P. (2017a). Mobile Application Development through Design-based Investigation. *International Journal of Reciprocal Symmetry and Physical Sciences*, 4, 14–19. Retrieved from <https://upright.pub/index.php/ijrps/article/view/58>
- Adusumalli, H. P. (2017b). Software Application Development to Backing the Legitimacy of Digital Annals: Use of the Diplomatic Archives. *ABC Journal of Advanced Research*, 6(2), 121-126. <https://doi.org/10.18034/abcjar.v6i2.618>
- Adusumalli, H. P. (2018). Digitization in Agriculture: A Timely Challenge for Ecological Perspectives. *Asia Pacific Journal of Energy and Environment*, 5(2), 97-102. <https://doi.org/10.18034/apjee.v5i2.619>
- Adusumalli, H. P. (2019). Expansion of Machine Learning Employment in Engineering Learning: A Review of Selected Literature. *International Journal of Reciprocal Symmetry and Physical Sciences*, 6, 15–19. Retrieved from <https://upright.pub/index.php/ijrps/article/view/65>
- Adusumalli, H. P., & Pasupuleti, M. B. (2017). Applications and Practices of Big Data for Development. *Asian Business Review*, 7(3), 111-116. <https://doi.org/10.18034/abr.v7i3.597>
- Donepudi, P. K., Ahmed, A. A. A., & Saha, S. (2020a). Emerging Market Economy (EME) and Artificial Intelligence (AI): Consequences for the Future of Jobs. *PalArch's Journal of Archaeology of Egypt/ Egyptology*, 17(6), 5562-5574. <https://doi.org/10.5281/zenodo.5562662>

- Donepudi, P. K., Ahmed, A. A. A., Hossain, M. A., & Maria, P. (2020b). Perceptions of RAI A Introduction by Employees on Employability and Work Satisfaction in the Modern Agriculture Sector. *International Journal of Modern Agriculture*, 9(4), 486–497. <https://doi.org/10.5281/zenodo.4428205>
- Donepudi, P. K., Banu, M. H., Khan, W., Neogy, T. K., Asadullah, ABM., & Ahmed, A. A. A. (2020c). Artificial Intelligence and Machine Learning in Treasury Management: A Systematic Literature Review. *International Journal of Management*, 11(11), 13–22. <https://doi.org/10.5281/zenodo.4247297>
- Fadziso, T., Adusumalli, H. P., & Pasupuleti, M. B. (2018). Cloud of Things and Interworking IoT Platform: Strategy and Execution Overviews. *Asian Journal of Applied Science and Engineering*, 7, 85–92. Retrieved from <https://upright.pub/index.php/ajase/article/view/63>
- Khan, W., Ahmed, A. A. A., Hossain, M. S., Neogy, T. K. (2020). The Interactive Approach to Working Capital Knowledge: Survey Evidence. *International Journal of Nonlinear Analysis and Applications*, 11(Special Issue), 379-393. <https://doi.org/10.22075/ijnaa.2020.4631>
- Madding, C., Ansari, A., Ballenger, C., Thota, A. (2020). Topic Modeling to Understand Technology Talent. *SMU Data Science Review*, 3(2), 1-18.
- Pasupuleti, M. B. (2016). Data Scientist Careers: Applied Orientation for the Beginners. *Global Disclosure of Economics and Business*, 5(2), 125-132. <https://doi.org/10.18034/gdeb.v5i2.617>
- Pasupuleti, M. B. (2017). AMI Data for Decision Makers and the Use of Data Analytics Approach. *Asia Pacific Journal of Energy and Environment*, 4(2), 65-70. <https://doi.org/10.18034/apjee.v4i2.623>
- Pasupuleti, M. B. (2018). The Application of Machine Learning Techniques in Software Project Management- An Examination. *ABC Journal of Advanced Research*, 7(2), 113-122. <https://doi.org/10.18034/abcjar.v7i2.626>
- Pasupuleti, M. B. (2020). Artificial Intelligence and Traditional Machine Learning to Deep Neural Networks: A Study for Social Implications. *Asian Journal of Humanity, Art and Literature*, 7(2), 137-146. <https://doi.org/10.18034/ajhal.v7i2.622>
- Pasupuleti, M. B., & Adusumalli, H. P. (2018). Digital Transformation of the High-Technology Manufacturing: An Overview of Main Blockades. *American Journal of Trade and Policy*, 5(3), 139-142. <https://doi.org/10.18034/ajtp.v5i3.599>
- Pasupuleti, M. B., & Amin, R. (2018). Word Embedding with ConvNet-Bi Directional LSTM Techniques: A Review of Related Literature. *International Journal of Reciprocal Symmetry and Physical Sciences*, 5, 9–13. Retrieved from <https://upright.pub/index.php/ijrsps/article/view/64>
- Pasupuleti, M. B., Miah, M. S., & Adusumalli, H. P. (2019). IoT for Future Technology Augmentation: A Radical Approach. *Engineering International*, 7(2), 105-116. <https://doi.org/10.18034/ei.v7i2.601>
- Rahman, M. M., Chowdhury, M. R. H. K., Islam, M. A., Tohfa, M. U., Kader, M. A. L., Ahmed, A. A. A., & Donepudi, P. K. (2020). Relationship between Socio-Demographic Characteristics and Job Satisfaction: Evidence from Private Bank Employees. *American Journal of Trade and Policy*, 7(2), 65-72. <https://doi.org/10.18034/ajtp.v7i2.492>
- Rahman, M. M., Pasupuleti, M. B., & Adusumalli, H. P. (2019). Advanced Metering Infrastructure Data: Overviews for the Big Data Framework. *ABC Research Alert*, 7(3), 159-168. <https://doi.org/10.18034/abcra.v7i3.602>