Machine Learning for Accurate Software Development Cost Estimation in Economically and Technically Limited Environments

Mohammad Alauthman, Department of Information Security, Faculty of Information Technology, University of Petra, Amman, Jordan

(iD) https://orcid.org/0000-0003-0319-1968

Ahmad al-Qerem, Computer Science Department, Faculty of Information Technology, Zarqa University, Jordan

Someah Alangari, Department of Computer Science, College of Science and Humanities, Shaqra University, Saudi Arabia

(iD) https://orcid.org/0000-0003-1308-1762

Ali Mohd Ali, Communications and Computer Engineering Department, Faculty of Engineering, Al-Ahliyya Amman University, Jordan

Ahmad Nabo, Software Engineering Department, Faculty of Information Technology, Zarqa University, Jordan

Amjad Aldweesh, College of Computing and IT, Shaqra University, Saudi Arabia*

Issam Jebreen, Computer Science Department, Faculty of Information Technology, Zarqa University, Jordan

Ammar Almomani, Skyline University College, UAE

Brij B. Gupta, CCRI, Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan & Symbiosis Centre for Information Technology (SCIT), Symbiosis International University, Pune, India & School of Computing, Skyline University, UAE, & Lebanese American University, Beirut, Lebanon & Center for Interdisciplinary Research, University of Petroleum and Energy Studies (UPES), Dehradun, India

ABSTRACT

Cost estimation for software development is crucial for project planning and management. Several regression models have been developed to predict software development costs, using historical datasets of previous projects. Accurate cost estimation in software development is heavily influenced by the relevance and quality of the cost estimation dataset and its suitability to the software development environment. The currently available cost estimation datasets are limited to North American and European environments, leaving a gap in the representation of other economically and technically constrained software industries. In this article, the authors evaluate the performance of regression models using the SEERA dataset, which highly represents these constrained environments. This study provides insights into selecting regression models for cost estimation in software development. It highlights the importance of using appropriate models based on the specific software development model and dataset used in the estimation process. In the performance evaluations of eight regression models, including elastic net, lasso regression, linear regression, neural network, RANSACRegressor, random forest, ride regression, and SVM, for cost estimation in different software models, along with correlation coefficients and accuracy indicators, were reported. The results showed that SVM and random forest indicated superior performance. However, the elastic net, lasso regression, linear regression, neural network, and RANSACRegressor models also demonstrated exemplary performance in cost estimation.

KEYWORDS

Constrained Environments, Cost Estimation, Regression Models, SEERA Dataset, Software Development

*Corresponding Author

1. INTRODUCTION

Cost estimation is a critical aspect of software development (Rankovic, Rankovic, Ivanovic, & Lazic, 2021; Rankovic, Rankovic, Ivanovic, & Lazic, 2021; Mukherjee & Malu, 2014), as it helps in predicting the resources required for the project and ensuring that the project is completed within budget and on time (Pandey et al., 2020). However, estimating the cost and effort for different software development models can be challenging due to their unique characteristics and requirements (Boehm, 2017; Kumar et al., 2020).

Several essential features must be considered when estimating the cost and effort for different software development models. One of the most crucial factors is the size of the project, which refers to the number of software components or functions that need to be developed. The larger the project, the more effort and resources it will require, ultimately impacting the cost estimation (Saavedra Martínez et al., 2020; Mahmood et al., 2021). The project's complexity is another critical feature affecting cost and effort estimation. The complexity of the software model can vary based on various factors, such as the number of interrelated components, the number of decision points, and the level of customization required. Developing more complex software models will require more effort and resources, resulting in higher costs (Mahmood et al., 2021).

The development team's expertise is another vital factor when estimating the cost and effort for different software models. The level of experience, knowledge, and skills of the team will significantly impact the development time and cost. A team with more experience and knowledge can develop a project more efficiently, resulting in lower costs (Nassif et al., 2019).

The development process also plays a crucial role in cost and effort estimation. The development process can be iterative or sequential, and each approach has advantages and disadvantages. The sequential approach, also known as the Waterfall model, is more structured, which can help ensure that each development phase is completed before moving on to the next. In contrast, the iterative approach, the Agile model, is more flexible and adaptable, allowing for changes throughout the development process.

Finally, the software development environment also affects cost and effort estimation. The environment can include hardware and software tools, such as integrated development environments, version control systems, and testing tools necessary to complete the project. The cost and availability of these tools and resources will impact the cost estimation for the project. As a result, several important features need to be considered when estimating the cost and effort for different software models. These include the project's size and complexity, the development team's expertise, the development process, and the software development environment. An accurate model for estimating the cost and effort will ensure the project is completed within budget and on time, providing significant benefits to the development team and the organization.

We observed first-hand the challenges that local software development teams faced due to limited resources and infrastructure constraints. Accurately estimating costs was critical for project planning and management under these conditions. However, existing cost estimation techniques and datasets did not adequately account for the realities of working in such constrained environments. We were motivated to address this research gap and help improve cost estimation practices for software teams operating under similar limitations.

Machine learning plays an essential role in determining the critical features that affect the cost and effort estimation of different software models (Safari & Erfani, 2020; Holtkamp et al., 2015; Casado-Lumbreras et al., 2014). With the help of machine learning algorithms, large and complex datasets can be analyzed to identify patterns and relationships between cost and effort variables and the factors that affect them. By utilizing machine learning techniques, such as regression analysis, decision trees, and neural networks, it is possible to identify the most significant variables that impact software development cost and effort.

Related Content

Smart Education Using Explainable Artificial Intelligence
Nikita Sharma (2025). *Internet of Behavior-Based Computational Intelligence for Smart Education Systems (pp. 95-130).*
www.irma-international.org/chapter/smart-education-using-explainable-artificial-intelligence/358977

Resource Scheduling in Fog Environment Using Optimization Algorithms for 6G Networks
Gaurav Goeland Rajeev Tiwari (2022). *International Journal of Software Science and Computational Intelligence (pp. 1-24).*
www.irma-international.org/article/resource-scheduling-in-fog-environment-using-optimization-algorithms-for-6g-networks/304440

Application of Functional Approach to Lists for Development of Relational Model Databases and Petri Net Analysis
Sasanko Sekhar Gantayatand B. K. Tripathy (2014). *Global Trends in Intelligent Computing Research and Development (pp. 407-444).*
www.irma-international.org/chapter/application-of-functional-approach-to-lists-for-development-of-relational-model-databases-and-petri-net-analysis/97067

Secure Image Processing and Transmission Schema in Cluster-Based Wireless Sensor Network
Mohamed Elhoseny, Ahmed Farouk, Josep Batle, Abdulaziz Shehaband Aboul Ella Hassanien (2017). *Handbook of Research on Machine Learning Innovations and Trends (pp. 1022-1040).*
www.irma-international.org/chapter/secure-image-processing-and-transmission-schema-in-cluster-based-wireless-sensor-network/180983

Computational Intelligence for Modelling and Control of Multi-Robot Systems
M. Mohammadian (2003). *Computational Intelligence in Control (pp. 122-135).*
www.irma-international.org/chapter/computational-intelligence-modelling-control-multi/6834