**MACHINE LEARNING ALGORITHMS FOR PREDECTING CONSTRUCTION DELAY**

Signature of supervisor

Sayed Arafat (1823027)

Department of Building Engineering and Construction Management, KUET

**Major**: Construction Engineering and Management

**ABSTRACT**: Construction delays heavily affect budgets, timelines, and the overall efficacy of the project. Conventional techniques for project delay prediction frequently depend on historical data and expert opinion, which can be imprecise and miss intricate relationships between different project components. This study thoroughly examines the use of machine learning methods to forecast building delays. At first, a review of the existing body of knowledge on influencing factors of construction project delay was used to survey experts to approach its quantitative data collection. Secondly several machine learning algorithms, such as Decision Tree, KNN, and Random Forest, are trained and assessed for their prediction abilities using a dataset that includes project parameters, such as size, location, complexity, weather, and contractor performance. Results from the evaluation metrics: accuracy score, confusion matrix, precision, recall, f1 score proved that ensemble algorithms are capable of improving the predictive force relative to the use of a single algorithm in predicting construction projects delay. KNN with optimized hyperparameters achieved accuracy of 83%. The method used to forecast the attribute underwent a thorough investigation, followed by a data analysis that revealed information that might be used to lessen the effects of the delay in construction in the future.

**INTRODUCTION**

Due to construction project delays, the construction industry is underperforming as seen by its low profitability, capital expenditure, and research and development, which leaves clients extremely dissatisfied with the sector's overall performance. According to certain studies, 9 out of 10 global megaprojects have delays, which typically lead to excessive cost overruns [1].

Machine learning algorithms are underutilized in the construction industry. To resolve these construction delay factor detection especially at the early stages of project planning. Machine Learning remains a new prospect within the construction sector despite its highly regarded advantageous potential [2]. The potential of machine learning (ML) techniques and algorithms in analyzing voluminous, complex, and interdependent data sets of varying structures. ML algorithms such as Decision Tree, Random Forest, KNN. This thesis introduces the prospect determine and create an effective predictive data analytics solution based on historical construction project data so that it may be used to assess and learn from objective delay risk factors. Achieving this goal will ultimately facilitate more accurate predictions of future projects.

**OBJECTIVES**

The objectives of this study are as follows:

1. To conduct literature review to identify common factors causing construction project delays, then interview experts to determine the most effective strategies.
2. To create hyperparameter-optimized prediction models, apply the factors that were determined in objective 1.
3. To combine the best predictive models from objective 2 to develop a high performant predictive model.

**RESEARCH METHODOLOGY**

* Conducting comparative studies as a literature review that shows the common factors causing construction project delays.
* Preparation of survey to identify the effective factors in construction field.
* Then analyzed the data, including a Cronbach's Alpha test to assess reliability.
* Imported into a Jupiter Notebook using Python programming language model.
* This raw data pre-processed and feature engineered dataset was split randomly into two in a ratio of 75% to 25% of training dataset and testing dataset respectively.
* Several machine learning algorithms, such as Decision Tree, KNN, and Random Forest, are trained and assessed for their prediction abilities.
* Finally, accuracy score to measure the model.

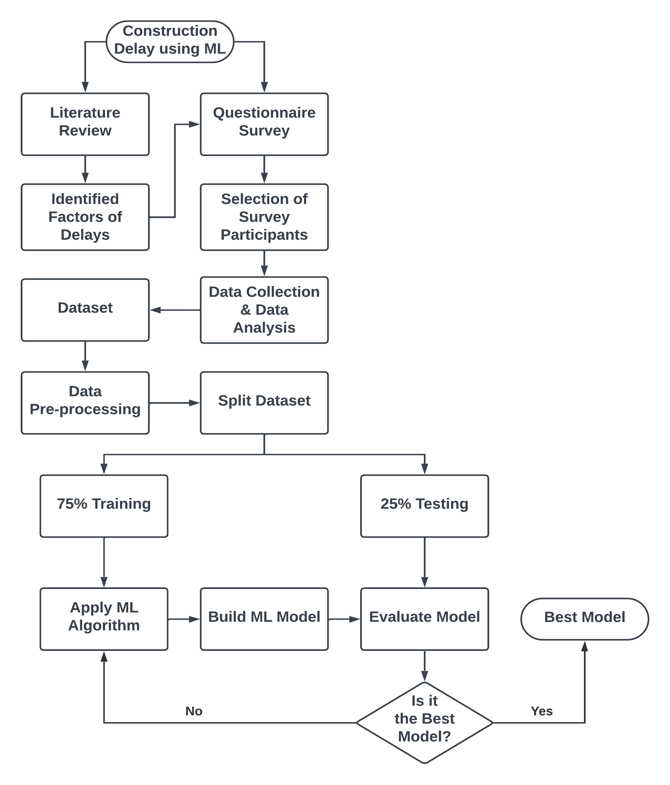


Figure 1. Flow Chart of the System

**RESULTS AND DISCUSSIONS**

* A questionnaire survey was done getting information about same space position held in construction industry.

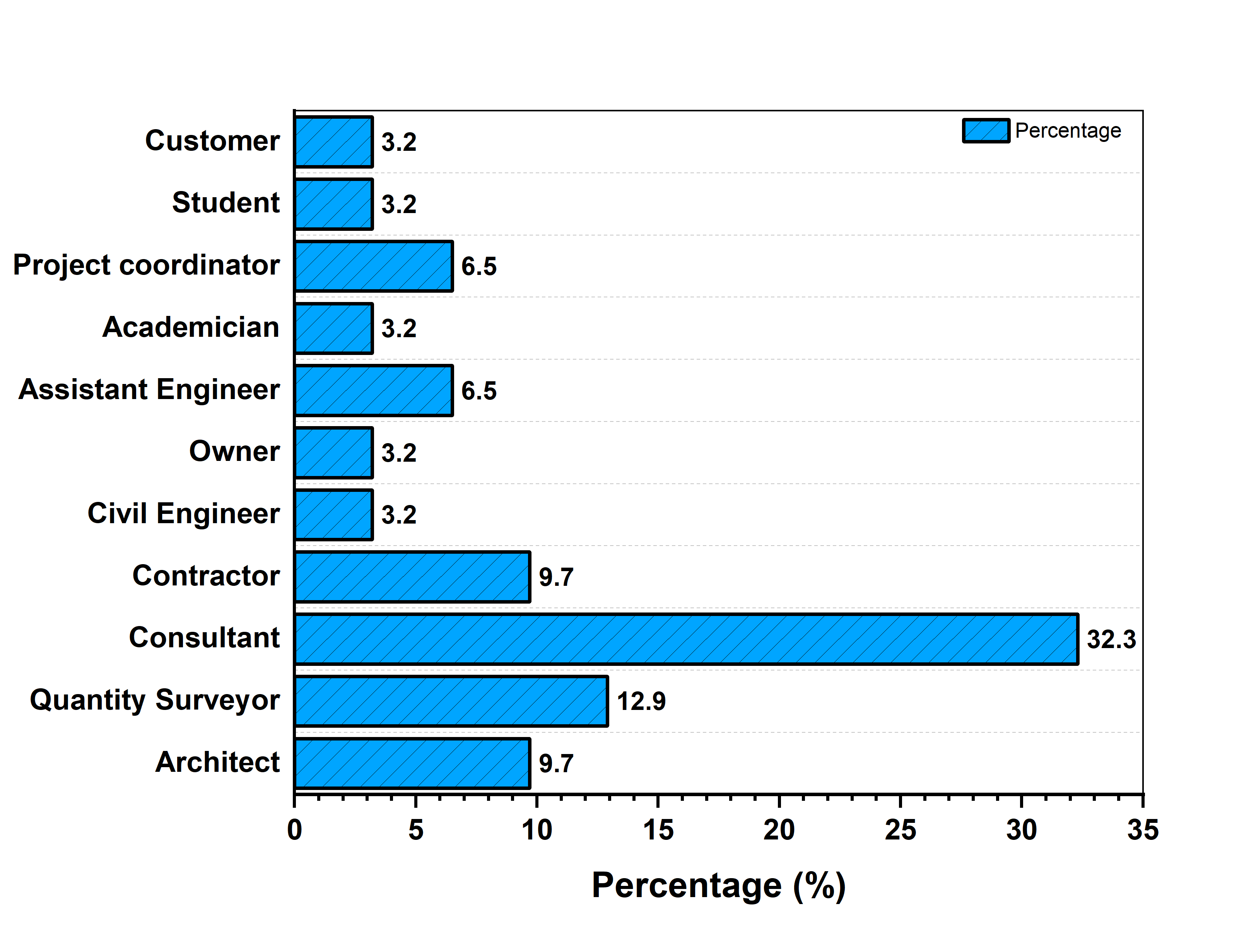
****

Figure 2. Analysis of Questionnaire Survey

* Analysis of questionnaire survey identified the most influential factors causing construction delays, providing insights into critical determinants impacting project timelines.

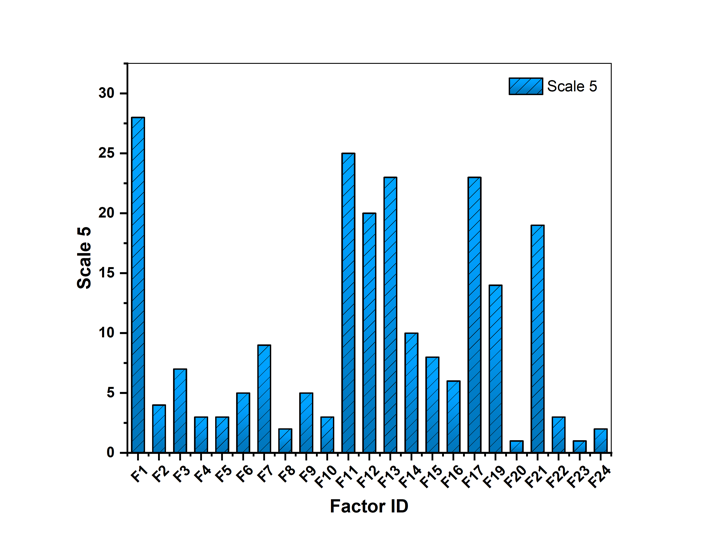


Figure 3. Analyzing Factors of Scale 5

* Through a survey on construction delay factors, the analysis revealed fewer effective reasons contributing to delays in construction projects.

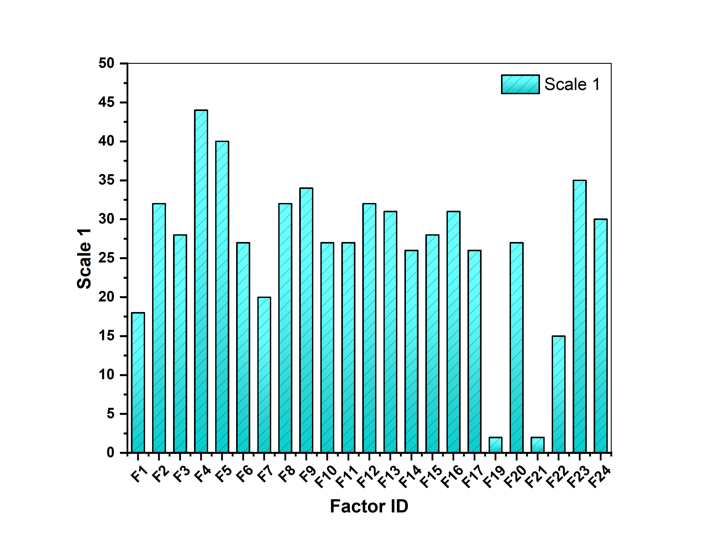
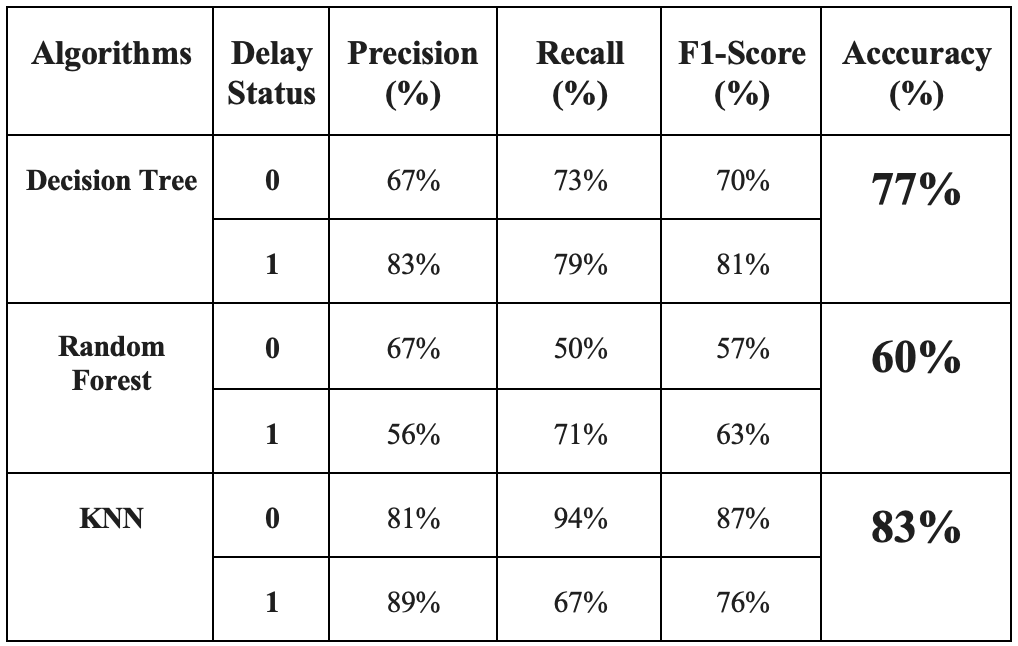


Figure 4. Analyzing Factors of Scale 1

* In my study on predicting construction delays, I implemented three machine learning algorithms: decision tree, random forest, and KNN. Among these, KNN yielded the highest accuracy, highlighting its effectiveness in forecasting construction delays.

Table I. Performance Metrics of Various ML Models



**CONCLUSION:**

* The study demonstrates the effectiveness of machine learning (ML) in predicting construction delays, aiding proactive decision-making and providing valuable insights to stakeholders. This integration enhances project success rates and risk management, enhancing planning and execution.

**REFERENCES**

1. Amos Darko, A. P. (2020). Building information modeling (BIM)-based modular integrated construction-Critical survey and future needs. *ELSIVIER*,7(3.4),128-143.
2. Sherif Abdelmageed, T. Z. (2020). A study of literature in modular integrated construction - Critical review and future directions. *ELSIVIER*,22(7),18-27.