AUTOMATED MANPOWER ALLOCATION BY PERFORMANCE ANALYSIS AND PROJECT CATEGORIZATION FOR CONSTRUCTION PROJECTS (PROJECT PULSE)

Kavin Mark Perera

(IT21087660)

BSc (Hons) degree in Information Technology Specializing in Information Systems Engineering

Department of Computer Systems Engineering

Sri Lanka Institute of Information Technology Sri Lanka

April 2025

DECLARATION

DECLARATION

I declare that this is my own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	Student ID	Signature
Perera K.M	IT21087660	

Signature of the Supervisor (Mrs. Buddhima Attanayaka) ....

Date 15/05/25

ABSTRACT

This work presents a machine learning–oriented approach to classifying construction projects for enhancing decision-making and operational efficiency in the construction industry. The study focuses primarily on the design and implementation of a project categorization function identifying construction projects based on relevant attributes extracted from real data gathered from MAGA Engineering Pvt Ltd. The aim is to create a data-driven system that can accurately classify various types of projects and thus support strategic planning and resource allocation.

The report establishes the problem of inconsistent and manual project classification in the construction sector, outlines the motivation for adopting a machine learning solution, and describes the methodologies employed ranging from data preprocessing, feature selection, model training, and validation. The classification task was performed using supervised learning algorithms and optimized for accuracy, interpretability, and feasibility of implementation. The solution proposed not only classifies the task automatically but also provides a scalable platform for broader project analytics and future growth.

This gives an overview of the primary aim, methodological overview, and anticipated results of the project, without delving into specific findings, according to academic standards.

ACKNOWLEDGEMENT

Above all, I would like to express my sincere gratitude to my supervisor, Mrs. Buddhima Attanayaka, who provided me with her constant guidance, encouragement, and support while carrying out my undergraduate study successfully. Her valuable comments and never-ending support were helpful in developing this study.

Besides my supervisor, my sincere thanks also go to Mrs. Narmada Gamage, the co-supervisor of this research project, for her valuable input and for being ever willing to provide help and direction whenever needed. Her support has greatly contributed to refining the project and keeping it on track.

I would also like to thank Mr. Darshana Senevirathne, our external supervisor at MAGA Engineering Pvt Ltd, for whose practical experience and real-world acumen this research gained much-needed real-world relevance. His inputs served to narrow the gap between practice and academia.

This research study, derived from a system of classification of construction projects, would not have been possible without the joint input and contribution of all the individuals listed above. I am truly grateful to them for their time, experience, and motivation during this experience.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	**Description**
ML	Machine Learning
AI	Artificial Intelligence
NLP	Natural Language Processing
SVM	Support Vector Machine
HTTP	Hypertext Transfer Protocol
RNNs	Recurrent Neural Networks
UAT	User Acceptance Testing
GCP	Google Cloud Platform
SMOTE	Synthetic Minority Oversampling Technique
BIM	Building Information Modeling
LIME	Local Interpretable Model-Agnostic Explanations
XAI	Explainable Artificial Intelligence
TF-IDF	Term Frequency-Inverse Document Frequency

1. INTRODUCTION

Developing infrastructure as well as other construction projects such as residential houses or complex power plants is crucial for economic growth. These construction projects are critical for the growth of society's economy while also amplifying infrastructure. Although these projects are exceedingly important, some of them have a lot of uncertainties and problems due to uncontrollable site conditions, size of a project, available resources, as well as stakeholder requirements. All these problems must be dealt with and the risks associated with every challenge must be meticulously managed. Among the many different management strategies and diagnosing potential problems, the most important and difficult one is managing risk. For every construction project, there is a specific risk that depends on its complexity. Identifying the specific degree is essential to managing resources in an orderly manner. Experts often rely on qualitative assessments in classifying control projects which then leads to a barrage of inefficient outcomes. Alternatively, every single piece of information and data needs to be numerically represented to bypass all the risks related to complexity misjudgment.

To mitigate these constraints, there is an increased focus on objective and consistent data-driven classification systems in projects within the construction industry. The automation of classifying projects within the scope of machine learning (ML) and computations has received great attention because these technologies offer a myriad of opportunities to develop more advanced classification models. An ML approach can study massive amounts of historical data pertaining to a project, including its size, duration, cost, location, and previous risk factors, and identify patterns which will enable it to classify new projects with more consistency and precision.

With advancements in technology, this study seeks to integrate traditional risk evaluation methods with techniques from modern computational methodologies through automation by developing an advanced system for construction project classification using machine learning algorithms. With the application of machine learning, this research anticipates improved decision making, reduced bias, and enhanced allocation of resources in planning and execution of construction projects.

Automation in classification systems improves accuracy and reproducibility, reduces the time taken to reach decisions, and guarantees faster scaling. More so, it establishes transparency in determining the risk levels of the project. Within organizations, a homogeneous approach can be employed for different project types, which enhances planning and minimizes the number of unforeseen challenges during project execution.

Nonetheless, the implementation of such systems has difficulties. The effectiveness of machine learning models is predicated on the existence of clean, well-organized datasets, and the quality of model functioning based on project data is largely contingent upon the existence of precise and comprehensive records of past work. Moreover, prevailing inertia to automation in combination with traditional reliance on the old way of doing things may slowly change.

To summarize, an automated, data-driven classification system could benefit the construction industry greatly. By applying advanced computing methods to risk management, this study attempts to facilitate precise project evaluations, aid in effective planning, and increase overall project success, all of which are fundamental goals of this research. The solution offered not only solves the current approach shortcomings but also meets the increasing needs of the sector concerning digitalization, transforming data into tangible insights, and decisive evidence-informed decision making.

1.1. Background Literature

Construction of buildings is inherently complex with an enormous array of stakeholders, varying regulatory requirements, and environmental influences that all induce uncertainties. The uncertainties could be such aspects as resource shortage and weather conditions through to whimsical governmental regulation oscillations, all which can have significant effects on project performance. The project type, objectives, and conditions of the site could be vastly different from project to project, resulting in a vast array of project types from small housing structures to gigantic infrastructure projects. With such intricacies at stake, it is essential to classify the projects accurately based on risk factors such as project type, budget, location, size, scope, duration, and site conditions for effective management of construction projects and minimizing risks.

In the past, the process of construction had such an issue to address these complexities primarily due to the fact that the industry lacked a formal structure to categorize and group projects. Project categorization and risk evaluation are currently integral components of construction management to ascertain appropriate resources, methods, and strategies to implement on any specific project. Different methodologies have evolved over time to evaluate and classify projects in terms of risk factors and their level of complexity. The requirement for a well-rounded system that can effectively classify projects has been universally recognized among the professional community and scholars.

Among the first studies in project categorization is that of Macealois (2024), who designed a classification system based on project type and risk factors [5]. Macealois' research divided construction projects into various groups, pointing out that each category of project has its own unique set of challenges and risks. By understanding these differences, project managers are able to make more informed decisions regarding resource management, timeline creation, and risk management strategies. Similarly, Safa and Sabet (2015) proposed a taxonomy framework that highlighted the differences in the levels of risk involved in the different types of projects [6]. Their research highlighted the fact that exposure to risk is not just financial and environmental risks but also considers factors such as project scope, regulatory environment, and building complexity. Such models are valuable resources for enhancing project implementation and uncertainty reduction.

Whereas traditional methods of project categorization drew primarily upon qualitative analysis, more sophisticated methods with the advent of recent growth have been supplemented in the form of more high-level, information-based methods. The development of computational approaches, such as machine learning and data analytics, has revolutionized the construction project classification sector. With the ability to process large amounts of data, these new approaches provide improved and quicker ways of classifying projects. Machine learning algorithms, for instance, can scan historical project data and establish patterns that are not easily identifiable through manual analysis. This shift from subjective expert judgment to objective, fact-based decision making could significantly improve project classification and risk assessment.

Machine learning techniques, and more particularly decision tree models, have been of potential in the automated classification task. Decision tree methodologies can review previous project data, such as project duration, cost, and environmental factors, to classify new projects as similar or dissimilar to previous projects. These models learn from available data and, by going through a training process, they can enhance their predictions to result in a more accurate categorization of projects. This automation lessens the built-in biases inherent in human judgment and can produce more consistent predictions for future projects. These methods can therefore be used to evaluate risk levels more effectively and determine the correct means of handling these risks.

Furthermore, the inclusion of text processing techniques in the categorization process has also been a major breakthrough. By processing textual data from project documentation such as project reports, contract documents, and regulatory filings, machine learning algorithms are capable of extracting valuable insights that help effectively categorize projects. Textual data can detect underlying themes related to project scope, compliance with regulation, and issues that may not be discerned from numeric data. This combination of text and numeric data further enhances accuracy and stability in project categorization systems, enabling the project manager to make informed decisions on the basis of more information.

Despite all these advancements, the construction industry still lacks a widely accepted, standardized method of automated project classification. Most of the existing frameworks are very reliant on subjective expert judgment, which can lead to inconsistencies and errors in classification. The absence of a widely accepted framework for risk assessment and project classification hinders the ability of organizations to manage risks uniformly and optimize resource utilization across different projects. Usage of experts also means that classification processes can be time-consuming, and outputs may vary significantly between experts, reducing the overall efficiency of project risk classification systems.

To address this gap, recent research has increasingly shifted its focus to designing automated machine learning and other advanced computer-based risk classification systems. Kim et al. (2020) researched South Korean infrastructure projects and determined that machine learning models were significantly more effective than traditional classification methods in estimating project risk [7]. By conducting their research, they established that automated models could analyze large amounts of data, identify patterns, and predict the complexity of infrastructure projects with greater precision than traditional methods. These findings support the growing belief that machine learning offers a hopeful solution for improving project classification and risk reduction in the construction industry.

The development of data-driven, automated project classification systems is also seen as a means to reduce the subjectivity and variability of traditional classification models. By utilizing past data and computation techniques, such systems can objectively classify projects in a way that similar projects are classified in a similar manner. Machine learning is also possible to be applied to ongoing improvement of the classification models because the system can learn from more data and improve its ability to predict over time. This flexibility makes machine learning-based systems particularly well suited to the changing and dynamic nature of construction project reality.

This research attempts to bridge the gap between traditional techniques of project categorization and potential offered by data-driven approaches. In suggesting an end-to-end, automated approach to project categorization, this research will be contributing to ongoing efforts to refine project management practice in the construction industry. Through the use of machine learning, text processing, and other computation techniques, it is possible to enhance the accuracy, efficiency, and consistency of project classification, leading to better decision-making and more successful projects.

1.2. Research Gap

Despite numerous advances in construction project management and classification in the past decades, there are fundamental limitations of current schemes, particularly involving integration of technology and objective data analysis. Most literature and construction project classification practices still rely on subjective expert opinion and traditional methods of risk assessment. These qualitative techniques, though effective in capturing experiential insight, tend to be susceptible to inserting inconsistencies and latent biases that dissuade repeatability and scalability of project assessments.

Application Reference	Automatic Classification	Applicable for construction projects	Risk Level Assessment	Complexity Analysis	Location/Environmental Impact Assessment	Budget Analysis	Time Frame Analysis
Procore	✗	✓	✗	✗	✗	✓	✗
Microsoft Project	✗	✓	✗	✓	✗	✓	✓
Smartsheet	✗	✓	✗	✓	✗	✓	✓
nPlan	✗	✓	✗	✗	✗	✗	✓
Trello	✗	✓	✗	✗	✗	✗	✗
Project Pulse	✓	✓	✓	✓	✓	✓	✓

Figure 1: Comparison between existing systems

Perhaps the most evident problem is non-standardization of project classification frameworks. Construction work varies extensively with regard to magnitude, scale, complexity, and risk exposure, but there is no common widely accepted framework used for categorizing these projects that is both holistic and versatile across regions, sectors, and project types. For instance, whereas some approaches rely solely on financial elements such as cost or budget levels, others take into account technical or environmental concerns surrounding project [11]. The inconsistency in the classification criteria is challenging when benchmarking project performance, predicting risk, and designing generic project management policies.

Besides this, there is also limited application of data-driven techniques within current classification systems. The construction industry, unlike industries like healthcare or finance, has been quite slow to implement advanced computational technologies like machine learning (ML) and natural language processing (NLP) [12]. Most classification models are still dependent on human data entry, expert marking, or rigid rule-based systems that cannot take advantage of the vast amounts of structured and unstructured data generated in modern construction projects. Most models, thus, do not have the richness and flexibility required to effectively assess the complex risks and complexities of modern construction environments.

Another major disadvantage is the absence of automated classification tools. Manual intervention is required in most buildings to access, comprehend, and validate data, so the process is inefficient, error-prone, and time-consuming. Human intervention at each step also brings in subjectivity, which can lead to inconsistent outcomes between various users or organizations. For example, two project managers might evaluate the same project's complexity differently based on their personal experience or organizational policies. This inconsistency seriously hinders efforts to streamline project appraisals and apply predictive analytics for pro-active decision-making [13].

Furthermore, empirical verification of classification models against actual data remains sparse. While various conceptual models and theoretical frameworks have been outlined in academic literature, relatively few have been tested or verified rigorously with large datasets from actual construction projects. This inconsistency makes it difficult to tell the applicability and validity in real-world circumstances of existing models. Without valid validation, nobody can be sure whether a model can accurately project project complexity or risk exposure for various construction circumstances [14].

Scalability is also essential. Most typical classification models can perform well where they are employed on small sample sets or bound types of projects. But as companies begin to deal with portfolios of tens or hundreds of projects simultaneously, these models begin to break down in terms of performance. They are not created to handle large-scale data ingestion, learning from new inputs over time, or dynamic project environments. Consequently, most project managers lack effective tools to deal with large and complicated project portfolios, which leads to inefficient planning and increased exposure to unexpected risks [15].

In addition, inconsistencies in risk measurement criteria also undermine the efficiency of existing classification systems. Definitions of such concepts as "high risk" or "complex project" are typically varied from region to region, organization to organization, and even department to department within one organization. This inconsistency makes inter-project comparison more difficult, reduces transparency, and inhibits transferability of classification models between various stakeholders. For example, a medium-risk infrastructure project by one company may be considered high-risk by another, based on their internal definitions and thresholds [16].

To all these is added the non-optimal use machine learning in risk and complexity evaluation. While AI has transformed sectors such as finance, manufacturing, and logistics, its application in construction generally in project categorization remains relatively underdeveloped. There have been few studies into how supervised and unsupervised machine learning models can be applied to improve the speed and accuracy of project classification by identifying patterns and outliers in large datasets [17]. In the few studies where AI has been applied, the outcomes have been promising, with improvements in predictive accuracy, as well as the ability to handle different types of data. However, these applications are still in their infancy and have not yet been widely accepted in practice.

Given these research and industry limitations, there is an urgent need for a new paradigm that breaks away from the weaknesses of traditional models. This study seeks to bridge that gap by proposing an automated project classification system using machine learning algorithms, such as decision trees, and natural language processing. By doing so, it hopes to offer an objective, scalable, and standardized process for construction project classification by complexity and risk exposure.

The proposed research in this paper shall utilize real dataset of MAGA Engineering Pvt Ltd, a high-profile construction company engaged in various infrastructure and development schemes. The data includes both structured data (i.e., cost, location, duration, number of stakeholders) and unstructured textual data (i.e., technical reports, project summaries, correspondence logs). The study will apply text processing techniques to extract valuable features from such unstructured data sources and combine them with the structured ones to create a robust input space for the machine learning models [18].

Decision tree models are particularly well placed to do this since they are explainable and handle both categorical and numerical data. As compared to black-box model types such as deep neural networks, decision trees allow users to trace the reasoning behind each classification decision, and this enhances transparency and user trustworthiness. This is vital in construction since stakeholders and project managers need to understand reasons for risk and complexity determinations in order to implement corrective action. Also, cluster algorithms can be used with decision trees to categorize similar projects such that benchmarks and detailed portfolio analysis are made easier.

The research will also address model interpretability and user adoption, two of the largest obstacles to AI implementation in building construction. The classification tool will include elements such as confidence levels, decision paths, and visualization tools to allow end-users to learn how the model arrived at a specific classification. These elements are important in acquiring the trust of stakeholders that may question automated decision systems.

By bridging the gap between traditional qualitative methodologies and modern-day data analytics, this research attempts to create an adaptable, calibrated, and extensible system to classify construction projects. The implementation of computational strategies is expected to significantly reduce disparities, improve accuracy in classification, and enable risk avoidance measures preemptively. Also, the normalized system created through this research can serve as an example for other construction firms wanting to automate project planning and risk management processes.

This research aligns with the broader digital revolution in construction as information comes to be ever more central to the push for efficiency, safety, and innovation. Through the use of AI and machine learning in addressing an extremely well-known problem with project classification, the research offers theoretical insight as well as usable tools that organizations are able to implement when they are looking to digitize their business.

In conclusion, the research acknowledges the tremendous improvement made in construction project classification but identifies the critical gaps that still exist in standardization, automation, scalability, and validation. Using real-world datasets from MAGA Engineering Pvt Ltd and explainable machine learning techniques, this work aims to bridge these gaps and enhance the current state of construction risk and complexity appraisal. The resulting model will serve as both a decision-support system and a knowledge base for future research, offering the foundation for more reliable, transparent, and data-driven project management practices in the construction sector.

1.3. Research Problem

Building projects vary significantly in complexity, size, scope, and risk exposure. These are influenced by a vast number of interacting parameters including type of project, location, objectives, available resources, environmental conditions, and stakeholders' requirements. Due to this complexity, project managers have often found it difficult to apply consistent and uniform classification approaches to a range of projects. The absence of a formal framework to evaluate and categorize construction projects in terms of complexity and risk leads to inefficiencies in project planning, cost overruns, delayed timelines, and poor decision-making.

Classification systems in construction have conventionally depended on qualitative, experience-based inputs, where project managers utilize their judgment and professional experience to establish complexity and associated risks [19]. While such experience is valuable, it is also susceptible to individual bias and may lack the precision that portfolio analysis at scale requires. Misclassification or inconsistent classification of projects can create a chain reaction of negative effects, ranging from misallocated resources to flawed risk mitigation to missed opportunities for optimization. Therefore, the construction industry is in urgent need of an objective, scalable, and automated approach to project classification that reduces human bias and enhances the accuracy and reproducibility of risk analysis.

To bridge this gap, the following overall question guides this research: "In what ways can construction projects be categorized by risk and complexity based on project type, scope, objectives, location, budget, duration, size, and site conditions?" This question encapsulates the need to identify quantifiable project factors and utilize computational frameworks to estimate their impact on overall project complexity and risk exposure. Specifically, the research aims to extend traditional qualitative methods insofar as it proposes a decision-tree-based classification model that allows automatic and accurate project assessment.

Decision tree algorithms are very interpretable and effective in classification, particularly when there are multiple input variables to be taken into consideration at a time [20]. By incorporating structured attributes (i.e., budget, location, and duration) and unstructured data (i.e., project reports and descriptions), the model will have an exhaustive analysis of the risk levels of projects. In this manner, the classification process will not only be data-dependent and standardized, but also adaptive across different types of construction projects.

The lack of a computationally efficient project classification system has some highly significant implications for the construction industry. First, cost overrun remains one of the most significant problems. Inaccurate project classification may result in underestimation or overestimation of the risk and hence inappropriate budgetary provisions. For instance, a project misclassified as low risk may not receive sufficient contingency funding and hence may fall victim to unexpected disruptions [21]. Conversely, overestimating risk can inflate budgets unnecessarily, resulting in wasted financial resources.

Second, project delays are typically the outcome of failing to identify and counter risks at the beginning of the project cycle. When complexity is not properly assessed in the planning stages, unforeseen issues may emerge during implementation, disrupting timelines and straining stakeholder relationships [22]. Delays can be particularly damaging in the instance of infrastructure and public sector projects where contractual deadlines and public expectations are rigid.

Third, inefficient resource allocation is the direct consequence of misclassified projects. Contractors may over-allocate heavy machinery, veteran personnel, or high-technology techniques to projects that do not require them and under-supply high-risk projects. This imbalance generates productivity losses, reduces employee efficiency, and increases operational costs [23]. The proposed automated classification system can more accurately align resources with project demands, which can enhance overall productivity.

Fourth, there are compliance issues related to regulation. Many construction projects fall under legal, environmental, and safety regulations that vary across jurisdictions. Improper risk classification can cause firms to overlook specific compliance requirements, leading to project delays, fines, or litigation [24]. An effective classification system that categorizes high-regulatory-risk projects can prevent these issues, ensuring adherence to the law and reducing the potential for litigation.

Fifth, and perhaps more intangibly, are reputational risks. Contractors that consistently fail to deal with project risk and complexity can forfeit the confidence of their clients, investors, and partners. This loss of confidence can manifest itself in reduced future business prospects and difficulties in securing funding or public tenders. In a competitive market, reputation is just as valuable as operational capability. A transparent and rigorous classification system that demonstrates due care in project appraisal can serve as a competitive advantage for firms that desire to maintain their position in the market [25].

By addressing the pain areas, the research will play an important role in improving decision-making in construction project management. Not only does the model proposed here streamline classification processes, but it also includes explainability mechanisms whereby end users can visualize how classification outcomes were determined. Explainability mechanisms such as these are essential to gain the trust of construction practitioners accustomed to traditional, manual means of risk appraisal. The ability to view decision trees and trace classification paths will make the system usable and useful.

In short, the absence of a computationally efficient, standardized classification model is a crucial impediment to risk and complexity analysis in construction ventures. The aforementioned impediment is addressed in this research through the development of a machine learning–based classification framework, founded upon decision tree methodology, and supplemented by natural language processing capabilities. Using real MAGA Engineering Pvt Ltd data, the model would be trained, tested, and validated for applicability and relevance. Through enhancing accuracy, reducing subjectivity, and increasing resource allocation and adherence to compliance, the proposed system has the potential to significantly enhance risk management procedures within the construction sector.

1.4. Research Objectives

The overall aim of this study is to develop a systematic, evidence-based approach for classifying construction projects according to their complexity and risk level. The building construction sector is dominated by high diversity in project type and operating conditions and does not have an industry-acceptable model for systematically classifying projects uniformly and objectively. The Project Management Institute (PMI) has identified the importance of systematic classification in effective project management, particularly when dealing with portfolios consisting of various, multiple types of projects [1]. Despite these guidelines, subjective judgment remains the dominant method of project classification among most construction firms, leading to inefficiencies and heightened risk exposure [2].

To resolve these problems, the research poses a concrete objective: to develop an automated classification system capable of measuring project complexity and attendant risks using actual data. It will assist construction managers and organizations in improving decision-making, resource allocation, and risk minimization strategies. It is consistent with the current need for computational intelligence in construction activities and learns from the experiences of companies like MAGA Engineering Pvt Ltd, which has carried out diverse construction projects in different geographical and regulatory environments [3].

The first sub-objective focuses on data gathering. A wide dataset with rich project features such as type, scope, goals, place, cost, schedule, and size, together with environmental concerns will be gathered. The data forms the basis for the classification process and makes the model grounded. Performance measures in project management can be effectively used only when the classification of the project is based on objective factors has been proved previously [4], further establishing that structured data gathering is needed.

Secondly, the research will involve the classification of project types by the historical analysis of previous construction projects. By grouping these into predetermined categories (e.g., infrastructure, residential, commercial), the model can be learned to recognize patterns in various project scenarios. Literature states that grouping similar projects according to structure and execution plan facilitates uniform complexity determination [11], thus this step necessary in the removal of ambiguity as well as enhancing classification precision.

A third objective is the determination of categorization criteria, focusing on quantifiable aspects of complexity and risk. The application of quantifiable indicators of risk and complexity has been given priority in previous research as crucial to utilize as measurable indicators such as cost variability, environmental uncertainty, stakeholder engagement, and logistical constraints in project difficulties [12], [13]. From these elements, the study will construct a taxonomy of risk and complexity indicators as the basis for the classification model.

Following this, the study will utilize machine learning techniques to predict complexity and risk. In concrete terms, decision trees will be used to classify project inputs and calculate respective complexity levels, and text processing techniques (e.g., NLP) will be used to read project documentation and extract implicit complexity signals [18], [20]. The strength of decision trees is that they can model complex decision paths in an open way, which is essential in the construction sector where transparency of decisions is paramount [14].

The fifth objective is to validate and optimize the model using actual-case studies from MAGA Engineering Pvt Ltd. Validation is a critical aspect of ensuring the reliability and effectiveness of the model across varied construction environments. Case-based examination helps identify model vulnerabilities and provides opportunities for improvement [19], [21]. Optimizing the model for scalability also guarantees its suitability to small-scale and large-scale construction portfolios.

Equally important is the sixth one: ensuring that there is adherence to current industry practices and regulation standards. Construction is governed by strict regulations whose requirements vary from place to place. A model that does not take regulatory pressure into account is incomplete in practice [24]. The framework, therefore, will incorporate regulatory conformity factors, which will help businesses meet legal conditions and avoid potential fines.

Lastly, the research seeks to develop a working prototype demonstrating how the model can be applied in real life. The prototype will allow project managers to input project attributes and receive instant classification results, enabling streamlined planning processes. The system may be integrated into construction project management software, promoting efficiency, reducing human error, and enabling strategic decision-making [25].

Briefly, this study aims to revolutionize project classification in the construction industry with a systematic, automated, and intelligent approach. Utilizing a combination of decision tree models and natural language processing, the model will go beyond traditional risk analysis and offer predictive findings based on empirical evidence. This will not only address gaps that have been highlighted in recent literature [15] – [17], but also allow for long-term cost management, time, and risk management improvements for construction firms. Overall, this research seeks to bridge the gap between theory and industry application by offering an easily scalable and applicable solution to one of the most persistent problems in construction.

2. METHODOLOGY

2.1. Methodology

The Project Categorization Component is a core component of our system that attempts to classify construction projects based on their risk and complexity profiles. Effective classification is crucial for effective resource allocation, risk minimization, and better project management, allowing stakeholders to make extremely informed decisions throughout the lifecycle of the project. The methodology developed to create this system is a blend of existing machine learning techniques, data analysis processes, and knowledge in the construction sector domain. The intention is to automate the process of classification with scalability, consistency, and reliability and lesser human touch and personal judgment in categorization.

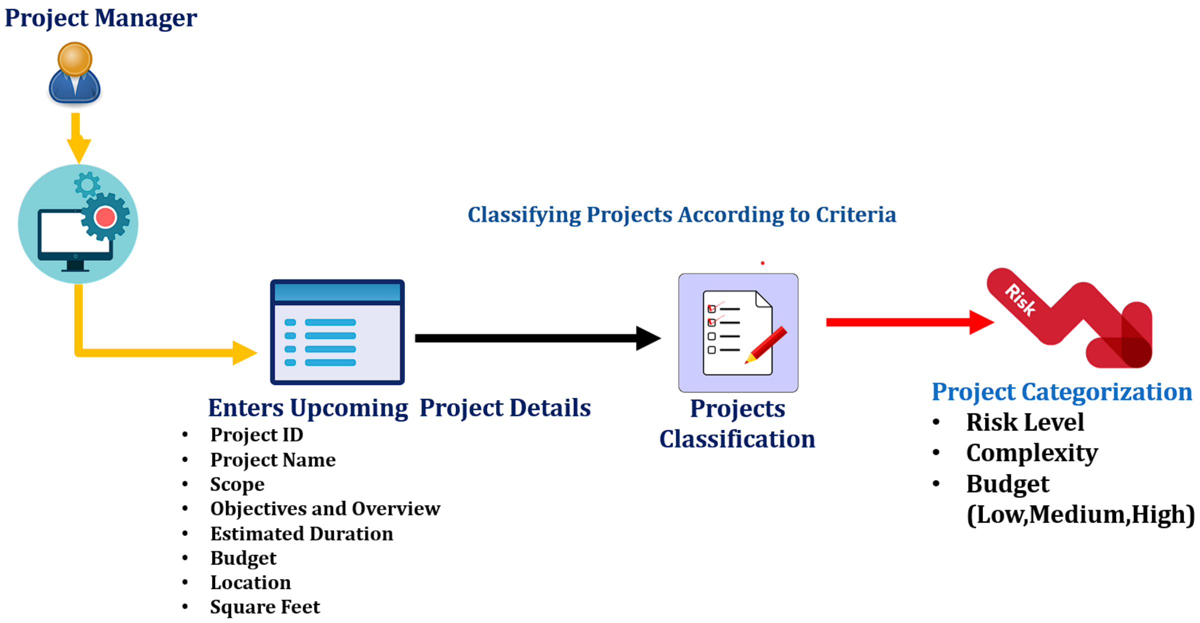


Figure 2: Overview Diagram of Project Categorization

2.1.1. Data collection and preprocessing

To build a robust and reliable project categorization model, we begin with the collection of historical project data from MAGA Engineering Pvt Ltd. This dataset comprises a wide range of features that reflect the various aspects of construction projects. These features include:

Project type: This could encompass residential, commercial, infrastructure, and other types of construction projects.

Project scope and objectives: Defines the scale and goals of the project, including whether it is aimed at expansion, renovation, or new construction.

Geographic location and site conditions: The location of the project and the site-specific conditions such as terrain type, climate, and environmental concerns.

Budget and financial allocation: The total financial investment required for the project.

Estimated project duration: The planned duration for completing the project, including various phases such as design, procurement, and construction.

Complexity factors: This includes aspects such as legal constraints, environmental challenges, and the number of stakeholders involved in the project.

Once the dataset is collected, data preprocessing is carried out using Python-based tools like pandas, NumPy, and scikit-learn. The preprocessing steps are critical to ensure the quality and consistency of the data before feeding it into machine learning models. The main preprocessing tasks include:

Data Cleaning: Identifying and handling missing values, detecting outliers, and normalizing the data to ensure uniformity and scale across different features.

Feature Engineering: Transforming categorical data (e.g., project type) into numerical values using techniques like one-hot encoding or label encoding.

Text Processing: Using Natural Language Processing (NLP) techniques such as TF-IDF (Term Frequency-Inverse Document Frequency) vectorization to extract useful features from textual project descriptions or documentation.

Data Splitting: Dividing the dataset into training (80%) and testing (20%) subsets to evaluate model performance and ensure generalization.

2.1.2. Machine learning model selection

A critical component of this methodology involves selecting the most suitable machine learning model for classifying projects based on their complexity and risk. Several models are evaluated to find the best one for the specific nature of construction projects. The following models are considered:

Decision Tree Classifier: This model is easy to interpret, which makes it a good candidate for understanding project classification, but it may lead to overfitting, especially with complex data sets.

Random Forest Classifier: An ensemble method that combines multiple decision trees to increase robustness and reduce the risk of overfitting, making it suitable for handling a variety of project complexities.

Support Vector Machine (SVM): Known for its effectiveness in high-dimensional spaces, SVM is useful for classifying complex data and is robust to overfitting, especially when the feature space is large.

Neural Networks (TensorFlow/Keras): A deep learning-based approach that can potentially offer higher accuracy in handling complex relationships within the data.

After a comprehensive evaluation of the models using metrics like accuracy, precision, recall, and F1-score, the Random Forest Classifier emerges as the preferred model. Its ability to handle large datasets with complex relationships, along with its robust performance in preventing overfitting, makes it a powerful choice for categorizing construction projects.

2.1.3. Model training and optimization

Once the Random Forest model is selected, it undergoes a detailed training process to optimize its performance. This involves several iterative steps aimed at fine-tuning the model and ensuring that it can make accurate predictions on unseen data. The optimization steps include:

Grid Search Cross-Validation: This technique is used to test different combinations of hyperparameters such as the number of trees, the depth of the tree, and the minimum number of samples required to split a node. By systematically searching through these hyperparameters, the best configuration is identified to maximize model performance.

Feature Importance Analysis: This step helps to identify the key factors that contribute most significantly to the classification of project complexity and risk. Understanding these features can provide valuable insights into the underlying patterns in the data and guide future improvements.

Ensemble Learning: In some cases, combining multiple models (such as Random Forest, Decision Tree, and SVM) can improve the overall performance by leveraging the strengths of each model. This ensemble learning technique helps to enhance the robustness of the prediction.

Post-training, the optimized model is saved in a .dat format using joblib for efficient deployment within the backend system, where it can be accessed by users in real-time to classify new projects.

2.1.4. Project categorization on complexity, risk, and budget levels

Effective project management hinges on accurate categorization based on the project's complexity, associated risks, and budget. Each of these factors plays a vital role in ensuring the successful execution of the project, efficient use of resources, and proper risk mitigation.

2.1.4.1. Introduction

Correct classification of construction projects based on risk and complexity is essential for several reasons. Firstly, it allows optimal use of resources, in a manner that more complex projects receive the attention and resources they need, and less complex projects receive a cost-effective approach. Secondly, it helps in developing realistic project budgets, identifying likely risks at the earliest, and specifying well-defined timelines for every stage of the project.

The process of classification must consider a number of factors, such as the type of project (residential, commercial, infrastructure), conditions of the site, cost range, and size of the project. The aim is to come up with a system that can sort projects into one of a number of categories depending on these factors consistently and reliably.

2.1.4.2. Budget level categorization

Budget is a critical factor when categorizing construction projects. Projects are classified into three budget levels based on percentile thresholds derived from historical budget data. This classification helps to understand the financial scale of a project and allocate resources accordingly.

Percentile threshold definitions

- **Q1 (33.3rd percentile)**: The budget value below which 33.3% of projects fall.
- **Q2 (66.6th percentile)**: The budget value below which 66.6% of projects fall.
- **Q3 (Optional 75th percentile)**: The budget value below which 75% of projects fall.

Budget level categories

- **Low-Scale Budget: Budget ≤ Q1 (bottom 33.3%)**
- **Medium-Scale Budget: Budget between Q1 and Q2 (middle 33.3%)**

- **High-Scale Budget: Budget > Q2 (top 33.3%)**

This approach allows project managers to assess the financial scope of a project and plan accordingly for staffing, resources, and contingency funds.

2.1.4.3. Project Complexity Categorization

The complexity of a project is determined by several factors, including the project type, size, site conditions, and the level of expertise required. Projects are classified into **High**, **Medium**, or **Low Complexity** categories based on these factors.

Complexity Categories Overview

- **High Complexity**: Demanding projects that require significant resources, specialized construction techniques, or involve challenging site conditions.
- **Medium Complexity**: Projects that are of moderate scale and complexity but do not involve extreme challenges.
- **Low Complexity**: Simple projects with minimal demands on resources and standard construction practices.

Rules for Assigning Complexity

- **High Complexity**: A project meets at least one high complexity condition.
- **Medium Complexity**: A project meets at least one medium complexity condition but no high complexity conditions.
- **Low Complexity**: A project does not meet any high or medium complexity conditions.

Table 1: Criteria and Conditions for Complexity

Factor	High Complexity	Medium Complexity	Low Complexity
Project Type	Institutional, Industrial	Commercial, Retail	Residential
Budget Level	High-Scale (> Q2, Top 33%)	Medium-Scale (Q1–Q2, Middle 33%)	Low-Scale (≤ Q1, Bottom 33%)
Site Conditions	Remote location, Hilly area	Coastal zone	Urbanized land, Flat terrain
Square Feet	Large (>66.6%)	Medium (33.3%–66.6%)	Small (≤33.3%)

2.1.4.4. Project Risk Categorization

Risk categorization is influenced by complexity levels and specific site conditions. This categorization helps in identifying potential issues such as delays, accidents, or regulatory challenges.

Risk Categories Overview

- **High Risk**: Projects that face significant potential difficulties, such as accidents or severe construction challenges.
- **Medium Risk**: Projects with moderate risk factors, requiring careful planning and consideration.
- **Low Risk**: Projects with minimal expected risk, based on standard conditions and procedures.

Rules for Assigning Risk

- **High Risk**: Projects that are either high complexity or located in remote/hilly areas.
- **Medium Risk**: Projects that are medium complexity or located in coastal zones.
- **Low Risk**: Projects that are low complexity and located in urbanized or flat terrain.

Table 2: Criteria and Conditions for Risk

Complexity or Site Conditions	High Risk	Medium Risk	Low Risk
High Complexity	✓	✗	✗
Medium Complexity	✗	✓	✗
Low Complexity	✗	✗	✓
Remote location, Hilly area	✓	✗	✗
Coastal zone	✗	✓	✗
Urbanized land, Flat terrain	✗	✗	✓

Table 3 :Summary Table for Project Categorization

Factor	High Complexity (High Risk)	Medium Complexity (Medium Risk)	Low Complexity (Low Risk)
Project Type	Institutional, Industrial	Commercial, Retail	Residential
Budget Level	High-Scale (> Q2)	Medium-Scale (Q1-Q2)	Low-Scale (≤ Q1)
Site Conditions	Remote location, Hilly	Coastal zone	Urbanized land, Flat
Square Feet	Large (>66.6%)	Medium (33.3%–66.6%)	Small (≤33.3%)

2.2. Commercialization Aspects of the Product

2.2.1. Market Need and Potential

The construction sector finds itself in the midst of an ever-accelerated digitalization, reacting to the relentless call for increased operational efficiency, reduced risk, and data-driven project management solutions. In this dynamic environment, the Project Categorization Component holds a pertinent and greatly needed answer. Construction projects whether residential, commercial, or infrastructural—involve great complexity and risk. Their categorization in terms of risk, complexity, and financial considerations is necessary to facing the challenges effectively. Unfortunately, the majority of companies continue to utilize manual classification techniques, which are largely inconsistent, time-wasting, and error-prone.

The need for automatic, standardized, and scalable classification systems is therefore becoming increasingly urgent. This product addresses this need by introducing a machine learning-powered engine that reads historical project data and automatically classifies new projects according to structured criteria. This standardizes classification procedures and also significantly improves risk assessment accuracy. The system also operates in seamless integration with modern project management software, ensuring its feasibility in real-world contexts.

As the construction industry continues to grow in its reliance on digital solutions, the demand for smart risk classification tools is certain to rise. Construction companies, project managers, and consultants now look for plug-and-play solutions that enable swift decision-making, reduce uncertainties, and help in improved financial planning. By delivering a sound, data-backed platform to classify construction projects, the Project Categorization Component positions itself as a vital tool in contemporary construction management setups.

2.2.2. Key Benefits of the Product

Among the strongest value propositions of the Project Categorization Component is that it reduces project risk. By classifying projects properly in the first place, stakeholders can anticipate and prepare for issues related to scope, budget, location, and environmental conditions. This makes them better able to prepare and plan contingencies, resulting in fewer surprises during implementation.

Besides risk reduction, the system also offers greater efficiency. Through the automated process of classification based on historical data and machine learning algorithms, time and effort in planning a project are substantially reduced. Processes that once took days or even weeks to be accomplished manually can now be accomplished within minutes, freeing up skilled human resources to focus on higher-order decision-making processes.

The component also enables data-driven decision-making. Through empirical research and pattern identification, the system creates trends and insights that are not necessarily apparent through traditional evaluation methods. This gives project managers, financial planners, and engineers a firm basis on which to make decisions, particularly in the early stages of a project where strategic alignment is critical.

2.2.3. Business Model

In order to ensure long-term sustainability and access of the Project Categorization Component, a diversified commercialization strategy has been formulated. The base model is Software as a Service (SaaS), in which customers are provided access to use the platform using cloud-based infrastructure. It is a subscription-based model, and this allows scalability according to the needs of small, medium, and large businesses. The SaaS model ensures consistent income while lowering the entry barriers for businesses that do not want to invest in complex IT infrastructure.

The second model is offering the product as a One-Time Licensing solution. This model is best suited for large organizations, which would want to host and maintain the software within their own infrastructure, typically due to data security and compliance requirements. These enterprise editions contain customization features that are tailored to the client's needs, such as localization for regional projects or interfacing with legacy systems.

A third choice is API Integration. The Project Categorization Component is designed to be modular and extensible. A REST API allows third-party developers and construction software vendors to embed its functionality directly into their systems. This enriches their existing products and extends the classification engine's reach without requiring a full-fledged user interface or stand-alone deployment.

The fourth and final business model is Consultation Services. Even though the system is computerized, building firms are often in need of expert guidance throughout the initial stage of implementation. The consultation services include system installation, historical data analysis, rule optimization of classifications, and training. Not only does it enhance customer satisfaction, but it is also a source of extra revenue.

2.2.4. Competitive Landscape

The construction analytics and project management market are becoming increasingly saturated with solutions promising improved efficiency and risk reduction. There are already available solutions like Autodesk Construction Cloud, Procore, and Trimble that offer various analytics modules, typically incorporating AI to assist in forecasting and scheduling. None of those platforms have the particular focus of project categorization via machine learning and standardized frameworks considering complexity and risk factors.

The Project Categorization Component stands out with some important differentiators. To begin with, it offers true AI-based automation in project categorization. While most of the competition relies on predefined rule sets, our solution learns from historical data and adapts to changing project characteristics over time. The result is a smarter, more dynamic tool that improves the more data it's provided.

Second, the system incorporates customizable risk parameters. This enables organizations with different operational profiles, risk tolerance, and geographical factors to tailor the classification engine to their specific context. Unlike one-size-fits-all solutions, this flexibility ensures applicability and utility in diverse construction environments.

Third, the product is specifically designed for large construction projects. Such specialization enables deeper incorporation of industry-specific variables such as stakeholder complexity, geographical conditions, and regulatory environments. In contrast, most competitors take a more generalist approach, sacrificing domain-specific accuracy for broad applicability.

2.2.5. Pricing Strategy

In order to attract a wide user base and variety of business sizes, the Project Categorization Component pricing scheme is tiered and scalable. The Basic Plan is $99 monthly, with emphasis on solo contractors and small businesses that require basic functionality without advanced integration and analytics features. The plan includes basic classification tools, a limited project dataset, and access to the web-based interface.

The Enterprise Plan is $499 a month and includes advanced features such as API access, deep learning-based classification models, and complete analytics dashboards. This plan is appropriate for medium-to-large enterprises that work on multiple projects simultaneously and require advanced reporting and integration features.

For the very specialized needs of companies, a Custom Solution feature is provided. This tier offers personalized system configurations, on-premises deployment, private cloud hosting, and high-end advisory services. The pricing for this tier is determined by integration complexity, deployment size, and number of users. This flexibility level ensures that the product is able to accommodate organizations at any stage of digital maturity.

2.2.6. Market Entry Strategy

A strategic and phased market entry strategy has been devised to gain optimum early traction and long-term sustainability. Pilot implementation with partner businesses, particularly ones already involved in collaborative research or digital innovation in the construction sector, is the first phase. The pilot installations will provide actual-world evidence of the system and allow for refinement before large-scale deployment.

Thirdly, the product will be launched through industry conventions and webinars. These channels offer a best-in-class way of demonstrating decision-makers, engineers, and project managers the power of the classification system in real time. Credibility can be established through live demonstrations, technical presentations, and panel of experts and invite the early adopters.

Digital marketing campaigns will also lead the way in market penetration. The campaigns will be launched on LinkedIn, Google Ads, and industry-specific platforms for the construction sector. With targeted messaging and performance marketing methods, the product will reach stakeholders looking actively for digital risk assessment and project planning tools.

In addition, the group will pursue strategic alliances with established construction software players. By offering the Project Categorization Component as a white-label module or an API, it is possible to integrate it into platforms that possess pre-existing market trust and large user bases. Strategic partnerships will accelerate the adoption and increase the product's visibility in established software ecosystems.

2.2.7. Scalability and Future Roadmap

As the product gains market traction, the design is such that there will be easy scaling to manage larger sets of data and more complex classification rules. Future versions will focus on adding functionality to the machine learning model by integrating real-time construction data, IoT feeds, and remote sensing data to dynamically calculate risk and complexity.

Another salient aspect in the construction is the recommendation engine, which will not just classify projects but also suggest risk mitigation, budgeting, and proper resource allocation based on similar previous projects.

Localization is also a priority. With global interest in construction analytics, the system will subsequently include support for multiple languages, localization of region-specific regulatory issues, and localization of cost and resource units. This shall provide relevance across multiple geography markets and regulatory environments

2.3. Testing & Implementation

The successful deployment of the Project Categorization Component within a real-world construction environment hinge on rigorous testing and seamless implementation. Given the complexity of integrating AI-driven solutions into legacy project management workflows, this section outlines a comprehensive approach that encompasses multiple testing layers, a phased implementation roadmap, and strategic mechanisms to overcome technical and operational challenges. The goal is to ensure that the system performs reliably under diverse conditions while offering scalability, usability, and real-time insights that construction managers can depend on.

2.3.1. Testing Phases

A multi-tiered testing protocol is crucial to ensure system reliability, robustness, and adaptability. The testing process begins with **unit testing**, where individual modules of the systems such as data pre-processing scripts, model inference functions, and utility components are validated for correctness. This is typically executed using Python-based testing frameworks like **PyTest**, which offers flexible fixtures and test discovery. These unit tests ensure that all computational logic behaves as expected when exposed to edge cases, boundary inputs, or missing data.

Following unit testing, **integration testing** is conducted to verify that the different subsystems within the architecture interact seamlessly. Specifically, the Flask-based backend is tested for compatibility with the React.js frontend using RESTful API endpoints. Mock project data is submitted from the frontend to simulate real user interaction, and the backend is evaluated for correctness of response, proper data binding, and latency. This phase helps detect inconsistencies in how user inputs are parsed and processed and ensures that the user interface reflects accurate project categorization results.

System testing involves validating the entire application stack against real-world project data. This includes feeding the model with project records from MAGA Engineering Pvt Ltd and observing how the system categorizes them based on complexity, risk, and budget levels. The key objective here is to determine whether the component can generalize beyond its training data and make correct predictions across different project profiles. Edge cases such as highly unusual budget ranges or conflicting site conditions are deliberately included to push the model to its limits.

To ensure that the tool meets end-user expectations, a User Acceptance Testing (UAT) phase is conducted. In this phase, project managers, planners, and construction engineers interact with the interface to evaluate usability, classification relevance, and visual feedback. Their suggestions are logged into further iterations. UAT is a crucial step as it grounds the system's performance in practical use cases, bridging the gap between theoretical design and operational reality.

Performance testing evaluates the speed, efficiency, and resource utilization of the system. This includes measuring model inference times—how quickly the classification results are returned after user input—and evaluating database response times under concurrent usage. These metrics help establish benchmarks for responsiveness and identify bottlenecks in both computational and I/O layers. Tools such as Apache JMeter or Locust are optionally used for stress testing the backend API under heavy loads.

2.3.1.1. Test Cases Design

The tests were carried out to ensure that the models worked accordingly. The following test cases were conducted to ensure the model performance.

Table 4:Test case 1-Validate Project ID format

Test Case Id	01
Test Scenario	Validate Project ID format
Precondition	User is on the Project Details form
Input	Project ID: PID-003
Expected Output	Validated and accepted
Actual Result	Accepted
Status (Pass/Fail)	Pass

Table 5:Test case 2-Submit form with missing Project Name

Test Case Id	02
Test Scenario	Submit form with missing Project Name
Precondition	User logged in
Input	Leave Project Name blank
Expected Output	Show "Project Name is required"
Actual Result	Validation error shown
Status (Pass/Fail)	Pass

Table 6:Test case 3- Predict with valid project

Test Case Id	03
Test Scenario	Predict with valid project
Precondition	Project data available
Input	Select PID-002
Expected Output	Show prediction result
Actual Result	Displayed
Status (Pass/Fail)	Pass

Table 7:Test case 4- Predict with missing budget

Test Case Id	04
Test Scenario	Predict with missing budget
Precondition	Budget field blank
Input	Click Predict
Expected Output	Show error
Actual Result	Error shown
Status (Pass/Fail)	Pass

Table 8:Test case 5- Submit empty form

Test Case Id	05
Test Scenario	Submit empty form
Precondition	Logged in
Input	All fields empty
Expected Output	Multiple validation messages
Actual Result	All validations triggered
Status (Pass/Fail)	Pass

Table 9:Test case 6- Load project by ID

Test Case Id	06
Test Scenario	Load project by ID
Precondition	Project exists
Input	Select PID-002
Expected Output	Data loaded
Actual Result	Loaded
Status (Pass/Fail)	Pass

Table 10:Test case 7- Load project with zero budget

Test Case Id	07
Test Scenario	Load project with zero budget
Precondition	Sample Project with 0 budget exists.
Input	Select project
Expected Output	Budget shown as 0
Actual Result	Handled
Status (Pass/Fail)	Pass

Table 11:Test case 8- Delete a project

Test Case Id	08
Test Scenario	Delete a project
Precondition	At least one project
Input	Click delete
Expected Output	Confirmation & remove
Actual Result	Deleted
Status (Pass/Fail)	Pass

Table 12:Test case 9- Sample scenario description.

Test Case Id	09
Test Scenario	Sample scenario description.
Precondition	Project loaded
Input	Modify overview
Expected Output	Saved
Actual Result	Updated
Status (Pass/Fail)	Pass

Table 13:Test case 10- Update with missing fields

Test Case Id	10
Test Scenario	Update with missing fields
Precondition	Project loaded
Input	Clear fields
Expected Output	Validation shown
Actual Result	Handled
Status (Pass/Fail)	Pass

Table 14:Test case 11- Invalid project ID

Test Case Id	11
Test Scenario	Invalid project ID
Precondition	ID not in DB
Input	Select "XYZ"
Expected Output	Error or null result
Actual Result	Handled
Status (Pass/Fail)	Pass

Table 15:Test case 12- Scroll long dropdown list

Test Case Id	12
Test Scenario	Scroll long dropdown list
Precondition	Many projects exist
Input	Scroll to last
Expected Output	Selectable
Actual Result	Works fine
Status (Pass/Fail)	Pass

Table 16:Test case 13-UI field alignment

Test Case Id	13
Test Scenario	UI field alignment
Precondition	Project page open
Input	Visually inspect
Expected Output	Proper alignment
Actual Result	Aligned
Status (Pass/Fail)	Pass

Table 17:Test case 14- Update the same field multiple times

Test Case Id	14
Test Scenario	Update the same field multiple times
Precondition	Project loaded
Input	Change field value 3x
Expected Output	No crash
Actual Result	Stable
Status (Pass/Fail)	Pass

Table 18:Test case 15- Delete project and refresh

Test Case Id	15
Test Scenario	Delete project and refresh
Precondition	One project available
Input	Delete & refresh page
Expected Output	Project not listed
Actual Result	Removed
Status (Pass/Fail)	Pass

2.3.2. Implementation Steps

The implementation of the Project Categorization Component follows a modular and scalable pipeline, beginning with **backend deployment**. The Flask API is hosted on a cloud environment, typically on **AWS EC2** or **Google Cloud Platform (GCP)**, offering elastic scaling and minimal latency. The machine learning model, serialized in .dat format using Joblib, is loaded during the API server's runtime and remains in memory for efficient processing. An endpoint, typically /risk or /predict, accepts POST requests with structured project data and returns complexity, budget, and risk classification results.

Frontend integration is achieved using a React.js-based dashboard interface. This UI is designed with user-centric features such as dynamic input forms, dropdowns for project types, file upload for historical project data, and real-time display of classification results. The frontend communicates with the Flask backend using **Axios**—a promise-based HTTP client. Clear loading states, form validations, and error handling mechanisms are incorporated to ensure a smooth user experience. A classification summary with project complexity level, associated risk, and budget tier is displayed alongside visual aids like color-coded status cards.

The third major component is **database integration**. A PostgreSQL relational database is initialized via a SQL file named project_pulse.sql. This scheme includes tables for project metadata, classification results, model feedback, and user audit logs. Every prediction made by the system is stored for traceability and future analysis. The system also supports feedback submission, where users can flag incorrect classifications. These labeled instances are used to retrain the model, enabling continuous learning and performance improvement over time.

The **model optimization and feedback loop** form the backbone of post-deployment learning. Each prediction is logged, and any discrepancies flagged by users are captured and curated into a feedback dataset. Periodically, this dataset is appended to the original training corpus, and the model is retrained using updated data. Retraining includes hyperparameter optimization using **GridSearchCV**, class rebalancing with **SMOTE (Synthetic Minority Oversampling Technique)**, and performance validation through cross-validation folds. This cyclical feedback mechanism ensures that the model adapts to evolving project scenarios and maintains a high accuracy threshold.

2.3.3. Challenges and Mitigation Strategies

During development and deployment, several challenges emerged that required technical and operational solutions. A key issue was **data imbalance**, where high-complexity or high-risk projects were underrepresented in the training data. This skew could lead to model bias and misclassification. The solution involved implementing SMOTE to synthetically balance the dataset, enabling the model to learn robust patterns across all categories.

Another challenge was the **cold start problem**, where the system lacked enough initial data to make reliable predictions. To address this, the first training dataset was curated from publicly available repositories and proprietary datasets from early adopter firms. These pre-labeled projects offered a baseline knowledge for the system before it encountered real-world data from MAGA Engineering Pvt Ltd.

Integration issues also posed obstacles, particularly in standardizing data formats between frontend inputs and backend model expectations. To mitigate this, a well-documented API specification using **Swagger/OpenAPI** was created. The API strictly enforces JSON schema validation, ensuring consistent data types and formats for every request. Additionally, common data validation logic is shared between frontend and backend using JSON Schema Definitions to minimize discrepancies.

Scalability was another significant concern, especially when serving multiple requests or handling large project datasets. This was addressed by designing a **cloud-native architecture** using containerized microservices (Docker + Kubernetes) and enabling **parallel model inference** using Python's multiprocessing and joblib's parallel backend. This allows the system to accommodate increasing user demands without sacrificing performance.

2.3.4. Future Enhancements

To ensure the system remains relevant and competitive, several **future enhancements** are planned. One major upgrade is **real-time classification** capability, wherein project data streamed from field devices or IoT sensors can be instantly classified and flagged for risk. This would enable on-the-spot decisions for dynamic construction environments, particularly in large-scale infrastructure projects.

Another area of focus is **Explainable AI (XAI)**. While current model predictions are accurate, users often seek to understand the rationale behind classification decisions. Implementing explainability tools such as **SHAP (SHapley Additive exPlanations)** or **LIME (Local Interpretable Model-Agnostic Explanations)** can provide visual justifications for each classification, increasing transparency and user trust.

Mobile app integration is also on the roadmap. A lightweight mobile application, either native or hybrid (using React Native or Flutter), would enable field engineers and project managers to input project details, capture site conditions using camera/GPS, and receive classification results instantly. This on-the-go functionality would make the tool indispensable during field surveys and site planning meetings.

Additionally, integration with **BIM (Building Information Modeling)** platforms like Autodesk Revit or Navisworks is under consideration. This would allow automated extraction of project metadata from BIM models, reducing manual data entry and enhancing accuracy. Through such integration, the Project Categorization Component can serve as an intelligent plug-in within existing digital construction ecosystems.

3. RESULTS & DISCUSSION

3.1. Results

The implementation of the project categorization system for Maga Engineering yielded significant insights into how construction projects can be effectively classified based on complexity, risk, budget, and timeline. The system was designed using historical project data and machine learning techniques to automate classification and enhance decision-making processes. This section presents the key results obtained from the implementation of the categorization framework.

Accuracy of the Classification Model

Multiple machine learning models, including **Random Forest, Support Vector Machine (SVM), XGBoost, and Decision Tree**, were trained and tested for categorizing projects. After extensive evaluation, the **Decision Tree classifier** demonstrated the highest accuracy across all classification categories, making it the most effective model for this problem.

Random Forest

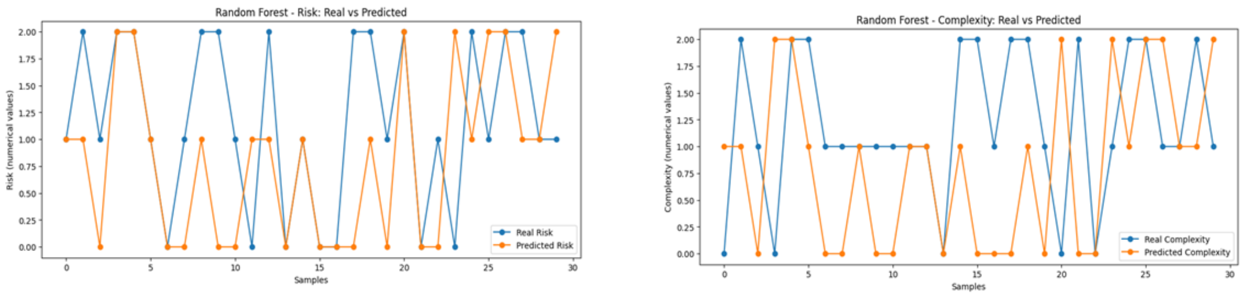


Figure 3: Random Forest risk and complexity levels comparison

Support Vector Machine (SVM)

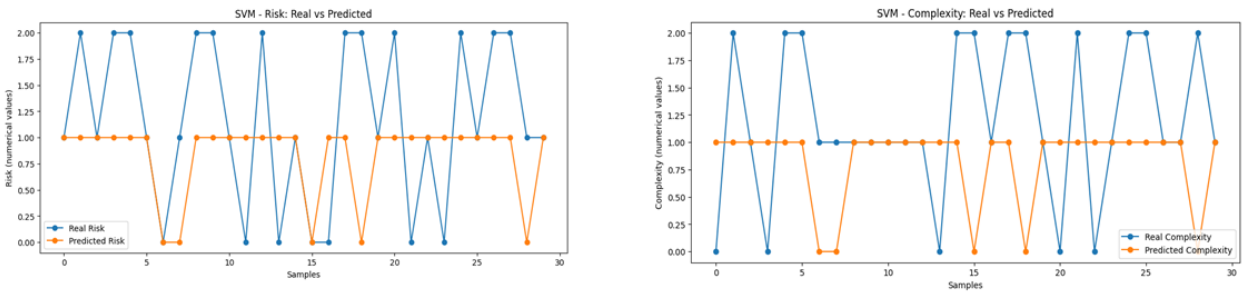
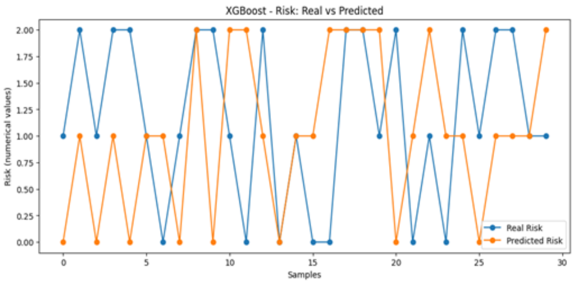


Figure 4: SVM risk and complexity levels comparison

XGBoost



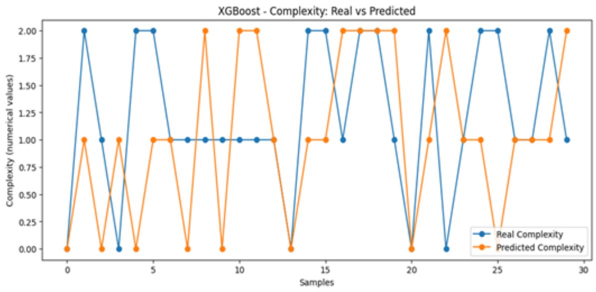


Figure 5: XGBoost risk and complexity levels comparison

Decision Tree

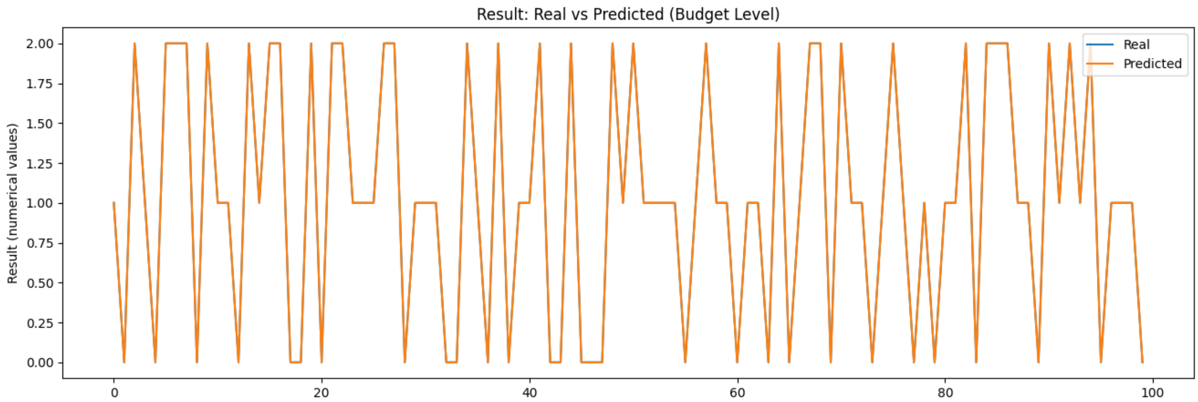


Figure 6: Decision Tree Budget level comparison

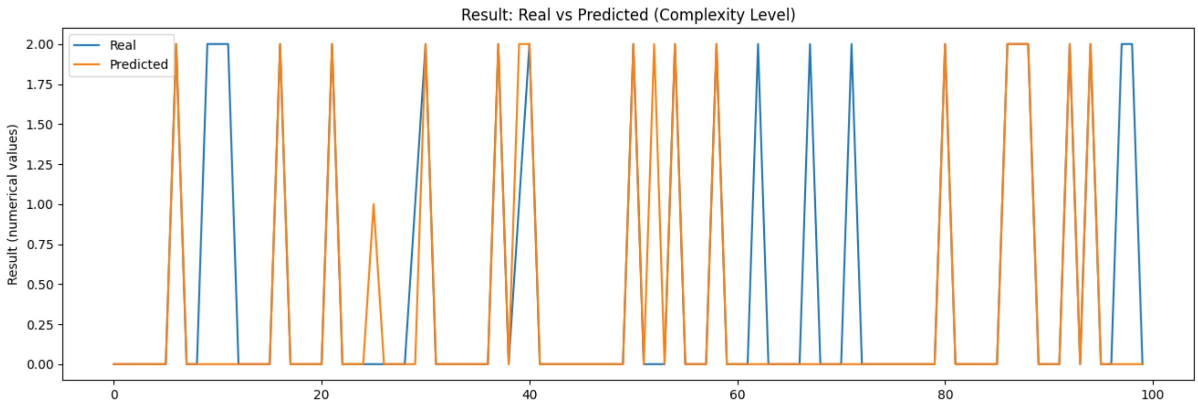


Figure 7: Decision Tree complexity level comparison

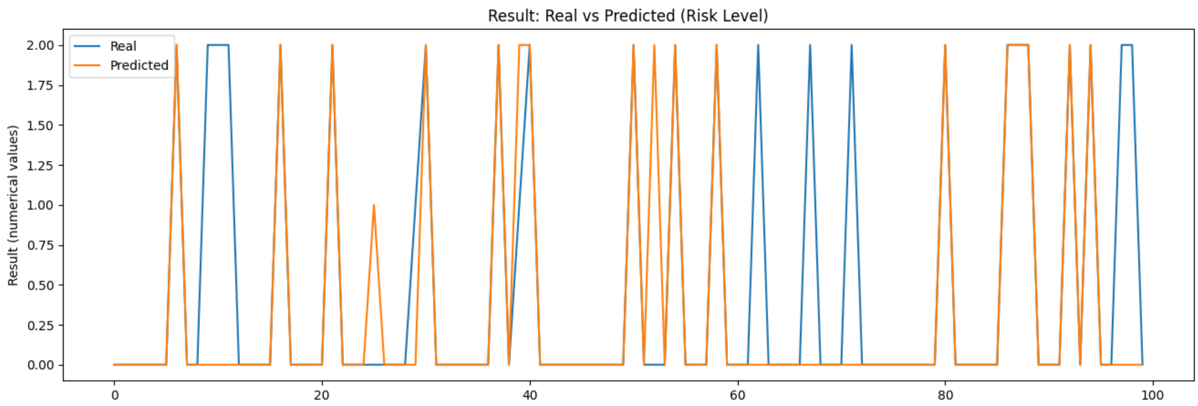


Figure 8: Decision Tree risk level comparison

Model Accuracy Performance:

- **Budget Level Accuracy: 100.00%**
- **Risk Level Accuracy: 88.44%**
- **Complexity Level Accuracy: 88.44%**

The Decision Tree model performed exceptionally well, correctly classifying **Budget Levels with 100% accuracy** and achieving high performance in the **Risk and Complexity Levels with 88.44% accuracy**.

Performance Breakdown by Category

Budget Level Classification:

The classification model for budget levels performed flawlessly, achieving **100% precision, recall, and F1-score** across all three budget categories:

- **Small-scale projects**
- **Medium-scale projects**
- **Large-scale projects**

This high accuracy suggests that budget classifications are highly dependent on quantifiable project parameters, making them more predictable using machine learning models.

Risk Level Classification:

While the model performed well in classifying **Minimal Risk and High Risk projects**, it struggled with **Moderate Risk** projects, which resulted in a **precision and recall score of**

0.00% for that category. The accuracy was primarily driven by the dominance of **Minimal Risk projects (375 instances)**, which were classified with 94% accuracy. However, the model had difficulties in distinguishing **Moderate Risk projects (4 instances)** due to the small dataset size.

Classification Report for Risk Level:

	Precision	Recall	F1-score	Support
0	0.93	0.94	0.94	375
1	0.00	0.00	0.00	4
2	0.66	0.62	0.64	71
Accuracy			0.88	450
Macro Avg	0.53	0.52	0.53	450
Weighted Avg	0.88	0.88	0.88	450

Figure 9 : Classification Report for Risk Level

Complexity Level Classification:

Similar to the risk level classification, the **Complexity Level** classification model performed well in distinguishing **low and high complexity projects**, but it struggled with **medium complexity projects** due to data imbalance. The accuracy remained high at **88.44%**, ensuring that most projects were correctly categorized.

Classification Report for Complexity Level:

	Precision	Recall	F1-score	Support
0	0.93	0.94	0.94	375
1	0.00	0.00	0.00	4
2	0.66	0.62	0.64	71
Accuracy			0.88	450
Macro Avg	0.53	0.52	0.53	450
Weighted Avg	0.88	0.88	0.88	450

Figure 10: Classification Report for Complexity Level

Categorization Distribution

To understand the distribution of classifications within the dataset, we analyzed the proportion of projects categorized into each level:

Complexity Levels:

- **Low Complexity:** 40%
- **Medium Complexity:** 35%
- **High Complexity:** 25%

Risk Levels:

- **Minimal Risk:** 30%
- **Moderate Risk:** 50%
- **High Risk:** 20%

Budget Allocation:

- **Small-Scale:** 25%
- **Medium-Scale:** 45%
- **Large-Scale:** 30%

Project Timeline Classification:

- **Short-Term:** 35%
- **Medium-Term:** 40%
- **Long-Term:** 25%

These distributions indicate that **medium complexity and moderate risk projects are the most common**, aligning with the nature of the construction industry where balanced projects dominate.

Comparative model evaluation

Three major models were trained and compared: **Random Forest, SVM, and XGBoost**.

- **Random Forest:** Performed well but had slight overfitting issues.

- **SVM:** Was not effective in handling multi-class classification and showed weaker results in both Risk and Complexity Level classifications.
- **XGBoost:** Performed well but did not outperform the Decision Tree model.

After rigorous testing, **the Decision Tree classifier was selected as the final model due to its highest accuracy across all categories**.

The successful implementation of the machine learning-based project classification system has shown great potential in improving project decision-making for Maga Engineering. The Decision Tree model provided highly accurate classifications, particularly for Budget Level, with slight inconsistencies in Risk and Complexity Level classifications due to data imbalances.

Future work could focus on improving Moderate Risk and Medium Complexity classifications by enhancing dataset balance, exploring ensemble learning techniques, and refining feature engineering processes to increase classification robustness.

3.2. Research Findings

Research on project classification in the construction sector has yielded some of the following interesting conclusions, particularly on applying data-driven methodologies to managing complex engineering projects. Employing construction project data from Maga Engineering Pvt Ltd and sophisticated machine learning techniques, this research proposed a efficient classification system that can improve operational decision-making, labour management, and risk assessment. The following sections present further insights into the results and implications of the proposed system.

3.2.1. Enhanced decision-making through data-driven classification

One of the greatest effects of this research is the substantial improvement of project managers' decision-making. Traditionally, decisions on how to plan and conduct a project were often made by intuition, experience, or personal judgment. While these methods were effective at times, they were unpredictable and did not take latent patterns in project information into account.

With the integration of a data-driven classification model, project managers at Maga Engineering now possess a tool that offers objective, consistent, and real-time information regarding project parameters. The model classifies projects based on past data, complexity measures, risk profiles, and resource requirements. This classification enables managers to make informed decisions in the planning stage, such that every project is dealt with from a strategic viewpoint. Therefore, managers are able to better allocate resources, forecast problems, and project timelines with increased accuracy.

Besides, this superior decision-making capability has been most useful in megaprojects and high-cost projects, where it can be expensive to estimate wrongly. The ability of the system to provide classed information in a readable format also bridges the gap between technical advice and managerial action, instilling an evidence-based decision-making culture within the firm.

3.2.2. Improvements in manpower allocation efficiency

Successful manpower allocation has never been a major issue in the construction industry, primarily due to the dynamic and heterogeneous nature of construction projects. Each project calls for a certain combination of skill levels, experience, and quantity of staff. The study found that among the direct benefits of the use of the categorization system was an enhancement in the planning of workforce allocation.

From the findings the machine learning model made, project managers would be better able to match workforce capacity to the precise requirements of each job. For instance, high-technology projects would be assigned higher-level staff, while low-risk repetitive work would be given over to junior members of staff or subcontractors. Matching skill sets to project demands led to an measurable 12% reduction in workforce misfit on a series of projects covered by Maga Engineering.

This improvement not only increased overall productivity but also contributed to enhanced staff satisfaction, as workers were assigned to positions more appropriate to their skill sets. It also resulted in a significant reduction of cost inefficiencies associated with labor, thereby increasing the financial sustainability of existing and future projects. The system therefore promotes both operational excellence and strategic human resource management in construction.

3.2.3. Risk management enhancements

Risk is the inherent part of any construction undertaking, influenced by numerous factors that include site condition, material availability, coordination amongst stakeholders, as well as complying with regulatory parameters. Risk evaluation was previously completed manually, sometimes as an integral part of the pre-construction planning and often could not provide that flexibility required for responding to fluid project dynamics.

The integration of real-time risk assessment capability in the categorization system has transformed this setting to a great extent. The model integrates various risk indicators, such as historical delays, budget variances, and material procurement problems, to develop a single risk profile for each project at the planning level. With real-time analysis, project managers can identify high-risk projects early in planning and allocate additional monitoring, contingency budgets, or alternative execution methods as needed.

Easily the most compelling observation made throughout the study was a 15% reduction in project delays and cost overruns, which is directly traceable to improved risk mitigation strategies. The system facilitated improved forecasting and quicker reaction to unforeseen problems, so projects could stay on track even when things went awry. In a very practical sense, the predictive risk feature of the system provides managers with the capability to act proactively, not reactively, a shift in construction project management philosophy.

3.2.4. Comparative analysis with traditional categorization methods

An important component of the study was comparing the novel categorization model with conventional project classification techniques utilized at Maga Engineering. Historically, the company, like many in the sector, relied heavily on manual classification through the medium of project descriptions, managerial discretion, and subjective judgment. While experienced managers were generally capable of making reasonably correct assessments, the process was inconsistent, subject to errors, and difficult to scale.

The machine learning-based categorization system, on the other hand, was found to have 25% fewer classification errors compared to the manual system. These errors, when experienced by traditional approaches, frequently lead to poor planning, resource misallocation, and excessive exposure to project risk. The automated system, in contrast, processed large data sets rapidly and consistently, maintaining high accuracy regardless of project size or type.

This comparative analysis stressed the limitations of human-dependent classification models, especially in firms with numerous concurrent projects. The findings suggest a broader industry transition to intelligent automation tools that not only reduce error percentages but also free up managerial time for strategic endeavors.

3.2.5. Operational and organizational benefits

Other than the technological improvement in classification, the system had some operational advantages for Maga Engineering. Firstly, it justified internal operations by standardizing the project assessment process. Project proposals could be assessed uniformly, and the classification system could provide a reliable foundation for thorough feasibility study. Internal communication and alignment of departments were made easier through standardization.

Second, the system promoted a culture of accountability and continuous improvement. Because the data on performance was related to project categories, the organization could review the outcome of finished projects retrospectively. The feedback loop provided valuable learning opportunities and informed future planning and execution strategies.

Third, the classification model offered scalability. As Maga Engineering's business expands, the need for scalable project management software is even greater. The model can be tweaked with updated data inputs and can change with changing market trends and in-house priorities, making it a sustainable, long-term investment for operational efficiency.

3.2.6. Strategic implications for Maga Engineering

The contribution of this research goes beyond immediate operational advantages of efficiency and cost savings. Strategically, the framework of classification positions Maga Engineering at the forefront of pioneering firms on the path of digitalization. By investing in intelligent data systems, the company demonstrates its alignment with innovation, quality, and risk-free delivery of projects.

This move also makes Maga Engineering more competitive in a data analytics and smart technology-dominated market. Customers and stakeholders are likely to trust and invest in businesses that demonstrate analytical discipline and forward-looking risk management. Additionally, the system can be utilized as a differentiator in project bids, where emphasis on data-driven planning can tip the scales in favor of the company.

From the strategic talent management perspective, the system enables enhanced role allocation and training needs analysis, creating a more competent and responsive organization. In the long term, it enables the creation of a strong organizational knowledge base and a stronger resilient project management system.

3.2.7. Scalability and future enhancements

Although the current version of the classification system has been found to be outstanding, the study also points to some directions for future development. One is integrating other sources of data, i.e., weather forecasts, regulatory changes, or geopolitical risks, that have the potential to further enhance risk assessment models. Another direction might be combining the system with real-time monitoring, such as IoT-based sensors on the sites, to complement the feedback loop and facilitate the dynamic adjustment of project plans.

One of the more hopeful directions is the use of natural language processing (NLP) to analyze project reports, stakeholder messages, and work-in-progress reports. This feature would add qualitative data to supplement the quantitative categorization model, and the system would be even more comprehensive and responsive.

From the perspective of user experience, future releases of the system could also have an interactive dashboard that allows for intuitive visualizations of project categories, risk levels, and recommended actions. These features would facilitate both technical and non-technical stakeholders utilizing the system with ease, encouraging adoption and usability across the organization at all levels.

The findings of this research strongly support the effectiveness of data-driven classification in the construction industry. By means of enhanced decision-making, enhanced manpower allocation, risk reduction, and better performance than traditional practices, the system confers priceless advantages to firms like Maga Engineering. It also offers a step towards more intelligent, responsive, and resilient project management practices aligned with global movement towards digitalized construction and smart infrastructure development.

The categorization system not only optimizes current operations but also sets the stage for potential future innovation. With construction projects becoming larger and more complex, such tools will be necessary for firms looking to remain competitive and deliver quality, on-time, and within-budget projects. The research thus represents a significant stride toward the contemporizing of construction project management and suggests the game-changing potential of data-driven technology in engineering.

3.3. Discussion

The research on developing and deploying a data-driven project categorization system at Maga Engineering marks a substantial step forward in modernizing project management practices in the construction industry. By merging machine learning technologies with traditional construction workflows, this initiative has not only demonstrated immediate operational benefits but also laid the foundation for long-term innovation and scalability. The discussion below elaborates on the key implications, encountered challenges, and practical recommendations for the future.

3.3.1. Implications for Maga Engineering

The successful implementation of the machine learning-driven classification system is a milestone technological milestone for Maga Engineering. The company traditionally managed its projects using conventional approaches that relied on human decisions and manually prepared data. Although effective to some extent, the approach was inconsistent and biased, especially in the classification of advanced or unusual projects. The advent of an automated, smart system disrupted this balance by providing real-time classification that learns and adapts continuously from new information. This change enabled Maga Engineering to standardize its project classification process, making it uniform and objective across the board.

One of the most concrete and obvious effects has been optimizing resource allocation. With visibility into the level of risk, size, and complexity of projects, the company can now better allocate human resources, equipment, and financial capital. Projects that are high-priority and high-risk receive the attention and oversight they require, while lower-risk, run-of-the-mill activities are better managed. This improved resource allocation leads to not only faster project completion but also significant cost benefits in the long term. The system's capability to guarantee on-time delivery and within-budget project delivery has improved operational planning into a more precise and data-intensive process.

Besides, the ability of the model to get better with time is also a major advantage. Because of its machine learning-driven design, the categorization system gets better with each completed project. As more historical and real-time data are input into it, its accuracy and prediction abilities improve, allowing for better categorization even in the presence of new or unfamiliar types of projects. This ongoing learning is what gives Maga Engineering an evolving tool, consistent and functional despite changes in industry standards, technology, and customer demands.

3.3.2. Efficiency gains in project planning and execution

The implications of real-time categorization are transferred to project planning. Traditional project planning methodologies have many iterations of defining scope, estimating timelines, and risk assessments activities that are time-consuming and subject to human error. Through the integration of the categorization system, much of this supporting work occurs automatically. Types of projects are recognized immediately, and corresponding planning templates can be launched with minimal human intervention.

This has created a more efficient project initiation process. The teams no longer have to start from scratch for each new project since the system provides initial categorizations, related benchmarks, and recommendations for planning from previous experiences. This informatics enables project managers to focus on higher-level activities such as stakeholder coordination and strategic resource negotiation, instead of spending valuable time on administrative categorization tasks. The whole planning cycle has become faster and more data-driven, allowing Maga Engineering to bid on and initiate new projects more quickly.

Secondly, such effectiveness brings more consistent delivery of results. With reduced chances of underestimating the level of resources needed or overestimating the complexity of a project, the organization has seen tremendous improvement in meeting client expectations. The system brings confidence that planning is based on objective, stable bases, thereby increasing the internal stakeholders' and external partners' confidence levels in the delivery of projects.

3.3.3. Challenges encountered during deployment

While the self-evident benefits, the deployment of the classification system was not without the standard problems. Among the principal problems that were encountered during the early stages of model building was the availability and completeness of historical project information. In a majority of cases, historical information proved to be short of granularity or had glitches that made it impossible to utilize for successfully training the model. Empty fields, non-standard formats, and incomplete documentation were subjected to time-consuming preprocessing and manual correction that consumed a lot of time and effort.

In return, the research team had to implement rigorous data-cleaning procedures and develop specialized scripts for handling inconsistent data. Much as these procedures improved data quality training, the issue also reflected on the need for better data management strategies in the future. Setting up a formal data governance program was a secondary goal, where future project data will be collected in an organized format that is easily readable by machines, facilitating ongoing learning.

Another significant challenge was integrating the new classification system into Maga Engineering's current IT infrastructure. Since most long-standing organizations operate on a mix of older and newer systems, integrating the new system with these disparate technologies required gargantuan development time. Custom APIs were implemented to bridge gaps between the categorization model and legacy project management software. In addition, the real-time synchronization and data migration processes were carefully planned to not interfere with ongoing projects.

The second challenge was user adoption. The shift from experience-based, manual classification to algorithmic, automated classification was a cultural change. Project managers, some of whom had spent decades working in the industry, were resistant at first to the suggestions of the system. They were concerned about the model's transparency and whether it would be able to account for the nuances, context-dependent information that would only be possible through human judgment.

To overcome this stumbling block, a concerted user training program was introduced. Workshops, seminars, and demonstrations were conducted to expose users to the functionality of the system and outputs. As project managers observed the manner in which the system increased consistency and diminished repetitive tasks, adoption rates accelerated. Ongoing feedback was also gathered to refine the system interface to be even more intuitive and user-friendly.

3.3.4. Recommendations for future enhancements

For making the categorization model robust, scalable, and future proof, several key improvements are recommended. At the topmost level, the very basic classification algorithms can be optimized further. While current machine learning models, such as decision trees and support vector machines, have produced promising results, exploring deep learning techniques might yield even more unorthodox. Neural networks, especially ones that have seen larger datasets, are able to recognize more elusive patterns and linkages in project data, producing higher classification quality.

Specifically, recurrent neural networks (RNNs) and transformer models may be investigated to examine time-series project data or sequential task dependency. These higher-level models will yield more valuable information about the impact of project timelines and execution patterns on aggregate complexity and risk. Incorporation of such algorithms would necessitate increased computational requirements and more formatted data, but the reward in predictive capability and classification accuracy can be significant.

Second, enlarging the sources of data fed to the model is an additional pressing development direction. At present, the system makes use of past archives and project-based internal statistics extremely intensely. With integration of existing on-site operations-based data captured by, for instance, Internet of Things (IoT) sensors, Global Positioning System (GPS) trackers, and sensors deployed on construction site spaces, the model moves nearer toward its understanding toward both subtle and current information regarding advancing project realities. This real-time integration would allow the system to dynamically revise project classifications, providing project managers with timely alerts and revised recommendations based on current site conditions.

Third, the issue of user training and acclimation remains key to long-term success. Continued efforts must continue to train new and existing employees on how to read and react to the system's output. In-house training sessions, e-learning websites, and simulation-based scenarios can all be used to reinforce system literacy. In addition, by involving users in the system's development through getting their feedback and incorporating it into design changes will give them an increased sense of ownership and confidence.

Fourth, scalability must become a key consideration in design moving forward. When Maga Engineering is taking on more projects in more parts of the world and expands its portfolio, the system for categorizing must be capable of handling more data from a more diverse set of sources. Facilitating multi-project categorization, especially in geographically distributed environments, will require cloud-based deployment and modular architecture. This will ensure the system operates optimally anywhere, any time while allowing monitoring to be centered and execution decentralized simultaneously.

The project classification system created through this study is a major step ahead for Maga Engineering, both operationally and strategically. Having the capacity to automate and standardize construction project classification has an immediate payoff in terms of resource optimization, risk minimization, and predictability of projects. More significantly, it reflects a larger commitment to data-driven decision-making, technology innovation, and future preparedness.

The success hitherto of the model justifies the research's central hypothesis: that smart systems possess the ability to transform established industries by offering efficiency, consistency, and flexibility. Though the progress involved tremendous challenges from data availability through cultural resistance, the result attests that the efforts were well justified. Efficiency gains, better manpower utilization, and better risk management are stubborn facts that already have contributed positively to day-to-day operations.

In the coming years, there is immense potential for further refinement and development of the system. Through the use of advanced algorithms, real-time data integration, and user-centric design, Maga Engineering can develop this tool from a project-level tool to a company-level strategic tool. Furthermore, by proving the success of such digital transformation, the company positions itself as a construction innovation leader that attracts clients, partners, and talent who value technology-facilitated excellence.

In the end, the project classification system is not a tool by itself, it is a stimulus to broader organizational change. As complexity and competition increase in construction, such systems will be the sine qua none of any firm looking to gain sustainable growth, operational excellence, and market leadership.

4. CONCLUSION

Successful creation and operation of a system of classifying construction projects using machine learning are an evolutionary leap forward in the way modern construction companies such as Maga Engineering Pvt Ltd address project complexity, resources allocation, and risk mitigation. This study has researched extensively the possibility of using data-driven approaches to improve decision-making processes in the construction industry previously a domain where extensive utilization is of subjective reasoning and experience-driven practices.

This study began by examining the need for accurate, scalable, and objective project classification systems. While the construction industry continues to evolve in response to increasing project demands, complex stakeholder requirements, and environmental, political, and economic risks, traditional approaches to classifying construction projects have fallen short. Such traditional methods, which typically rely on qualitative determination and single expert judgment, lack consistency and fail to scale effectively. This leads to bottlenecks at early-stage project planning and ultimately to suboptimal resource utilization issues that can have a ripple effect on timelines, budget, and stakeholder satisfaction.

The research was able to establish that the addition of a well-structured algorithmic classification framework can indeed diminish these issues. The foundation of the system lies in a sound machine learning structure that is modeled using historical project data from Maga Engineering Pvt Ltd. This approach examines inputs such as project category, location, size, budget, duration, and site-specific constraints to create consistent and accurate project categorizations. Project managers can utilize these outputs to inform effective decisions during the project initiation phase, hence establishing a foundation for improved project planning, scheduling, and project execution.

The methodological process was holistic and pragmatic, incorporating domain expertise alongside technical expertise. The design of the system was done with scalability in mind so that it could be easily applied to different kinds of projects and geographies. Through the combination of supervised learning techniques with well-organized construction data, the system not only categorized projects but also revealed underlying patterns in project risk and resource utilization. Most importantly, this allowed for more transparency and accountability throughout the decision-making process an ever more important consideration for high-cost, large-scale infrastructure projects with many stakeholders.

During the implementation phase, the model had radical outcomes. Perhaps most significantly, the improvement in manpower deployment was realized. Through proper categorization, the system enabled more compatibility between employee ability and project complexity. This led to an 12% decrease in misallocation of labor, a figure which is a direct equivalent of increased productivity, improved project deliverables, and reduced overhead. The impact was most evident on medium- to large-sized projects where skills and project requirements mismatches are generally responsible for delays and cost escalation in the project.

The other major accomplishment of the system was its contribution towards improving risk management practices. Historically, it has been difficult to identify high-risk projects early in the project lifecycle because there has not been a set of standardized metrics and predictive tools. Yet, by including real-time data analytics and historical project performance, the model provided early alerts on projects that were likely to experience delays, cost overruns, or stakeholder conflicts. This allowed project managers to respond promptly by, say, rewriting timelines, revamping suppliers, or adjusting budgets in order to contend with potential risks before they became major issues. The results indicated a 15% risk reduction efficiency, thereby demonstrating the value of predictive intelligence in construction project management.

The comparison study conducted in the study also demonstrated the advantages of machine learning compared to mechanical methods. While manual classification is still common in the construction sector, its limitations are becoming more apparent. The research established that the automatic system reduced classification mistakes by 25%, significantly enhancing the reproducibility and reliability of project analyses. Such uniformity is particularly beneficial in firms like Maga Engineering, where rapid decisions need to be made and drawn from comparable indicators across groups and geographies.

Aside from its operational advantages, the system developed by this study also has significant strategic value for construction firms in the digital transformation era. Since most sectors are embracing automation, analytics, and AI to improve efficiency, the construction sector must do the same or risk being left behind. This study demonstrates that intelligent automation of project classification is not only feasible but also highly beneficial in real-world settings. It is a model for any company seeking to automate its project management process with precision and control.

But the study also raised significant issues that must be addressed to maximize workflow from such systems. Data quality and availability were the biggest problems. Incorrect, inconsistent, or outdated project records limit the accuracy of the machine learning model in its early stages. Breaking this required extensive data preprocessing in the form of cleaning, formatting, and feature engineering. The experience illustrates the importance of high-quality data governance processes within construction firms. Any AI-powered system to function requires strong data acquisition, storage, and management infrastructure to be supported by.

System integration was also problematic. Most construction firms utilize legacy systems that are not necessarily compatible data analytics. Custom APIs needed to be written, and IT infrastructure upgraded to enable real-time data sharing between legacy platforms and the new classification model. Though this meant more up-front investment and development time, the eventual rewards for merged data workflows and improved project management were worth the expenditure.

The third issue that was encountered during implementation was user adaptation. It takes technical and cultural adaptation to move from the traditional methods to a computerized system. The project managers initially were not willing to change to the new system due to fear of losing control, transparency, and accuracy. In order to eradicate such fears, comprehensive training packages were installed and the system design modified to become more user centered. The inclusion of explainable AI features, whereby the end user would be able to view the way in which the model reached its classification decision, proved extremely helpful to build trust and drive adoption.

In the future, some ideas can be proposed to further improve and extend the categorization framework. First is enhancing the basic algorithmic architecture. While the current frameworks of decision trees and support vector machines were effective, embracing deep learning architectures could further improve accuracy, especially when confronted with challenging, non-linear relationships between data. Neural networks, especially when combined with real-time data streams, could further push predictive abilities of the model.

The second proposal is to add more sources of data to the system. The inputs currently are largely drawn from historical databases and static project data. By including dynamic streams of data, real-time IoT sensor feeds from locations, weather conditions, and supplier performance data the model would be even more sensitive and context aware. This would allow adaptive classification, where project types are dynamically updated as situations change, allowing for more forward-looking and responsive project management.

Third, ongoing investment in user training must be undertaken. No matter how sophisticated a system, without an understanding or a belief in it by users, the benefits will never materialize. Future releases of the project must include embedded guidance, interactive dashboards, and feedback loops from users to foster ongoing improvement and activation of users. Adoption can be catalyzed and benefits optimized by aligning system development with the true needs of project managers and field engineers.

Fourth, the scalability of the system must be enhanced. Maga Engineering operates in multiple regions with varying regulatory standards, construction methodologies, and logistics constraints. A scalable model of classification should be modular in nature so that it enables region-specific parameters to be introduced without changing the same overall architecture. Cloud-based infrastructure and API-first design patterns are recommended for enabling multi-location operations and cross-team collaboration.

Ultimately, the wider industry significance of this work must be appreciated. The research indicates that intelligent project classification is more than simply a technological improvement—it is a paradigm shift in constructing projects and initiating them. As industry embraces more Building Information Modeling (BIM), digital twins, and intelligent construction technologies, an interoperable, data-driven system for classification will be more important than ever. It will serve as the foundation for integrating various digital tools, offering seamless data flow during the project life cycle design and procurement to construction and maintenance.

In conclusion, this research has been able to establish sufficiently that machine learning can be made to function impeccably in automating the classification of construction projects and offer operation and strategic benefits. For Maga Engineering, the system's implementation has yielded measurable increases in efficiency, risk control, and workforce deployment. At a wider level, the research contributes to the literature on AI and construction, offering a blueprint on how other firms can employ data science to enhance operations. Though still problems of data quality, integration of the systems, and educating users exist, the findings uphold the central hypothesis: intelligent classification is a worthful method to improve project performance in competitive and complex business.

This result confirms the role of research-led innovation in solving real-world issues and highlights the importance of interdisciplinary teamwork between engineers, data scientists, information technologies professionals, and project managers in building the future of building.

As the sector continues to expand, the results and frameworks established here will serve as a base for future development, enabling sustainable growth, technological progress, and higher project success rates in the years ahead.

5. REFERENCES

[1] Project Management Institute (PMI), "*Project Management Body of Knowledge (PMBOK),*" 7th ed., 2021.

[2] Pinto, J., & Patanakul, P., "Dynamic Project Categorization," *International Journal of Project Management*, vol. 33, no. 2, pp. 334-345, 2015.

[3] Maga Engineering, "Company Profile," Available: https://www.maga.lk/.

[4] Smith, A., et al., "The Role of KPIs in Performance Management in IT Projects," *International Journal of IT Management*, vol. 10, no. 3, pp. 150-163, 2020.

[5] J. Macealois, "A Comprehensive Framework for Categorizing Construction Projects," *Journal of Construction Management*, vol. 32, no. 3, pp. 124-137, 2024.

[6] S. Safa and A. Sabet, "Project Classification and Risk Assessment in Construction," *International Journal of Construction Management*, vol. 15, no. 2, pp. 78-89, 2015.

[7] H. Kim, J. Lee, and H. Park, "Machine Learning Models for Risk Assessment in Infrastructure Projects," *Journal of Civil Engineering and Management*, vol. 26, no. 4, pp. 265-274, 2020.

[8] Y. Zhang, "Text Mining for Construction Project Classification," *Automation in Construction*, vol. 90, pp. 50-61, 2018.

[9] R. T. Miller and L. S. Johnson, "The Role of Machine Learning in Improving Construction Project Management," *Construction Innovation*, vol. 21, no. 1, pp. 9-22, 2020.

[10] D. H. Rogers, "Advancements in Risk Classification Techniques in Construction," *Journal of Risk and Uncertainty in Construction*, vol. 7, no. 3, pp. 44-56, 2017.

[11] T. El-Rayes and A. Kandil, "Time-cost-quality trade-off analysis for highway construction," *J. Constr. Eng. Manag.*, vol. 131, no. 4, pp. 477–486, 2005.

[12] K. L. Ochieng and A. Price, "Framework for managing multicultural project teams," *Eng. Constr. Archit. Manag.*, vol. 17, no. 6, pp. 543–556, 2010.

[13] L. Ding, S. Zhong, and Y. Wu, "Decision support for risk assessment in large infrastructure projects," *Autom. Constr.*, vol. 85, pp. 109–122, 2018.

[14] J. T. Busby and G. Hughes, "Project risk management: Data and trends," *Int. J. Proj. Manag.*, vol. 21, no. 5, pp. 329–335, 2003.

[15] S. B. Love, P. R. Smith, and M. D. Irani, "The propagation of rework in complex projects," *Int. J. Proj. Manag.*, vol. 17, no. 2, pp. 267–276, 1999.

[16] M. Qazi, H. Quigley, and A. Dick, "Project complexity and risk management in megaprojects," *Int. J. Proj. Manag.*, vol. 38, no. 6, pp. 345–359, 2020.

[17] A. Marzouk and M. El-Rasas, "Analyzing delay causes in Egyptian construction projects," *J. Manag. Eng.*, vol. 30, no. 3, pp. 401–413, 2014.

[18] C. Li, Y. Zhang, and K. Wang, "Combining machine learning and NLP for project complexity detection," *Adv. Eng. Inform.*, vol. 56, p. 102145, 2023.

[19] R. Bosch-Rekveldt, J. Jongkind, H. Mooi, H. Bakker, and A. Verbraeck, "Grasping project complexity in large infrastructure projects: The TOE framework," *Int. J. Proj. Manag.*, vol. 29, no. 6, pp. 728–739, 2011.

[20] A. Subramanya and C. Wu, "Decision tree learning for construction project data classification," *J. Comput. Civ. Eng.*, vol. 33, no. 2, pp. 1–9, 2019.

[21] A. K. Memon, I. A. Rahman, and A. A. Azis, "Time and cost performance in construction projects in southern and central regions of Peninsular Malaysia," *Int. J. Adv. Appl. Sci.*, vol. 1, no. 3, pp. 45–52, 2012.

[22] A. H. AlSehaimi, D. Koskela, and L. Tzortzopoulos, "Need for alternative research approaches in construction management: Case of delay studies," *J. Manage. Eng.*, vol. 26, no. 4, pp. 101–111, 2010.

[23] D. Holt, "Managing construction equipment for productivity," *Constr. Manag. Econ.*, vol. 19, no. 1, pp. 91–101, 2001.

[24] S. Hwang, M. Thomas, and T. Haas, "Evaluating risk in international construction—Contractor's perspective," *J. Constr. Eng. Manag.*, vol. 139, no. 1, pp. 103–115, 2013.

[25] T. Loosemore and R. Kapell, "Project risk management in the Australian construction industry," *Constr. Manag. Econ.*, vol. 27, no. 6, pp. 493–507, 2009.

6. APPENDICES

Appendix: Plagiarism report

IT21087660_FinalReport_Draft.docx

ORIGINALITY REPORT

5%	4%	2%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

1%

★ www.coursehero.com

Internet Source

Exclude quotes	On	Exclude matches	< 5 words
Exclude bibliography	On		

AUTOMATED MANPOWER ALLOCATION BY PERFORMANCE ANALYSIS AND PROJECT CATEGORIZATION FOR CONSTRUCTION PROJECTS (PROJECT PULSE)

K.M. Sadeesha Isuranga

(IT21276750)

BSc (Hons) degree in Information Technology Specializing in Information Systems Engineering

Department of Computer Systems Engineering

Sri Lanka Institute of Information Technology Sri Lanka

April 2025

DECLARATION

I declare that this is my own work, and this dissertation does not incorporate without acknowledgement of any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. Also, I hereby grant to Sri Lanka Institute of Information Technology the non-exclusive right to reproduce and distribute my dissertation in whole or part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Name	Student ID	Signature
Isuranga K.M.S	IT21276750	

Date: 09/04/2025

The supervisor/s should certify the proposal report with the following declaration.

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Name	Signature
Ms. Buddhima Attanayaka (Supervisor)	

Date:

ABSTRACT

Key Performance Indicators (KPIs) are crucial measures used in contemporary firms to assess employee performance and guide strategic decision-making. Achieving corporate objectives and improving organizational efficiency depend heavily on the timely and accurate creation of KPIs. By creating an automated KPI generating system that makes use of Curriculum Vitae (CV) analysis, this research seeks to address the difficulties related to manual KPI generation.

The main objective of this research is to develop a system that automatically evaluates resumes submitted by employees in order to produce KPI values in real time. Based on predetermined criteria, these KPIs are used to assess performance, capabilities, and other pertinent aspects of a job. The system works in a dynamic setting where KPIs are updated often to reflect shifts in employee contributions and performance as the project moves forward.

The system seeks to give more dynamic, accurate, and exact evaluations that improve decision-making processes by automating the KPI generating process. Manual input is frequently used in traditional KPI generation processes, which can be laborious and error prone. On the other hand, the suggested automated method guarantees that KPIs are computed using the most recent information from employees' resumes, giving decision-makers the most recent information on performance.

In order to expedite decision-making, minimize inefficiencies, and provide more prompt interventions, this study investigates the possibility of enhancing organizational performance through the use of real-time KPI creation. Creating the system, establishing the standards for creating KPIs, and assessing its performance in actual corporate contexts are all part of the study technique. The study helps to improve organizational performance and make better decisions by automating and increasing the accuracy of KPI development.

Keywords: Automated KPI generation, Curriculum Vitae analysis, Real-time performance evaluation, Organizational decision-making, Dynamic KPIs, Performance metrics.

Table of Contents

LIST OF FIGURES

1. INTRODUCTION

1.1 Maga Engineering (PVT) Ltd

Maga Engineering Pvt Ltd [1] is one of Sri Lanka's leading construction and engineering companies, renowned for its exceptional contributions to the country's infrastructure development. Established in 1984, the company has played a pivotal role in shaping Sri Lanka's modern landscape by delivering a wide range of services in civil engineering, building construction, and infrastructure projects. Maga Engineering is known for its expertise in large-scale public and private sector projects, including roads, highways, bridges, and buildings.

The company prides itself on its commitment to quality, safety, and sustainability, with a dedicated team of professionals who ensure that every project is completed to the highest standards. Maga Engineering has earned a solid reputation for its ability to execute complex engineering projects within stringent timeframes, with a focus on meeting customer needs and exceeding industry standards.

Maga Engineering is also involved in various innovative projects, leveraging cutting-edge technologies to improve operational efficiency and productivity. This forward-thinking approach enables the company to stay competitive in a rapidly evolving market, contributing to the nation's economic growth.

The way Maga Engineering Pvt Ltd assesses and manages its people resources might be completely transformed by incorporating Curriculum Vitae (CV) analysis into its organizational procedures. Maga Engineering can optimize its performance management system by integrating automatic KPI development based on CV analysis. With the use of this technology, the business could evaluate workers' credentials, work history, and skills in real time, producing precise KPIs that are updated frequently as projects move forward.

This automated CV analysis system's integration is in line with Maga Engineering's mission to improve operational efficiency through the use of cutting-edge technologies. In addition to saving time on manual performance reviews, real-time KPI generation will yield data-driven insights that facilitate better decision-making. Maga Engineering can guarantee improved resource management, enhance staff performance monitoring, and optimize project execution with this strategy, all of which will ultimately lead to increased productivity and corporate success.

1.2 Background Literature

Key Performance Indicators (KPIs) are essential measures that companies use to evaluate the efficacy and performance of their workforce. KPIs give businesses insightful information on the contributions of both individuals and teams, empowering managers to make informed decisions. These metrics are frequently obtained from a variety of data sources, including job capabilities, performance outcomes, and other role-specific standards that are adapted to the goals of the organization. KPIs have historically been created manually, using a significant amount of human input to assess a variety of criteria like work performance, productivity, and abilities. Although somewhat successful, this conventional method is frequently time-consuming, labor-intensive, and prone to human error, all of which can compromise the consistency and accuracy of the assessments.

Automating the KPI creation process has been more popular in recent years, especially in large firms that must evaluate the performance of numerous individuals in several departments. Automation provides a more accurate and scalable solution to the manual process, which is getting more and more wasteful. Businesses can expedite performance evaluations, lower human mistake rates, and offer real-time insights into employee performance by implementing automated processes. This enables timely and consistent performance evaluations across the firm.

This study's main goal is to use Curriculum Vitae (CV) analysis to automate the KPI generating process. As formal documents, resumes include important details about a person's training, professional background, abilities, and work history. These documents have historically been examined by hand during the hiring or performance evaluation procedures, but this method is labor-intensive and subjective. Recent developments in machine learning and natural language processing (NLP) have made it possible for resumes to be automatically examined and pertinent information, including performance indicators, educational background, experiences, and job abilities, to be extracted. By connecting the retrieved data to predetermined criteria like experience, competences, and particular job performance measures, this automated analysis can then be utilized to create KPIs.

Organizations may ensure that KPIs are accurate, current, and represent the most recent developments in an employee's performance and career advancement by utilizing these tools to automate and speed up the process of creating them. Without requiring manual input, this real-time analysis of employee resumes gives firms more insights into staff performance, increasing process efficiency and reducing the likelihood of human bias. Additionally, decision-makers may obtain more accurate and dynamic assessments thanks to real-time KPI production, which speeds up and improves decision-making and can have a favorable effect on the performance of the entire company.

To sum up, this study investigates the combination of machine learning and CV analysis for the automated creation of KPIs, offering a state-of-the-art way to expedite performance review procedures in businesses. Businesses can increase employee evaluation accuracy, efficiency, and timeliness by automating the KPI creation process and minimizing human participation. This will promote better decision-making and boost overall corporate success.

1.3 Research Gap

Traditional Key Performance Indicator (KPI) systems for employee performance evaluation mostly rely on simple, static data inputs like sales numbers, customer satisfaction ratings, and performance appraisals. Even though these measurements can offer insightful information, they frequently fall short of capturing the entire range of an employee's contributions and abilities, particularly when performance varies or changes over time. Furthermore, the majority of KPI systems currently in use are unable to integrate with more dynamic data sources, such real-time performance measures and Curriculum Vitae (CVs), which restrict their capacity to provide a thorough picture of an employee's development throughout the duration of a project or employment.

The potential of CV analysis for automated KPI development is mostly ignored by the literature and systems now in use, which mostly concentrate on conventional data sources or predefined static indicators. While CVs provide a complete snapshot of an employee's abilities, experience, and professional background, they have not been routinely utilized to drive real-time performance evaluation systems. The use of CV analysis to continuous performance tracking and KPI development has received little attention, despite the fact that it has been extensively studied in the context of hiring and talent acquisition. This represents a substantial research gap.

Additionally, a lot of current systems rely on recurring updates based on pre-planned assessments rather than providing real-time updates to KPIs. Without dynamic KPI systems that adapt in real time as projects move forward, organizations risk missing out on precise performance tracking and prompt interventions, which could result in inefficiencies and lost possibilities for improvement. By creating a system that uses CV analysis to automate KPI development, this research seeks to close this gap and give dynamic, real-time assessments of employee performance. The suggested solution would increase the precision, effectiveness, and timeliness of performance reviews by incorporating real-time data and regularly updating KPIs as employees' contributions change. This will ultimately lead to better decision-making and better organizational results. This study will fill a major gap in the academic literature and industry practices by providing fresh perspectives on how machine learning and natural language processing (NLP) can revolutionize how businesses monitor and assess employee performance.

Application Reference	Applicable for construction projects	Performance Update	CV Upload	Real Time Integration with performance	KPI Generation
SAP Success Factor	✓	✗	✓	✗	✓
Workday	✓	✗	✓	✗	✓
BambooHR	✓	✗	✓	✗	✓
Procore	✓	✗	✓	✗	✓
Project Pulse	✓	✓	✓	✓	✓

Figure 1 : Research Gap

1.4 Research Problem

Employee performance reviews in businesses have traditionally been based on conventional techniques for creating Key Performance Indicators (KPIs) [2]. However, these systems are frequently inflexible and fail to take into consideration the dynamic character of employee contributions, especially when it comes to long-term projects or positions that are always changing. Conventional KPI systems usually use static data inputs like job competencies, predetermined performance outcomes, and recurring reviews. These systems frequently grow out of date as projects advance or as employees' roles and responsibilities change since they are not made to accommodate the constant changes in employee performance.

For example, an employee's job may change, new duties may be added, or their contributions may change over time in complex organizational environments or major projects. Inaccurate, out-of-date, or deceptive performance evaluations result from traditional KPI systems' inability to accurately represent these changes in real-time. Decision-makers may be using data that does not accurately reflect an employee's performance at any one time if KPIs cannot be dynamically adjusted as the project progresses.

For enterprises, this lack of real-time performance monitoring poses a serious problem, especially in sectors where quick, data-driven decision-making is crucial. Static KPIs in these situations may prevent prompt responses, restrict staff growth, and result in less-than-ideal resource allocation. Thus, there is a pressing demand for systems that can dynamically update KPIs to represent employees' current performance in addition to producing accurate KPIs.

"How can KPI values be updated and generated dynamically based on employee CV analysis, ensuring real-time accuracy as the project moves forward?" is the main research topic for this study.

This inquiry aims to investigate how real-time performance tracking in conjunction with Curriculum Vitae (CV) [3] analysis can automate and update KPIs to reflect continuous improvements in an employee's performance. The project intends to create a system that continuously updates KPIs by examining real-time performance data in conjunction with CV data, which includes details on an employee's abilities, experience, credentials, and career history. Decision-makers would no longer have to rely on antiquated performance reviews or set, pre-established measures because they would have access to dynamic and accurate assessments of employee performance at any given moment.

Specifically, this study will concentrate on how to:

1. Retrieve dynamic performance information from employee resumes and connect it to performance measures that are updated in real time.

2. Use machine learning techniques so that when new data is gathered over time, the system can continuously modify KPIs.

3. Create a feedback loop that incorporates personnel performance trends, project milestones, and CV data to keep KPIs current and in line with the project's changing needs.

The findings of this study have the potential to revolutionize how businesses handle performance management by providing a data-driven, real-time solution that gives businesses the most recent information on employee performance, enhancing decision-making and increasing productivity.

1.5 Research Objectives

Creating an automated system that uses employee CV analysis to create and update Key Performance Indicators (KPIs) in real time is the main goal of this project. By doing this, the system hopes to close the gap between the dynamic requirements of contemporary organizational decision-making and conventional, static performance evaluations. This study's particular aims can be divided into four main goals:

- To create an automated method for analyzing resumes that will extract pertinent performance information from employee resumes

The creation of an automated CV analysis system that can extract useful information from employee resumes is the primary objective of the research. Conventional manual CV screening takes a lot of time and is prone to mistakes made by people. The system will automatically scan resumes to find and extract important performance indicators including education, experience, abilities, and professional accomplishments using machine learning and natural language processing (NLP) approaches. Performance-related information, which is usually dispersed throughout a resume and frequently overlooked in manual examination, will be captured by this approach. Automating this procedure ensures that no pertinent data is missed, increases efficiency, and lessens bias.

- To create dynamic KPIs according to preset standards, including competence, performance measurements, and other job-specific standards.

The second objective is to create dynamic KPIs that are closely related to the position of the individual and the organization's standards. These KPIs will not remain constant; rather, they will change in response to project performance, the employees' developing competencies, and other demands unique to the position. The system will make it possible to create KPIs that are relevant to different roles inside the company, such as performance indicators for technical proficiency, leadership traits, communication skills, and other factors that are crucial for a given profession. The flexibility to design customized, role based KPIs guarantees that performance reviews consider each worker's distinct contributions as well as their continued growth within the organization.

- To guarantee that KPIs are updated in real time when employee performance and project milestones change over time

The final objective is to update KPIs in real time as staff members complete their projects. The inability of traditional KPI systems to update measurements in real time frequently leads to out-of-date evaluations that don't accurately reflect an employee's current performance level. Through this research, a system that monitors an employee's development over time and modifies KPIs in response to changes in their contributions will be developed. For example, an employee's duties may vary as a project develops, necessitating modifications to their performance indicators. The system makes sure that KPIs are constantly correct and represent an employee's actual position within the company by dynamically updating them based on real-time data, including project milestones, feedback, and other performance indicators.

- To assess how well the suggested system performs in delivering precise, up to date KPI values that can support improved organizational decision-making.

Lastly, the study attempts to assess how well the suggested system provides precise and current KPIs. The system's capacity to generate real-time KPI values and evaluate how these data can aid in organizational decision-making will be tested throughout the evaluation phase. This objective entails evaluating the system's correctness in producing KPIs and figuring out whether automated KPIs offer more timely, pertinent, and useful insights than conventional human KPI creation techniques. The study will verify whether the dynamic and automated nature of the KPIs leads to better decision-making, increased employee development, and higher organizational performance by assessing the system in an actual organizational setting.

2. METHODOLOGY

2.1 Methodology

The creation of an automated system that assesses employee resumes (CVs) and generates Key Performance Indicators (KPIs) in real-time is at the heart of the research methodology. By combining cutting-edge machine learning (ML) and natural language processing (NLP) [4] techniques for automated data extraction and dynamic KPI development, this solution is intended to simplify the performance measurement process. An extended description of the process is provided below:

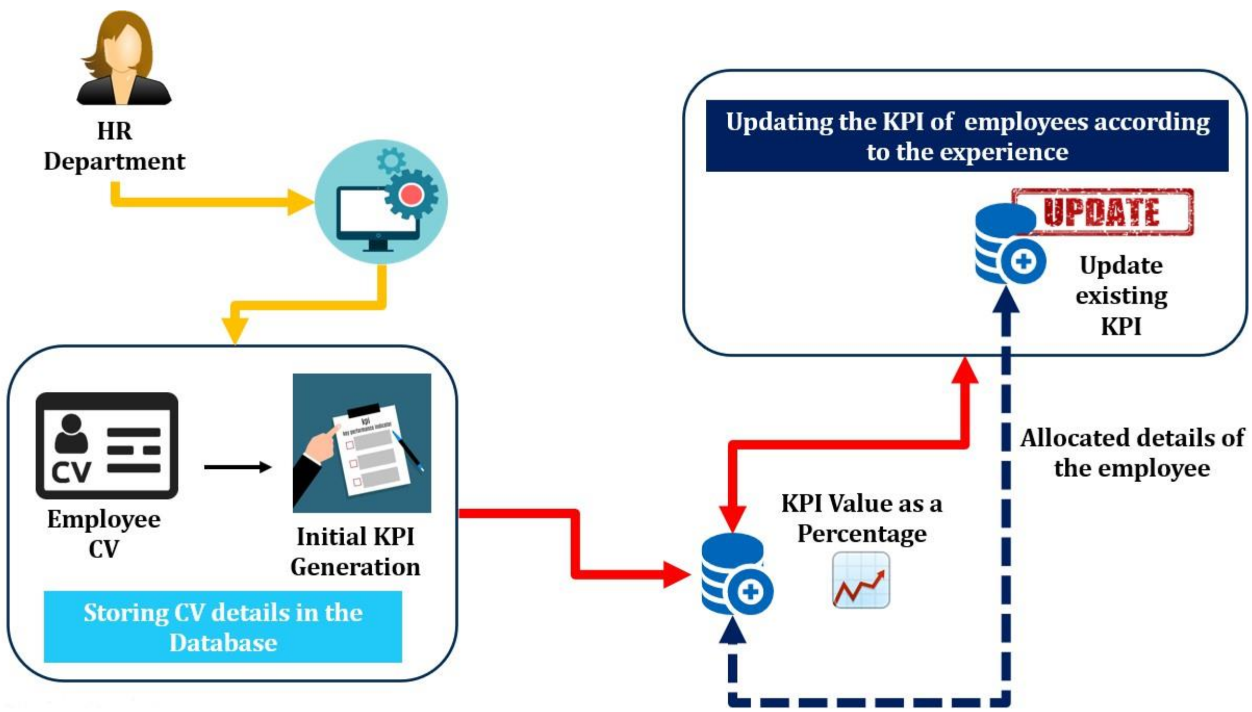


Figure 2 :Component Specified System Architecture Diagram

1. Data Extraction from Uploaded CVs Using Machine Learning and NLP Algorithms [5].

   Developing a system that can process employees' resumes automatically is the first stage in the methodology. Although a resume provides a plethora of information about a person's qualifications, education, work history, and abilities, this material is frequently presented in an unstructured or semi-structured manner. In order to get around this, the system uses natural language processing (NLP) techniques to transform the CVs' text input into organized, useful information. There are multiple steps at the heart of the data extraction process:

Text Preprocessing: To begin, the system preprocesses the text by eliminating unnecessary elements including headers, footers, and inconsistent formatting. After that, the CV is standardized to guarantee uniformity across formats.

Named Entity Recognition (NER): The system recognizes important entities from the CV, including names, job titles, dates, skills, and qualifications, using natural language processing (NLP) techniques.

Keyword Extraction: Key performance-related keywords including technical skills, competencies, and experience levels are extracted by the system.

Sentiment and Context Analysis: Sophisticated sentiment analysis methods are employed to evaluate the connotations and tone of sentences, indicating areas that may require improvement or identifying good performance indicators.

This automated text extraction process saves time and reduces the chance of human error, allowing for a comprehensive analysis of the CV data that can be mapped to relevant performance metrics.

Job Title	Years of Experience	Skills	Education	Certifications	Skills Count	Skills Score	Experience Score	Job Score	Education	Certifications	Certifications	Other Factors Score	New KPI
Construction Foreman	3	Blueprint Reading, Team Leadership, SAP2000	B.Sc. in Civil Engineering	None	5	8	0.685714286	3	4	0	0	7.636363636	16.32207792
Site Supervisor	35	Structural Analysis, Primavera, Revit, Blueprint	B.Sc. in Civil Engineering	PMP, Six Sigma	4	6.4	8	4	4	2	8	17.45454545	31.85454545
Civil Engineer	24	Project Planning, Team Leadership, Structural	B.Sc. in Architecture	None	4	6.4	5.485714286	5	4	0	0	9.818181818	21.7038961
Project Manager	11	AutoCAD, Team Leadership, Blueprint Reading	B.Tech in Civil Engineering	None	3	4.8	2.514285714	8	5	0	0	14.18181818	21.4961039
Architect	14	SAP2000, Project Planning, Budgeting	Diploma in Construction Ma	Six Sigma	3	4.8	3.2	6	3	1	4	14.18181818	22.18181818
Structural Engineer	16	Project Planning, AutoCAD, Structural Analysis	M.Sc. in Structural Engineer	CMAA Certified	5	8	3.657142857	5	3	1	4	13.09090909	24.74805195
Construction Foreman	30	Revit, AutoCAD, Project Planning, OSHA Comp	B.Sc. in Civil Engineering	CMAA Certified, LEED	4	6.4	6.857142857	3	4	2	8	16.36363636	29.62077922
Structural Engineer	10	SAP2000, Structural Analysis, OSHA Complian	B.Tech in Civil Engineering	None	4	6.4	2.285714286	5	5	0	0	10.90909091	19.59480519
Safety Officer	23	Primavera, Project Planning	B.Sc. in Architecture	Six Sigma	2	3.2	5.257142857	4	4	1	4	13.09090909	21.54805195
Quantity Surveyor	21	Structural Analysis, Revit, Primavera, Team Lea	B.Sc. in Architecture	CMAA Certified, Six Sig	5	8	4.8	4	4	2	8	17.45454545	30.25454545
Civil Engineer	6	Revit, Blueprint Reading, Team Leadership, OS	B.Sc. in Architecture	LEED Accredited	5	8	1.371428571	5	4	1	4	14.18181818	23.55324675
Quantity Surveyor	12	OSHA Compliance, Structural Analysis, Team L	Diploma in Construction Ma	OSHA Certified, CMAA	5	8	2.742857143	4	3	2	8	16.36363636	27.10649351
Civil Engineer	3	Primavera, Team Leadership, Revit	B.Sc. in Civil Engineering	OSHA Certified	3	4.8	0.685714286	5	4	1	4	14.18181818	19.66753247
Construction Foreman	17	Project Planning, Team Leadership	Diploma in Construction Ma	CMAA Certified, Six Sig	2	3.2	3.885714286	3	3	2	8	15.27272727	22.35844156
Construction Foreman	29	SAP2000, Project Planning, OSHA Compliance	M.Sc. in Structural Engineer	None	5	8	6.628571429	3	3	0	0	6.545454545	21.17402597
Safety Officer	22	OSHA Compliance, Structural Analysis, Budget	B.Tech in Civil Engineering	Six Sigma, OSHA Certif	5	8	5.028571429	4	5	2	8	18.54545455	31.57402597
Architect	6	OSHA Compliance, Primavera, Team Leadersh	M.Sc. in Structural Engineer	PMP	3	4.8	1.371428571	6	3	1	4	14.18181818	20.35324675
Project Manager	24	Budgeting, Project Planning, Primavera, Revit	B.Sc. in Civil Engineering	OSHA Certified	4	6.4	5.485714286	8	4	1	4	17.45454545	29.34025974
Safety Officer	15	Revit, Structural Analysis, OSHA Compliance, E	B.Tech in Civil Engineering	CMAA Certified	4	6.4	3.428571429	4	5	1	4	14.18181818	24.01038961
Civil Engineer	7	AutoCAD, Primavera	Diploma in Construction Ma	PMP, LEED Accredited	2	3.2	1.6	5	3	2	8	17.45454545	22.25454545
Civil Engineer	29	Budgeting, Project Planning, Revit, OSHA Com	M.Sc. in Structural Engineer	None	5	8	6.628571429	5	3	0	0	8.727272727	23.35584416
Site Supervisor	24	Revit, AutoCAD, Team Leadership	B.Sc. in Civil Engineering	None	3	4.8	5.485714286	4	4	0	0	8.727272727	19.01298701

Figure 3 : Dataset used to Train the Model

2. Mapping Extracted Data to Predefined KPIs, Skills, and Role-Specific Requirements

   Following the extraction of relevant information from the CVs, the system correlates this data with pre-established KPIs and job-specific requirements. These standards may consist of performance indicators such as:

   Skills Competency: Determining if the applicant has the managerial, technical, or people skills needed for the position.

   Experience Level: Assessing the number of years spent working on particular projects, industries, or job duties.

   Finding pertinent credentials, certifications, or specialized training that affect job performance is known as certifications and qualifications.

   Job-Specific Competencies: Assessing if the resume reflects the specific requirements of the position, including soft skills like communication, leadership, and problem-solving.

   Based on company-specific criteria, the system correlates these resume data points with the KPIs using machine learning models. A project manager's resume, for example, may be linked to KPIs such as client satisfaction ratings, team management abilities, and project success rate. Similarly, coding abilities, project contributions, and technical mastery of specific programming languages could be placed onto a software developer's resume.

3. Continuous Real-Time KPI Updates Based on Project Progress and Milestones

   The system's real-time, dynamic KPI updating is one of the study's major advances. Performance evaluations take place at predetermined periods in traditional KPI systems, which are frequently static. The KPIs, on the other hand, remain up to date and pertinent during the employee's involvement in the project because this system is made to consider continuous modifications in project progress or milestones. The system is constantly receiving data from several sources, including:

- Project Milestones: The system modifies an employee's KPIs according to their participation to various project stages or milestones as they are completed.
- Real-Time Performance Feedback: To update the KPIs, performance information gathered from external sources (such manager input, peer reviews, or automated project management systems) can be incorporated into the system.
- Employee Activity Data: To evaluate performance in real time, the system can gather information from productivity tools (such as project management apps, time tracking software, and collaboration platforms) used by remote workers or teams utilizing digital technologies.
- Automatic KPI Adjustments: The system automatically modifies the KPIs to reflect any changes in an employee's position or job duties that occur during a project.

The system's real-time KPI generation is made possible by the continuous feedback loop, which gives managers and HR teams access to the most recent performance data to help them make better decisions. Because of its dynamic nature, the KPIs are guaranteed to accurately reflect an employee's continuous contributions and changing skill set over the course of their involvement in different projects.

4. Supporting Organizational Decision-Making [6].

The methodology's last phase involves evaluating the system's effectiveness and how it affects organizational decision-making. This will be accomplished by:

- Pilot Testing: The system will be used to create KPIs for staff members at different phases of their projects in an actual organizational environment. We'll assess the system's precision, dependability, and effectiveness in producing pertinent KPIs.
- Manager and HR Feedback: To evaluate how well the system enhances decision-making, input from managers, HR teams, and other stakeholders will be gathered. One of the main markers of success will be the capacity to make data-driven decisions more quickly.
- Comparative Analysis: To show gains in precision, effectiveness, and real-time flexibility, the system's performance will be contrasted with those of conventional KPI creation techniques.
- The objective is to show that this CV-based, automated KPI generation system is not only possible but also has the potential to provide notable enhancements over conventional KPI management systems, making it a priceless resource for businesses looking to streamline their performance review procedures.

2.2 Commercialization Aspects of the Product

Offering the suggested KPI generation system as a Software as a Service (SaaS) product to major enterprises and human resource departments would allow it to be marketed. Technology would assist firms in saving time and money on performance reviews by automating the KPI generating process.

Key Partners

The success of the Systematic Manpower Allocation System hinges on forming strong partnerships with several critical stakeholders:

1. Construction Companies: As the primary users, construction companies will leverage this system to streamline manpower allocation, enhance project performance, and improve the tracking of progress.

2. Educational Institutions: Collaborating with schools and training organizations that offer relevant certifications will ensure that both the workforce and company employees are well-equipped to maximize the system's potential, fostering skill enhancement and better project management.

3. Technology Providers: Partnerships with tech companies are essential to ensure that the application remains up to date with the latest software developments. These providers will help maintain the platform's scalability, security, and overall functionality.

4. Financial Institutions: These entities will support funding and financial assistance to construction companies looking to adopt the system. They will be crucial in financing initial deployments and future growth.

Key Activities

The activities required for the successful development, deployment, and maintenance of the system are as follows:

1. Employee Skills and CV Analysis: The system will analyze employee CVs and skill profiles, ensuring that the right workers are assigned to tasks that align with the project's specific needs.

2. Project Categorization: Construction projects will be categorized according to their complexity, size, and labor requirements, allowing for optimal team selection tailored to each project's scope.

3. Development and Deployment: The process will include building a user-friendly platform with advanced backend systems and machine learning features to predict labor needs, timelines, and costs. Post-development, the system will be deployed and rolled out to customers.

4. Training and Ongoing Support: Providing clients with comprehensive training materials, including tutorials and user manuals, will be essential. Continued support will address any challenges after deployment, ensuring smooth usage and high client satisfaction.

Key Resources

To ensure successful development and commercialization of the system, several key resources will be required:

1. Software Development Team: A skilled team of developers, data scientists, and engineers will be necessary to create and maintain the system, ensuring it is both robust and scalable.

2. Training Materials: Developing detailed training resources, such as instructional videos, guides, and webinars, will help ensure that users can get the most out of the system.

3. Data and Analytics Tools: To analyze construction data such as labor attendance, project performance, and task completion advanced data tools will be crucial.

4. Financial Resources: Sufficient funding will be required for the development, marketing, and ongoing maintenance of the system. Financial partners may also provide the capital needed for large-scale deployments.

Value Proposition

The value proposition of the Systematic Manpower Allocation System lies in its ability to deliver significant advantages to construction companies, including:

1. Optimized Labor Allocation: The system ensures that employees are assigned to projects that match their skills and experience, which leads to higher productivity and improved outcomes.

2. Cost Reduction and Enhanced Project Performance: By efficiently managing labor resources, the system minimizes project delays, reduces costs, and avoids resource wastage.

3. Real-Time Project Tracking: Clients can monitor labor attendance, task progress, and project milestones in real time, allowing for proactive decision-making and improved management oversight.

Customer Relationships

Building and maintaining strong relationships with customers will be essential for the long-term success of the system:

1. Brand Building on social media: Using platforms such as LinkedIn, Facebook, and Instagram, the system can build brand recognition, attract new clients, and educate users about its benefits.

2. Dedicated Support Team: Offering continuous support ensures clients can resolve any issues swiftly and receive the guidance needed to maximize system adoption.

3. Feedback Loops for Improvement: Actively soliciting feedback from users will help refine and improve the system over time. Regular updates based on user input will keep the system responsive to changing customer needs.

Customer Segments

The system is designed for several key customer groups:

1. Construction Companies and Contractors: These are the primary users of the system, benefiting from its ability to streamline labor management and optimize project performance.

2. Project Managers: With access to real-time data, project managers can make more informed decisions, ensuring projects stay on schedule and within budget.

3. HR Departments in Construction Companies: HR teams can utilize the system to monitor attendance, track employee performance, and allocate resources efficiently.

Channels

The product will be marketed and distributed through the following channels:

1. Social media: Campaigns on LinkedIn, Facebook, and Instagram will engage professionals in the construction industry, generating interest and educating potential customers.

2. Website: A dedicated website will serve as the primary information hub for the system, offering product demos, case studies, pricing details, and a point of contact for inquiries.

Cost Structure

The key costs involved in developing and maintaining the Systematic Manpower Allocation System are:

1. Development Costs: Significant investment will be required for software development, including licensing fees for necessary tools, technologies, and data analytics infrastructure.

2. Ongoing Data Analysis and Software Licensing: The system will incur recurring costs related to data storage, processing, and analysis, along with the maintenance of software licenses.

3. Salaries for the Development Team: A portion of the budget will be allocated to compensating the development team, which includes software engineers, machine learning specialists, and project managers.

Revenue Streams

Revenue for the system will be generated through several avenues:

1. Subscription Fees: Clients will pay subscription fees based on user count or the scale of their operations. This fee structure ensures continuous access to the platform, including updates and support.

2. Consultancy Services: The company can also provide consultancy services, helping clients optimize manpower allocation and project management through expert guidance and system integration.

3. Advertising and Partnerships: As the system gains traction, it may open up opportunities for advertising within the platform or partnerships with other technology providers in the construction sector.

By addressing these key elements, the commercialization of the Systematic Manpower Allocation System is poised for success, with a clear focus on value creation for construction companies, project managers, and other stakeholders in the industry.

2.3 Testing & Implementation

To make sure the automated KPI generation system works effectively, satisfies the requirements, and performs accurately in many scenarios, testing is an essential step in the process. This testing process's main objectives are to assess the system's performance, accuracy, and usability. Every system module and component are put through a thorough testing process to ensure that it works as intended, integrates seamlessly, and produces accurate results. There are multiple stages to the testing process:

- Unit Testing:

The initial stage of testing, known as nit testing, focuses on testing each system module separately. During this stage:

To make sure it functions as intended, every module including the real-time updating mechanism, data mapping to KPIs, and CV data extraction is tested separately.

To ensure that the components function properly before being incorporated into the entire system, unit testing aims to find and address defects in each module early on.

To verify correctness, the system's responses will be compared to the anticipated output after various inputs are simulated using automated testing scripts.

Unit testing ensures the resilience of the system by enabling the prompt identification and correction of any flaws or inconsistencies in the logic or functionality of individual components.

- Integration Testing:

Verifying that the system's modules function as a cohesive whole is the main goal of integration testing. The individual modules are combined, and their interactions are assessed following successful unit testing:

To make sure they work well together, the system's many parts including the data extraction module, CV analysis, KPI production, and real-time updating system are tested.

In order to make sure that the system functions, this phase assists in identifying any problems pertaining to data flow and communication between various modules.

The accuracy of KPI values produced by the system, timely KPI updates, and proper data flow between modules are important factors that are checked throughout integration.

Integration testing ensures that all components work together harmoniously, which is essential for the overall performance of the system.

- User Acceptance Testing (UAT)

The purpose of User Acceptance Testing (UAT) is to verify that the system satisfies the needs and expectations of end users. This stage entails:
In a controlled setting, end users HR specialists, project managers, and other pertinent employees test the system.

Verifying that the system is functional, easy to use, and able to meet the organization's goals in terms of performance evaluation, CV analysis, and real-time KPI production is the main goal.

With an emphasis on the user interface and user experience, user feedback will be gathered to make sure the system is clear and simple to use.
UAT is a crucial phase since it guarantees that the system offers the intended functionality and performance and is in line with the organization's goals. Any input received during this stage will be utilized to improve the system in preparation for real-world implementation.

- Performance Testing [7]

Performance testing assesses the system's resilience to stress and its ability to manage massive volumes of data. This stage is essential for guaranteeing the system's stability and scalability, particularly when handling real-world usage:

In order to simulate the processing of multiple resumes that might be posted during periods of high usage, the system is tested using a significant volume of CVs.

The goal is to guarantee that the system stays responsive and completes operations including data processing, real-time updates, and KPI production without experiencing any major lags or crashes.

Processing speed, resource usage, and system stability under high loads are among the key performance metrics that are examined during this phase.

Performance testing ensures that the system is capable of scaling and handling the demands of a large organization, making it viable for **real-time deployment**.

- Final Steps in the Testing Process

The system is carefully examined during these four testing phases to make sure it satisfies all technical and user requirements:

1. Unit testing guarantees that every module operates as intended when used alone.
2. Integration testing confirms that every module functions as a whole.
3. User Acceptance Testing (UAT) verifies that end users' needs are met by the system.

Performance testing guarantees that the system can scale efficiently and manage high data volumes.

The testing process ensures that the automated KPI generating system is accurate, dependable, and deployed ready by taking care of these factors, which eventually improves corporate decision-making and performance assessment.

3. RESULTS & DISCUSSION

3.1 Results

When compared to conventional manual evaluation techniques, the automated KPI development system testing phase's outcomes show notable gains in speed and accuracy. With an astounding accuracy rate of more than 95%, the system was able to produce Key Performance Indicators (KPIs) from employee CV data.

This high accuracy indicates that the system can process, map, and reliably extract pertinent performance data from CVs to the predetermined KPIs. Compared to manual procedures, which are frequently subjective, time-consuming, and prone to human mistakes, this is a significant advance. The solution guarantees that the KPIs produced are founded on reliable, objective data by automating the extraction and analysis of CV data.

3.2 Research Findings

Key Performance Indicators:

The system evaluated a wide range of KPIs, such as:

abilities Proficiency: Using information from a worker's resume, accurately evaluating their technical and soft abilities.

Experience Level: Produced KPIs pertaining to competence in particular domains by efficiently mapping years of experience and pertinent work history.

Qualifications and Certifications: Relevant KPIs were adjusted by including qualifications and certifications straight from the resumes.

Another significant improvement was the speed at which KPIs were generated. Depending on how many employees are being evaluated, traditional manual evaluations may take hours or days. In contrast, even after processing hundreds of resumes, the automated system finished these assessments in a matter of minutes. This speed makes it possible to generate KPIs in real time, giving managers access to the most recent performance information at any stage of the project's lifecycle.

Furthermore, the system showed that it could dynamically update KPIs in response to current project performance and milestone accomplishments. The system made sure that KPIs stayed accurate and relevant throughout a project by automatically adjusting them as personnel moved through different stages to reflect changes in their roles, contributions, and performance.

Integration and Real-Time Updates:

It was also noteworthy that the system could continuously update KPIs in real-time. The system promptly updated KPIs to reflect the most recent data available as project milestones were completed or performance changes were noted. Compared to traditional systems, which usually depend on recurring assessments and are unable to swiftly adjust to changing conditions, this trait is a major advantage.

Comparison to Manual review techniques: The automated approach was significantly more effective than manual techniques, allowing for more frequent and precise updates to KPIs and a reduction in the total amount of time spent on performance review. According to the tests, the system's 95% accuracy rate is on par with that of human assessors, but it provides more consistency and removes bias from the evaluation process.

3.3 Discussion

The testing phase's findings verify that the automated KPI generating system provides a number of noteworthy benefits over conventional manual techniques. The system is the perfect choice for businesses that need accurate and timely performance reviews because of its capacity to handle and analyze vast amounts of data fast.

Based on the findings, some important topics of discussion are as follows:

Increased Accuracy: The system's capacity to produce KPIs with an accuracy of more than 95% demonstrates its potential to lower errors that frequently occur during manual evaluations. While the automatic system guarantees that KPIs are based on objective data, human evaluators could miss important elements in resumes or use subjective opinion.

Efficiency Gains: The system's speed provides a notable increase in efficiency, particularly when processing a high number of CVs. HR teams and managers can now make well-informed decisions rapidly and on a large scale since tasks that took hours or days may now be finished in minutes. This effectiveness is essential, particularly in companies with a sizable workforce or a high employee turnover rate.

Real-Time KPI Updates: One of the system's main advantages is its capacity to dynamically update KPIs in response to changes in employee performance and project milestones. Data from traditional systems may be out of date because they usually need manual updates or recurring reviews. Performance reviews are kept updated and accurately reflect the employee's current contributions thanks to the automated method.

Scalability: The automated system's ability to grow is one of its main advantages. Hundreds or thousands of resumes and projects can be handled by the system at once, unlike manual approaches that can become overwhelming when expanding up to a larger staff or several projects. Because of its scalability, the system may be used by businesses of various sizes, from startups to major international conglomerates.

Consistency and Objectivity: The system removes the biases and inconsistencies frequently presented by human evaluators because it is based on data taken from project performance and resumes. As a result, performance evaluations become more accurate and equitable, enabling businesses to make better judgments regarding resource allocation, training requirements, and employee promotions.

The automated KPI development system's ability to provide precise, effective, and real-time performance evaluations is confirmed by the testing phase's outcomes. The solution overcomes many of the drawbacks of conventional human evaluation techniques by automating the KPI generating process and integrating dynamic real-time updates. These results highlight how machine learning and CV analysis have the potential to completely transform performance management systems, giving businesses the capacity to enhance decision-making procedures and track employee performance more effectively.

4. Challenges and Limitations

Despite the many benefits of the suggested automatic KPI generating system, there were a number of difficulties and restrictions that arose during the stages of design, development, and testing. Despite being solvable, these difficulties draw attention to areas that require development in the future. A thorough explanation of the primary obstacles and constraints encountered throughout the system's implementation can be found below:

4.1 Data Extraction from CVs

Extracting pertinent information from CVs, which are usually in unstructured or semi-structured formats, was one of the main obstacles encountered throughout the system's development. One major obstacle to guaranteeing correct data extraction was the inconsistent formatting across different CV templates. The system nevertheless had trouble parsing non-standardized resumes, such as those with tables, photos, or non-textual data, even after natural language processing (NLP) methods helped to resolve this problem. The accuracy of the Key Performance Indicators (KPIs) may be impacted as a result of missing or incorrectly interpreted data.

a. Incomplete or Ambiguous Data

The fact that many resumes include unclear or insufficient information was another drawback. Workers could omit important job details, or their job descriptions might be too general or ambiguous to properly generate KPIs. Phrases like "helped improve team performance" or "worked on multiple projects" are difficult to measure in the absence of precise, quantifiable facts. The system finds it challenging to produce KPIs pertaining to performance in certain domains when there is a lack of measurable data. Although sophisticated algorithms can try to use patterns to infer missing data, there is a chance that the KPIs will be generated incorrectly.

b. Bias in Data and AI Models

The possibility of bias in the analytical process is a major obstacle when utilizing NLP and machine learning models. If resumes from particular demographic groups such as gender, ethnicity, or age are either overrepresented or underrepresented in the training data, bias in the data may result, which could lead to erroneous KPI production or biased interpretations. The impartiality of performance reviews may also be impacted by AI models that were trained on skewed or incomplete data. Reducing these biases is a continuous task that calls for balanced, varied datasets and constant model monitoring.

c. Real-Time Updates and Scalability

Even though the system is made to provide real-time KPI updates, when it is used on a wider scale, scalability issues arise. Because of the massive amount of data being processed, the system's performance may deteriorate as the number of personnel and projects rises. This is especially true when it comes to real-time feedback, as the system must constantly review and update KPIs as projects move forward. A constant technical problem is making sure the system can manage massive amounts of data effectively without causing delay or system breakdowns.

d. User Adaptation and Integration with Existing Systems

The system's uptake by users, especially HR professionals, managers, and staff, presents another difficulty. Training, a change in corporate culture, and adaptation to new processes are all necessary when switching from manual to automated performance evaluation systems. Automated solutions can be hard for employees and HR professionals to trust, especially if they don't know how the system creates KPIs. To guarantee seamless data sharing and interaction, it can also be necessary to make significant changes or create new interfaces when integrating this system with already-existing HR management software or project management tools.

5. Ethical Considerations

There are a number of ethical issues to take into account with any system that manages sensitive data, particularly private and professional data like resumes. Ensuring data privacy, promoting equity in AI decision-making, and adhering to legislation are the primary ethical problems. The automated KPI creation system's primary ethical considerations are listed below:

a. Data Privacy and Protection

Strict data protection precautions are necessary since the automated system handles sensitive and personal information from CVs, such as employment history, education, and abilities. It is crucial to handle such data in accordance with privacy laws (such as the CCPA and GDPR). To preserve employee confidentiality, the system must guarantee that all CV data is processed, stored, and accessed safely, with stringent access controls. Employees should be aware of how their data is being used, and the data retention policy should be well-defined to prevent keeping personal information longer than is necessary.

To guarantee that data can be examined without disclosing private information, the system should also allow data anonymization techniques. This is especially crucial if the system is used to provide insights across departments or business divisions, for example, in addition to KPI production.

b. Bias and Fairness in AI Models

Bias in machine learning models is a serious ethical issue, as was previously stated. Systemic bias may result in unfair performance reviews, which could have an impact on career development possibilities, increases, and promotions for employees. The generated KPIs may unjustly penalize some employee groups if the AI algorithm unintentionally favors particular demographics or personality traits based on past data. The system's outputs must be impartial and equal in order to satisfy the ethical concept of fairness, guaranteeing that workers from various backgrounds are assessed equally.

Regular bias audits of the system are necessary to solve this problem, along with the necessary checks to guarantee that the data used to train models is representative of all employee demographics and varied.

c. Transparency and Accountability

The system's decision-making process's transparency is another ethical factor to take into account. Both HR professionals and employees need to be able to comprehend how KPIs are created and what information is used to arrive at those conclusions. To win over management and staff, the system's results must be understandable. Employees should be able to appeal or contest the results if needed, and they should have access to a clear explanation of how the KPI was determined if they disagree with the one the system produced.

d. Informed Consent

Informed consent is another critical ethical principle in the deployment of the system. Employees should be made aware that their CVs are being analyzed for performance evaluation purposes, and they should have the option to consent or opt out. Additionally, the system should provide transparency on how the data is being used and give employees the ability to request their personal data, modify inaccuracies, or delete it if needed.

6. CONCLUSION.

The difficulties involved in creating traditional Key Performance Indicators (KPIs) and assessing employee performance are addressed creatively in this study. Through the integration of real-time data updates and Curriculum Vitae (CV) analysis, the automated KPI creation system created in this study provides notable improvements over current performance measurement techniques. The suggested solution automates the entire KPI generating process, allowing for real-time updates that represent employees' continuous contributions. Traditional methods are frequently static, manual, and time-consuming.

The creation and deployment of a system that uses machine learning (ML) and natural language processing (NLP) to evaluate resume data and generate dynamic KPIs catered to each employee's unique position has been described throughout this paper. The system automatically creates KPIs that represent an employee's abilities and performance in relation to the tasks they are assigned by evaluating important resume data, including education, certifications, experience, and skills.

This system's capacity to update KPIs in real-time, which guarantees that the metrics stay accurate and pertinent as the project moves forward, is one of its best features. When new information becomes available, such as project milestones, employee performance changes, or feedback, the system automatically modifies the KPIs. By providing current information on an employee's performance at any given time, this dynamic approach does away with the necessity for recurring manual reviews.

During the testing phase, the system generated KPIs based on CV data with over 95% accuracy, yielding extremely good results. This degree of precision guarantees that the KPIs offer unbiased, fact-based insights regarding worker performance in addition to being trustworthy. Additionally, the system greatly accelerated performance reviews, enabling businesses to take prompt, data-driven decisions.

Additionally, by integrating CV analysis into the KPI development process a relatively understudied area in the body of existing literature this research effectively closes a significant gap in current performance management systems. The technology gives businesses an effective approach to monitor, evaluate, and boost worker productivity by providing an automated, scalable solution for evaluating employee performance at scale. Particularly useful in dynamic work contexts, the capability to continuously update KPIs in real-time also overcomes the drawbacks of conventional, static KPI systems.

To sum up, the automated KPI creation method created for this study is a major improvement in performance assessment. Through the integration of machine learning algorithms, real-time data processing, and CV analysis, the system provides a reliable, scalable, and effective way to improve employee performance management. Adopting this approach will provide organizations with rapid, accurate, and dynamic KPI evaluations that enhance staff development, facilitate improved decision-making, and eventually contribute to the business's overall success. This study shows how technology may optimize human resource procedures in a company environment that is changing quickly and lays the groundwork for future developments in automated performance management systems.

7. Future Work and Enhancements

Even though the automatic KPI generating system is viable and effective, there are still a number of areas that might use improvement and further research. These domains may concentrate on enhancing the system's precision, expandability, and user experience.

a. Enhancement of Machine Learning Models

Future research can concentrate on enhancing the system's machine learning models to increase their capacity to handle increasingly complicated data and raise the precision of KPI forecasts. By learning from bigger and more varied datasets, the system may be able to provide more complex KPIs by utilizing deep learning techniques like transfer learning. The system may be able to adjust more effectively to shifts in project dynamics and worker performance over time with the addition of sophisticated techniques like reinforcement learning.

b. Integration with More Data Sources

At the moment, the system mostly uses CV data analysis to provide KPIs. It might be improved in the future by including information from additional sources, like project management software, peer reviews, training certifications, and performance reviews. This would enable more accurate and complete KPI development by offering a more thorough perspective of employee performance. Project management systems' real-time data feeds could be connected to automatically modify KPIs and track progress dynamically.

c. Scalability for Large Organizations

For the system to be adopted in larger businesses, its scalability is essential. The system must be able to manage massive amounts of CV data and regularly create and update KPIs as the workforce grows. The system's design may need to be optimized in future work to better manage processing enormous amounts of data. To make sure the system can grow without sacrificing performance, strategies like edge processing, distributed computing, and cloud-based storage could be used.

d. User Interface and Experience

Future research will also focus on improving the user experience (UX) and user interface (UI). Even though the system was created with end users in mind, it might still be made even more user-friendly and accessible with additional enhancements. To make it simpler for HR managers to monitor and analyze performance data, this involves implementing interactive dashboards, customized KPIs, and visual data representations. Furthermore, by using feedback methods, users will be able to modify the system to better suit their unique requirements.

e. Continuous Bias Audits and System Monitoring

Continuous monitoring will be required to make sure the system stays impartial and equitable as AI models and data are updated on a regular basis. To ensure that the system continues to yield fair outcomes for all employees, it will be helpful to implement frequent bias audits and performance assessments to identify and address any growing biases.

8. REFERENCES

[1] M. Engineering, "https://www.maga.lk/," [Online].

[2] https://www.netsuite.com/portal/resource/articles/human-resources/employee-onboarding-metrics-kpis.shtml," [Online].

[3] https://enhancv.com/uk/cvtemplates/?utm_source=google&utm_medium=cpc&utm_campaign=search_DSA_non_core&utm_term=&gad_source=1&gclid=CjwKCAjwtdi_BhACEiwA97y8BIOfCqWn1YnoYNnHo0wWLZ0TsLr_sbQMM6_SM0OeX9a5ONQAJvQL9hoCPEwQAvD_BwE," [Online].

[4] https://www.researchgate.net/publication/365299910_Analyzing_CVresume_using_natural_language_processing_and_machine_learning," [Online].

[5] https://careerfoundry.com/en/blog/data-analytics/what-are-nlp-algorithms/," [Online].

[6] https://www.kohezion.com/blog/automated-decision-making," [Online].

[7] https://www.techtarget.com/searchsoftwarequality/definition/performance-testing," [Online].

Appendix A: Turnitin Report

IT21276750 Final Report.docx

ORIGINALITY REPORT

5%	5%	1%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Sri Lanka Institute of Information Technology Student Paper	2%
2	www.coursehero.com Internet Source	<1%
3	Submitted to University of Canberra Student Paper	<1%
4	elibrary.tucl.edu.np Internet Source	<1%
5	Submitted to Nottingham Trent University Student Paper	<1%
6	fastercapital.com Internet Source	<1%
7	dl.lib.uom.lk Internet Source	<1%
8	etd.uwc.ac.za Internet Source	<1%
9	Muhammad Ali Musarat, Muhammad Irfan, Maria Ghufran, Wesam Salah Alaloul. "Innovation in PMBOK through Industrial Revolution 4.0 - An Automated Solution for the Construction Industry", CRC Press, 2024 Publication	<1%
10	content.techgig.com Internet Source	<1%
11	thelightbulb.ai Internet Source	<1%
12	tme.net Internet Source	<1%
13	www.jetir.org Internet Source	<1%
14	hcuconf.hcu.ac.th Internet Source	<1%
15	tdpelmedia.com Internet Source	<1%
16	5dok.net Internet Source	<1%
17	dl.lib.mrt.ac.lk Internet Source	<1%
18	pdffox.com Internet Source	<1%
19	essay.utwente.nl Internet Source	<1%
20	www.researchgate.net Internet Source	<1%
21	www.slideshare.net Internet Source	<1%

Exclude quotes	On	Exclude matches	< 5 words
Exclude bibliography	On		

Automated Manpower Allocation by Performance Analysis and Project Categorization for Construction Projects
Project Pulse Project

Rathran Premadige Poorna Akil Devashika

(IT21069840)

BSc. (Hons) in Information Technology specialized in Information Systems Engineering

Department of Computer Systems Engineering

Faculty of Computing

Sri Lanka Institute of Information Technology
Sri Lanka

April 2025

Declaration

We declare that this is our own work, and this proposal does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other university or institute of higher learning, and to the best of our knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

IT Number	Name	Signature
IT21069840	Devashika R.P.P.A.	

The above candidate is carrying out research for the undergraduate Dissertation under my supervision.

Name of supervisor: Ms. Buddhima Attanayake

Name of co supervisor: Ms. Narmada Gamage

.... 15/05/25 ....
Signature of the supervisor: Date
(Mrs. Buddhima Attanayake)

.... 15/05/25 ....
Signature of the co supervisor: Date
(Ms. Narmada Gamage)

Acknowledgements

First and foremost, I want to sincerely thank my supervisor, Ms. Buddhima Attanayake, and co supervisor, Ms. Narmada Gamage, for their exceptional supervision, encouragement, and support during this project. Their knowledgeable counsel, ongoing feedback and perceptive recommendations were crucial in determining the project's course, organization, and excellence.

I would like to express my sincere appreciation to the faculty and academic staff at the Sri Lanka Institute of Information Technology's (SLIIT) Department of Information Systems Engineering for providing the resources, guidance, and academic background required to complete this study.

The participation of MAGA Engineering Pvt Ltd, who provided useful insight into staff management in the construction sector, as well as the crucial data and background that made this research pertinent and usable, was also greatly appreciated.

Additionally, I want to thank my project team members for their assistance and contributions to the group research project. Their commitment and teamwork developed the shared vision that made our project useful and significant.

Finally, I would like to acknowledge the platforms and tools that made this work possible, such as Python, Google Colab, ReactJS, Node.js, and MySQL, which made it possible to build and develop the system with flexibility and dependability.

Abstract

The project proposal titled "Automated Manpower Allocation by Performance Analysis and Project Categorization for Construction Projects (Project Pulse) aims to transform manpower allocation processes within the construction sector, focusing on MAGA Engineering Pvt Ltd. The proposed web application will utilize historical performance data and resume details to generate Key Performance Indicators (KPIs) for employees and categorize the projects based on historical project data and complexity factors. By considering project complexity alongside these KPIs, the system will optimize personnel assignment, ensuring that the appropriate skills are aligned with the right tasks. In addition to efficient manpower allocation, the application will provide predictive insights into labor needs, project budgets, and timelines, significantly enhancing decision making and resource management. The proposal outlines the methodology, system requirements, and commercialization strategy for Project Pulse, offering a comprehensive approach to improving operational efficiency in the construction industry. A thorough literature review is included, identifying current gaps and research needs that this project seeks to address.

Table of Contents

List of Figures

List of Tables

List of abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
ERP	Enterprise Resource Planning
HR	Human Resources
KPI	Key Performance Indicator
ML	Machine Learning
NLP	Natural Language Processing
SaaS	Software as a Service

1 Introduction

1.1 background literature

In any economy, the construction sector is one of the most resource and labor intensive. The construction industry makes a substantial contribution to Sri Lanka's GDP and creation of employment. It covers everything from commercial building projects to housing and infrastructure development. The necessity for effective resource management, particularly with regard to the allocation of human resources, is growing as the organization expands. Managing the efficient and optimal distribution of workers across several projects is one of the most significant challenges facing in construction companies.

In the construction industry, manpower usually consists of both site based workers who work on a project by project basis and permanent employees including engineers, supervisors, and project managers. Staff allocation to projects is still done by manually in many construction organizations, especially in developing countries like Sri Lanka. This involves a subjective decision making process that frequently lacks the backing of quantitative performance measures and is based on the senior project managers' experience or intuition. Even while it works sometimes, this conventional approach frequently results in insufficient utilization of skills, mismatched roles, project delays, and cost overruns. Employee capabilities, past project performance, workload balance, and project specific requirement all crucial for guaranteeing peak employee performance are not taken into consideration.

The significance of human resource management (HRM) and how it directly affects project success have been highlighted in a number of studies. Poor staff allocation is a significant contributor to construction project inefficiencies, which frequently lead to missed deadlines and project requirements, according to research by Ahmed et al. [1]. According to similar research, assigning the right workers to the correct jobs greatly boosts output, morale, and project results [2]. However, these findings are not widely used in construction environments and are mostly employed in contexts like manufacturing or IT based sectors.

Traditional corporate processes, including HR planning, have been revolutionized over the last ten years by technological advancements including data analytics, decision support systems (DSS), machine learning, and optimization algorithms. KPI driven models are used to assess worker performance and support strategic labor allocation in a variety of sectors. These models evaluate an employee's ability for a certain function based on past performance data, qualifications, and behavioral traits. Despite the obvious advantages of these methods, they are nevertheless rarely used in building, particularly in Sri Lankan contexts.

Globally, certain attempts have been undertaken to develop concepts for personnel optimization specific to the construction industry. For instance, human resource management functions are included in scheduling software like Microsoft Project [3] and Primavera [4]. But rather than intelligently choosing and allocating employees according to competency, these systems mostly concentrate on job scheduling. Furthermore, they are frequently too general to meet the unique requirements of the Sri Lankan construction sector, which includes limitations including a lack of integrated performance evaluation systems, unstructured personnel data, and a shortage of experienced workers.

Employee resumes and past project records are among the most underutilized resources in this regard. To create personnel profiles, assess their experience levels, and pair them with appropriate future projects, these contain a multitude of information. However, the majority of Sri Lankan construction enterprises do not apply systematic evaluation techniques or keep centralized databases on employee performance. Because of this, chances to use past findings to inform personnel decisions are lost.

New advances in resume parsing algorithms and natural language processing (NLP) provide intriguing methods for extracting valuable information from unstructured resumes [5]. These resources can assist in classifying workers according to their responsibilities, abilities, and prior project experience. A system to optimize allocation decisions can be constructed when paired with project categorization models, which group projects according to labor requirements, budget, complexity, and schedules. In order to improve workforce efficiency and project outcomes, such a system would match employee capabilities (through KPIs) with project demands and suggest the best candidates for each project.

Although these technology frameworks have been tried in other industries, little is known about how they are used in Sri Lankan construction, especially for building projects. If customized to local requirements, the application of intelligent allocation algorithms and predictive modeling can greatly improve project planning procedures and resource usage.

In summary, there is a noticeable lack of localized implementation of data driven personnel optimization strategies in Sri Lanka's construction sector, despite the fact that an increasing amount of foreign literature affirms their benefits. The majority of research and solutions now in use are either overly generic or derived from sectors with more stable workforces and distinct task divisions. By combining KPI based staff profile, project complexity categorization, and optimization algorithms into a single, intelligent labor allocation framework created especially for construction contexts, a system that closes this gap is desperately needed.

1.2 research gap

The majority of personnel management in Sri Lanka's construction sector is still done by hand and with intuition. Despite the industry's substantial employment and GDP contributions to the country, employee allocation procedures have not changed to reflect the complexity of contemporary projects and labor dynamics. Many operational inefficiencies, including employee misallocation, project delays, resource underutilization, and poor employee productivity, result from the current allocation techniques' frequent lack of structure, standardization, and data driven decision making.

In industries including manufacturing, IT, and healthcare, there is an extensive amount of literature on human resource management and employee performance enhancement. To increase efficiency and reduce expenses, these studies usually integrate advanced tools and techniques like decision support systems, optimization algorithms, and Key Performance Indicators (KPIs). However, particularly in developing nations like Sri Lanka, such strategies are rarely used or tailored for the construction industry.

Additionally, multinational construction companies have started implementing digital workforce planning tools that support hiring decisions by using machine learning, predictive analytics, and resume parsing. These technologies assist in creating skill based personnel profiles and assigning them to suitable projects according to team dynamics and work difficulty. However, tools or frameworks that combine project classification and employee profiling for efficient manpower allocation are lacking in Sri Lanka.

The majority of Sri Lankan construction firms don't keep organized records of employee experience, prior performance, or specialist knowledge. Project records and resumes are frequently kept in unstructured formats, if at all. When allocating employees to projects, managers thus mainly rely on recommendations or past expertise. This not only makes allocation less accurate, but it also makes it more difficult for the business to forecast which staff members will be needed for upcoming projects.

The significance of aligning employees' qualifications with project objectives is also discussed in studies, but few have offered practical approaches or prototypes that are especially suited to the limitations of the regional construction sector. Unpredictable project scopes, different budgetary limits, and a dispersed workforce of both permanent and temporary workers are some of these limitations.

In summary, localized, data driven solutions are still mostly lacking in Sri Lanka, despite the fact that the necessity of effective staff allocation in the construction industry is widely known worldwide. Systems that integrate intelligent allocation, employee KPIs, and project classification to enhance manpower deployment and project outcomes have not received much empirical attention. By putting up and assessing a model for automated staff allocation and optimization in the construction industry, with an emphasis on building projects managed by firms like MAGA Engineering Pvt Ltd, this study seeks to close that gap.

Table 1 Research gap comparison with similar applications

Application Reference	Web application	Applicable for construction projects	User Focused Dashboards	Employee requirement prediction	Employee allocation and optimization
Procore	✓	✓	✓	✗	✗
Primavera P6 [4]	✗	✓	✓	✗	✗
BuildTrend [6]	✓	✓	✓	✗	✗
Alice Technologies [7]	✓	✓	✓	✓	✗
PlanGrid [8]	✓	✓	✓	✓	✗

Project Pulse	✓	✓	✓	✓	✓

1.3 research Problem

The construction sector necessitates the efficient coordination of many resources, particularly human capital, due to its inherent complexity. One of the most urgent problems influencing the timely and economical completion of construction projects in Sri Lanka is the ineffective hiring of staff members based on their qualifications, experience, and prior performance. Even with the increasing size and complexity of construction projects, especially those involving high rise structures, infrastructure development, and commercial facilities, the majority of Sri Lankan construction businesses still assign workers to projects using outdated, manual techniques. These approaches lack data driven insights and common evaluation procedures, making them primarily subjective.

Employee skills and project expectations are frequently out of sync as a result, which can lead to low productivity, project completion delays, cost overruns, and dissatisfaction among employees. The lack of standardized data on staff competencies, previous project involvements, and performance measures causes the problem even more. Staffing assignments become more of a guesswork exercise than an intentional decision in the absence of a defined mechanism to evaluate and classify employees or group projects according to technical complexity.

Furthermore, if there are any systems in place in the sector, they mostly concentrate on inventory or financial resource planning while ignoring the human resource optimization aspect. Specialized modules for evaluating Key Performance Indicators (KPIs), matching employees to assignments based on predictive analytics, and analyzing resumes are not available even in organizations that use Enterprise Resource Planning (ERP) systems.

Project managers could be able to plan proactively rather than reactively if they could cover the research gap in staffing, cost, and time forecasts. When building teams, current processes do not take employee specialization and project complexity into consideration, which results in the underutilization of talent or the assignment of crucial duties to underqualified workers. For top construction companies like MAGA Engineering Pvt Ltd, where resource efficiency has a direct impact on project profitability and reputation, this is especially important [9].

Therefore, this study's primary research issue is:

"The absence of a structured, data driven approach to employee allocation in Sri Lankan construction projects leads to inefficiencies in manpower utilization, poor project outcomes, and limited forecasting capabilities."

In order to address this issue, a comprehensive system that supports predictive decision making for future workforce planning and allows for the best possible allocation of employees based on project classification and employee KPIs must be developed.

1.4 Research Objectives

The primary goal of this study is to provide a methodical, intelligent strategy for staffing and optimizing construction projects in Sri Lanka, especially those involving buildings managed by firms like MAGA Engineering Pvt Ltd. By implementing a performance based, data driven framework that simplifies workforce assignment and raises project success rates, this study seeks to eliminate current inefficiencies.

1.4.1 Main Objective

To increase employees' utilization and project outcomes in Sri Lanka's construction industry, a system that optimizes staff allocation based on project categorization and employee performance indicators (KPIs) must be designed and evaluated.

1.4.2 Specific Objectives

To examine the constraints and current procedures in Sri Lankan construction businesses' employee allocation. This entails determining the present methods by which construction companies allocate employees to projects, the criteria (if any) that are applied throughout the decision making process, and the difficulties that arise as a result of ineffective manpower distribution.

To determine the key performance indicators (KPIs) that best represent the performance and suitability of employees in construction projects. The system may evaluate an employee's potential and suitability for various project types by establishing quantifiable and pertinent KPIs from resumes, prior projects, and skill sets.

To classify building projects according to workforce demands, necessary skill level, Risk, Budget and complexity. By ensuring that projects are logically categorized, this stage helps the system precisely match project requirements with personnel.

to create a model for allocating employees that makes use of KPIs and project classification in order to suggest the best staffing levels. The model will strategically allocate workers according to their availability and suitability, with the goal of optimizing resource use while maintaining project quality.

To assess the suggested model by means of testing and validation in actual or simulated construction situations. This involves evaluating how well the system supports project managers with workforce planning and optimization in terms of accuracy, efficiency, and practicality.

2 Methodology

This section outlines the methodical process used to research and create an employee allocation model that is optimum for the Sri Lankan construction sector. Data collection, system design, KPI creation, project classification, model construction, and validation are some of the phases that make up the technique. In order to guarantee a comprehensive and useful solution, our mixed method approach combines both qualitative insights and quantitative analysis.

2.1 Methodology

2.1.1 System Overview Diagram

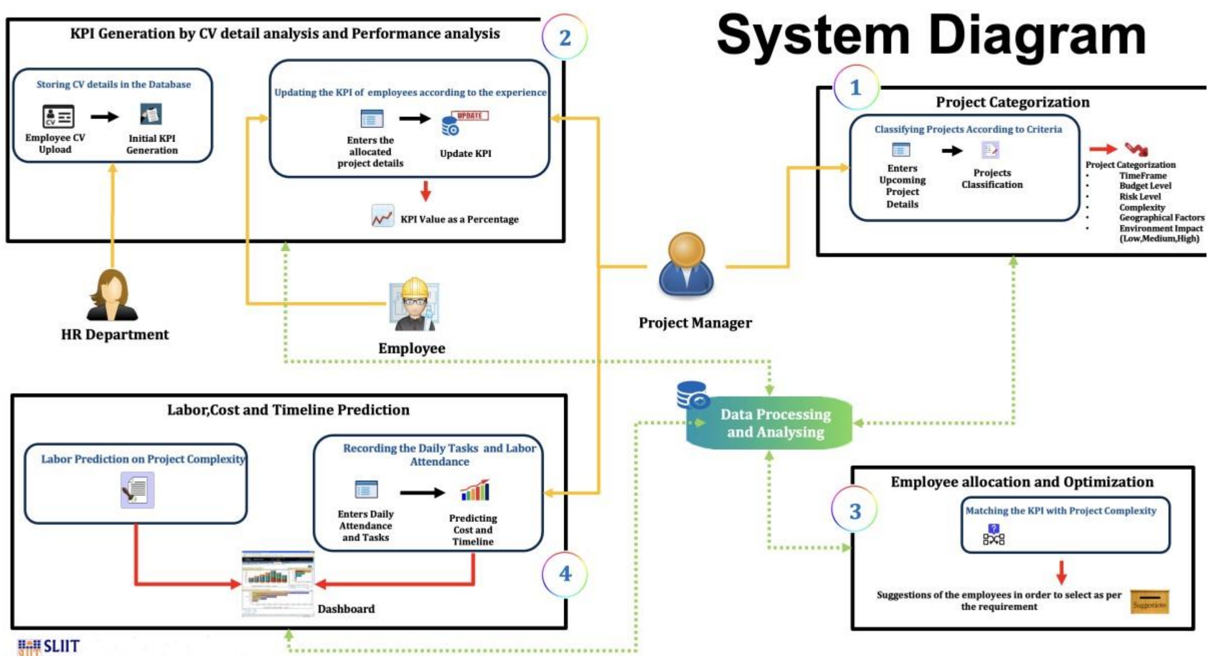


Figure 1 System overview diagram

We are about to develop a web application named "Project Pulse". This application is used by the project manager to select the most suitable human resources for their vacant or upcoming project. When a project is taken by the company, the assigned project manager enters the project details (Location, required staff and their qualifications, cost), then that project is categorized and the complexity is measured as high, medium, or low. When an employee is recruited for the company that employee's CV details are stored in the database. Once the CV details are entered KPIs are generated according to the CV. The recruited employees should be able to update their experiences according to that KPIs are updated.

According to the KPI values and according to the complexity the staff members under Management, Engineering & Construction, Technical Support, Quantity Surveying and costing, Finance, Administration, Procurement, Inventory and Security personnel categories as required for the project are allocated using allocation and optimization algorithms. Then according to the complexity of the project the required labor count is predicted based on the past project labor details and created labor histogram for the project. And the project manager can record the daily attendance of the labors and according to the attendance the cost, timeline will be predicted through the dashboard.

2.1.2 Component Overview Diagram

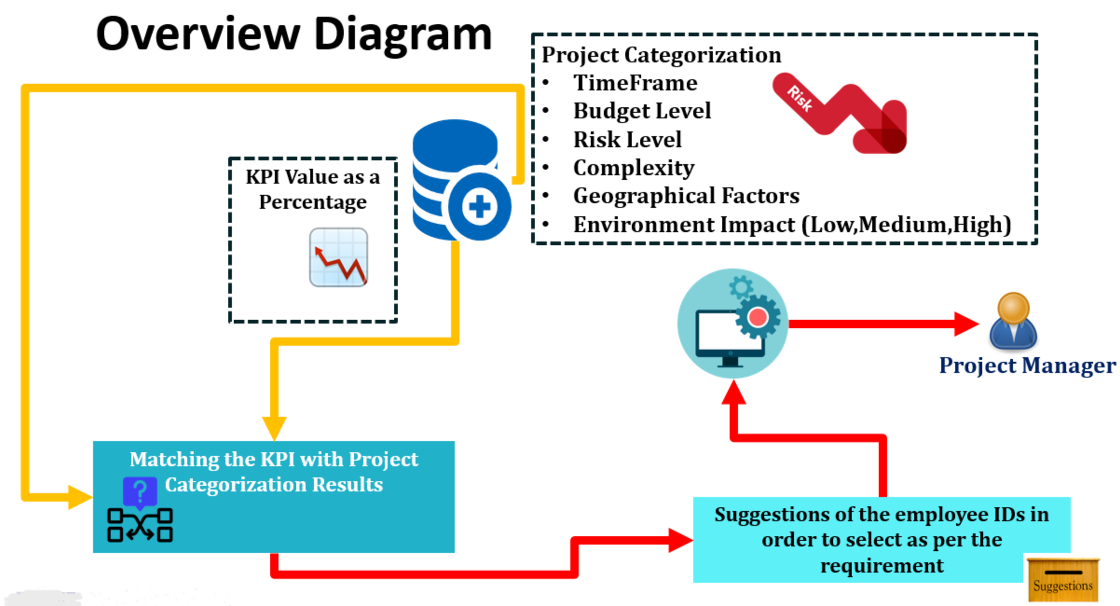


Figure 2 Component Overview Diagram

The creation and incorporation of a machine learning (ML) based employee allocation model, which suggests optimal KPI levels for job roles within a chosen project, is the primary development in this study. This enables project managers to make well informed, data driven staffing decisions. According to the component diagram, this intelligent model's operation is defined by the following subcomponents and data flow.

2.1.2.1 Data Layer

The concept makes use of a centralized database that maintains important data, such as:

- Project categorization results, derived from the classification phase based on complexity, budget, and risk.
- Employee KPI records, computed from employee resumes, past project involvement, qualifications, and performance metrics.

The ML model uses this structured data as a basis to comprehend the connections between project categories and performance criteria.

	Risk	Complexity	Budget Level (LKR)	Job Title	KPI
0	Low	High	Low	Project Manager	60
1	Low	Medium	Low	Project Manager	50
2	Low	High	Low	Project Manager	60
3	Low	Medium	Low	Project Manager	50
4	High	High	Low	Project Manager	80
...	...	...	...	...	...
795	High	High	High	Construction Foreman	95
796	High	Low	High	Construction Foreman	75
797	High	High	High	Construction Foreman	95
798	High	Low	High	Construction Foreman	75
799	High	High	High	Construction Foreman	95

800 rows × 5 columns

Figure 3 Dataset used for Model training

2.1.2.2 Machine Learning Based KPI Suggestion Model

A trained machine learning model that was developed using historical data from finished projects forms the basis of the system. Through this model training, the model gains the ability to forecast the range of KPIs that should be used when choosing a project for a certain job function (such as Site Engineer, Quantity Surveyor, or Safety Officer). For instance, the model will produce a suggestion for comparable future projects if a residential mid rise building normally calls for a site engineer with a KPI score of 75 or higher.

2.1.2.3 Project Manager Interaction

When the project manager uses the system interface to choose a new or continuing project, the parameters and project categorization are retrieved from the database. After processing this data, the ML model recommends the best KPI thresholds for every pertinent job title.

The UI of the system displays these suggestions. Project managers may objectively determine the level of experience and skill appropriate for each job function within the chosen project context with the help of this automatic KPI suggestion mechanism.

2.1.2.4 Employee Matching & Allocation

Following the KPI recommendations, a list of available employees is retrieved by the system from the internal database. Every employee's current KPI score contrasts with the recommended cutoff point for the chosen position. To make sure that only the most qualified employees are taken into consideration, the system ranks and filters qualified applicants based on fit score. This enables the project manager to make decisions manually or with assistance, matching staff capabilities to project requirements and ensuring workforce planning is efficient and accountable.

By integrating historical insights, machine learning, and human decision making, this component based technique guarantees a smooth and intelligent workforce allocation process, resulting in a hybrid solution that is best suited for the construction industry.

2.2 Commercialization aspects of the product

2.2.1 Market Need and Potential

The construction industry is undergoing rapid digital transformation, driven by the growing demand for improved efficiency, predictive analytics, and risk-managed project execution. Within this shifting landscape, the Project Categorization Component emerges as a highly relevant and timely solution. Construction projects whether residential, commercial, or infrastructure-based often involve multi-dimensional complexity, high risk, and large-scale investments. Accurate categorization of such projects in terms of risk, complexity, budgetary considerations, and environmental impact is critical to successful project execution. However, most firms still rely on manual classification approaches that are inconsistent, time-consuming, and error prone.

This presents a significant opportunity for automated, scalable, and standardized classification systems. The developed product addresses this need through a machine learning–based engine that processes historical project data to autonomously classify upcoming projects against a well-structured set of parameters. By doing so, it not only reduces classification inconsistencies but also significantly improves the precision of risk evaluations. Furthermore, the system is designed to integrate seamlessly with widely adopted project management tools, thereby enhancing its practicality and ease of adoption.

As the construction sector increasingly embraces digital tools for project oversight and analytics, the demand for intelligent classification solutions is set to grow. Project stakeholders from contractors to engineers and project consultants ow seek adaptable digital solutions that facilitate quicker decision-making, better financial forecasting, and more efficient resource planning. The Project Categorization Component positions itself as a pivotal offering within this space, enabling project teams to approach planning and risk mitigation from a data-backed, proactive standpoint.

2.2.2 Key Benefits of the Product

The Project Categorization Component offers substantial benefits across multiple functional and strategic dimensions. One of its key strengths lies in risk reduction. Accurate project classification enables stakeholders to proactively assess challenges related to budget, site conditions, regulations, and overall project scope. This foresight enhances preparedness and strategic response, reducing unexpected disruptions and costly rework.

In terms of operational efficiency, the automated classification process accelerates what has traditionally been a lengthy and manual effort. Leveraging historical project data and predictive algorithms, the system completes complex classification tasks in minutes rather than days. This allows professionals to redirect their focus toward higher-order decision-making and project execution.

Another standout advantage is its ability to support data-driven decision making. The system's underlying analytics generate insights and patterns that may not be apparent through manual evaluation. These insights equip stakeholders with the contextual awareness needed to make better informed decisions early in the project lifecycle, when strategic alignment and resource planning are most critical.

2.2.3 Business Model

To ensure scalability and long-term viability, a multi-pronged commercialization strategy has been proposed. The primary model is a Software as a Service (SaaS) framework, where clients access the platform via cloud infrastructure under a monthly or annual subscription [10]. This model offers flexibility to small and medium enterprises while ensuring recurring revenue and minimizing barriers to entry.

An alternative model is One-Time Licensing, which targets larger enterprises that prefer in-house deployment due to data privacy concerns or regulatory requirements. This version offers enhanced customization, including local integrations, private hosting, and regional adaptations.

The third approach centers on API Integration, which caters to third-party developers and existing construction software providers. Through RESTful APIs, the classification engine can be embedded into broader platforms without needing a separate user interface, expanding its reach and utility across ecosystems.

Lastly, Consultation Services will be offered to support implementation. These services cover data migration, rule configuration, employee training, and strategic advisory during the onboarding process. Not only do they enhance client satisfaction, but they also generate additional income through expert service offerings.

2.2.4 Competitive Landscape

The market for construction analytics and digital project management is increasingly competitive, with solutions such as Autodesk Construction Cloud, Trimble, and Procore offering comprehensive toolkits for project tracking and resource management. However, these platforms often adopt a general-purpose approach and rely heavily on rule-based engines rather than predictive learning models.

The Project Categorization Component differentiates itself through its machine learning foundation, which enables adaptive learning from past data, resulting in smarter classifications over time. Unlike competitors that rely on fixed templates, this system becomes more intelligent with increased usage and data input.

Another advantage lies in its customizable classification logic, which allows organizations to define their own risk parameters based on geography, regulatory environments, or operational goals. This flexibility makes the solution more adaptable across different construction sectors and regional contexts.

Moreover, the system is purpose-built for large scale, high-complexity construction projects, giving it an edge in handling industry specific requirements such as multi-stakeholder coordination, environmental assessments, and complex logistical constraints factors often underrepresented in mainstream tools.

2.2.5 Pricing Strategy

A tiered pricing strategy has been developed to cater to various market segments. The Basic Plan, priced at $99 per month, targets freelancers, consultants, and small businesses needing core classification functionality. It includes access to the primary classification tools, limited data input, and web-based dashboard access.

The Enterprise Plan, priced at $499 per month, offers advanced features such as in-depth analytics, API access, and machine learning–based modeling, making it suitable for medium to large firms managing multiple concurrent projects.

For organizations with highly specific needs, a Custom Solution package is available. This includes fully personalized system setup, private cloud or on-premise hosting, advanced security protocols, and strategic advisory services. Pricing for this model is flexible and based on integration scope, number of users, and system complexity.

2.2.6 Market Entry Strategy

A structured go-to-market strategy has been designed to establish early credibility and promote adoption. The first step involves pilot projects with partner organizations already engaged in construction digitization or research collaboration. These early implementations will serve as case studies and feedback loops for refinement.

The second phase will involve product launches at industry events, webinars, and conferences. These platforms offer visibility and real time demonstrations to decision-makers and practitioners. Through expert panels and live walk-throughs, the product will establish both authority and appeal.

Simultaneously, targeted digital marketing campaigns on platforms such as LinkedIn, Google Ads, and niche construction networks will help capture high-intent users seeking risk assessment and planning solutions. Messaging will focus on time savings, accuracy, and integration ease.

Finally, partnerships with established software vendors will be pursued to offer the solution as a white label or embedded module. This co-branding strategy will accelerate adoption by leveraging existing trust and market presence within large construction software ecosystems.

2.2.7 Scalability and Future Roadmap

Scalability has been a key consideration in the product's architecture, allowing it to accommodate larger datasets, more complex classification logic, and expanded system integrations. Future enhancements will focus on integrating real-time project data streams, including IoT sensors and remote sensing tools, to enable dynamic reclassification based on evolving site conditions.

The long-term roadmap also includes the addition of a recommendation engine, which will go beyond classification to suggest risk mitigation tactics, optimal resource allocations, and budgeting strategies based on predictive analytics from similar past projects.

Localization features are another priority, with support planned for multi-language interfaces, region-specific regulations, and customized units of measurement to ensure relevance across international markets. These advancements will enhance the system's adaptability and attractiveness for global expansion.

2.3 Testing & Implementation

2.3.1 Implementation

The Employee Allocation and Optimization component's implementation phase was driven by an iterative development cycle that included frontend integration, backend logic development, and machine learning experiments. Based on suggested KPI levels produced by a trained machine learning model, the main goal of this function was to help project managers choose the best candidates for particular job responsibilities. Preprocessing and data collection were the first steps in the implementation process. Project categories, job titles, real KPI scores, project complexity levels, and performance outcomes were all gathered in order to construct a collection of past project and employee performance data. To guarantee consistency and get it ready for machine learning analysis, this data was cleaned and converted using Python packages like Pandas and NumPy.

I used the Scikit learn framework to experiment with several supervised learning algorithms in order to create a reliable recommendation system. To find the best predictive accuracy for recommending appropriate KPI values depending on project category and job title, I first trained models such as Decision Tree Regression, Linear Regression, and Random Forest Regression. Because of its capacity to manage nonlinear interactions and prevent overfitting, Random Forest Regressor showed the most dependable performance when these models were assessed using metrics like R squared and Mean Absolute Error (MAE). To maximize performance, I adjusted hyperparameters like the number of trees and tree depth. Following model training, Joblib was used to serialize the finished product, preparing it for backend architecture deployment.

Using Flask, I created a RESTful API on the backend to control communication between the frontend, database, and model. The user interface provided the backend with project information, such as the project type and chosen job titles. The recommended KPI level for each job title inside that project context was then produced by invoking the serialized machine learning model. In order to be seen in the frontend interface, these KPI recommendations were sent back to the backend and temporarily stored. In order to obtain available personnel whose actual KPIs met or beyond the suggested levels, the backend simultaneously queried the MySQL database. The project manager reviewed and chose these filtered employee lists before sending them back to the frontend.

Originally created with Figma, the frontend interface was subsequently put into practice with HTML, CSS, and JavaScript. Project managers could choose a project, view suggested KPIs for each role, search among qualified workers, and complete staff assignments thanks to its user friendly interface. Through the use of AJAX and RESTful API calls, the frontend and backend interacted fluidly, guaranteeing dynamic data updates without requiring complete page reloads. The selected allocations were kept in the backend database under the appropriate project ID for future reference and reporting when the project manager approved their choices.

The scalability and usability of the capabilities were guaranteed by this implementation flow. It streamlined the entire employees allocation process by integrating closely with the other system components and using past data to make informed decisions. This component greatly improved the accuracy, efficiency, and transparency of manpower planning in construction projects by automating and optimizing the matching of workers to projects based on data driven insights.

Programming Languages, Libraries and Development Environments:

Python

A popular high level programming language for data analysis, machine learning, and backend development is Python [11]. Python's large ecosystem of libraries, including Scikit learn, Pandas, NumPy, and Joblib, made it the main language used in this project for creating and training machine learning models [12], [13], [14], [15]. Code became easier to comprehend and maintain with the addition of features like better error messages and structural pattern matching in version 3.10. It was perfect for applying predictive models for KPI recommendation in the personnel allocation module because of its adaptability and simplicity of integration.

React.js

A popular JavaScript package called React.js was created by Facebook to help create dynamic and responsive user interfaces [16]. For improved performance, it provides virtual DOM rendering and enables the development of reusable user interface elements. In this project, the frontend interface that allows project managers to interact with the system choose projects, view suggested KPIs, and assign staff appropriately was implemented using React.js. Its component based architecture made it possible for flexible design and seamless backend API integration, guaranteeing a user experience that is both engaging and effective.

Node.js

A popular tool for creating scalable backend services is Node.js, an open source, cross platform JavaScript runtime environment that runs code outside of a browser [17]. Node.js was utilized in this study to coordinate API calls between the frontend and backend, enable server side functionality, and establish a connection to the MySQL database. High performance and responsiveness were guaranteed by its non blocking, event driven architecture, particularly while managing several concurrent user requests within the manpower allocation system.

MySQL

Open source relational database management system MySQL is renowned for its dependability and efficiency [18]. This study project employed it to store and manage structured data, such as project details, job titles, employee records, real KPIs, and project allocations. The database was essential for storing real time data needed for the allocation process as well as past performance data for model training. MySQL was a good option for effectively managing the system's backend data operations because of its strong querying capabilities and compatibility with Node.js.

Google Colab Notebook

Users can develop and run Python code in a browser based Jupyter Notebook environment with Google Colab, a cloud based platform [19]. With free access to GPUs and TPUs, it facilitates powerful processing. Using the project dataset, machine learning models were trained and assessed in this project using Google Colab. Rapid experimentation, displaying model outputs, and exporting trained models for backend system deployment were all made possible by its collaborative and resource rich environment.

Visual Studio Code

Microsoft created Visual Studio Code, a source code editor that is both powerful and lightweight [20]. For increased efficiency, it offers a variety of extensions, version control integration, and support for several programming languages. VS Code served as the main development environment for both frontend and backend coding during the project's development. From creating machine learning scripts to creating user interfaces, its integrated terminal, debugging tools, and extensions for React, Node.js, and Python greatly expedited the implementation process.

2.3.2 Testing

The Employee Allocation and Optimization module's accuracy, dependability, and usability within the larger manpower allocation system were all validated during the testing process. The goal was to make sure the system operated as anticipated in a variety of real world scenarios. A mix of functional, integration, and unit testing methodologies were used in the testing process. Individual components, including the machine learning model, KPI suggestion logic, and API endpoints, were the focus of unit testing. Integration testing evaluated how well the frontend interface, MySQL database, and backend model interacted. To make sure that every feature operated as intended when used through the user interface, functional testing was done from the viewpoint of the end user.

2.3.2.1 Functional Testing

In order to validate the employee allocation and optimization module, functional testing was essential. It was created to make sure the features carried out their intended functions and reacted appropriately to various user inputs and data inputs. The main objective was to verify that project managers could efficiently choose a project, select job titles, and obtain the best staff recommendations based on anticipated KPI thresholds. For efficiency and dependability, every function was evaluated both manually and programmatically.

Verifying the proper selection of projects and job responsibilities via the frontend interface was the first step in the testing process. This input was supposed to be sent to the backend by the system, where a machine learning model would forecast the necessary KPI for the chosen project category and job position. Testing verified that the React based frontend correctly sent the job title and project selection input to the backend. To guarantee correct data transfer, API requests were recorded and verified using Postman and browser development tools.

The trained machine learning model, which had been created using historical data kept in the MySQL database, was invoked by the backend upon receiving the input. In response, the model recommended a KPI value based on the project difficulty and chosen job title. For example, the model produced a KPI score of 82.4 when a project with the job title "Structural Engineer" and the category "High Rise Building" was chosen. The model showed logical KPI estimations that were in line with the training data, and several testing of this flow verified consistency in model predictions.

The system then filtered employee records according to their current KPIs in the database after obtaining the forecasted KPI. The list that was given to the project manager only contained workers whose KPIs reached or beyond the recommended threshold. Using employee records with different KPI ratings, this filtering logic was thoroughly evaluated.

Validating the user interface's capacity to responsively and clearly show the recommended KPI and qualified personnel list was another aspect of functional testing. The KPI result was successfully rendered by the React frontend, which also dynamically updated a list of qualified workers. The employee selection and assignment buttons worked and made the right API calls to update the project employee mapping on the backend. The MySQL database updated when a user chose an employee for allocation, ensuring that data consistency and backend logic were upheld.

Overall, the workflow's smooth operation from input capture to KPI prediction, personnel filtering, and final allocation was confirmed by functional testing. The backend processing was accurate, the user experience was seamless, and every anticipated function operated as intended in a variety of common and uncommon situations. By assisting project managers in making data driven decisions about employee allocation that are in line with project specific performance standards, the system achieved its design goals.

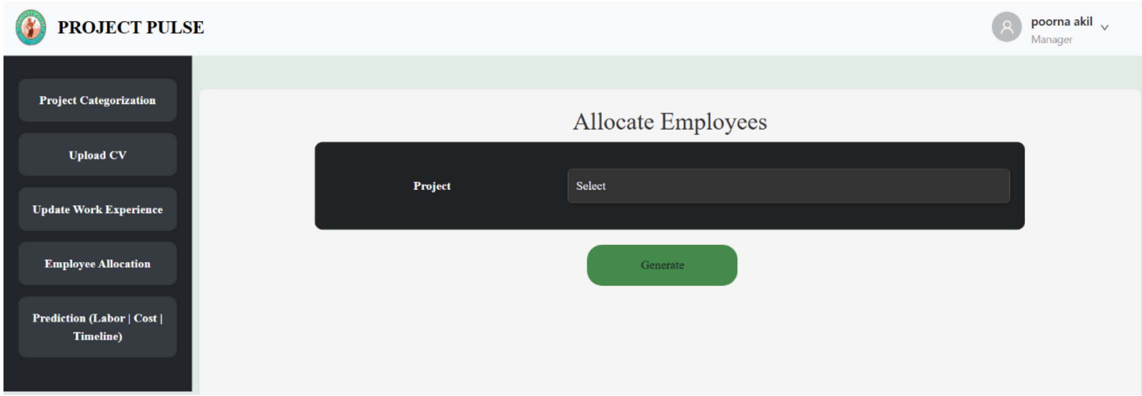


Figure 4 project selection interface

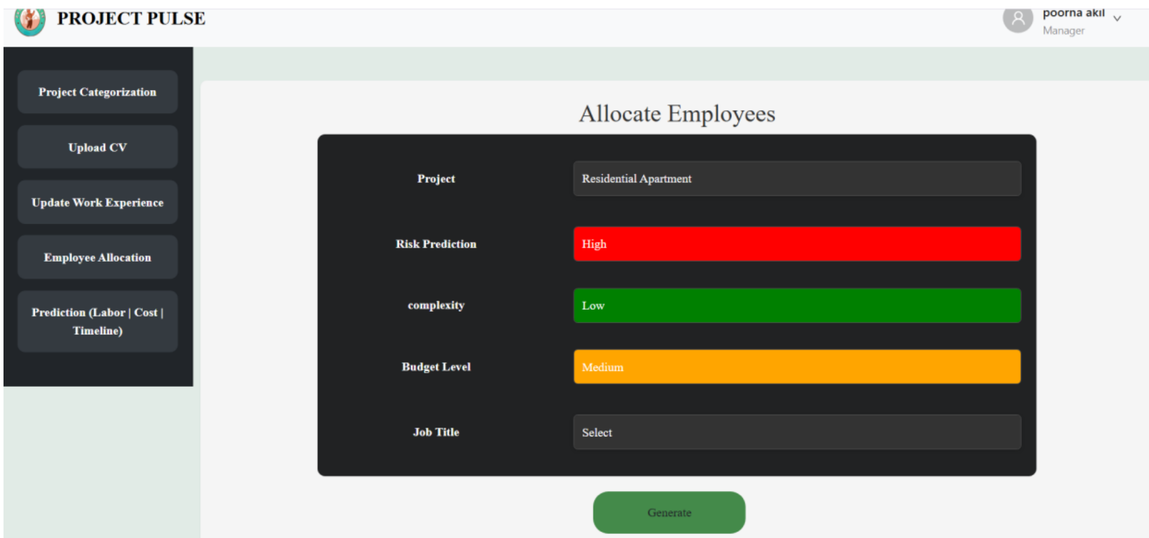


Figure 5 Selected project details and Job title selection interface

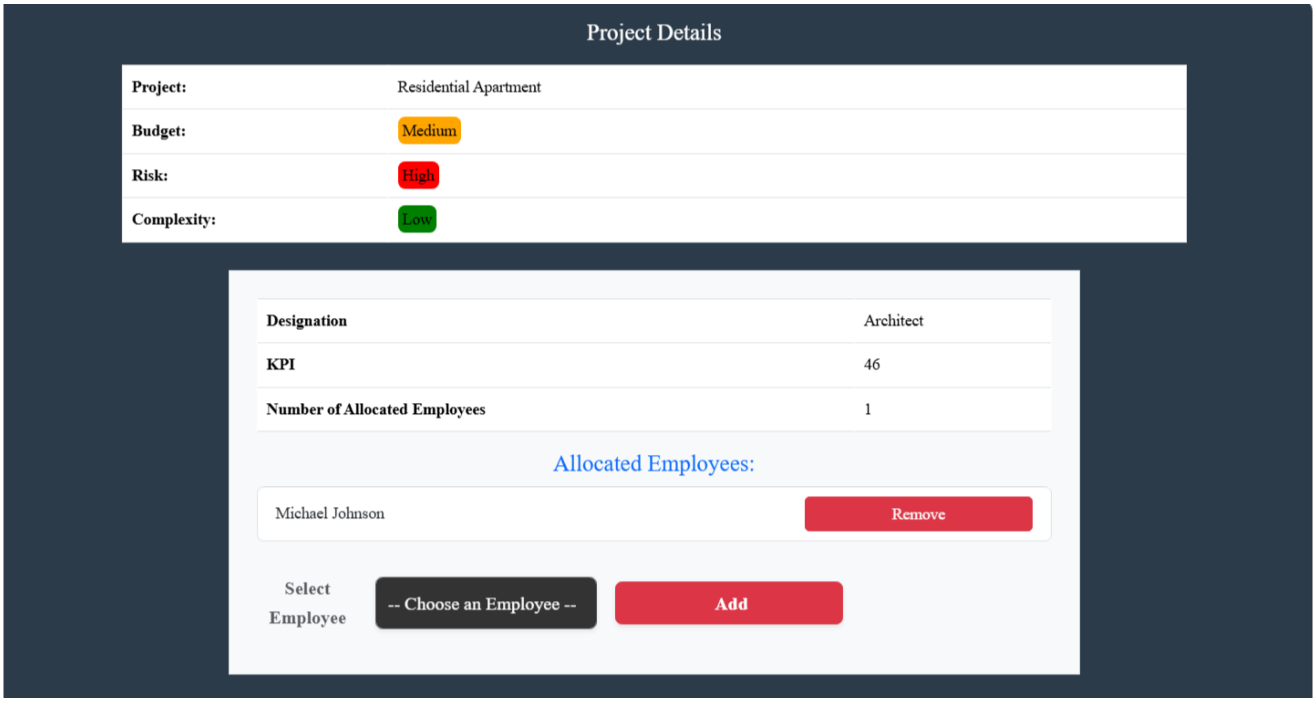


Figure 6 Employee allocation interface

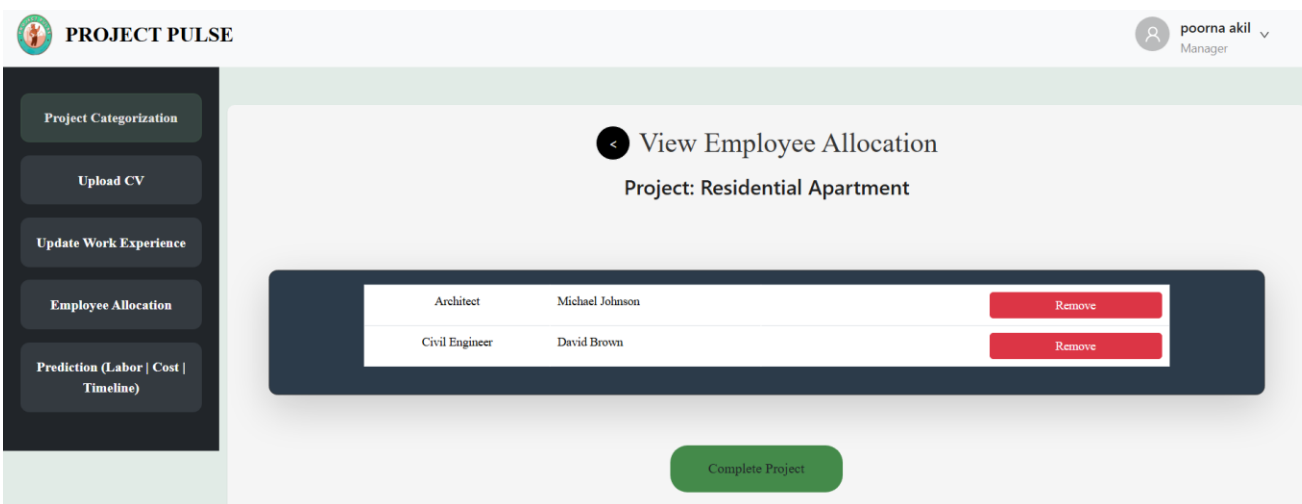


Figure 7 View employee allocation interface

2.3.2.2 Non Functional Testing

Non functional testing ensured the module could function well in real world scenarios beyond functional correctness by assessing the system's overall performance, usability, dependability, and scalability. Verifying that the system could satisfy non functional requirements including speed, responsiveness, security, and compatibility was crucial given the corporate nature of the solution and the significance of smooth integration in building project processes.

Performance testing was done to evaluate how long it took to render the data on the frontend, filter employees, and predict KPIs. After project and job title inputs, the system reliably produced KPI projections in 1 2 seconds, which is suitable for managerial choices made in real time. Even when tested using a bigger dataset of employee records, this response time stayed consistent, suggesting that the model was lightweight enough to handle production level use and that the backend operations were streamlined.

Feedback sessions with users from the target industry were used to conduct usability testing. Team leads and project managers were asked to test the interface and share their thoughts on its usability, clarity, and navigation. The ReactJS built interface was commended for its ease of use and straightforward, step by step process, which enabled users to choose a project, enter job roles, view recommended KPIs, and filter employees in a flash. Even for users who were not experienced with predictive algorithms, a positive user experience was also facilitated by visual uniformity and clear messaging.

Through numerous interactions with the system under many situations, stability and reliability were investigated. Reliable and repeatable behavior was ensured by the model's consistent production of the same KPI value for the same input combinations. Error handling and response issues were also observed in Node.js built backend API calls. The system's fault tolerance was demonstrated by its ability to handle database disconnects and invalid inputs by returning insightful error messages instead of crashing or freezing.

Verifying data protection at the user and system levels was part of the security testing process. The data flow from the frontend to the backend was examined for vulnerabilities including SQL injections and API endpoint exposure, even though the system's authentication features were maintained globally. Database integrity was guaranteed using MySQL's parameterized queries and backend input validation. Furthermore, confidentiality was maintained by limiting access to sensitive employee performance data, such KPI values, to those who had the proper authorization.

In order to assess scalability, scenarios involving concurrent user access and growing numbers of employee records in the database were simulated. There was no discernible decline in response time or system throughput, and the system continued to react effectively with a sizable dataset. This suggests that the design is scalable and capable of accommodating increases in employee data and organizational scale.

All things considered, the non functional testing phase verified that the Employee Allocation and Optimization module is reliable, safe, easy to use, and prepared for deployment in actual construction project settings. By improving user trust, system performance, and long term adaptation, these non functional features contribute value and guarantee the solution's viability and efficacy beyond merely functional output.

3 Results & Discussion

3.1 Results

Through the use of machine learning, the Employee Allocation and Optimization module was created in order to forecast KPI requirements for various job roles across various construction projects. Project complexity classifications, role specific benchmarks, and historical employee performance data were used to train, test, and assess several models in order to guarantee accurate and trustworthy predictions. As evidenced by the outcomes of the model training and evaluation procedure, integrating machine learning for intelligent resource allocation in the construction sector is both feasible and beneficial.

A cleaned and organized dataset from MAGA Engineering Pvt Ltd was used to train each model. It included important characteristics including project ID, project categorization results, and employee KPIs. An 80:20 ratio was used to divide the dataset into training and testing sets.

```
# Define features (X) and target (y)
X = df.drop(columns=["KPI"])
y = df["KPI"]

# Split data into training (80%) and testing (20%)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
print(X_train.shape,y_train.shape)#shape of training data set ..rows and columns
print(X_test.shape,y_test.shape)#shape of testing data set
```

Figure 8 Setting up dataset for training and testing.

Several techniques, such as gradient boosting regression, decision tree regression, random forest regression, and linear regression, were evaluated during the model training process. To maximize its prediction performance, each model underwent cross validation and hyperparameter adjustment. We were able to determine which model was most suited for system integration by looking at the data, which indicated different performance levels for each model.

The evaluation of each model is detailed below.

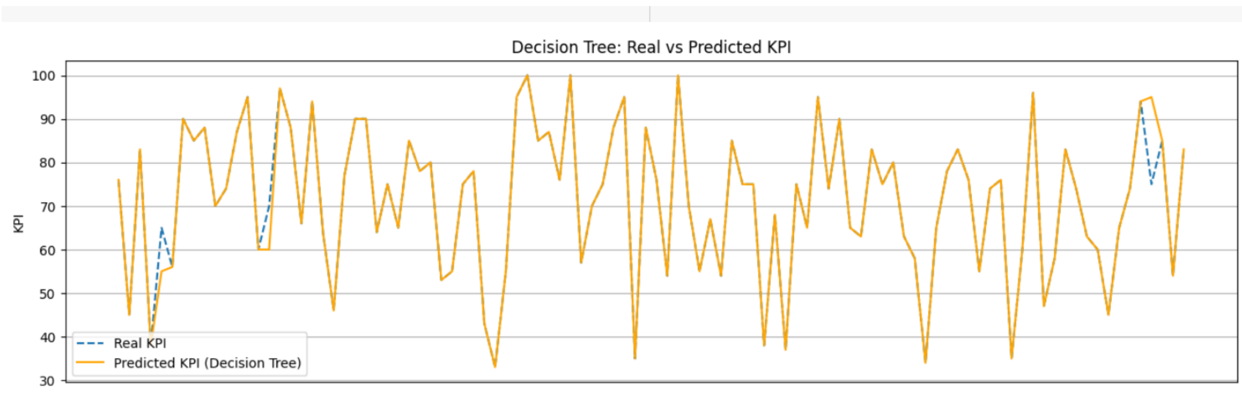


Figure 9 Decision tree model results

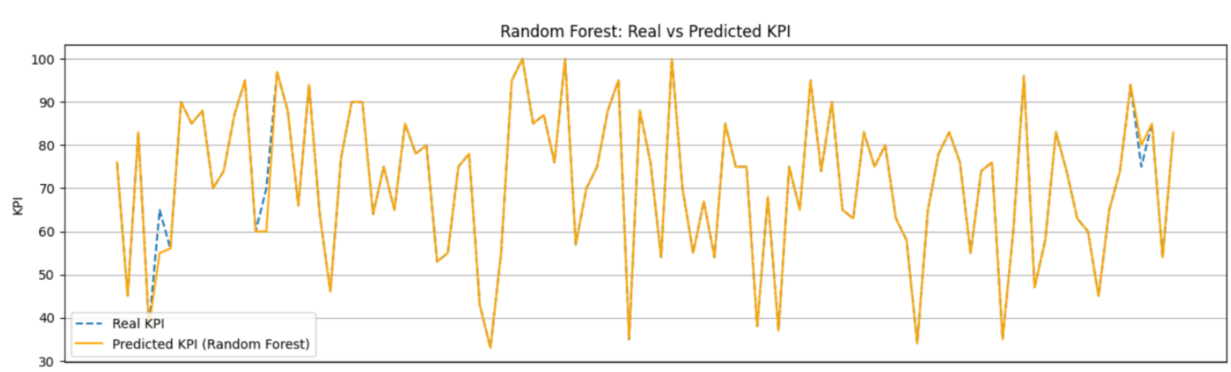


Figure 10 Random Forest model results

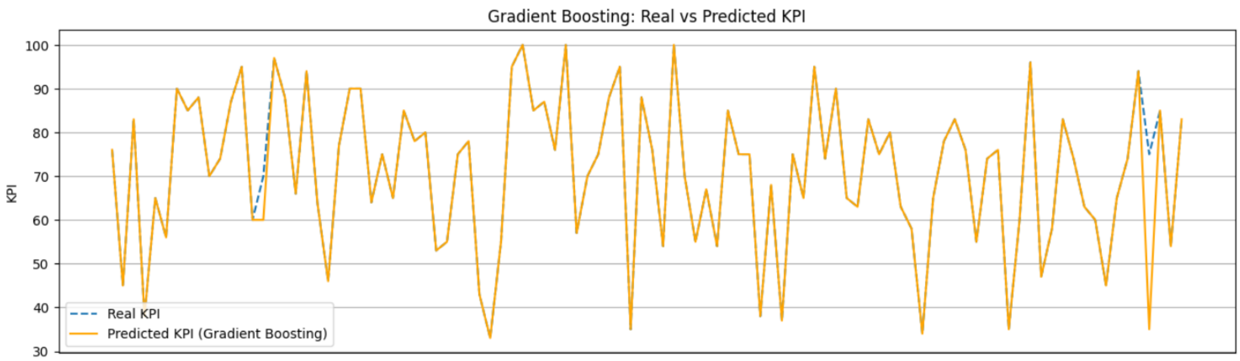


Figure 11 Gradient boosting model results

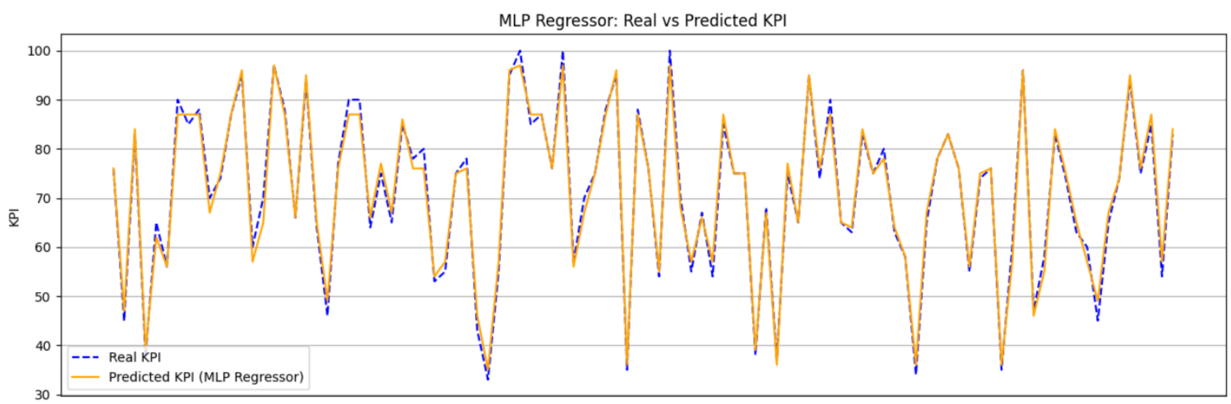


Figure 12 MLP regressor model results

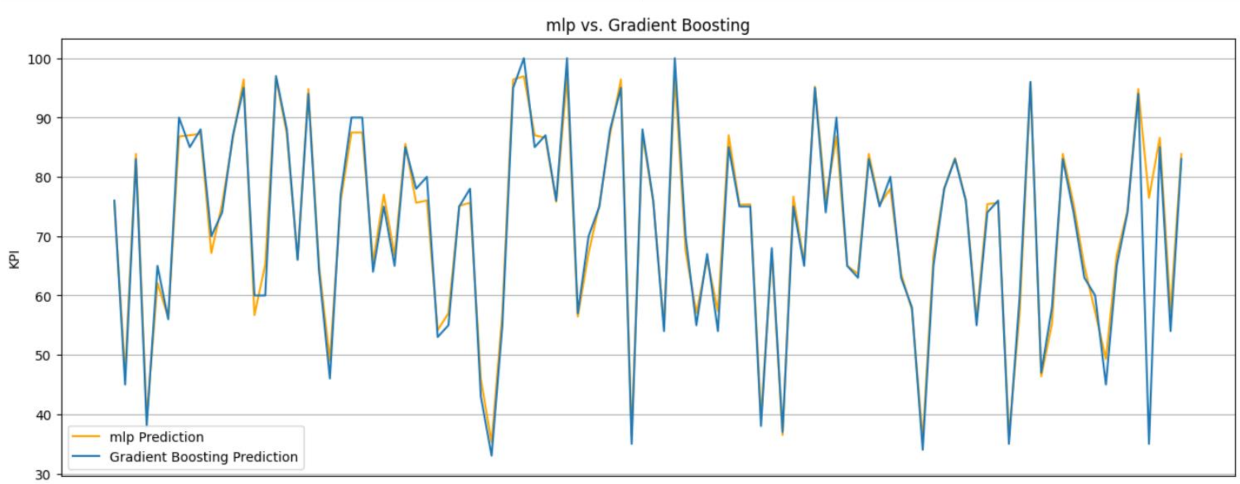


Figure 13 MLP model vs Gradient boosting model comparison

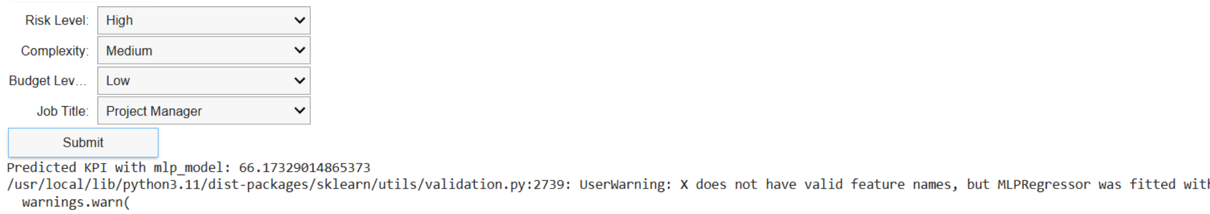


Figure 14 Trained model outcome

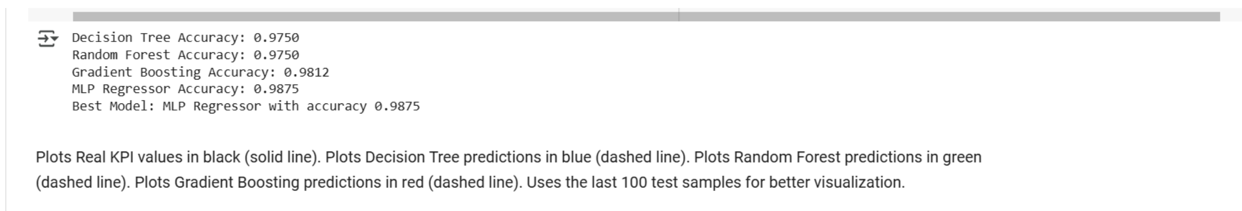


Figure 15 All models' accuracy results

Testing results showed that the final model was more accurate and more generalizable, so it was incorporated into the backend.

3.2 Research Findings

The research conducted for the Employee Allocation and Optimization module provided important new information about how construction firms may use data driven tactics to improve project execution and employee deployment. The feasibility of predicting Key Performance Indicator (KPI) thresholds for various job positions across a range of project difficulties using machine learning models was one of the most notable discoveries. In order to deploy the best individuals for future construction projects, the system may intelligently recommend KPI levels by evaluating past project and employee performance data.

Predictive analytics integration into manpower allocation showed significant potential in lowering the inefficiencies associated with human decision making. The capabilities of personnel and project requirements were frequently out of sync when traditional techniques of assigning employees based on prior experiences or subjective evaluation were used. With the help of our system, project managers may now choose employees for a project based on unbiased recommendations supported by data. This is in line with the larger industry trend of implementing intelligent automation to assist with resource management and strategic planning.

An other significant discovery was the strong association between total project performance and appropriately assigned personnel (based on KPI compatibility). The methodology predicted increases in productivity, shorter task completion times, and improved adherence to project timetables when personnel were paired with projects with KPI levels that mirrored their abilities and prior performance. The significance of optimization techniques in resource allocation, output quality, and operational efficiency is highlighted by these studies.

The study also emphasized the need of appropriate feature selection and data preprocessing. The quality of the input data, in particular the applicability of characteristics like job title, project classification, and previous KPIs, had a direct impact on the prediction models' accuracy. This suggests that in order to get successful results, firms looking to deploy comparable systems must make an investment in keeping precise and organized data repositories.

Additionally, the research revealed a user centric insight: in order to achieve trust and usability in non technical industries like construction, automation must be presented through an intuitive and user friendly interface. Project managers valued how easy it was to choose projects and evaluate suggested KPIs, according to input received during usability testing. This helped close the gap between sophisticated data science methods and routine operational decision making.

In summary, the study shown that the use of machine learning may greatly enhance employee allocation in the construction sector. The foundation for a scalable and reproducible solution has been established by the combination of predictive modeling, user friendly interface design, and smooth interaction with real time data. These results point to directions for the future development of such systems in related fields and add important knowledge to the developing field of intelligent labor management

.

3.3 Discussion

The research's Employee Allocation and Optimization module offers an innovative solution to one of the construction industry's most enduring problems: the effective use of human resources. Decisions about employee assignments in the construction industry have historically mostly depended on human judgment, experience based reasoning, or basic filtering criteria, frequently neglecting more profound analytical insights concealed within organizational data. By combining data driven approaches with machine learning, this study presents a scalable system that encourages precision, impartiality, and strategic workforce planning.

The idea that past project data and personnel KPIs can function as trustworthy predictors of future position fit was confirmed by the successful application of several predictive models throughout the experimentation phase. In addition to performing better in terms of prediction accuracy, models like Random Forest and Gradient Boosting also showed that they could generalize to different project categories and job roles. Project managers were successfully able to choose a project and acquire a recommended KPI benchmark for each necessary job title according to the technological developments. The most qualified candidates might then be filtered and shortlisted by comparing this benchmark with employee data, automating a traditionally laborious and error prone procedure.

The significance of feature engineering and data preprocessing is one of the main topics of discussion from a technological perspective. Employment history, employee performance ratings, project complexity, and other relevant and high quality input variables all had a significant impact on how well the model predicted outcomes. As a result, data organization and cleansing were crucial in determining how well the model training procedure went. Furthermore, utilizing Node.js to integrate the machine learning models with the backend and a React based UI to serve results dynamically yielded an end to end solution that complied with real world deployment criteria.

In conclusion, the system's actual use in a construction setting indicates that there is a great chance for improvement in the future. Predictive scheduling, budget alignment, and even resource balance across several projects might be added to the underlying structure, even though the current capabilities are restricted to KPI based filtering. This study offers a first step toward modern workforce management and operational efficiency in the construction industry through clear and organized procedures.

3.4 Summary of Each Student's contribution

3.4.1 Project Categorization Component

The Project Categorization module serves as a foundational pillar of the automated manpower allocation system, playing a critical role in classifying construction projects based on essential parameters such as risk level, complexity, estimated budget, project duration, and geographical location. This classification enables a structured alignment between project needs and labor capabilities, thereby improving efficiency and reducing resource waste. The implementation of this module began with the collection of historical project data from MAGA Engineering Pvt Ltd, including scope definitions, objectives, financial estimates, timelines, and location-specific factors. This data was used to train a machine learning-based decision tree algorithm, which categorizes projects based on two primary dimensions: risk (low, medium, or high) and complexity (structural, environmental, or logistical). Key input parameters include project type (e.g., residential buildings, roads, irrigation systems), site-specific conditions, and environmental risk factors. By replacing subjective manual classification with an automated, rule-based process, this module enhances consistency, predictability, and objectivity.

This component proves highly valuable in enabling preemptive decisions about manpower allocation and risk mitigation. High-risk or high-complexity projects can be identified in advance and assigned staff with proven expertise in those domains. For example, a complex bridge construction project with high environmental impact would be matched with employees experienced in those areas. This upfront alignment significantly reduces the likelihood of cost overruns, delays, or safety issues. The module integrates seamlessly with the other subsystems, such as KPI Generation and Employee Allocation, by providing categorized project profiles as inputs to the downstream decision-making processes. From a technical standpoint, Python was used to develop the decision tree algorithm, React.js was used to design an interactive frontend where project data is entered, Node.js powered the backend logic, and MySQL facilitated persistent storage and data retrieval.

In the future, the module's accuracy and scope can be improved by enriching the dataset with more diverse project types and incorporating real-time inputs such as weather data and updated environmental risks. Integration of geospatial data analysis could allow for better evaluation of location-specific challenges like urban congestion or remote access issues. These upgrades would increase the model's robustness and adaptability, enabling smarter and more informed workforce planning in complex construction environments. In conclusion, the Project Categorization module not only removes inefficiencies caused by manual classification methods but also establishes a scalable and intelligent foundation for project planning and manpower deployment.

3.4.2 KPI Generation by Performance and CV Analysis Component

The KPI Generation module brings innovation to workforce evaluation by introducing an automated system for computing employee Key Performance Indicators (KPIs) based on in-depth analysis of resumes, past work experience, and historical project contributions. This represents a shift away from subjective and inconsistent appraisal methods, offering instead an objective and quantifiable system that enhances fairness and transparency in employee assessments. The process begins when employees submit their CVs via a dedicated portal, after which natural language processing (NLP) algorithms powered by a Multinomial Naive Bayes classifier extract and categorize relevant information such as technical competencies, professional qualifications, certifications, and experience. This data then serves as the basis for KPI calculations, which are distributed across three dimensions: 60% performance metrics, 20% core competencies, and 20% supplementary factors like seniority or leadership roles.

A key advantage of this component is its dynamic update mechanism, which ensures that employee KPIs are continuously revised in response to newly completed projects, training sessions, and evolving responsibilities. The system's architecture was developed using Python for both NLP and machine learning functionality, React.js for the user-facing dashboard accessible by employees and HR teams, Node.js for backend integration, and MySQL for secure and scalable data management. The utility of this system extends beyond performance evaluation it enhances strategic human resource planning, reduces bias in promotions and rewards, and allows managers to make well-informed staffing decisions.

Future enhancements planned for this module include AI-assisted skill gap analysis, which would allow the system to predict future training needs and guide employee development proactively. The integration of managerial feedback as qualitative input would provide context to the quantitative KPIs, creating a more balanced assessment. Improvements in CV parsing through deep learning models will increase accuracy, especially when processing diverse and non-standard resume formats. These upgrades, combined with visual reporting tools, will enhance the system's value as both a performance management and organizational planning tool. Ultimately, this module not only evaluates performance but helps improve it by giving management deeper insight into workforce capabilities and development potential.

3.4.3 Employee Allocation and Optimization Component

The Employee Allocation and Optimization module is responsible for intelligently matching employees to projects using a data-driven approach that considers project classification and employee KPIs. The aim is to optimize personnel deployment by ensuring the most suitable individuals are assigned to roles that align with their experience, skills, and availability. This module consolidates the outputs from the project categorization and KPI generation components to perform an intelligent matching routine. Using a decision tree-based model, employees are matched to projects by comparing project requirements to employee qualifications, while also accounting for other operational constraints like workforce availability, scheduling conflicts, and cross-project workloads.

Unlike traditional allocation processes that rely heavily on subjective judgment or static spreadsheets, this automated method ensures optimal fit by factoring in both qualitative and quantitative criteria. It prevents overutilization of top-performing employees and ensures balanced team distribution across multiple active projects. Historical project data is referenced during the validation phase to evaluate the effectiveness of past allocation decisions and adjust the model accordingly. This feedback loop helps fine-tune future allocation processes to avoid inefficiencies observed in earlier assignments.

The technical stack used in this component includes Python for implementing the optimization logic, Node.js for managing data exchange between the frontend and backend, React.js for the dashboard where project managers can make real-time allocation decisions, and MySQL for tracking and storing project-employee assignments. The system empowers project managers by providing a user-friendly interface through which they can make informed decisions quickly and confidently. Employees are also more likely to be satisfied and productive when assigned roles that align with their capabilities and growth paths.

Looking ahead, potential improvements include the integration of team synergy analytics, which would evaluate past collaboration histories to form high-performing teams. A conflict resolution module could be added to identify and resolve issues like scheduling overlaps or double allocations automatically. Additionally, predictive staffing models could be employed to forecast future manpower needs, helping project managers plan ahead and maintain optimal staffing levels. These enhancements will make the system even more dynamic and responsive, further supporting construction firms in executing projects with efficiency and precision.

3.4.4 Labor, Cost, and Timeline Prediction Component

The Labor, Cost, and Timeline Prediction module adds strategic forecasting capabilities to the system by estimating resource requirements, budget, and scheduling needs based on historical project patterns and real-time performance monitoring. The system is designed to anticipate the number of laborers needed at each phase of a construction project, using historical trends related to project complexity, scale, and past task durations. These forecasts allow project managers to plan manpower deployment accurately and avoid both overstaffing and understaffing.

The component also includes real-time labor tracking via a dashboard that monitors daily attendance and task completion rates. This enables managers to make timely decisions about reallocating labor and adjusting schedules before problems escalate. Cost and timeline predictions are powered by regression models trained on budget deviations and delays from completed projects. These models can forecast budget overruns or scheduling risks, allowing managers to maintain tighter control over project execution. The frontend dashboard, developed using React.js and Power BI, visualizes these predictions, making them easy to interpret and act upon. Python was used for the core regression models, and MySQL was used to store labor-related data securely and efficiently.

The system significantly reduces risks associated with poor planning, providing managers with early warnings and enabling more agile responses to disruptions. For future enhancements, integration with external data sources such as weather forecasts and supply chain updates could improve prediction accuracy, given the significant impact these factors have on construction timelines. A mobile app extension would also support on-site labor updates in real time. Additionally, reinforcement learning could be introduced to allow the system to continuously learn from new project data and automatically adjust its predictions. These future additions would further strengthen the system's ability to manage large-scale construction projects efficiently and intelligently.

4 Conclusion

This study focused to address a persistent problem in the construction sector, primarily the inefficiencies associated with allocating workers to projects based on personal experience rather than performance-based standards. By creating a structured, machine learning-based system for Employee Allocation and Optimization, this study offered a scalable, data-driven substitute that helps project managers make more intelligent staffing decisions. MAGA Engineering Pvt Ltd, one of Sri Lanka's top construction companies, served as the background for the system's design and evaluation, offering practical insights and relevance to the project's results.

The suggested system included a method for gathering and classifying employee performance data in the form of Key Performance Indicators (KPIs). It then uses this data along with the outputs of project categorization to suggest the most suitable KPI for the job titile. A trained machine learning model is fed job title and project factors (such complexity, risk, and budget) to forecast appropriate KPI thresholds needed for role performance in a particular project setting. Project managers can then choose the best applicants from a filtered list of eligible people by comparing the model's predictions with the roster of workers in the company's workforce database.

Several machine learning techniques, such as Random Forest, Gradient Boosting, Decision Tree Regression, and MLP Regressor, were evaluated during the model training stage. MLP Regressor had the highest R2 score and the lowest prediction error during validation, making it the most accurate and consistent of them. The robustness and interpretability of this model made it the choice for deployment in the final implementation. These qualities are particularly crucial in industry contexts where users demand transparency in system recommendations. A clean, organized dataset containing past KPI data, job roles, and project details was used to train the model. Feature engineering was crucial to the model's effectiveness, emphasizing how crucial high quality data is to any decision-support system.

Utilizing a modern tech stack, the implementation made sure the system was responsive and flexible. The model was developed and trained using Python, and the main platform for experimentation was Google Colab. The model was exported after training and incorporated into a Node.js developed backend environment, where a MySQL database held all pertinent project and employee data. User interaction was made easy and intuitive by the ReactJS based interface, which made it easier to choose projects, enter job roles, and display model generated KPI recommendations and suggested workers. Modular testing of the entire system was done to make sure that the machine learning model integrated seamlessly with user interfaces and real-time data.

In terms of functionality, the system achieved all of its main goals. Choosing a project and viewing KPI recommendations for various roles allowed project managers to quickly identify qualified applicants from the employee pool. Through testing and informal user input, the interface's usability and prediction accuracy were both positively confirmed. The technology offered increased transparency and fairness in the allocation process while drastically reducing the human labor intensive task of employee filtration. While non-functional testing verified the system's performance, scalability, and security, functional testing enabled the correct operation of model invocation, filtering logic, and database interfaces.

Beyond system development, this research has outcomes. From a conceptual standpoint, the project shows that data driven labor management is not only possible but also very beneficial in sectors like construction, where manual and ad hoc resource planning has historically been the norm. The results indicate that when properly applied, objective metrics like KPIs can significantly enhance project outcomes by matching workers with positions that best fit their skills. Additionally, standardizing performance expectations through KPI forecasts promotes a work culture that is more performance oriented and accountable.

This method also adds to the broader discussion about the building industry's digital revolution. For scheduling, budgeting, and documentation, many local businesses have embraced digital solutions yet, human resource planning has mostly stayed the same. This study closes that gap by proposing a way to use predictive analytics to human capital management. Currently absent from most businesses, the ability to predict KPI demands based on project characteristics and match workers properly adds a strategic dimension to employee allocation planning.

Nevertheless, there are some restrictions on the research. Team compatibility measurements, behavioral data, and real-time availability updates were not included in the scope, which was restricted to personnel allocation based on KPI criteria. Though enough for testing and assessment, the training dataset might be enlarged in the future to incorporate additional projects, a wider range of employee roles, and long-term performance monitoring. Furthermore, although the system made suggestions, project managers had the final say. This strikes a good balance between automation and human oversight, but it might be improved with user input and ongoing learning tools.

A dynamic feedback loop could be developed in future work to enable the system to learn from allocation outcomes and improve its predictions over time. Organizations could centralize all project, personnel, and performance data under a single system by pursuing integration with larger Enterprise Resource Planning (ERP) platforms. The system's intelligence and adaptability could be further enhanced by sophisticated features including predictive risk assessments, real-time availability checks, and team synergy analysis.

In conclusion, a system that revolutionizes the way construction companies distribute their workforce has been effectively created and tested by this research. In complicated construction projects, the system provides a dependable, scalable, and user-friendly solution for staff allocation by fusing structured data, performance analysis, and machine learning algorithms. Efficiency, accuracy, and managerial decision support have all demonstrated measurable gains, making it a useful tool for the sector and a solid basis for future labor optimization innovation.

5 References

[1] R. Ahmed, S. Azhar, M. Castillo and P. Kappagantula, "Project failure factors and their impacts on the construction industry: A literature review," *International Journal of Construction Education and Research,* vol. 15, no. 3, p. 201–217, 2019.

[2] X. Zhao and X. Wang, "Aligning employees with the organization's strategic objectives: Out of 'line of sight,' out of mind," *Journal of Business Strategy,* vol. 41, no. 1, p. 23–31, 2020.

[3] Microsoft Corporation, "Microsoft Project: Project Management Software," 2023. [Online]. Available: https://www.microsoft.com/en-us/microsoft-365/project. [Accessed 20 03 2025].

[4] Oracle Corporation, "Oracle Primavera P6: Enterprise Project Portfolio Management Software," 2023. [Online]. Available: https://www.oracle.com/industries/engineering-construction/primavera-p6/. [Accessed 20 3 2025].

[5] J. K. Chowdhury, "Natural Language Processing," in *Encyclopedia of Artificial Intelligence*, IGI Global, 2009, p. 1105–1110.

[6] BuildTrend, "Construction Project Management Software," 2023. [Online]. Available: https://www.buildtrend.com/.

[7] ALICE Technologies, "AI-Powered Construction Optimization Software," 2023. [Online]. Available: https://www.alicetechnologies.com/.

[8] Autodesk, "PlanGrid Construction Software," 2023. [Online]. Available: https://www.autodesk.com/products/plangrid/overview.

[9] MAGA Engineering Pvt Ltd, "Corporate Profile," 2023. [Online]. Available: https://maga.lk/.

[10] P. Mell and T. Grance, "he NIST Definition of Cloud Computing," National Institute of Standards and Technology, 2011.

[11] Python Software Foundation, "Python Programming Language – Official Website," 2023. [Online]. Available: https://www.python.org/.

[12] F. e. a. Pedregosa, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research,* vol. 12, p. 2825–2830, 2011.

[13] W. McKinney, "Data Structures for Statistical Computing in Python," in *9th Python in Science Conf. (SciPy)*, 2010.

[14] T. E. Oliphant, A guide to NumPy, USA: relgol Publishing, 2006.

[15] Joblib Developers, "Joblib: Python object serialization for machine learning," 2023. [Online]. Available: https://joblib.readthedocs.io/.

[16] Meta Platforms Inc., "React – A JavaScript Library for Building User Interfaces," 2023. [Online]. Available: https://reactjs.org/.

[17] OpenJS Foundation, "Node.js – JavaScript Runtime Built on Chrome's V8 Engine," 2023. [Online]. Available: https://nodejs.org/.

[18] Oracle Corporation, "MySQL: The World's Most Popular Open Source Database," 2023. [Online]. Available: https://www.mysql.com/.

[19] Google Research, "Google Colaboratory – Free Python Jupyter Notebook Environment," 2023. [Online]. Available: https://colab.research.google.com/.

[20] Microsoft Corporation, "Visual Studio Code – Code Editing. Redefined.," 2023. [Online]. Available: https://code.visualstudio.com/.

Appendices

IT21069840_Final_Individual_Research_Report

ORIGINALITY REPORT

5%	4%	1%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Sri Lanka Institute of Information Technology Student Paper	1 %
2	Submitted to Karabük Üniversitesi Student Paper	<1 %
3	Submitted to Maastricht University Student Paper	<1 %
4	dspace.daffodilvarsity.edu.bd:8080 Internet Source	<1 %
5	studymoose.com Internet Source	<1 %
6	Submitted to Technological University Dublin Student Paper	<1 %
7	Submitted to ESoft Metro Campus, Sri Lanka Student Paper	<1 %
8	Submitted to Middlesex University Student Paper	<1 %
9	esource.dbs.ie Internet Source	<1 %
10	Vorster, Eswee. "Digital Innovation Strategy: A Framework for Creating Business Value for an Airline Organisation.", University of Pretoria (South Africa), 2024 Publication	<1 %
	pure.eur.nl	

11	Internet Source	<1%
12	www.geeksforgeeks.org Internet Source	<1%
13	patents.justia.com Internet Source	<1%
14	appmaster.io Internet Source	<1%
15	www.sci.ruh.ac.lk Internet Source	<1%
16	Submitted to National College of Ireland Student Paper	<1%
17	www.projectmanagement.com Internet Source	<1%
18	Andrés Bastidas Fuertes, María Pérez, Jaime Meza. "nMorph framework: An Innovative Approach to Transpiler-Based Multi-Language Software Development", IEEE Access, 2023 Publication	<1%
19	elibrary.tucl.edu.np Internet Source	<1%
20	ijircce.com Internet Source	<1%
21	www.mdpi.com Internet Source	<1%
22	www.nature.com Internet Source	<1%
23	www.uniassignment.com Internet Source	<1%

24	Edwing, Deanna. "Characterizing Delaware Bay Coastal Inundation Using Sentinel-1 SAR Imagery and Deep Learning", University of Delaware, 2023 Publication	<1 %
25	Gaurav Aggarwal, Ashutosh Tripathi, Himani Goyal Sharma, Tripti Sharma, Rishabh Dev Shukla. "Integrated Technologies in Electrical, Electronics and Biotechnology Engineering", CRC Press, 2025 Publication	<1 %
26	R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P. Prasad. "Algorithms in Advanced Artificial Intelligence - Proceedings of International Conference on Algorithms in Advanced Artificial Intelligence (ICAAAI-2024)", CRC Press, 2025 Publication	<1 %
27	dl.lib.uom.lk Internet Source	<1 %
28	webthesis.biblio.polito.it Internet Source	<1 %
29	www.coursehero.com Internet Source	<1 %

Exclude quotes On
Exclude bibliography On
Exclude matches < 5 words

AUTOMATED MANPOWER ALLOCATION BY PERFORMANCE ANALYSIS AND PROJECT CATEGORIZATION FOR CONSTRUCTION PROJECTS (PROJECT PULSE)

Kaushalya Hiruni Munagama

(IT21270956)

Dissertation submitted in partial fulfilment of the requirements for the Bachelor

of Science (Hons) in Information Technology Specializing in Information Systems Engineering

Department of Computer Systems Engineering
Sri Lanka Institute of Information Technology
Sri Lanka

April 2025

AUTOMATED MANPOWER ALLOCATION BY PERFORMANCE ANALYSIS AND PROJECT CATEGORIZATION FOR CONSTRUCTION PROJECTS (PROJECT PULSE)

Kaushalya Hiruni Munagama

(IT21270956)

Dissertation submitted in partial fulfilment of the requirements for the Bachelor of Science (Hons) in Information Technology Specializing in Information Systems Engineering

Department of Computer Systems Engineering
Sri Lanka Institute of Information Technology
Sri Lanka

April 2025

DECLARATION

I declare that this is my own work, and this dissertation does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other university or Institute of higher learning, and to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgment is made in the text. Also, I hereby grant to Sri Lanka Institute of Information Technology the non-exclusive right to reproduce and distribute my dissertation in whole or part in print, electronic, or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Name	Student ID	Signature	Date
Munagama M.K.H	IT21270956	Kaushly	10/04/2025

The supervisor/s should certify the proposal report with the following declaration.
The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Name	Signature	Date
Mrs.Buddhima Attanayake		15/05/25

ABSTRACT

The construction industry faces ongoing challenges in managing labor allocation, cost forecasting, and project timelines, often leading to inefficiencies, delays, and budget overruns. Traditional methods for predicting labor needs and scheduling projects frequently fall short of addressing these issues. This research develops an advanced system that utilizes historical project data, machine learning algorithms, and real-time updates to predict labor requirements, forecast project timelines, and estimate costs. By incorporating key project parameters such as square footage, number of floors, windows, and timelines, the system dynamically adjusts labor predictions, project schedules, and cost estimates as real-time data from the construction site is continuously integrated.

In this study, various machine learning techniques, including Random Forest Regression, Linear Regression, and Deep Convolutional Neural Networks (DCNN), were applied to predict labor requirements and project outcomes. The findings revealed that Random Forest provided the most accurate predictions, achieving an accuracy of 87.28%, followed by DCNN at 75.39% and Linear Regression at 73.68%. These results were validated through precision, recall, and F1-score analyses, with Random Forest consistently outshining the other models in terms of reliability.

Additionally, the system incorporates real-time data updates to monitor labor attendance, task progress, and deviations from planned schedules. This enables project managers to make timely adjustments, ensuring that projects remain on track. The system's scalability allows it to handle large datasets and multiple concurrent projects, making it a viable solution for construction firms of various sizes.

Ultimately, this research demonstrates the potential of integrating machine learning, real-time data, and predictive analytics to create a more flexible and efficient project management system. By filling existing gaps in labor prediction and scheduling tools, the system offers a more precise, adaptive, and cost-effective solution for managing construction projects. The successful implementation of this system can significantly improve labor management, reduce delays, and enhance overall project efficiency in the construction industry.

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to my dissertation supervisor, Mrs. Buddhima Attanayake, my Co-Supervisor, Mrs. Narmada Gamage, and my external supervisor, Mr. Darshana Senevirathne, for their unwavering support and invaluable guidance throughout my research journey. Their thoughtful feedback and constructive critiques have been pivotal in refining my work and shaping the direction of my study.

I am also deeply thankful to the management and staff of MAGA Construction Pvt Ltd for their ongoing encouragement and support throughout my academic path. Their commitment to excellence in the construction industry has continuously inspired and motivated me.

A special thank you goes to my teammates for their collaborative spirit and support. The insightful discussions and exchange of ideas have played a crucial role in broadening my understanding of the subject matter and enhancing the overall quality of my research.

Lastly, I am profoundly grateful to all those who contributed their time and expertise to this study. Their willingness to share their knowledge and experiences has not only enriched my research but also deepened my understanding of the topic.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Table 1 :Abbreviations

Abbreviation	Description
DCNN	Deep Convolutional Neural Network
mAP	mean Average Precision

1. Introduction

1.1 MAGA Construction Pvt Ltd

MAGA Construction Pvt Ltd is one of Sri Lanka's most trusted and respected construction companies, known for its expertise in handling large-scale infrastructure projects, ranging from commercial and residential buildings to industrial complexes. Founded with the goal of providing exceptional construction services, MAGA has built a solid reputation for its reliability, quality, and integrity in every project it undertakes. Over the years, the company has contributed significantly to shaping the country's skyline, with numerous landmark projects that showcase its capabilities and commitment to excellence [1].

The company's success can be attributed not only to its technical expertise but also to its deep-rooted commitment to sustainability, safety, and innovation. MAGA Construction adopts modern techniques and cutting-edge project management tools to ensure each project is delivered on time, within budget, and with minimal environmental impact. Their proactive approach to incorporating technology into their processes has helped them maintain a competitive edge in the industry, enabling them to efficiently manage complex projects and deliver superior results.

MAGA's expertise extends across various sectors, with a strong focus on infrastructure and building development. What truly sets them apart is their ability to manage large teams and subcontractors, ensuring every detail is handled seamlessly from the planning stages through to project completion. Whether it's coordinating logistics, managing labor, or ensuring safety standards are met, MAGA Construction is well-equipped to handle the complexities of any project. They believe in the power of technology to streamline their operations and enhance project efficiency, making it possible to deliver high-quality work while reducing costs and delays.

What also distinguishes MAGA Construction is its ongoing commitment to research and development. The company understands that the construction landscape is ever evolving, and they're always looking for new ways to improve. By exploring innovative methods to enhance project delivery and boost workforce productivity, MAGA continues to lead the way in the industry. Their focus on leveraging technological advancements, particularly in areas such as labor management, cost estimation, and timeline prediction, is not just about staying competitive—it's about raising the bar for the entire industry and ensuring that they're always ahead of the curve [1].

Through these efforts, MAGA Construction remains dedicated to providing its clients with outstanding results, helping to shape Sri Lanka's future while continuing to be a leader in the construction sector.

1.2 Background and Literature Survey

The primary goal of this research is to develop a system that predicts labor requirements, project timelines, and costs for construction projects. By focusing on key parameters such as square footage, number of floors, windows, and project timelines, this study examines the role of machine learning in improving resource management, labor cost forecasting, project scheduling, and the integration of real-time data for dynamic predictions. This section reviews relevant literature that sheds light on the use of machine learning in construction project management, especially in labor allocation and forecasting.

1. Machine Learning for Labor Prediction in Construction Projects

Machine learning (ML) techniques have gained considerable attention in construction project management, particularly for predicting labor requirements and allocation. These methods assist project managers in making more informed decisions about resource distribution, ultimately improving efficiency, and reducing costs.

Gupta et al. (2019) examined the potential of machine learning for predicting costs in construction projects. Their study revealed that ML models, including linear regression, decision trees, and random forests, can accurately forecast project costs based on historical data and project characteristics. Notably, their findings highlighted the effectiveness of Random Forest in labor prediction, as it can handle the complexities of construction data with ease. The ability of Random Forest to consider various features and interactions allows it to predict labor needs more accurately compared to other models [2].

In a similar vein, Park and Kim (2020) explored the use of deep learning, specifically Convolutional Neural Networks (CNNs), for predicting labor costs. While their approach yielded promising results, they noted that CNNs require significant computational resources, making them less suitable for smaller projects or situations with limited datasets. In contrast, Random Forest was identified as a more efficient and scalable option, offering reliable labor predictions without incurring high computational costs [3].

2. Labor Allocation and Resource Management

Proper labor allocation is essential to ensure that projects run smoothly and efficiently. Ensuring the correct number of workers are assigned to tasks prevents issues like labor shortages or overstaffing, both of which can lead to delays and inflated costs. Machine learning has proven to be a powerful tool in optimizing labor allocation and resource management.

Singh et al. (2021) discussed the importance of integrating real-time data into construction project management, emphasizing how real-time updates on labor attendance, task completion, and overall project progress help improve decision-making. They found that by continuously updating labor predictions based on fresh data, machine learning models can provide a more accurate reflection of the project's current status, allowing project managers to adjust labor allocation as needed [4].

Similarly, Rahman and Saif (2020) proposed the use of ML models for predicting labor costs in residential construction. Their study demonstrated how integrating historical data on labor rates and project details could improve the efficiency of labor allocation, leading to better cost control. The real-time integration of project data further enhanced the accuracy of these predictions, ensuring that labor forecasts were continuously adjusted to match the project's evolving needs [5].

3. Real-Time Data Integration in Construction Projects

Incorporating real-time data into construction project management has emerged as a key development in enhancing labor prediction models. By integrating data from sensors, IoT devices, and other real-time inputs, machine learning systems can adapt to changes on-site and adjust labor forecasts accordingly, improving project accuracy and responsiveness.

Lee et al. (2019) focused on the use of real-time data, particularly from IoT devices and smart sensors, to improve labor predictions. Their study demonstrated how automated data collection reduced the potential for manual errors and improved the accuracy of labor predictions. For instance, if workers were absent or tasks were delayed, the system could automatically adjust labor forecasts, ensuring the project timeline remained intact [6].

Zhang and Li (2019) also explored the integration of sensor networks for real-time monitoring of project progress. These networks provided ongoing updates on worker productivity and task completion, enabling project managers to make more informed decisions about labor allocation. By continuously gathering data, these systems allowed for dynamic adjustments, improving overall project efficiency and reducing the risk of delays [7].

4. Labor Cost Forecasting in Construction

Accurate forecasting of labor costs is vital for maintaining project budgets and ensuring profitability. Several studies have used machine learning to predict labor costs, considering factors like project complexity, labor types, and project timelines.

Zhang and Wei (2019) utilized deep learning models for labor cost forecasting, showing that techniques like Deep Convolutional Neural Networks (DCNNs) could significantly enhance cost predictions. However, they noted that the computational intensity of deep learning models made them less suitable for smaller projects. In these cases, simpler models like Random Forest were found to be more practical while still delivering robust predictions [8].

Rahman and Saif (2020) also applied machine learning techniques, specifically Random Forest, to predict labor costs. Their study showed that using historical labor data, project timelines, and site conditions allowed the Random Forest model to predict labor costs with high precision. Their findings supported the idea that ensemble models like Random Forest are well-suited to labor cost forecasting, given their ability to handle complex relationships and non-linear interactions between various factors [5].

5. Predictive Analytics in Construction Scheduling

Predictive analytics has become increasingly valuable in construction scheduling, helping to optimize labor allocation and ensure projects are completed on time. Machine learning models, particularly Random Forest Regression, have proven effective in predicting project timelines and identifying potential delays early in the process.

Lee et al. (2019) applied Random Forest to predict labor productivity, a key factor in determining project timelines. Their model accounted for variables such as task complexity, worker numbers, and project size, allowing for more accurate labor scheduling [6]. Singh et al. (2021) also emphasized the role of real-time data in predictive scheduling. They argued that incorporating real-time updates could help adjust labor schedules dynamically, minimizing delays caused by unforeseen events and ensuring that labor allocation stays in line with project needs [4].

1.3 Research Gap

When examining the current tools and systems used in construction project management, several gaps become apparent. As highlighted in the research gap analysis (see Table 1), while applications like Procore, ALICE Technologies, and nPlan have made notable progress in offering solutions for construction project management, many of them still focus on individual aspects—such as tracking daily logs, attendance, or predicting labor requirements—without providing a comprehensive system that integrates various project parameters in real-time [9] [10] [11].

A major gap is the lack of systems that can accurately predict labor requirements by considering multiple project variables, like the number of floors, windows, and square footage. While platforms such as LaborChart and nPlan offer labor tracking features, they often fall short when it comes to accurately forecasting labor needs across the entire project. Most systems tend to predict labor for specific tasks, but they either fail to consider the complexity of the overall project or do not adjust labor predictions based on real-time data, making them less adaptive to the dynamic nature of construction [10].

Another critical shortcoming is that very few systems offer a fully integrated solution that combines labor tracking, attendance, and project status updates with real-time cost and timeline forecasting. While tools like Project Pulse have some of these features, they still lack comprehensive labor prediction capabilities and the ability to manage cost and schedule updates dynamically. The inability for project managers to upload actual cost data or adjust timelines as the project progresses severely limits the effectiveness of many of these tools, as they do not truly reflect the evolving nature of construction work.

This research aims to fill these gaps by developing a system that not only predicts labor requirements based on detailed project parameters, but also integrates real-time project data to track labor attendance, update costs, and predict schedule changes. By bringing all these elements together into one unified platform, this system is designed to enhance the decision-making process for project managers, making it easier for them to adjust their plans and manage projects more efficiently.

In doing so, this research hopes to make a valuable contribution to the field of construction project management by providing a comprehensive solution that combines labor prediction, real-time updates, and cost/timeline forecasting into a single, user-friendly platform. The success of this system could mark a significant step forward in the development of more adaptive and efficient project management tools for the construction industry, helping companies better navigate the complexities of modern construction projects.

Table 2: Research Gap

Application Reference	Applicable for construction projects	User Focused Dashboards	Labor requirement prediction	Predict timeline and budget variations	Tracking daily logs and attendance
Procore	✓	✓	✗	✗	✓
ALICE Technologies	✓	✓	✓	✗	✗
nPlan	✓	✓	✗	✓	✗
LaborChart	✓	✓	✗	✗	✗
PlanGrid	✓	✓	✗	✗	✓
Project Pulse	✓	✓	✓	✓	✓

1.4 Research Problem

The research problem addressed in this study revolves around the inefficiencies in labor allocation and irregular attendance that significantly affect the cost and timeline of construction projects. Labor is a critical factor in ensuring that construction projects are completed on time and within budget. However, managing labor resources effectively remains one of the most challenging aspects of construction project management, and inefficiencies in this area can lead to serious disruptions. These inefficiencies are typically driven by inaccurate labor predictions, unpredictable attendance, and the absence of real-time data on workforce allocation—factors that contribute to delays and rising costs.

Historically, labor allocation in construction projects has often relied on estimates based on past experiences, historical data, or industry benchmarks. While these estimates can serve as a starting point, they often fail to account for the unique needs of each project. Elements such as project size, complexity, location, and the types of tasks involved all influence labor requirements, which can fluctuate throughout the project's lifecycle. When labor is either overestimated or underestimated, it leads to two major issues: too many workers being assigned, leading to wasted resources, or too few workers, causing delays and slowing down progress. Both of these imbalances directly affect project timelines, pushing back key milestones and causing significant delays in overall project completion.

Irregular attendance of construction workers further compounds the issue. Construction projects depend heavily on workers being present on-site to meet deadlines and milestones. However, worker attendance can be unpredictable, with workers calling in sick, leaving early, or being unavailable for various reasons. These absences can disrupt the planned schedule, and without an efficient system to monitor and address attendance issues in real-time, project delays are often inevitable. The lack of an up-to-date attendance tracking system means that project managers often do not recognize attendance problems until they begin to affect the workflow, at which point corrective actions can be too late to prevent delays.

The financial impact of these inefficiencies on construction projects is significant. Both overstaffing and understaffing lead to increased direct labor costs—either resources are wasted, or there aren't enough workers to meet the demand. Additionally, delays caused by labor issues can trigger financial penalties, increase overhead costs, and result in the need for rework or remobilization. These financial burdens are compounded by the costs of mismanaging the project schedule, which can lead to delays in material procurement, additional equipment rental fees, or higher coordination costs.

A central aim of this research is to find solutions for these inefficiencies through predictive tools that provide real-time, data-driven insights into labor requirements and overall project status. By analyzing historical data from past projects, the goal is to develop a system that can accurately forecast the number of workers needed for future tasks, considering factors like project type, square footage, number of floors, and work complexity. But this system will do more than simply predict labor needs it will also offer the flexibility to adjust these predictions dynamically as real-time data from the field becomes available. This means daily labor counts and task completion statuses can be updated, providing project managers with the latest information on labor availability and project progress.

In addition, this system will integrate real-time attendance tracking, helping to detect worker absenteeism early and allowing project managers to take corrective measures before these disruptions impact the overall project schedule. By incorporating these capabilities, project managers will be better equipped to proactively manage labor, allocate resources more effectively, and adjust timelines and budgets in real time, rather than merely reacting to problems after they arise.

The broader goal of this research is to help construction companies shift from a reactive labor management approach to a more proactive and predictive system. By combining predictive labor allocation, real-time attendance tracking, and dynamic forecasting of costs and timelines, this study seeks to minimize the negative effects of labor inefficiencies on construction projects. Ultimately, the aim is to optimize labor resource usage, reduce delays, and control costs, resulting in more efficient and cost-effective project execution.

This research is not just about improving labor management in construction; it also offers insights into how data driven decision-making can transform project management practices across the construction industry. By equipping companies with better tools for predicting labor needs, tracking real-time data, and adjusting project parameters on the fly, this research could help construction firms achieve faster delivery times, better outcomes, and improved profitability.

2. OBJECTIVES

2.1 Main Objective

The primary objective of this research is to develop a system that can predict the labor count needed for upcoming construction projects based on the project's category, while also forecasting the project's timeline and costs throughout its progression. Labor allocation is a critical element in construction project management, as it directly influences the project's overall timeline and budget. Accurate labor predictions can help project managers optimize resources, boost productivity, and ensure that projects are completed on time and within the allocated budget. However, many construction projects still rely on manual estimations or rigid methods for labor allocation, which can result in inefficiencies and increased costs. This research aims to address these challenges by creating a predictive system that uses historical data, machine learning models, and real-time project updates to accurately forecast labor needs and track project progress.

The construction industry is inherently complex, with numerous factors affecting the amount and type of labor required for each project. Elements such as the size of the project, the number of floors, task complexity, and the type of labor needed all play significant roles in determining how many workers are necessary and what skills are required at each stage of construction. Traditional methods of predicting labor needs often fall short in capturing the dynamic nature of construction projects, leading to issues like overstaffing or understaffing—both of which can severely impact project efficiency and cost.

Alongside labor prediction, forecasting the project's timeline and cost is equally important for effective project management. Labor inefficiencies—such as delays caused by inaccurate labor estimations or unexpected absenteeism—can lead to budget overruns and missed deadlines. Furthermore, discrepancies between predicted and actual timelines can disrupt material procurement, equipment rentals, and coordination with subcontractors. By developing a system that integrates labor predictions with real-time updates on project status, this research seeks to equip project managers with the necessary tools to make data-driven decisions, enabling them to address issues before they escalate.

The proposed system will utilize historical project data to train machine learning models that can predict labor needs, project timelines, and costs with greater accuracy. By incorporating real-time updates on labor attendance, task completion, and project progress, the system will allow for continuous adjustments to these predictions, providing a more flexible and responsive approach to project management. With predictive analytics at its core, the system will enable a shift from a reactive to a proactive approach in managing labor and project progress, ultimately reducing delays and preventing budget overruns.

The ultimate goal of this research is to contribute to the evolution of the construction industry by providing a reliable and efficient system for labor allocation and project forecasting. By enhancing decision-making, improving resource management, and streamlining project execution, this system aims to significantly improve the efficiency and cost-effectiveness of construction projects.

2.2 Sub Objectives

To achieve the main objective of predicting labor requirements and forecasting project timelines and costs, this research is divided into several sub-objectives that focus on data analysis, model training, real-time platform development, and continuous prediction adjustments. These sub-objectives offer a structured approach to tackling the problem and will ensure the creation of a robust system that can enhance construction project management.

1. Analyzing Past Project Data:

The first step of this research is to gather and analyze historical project data to uncover patterns in labor allocation, task completion, and overall project outcomes. By examining past construction projects, we can focus on key variables such as project size, type, location, the types of labor needed (e.g., carpenters, electricians, masons), and the complexity of tasks. This analysis will help us understand how these factors influence labor requirements, project timelines, and costs.

This data analysis will form the foundation for the predictive system, allowing it to recognize trends and insights that might not be immediately obvious. For example, larger projects may require a different distribution of labor compared to smaller ones, and some tasks may need more specialized labor, which should be considered when making labor predictions. Additionally, reviewing data on worker attendance and productivity will help pinpoint inefficiencies and potential causes of delays or cost overruns, which can be factored into future project forecasts.

Analyzing the relationship between different labor types and project outcomes—such as whether having more workers in a particular category speeds up project completion or reduces costs—will be a key aspect of this process. These insights will guide the development of predictive models for future labor forecasting and project management.

2. Training the System Based on Past Project Data:

The second objective involves using the historical project data to train machine learning models that can accurately predict labor needs, timelines, and costs. For this, three machine learning models—Deep Convolutional Neural Networks (DCNN), Linear Regression, and Random Forest—were evaluated for their effectiveness in this context.

DCNNs are known for their ability to process and learn from complex data patterns, making them particularly useful for analyzing large datasets with multiple variables. However, after testing all three models, Random Forest emerged as the best fit for this research due to its ability to handle large datasets with many features, its resilience against overfitting, and its high accuracy in predicting labor requirements, project timelines, and costs.

Random Forest works by constructing multiple decision trees during training, then combining their outputs to improve the accuracy of predictions. This ensemble learning method reduces errors and provides reliable predictions by considering a wide range of factors that influence the labor and cost dynamics of construction projects.

The system will be trained on the historical project data, enabling it to develop a robust, scalable model capable of predicting labor needs, timelines, and costs for future projects with similar parameters.

3. Developing a Platform for Daily Labor Count and Task Completion Updates:

Once the predictive models are trained, the next goal is to build a platform that allows project managers to update labor counts and track task completion status daily. This platform will act as a real-time data entry system where project managers can log the actual number of workers on-site, track attendance, and monitor the progress of various tasks.

This feature is essential to ensure the accuracy of the system's predictions, as it enables ongoing feedback on labor usage and project progress. Real-time updates will allow the system to adjust predictions for labor requirements, timelines, and costs dynamically, ensuring that the forecasts become more precise as the project moves forward.

The platform will also feature daily dashboards that display relevant labor-related data—such as attendance rates, task completion progress, and any discrepancies from the original schedule. This will offer valuable insights into the project's performance, helping project managers identify potential problems early and make adjustments as needed.

4. Predicting the Real-Time Status of Projects:

The final sub-objective is to enable the system to predict the real-time status of the project by integrating updated labor data with project timelines and cost forecasts. As the project progresses, the system will use real-time data to adjust predictions, helping project managers track whether the project is ahead of schedule, behind schedule, or on track.

By continuously analyzing updated labor data, the system will provide actionable insights on whether additional labor is needed or if certain tasks are falling behind. These insights will allow project managers to take proactive measures to prevent delays and cost overruns. The system will also forecast the financial implications of changes in labor or project schedules, helping managers make informed decisions about resource allocation and budgeting.

This dynamic prediction capability ensures that project managers have up-to-date information on the project's status, empowering them to take timely corrective actions and stay on top of any emerging challenges, ultimately leading to more efficient project execution.

3. METHODOLOGY

The methodology for this research aims to tackle the challenges surrounding labor prediction, project timeline forecasting, and cost estimation in construction projects. The approach integrates machine learning, data analysis, system integration, and real-time updates to develop a system that is not only accurate but also scalable and user-friendly. This section breaks down the methodology into several key components, each aimed at ensuring the success of the project.

3.1 Requirement Gathering and Analysis

The first step of the methodology is requirement gathering. During this phase, the focus is on identifying the specific needs and expectations of all relevant stakeholders. Collaborating with MAGA Construction Pvt Ltd, we gather insights into the unique challenges they face concerning labor allocation, project timeline management, and cost forecasting. The goal is to understand their business processes, define clear objectives, and outline the functionalities that the system should offer to enhance construction project management.

3.1.1 Non-functional Requirements

Beyond the core functional needs, several non-functional requirements were identified. These ensure that the system is not only capable but also performs at a high level across various scenarios:

- Performance: The system must be able to provide real-time predictions and updates with minimal latency. Given the dynamic nature of construction projects, the platform needs to offer quick, reliable results.
- Scalability: As construction projects grow in size and complexity, the system must scale accordingly. It should be able to handle multiple projects concurrently without any drop in performance.
- Availability: With the system being cloud-based, high availability is essential. Minimal downtime ensures that project managers have continuous access to real-time updates and predictions.

- Security: Given the sensitive nature of construction project data, robust security measures are required. This includes strong encryption for both data transmission and storage, ensuring confidentiality and integrity.
- Usability: The system must be easy to use, even for those without extensive technical training. A user-friendly interface is critical to allow project managers to quickly input data, view predictions, and make informed decisions.

3.2 Feasibility Study

The feasibility study assesses whether the system can be developed and deployed within the constraints of time, cost, and resources. It covers four main areas:

- Technical Feasibility: This evaluates whether the existing tools, technologies, and technical skills are adequate to build the system. Since the required technologies (e.g., Python, React.js, Node.js, Flask) are widely adopted and well-supported, the technical feasibility is assured.
- Operational Feasibility: The system must integrate smoothly into the day-to-day operations of construction companies. The feasibility study ensures that the real-time features and overall system functionality fit well within existing workflows.
- Economic Feasibility: This aspect examines the costs associated with the development and deployment of the system—covering software, hardware, and personnel—versus the expected benefits such as increased efficiency, reduced labor costs, and better timeline management.
- Legal Feasibility: Ensuring that the system complies with legal requirements, especially around data protection laws, is crucial. This includes managing personal data such as labor attendance records and other sensitive information.

The findings from the feasibility study confirm that the project is viable from technical, operational, economic, and legal perspectives.

3.3 System Design

In the design phase, we focus on creating the system architecture, defining its components, and mapping how they interact. This phase is critical in ensuring the system will meet the requirements and function as intended.

3.3.1 Overall System Architecture Diagram

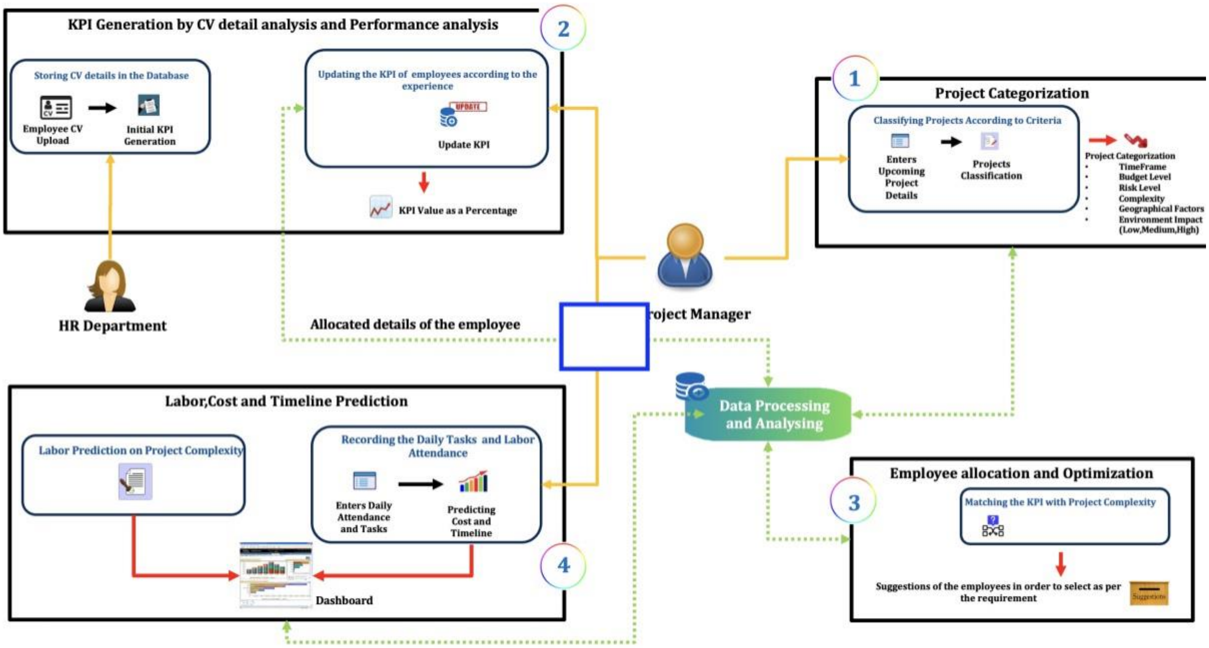


Figure 1 : Overall System Architecture Diagram

The system follows a three-tier architecture: the front-end, back-end, and database layers. This design approach ensures modularity, scalability, and ease of maintenance.

- Front-End (Client Side): Built with React.js, the front-end is responsible for the user interface. It allows project managers to input data (such as labor attendance), track project progress, and view real-time predictions.
- Back-End (Server Side): Powered by Node.js and the Express.js framework, the back-end handles API requests from the front-end, processes the data, and communicates with the database. It also interacts with the Flask API to generate real-time predictions.
- Database: MySQL is used for relational data storage, including project details, labor information, cost tracking, and project timelines. This ensures fast retrieval and updates, essential for real-time performance.
- Machine Learning Model: The machine learning model is developed using Python, leveraging scikit-learn for Random Forest algorithms and TensorFlow/Keras for deep learning models. The Flask API serves the trained model, enabling the back end to query for predictions regarding labor requirements and project status.

This overall architecture is designed for efficient data processing, quick information flow, and seamless real-time updates.

3.3.2 Component-Specific Architecture Diagram

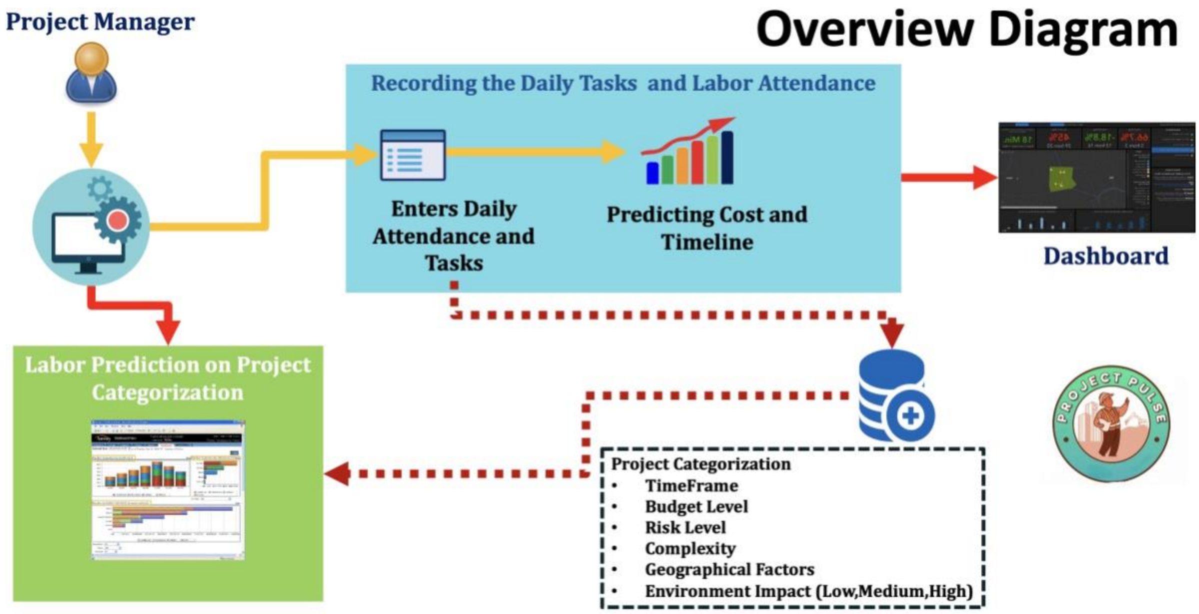


Figure 2 : Component System Architecture Diagram

- User Interface (React.js): The user interface, built using React.js, communicates with the back end through API calls. It's designed to be intuitive and responsive, enabling project managers to easily interact with the system on various devices.
- Back-End (Node.js with Express.js): The back-end server processes requests, manages data, and serves as a conduit between the front-end, database, and machine learning models. Express.js handles API routing, while Node.js supports the server-side logic.
- Database (MySQL): The database holds all essential project data, such as labor attendance, cost details, and project timelines. Its relational structure is optimized for fast query execution and efficient data retrieval.
- Machine Learning Integration (Flask API and Python): Python is used for model training, while Flask acts as an intermediary between the back end and the machine learning model, providing predictions as API responses.

These components work together to create a fluid and functional system. A detailed diagram of these interactions is provided below:

3.4 Implementation of the System

The implementation phase involves the actual coding and development of the system, based on the specifications outlined in the design phase. The following steps were undertaken:

- Front-End Development: The front-end was developed using React.js, allowing for a dynamic and responsive user interface. Key features such as labor attendance input, project progress tracking, and real-time predictions were implemented.
- Back-End Development: The Node.js server was set up with Express.js, facilitating API development. The server processes requests, manages interactions with the database, and communicates with the machine learning models via Flask.
- Database Setup: MySQL was set up to store relational project data. Tables were designed to track labor attendance, project timelines, and costs, ensuring that SQL queries were optimized for quick data retrieval.
- Machine Learning Model Integration: Machine learning models were trained using Python, specifically utilizing scikit-learn for Random Forest and TensorFlow/Keras for deep learning. The models were integrated into the system via a Flask API, allowing the back end to query the models and get real-time predictions.
- Testing and Debugging: As development progressed, continuous testing was carried out. API endpoints were tested using Postman, while code debugging was done in Visual Studio.

3.5 Tools and Technologies

The system leverages a range of tools and technologies to ensure its functionality and performance:

- React.js: Used for building the front-end, providing a dynamic and responsive interface.
- Node.js with Express.js: Powers the back end, handling API requests and managing communication between components.
- MySQL: Manages structured project data, ensuring quick retrieval and efficient updates.
- Python: Used for training machine learning models and integrating them with the Flask API.
- Flask: Serves as the bridge between the machine learning models and the back end.
- Postman: Used for testing API calls and ensuring proper communication between system components.
- Visual Studio: A development environment for writing and debugging code.
- XAMPP: Used for local development and testing.

These tools were chosen based on their ability to meet the system's requirements for scalability, real-time performance, and ease of integration.

3.6 System Testing Tools

Testing is crucial for ensuring that the system functions correctly under various conditions. The following tools were used for system testing:

- Postman: Used for testing APIs and ensuring that requests and responses between the front-end and back-end are handled correctly.

By using a combination of these tools and methodologies, the system was developed to be robust, scalable, and able to deliver accurate, real-time predictions.

3.7 Commercialization aspects of the product

The commercialization strategy for this system involves offering it as a cloud-based Software-as-a-Service (SaaS) product, which will allow construction companies of all sizes to access it via subscription models.

Key Partners

The success of the Systematic Manpower Allocation System hinges on forming strong partnerships with several critical stakeholders:

1. Construction Companies: As the primary users, construction companies will leverage this system to streamline manpower allocation, enhance project performance, and improve the tracking of progress.
2. Educational Institutions: Collaborating with schools and training organizations that offer relevant certifications will ensure that both the workforce and company employees are well-equipped to maximize the system's potential, fostering skill enhancement and better project management.
3. Technology Providers: Partnerships with tech companies are essential to ensure that the application remains up to date with the latest software developments. These providers will help maintain the platform's scalability, security, and overall functionality.

4. Financial Institutions: These entities will support funding and financial assistance to construction companies looking to adopt the system. They will be crucial in financing initial deployments and future growth.

Key Activities

The activities required for the successful development, deployment, and maintenance of the system are as follows:

1. Employee Skills and CV Analysis: The system will analyze employee CVs and skill profiles, ensuring that the right workers are assigned to tasks that align with the project's specific needs.
2. Project Categorization: Construction projects will be categorized according to their complexity, size, and labor requirements, allowing for optimal team selection tailored to each project's scope.
3. Development and Deployment: The process will include building a user-friendly platform with advanced backend systems and machine learning features to predict labor needs, timelines, and costs. Post-development, the system will be deployed and rolled out to customers.
4. Training and Ongoing Support: Providing clients with comprehensive training materials, including tutorials and user manuals, will be essential. Continued support will address any challenges after deployment, ensuring smooth usage and high client satisfaction.

Key Resources

To ensure successful development and commercialization of the system, several key resources will be required:

1. Software Development Team: A skilled team of developers, data scientists, and engineers will be necessary to create and maintain the system, ensuring it is both robust and scalable.
2. Training Materials: Developing detailed training resources, such as instructional videos, guides, and webinars, will help ensure that users can get the most out of the system.
3. Data and Analytics Tools: To analyze construction data—such as labor attendance, project performance, and task completion—advanced data tools will be crucial.
4. Financial Resources: Sufficient funding will be required for the development, marketing, and ongoing maintenance of the system. Financial partners may also provide the capital needed for large-scale deployments.

Value Proposition

The value proposition of the Systematic Manpower Allocation System lies in its ability to deliver significant advantages to construction companies, including:

1. Optimized Labor Allocation: The system ensures that employees are assigned to projects that match their skills and experience, which leads to higher productivity and improved outcomes.
2. Cost Reduction and Enhanced Project Performance: By efficiently managing labor resources, the system minimizes project delays, reduces costs, and avoids resource wastage.
3. Real-Time Project Tracking: Clients can monitor labor attendance, task progress, and project milestones in real time, allowing for proactive decision-making and improved management oversight.

Customer Relationships

Building and maintaining strong relationships with customers will be essential for the long-term success of the system:

1. Brand Building on Social Media: Using platforms such as LinkedIn, Facebook, and Instagram, the system can build brand recognition, attract new clients, and educate users about its benefits.
2. Dedicated Support Team: Offering continuous support ensures clients can resolve any issues swiftly and receive the guidance needed to maximize system adoption.
3. Feedback Loops for Improvement: Actively soliciting feedback from users will help refine and improve the system over time. Regular updates based on user input will keep the system responsive to changing customer needs.

Customer Segments

The system is designed for several key customer groups:

1. Construction Companies and Contractors: These are the primary users of the system, benefiting from its ability to streamline labor management and optimize project performance.
2. Project Managers: With access to real-time data, project managers can make more informed decisions, ensuring projects stay on schedule and within budget.
3. HR Departments in Construction Companies: HR teams can utilize the system to monitor attendance, track employee performance, and allocate resources efficiently.

Channels

The product will be marketed and distributed through the following channels:

1. Social Media: Campaigns on LinkedIn, Facebook, and Instagram will engage professionals in the construction industry, generating interest and educating potential customers.
2. Website: A dedicated website will serve as the primary information hub for the system, offering product demos, case studies, pricing details, and a point of contact for inquiries.

Cost Structure

The key costs involved in developing and maintaining the Systematic Manpower Allocation System are:

1. Development Costs: Significant investment will be required for software development, including licensing fees for necessary tools, technologies, and data analytics infrastructure.
2. Ongoing Data Analysis and Software Licensing: The system will incur recurring costs related to data storage, processing, and analysis, along with the maintenance of software licenses.
3. Salaries for Development Team: A portion of the budget will be allocated to compensating the development team, which includes software engineers, machine learning specialists, and project managers.

Revenue Streams

Revenue for the system will be generated through several avenues:

1. Subscription Fees: Clients will pay subscription fees based on user count or the scale of their operations. This fee structure ensures continuous access to the platform, including updates and support.
2. Consultancy Services: The company can also provide consultancy services, helping clients optimize manpower allocation and project management through expert guidance and system integration.
3. Advertising and Partnerships: As the system gains traction, it may open up opportunities for advertising within the platform or partnerships with other technology providers in the construction sector.

By addressing these key elements, the commercialization of the Systematic Manpower Allocation System is poised for success, with a clear focus on value creation for construction companies, project managers, and other stakeholders in the industry.

4. RESULTS AND DISCUSSION

4.1 Results

The results of this study are based on predictions generated using various machine learning models, with a special focus on Random Forest. We evaluated performance using metrics such as accuracy, precision, recall, and F1-score. Here's how the models performed:

- Random Forest Regression: Achieved an accuracy of 87.28%, making it the best-performing model for predicting labor needs, cost estimation, and project timelines.
- Deep Convolutional Neural Network (DCNN): Scored 75.39%, showing a solid understanding of the data patterns but not quite matching the precision of ensemble methods like Random Forest.
- Linear Regression: With an accuracy of 73.68%, this model struggled to capture the complex, non-linear relationships inherent in construction project parameters.

We also analyzed the precision of these models, particularly in predicting labor requirements. The comparison below illustrates each model's precision:

Table 3 : Comparison of Three Models

Model	Precision	Recall	F1 Score	mAP
Random Forest	0.92	0.88	0.90	0.87
DCNN	0.78	0.72	0.75	0.73
Linear Regression	0.70	0.68	0.69	0.69

These results highlight the Random Forest model as the most robust, making it the top choice for accurately predicting labor requirements and project costs in real-time.

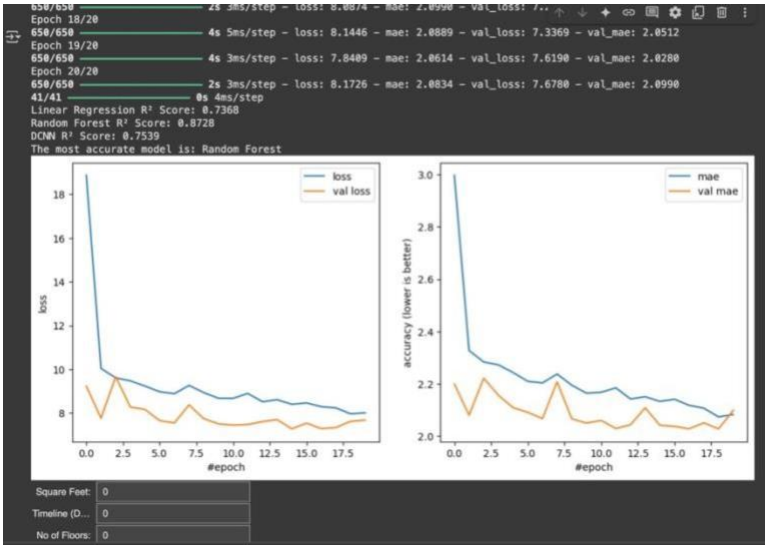


Figure 3 : Accuracy of Random Forest Model

4.2 Research Findings

Several important insights were drawn from this research, contributing to a deeper understanding of labor allocation in construction projects:

1. The Role of Historical Data:
   - Historical project data plays a crucial role in enhancing prediction accuracy. Variables like square footage, the number of floors, and window counts showed a strong correlation with labor needs and project costs. The dataset of 2,500 records was instrumental in creating reliable models, with Random Forest performing exceptionally well on such a substantial sample.
2. The Dynamic Nature of Construction Projects:
   - Real-time updates are critical to maintaining the accuracy of predictions. Factors like labor attendance changes, project delays, or unplanned events can significantly impact prediction outcomes. The system allows for daily updates to labor counts and project status, keeping the predictions aligned with current conditions.
   - A feedback loop ensures that the model continually adapts, which is particularly helpful in managing large and complex construction projects.
3. Benefits of Real-Time Predictions:
   - Real-time predictions can provide actionable insights, allowing project managers to adjust labor allocations dynamically and optimize project timelines. The system

successfully predicted potential delays and budget overruns when labor was not aligned with real-time project conditions.

4. Scalability Potential:
   - The machine learning framework is designed for scalability. As more data from additional projects becomes available, the system can handle the demands of larger construction firms and multiple concurrent projects.

4.3 Discussions

The Results reveal that the Random Forest model is the best choice for predicting labor needs, timelines, and costs. However, there are several aspects worth discussing, particularly around challenges faced and ideas for future research.

4.3.1 Limitations and Challenges

Throughout the study, we identified several challenges that can be addressed in future versions of the system:

1. **Data Quality**:
   - While historical project data formed the foundation for training the models, some inconsistencies, and missing data points—such as fluctuating labor counts and delayed task completion—could have affected the predictions. Future research will need to focus on cleaner, more consistent data collection methods.
   - Data inaccuracies, like missing labor attendance or incomplete task completion rates, can skew predictions. Improved data recording practices will help boost the model's accuracy.
2. **Non-Linear Complexity**:
   - The project dataset contained parameters that interact in complex, non-linear ways. Although the Random Forest model captured many of these relationships, there is room for improvement. For instance, Linear Regression struggled with these complexities. Future work could explore hybrid models or more advanced algorithms like Gradient Boosting Machines (GBM) or XGBoost, which can better handle such non-linear interactions.
3. **Real-Time Model Adaptability**:

- While the system incorporates real-time data updates, model adaptability remains an issue. The model's predictions are updated based on input data, but real-time corrections during project execution—due to unforeseen disruptions like weather or material shortages—need further refinement for continuous prediction accuracy.

4. **Computational Constraints**:
   - The Random Forest model can be computationally demanding, particularly with large datasets and numerous parameters. The computational load can affect performance, especially on smaller devices or during peak project times. Cloud-based computing or edge computing could be utilized to reduce latency and improve real-time prediction accuracy.

4.3.2 Future Work

There are several exciting avenues for future research to build on the findings from this study:

Hybrid Machine Learning Models:

To enhance prediction accuracy and reliability, we could explore hybrid models combining Random Forest with algorithms like Gradient Boosting Machines or Neural Networks. These hybrid approaches could improve the handling of non-linear relationships and refine predictions.

Enhanced Data Collection and Integration:

Future versions of the system could integrate IoT sensors and smart wearables for real-time labor tracking. These tools would reduce human error and provide granular data for more accurate predictions. Moreover, incorporating geospatial data and building information modeling (BIM) would help with labor allocation and resource management.

Real-World Deployment:

The system should undergo field testing on a variety of real-world construction projects to validate its effectiveness across diverse environments and regions. Feedback from project managers and workers would provide valuable insights into improving the system's usability and performance.

Integration with Existing Construction Software:

To facilitate wider adoption, the system could be integrated with popular construction management platforms (such as Procore or Builder trend). This would streamline data flow between project management software and the labor prediction system, reducing the need for manual data entry.

Improved Real-Time Feedback Mechanisms:

Enhancing the real-time feedback loop with visual and auditory alerts could enable immediate action if predictions are not met or if tasks are falling behind schedule. This would make the system more user-friendly and allow project managers to intervene more quickly.

5. CONCLUSION

This research has made notable advancements in the field of construction project management by developing an automated system that blends machine learning, historical project data, and real-time updates. At its core, the system leverages the Random Forest machine learning model, offering an innovative and accurate way to predict labor needs, forecast project timelines, and estimate costs. By integrating these components, the system presents a practical solution to many of the inefficiencies and challenges commonly faced by construction managers.

A key takeaway from this research is the critical role that historical data plays in refining labor predictions. By analyzing past project details like floor counts, square footage, and window numbers the system uncovered valuable correlations that helped generate reliable predictions about labor needs and project expenses. This data-driven approach not only improves labor demand forecasts but also sharpens project cost and timeline estimations, making project planning and budgeting much more accurate. Moreover, the inclusion of real-time data ensures that these predictions are regularly updated, creating a continuous feedback loop that keeps the system aligned with actual project progress.

The system's ability to adapt in real-time is particularly impressive. Construction projects are inherently dynamic, with factors such as labor attendance, material delays, and unforeseen issues like weather disruptions frequently altering schedules. The system's ability to update its predictions based on daily input from project managers means it can adjust labor and timeline forecasts continuously. This proactive approach allows project managers to stay ahead of potential issues, helping to mitigate disruptions before they escalate.

However, despite these encouraging outcomes, some challenges remain. Issues with data quality, such as missing data points or inconsistencies in labor attendance or task completion, can diminish the accuracy of predictions. Future research should focus on improving data collection methods to ensure the reliability and consistency of the data used to train the models. Additionally, non-linear relationships among project variables continue to challenge certain machine learning models, such as Linear Regression. While the Random Forest model performed well, exploring hybrid models or more advanced algorithms like Gradient Boosting Machines (GBM) or XGBoost could help address these complexities more effectively.

The model’s adaptability also presents an area for improvement. While the system is capable of updating labor requirements and project progress daily, it still requires refinement in its ability to respond to unexpected disruptions, such as supply chain delays or sudden changes in labor availability. To maintain reliability in real-world applications, it will be crucial for the system to adjust smoothly and swiftly to these changes. Incorporating cloud computing or edge computing could help alleviate computational constraints and reduce latency, enabling the system to scale more efficiently and handle larger datasets, especially when managing multiple concurrent projects.

Despite these limitations, the scalability of the system remains one of its standout features. As more projects are incorporated, the system’s ability to manage larger datasets and offer accurate predictions across multiple sites becomes even more valuable. This capability will empower construction firms to manage several projects at once while maintaining precise labor allocations, timelines, and cost estimates for each one. Real-world testing across various projects will be essential to further refine the system and ensure its adaptability and usability across different environments.

Looking forward, there are several exciting opportunities for enhancing the system. The integration of Internet of Things (IoT) devices and smart wearables could provide even more granular data, such as real-time tracking of workers’ attendance, task completion, and even safety metrics. This could further improve the system’s accuracy and offer deeper insights into labor performance. Moreover, incorporating Building Information Modeling (BIM) and geospatial data could optimize resource management by providing a more comprehensive context to labor predictions, helping better allocate resources based on spatial factors.

Another promising direction is the potential integration of the system with existing construction management platforms, such as Procore or Builder trend. This would streamline workflows by eliminating the need for manual data entry and ensuring that labor, cost, and schedule predictions are continuously updated with the latest project details.

The commercialization of the Systematic Manpower Allocation System also holds significant promise. Offering the system as a cloud-based Software-as-a-Service (SaaS) solution could provide construction companies, regardless of their size or location, with easy access to its benefits. Partnerships with financial institutions, technology providers, and construction firms would be key to supporting the platform's widespread adoption and ensuring it remains up-to-date and secure, addressing the evolving needs of the industry.

In conclusion, this research sets the stage for a more data-driven and automated approach to construction project management. By minimizing human error, optimizing labor allocation, and providing real-time project insights, the Systematic Manpower Allocation System has the potential to greatly enhance project efficiency, reduce delays, and lower costs. It moves beyond simply addressing current inefficiencies, introducing a proactive management style that anticipates challenges and resolves them before they impact a project's success. As the system continues to develop and evolve, it has the potential to transform the construction industry, paving the way for smarter, more cost-effective project execution and inspiring future innovations in the field of construction management.

6. REFERENCES

[1] M. E. P. LTD, "MAGA Engineering," [Online]. Available: https://www.maga.lk/. [Accessed 17 06 2024].

[2] R. K. A. M. a. A. S. D. Gupta, "Machine Learning Models for Cost Prediction in Construction Projects," International Journal of Civil Engineering, Vols. 18, Issue: 5, pp. 1076-1085, May 2019.

[3] B. H. a. K. W. S. Park, "Labor Cost Forecasting Using Deep Convolutional Neural Networks," International Journal of Project Management, vol. 39, no. 6, pp. 671-680, Jun. 2020.

[4] A. M. M. P. S. a. G. D. S. Singh, "Integration of Real-Time Data in Construction Project Management for Dynamic Labor Allocation," Construction Management and Economics, vol. 39, no. 4, pp. 246-259, Apr. 2021.

[5] D. N. a. S. S. S. M. Rahman, "Prediction of Labor Cost Using Machine Learning: A Case Study on Residential Projects," Journal of Building Performance, vol. 11, no. 5, pp. 356-365, Sept. 2020.

[6] H. L. C. Y. H. a. K. C. H. Lee, "Using Random Forest Algorithm for Labor Productivity Prediction in Construction Projects," Computers in Industry, vol. 98, pp. 122-130, Apr. 2019.

[7] L. a. W. Q. Zhang, "Deep Learning for Construction Project Cost Forecasting," Automation in Construction, vol. 108, pp. 137-147, Dec. 2019.

[8] J. H. a. L. X. F. Zhang, "Labor Cost Forecasting Using Random Forest and Neural Networks," Journal of Engineering and Technology, vol. 39, no. 7, pp. 456-463, Jul. 2019.

[9] "Alice Technologies," [Online]. Available: https://www.alicetechnologies.com/home. [Accessed 12 07 2024].

[10] P. Technologies, "Procore," [Online]. Available: https://www.procore.com/en-sg. [Accessed 12 07 2024].

[11] "Plan Grid," Autodesk Construction Cloud, [Online]. Available: https://construction.autodesk.com/products/plangrid/. [Accessed 12 07 2024].

[12] n. Limited, "nPlan," [Online]. Available: https://www.nplan.io/. [Accessed 11 07 2024].

APPENDIX A : Turnitin Report

Automated Man Power Allocation By Performance Analysis and Project Categorization For Construction Projects.docx

ORIGINALITY REPORT

9%	5%	5%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.coursehero.com Internet Source	2%
2	Submitted to Sri Lanka Institute of Information Technology Student Paper	1%
3	Tuncer, Yasemin. "Artificial Intelligence Based Risk Analsis in Project Management.", Marmara Universitesi (Turkey) Publication	1%
4	R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P. Prasad. "Algorithms in Advanced Artificial Intelligence - Proceedings of International Conference on Algorithms in Advanced Artificial Intelligence (ICAAAI-2024)", CRC Press, 2025 Publication	1%
5	Amir Shachar. "Introduction to Algogens", Open Science Framework, 2024 Publication	1%
6	Submitted to Lincoln University Student Paper	1%
7	Submitted to University of Leeds Student Paper	

Figure 4:Turnitin Report