Prediction Model for Contractor Bid Price Using Machine Learning Approach

Authors

Syed Khizar Hayat Jan	CU-985-2020
Malik Noman Shafiq	CU-952-2020
Alamgir Khan	CU-964-2020
Naveed Ahmad	CU-982-2020

Supervisor

Engr. Sultan Shah

Co-Supervisor

Engr. Muhammad Waqas

DEPARTMENT OF CIVIL ENGINEERING

CECOS University of IT and Emerging Science Hayatabad Peshawar

June 2024

Prediction Model for Contractor Bid Price Using Machine Learning Approach

Authors

Syed Khizar Hayat Jan	CU-985-2020
Malik Noman Shafiq	CU-952-2020
Alamgir Khan	CU-964-2020
Naveed Ahmad	CU-982-2020

A thesis submitted in partial fulfilment of the requirement for the degree of

B.Sc. Civil Engineering

Thesis Supervisor:

Engr. Sultan Shah

Associate Professor CECOS University

Co-Supervisor

Engr. Muhammad Waqas

Associate Professor CECOS University

Thesis Supervisor Signature

Chairman CED, Signature

DEPARTMENT OF CIVIL ENGINEERING

CECOS University of IT and Emerging Science Hayatabad Peshawar

June 2024

UNDERTAKING

I certify that research work titled "Prediction Model for Contractor Bid Price Using Machine Learning Approach" is my work.

The Work has not presented elsewhere for assessment. Where material has been used from other sources it has been properly acknowledged/referred.

Syed Khizar Hayat Jan	CU-985-2020	________________
Malik Noman Shafiq	CU-952-2020	________________
Alamgir Khan	CU-964-2020	________________
Naveed Ahmad	CU-982-2020	________________

ACKNOWLEDGEMENT:

Deepest praises and gratitude to ALLAH Almighty, for His endless blessings throughout this research work to complete it successfully within set forth time. Our gratified thanks to our supervisor Engr. Sultan Shah and Engr. Muhamad Waqas, Department of Civil Engineering, CECOS University of IT and Emerging Science, Hayatabad Peshawar for the perpetual support, guidance, and special thanks for illuminating us with scholarly knowledge during this research. Without their scholarly, enlightened, and intellectual advices, this thesis could not have reached to its present form. We are also indebted to the teachers who gave us moral support.

We offer devout thanks to our beloved parents for their support, tireless efforts, and patience for this study and throughout every moment of our life.

We would like to express our gratitude to our siblings for their unconditional support and encouragement during this study. Our solemn thanks to the friends who supported us.

We owe special thanks to Transportation Lab staff of Department of Civil Engineering, CECOS University of IT and Emerging Science, Hayatabad Peshawar for their assistance and cooperation during the testing conducted for their research work.

TABLE OF CONTENT:

Chapter 1

Introduction

1.1. Understanding Bid Price Estimation in the Construction Industry:

1.1.1. Importance of Accurate Bid Price Estimation:

In Pakistan's construction sector, precisely estimating bid prices is important for profitability and project success **[1]**. Bid price estimation is a complex process that considers numerous factors such as overhead and profit margins, equipment costs, project duration, labor costs and material expanses **[2]**. This approach ensures that contractors provide competitive bids that meet project costs while maintaining profitability **[3]**. In contrast, clients rely on precise bid costs to make informed decisions regarding project feasibility, contractor selection and budget allocation **[4]**.

Furthermore, precise bid price estimation allows contractors to reduce the financial risks associated with projects **[5]**. Contractors can efficiently minimize cost overruns by thoroughly assessing expanses throughout the estimation phase **[6]**. This proactive strategy enhance profitability, stability and financial health of construction sector in Pakistan **[7]**.

1.2. Navigating Competitive Bidding Dynamics in construction Procurement:

The process of procuring construction works involves selecting the most suitable bid that meets the investor's requirements while ensuring competitive pricing **[8]**. Traditionally, the lowest bid (LB) model has been favored, particularly in scenarios where clients prioritize cost transparency and enforcement of completion time and payment conditions **[9]**. However, relying solely on the lowest price often leads to compromises in quality, disputes, and time overruns, especially during economic downturns **[10]**. Despite its simplicity, the LB model tends to promote bid rigging and necessitates strict controls to mitigate risks. Consequently, it may result in increased costs and lower overall project quality **[11]**.

To address the shortcomings of the LB model, alternatives like the average bid method and the best value or most economically advantageous tender model have been introduced. The average bid method aims to prevent the selection of unrealistically low bids by comparing them to the average, although it does not eliminate the potential for collusion. On the other hand, the best value model considers a combination of factors such as whole-life cost, quality, and environmental aspects, allowing clients to make more informed decisions beyond just price. However, this approach demands significant effort from the client in defining criteria and their relative importance, potentially complicating the procurement process **[12]**.

In local construction markets, where competition is fierce and cost reduction opportunities are limited, contractors often rely on adjusting their markup to maximize long-term profits and stay competitive **[13,14]**. Decision-making support models for such markets need to account for the correlation between competitors' bid prices, as pricing decisions are influenced by similar market conditions and tender procedures. By understanding these dynamics, stakeholders can better navigate bidding strategies to enhance their chances of winning contracts while maintaining profitability **[15]**.

1.3. Challenges in Bid Price Estimation in Pakistan:

The bid price estimation process in Pakistan’s construction industry faces various challenges **[16]**. The construction sector is characterized by its dynamic nature, with factors such as fluctuating material prices, labor shortages, regulatory constraints and market dynamics influence bid prices **[17]**. Additionally, reliance on conventional estimation methods may lead to inaccuracies and inefficiencies, highlighting the need for innovative approaches **[18]**.

Furthermore, the geopolitical situation in Pakistan can also influence bid price estimation in the construction sector **[19]**. Regional conflicts, political instability and changes in government policy can all present risks that contractors must take into account while bidding on projects **[20]**. These factors can affect project budgets, costs and overall feasibility, which complicates the bid price estimation process further **[21]**.

The dynamics of competitive bidding in construction procurement highlight the need for innovative approaches to enhance decision-making processes **[22]**. As stakeholders seek to optimize bid selection while balancing cost, quality, and market competitiveness, emerging technologies offer promising solutions **[23]**. One such avenue is the application of machine learning algorithms to predict contractor bid prices **[24]**. By leveraging historical bid data, project characteristics, and market trends, these prediction models can provide valuable insights into pricing strategies and competitive positioning **[25,26]**. Integrating such advanced tools into the bidding decision support framework can empower stakeholders to make informed decisions, mitigate risks, and optimize project outcomes **[27]**. As the construction industry continues to evolve, embracing predictive analytics holds the potential to revolutionize bidding practices and drive efficiency across the procurement lifecycle **[28]**.

1.4. Identifying Influential Factors Affecting Contractor Bid Prices in Pakistan:

1.4.1. Economic Factors:

Economic factors have a considerable impact on bid prices in Pakistan. Variations in interest rates, currency values and inflation rates directly affect material costs, equipment expanses and labor wages, thereby impacting bid price estimations **[29]**.

Moreover, currency values play a pivotal role in shaping bid prices for construction projects in Pakistan **[30]**. The depreciation or appreciation of the Pakistani Rupee against major currencies can directly influence the cost of imported materials and equipment, as well as impact the profitability of contractors **[31]**. A depreciating Rupee may lead to higher material and equipment costs, compelling contractors to adjust bid prices to maintain profit margins **[32]**. Conversely, a strengthening Rupee can lower import costs, allowing contractors to offer more competitive bids **[33].**

Inflation rates also exert a significant influence on contractor bid prices in Pakistan **[34]**. Rising inflation can drive up the prices of construction materials, equipment, and labor wages, thereby increasing project costs and bid prices **[35].** Contractors often incorporate inflation forecasts into their pricing strategies to mitigate the risk of cost overruns during project execution **[36].** Additionally, inflation expectations can influence contractors' perceptions of future project profitability, shaping their bidding behavior accordingly **[37].**

Furthermore, macroeconomic policies and government initiatives aimed at economic development can indirectly impact contractor bid prices in Pakistan **[38]**. For instance, infrastructure projects funded by government investments or international aid programs can create opportunities for contractors, stimulating competition in the bidding process **[39]**. Conversely, economic downturns or policy uncertainties may lead to subdued demand for construction services, prompting contractors to adopt more conservative pricing strategies to mitigate risks associated with project delays or cancellations **[40]**. Thus, a comprehensive understanding of economic factors is essential for contractors to formulate competitive bid prices and navigate the dynamic landscape of Pakistan's construction industry **[41]**.

1.4.2. Regulatory Environment:

The regulatory framework of Pakistan, including permit requirements, building codes and zoning laws significantly affects construction project timelines and expanses **[42]**. Adherence to regulatory standards increases project expanses, which in turn influence bid prices **[43]**.

Contractors must allocate resources to ensure compliance with these regulatory standards, which often entails obtaining multiple permits, certifications, and approvals from various government agencies **[44]**. Delays in securing necessary permits or addressing regulatory compliance issues can prolong project timelines and increase construction costs, as contractors may incur additional expenses for labor, equipment, and overhead during the extended duration **[45]**.

Furthermore, the enforcement of building codes and zoning laws can significantly impact project designs and construction methodologies, thereby influencing bid prices **[46]**. Strict adherence to safety and structural requirements outlined in building codes may necessitate the use of specific materials, construction techniques, or design modifications, all of which can add to project costs **[47, 48]**. Similarly, zoning regulations dictating land use and development restrictions may limit the available options for site selection or require special considerations for projects located in designated zones, such as heritage sites or environmentally sensitive areas **[49, 50]**.

1.4.3. Market Trends and Demands:

Market trends and demands, such as the material supply chain disruptions, availability of skilled labors and industry competition, significantly influence bid prices **[51]**. Understanding various market trends for contractors is crucial to maintain competitiveness and accurately estimate bid prices **[52]**.

Industry competition is another key consideration influencing bid prices in Pakistan's construction market **[53]**. The level of competition among contractors vying for the same projects can impact pricing dynamics, with intense competition often driving bid prices down as contractors seek to undercut rivals to secure contracts **[54, 55]**. Conversely, in less competitive markets or for specialized projects requiring unique expertise, contractors may have more flexibility to price bids at higher margins to reflect their value proposition or technical capabilities **[56]**.

Furthermore, emerging market trends such as the adoption of sustainable construction practices, digital technologies, or prefabrication methods can influence contractors' cost structures and bidding strategies **[57]**. Contractors who invest in innovative solutions to enhance project efficiency, reduce waste, or improve environmental sustainability may differentiate themselves in the market and command premium prices for their services **[58]**.

1.5. Machine Learning: A Promising Approach for Bid Price Estimation:

1.5.1. Understanding Machine Learning (ML):

Machine Learning (ML) is one of the powerful tool for improving bid price estimation in the construction industry **[59]**. ML algorithms analyze large datasets, generate predictive models and identify patterns. Unlike conventional methods which rely on predetermined rules and static algorithms, ML techniques can adapt to changing conditions and learn from new data, making them well-suited for dynamic industries like construction **[60, 61]**.

1.5.2. Advantages of Machine Learning in Bid Price Estimation:

The use of machine learning (ML) in bid price estimation have several advantages, changing conventional practices and improving the precision, transparency efficiency of the estimation process **[62]**.

1.5.2.1. Enhanced Accuracy:

ML models leverage advanced statistical techniques and algorithmic complexity to generate bid price estimates with unprecedented accuracy **[63]**. By assimilating vast datasets and discerning subtle patterns, ML algorithms can provide reliable predictions, mitigating the risk of underestimation or overestimation inherent in conventional methods **[64]**.

Moreover, ML models continuously learn and adapt to new data, further enhancing their predictive power over time **[65]**. This dynamic nature allows them to capture evolving market trends and adjust bid price estimates accordingly, offering businesses a competitive edge in dynamic and unpredictable market environments **[66]**. Additionally, the transparency and interpretability of ML models enable stakeholders to understand the rationale behind bid price predictions, fostering trust and confidence in the decision-making process **[67]**. As a result, organizations can make more informed and strategic decisions, optimizing resource allocation and maximizing profitability **[68]**.

1.5.2.2. Adaptability to Changing Conditions:

The dynamic nature of construction projects necessitates bid price estimation methodologies capable of accommodating evolving conditions and unforeseen variables **[69]**. ML algorithms in this regard, as they possess the inherent capacity to adapt to changing circumstances, learn from new data and refine predictions accordingly. This adaptability ensures that bid price estimates remain robust and reflective of real-time projects dynamics **[70]**.

One of the key strengths of ML algorithms lies in their ability to continuously update and improve their models as new information becomes available **[71]**. This adaptability ensures that bid price estimates remain robust and reflective of real-time project dynamics **[72]**. For instance, if unexpected delays or material shortages occur during a project, ML algorithms can quickly integrate this information into their calculations and adjust the bid price estimate accordingly, helping contractors maintain profitability and project viability **[73]**.

Furthermore, ML algorithms can analyze historical project data to identify patterns and trends, enabling them to anticipate potential challenges and adjust bid prices preemptively **[74]**. This proactive approach not only enhances the accuracy of bid price estimates but also helps mitigate risks associated with project uncertainties **[75]**.

Moreover, the adaptability of ML algorithms extends beyond individual projects to encompass broader market trends and economic conditions **[76]**. By continuously monitoring and analyzing data from various sources such as industry reports, economic indicators, and supply chain dynamics, ML algorithms can provide valuable insights into market fluctuations and their potential impact on bid price estimation **[77]**. This foresight allows contractors to adjust their bidding strategies accordingly, optimizing their competitiveness and profitability in dynamic market environments **[78]**.

1.5.2.3. Automation and Efficiency:

ML based bid price estimation automates labor-intensive tasks, streamline data analysis and expediting decision-making processes **[79]**. By automating repetitive tasks such as data preprocessing, model training and feature processing, ML models liberate valuable resources and time, enabling contractors to focus on risk assessment and strategic planning **[80].**

One of the primary advantages of ML-based systems is their ability to automate repetitive tasks that traditionally consumed significant amounts of time and resources **[81]**. Tasks such as data preprocessing, model training, and feature processing can now be efficiently handled by ML algorithms, freeing up valuable human resources to focus on more strategic endeavors **[82]**. By automating these mundane tasks, ML models liberate contractors from manual labor, allowing them to allocate their time and expertise to critical activities such as risk assessment, project planning, and client engagement **[83]**.

1.5.2.4. Transparency and Interpretability:

ML approaches facilitate interpretability and transparency in bid price estimation by providing explicit insights into the underlying factors affecting predictions **[84]**. Modern ML algorithms provide a transparency through interpretability tools, feature importance analysis and model visualization approaches. This transparency enhances confidence of stakeholders in the bid estimation process, fostering trust and accountability across the construction sector **[85].**

Additionally, model visualization approaches play a crucial role in enhancing interpretability. Through intuitive visualizations, ML models can depict complex relationships between input variables and bid prices, making it easier for stakeholders to grasp the underlying patterns and trends **[86]**. Visualization techniques such as decision trees, partial dependence plots, and heatmaps provide stakeholders with a comprehensive overview of how different factors interact to influence bid price estimates **[87]**.

1.6. Future Directions and Opportunities in Machine Learning for Bid Price Estimation:

Looking ahead, the integration of machine learning into bid price estimation processes presents exciting opportunities for innovation and advancement in the construction industry **[88]**. As ML algorithms continue to evolve and improve, contractors can leverage predictive analytics to optimize bidding strategies, mitigate risks, and enhance project outcomes **[80]**. Furthermore, the growing availability of cloud computing infrastructure and data analytics platforms democratizes access to ML tools, enabling even small and medium-sized contractors to harness the power of data-driven decision-making **[89]**. Additionally, advancements in artificial intelligence (AI) techniques, such as natural language processing (NLP) and computer vision, hold promise for augmenting bid price estimation capabilities by extracting insights from unstructured data sources and automating manual tasks **[90]**. Embracing these emerging technologies fosters a culture of innovation and competitiveness within the construction industry, driving efficiency, sustainability, and resilience in project delivery **[91]**.

1.7. Problem Statement:

Developing a reliable prediction model for contractor bid prices with the help of a machine learning approach is a major problem in the construction sector. The complexities of construction projects, that includes diverse variables including material expanses, project specifications, labor costs and market conditions, necessitates an adaptable and accurate model. Addressing this problem requires overcoming hurdles including data complexity, selecting relevant features, ensuring model accuracy across diverse project types and regions, maintaining interpretability and enabling real time application. The resolution of these problems will improve project feasibility, bid price estimation and increasing competitiveness in the construction bidding process.

1.8. Aim and Objectives

The objectives of the current study were;

1. To determine the influential factors affecting contractor bid prices in Pakistan.
2. To develop a machine learning (ML) model capable of predicting optimum bid price for contractor.

1.9. Significance of Research:

1.9.1. Enhancing Bid Price Accuracy:

Implementing a prediction model based on machine-learning for contractor bid prices significantly enhance accuracy as compared to conventional estimation approaches. This improvement is crucial for ensuring project feasibility and fair competition in the construction sector **[92]**.

1.9.2. Advancing Data-Driven Decision-Making:

Machine-learning approaches contributes to advancing data-driven decision-making processes within the construction sector. The model enables risk assessment, informed bidding strategies and resource allocation, resulting in improved profitability and enhance efficient project management for contractors **[92]**.

1.9.3. Promoting Transparency and Trust:

The development of a reliable and transparent bid price prediction model builds trust among their clients and stakeholders. The model promotes accountability and transparency in the construction bidding process, eventually enhancing the reputation of the industry by providing interpretable insights into the elements influencing bid costs.

1.9.4. Enabling Strategic Planning and Innovation:

The insights offered by the prediction model provide valuable opportunities for strategic planning, resource management and collaboration with contracting firms. Contractors might use the analysis of the proposed model of project dynamics and market trends to identify growth opportunities, guide business development efforts and remain ahead of competitors **[93]**.

1.9.5. Driving Sustainable Growth and Development:

Overall, the research on establishing a prediction model for contractor bid prices with a machine learning approach contributes to driving long-term growth and development of the construction industry. The proposed model contributes significantly to a more efficient, competitive and resilient construction sector by simplifying the bidding processes, encouraging innovation and enhancing project outcomes **[94]**.

1.10. Expected Outcomes:

1. The implementation of the machine learning model would result in noticeably improved accuracy in bid price estimation, ensuring competitive and fair pricing for construction projects.
2. Increased efficiency in bid preparation enables quicker response of contractors to tender invitations and project queries, thereby enhancing the competitiveness in the market.

3. The insights provided by the proposed model will empower contractors to make better informed decisions, leading to improved risk management strategies and reduced the likelihood of under-bidding and over-bidding.

Chapter 2

Literature Review

2.1. Introduction:

The construction sector is vital to the global economy, with contractor bid price estimation being an important aspect of project planning and execution. Accurate prediction of bid prices fosters fair competition among contractors and assists project owners in selecting the most suitable bid for their projects **[16]**. Conventional methods of bid price prediction method rely primarily on historical data, manual calculations and experts' judgements, often leading to inefficiencies and inaccuracies. In recent years, using machine learning techniques into bid price prediction has emerged as a viable strategy for enhancing efficiency, accuracy and transparency in the bidding process **[4]**.

2.2. The Role of Machine Learning in Bid Price Prediction:

Machine learning (ML) approaches provide a data-driven approach to bid price prediction, identify patterns and make predictions based on learned patterns. By automating the estimation process, ML models can manage large volumes of data efficiently while adapting to changing project parameters. Furthermore, ML algorithms may include aspects other than historical data, such as materials pricing, market trends and labor availability, resulting in more accurate and bid price estimates. Moreover, ML models have the potential to enhance accountability and confidence among stakeholders **[95]**.

2.3. Benefits of Machine Learning in Bid Price Estimation:

The use of machine learning techniques into bid price estimate provides several benefits to stakeholders in the construction sector. Firstly ML models can enhance the accuracy of bid price predictions by taking into account several factors and capturing complicated interactions between variables. This allows contractors to submit competitive bids that represent the real cost of project execution, limiting the possibility of over or under-bidding. Furthermore, ML-based estimations eliminate reliance on expert judgement and manual computations, speeding up the bidding process and saving time for project owners and contractors. Moreover, ML models provide a more fair and competitive bidding environment by offering transparent and data driven insights, allowing decisions to be made based on objective standards rather than judgements **[96]**.

2.4. Advancements in Machine Learning for Construction Cost Prediction:

Shehadeh et al., 2021 [97] highlights the challenges in making precise models for estimating the cost of heavy construction equipment using traditional methods. In this study, supervised machine learning algorithms were used to explore the datasets during various phases, including training, testing, cross validation and modeling. To analyze and compare the accuracy of the algorithms, four performance metrics such as Mean Absolute Error (MAE), Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Mean and Coefficient Determination R^2 were used. On the basis of coefficient determination results, the MDT algorithm showed the highest prediction accuracy of 0.9284, versus the LightGBM showed an accuracy of 0.8765, followed by XGBoost, obtaining an accuracy of 0.8493. The MDT might be used as a managerial decision support tool for equipment buyers, sellers and owners to perform equipment life cycle analysis and take equipment selling, repairing, purchasing, overhauling, replacing and disposing decisions. Thus, this study supports machine learning's ability to advance automation as a coherent area of study in the construction sector.

According to Alshboul et al., 2022 **[98]** accurate cost prediction in building construction is crucial, especially in sustainable projects particularly, green buildings. Construction contracts for Green buildings are still in their infancy in the construction industry and stakeholders sometimes lack the expertise necessary to accurately estimate the costs these kinds of projects. In contrast to conventional construction methods, green buildings are designed to incorporate innovative technologies with the goal of mitigating their environmental societal impacts during operation. Consequently, the bidding and awarding processes for green building projects have become more intricate due to the challenges in projecting the initial construction costs and developing criteria for selecting winning bidders. In the present study, machine learning (ML) based algorithms, such as Extreme

Gradient Boosting (XGBOOST), Deep Neural Network (DNN) and Random Forest (RF) have been introduced to estimate the cost associated with green building projects. To evaluate the accuracy of these algorithms, several evaluation metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE) and R^2 have been applied. Among these models, XGBOOST was the most accurate, achieving a high accuracy score of 0.96, whereas DNN showed an accuracy of 0.91, followed by RF model with an accuracy of 0.87.

Cao et al., 2018 **[99]** introduced an ensemble learning model design to predict unit bids price. Data on bidding prices for over 1400 projects in the past 9 years was collected, which included information on 57 related variables. After conducting Boruta feature analysis, 20 relevant variables were selected to train and test the proposed model. The study involved comparing the results obtained from the proposed ensemble learning model with a baseline Monte Carlo simulation and a multiple linear regression model. The comparison showed that the ensemble learning model performed better than the other two models. In particular, the mean absolute percentage error of the ensemble learning model was 7.56%.

Kim et al., 2019 **[100]** discussed the high dimensionality and challenges of multicollinearity in empirical analysis, specifically with regard to itemized bids in highway procurement auctions. The study uses regularized linear regression to predict more accurate intervals for project winning bids in order to address these problems. The efficacy of this approach is assessed by analyzing actual data from Vermont highway procurement auctions. First, the study uses the random forest variable selection approach to identify a set of key tasks that influence bidder bid amounts. With these tasks identified, the study proceeds to forecast project bids. A comparison is made between the suggested methodology and the conventional least squares linear model based on bias and standard root mean square error of bid estimates. The results suggest that the proposed approach provides superior forecasts for an interval of winning bids over the competing model.

Takano et al., 2018 **[101]** focus on an important component of competitive bidding in project contracts, where contractors set bid prices by adding a markup to the expected project cost. The accuracy of this estimate is critical to bid success, requiring proper resource allocation for estimation. The research presents a novel optimization model for estimating costs by concurrently determining bid markups and resource allocations. Initially, the work defined optimality assumptions for this simultaneous optimization model and provided numerical examples based on a single competitor and evenly distributed estimate errors. The research then investigates computational solutions for the model in order to evaluate a more realistic scenario. Through these analysis, the study investigates into the influence of bid markup decisions and resource allocation on contractors' predicted profit, emphasizing their importance in competitive bidding.

Zaqout et al., 2022 **[102]** introduced a novel Fuzzy Interference System (FIS) for identifying crucial factors and estimating markups in construction projects. Expert weights are first determined by this model based on experience and academic qualifications. Then, it assesses the weights of frequency, severity and importance of several factors, identifying crucial factors such as current workload, labor availability and project size. Finally, considering these variables and contractor assessments, the model predicts bid mark-ups. It has been tested on real projects and produces consistent results, assisting contractors with bidding decisions and proper risk management. Future study might enhance the model by integrating competitor bids and improving prediction accuracy for optimal prices.

Khan et al., 2022 **[103]** investigated the effect of stabilizing alkali-contaminated soil by using fly ash. The influence of alkali concentration (2 N and 4 N) and curing period (up to 28 days) on the unconfined compressive strength (UCS) of fly ash (FA)-treated (10%, 15%, and 20%) alkali contaminated kaolin and black cotton(BC) soils was investigated. The effect of incorporating different dosages of FA (10%, 15%, and 20%) on the UCS kaolin and UCSBC soils was also studied. Sufficient laboratory test data comprising 384 data points were collected, and multi expression programming (MEP) was used to create tree-based models for yielding simple prediction equations to compute the UCS kaolin and UCSBC soils. The experimental results reflected that alkali contamination resulted in reduced UCS (36% and 46%, respectively) for the kaolin and BC soil, whereas the addition of FA resulted in a linear rise in the UCS. The optimal dosage was found to be 20%, and the increase in UCS may be attributed to the alkali-induced pozzolanic reaction and subsequent gain of the UCS due to the formation of calcium-based hydration compounds (with FA addition). Furthermore, the developed models showed reliable performance in the training and validation stages in terms of regression slopes, R, MAE, RMSE, and RSE indices. Models were also validated using parametric and sensitivity analysis which yielded comparable variation while the contribution of each input was consistent with the available literature.

According to Chao et al. 2016 **[104]** aimed to develop an improved approach to determining the combined rate of overhead and markup in the bid price for a project. Four factors, i.e., direct cost, duration, type of work, and location, were used as inputs to build a regression model from cost and bid data of collected projects for predicting the overhead and markup rate in the winning bid for a project, which, together with the model error, is used to estimate the probability of winning for a bid level. Then, based on minimization of overall loss risk proposed by a previous research, the bid preventing over-cuts in price competition is determined by using the model, the probabilistic estimates of project cost, and the probability of recovering costs if losing the bid. The approach is illustrated using two cases and the suggested bids for the cases are compared with those from other models.

2.5. ANALYZING BID MARKUPS AND OVERHEAD RATE DETERMINATION IN CONSTRUCTION SECTOR:

Existing models for bidding in construction focus on determination of the markup rate in the bid price. In conventional models the optimum markup rate is suggested as one with the maximum expected profit, where the expected profit for a markup rate is defined as the product of it and its probability of winning that is estimated using statistics of past bids. Theoretically such a markup rate will achieve the highest profit in the long term, but it tends to give too low a chance of winning for contractors in intense competition, who often sacrifice profit in order to raise the chance of winning. As an illustration, a zero markup will never be recommended by conventional models since its expected profit of zero is always less than a positive markup’s, but a zero markup is not uncommon in construction **[105]**.

Ahmad and Minkarah **[106]** took the lead in conducting a comprehensive study of factors influencing the markup decision that are grouped into the environment, company, and project aspects. Others followed, identified key factors affecting bid-reasoning sub-goals. Meanwhile, various multi-criteria markup models built upon identified factors have been proposed, including the case-based reasoning model, and the fuzzy neural network model. They offered various methods for producing an optimum markup for a project, yet they did not determine how low a bid could be and provide a rational solution in line with market conditions in which bidding competitively is imperative to survive **[104]**.

Chapter 3

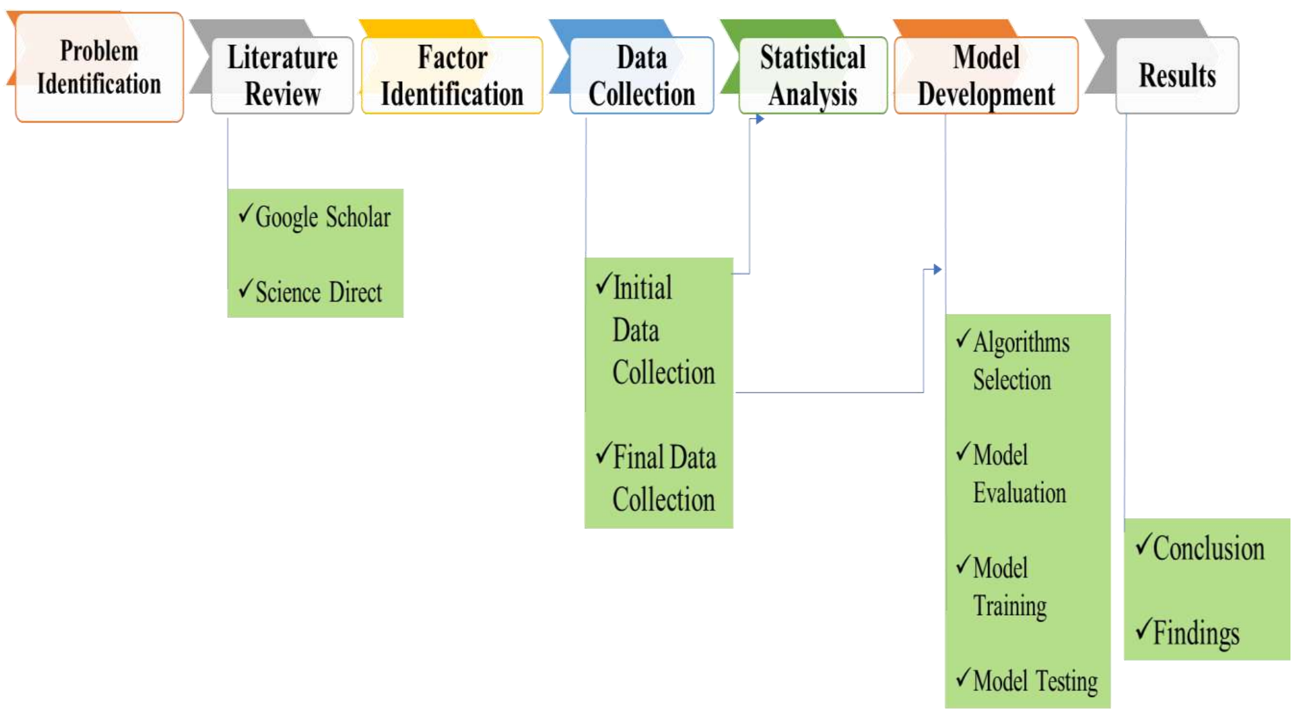
Methodology

3.1. Sustainable Development Goals (SDGs):

The project targeted the following (SDGs):

- ✓ SDG-9: Sustainable Cities and Communities
- ✓ SDG-11: Industry Infrastructure and Innovation

3.2. Study Design:



3.3. Data Collection:

3.3.1. Initial Data Collection:

The collection of relevant data was the first step in developing a prediction model. The data included historical bid prices, location, labor costs, materials and final bid prices.

Questionnaires were designed through Google Forms to collect the relevant data. The questionnaires were distributed via links and meetings to collect responses from contractors, academia, consultants and professional engineers. Responses were collected to evaluate the factors impacting bid prices **[107]**.

3.3.2. Secondary Data Collection:

The designed questionnaires were printed in sufficient quantity and distributed among contractors. The responses were collected and analyzed further **[107]**.

3.4. Factor Ranking:

Each factor was ranked in response to the questionnaire based on its importance using a Likert scale. The responses were assessed to determine the importance of each factor in determining bid prices **[108]**.

3.5. Response Collection:

A total 200 factors were identified in initial stage of literature review. From those 200 factors most were overlapping, while some of them were least important and the factors were reduced to almost 60 factors. By discussing these with our supervisor, experienced engineers and contractors they were further refined to 39 influential factors.

3.6. Statistical Analysis:

On the basis of responses received, 39 influential factors were identified and statistical analysis was performed to determine the importance of each factor in determining bid prices. Relative Importance Index (RII) analysis was performed to identify the most relevant factors. Each factor was rated on the basis of its importance index. Factors with a threshold of greater than 0.8 was selected. Nine factors out of 39 showed Importance Index of above 0.8. These factors were involved in impacting bid prices and were analyzed further **[109]**.

3.7. Data Preprocessing:

After identifying the most important factors, the data was reprocessed, cleaned and prepared for further analysis. Reprocessing involved handling of missing values, eliminating duplicated values and addressing outliers. To improve the performance of the model, numerical variables were scaled on a similar scale, thereby improving model performance **[110]**.

3.8. Feature Engineering:

The most significant features for bid price estimation was determined based on the dataset, considering several factors including project size, contractor experience, location, complexity, material costs, economic indicators and labor costs. The importance of all features in estimating bid prices was analyzed through recursive feature elimination and correlation analysis. Duplicated and less impacting features were deleted to simplify the model and enhance its efficiency. New features were developed when needed, by manipulating and merging existing features to detect underlying patterns in the data. It was assured that the selected features showed the variability in bid prices across several projects and market conditions **[111]**.

3.9. Model Selection:

The most suitable machine learning algorithm for building an accurate prediction model involved considering several options including linear regression, ensemble methods and neural networks. Different algorithms and hyper-parameters were assessed using approaches such as grid search cross-validation to identify the most accurate performing model on the basis of nature of data, trade-off between bias and variance and desired interpretability level.

3.9.1. Gene Expression Programming (GEP):

Along with the linear regression, ensemble methods and neural networks, Gene Expression Programming (GEP) was also selected as a model for bid price estimation. Gene Expression Programming was selected on the basis of its capacity to handle complex interactions and non-linearity among different variables in bid price estimation. The GEP was then applied to the preprocessed data sets. This approach used mutation, selection and recombination to evolve the mathematical expressions (chromosomes) of a population. The expressions represented hypothetical model that linked input features, i-e., contractor experience, project size and location to the targeted variable, which is a bid price in this case.

A particular initialization approach was used to create the initial population of candidate expressions. The algorithms developed the population by applying genetic operators (mutations and recombination) on the basis of their fitness. The expression of each candidate was assessed on the basis of how well it estimated the bid prices in comparison to the actual bid prices in the training data sets. This evolutionary process continued till termination criteria was met, i-e., achieving satisfactory performance and maximum number of generations. GEP algorithm parameters including mutation rate, population size and cross over rate was optimized to enhance the performance of model. The performance of this trained GEP model was assessed through metrics including mean absolute error (MAE), mean squared error (MSE) and root mean square deviation (RMSE) score. These metrics analyzed the difference between actual and estimated bid prices.

3.10. Model Training:

The selected model was run on the pre-processed data to assess the underlying patterns and relationships between bid price (target data) and input features. The data set was split into validation sets and training to analyzed the performance of selected model during training. The parameters of model were adjusted through optimization algorithms including gradient descent to minimize estimation error.

3.11. Model Evaluation:

The generalization ability and accuracy of the trained model was analyzed through performance metrics including mean absolute error (MAE), mean squared error (MSE) and root mean square deviation (RMSE) score. The strength and weakness of the model was assessed by scatter plots of actual bid prices and predicted bid prices. Further iterations including tuning hyper-parameters, revisiting feature engineering and selecting several algorithms were conducted if the proposed model didn't meet desired performance criteria **[112].**