Jurnal E-Komtek
Vol. 10, No. 1. (2026) pp. 324-337
https://jurnal.politeknik-kebumen.ac.id/index.php/E-KOMTEK
p-ISSN : 2580-3719 e-ISSN : 2622-3066

Development of an Electrical Work Unit Price Analysis System with Random Forest Regressor Approaches

Gita Winanda Pramesthi[1], Naufal Abdillah[2]
[1-2]Department of Informatics, Universitas 17 Agustus 1945, Indonesia, 60118

naufal@untag-sby.ac.id

https://doi.org/10.37339/e-komtek.v10i1.3302

Published by Politeknik Piksi Ganesha Indonesia

Artikel Info
Submitted:
07-07-2026
Revised:
08-07-2026
Accepted:
10-07-2026
Online first :
10-07-2026

Abstract

The preparation of Cost Budget Plans (RAB) for electrical installation projects is still commonly performed manually using Microsoft Excel, making cable requirement calculations and Unit Price Analysis (AHS) time-consuming and prone to errors. This study develops a machine learning-based system to predict the cable length required for lighting installations using room characteristics, including the number of rooms, floor area, room height, and the number of spotlight, downlight, and pendant lighting points. Two prediction models, Random Forest Regressor and Artificial Neural Network (ANN), were trained using historical residential electrical installation project data with an 80:20 training-testing split. Model performance was evaluated using R^2, MAE, MSE, RMSE, and MAPE. The Random Forest Regressor achieved superior performance with an R^2 of 0.982 and a MAPE of 18.94%, outperforming the ANN (R^2 = 0.942; MAPE = 19.34%). The best model was integrated into a Streamlit-based web application to support faster, more accurate, and consistent cable estimation for efficient AHS and RAB preparation.

Keywords: Unit Price Analysis; Electrical Installation; Machine Learning; Random Forest Regressor; Artificial Neural Network

Abstrak

Penyusunan Rencana Anggaran Biaya (RAB) pekerjaan instalasi listrik masih banyak dilakukan secara manual menggunakan Microsoft Excel, sehingga proses perhitungan kebutuhan kabel dan Analisis Harga Satuan (AHS) menjadi kurang efisien serta berpotensi menimbulkan kesalahan. Penelitian ini bertujuan mengembangkan sistem berbasis machine learning untuk memprediksi panjang kabel instalasi penerangan berdasarkan karakteristik ruangan, meliputi jumlah ruangan, luas lantai, tinggi ruangan, serta jumlah titik lampu spotlight, downlight, dan lampu gantung. Dua model dikembangkan dan dibandingkan, yaitu Random Forest Regressor dan Artificial Neural Network (ANN), menggunakan data historis proyek instalasi listrik bangunan sederhana dengan pembagian data 80% untuk pelatihan dan 20% untuk pengujian. Kinerja model dievaluasi menggunakan R^2, MAE, MSE, RMSE, dan MAPE. Hasil penelitian menunjukkan bahwa Random Forest Regressor memberikan performa terbaik dengan nilai R^2 sebesar 0,982 dan MAPE 18,94%, lebih baik dibandingkan ANN. Model terbaik kemudian diimplementasikan pada aplikasi berbasis Streamlit untuk mendukung penyusunan AHS dan RAB yang lebih cepat, akurat, dan konsisten.

Kata Kunci: *Analisis Harga Satuan; Instalasi Listrik; Machine Learning; Random Forest Regressor; Artificial Neural Network*

1. Introduction

Electrical installation acts like the neuron system in a building, playing a crucial role in supporting safe and efficient operational functions [1]. Electrical installation work involves lighting engineering design, light intensity, distribution systems, the number of lighting points or socket outlets, material types, and light reflection factors, thereby requiring high precision from electrical contractors and subcontractors. The larger the building and the more electrical installation points planned, the more complex the calculation of material requirements (such as cables, fittings, switches) and labor estimations becomes. This ultimately consumes considerable time in the decision-making process, particularly during project bidding processes that operate under strict time constraints [1].

Most of the Cost Budget Plan (BoQ) calculations for electrical construction work are still performed manually by estimators using Microsoft Excel [2]. While this tool helps with data processing, input errors remain frequent in practice. An estimator is also responsible for creating the Unit Price Analysis (AHS) to provide an accurate valuation of the work. The official reference used in preparing the AHS for construction work in Indonesia, including lighting systems, is the Regulation of the Minister of Public Works and Public Housing No. 1 of 2022 concerning Guidelines for the Preparation of Construction Work Cost Estimates [3]. This electrical work unit price analysis is influenced by the building floor area, which determines the estimated cable length, as well as the number of lighting points and sockets, which determine the need for fittings and switches. These quantities are derived by measuring the existing blueprints and aligning them with the RAB format. Because this calculation is still done manually, there is a high probability of errors, which drains a significant amount of time and effort.

Developments in information technology have driven computerization across various sectors, including construction [1], However, currently available applications are still limited to conventional calculations without automated prediction capabilities [4].

Machine learning-based approaches have proven effective in improving the accuracy of cost and duration estimations in construction projects using algorithms such as the Random Forest Regressor and Artificial Neural Networks (ANN), through stages of data preprocessing, feature selection, and model training [5]. Therefore, an intelligent system is needed to predict electrical installation material needs—specifically cable length—using only simple room parameters, enabling the estimation process to be conducted more quickly and consistently.

Related studies conducted by previous researchers such as, [1] developed a client-server application using Visual Basic 6.0, MySQL, and Crystal Reports to calculate residential electrical installation costs, but it remained manual without predictive capabilities [2]. Developed a web-based RAB software using PHP Laravel that supports integrated calculations and validation against the 2022 SNI standards, but it still requires detailed inputs of work volumes and wage rates [4]. Applied Random Forest Regressor and ANN for general building construction cost and time estimations, but did not focus specifically on electrical installation AHS. Meanwhile, [5] applied a hybrid machine learning approach for the conceptual cost estimation modeling of school building projects with high accuracy, but it was conducted on a large building scale, which differs in context from residential installations. [6] applied the Random Forest algorithm to predict general construction costs through a web platform, demonstrating that the ensemble tree approach is reliable for construction data with limited sample sizes.

Based on the research gaps identified above, it is evident that no system currently integrates machine learning with simple room parameter inputs—such as the number of rooms, surface area, room height, and the number of lighting points—to specifically predict lighting installation cable requirements. Previous studies generally still require detailed work volume inputs or focus on general building construction cost estimates, meaning they cannot be directly utilized by electrical estimators for residential and simple building lighting installations.

This study aims to develop and compare two machine learning models, namely the Random Forest Regressor and Artificial Neural Network, in predicting lighting installation cable length based on the parameters of room count, surface area, room height, and the number of spotlight, downlight, and pendant lighting points. The best-performing model from the comparison will then be implemented into a Streamlit-based web application for practical use by estimators and electrical contractors. This research is expected to accelerate the AHS calculation process, increase the accuracy of material requirement estimations, and serve as an alternative calculation tool for estimators to support a more efficient electrical installation RAB preparation.

2. Method

This research utilizes a system development approach based on the Waterfall model, which consists of five main stages: requirements analysis, system design, data collection and preprocessing, model implementation, and testing and evaluation. This approach was chosen due to its linear and well-documented workflow, as applied in similar application development research [1], [2]. The general workflow of this research can be seen in **Figure 1**.

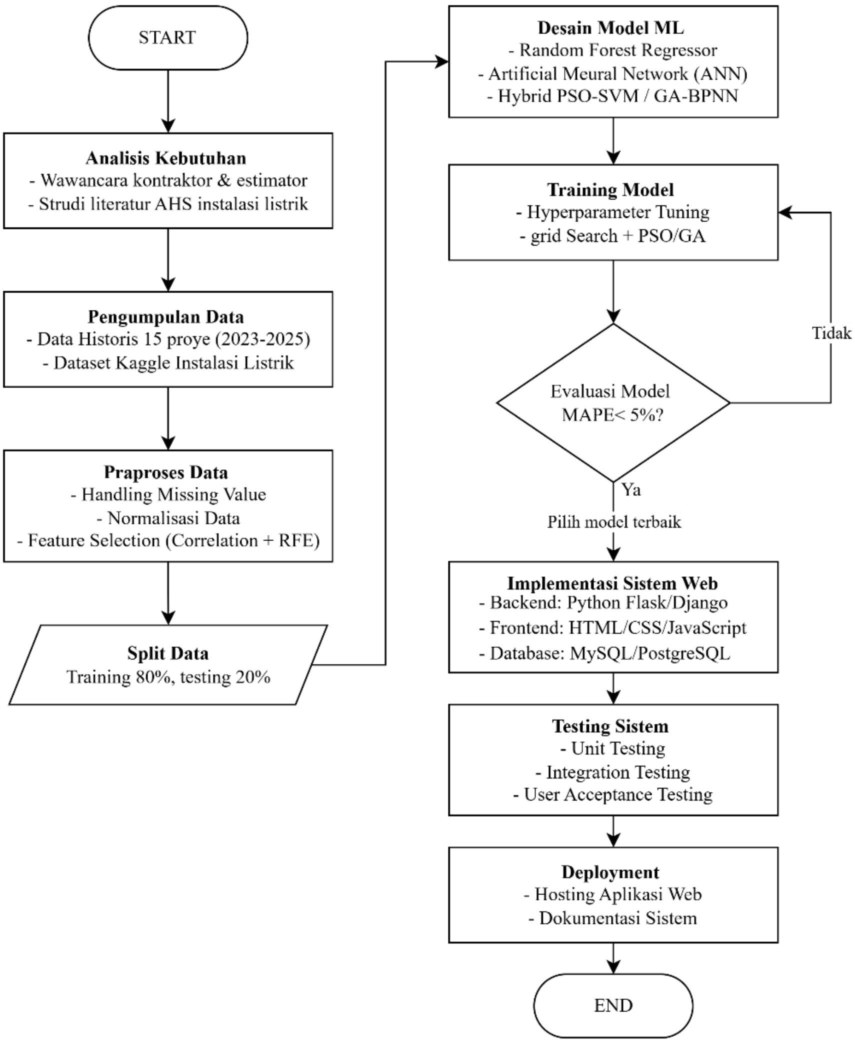

Figure 1. Flowchart Detailing the Research Workflow from Requirements Analysis to Deployment

The research workflow begins with a requirements analysis through interviews with contractors and estimators, followed by the collection of historical electrical installation project datasets, data preprocessing, splitting data into training and testing sets, training the Random Forest Regressor and ANN models, evaluating both models' performances using standard metrics, and finally implementing the best model into a web application.

Based on the literature review, the appropriate machine learning approach for estimating lighting installation cable lengths is a regression-based model, since the task involves predicting a numerical value from numerical inputs of room parameters, as commonly applied in machine learning-based construction cost estimation models [7]. This approach can handle both linear and non-linear relationships between variables, such as cable lengths that depend on room area, room height, and the number of lighting points.

2.1 Data Collection and Preprocessing

The dataset used originates from a collection of lighting installation calculations across 15 residential and simple building projects from the 2022 to 2026 period, obtained through measurements of working drawings using the AutoCAD application. Each project was recorded in the form of features: number of rooms (`rooms`), room surface area (`surfaces`), room height (`height`), number of lighting points based on type, namely `spotlight`, `downlight`, and `pendant`, total number of lamps (`totLamps`), and the target variable, which is the lighting installation cable length (`cableLength`). The cable length was calculated by drawing lines between lighting points and the switches, then adding the estimated building height as a vertical factor.

Each row in the dataset represents one room, meaning that a single project can contribute several rows of data according to the number of rooms it contains. This room-by-room recording approach was chosen so that the model could capture variations in room characteristics in greater detail compared to data recorded only at the overall project level.

Data preprocessing was carried out using the Python programming language on the Google Colab platform with the help of the Pandas, NumPy, and Scikit-learn libraries. The preprocessing stages included ensuring numerical data types across all columns, filling missing values with 0 if there were rows without data, and separating the `cableLength` column as the target variable from the other feature columns. A proper preprocessing stage, including handling missing values and appropriate feature selection, is reported to significantly increase the accuracy of machine learning models compared to using raw data without preprocessing [8]. The NumPy library was also used in the subsequent phase to numerically calculate model evaluation parameters, such as the mean error and the root mean squared error.

2.2 Data Splitting and Model Training

The preprocessed dataset was then split into 80% training data and 20% testing data using the `train_test_split` function from Scikit-learn. This division aims to evaluate the model's performance objectively on data it has never seen before.

The first model built was the Random Forest Regressor, an ensemble learning method that combines many decision trees to improve prediction accuracy and reduce variance compared to a single decision tree [9]. The model was built with the `n_estimators` parameter set to 100, which determines the number of decision trees, and `random_state` set to 42 to lock random values so that the results can be reproduced. The model fitting process used the training data feature matrix, including `rooms`, `surfaces`, `height`, `spotlight`, `downlight`, `pendant`, and `totLamps`, with `cableLength` as the target. Each tree was trained on a data subset generated via bootstrap sampling and considered only a portion of the features at each split. This bagging mechanism helps reduce the risk of overfitting. The model's final prediction is the average of all tree predictions, as formulated in equation (1).

$$\hat{y} = (\mathbf{1/N}) \times \boldsymbol{\Sigma}\, \mathbf{tree}_i(\mathbf{x}) \qquad (1)$$

Where N s the number of trees and $_i$(x) is the prediction result of the-i tree.

The second model was an Artificial Neural Network (ANN) implemented using Scikit-learn's MLPRegressor, which is a Multilayer Perceptron implementation for regression tasks. The model architecture used one hidden layer with 100 neurons (hidden_layer_sizes = 100), a `max_iter` parameter of 1000 to limit the number of training iterations, and a `random_state` of 42 for reproducibility. The training process consisted of forward propagation to generate predictions, loss calculation using Mean Squared Error, and backpropagation to update network weights based on the error gradient, which was repeated until convergence or the maximum iteration limit was reached.

2.3 Model Evaluation

The performance of both models was evaluated using five metrics: R^2 Score, Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), which are standard metrics commonly used in artificial intelligence research for construction project cost estimation [10]. MAPE was selected because it provides an easily understood percentage representation of error, using the formula in equation (2).

$$\mathbf{MAPE} = (\mathbf{1/n}) \times \boldsymbol{\Sigma}\, |(\mathbf{Aktual} - \mathbf{Prediksi})/\mathbf{Aktual}| \times \mathbf{100\%} \quad (2)$$

According to Lewis's interpretation [11], a MAPE value below 10% is categorized as highly accurate, 10–20% is categorized as accurate, 20–50% is categorized as reasonably accurate, and above 50% is categorized as inaccurate. The R^2 Score measures how well the model explains the data variability, with a range of values from 0 to 1, while RMSE and MAE are used to observe the magnitude of prediction errors in meters, where smaller values indicate better performance.

2.4 System Implementation

The top-performing model from the evaluation phase was integrated into a web application using the Streamlit framework. Streamlit was selected because it enables rapid development of data-driven interfaces without requiring extensive front-end expertise, allowing the machine learning model to be deployed as an interactive app with minimal Python code [12]. To bypass the need for retraining, the trained model was preserved in a pickle (.pkl) format. The project architecture comprises app.py for the primary user interface and model invocation logic, model_rf_ahs.pkl for the trained weights, and requirements.txt to manage Python library dependencies. To optimize user experience, the interface utilizes an equal two-column layout (1:1 ratio) that places the room parameter inputs directly alongside the cable length prediction results, enabling side-by-side data comparison without scrolling.

3. Results and Discussion

3.1 Dataset Characteristics

The dataset comprises historical electrical installation data for lighting across 15 residential and simple building projects. Each data row represents an individual room, with the number of rooms per project ranging from 1 to 5. The dataset includes architectural and lighting fixture features: the surfaces column denotes the room's surface area, the height column indicates the ceiling height, while the spotlight, downlight, and pendant columns specify the types of luminaires installed. For instance, projects utilizing fluorescent lamps (TL) are recorded under downlight, track lights or surface-mounted spotlights are logged under spotlight, and decorative hanging lights are categorized under pendant. The target variable, cableLength, was derived from manual cable measurements on AutoCAD blueprints, supplemented by the estimated building height to account for the vertical installation component.

3.2 Model Training and Testing Results

Following the preprocessing stage, the dataset was split into an 80% training set and a 20% testing set. Both models, namely the Random Forest Regressor and ANN, were trained using the same data split to ensure a fair comparison. The Random Forest model was trained with 100 decision trees using the features \`rooms\`, \`surfaces\`, \`height\`, \`spotlight\`, \`downlight\`, \`pendant\`, and \`totLamps\` to predict the \`cableLength\` target. Because each constituent tree was trained on a unique data subset generated via bootstrap sampling, the final ensemble prediction was computed as the average of all trees, rendering this architecture inherently robust against data outliers, such as atypical cable length measurements.

The ANN model was configured with a single hidden layer containing 100 neurons and trained for a maximum of 1,000 iterations. This training pipeline encompassed forward propagation for initial prediction generation, Mean Squared Error (MSE) calculation for loss assessment, and backpropagation for iterative weight updates. To ensure consistency across the evaluation metrics, both the Random Forest Regressor and the ANN models were evaluated using the exact same testing dataset, allowing for a rigorous comparative analysis.

3.3 Model Evaluation and Comparison

The evaluation results for both models on the testing data are presented in **Table 1**.

Table 1. Model Evaluation Comparison Results

Metrik	Random Forest	ANN	Acuan
R^2 Score	0,982	0,942	> 0,80 = baik
MAE (meter)	3,57	7,23	semakin kecil semakin baik
MSE	28,991	94,558	semakin kecil semakin baik
RMSE (meter)	5,385	9,724	semakin kecil semakin baik
MAPE	18,94%	19,34%	< 20% = baik

Based on Table 1, the Random Forest Regressor model demonstrates superior performance compared to the ANN across all evaluation metrics. The R^2 value of 0.982 for Random Forest indicates that the model can explain approximately 98.2% of the variability in cable length data, whereas the ANN only reached 0.942. The MAE and RMSE values for Random Forest are also significantly lower at 3.57 meters and 5.385 meters, respectively, compared to the ANN, which produced an MAE of 7.23 meters and an RMSE of 9.724 meters. This shows that the average prediction error for cable length in the Random Forest model is much smaller than in the ANN.

Regarding the MAPE metric, both architectures demonstrated performance within the "accurate" threshold based on Lewis's classification **[11]**, yielding 18.94% for the Random Forest and 19.34% for the ANN. Despite these acceptable results, neither model achieved the initial study benchmark of a MAPE below 5%. Although the margin in MAPE between the two approaches is marginal, the Random Forest's resilience to outliers and its capacity to model non-linear relationships among room dimensions, lighting fixtures count, and cable length positioned it as the more viable candidate for subsequent deployment. This strength corroborates existing literature on Random Forest behavior, which indicates that tree-based ensemble frameworks regularly exhibit superior stability on small-to-medium datasets compared to artificial neural networks that typically mandate extensive data scales for optimization [4], [5].

The observed performance variance between the two models stems from the core algorithmic differences inherent to each approach. Random Forest builds an ensemble of independent decision trees via bootstrap sampling and feature subsampling, aggregate-averaging diverse viewpoints to formulate its final prediction. This structural mechanism provides the model with notable resilience when operating under constrained training datasets, a critical advantage given that this study relied on only 15 foundational projects. In contrast, the Multilayer Perceptron-based ANN depends on backpropagation to iteratively fine-tune network weights—a optimization process that typically demands a vast data footprint for robust pattern generalization. Consequently, when constrained to small-to-medium datasets, ANNs exhibit a higher susceptibility to either overfitting or underfitting, which explains why the empirical outcomes of this research align precisely with these theoretical limitations.

These findings strongly align with a previous study on highway construction cost estimation in Ethiopia, which reported that the Random Forest model generated a lower RMSE compared to both ANN and Support Vector Machine frameworks when trained on data-scarce construction project portfolios [13]. Similarly, in an investigation regarding residential building price predictions in Egypt, both Random Forest and ANN architectures demonstrated highly accurate predictive capabilities; however, their respective performance profiles remained closely tied to the underlying volume and structural complexity of the datasets. Furthermore, within the domain of power plant cost estimation, Random Forest has been documented to systematically outperform alternative baselines—such as Support Vector Regression and K-Nearest Neighbors—across key evaluative dimensions, including R^2, MAE, and RMSE. Collectively, these comparative benchmarks reinforce the broader empirical consensus that Random Forest exhibits superior stability and reliability when tasked with construction cost estimation problems involving limited historical records.

Compared to study [4] which deployed Random Forest and ANN for general building construction cost estimation, this research exhibits a similar trend, demonstrating that Random Forest tends to perform superiorly on small-scale historical construction datasets. Meanwhile, compared to [5] which applied a hybrid ANN-PSO approach for estimating school building costs with higher accuracy, this study has not yet implemented advanced hyperparameter optimization such as Particle Swarm Optimization. Consequently, the obtained MAPE value is still in the accurate category but has not yet reached the highly accurate category. This presents an opportunity for further development in future research, for instance, by applying a hybrid PSO-SVM or GA-BPNN as planned in this study's problem boundaries to increase prediction accuracy closer to the target MAPE below 5%..

Based on the evaluation results, it can be concluded that the Random Forest Regressor was selected for web application deployment, as it consistently delivered higher predictive accuracy and lower error rates than the ANN in forecasting lighting installation cable lengths.

3.4 Web Application Implementation

The trained Random Forest model was saved in pickle format and integrated into a web application using the Streamlit framework. The application was developed on the Google Colab platform and managed within a GitHub repository to facilitate version control and online integration with Streamlit hosting services. The application structure consists of the `app.py` file for interface logic, `model_rf_ahs.pkl` as the trained machine learning model, and `requirements.txt` as the dependency list.

Figure 2. Shows the Web Application Interface with input parameters on the left and prediction outputs on the right

The application interface went through several design iterations. In the initial design, all input and output elements were displayed vertically in a single column, but this approach was considered less flexible when users needed to directly compare input parameters with prediction results. The display was then changed to a two-column layout with a 1:1 ratio, featuring an input column for parameters like room count, surface area, building height, and the quantity of each lamp type, alongside a results column displaying the total predicted cable length highlighted in green as the primary output indicator. The interface elements were arranged within container components so that each section had a clear visual boundary and was easy to maintain. The web application results are presented in **Figure 2.**

3.5 Discussion of Implications

The results of this study demonstrate that a machine learning approach, specifically the Random Forest Regressor, can be used to predict lighting installation cable lengths using only simple room parameter inputs, without requiring detailed manual measurements for every installation point. This approach has the potential to shorten the time needed by estimators to prepare the AHS, particularly during the early stages of project bidding, where time is a critical factor in the tender process. Compared to manual calculations using Microsoft Excel, which require detailed inputs of work volumes [2], the system developed in this study only requires seven input parameters that are relatively easy to extract from architectural working drawings: the number of rooms, surface area, room height, and the quantities of each type of lighting point.

Another implication of this study's findings is the importance of selecting a machine learning algorithm that matches the characteristics of the available dataset. In the case of relatively small historical construction datasets like the one used in this study, ensemble tree-based models like Random Forest tend to outperform neural network models, which generally require larger volumes of data to achieve optimal performance. A similar trend was reported in property price prediction research in Indonesia, where Random Forest provided good predictive results on tabular data with relatively limited features and samples **[14].** This finding is relevant not only for electrical installation estimates but can also serve as a consideration for other construction cost estimation research with similar dataset characteristics, as also found in previous Random Forest-based construction cost estimation studies **[6].**

Nevertheless, the MAPE value obtained in this study—18.94% for the Random Forest—falls within the "accurate" threshold but has not yet attained the "highly accurate" classification according to Lewis's benchmarks **[11]** remaining above the initial research target of a MAPE below 5%. This performance profile indicates that the currently developed system is more appropriate as a decision-support tool for preliminary estimation to accelerate the calculation process, rather than a definitive replacement for the detailed and validated Analytical Unit Price Estimation (AHS) produced by professional estimators. Consequently, estimators must still audit and verify the system's predictive outputs, particularly when encountering input parameter combinations that were underrepresented within the training dataset.

3.6 Application Testing

Application evaluation was conducted through usability testing utilizing the System Usability Scale (SUS) framework, involving a cohort of 35 respondents comprised of cost estimators, electrical contractors, and project managers. This evaluation was specifically designed to verify that the application delivers a highly user-friendly experience and aligns with professional workflow requirements. The SUS score results are presented in **Table 2.**

Table 2. SUS Assessment Results

Responden	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Ganjil	Genap	Total
R1	4	2	4	1	4	2	4	2	4	2	15	16	77.5
R2	5	1	5	1	5	1	5	1	5	1	20	20	100
R3	4	1	4	1	4	1	4	1	4	1	15	20	87.5
R4	5	1	5	1	5	1	5	1	5	1	20	20	100
R5	3	3	3	3	3	3	3	3	3	3	10	10	50
R6	3	2	3	2	3	2	3	2	3	2	10	15	62.5
R7	4	3	4	3	4	3	4	3	4	3	15	10	62.5
R8	5	1	5	3	5	1	5	1	5	4	20	15	87.5
R9	5	1	5	1	5	1	5	1	5	1	20	20	100
R10	4	2	4	1	4	2	4	1	4	2	15	17	80
R11	5	3	5	3	5	2	5	3	5	2	20	12	80
R12	5	2	5	4	4	2	5	1	4	3	18	13	77.5
R13	4	2	4	2	4	2	4	2	4	2	15	15	75
R14	5	1	5	1	5	1	5	1	5	1	20	20	100
R15	3	3	3	3	3	3	4	3	4	3	12	10	55
R16	4	1	4	1	4	1	4	1	4	1	15	20	87.5
R17	5	2	5	1	5	2	5	1	5	2	20	17	92.5
R18	3	1	3	1	3	1	3	1	3	1	10	20	75
R19	5	2	5	1	5	2	5	1	5	2	20	17	92.5
R20	4	1	4	2	4	1	4	2	4	1	15	18	82.5
R21	5	1	5	2	5	1	5	2	5	1	20	18	95
R22	4	3	4	3	4	3	4	3	4	3	15	10	62.5
R23	5	3	5	3	5	3	5	3	5	3	20	10	75
R24	5	2	5	2	5	2	5	2	5	2	20	15	87.5
R25	4	1	4	1	4	1	4	1	4	1	15	20	87.5
R26	4	2	4	2	4	2	4	2	4	2	15	15	75
R27	5	1	4	1	5	1	4	1	5	1	18	20	95
R28	5	1	5	1	5	1	5	1	5	1	20	20	100
R29	5	1	5	1	5	1	5	1	5	1	20	20	100
R30	5	1	5	1	5	1	5	1	5	1	20	20	100
R31	5	2	5	2	5	2	5	2	5	2	20	15	87.5
R32	5	1	5	1	5	1	5	1	5	1	20	20	100
R33	5	1	5	1	5	1	5	1	5	1	20	20	100
R34	5	1	5	1	5	1	5	1	5	1	20	20	100
R35	5	1	5	1	5	1	5	1	5	1	20	20	100
											RATA-RATA NILAI SUS		**85.4285**

The evaluation yielded an average SUS score of 85.428, indicating that the application is highly feasible for implementation and well-received by its target users. However, the granular

SUS analysis reveals that targeted enhancements are necessary to address specific user requirements highlighted in question Q4, which will further improve user adoption and maximize the application's utility in professional workflows.

4. Conclusion

This study developed and compared two machine learning models, namely the Random Forest Regressor and Artificial Neural Network, to predict lighting installation cable lengths based on parameters including room count, surface area, room height, and the quantity of spotlight, downlight, and pendant lighting points. Evaluation results on the testing data show that the Random Forest Regressor performs better than the ANN, yielding an R^2 value of 0,982, an MAE of 3,57 meters, an RMSE of 5,385 meters, and MAPE of 18,94%, whereas the ANN generated an R^2 of 0,942, an MAE of 7,23 meters, an RMSE of 9,724 meters, and a MAPE of 19,34%. Based on these results, the Random Forest Regressor was selected as the model implemented into the Streamlit-based web application to assist estimators and contractors in calculating lighting installation cable needs more quickly and consistently compared to manual calculations.

This study still has limitations, including a relatively small dataset size of 15 projects, which prevented the obtained MAPE value from meeting the initial target of under 5%. Additionally, this study focuses strictly on predicting lighting installation cable lengths and does not yet encompass comprehensive AHS and RAB calculations, including labor wages and other material components. Future research is advised to expand the dataset by increasing the number of historical projects, explore alternative models such as XGBoost or hybrid PSO-SVM approaches, and perform comprehensive functional testing like Extreme Testing to validate the application's feasibility before it is widely deployed by estimators and contractors in the field.

5. *Acknowledgement*

The authors would like to express their gratitude to the thesis advisors and partner contractors who supported the provision of historical electrical installation project data for this study.

References

[1] S. Tinggi Teknologi Kedirgantaraan Yogyakarta, “PENGEMBANGAN APLIKASI PERHITUNGAN BIAYA INSTALASI LISTRIK RUMAH TINGGAL BERBASIS CLIENT-SERVER,” *Jurnal Teknika STTKD*, vol. 3, no. 2, p. 44, 2016.

[2] A. Supriatna, Y. Yulianto, and E. Setiadi, “MESA (Teknik Mesin, Teknik Elektro, Teknik Aipil, Teknik Arsitektur) Perancangan Perangkat Lunak Rencana Anggaran Biaya (RAB) Pekerjaan Berbasis Web Application”.

[3] “Permen PUPR_1 Tahun 2022”.

[4] A. Subrata and M. Mughnyanti, “PENGGUNAAN MACHINE LEARNING UNTUK ESTIMASI BIAYA DAN WAKTU KONSTRUKSI BANGUNAN GEDUNG,” 2025. [Online]. Available: http://jurnal.goretanpena.com/index.php/JSSR

[5] O. Fuk Jin and dan Kevin Pratama Surya, “PEMODELAN ESTIMASI BIAYA KONSEPTUAL PROYEK GEDUNG SEKOLAH DENGAN MENGGUNAKAN ARTIFICIAL NEURAL NETWORK,” *Oktober*, vol. 5, no. 2, pp. 59–71, 2024.

[6] D. Puspitasari, N. Qotrun Nada, and A. Harjanta, “Penerapan Algoritma Random Forest Untuk Prediksi Biaya Kontruksi Berbasis Web,” *Jurnal Ilmiah ILKOMINFO - Ilmu Komputer & Informatika*, vol. 8, pp. 314–326, Jul. 2025, doi: 10.47324/ilkominfo.v8i2.385.

[7] F. Elghaish, M. R. Hosseini, S. Talebi, S. Abrishami, I. Martek, and M. Kagioglou, “Factors driving success of cost management practices in integrated project delivery (IPD),” *Sustainability (Switzerland)*, vol. 12, no. 22, pp. 1–14, Nov. 2020, doi: 10.3390/su12229539.

[8] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, “Data Preprocessing for Supervised Learning,” *Int. J. Comp. Sci.*, vol. 1, pp. 111–117, Jan. 2006.

[9] L. Breiman, “Random Forests,” 2001.

[10] N. Rankovic, D. Ranković, M. Ivanovic, and L. Lazic, *Recent Advances in Artificial Intelligence in Cost Estimation in Project Management*. 2024. doi: 10.1007/978-3-031-76572-8.

[11] M. S. Başar and H. Küçükönder, “Measuring the Correlation between Commercial and Economic States of Countries (B2G Relations) and the E-Government Readiness Index by Using Neural Networks,” *Open Journal of Business and Management*, vol. 02, no. 02, pp. 110–115, 2014, doi: 10.4236/ojbm.2014.22014.

[12] Streamlit Inc, “Streamlit: A Faster Way to Build and Share Data Apps.” Accessed: Jun. 30, 2026. [Online]. Available: https://streamlit.io/

[13] M. G. Meharie and N. Shaik, “Predicting highway construction costs: Comparison of the performance of random forest, neural network and support vector machine models,” *Journal of Soft Computing in Civil Engineering*, vol. 4, no. 2, pp. 103–112, 2020, doi: 10.22115/SCCE.2020.226883.1205.

[14] A. Fauzi, N. Maulidah, R. Supriyadi, H. Nalatissifa, and S. Diantika, “Prediksi Harga Properti Di Indonesia Menggunakan Algoritma Random Forest,” *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 1, pp. 43–49, Apr. 2025, doi: 10.31004/riggs.v4i1.367.