Machine Learning Based Automated Construction Planning System for SriLanka

Sathurjan K
Department of Information Technology
Sri Lanka Institute of Information Technology
Malabe, Srilanka
it21188718@my.sliit.lk

R A Ahamed
Department of Information Technology
Sri Lanka Institute of Information Technology
Malabe, Srilanka
it21158018@my.sliit.lk

Linganathan J
Department of Information Technology
Sri Lanka Institute of Information Technology
Malabe, Srilanka
it21223808@my.sliit.lk

Silva A.A.I
Department of Information Technology
Sri Lanka Institute of Information Technology
Malabe, Srilanka
it21301254@my.sliit.lk

Mr.Supunya Swarnakantha
Department of Information Technology
Sri Lanka Institute of Information Technology
Malabe, Srilanka
ravi.s@sliit.lk

Dr.Dharshana Kasthurirathna
Department of Computer Science and Information Technology
Sri Lanka Institute of Information Technology
Malabe, Srilanka
dharshana.k@my.sliit.lk

***Abstract* - Decision-making processes in multiple commercial sectors depend primarily on traditional methodologies which present both time-intensive operations together with manual processing. The proposed framework uses machine learning alongside predictive analytics to create an AI system for improving four specific operations regarding solar energy recommendations as well as real estate price prediction and construction cost estimation and interior design collaboration. Through machine learning models the Solar Recommendation System evaluates land dimensions combined with site position and environmental data to suggest suitable solar panel designs as well as forecast system output. The Land and House Price Prediction System uses past and present Colombo market data to generate property price forecasts which help developers alongside investors and buyers make data-based choices. The Cost Estimation for Construction Projects automatically generates project budgets from material expenses and labor counts and project environment changes while reducing overall financial uncertainties. The Interior Design Collaboration Platform facilitates seamless interaction between clients and designers through AI-powered recommendations and interactive visualization.**

Keywords – *Construction, Planning, Solar, Prediction, Cost, Design, Real Estate, Automation*

I. INTRODUCTION

The Sri Lankan construction sector experienced rapid development during recent years due to rising requirements for improved planning and decision platforms. Multiple processes in construction project management operate using traditional methods that foster human mistakes and exhibit insufficient use of modern technologies. This study adopts machine learning (ML) techniques to build automatic optimization solutions for construction planning elements including solar energy suggestions together with land value and house pricing forecasts and construction expense assessments and interior design groupwork capability. These domains enable improvement of sustainability performances and reduce costs while boosting communication channels across the Sri Lankan construction field.

A key element in this study requires developing the Solar Recommendation System that implements ML algorithms to suggest solar panel installations while taking land dimensions alongside weather conditions and site location into account. The system aims to assist both residential and business stakeholders in Sri Lanka toward adopting renewable power solutions which combine financial efficiency with environmental sustainability.

The Land and House Price Prediction System utilizes historical along with real-time market data to generate precise property value estimations which help developers and investors and potential buyers in their investment planning process. An ML-based system will examine Colombo real estate market data about location and land size and amenities and market trends to forecast upcoming property values in this busy Sri Lankan market [1] [2].

The Cost Estimation for Construction Projects system brings an intelligent solution to control project budgets through its features [3]. The system leverages historical datasets and real-time feed to adjust cost predictions through ML models that monitor material prices and labor costs and market changes thus minimizing budget overruns for construction projects.

The Interior Design Collaboration Platform creates connections between designers and clients through AI design recommendations for fast work coordination. Design collaboration benefits from this interactive platform because it uses an efficient process which prevents misinterpretation to develop superior design solutions

II. LITERATURE REVIEW

Construction cost estimation together with land and house price prediction and solar panel recommendation with interior design collaboration form essential decision-making components in modern business operations. Advanced operation methods are necessary to make machine learning models function at multiple efficiency targets alongside accuracy improvement.

The construction sector mainly depends on Quantity Rate Analysis together with other traditional estimating methods to evaluate project tasks through individual components [4]. The evaluation methods require large data collection efforts along with significant time investment in processing

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

procedures. By applying SVM and Random Forests together with ANN under a machine learning framework it becomes possible to achieve better prediction accuracy through analyzing big data sets for pattern recognition. The combination of SVM and ANN under machine learning methodology resulted in improved construction cost analysis according to Pandi and Nampoothiri (2019) and Samphaongoen (2010) demonstrated parametric CAD software that connects to database resources through modeling [5] [6] [7].

The process of predicting land and house prices now uses machine learning methods instead of traditional heuristic models [1] .Through evaluation of construction costs and demographics information SVR (Support Vector Regression) by Rafiei and Adeli (2016) predicted home prices. Research conducted by Koktashev et al [8]. (2019) and Sankar et al [9]. (2024) demonstrate that ensemble methods, especially Random Forests and Gradient Boosting achieve high property price prediction quality because they reveal complex variable relationships [10]. Model accuracy has received boost through the application of demographic and geographic data making these methods outperform traditional approaches in predictive reliability.

The solar panel recommendation system leverages machine learning to optimize panel selection based on environmental conditions, overcoming the limitations of traditional rule-based methods [11] [12]. Studies show that Gradient Boosting and Random Forest models enhance efficiency, scheduling, and cost reduction by adapting to real-time data [13] [14]. However, most existing systems focus on energy forecasting without integrating cost estimation, limiting practical adoption [15] [16] [17]. This study bridges the gap by developing a comprehensive recommendation and pricing system, ensuring accurate selection while considering economic feasibility for real-world applications [18].

Technical progress in interior design collaboration makes it possible for AI to transform the design process between clients and interior designers [19]. These AI platforms have replaced conventional hand-made design methods through their evaluation of client input together with spatial designs to automatically create solutions [20]. AI instruments change design processes to deliver both improved specific solutions and faster production rates according to López et al. (2020) and Fu et al. (2021) [21] [22]. These platforms enhance their recommendation framework through time due to their implementation of reinforcement learning systems that collect client engagement metrics. The proposed environmentally friendly layout features of AI offer sustainable recommendation solutions as per Chen and Zhang (2022).

III. METHODOLOGY

A Machine Learning Based Automated Construction Planning System includes four operational components which provide combined functionality for solar energy system suggestions and land-house price calculation and interior design co-creation features. The system uses machine learning to assess geographic and environmental factors, providing cost-effective solar panel recommendations. Time series forecasting combined with ensemble learning should be used to forecast land and house prices through analysis of previous data points and market condition patterns.

The system combines linear regression with Random Forest to make flexible estimates before using real-time material and labor cost updates for budgeting purposes. Within the context of interior design proposal collaboration, the system will combine NLP and computer vision to interpret client design preferences thus enabling designers and clients to create design ideas together in real-time. By combining algorithm analysis with learning capabilities, the system achieves high accuracy and enhances overall efficiency.

A. Data collection and preprocessing

The first step involves gathering a diverse and extensive dataset for each of the four components, each sourced from various platforms and databases.

For the Construction Cost Estimation System, datasets were collected from industry sources, including Kaggle, construction websites, and civil engineering reports. These datasets include historical costs, material prices, labor rates, and project sizes across various construction projects. The data was cleaned by handling missing values using statistical imputation and outlier detection methods. Categorical variables, such as project type and location, were encoded using one-hot encoding, while continuous variables like material costs and labor hours were normalized for better model performance.

For the Land and House Price Prediction, we collected the dataset from real estate platforms, Kaggle, and publicly available property records. The dataset included important features like property size, location, number of rooms, nearby amenities, and historical price data. During preprocessing, we handled missing values through imputation, identified and dealt with outliers, and applied feature scaling to continuous variables like price and size. To clean the data, we removed duplicates and irrelevant entries, ensuring everything was in top shape for model training.

The Solar Recommendation System collected its data through combined input from solar panel manufacturers and energy providers and public database records. The data sources included information about power specifications and efficiency levels and panel prices as well as different types of panels together with power limits and site information. Prior to addressing missing values and categorical variable conversion the data needed normalization procedures. The project applied feature engineering approaches that led to improved performance variables including land-area-based energy output determination.

The Interior Design Collaboration Platform obtained its datasets by collecting information from online design repositories as well as user input forms. Different datasets with information about room dimensions together with both furniture preferences and design styles and client feedback existed in the system. Publicly accessible interior design website content containing furniture catalogs and layout designs became part of the system. The processing pipeline applied lemmatization to textual data after tokenization and cleaning as well as normalization and room type and furniture specification tagging to image data. The model obtained better design style generalization through data augmentation where rotation procedures and flipping treatments and color modification techniques were integrated [23].

All preprocessing procedures produce datasets which are clear of errors and optimal for deployment in machine learning models to generate predictive and recommender outputs. All datasets use tokenization together with normalization and imputation and data augmentation to maintain dataset consistency along with improved model performance thus creating a robust system for real-world applications.

B. *Model selection*

During this stage machine learning along with deep learning models were used to tackle the specific issues in predicting construction costs as well as land and house prices alongside solar panel suggestions and interior design choices. Different models were selected because they matched the requirements of each component part. The Construction Cost Estimation System implemented Linear Regression because it established simple connections between project material expenses along with labor hours to generate final construction prices. Decision Tree Regressor brought effectiveness in detecting complicated cost data interactions thus delivering precise predictions when cost elements exhibit complex patterns

The Linger regression model serves as a choice for dealing with extensive large construction project datasets because traditional processing systems lack sufficient power to detect hidden cost variables dependencies. The predictive models underwent extra regularization adjustments to enhance predictions while avoiding overfitting inaccuracy.

The Land and House Price Prediction System chose Gradient Boosting Regressor for its main predictive model because it showed outstanding accuracy while evaluating properties using area features such as location properties and size information with surrounding amenity factors. The Random Forest Regressor system united efficient big data analysis with strong overfitting resistance to perform land and house market predictions across multiple regions. The baseline model for users who needed to study basic linear price patterns was Linear Regression.

Random Forest Regressor served as the recommendation tool for most efficient solar panels because it optimizes non-linear prediction along with panel-feature-location-environment connections. Solar panel efficiency estimation relied on Gradient Boosting Regressor because this model excels at sequential prediction improvements as well as low error reduction. Linear Regression performed price estimation because it effectively detects linear price-to-specification relations.

The design application utilizes Gated Recurrent Unit (GRU) from the Recurrent Neural Network (RNN) family to process the written information provided by customers. The text data processing capabilities of GRUs enable them to understand extended dependencies while handling sequences of different sizes which makes this model suitable for design preference comprehension and recommendation creation. Xception architecture served as the Convolutional Neural Network (CNN) approach for image-based content examination through depth-wise separable convolutions which maintain high deep learning capability. The system utilized this robust design because it could effectively extract and suggest elements in images by applying transfer learning from pre-trained models that analyzed ImageNet database information.

C. *Model Training*

Our methodology allows machine learning models to undergo precise training using excerpts from preprocessed data during the training process to achieve optimal accuracy when forecasting construction prices together with solar panel efficiency and property values and interior design advice.

The model training process for the Construction Cost Estimation System includes the implementation of Neural Network (MLP) together with Decision Tree Regressor and Linear Regression models. The data allocation divides the set into 80% training data and 20% validation data. Training periods depend on three factors: the size of dataset and model intricacy and system capabilities under which training occurs. Concentrated training scenarios result from modest dataset volumes but MLP along with extensive datasets demand thorough training duration. A learning rate of 0.001 defines the Adam optimizer to optimize the loss function while mean squared error operates as the main evaluation metric.

The first phase in training the Land and House Price Prediction System includes dividing preprocessed data into 80-20 training and validation subsets. Training of the Gradient Boosting Regressor and Random Forest Regressor as well as Linear Regression occurs on this data. The Adam optimizer serves to optimize the models while evaluation depends on R^2 together with mean absolute error (MAE). Such models learn to detect patterns found within multiple features of properties including their place of residence along with their dimensions and supplemental features. The validation data exists to help the development team readjust their models during the overfitting prevention process.

In the Solar Recommendation System training consists of Random Forest Regressor and Gradient Boosting Regressor together with Linear Regression models. Performance is measured through RMSE and R^2 as the data sets into training and validation parts during processing. The training duration for these models relies on the combination of dataset dimensions alongside panel specifications complexity. Adam optimizer runs at 0.001 learning rate for the optimization process. The recommendation system teaches these models to identify the top solar panels using criteria consisting of power output, efficiency and pricing.

When users attend training at the Interior Design Collaboration Platform they must establish two separate data sets for training and validation. Client preference texts undergo training through GRU which belongs to the RNN category alongside Exception-CNN that operates on image-based design inputs. Architectural patterns get recognized through training both systems with text and visual data. Performance evaluation uses accuracy together with F1-score metrics whereas cross-entropy loss functions serve as the loss function. The CNN benefits from pre-trained ImageNet models utilized for pattern recognition enhancement. The Adam optimizer adjusts weights for GRU so this model optimizes its capacity to process text-based preferences. Several epochs of training take place until validation performance reaches its plateau thus requiring early stopping to protect against overfitting. The trained model weights receive storage for subsequent deployment in real-world applications because of their established method of archiving.

D. Model testing and evaluation

The analysis of constructed cost predictions against real values depends on Mean Absolute Error and Root Mean Squared Error in combination with R^2 for Linear Regression, Decision Tree Regressor and Neural Network (MLP). The extension of feature engineering techniques leads to enhanced model precision because it includes all essential cost aspects from both materials expenditure and manpower requirements. The combination between grid search and random search optimizes hyperparameters and the cross-validation method establishes generalization capability for new data points. The mechanism of regularization works as an automatic defense system which protects model accuracy during multiple project evaluations. The chosen predictive model serves for construction cost estimation because its evaluations show accurate predictive capacity. The real-life operational version of the selected model will perform as intended.

The Land and House Price Prediction System maintains Gradient Boosting Regressor and Random Forest Regressor and Linear Regression as predictive models for analyzing predictive accuracy through R^2 and Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE). The models integrated properties' location features with their dimensions and service features during data processing for training and testing. The models acquire market functionality across various conditions because of cross-validation which enables them to predict different property types. Data augmentation techniques in the training set use artificial property information generation for improving generalization capacity. Optimizing real estate market predictions across different conditions yields enhanced accuracy levels through parameter adjustments by hyperparameter search methods.

The solar panel recommendation system was evaluated using ensemble learning and multiple regression models, relying on RMSE and R^2 scores for accuracy. The Random Forest Regressor achieved perfect panel count prediction (RMSE 0.0000), while the Gradient Boosting Regressor ensured precise efficiency estimates (RMSE 2.38). The Linear Regression model demonstrated strong cost estimation accuracy (R^2 = 0.98). Test results and prediction graphs (Figure 3:Scatterplot for Actual vs Predicted Number of Panels,Figure 4:Scatterplot for Actual vs Predicted Panelsize,) confirm the models' accuracy by comparing actual and predicted values. Clustering analysis further categorizes solar panels based on efficiency and cost, aiding decision-making. The model was fully validated for real-world solar system implementation.

Evaluation for the Interior Design Collaboration Platform entails GRU model and Exception architecture testing for processing text-based input and image-based content respectively. The evaluation metrics for performance assessment include accuracy together with F1-score and mean squared error (MSE). The GRU text model alongside the Exception image model receives training from labeled datasets through which they generate precise design recommendations. Cross-validation Testing allows the models to function effectively with multiple designs formats. The augmentation of training data occurs through image-based operations that include rotation and color transformation methods to boost data variety.

User preferences become accurately captured through both models following data testing with real-world design examples to deliver proper interior recommendations. Testing operational models using different assessment approaches and validation methods as well as data augmentation procedures ensures that hyperparameter optimization brings increased operational efficiency. A model testing phase measures both its ability to use unfamiliar data points and its deployment consistency across different platforms.

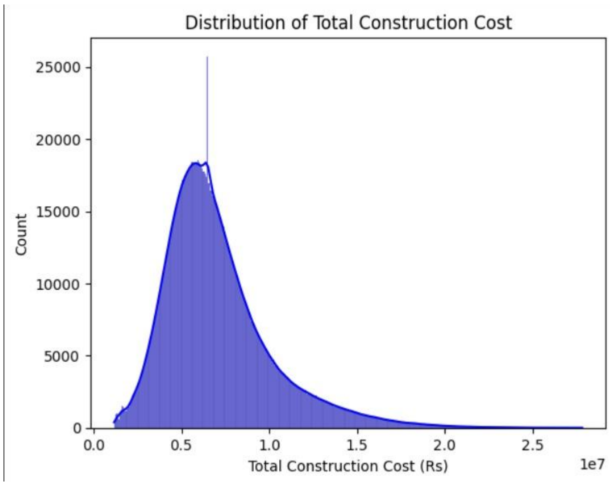

Figure 1:Hisplot for construction cost price calculation

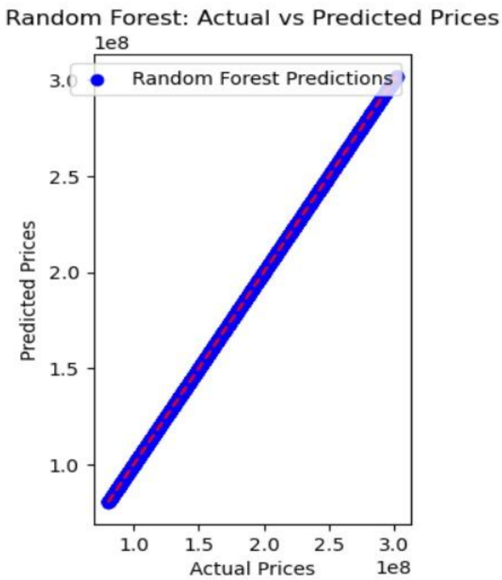

Figure 2:Matplotlib for house price calculation

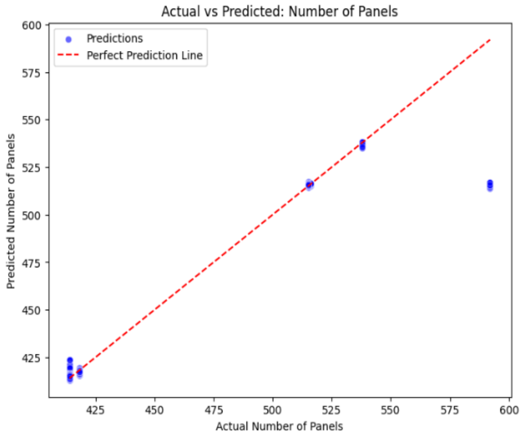

Figure 3: Scatterplot for Actual vs Predicted Number of Panels

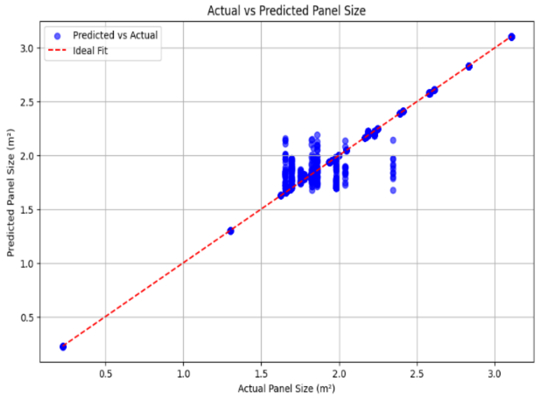

Figure 4:Scatterplot for Actual vs Predicted Panel Size

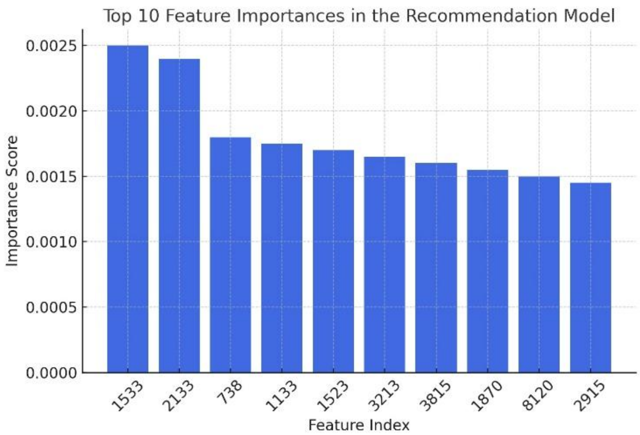

Figure 5:Feature importance graph for image analysis

E. *Integration of the models*

In the integration phase, we bring together the machine learning models from the Construction Cost Estimation System, Land and House Price Prediction System, Solar Recommendation System, and Interior Design Collaboration Platform into one cohesive system. By combining data from all these components, the system provides smooth and uninterrupted predictive services.

The system develops dedicated models using preprocessed data for every task domain. Each model applies output standardization which standardizes data formats so operators can aggregate and make decisions without difficulty. A content routing mechanism links up with an integration framework to facilitate quick processing that supports cost estimation along with price prediction and solar recommendations as well as interior design suggestions.

The real-time detection process benefits from parallel processing methods. The system operates more accurately through the inclusion of feedback loops and output aggregation together with decision logic. The system operation benefits from user privacy maintenance and bias handling through ethical decision-making processes that establish transparency and trustworthiness.

IV. Results

The project for Construction Cost Estimation and Land and House Price Prediction and Solar Recommendation and Interior Design Collaboration progressed through two fundamental stages. During model training, we established four distinct specialized models that targeted operational needs. All models received carefully cleaned data to reach a high accuracy standard for their predictive functions. The main system objective involved striving for accuracy because the system delivered superior results compared to existing approaches.

The second phase of our work involved building crucial elements that included output standardization procedures and modular frameworks together with content routing systems, parallel processing capabilities, and result aggregation functions. All components of the system work together seamlessly to ensure smooth operations and efficient processing while also providing flexibility to scale and adapt to future needs with real-time forecasting capabilities.

V. Conclusion

The development of a system that combines Construction Cost Estimation with Land and House Price Prediction and Solar Recommendations and Interior Design Collaboration stands as the main achievement of our research. Our system uses machine learning to create an advanced tool that accurately estimates construction costs and property values while also assessing solar power output and providing personalized design suggestions. Based on our research, these machine learning techniques have delivered strong and reliable results. Moving forward, we aim to improve model accuracy, expand our dataset, and refine the system’s ability to adapt to changing market conditions

VI. References

[1] M.shankar, "Analysis of Landlord's Land Price Prediction using Machine Learning," in *IEEE*.

[2] M. Ma, Y. Zhang, and X. Xu, " Urban-scale land value prediction using non-linear machine learning approaches," *Journal of Urban Planning and Development*, vol. vol. 146, no. no. 4, 2020.

[3] Zhang, X., Li, Q., & Wang, J., "IoT-Driven Construction Cost Estimation Using Machine Learning: A Real-Time Approach.," *Automation in Construction,* p. 112, 2020.

[4] Kim, J., Lee, H., & Park, S., "Application of Artificial Neural Networks in Construction Cost Estimation: A Case Study of High-Rise Building Projects.," *Journal of Construction Engineering and Management,*, 2019.

[5] Son, H., & Kim, C. , "Hybrid Principal Component Analysis and Support Vector Machine Model forPredicting the Cost Performance of Commercial Building Projects Using Pre-Project Planning Variables.," *Automation in Construction,* 2010.

[6] Kim, G. H., An, S. H., & Kang, K. I., "Comparison of Construction Cost Estimating Models Based on Regression Analysis, Neural Networks, and Case-Based Reasoning.," *Building and Environment,* 2004.

[7] McMahon, S., & Smith, R., " Enhancing Construction Cost Estimation Accuracy Through Machine Learning and Big Data Integration.," *Journal of Construction Research,* 2018.

[8] S. Rafiei and M. Adeli, "Predicting home prices using Support Vector Regression," *Computational Intelligence and Neuroscience,* 2016.

[9] A. Koktashev, I. K. Koc, and H. A. Aslantas, "A comparative study on housing price prediction," *Applied Soft Computing Using machine learning algorithms* , 2019.

[10] M. Park and H. Bae, "Real estate price prediction using machine learning algorithms," *Computers, Environment and Urban Systems,* vol. 49, pp. 33-44, 2015.

s[11] T. Huld, R. Müller, and A. Gambardella, "A new solar radiation database for estimating PV performance in Europe and Africa," *Solar Energy,* vol. 86, pp. 1803-1815, 2012.

[12] B. M. Kumar and K. Sudhakar , "Performance analysis of 10 MW grid connected solar photovoltaic power plant in India," *Energy Reports,* pp. 184-192, 2015.

[13] K. R. Khyati, A. K. Tripathi, and V. Khare, "Predictive modeling of solar photovoltaic efficiency using machine learning techniques," *IEEE,* vol. 9.

[14] M. Eltawil and Z. Zhao, "Grid-connected photovoltaic power systems: Technical and potential problems—A review," *Renewable and Sustainable Energy Reviews,* vol. 14, 2010.

[15] P. F. Katiraei, P. L. Ping, and W. Li, "Machine learning-based prediction models for solar panel performance evaluation," *in Proc. IEEE Int. Conf. Smart Grid Communications (SmartGridComm,* pp. 162-169, 2020.

[16] S. Das, A. Dutta, and A. B. Mandal, "Cost-effective solar power system design using predictive analytics," *IEEE Transactions on Sustainable Energy,* vol. 12, pp. 2942-2951, 2021.

[17] J. Zhang, H. Guo, and T. Zhang, "Short-term solar power forecasting based on deep learning," *IEEE Transactions on Smart Grid,* vol. 11, pp. 2816-2825, 2020.

[18] L. Xu, X. Wang, and C. Liu, "Optimal sizing and economic analysis of a photovoltaic-battery system considering time-of-use pricing," *Applied Energy ,* vol. 280, 2020.

[19] Y. . Z. D. T. a. J. W. Zhe Xu, "Architectural Style Classification Using Multinomial Latent Logistic Regression," 2014. [Online]. Available: https://www.researchgate.net/publication.

[20] M. . S. M. R. E. a. T. H. K. Nima Norouzi, "The Architect, the Client and Effective Communication in Architectural Design Practice," 2015.

[21] M. . S. M. R. B. E. a. T. H. K. Nima Norouzi, "The Architect, the Client and Effective Communication in Architectural Design Practice," 2015.

[22] S. . I. M. H. W. a. W. N. M. W. M. R. Hala Taleb, "Communication management between architects and clients," 2017.

[23] "UMITRUX, Architectural Styles," 2020. [Online]. Available: https://www.kaggle.com/datasets/dumitrux/architectural-styles-dataset..