Proceedings of the International Conference on Sustainable Communication Networks and Application (ICSCNA-2024)
IEEE Xplore Part Number: CFP24DW8-ART; ISBN: 979-8-3315-3001-3

Optimizing Building Energy Consumption using Smart Grid Data

Viduran Neelakandan
Department of Networking and Communications
School of Computing
College of Engineering and Technology
SRM Institute of Science and Technology
Kattankulathur-603203, India
vn3046@srmist.edu.in

Sai Suraj Voleti
Department of Networking and Communications
School of Computing
College of Engineering and Technology
SRM Institute of Science and Technology
Kattankulathur-603203, India
sv2237@srmist.edu.in

Dhanush Srinivasan
Department of Networking and Communications
School of Computing
College of Engineering and Technology
SRM Institute of Science and Technology
Kattankulathur-603203, India
ds7068@srmist.edu.in

Gouthaman. P*
Department of Networking and Communications
School of Computing
Faculty of Engineering and Technology
SRM Institute of Science and Technology
Kattankulathur-603203, India
gouthamp@srmist.edu.in

2024 International Conference on Sustainable Communication Networks and Application (ICSCNA) | 979-8-3315-3001-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICSCNA63714.2024.10863865

***Abstract*—This research discusses how data analytics techniques can be applied in the enhancement of building energy management. It is therefore with the purpose of stimulating the right utilization of the practically inexhaustible quantity of data obtained through smart grids that this research seeks to design original approaches to energy management, building performance, and counteracting of negative impacts on the environment. Since Apache spark process is very efficient we do an excellent job in pre-processing and analyzing the data collected. Combining high-level artificial intelligence decision making tools like linear regression as well as the random forest, the relationship was established between energy utilization and determinants, forecast patterns and create accurate prognosis models. In order to present the results in a format that can be easily understood with the help of Tableau to build a data visualization where depending on the values chosen by the stakeholders further steps can be taken in achieving a more efficient use of energy and less expenditure. In our work, we strive to generate knowledge that will support improved energy efficiency of the built environment, and therefore a more sustainable future. The implications of our work have the ability to affect the processes within and energy usage of buildings, saving energies and making cities greener. Furthermore, our work can help with the creation of further reference points for future research in the field of building energy optimization.**

***Keywords*—Big Data, Smart Grids, Energy Optimization, Machine Learning, Predictive Analytics, Building Efficiency, Sustainability**

I. INTRODUCTION

The intention to reduce energy costs and conserve the environment has boosted a lot of research in building energy use. Generally in building energy management, a conventional approach has been applied which mainly involved some techniques and limited data, hence operating at the lowest level possible. Over the last few decades, the smart grid infrastructure has given a lot of data that can be used to improve the building energy efficiency [1], [2].

This research work focuses on the use of Big data analysis methods to enhance energy consumption of buildings. Our project involves using the tremendous amount of data produced by smart grids in order to identify novel approaches to decreasing energy consumption, enhancing the functionality of our buildings, and ultimately minimizing the effects of these actions on the environment [3][4]. The proposed integration of data analytics in building energy management is a potential area for attaining high energy saving as well as environmentally sustainable city planning [5].

The research objectives and the method of the study are laid down in this section. The objectives of this research are to: They include: (1) building the quantitative model that enables the utilization of energy data to achieve efficient consumption rates, (2) establishing a quantifiable approach identifying determinants of energy consumption, (3) estimating future energy consumption rates, and (4) assessing the efficiency of implemented optimization methods[6]. The approach includes data gathering and cleansing from smart grids, employing the Analytics of Things techniques and constructing prognosis models.

A literature survey is performed to establish an understanding of the current and past studies in the area of BEOP and big data processing [7], [8]. The scope of the paper is to show the state of the art for the current topic and give indications of possible research directions. After that, data collection and preprocessing take place, which consist in obtaining data related to smart grids and their further quality and coherence assessment.

The randomness and fluctuations in energy consumption characteristics have emerged as major challenges of constructing an efficient building energy management system. Indications like building occupancy, climate, and use of appliances have a considerable impact on energy requirement. With data analytical tools, it is possible to determine these patterns and, therefore, come up with better approaches that would solve the exact and optimum usage of energy.

In addition, coupling of a data analysis solution with the building automation conceptual systems, it would be possible to regulate energy usage in real time basis with appropriate manageability. This can enhance the improvement of building systems operation and prevents energy wastage [9]. Further, the data is highly effective in determining where energy-saving upgrades and retrofit are needed, for instance efficient equipment or enhancing insulation.

II. LITERATURE SURVEY

The author [10] designs a federated learning paradigm with blockchain as an addition to carry out the training of the models in a collaborative manner by using a collective of distributed datasets while ensuring the sensitive data remain private to the owners. In this paper, certain aspects of the energy consumption of the proposed framework are described and discussed with respect to the energy consumption of traditional centralized solutions. Performance analysis proves the efficiency of the proposed federated learning technique along with blockchain as a significant contributor to energy efficiency.

The work [11] suggests an integrated approach of power demand forecast for large scale building power demands. The method uses energy simulations based on EnergyPlus physics and a generative adversarial network. Performance evaluation of the research presented herein shows that the proposed method successfully aids in the determination of building power demands and saves computational time.

A deep learning model suggested [12] for the prediction of electric energy consumption especially in the office buildings. The model contrasts the Deep Neural Network (DNN) with Support Vector Regression (SVR) and Random Forest (RF) and shows that DNN works more accurately.

The proposed work [13] designs an effective scheduling strategy of supply and battery usage in residential buildings for the grid system. The proposed algorithm integrates Q Learning and Fuzzy Logic Control for energy management. The experimental findings show that the suggested algorithm can minimize electricity costs and increase the battery's lifespan.

The author [14] suggests a part-b Newsletter model to improve energy efficiency of smart grids. The model is composed of temporal convolutional networks (TCN), bidirectional gated recurrent units (BIGRU), and an attention gate. Experimental outcome shows that implementation of the proposed model facilitates energy efficiency and optimal grid management.

The proposed framework builds upon these features of big data storage, processing, and analysis; among the components are HBase for storage, Apache Spark for processing with machine learning techniques and visualization tools. This offers a realistic working system for managing the energy consumption of buildings and it can handle historical big data and real building data for more effective building energy optimizing for all sectors influenced by building energy situations.

III. PROPOSED WORK

The system architecture comprises the proposed system that is intended to process smart grid data on energy consumption in buildings as envisaged. The following elements are present in the system architecture:

Data Ingestion

The first step of the process is data ingestion where a lot of data from a myriad of sources are then fed into Apache Spark. This can include, smart meters, weather sensors, or building automation systems to include data gathering from the indicated sources. The data collected is then loaded into Spark using correct methods of data ingestion, while following the right standards in data quality and quantity.

Data Storage

After that, the data gets accumulated in an appropriate data warehousing system. In this case, it is possible to make use of the Apache Spark to store the data off of the distributed memory. In-memory data organization at Spark prevents frequent consultations of external storage facilities to access data and analyze queries. But for handling large amounts of data or data logged into the long term, the data can be stored using a distributed file system such as Hadoop Distributed File System (HDFS).

Data Processing

Data processing then entails acquisition and the grooming which incorporates cleaning, transformation and preparing of the ingested data for analysis and warehousing. part of the preprocessing step is how to deal with missing values, outliers and any inconsistency that may be present in the dataset we want to use; feature scaling and feature normalization is also part of this process of preprocessing. Another example is that while developing features, new features may be developed or the existing features may be transformed in order enhance the efficiency of the model. This information is the

preprocessed data ready for more refined analysis such as through the use of Machine Learning algorithms.

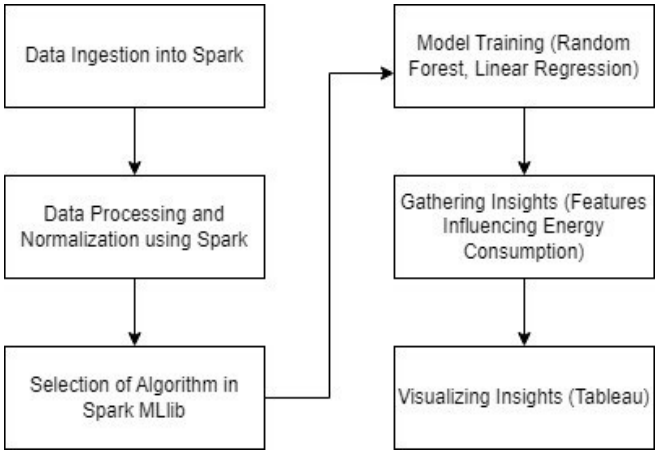

Fig. 1. Workflow diagram

Step 1: Data Ingestion into Spark

Data ingestion are the first step where the data required is collected and generated and loaded to Apache Spark framework. This involves determining the data that is available, what other data source would be able to contribute to the forecasting function and collecting data from these sources like smart meters, weather sensors and Building automation systems. The data collected is then processed and brought into Spark using proper data ingestion methods to check the overall quality of the data.

Step 2: Data Processing and Normalization using Spark

After data ingestion, data is preprocessed and normalized to a level that is amenable to analysis. Here, geometric means are calculated and missing values and other inconsistencies are addressed. The data is then converted into a suitable format for analysis including developing time series or numerical features.

Normalization is a crucial step in data preprocessing that brings all numerical features to the same scale, preventing skewed or risky comparisons. This is particularly important when features have vastly different ranges or units. Normalization techniques such as min-max scaling, standardization, and robust scaling are commonly used. The choice of normalization technique depends on the specific characteristics of the data and the requirements of the machine learning algorithm. Careful normalization can significantly improve the performance and stability of the model.

Step 3: Selection of Algorithm in Spark MLlib

The Specific selection of the machine learning evaluation model is important in order to make correct predictions and drive the insights. In this step, as will be described, suitable algorithms are applied depending on the type of data and research hypothesis or purpose. Issues like the type of data (for instance time series, regression) the degree of model intricacy, the available computation power are taken into account. Such algorithms may include the linear regression, the random forest and the time series. Closer to the selection of an algorithm to solve a given problem, additional tuning of its hyperparameters is conducted for better performance.

Step 4: Model Training (Random Forest, Linear Regression)

The selected algorithm is then trained on the processed data for developing the models of prediction. Again, it entails introducing the data into the algorithm so that working goes on to identify relationship patterns. Training in neural networks is the process of consistently using an iterative approach to impact the model's parameters with an eye on minimizing the gap between the prediction and the actual value. The trained models allow to forecasting energy consumption and define the main factors which affect it.

Step 5: Gathering Insights (Features Influencing Energy Consumption)

Once the models are fitted, inference is made on the models with a view of identifying factors that contribute to energy consumption. These include the identification of the feature importance scores, which tells of relative importance of a given feature concerning the model under analysis. In this case, the topological features define the core variables that require attention for driving energy efficiency and suggest approaches for optimization.

Step 6: Visualizing Insights (Tableau)

To share the results, the data visualizations are made using Tableau. This step involves pre-processing the data, collection of key values and then choosing the right figures to represent on a chart. Through the use of the dashboards and an interaction interface, the various patterns, trends and correlation that need to be presented can be easily deciphered by the various stakeholders.

1. Data Processing: Spark for Batch Analysis

We utilize Apache Spark for batch processing of building energy data. Spark's in-memory processing capabilities significantly improve performance compared to traditional batch processing frameworks.

2. Batch Processing

Spark is well-suited for analyzing large datasets stored in HDFS. We can use Spark to perform various data analysis tasks, such as aggregating data, calculating statistics, and applying machine learning algorithms. Spark's distributed processing capabilities allow us to handle large-scale data analysis efficiently.

IV. DATA SOURCE AND METHODOLOGY

A. Data Source

The dataset employed in this research was sourced from Kaggle, an open-access platform. The specific dataset utilized, "Bangalore area smart grid data," was

selected due to its comprehensive representation of building energy consumption variables, including temperature, humidity, occupancy, and equipment usage. Additionally, the dataset incorporates controllable variables such as HVAC and renewable energy consumption. This dataset, specifically "building_energy_consumption_datasets.csv," comprises 365 daily energy consumption records for a building, providing valuable insights into energy consumption patterns and enabling the identification of potential energy-saving strategies.

B. Data Analysis and Visualization:

To analyse the features influencing energy consumption, a combination of statistical analysis and machine learning techniques was employed. Descriptive statistics, correlation analysis, and hypothesis testing were used to understand data distributions and relationships between variables like temperature, occupancy, and HVAC usage. Machine learning models, such as Random Forest, were utilized to determine the most influential features for predicting HVAC status. The computational complexity of the proposed model is influenced by the use of Random Forest and the size of the dataset. While the dataset size is moderate, the number of trees and their depth in the Random Forest model can impact computational cost. However, the use of Apache Spark for data processing significantly reduces the overall computational time. By combining these techniques and addressing computational considerations, a comprehensive understanding of energy consumption patterns was gained, enabling the development of effective optimization strategies.

V. IMPLEMENTATION

A. Data Ingestion and Preprocessing

The first stage of the project involved data extraction and cleaning, which were key to the second and subsequent state of analysis and modelling. The information was obtained from Kaggle and transferred to the programming tool, Python, with the help of the Pandas data analysis system. Preprocessing was an important part on the data quality since it involved operations like data cleaning. All the data management required some measures to be taken in order to handle the missing values in the data set: means and/or median imputation. Any cases that fell outside the normal range within the specified data set were then dealt with employing statistical tools or else visualizations to ensure they did not distort the analysis. Feature preprocessing was used in two forms: feature construction wherein new features were created from the existing features to create better features or feature transformation wherein the existing features were modified in order to enhance their ability to improve model performance. For instance, new features were developed to capture relations of simultaneous occurrence, such as the temperature and the relative humidity, or the density of occupation per area unit. These are features that augmented more information to the model and therefore improving the predictive ability of the model

TABLE I. MACHINE LEARNING ALGORITHM ACCURACY

Machine Learning Algorithm	**Accuracy**
Random Forest Algorithm	93.408%
Logistic Regression	86.032%
Support Vector Machine (SVM)	83.732%

B. Model Training and Evaluation

After data preparation, relevant machine learning models like Random Forest were trained and evaluated on a portion of the data. Feature importance analysis was conducted to identify the most significant factors influencing energy consumption. The model's performance was assessed using metrics such as accuracy, precision, recall, and F1-score. By analyzing the model's predictions and the importance of features, valuable insights were gained into energy consumption patterns, enabling the identification of potential energy-saving strategies. A Random Forest algorithm was chosen due to its ability to handle both numerical and categorical data, as well as its robustness to overfitting. The set of models was calibrated on a part of the data, which is called the training data, whereas the rest of the data is used for testing.

1. Visualization

In order to convey the message behind the discovery, a depiction was made using Tableau and Python Matplotlib. The knowledge obtained from the models were explained in simple manners with the use of charts and graphs. This was because the energy consumption patterns in the buildings could be broken down and clearly explained to the other stakeholders, along with an assessment of possible savings.
Key visualizations included: Scatter plots: In order to plot energy consumption against influences, such as temperature or occupancy. (Fig 2)

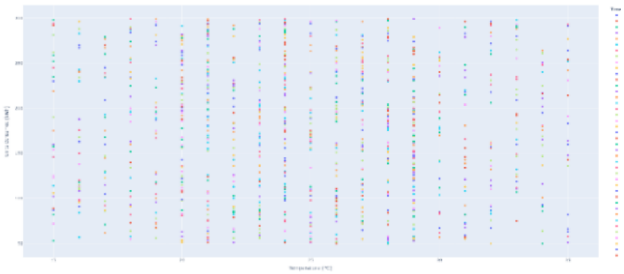

Fig. 2. Scatter Plot of Units Consumed vs Temperature

Line charts: In the study of energy consumption with respect to time. Bar charts - To evaluate the energy use of buildings of varying types, building occupancies, or periods of time. (Fig 3)

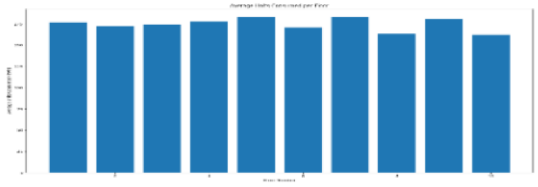

Fig 3. Bar Chart of Units Consumed vs Floor Number

Heatmaps: In order to help visualize the relationships between different features and check for issues of multicollinearity (Fig. 4)

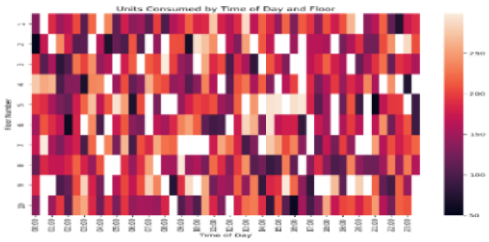

Fig. 4. Heatmap of Units Consumed in Floor Number and Time of day

2. Additional Considerations

Time Series Analysis: If the data is time-series based, special procedure like the ARIMA or SARIMA models are used to model temporal dependencies adequately for accurate forecasts. Ensemble Methods: Ensemble of two or more models such as random forest, gradient boosting often leads to better accuracy and also is more resistant.

3. Model Deployment and Monitoring

After that, the models were deployed into production as real time or near real time prognosis integrated into production platform. This meant that the models had to be plugged into building automated systems or an energy managing system.

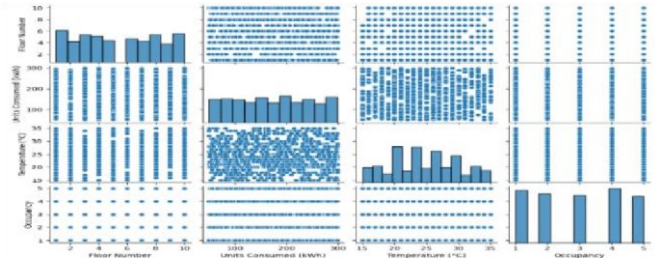

Fig. 5. Consolidated Charts for all Features

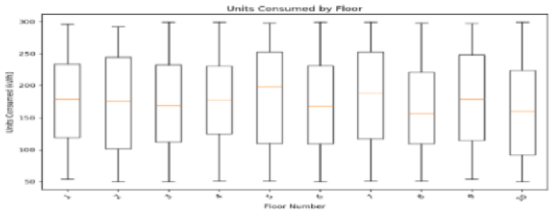

Fig. 6. Box Plot of Units Consumed by Floor

These included monitoring the experiential test results of the model, that is, a comparison of the level of energy consumption as proposed in the model to that perceived in the actual environment, as well as the degree of agreement in recommendations offered by the model.

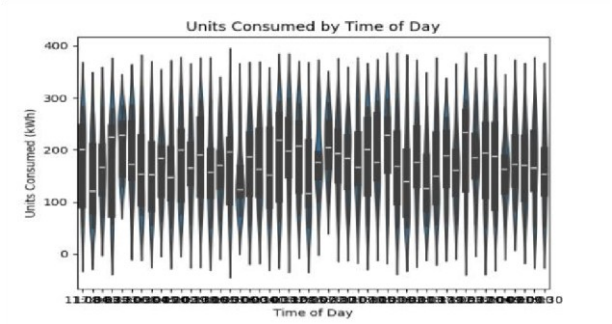

Fig. 7. Violin Plot of Units Consumed by Time of Day

4. Case Study: Application of Building Energy Optimization

In order to explicate this practical application of this methodology, we can use an example of a case study. An example may refer to a medium or big commercial facility that on purpose decided to minimize energy use. By using the indicated approach, the building manager can define factors which affect energy consumption, design effective optimisation measures and control the effectiveness of implemented actions.

5. Challenges and Future Directions

This research presents a model for constructing energy optimizations. However, challenges related to data accessibility, quality, and the temporal and spatial randomness of energy consumption in buildings persist. Future research should explore advanced techniques, such as deep learning and reinforcement learning, to achieve high precision in energy consumption predictions. By improving data quality, employing robust data preprocessing, and considering factors like building occupancy, climate, and equipment usage, we can enhance model performance and contribute to sustainable energy management.

VI. CONCLUSION

The analysis revealed key factors influencing HVAC usage, including floor number, time of day, and units consumed. These findings underscore the importance of building design, occupancy patterns, and external conditions for optimizing HVAC systems. Machine learning models, such as Random Forest, demonstrated promising performance in predicting HVAC status, providing valuable insights for energy management strategies. To further enhance performance, future research could explore advanced techniques, robust data preprocessing, and the integration of additional factors. By leveraging techniques like feature engineering, hyperparameter tuning, and ensemble methods, we can improve model accuracy

and generalization. Additionally, addressing data quality issues, sensor inaccuracies, and dynamic changes in building conditions is crucial for real-world applicability. Incorporating renewable energy sources and promoting energy-conscious behavior can further enhance energy efficiency and sustainability.

REFERENCES

[1] "Hybrid_method_for_building_energy_consumption_prediction_based_on_limited_data".

[2] A. Almalaq and J. J. Zhang, "Evolutionary Deep Learning-Based Energy Consumption Prediction for Buildings," *IEEE Access*, vol. 7, pp. 1520–1531, 2019, doi: 10.1109/ACCESS.2018.2887023.

[3] L. Yu, S. Qin, M. Zhang, C. Shen, T. Jiang, and X. Guan, "A Review of Deep Reinforcement Learning for Smart Building Energy Management," Aug. 01, 2021, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/JIOT.2021.3078462.

[4] D. Kolokotsa, "The role of smart grids in the building sector," Mar. 15, 2016, *Elsevier Ltd*. doi: 10.1016/j.enbuild.2015.12.033.

[5] C. Li, Z. Ding, J. Yi, Y. Lv, and G. Zhang, "Deep belief network based hybrid model for building energy consumption prediction," *Energies (Basel)*, vol. 11, no. 1, 2018, doi: 10.3390/en11010242.

[6] L. Tang, H. Xie, X. Wang, and Z. Bie, "Privacy-preserving knowledge sharing for few-shot building energy prediction: A federated learning approach," *Appl Energy*, vol. 337, May 2023, doi: 10.1016/j.apenergy.2023.120860.

[7] E. Mocanu *et al.*, "On-Line Building Energy Optimization Using Deep Reinforcement Learning," *IEEE Trans Smart Grid*, vol. 10, no. 4, pp. 3698–3708, Jul. 2019, doi: 10.1109/TSG.2018.2834219.

[8] K. Dai, F. Ji, L. Bai, and H. Li, "Research on Building Energy Consumption Prediction based on BP Neural Network," in *2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications, AEECA 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 476–480. doi: 10.1109/AEECA55500.2022.9918918.

[9] F. Wurtz and B. Delinchant, "'Smart buildings' integrated in 'smart grids': A key challenge for the energy transition by using physical models and optimization with a 'human-in-the-loop' approach," Sep. 01, 2017, *Elsevier Masson s.r.l.* doi: 10.1016/j.crhy.2017.09.007.

[10] N. Romandini, C. Mazzocca, and R. Montanari, "Federated Learning Meets Blockchain: a Power Consumption Case Study," in *Proceedings - 2023 31st Euromicro International Conference on Parallel, Distributed and Network-Based Processing, PDP 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 206–211. doi: 10.1109/PDP59025.2023.00040.

[11] C. Tian, Y. Ye, Y. Lou, W. Zuo, G. Zhang, and C. Li, "Daily power demand prediction for buildings at a large scale using a hybrid of physics-based model and generative adversarial network," *Build Simul*, vol. 15, no. 9, pp. 1685–1701, Sep. 2022, doi: 10.1007/s12273-022-0887-y.

[12] R. Panigrahi, N. R. Patne, S. Pemmada, and A. D. Manchalwar, "Prediction of Electric Energy Consumption for Demand Response using Deep Learning," in *2022 International Conference on Intelligent Controller and Computing for Smart Power, ICICCSP 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICICCSP53532.2022.9862353.

[13] A. Selim, H. Mo, and H. Pota, "Optimal Scheduling of Grid Supply and Batteries Operation in Residential Building: Rules and Learning Approaches," in *SCEMS 2022 - 2022 IEEE 5th Student Conference on Electric Machines and Systems*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/SCEMS56272.2022.9990764.

[14] F. Hamayat, Z. Akram, and S. Zubair, "Deep Learning-based Predictive Modeling of Building Energy Usage," in *2023 6th International Conference on Energy Conservation and Efficiency, ICECE 2023 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICECE58062.2023.10092502.

[15] D. Durairaj, Ł. Wróblewski, A. Sheela, A. Hariharasudan, and M. Urbański, "Random forest based power sustainability and cost optimization in smart grid," *Production Engineering Archives*, vol. 28, no. 1, pp. 82–92, Mar. 2022, doi: 10.30657/pea.2022.28.10.