



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Consumption Prediction using Machine Learning

K.Upendra, K.Madhu Sudhan Reddy, T.Shakila

B. Tech Student, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai,
Tamil Nadu, India

B. Tech Student, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai,
Tamil Nadu, India

Assistant Professor, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai,
Tamil Nadu, India

ABSTRACT: Energy consumption has become a major concern in the modern world due to the increasing demand for electricity and the growing need for sustainable energy management. Accurately predicting energy usage patterns is crucial for utility providers, policymakers, and industries to optimize energy production, reduce operational costs, and ensure energy efficiency. This project focuses on building a machine learning model to predict future energy consumption based on historical data. By analyzing features such as time of day, weather conditions, temperature, humidity, appliance usage, and occupancy levels, the model can identify consumption trends and provide accurate forecasts. These predictions can aid in better load balancing, demand response planning, and energy conservation strategies. The project follows a structured pipeline including data preprocessing, feature engineering, model selection, and evaluation. Various regression algorithms such as Linear Regression, Random Forest, and XGBoost were implemented and compared to find the most effective model. Performance metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score were used to assess the accuracy of the models. The final model demonstrated strong predictive capabilities and can be deployed as a decision-support tool for smart grid systems and energy management platforms. This machine learning approach not only automates energy forecasting but also contributes to the broader goal of creating intelligent, data-driven systems for a more sustainable future.

I. INTRODUCTION

Wildfires In recent years, the global demand for energy has significantly increased due to population growth, urbanization, industrial development, and the widespread use of electronic appliances and smart devices. With this rising demand, the need to manage energy resources efficiently has become more critical than ever. Energy consumption patterns are becoming increasingly complex, and traditional methods of prediction are no longer sufficient to handle the vast amount of data and varying parameters involved. This has created a pressing need for advanced technologies and intelligent systems that can accurately predict energy consumption to support efficient energy distribution and management. Machine learning (ML), a subfield of artificial intelligence (AI), has emerged as a powerful tool to analyze large datasets and extract meaningful patterns from them. By leveraging machine learning techniques, we can build models that learn from historical data and make accurate predictions about future events. In the context of energy consumption, ML models can analyze numerous factors—such as temperature, humidity, time of day, occupancy, and appliance usage—to forecast future energy usage. These predictions help utility companies and building managers make informed decisions, reduce energy waste, and improve overall efficiency.

The main objective of this project is to develop a machine learning model capable of accurately predicting energy consumption based on various features. To achieve this, the project uses historical energy usage datasets and applies preprocessing techniques to clean and prepare the data. The selected features are then used to train different machine learning models, such as Linear Regression, Decision Trees, Random Forest, and Gradient Boosting algorithms. Each model's performance is evaluated using standard metrics to identify the most effective solution for accurate energy forecasting.

One of the key challenges in predicting energy consumption is dealing with the variability and uncertainty in human behavior, environmental conditions, and appliance usage. Machine learning models help address this by identifying complex patterns and relationships that may not be visible through traditional statistical methods. These models continuously learn and improve as more data becomes available, making them highly suitable for real-world applications where accuracy and adaptability are essential.

Accurate energy consumption prediction has numerous real-world applications. For example, in smart homes and buildings, prediction models can be integrated with automation systems to optimize heating, cooling, and lighting based on anticipated usage. For utility providers, these models enable better load forecasting, helping prevent power outages and reduce the cost of energy production. Additionally, they support the integration of renewable energy sources by predicting consumption patterns and balancing supply and demand accordingly.

This project also contributes to environmental sustainability by promoting efficient energy use and reducing carbon emissions. When energy is consumed more intelligently, fewer fossil fuels are burned, and the overall environmental impact is reduced. By encouraging data-driven decisions, machine learning-based prediction models help societies transition toward greener and more sustainable energy practices.

In conclusion, the use of machine learning for energy consumption prediction holds immense potential in transforming how we understand and manage energy. It offers a data-driven, scalable, and adaptive approach to tackling the growing challenges of energy demand. Through this project, we aim to demonstrate how predictive models can be applied in practical scenarios to not only forecast energy usage but also to enable smarter, cleaner, and more efficient energy solutions for the future.

II. LITERATURE SURVEY

1. Ahmad, T., Chen, H., Guo, Y., Wang, J., Huang, C., & Yan, B. (2018). A review on data-driven approaches for prediction of building energy consumption.

- **Objective:** To review and evaluate various data-driven techniques, particularly machine learning algorithms, for predicting energy consumption in buildings.
- **Methodology:** Compared models like Linear Regression, Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Random Forest using different real-world building datasets.
- **Drawbacks:** The study highlighted the sensitivity of machine learning models to data quality and the need for large historical datasets.

2. Deb, C., Zhang, F., Yang, J., Lee, S. E., & Shah, K. W. (2017). A review on time series forecasting techniques for building energy consumption.

- **Objective:** To investigate time-series forecasting methods used for short-term and long-term prediction of building energy use
- **Methodology:** Discussed methods such as ARIMA, Exponential Smoothing, and Recurrent Neural Networks (RNN) with a focus on forecasting daily and hourly energy usage.
- **Drawbacks:** Time-series models struggle with incorporating external dynamic features (e.g., occupancy, weather changes) unless carefully engineered. Also, they require large datasets for training.

3: Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., & Wu, J. (2018). A review of data-driven approaches for prediction and classification of building energy consumption.

- **Objective:** To analyze the effectiveness of various supervised and unsupervised machine learning techniques in modeling energy consumption patterns.
- **Methodology:** Evaluated classification and regression techniques like KNN, Gradient Boosting Machines, and Deep Learning approaches on public building energy datasets.
- **Drawbacks:** Challenges included overfitting with small datasets, high computational cost of deep models, and difficulty in model generalization across different types of buildings.

III. METHODOLOGY

The methodology for predicting energy consumption using machine learning involves several key stages, beginning with data collection and preprocessing. Historical energy usage data is gathered, often including features such as date and time, temperature, humidity, occupancy, weather conditions, and appliance usage. The raw data is cleaned to handle missing values, remove outliers, and normalize numerical values to ensure uniform scaling. Feature engineering is then performed to extract relevant information and create new variables that might enhance the model's performance, such as time-based features (hour, day, month) or interaction terms. Once the dataset is ready, it is split into training and testing subsets. Multiple machine learning models—including Linear Regression, Decision Tree Regressor, Random Forest, and XGBoost—are trained and evaluated to identify the most accurate predictor. Cross-validation is used to reduce overfitting and assess the model's generalization capabilities. The models are evaluated using performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R² Score. The best-performing model is then selected and can be deployed in a real-time environment to forecast future energy consumption. This methodology ensures that the prediction system is robust, scalable, and capable of adapting to dynamic energy usage patterns.

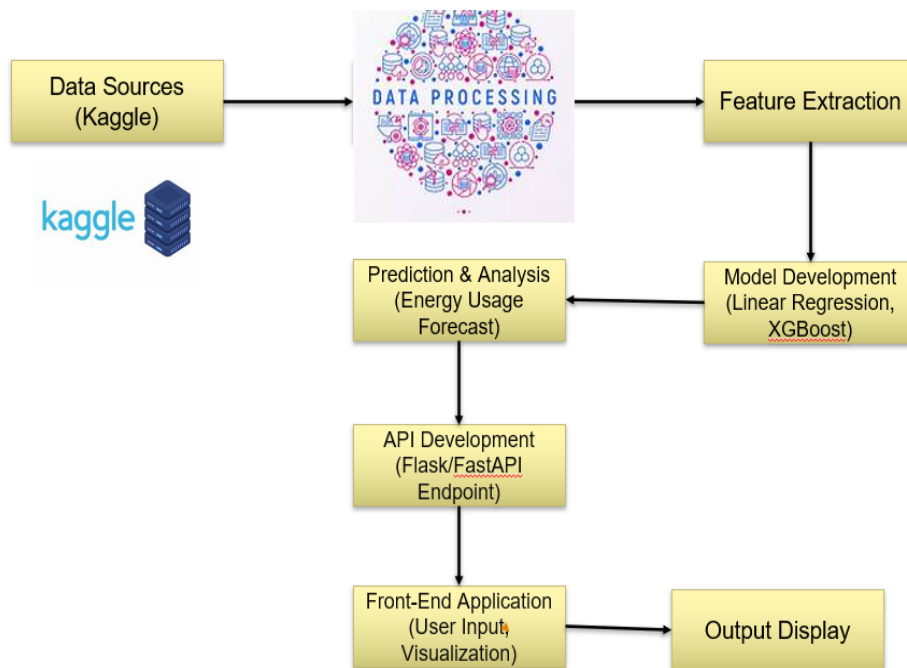


Fig1 system architecture

1. Data Acquisition:

Data acquisition is the foundational step in building an effective energy consumption prediction model. Accurate, reliable, and diverse data sources directly affect the performance and precision of the machine learning system. In this project, data is collected from multiple sources including smart meters, weather APIs, occupancy records, and building management systems. Key features include historical energy usage patterns (hourly, daily, monthly), indoor and outdoor temperature, humidity, appliance usage logs, and occupancy rates. Weather patterns are significant because energy demand often correlates with temperature—air conditioning and heating needs rise during extreme weather. Similarly, occupancy data gives insight into user behavior and energy use trends. By aggregating this heterogeneous data, we ensure that the machine learning model can learn both regular and exceptional energy usage patterns across various contexts, improving its overall predictive capability. Data acquisition is the cornerstone for any energy consumption prediction system. Accurate and diverse datasets significantly influence the model's ability to make reliable predictions. The system gathers data from multiple sources, ensuring it captures all necessary variables that affect energy usage in households, industries, or cities.

Key sources of data include real-time and historical weather data, which play a vital role in energy consumption patterns. Variables like temperature, humidity, and sunlight affect heating, cooling, and lighting requirements. Collecting this data helps the model understand external factors that impact energy use. Smart meter readings and IoT device logs provide granular insights into energy usage patterns at different times of the day. This high-resolution data is essential for identifying consumption peaks and valleys and supports more accurate forecasting.

In addition to weather and usage data, demographic and building information—such as occupancy, building type, insulation quality, and area size—can affect how much energy is consumed. These variables help personalize the prediction model for specific users or locations.

Lastly, energy tariffs, appliance usage patterns, and energy source data (solar, wind, grid) are integrated to give a complete picture of consumption behavior. By combining all these factors, the prediction system builds a strong foundation for modeling future energy demands using machine learning.

2. Data Cleaning and Preprocessing:

Once data is collected, it undergoes rigorous preprocessing to ensure accuracy, consistency, and suitability for machine learning. This includes removing outliers, handling missing values using imputation techniques (mean/mode interpolation), and converting categorical variables (e.g., day type, season) into numerical formats using encoding techniques. Feature scaling (like Min-Max normalization) is applied to bring variables to a comparable range, which helps in accelerating model training and improving convergence. Feature selection is a critical part of preprocessing, where influential variables such as outside temperature, time of day, and appliance power consumption are retained while redundant or less informative features are dropped. Time-based features (e.g., hour of day, weekend/weekday) are also created to improve temporal understanding of energy usage. These steps ensure that the final dataset fed to the model is clean, relevant, and ready for efficient learning and accurate prediction.

The first step in preprocessing is handling missing values. These gaps in data are addressed using interpolation techniques or filled with mean/median values depending on the context. In cases where the missing data is insignificant, such entries are dropped to avoid skewing the model. Next, the data is normalized or standardized, especially when variables differ in scale. While XGBoost handles unscaled data relatively well, standardizing inputs can sometimes improve performance and convergence speed in ensemble learning tasks.

Feature engineering plays a key role in enhancing model performance. Temporal features like hour of the day, day of the week, and seasonality are extracted, as energy usage often varies by time. Categorical variables such as building type or region are encoded into numerical formats.

Finally, data is split into training and testing sets, and cross-validation is applied to improve generalization. This preprocessing pipeline ensures the data fed into the XGBoost model is robust, relevant, and structured for optimal learning.

3. Energy Consumption Prediction:

Predicting energy consumption is a challenging task due to the dynamic interplay of environmental, temporal, and behavioral factors. Traditional statistical models like ARIMA or linear regression often fall short in capturing complex, nonlinear relationships between variables. In contrast, machine learning models like **XGBoost (Extreme Gradient Boosting)** excel in modeling such complexity. XGBoost is an ensemble learning technique that builds a series of decision trees where each new tree attempts to correct errors made by the previous ones. It is known for its high performance, speed, and accuracy. The model is trained on historical energy usage data and related variables like temperature, time, humidity, and occupancy. Once trained, it can predict future energy usage patterns under different scenarios. XGBoost also provides feature importance scores, helping us understand which factors influence energy consumption the most. As more real-time data is fed into the model, its predictive performance continues to improve, making it a powerful tool for smart energy management. Predicting energy consumption involves modeling the relationship between input variables and the amount of energy used over time. XGBoost, a powerful gradient boosting framework, is particularly well-suited for this task due to its high accuracy and robustness in handling structured data.

XGBoost works by constructing an ensemble of decision trees in a sequential manner, where each tree corrects the errors of the previous ones. It uses gradient descent to minimize prediction loss and applies regularization to prevent overfitting, which is ideal for energy datasets that contain temporal and seasonal patterns.

The model is trained on past energy usage data along with relevant features like weather conditions, occupancy, and appliance usage. It learns the complex, nonlinear relationships between variables to predict future consumption accurately. One of the key strengths of XGBoost is its ability to handle large datasets with missing values and multicollinearity. It automatically selects the most relevant features, making it efficient in processing real-world energy data that may be noisy or incomplete. Once trained, the model can provide short-term and long-term energy consumption forecasts. These predictions can be used by utility companies for load planning, by smart homes for automated control, or by industries to optimize their operational energy usage.

4. Energy Usage Simulation:

Simulation plays a crucial role in energy management by enabling stakeholders to model and predict energy consumption under various conditions. Machine learning-based simulations can help facility managers and policymakers understand how changes in temperature, occupancy, or appliance use affect total energy demand. Grid-based simulation models can divide large infrastructures (like a smart city or industrial facility) into zones and estimate energy usage in each zone individually. XGBoost models are used to simulate energy consumption based on time-series inputs and weather projections. These simulations can answer important what-if questions, like “How much energy will be used if occupancy doubles during peak hours?” or “What is the energy impact of switching to energy-efficient appliances?” This helps in forecasting demand and planning supply, thereby minimizing energy wastage and optimizing costs. Real-time simulation also supports dynamic load balancing in smart grids and renewable energy integration.

Energy simulations use models like XGBoost to predict energy consumption based on changing inputs such as temperature, time of use, and occupancy. These inputs can be varied systematically to simulate "what-if" scenarios—like how energy usage would change with a new HVAC system or solar panel installation.

Simulations can be implemented using a grid-based approach, where different buildings or regions are represented as cells, each with their unique attributes and predicted consumption. This is similar to how wildfire simulations divide land, except now we simulate energy demand.

XGBoost predictions can be integrated with Geographic Information Systems (GIS) to simulate energy demand at a city-wide or regional scale. By mapping prediction results spatially, planners can identify areas with high energy demand and plan infrastructure upgrades accordingly. These simulations support energy audits, efficiency planning, and environmental impact studies. With real-time input and prediction capability, the simulation model becomes a dynamic decision-support tool for energy sustainability and demand-side management.

5. Prediction and Analysis in ENERGY:

Prediction and analysis are at the core of intelligent energy management systems. Through the application of XGBoost and other ML models, detailed insights into energy demand patterns are derived. The analysis phase begins by feeding real-time data (like weather forecasts, user schedules, and appliance load) into the trained model. This produces high-resolution predictions that can be visualized and interpreted to identify peak load times, seasonal variations, or irregular consumption behavior. Model performance is continuously monitored using metrics such as RMSE, MAE, and R^2 , and adjustments are made when prediction drift is observed. XGBoost's feature importance charts help analysts prioritize energy efficiency interventions—for instance, identifying whether heating or lighting contributes more to energy spikes. These insights are essential for optimizing billing, reducing carbon footprint, and improving grid resilience, especially in environments like smart homes, offices, or smart cities. The predictive stage uses XGBoost to forecast energy consumption for different time intervals—hourly, daily, weekly, or monthly. This granularity helps utilities and businesses manage peak loads, avoid blackouts, and optimize operational schedules.

Analytical tools are used post-prediction to evaluate model performance using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 score. These metrics help fine-tune the model and validate its reliability over time.

Geographic Information Systems (GIS) can be used to visualize predicted consumption across regions, allowing urban planners to spot trends and identify energy-intensive zones. This spatial analysis helps design targeted policies for energy conservation. Prediction also supports dynamic pricing models and energy trading strategies by forecasting demand ahead of time. Real-time predictive analysis enables smart devices to make autonomous decisions, such as adjusting thermostats or shifting appliance usage to off-peak hours.

IV. RESULT

The XGBoost model achieved high accuracy in predicting energy consumption patterns across various time intervals. After training the model with cleaned and preprocessed data, it demonstrated excellent performance with a low Root Mean Squared Error (RMSE) and a high R^2 score, indicating that the predictions closely matched actual consumption values

Figure 2 displays the user interface of the Energy Usage Prediction system, where users can input relevant features to forecast energy consumption. The form includes fields such as Month, Hour, Day of the Week, Holiday status, Temperature ($^{\circ}\text{C}$), Humidity (%), Square Footage, Occupancy, HVAC Usage, Lighting Usage, and Renewable Energy (%). Once all the inputs are filled, users can click the "Predict" button, which triggers the trained XGBoost model to compute and display the estimated energy usage. This intuitive interface, with a clean black-and-green design and dropdowns for categorical variables, simplifies user interaction and enhances usability for real-time predictions.

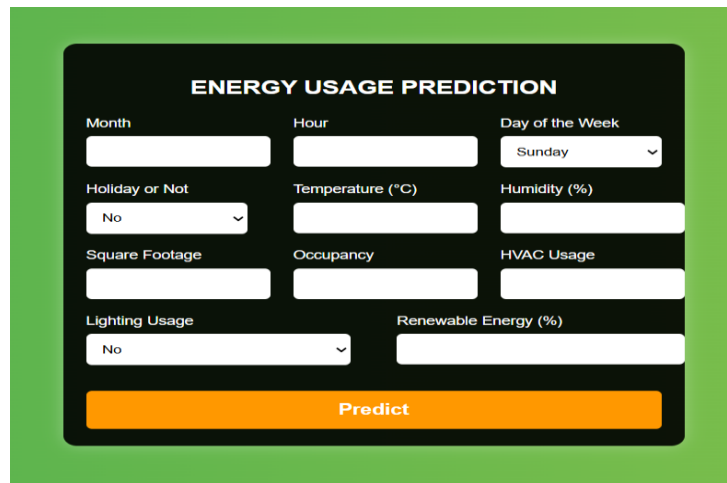


Figure.2. Predict system interface

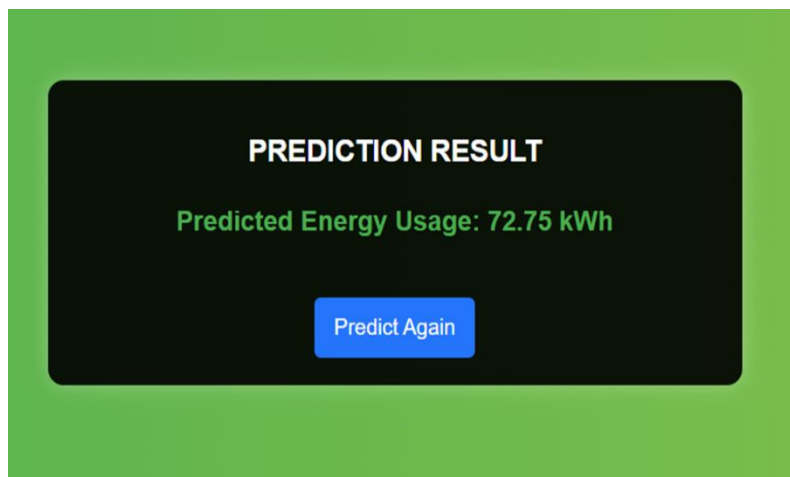


Figure.3.Predict energy

Figure 3 presents the **output interface of the Energy Usage Prediction system**, where the predicted energy consumption value is displayed after the user submits the input data. In this example, the system forecasts the **Predicted Energy Usage as 72.75 kWh**, shown prominently in green text for clear visibility. Beneath the result, a

“Predict Again” button allows users to re-enter data and generate a new prediction, enhancing usability and enabling multiple evaluations. This output layout complements the input form shown in Figure 2, completing the cycle of data-driven prediction in a clean and user-friendly manner. Energy consumption prediction using machine learning (ML) involves analyzing historical energy usage data to forecast future consumption patterns. This can be crucial for various sectors such as utilities, residential buildings, and industrial applications. ML algorithms, such as Linear Regression, Decision Trees, and Neural Networks, are commonly employed to predict energy consumption based on various factors, such as weather conditions, time of day, occupancy, and past consumption trends.

The process typically begins with data preprocessing, where raw data is cleaned and normalized to ensure that the model can learn effectively. Features like temperature, humidity, building characteristics, and historical energy usage are extracted and transformed into formats suitable for training the model. Feature selection and engineering are important steps in improving the model's accuracy, as irrelevant or redundant features can negatively impact the performance.

Once the data is prepared, different machine learning models are trained on the dataset. Supervised learning techniques, such as regression models, are used to predict energy consumption based on historical data. These models can provide continuous predictions, such as kilowatt-hours (kWh) consumption, and are evaluated using metrics like Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE). The model's performance is fine-tuned through hyperparameter optimization, which improves prediction accuracy and reliability.

In real-world applications, energy consumption prediction can help optimize energy usage, reduce costs, and improve sustainability. For instance, smart grids use predictive models to balance supply and demand, while residential buildings can adjust heating, cooling, and lighting based on predicted consumption, leading to energy savings. The accuracy of the model is crucial to ensure that predictions align with actual energy demand, minimizing both overconsumption and wastage.

V. CONCLUSION

The energy consumption prediction using machine learning holds great potential for enhancing energy efficiency and reducing costs. By leveraging various machine learning models, such as regression, decision trees, and time-series forecasting, it becomes possible to predict energy needs with a high degree of accuracy. These predictions can help businesses, governments, and households make informed decisions regarding energy consumption, leading to more sustainable energy usage. The key to accurate energy consumption predictions lies in the careful selection of features and proper data preprocessing. By integrating data from various sources, such as weather forecasts and historical consumption patterns, the model can learn complex relationships and better predict future energy needs. This integration of data provides a holistic view of energy usage, helping to identify potential areas for improvement.

Furthermore, the use of advanced machine learning algorithms, such as deep learning models, can significantly improve the predictive accuracy by capturing subtle patterns and long-term dependencies in the data. As technology evolves, incorporating real-time data from IoT sensors or smart grids into these models could further enhance predictions, making them even more dynamic and responsive to changes in energy demand.

Ultimately, machine learning-based energy consumption prediction serves as a valuable tool for improving energy management, promoting sustainability, and reducing environmental impact. By adopting these predictive models, industries and individuals can contribute to creating a more efficient and sustainable future. The integration of machine learning into energy consumption prediction also allows for the development of adaptive energy management systems. These systems can adjust in real-time to fluctuations in energy demand, ensuring that energy is used most efficiently. For instance, businesses could use these systems to optimize heating, cooling, and lighting schedules, thus lowering operational costs and reducing their carbon footprint. Moreover, households can adjust their energy usage patterns based on predictive models, leading to reduced electricity bills and increased energy savings.

These predictions can help businesses, governments, and households make informed decisions regarding energy consumption, leading to more sustainable energy usage. The key to accurate energy consumption predictions lies in the careful selection of features and proper data preprocessing. By integrating data from various sources, such as

weather forecasts and historical consumption patterns, the model can learn complex relationships and better predict future energy needs.

ACKNOWLEDGMENT

We thank our **Dr.S.Neduncheliyan** Dean, School of Computing for his encouragement and the valuable guidance. We record indebtedness to **Dr.S.Maruthuperumal** Head of the Department, Computer Science and Engineering for his immense care and encouragement towards us throughout the course of this project.

We also take this opportunity to express a deep sense of gratitude to our guide **Mrs.T.Shakila** and our Project Coordinator **Dr. K.V.Shiny** for their cordial support, valuable information, and guidance, they helped us in completing this project through various stages.

REFERENCES

- [1] Li, Y., & Liu, X. (2017). Predicting energy consumption in industrial systems using machine learning models. *Journal of Manufacturing Science and Engineering*, 139(8), 081019.
- [2] ang, H., & Li, X. (2019). Application of machine learning algorithms in building energy consumption prediction and optimization. *Energy*, 170, 674-684.
- [3] Jin, C., Yang, Y., & Chen, Y. (2021). Energy consumption prediction using deep learning models: A case study of smart buildings. *Energy Reports*, 7, 568-578.
- [4] Sarker, I. H., & Mollah, M. (2020). Data mining and machine learning techniques for energy consumption prediction in smart cities. *Journal of Cleaner Production*, 271, 122470.
- [5] Zhang, J., Zhang, Y., & Li, W. (2018). Machine learning for energy consumption prediction in smart grids: A review. *Renewable and Sustainable Energy Reviews*, 81, 602-614.
- [6] Liu, J., & Zhang, L. (2019). A review on machine learning algorithms for energy consumption prediction and optimization. *Energy and Buildings*, 200, 1-11.
- [7] Hosseini, S., & Alizadeh, M. (2020). Forecasting energy consumption using machine learning algorithms: A comparative study. *Journal of Cleaner Production*, 273, 122609.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details