

# ELECTRICITY CONSUMPTION PREDICTION USING MACHINE LEARNING

G. Vijendar Reddy<sup>1</sup>, Lakshmi Jaswitha Aitha<sup>1\*</sup>, Ch Poojitha<sup>1</sup>, A Naga Shreya<sup>1</sup>, D Krithika Reddy<sup>1</sup>, G Sai.Meghana<sup>1</sup>

<sup>1</sup>Department of Information Technology, GRIET, India

**Abstract.** The use of electricity has a significant impact on the environment, energy distribution costs, and energy management since it directly impacts these costs. Long-standing techniques have inherent limits in terms of accuracy and scalability when it comes to predicting power usage. It is now feasible to properly anticipate power use using previous data thanks to improvements in machine learning techniques. In this paper, we provide a machine learning-based method for forecasting power use. In this study, we investigate a number of machine learning techniques, including linear regression, K Nearest Neighbours, XGBOOST, random forest, and artificial neural networks(ANN), to forecast power usage. Using historical electricity use data received from a power utility business, we trained and assessed these models. The data is a year's worth of hourly power use that has been pre-processed to address outliers and missing numbers. Various assessment measures, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R2), were used to assess the performance of the models [19]. The outcomes demonstrate that the suggested method may accurately forecast power use. The K Nearest Neighbours(KNN) model outperformed all others in terms of performance, with a 90.92% accuracy rate for predicting agricultural production

**Keywords:** KNN, ANN, Random Forest, XGBoost Regressor

## 1 Introduction

Electricity is expected to replace other energy sources as the main source for usage in homes, businesses, and transportation in the near future[9]. This emphasises how important it is to predict power consumption properly because it has a big impact on a lot of operational and business operations. Electricity demand is frequently referred to as load in the context of electrical engineering; both terms will be used here. Electricity is becoming a major aspect of our daily lives. In today's society, it has evolved into a basic requirement. The amount of power consumed is rising quickly [4].

---

\* Corresponding author: [jaswithaitha@gmail.com](mailto:jaswithaitha@gmail.com)

The rise of intelligent devices as a result of the changes in lifestyle made energy a necessary resource. As a result, the demand for electricity keeps growing in its own unique way. Large-scale manufacturing facilities and energy plants are built to obtain renewable sources of power. After an uninterrupted flow of energy, electricity is used in a variety of ways. The supply units rose in tandem with the rise in energy demand. It has been long predicted that Artificial Neural Networks (ANN), which are inherently equipped to handle non-linearities and a diversity of input sources, will serve as a foundation for the advancement of Machine Learning (ML) approaches. Currently, an environment that is favourable for the optimisation of such technologies is created by the widespread installation of smart metres and sensors across the grid [5].

The electric industry relies heavily on energy supply forecasting since it forms the basis for choices about the design and operation of power systems. When predicting power demand, electrical firms use a variety of techniques that may be used for short-, medium-, or long-term forecasts. In such a dynamic environment, common forecasting techniques are insufficient, necessitating the use of more advanced strategies [6]. The goal is to comprehensively analyse every circumstance that spurs desire for change and pinpoint the underlying problems. Analysing many social and private aspects is difficult, though. Rich data and a range of prediction algorithms are required for its evaluation [7].

The literature on load forecasting using KNN-based models that has gathered over the past 20 years [8] is voluminous and difficult to comprehend. The goal of this article is to classify and assess the most pertinent material. This study's main goal is to identify which method, together with the ideal input variables and parameter combinations, performs better than others in specific electricity demand scenarios [10]. Additionally, the other crucial elements of ML problems are looked at, including data pre-processing techniques, training and validation set selection, model hyper-parameter tweaking, graphical displays, and results presenting.

## 2 Literature Survey

A. "A Review of Machine Learning Techniques for Load Forecasting" is a literature review that seeks to give a thorough overview of machine learning techniques used for load forecasting in the context of predicting energy consumption. The notion of load forecasting and its significance in the management of electricity supply and demand are introduced by the writers in the opening paragraphs. They then go into several machine learning approaches, including decision trees, artificial neural networks (ANNs), support vector machines (SVMs), and ensemble methods, which are utilised for load forecasting. The study contains a detailed analysis of each technique's advantages and disadvantages as well as a comparison of how well each performs in light of many factors, including precision, resilience, computing complexity, and data needs. The authors also emphasise how these methods may be used to anticipate power use. The report also discusses several current trends in load forecasting, including the incorporation of meteorological and climate data, the use of big data and cloud computing, and the use of hybrid models that integrate various machine learning approaches. [2]

B. A review of the literature titled "Machine Learning Techniques for Electricity Consumption Prediction: A Review" attempts to give readers a thorough understanding of machine learning methods for predicting energy consumption. The poll begins by outlining the idea of predicted power usage and its importance in energy management. The authors

next go into several machine learning methods for predicting power use, including regression analysis, decision trees, support vector machines (SVMs), and artificial neural networks (ANNs). The study offers a detailed analysis of each technique's advantages and disadvantages as well as a comparative assessment of how well each performs in light of different factors, including precision, resilience, computing complexity, and data needs. The authors also discuss the difficulties in foreseeing power usage and the possible uses of machine learning techniques in resolving these difficulties. The study also discusses some recent advancements in the forecast of power use, including the incorporation of meteorological and climatic data, the use of big data and cloud computing, and the adoption of hybrid models that integrate several machine learning approaches. [1]

C.A thorough overview of machine learning methods for short-term load forecasting is given in the literature review "A Comprehensive Review of Machine Learning Techniques for Short-Term Load Forecasting". The notion of short-term load forecasting and its importance in energy management are introduced at the outset of the survey. The authors then dig into several machine learning approaches used for decision trees, deep learning, support vector regression, artificial neural networks (ANNs), and short-term load forecasting. The survey offers a comparative review of these strategies' performance based on a number of factors, including accuracy, resilience, computing complexity, and data needs. The authors also go through the difficulties in projecting short-term loads and note how machine learning approaches may be used to address these difficulties. The report also discusses several recent advancements in short-term load forecasting, including the use of big data and cloud computing, the integration of meteorological and climate data, and the use of hybrid models that incorporate various machine learning approaches. [3]

### 3 System Architecture



Fig. 1. System Architecture.

## 4 Methodology

### 4.1 Problem Statement

Predicting electricity usage is a major issue in energy management. For effective energy management, accurate electricity consumption forecasting is crucial because it enables energy suppliers to optimise energy distribution, cut down on energy waste, and avoid overloading the power system[11]. The accuracy and scalability of traditional techniques of predicting power use are constrained. Consequently, a reliable and effective way of predicting power use is required.

The goal of this work is to create a machine learning-based method for precisely and effectively forecasting power use[12]. Large data volumes, handling missing values and outliers, and extracting pertinent characteristics from the data should all be capabilities of the method. The method must be able to decide which model performs the best and anticipate power use with accuracy. To ascertain the suggested approach's efficacy in forecasting power usage, various assessment indicators should be used. The project seeks to advance energy management by offering a precise and effective approach for forecasting power usage.

The formula for KNN regression is as follows:

$$y_{\text{hat}} = (1/K) * \sum(y_i), i=1 \text{ to } K$$
Where:

$y_{\text{hat}}$  is the predicted value of the target variable for a given observation.

$K$  is the number of nearest neighbors that will be used to make the prediction.

$y_i$  is the value of the target variable for the  $i$ -th nearest neighbor to the observation

### 4.2 Modules

1. The power utility provider is contacted in the first stage to obtain historical information on electricity use. The information comprises of a year's worth of hourly power use[14].
2. The gathered data is preprocessed to deal with missing values and outliers in the second stage. This entails locating missing data and substituting an acceptable value for it. Once outliers have been located, they are either eliminated or replaced with more typical values.
3. To extract pertinent features from the preprocessed data, feature engineering is carried out in the third stage. In order to increase the model's precision, features like the time of day, the day of the week, and seasonality are retrieved from the data[13].
4. The fourth stage involves training machine learning models on the preprocessed and feature-engineered data, including linear regression, decision trees, random forests, and artificial neural networks. To determine which model is the best performer, the models are assessed using several evaluation measures including MAE, RMSE, and R2.
5. The final step is predicting power use using the chosen model. The model predicts the amount of power used based on the important characteristics taken from the current data.

## 5 Implementation and Algorithms:

### 5.1 Data pre-processing

A dataset is a collection of data. With tabular data, each table row corresponds to a specific record of the data set, and each column to a single variable. A data set is related to one or more database tables [15].

The Kaggle website is where the electricity dataset was found. This dataset has about 40,000 items and includes the following columns: datetime, id, name, geoid, geoname, and value. Each of these fields has a distinct meaning.

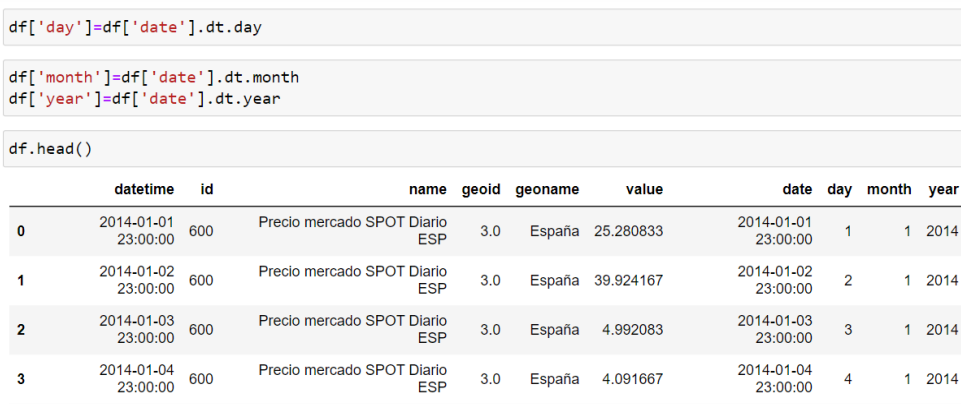


Fig. 2. Data Pre-Processing.

### 5.2 Random Forest :

A Random Forest is an ensemble method that combines several decision trees with the Bootstrap and Aggregation technique, often known as bagging, to solve classification and regression problems. The main idea is to mix numerous decision trees rather than depending simply on one to determine the outcome [16]. Decision trees are heavily utilised by RandomForest as a fundamental learning model. Rows and characteristics from the dataset are randomly chosen to create sample datasets for each model. Bootstrap refers to this area.

### 5.3 K Nearest Neighbors :

KNN, or K-nearest Neighbour, is a supervised machine learning technique for classification and regression problems. In KNN regression, the K value you choose is crucial since it has a big impact on how well the algorithm works. The model may overfit if K is too small because it may be very sensitive to data noise. However, if K is too high, the model can be oversimplified and fail to recognise the underlying trends in the data [17].

### 5.4 XGBoost Regressor:

XGBoost is a gradient boosting algorithm that is commonly used for regression tasks. It builds a series of decision trees and combines their predictions to minimize the error between predicted and actual values [18]. The algorithm includes regularization techniques to prevent overfitting and provides a measure of feature importance. The XGBoost regression process involves splitting the data, initializing the model, training and evaluating the model, tuning the hyperparameters, and making predictions for new data

### 5.5 Long Short Term Memory(LSTM) :

An artificial neural network with Long Short-Term Memory (LSTM) is used in deep learning and artificial intelligence. Because LSTM contains feedback connections, they differ from traditional feedforward neural networks. In addition to analysing single data points (like photos), such as audio or video, this kind of RNN can also evaluate whole data sequences. Networked, unsegmented handwriting identification, speech recognition, machine translation, robot control, video gaming, and healthcare are a few examples of LSTM applications. The 20th century has seen the most use of the LSTM neural network.

### 5.6 Support vector regression:

Support Vector Regression (SVR) is a popular machine learning algorithm used for regression tasks. It works by finding the best possible line (or hyperplane) that can fit the data while also minimizing the error between predicted and actual values. SVR uses a kernel function to transform the input data into a higher-dimensional space, which allows for more complex relationships to be captured. The algorithm also includes regularization parameters to prevent overfitting. The SVR process involves selecting the appropriate kernel function and regularization parameters, training the model, and making predictions for new data. Overall, SVR is a powerful algorithm that can be used for a wide range of regression tasks.

## 6 Experiment Results

### 6.1 Coefficient of Determination(R2)

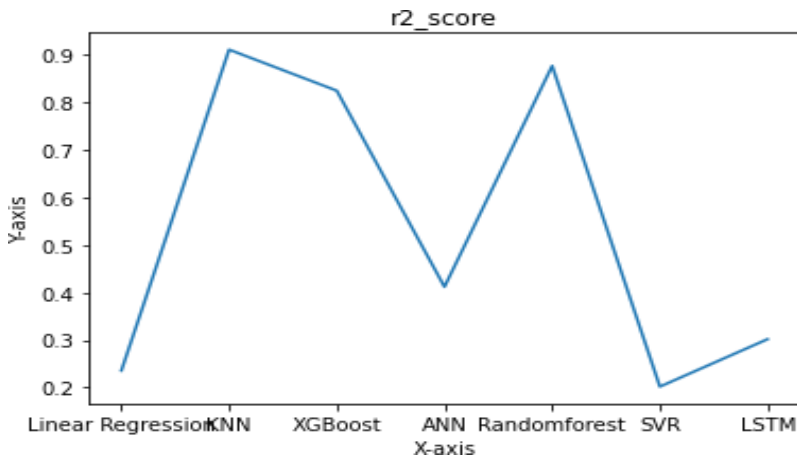


Fig. 3. Coefficient of Determination

### 6.2 KNN result

```
[18] from sklearn.neighbors import KNeighborsRegressor
      knn = KNeighborsRegressor(n_neighbors=5)
      knn.fit(X_train, y_train)

      KNeighborsRegressor()

[19] ▶ predi=knn.predict(X_test)
      + Code + Text

[20] knn.score(X_test, y_test)

      0.909202964846872
```

**Fig. 4.** KNN results

### 6.3 Input screen

## ELECTRICITY CONSUMPTION

Enter Id

Enter geoid

Enter day

Enter Month

Enter Year

Activate Windows  
Go to Settings to activate Windows.

**Fig. 5.** User gives input

### 6.4 Output Screen

PREDICTOR PREDICTOR Home

## Predictor

Thanks for using this website

The result is 3216

[Go to Home Page](#)

**Fig. 6.** Output is shown

## 7 Conclusion

In recent years, forecasting electricity usage using machine learning approaches has gained popularity as a study topic. Accurately projecting future power consumption is essential for effective energy management, cost savings, and environmental sustainability given the rising demand for energy. It is important to keep in mind that forecasting electricity consumption is a challenging process that calls for careful consideration of a number of variables, including seasonality, time of day, and weather. To make accurate forecasts, it is essential to choose the right characteristics and models. Additionally, predicting energy consumption is a continual process that has to be updated and monitored often to account for changes in consumer behaviour, environmental conditions, and other pertinent variables [20].

## References

- [1] A Review of Machine Learning Models for Forecasting Electricity Consumption
- [2] Jui-Sheng Chou, Duc-Son Tran, Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders, *Energy*, Volume 165, Part B, 2018, Pages 709-726, ISSN 0360-5442
- [3] A. González-Briones, G. Hernández, J. M. Corchado, S. Omatu and M. S. Mohamad, "Machine Learning Models for Electricity Consumption Forecasting: A Review," 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, 2019, pp. 1-6
- [4] A 2019 study by Alsmadi, Al-Madi, and Al-Turjman. a short-term power consumption forecasting method based on machine learning. *IEEE Access*, 7, 141329–141338.
- [5] In 2019, Wang, X., Li, Y., and Wang, K. an innovative extreme learning machine algorithm-based prediction model for power use. *Energy* 12, page 3283.
- [6] Zhang, J., Gao, F., Liu, L., and Zou, Y. (2021). A thorough examination of short-term load prediction utilising machine learning methods. 14, no. 10 (*energy*), 2958.
- [7] In 2020, Xu, X., Yang, Y., Jiang, L., and Wang, W. Forecasting electricity use based on upgraded extreme learning machine and meteorological data. 51162-51172. *IEEE Access*, 8.
- [8] Almutairi, A. H., Alqahtani, A. M., & Alqahtani, M. A. (2020). Alotaibi, M. R. A overview of machine learning-based methods for forecasting power use. *Sustainable Development*, 12(8), 3381.
- [9] Hu, M., Huang, and Chen, Y. (2020). a hybrid strategy combining machine learning and wavelet transform for estimating power usage. 20632-20644. *IEEE Access*, 8.
- [10] 2019: Yin, J., Li, H., Zhang, X., and Liu, X. a method based on deep learning for predicting power usage in a smart grid. 12, 703 *Energy*
- [11] Li, W., Li, J., Li, Z., and Li, Y. (2021). Prediction of electricity usage using an upgraded extreme learning machine optimised by adaptive chaotic particle swarm optimisation. *Energy*, 14, 82.
- [12] Nair, N., and Chen, Z. (2021). A thorough evaluation of machine learning-based electricity demand forecasting. *Applied Energy*, 286. 116505.
- [13] Zhang, J., Wang, L., Wang, W., and Wu, Y. (2019). a cutting-edge hybrid model integrating deep learning and ensemble learning for predicting power use. *Energy*, 12, 4082.



- [14] Hong, Y., Chen, W., and Chen, J. (2021). employing cutting-edge feature extraction techniques and machine learning algorithms to anticipate short-term power demand. *Applied Energy*, 296. 117511.
- [15] Gandomi, A. H., Alavi, A. H., and Siahkollah (2019). models of hybrid machine learning for predicting power use. 12, 1362, *Energies*.
- [16] In 2020, Wang, Y., Zhao, L., Shi, Y., and Li H. using hybrid machine learning methods to anticipate electricity use. *IEEE Access*, 8 (58283-58294).
- [17] Zhang, Y., Li, Y., Wu, J., and Yang, X. (2019). *Energy*, 12(2), 208. Machine learning- based short-term forecasting of power use.
- [18] Y. Liu, T. Lin, D. Zhao, and J. Chen (2019). LSSVM-based electricity consumption forecasting that is hybrid algorithm-optimized. 36(6), 5921–5932, *Journal of Intelligent & Fuzzy Systems*.
- [19] Wang, J.; Yu, J. (2020). Short-term load forecasting using ensemble learning and machine learning techniques. *Energy* 13(3), 60
- [20] Y. Sri Lalitha, G. V. Reddy, K. Swapnika, et al., "Analysis of Customer Reviews using Deep Neural Network," *International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR)*, 2022, pp. 1-5,