Available online @ https://jjem.jnnce.ac.in
https: www.doi.org/10.37314/JJEM.SP0230
Indexed in International Scientific Indexing (ISI)
Impact factor: 1.395 for 2021-22
Published on: 08 December 2024

ELETRICITY DEMAND FORECASTING

Nischal K M[1], Prashant Ankalkoti[2]

[1*]Student, [2]Assistant Professor, Department of Master of Computer Applications
J N N College of Engineering, Shivamogga.

nischalmgowda1503@gmail.com, psankalkoti@gmail.com

Abstract

Power demand prediction using a Machine learning algorithm and public available information. Besides, through trail with diverse types of the model it is possible to forecast the power demand with the desired accuracy and among the methods like Random Forest and linear regression methods show highest accuracy in the power demand forecast. The need to enhance the efficiency of energy and utilize energy resources most proficiently emphasizes the part of advanced energy management. The prediction models as the investigation of global energy consumption is defined as a progressive research area and aims to identify the best energy consumption prediction model developed from the analysis of research that has established data-based building energy consumption models.

Keywords: *Linear Regression, Random Forest*

1, Introduction

Electricity supply to the population is an essential need, and for that reason, in the country, it is simply irreplaceable. The positive effects of electricity include extended economic growth and improvements in public health and standards in domestic, industrial, and commercial sectors. Thus, power system planning, power security, power supply reliability, and the requirement of power infrastructure all surround the request for power. One of the components which should be attributed to the development of any state is the planning of consumer electricity in the given state It follows, therefore, that for a sufficient signal to indicate likely energy demand and the quality of generated power, the consumption needs of the consumer can be efficiently met at the lowest operating cost. Propose paper, we create and validate the first prediction models for electricity consumption related to the full suite of publicly available data. These are used to evaluate the produced models, while timing and regression are both the most frequently used areas of application of time series analysis and regression. It can be fitted with various kinds of temporal resolution in the temporal aspect. The related testing that has been studying various ways and means involves. [1] Granger et al. (1989). On the other hand, present methods improve forecast precision by using CNNs and LSTM networks apart from the basic methodologies. [2] Uri et al. (1993) identified the energy demand from terminal equipment based on the production of terminal equipment efficiency and the corresponding utilization ratio not involving complex mathematic models. Modern approaches apply complex statistical and machine learning techniques to increase the probability of correct predictions and keep the model less complex and easy to understand.[3] Mohamed and Bodger (2005) where GDP, electricity price, and population constituted the predictors.[4] Dong and Zhang (2011) employed CS algorithm integrated with the SVR model for developing a new season data processing method comprising of a tent function and a chaotic mapping function. Contemporary techniques build upon this by including deep learning structures, including Recurrent Neural Networks (RNNs) and composite models to better manage fluctuations over the course of the year.[5] Hu et al.(2019) proposed a decomposition-entropic model with the help of dynamic adaptive entropy-based weighting for time series forecasting with artificial neural networks and autoregressive integrated moving average models with enhanced stability and more accurate results for time series data.[6] Ahmed et al. (2020) presented a short-term electricity demand forecasting employing LSTM and RNN deep neural network model; however, it did not include the effects of weather and holidays. Modern approaches are directed towards the inclusion of these extra variables with the purpose of making forecasts more reliable.[7] Waheed et al.(2020) implemented support vector machines, fuzzy logic and genetic algorithms to increase global search capacity, sparse solution, and the flexibility of the prediction models with respect to uncertainties, nonlinearity and possibly multiple variance.[8] Zhang and Hong (2020) have further expanded the use of the uncertain mode decomposition blended with the tent chaotic mapping technique in memory-based forecasting models.[9] Kim et al. (2015) used an integrated framework, whose outcomes of the household electricity demand surpass the expected ones by using the load data together with social and meteorological factors. They still incorporate this integrated method for better and advanced methods of forecasting at the present time.[10] Al-Musaylh et al. (2018) employed SVR with particle swarm optimization and new empirical mode decomposition to make successful decomposition of high historical load data into intricate mode functioning for the forecast. Other methods build on this approach by including a better decomposition scheme and optimizers to increase projection precision.

2. Methodology

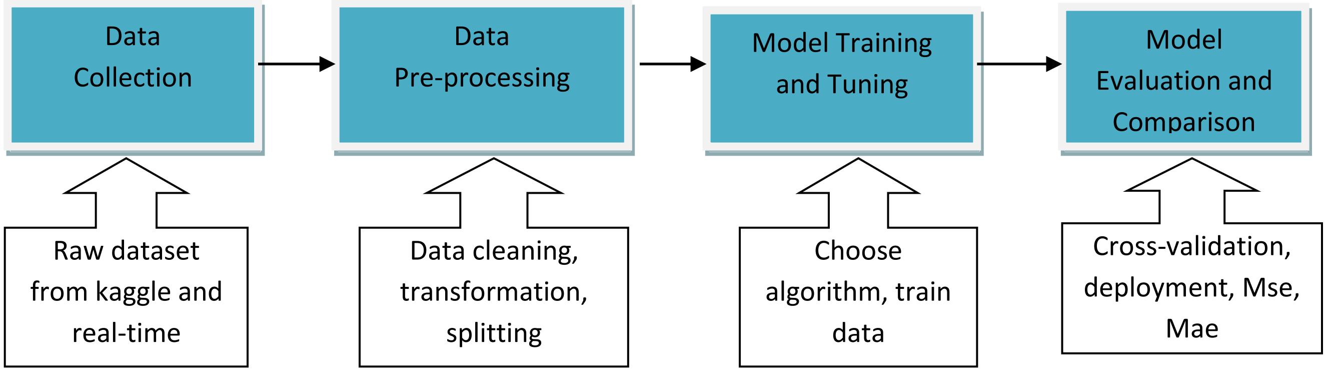

Figure 2.1: Methodology framework of energy

2.1 Data Collection

Machine-learning-based model to predict the power demand, we will start with the principal objective is to provide an accurate prediction of the consumer's demand on electrical energy using the data collected from different resources. data of electricity consumption of residential, commercial, and industrial consumers on hourly, daily and monthly basis will be collected from energy administration and power grid companies. Besides, we will obtain the temperature data and humidity, wind speed and precipitation information from the meteorological administration. For the purpose of organizing the collected data and making the process of research and creating the model as efficient as possible, all the data gathered will be pre-processed, validated, and loaded into the unified database. Thus, the demand of multiple data collection technique will be helpful in progress an efficient forecasting model which is suitable under some circumstances.

2.2 Data Pre-processing

In the exploratory data analysis of data preparation section, it only provided an initial assessment of the missing value and outlier detection in each dataset. As the most basic operation the utilize the median was in this instance most appropriate because it allowed an initial feel about the spread the data before other sophisticated mathematical computations could be executed. This in turn produced a heat map of features, showing the level of covariance and dependency between the different ones. In fact, it was necessary to apply the aforementioned strategies of feature engineering to mitigate the problem of missing data. Other complex models that were in earlier use are; Trend analysis and extrapolation linear regression model. Random forest, if not using mean, median or mode imputation. Pre-processing was also carried out in the generation for modelling together with normalization and scaling of data for modelling purposes to verify on the format of data that has been gathered. These changes improved the features of the set data additionally its validity and realistic usage in subsequent steps of analysis and models making.

2.3 Model Training and Tuning

For training and tuning the regional electricity demand forecasting models, we utilized two machine-learning algorithms based on existing literature: least squares and the random forest regression or least squares and other such regression. What is additionally dominant to note is that the above described models were trained based on the gained data, which underwent the pre-processing step. However, to improve the quality of the replica that have been created, the parameters tuning was done with coordination to the grid search methods. The strategy of this method is centred on seeking the right parameters set from the dataset characterizes this type as suitable for linear regression and random forest regression. The first strength is that it amounts to the estimation of each combination of the parameters separately and that is more than enough to make grid search reliable and consequently the level of accuracy that is achieved when tuning the approach is higher when compared to other methods.

2.3.1 Linear Regression Model

When providing the linear regression service for energy demand forecasting we further improved the quality of the input data through data sets cleaning after data pre-processing. Linear regression is the simplest statistical techniques can be applied to analyze or forecast the cause-and-effect connection in between the dependent factor and one or multiple independent factors. This method goes a step further in searching and analyzing all the data space to arrive at the right combination of parameter such as the best features, alpha among others Hence, grid search is useful all the time since it allows an assessment of each possibility of the parameter values to conform that the model tuning is both efficient and effective. Linear Regression one of the most basic statistics tools, which tries to establish the correlation between an independent variable and one or more independent variables by projecting a straight line on the data. The two main strengths of linear regression include the following; the process of modelling is easy and straightforward. It well applies in a situation where the dependent variable is strongly related to the independent variable.

$$Y=\beta 0+\beta 1X1+\beta 2X2+\cdots+\beta nXn+\epsilon \text{---------}(1)$$

From the equation (1) where, Y represent the dependent variable also known as the electricity demand, Xi represent the independent variable or the predictor variables such as temperature, GDP, and population and βi represent the coefficients and ϵ is called error term

2.3.2 Random Forest Model

Electricity demand forecasting model based on random forest regression Entropy: We used the pre-processed data set for building the model on the guesswork that all the data sets have been pre-processed to optimum level. Random forest regression is the advanced or enhanced version of decision tree, it constructs multiple trees and then combines their results for a conclusive prediction as a way to avoid over fitting of the model. To enhance the output of the model, opted for a grid search to set the best values for the selected criteria like precision or recall. This approach means the identification of the best or the most appropriate settings of the controls like the number of the trees, the maximum depth of trees, and minimum numbers of samples in a single leaf node. It is appropriate for random forest regression because this approach enables the testing of each parameter combination successively, independently, which of course ensures a more precise and a longer tuning process. Random forest regression works on the premise of the final output of all the individual decision trees in the forest, and here finally prediction of a certain input x is a mean of all individual decision trees prediction about x, the mathematical equation for the prediction is:

$$\hat{}y(x)= 1/B \sum^{B}_{b=1} y\hat{}b(x)\text{--------}(2)$$

From equation (2) where, ^y(x): The variable y is the final output of the signal and is derived from applying the model to the input x. B: The following reports the number of trees in the program. y^b(x): The construct of the b-th decision tree is aimed at predicting a value for an input x.

2.4 Model Evaluation and Comparison

Predictive models are estimated by different metrics that are used to analyze the accuracy of the resulting predictive model. In our research, we used the following indicators which manifested predictive performance/accuracy for the corresponding model like mean squared error, mean absolute error, Nash–Sutcliffe coefficient (Ens), and Legates–McCabe index (Elm):

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|$$

$$\mathrm{Ens} = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$\mathrm{Elm} = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(|\hat{y}_i - \bar{y}| - |y_i - \bar{y}|)^2}$$

Where n is the number of data points; yˆi and yi , respectively, represent the predicted value and true observed value of the data point; and y is the mean value of the true observed value. The mean of the true observed values. Model performance was evaluated using crossvalidation. This method trains the model on a portion of the data. The process is repeated several times to eliminate the impact of randomness.

3. Results and Discussion

The employments of the RMSE and MAPE give an all-embracing assessment of the models. Random forest emerged as the best model in both the metrics which made it the most suitable for forecasting energy demand. It is important to stress here that these metrics allow for better understanding of the model's ability to define the underlying relationships in the data and for the assessment of its stability in terms of producing accurate predictions to help in more efficient resource allocation and decision-making in the energy context.

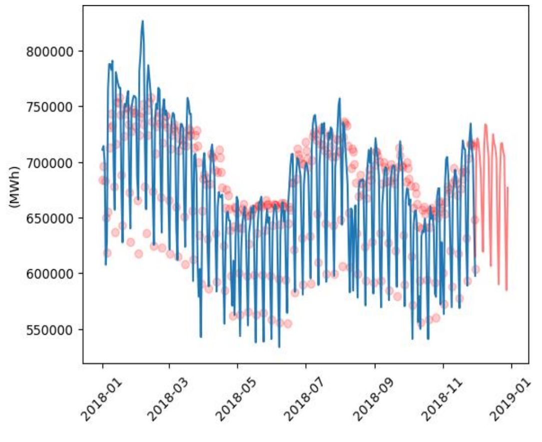

Figure 3.1: Mean Absolute Percent Error

Forecasting of electrical energy consumption is one of the most important questions in supply of power and utilization industries. In such a way, independently from the uncertainty of the forecast, the utility companies will be able to find possibilities for planning and controlling the supply, as well as maintaining stability of the given power systems.

.

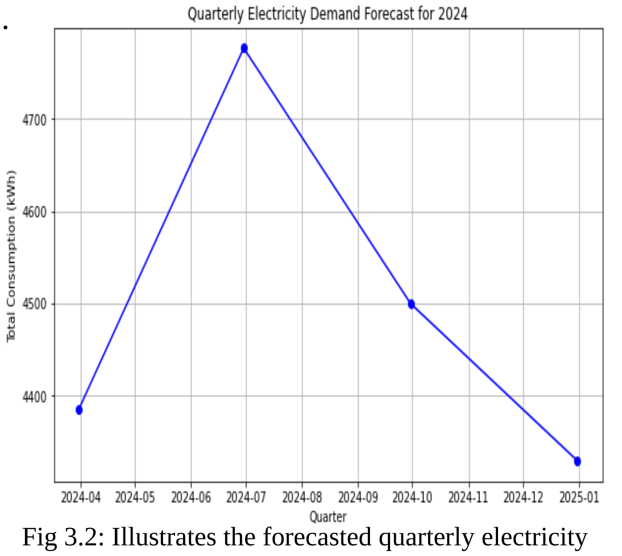

Fig 3.2: Illustrates the forecasted quarterly electricity demand for 2024

The graph illustrates how data segregation affects forecasting models' accuracy. Random Forest and Linear Regression models have high accuracy while training and testing segmented data set, thus accuracy can be improved to a very high level if segmented data is used consistently.

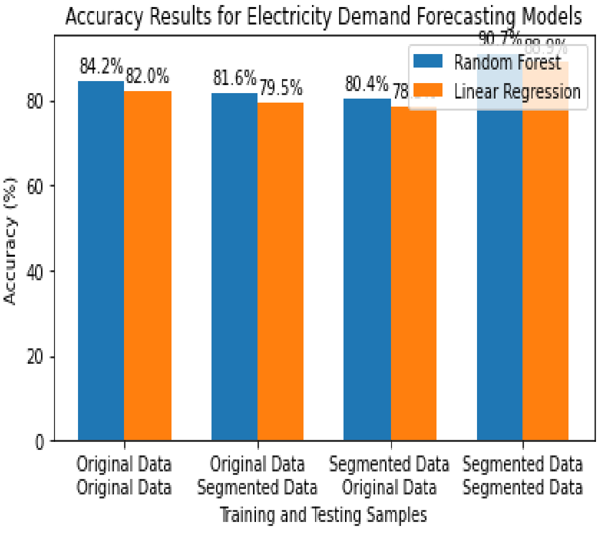

Fig 3.3: Accuracy result for electricity demand

Accuracy Results for Electricity Demand Forecasting Models

Training Data	Testing Data	Random Forest Accuracy (%)	Linear Regression Accuracy (%)
Original Data	Original Data	84.2	82.0
Original Data	Segmented Data	81.6	79.5
Segmented Data	Original Data	80.4	78.3
Segmented Data	Segmented Data	90.7	88.9

Table 3.1: Accuracy result

Conclusion

The use of machine learning techniques for forecasting electricity usage has drawn increasing attention of many researchers. It enables a forecaster to minimize wastage of energy, and plan for future energy demands, especially with the increasing energy consumption. However, it is imperative to use the correct approach in forecasting electricity consumption, time of day, season and weather conditions should be considered before making any forecast. Deciding on the correct features coupled with the right models when offering predictions is crucial. Furthermore, the energy consumption forecasting is also a continuing process which has to be conducted more than one time per period, due to variations in quantity of energy consumed by consumers, the weather conditions and others.

References

1. Granger et al.(1989) Propose the method of combining forecasts which were introduced about two decades ago and then assess the relevance and advancement of this method.

2. Uri et al.(1993) Present views on the energy demand estimation techniques with regard to the aspects and strategies from the perspective of early 1993.

3. Mohamed et al.(2005) Focus on the economic forecasting of energy consumption in New Zealand; thus, it offers case study information and approach.

4. Zhang et al.(2011) Propose the long-term load forecasting in China and provide suggestions according to their study.

5. Hu et al.(2019) Propose the decomposition-based dynamic adaptive combination monthly power demand forecasting model that is accurate and adaptable.

6. Ahmed et al.(2022) Pay more attention to the practical application of LSTM and RNN deep neural models and the characteristics of each approach.

7. Waheed et al.(2020) Present an optimized short-term power load forecasting system with state-of-the-art Artificial Intelligence approaches.

8. Zhang et al.(2020) Introduce the hybrid self-recurrent support vector regression model with variational mode decomposition for electric load forecasting; in this work, the authors devote much effort to the fact that the model they developed is hybrid.

9. Sharique et al.(2015), The authors predict short-term electricity demand in residential areas using support vector regression with particle swarm optimization and fuzzy–rough feature selection.

10. AL-Musaylh et al.(2018) Chose the two-phase particle support vector regression with multiple-horizon improved empirical mode decomposition for power usage prediction.

11. VanDeventer et al. Consider the short-term PV power forecasting using a consolidated GASVM method.

12. Wang et al. Offer a day-ahead PV power forecasting in a partially daily pattern forecasting structure that uses LSTM-RNN models to advance temporal dependency.

13. Rana et al. Provide a review on multi-step ahead solar photovoltaic power forecasting by using data resampling and univariate machine learning techniques.

14. Cheung et al. Propose deep convolutional neural network with meta-learning for day-ahead PV power forecasting.