❒ 2233

Performance analysis of demand forecasting in energy consumption based on ensemble model

Dhanalakshmi Jaganathan[1], Ayyanathan Natarajan[2]
[1]Department of Computer Science and Engineering, B. S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India
[2]Department of Computer Applications, B. S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India

Article Info

ABSTRACT

Over the previous decade, energy usage has increased exponentially all over the world. The machine learning algorithms are used to classify the demand and requirement of off and evening peak load of southern regional load dispatch centre (SRLDC) data. In this paper, data are classified based on demand and requirement of both evening and off peak of day wise southern regional grid of Andhra Pradesh, Karnataka, Kerala, Tamilnadu, and Pondicherry of different states are proposed. The machine learning algorithms like k-nearest neighbors (KNN), random forest, and logistic regression have been adopted to classify the model. To improve this model efficiency, an ensemble learning method is used to increase the accuracy. The performance measure of state-wise outcome is determined by classifying its demand and requirement needs over its state energy consumption and with different classification algorithms and it is improved by using a combined method of ensemble model with accuracy of 86%.

Corresponding Author:

Dhanalakshmi Jaganathan
Department of Computer Science and Engineering, B. S. Abdur Rahman Crescent Institute of Science and Technology
Vandalur, 600048, India
Email: dhanamedha@gmail.com

1. INTRODUCTION

The big data strategy is used in various technologies for implementing the large volume of time series forecasting. These insights are used for energy usage patterns, demand forecasting, and optimizing the energy usage in the time series streams. This will prevent outages from smart meters for data driven decision making system for power system. The consumer behavior is based on seasons and weather for influencing the energy consumption. Therefore, the design models are capable for analyzing the time series from smart meters for energy consumption [1]. The energy forecasting model is one of the components used for optimizing the energy model. This building energy model is used for predicting the energy consumption for weather conditions. A number of approaches are explained for demand forecasting to precise energy model. Three approaches are classified for predicting the building energy model for qualifying the work of smart meters [2].

The most challenges for energy utilities for organization are based on voltage security. The voltage level occurs in energy systems for power transmission for heavily loaded energy model. This power transmission is reacted over the voltage stability for shortage of reactive energy model in power system [3]. The energy efficiency of buildings are used more in European parliament and council for cost effective in energy savings. As a result, the energy efficiency is increased in day to day life for securing the public

buildings and national government. The energy consumption data are loaded in web application for accessing the energy efficiency in power transmission [4].

The renewable energy is used in various global locations for transmitting the several energies in power system. It requires the conventional energy for developing the energy systems throughout the world for applying the time series demand data for developing the power emissions in solar energy [5]. This solar energy review is done in various research and development for analyzing the regional innovation for investigating the global patent of energy analyses [6]. Energy consumption is used in different sources like renewable, fossil fuels, and nuclear resources for tracking the consumption of energy in different area or locations. This usage is rapidly increasing in industrial process for developing the country in power transmission. Therefore the energy is used in all over the country for developing the energy model in various sectors [7].

The demand data are used in time series energy forecasting for manipulating the data driven model for transmitting the energy data into a regional characteristics for smart grid. This predictive model utilizes the energy consumption for demand side data for improving the energy usage in power grid system. This energy utility is providing the sufficient data for increasing the forecast model for real time changes on both demand and supply for energy data [8]. This characteristic is used in different countries for developing the power conditions and the production of energy consumption. Furthermore the power companies are used and maintained the weather conditions for calculating the amount of energy for satisfying the customer in energy production. The power companies are used in advanced for predicting the energy data for electricity production in power consumption [9].

This approach is used in several sources in sensors and the internet for collecting the data to predict the energy model for decision making. Based on this model, the power transmission will be high in implementing the computing operations in power usage. It will easily rectify the power transmission for predicting the future results in various applications for power operation [10]. The power transmission is carried out in various energy sectors for calculating the time series demand for energy usage in both supplies as well as in demand. The energy usage is increased in several areas for averaging the demand in various streams in power grid system [11]. The ensemble learning method is used to co-operate the combination of various algorithms for implementing the new input patterns for predicting the higher accuracy measurement of different models. This model represents the generative algorithm to improve the accuracy of base learners [12].

The working model is done by using clustering algorithms for forecasting the energy usage of peak and non-peak hours of demand. It focuses on determining the consumption of demand data for smart meters applications [12]. The energy model studies the investigation of energy prediction in building energy model. The artificial neural network (ANN) algorithm is proposed for predicting the energy consumption for computing the energy data driven model. It tolerates the energy transmission for classifying the control and diagnosis of energy domain [13].

The voltage measurement is used for power system parameters for calculating the stability of variable measurement of demand data. This measurement can be used with computational technique for calculating the time series streaming data for demand met in peak hours. This measurement is done by day wise calculation of voltage stability of various countries for deploying the power system in energy consumption [14]. This paper examines the pre-processing methods for intelligent system for power quality distribution. It is possible for implementing the several algorithms for determining the energy needs in public as well as in private sectors. The predictive energy model is used in convenience of energy consumption in proposed architecture [4].

An iterative approach is implemented for developing the energy efficiency of various energy domains for solar energy. This photovoltaic infrastructure is used in large scale utilization of energy needs in various sectors. The physical principles are used in power distribution for determining the various power quality machines. This infrastructure is increasing on energy precise for demand needs [15]. The heating, ventilation and air conditioning (HVAC) model is determined for energy consumption of energy needs for various power sectors for both private and public consumption. The neural network algorithm is implemented for calculating the needs of demand in both peak hours of energy consumption [16].

The clustering methods are used in this study for deploying the power consumption of energy infrastructure in residential buildings. The swarm optimization technique is deployed for calculating the energy in various energy needs. This calculation is done and verified by using machine learning algorithms [7]. The weather data is used for pre-processing the needs of day and off peak of energy demand in power system. These data are further compared with random forest and deep learning algorithms for weather demand forecast [8]. The smart meters of different areas are used for data driven modeling for power harvesting of energy needs. These energy consumption of data are used in power quality distribution in European council power development of demand data [9]. The classification of building energy model is

used for developing the various energy demand and requirement needs of various forecasting methods by implementing the random forest (RF) algorithm. This approach is done using deep learning algorithms for deploying the quality of power maintenance in national government [11].

The energy data is maintained in data warehouse for calculating the energy efficiency of predictive modeling of power synchronization. This data can be easily maintained in a database for building the energy model. The energy utilization is done by using the ANN algorithms for maintaining the power consumption in various energy sectors [13], [14]. Both supervised and unsupervised algorithms are used for deploying the power maintenance for various distributions of demand and requirement needs of power quality problems. It assures the predictive model for analyzing the various needs of energy [17].

The correlation coefficient of statistical database is distributed in energy needs for calculating the mean and median of and demand data needs. This study examines the relationship of energy demand in various power quality distributions. This technique is used in various statistical methods for developing the requirement of energy distribution [15]. The big data strategy of IoT devices are used in power distribution for supporting the power quality problems in demand [16].

. The energy efficiency of demand side infrastructure is used for calculating the energy demand in energy management system [18]. The study describes about the energy consumption of demand data for calculating the data driven for decision making for power transmission analysis [19]. The architecture used in this study is unified process model for developing the big data in internet of thing (IoT) [20]. In addition, the anomaly detection of power consumption is calculated for energy consumption for predictive big data modeling. Thus the infrastructure behind the energy needs is measured base on the demand and its requirements [21].

The meteorological power distribution parameters are also used for measuring the energy demand prediction for data modeling techniques. The deep learning algorithms are used for implementing the demand of energy consumption of data [22]. To predict the renewable and non-renewable of power energy sources is used for measuring the power quality problems in energy sectors [23].

The algorithms are implemented and root mean square has been calculated for measuring the demand needs in power generation. The application is used for determining the requirement of power distribution in various energy distributions [24], [25]. The ANN model is used for predicting the power consumption for quick estimating the energy needs in both demand and requirement of energy variables. Thus the energy model has been determined for transmitting the energy needs in power distribution system [26].

The non-parametric model is proposed in this study for analyzing the new instances of classification model. KNN model is used for predicting the power quality problems in residential buildings. The hourly consumption of data is used for testing the power problems to increase the impact of energy consumption in better way [27]. The Euclidean distance is calculated for power load data from national electricity market. The k-nearest data points are measured for reducing the regulations in electricity. KNN model is achieved and the algorithms produce the good classification for fitting the energy variation [28]. The electrical grid in power transmission is used in power distribution of demand side of energy data. This energy data is implemented in various energy generation for calculating the load forecasting of streaming power distribution data [29]. The energy management system is used for transmitting the need of energy for increasing the energy needs in various energy purposes. Therefore, the development of power transmission will be increased in day today life [30]. According to the study, machine learning algorithms are used for classifying the demand prediction of power transmission in physical model [31].

The main objective is to demonstrate the various classification algorithms for classifying the energy demand and requirement for both evening and off peak of energy consumption. The proposed algorithm is used for classifying the needs using random forest, KNN, and logistic regression. The proposed workflow model of data is obtained from southern regional dispatch centre (SRLDC) of demand met and its requirement of various states of southern region. The classification algorithm is implemented in Google Colab using python. The main focus is to demonstrate the energy consumption data to calculate it's demand and requirement of overall energy during evening and off peak.

2.   METHODS
2.1. Dataset description

The data set was obtained from the web sources of SRLDC. This operational hub, which screens the conditions of Karnataka, Tamil Nadu, Andhra Pradesh, Telangana, Kerala, Pudhuchery, and covering the central segment, state sector and independent power is delivering stations, an all-out zone of 6,51000 sq.kms. It is one of the five regional load dispatch centers, working under the national load dispatch center. National load despatch centre (NLDC) is claimed, worked and kept up by power system activity corporation of india limited (POSOCO), an auxiliary organization of power grid corporation of India limited [32], [33]. The data

file consists of 2124 instances under the variable names demand met in the evening peak, off-peak, and requirement in evening and off-peak. The data are mainly worked under the state demand met of day wise parameters like demand met at evening peak, requirement at evening peak, demand met at off peak, and requirement at off peak of day energy. Based on this calculation, the values are collected from the SRLDC. All the values for these particular variables are taken in real time. Each and every variable of demand met data is described as shown in. Designing the machine learning algorithms based on energy productivity of demand and requirements, the following algorithms are described as shown in Table 1.

Table 1. Description of various energy procedures based on demand and requirement

Group of attributes	Attribute name	Descriptive statistics
Demand met at evening peak	Demand of overall energy during evening.	Numerical data based on time series demand.
Requirement at evening peak	Requirement of energy during evening peak	Numerical data based on time series requirement–unstructured data.
Demand met at off-peak	Demand met during off peak.	Numerical data based on time series at off-peak demand.
Requirement at off-peak	Requirement met during off peak.	Numerical data based on time series at off-peak requirement-unstructured data.

2.2. Logistic regression

Logistic regression is used for calculating the discrete set of classes. Binary model and multi linear functions are the two models used mainly for logistic regression. It can easily solve the classification problems to a predictive analysis based on the concept of probability [34]. The cost function can be analysed using sigmoid function. The hypothesis of logistic regression tends to a limit for determining the cost function between the value of 0 and 1.

$$0 \leq h_{(\theta)}(y) \leq 1 \tag{1}$$

In order to map the probability function for a predicted values, the function maps of a real value data to calculate another limits of 0 and 1 [35]. From the limits, the regression value is used for hypothesis testing for linear regression formula is:

$$h \ominus (y) = \beta_0 + \beta_1(y) \tag{2}$$

Whereas for the logistic regression, the little bit modification of formula will be to calculate the hypothesis is,

$$\sigma(x) = \sigma(\beta_0 + \beta_1 X) \tag{3}$$

Now the expected values of hypothesis will give values between 0 and 1.

$$Z = \beta_0 + \beta_1(y) \tag{4}$$

$$h \ominus (y) = Sigmoid(X) \tag{5}$$

$$(i.e) \ h \ominus (y) = \frac{1}{1+e^{-(\beta_0+\beta_1(y))}} \tag{6}$$

From the above result, the decision boundary has been calculated for threshold of returning a probability score between 0 and 1 [36]. Finally, the cost function represents the optimization of convex function to minimize the cost value and finding the global minimum of logistic regression.

2.3. Random forest

Random forests are a directed learning calculation. It might be used both for portrayal and backslide. It is moreover the most versatile and easy to use calculation [37]. Random forests suit the choice trees on randomly picked data tests, get estimate from each tree and pick the best course of action of characterization model shown in Figure 1. It moreover gives a respectable marker of the component importance. The algorithm carried out in the below steps such as:
Step 1: Select random examples from a given dataset.
Step 2: Develop a decision tree for each example and get a forecast result from every choice tree.

Step 3: It performs voting for a predicted outcome.
Step 4: Select the forecast result with the most votes as the final prediction.

Random forests are considered as a profoundly exact and strong technique in light of the quantity of decision tree model [38]. The fundamental clarification is that it takes the typical of the general huge number of figures, which balances the steady factors. There are two distinct approaches to manage these: using centre characteristics to replace consistent factors, and handling the closeness weighted normal of missing qualities.
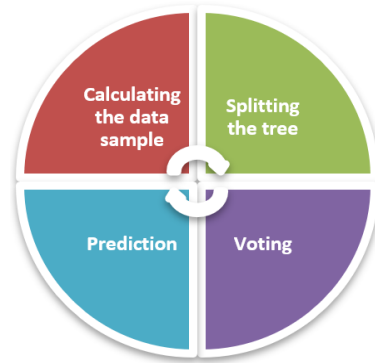

Figure 1. RF classification model

The importance of random tree of each feature selection is divided into the smaller subsets such as,

$$fsi_i = \sum j = \frac{\text{node j splits on feature selection i } n_j^i}{\sum k \,\epsilon \text{ all nodes } n_k^i} \tag{7}$$

where feature selection of sub (i) belongs to the feature of i.Node sub (j) belongs to the importance of node j. for each random forest iteration the binary tree is calculated the value between 0 and 1 by dividing the feature selection of values,

$$normfsi_i = \frac{fsi_i}{\sum j \,\epsilon \text{ all features } f_{ij}} \tag{8}$$

The sum of feature importance of random tree model is divided by the total number of trees,

$$RFfi_i = \frac{\sum j \,\epsilon \text{ all trees norm } fi_{ij}}{T} \tag{9}$$

where $RFfi_i$ is calculated from all trees in random forest model. T represents the total number of trees, norm fi_{ij} belongs to the feature importance for i in tree j.

2.4.. KNN classifier

KNN is a supervised machine learning algorithm which can be used for both regression and classification analysis [39]. KNN is used for implementing the feature similarity to predict the value for new data point. It can also match the training dataset for analysing state assumptions [40]. Working procedure for KNN:
Step 1: First step is to work KNN algorithm is to loading the data.
Step 2: After loading initialise the value of K in chosen neighbours
Step 3: Now calculate the distance of data points for each K Neighbours
Step 4: From the distance calculations add the indices from K entries
Step 5: Get the variable of selected K entries.
Step 6: Next chose the top K list of an element.
Step 7: finally assign the test points based on the selected variables.

3. RESULTS AND DISCUSSION

The input data is segregated based on Table 2. The state demand met data is wrapped from online SRLDC report. The required data is demand and requirement of evening peak and off peak. In this model, the

data set is divided into four sub data sets which are demand based on evening and off peak, requirement based on evening and off peak, evening peak based on demand and requirement, and off-peak based on demand and requirement. In this classification, the target variable and class variables are divided based on the fields. By analysing the classification accuracy, the ensemble model is proved to shows the best accuracy in energy model [41], [42].

Table 2. Generation of dataset

Demand		Requirement		Off		Evening	
Numeric Data	Class	Numeric Data	Class	Numeric Data	Class	Numeric Data	Class
Evening	0	Evening	0	Demand	0	Demand	0
Off	1	Off	1	Requirement	1	Requirement	1

The main task after classification is to identify how far, the classification algorithm works on different set of same data. Initially, the classification starts with evening peak on demand and requirement followed with off peak on demand and requirement. Finally, it ends with requirement and demand of evening and off peak [43]. The class values are rearranged based on categorical data. In evening peak, the class value of demand is set as o and the requirement is set as 1 and the target value is evening peak. In off peak data, the class value of demand is set as 0 and the requirement is set as 1 and the target value is off peak. In requirement data, the class value of evening is set as 0 and the off value is set as 1 and the target value is requirement. In demand data, the class value of evening is set as 0 and the off value is set as 1 and the target value is demand.

Once the performance is classified using both demand and requirement next parameter is to classify the evening peak and off peak. The evening peak is set as a target variable the class fields are classified using the three algorithms and the classification accuracy gives low performance in evening peak in KNN as 34%, random forest gives 27% and logistic regression gives 50% of accuracy. Once the algorithms are implemented, the ensemble model is performed and it gives 35% of accuracy in evening peak. Next, the off peak is classified using the same algorithms and gives the classification accuracy of 40% in KNN, 37% in random forest, 50% in logistic regression. By combining these three algorithms in ensemble model and it shows the accuracy as 40% in off peak. The Table 3 shows the evening and off peak of accuracy result of energy data. Figure 2 explains the contribution of each classification algorithm based on evening and off peak.

Table 3. Accuracy rate of evening and off peak

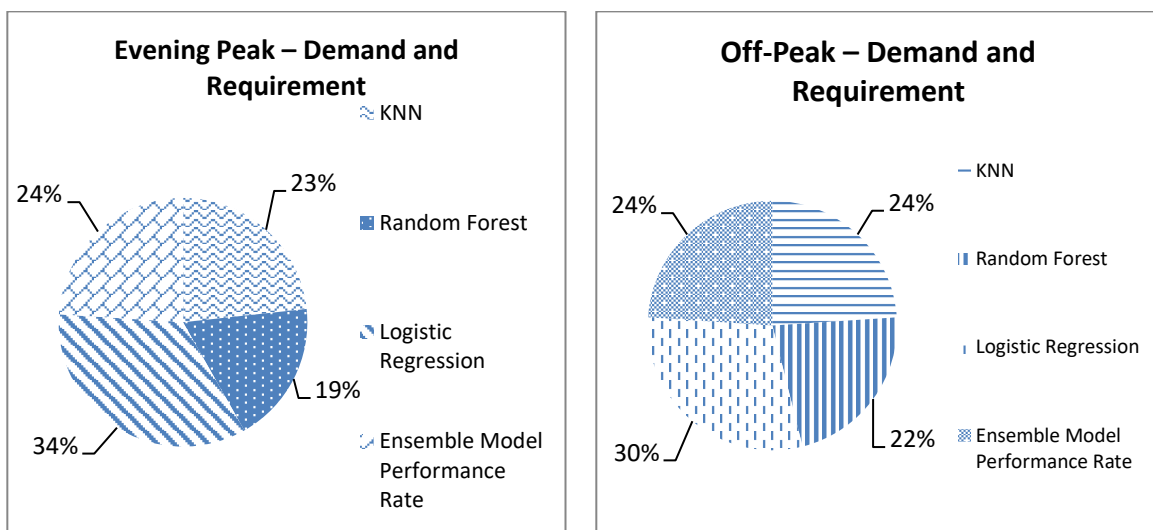| Target Data	Algorithm Used			Ensemble Model Performance Rate
	KNN	Random Forest	Logistic Regression	
Evening Peak – Demand and Requirement	34%	27%	50%	35%
Off-Peak – Demand and Requirement	40%	37%	50%	40%

Figure 2. Contribution of each classification algorithm based on evening and off peak

After performing the demand data, next data is classifying using requirement as a target variable. The classification algorithm performs and gives the accuracy rate of 74% in KNN, 69% in random forest and 50% in logistic regression. From this classification accuracy, ensemble model is performed and gives the accuracy rate of 75%. The Table 4 describes the output result of requirement data and demand data. The Figure 3 shows the Contribution of each classification algorithm based on requirement data and demand data.

Table 4. Accuracy rate of requirement data and demand data

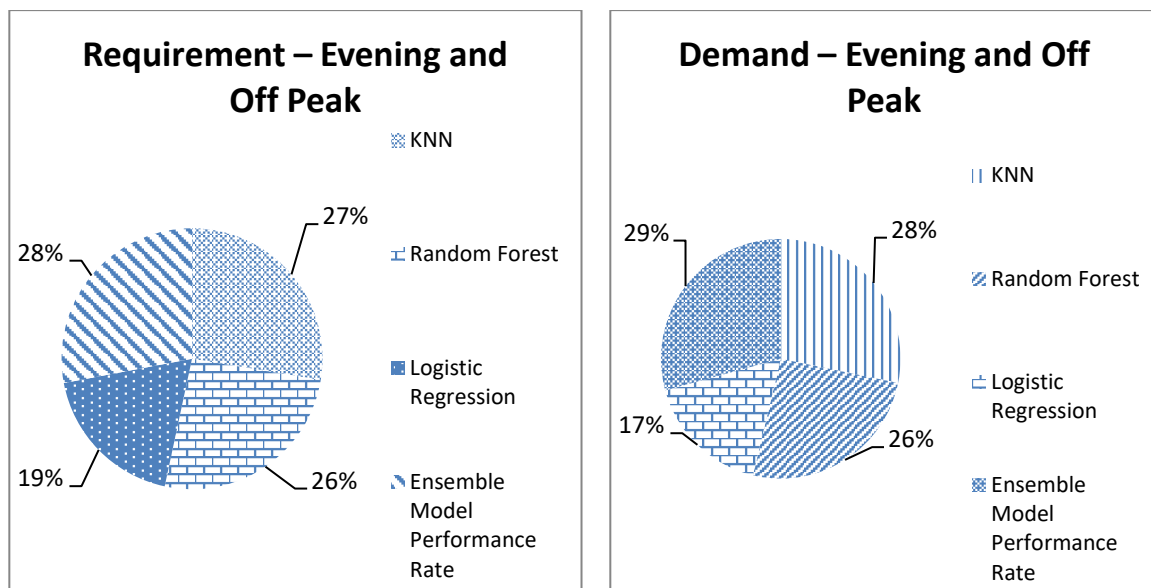| Target Data | Algorithm Used | | | Ensemble Model Performance Rate |
	KNN	Random Forest	Logistic Regression	
Requirement – Evening and Off Peak	74%	69%	50%	75%
Demand – Evening and Off Peak	83%	75%	50%	86%

Figure 3. Contribution of each classification algorithm based on requirement data and demand data

First, from the demand data classification, the KNN algorithm is analysed and gives the accuracy rate of 83%. Next, the data are classified using logistic regression and random forest. Random forest gives the accuracy rate of 75% whereas the logistic regression gives the 50% of classification accuracy. From these classification rates, the data are ensembled by combining the KNN, RF, and LR and give the new instances of accuracy is 86% in ensemble model.

From the overall performance, the data are classified with four parameters. By showing of the result, logistic regression gives equal output for demand, requirement, evening and off peak whereas the random forest is low and high based on the data and KNN shows the good accuracy in all energy parameters than the logistic regression and random forest. From this classification accuracy, the ensemble model is performed and gives the good precision in demand data. In the evening and off peak the algorithm as well as ensemble model gives low accuracy. So the demand data is very high and performs the good accuracy of 80% in ensemble modelling using KNN, logistic regression and random forest. The Table 5 represents the overall performance of ensemble modelling classification. Figure 4 shows the complete graphical analysis of all classification based on input dataset.

Table 5. Accuracy value of all classification algorithms based on each dataset

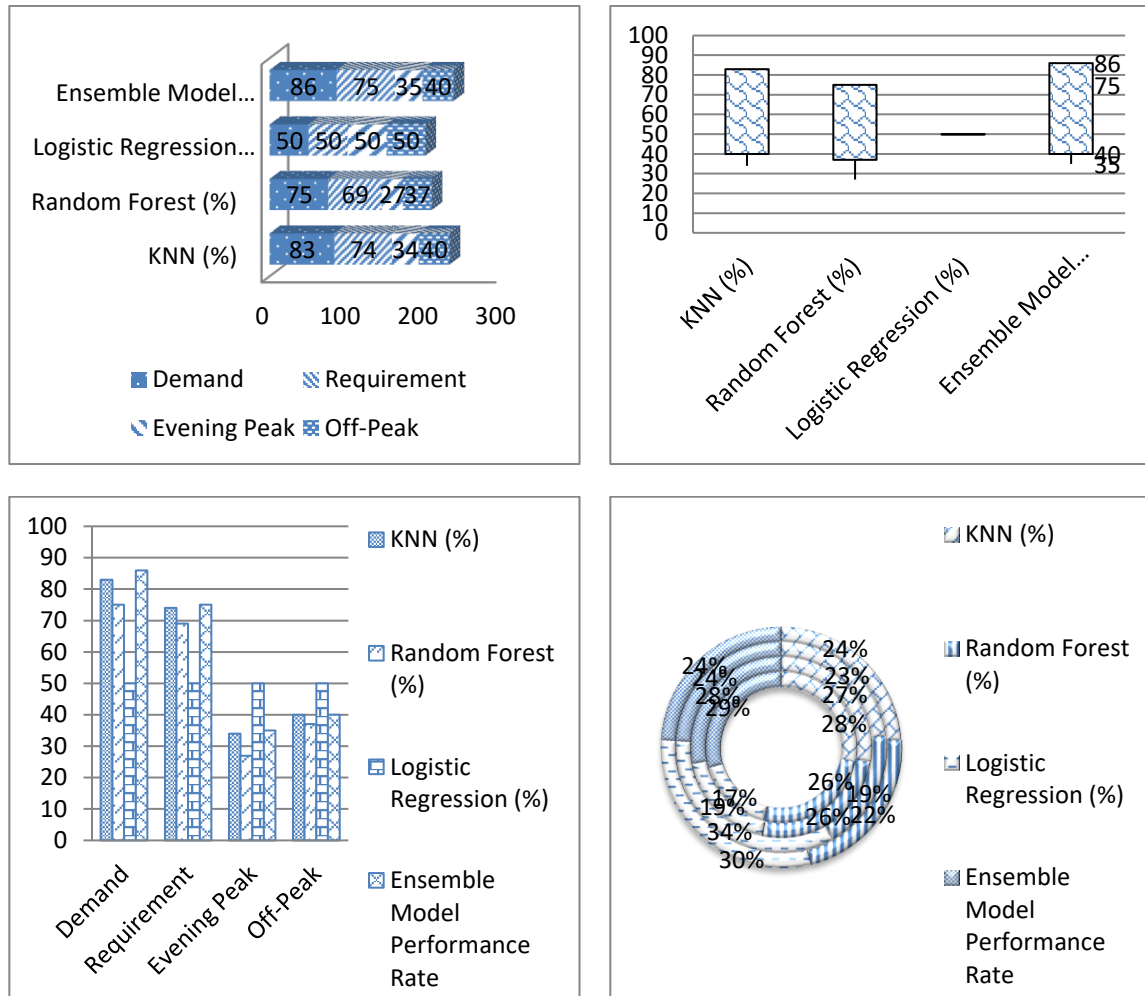| Target Data | Algorithm Used | | | Ensemble Model Performance Rate |
	KNN (%)	Random Forest (%)	Logistic Regression (%)	
Demand	83	75	50	86
Requirement	74	69	50	75
Evening Peak	34	27	50	35
Off-Peak	40	37	50	40

Figure 4. Complete graphical analysis of all classification based on input dataset

4. CONCLUSION

This paper deals with the classification of energy consumption of both demand and requirement data based on evening and off peak of Southern Indian Power Distribution System. The machine learning algorithms are used on time series data for classifying the energy needs in evening and off peak hours based on demand and requirement. The algorithms KNN, Random Forest and Logistic Regression are used to classify the data. Furthermore, classified data is taken to ensemble model to improve the accuracy of classification. The results show that the accuracy is achieved by ensemble model which produces 86% than the remaining machine learning algorithms. The benefits levels of energy consumption data is achieved more in demand based on off peak and evening peak.

REFERENCES

[1] L. Vanfretti and V. S. N. Arava, "Decision tree-based classification of multiple operating conditions for power system voltage stability assessment," *International Journal of Electrical Power & Energy Systems*, vol. 123, p. 106251, Dec. 2020, doi: 10.1016/j.ijepes.2020.106251.
[2] M. Zekic-Susac, S. Mitrovic, and Adela Has, "Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities," *International Journal of Information Management*, vol. 58, p. 102074, Jun. 2021, doi: 10.1016/j.ijinfomgt.2020.102074.
[3] D. Harwalkar, S. P. Gahana, T. Pilakkal, and S. G. Sreeganesan, "Analytical Study of Correlation Between Demand and Renewable Energy Forecasting Using Data Mining/Analytics," *Journal of Emerging Technologies and Innovative Research*, vol. 7, pp. 72-78, 2020.
[4] P. W. Khan, Yung-Cheol Byun, Sang-Joon Lee, Dong-Ho Kang, Jin-Young Kang, and Hae-Su Park, "Machine Learning-Based Approach to Predict Energy Consumption of Renewable and Nonrenewable Power Sources," *Energies*, vol. 13, no. 18, p. 4870, Sep. 2020, doi: 10.3390/en13184870.

[5]   H. Hermansah, D. Rosadi, A. Abdurakhman, and H. Utami, "Automatic time series forecasting using nonlinear autoregressive neural network model with exogenous input," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 5, pp. 2836-2844, Oct. 2021, doi: 10.11591/eei.v10i5.2862.

[6]   A. P. Chakkaravarthy, R. pugalendhi, R. Jayaraman, and J. Dhanalakshmi, "An automatic Classification of Covid with J48 and Simple K Means using Weka," *International Journal of Future Generation Communication and Networking*, vol. 13, no. 3, pp. 490-500, 2020.

[7]   E. García-Martín, C. F. Rodrigues, G. Riley, and Håkan Grahn, "Estimation of energy consumption in machine learning," in *Journal of Parallel and Distributed Computing*, vol. 134, pp. 75-88, Dec. 2019, doi: 10.1016/j.jpdc.2019.07.007.

[8]   C. E. Kontokosta and C. Tull, "A data-driven predictive model of city-scale energy use in buildings," *Applied Energy*, vol. 197, pp. 303-317, Jul. 2017, doi: 10.1016/j.apenergy.2017.04.005.

[9]   A. J. C. Trappey, P. P. J. Chen, C. V. Trappey, and L. Ma, "A Machine Learning Approach for Solar Power Technology Review and Patent Evolution Analysis," *Applied Sciences*, vol. 9, no. 7, pp. 1-25, Apr. 2019, doi: 10.3390/app9071478.

[10] M. K. M. Shapi, N. A. Ramli, and L. J. Awalin, "Energy consumption prediction by using machine learning for smart building: Case study in Malaysia," *Developments in the Built Environment*, vol. 5, p. 100037, Mar. 2021, doi: 10.1016/j.dibe.2020.100037.

[11] S. Sohangir, D. Wang, A. Pomeranets, and T. M. Khoshgoftaar, "Big Data: Deep Learning for financial sentiment analysis," *Journal of Big Data*, vol. 5, no. 1, pp. 1-25, 2018, doi: 10.1186/s40537-017-0111-6.

[12] A. J. del Real, F. Dorado, and J. Duran, "Energy Demand Forecasting Using Deep Learning: Applications for the French Grid," *Energies*, vol. 13, p. 2242, May 2020, doi: 10.3390/en13092242.

[13] H. Cai, S. Shen, Q. Lin, X. Li, and H. Xiao, "Predicting the Energy Consumption of Residential Buildings for Regional Electricity Supply-Side and Demand-Side Management," in *IEEE Access*, vol. 7, pp. 30386-30397, 2019, doi: 10.1109/ACCESS.2019.2901257.

[14] B. Carrera and K. Kim, "Comparison Analysis of Machine Learning Techniques for Photovoltaic Prediction Using Weather Sensor Data," *Sensors*, vol. 20, no. 11, pp. 1-14, 2020, doi: 10.3390/s20113129.

[15] M. Bourdeau, X. Q. Zhai, E. Nefzaoui, X. Guo, and P. Chatellier, "Modeling and forecasting building energy consumption: A review of data-driven techniques," *Sustainable Cities and Society*, vol. 48, p. 101533, 2019, doi: 10.1016/j.scs.2019.101533.

[16] Seung-Jun Shin, J. Woo, S. Rachuri, and W. Seo, "An energy-efficient process planning system using machine-monitoring data: A data analytics approach," *Computer-Aided Design*, vol. 110, pp. 92-109, May 2019, doi: 10.1016/j.cad.2018.12.009.

[17] H. Habizadeh, C. Kaptan, T. Soyata, B. Kantarci, and A. Boukerche, "Smart City System Design: A Comprehensive Study of the Application and Data Planes," *ACM Computing Surveys*, vol. 52, no. 2, pp. 1-38, 2019, doi: 10.1145/3309545.

[18] T. Anagnostopoulos, G. L. Kyriakopoulos, S. Ntanos, E. Gkika, and S. Asonitou, "Intelligent Predictive Analytics for Sustainable Business Investment in Renewable Energy Sources," *Sustainability*, vol. 12, no. 7, p. 2817, 2020, doi: 10.3390/su12072817.

[19] S. Singh and A. Yassine, "IoT Big Data Analytics with Fog Computing for Household Energy Management in Smart Grids," *International Conference on Smart Grid and Internet of Things*, vol. 256, pp. 13-22, Dec. 2018, doi: 10.1007/978-3-030-05928-6_2.

[20] U. Ali, A. Rauf, U. Iqbal, and I. A. Shoukat, "Big Data Analytics for a Novel Electrical Load Forecasting Technique," *International Journal on Information Technologies & Security*, vol. 11, no. 3, pp. 33-40, 2019.

[21] D. A. C. Narciso and F. G. Martins, "Application of machine learning tools for energy efficiency in industry: A review," *Energy Reports*, vol. 6, pp. 1181-1199, Nov. 2020, doi: 10.1016/j.egyr.2020.04.035.

[22] Y. Himeur, A. Alsalemi, F. Bensaali, and A. Amira, "A Novel Approach for Detecting Anomalous Energy Consumption Based on Micro-Moments and Deep Neural Networks," *Cognitive Computation*, vol. 12, pp. 1381–1401, Sep. 2020, doi:10.1007/s12559-020-09764-y

[23] J. Dhanalakshmi and N. Ayyanathan, "An Implementation of Energy Demand Forecast using J48 and Simple K Means," *2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, 2019, pp. 174-178, doi: 10.1109/ICONSTEM.2019.8918883.

[24] J. Dhanalakshmi, N. Ayyanathan, and N. S. Pandian, "Energy Analytics and Comparative Performance Analysis of Machine Learning Classifiers on Power Boiler Dataset," *2019 11th International Conference on Advanced Computing (ICoAC)*, 2019, pp. 72-78, doi: 10.1109/ICoAC48765.2019.246819.

[25] T. A. Naqishbandi and N. Ayyanathan, "Clinical Big Data Predictive Analytics Transforming Healthcare: -An Integrated Framework for Promise Towards Value Based Healthcare," *Advances in Decision Sciences, Image Processing, Security and Computer Vision. Learning and Analytics in Intelligent Systems*, vol. 4, pp. 545-561, Jul. 2019, doi: 10.1007/978-3-030-24318-0_64.

[26] S. Singh and A. Yassine, "Big Data Mining of Energy Time Series for Behavioral Analytics and Energy Consumption Forecasting," *Energies*, vol. 11, no. 2, p. 452, 2018, doi: 10.3390/en11020452.

[27] F. Wahid and D. H. Kim, "A Prediction Approach for Demand Analysis of Energy Consumption Using K-Nearest Neighbor in Residential Buildings," *International Journal of Smart Home*, vol. 10, no. 2, pp. 97-108, 2016, doi: 10.14257/ijsh.2016.10.2.10.

[28] Guo-Feng Fan, Yan-Hui Guo, Jia-Mei Zheng, and Wei-Chiang Hong, "Application of the Weighted K-Nearest Neighbor Algorithm for Short-Term Load Forecasting," *Energies*, vol. 12, no. 5, pp. 1-19, Mar. 2019, doi: 10.3390/en12050916.

[29] F. Wahid, R. Ghazali, A. S. Shah, and M. Fayaz, "Prediction of Energy Consumption in the Buildings Using MultiLayer Perceptron and Random Forest," *International Journal of Advanced Science and Technology*, vol. 101, pp.13-22, 2017, doi: 10.14257/ijast.2017.101.02.

[30] M. W. Ahmad, M. Mourshed, and Y. Rezgui, "Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption," *Energy and Buildings*, vol. 147, pp. 77-89, Jul. 2017, doi: 10.1016/j.enbuild.2017.04.038.

[31] N. Ayyanathan, A. Shilpa, A. B. Rekha, and P. Raj, "Comparative Performance Analytics of Machine Learning Classifiers in the Countrywise Foreign Tourists arrival pattern and Demand Modeling of Indian Tourism*," 2016 IEEE Eight International Conference on Advanced Computing*, 2016.

[32] J. Dhanalakshmi and N. Ayyanathan, "State Wise Modeling and Descriptive Analytics with Data Visualization of Indian Southern Power Grid Operation," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 10, pp. 5931-5944, 2021, doi: 10.17762/turcomat.v12i10.5414.

[33] J. Dhanalakshmi and N. Ayyanathan, "A systematic review of big data in energy analytics using energy computing techniques," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 4, p. e6647, 2021, doi: 10.1002/cpe.6647.

[34] T. T. Nguyen, T. T. Ngoc, T. T. Nguyen, Thanh-Phuc Nguyen, and N. A. Nguyen, "Optimization of location and size of distributed generations for maximizing their capacity and minimizing power loss of distribution system based on cuckoo search algorithm," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 4, pp. 1769-1776, Aug. 2021, doi: 10.11591/eei.v10i4.2278.

[35] A. Y. Ahmed, M. A. Kahya, and S. A. Altamir, "Classification improvement of gene expression for bipolar disorder using weighted sparse logistic regression," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 2, pp. 1062-1068, Apr. 2022, DOI: 10.11591/eei.v11i2.3594.

[36] A. I. Abdurrahman and A. Zahra, "Spoken language identification using i-vectors, x-vectors, PLDA and logistic regression," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 4, pp. 2237-2244, Aug. 2021, doi: 10.11591/eei.v10i4.2893.

[37] A. Y. Dawod and M. A. Sharafuddin, "Assessing mangrove deforestation using pixel-based image: a machine learning approach," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no .6, pp. 3178-3190, Dec. 2021, doi: 10.11591/eei.v10i6.3199.

[38] M. A. Rasyidi, T. Bariyah, Y. I. Riskajaya, and A. D. Septyani, "Classification of handwritten Javanese script using random forest algorithm," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 3, pp. 1308-1315, Jun. 2021, DOI: 10.11591/eei.v10i3.3036.

[39] E. B. B. Palad, M. J. F. Burden, C. R. D. Torre, and R. B. C. Uy, "Performance evaluation of decision tree classification algorithms using fraud datasets," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 6, pp. 2518-2525, Dec. 2020, doi: 10.11591/eei.v9i6.2630.

[40] U. A. Khan, S. M. U. Din, S. A. Lashari, M. A. Saare, and M. Ilyas, "Cowbree: A novel dataset for fine-grained visual categorization," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 5, pp. 1882-1889, 2020, doi: 10.11591/eei.v9i5.2443.

[41] S. M. C. M. Nor, S. M. Shaharudin, S. Ismail, N. H. Zainuddin, and M. L. Tan, "A comparative study of different imputation methods for daily rainfall data in east-coast Peninsular Malaysia," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 2, pp. 635-643, Apr. 2020, doi: 10.11591/eei.v9i2.2090.

[42] D. K. Hashim and L. A. N. Muhammed, "Performance of K-means algorithm based an ensemble learning," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 1, pp. 575-580, Feb. 2022, doi: 10.11591/eei.v11i1.3550.

[43] Duc-Duong Nguyen, Minh-Thuy Le, and Thanh-Long Cung, "Improving intrusion detection in SCADA systems using stacking ensemble of tree-based models," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 1, pp. 119-127, Feb. 2022, doi: 10.11591/eei.v11i1.3334.

BIOGRAPHIES OF AUTHORS

Dhanalakshmi Jaganathan received the B.E. degreein Computer Science and Engineering and the M.E. degree in Computer Science and Engineering from Anna University, Tamilnadu, India, in 2015 and 2017 respectively. She have secured 31st university rank holder in Anna University for Master Degree in Computer Science and Engineering. She is currently pursuing the Ph.D. degree in the School of Computer Science and Engieering, B. S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India. Since 2019, she has been a full-time Research Scholar. She can be contacted at email: dhanamedha@gmail.com.

Ayyanathan Natarajan received the Ph.D. degree from Anna University, in 2015. He has 27 Yrs of experience in Teaching. He is currently working as the Professor with the School of Computer Information and Mathematical Sciences in B. S. Abdur Rahman Crescent Institute of Science and Technology. He has ten research articles in various journals, including Scopus and web of science, and has presented seventeen articles in several international proceedings and conferences. He has guided several undergraduate and postgraduate students and is also guiding four Ph.D. research scholars. Additionally, he has conducted and organized several workshops and conferences. His research interests include computational business intelligence, big data, energy analytics, data analytics, and mathematical modeling. He can be contacted at email: n.ayyanathan@crescent.education.