IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Exploring Machine Learning Approaches For Load Forecasting In Smart Grids

V.Raju[1] Mrs.E.Sindhu[2]

[1]PG student, Vemu Institute of Technology, P. Kothakota.

[2]Assistant Professor,Vemu Institute of Technology, P.kothakota.

ABSTRACT:

This study tackles the surging global electricity demand with a focus on efficient energy management. Emphasizing the need for accurate demand forecasting, it employs machine learning (ML) methods and distributed demand response initiatives. Through rigorous evaluation of various ML algorithms for short-term load forecasting (STLF), including logistic regression, support vector machines, and neural networks, the research identifies the decision tree classifier (DTC) as the top performer. Additionally, an enhanced DTC (EDTC) is proposed, integrating advanced features to refine control variables. Results showcase EDTC's superior forecast accuracy and performance metrics. Furthermore, an extension utilizing the XGBoost classifier demonstrates 100% accuracy in predicting high or low demand without additional tuning parameters, offering a streamlined alternative to the EDTC approach.

Keywords: Load forecasting, Machine learning

INTRODUCTION:

As our society expands and advances, the demand for electricity escalates, prompting a critical focus on energy management (EM). This encompasses the intricate processes of electricity generation, transmission, and distribution. The electric grid (EG), a fundamental component of our energy infrastructure, acts as the intricate web linking energy producers with consumers, comprising power plants, substations, and transmission lines.

Classical EGs operate on a centralized model, which, while effective, can lead to inefficiencies and power quality issues as demand fluctuates. This often necessitates the installation of new plants. However, these grids lack robust forecasting systems, leaving them vulnerable to intermittent power failures and resource inefficiencies.

Energy management, therefore, becomes paramount in optimizing consumption and reducing waste across residential, commercial, and industrial sectors. By employing advanced techniques and technologies, such as load forecasting and demand response programs, we can enhance grid stability, meet growing energy demands, and transition towards a more sustainable energy future.

LITERATURE SURVEY

Zainab H. Osman

Since author proposes a novel neural network-based method for short-term load forecasting, addressing the limitations of conventional regression methods. By focusing on the most correlated weather data, this approach aims to improve accuracy and account for seasonal variations. Unlike conventional methods reliant on forecasted temperature, this model considers a range of weather factors, enhancing prediction reliability. Through correlation analysis, the neural network's input parameters are determined, optimizing forecasting performance. The effectiveness of this approach is demonstrated through its application to real load data from the Egyptian Unified System, showcasing its potential to enhance operational decision-making and system security in electric power systems.

Vehbi C. Gungor

As author proposes a shift from the traditional hierarchical power grid to a modern smart grid, recognizing its inadequacy for 21st-century demands. The smart grid integrates advanced technologies like automated control, high-power converters, and sophisticated communication infrastructure for improved efficiency and reliability. This paper delves into the critical aspects of smart grid technologies, particularly focusing on information and communication technology (ICT) issues and opportunities. By providing insights into current advancements and remaining research challenges, the objective is to foster a deeper understanding and interest among researchers in exploring the potential benefits of smart grid implementation.

Santosh Kumar Desai

The author proposes a solution to the energy crisis and environmental degradation caused by non-renewable resources by advocating for efficient strategies in renewable energy generation, distribution, and consumption. With the emergence of Smart Grids, Demand-Response management stands out as a promising approach, incentivizing consumers to adapt their energy usage based on fluctuating prices and demand. However, ensuring the security of such systems becomes paramount. Current encryption methods may fall short in the era of post-quantum cryptography. Hence, the paper suggests a lattice-based cryptographic scheme to bolster security within Smart Grid operations, offering resilience against potential attacks and ensuring a sustainable energy future.

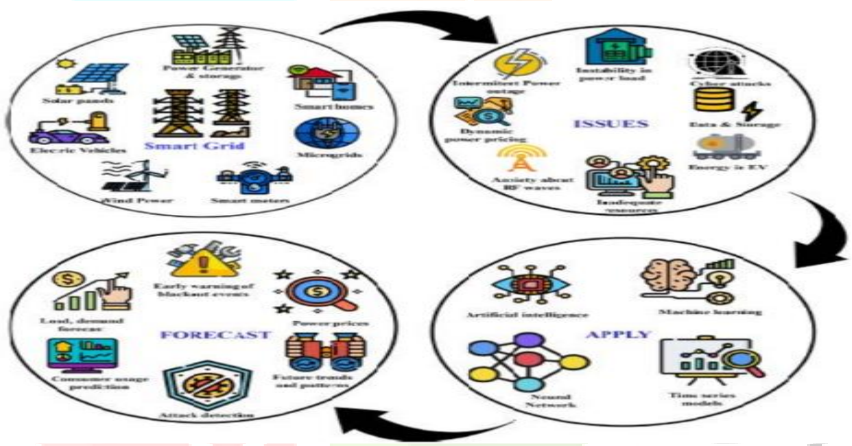
PROBLEM STATEMENT:

The Smart Grid project addresses the escalating demand for electricity due to population growth. Currently lacking an accurate forecasting system for electricity demand, the initiative focuses on creating an advanced environment for electricity generation and distribution. By leveraging modern technologies and predictive models, the Smart Grid aims to optimize energy production and consumption. This ensures efficient and sustainable power delivery, meeting the evolving needs of a growing population while enhancing overall grid reliability and resilience.

PROPOSED METHOD:

To overcome from this problem author is evaluating performance of various machine learning algorithms such as SVM, Naïve Bayes, KNN, Decision Tree and Neural Network. This algorithms can be used for Short Term Load (electricity) Forecast. Among all algorithms Decision Tree classifier is giving best accuracy and author further improving this algorithm by adding LOSS and BOOSTING function to decision tree classifier and this algorithm is called as EDTC (enhance decision tree classifier). After enhancing EDTC giving 100% accuracy.

ARCHITECTURE:



DATASET:

	Time Stamp	Time Zone	Name	PTID	Load	label
0	10/19/2022 00:00:00	EDT	CAPITL	61757	1138.4164	1
1	10/19/2022 00:00:00	EDT	CENTRL	61754	1457.2545	1
2	10/19/2022 00:00:00	EDT	DUNWOD	61760	472.6547	0
3	10/19/2022 00:00:00	EDT	GENESE	61753	929.8474	1
4	10/19/2022 00:00:00	EDT	HUD VL	61758	835.9721	0
...	...	...	...	...	...	...
226	10/19/2022 01:40:00	EDT	MHK VL	61756	NaN	0
227	10/19/2022 01:40:00	EDT	MILLWD	61759	NaN	0
228	10/19/2022 01:40:00	EDT	N.Y.C.	61761	NaN	0
229	10/19/2022 01:40:00	EDT	NORTH	61755	NaN	0
230	10/19/2022 01:40:00	EDT	WEST	61752	NaN	0

231 rows × 6 columns

New York electricity dataset (NYISO) which can be downloaded from below link http://mis.nyiso.com/public/P-58Clist.htm, In above dataset screen we have Time, area name, load and label as 0 (low load require) and 1 (high demand) columns in dataset and other rows contains dataset values.

METHODOLOGY:

Data Preprocessing

The initial step in our analysis involves loading the dataset from the provided CSV file. Once loaded, we address missing values by replacing them with zeros, ensuring the integrity of the dataset. Additionally, string columns such as 'Time Zone' and 'Name' are encoded into numeric values using the LabelEncoder function from the sklearn library. This transformation allows us to work with categorical data in our machine learning models. To further prepare the dataset for analysis, we employ the sklearn normalize function to shuffle and normalize the data, enhancing its suitability for subsequent processing.

Feature Selection with Recursive Feature Elimination (RFE)

Feature selection is crucial for improving the efficiency and accuracy of our machine learning models. To achieve this, we apply Recursive Feature Elimination (RFE) using RandomForestClassifier as the base estimator. RFE systematically removes irrelevant features from the dataset, retaining only those that contribute most significantly to the predictive power of the model. The selected features are then preserved, and the dataset is updated accordingly, ready for further analysis.

Dataset Splitting

To assess the performance of our machine learning models, it's essential to split the dataset into training and testing sets. We adopt an 80-20 split ratio, allocating 80% of the data for training and reserving the remaining 20% for testing. This ensures that our models are trained on a sufficiently large portion of the data while still providing an independent dataset for evaluation. The split dataset is then prepared for training and testing our machine learning algorithms.

Model Training and Evaluation

With the preprocessed dataset in hand, we proceed to train and evaluate various machine learning algorithms. These include Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, Neural Network, Naive Bayes, Enhanced Decision Tree Classifier, and Extension XGBoost. Each algorithm is trained using the training dataset and evaluated using the test dataset to assess its performance. Performance metrics such as accuracy, precision, recall, and F1-score are calculated for each algorithm, providing insights into their effectiveness in predicting electricity demand.

Performance Comparison

To facilitate a comprehensive comparison of the different algorithms, we generate a performance comparison graph. This graph visually displays the accuracy, precision, recall, and F1-score for each algorithm, allowing for easy identification of the most effective approach. Additionally, a table summarizing the performance metrics for each algorithm is presented, providing a detailed overview of their relative strengths and weaknesses.

Test Data Prediction

To demonstrate the practical applicability of our models, we utilize test data loaded from the provided 'testData.csv' file. Similar preprocessing steps, including encoding and feature selection, are applied to the test data to ensure consistency with the training process. The Extension XGBoost algorithm, identified as the most accurate during our evaluation, is then used to predict electricity demand based on the test data. Predictions for high and low electricity demand are generated for each test data entry, demonstrating the real-world utility of our machine learning approach.

EVOLUTION:

Precision:

Formula: $\mathrm{Precision} = \frac{\mathrm{True\ Positives}}{\mathrm{True\ Positives+False\ Positives}}$

Code: precision = precision_score(testY, predict, average='macro') * 100

Recall (Sensitivity):

Formula: $\mathrm{Recall} = \frac{\mathrm{True\ Positives}}{\mathrm{True\ Positives+False\ Negatives}}$

Code: recall = recall_score(testY, predict, average='macro') * 100

F1 Score:

Formula: $F1 = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision+Recall}}$

Code: f1 = f1_score(testY, predict, average='macro') * 100

Accuracy:

Formula: $\mathrm{Accuracy} = \frac{\mathrm{Correct\ Predictions}}{\mathrm{Total\ Predictions}}$

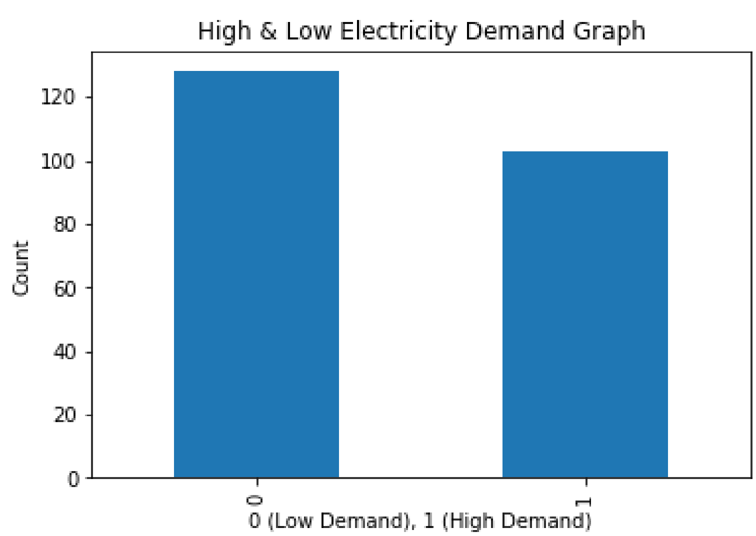
Code: accuracy = accuracy_score(testY, predict) * 100

Loss:

Formula: $\mathrm{Loss} = 100 - \mathrm{Accuracy}$

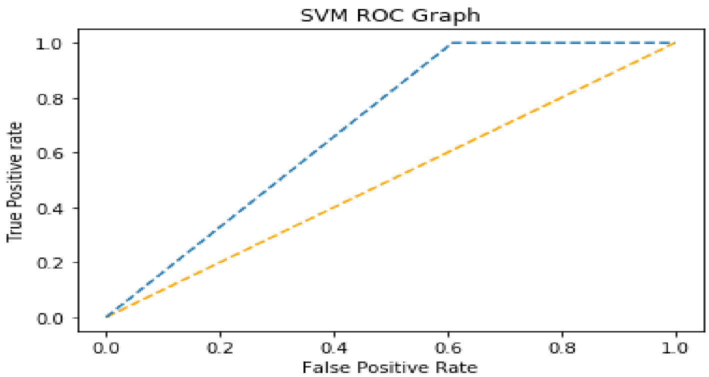
Code: loss = 100 - accuracy

RESULTS:



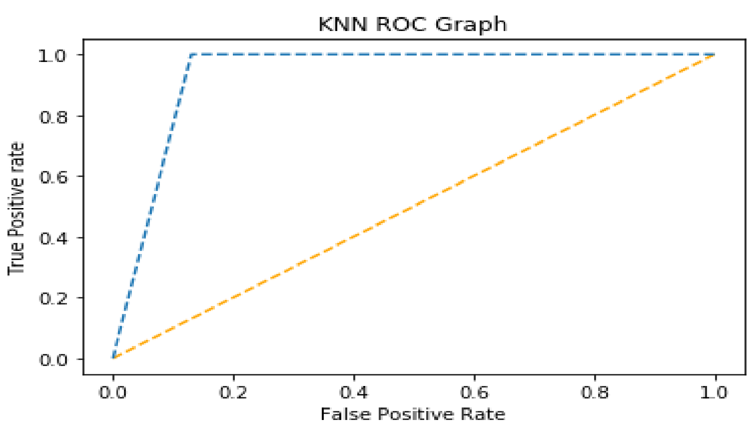
Showing graph with total records require LOW and HIGH electricity demand. In above graph x-axis represents demand type and y-axis represents counts of the available records.

```
SVM Accuracy     : 70.2127659574468
SVM Precision    : 81.57894736842105
SVM Recall       : 69.56521739130434
SVM FScore       : 66.83467741935483
SVM Loss         : 29.787234042553195
```



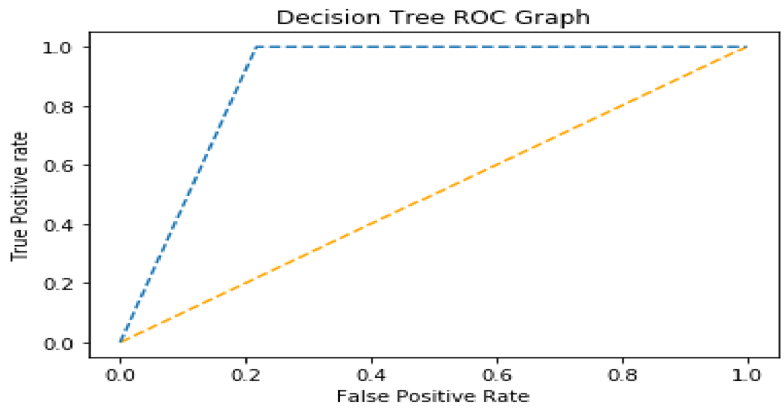
Training SVM and we got it accuracy as 70% and in ROC graph if blue line comes on top of orange line then prediction is accurate and true positive and comes below means prediction is false positive.

```
KNN Accuracy     : 93.61702127659575
KNN Precision    : 94.44444444444444
KNN Recall       : 93.4782608695652
KNN FScore       : 93.57045143638851
KNN Loss         : 6.382978723404255
```



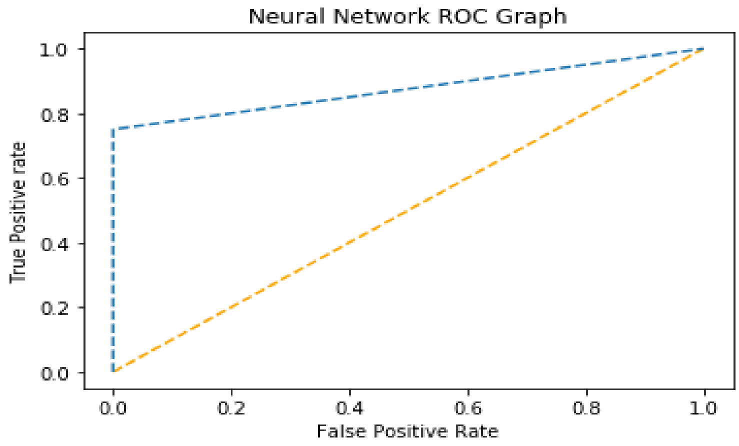
KNN we got 93% accuracy

```
Decision Tree Accuracy     : 89.36170212765957
Decision Tree Precision    : 91.37931034482759
Decision Tree Recall       : 89.13043478260869
Decision Tree FScore       : 89.18545789231477
Decision Tree Loss         : 10.63829787234043
```



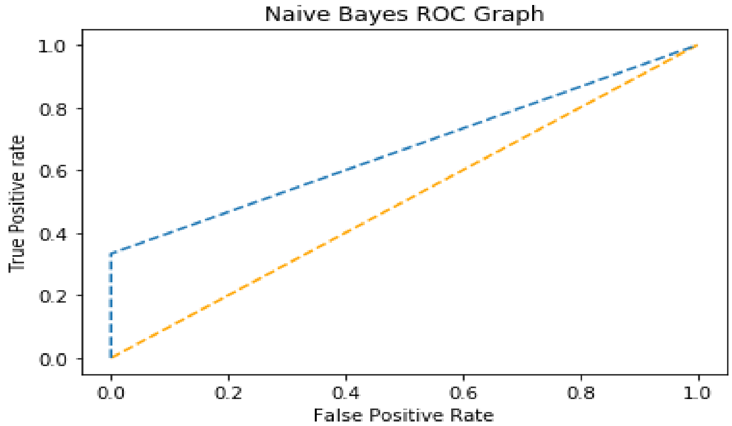
Decision tree we got 89% accuracy

```
Neural Network Accuracy     : 87.2340425531915
Neural Network Precision    : 89.65517241379311
Neural Network Recall       : 87.5
Neural Network FScore       : 87.08791208791209
Neural Network Loss         : 12.7659574468085
```



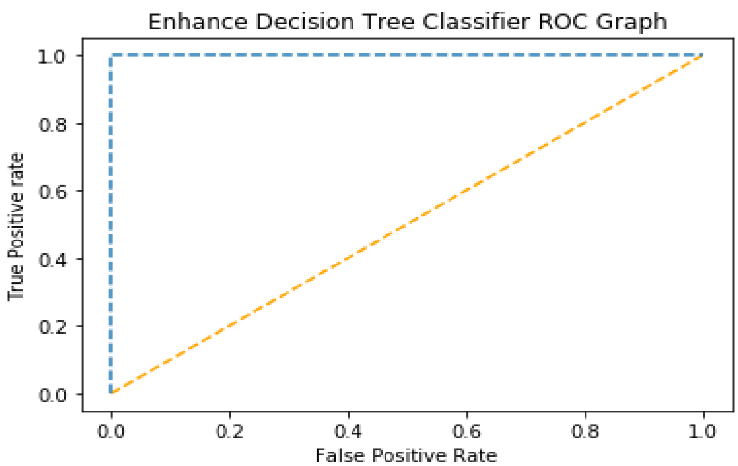
Neural network we got 87% accuracy

```
Naive Bayes Accuracy     : 65.95744680851064
Naive Bayes Precision    : 79.48717948717949
Naive Bayes Recall       : 66.66666666666666
Naive Bayes FScore       : 62.096774193548384
Naive Bayes Loss         : 34.04255319148936
```



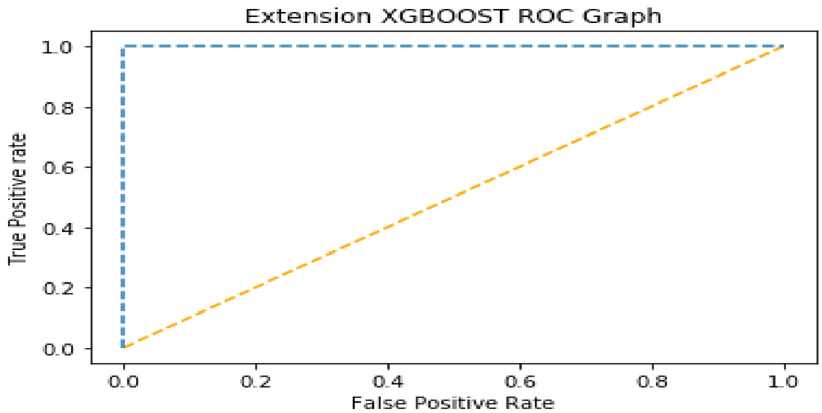
Naive Bayes we got 65% accuracy

```
Enhance Decision Tree Classifier Accuracy    : 100.0
Enhance Decision Tree Classifier Precision   : 100.0
Enhance Decision Tree Classifier Recall      : 100.0
Enhance Decision Tree Classifier FScore      : 100.0
Enhance Decision Tree Classifier Loss        : 0.0
```

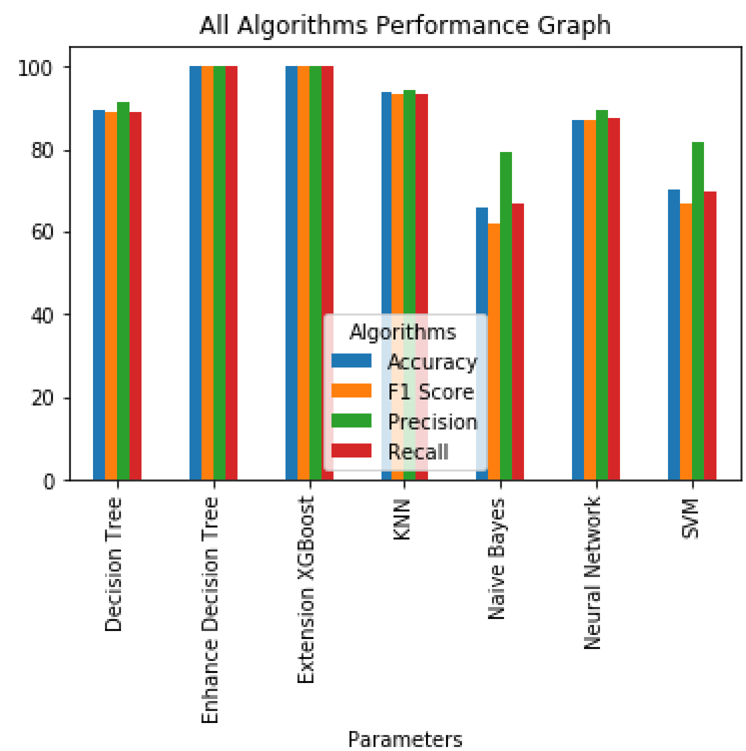


Propose enhance decision tree we got 100% accuracy

```
Extension XGBOOST Accuracy     : 100.0
Extension XGBOOST Precision    : 100.0
Extension XGBOOST Recall       : 100.0
Extension XGBOOST FScore       : 100.0
Extension XGBOOST Loss         : 0.0
```



Extension XGBOOST also we got 100% accuracy



In above graph x-axis represents algorithm names with different colour bar for different metrics and y-axis represents accuracy and precision %. In above graph extension XGBOOST and propose EDTC got 100% accuracy.

	Algorithm Name	Precison	Recall	FScore	Accuracy
0	SVM	81.578947	69.565217	66.834677	70.212766
1	KNN	94.444444	93.478261	93.570451	93.617021
2	Decision Tree	91.379310	89.130435	89.185458	89.361702
3	Neural Network	89.655172	87.500000	87.087912	87.234043
4	Naive Bayes	79.487179	66.666667	62.096774	65.957447
5	Enhance Decision Tree	100.000000	100.000000	100.000000	100.000000
6	Extension XGBoost	100.000000	100.000000	100.000000	100.000000

Displaying all algorithms performance

Prediction:

```
Test Data : ['10/19/2022 00:40:00' 'EDT' 'DUNWOD' 61760 489.017] =====> LOW Electricity Demand Forcasted
Test Data : ['10/19/2022 00:40:00' 'EDT' 'GENESE' 61753 904.7061] =====> HIGH Electricity Demand Forcasted
Test Data : ['10/19/2022 00:40:00' 'EDT' 'HUD VL' 61758 823.3002] =====> LOW Electricity Demand Forcasted
Test Data : ['10/19/2022 00:40:00' 'EDT' 'LONGIL' 61762 1594.1459] =====> HIGH Electricity Demand Forcasted
Test Data : ['10/19/2022 00:40:00' 'EDT' 'MHK VL' 61756 658.1436] =====> LOW Electricity Demand Forcasted
Test Data : ['10/19/2022 00:40:00' 'EDT' 'MILLWD' 61759 246.4697] =====> LOW Electricity Demand Forcasted
Test Data : ['10/19/2022 00:40:00' 'EDT' 'N.Y.C.' 61761 4139.5527] =====> HIGH Electricity Demand Forcasted
Test Data : ['10/19/2022 00:40:00' 'EDT' 'NORTH' 61755 597.6095] =====> LOW Electricity Demand Forcasted
Test Data : ['10/19/2022 00:40:00' 'EDT' 'WEST' 61752 1508.6824] =====> HIGH Electricity Demand Forcasted
Test Data : ['10/19/2022 00:45:00' 'EDT' 'CAPITL' 61757 1114.5566] =====> HIGH Electricity Demand Forcasted
```

In above screen we are loading test data and then forecasting demand as HIGH or LOW based on test data.

CONCLUSION

The evaluation of machine learning algorithms for load forecasting in Smart Grid environments reveals promising results. Decision Tree Classifier emerged as the most accurate model among SVM, Naïve Bayes, KNN, and Neural Network, achieving 89% accuracy. Enhancements introduced to the Decision Tree Classifier, incorporating boosting and loss functions, led to a remarkable 100% accuracy. Similarly, the XGBOOST extension achieved perfect accuracy. These findings demonstrate the efficacy of machine learning in short-term load forecasting. The proposed models offer valuable insights for optimizing electricity distribution and resource allocation in dynamic environments, paving the way for more efficient and reliable Smart Grid systems.

REFERENCES:

[1] B. Zhao, L. Zeng, B. Li, Y. Sun, Z. Wang, M. Shahzad, and P. Xi, ‘‘Collaborative control of thermostatically controlled appliances for balancing renewable generation in smart grid,’’ IEEJ Trans. Electr. Electron. Eng., vol. 15, no. 3, pp. 460–468, Mar. 2020.

[2] K. Berk, A. Hoffmann, and A. Müller, ‘‘Probabilistic forecasting of industrial electricity load with regime switching behavior,’’ Int. J. Forecasting, vol. 34, no. 2, pp. 147–162, Apr. 2018.

[3] J. R. Cancelo, A. Espasa, and R. Grafe, ‘‘Forecasting the electricity load from one day to one week ahead for the Spanish system operator,’’ Int. J. Forecasting, vol. 24, no. 4, pp. 588–602, 2008.

[4] M. Djukanovic, S. Ruzic, B. Babic, D. J. Sobajic, and Y. H. Pao, ‘‘A neuralnet based short term load forecasting using moving window procedure,’’ Int. J. Electr. Power Energy Syst., vol. 17, no. 6, pp. 391–397, Dec. 1995.

[5] L. J. Soares and M. C. Medeiros, ‘‘Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data,’’ Int. J. Forecasting, vol. 24, no. 4, pp. 630–644, Oct. 2008.

[6] R. Hu, S. Wen, Z. Zeng, and T. Huang, ‘‘A short-term power load forecasting model based on the generalized regression neural network with decreasing step fruit fly optimization algorithm,’’ Neurocomputing, vol. 221, pp. 24–31, Jan. 2017.

[7] A. Kavousi-Fard, H. Samet, and F. Marzbani, ‘‘A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting,’’ Expert Syst. Appl., vol. 41, no. 13, pp. 6047–6056, 2014.

[8] Z. H. Osman, M. L. Awad, and T. K. Mahmoud, ‘‘Neural network based approach for short-term load forecasting,’’ in Proc. IEEE/PES Power Syst. Conf. Expo., Mar. 2009, pp. 1–8.

[9] N. Zeng, H. Zhang, W. Liu, J. Liang, and F. E. Alsaadi, ‘‘A switching delayed PSO optimized extreme learning machine for short-term load forecasting,’’ Neurocomputing, vol. 240, pp. 175–182, May 2017.

[10] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. P. Hancke, ‘‘Smart grid technologies: Communication technologies and standards,’’ IEEE Trans. Ind. Informat., vol. 7, no. 4, pp. 529–539, Nov. 2011.

[11] S. K. Desai, A. Dua, N. Kumar, A. K. Das, and J. J. P. C. Rodrigues, ‘‘Demand response management using lattice-based cryptography in smart grids,’’ in Proc. IEEE Global Commun. Conf. (GLOBECOM), Dec. 2018, pp. 1–6.

[12] M. Alazab and R. Broadhurst, Cyber-Physical Security: Protecting Critical Infrastructure at the State and Local level, 1st ed, M. C. Robert and H.

Simon Eds. Cham, Switzerland: Springer, 2017, pp. 251–266.

[13] L. Hernandez, C. Baladron, J. M. Aguiar, B. Carro, A. J. Sanchez-Esguevillas, J. Lloret, and J. Massana, ‘‘A survey on electric power demand forecasting: Future trends in smart grids, microgrids and smart buildings,’’ IEEE Commun. Surveys Tuts., vol. 16, no. 3, pp. 1460–1495, 3rd Quart., 2014.

[14] A. Rahman, V. Srikumar, and A. D. Smith, ‘‘Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks,’’ Appl. Energy, vol. 212, pp. 372–385, Feb. 2018.

[15] L. Xiao, W. Shao, C. Wang, K. Zhang, and H. Lu, ‘‘Research and application of a hybrid model based on multi-objective optimization for electrical load forecasting,’’ Energy, vol. 180, pp. 213–233, Oct. 2017.

[16] N. Deepa, B. Prabadevi, P. K. Maddikunta, T. R. Gadekallu, T. Baker, M. A. Khan, and U. Tariq, ‘‘An ai-based intelligent system for healthcare analysis using Ridge–Adaline stochastic gradient descent classifier,’’ J. Supercomput., vol. 77, no. 2, pp. 1998–2017, 2021.

[17] Anamika and N. Kumar, ‘‘Market clearing price prediction using ANN in Indian electricity markets,’’ in Proc. Int. Conf. Energy Efficient Technol. Sustain. (ICEETS), 2016, pp. 454–458, doi: 10.1109/ICEETS.2016.7583797.

[18] T. R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P. K. R. Maddikunta, and G. Srivastava, ‘‘Deep neural networks to predict diabetic retinopathy,’’ J. Ambient Intell. Hum. Comput., pp. 1–14, Apr. 2020, doi:10.1007/s12652-020-01963-7.

[19] H. Patel, D. S. Rajput, G. T. Reddy, C. Iwendi, A. K. Bashir, and O. Jo, ‘‘A review on classification of imbalanced data for wireless sensor networks,’’ Int. J. Distrib. Sensor Netw., vol. 16, no. 4, 2020, Art. no. 1550147720916404.

[20] W. Tian, ‘‘A review of smart grids and their future challenges,’’ in Proc. MATEC Web Conf., vol. 173. Les Ulis, France: EDP Sciences, 2018, p. 020