International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/CSEIT23903112

Evaluating Machine Learning Algorithms for Load Forecasting in Smart Grid

[1]Gopi Tharun, [2]Mr. G. Lokesh

[1] PG Scholar, [2]Associate Professor

Department of CSE, Vemu Institute of Technology, P. Kotha Kota, Chittoor, Andhra Pradesh, India

ARTICLEINFO

Article History:

Accepted: 05 July 2023

Published: 21 July 2023

Publication Issue

Volume 9, Issue 4

July-August-2023

Page Number

159-165

ABSTRACT

Because it makes computing much simpler and eliminates the need to purchase the actual hardware required for computations, cloud computing is quickly replacing on-premise computing in the information technology sector. These businesses depend on the availability of a reliable and affordable electrical power supply because they house numerous computers and servers whose primary power source is electricity. Cloud centers use a lot of energy. With recent increases in electricity prices, one of the biggest obstacles in designing and efficiently placing data and scheduling nodes to unload or transfer storage is one of the upkeep of such centers. Another difficulty is to reduce the amount of electricity that data centers use and conserve energy. In this project, we suggest using an Extreme Gradient Boosting (XGBoost) model to offload or transfer storage, forecast electricity prices, and as a result cut data center energy expenses. On a real-world dataset provided by the Independent Electricity System Operator (IESO) in Ontario, Canada, the effectiveness of this strategy is assessed in order to offload data storage in data centers and effectively reduce energy consumption. 70% of the data is used for training and 30% for testing.

Keywords : Machine learning, XG Boost, CatBoost and ANN. And ML techniques, evaluation.

I. INTRODUCTION

As a storage platform, cloud computing is being used more and more frequently, which lowers hardware investments and procurement costs. Data Centers (DCs) are in high demand due to the exponential growth in the demand for information. Data Centers (DCs) use 2% of the world's energy, which is a significant amount. It is anticipated to increase by 12% annually. A little over 39% of energy is utilized for cooling, 45% for powering IT infrastructure, and 13% for lighting. The cost of this level of consumption to the economy is high. In order to ensure reliability through replication, DC operators typically have a few DCs dispersed throughout different areas. Being near the clients will satisfy the latency specifications. Distributed DCs, however, might result in cost unpredictability because power markets have fluctuating prices. These power markets have significant cost flexibility. As a result, DC manufacturers would construct DCs in regions with low temperatures and affordable electricity. Companies like Netflix employ content delivery networks (CDNs). In order to reduce the need for long-distance data transmission and improve Quality of Service, they would situate the data center closer to the clients . The capacity of centralized DCs might potentially be offloaded using this technique to hubs at the system's edge, allowing businesses to use less energy overall and cut their energy bills as a result. Due to their significance and pressing necessity, green energy and the environment have gained a lot of attention in recent years. The problems have been solved by a number of researchers using both cutting-edge and conventional methods. Due to the fact that costs range from one geographic region to another, several researchers recommend conducting a market survey on the expense of setting up servers at various places. Similarly, many researchers have focused on the diverse effects of machine learning methods on modeling, designing, and forecasting electricity price, particularity in global market. Generally two machine learning techniques are mostly used where the first one is for forecasting electricity price and the later one is for the energy systems.Most of the previous works on electricity price prediction are still in their infancy and lack in terms of accuracy, computational overhead or unable to prove results on real-time data. Python-Flask is employed as front end which is used to craft the user interface. MySQL is employed as back end and used to craft the database and save the particulars.

II. RELATED WORKS

Collaborative control of thermostatically controlled appliances for balancing renewable generation in smart grid: This paper proposes a collaborative control strategy for thermostatically controlled appliances to address supply-demand imbalances caused by the integration of uncertain and variable renewable generation resources into the smart grid. The strategy consists of four key components: state awareness, in-the-moment analysis, rational judgment, and exact execution. The state awareness component utilizes a responsive task-type model to ascertain the appropriate executable task type for each appliance. In the scientific decision-making portion, a multi-objective 0-1 programming method is employed to identify the optimal control combination. During the exact execution phase, a thermostat set-point control rule is utilized to generate control signals for the appliances. To showcase the efficacy of the suggested strategy, four case studies are meticulously designed, and simulation results are obtained to validate its numerous advantages. By implementing this strategy, the paper aims to mitigate power quality issues, enhance power delivery reliability, and ensure the smooth integration of renewable energy sources.

Performance Comparison of Machine Learning Algorithms in Classifying Information Technologies Incident Tickets: When it comes to classifying information technology incident tickets, various machine learning algorithms can be employed, each with its strengths and weaknesses. An evaluation of these algorithms can help identify the most effective approach. Decision trees are intuitive and easy to interpret but may struggle with complex relationships. Random forests combine multiple decision trees, resulting in improved accuracy and robustness. Support Vector Machines (SVMs) can handle high-dimensional data well but may be computationally expensive. Naive Bayes classifiers assume independence between features and perform well with limited training data. Logistic regression is a widely used algorithm that models the probability of a certain class. It is simple and efficient but may not capture complex patterns as effectively. Gradient Boosting Machines (GBMs) create an ensemble of weak learners to form a strong learner, achieving high accuracy but requiring careful parameter tuning. Ultimately, the choice of algorithm depends on the specific dataset and requirements. Conducting thorough performance evaluations and considering factors such as accuracy, interpretability, and computational efficiency will aid in selecting the most suitable algorithm for classifying information technology incident tickets.

Forecasting the electricity load from one day to one week ahead for the Spanish: The building process and models used by Red Eléctrica de Espaa (REE), the Spanish system operator, for projecting the short-term electricity load are thoroughly examined in this article. The forecasting method used by REE consists of a daily model and 24 hourly models, all of which have a similar framework. The study focuses on two essential forecast categories for REE: forecasts for hourly data one day in advance and predictions for daily data several days in advance. The paper examines historical real-time forecasting errors for daily and hourly data throughout the year 2006 to assess how accurate these forecasts are, offering insights into the forecasting performance in relation to various factors like the day of the week, season, and type of day. Furthermore, the study investigates additional aspects of the prediction problem, including as the necessity for an appropriate treatment of special days.

A neuralnet based short term load forecasting using moving window procedure," Int. J. Electr. Power Energy Syst: A novel approach combining unsupervised and supervised learning techniques is proposed for improved short-term electric load forecasting. The method utilizes a "moving window" procedure to create a training set database based on recent load and weather information. To meet the operational requirements of real electric utility practices, a forecasting lead time ranging from 16 to 88 hours is introduced. Initially, unsupervised learning (UL) is employed to identify days with similar daily load patterns. Subsequently, a feed-forward three-layer neural network is designed to predict 24-hour loads using the supervised learning (SL) phase. The effectiveness of the proposed methods is demonstrated by comparing the forecasted hourly loads for each day in 1991 with the realized data in the Electric Power Utility of Serbia (EPS) during the same period. By employing a more suitable selection procedure for input features and training sets, along with a flexible neural network structure and re-training procedure, significant improvements in forecasting results are achieved compared to the previous UL/SL concept.

Modeling and forecasting short-term electricity load:A comparison of methods with an application to Brazilian data: This paper presents a forecasting model designed to predict the hourly electricity load in a utility service area located in the southeastern region of Brazil. The model employs a unique approach, wherein each hour of the day is represented by a distinct forecasting model. These models utilize a two-component decomposition technique to analyze the daily series for each hour. The first component encompasses deterministic factors such as trends and seasonality, while also accounting for special days' effects. The second component is a stochastic model that incorporates a linear autoregressive framework, with additional consideration for nonlinear alternatives. a comparison is conducted against a benchmark model. The findings demonstrate that the proposed model outperforms the

benchmark, highlighting its efficacy in forecasting electricity load in tropical regions. This research contributes to the advancement of accurate load forecasting techniques, which are crucial for efficient planning and management of electric utility services. By capturing the specific characteristics and fluctuations of hourly load patterns, the proposed model offers valuable insights for decision-making and resource allocation in the context of electricity supply and demand.

III. Methodology

Proposed system:

To classify data, a number of machine learning techniques have been developed, but none have adequately addressed the problem of inaccurate diagnoses. One tool for this is Stevens Multi Performance Comparison of Machine Learning Algorithms for Load Forecasting in Smart Grid.

Additionally, the variability and size of the data are often not taken into consideration in comparable studies that have proposed strategies for evaluating these malignancies. In order to remove bias and deviation from stability, we propose a machine learning-based approach that combines a novel method of preprocessing the data for feature transformation, CatBoost, and XG Boost give the best score techniques, and running classifier tests based.

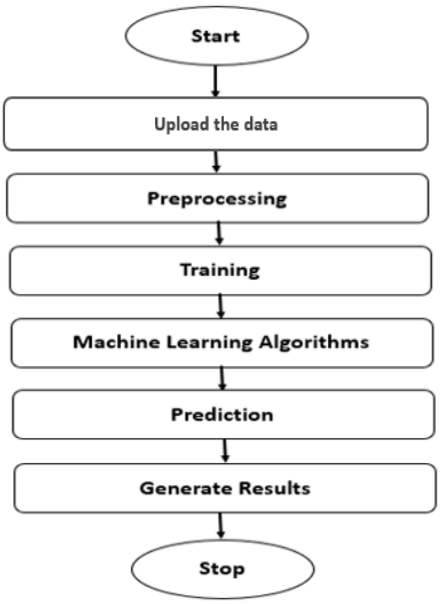


Figure 1 : Block diagram

IV. Implementation

The algorithms listed below were used to complete the project.

1. ANN:

An Artificial Neural Network (ANN) is a computational model inspired by the structure and functioning of biological neural networks in the human brain. It is a machine learning technique that processes information by mimicking the interconnected neurons in the brain. ANNs consist of interconnected nodes, called artificial neurons or units, organized in layers. The three main types of layers are the input layer, hidden layers, and output layer. Each neuron receives inputs, performs a weighted sum, applies an activation function, and passes the output to the next layer. The weights are adjusted through a process known as training, which involves presenting the network with labeled training examples and updating the weights to minimize the error. ANNs are capable of learning complex patterns, making them suitable for various tasks such as pattern recognition, classification, regression, and even decision-making. They have been successfully applied in diverse fields including computer vision, natural language processing, speech recognition, and finance. With their ability to learn from data and generalize to new inputs, artificial neural networks have become a powerful tool for solving complex problems and have revolutionized the field of machine learning.

2.XG Boost:

XGBoost (Extreme Gradient Boosting) is a powerful machine learning algorithm that has gained significant popularity in the field of data science and predictive modeling. It is known for its exceptional performance and efficiency in handling structured and tabular data. XGBoost is an implementation of the gradient boosting framework, which combines multiple weak predictive models to create a strong predictive model. One of the key advantages of XGBoost is its ability to handle a variety of data types, including numerical and categorical features. It employs a gradient boosting approach, where subsequent models are built to correct the errors made by previous models. This iterative process allows XGBoost to optimize the objective function, which can be customized based on the specific problem at hand. XGBoost incorporates several techniques to improve model performance, such as regularization, tree pruning, and parallel processing. It also provides important features like feature importance ranking, which helps in understanding the contribution of each feature towards the model's predictions ue to its exceptional accuracy and scalability, XGBoost has been successfully applied in various domains, including finance, healthcare, and online advertising. Its efficiency and ability to handle large datasets make it a popular choice for data scientists and machine learning practitioners.

3. Cat Boost:

Cat Boost is a machine learning algorithm that excels in gradient boosting frameworks, specifically designed to handle categorical features efficiently. It was developed by Yandex, a leading technology company. Cat Boost offers several advantages, making it a popular choice for various applications. One key feature is its ability to process categorical data directly, eliminating the need for extensive preprocessing. This algorithm also incorporates novel techniques like ordered boosting, which enhances the learning process by considering the natural ordering of categorical variables. Additionally, Cat Boost implements a unique algorithm to handle missing data effectively. It provides excellent predictive accuracy, robustness against outliers, and fast training speed. Cat Boost has gained significant popularity due to its versatility and ease of use, finding applications in various domains like finance, healthcare, and marketing. Its open-source nature and comprehensive documentation make it accessible to both beginners and experienced data scientists, enabling them to leverage its power for developing high-performing machine learning models.

V. Results and Discussion

The following screenshots are depicted the flow and working process of project.

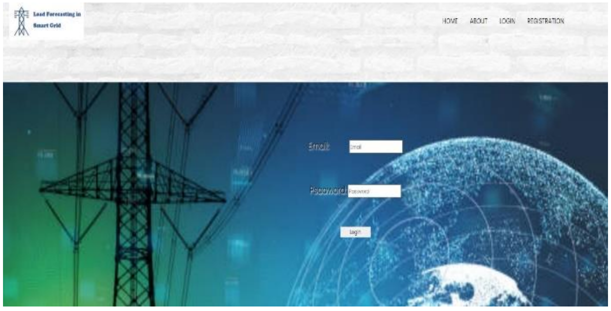
Home Page: Here user view the home page for A performance comparison appellation.

About page:

In the about page, users can learn more about A performance comparison

Login page :

User Login page



Registration page:

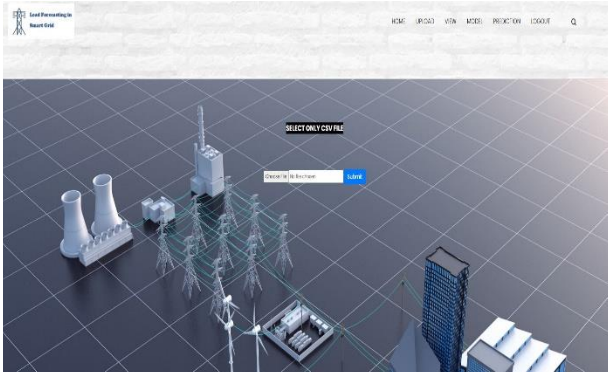
User Registration page

User Home Page:

User After login will view The user home page

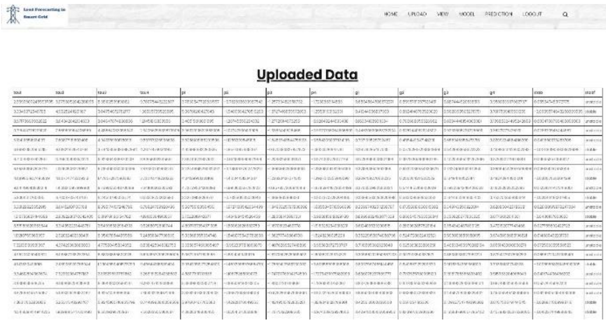
Load Page:

User will Load The Data set



View Page:

User View The Data

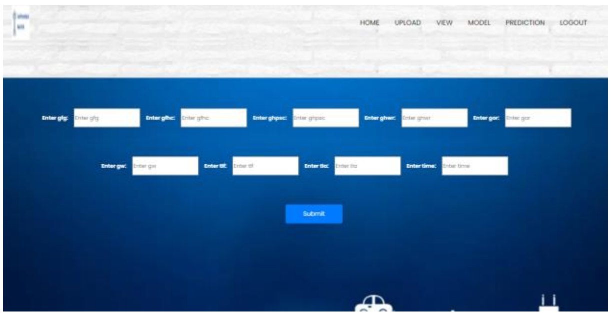


Model:

User will View the accuracy on every algorithm.

Prediction page:

User will give a proper input and view the result



VI. CONCLUSION

The implementation of the Xgboost model has proven to be effective in optimizing data placement and node scheduling to offload or move storage in cloud data centers. By accurately predicting electricity prices using the real-world dataset provided by the Independent Electricity System Operator (IESO) in Ontario, Canada, the XGBoost model enables data centers to make informed decisions on when to perform offloading, thus reducing energy consumption costs significantly. This approach demonstrates the potential to alleviate the energy-hungry nature of cloud centers, promoting a more sustainable and cost-efficient cloud computing infrastructure. The successful results achieved through this project highlight the promising future of data center energy optimization with machine learning techniques.

Algorithms	R2_score
XG Boost	0.918281418
Cat Boost	85.16375660615508
ANN	15.9301513

VII. 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.

Cite this article as :

Gopi Tharun, Mr. G. Lokesh , "Evaluating Machine Learning Algorithms for Load Forecasting in Smart Grid", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 4, pp.159-165, July-August-2023.