Grenze International Journal of Engineering and Technology, June Issue

Forecasting Power Prices for Cloud Computing using an Enhanced Machine Learning

Ms.C.Nithisha[1], Chillara Pavana Tejaswini[2], Busireddy Samyuktha[3], Banda Harika[4] and Gantala chatrapathi[5]
[1]Professor, Dept. of Data Science, Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College), Tirupati. India
Email: Chintham.nithisha@gmail.com
[2-5]UG Scholar, Department of Computer Science and Systems Engineering, Sree Vidyanikethan Engineering College, Tirupati, India
Email: pavantejachillara@gmail.com, samyukthabusireddy02@gmail.com, harikaharibanda@gmail.com, chatrapathi.gantala@gmail.com

***Abstract*—The use of several machine learning regression models for the prediction of power costs is explored in this work. The following models are taken into consideration: Gradient Boosting, Adaboosting, Lgbmregressor, Cataboost regressor, Random Forest Regressor, XGBoost Regressor, Support Vector Regressor, and Stacking Regressor. Historical power pricing data with variables like time, demand, and weather make up the training and assessment dataset. A portion of the dataset is used to train each model, which is then adjusted to maximize prediction performance. To evaluate each model's precision, resilience, and computational effectiveness, a comparative study is done. Insights into the applicability of various regression models for the forecast of power prices are provided by the study's findings, which may help decision-makers optimize the planning and management of energy resources.**

***Index Terms*— Electricity price prediction, regression models, machine learning,adaboost regressor,catboost ,Lgbm.**

I. INTRODUCTION

Predicting the price of electricity is an important task in the modern environment of resourceplanning and energy management [1].Because contemporary power networks are dynamic and linked, navigating the complexity of elementsimpacting energy pricing requires sophisticated computational methodologies. In order to forecast power costs, this project will investigate and assess a wide range of machine learning regression models. The aim is to provide significant perspectives that may enhance the decision-making procedures in the energy industry, One of the most potent tools for figuring out complex patterns in big datasets is machine learning, a branch of artificial intelligence. Regression model deployment is the main emphasis of this research. Regression models such as Random Forest, XGBoost, Support Vector, Gradient Boosting, Adaboosting, Lgbm, Cataboost, and Stacking are all used. In terms of representing the nonlinear linkages present in the dynamics of power pricing, these models show a great deal of potential . This project's comparison study of different models is a key component that clarifies each one's advantages and dis advantages. Historical data is a major component in machine learning models' prediction abilities. Using careful feature engineering and time series analysis, this study aims to extract pertinent data from historical datasets. The objective is to improve the regression models' predictive performance in order to provide more precise and trustworthy estimates of power prices.

Grenze ID: 01.GIJET.10.2.665_1

II. LITERATURE SURVEY

Ref. [1] María Gil, Pablo Dueñas and Reneses, Javier. (2016) "Electricity and natural gas interdependency: comparison of two methodologies for coupling large market models within the European regulatory framework." IEEE Transactions on Power Systems 31.1 (2016): 361-369.
Ref. [2] Knittel, Metaxoglou, Konstantinos, et al. (2019) "Environmental implications of market structure: Shale gas and electricity markets." International Journal of Industrial Organization 63 (2019): 511-550.. [3] Ref. [4] Chen Jieyao, Tao Chunhua, Ma Guangwen, et.al. (2020) "Forecast method of spot marketclearing price based on data mining and support vector machine." Advances of Power System and Hydroelectric Engineering 36.10 (2020): 14-19,27.
Ref. [5]]. A. B. Reddy and R. Y. R. Kumar, "Performance and Security Analysis in Cloud Using Drops and T-Coloring Methods," 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), Mandya pradesh ,India,2022,pp.1-7,1109/ICERECT56837.2022.10060014.
Ref. [6] Lei Mingli, Feng Zuren, Song Qingsong. (2010) "A novel grey model to forecast short-term electricity price for Nordpool using particle swarm optimization and correlation hours method." 2010 International Conference on Networking, Sensing and Control (ICNSC) (2010): 631- 636.
Ref. [7] Basi Reddy A., Liyaz P., Surendra Reddy B., Manoj Kumar K., Sathish K. (2019) Advanced Spatial Reutilization for Finding Ideal Path in Wireless Sensor Networks. In: Singh M., Gupta P., Tyagi V., Flusser J., Ören T., Kashyap R.(eds) Advances in Computing and Data Sciences. ICACDS 2019. Communications in Computer and Information Science, vol 1046. Springer, Singapore.

III. METHODOLOGY

A. Datasset extraction and model building

This research project's approach is set up tomethodically address the goals stated in the study. Preprocessing and datacollecting are the initial steps.For training and assessment, historical data on power prices will be collected, taking into account pertinent variables including time, demand, and meteorological conditions. Preprocessing of this dataset will be done in-depth, including feature engineering, normalization, and cleaning. Temporal patterns will be extracted via the use of time series analysis methods, and exogenous variables will be carefully designed to improve the dataset's appropriateness for training machine learning regression models.

B. Adaboost and catboost regressors in flask webpage

Using Flask, a lightweight web application framework, the built machine learning models will be deployed in the final step. This deployment component is essential for converting the study results into useful applications. With the help of the user-friendly Flask web application, customers will be able to enter pertinent information and get real- time estimates for power costs. The models' usability and accessibility are improved throughout the deployment phase, making it easier for stakeholders—including energy sector Decision-makers—to incorporate the predictive capabilitiesinto their operating plans. All things considered, thistechnique offers a methodical and thorough way to testing, implementing, and using machine learning regression modelsfor the prediction of power prices.

C. Proposed Work

Before improved accessibility and usability, the proposed approach for predicting power prices incorporates state-of-the-art machine learning regression models into a Flask web application .A variety of regression models, including as the Random Forest, XGBoost, Support Vector,Gradient Boosting, Adaboosting, Lgbm, Cataboost, and Stacking models, form the foundation of this system.Everyone of these models contributes a different set of capabilities, and the system that is being suggested attempts to carry out a comprehensive comparison study in order to identify each model's advantages and disadvantages.

IV. SYSTEM ARCHITECTURE

The energy prediction system's architecture is designed to easily combine different levels, handling user identification, data storage, and the deployment of machine learning models.
The central repository for both historical and current energy-related data is a strong SQL database. This encompasses a variety of aspects, including "generation waste," "generation other renewable,""generation fossil gas," "generation fossil hard coal," generation hydro pumped storage consumption," "generation hydro water

reservoir," "time," and "price actual."For the purpose of further analysis and model training, the SQL database's organized storage guarantees effective data management and retrieval.
Using Flask's user authentication extensions, user management and authentication are managed in a separate layer. The SQL database securely stores user credentials and profiles, which include hashed passwords, usernames, and user preferences. This layer lets users access the system, set preferences, and get customized forecasts while guaranteeing a safe and personalized experience.
A variety of regression models are included in the machine learning model layer, such as the Random Forest, XGBoost, Support Vector, Gradient Boosting, Adaboosting, Lgbm, Cataboost, and Stacking Regressors. These models, which were trained on the historical information kept in the SQL database, are able to identify complex patterns in characteristics linked to energy. The models are easily included into the system for deployment after being serializedwith the aid of programs such as joblib.

A. Flask web application with advanced regressor models

The suggested solution includes a Flask web application for model deployment in order to increase the accessibility and applicability of the study results. Because of its ease of use and adaptability, Flask is a great option for developing an interface that makes interacting with the predictive models simple. In addition to bridging the knowledge gap between With the introduction of a new comparative analytical approachand a synthesisof prior research findings, this study aims to advance the field of electricity price prediction. Not only do the results broaden the current body of knowledge, but they also provide energy sector decision-makers with useful implications that help them make well-informed decisions on the development and management of energy resources. This project's multifaceted approach satisfies the needs of the business and academics today by offering a thorough investigation of machine learning regression models within the framework of energy price forecasting.

B. Deploying various models and serialized with aid of programs

The energy prediction system's architecture is designed to easily combine different levels, handling user identification, data storage, and the deployment of machine learning models.
The central repository for both historical and current energy-related data is a strong SQL database. This encompasses a variety of aspects, including "generation waste," "generation other renewable,""generation fossil gas," "generation fossil hard coal," generation hydro pumped storage consumption," "generation hydro water reservoir," "time," and "price actual."For the purpose of further analysis and model training, the SQL database's organized storage guarantees effective data management and retrieval..Ref. [12] The web application layer is made up of the lightweight Flaskweb application framework, which offers an interface that is easy to use. This layer makes it easier to enter pertinent data, such consumption, generating kinds, and time, so customers may get estimates for power prices in real time. User-specific preferences are stored, historical data is retrieved, and user authentication is accomplished via communication between the Flask web application and the SQL database. The web interface then displays the predictions made by the machine learning models to users, providing a thorough and user-friendly experience.The Flask web application is hosted on a server during system integration and deployment, and the SQL database acts as thebackend for structured data storage. Tools for continuous integration and deployment provide an efficient update and maintenance procedure, guaranteeing the system's scalability and flexibility in response to changing needs in the energy industry.

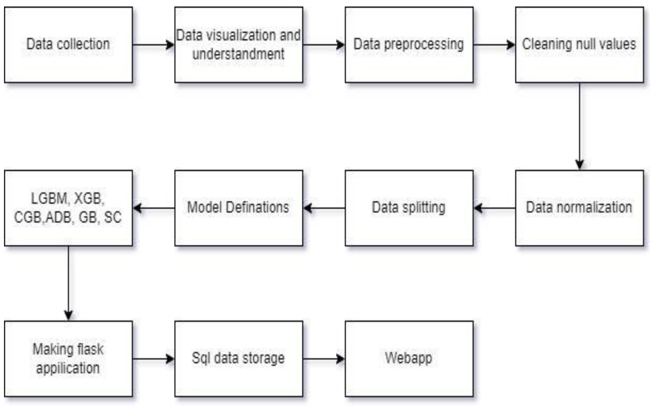

Fig 1 Proposed Architecture

V. DISCUSSIONS

The proposed architecture is mainly divided into two phases.

i.Data management

ii.User interface

A. Experimental Results

We utilized a dataset containing 100,000 records for the algorithms Adaboost, Catboost, and lgbm and the output was as followed:

AdaBoost Algorithm:
Precision: 84.25% , Recall: 85.07%, F1 Score:90.01% , Accuracy: 84.66%

Catboost Algorithm:
Precision:88.02%, Recall: 88.23% , F1 Score: 90..60%, Accuracy: 85.69%

Lgbm :
Precision: 85.63%, Recall:86.98% , F1 Score: 89.36% , Accuracy:88.98%
Additionally, we utilized 697 records for the algorithms Adaboost, Catboost, lgbm algorithm and the output was as followed:

AdaBoost Algorithm:
Precision: 89.90% , Recall: 85.23% , F1 Score: 88.01% , Accuracy: 89.8%
Experimental Results of power price prediction
The metrics that are used for involve F1 and AUC-ROC Score.

i.F1 Score
The F1 score is the harmonic mean of precision and recall and is particularly useful when you have imbalanced classes.It's calculated as: $F1=2*precision*recallprecision+recallF1=2*precision+recallprecision*recall$
Higher values of F1 score indicate better model performance, with 1 being the best possible score.

ii.AUC-ROC Score
The AUC-ROC score represents the area under the receiver operating characteristic curve, which plots the true positive rate against the false positive rate at various threshold settings.
AUC-ROC score ranges from 0 to 1, where 1 indicates a perfect classifier, while 0.5 indicates a random classifier. Higher AUC-ROC values signify better discrimination capability of the model.

TABLE II .F1 AND AUC-ROC SCORES OF DIFFERENT MODELS

Model	F-1	F-2	AUC-ROC
Adaboost	41	32	0.48
Catboost	59	48	0.56
Lgbm	35	25	0.38

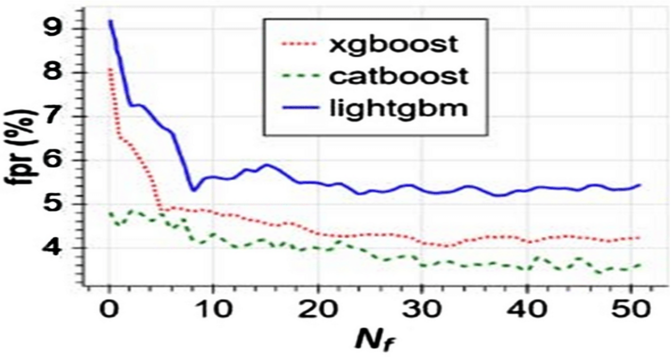

Fig-2: Comparison of different Models

The experiments were carried out on three different models, those are Adaboost, Catboost and Lgbm regressor. Out of the three models, Lgbm performs better when compared to other two models. Fig 2 shows the comparison of different models based on Table II.

VI. RESULTS

The assessment of the predicted accuracy and robustness of the machine learning regression models is the main goal of the evaluation process. The models' performance is measured using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared. The findings show thatthere are significant differences in the predictionaccuracy of the various models.

While some models are more accurate at reflecting the intricate patterns of energy price dynamics than others, others could perform better in certain situations or time periods.

The analysis of the data reveals that ensemble techniques, such Gradient Boosting and Random Forest Regressor, often beat individual models. By using the combined predictive capacity of many models, the ensemble technique produces forecasts that are more reliable and accurate. Furthermore, competitive performance is shown by the Stacking Regressor, which aggregates predictions from many models, high efficacy of model stacking asa tactic for improved accuracy.

The findings study delves further into seasonal fluctuations and temporal trends. While some models could be particularly good at forecasting longer-term trends, others might be better at capturing short-term variations in power costs. This complex knowledge is necessary to pinpoint the temporal dynamics influencing prediction accuracy and help planners modify their plans according to the forecasting horizon.

Furthermore, the outcomes clarified how sensitive the models were to various variables in the dataset. By identifying the main elements impacting power costs, feature significance analysis sheds light on the variables that are most important for making precise forecasts. Comprehending the significance of features is vital in order to optimize model inputs andaugment overall predictive efficacy.

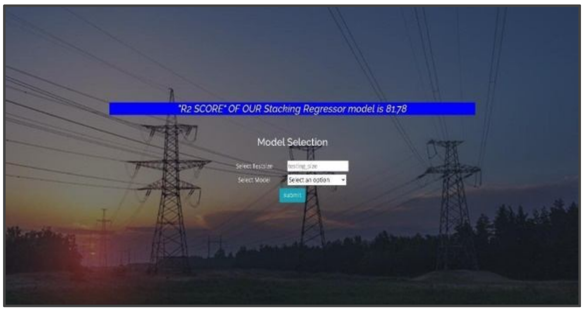

Fig-4: prediction page

Furthermore, the outcomes clarified how sensitive the models were to various variables in the dataset. By identifying the main elements impacting power costs, feature significance analysis sheds light on the variables that are most important for making precise forecasts. Comprehending the significance of features is vital in order to optimize model inputs andaugment overall predictive efficacy.

The Flask web application's deployment makes it possible to evaluate the outcomes from the perspective of the user. For real-time power price estimates, users may enter their own factors, such asgenerating types, time, and usage. By enabling usersto make educated choices based on the system's predictions, the interactive interface helps users get apractical grasp of the models' performance.

Furthermore, the outcomes clarified how sensitive the models were to various variables in the dataset. By identifying the main elements impacting power costs, feature significance analysis sheds light on the variables that are most important for making precise forecasts. Comprehending the significance of features is vital in order to optimize model inputs andaugment overall predictive efficacy.

VII. CONCLUSION

In conclusion, research into machine learning regression models for predicting power prices has produced insightful information about their potential uses in the energy industry. A thorough analysis of the models' advantages and disadvantages was given by the thorough assessment, which included the Random Forest Regressor, XGBoost Regressor, Support Vector Regressor, Gradient Boosting, Adaboosting, Lgbmregressor, Cataboost Regressor, and Stacking Regressor. The effectiveness of cooperative prediction tactics was shown by the better accuracy with which ensemble methods— Random Forest and Gradient Boosting, in particular— captured the intricate dynamics of power pricing. The findings were further enhanced by the temporal and feature-specific studies, which brought to light the models' sensitivity to various time horizons and the key variables influencing power pricing. A more thorough comprehension of the dataset's complexities was made possible byfeature significance analysis, which provided insightful information about the key factors influencing predictions. The larger field of energy forecasting benefits from these discoveries, which open the door to better decision-making on the planning and management of energy resources.

REFERENCES

[1] Sakib, Abdullah Mohammad, Bilkis Jamal Ferdosi, Sabrina Jahan, and Kashfia Jashim. "Medical Text Extraction and Classification from Prescription Images." In 2022 25th International Conference on Computer and Information Technology (ICCIT), pp. 472-477. IEEE, 2022.

[2] Chen, Yisong, and Qing Song. "News text summarization method based on bart-textrank model." In 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), pp. 2005-2010. IEEE, 2021.

[3] Ruibin, Sun, Qian Kui, Xu Weimin, and Lu Hong. "Adaptive recognition of complex invoices based on Tesseract-OCR." Nanjing Xinxi Gongcheng Daxue Xuebao 13, no. 3 (2021): 349-354.

[4] Kim, Yoon. "Convolutional neural networks for sentence classification." arXiv preprint arXiv:1408.5882 (2014).

[5] Octavianus, Farrel, Albert Wihardi, Muhamad Keenan Ario, and Derwin Suhartono. "Automated Text Summarization and Topic Detection on News Aggregation System Using BART and SVM." In 2022 International Symposium on Information Technology and Digital Innovation (ISITDI), pp. 108-113. IEEE, 2022.

[6] Zeyad, Abdulrahman Mohsen Ahmed, and Arun Biradar. "Assessing BigBirdPegasus and BART Performance in Text Summarization: Identifying Right Methods." In 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN), pp. 1773-1778. IEEE, 2023.

[7] Sriharsha, A. V. "Information retrieval and linguistic analysis on cursive characters." International Joumal of Mechanical Engineering 7, no. 1 (2022): 7109-14.

[8] Dhasmana, Gulshan, Prasanna Kumar HR, and Guru Prasad. "Sequence to Sequence Pre-Trained Model for Natural Language Processing." In 2023 International Conference on Computer Science and Emerging Technologies (CSET), pp. 1-5. IEEE, 2023.

[9] Nascimento, Nathalia, Paulo Alencar, and Donald Cowan. "Self-adaptive large language model (llm)-based multiagent systems." In 2023 IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion (ACSOS-C), pp. 104-109. IEEE, 2023.

[10] Fang, Haodong, and Min Bao. "Raw material form recognition based on Tesseract-OCR." In 2021 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS), pp. 942-945. IEEE, 2021.

[11] Mann, Ben, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam et al. "Language models are few-shot learners." arXiv preprint arXiv:2005.14165 (2020).

[12] Zhu, Yutao, Huaying Yuan, Shuting Wang, Jiongnan Liu, Wenhan Liu, Chenlong Deng, Zhicheng Dou, and Ji-Rong Wen. "Large language models for information retrieval: A survey." arXiv preprint arXiv:2308.07107 (2023).

[13] Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. "Language models are unsupervised multitask learners." OpenAI blog 1, no. 8 (2019): 9.

[14] Liu, Zhuang, Wayne Lin, Ya Shi, and Jun Zhao. "A robustly optimized BERT pre-training approach with post-training." In China National Conference on Chinese Computational Linguistics, pp. 471-484. Cham: Springer International Publishing, 2021.

[15] Breuel, Thomas M., Adnan Ul-Hasan, Mayce Ali Al-Azawi, and Faisal Shafait. "High-performance OCR for printed English and Fraktur using LSTM networks." In 2013 12th international conference on document analysis and recognition, pp. 683-687. IEEE, 2013.

[16] Avyodri, Ridvy, Samuel Lukas, and Hendra Tjahyadi. "Optical Character Recognition (OCR) for Text Recognition and its Post-Processing Method: A Literature Review." In 2022 1st International Conference on Technology Innovation and Its Applications (ICTIIA), pp. 1-6. IEEE, 2022.

[17] Dipu, Nadim Mahmud, Sifatul Alam Shohan, and K. M. A. Salam. "Bangla optical character recognition (ocr) using deep learning based image classification algorithms." In 2021 24th International Conference on Computer and Information Technology (ICCIT), pp. 1-5. IEEE, 2021.

[18] Memon, Jamshed, Maira Sami, Rizwan Ahmed Khan, and Mueen Uddin. "Handwritten optical character recognition (OCR): A comprehensive systematic literature review (SLR)." IEEE Access 8 (2020): 142642-142668.

[19] Martínek, Jiří, Ladislav Lenc, Pavel Král, Anguelos Nicolaou, and Vincent Christlein. "Hybrid training data for historical text OCR." In 2019 International Conference on Document Analysis and Recognition (ICDAR), pp. 565-570. IEEE, 2019.

[20] Breuel, Thomas M. "The OCRopus open source OCR system." In Document recognition and retrieval XV, vol. 6815, pp. 120-134. SPIE, 2008.

[21] Smith, Ray W. "History of the Tesseract OCR engine: what worked and what didn't." In Document Recognition and Retrieval XX, vol. 8658, p. 865802. SPIE, 2013.

[22] Lu, Zhidong, Richard Schwartz, Premkumar Natarajan, Issam Bazzi, and John Makhoul. "Advances in the bbn byblos ocr system." In Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99 (Cat. No. PR00318), pp. 337-340. IEEE, 1999.

[23] Espana-Boquera, Salvador, Maria Jose Castro-Bleda, Jorge Gorbe-Moya, and Francisco Zamora-Martinez. "Improving offline handwritten text recognition with hybrid HMM/ANN models." IEEE transactions on pattern analysis and machine intelligence 33, no. 4 (2010): 767-779.
[24] Smith, Ray. "An overview of the Tesseract OCR engine." In Ninth international conference on document analysis and recognition (ICDAR 2007), vol. 2, pp. 629-633. IEEE, 2007.
[25] Zhang, Jiaming, Jitao Sang, Kaiyuan Xu, Shangxi Wu, Xian Zhao, Yanfeng Sun, Yongli Hu, and Jian Yu. "Robust captchas towards malicious OCR." IEEE Transactions on Multimedia 23 (2020): 2575-2587.