ICEST-2K24 [International Conference on Engineering, Science and Technology]
In Association with
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN - 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

AI for Energy Efficiency

Mayura Shelke, Hemant Bhase, Rushikesh Phalake, Alam Shaikh
Department of Artificial Intelligence & Data Science, AISSMS Institution of Information Technology, Pune, Maharashtra, India

ABSTRACT

Efficient energy consumption management is imperative amidst escalating global energy demands and environmental concerns. This paper explores the application of Artificial Intelligence (AI) techniques in predicting real-time global active power consumption, focusing on diverse machine learning models and comprehensive parameter sets. Leveraging a household power consumption dataset, we evaluate the efficacy of XGBoost regressor, Linear regression, Random Forest, and ridge regressor models. Results shows the effectiveness of ensemble methods like XGBoost and Random Forest, LSTM model, offering lower errors and higher predictive accuracy. Additionally, we propose a user-friendly web interface featuring a chatbot for seamless access to real-time energy consumption predictions, aiding in energy efficiency strategies. Through empirical insights and practical solutions, this research contributes to the field of AI-driven energy optimization.

Keywords- Energy consumption prediction, Artificial Intelligence, Machine learning models, Global active power, Household power consumption dataset, Linear regression, XGBoost regressor, Random Forest, Ridgeregressor, LSTM model, Ensemble methods, Web interface, Chatbot, Energy efficiency strategies.

I. INTRODUCTION

The pursuit of efficient energy consumption and optimization has become a critical endeavor in the face of mounting global energy demands and environmental concerns [1]. Leveraging Artificial Intelligence (AI) techniques presents a promising avenue for addressing these challenges by enabling predictive analytics and strategies. In this research paper, we delve into the realm of AI driven energy optimization, focusing especially on the real time prediction of global active power consumption in real-time or over specified time intervals. Our research aims to contribute to the advancement of AI-driven energy efficiency solutions for practical implementation. By harnessing the power of AI, we endeavor to not only optimize energy consumption but also reduce carbon emissions and mitigate the environmental impact associated with energy production and usage. Through empirical insights and practical solutions, we aspire to pave the way for a sustainable and greener future. Prior studies have explored various methodologies for predicting energy consumption [2][3][5], comprising sophisticated machine learning methods as well as conventional statistical methods, it includes linear regression, XGBoost regressor, random forest, and ridge regressor, etc. Significant contributions have been shown where tree-based models among other models used for energy consumption prediction have shown promising results [2], but there are also cases where the XGB regressor was outperforming the tree-based model and other models [3].

Depending on the data, the results may occur differently, so the quality of data is always a concern while training any model. However, a notable gap persists in the literature concerning the integration of AI models tailored specifically for energy prediction, utilizing comprehensive sets of parameters. While some studies have touched upon predictive modeling for energy consumption, few have thoroughly investigated the comparative performance of diverse machine learning models in this context. Furthermore,

the existing literature lacks a comprehensive examination of the impact of minute-level data on predictive accuracy and the scalability of such models for practical deployment.

The primary objective of our research is twofold: first, to evaluate the efficacy of various machine learning models in predicting global active power consumption based on a rich feature set encompassing global and sub-meter parameters, we will be using individual household power consumption dataset as it provides more flexible results [4]. Second, we aim to design a user-friendly interface that enables users to seamlessly access real-time energy consumption predictions or upload csv data file for customized predictions over specified time horizons. Through this research, we seek to bridge the identified knowledge gap by providing empirical insights into the performance of different AI models for energy prediction and by offering monitoring power consumption features as per the request of user [4].

To achieve our objectives, we will conduct a comparative analysis of the aforementioned machine learning models using a dataset comprising global active power and associated parameters. We will preprocess the data to handle missing values, normalize features, and engineer relevant features for model training. Subsequently, we will train each model using a portion of the dataset and evaluate their performance on a held-out test set using appropriate metrics such as mean absolute error and root mean squared error. Based on the results, we will select the most effective model for predicting global active power consumption. Additionally, we will develop a user-friendly web interface featuring a chatbot to facilitate interaction and provide intuitive access to energy consumption predictions, also providing users with useful suggestions in the interface including carbon emission and monthly cost based on hourly predictions of global active power or based on csv data, energy consumption patterns can be generated. Through this approach, we aim to contribute to the field of AI-driven energy efficiency while offering a practical solution for real-world implementation.

II. LITERATURE SURVEY

The paper titled "Prediction and Analysis of Household Energy Consumption by Machine Learning Algorithms in Energy Management" explores the application of machine learning algorithms to forecast household energy consumption based on various factors such as temperature, humidity, and time of day. The research emphasizes the importance of accurate energy consumption prediction for effective energy management in residential and commercial properties. The study delves into the significance of short term energy consumption forecasting in the context of power distribution system planning and operations. Various machine learning algorithms, including Linear Regression, Lasso Regression, Random Forest, Extra Tree Regressor, and XG Boost, are deployed to analyze and predict household energy consumption patterns. The paper highlights the impact of weather, time, and socio-economic factors on energy consumption and the challenges of accurately anticipating consumption due to these diverse influences. The results show that tree-based models, particularly Extra Trees Regressor, yield the best performance in predicting household energy consumption, with a suggested R square value of 74.5% after hyper parameter tuning. Random forest and XGboost models also did well compared to linear and ridge regression models [2].

The research paper "Electricity Consumption Prediction Using Machine Learning" explores using machine learning to predict power usage. It highlights how important it is to predict electricity consumption correctly because it affects the environment, energy management, and costs. The study looks at different machine learning models, like linear regression, K Nearest Neighbors (KNN), XGBOOST, random forest, and artificial neural networks (ANN), to predict power usage. They judge these models based on how close their predictions are to the real values. They found that the KNN model did the best, getting it right about 91% of the time when predicting agricultural production. XGBoost also did well but not as good as KNN. One big challenge they faced was considering lots of factors like the time of year, time of day, and weather when predicting power usage. It's crucial to get these predictions right to manage energy well, save money, and help the environment because the demand for energy keeps going up [6].

The paper "Application of XGboost in Electricity Consumption Prediction" introduces a technique to predict power consumption using the XGBoost algorithm. Its goal is to address the challenges in predicting regional power consumption among various users. Experiments conducted with data from an industrial park demonstrate the effectiveness of the method. XGBoost is applied to develop prediction models for different user categories, leading to more accurate short-term load forecasts. Results indicate that XGBoost outperforms the random forest algorithm in reducing prediction errors, making it the preferred choice. Furthermore, the study employs the Maximum Information Coefficient (MIC) to analyze the relationship between different factors and power consumption [7].

III. PROPOSED SYSTEM

The proposed system aims to enhance energy optimization by integrating advanced AI techniques tailored for predicting global active power consumption. Unlike current approaches, our system utilizes a diverse feature set, including global and sub-meter parameters, leveraging individual household power consumption data for enhanced flexibility and accuracy. Through a comprehensive comparative analysis of machine learning models, we will identify the most effective model for precise predictions. Additionally, our system will feature a user-friendly interface with a chatbot, facilitating easy access to real-time energy consumption predictions or customized predictions based on uploaded CSV data. This interface will not only provide energy consumption predictions but also offer insights into carbon emissions, monthly costs, and energy consumption patterns, empowering users with valuable information for efficient energy management. This innovative approach aims to address existing gaps in knowledge and contribute to the advancement of AI-driven energy efficiency solutions for practical implementation.

IV. OPERATIONAL WORKFLOW

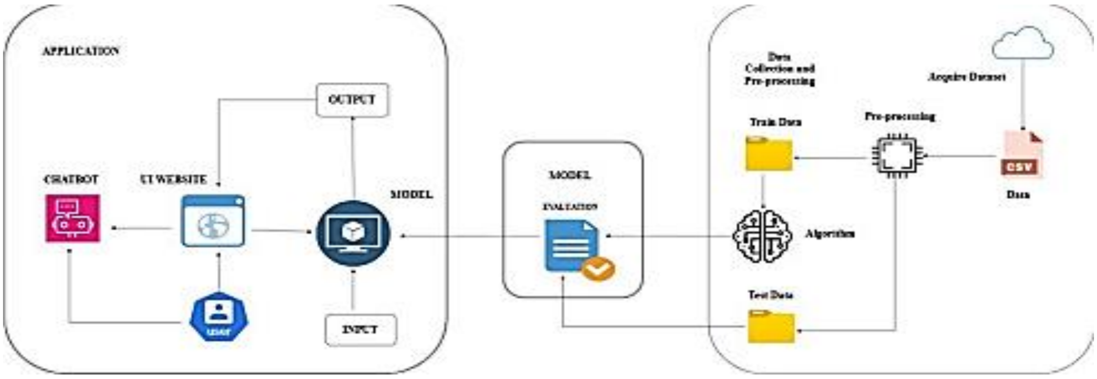

Fig. [1] Operational Workflow

The operational workflow of the tool is depicted in Figure [2], illustrating three main modules:

- **Acquire Dataset:** This step involves obtaining the dataset from an external source. The dataset could be related to user interactions, historical data, or any other relevant information.
- **Data Collection:** In this phase, raw data is collected. It could be user-generated content, sensor readings, or any other relevant data.
- **Preprocessing:** Data preprocessing is crucial for cleaning and preparing the dataset. Common preprocessing steps include handling missing values, removing duplicates, and converting data types. Any nan value present in the dataset has been dropped to make data more suitable for model. The preprocessing task include,
  - Feature Extraction: Creating new features from existing ones (e.g., extracting keywords from text). • Feature Selection: Choosing relevant features to improve model performance.
  - Data visualization: visualizing the data using different types of plots and heatmaps to understand the relationship between the variables.
- **Model Training:** We train our models based on pre-processed data the Sklearn library is useful for this work as it includes wide range of models that we can use
- **Algorithms**: Algorithms are the heart of the model. They learn patterns from the data.
- **Test Data:** Test data is a subset of the dataset that is set aside to assess the performance of a trained model. It is used to evaluate the model's ability to generalize to new, unseen data.
- **Model:** The model processes input data based on the learned patterns. It could be a recommendation model, sentiment analysis model, fraud detection model, etc.
- **Application:** The user-facing part of the system.
  - Chatbot: Interacts with users through natural language. Responds to queries, provides information, and performs tasks.
  - Website: Offers a graphical interface for users. Displays recommendations, visualizations, or any relevant content. Users can interact directly with the website.

V. METHODOLOGY

5.1 Dataset

The Household Power Consumption dataset is downloaded from UCI repository as shown in Figure [1], It is a valuable resource for studying energy efficiency and consumption patterns. It includes several key columns:

Date and Time: Provide timestamps for each data point, enabling trend analysis over time. Global_ active_ power: Indicates total active power consumed, useful for understanding overall energy usage trends and identifying optimization opportunities. Global_ reactive_ power Helps analyze the reactive power component, important for assessing power quality and system efficiency. Voltage: Provides voltage levels, which can impact appliance efficiency and indicate electrical system issues. Global_ intensity: Represents total current intensity, closely linked to active power and voltage levels, used to assess system load. Sub_metering_1, Sub_metering_2, Sub_metering_3: Show active energy consumption in specific areas, aiding in identifying energy-intensive appliances and areas for energy-saving measures. This dataset facilitates insights into household energy usage, aiding in consumption optimization and promoting energy efficiency practices.

The Household Power Consumption dataset in CSV format is a valuable asset for AI applications in energy efficiency projects [2]. It aids in preparing and refining data for AI analysis, and in creating new features for better model performance. AI models trained on this dataset can predict energy consumption patterns, recommend energy-saving strategies, and detect anomalies in usage. By analyzing this data, AI systems can identify peak usage times, suggest ways to save energy, and improve overall energy efficiency.

5.2 Machine Learning Model

Linear regression is a fundamental tool in statistical modelling, it provides the relationship between the variables. At its core, it establishes a linear association between a dependent variable, often denoted as y, and one or more independent variables, typically represented as x [8]. Imagine it as fitting a line through a scatterplot of data points, aiming to capture the overall trend in the observations. Through this method, we seek to discern how changes in the independent variable(s) correspond to alterations in the dependent variable's value. To achieve this, linear regression employs various techniques, such as ordinary least squares (OLS) or gradient descent [9]. These methods iteratively adjust coefficients to minimize the discrepancy between predicted and observed values, it can be evaluated by metrics like the Mean Squared Error (MSE). The ultimate goal is to find the line that best represents the data, to make prediction accurately we evaluate the model.

Ridge regression emerges as a refinement of traditional linear regression, designed to address multicollinearity issues inherent in datasets with correlated predictors [10]. By introducing a penalty term, often in the form of the L2 norm of the coefficient vector, ridge regression effectively constrains the magnitude of coefficients, thus reducing their sensitivity to fluctuations in the data. This regularization technique enhances the stability and generalizability of the model, particularly when dealing with high-dimensional datasets.

In the realm of predictive modelling, Extreme Gradient Boosting (XGBoost) has garnered significant attention for its remarkable efficiency and effectiveness, especially in regression tasks [11]. By iteratively improving upon weak learners, typically shallow decision trees, XGBoost minimizes a chosen objective function through gradient descent optimization. This approach, which incorporates both first-order and second-order gradients, allows for precise parameter updates, resulting in models with superior predictive power. We also used LSTM model for data analysis and prediction of power consumption.

Another powerful tool in the regression arsenal is Random Forest Regression, an ensemble learning technique that harnesses the collective wisdom of decision trees [12]. By aggregating the predictions of multiple trees trained on random subsets of the data, Random Forests mitigate overfitting and enhance model robustness. Moreover, their ability to handle missing values and outliers makes them particularly well-suited for real-world regression challenges.

These methodologies represent just a glimpse into the vast landscape of regression analysis, each offering unique strengths and applications in various domains. Whether it's uncovering hidden patterns in financial data, predicting housing prices based on demographic variables, or understanding the factors driving customer satisfaction, regression techniques provide invaluable insights that drive decision-making and innovation.

5.3 Training-Validation-Test Split

We divided our household dataset into 2 parts one will go for training and another for testing purpose we can use python library for this task, we can choose the partition of dataset such as 70-30 or 80-20 percentage as per our need we choose, the split ensures that the models are trained on a sufficient dataset and evaluated on unseen data [13].

5.3.1 Performance Metrics

The performance metrics helps in finding the best model among other models which will further help for prediction task we use MAE and MSE for comparing models which are simple to understand. MSE calculates the average of the squared errors, giving more weight to larger errors. This makes it useful for penalizing models that have large deviations from the actual values, providing a more comprehensive view of the model's performance. RMSE help in providing an interpretable measure of the average error magnitude. It is useful for understanding the typical size of the errors in the model's predictions. Coefficient of Determination (R squared): R2 measures how much of the variance in energy consumption is explained by the model's features. Higher values indicate a better fit. These metrics aid in refining models for more accurate energy consumption predictions, crucial for effective energy efficiency strategies.

These metrics are crucial for evaluating AI models' accuracy in predicting energy consumption, aiding in model selection and performance improvement. They also help detect overfitting and guide adjustments to enhance predictive power. Ultimately, these metrics inform the optimization of energy efficiency strategies by assessing the models' effectiveness in predicting consumption patterns.

5.3.2 Comparative Analysis

We initially assessed each model's performance utilizing the specified metrics. Afterward, we implemented a five-fold cross validation to measure the models' ability to predict global active power usage [14]. To gauge the precision of the models' predictions on the test data, we carefully examined the MAE, MSE, RMSE, and R2 values linked with each model.

5.4 Development of User Interface

Flask, a lightweight WSGI app framework, provides developers the ability to rapidly develop web applications by focusing solely on core functionalities. Whether utilized for backend or frontend purposes, it provides essential features like an interactive debugger, routing system for endpoints, HTTP utilities for managing entity tags, cache controls, dates, cookies, and more [15].

At the front end on website, one should capable of working with different front end languages such as Html, CSS and JavaScript and additionally developer should have knowledge of Bootstrap, Angular JS, React JS to make the development easier and attractive. Using frameworks the development of webpages have become much easier and has become more dependent on frameworks for making different kinds of websites [16]. In the realm of frontend development, where the visitor's interaction with the website is crafted, proficiency in three fundamental languages HTML, CSS, and JavaScript is indispensable. However,[i](Zhang et al., 2022) mastery of additional frameworks and libraries, such as Bootstrap, AngularJS, ReactJS, is equally essential. These tools empower developers to craft visually captivating content adaptable across diverse devices. The proliferation of frameworks has undeniably streamlined web programming, ushering in an era where full-stack web design is within reach. Harnessing frameworks, full-stack developers navigate the landscape of website creation, optimization, and management with unparalleled efficiency and efficacy.

AI chatbots are used to engage with users and help them with queries asked, it has ability to understand the human text and provide the answers accordingly there are three types of categories: first is chatbot that uses deep learning, second one is used for end to end system and third is sequence to sequence models [17]. Real-time data access entails the capability to seamlessly retrieve and process live data streams as they are generated. In the context of energy efficiency, this translates to instant access to critical metrics such as current energy consumption levels, power fluctuations, and equipment performance indicators. By leveraging real-time data, AI algorithms can promptly detect anomalies, identify inefficiencies, and trigger automated responses to optimize energy usage in real

time. Conversely, historical data access involves delving into archived data sets spanning days, weeks, months, or even years. This historical perspective offers valuable insights into long-term energy consumption patterns, seasonal variations, and the efficacy of past energy-saving initiatives. Analysing historical data enables AI algorithms to identify recurring trends, forecast future energy demands, and prescribe proactive measures for sustainable energy management. Combining real-time and historical data access capabilities empowers AI systems to operate with unparalleled foresight and agility in optimizing energy efficiency.

By continuously assimilating real-time data inputs and cross-referencing them with historical trends and patterns, AI algorithms can adaptively refine their energy-saving strategies over time. In the realm of electrical data analysis for energy efficiency, the seamless integration of real-time and historical data access mechanisms is paramount. It enables AI systems to function as proactive stewards of energy resources, dynamically adjusting operational parameters, and fostering a sustainable approach to energy consumption and management [18].

VI. RESULTS

When comparing the performance of different regression models as shown in Figure [3] , it's evident that XGB Regressor and Random Forest outperform Linear Regression and Ridge Regression across multiple evaluation metrics. Both XGB Regressor and Random Forest demonstrate lower MAE, MSE, RMSE, and higher R-Square values, indicating better predictive accuracy and goodness of fit. This suggests that ensemble methods like XGB Regressor and Random Forest might be more suitable for the given dataset compared to traditional linear regression approaches.

Models	MAE	MSE	RMSE	R-Square
Linear Regression	0.0281	0.0019	0.0438	0.9985
XGB Regressor	0.0209	0.0012	0.0349	0.9990
Random Forest	0.0211	0.0013	0.0369	0.9989

Ridge Regression	0.0281	0.0019	0.0438	0.9985

Fig. [2] Model performance Evaluation

Fig [3] Input figure

The simple interface of the web page ensures that users can easily enter the required parameters, which are crucial for accurate predictions and calculations. By focusing on key values such as reactive power and sub-meter readings in watts, and global intensity in ampere units, the interface streamlines the data entry process. This approach not only enhances user experience but also promotes a deeper understanding of the energy consumption metrics involved. By presenting these complex concepts in a straightforward manner, users can make more informed decisions regarding their energy usage, leading to greater efficiency and cost savings.

Fig [4] Output Page

After inserting parameter values, as we shown in fig [5] the result page provides a comprehensive analysis of energy usage. The predicted global active power in watt units allows users to estimate their carbon footprint accurately. This estimation is crucial for understanding the environmental impact of energy consumption. Additionally, the calculation of monthly costs in Indian rupees provides users with a practical understanding of their energy expenses over time. This information can be valuable for budgeting and planning purposes, enabling users to make informed decisions about their energy usage.

Overall, the result page offers a holistic view of energy consumption, combining technical details with practical implications, to empower users in managing their energy resources efficiently.

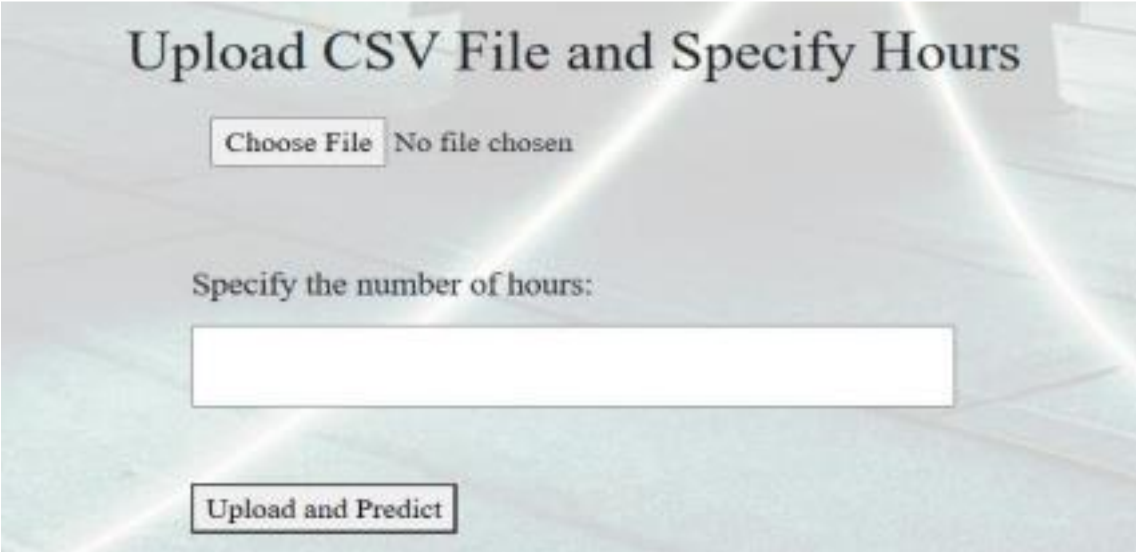

Fig [5] CSV file as an input

For getting the output in hour we can make an CSV file which must include the parameters that we used in input form and we can now choose that file to add as shown in Figure [6], we next specify the number of hours we want to predict power consumption i.e global active power, lastly, we click upload and predict to get the results.

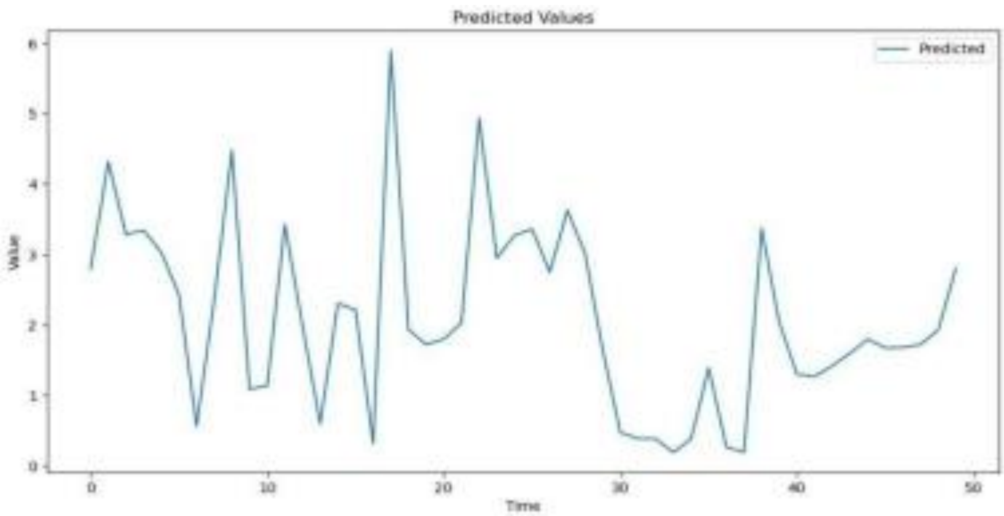

Fig [6] CSV Output

At last a user can see the power consumption as they requested for specific hours from this they can get an idea of their daily power consumption pattern and find if there is any anomaly present in their grid if the graph is abnormal, x-axis contains the time step and y axis has power consumption values in unit. This can help user to make new decision, rules, policies regarding their energy usage.

VII. CONCLUSION

In conclusion, this study demonstrates the effectiveness of AI-driven models, particularly ensemble methods like XGBoost and Random Forest, in predicting real-time energy consumption. Leveraging a household power consumption dataset, these models outperform traditional linear regression approaches, offering lower errors and higher predictive accuracy. Additionally, our proposed user-friendly web interface with a chatbot facilitates seamless access to energy consumption predictions, aiding in energy efficiency strategies. By bridging advanced AI algorithms with practical user interfaces, we contribute to the advancement of sustainable energy management practices.

VIII. REFERENCES

[1]. Aderibigbe, Adebayo & Ani, Emmanuel & Ohenhen, Peter & Ohalete, Nzubechukwu & Daraojimba, Donald. (2023). ENHANCING ENERGY EFFICIENCY WITH AI: A REVIEW OF MACHINE LEARNING MODELS IN ELECTRICITY DEMAND FORECASTING. Engineering Science & Technology Journal. 4. 341-356. 10.51594/estj.v4i6.636.

[2]. Rambabu, Muppidi & Nuvvula S S, Ramakrishna & Polamarasetty, P. (2022). Prediction and Analysis of Household Energy Consumption by Machine Learning Algorithms in Energy Management. E3S Web of Conferences. 350. 02002. 10.1051/e3sconf/202235002002.

[3]. Abdelfattah, E., & Bowlyn, K. (2023). Application of Machine Learning Models on Individual Household Electric Power Consumption. 2023 IEEE World AI IoT Congress (AIIoT), 0143-0146.

[4]. A. Loginov, M. I. Heywood and G. Wilson, "Benchmarking a coevolutionary streaming classifier under the individual household electric power consumption dataset," 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 2016, pp. 2834-2841, doi: 10.1109/IJCNN.2016.7727557.

[5]. Elbeltagi, Emad. (2021). Predicting energy consumption for residential buildings using ANN through parametric modeling. Energy Reports. 7. 10.1016/j.egyr.2021.04.053.

[6]. Reddy, G. & Aitha, Lakshmi & Poojitha, Ch & Shreya, A. & Reddy, D. & Meghana, G.. (2023). Electricity Consumption Prediction Using Machine Learning. E3S Web of Conferences. 391. 10.1051/e3sconf/202339101048.

[7]. Dong, Danhuang & Wen, Fan & Zhang, Yihong & Qiu, Wenxin. (2023). Application of XGboost in Electricity Consumption Prediction. 1260-1264. 10.1109/ICETCI57876.2023.10176934.

[8]. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. Springer.

[9]. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

[10]. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.

[11]. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).

[12]. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32

[13]. Toleva, Borislava. (2021). The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems. Business Systems Research Journal. 12. 10.2478/bsrj-2021-0015.

[14]. Nti, Isaac & Nyarko-Boateng, Owusu & Aning, Justice. (2021). Performance of Machine Learning Algorithms with Different K Values in K-fold Cross-Validation. International Journal of Information Technology and Computer Science. 6. 61-71. 10.5815/ijitcs.2021.06.05.

[15]. Devndra Ghimire , Comparative study on Python web frameworks: Flask

[16]. Wijaya, Eric & Kosasi, Sandy & David,. (2021). Implementasi Aplikasi Web Full Stack Pendataan Cloversy.id. Jurnal Sisfokom (Sistem Informasi dan Komputer). 10. 320-327. 10.32736/sisfokom.v10i3.1293.

[17]. Mahmoud, Maha & Kumar, Ramakrishna. (2020). A Review on Chatbot Design and Implementation Techniques. International Journal of Engineering and Technology. 7. 2791.

[18]. Zhang, Y & Sun, S & Yu, P & Wang, Y & Li, S. (2018). Research on remote test data access method of new energy grid connected generation based on cloud platform. IOP Conference Series: Earth and Environmental Science. 188. 012114. 10.1088/1755-1315/188/1/012114.