Easy Power Perdictor:Machin learning model to predict electricity consumption

Kanchan A.Khedikar Assistant Professor,Walchand Institute of Technology,Solapur,Maharashtra,India Email: [kakhedikar@witsolapur.org](mailto:kakhedikar@witsolapur.org)

Mr.Shreyansh Kahate UG Scholars, Walchand Institute of Technology, Solapur, Maharashtra, India Email: [shreyanshk29@gmail.com](mailto:shreyanshk29@gmail.com)

Mr.Vipul Ghodke UG Scholars, Walchand Institute of Technology, Solapur, Maharashtra, India Email: [vipulghodke199@gmail.com](mailto:vipulghodke199@gmail.com)

Miss.Shreya Gandhi UG Scholars, Walchand Institute of Technology, Solapur, Maharashtra, India Email: [shreyakishorgandhi@gmail.com](mailto:shreyakishorgandhi@gmail.com)

Miss.Nisha Gurav UG Scholars, Walchand Institute of Technology, Solapur, Maharashtra, India Email: [nishagurav79@gmail.com](mailto:nishagurav79@gmail.com)

Abstract:

"Easy Power Predictor" is a machine learning model designed to forecast electricity consumption by integrating various weather and holiday factors. The objective is to enhance electricity distribution efficiency, reducing losses, and increasing cost-effectiveness by supplying optimal power levels. The dataset, sourced from Kaggle, underwent thorough preprocessing, merging, and cleaning to integrate data from diverse sources, including weather and holiday data. The project utilized machine learning algorithms such as Linear Regression, Random Forest, and XGBoost for prediction tasks. These models were trained, validated, and evaluated using different data subsets, employing metrics like Mean Absolute Error , R-squared Value , Mean Squared Error and Root Mean Squared Error to assess performance. Through a combination of visual analysis and rigorous model evaluation, the project identified Random Forest as the most effective model for accurate electricity consumption forecasting. This predictive tool holds significant potential for optimizing power distribution strategies, reducing energy costs, and enhancing overall efficiency in electricity management.

Keywords— Easy power predictor,machine learning model,random forest algorithm.

I Introduction

The "Easy Power Predictor" project aims to address the complexities of electricity production and consumption dynamics in today's energy landscape, particularly with the rising integration of renewable energy sources. By leveraging machine learning techniques, this project seeks to develop a predictive model that accurately forecasts energy demand, bridging the gap between electricity supply and demand. The predictive model considers various influencing factors such as weather conditions, time of day, and holiday indicators, which play pivotal roles in determining electricity usage.

To achieve this, the project begins with comprehensive data acquisition and processing from diverse sources like smart meter readings, weather data, and holiday schedules. The initial raw data, often unstructured and disorganized, undergoes thorough cleaning and transformation to create a coherent dataset suitable for analysis. This involves handling missing values, transforming date and time data, and merging data from different sources to form an integrated dataset.

Following data preparation, the project employs exploratory data analysis (EDA) to uncover patterns, correlations, and trends in the data. Visualizations like histograms, scatter plots, and correlation heatmaps offer insights into the relationships between different variables, shedding light on the impact of weather and holidays on electricity consumption.

With a refined dataset ready for modeling, the project explores various machine learning algorithms, including Linear Regression, Random Forest, and XGBoost, to forecast energy usage. Each model is trained and evaluated using performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared values. Through feature engineering and model development, the project aims to optimize energy management strategies, promising proactive resource allocation, reduced energy wastage, and enhanced efficiency in utility operations.

Ultimately, the project's goal is to identify the most effective model for predicting electricity consumption, providing power providers with valuable insights to make informed decisions about energy management and distribution. The outcomes of the "Easy Power Predictor" project highlight its significance in the context of energy optimization, sustainability, and the efficient distribution of electricity to various regions, ensuring reliable and cost-effective energy supply.

## 

## **II Related Work**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sr.No | Title | Year | Publisher | Approach |
| 1 | Empirical Mode Decomposition Based Deep Learning for Electricity Demand Forecasting | 2018 | IEEE | Significance of electricity demand and prediction using Deep Learning and ANN |
| 2 | Investigation of Performance of Electric Load Power Forecasting in Multiple Time Horizons with New Architecture Realized in Multivariate Linear Regression & Feed-Forward Neural Network Techniques | 2020 | IEEE | Load forecasting using multivariate linear regression and feed-forward neural network techniques |
| 3 | Electricity consumption forecasting based on ensemble deep learning with application to the Algerian market | 2021 | ScienceDirect | Deep Learning with LSTM ,GRU, Models |
| 4 | Electric Heating and the Effects of Temperature on Household Electricity Consumption in South Africa | 2019 | Energy Institute at Haas | Log-linear correlation between cold temperature and electricity consumption |
| 5 | Research on Deep Learning Energy Consumption Prediction Based on Generating Confrontation Network | 2019 | IEEE | Deep learning, convolutional neural network model |

**Table 1 Comparison on Existing techniques**

**III Methodology**

1. Data Collection and Integration:

* Data Sources: Collect power consumption data from various sources, such as smart meters historical usage records.
* Data Aggregation: Aggregate data from multiple sources to create a comprehensive dataset. This involves combining readings at different time intervals.

2. Data Preprocessing:

* Data Cleaning: Remove noise, handle missing values, and correct errors in the dataset. This step ensures data quality and reliability.
* Normalization: Scale the data to ensure consistency across different ranges and units, improving model performance.
* Feature Engineering: Extract relevant features such as time of day, weather conditions, and device types. Create new features that could potentially enhance model accuracy.

3. Exploratory Data Analysis (EDA):

* Descriptive Statistics: Calculate summary statistics to understand the dataset's distribution.
* Visualization: Use plots and charts to identify patterns, trends, and outliers in the data.
* Correlation Analysis: Assess the relationships between different features and the target variable to identify significant predictors.

4. Model Building:

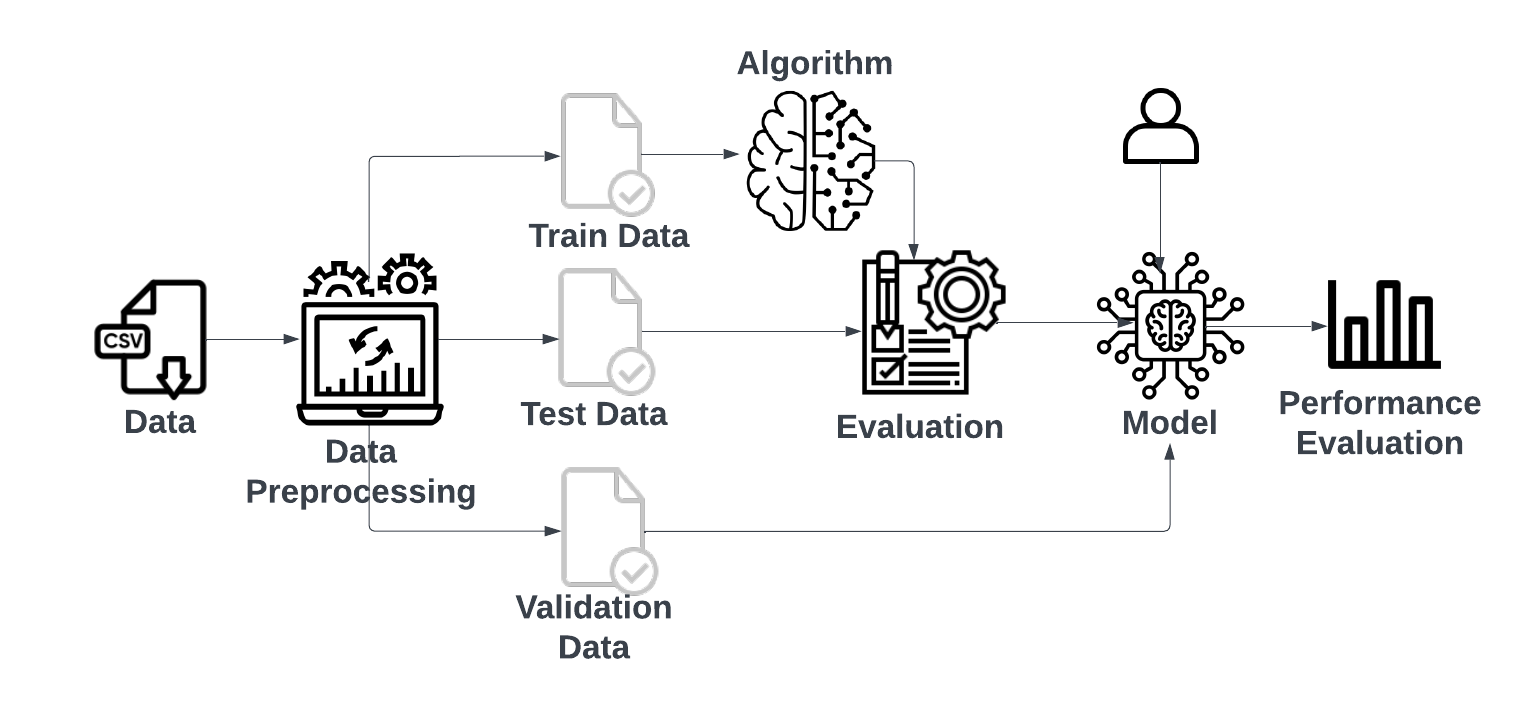
* Algorithm Selection: Experiment with various machine learning models such as linear regression, random forests, and XGBOOST
* Training and Testing: Split the dataset into training and testing sets to evaluate model performance. Train models on the training set and test their accuracy on the testing set.

5. Model Evaluation:

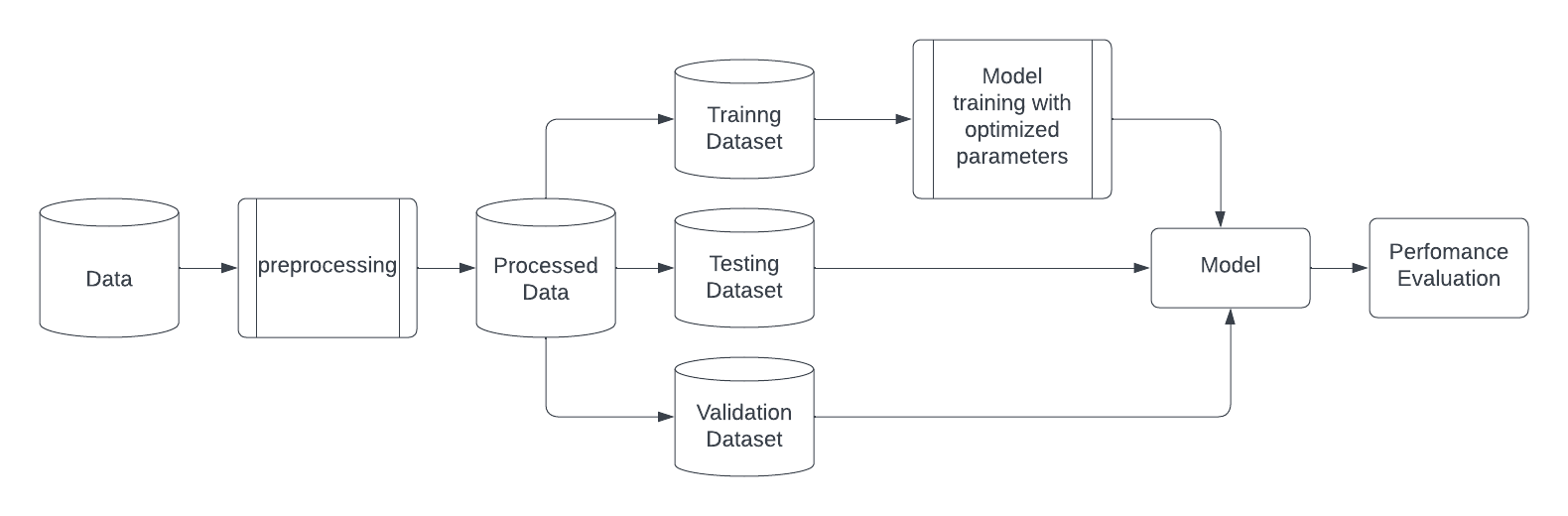
* Performance Metrics: Evaluate models using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared. These metrics help in comparing models and selecting the best one.

6. Model Selection:

* Comparison: Compare the performance of different models based on the evaluation metrics.
* Selection: Choose the best-performing model that provides the most accurate and reliable power consumption predictions.



**Fig 1:System Architecture of Easy power predictor**



**Fig 2:Flowchart of Easy power predictor**

**IV Implementation**

The implementation phase of the project was characterized by iterative development and evolving methodologies. This section outlines the detailed steps and procedures followed, including data preprocessing, Exploratory Data Analysis (EDA), model development and training, evaluation, and insights gained from the project's progression.

1. Data Preprocessing

Data Cleaning:

The dataset underwent cleaning to handle missing values, and outliers were removed to ensure data quality and reliability.

Feature Engineering:

Selected relevant features such as temperature, humidity, wind speed, and others for model input.

Date-wise Energy Sum Calculation:

Aggregated daily energy consumption by summing the energy values for each date.

2. Exploratory Data Analysis (EDA)

Data Visualization:

Visualized data distributions, relationships between features, and patterns in electricity consumption using graphs and charts.

Statistical Analysis:

Conducted statistical tests to identify correlations between features and electricity consumption, providing insights into potential relationships and patterns.

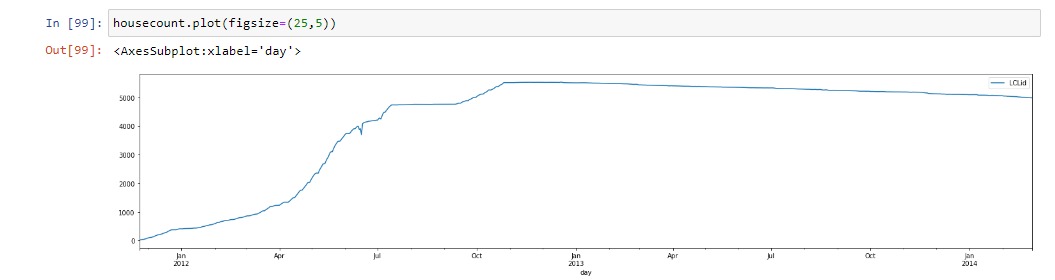


Fig 3:

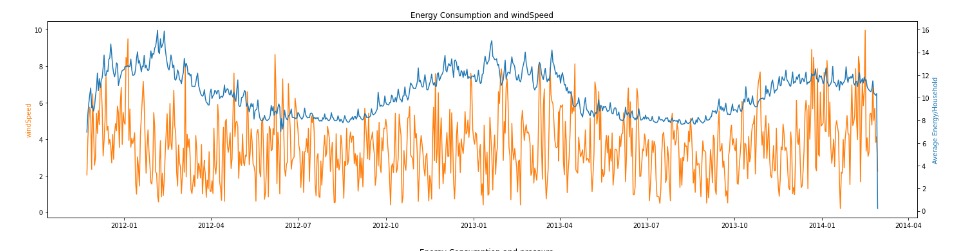


Fig 4: Energy consumption and windSpeed

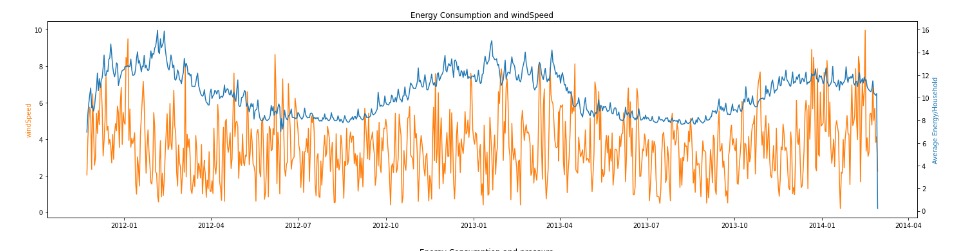


Fig 5: Energy consumption and windspeed

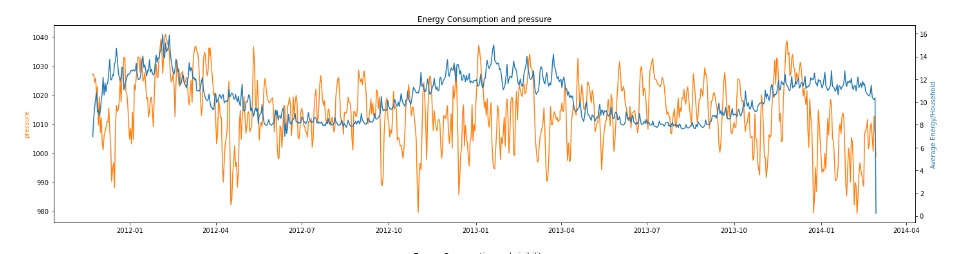


Fig 6: Energy consumption and pressure

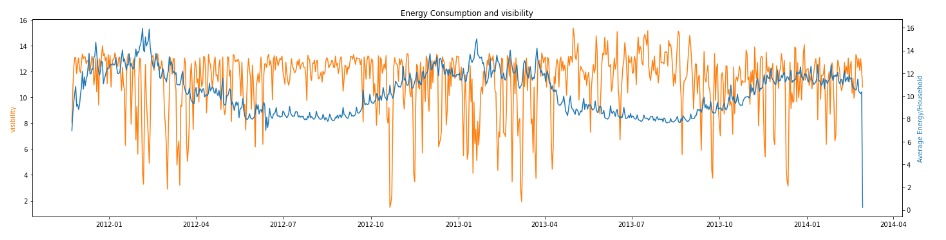


Fig 7: Energy consumption and visibility

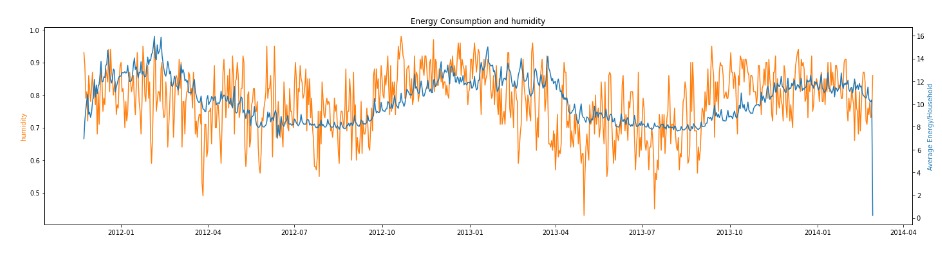


Fig 8: Energy consumption and humidity

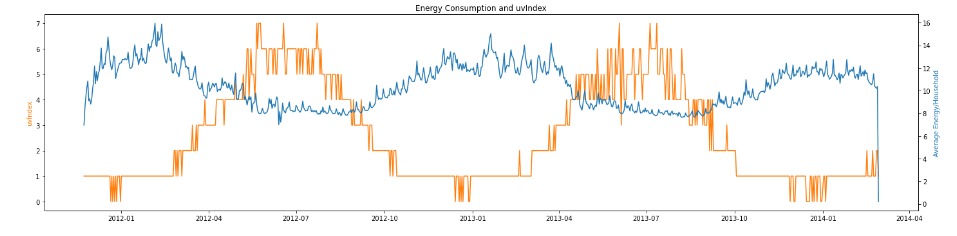


Fig 9: Energy consumption and uvIndex

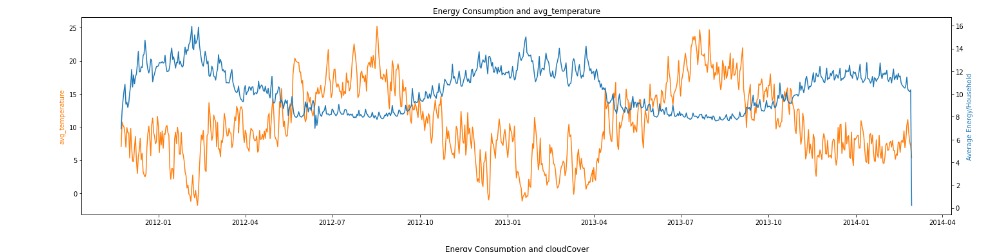


Fig 10: Energy consumption and average temperature

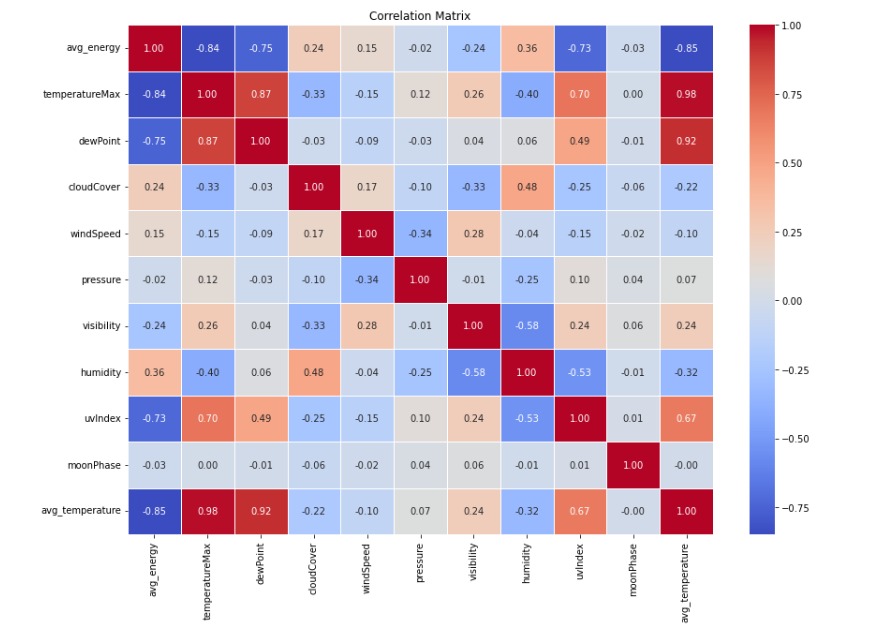


Fig 11: Correlation matrix between parameters

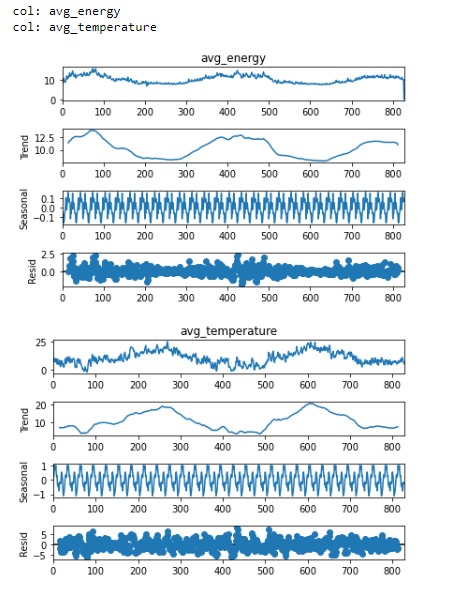


Fig 12:

3. Model Development and Training

3.1 Individual Prediction Approach

Description:

Initially, the focus was on predicting electricity consumption for individual smart meter IDs (LCLid) using their historical data.

Challenges:

This approach faced significant challenges due to the complexity and variability in individual consumption patterns, leading to less accurate predictions.

Insights:

It became evident that the individual prediction approach was not suitable for the given dataset's characteristics and would require more sophisticated techniques like deep learning to capture the underlying patterns.

3.2 Overall Prediction Approach

Description:

To overcome the limitations of the individual prediction approach, the focus shifted to predicting the total energy consumption for a given date by considering the number of active smart meter IDs.

Advantages:

This approach aimed to capture the collective consumption pattern rather than individual variations, resulting in more stable and accurate predictions.

Model Selection:

Tested different machine learning models, including Linear Regression, Random Forest, and XGBoost, to identify the most suitable model for this approach.

4. Model Evaluation and Validation

Validation Approach:

The project employed a simple hold-out validation method, splitting the data into training, validation, and test sets, allowing effective assessment of model performance on unseen data.

Model Performance Metrics:

The models were evaluated on the test dataset using performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2).

5. Results and Insights

5.1 Best Fit Model: Random Forest

5.2 Performance:

Random Forest has proven to be the best-performing model overall, demonstrating strong performance across all datasets (train, test, and validation).

5.3 Error Metrics:

The model maintains consistently low error metrics (Mean Absolute Error, Mean Squared Error, and Root Mean Squared Error) and high R-squared values across all data splits.

5.4 Generalization:

Random Forest shows good generalization, avoiding overfitting while still achieving high accuracy.

5.5 Comparison with Other Models:

Although XGBoost performs exceptionally well on the training data, its performance on the test and validation data is slightly lower than that of Random Forest. This indicates a potential for overfitting in XGBoost, which may compromise its performance on unseen data.

**V Testing Report**

* Dataset Split: The dataset was split into training, validation, and testing sets to ensure a comprehensive evaluation of the models' performance on different data subsets.
* Model Evaluation Metrics: The performance of the models was assessed using various metrics, including:
* Mean Absolute Error (MAE): Measures the average absolute difference between predicted and actual values.
* Mean Squared Error (MSE): Measures the average squared difference between predicted and actual values.
* Root Mean Squared Error (RMSE): The square root of MSE, providing a more interpretable measure of error.
* R-squared (R²): Indicates the proportion of variance in the dependent variable explained by the model.
* Model Performance and Findings
* Linear Regression

Training Performance: Linear Regression serves as a baseline model, providing reasonable predictions on the training data.

Testing and Validation Performance: The model performs adequately on the testing and validation data, but its simplicity may limit its ability to capture complex relationships between features and electricity consumption.

Conclusion: Linear Regression is a useful baseline model but may not be the most effective choice for this project.

* Random Forest

Training Performance: Random Forest exhibits strong performance on the training data, accurately capturing patterns in the data and avoiding overfitting.

Testing and Validation Performance: The model maintains consistent performance across testing and validation data, with low error metrics and high R-squared values.

Conclusion: Random Forest's balanced performance and generalization make it the most reliable and best-fit model for this project.

* XGBoost

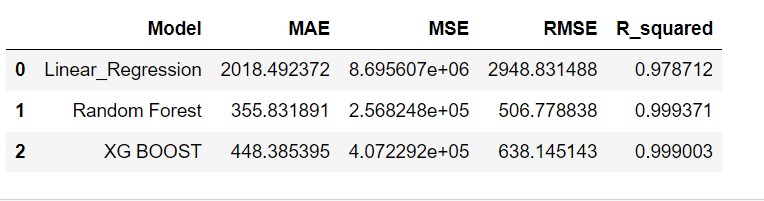
Training Performance: XGBoost demonstrates excellent performance on the training data, effectively learning complex relationships and patterns.

Testing and Validation Performance: While XGBoost performs well, its performance on testing and validation data is slightly lower than Random Forest, indicating a potential for overfitting.

Conclusion: Although XGBoost is a strong contender, its potential overfitting on unseen data may limit its effectiveness in this project.

VI Experimental Results

* Performance on Trained Data



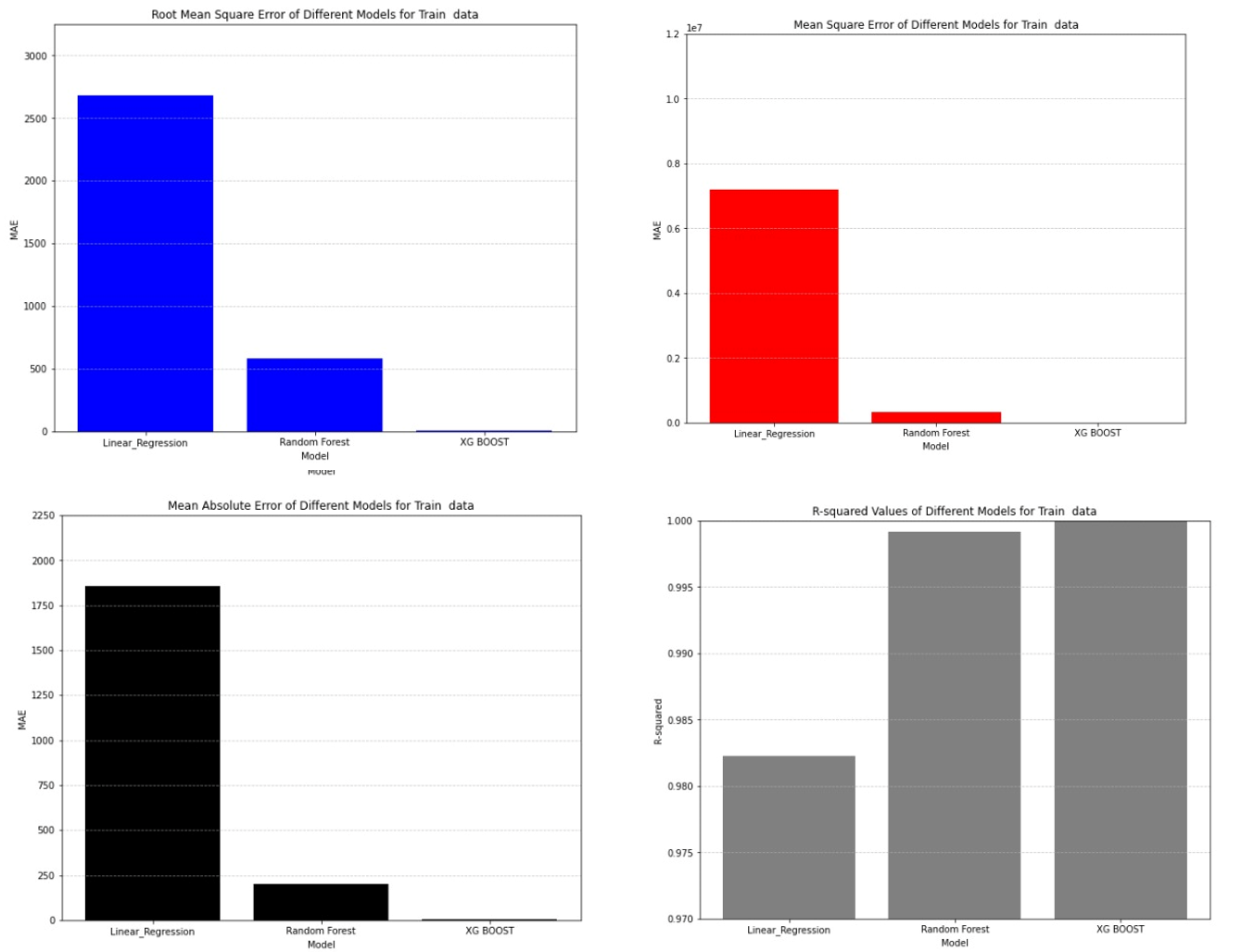
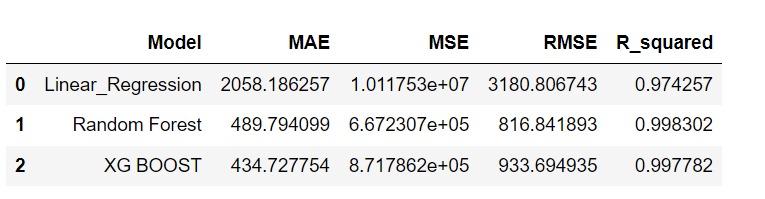


Fig 13: Performance on Trained data

* Performance on Test Data



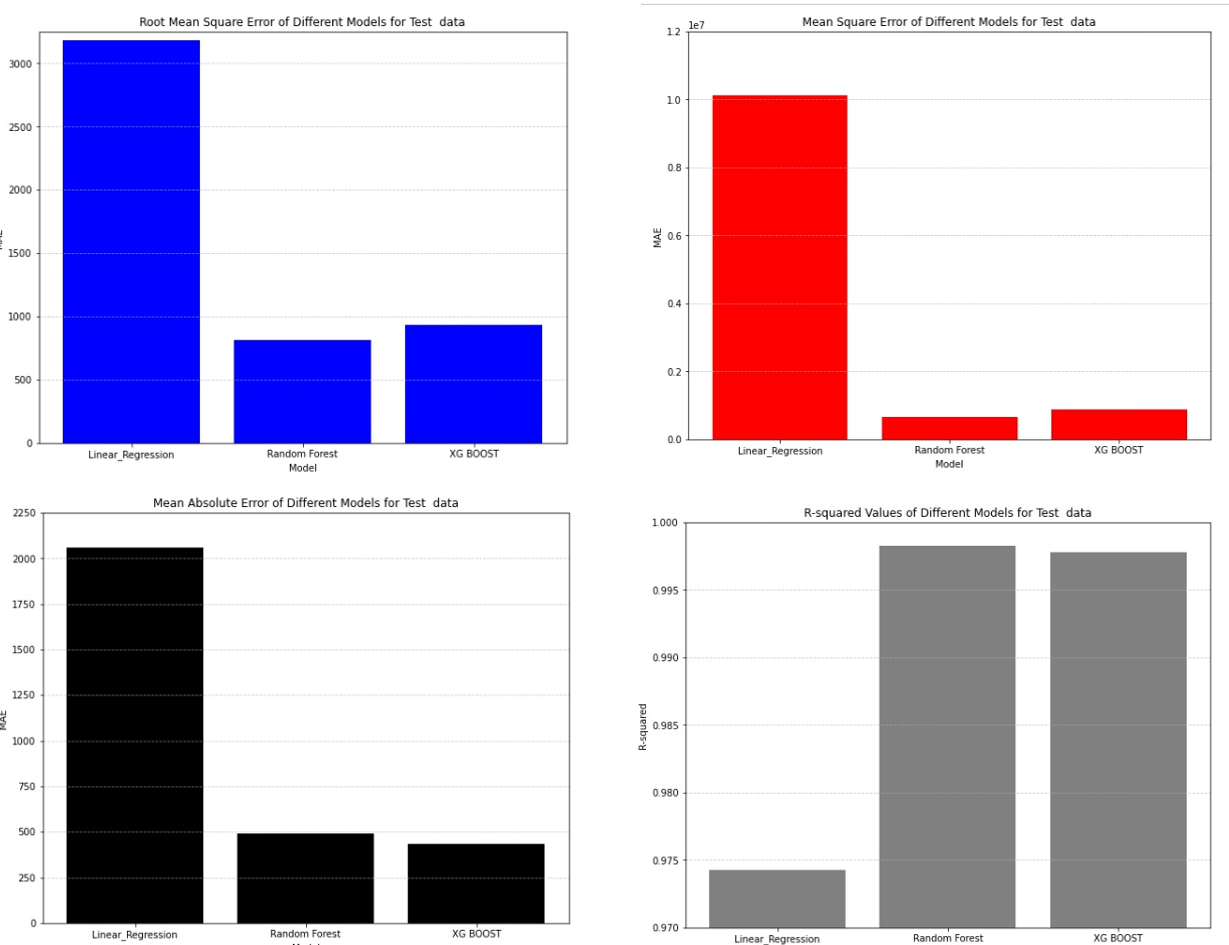
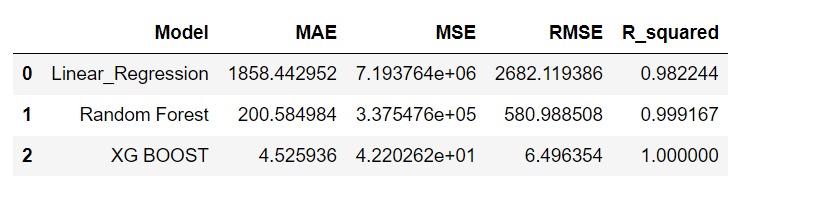


Fig 14: Performance on Test data

* Performance on Validation Data



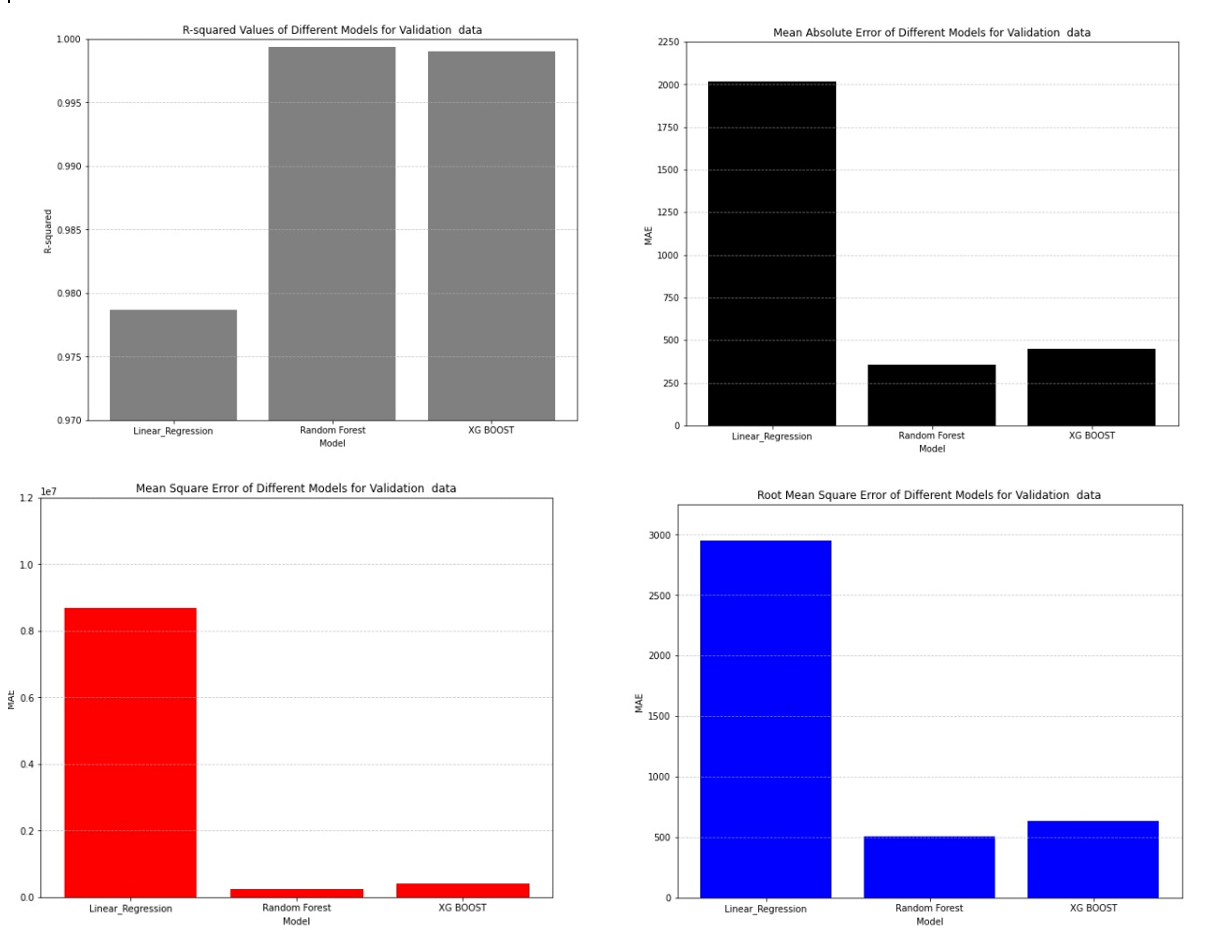


Fig 15: Performance on Validation data data

**VII Technology Stack**

The "Easy Power Predictor" project leverages a combination of technologies and tools to develop and deploy the predictive model for electricity consumption forecasting. The key technologies used include:

* Python: Python serves as the primary programming language for data preprocessing, modeling, and implementation of the predictive model. Python's extensive libraries and frameworks are utilized for machine learning tasks.
* Pandas: Pandas is used for data manipulation and preprocessing tasks, including handling of time-series data, feature extraction, and data cleaning.
* Scikit-Learn (sklearn): Scikit-Learn is employed for building and training machine learning models such as linear regression, decision trees, and ensemble methods. It provides efficient implementations of various algorithms and tools for model evaluation.
* Matplotlib and Seaborn: These libraries are used for data visualization, allowing for the creation of informative plots and graphs to analyze relationships between variables and visualize model outputs.
* Jupyter Notebook: Jupyter Notebook provides an interactive environment for code development, data exploration, and model prototyping. It facilitates a seamless workflow for experimenting with different approaches and visualizing results.
* Machine Learning Algorithms: Various regression algorithms (e.g., linear regression, decision trees) are implemented using Scikit-Learn to build the predictive model based on historical electricity consumption data and relevant features.

**VIII** **Conclusion**

In this project, we explored predictive analysis for electricity power consumption, focusing on its relationship with weather factors and holiday data. The goal was to optimize electricity distribution and make it more cost-effective by predicting energy demand accurately. We cleaned and preprocessed data from multiple sources, including Kaggle, weather, and holiday data, to create a cohesive dataset for analysis. Through visual analysis and feature engineering, we identified key variables such as temperature, humidity, wind speed, and holiday information. We trained and evaluated various machine learning models, including Linear Regression, Random Forest, and XGBoost. Among these, Random Forest proved to be the best-performing model, demonstrating strong generalization and low error metrics across train, test, and validation datasets. This project highlights the potential of using machine learning techniques to optimize electricity distribution, minimize wastage, and achieve cost savings for utility providers.

Random Forest Fitting

Trained data

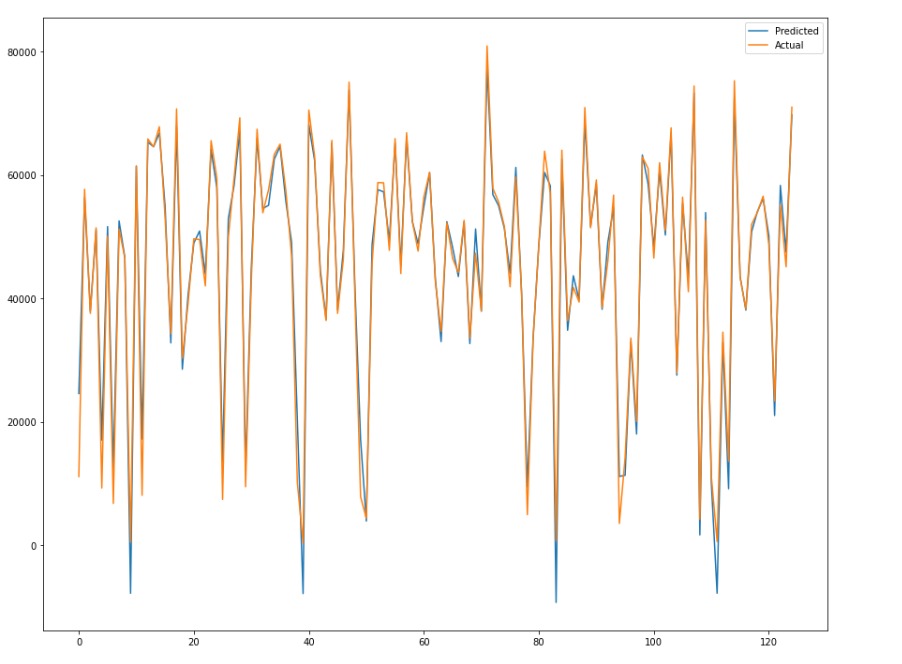


Fig 16: Random Forest fitting on Trained data

Test Data

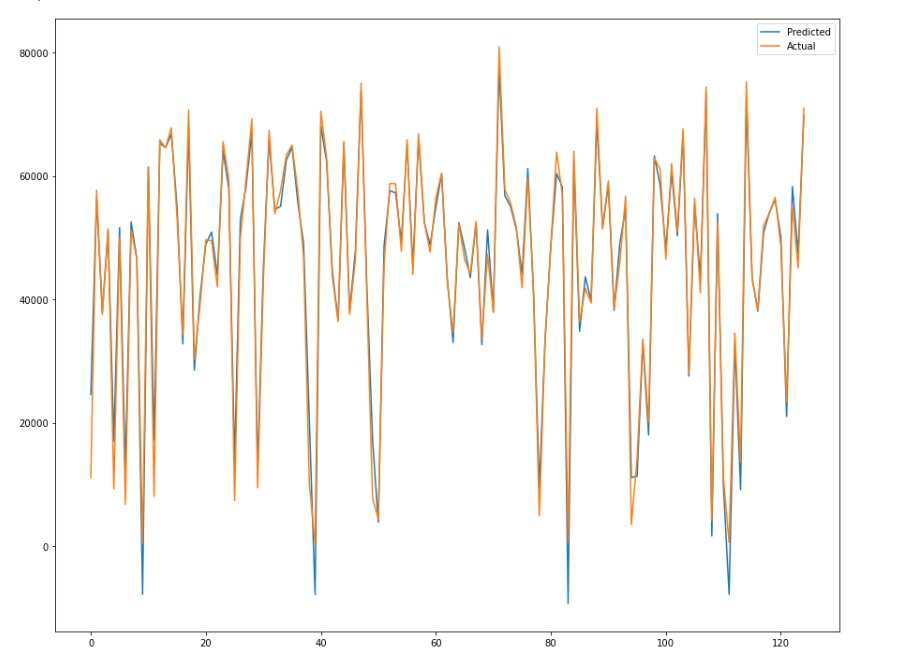


Fig 17: Random Forest fitting on Test data

Validation data

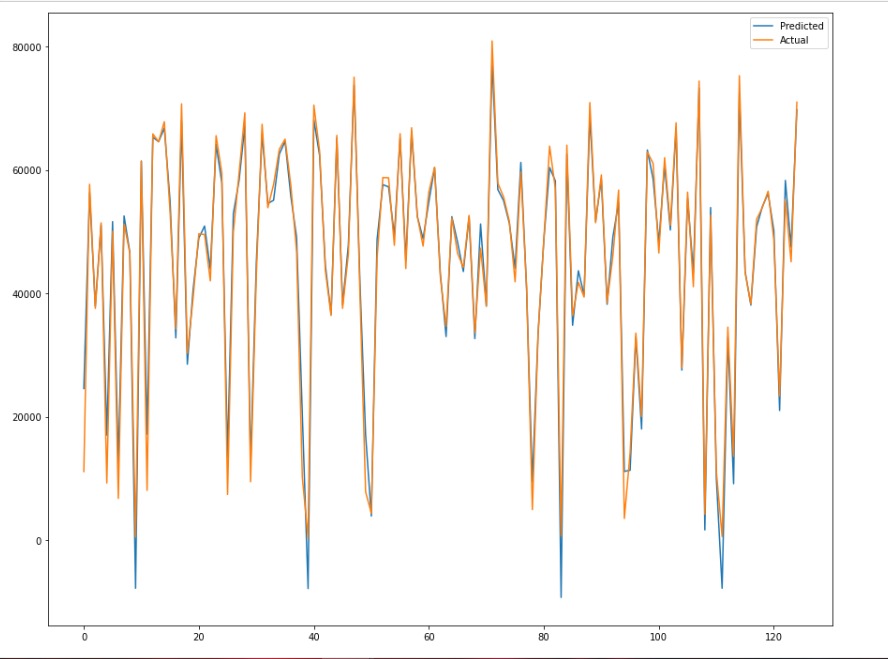


Fig 18: Random Forest fitting on Validation data

**References**

1. JATIN BEDI AND DURGA TOSHNIWAL, (Member, IEEE) Department of CSE, IIT Roorkee, Roorkee 247667, India,“ Empirical Mode Decomposition Based Deep Learning for Electricity Demand Forecasting”,Published:August 29, 2018.

Source: <https://ieeexplore.ieee.org/document/8449937>

1. M Vetri Selvi Research Scholar Electrical Engineering Department Indian Institute of Technology Delhi, India. Sukumar Mishra Professor & IEEE Member Electrical Engineering Department Indian Institute of Technology Delhi, India,“ Investigation of Performance of Electric Load Power Forecasting in Multiple Time Horizons with New Architecture Realized in Multivariate Linear Regression & Feed-Forward Neural Network Techniques”,Published in 2020

Source: <https://ieeexplore.ieee.org/document/9140425>

1. D. Hadjout a, 1 , J.F. Torres a, 1 , A. Troncoso a , A. Sebaa b, c , F. Martínez-Alvarez a, \* a Data Science and Big Data Lab, Universidad Pablo de Olavide, ES-41013, Seville, Spain b LIMED Laboratory, Faculty of Exact Sciences, University of Bejaia, 06000, Bejaia, Algeria c Higher School of Sciences and Technologies of Computing and Digital, Amizour, 06000, Bejaia, Algeria,“ Electricity consumption forecasting based on ensemble deep learning with application to the Algerian market”,Published in 2021

Source: :<https://www.sciencedirect.com/science/article/pii/S0360544221033090>

1. SHILONG FAN School of Educational Science, Jilin Normal University, Changchun 10203, China,“ Research on Deep Learning Energy Consumption Prediction Based on Generating Confrontation Network”,Published in October 23,2019

Source: <https://ieeexplore.ieee.org/document/8880551>

1. Susanna Berkouwer,“ Electric Heating and the Effects of Temperature on Household Electricity Consumption in South Africa”,Published on February 5, 2019

Source: <https://haas.berkeley.edu/wp-content/uploads/WP299.pdf>

1. Yi Wang, Student Member, IEEE, Qixin Chen, Senior Member, IEEE, Chongqing Kang, Senior Member, IEEE, and Qing Xia, Senior Member, IEEE ,” Clustering of Electricity Consumption Behavior Dynamics toward Big Data Applications” Published on 06 April 2016

Source: <https://ieeexplore.ieee.org/document/7448460>

1. G. Vijendar Reddy, Lakshmi Jaswitha Aitha, Ch Poojitha, A Naga Shreya, D Krithika Reddy,G Sai.Meghana Department of Information Technology, GRIET, India” ELECTRICITY CONSUMPTION PREDICTION USING MACHINE LEARNING” .Published by E3S Web of Conferences 391,01048 (2023) Source:<https://www.e3s-conferences.org/articles/e3sconf/pdf/2023/28/e3sconf_icmed-icmpc2023_01048.pdf>
2. P. Balachandra, Indian Institute of Science,“ Indian Electricity System Transition and Its Challenges: A Review”, Published on January 2022Source: <https://www.researchgate.net/publication/357891974_Indian_Electricity_System_Transition_and_Its_Challenges_A_Review>