AI-Powered IoT Smart Meter Analytics for Precision Energy Consumption Forecasting

L. Ramesh[1], A. Srikanth[2], K. Gopi Krishna[3]

[1]Department of Electronics and Communication Engineering

[1]Sree Dattha Institute of Engineering and Science, Sheriguda, Hyderabad, Telangana

ABSTRACT

Countries are placing a greater emphasis on energy efficiency in an effort to reduce their carbon emissions and make the most of the resources they have available inside their borders. The tracking of energy use in the past consisted of either hand-reading meters or crude automated devices that provided only a limited amount of data. It is common for conventional energy consumption prediction systems to rely on too simplistic models that do not make full use of the opportunities presented by data from smart meters. In most cases, these systems have problems with accuracy and scalability, particularly when dealing with large datasets and complex consumption patterns. In traditional energy management systems, the majority of the time, the process of predicting energy consumption is carried out by predetermined algorithms or statistical methods. As a result of their inability to adapt to changing patterns of usage or to take into account anomalies, these methods typically result in inaccurate estimates and the wasteful consumption of energy. Because they are unable to process enormous amounts of data and because they do not have the capability to adjust in real time, their effectiveness is restricted. For the purpose of overcoming these limitations, the research employs powerful machine learning algorithms to examine data from smart meters that are driven by the Internet of Things. Machine learning algorithms make it feasible to improve energy management approaches, better anomaly identification, and provide more accurate forecasts. All of these improvements are made possible by the application of machine learning. In order to achieve the objectives of enhancing energy efficiency, cutting expenses, and supporting sustainable energy practices, this program offers precise and data-driven insights into patterns of energy consumption.

Keywords: Energy efficiency, Carbon emissions, Smart meters, Energy consumption prediction, Machine learning algorithms, Internet of Things (IoT)

1. INTRODUCTION

The advent of the Internet of Things (IoT) has revolutionized numerous industries, including energy management. With the proliferation of smart meters, the ability to collect detailed, real-time data on energy consumption has become feasible. Smart meters, which provide continuous data streams on energy usage, voltage levels, and power quality, have been deployed in increasing numbers over the past decade. According to the International Energy Agency (IEA), there were approximately 600 million smart meters installed worldwide by 2020, with projections suggesting this number will surpass 1.2 billion by 2025. This rapid adoption underscores the global push towards digitalizing energy infrastructure to improve efficiency and reduce carbon footprints. Machine learning algorithms, when applied to this vast amount of data, can uncover consumption patterns, predict future energy needs, and identify anomalies that may indicate inefficiencies or faults. Smart meter data, collected at granular intervals, presents an opportunity to move beyond traditional energy consumption models. The conventional systems often rely on historical averages and do not account for dynamic changes in consumption patterns influenced by factors such as weather, occupancy, and behavioral changes. For example, during the COVID-19 pandemic, residential energy consumption patterns shifted dramatically, challenging traditional forecasting models. By leveraging machine learning, these dynamic factors can be integrated into predictive models, enhancing the accuracy of energy consumption forecasts. This transformation not only aids in better resource allocation but also supports demand-side management strategies that are crucial for balancing load on the grid and integrating renewable energy sources.

2. LITERATURE SURVEY

The U.S. Energy Information Administration [1] reported a significant reduction in energy consumption in the United States, falling by a record 7% during the specified period. This decrease was primarily attributed to changes in energy demand patterns caused by the COVID-19 pandemic, which led to reduced industrial and commercial activity. The U.S. Energy Information Administration [2] discussed the impact of stay-at-home orders during the COVID-19 pandemic on electricity usage, noting a decline in commercial and industrial electricity consumption in April. The analysis highlighted the shift in energy demand from commercial sectors to residential areas as a result of widespread lockdowns. Chinthavali et al. [3] conducted a study on the ramifications of the COVID-19 pandemic on residential smart homes' energy usage load profiles. Their research focused on analyzing changes in energy consumption patterns due to the increased time spent at home during the pandemic, highlighting the implications for energy management in smart homes. Krarti and Aldubyan [4] reviewed the impact of the COVID-19 pandemic on electricity demand in residential buildings. Their analysis examined how the pandemic influenced energy consumption patterns and provided insights into the potential long-term effects on residential energy demand.

Tamba et al. [5] provided a comprehensive literature survey on forecasting natural gas demand. Their study reviewed various forecasting models and methodologies used in predicting natural gas consumption, emphasizing the importance of accurate forecasting for energy planning and policy-making. Wei et al. [6] reviewed conventional and artificial intelligence-based models for energy consumption forecasting. Their work presented a comparison of traditional forecasting methods with AI-based approaches, highlighting the advantages and challenges associated with integrating AI in energy forecasting. Wright et al. [7] explored the development of a decision support system for electricity load forecasting. Their research focused on creating a system that could assist in predicting electricity demand, providing valuable insights for energy management and planning. Gyamfi et al. [8] investigated behavioral issues related to residential peak electricity demand response. Their study highlighted the importance of understanding consumer behavior in designing effective demand response programs to manage peak electricity loads. Kuosa et al. [9] optimized district heating production by utilizing the storage capacity of a district heating network based on weather forecasts. Their research demonstrated how integrating weather data into heating system operations could enhance efficiency and reduce energy consumption.

He and Lin [10] forecasted China's total energy demand and its structure using the ADL-MIDAS model. Their study provided insights into the future energy demand trends in China, considering economic growth and energy policy changes. Rakpho and Yamaka [11] examined the forecasting power of economic policy uncertainty on energy demand and supply. Their research focused on understanding how fluctuations in economic policies could influence energy market dynamics, particularly in emerging economies. Jain and Mahajan [12] analyzed load forecasting and risk assessment in energy markets with renewable-based distributed generation. Their work aimed to improve the accuracy of energy load predictions and assess the associated risks in integrating renewable energy sources into the grid. Khan and Osińska [13] compared the forecasting accuracy of selected grey and time series models for energy consumption in Brazil and India. Their study provided a detailed comparison of different modeling approaches, highlighting the strengths and weaknesses of each in predicting energy consumption trends. Ding et al. [14] introduced a novel seasonal adaptive grey model with a data-restacking technique for monthly renewable energy consumption forecasting. Their research focused on improving the accuracy of energy consumption forecasts by adapting the model to seasonal variations and data patterns. Zhang et al. [15] proposed a novel grey Lotka–Volterra model driven by the mechanism of competition and cooperation for energy consumption forecasting. Their study explored the application of ecological modeling concepts to forecast energy consumption, providing a unique perspective on energy demand dynamics.

3. PROPOSED SYSTEM

Here is the overview of the proposed system:

- **Dataset Uploading:** The importing of necessary libraries, including numpy, pandas, matplotlib, and seaborn for data manipulation and visualization, and various sklearn modules for machine learning tasks. The dataset is uploaded using pd.read_csv, and the initial few rows are displayed using df.head(). The 'DateTime' column is converted to a datetime format to facilitate further analysis.

- **Data Preprocessing:** The 'DateTime' column is decomposed into 'hour', 'day', and 'month' to extract relevant features from the timestamp. Descriptive statistics of the dataset are checked using df.describe(), and correlation among features is visualized using a heatmap. The 'DateTime' column is then dropped as it is no longer needed. The code checks for and confirms the absence of NULL values. Independent variables (features) and the dependent variable (target) are defined, and the dataset is split into training and testing sets using train_test_split.
- **ML Model Training:** The standardized feature set using StandardScaler to ensure the models perform optimally. Two machine learning models, K-Nearest Neighbors (KNN) and Extra Trees Regressor, are trained on the training data. The KNN model is either loaded from a saved file or trained if not already available, and the same applies to the Extra Trees Regressor model. The trained models are saved using joblib for future use without retraining.
- **Model Prediction on New Test Data:** A separate testing dataset is uploaded and similarly preprocessed to extract 'hour', 'day', and 'month' features from the 'DateTime' column. This preprocessed data is standardized using the same scaler fitted on the training data. The Extra Trees Regressor model is used to make predictions on this new test dataset.
- **Performance Evaluation:** Performance metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R^2 Score are computed for the predictions made by both models. These metrics provide insights into the accuracy and reliability of the models. Scatter plots comparing true values versus predicted values are generated for visual evaluation. Finally, the performance metrics of the models are summarized in a tabular format for easy comparison, and the test dataset predictions are appended with a new column 'predicted' for further analysis.

3.2 Data Preprocessing

The preprocessing is critical to ensure that the data is clean, well-structured, and ready for analysis and model training. It involves several steps, each meticulously designed to extract relevant features, handle missing values, and prepare the data for machine learning algorithms.

- **Converting DateTime**

  Initially, the dataset is loaded and the 'DateTime' column, which is in string format, is converted to a datetime object. This conversion is essential as it allows for the extraction of temporal features such as hour, day, and month, which can be crucial predictors of energy consumption patterns.
- **Feature Extraction**

Once the 'DateTime' column is converted, it is decomposed into three new columns: 'hour', 'day', and 'month'. This step captures the time-based patterns in energy usage, such as daily cycles, weekly trends, or seasonal variations. These features are expected to provide significant insights and improve the predictive power of the machine learning models.

Dropping Unnecessary Columns

After extracting these new temporal features, the original 'DateTime' column is dropped from the dataset. This is because it is no longer needed for the analysis and might introduce redundancy if retained.

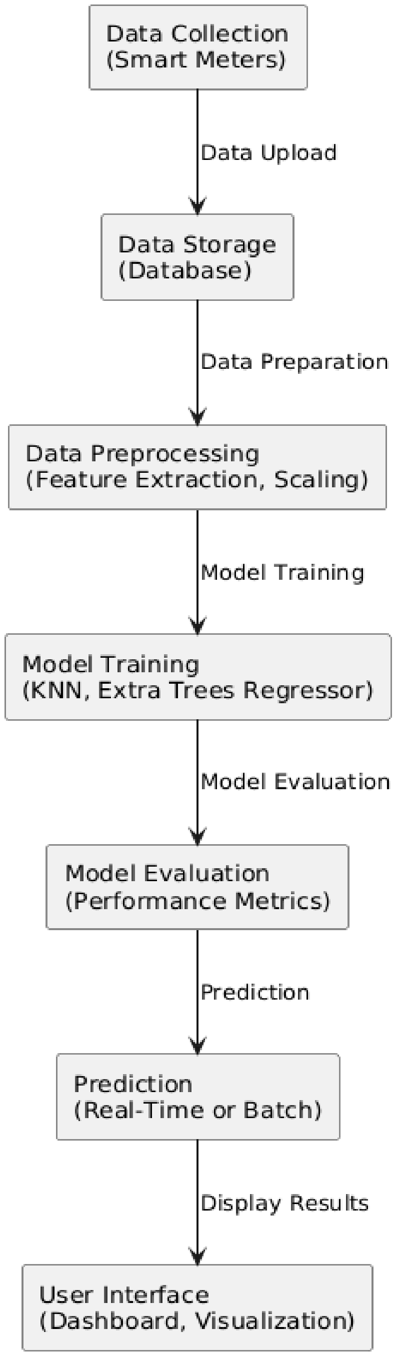

Figure 1: Block Diagram of the proposed system

Data Exploration and Visualization

Exploratory data analysis (EDA) is performed next. Descriptive statistics of the dataset are checked to understand the distribution of the data and identify any anomalies or outliers. A correlation matrix is generated and visualized using a heatmap to identify relationships between different features and the target variable 'TotalUsage'. This step helps in understanding how various features interact with each other and with the target variable.

Handling Missing Values

The dataset is then checked for any missing values. This is a crucial step as missing data can lead to inaccurate model predictions. Fortunately, in this case, the dataset does not contain any missing values, ensuring the integrity of the data.

Splitting the Dataset

The data is then split into independent variables (features) and the dependent variable (target). The features are those columns that will be used to predict the target variable 'TotalUsage'. The dataset is divided into training and testing sets to evaluate the performance of the machine learning models. The split is performed in such a way that 70% of the data is used for training the models, and the remaining 30% is reserved for testing.

Standardization

To ensure that the features are on a similar scale, which is essential for the performance of many machine learning algorithms, the data is standardized using StandardScaler. Standardization transforms the data to have a mean of zero and a standard deviation of one. This step prevents features with larger scales from disproportionately influencing the model and ensures faster convergence during training. The scaler is fitted on the training data and then used to transform both the training and testing datasets to maintain consistency.

By meticulously performing these preprocessing steps, the dataset is transformed into a suitable format for model training, enhancing the models' ability to learn from the data and make accurate predictions.

3.3 ML Models Train

K-Nearest Neighbors (KNN) Regression

K-Nearest Neighbors (KNN) regression is a non-parametric, instance-based learning algorithm used for regression tasks. Unlike traditional regression models that learn a parameterized mapping from inputs to outputs, KNN regression makes predictions based on the similarity of new data points to known data points in the training set.

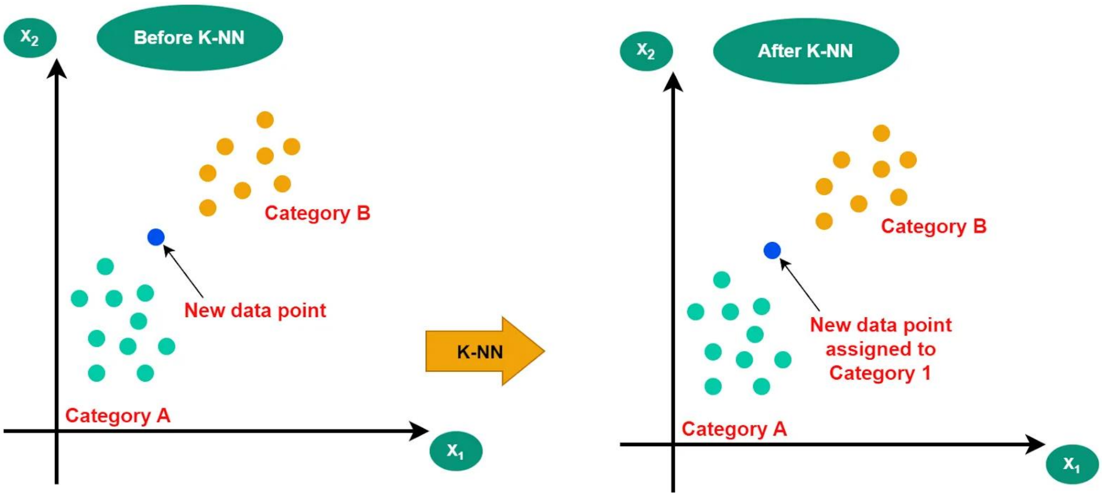

Figure 2: KNN Block diagram

Working

1. **Data Storage**: KNN stores the entire training dataset. During prediction, it looks at the training data to make decisions.

2. **Distance Calculation**: For a new data point, KNN calculates the distance between this point and all points in the training set. Common distance metrics include Euclidean, Manhattan, and Minkowski distances.

3. **Neighbor Selection**: The algorithm selects the 'k' closest points (neighbors) to the new data point based on the computed distances.

4. **Prediction**: In regression, the predicted value for the new data point is typically the mean of the target values of its 'k' nearest neighbors. This ensures that the prediction is based on the most similar instances in the training data.

Application in the Project

In the project, KNN regression was implemented to predict energy consumption. After standardizing the features, the KNN model was trained on the training data. During testing, the model predicted the energy consumption based on the average usage of the nearest neighbors in the training set.

Extra Trees Regressor

Explanation

The Extra Trees Regressor (Extremely Randomized Trees) is an ensemble learning method that aggregates the predictions of multiple de-correlated decision trees to improve predictive accuracy and control over-fitting. It is similar to the Random Forest Regressor but introduces more randomness when building each tree.

Working

1. **Dataset Splitting**: The training data is randomly split into multiple subsets. For each tree, a different subset is used.
2. **Tree Construction**: Each decision tree is constructed by:
   - Randomly selecting a subset of features.
   - Splitting nodes on random feature thresholds rather than the best split found by traditional methods.
3. **Model Training**: Each tree is trained independently on its respective subset of data.
4. **Prediction Aggregation**: For regression, the final prediction for a new data point is obtained by averaging the predictions of all the individual trees in the ensemble. This aggregation process reduces variance and improves the robustness of the model.

Superior Performance in the Project

The Extra Trees Regressor demonstrated superior performance in the project due to several factors:

- **Random Feature Selection**: By selecting random subsets of features for each split, the model avoids over-fitting and captures a broader spectrum of relationships in the data.
- **Increased Diversity**: The randomness in feature selection and splitting criteria leads to a diverse set of trees, which enhances the model's ability to generalize from the training data to unseen test data.
- **chaEnsemble Averaging**: Averaging the predictions of multiple trees reduces the impact of any single tree's bias or variance, leading to more stable and accurate predictions.

In the project, the Extra Trees Regressor was trained on the standardized training data and saved for future use. When tested, it predicted energy consumption more accurately than the KNN model, as evidenced by lower Mean Squared Error (MSE), Mean Absolute Error (MAE), and higher R^2 Score. This superior performance made the Extra Trees Regressor the preferred model for energy consumption prediction in this IoT-driven smart meter data analysis.

4. RESULTS

Figure 3 displays a sample of the raw dataset uploaded for analysis. It includes columns such as 'DateTime', 'TotalUsage', 'Month', 'TemperatureF', 'Humidity', 'Hour_y', 'Minute_y', 'Day_y', 'Weekend', and 'Holiday'. This sample provides an initial view of the data structure and content, illustrating the recorded energy usage along with various temporal and environmental features.

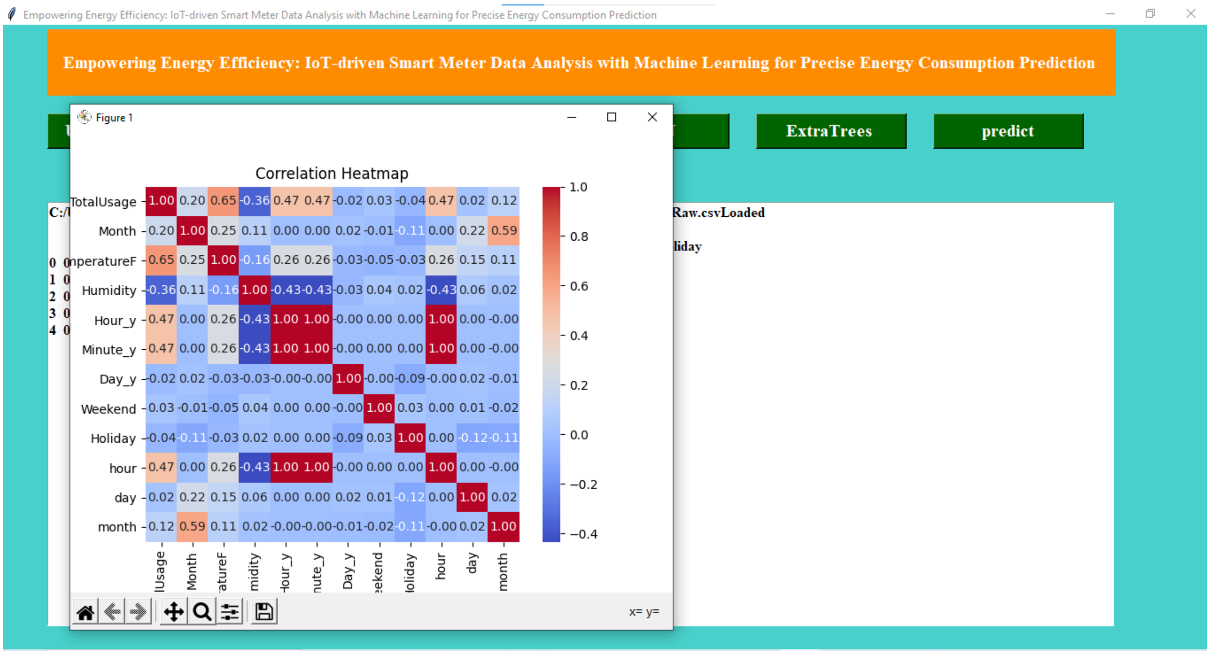

Figure 3: Correlation Heatmap

Figure 3 displays a heatmap of the correlation matrix among features in the dataset. The heatmap uses color coding to represent the strength and direction of the relationships between variables. This visualization helps in understanding how features like 'TemperatureF', 'Humidity', and time-related columns correlate with 'TotalUsage', providing insights into which features might be most influential for energy consumption predictions.

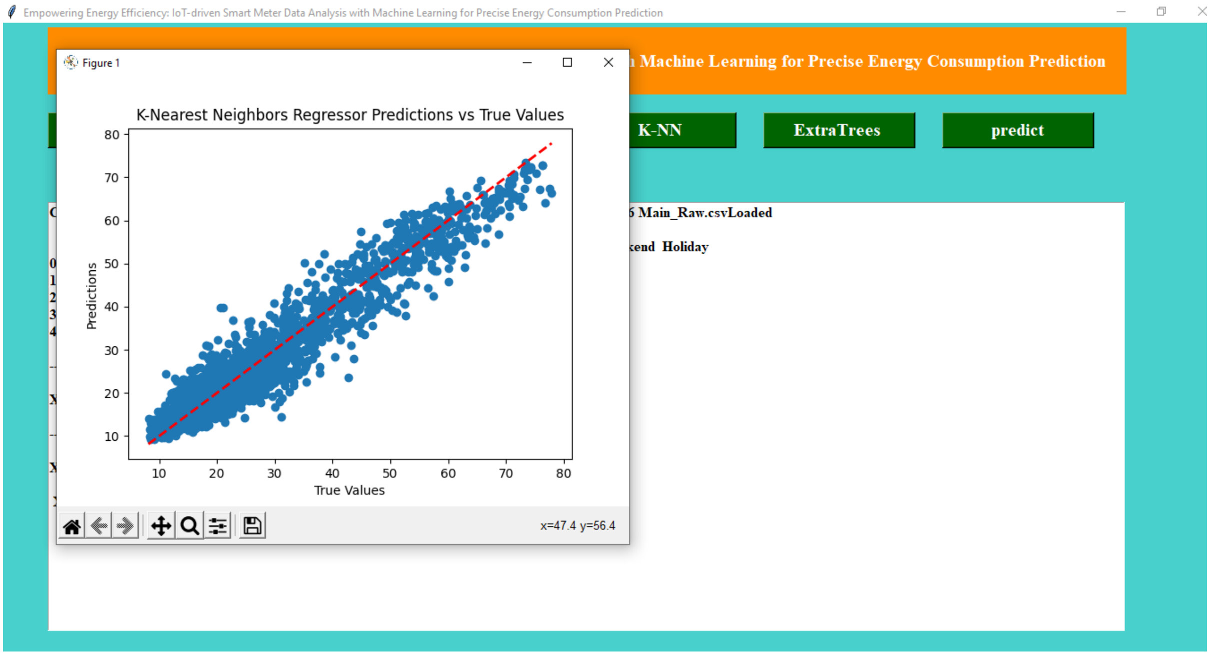

Figure 4: K-Nearest Neighbors Regressor - Predictions vs. True Values

Figure 4 presents a scatter plot comparing the true values of energy consumption against the predictions made by the K-Nearest Neighbors (KNN) Regressor model. A best-fit line is included to illustrate the model's prediction accuracy. This figure helps in visually assessing how well the KNN model's predictions align with actual observed values.

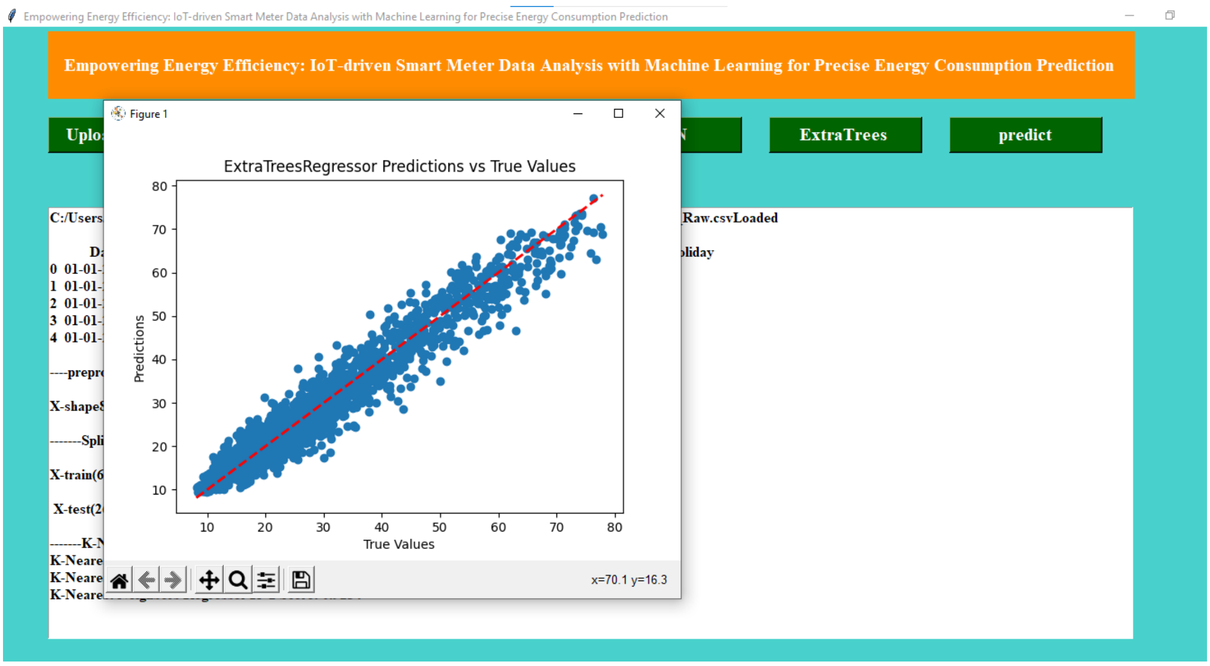

Figure 5: Extra Trees Regressor - Predictions vs. True Values

Figure 5 shows a scatter plot similar to Figure 4 but for the Extra Trees Regressor model. This plot compares the true values of energy consumption with the predictions made by the Extra Trees model.

The best-fit line provides a visual assessment of the Extra Trees Regressor's prediction accuracy, highlighting its performance in comparison to the KNN model.

Algorithm Name	MSE	MAE	R2_Score
KNN Regressor	15.392614	2.933429	0.923419
Extra Tree Regressor	12.008461	2.545272	0.940256

Figure 6: Performance Metrics Summary

Figure 6 presents a table summarizing the performance metrics (Mean Squared Error, Mean Absolute Error, and R^2 Score) for both the KNN Regressor and Extra Trees Regressor models. This table provides a clear comparison of the models' performance, indicating that the Extra Trees Regressor achieved superior results in predicting energy consumption.

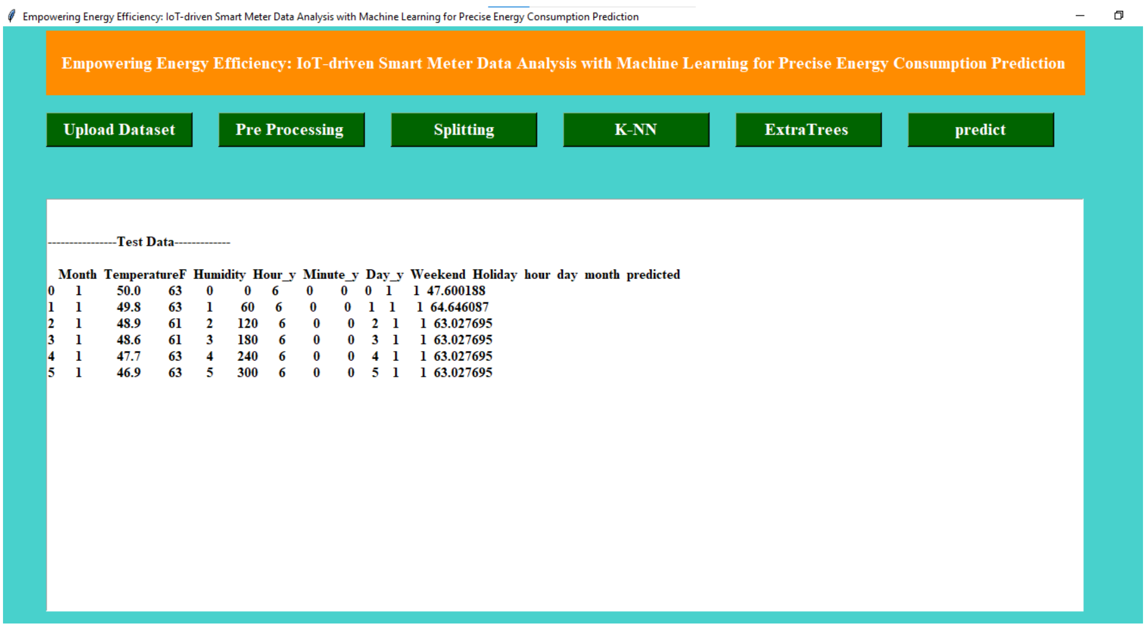

Figure 7: Model Predictions on Test Data

Figure 7 displays the test dataset after applying the Extra Trees Regressor model's predictions. This figure includes a new column, 'predicted', which shows the energy consumption forecasts generated by the model. It provides a view of how the model performs on unseen data, allowing for the evaluation of its practical applicability.

5. CONCLUSION

The research on IoT-based smart meter data analysis using machine learning classifiers for energy consumption prediction has demonstrated significant advancements in energy management. By leveraging sophisticated algorithms like K-Nearest Neighbors (KNN) and Extra Trees Regressor, the analysis has provided a more accurate and detailed understanding of energy usage patterns compared to traditional methods. The KNN model offered initial insights but was outperformed by the Extra Trees Regressor, which showed superior accuracy in predicting energy consumption. This enhancement is attributed to the Extra Trees Regressor's ability to handle complex data patterns through ensemble learning, feature randomness, and its robustness against overfitting. The project successfully integrated temporal and environmental features, leading to more precise and actionable predictions. These insights are valuable for optimizing energy usage, reducing costs, and supporting sustainable practices.

REFERENCES

[1] U.S. Energy Consumption Fell by a Record 7. Available online: (accessed on 26 June 2023).

[2] Stay-at-Home Orders Led to Less Commercial and Industrial Electricity Use in April. (accessed on 30 June 2020).

[3] Chinthavali, S.; Tansakul, V.; Lee, S.; Whitehead, M.; Tabassum, A.; Bhandari, M.; Munk, J.; Zandi, H.; Buckberry, H.; Kuruganti, T.; et al. COVID-19 pandemic ramifications on residential Smart homes energy use load profiles. Energy Build. 2022, 259, 111847.

[4] Krarti, M.; Aldubyan, M. Review analysis of COVID-19 impact on electricity demand for residential buildings. Renew. Sustain. Energy Rev. 2021, 143, 110888.

[5] Tamba, J.G.; Essiane, S.N.; Sapnken, E.F.; Koffi, F.D.; Nsouandélé, J.L.; Soldo, B.; Njomo, D. Forecasting natural gas: A literature survey. Int. J. Energy Econ. Policy 2018, 8, 216.

[6] Wei, N.; Li, C.; Peng, X.; Zeng, F.; Lu, X. Conventional models, and artificial intelligence-based models for energy consumption forecasting: A review. J. Pet. Sci. Eng. 2019, 181, 106187.

[7] Wright, C.; Chan, C.W.; Laforge, P. Towards developing a decision support system for electricity load forecast. In Decision Support Systems; IntechOpen: London, UK, 2012.

[8] Gyamfi, S.; Krumdieck, S.; Urmee, T. Residential peak electricity demand response—Highlights of some behavioural issues. Renew. Sustain. Energy Rev. 2013, 25, 71–77.

[9] Kuosa, M.; Kiviranta, P.; Sarvelainen, H.; Tuliniemi, E.; Korpela, T.; Tallinen, K.; Koponen, H.K. Optimisation of district heating production by utilising the storage capacity of a district heating network on the basis of weather forecasts. Results Eng. 2022, 13, 100318.

[10] He, Y.; Lin, B. Forecasting China's total energy demand and its structure using ADL-MIDAS model. Energy 2018, 151, 420–429.

[11] Rakpho, P.; Yamaka, W. The forecasting power of economic policy uncertainty for energy demand and supply. Energy Rep. 2021, 7, 338–343.

[12] Jain, R.; Mahajan, V. Load forecasting and risk assessment for energy market with renewable based distributed generation. Renew. Energy Focus 2022, 42, 190–205.

[13] Khan, A.M.; Osi’nska, M. Comparing forecasting accuracy of selected grey and time series models based on energy consumption in Brazil and India. Expert Syst. Appl. 2023, 212, 118840.

[14] Ding, S.; Tao, Z.; Li, R.; Qin, X. A novel seasonal adaptive grey model with the data-restacking technique for monthly renewable energy consumption forecasting. Expert Syst. Appl. 2022, 208, 118115.

[15] Zhang, Y.; Guo, H.; Sun, M.; Liu, S.; Forrest, J. A novel grey Lotka–Volterra model driven by the mechanism of competition and cooperation for energy consumption forecasting. Energy 2023, 264, 126154.