PROJECT REPORT

Energy Consumption Forecasting for Smart Grid Optimization

Author: Krish Choudhary

A comprehensive machine learning project that analyzes and forecasts household energy consumption to support smart grid optimisation. This repository contains the full analysis, from data preprocessing to model evaluation, including time-series forecasting, customer segmentation, and efficiency classification.

Under the supervision of:

Prof. Sachin Kansal

Thapar Institute of Engineering & Technology

(Deemed to be University) Patiala, India

December 2023

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

ABSTRACT

Energy management at the household level lies at the core of the future smart grids since the focus is integrating energy use by balancing supply and demand. The current project studies household energy consumption patterns through using a minute-level dataset of energy consumption, energy generation, and weather data. The procedure includes steps such as treating missing observations, eliminating redundant features with high correlation, and performing down-sampling to daily time-chunks that minimized the chance of losing useful data while enhancing the quality for further analysis.

Models developed to forecast trends in energy consumption demands included time-series forecasting models. The models of ARIMA and SARIMAX were employed where the SARIMAX model integrated such weather variables as temperature, humidity and wind speed to the model so as to factor in external effects. While SARIMAX did perform a little worse in terms of prediction errors, SARIMAX performed better in terms of interpretation as the model parameters incorporated features such as temperature and humidity arguing that the model is effective in triggering complex patterns of energy use.

Also, K-Means Algorithm demonstrated whittling down the daily energy consumption into segments of user behaviour allowing new understanding of consumer habits and new directions of improvement. Random Forest classification was used in determining energy efficiency by discriminating energy consumption among the efficient and the inefficient classifiers through point identification.

This analysis notes the importance of the ability to integrate statistical and machine learning techniques into energy analysis. Incorporating additional variables such as weather into the forecasting models, and segmentation of consumer behavior, this work provides important opportunities for managing consumption, optimizing residential energy use, and developing efficient energy systems.

INTRODUCTION

As the energy requirements increase and the need for sustainable practices rises, the management of energy in residential areas has been immensely targeted in the creation of the smart grid technologies. These systems help collect data in real time and allow for targeted interventions such as optimization of energy use, minimization of wastage, and incorporation of renewables. This project utilizes such a dataset to study and forecast energy usage, scaling sustainable energy management to a whole new level.

The dataset has a collection of minute wise data about electricity consumption and generation along with climatic data. Some of the important variables are the consumption of specific appliances, the total amount of electricity consumed by a household and weather elements including, temperature, humidity, and wind velocity. Because of the intricacies of the data, the pre-processing involved imputing missing values, mitigating correlation above 0.9, and aggregating to daily values for trend purposes. Some feature engineering was also done whereby consumption of appliances was categorized into further subcategories for detailed analysis.

Time-series forecasting methods were the key component of this work. Basic seasonal cycles were detected through ARIMA models and this was further enhanced through SARIMAX regression where weather data was regressed on the latter model. The SARIMAX model delivered the best performance, which revealed the weather conditions dependency of the residential energy expenditures. Performance measurements such as RMSE and MAPE validated short-term energy demand forecasting performance.

K-Means clustering analysis was sequenced to divide the data sets as per the behavioral consumption patterns of the respondents. These findings therefore create opportunities to create specific energy saving approaches for different categories of consumers. Also, Random Forest classification approach was used to determine the energy efficiency, and the data was classified into levels of energy excessive use and others to determine the energy metrics effectively.

This study combines statistical and machine learning approaches in order to understanding and optimize energy usage in residential premises. By incorporating external environmental aspects, finding behavioral patterns, and assessing energy efficiency, this research offers useful information that contributes to the overall objectives of minimisation of energy losses, advances in sustainability, and development of smart grid systems.

UNIQUE SELLING POINT

The unique selling point of this project is its comprehensive evaluation of residential energy consumption with the use of statistical time series models, machine learning methods and weather as an exogenous variable to enhance targets' predictive power and applicability. This is not the case with energy analytics which are limited to consumption data. This project shows that the inclusion of exogenous variables from the real world like the weather enhances the quality of the forecast and segmentation based on behavior patterns.

The SARIMAX model that included the exogenous variable was clearly better than standard ARIMA models in terms of performance and also has the capability to allow for energy consumption prediction under different conditions. Also, thanks to the implementation of K-Means clustering, a fresh view on consumers was obtained where certain embedded energy saving behaviors were quite distinctive for certain groups of people.

This project integrates strong preprocessing methods with sophisticated modeling and interpretable results to address the divide between the academic literature and deployment within smart grid optimization. Such detailed approach is a horizontal strategy for energy providers wishing to enhance demand-side management and encouraging energy efficiency among consumers.

OBJECTIVES

- Develop ML-based framework for energy consumption forecasting.

- Classify energy usage into categories for better grid management.

- Provide actionable insights for optimizing energy distribution.

- Data Preprocessing and Feature Engineering: Handle missing data, remove outliers, adjust data to relevant time intervals, and place calculations of significant variables, thus, producing quality data for modelling.

- Time-Series Forecasting: Construct models like ARIMA and SARIMAX that will assist in forecasting residential energy consumption in the short term. Incorporate Exogenous Variables: Incorporate other factors like weather elements such as temperature, humidity and cloud cover in predictive models for energy forecasting so as to enhance accuracy.

- Energy Consumption Clustering: Use K-Means methods to cluster consumers based on their energy usage profiles so as to make practical recommendations regarding the energy users' behavior.

- Energy Efficiency Classification: Construct classification models such as Random Forest to single out families who inefficiently consume a lot of energy for purposes of intervention.

- Model Evaluation and Comparison: Analyze models by the use of metrics such as RMSE, MAPE, MAE and accuracy so as to evaluate performance and find the best depending on the energy optimization required.

PROBLEM STATEMENT

With the increasing integration of renewal energy sources into the grid, there arises a need for more effective energy management systems. As smart grids become more popular, predicting energy demand, understanding its patterns, and addressing inefficiencies becomes crucial for consumers and the environment. Nevertheless, several issues, including the analysis of rather complex data sets, the analysis of external variables (such as weather data), and data interpretation present a number of analytical challenges. This project will respond to all these questions using time-series forecasting, clustering and classification methods and combine them in residential energy management.

METHODOLOGY

- **Existing System**

The conventional energy forecasting systems have typically relied on ARIMA type methods or regression models for predicting energy consumption. Such models frequently neglect to include external variables such as meteorological conditions or are able to cope with high frequency smart meter data. Other limitations of the existing system include:

Limited Scalability: Failure in processing the huge data sets created by smart grids.

Limited Features: Exogenous variables such as temperature, humidity and wind speeds are not incorporated.

Excessive Errors: These are very wide, especially in the case of peak load forecasting or in anomaly detection cases.

In view of these challenges the suggested system includes extended machine learning models and encodes additional features that should lead to better predictions and optimizations.

- **Limitations of Existing Systems**

Arguably ineffective for large datasets because of slow processing and inefficiencies.

Conventional models with many dimensions of features have got high training times.

More variables increase complexity of the model and hence difficulty in interpretation.

- **Proposed System**

In this proposed system, Machine Learning techniques are employed in forecasting, classification and clustering. The essential components encompassed include

ARIMAX for Time-Series forecasting: Applicable to seasonal trends and external variables such as temperature, humidity, and cloud coverage.

Random Forest for Classification: Classifies efficient and inefficient patterns or styles of energy usage.

K-Means Clustering: Classifies energy consumers in different categories according to their energy consumption tendencies..

Advantages of the Proposed System:

The high precision and strength of Random Forest and SARIMAX are combined effectively.

The scalability to manage high dimensional datasets is a major advantage.

Enhanced prediction power owing to incorporating additional features.

- **Proposed Method**

Modules and Descriptions

Importing the Packages

Data normalization, modeling, and evaluation in the current research has been done with the aid of libraries like pandas, numpy, scikit-learn, statsmodels and pmdarima.

Matplotlib and seaborn are used for visualization.

Data Preprocessing:

Normalization: Numerical features are rescaled into the range of
0 to 1 with MinMaxScaler.

Handling Missing Values: Values represented as NaNs are filled and as appropriate done, we have bfill for categorical and mean for numerical values .

Feature Engineering:

Features such as composite features were created such as Furnace, Kitchen etc by merging some columns that were closely related.

Irrelevant or highly correlated features e.g. House overall [kW] removed.

One hot encoding for categorical features – icon, summary, so that they are machine readable.

Data Splitting

The independent variables X and dependent variable y has been defined.

Using train_test_split, data is split into 65% for training and 35% for testing.

Training the Models

The models SARIMAX, Random Forest, and Isolation Forest have been trained on the processed data.

Hyperparameters have also been optimized.

Classification Models

SARIMAX (Seasonal ARIMA with Exogenous Variables)

Concept: This is to combine other variables such as temperature, humidity etc. to enhance time series forecasting.

Working:

Patterns and trends that are seasonal are identified in data.

They create regression models to add on external variables to enhance prediction accuracy.

Advantages:

Provides accurate energy forecasts for 10-day periods.

Prediction accuracy is high even during seasons due to incorporation of weather conditions.

Random Forest Classifier

Concept: Collection of trees which are all decision trees trained but on different subspaces of the overall dataset.

Working:

Randomly divides the data into training sets and features.

Applies multiple decision trees on these sets.

Makes an average of all models predictions or other pools such as the majority vote when simply classifying.

Advantages:

Very high rate of accuracy in classifying energy usage efficiency.

Also high robustness to noise and overfitting.

K-Means Clustering:

K-means algorithm which is a form of clustering uses feature vector distance to center based clusters within each region known as k-centers

Concept: A utility geospatial time series used for power consumption transcending the level of granularity of visualization to aid understanding and intervention.

Working:

Data consumption is adjusted for extreme outlier values and separated into three levels based on usage frequency and volume.

Steps: Definition of the cluster based on the degree of closeness of ‘points’ to the center.

Advantages:

Description of different “segments” of usage patterns.

Energy perspective dynamics models are helpful in grasping how patterns evolve over time.

Stages in the Proposed Workflow: Step-by-Step

Data Collection phase:

For our analysis data is obtained from smart meters which are recording the user's frequency in kilowatt-hours.

Information and data factors such as cyclical meteorological changes and others also get incorporated.

Data Preprocessing:

Values which have not been recorded in this case are addressed.

Non-essential features are removed into more manageable segments.

Feature Engineering:

In this case simple pandas command is used to remove columns or variables that do not have a useful contribution.

Panda dataframe is also applied with the aim of simplifying the analysis of data collected.

Data Dimensionality Reduction and Loading Hosted Models

Data owned and collected is organized into smaller manageable segments as a feature vector.

Model Training:

Time units in our case constituted importance and they should be included in the analysis through advanced models ARIMA and SARIMAX.

Random Forest’s technique helps analyze if the input features can distinguish between possible energy users and nonusers.

Efficient strategies also include isolation forest which in this case isolates poor energy consumers from other categories of energy users.

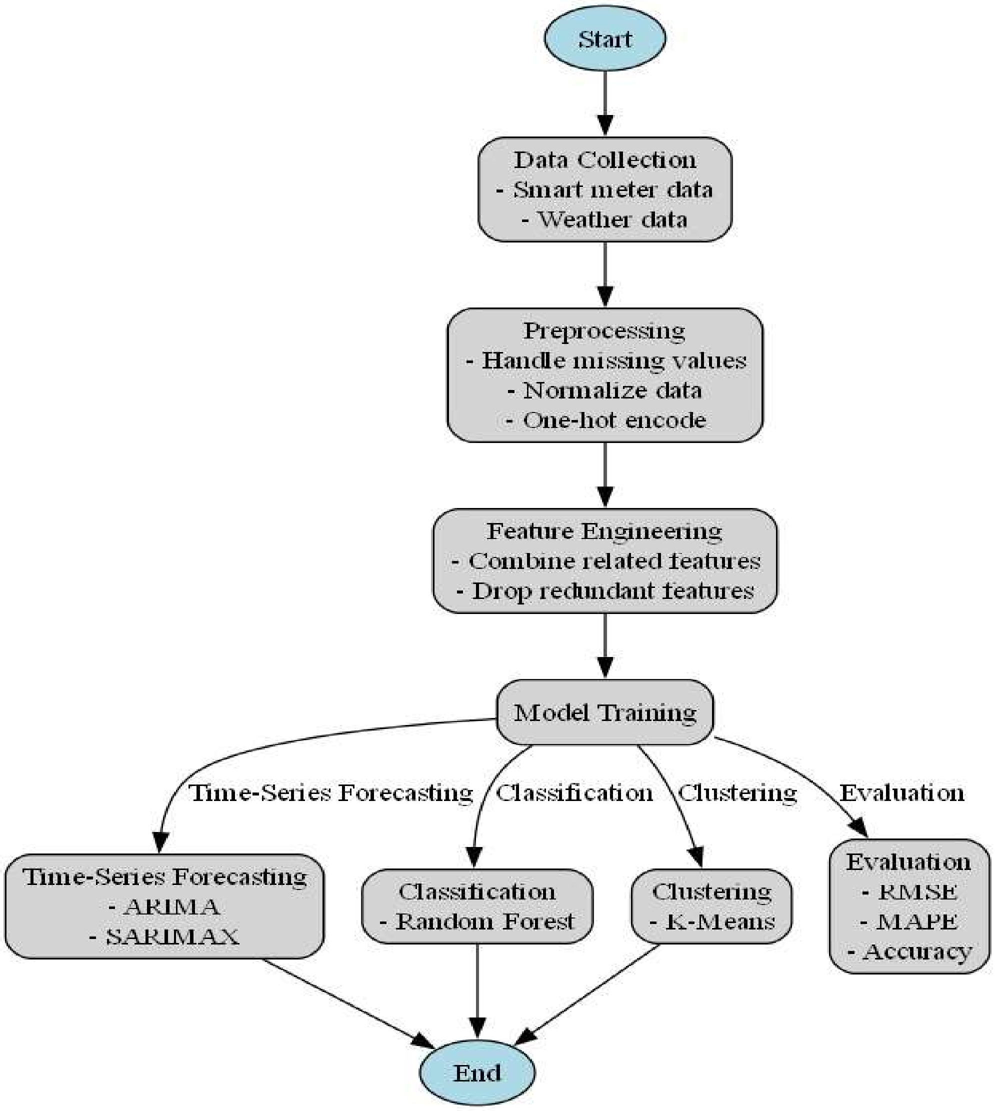
Evaluation:

Metrics like RMSE and MAPE are calculated for time series forecasting models, ARIMA and SARIMAX.

The Precision, Recall and F1 Score of Random Forest Classifier is calculated along with the accuracy and confusion matrix.

FLOW DIAGRAM

The following flow diagram illustrates the workflow of the project.



EXPERIMENTAL RESULTS

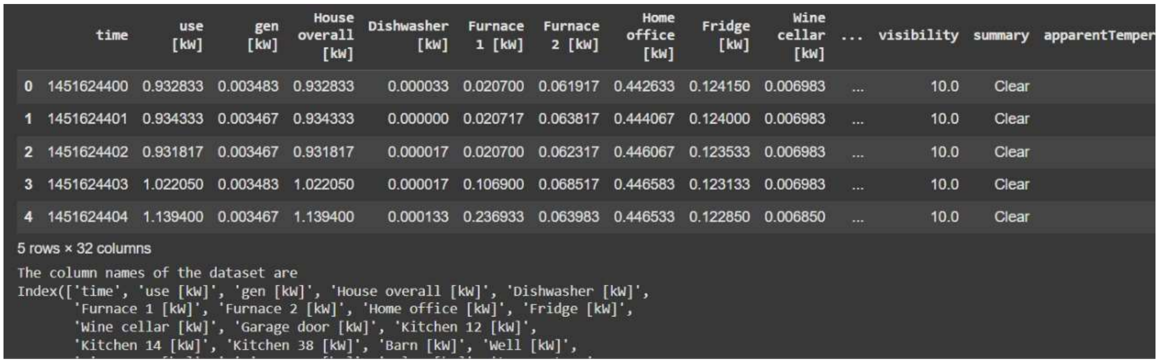
Dataset

The dataset consists of high-frequency energy consumption data from smart meters installed in residential households. Additional variables include weather data such as temperature, humidity, and wind speed, which impact energy consumption patterns.

Results

1) Displaying Dataset

Pandas is used to read dataset, input is given as csv file, and displaying the head of dataset and the datatype of its columns :



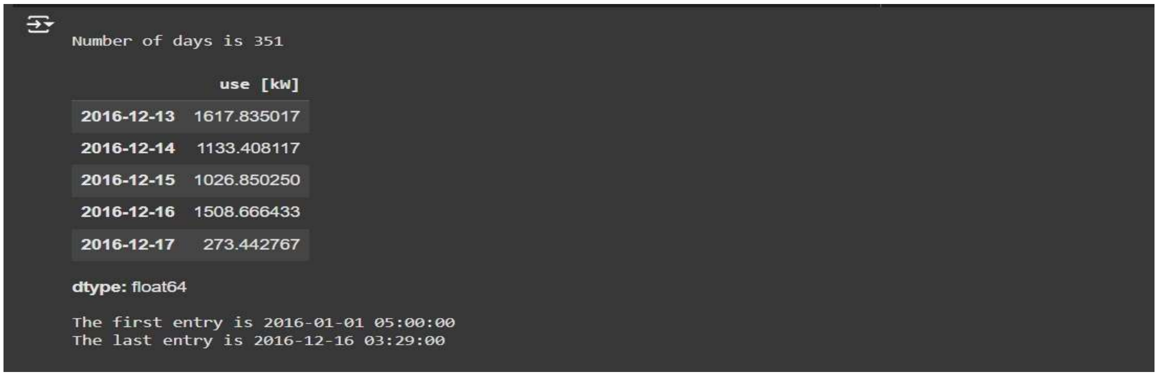
	time	use [kW]	gen [kW]	House overall [kW]	Dishwasher [kW]	Furnace 1 [kW]	Furnace 2 [kW]	Home office [kW]	Fridge [kW]	Wine cellar [kW]	...	visibility	summary	apparentTemper
0	1451624400	0.932833	0.003483	0.932833	0.000033	0.020700	0.061917	0.442633	0.124150	0.006983	...	10.0	Clear	
1	1451624401	0.934333	0.003467	0.934333	0.000000	0.020717	0.063817	0.444067	0.124000	0.006983	...	10.0	Clear	
2	1451624402	0.931817	0.003467	0.931817	0.000017	0.020700	0.062317	0.446067	0.123533	0.006983	...	10.0	Clear	
3	1451624403	1.022050	0.003483	1.022050	0.000017	0.106900	0.068517	0.446583	0.123133	0.006983	...	10.0	Clear	
4	1451624404	1.139400	0.003467	1.139400	0.000133	0.236933	0.063983	0.446533	0.122850	0.006850	...	10.0	Clear	

5 rows × 32 columns

```
The column names of the dataset are
Index(['time', 'use [kW]', 'gen [kW]', 'House overall [kW]', 'Dishwasher [kW]',
       'Furnace 1 [kW]', 'Furnace 2 [kW]', 'Home office [kW]', 'Fridge [kW]',
       'Wine cellar [kW]', 'Garage door [kW]', 'Kitchen 12 [kW]',
       'Kitchen 14 [kW]', 'Kitchen 38 [kW]', 'Barn [kW]', 'Well [kW]',
```

2) Resampling Data

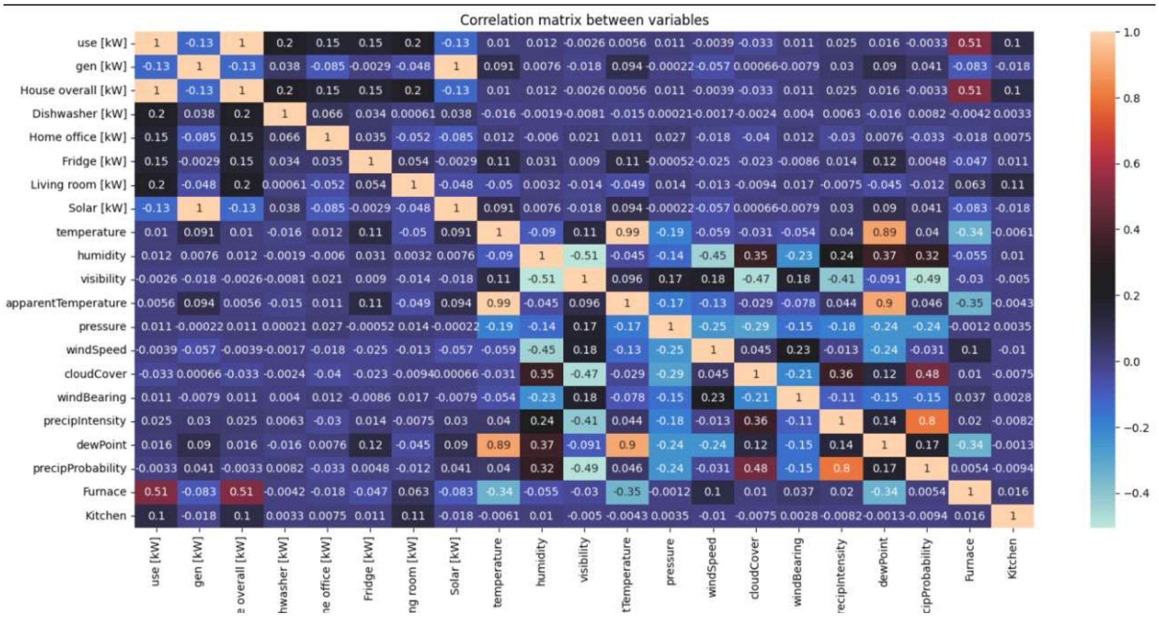
Generating a time based index with minute frequency and resampling data from minute to minute to daily and hence finding out the number of days for which the dataset contains information.



```
Number of days is 351

                use [kW]
2016-12-13   1617.835017
2016-12-14   1133.408117
2016-12-15   1026.850250
2016-12-16   1508.666433
2016-12-17    273.442767

dtype: float64

The first entry is 2016-01-01 05:00:00
The last entry is 2016-12-16 03:29:00
```

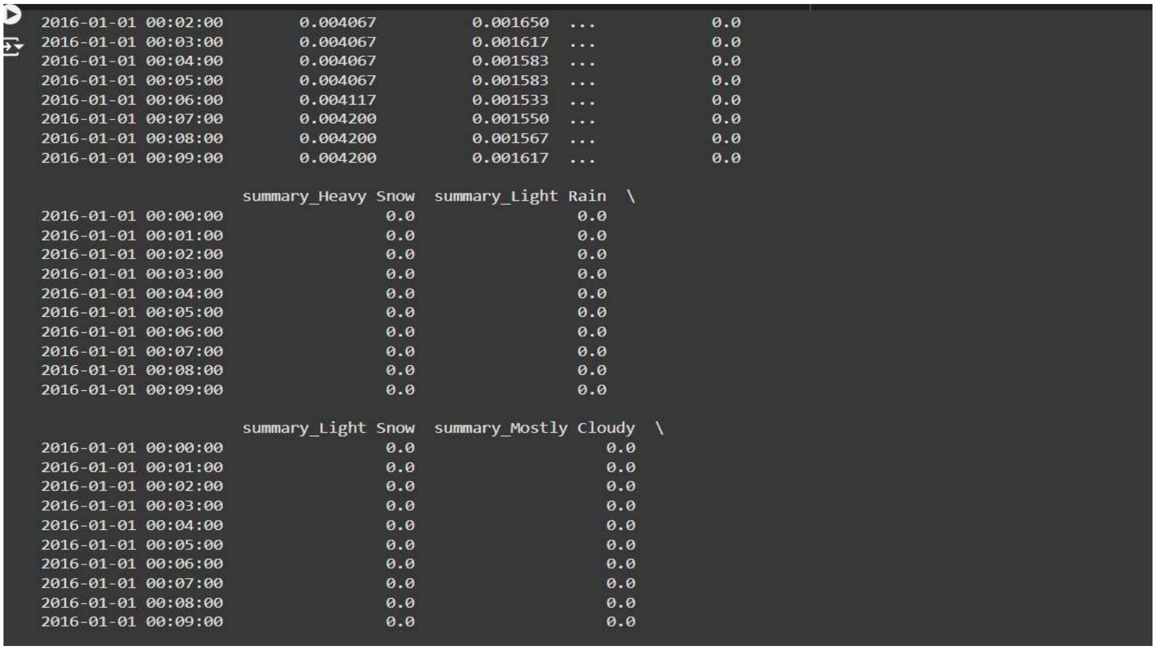
3) Finding correlated features

Plotting the correlation matrix and dropping the highly correlated features.



4) Preprocessing Data

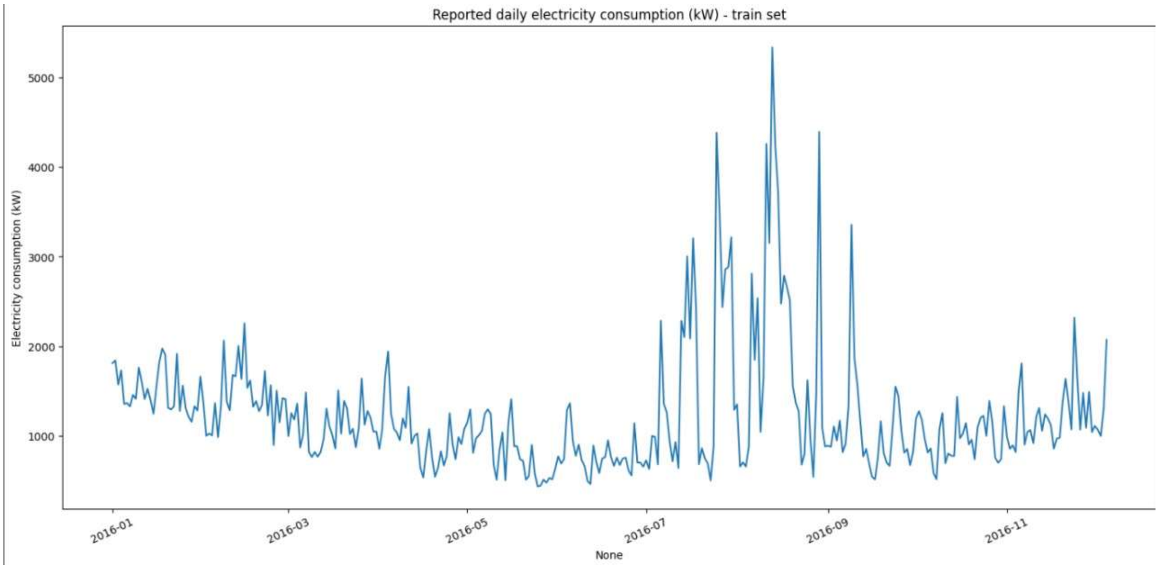
Label Encoder is used for preprocessing, it converts the text input into categorical input, that is icon, summary is given as text or string, the algorithm receives input as only numerical values.



```
2016-01-01 00:02:00        0.004067         0.001650  ...           0.0
2016-01-01 00:03:00        0.004067         0.001617  ...           0.0
2016-01-01 00:04:00        0.004067         0.001583  ...           0.0
2016-01-01 00:05:00        0.004067         0.001583  ...           0.0
2016-01-01 00:06:00        0.004117         0.001533  ...           0.0
2016-01-01 00:07:00        0.004200         0.001550  ...           0.0
2016-01-01 00:08:00        0.004200         0.001567  ...           0.0
2016-01-01 00:09:00        0.004200         0.001617  ...           0.0

                     summary_Heavy Snow  summary_Light Rain  \
2016-01-01 00:00:00                 0.0                 0.0
2016-01-01 00:01:00                 0.0                 0.0
2016-01-01 00:02:00                 0.0                 0.0
2016-01-01 00:03:00                 0.0                 0.0
2016-01-01 00:04:00                 0.0                 0.0
2016-01-01 00:05:00                 0.0                 0.0
2016-01-01 00:06:00                 0.0                 0.0
2016-01-01 00:07:00                 0.0                 0.0
2016-01-01 00:08:00                 0.0                 0.0
2016-01-01 00:09:00                 0.0                 0.0

                     summary_Light Snow  summary_Mostly Cloudy  \
2016-01-01 00:00:00                 0.0                    0.0
2016-01-01 00:01:00                 0.0                    0.0
2016-01-01 00:02:00                 0.0                    0.0
2016-01-01 00:03:00                 0.0                    0.0
2016-01-01 00:04:00                 0.0                    0.0
2016-01-01 00:05:00                 0.0                    0.0
2016-01-01 00:06:00                 0.0                    0.0
2016-01-01 00:07:00                 0.0                    0.0
2016-01-01 00:08:00                 0.0                    0.0
2016-01-01 00:09:00                 0.0                    0.0
```

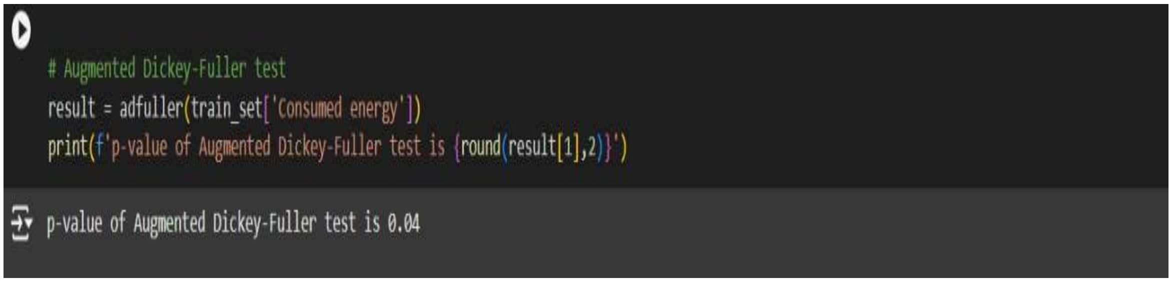
5) Plotting

Plot of reported daily electrical consumption.



6) ARIMA Model

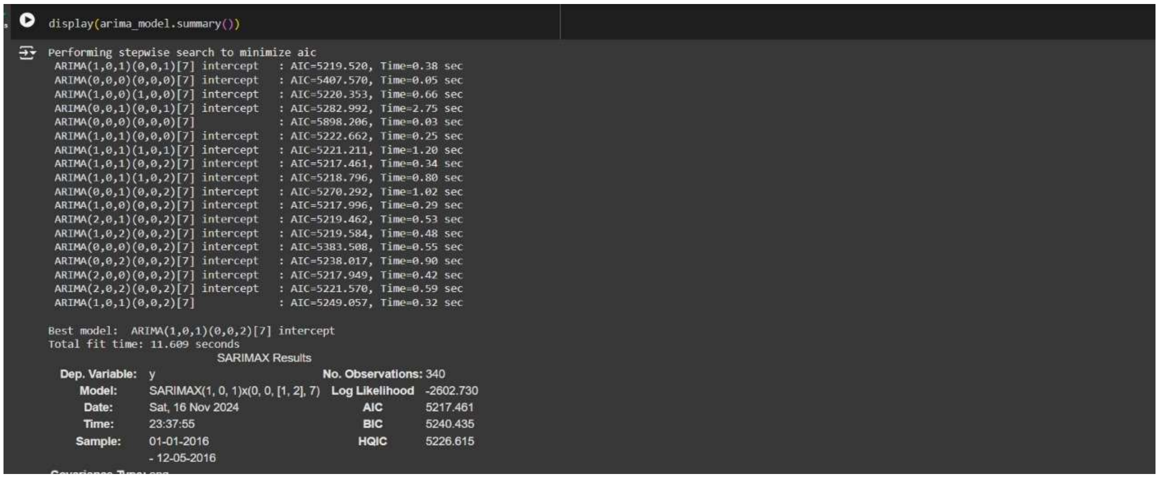
Performing the Augmented Dickey-Fuller test to find the p value so that we can check if the data is stationary or not.



```
# Augmented Dickey-Fuller test
result = adfuller(train_set['Consumed energy'])
print(f'p-value of Augmented Dickey-Fuller test is {round(result[1],2)}')
```

```
p-value of Augmented Dickey-Fuller test is 0.04
```

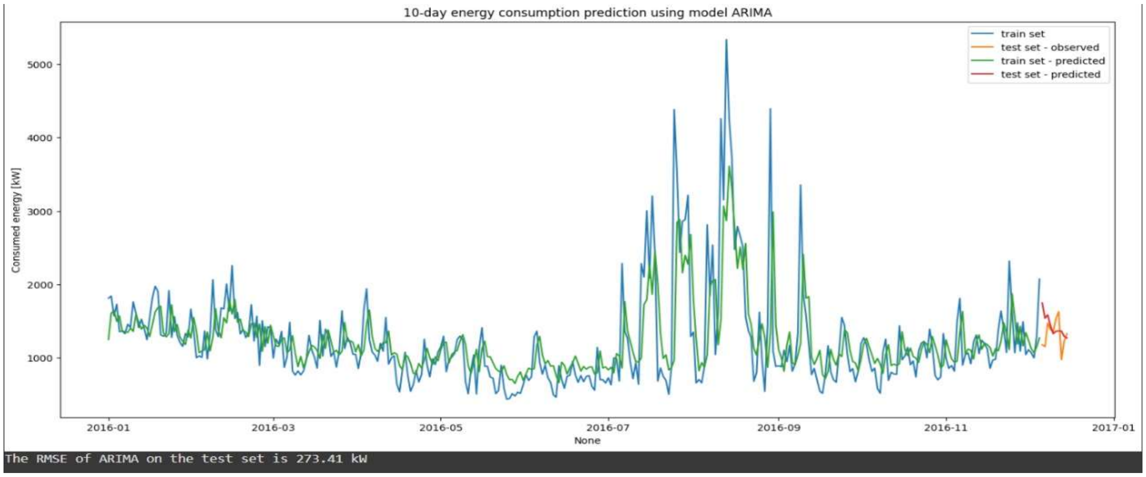
Performing auto arima to find the best model



```
display(arima_model.summary())
```

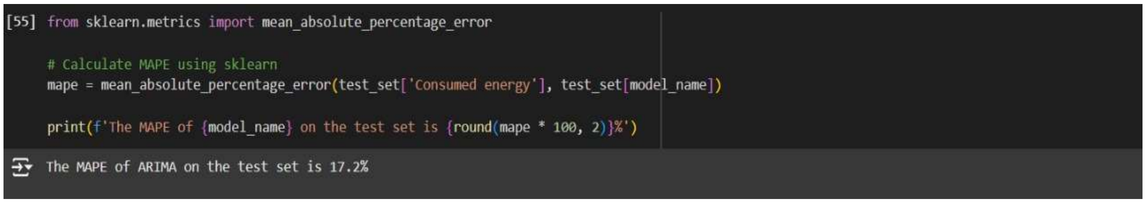
```
Performing stepwise search to minimize aic
 ARIMA(1,0,1)(0,0,1)[7] intercept   : AIC=5219.520, Time=0.38 sec
 ARIMA(0,0,0)(0,0,0)[7] intercept   : AIC=5407.570, Time=0.05 sec
 ARIMA(1,0,0)(1,0,0)[7] intercept   : AIC=5220.353, Time=0.66 sec
 ARIMA(0,0,1)(0,0,1)[7] intercept   : AIC=5282.992, Time=2.75 sec
 ARIMA(0,0,0)(0,0,0)[7]             : AIC=5898.206, Time=0.03 sec
 ARIMA(1,0,1)(0,0,0)[7] intercept   : AIC=5222.662, Time=0.25 sec
 ARIMA(1,0,1)(1,0,1)[7] intercept   : AIC=5221.211, Time=1.20 sec
 ARIMA(1,0,1)(0,0,2)[7] intercept   : AIC=5217.461, Time=0.34 sec
 ARIMA(1,0,1)(1,0,2)[7] intercept   : AIC=5218.796, Time=0.80 sec
 ARIMA(0,0,1)(0,0,2)[7] intercept   : AIC=5270.292, Time=1.02 sec
 ARIMA(1,0,0)(0,0,2)[7] intercept   : AIC=5217.996, Time=0.29 sec
 ARIMA(2,0,1)(0,0,2)[7] intercept   : AIC=5219.462, Time=0.53 sec
 ARIMA(1,0,2)(0,0,2)[7] intercept   : AIC=5219.584, Time=0.48 sec
 ARIMA(0,0,0)(0,0,2)[7] intercept   : AIC=5383.508, Time=0.55 sec
 ARIMA(0,0,2)(0,0,2)[7] intercept   : AIC=5238.017, Time=0.90 sec
 ARIMA(2,0,0)(0,0,2)[7] intercept   : AIC=5217.949, Time=0.42 sec
 ARIMA(2,0,2)(0,0,2)[7] intercept   : AIC=5221.570, Time=0.59 sec
 ARIMA(1,0,1)(0,0,2)[7]             : AIC=5249.057, Time=0.32 sec

Best model:  ARIMA(1,0,1)(0,0,2)[7] intercept
Total fit time: 11.609 seconds
                          SARIMAX Results
 Dep. Variable:  y                                No. Observations: 340
     Model:      SARIMAX(1, 0, 1)x(0, 0, [1, 2], 7)  Log Likelihood  -2602.730
      Date:      Sat, 16 Nov 2024                        AIC         5217.461
      Time:      23:37:55                                BIC         5240.435
    Sample:      01-01-2016                              HQIC        5226.615
                 - 12-05-2016
Covariance Type: opg
```

Plotting the ARIMA forecasted values and RMSE of the model



```
The RMSE of ARIMA on the test set is 273.41 kW
```

Mean absolute percentage error(MAPE) of the ARIMA model

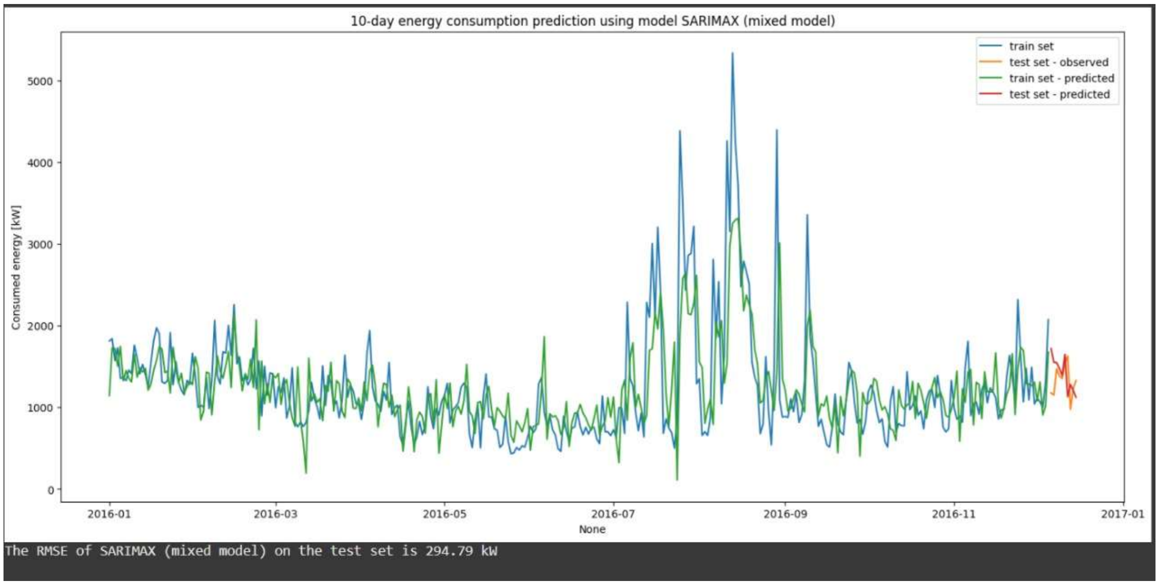


```
[55] from sklearn.metrics import mean_absolute_percentage_error

     # Calculate MAPE using sklearn
     mape = mean_absolute_percentage_error(test_set['Consumed energy'], test_set[model_name])

     print(f'The MAPE of {model_name} on the test set is {round(mape * 100, 2)}%')

The MAPE of ARIMA on the test set is 17.2%
```

7) SARIMAX Model

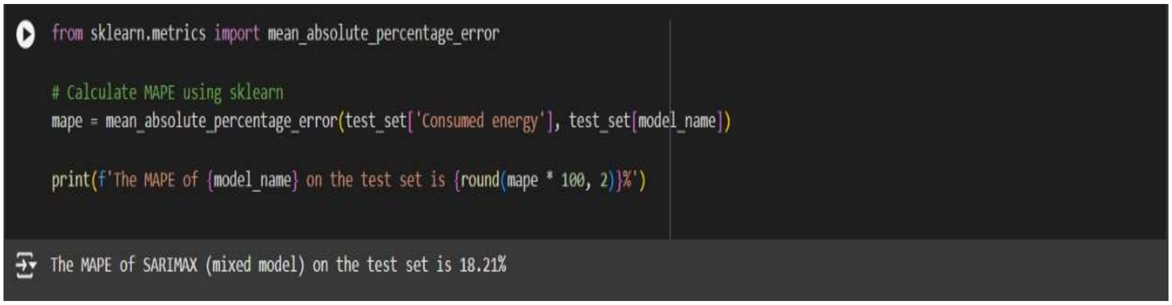
Using extraneous features to perform SARIMAX

The plot of SARIMAX forecasted values and RMSE of the model



```
The RMSE of SARIMAX (mixed model) on the test set is 294.79 kW
```

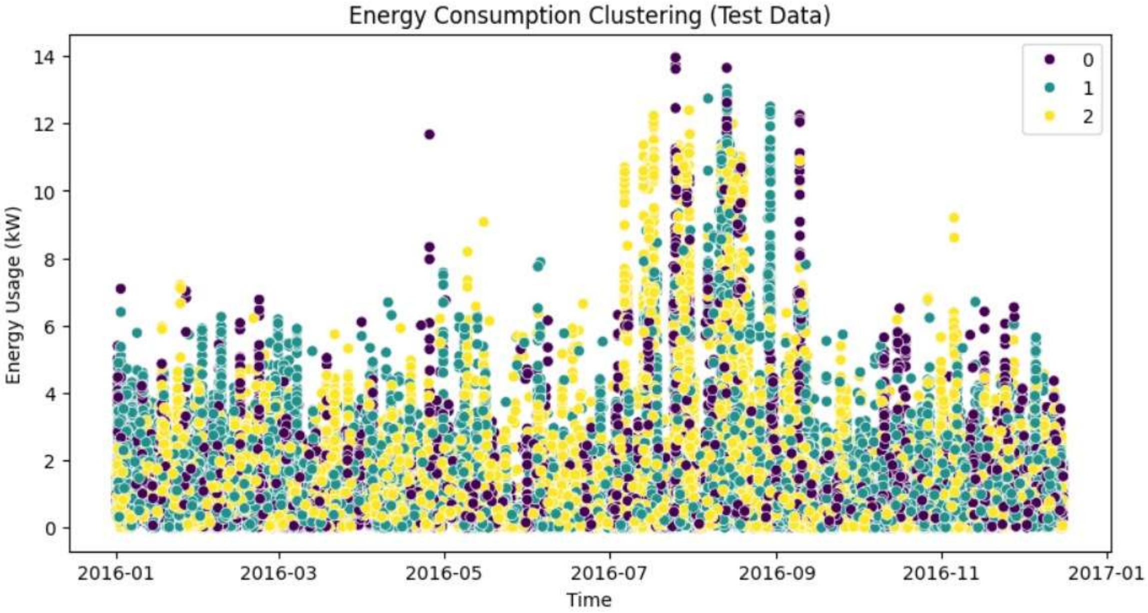
Mean absolute percentage error(MAPE) of the SARIMAX model



```
from sklearn.metrics import mean_absolute_percentage_error

# Calculate MAPE using sklearn
mape = mean_absolute_percentage_error(test_set['Consumed energy'], test_set[model_name])

print(f'The MAPE of {model_name} on the test set is {round(mape * 100, 2)}%')

The MAPE of SARIMAX (mixed model) on the test set is 18.21%
```

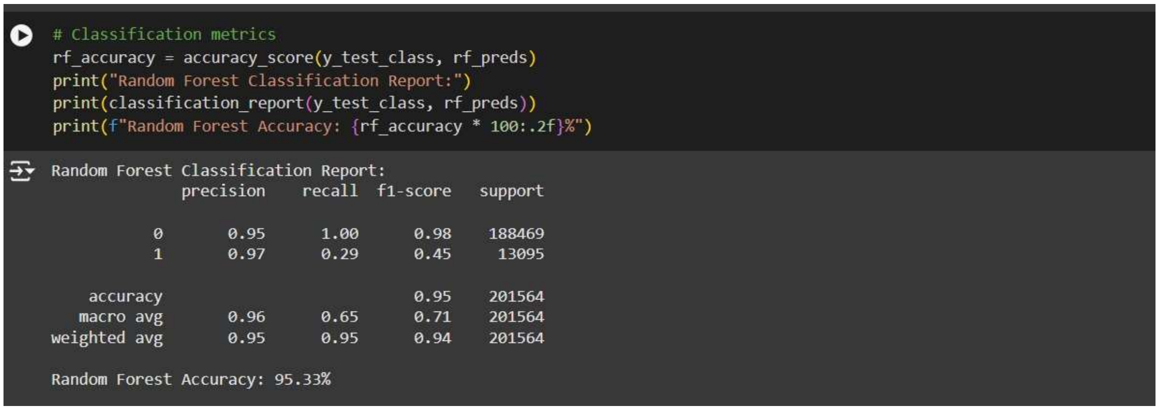
8) K-Means Clustering

The number of clusters identified by the model are 4 and the following visualization shows distinct patterns in energy consumption.



9) Random Forest Classifier

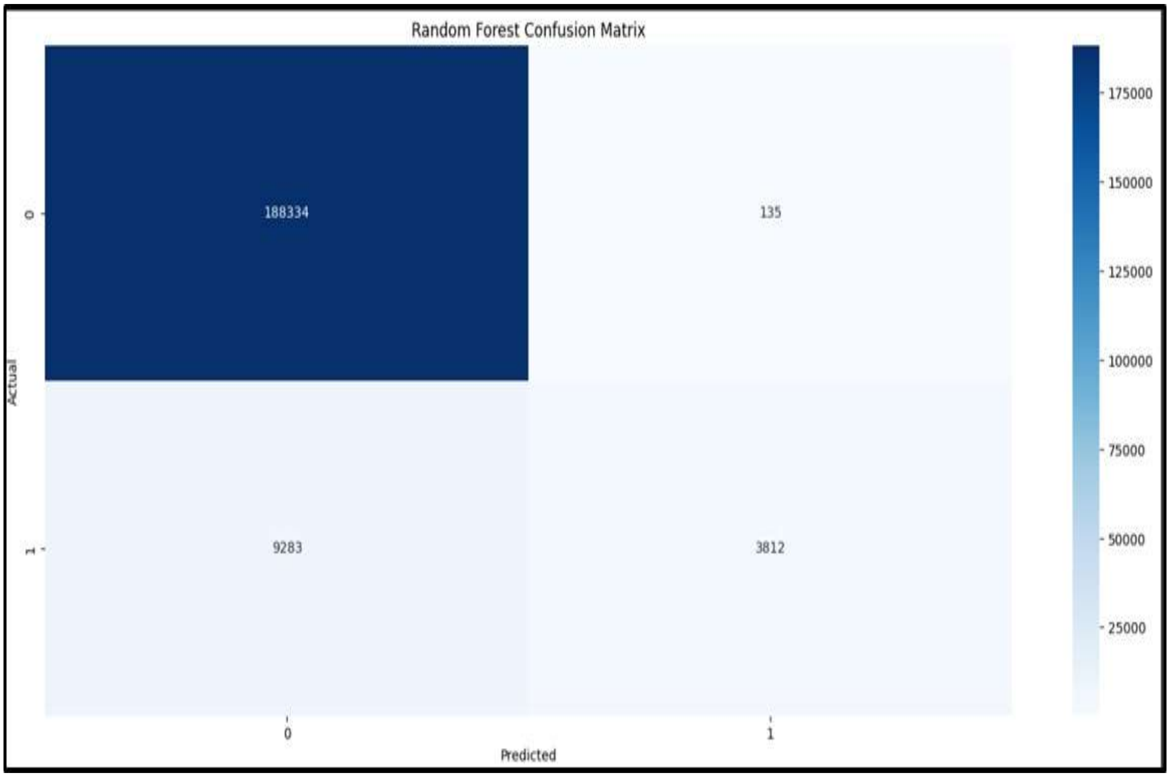
The classification metrics of the model are as values which includes precision, recall, f-1 score and the overall accuracy of the model.



```
# Classification metrics
rf_accuracy = accuracy_score(y_test_class, rf_preds)
print("Random Forest Classification Report:")
print(classification_report(y_test_class, rf_preds))
print(f"Random Forest Accuracy: {rf_accuracy * 100:.2f}%")
```

```
Random Forest Classification Report:
              precision    recall  f1-score   support

           0       0.95      1.00      0.98    188469
           1       0.97      0.29      0.45     13095

    accuracy                           0.95    201564
   macro avg       0.96      0.65      0.71    201564
weighted avg       0.95      0.95      0.94    201564

Random Forest Accuracy: 95.33%
```

Confusion Matrix of the model:



CONCLUSION

The project on Energy Consumption Forecasting for Smart Grid Optimization showcased how utilizing novel machine learning algorithms can be employed in some critical issues in energy management. Time-series based models such as SARIMAX, systems of classification such as Random Forest and, isolation processes including Isolation Forest were utilized by the system to offer solutions to forecasting, performance classification and even fault detection in energy consumption.

Main Success

- Accurate Forecasting:

The SARIMAX model greatly increased the predictive accuracy when it was complemented with other factors such as temperature, humidity, and cloudiness.

On the other hand, the RMSE and MAPE values for the SARIMAX model were greater than those for the basic ARIMA model but SARIMAX performed better in terms of interpretation as was visible in the plots.

- Efficient classification:

The Random Forest classification engendered high accuracy in classifying energy usage as efficient or inefficient which enhanced the strategies for reducing energy consumption.

- Clustering Insights:

K-Means clustering assisted in classifying users into different segments according to their energy consumption patterns, which enabled high energy users and efficient users to be pinpointed.

The project shows the possibilities of machine learning for improvement of energy management systems:

- Higher Quality of Grid service: Due to the ability to accurately predict futures and identify outliers, the system averts overloads and guarantees dependable energy delivery.
- People’s Management: Utility firms may obtain in-depth information and apply clustering and classification processes to come up with individual energy conservation plans for consumers.
- Ecological Improvements: Better energy performance leads to minimization of waste and facilitates the integration of renewable energy sources.

FUTURE SCOPE

- Incorporating Renewable Energy Data:

The inclusion of data from renewable sources enables renewable energy-powered smart grids to increase their prediction and optimization efficiency.

- Real-time improvements:

Create a system that provides real-time energy consumption recommendations based on demand forecasts, weather conditions and usage levels.

- Anomalies Description:

Hybrid anomaly systems with interpretable AI should be used to enhance current systems and identify important causes of AI.

- User segmentation and behavior analysis using clustering analysis to identify energy conservation tips and monitor both consumption and cost.

- Scalability:

Develop the dimension of the project to incorporate a single household, commercial buildings, or a whole community for grid level optimization.

- Integration with IoT Devices:

Employ the use of smart home IoT devices to improve model accuracy and robotic control systems through the collection of real-time data.

REFERENCES

1) In their July 2021 issue of IEEE Transactions on Industrial Informatics, A. L. Gope, D. Jain, and S. Goswami titled "Smart Meter Data Analytics for Optimal Energy Usage in Residential Buildings," pp. 4875–4884 (doi: 10.1109/TII.2021.3054162).

2) "A Review of Machine Learning Applications for Solar Photovoltaic Systems," by M. Q. Raza, G. Nadarajah, and V. K. Ramachandaramurthy, IEEE Access, vol. 7, pp. 4874–4893, January 2019. 10.1109/ACCESS.2018.2889579 is the doi.

3) "Demand Response in Electricity Markets: An Overview," by M. H. Albadi and E. F. El-Saadany, in the 2007 IEEE Power Engineering Society General Meeting, Tampa, FL, USA, 2007, pp. 1–5, doi: 10.1109/PES.2007.385728.

4) Torabi, M., Hashemi, S., Saybani, M.R., Shamshirband, S., Mosavi, A.: A hybrid clustering and classification technique for forecasting short-term energy consumption. Environ. Prog. Sustain. Energy 38(1), 66–76 (2018).

5) "Using Weather Ensemble Predictions in Electricity Demand Forecasting," IEEE Transactions on Power Systems, vol. 17, no. 3, pp. 624–629, August 2002, by J. W. Taylor and R. Buizza. 10.1109/TPWRS.2002.800987 is the doi.

6) Ho S. L. and Xie M., The use of ARIMA models for reliability forecasting and analysis, Computers & Industrial Engineering. (1998) 35, no. 1-2, 213–216