FORECASTING ELECTRICAL ENERGY CONSUMPTION

Submitted by

Marmik Parashar [240810125002]

Vansh Sachdeva [240810125003]

Ayushi Dadhich [240810125004]

Shashwat Mishra [240810125007]

For the award of the

Post Graduate Diploma in Big Data Analytics (PG-DBDA)

Under the supervision of

Name: Mr. Anurag Singh

Designation: Project Engineer

CENTRE FOR DEVELOPMENT OF ADVANCED COMPUTING

Session: AUG-2024 To FEB-2025

INSTITUTIONAL AREA, JASOLA, NEW DELHI – 110025

DECLARATION

We declare that this written submission represents our ideas in our own words and where other's ideas or words have been included, we have adequately cited and referenced the original sources.

We also declare that we have adhered to all the principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will cause for disciplinary action by the Institute and so evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed. We further declare that if any violation of the intellectual property right or copyright occurs my supervisor and Institute should not be held responsible for the same.

Name : Marmik Parashar

PRN No : 240810125002

Place : New Delhi

Date :

Signature :

Name : Vansh Sachdeva

PRN No : 240810125003

Place : New Delhi

Date :

Signature :

Name : Ayushi Dadhich

PRN No : 240810125004

Place : New Delhi

Date :

Signature :

Name : Shashwat Mishra

PRN No : 240810125007

Place : New Delhi

Date :

Signature :

ACKNOWLEDGEMENT

"Enthusiasm is the feet of all progress, with it there is accomplishment and Without it there are only slits alibis."

Acknowledgment is not a ritual but is certainly an important thing for the successful completion of the project. At the time when we were made to know about the project, it was tough to proceed further as we were to develop the same on a platform, which was new to us. More so, the coding part seemed tricky that it seemed to be impossible for us to complete the work within the given duration.

We really feel indebted in acknowledging the organizational support and encouragement received from the CDAC Delhi.

The task of developing this system would not have been possible without the constant help of our faculty members and friends. We take this opportunity to express our profound sense of gratitude and respect to those who helped us throughout the duration of this project.

We express our gratitude to our supervisors Mr. Anurag Singh and Mr. Sunil Kumar for giving their valuable time and guidance to us.

Name:

Date:

Name:

Date:

Name:

Date:

Name:

Date:

ABSTRACT

This project is focusing on developing an advanced electricity consumption forecasting model using machine learning techniques to enhance the versatility and accuracy of the forecasts. Using traditional statistical models may result in failing to capture the non-linear nature of modern energy system, particularly when they are influenced by weather, global crises such as COVID-19 and economic changes.

This project is being developed to cap this gap by integrating different machine learning models such as TBATS, ARIMA and SARIMA to analyse huge datasets and reveal hidden patterns in different power consumption zones of Morocco.

The approach is delivering high scalability and adaptability, so it can be used on different economic, climatic and geographical conditions. It also ensures reliability among all datasets by using some special techniques for augmenting the data (synthetic data generation) and feature engineering. This methodology includes data collection, data pre-processing, model selection and evaluation then final deployment using web interface.

ARIMA, TBATS and SARIMA are the major machine learning models used, all of these are evaluated for their performance. In this project, they will contribute to more efficient energy production and more sustainable energy management practices.

TABLE OF CONTENT

LIST OF FIGURES

OBJECTIVE

The objective of this project is to make an advanced forecasting model for electrical energy consumption by using machine learning techniques to improve prediction accuracy and adaptability. Statistical models that are traditional often fail to justify the dynamic and non linear nature of modern energy systems, particularly with the control of external factors such as natural factors and epidemics like COVID-19, weather and economic changes.

This project aims to bridge this gap by integrating various machine learning models like ARIMA and SARIMA to analyse large datasets and uncover hidden patterns in energy consumptions from total power consumption zone of Morocco.

Our project is designed to provide highly adaptable and scalable energy consumption forecasting irrespective of the dataset used. Although we are currently utilizing a dataset from Morocco, the methodologies, algorithms, and synthetic data generation techniques we implemented ensures that our approach can be applied to diverse geographical, economic, and climatic conditions.

This project aims to improve forecasting accuracy, providing a base for more efficient energy production and sustainable energy management practices and tools. Using techniques like synthetic data generation and feature engineering this project ensures reliability across any dataset and the techniques applied make it scalable and highly adaptable for global deployment.

CHAPTER 1

INTRODUCTION

The complexity and uncertainty of energy use patterns is one of the main barriers to energy forecasting. Rapid changes in consumption are driven by a number of factors, including urbanisation, fuel price fluctuations, industrial demand, climate conditions, and worldwide crises like COVID-19. As a result, standard models find it challenging to produce accurate estimates. Because energy output from renewable energy sources (solar, wind, and hydro) is inherently intermittent, integrating them into the power system increases unpredictability. Energy supply chain disruptions and financial losses can arise from overproduction or shortages caused by inaccurate energy projections.

Furthermore, the energy industry's quick automation is creating new data sources like satellite imaging, smart meters, and Internet of Things-enabled gadgets. For predictive modelling, these various data streams present both possibilities and difficulties. Large-scale energy consumption data makes prediction more possible, however model performance may be impacted by noisy, insufficient, or biassed data. Advanced computational methods that can handle massive information, identify significant patterns, and adjust to changes in real time are needed to overcome these obstacles.

As deep learning (DL) and machine learning (ML) methods have become more popular, researchers have created highly complex algorithms that can analyse vast amounts of real-time and historical data to identify intricate patterns in energy usage. Nonlinear and dynamic trends in energy use have been successfully handled by machine learning models including ARIMA, SARIMA, and T-BATS. In contrast to conventional techniques, these models have the ability to combine several influencing aspects, including temperature, economic indicators, and consumer behaviour, as well as learn from past data and adjust to changing situations.

Anomaly detection and deep transfer learning are some new methods along with conventional ML models. These methods present new chances to improve energy consumption forecasts. Even in situations with little labelled data, transfer learning enhances forecast accuracy by enabling models trained on a dataset to generalise to other geographic locations and consumption patterns. In order to ensure a reliable and flexible forecasting

system, anomaly detection approaches based on transformer and diffusion models assist in identifying infrequent consumption peaks and fluctuations brought on by outside disruptions. The goal of this study is to use cutting-edge machine learning techniques to anticipate electric energy usage more accurately and adaptably. Through the use of real-time inputs, integration of data from many sources, and big datasets, the suggested model seeks to enhance decision-making regarding energy generation, distribution, and regulation. Additionally, the validity of prediction models is maintained across a variety of datasets, including the Moroccan dataset, through the use of data augmentation approaches. By improving the generalisation of models, synthetic variations enable them to be applied outside geographical constraints.

The significance of this study extends beyond forecast accuracy; it also advances the more general objectives of environmental stewardship, energy efficiency, and sustainability. Forecasts that are more accurate can help industry and governments cut waste, optimise energy consumption, and transition to a cleaner energy ecosystem. Through the use of machine learning-based forecasting, this study supports the global trend towards automated energy management, smart grids, and sustainable development.

In addition, this initiative will enhance energy planning and offer insightful information to reduce the dangers brought on by erratic energy consumption. Modern AI-driven forecasting models will let energy providers anticipate peak loads more precisely, modify grid operations in real time, and lessen their dependency on fossil fuels. In the end, the move to data-driven decision-making will foster the widespread use of renewable resources and climate-smart energy policy.

CHAPTER 2

METHODOLOGY

The procedure for electrical energy consumption prediction using machine learning follows a methodical approach that have steps such as data collection, model selection, pre-processing, training, evaluation and model deployment. The aim is to implement a highly adaptive and accurate predictive model that can handle sudden consumption shifts, non-linear trends and some external factors such as economic changes and climate fluctuations.

2.1. DATA COLLECTION AND PREPROCESSING

Step one is gathering historical data from various sources like energy suppliers, weather forecasts, smart meters and government databases etc. These datasets have information related to energy consumption, external factors affecting energy demand and economic indicators and weather conditions. In real world, data is usually incomplete or with full of inconsistencies like null values, duplicate values etc, so it requires preprocessing techniques such as Exploratory Data Analysis which in end helps to remove duplicates, handling missing values and smoothening of noise to ensure high data veracity.

Synthetic Data Generation:

For robustness of model and improved accuracy of the forecasts synthetic data is generated by adding noise around the standard deviation. It will help to simulate the changes in energy consumption and will help to reduce overfitting especially when training is being done on a limited dataset.

2.1.1 FEATURE SELECTION AND ENGINEERING

Refining input variables to improve model performance can be done by feature selection and engineering. This project is highlighting transforming the data distribution after handling outliers to get a Gaussian distribution, instead of using highly mathematical techniques like Recursive Feature Elimination and Principal Component Analysis.

Managing Outliers and Data Distribution :

- Statistical methods like Inter Quartile Range(IQR) visualized through box plots and Z score analysis are helpful in outlier detection.

- Box-cox and Yeo-Johnson transformation are used to normalize skewed distributions, both of these are part of data transformation techniques used in this project.

- Extreme values should be handled with caution as they are responsible for data sensitivity, so to make data less sensitive Robust Scaling is used. This scaling is used before Standard scaling for uniform distribution of data across all the features.

The steps mentioned above helps to improve the stability of the machine learning models and makes them more resistant to changes in energy consumption trends.

2.2 MODEL SELECTION

Selecting the right forecasting model is crucial to forecast electrical energy consumption with high accuracy. Considering the complex and dynamic nature of energy demand patterns, this project uses a combination of time series models (ARIMA, SARIMA, TBATS) suited for different aspects of forecasting.

1. Autoregressive Integrated Moving Average (ARIMA):

- ARIMA proves effective at finding linear relationships between time series data points because of its wide use as a statistical model. The data collection works effectively for short-term power predictions in situations with predictable patterns. Due to its straightforward nature this model functions as a dependable tool which makes it easily understandable.

- The main drawback of using ARIMA occurs when researchers need to analyse unsteady data because the model fails to process data affected by seasonal changes and weather patterns. The technique does not provide immediate responses to either sudden demand changes or lengthy interdependent relationships.

- ARIMA represents the fundamental approach to model comparison because it helps establish advanced prediction methods to handle short-term pattern fluctuations in energy usage.

2. Seasonal ARIMA(SARIMA):

- ARIMA forecasts become more effective through SARIMA because it includes seasonal pattern analysis which suits energy use prediction of cyclical data patterns. The method enhances forecasting precision in datasets that display repetitive pattern

consumption over daily or weekly or annual cycles but it keeps the interpretability features of ARIMA.

- Adjusted use of SARIMA models with several seasonal components might lead to increased computational complexity. It proves less effective than alternative methods for dealing with unexpected changes which do not follow established seasonal patterns.

- The modelling of repeated seasonal trends in SARIMA produces better results than ARIMA since it provides precise estimates for electricity use.

3. Trigonometric Functions, Box-Cox Transformation, ARMA Error, Trend and Seasonal Components(TBATS)

- This model is specifically designed for handling multiple non-linear trends and seasonality. It is best suitable for estimating energy usage and long-term patterns. Variations occurred by factors like climate changes and economical changes are managed by TBATS.

- TBATS requires a lot of computing power and has longer training times, especially for large datasets. Its hyperparameters must also be changed for performance to be optimized.

- Computing power is much needed by TBATS and it has longer training time for large datasets. Hyperparameters must also be varied for performance optimization. Seasonal energy usage is affected by daily peaks and weekly demand cycles, this model is important for collecting these issues and helps to enhance prediction accurately.

By using these three models, our project provides a complete comparative analysis by selecting the best approach for precise energy consumption projections.

2.3 MODEL CREATION

After selecting model, next stage is achieving unbiased evaluation for this it is necessary to split the pre-processed dataset into training and test sets. To improve the robustness of the model, this energy consumption dataset is transformed using synthetic data generation. For the augmentation of data, noise around standard deviation is used.

Data Scaling and Transformation

Outlier detection and data distribution transformation are applied to improve the model performance. The data is transformed into a Gaussian distribution using the following methods:

- **Robust Scaling** : reduces the impact of outliers.
- **Power Transformation (Yeo-Johnson and Box-Cox):** stabilizes the variance and normalizes skewed data.
- **Default scaling:** Ensures that the data is evenly distributed before the model is trained.

Hyperparameter Tuning

Models are fine-tuned using techniques like auto-ARIMA to optimize parameters such as p, d, q (ARIMA/SARIMA) and seasonal components (SARIMA/TBATS).

Evaluation metrics include:

- Mean Absolute Error (MAE)
- Mean Square Error(MSE)
- Root Mean Squared Error (RMSE)
- R^2 Score

This structured model training approach guarantees a fair comparison between different forecasting techniques.

2.3.1 TRAIN AND TEST SPLIT

The dataset is then split into training and test data to make sure that the model generalizes very well to new, unknown data. This makes the ratio between the train/test split as follows: 80:20%. This implies that 80% of the historic energy consumption dataset is used in training and the remaining 20% for the validation and the test phase, respectively.

Time Series Allocation :

The consumption of energy is time dependent. Thus, simple random allocation is avoided. Instead, the dataset is split into a time series where older data is used for training and newer data is used for testing. This allows the model to learn past trends and apply them to predict the future.

2.3.2 MODEL EVALUATION

After training, the models are rigorously evaluated using multiple metrics to assess accuracy, robustness, and predictive power. The following key evaluation metrics are considered:

1. MAE, or mean absolute error :

It calculates the mean absolute difference (MAE) between the actual and predicted energy consumption figures. It provides a straightforward metric of prediction accuracy by ignoring the direction of the errors. Since it makes fewer predictions, a model with a lower MAE value is more accurate.

2. Root mean squared error, or RMSE :

RMSE surpasses MAE because it penalizes larger deviations more harshly. Because RMSE squares the errors before averaging them, it magnifies the impact of significant errors and is hence particularly helpful for detecting sharp variations in energy usage. A model is more predictive and less likely to make large forecasting errors if its RMSE is smaller.

3. The R2 score error:

This is often known as the coefficient of determination evaluates how well a model takes dataset volatility into account. It determines the proportion of variability in real energy use that the prediction model can explain. The model is highly dependable in forecasting when its R2 value is close to 1, successfully accounting for most changes.

4. Analysing Model Performance :

Each forecasting model has unique benefits and drawbacks that influence how well it matches different patterns in energy consumption:

- For short-term forecasting, the Auto-Regressive Integrated Moving Average, or ARIMA, is a useful tool since it effectively captures linear trends. It struggles with intricate seasonal correlations and long-term fluctuations, though.

- By adding seasonal patterns, SARIMA (Seasonal ARIMA), an extension of ARIMA, improves forecasting accuracy and is better suited for datasets where consumption is greatly influenced by periodic shifts.

- Considering that it effectively manages numerous seasonal patterns, non-linearity, and unpredictable variations in energy demand, the TBATS model is anticipated to produce the most accurate projections.

The most accurate and trustworthy forecasting technique is found for practical application by examining these performance indicators and model behaviours, guaranteeing better energy management and planning.

2.4. DEPLOYMENT

The real-time forecasting system receives the best models from model evaluation and training procedures subsequent to model deployment. The deployment process includes:

- **Streamlit UI integration:** Streamlit is capable of fetching real-time changes in data as it directly fetching data from the python code and any changes that occurs in data or code are reflected in real-time .By using Streamlit we can provide a highly interactive and user friendly Web App .

- **Automatic model updates:** The forecasting models get trained intermittently using latest consumption data to adapt to evolving energy pattern changes.

- **Scalability and Adaptability:** The forecasting system adopts scalable design principles which permit easy adaptation into smart grid and energy management platforms.

The adopted methodology enables the project to develop a forecasting system with high accuracy that enables adaptation and efficiency to benefit energy optimization initiatives and sustainability goals.

CHAPTER 3

REQUIREMENT SPECIFICATIONS

SOFTWARE

- **OS:** Linux, Windows 11 Home

- **Text Editors**: VSCode, Anaconda Jupyter

- **Data Visualization:** Power BI, Spreadsheet Viewer

- **Browser:** Google Chrome or compatible browser

HARDWARE

- **Processor:** Intel i7 (8th Gen) or AMD Ryzen 7

- **RAM**: 12GB DDR4

- **Storage:** 100GB SSD

- **GPU**: WhiskeyLake-U GT2 [UHD Graphics 620]

CHAPTER 4

FLOW CHART(Project Flow)

DATA FLOW DIAGRAM

A Data Flow Diagram (DFD) is a graphical representation of how data moves through a system. It illustrates inputs, processes, storage, and outputs to efficiently understand data processing flows.
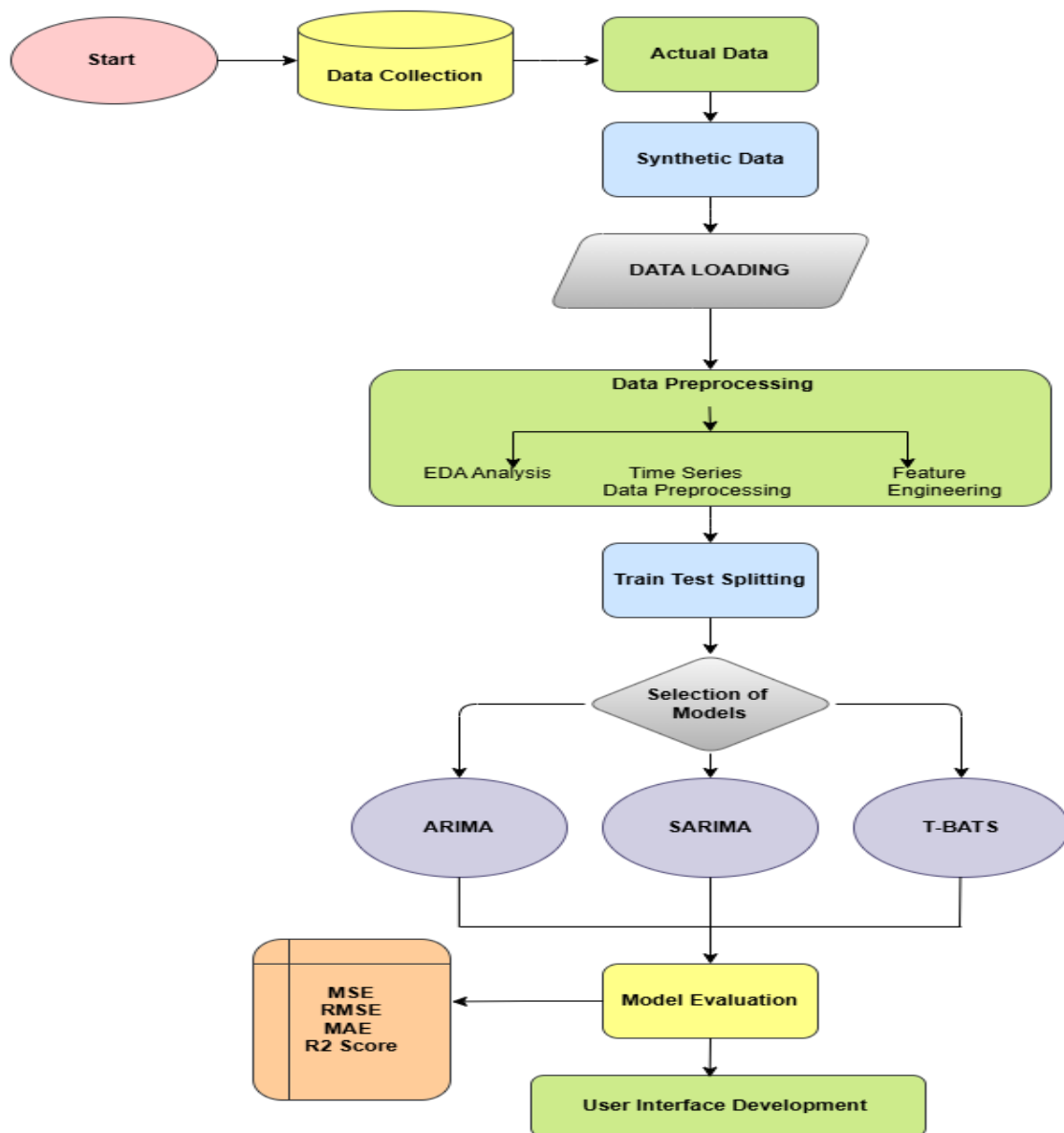

Figure 1 : Data Flow diagram for energy consumption project

ACTIVITY DIAGRAM

A system or process flowchart called an activity diagram shows both sequential activities and parallel actions and decision points as it presents complete workflow information.
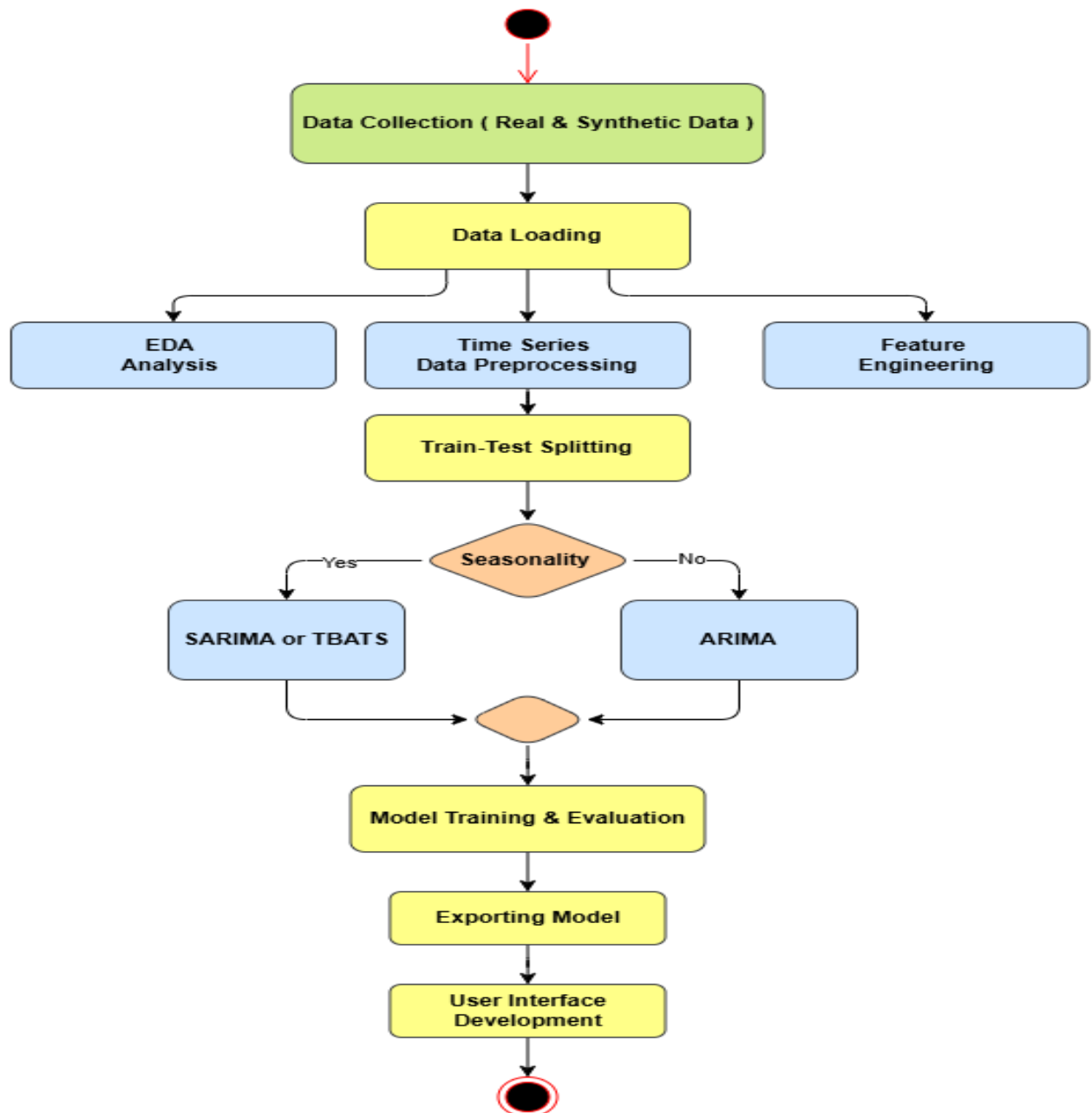

Figure 2 : Activity Diagram

GANTT CHART

The project management tool Gantt chart shows project plans through horizontal bars to display tasks with their time requirements together with task relationships and ongoing progress milestones.
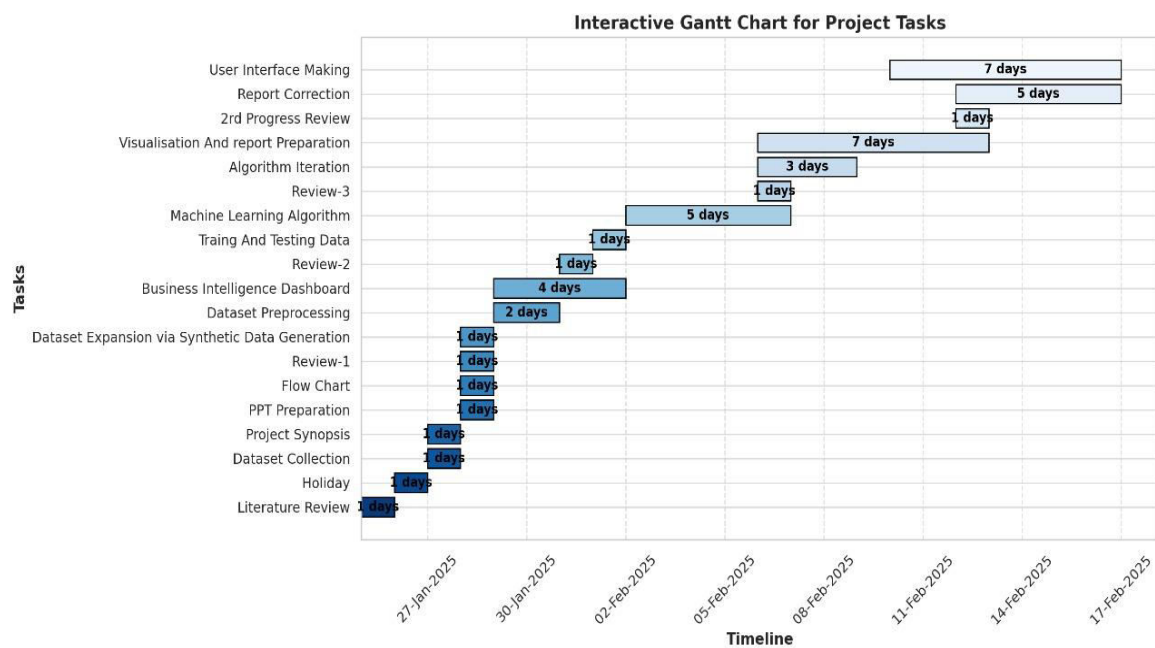

Figure 3 : Gantt Chart showcasing timeline for complete project

CHAPTER 5

TECHINCAL CODE

Importing the required libraries

All the required libraries like standard libraries of pickle and os, third party library imports for statistical calculations, numpy for numerical operations and pandas for data manipulation and analysis and for visualization part we have used seaborn and matplotlib.

For achieving stationarity in our data we have imported two statistical tests Adfuller and KPSS (Kwiatkowski–Phillips–Schmidt–Shin) test and relevant machine learning models like ARIMA, SARIMA and TBATS for predictive analysis.

Reading the dataset

The dataset is taken from Kaggle named as Morocco Energy Consumption. The initial size of this dataset was not suitable for better insights so, after implementing synthetic data generation size of dataset is increased. The expanded dataset's size is around 13 MB which is ready for analysis.

Feature Engineering

Datetime column is used as index and value is set to True and instead of analysing power consumption from different zones a single new column named as Total Power Consumption is made after dropping the individual power consumption zone columns. This feature engineering technique helps to reduce the number of classes to be predicted which saves computational memory.

Synthetic data generation

To increase the size and variance of our dataset, we have used synthetic data generation. In this technique noise around standard deviation has been introduced to increase the number of records in dataset for better learning from the dataset for machine learning models. Synthetic data generation in time series models like ARIMA, SARIMA and TBATS helps to improve forecast stability and accuracy. There are chances of model overfitting if data is trained through limited size so it helps in reducing overfitting too. To handle seasonality in different cycles, adding noise around standard deviation is helpful for models like TBATS, which is specifically designed for datasets with various seasonal trends.

Outlier Detection

In energy forecasting some factors like wind speed and temperature are the extreme values that affects the consumption forecast. So outlier detection using box plots helps to rely on regular patterns which ensures cleaner data and improving model performance. After carefully handling outliers the data becomes more structured and smooth, which also helps to reduce RMSE and MAE scores during evaluation. Outlier detection reduce the risk of biased predictions caused by extreme values.

Scaling

Two scaling methods are implemented for this project to improve model accuracy and stability. For skewed data handling and outliers robust scaling is used which detects outliers by transforming features on the basis of IQR(Inter Quartile Range). Standard scaling keeps the mean at zero and variance at one, that helps in conversion of dataset into Gaussian distribution. Both of the scaling steps ensures stable forecasting and reducing computational complexity of models.

Model Building

The model building phase for this time series project is using ARIMA, SARIMA and TBATS models. ARIMA is useful in storing linear trends by finding the best p, d and q values. Statistical test like Adfuller is used to calculate the value of d. Seasonal ARIMA extends ARIMA models by adding seasonal components to handle repetitive trends which helps to make the model more preferable with seasonal fluctuations. TBATS model is perfect for handling complex seasonality and irregular patterns. The code for model building checks for pre saved model to avoid redundant training and then selects the best model using auto_arima() from ARIMA and SARIMA model. Selecting the best model helps to increase the accuracy and also maintains the computational efficiency.

Model Evaluation

It is a process of checking a machine learning model's performance using common evaluation metrics for regression models like MSE, RMSE, MAE and R square score and accuracy, precision, F1 score, recall for classification models. A well evaluated model balances bias and variance and also prevents overfitting. In this project we have used ARIMA, SARIMA and TBATS models and their evaluation is mentioned in the next chapter.

CHAPTER 6

RESULT ANALYSIS

The research evaluated three time series forecasting approaches including ARIMA, SARIMA and TBATS to forecast future patterns. The ARIMA model demonstrated solid results when analysing stationary data although it works best when working with data that lacks substantial seasonality and trends. The model produced forecasts that were less accurate because of its difficulty in dealing with datasets that presented strong seasonal patterns.

Model Comparison

Model Name	Mean Square Error	Mean Absolute Error	Root Mean Squared Error	R-squared
ARIMA	0.095	0.274	0.309	-3.389
SARIMA	0.095	0.274	0.309	-3.39
TBATS	0.026	0.013	0.016	-0.381

The SARIMA analysis achieved superior results than ARIMA when dealing with seasonal time series. SARIMA achieved better forecasting accuracy because it integrated seasonal variance to detect underlying seasonal patterns. ARIMA failed to produce satisfactory results within patterns exhibiting monthly or quarterly seasonality until the dataset demonstrated these periodic features.
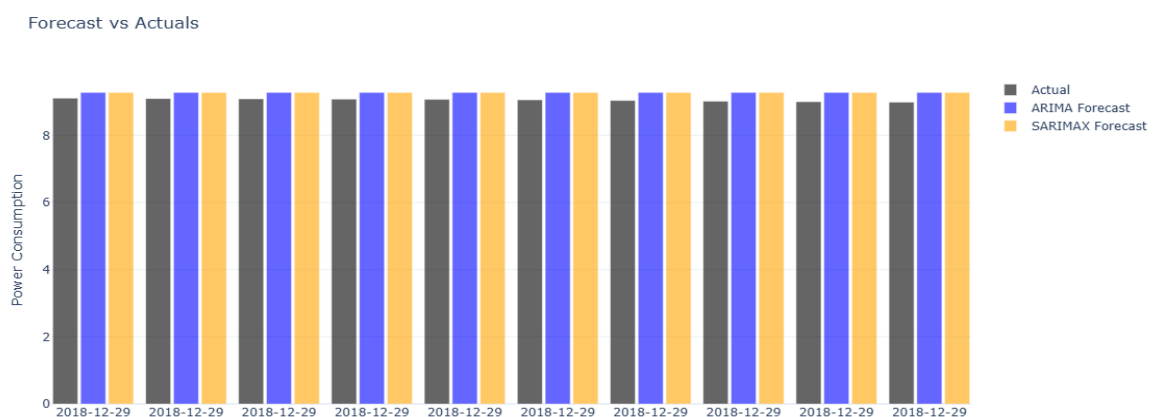

Figure 4: Box plot Showcasing Comparison between actual and predicted values

The TBATS model generated the superior results among these time series models because it handles multiple seasonal cycles and incorporates nonlinear trends on complex seasonality data. The TBATS model efficiently handled multiple seasonal periods by allowing complex seasonal effects modelling and produced better forecasts than both ARIMA and SARIMA during time-dependent seasonality changes.
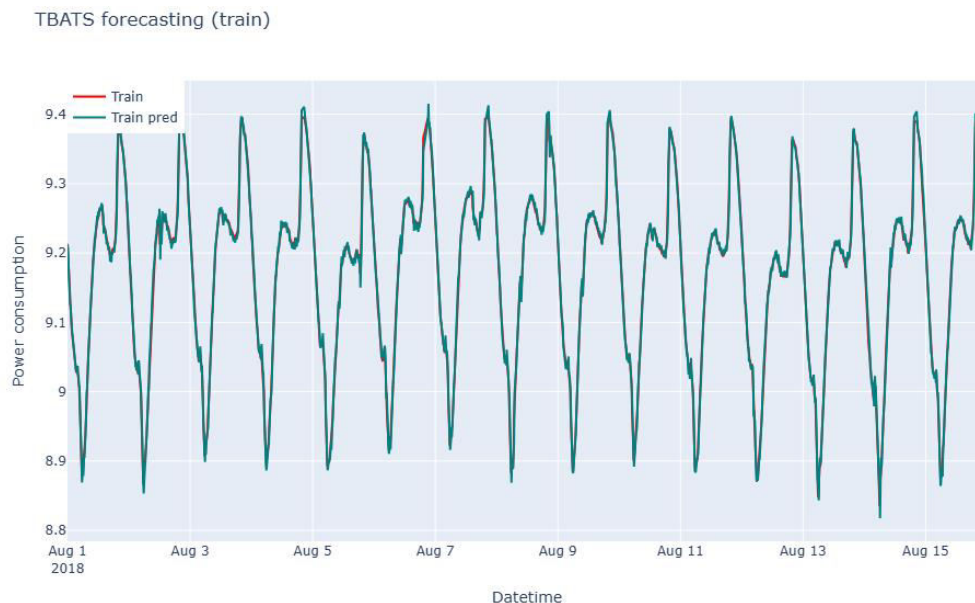

Figure 5 : Line Chart Comparison Between Actual And TBATS Predictions

TBATS produced the minimum RMSE from all models tested followed by SARIMA while ARIMA had the maximum error values. The analysis indicates that TBATS demonstrates the best capability to analyse intricate time series datasets.

UI deployment

For UI deployment we have used Streamlit which allows the development of creative, dynamic and real-time data driven apps using external data. It has been designed to display the forecasting results of electric power consumption using various time series models.

This web application sets up a page with a sidebar menu that helps with navigation through different sections of visualizations like Boxplots , Data Distributions and time series models like ARIMA , SARIMA and TBATS.

- Boxplots are used to visualize the distribution of different features in the dataset with and without scaling and skewness.
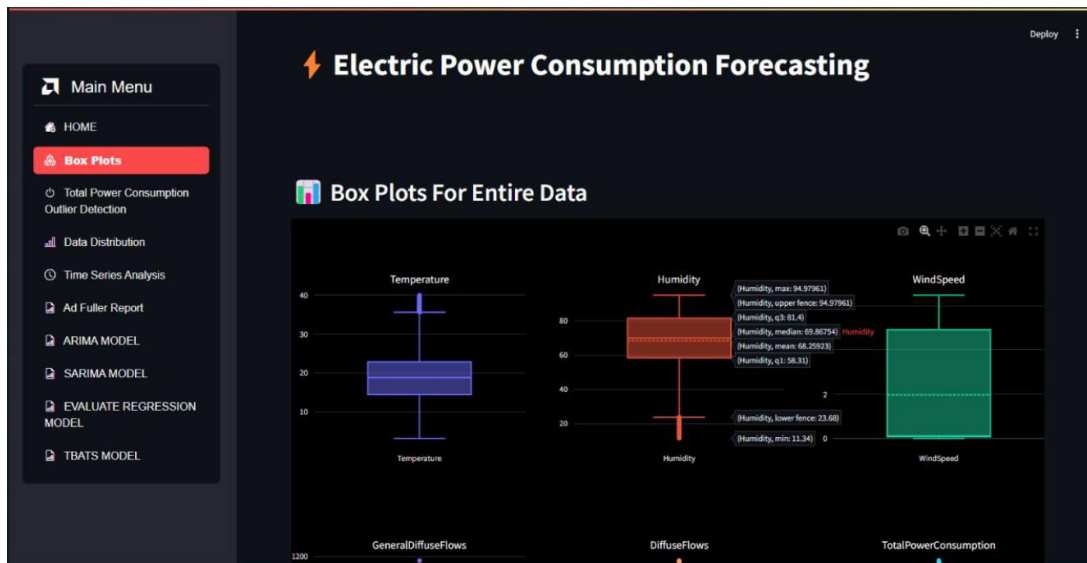

Figure 6 : Box Plots from Streamlit UI

- The Data Distribution Plots for various features are used for showing that transformation using boxcox and power transformer (yeo-johnson) can reduce skewness and present data in a normally distributed manner.

Figure 7 : Distribution Plots from Streamlit UI

- The Time Series Analysis plot gives the total power consumption over time and analyse the autocorrelation to understand patterns and seasonality.
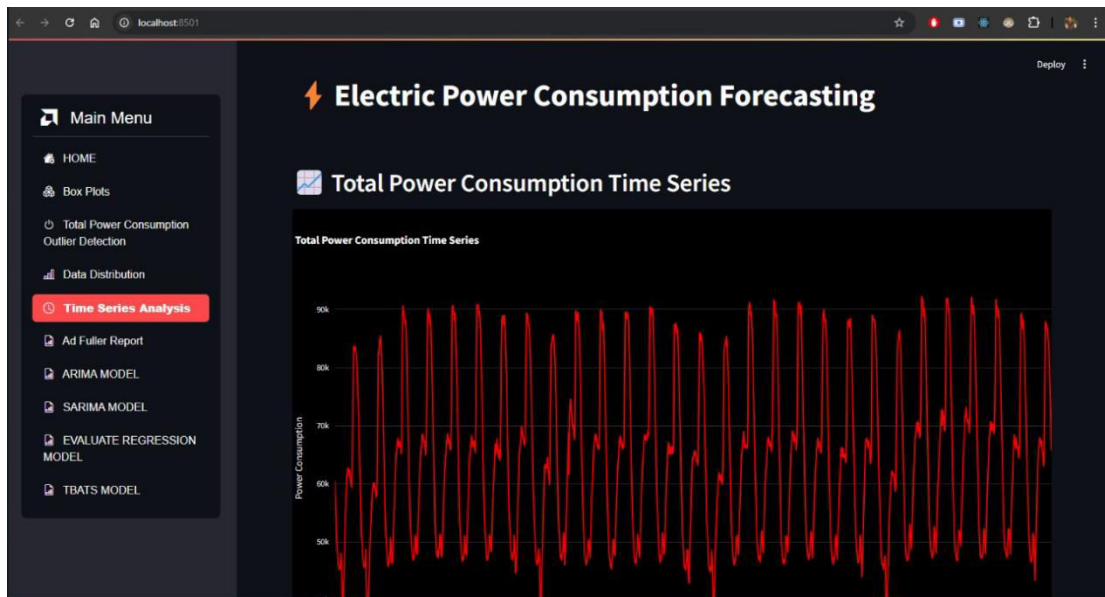

Figure 8 : Total Power Consumption Trend

- The Stationary Tests are performed using ADF and KPSS tests to check if time series data is stationary for the purpose of accurate forecasting.

Forecasting Models :

Models like ARIMA , SARIMA and TBATS are used for making predictions and evaluating performance metrics like MSE , MAE , RMSE and R-squared.

Figure 9 : ARIMA Model Predictions using Streamlit

The Web application has used Plotly library for creating interactive and visually appealing plots that are user friendly and provides summary of model performance.

Power BI Dashboard

Power BI dashboard represents zone wise power consumption trend with interactive filters and visuals, also we have used seasonal forecasting of next sixteen months using line charts.
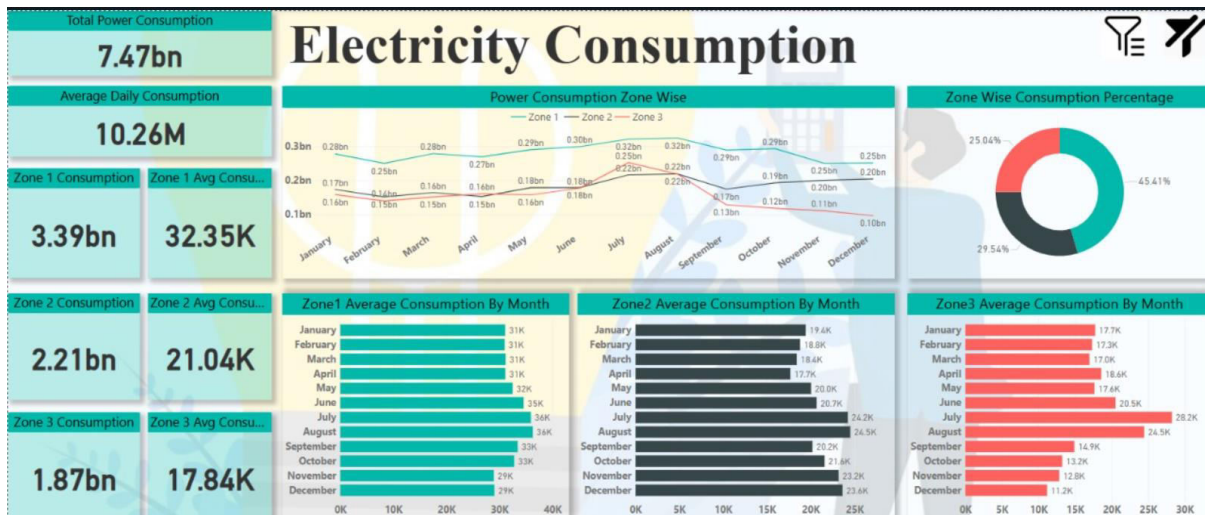

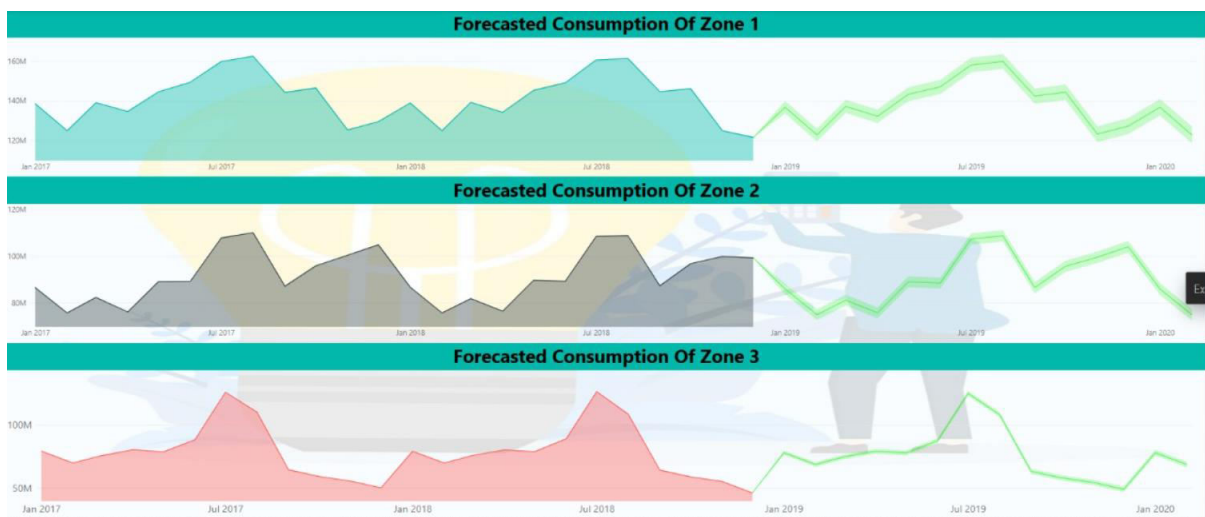*Figure 10 : Electricity Consumption Power BI dashboard*

Figure 11 : Forecasting Consumption of different zones using Power BI

CONCLUSION

The aim of this project was to predict time series data using different models and evaluate their accuracy and suitability for different types of datasets. Comparing ARIMA with other models revealed that it was effective for stationary datasets and cannot handle seasonality. SARIMA excels in handling seasonal data and provided better forecasting results than ARIMA specially for seasonal changes. Best model is TBATS, as it outperformed both ARIMA and SARIMA in more complex scenarios. For datasets with rigid seasonal fluctuations and provided the lowest RMSE values.

In closure, opting forecasting model depends on the type of the data. ARIMA remains a solid choice for non seasonal data, and other two provided more accurate solutions for data with seasonality or multiple seasonal cycles. Therefore, TBATS is recommended as perfect choice for complex time series predictive tasks because of its better accuracy and flexibility.

FUTURE SCOPE

For future usage, all the models discussed can be further improved by considering hybrid models. Such as strengths of ARIMA, SARIMA and TBATS can be combined that will lead to improve forecasting performance. Deep learning models such as Prophet and LSTM(Long Short-Term Memory) are known for handling complex and non linear relationships successfully in large datasets.

Model tuning like optimizing hyperparameters using random search techniques and grid search will yield better results. In addition, integrating external regressors (such as economic indicators and other time-dependent data) may improve forecasting accuracy, especially in real business applications.

REFERENCES

[1] P. Yan *et al.*, "A Comprehensive Survey of Deep Transfer Learning for Anomaly Detection in Industrial Time Series: Methods, Applications, and Directions," *IEEE access*, pp. 1–1, Jan. 2024, doi: https://doi.org/10.1109/access.2023.3349132.

[2] P. Yuan *et al.*, "Electricity consumption and load prediction method for Chinese rural residences based on the randomness and seasonality in electricity usage behavior," *Energy and Buildings*, vol. 279, p. 112647, Nov. 2022, doi: https://doi.org/10.1016/j.enbuild.2022.112647.

[3] S.-Y. Shin and H.-G. Woo, "Energy Consumption Forecasting in Korea Using Machine Learning Algorithms," *Energies*, vol. 15, no. 13, p. 4880, Jul. 2022, doi: https://doi.org/10.3390/en15134880.

[4] A. Bandlamudi, "Household Energy Consumption Forecasting Using LSTM and ARIMA Models: A Comprehensive Analysis," Jan. 2025, doi: https://doi.org/10.21203/rs.3.rs-5916961/v1.

[5] S. Xu *et al.*, "Impact of the COVID-19 on electricity consumption of open university campus buildings – The case of Twente University in the Netherlands," *Energy and Buildings*, vol. 279, p. 112723, Jan. 2023, doi: https://doi.org/10.1016/j.enbuild.2022.112723.

[6] Saleh Albahli, "LSTM vs. Prophet: Achieving Superior Accuracy in Dynamic Electricity Demand Forecasting," *Energies*, vol. 18, no. 2, pp. 278–278, Jan. 2025, doi: https://doi.org/10.3390/en18020278.

[7] S. Kumari and S. K. Singh, "Machine learning-based time series models for effective CO2 emission prediction in India," *Environmental Science and Pollution Research*, Jul. 2022, doi: https://doi.org/10.1007/s11356-022-21723-8.

[8] S. Rajesh, Shalini Justin R, Rajesh Shankaranarayana, and G. N. Jetty, "Sustainability Index Analysis using ARIMA & TBATS Approach," *Research Square (Research Square)*, May 2024, doi: https://doi.org/10.21203/rs.3.rs-3994605/v1.

[9] X. Wu *et al.*, "Improving Synchronization and Stability in Integrated Electricity, Gas, and Heating Networks via LSTM-Based Optimization," *Energies*, vol. 18, no. 3, p. 749, 2025, doi: https://doi.org/10.3390/en18030749.