Load Forecasting using ARIMA

2020UEE4611 Divy Kant Sharma

2020UEE4623 Abhishek Yadav

2020UEE4630 Deepender

2020UEE4647 Ansham

SUPERVISOR:

DR. MANISHA

Contents

Introduction

- **Load forecasting** is essential in several disciplines, including energy management, resource allocation, and financial planning. It entails forecasting future demand using past data and a variety of affecting factors. The **Autoregressive Integrated Moving Average (ARIMA)** model, which uses statistical techniques to discover patterns and trends in time series data, is a prominent approach to load forecasting. However, in real-world scenarios, load data can exhibit complex non-linearities and seasonal variations that ARIMA models may struggle to capture accurately.

- This is where **clustering algorithms** come into play. By grouping similar data points into clusters, we can effectively segment the data based on underlying patterns and characteristics. This allows for hybrid models that combine the strengths of both ARIMA models and clustering algorithms.

ARIMA Model: The ARIMA model is a time series forecasting technique that combines autoregression (AR), differencing (I), and moving averages (MA). The three components of the model work together to capture different patterns and trends in time series data.

- **AutoRegressive (AR):** This component accounts for the relationship between the current value and its past values. It assumes that the future value of a variable depends on its previous values.
- **Integrated (I):** This component involves differencing the time series data to make it stationary. Stationarity is important for ARIMA models as they work better with time series data that exhibit constant statistical properties over time.
- **Moving Average (MA):** This component considers the relationship between the current value and a residual error from a moving average of past values. It helps capture short-term fluctuations in the time series data.
- The ARIMA process is written with the notation **ARIMA(p,d,q)**, where **(p)** denotes the number of autoregressive orders in the model. Autoregressive orders specify the previous values from the series which are used to predict current values; difference **(d)** specifies the order of differencing applied to the series before estimating models; and moving average **(q)** specifies how deviations from the series mean for previous values are used to predict current values.

Literature Survey

- Shilpa G.N. and G.S. Sheshadri's presentation, which will be presented at the IEEE Bangalore Humanitarian Technology Conference in 2020, focuses on using time series analysis for electrical demand forecasting. It is anticipated to investigate methods and models for forecasting electricity consumption, providing insights useful to the power industry and related technology.

- Mahmoud A. Hammad, Borut Jereb, Bojan Rosi, and Dejan Dragan authored a 2020 paper providing a thorough review of methods and models in electric load forecasting. It likely covers a wide array of techniques, offering a comprehensive overview useful for understanding and implementing effective forecasting strategies in the energy sector.

- "The paper by Nepal, Yamaha, Sahashi, and Yokoe investigates building electricity usage patterns using the K-means clustering algorithm." It focuses on improving centroid determination and cluster quantity determination for improved analysis, providing insights into optimising these critical areas in understanding building electricity use."

Literature Survey

- "Han, Kamber, and Pei's foundational reference "Data Mining: Concepts and Techniques" covers data pretreatment, classification, clustering, and advanced approaches for mining varied data types. It is a comprehensive resource that offers methods and real-world examples, and it was published in 2012. It is essential for students and professionals in data mining and related topics.

- Chicco's paper provides an evaluation of clustering methods for grouping electrical load patterns. It offers an extensive overview and performance assessment of these methods, analyzing their effectiveness in handling load data. Crucial insights are provided for understanding and optimizing clustering techniques in electrical load pattern analysis.

- Zhang's paper offers a hybrid model for time series forecasting that combines ARIMA and neural networks. It investigates how these methods might be used to improve prediction accuracy, providing insights into the applicability and potential benefits of merging classic statistical approaches with neural networks in time series forecasting applications.

Gaps Identified

- The ARIMA (AutoRegressive Integrated Moving Average) model, a widely-used method for time series forecasting, does have its limitations. Notably, it assumes a **linear relationship** between past observations and future predictions, however many real-world time series data have complex relationships that cannot be captured by a linear model.

- ARIMA models can struggle with complex and evolving seasonal patterns, and they treat the entire time series as a homogeneous sequence, lacking the inherent capability to identify and adapt to different clusters or groups within the data.

- It does not take into account external factors that may affect the time series data, such as changes in the economy or weather patterns.

Objective of the Project

Load forecasting using ARIMA models with clustering aims to enhance the accuracy and efficiency of predicting future load values when compared to using ARIMA models alone. This is accomplished by combining the advantages of ARIMA and clustering.

Overall, the objective of this hybrid approach is to achieve more accurate, efficient, and reliable load forecasts, which can be valuable for a variety of applications, including:

1. **Energy management in buildings and smart grids:** Accurate load forecasts can help optimize energy consumption and reduce energy costs.
2. **Resource planning in power systems:** Load forecasts are crucial for planning the generation and transmission of electricity to meet future demand.
3. **Demand response programmes:** Accurate projections can aid utilities in the development of effective programmes that encourage consumers to adjust their electricity consumption in response to price signals.

Methodology

This methodology combines the strengths of the ARIMA model for time series forecasting with the ability of clustering algorithms to identify groups of similar data points. Here's a breakdown of the key steps:

1. **Data Preprocessing:**
   - Collect historical load data: This could be hourly, daily, or even monthly data depending on your desired forecast horizon.
   - Cleaning and filtering: Check for missing values, outliers, and inconsistencies in the data.
2. **Clustering:**
   - Cluster historical data: Group similar load patterns together based on their features.
   - Identify the cluster of the forecasting day: This cluster will be used for training the ARIMA model.

3. ARIMA Model Training:

- Select the appropriate ARIMA order (p,d,q): Analyze the autocorrelation and partial autocorrelation plots to identify the order of the autoregressive (p), integrated (d), and moving average (q) components.
- Train the ARIMA model: Use the historical data within the forecasting day's cluster to train the ARIMA model.

4. Validate the model:

- Evaluate the performance of the model using metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

5. Forecasting:

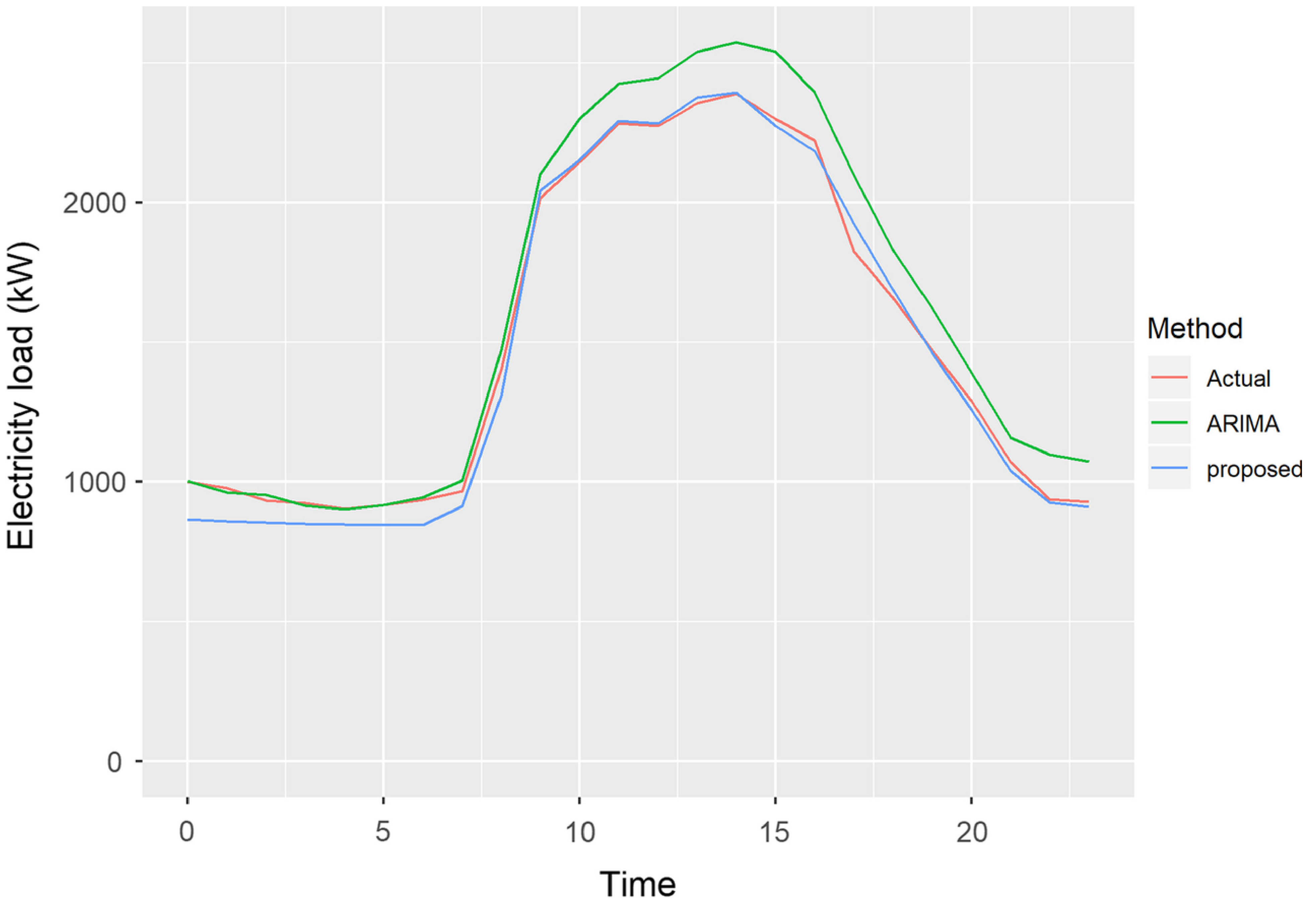
- Input the features of the forecasting day into the trained ARIMA model.
- Obtain the forecast load value for the desired horizon.
- Repeat steps 3 and 4 for each forecasting period.

Estimated Output

- The comparison between forecasting result of ARIMA model and proposed method in comparison to actual data on July 27, 2018.

- The graph of proposed model is more **closer** to graph of actual value as compared to ARIMA model.

- This shows that the proposed model has more **accuracy** as compared to ARIMA model.



Performance Of The Model

The electricity load forecasting of each day using the ARIMA model and the proposed method from April 1, 2018 till March 31, 2019 was performed. The Table shows the one-year average value of **MAPE**, **RMSE**, and **MAE** between the actual and the forecasted value using the ARIMA model and the proposed method by day of the week.

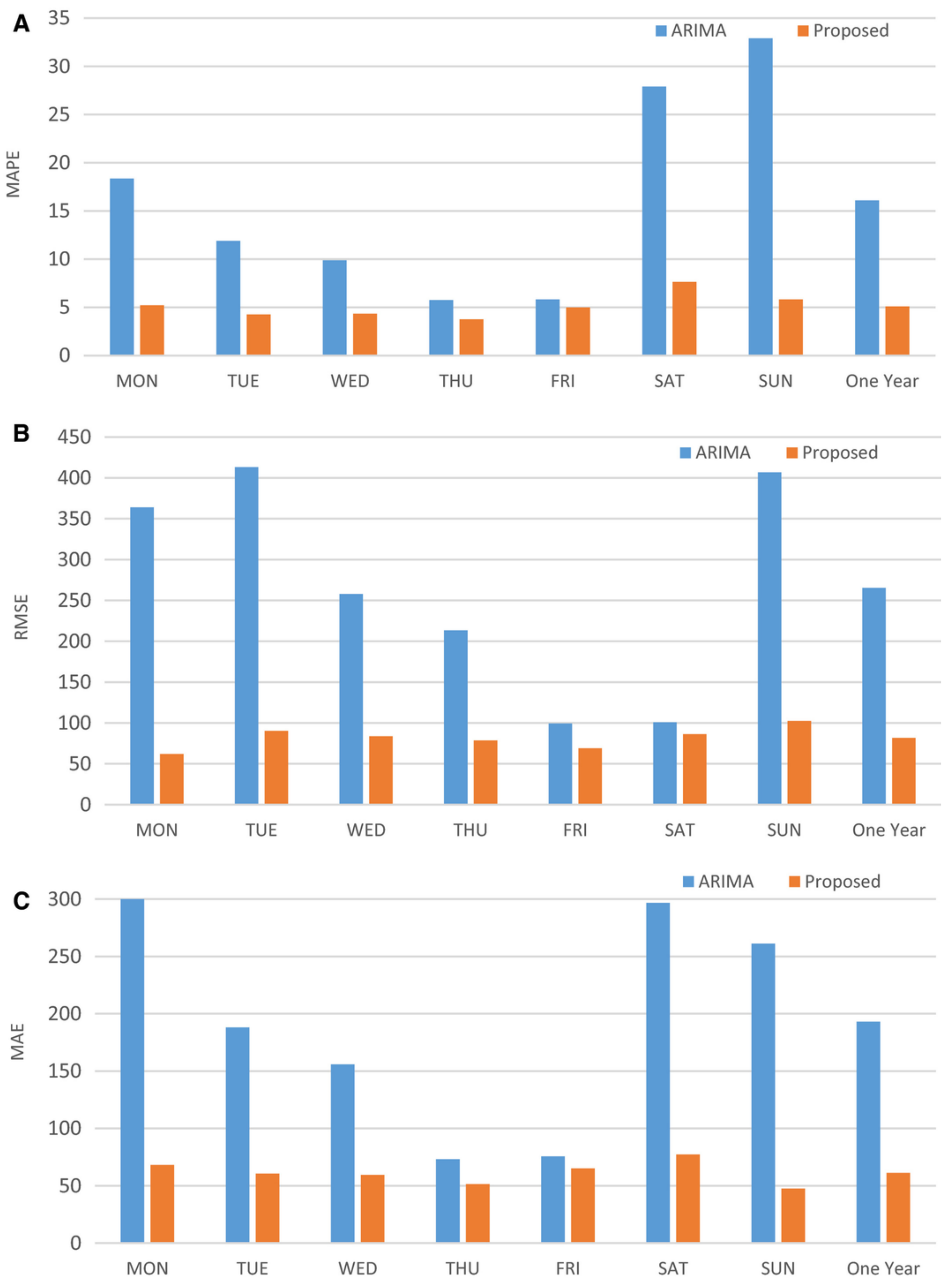
	ARIMA model			Proposed method		
	MAPE	RMSE	MAE	MAPE	RMSE	MAE
SUN	32.9	364	261.2	5.8	61.9	47.6
MON	18.4	413.1	299.9	5.2	90.4	68.1
TUE	11.9	258	188	4.3	83.8	60.7
WED	9.9	213.4	155.8	4.3	78.7	59.5
THU	5.8	99.4	73.2	3.8	69.0	51.4
FRI	5.8	100.9	75.6	5.0	86.2	65.2
SAT	27.9	406.8	296.7	7.6	102.4	77.4
One year	16.1	265.3	193.1	5.1	81.7	61.3

MAE, mean absolute error; MAPE, mean absolute percentage error; RMSE, root mean square error.

- The values of **MAE, RMSE,** and **MAPE** of the proposed method are found to be small in comparison to that of the ARIMA model.
- The values of **MAE, RMSE,** and **MAPE** of the proposed method shows that the proposed method produces **better results** than the ARIMA model.

Bar Graph

- The Figure represent the bar graph showing the values of **MAPE, RMSE,** and **MAE** for the ARIMA and proposed method, respectively.



Conclusion

In this project, the focus was on analyzing electricity usage data from Chubu University. The dataset was subjected to two main analyses: ARIMA modeling and clustering algorithms. Specifically, the ARIMA model was applied to the electricity usage data, with a particular emphasis on forecasting for the time period from 6 to 9 am. Simultaneously, the one-year dataset was employed for clustering using the K-means algorithm. The clustering process aimed to categorize the electricity usage patterns into six distinct clusters, facilitating a more granular understanding of the data. By combining ARIMA modeling and clustering, the project aimed to draw meaningful conclusions and insights from the Chubu University electricity consumption dataset, providing a comprehensive approach to analyze and understand the temporal and structural variations in the university's energy usage patterns.

Future work to be done

While ARIMA with clustering has shown promising results in load forecasting, several areas deserve further research and development:

1. Advanced Clustering Techniques:

- Exploring alternative clustering methods like hierarchical clustering, fuzzy clustering, and density-based clustering for improved group identification.
- Investigating dynamic clustering techniques that adapt to changing load patterns over time.

2. Hybrid ARIMA Models:

- Combining ARIMA with other forecasting methods like machine learning or deep learning algorithms for boosted accuracy.
- Utilizing hybrid models to capture both short-term and long-term load patterns.

3. Real-Time Implementation:

- Work on implementing the ARIMA with clustering model in real-time or near-real-time scenarios. This is particularly important for applications where timely and accurate forecasting is critical.

Additionally, future research could focus on:

- Investigating the effectiveness of transfer learning and knowledge sharing between similar load forecasting domains.
- Developing automated feature engineering techniques tailored to load forecasting tasks.
- Applying advanced optimization algorithms for parameter estimation in ARIMA models.
- Evaluating the economic benefits and cost-effectiveness of ARIMA with clustering for different application scenarios.

Researchers can increase the accuracy and reliability of load forecasting using ARIMA models with clustering by pursuing these paths, leading to major gains in energy management, grid planning, and operational efficiency across diverse industries.

References

[1] Shilpa. G. N, G. S. Sheshadri "Electrical Load Forecasting Using Time Series Analysis" IEEE Bangalore Humanitarian Technology Conference (2020) **DOI: 10.1109/B-HTC50970.2020.9297986**

[2] Mahmoud A. HAMMAD, Borut JEREB, Bojan Rosi, Dejan DRAGAN2 "Methods and Models for Electric Load Forecasting: A Comprehensive Review" (2020) **DOI: 10.2478/jlst-2020-0004**

[3] Ewa Chodakowska, Joanicjusz Nazarko and Łukasz Nazarko "ARIMA Models in Electrical Load Forecasting and Their Robustness to Noise" (2021) **DOI: https://doi.org/10.3390/en14237952**

[4] Jinliang Zhang a, Yi-Ming Wei b, Dezhi Li c, Zhongfu Tan a, Jianhua Zhou d "Short term electricity load forecasting using a hybrid model" (2018) **DOI: https://doi.org/10.1016/j.energy.2018.06.012**

[5] Sabu M. Thampi, El-Sayed M. El-Alfy and Ljiljana Trajkovic "Hybrid online model based multi seasonal decompose for short-term electricity load forecasting using ARIMA and online RNN"(2021) **DOI: 10.3233/JIFS-189884**

References

[6] Han J, Kamber M, Pie J. "*Data Mining Concepts and Technique*", Waltham, MA: Academic Press; Morgan Kaufmann Publisher (2018).

[7] Chicco G. Overview and performance assessment of the clustering methods for electrical load pattern grouping. Energy (2018) **DOI:10.1016/j.energy.2018.12.031**

[8] B Nepal, M Yamaha, H Sahashi "Energy conservation in university buildings by energy pattern analysis using clustering technique" (2021) **DOI: 10.2495/EQ-V4-N2-158-167**

[9] Haihong Bian, Yiqun Zhong, Jianshuo Sun, Fangchu Shi "Study on power consumption load forecast based on K-means clustering and FCM–BP model" (2020) **DOI: https://doi.org/10.3390/en12122451**

[10] Nepal B, Yamaha M, Sahashi H, Yokoe A "Analysis of building electricity use pattern using K-means clustering algorithm by determination of better initial centroids and number of clusters" (2021) **DOI : https://doi.org/10.3390/en12122451**

THANK YOU