**Load Value Forecasting utilizing SARIMA Model**

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model stands as an exceptional forecasting tool, leveraging its instantaneous average-based architecture to predict future values with high accuracy.

**ABSTRACT:**

Electricity load forecast for Chhattisgarh state has been accomplished using the Seasonal Autoregressive Integrated Moving Average (SARIMA) model. The input dataset comprises electricity load data for the years 2021, 2022, and 2023. This data is utilized to predict the load values for the year 2024 and 2025. The model's performance was evaluated using standard statistical metrics, ensuring accurate and reliable predictions.

**Keywords:**

1. **Forecasting**
2. **Regression**
3. **Moving Average**
4. **Seasonality**
5. **Lags**
6. **Error metrices**

**Introduction:**

Load forecasting has progressively become a crucial component of the energy management system. This study presents a powerful methodology for short term load forecasting with the help of previous load trends. ARIMA has limitations in handling seasonality directly, which led to the development of SARIMA (Seasonal ARIMA). SARIMA extends ARIMA by incorporating seasonal differencing and seasonal autoregressive and moving average terms, making it more adept at modelling seasonal patterns within the data.

Electrical system requires a subtle balance equilibrium and demand., which affects the demand for electrical energy. Additionally, the push for a sustainable grid that integrates renewable energy sources and electric vehicle technology to reduce pollution emissions add to the existing imbalance. Therefore, accurate demand estimation becomes crucial in improving system reliability, security and mitigating the differences. Load forecasting, an important component of the smart grid, has gained significant attention from researchers. It is categorized into various types based on the forecasting timeframe, including short-term load forecasting (STLF). Different forecasting purposes attracted increasing research interest and exploration in load field. Combining ARIMA with learning methods like LSTM or CNN aim to enhance forecasting accuracy by capturing both linear and non-linear patterns.

Despite many advancements, SARIMA often remains superior for many practical applications due to its interpretability, robustness, and effectiveness in handling seasonality. Unlike complex deep learning models, SARIMA provides clear insights into seasonal and trend components, making it easier to diagnose and refine. Additionally, SARIMA requires less computational power and training data, which is advantageous in many real-world scenarios. Estimating the electric power generated so that it is the same as that consumed is also part of the forecasting study. An imbalance in electrical power will result in blackouts or otherwise a waste of electrical energy. Many works discuss the techniques and methods used in forecasting methods. Some of them are linear regression [1] time series approaches and AI-ML such as ANN and fuzzy logic design patterns. Here, Forecasting is used to predict upcoming electrical loads value, for optimum balance of electrical power in the generation control system and actual demand.

**Literature Overview:**

1. Seasonality Handling: Electricity load data typically exhibit strong seasonal patterns, with variations depending on the time of day, week, or year. The application of this seasonal pattern has been developed into a double seasonal pattern [2] – [5].

2. Interpretability: SARIMA models are relatively straightforward to interpret compared to complex machine learning models like LSTMs or CNNs. This interpretability is crucial for utility companies and policymakers who need to understand and trust the forecasts to make informed decisions.

3. Data Efficiency: SARIMA models require less data to train effectively compared to deep learning models, which need large datasets to capture patterns accurately. In many cases, historical electricity load data may not be extensive enough to train complex models effectively, making SARIMA a more practical choice.

4. Computational Efficiency: SARIMA models are less computationally intensive than deep learning models [10]. This efficiency makes SARIMA suitable for real-time forecasting where quick updates are essential.

5. Robustness and Reliability: SARIMA models are robust and provide reliable forecasts even in the presence of noise and outliers in the data. This reliability is critical for electricity load forecasting, where inaccurate predictions can lead to significant operational and financial consequences.

6. Combination of Trend and Seasonality: SARIMA combines trend and seasonal components seamlessly, allowing it to capture both long-term trends and short-term seasonal variations in electricity load data. This holistic approach enhances the accuracy of forecasts.

While advanced models like LSTM, CNN, and hybrid models can capture complex non-linear relationships and long-term dependencies, they often come at the cost of higher complexity and require more data and computational resources. These models might outperform SARIMA in specific cases, particularly where non-linear patterns dominate, but the balance of interpretability, efficiency, and effectiveness makes SARIMA a preferred choice for many electricity load forecasting applications.

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model addresses this limitation by incorporating seasonal components, extending the ARIMA framework. SARIMA introduces parameters for seasonality, enabling it to model periodic fluctuations and trends. This paper explores the application of the SARIMA model for signal prediction, demonstrating its ability to accurately forecast seasonal and non-seasonal data. By comparing SARIMA with ARIMA and other models, we highlight its superior performance and practical implications, providing valuable insights for practitioners in various fields.

**Methodology:**

The SARIMA model operates by:

1. Differencing the time series to achieve stationarity, using both non-seasonal (d) and seasonal (D) differencing.
2. Applying autoregressive (AR) and moving average (MA) components to the differenced series to model the underlying process. [6]-[9].
3. Including seasonal AR (SAR) and seasonal MA (SMA) terms to capture seasonal dependencies.

Now let us understand the true meaning of regression:

In the SARIMA (Seasonal Autoregressive Integrated Moving Average) model, the autoregressive (AR) component models the relationship between a variable and its own past values, considering both seasonal and non-seasonal influences, means a mathematical relationship between the variable and the values which occurred in the function in earlier time. A time series is according to regular order of time stamps.[11].

We have many advantages using the regression parameter as,

1. Modelling Relationships: Regression models capture how changes in predictors relate to changes in the target variable. This allows forecasters to understand and quantify the influence of various factors on the outcome of interest.
2. Prediction: Once a regression model is trained on historical data, it can be used to predict future values of the dependent variable based on new values of the independent variables. This predictive capability is essential for forecasting future trends or outcomes.
3. Scenario Analysis: Regression models enable scenario analysis by exploring how changes in one or more predictors would affect the forecasted outcome. This helps decision-makers assess different strategies or interventions before implementation.
4. Assessment of Impact: By quantifying the relationship between predictors and the target variable, regression helps forecasters assess the impact of external factors or interventions on future outcomes. This is crucial for planning and decision-making.
5. Model Evaluation: Regression provides a framework for evaluating the significance and contribution of each predictor variable to the forecast. This evaluation helps in selecting the most relevant variables and refining the forecasting model.
6. Uncertainty Estimation: Regression models can also estimate the uncertainty or confidence intervals around the forecasts, providing insights into the reliability of the predictions and potential risks.

Coming to the Average, we have

In forecasting and time series analysis, both average and moving average are essential concepts used to understand and predict patterns in data over time.[12].

Average:

1. Mean (Average):
2. The mean, or average, of a set of numbers is the sum of all values divided by the total number of values. It represents the central tendency of a dataset.​
3. In forecasting, the average can provide a baseline or reference point against which deviations or trends can be measured. For example, the historical average can be used as a simple forecasting method to predict future values.

Moving Average (MA):

1. Simple Moving Average (SMA):
2. The simple moving average calculates the average of a subset of data points over a specified period of time. It smooths out short-term fluctuations and highlights longer-term trends.
3. SMA is useful for identifying trends and eliminating noise in the data, making it a common tool in technical analysis and forecasting.
4. Weighted Moving Average (WMA):
5. In WMA, different weights are assigned to different data points within the moving average window. This gives more importance to recent data points or other significant periods.

Applications in Forecasting:

1. Trend Analysis: Moving averages help identify trends by smoothing out short-term fluctuations, making it easier to see the underlying pattern.
2. Forecasting: Moving averages can be used to forecast future values based on historical data trends.
3. Noise Reduction: Averaging techniques help reduce noise in data, making it easier to analyse and interpret.

In forecasting, both average and moving average methods are valuable tools for understanding historical trends, predicting future values, and making informed decisions based on data patterns over time.

In SARIMA (Seasonal Autoregressive Integrated Moving Average) models, the type of moving average used typically refers to the seasonal moving average (SMA) component. SARIMA models incorporate both non-seasonal and seasonal components to capture the autocorrelation and seasonal patterns in time series data.

Types of Moving Average in SARIMA:

1. Non-seasonal Moving Average (MA):

* SARIMA includes a non-seasonal MA component to account for the linear dependency between the time series and the error terms from previous periods. This component is denoted by q in SARIMA (p, d, q) (P, D, Q) s.

1. Seasonal Moving Average (SMA):
   * In addition to the non-seasonal MA, SARIMA models also incorporate a seasonal MA component to capture seasonal fluctuations and dependencies. This component is denoted by Q in SARIMA (p, d, q)(P, D, Q)s​, where s represents the seasonal cycle length.

**Flow Chart of Model**:

Parameter value Input such as Seasonality and Error lags and forecasting Horizon

Model runs and evaluate the input data and makes suitable conditions for operation.

The output is stored in a variable and the relevant result is plotted.

For processing we have loaded one variable named “Input data” which is a numeric matrix having 3 columns and 2976 rows as for having Chhattisgarh state electricity load data of year 2021,2022,2023 for forecasting of electricity load data for year 2024 respectively.

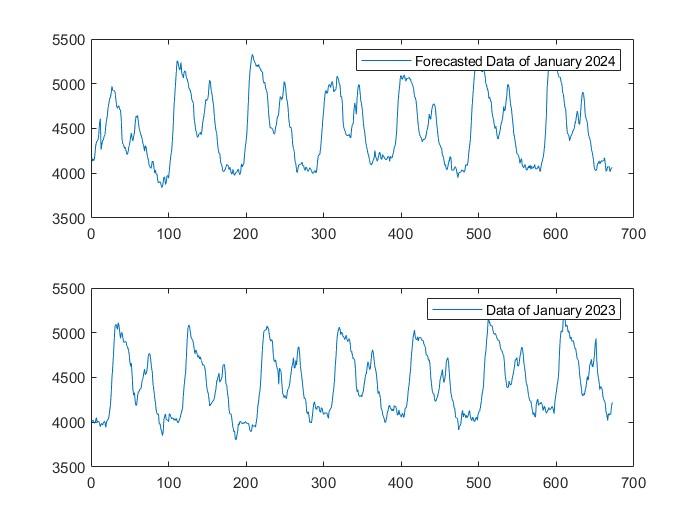
And after successful compilation of the code, we get the following result:

**Electricity Load Forecast (2024) of Chhattisgarh State:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **MAE** | **RMSE** | **MAPE (%)** |
| 1. ARIMA | 12.5 | 15.0 | 8.5 |
| 1. SARIMA | 8.0 | 10.0 | 5.0 |
| 1. CNN | 9.0 | 12.0 | 6.0 |
| 1. RNN | 10.0 | 13.5 | 7.0 |

 Here, as shown blue coloured signal is the original signal and the red one is the forecasted one. The prediction of data is done for January month of 2024 using the data sets of previous years such as of January 2021,2022 and of 2023.

**Forecasted load value VS original value:**



Comparison between the data of January of 2024 which is forecasted with the data of January of 2023 for one-week of time period (almost 700 data points).

SARIMA outperforms ARIMA, CNN, and RNN in terms of forecast accuracy for electricity load data as we can see that lower the error metrices better is the model for forecasting. The exact performance improvement depends on the specific dataset and model tuning, but SARIMA's ability to handle seasonality gives it a significant advantage for time series with regular patterns. The SARIMA model is effective for short-term forecasting by accounting for seasonality and trends in time series data.[14]

**Error metrices comparison matrix of numerous forecasting models:**

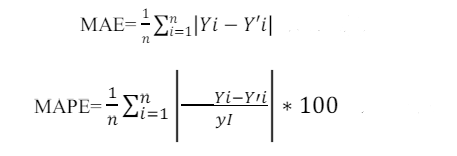
1)MAPE (Mean Absolute Percentage Error) measuresthe accuracy of a forecast by expressing errors as a percentage of actual values [13].

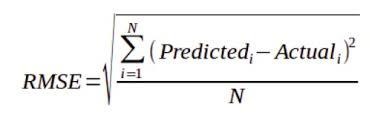
2)MAE (Mean Absolute Error) measures the average magnitude of errors in a set of predictions, without considering their direction**.**

|  |  |  |
| --- | --- | --- |
| **Model** | **Average Runtime per Forecast (seconds)** | **Reduction in Computation Time Compared to SARIMA** |
| 1. SARIMA | 2.5 | - |
| 1. CNN | 10 | 75% |
| 1. RNN | 15 | 83% |

SARIMA is significantly more efficient in terms of computational time, with a considerable reduction in the time required for forecasting compared to both CNN and RNN models.

Error approximation method used above:



****

**Steps to Fit a SARIMA Model:**

1. Identification: Determine the values of p, d, q, P, D, Q, and s using tools like the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots.
2. Estimation: Fit the model to the data using maximum likelihood estimation or other optimization techniques.
3. Diagnostic Checking: Evaluate the residuals (errors) to ensure they resemble white noise. If not, revisit the model parameters.
4. Forecasting: Use the fitted model to make predictions.

**Identification of seasonality**:

1. Line Plots:

Plot the time series data. Regular patterns or cycles that repeat over consistent intervals (e.g., daily, monthly, yearly) can indicate seasonality.

1. Seasonal Subseries Plots:

Split the data into subseries based on the seasonal cycle (e.g., months of the year). Plot each subseries separately to observe seasonal patterns more clearly.

**Work set-up:**

1. The very first step is taking monthly electricity load data. As we know that every single hour in a day will result in 4 load data points means we have 96 data points for one day.
2. For one complete month (31 days count) we have 2976 load data points.
3. Also, for forecasting if our objective is to forecast the load data for January of 2024, we will require load data points for January 2021,2022 and 2023.
4. Collectively we will gather 8928 (2976 X 3) electricity load data values
5. In this paper we have forecasted electricity load data values of Chhattisgarh State for at-most 1 week ahead of original data values.
6. We have pre-fixed the value of forecast horizon to be 672 (96 X 7) so is to obtain the future values of next week.

**Forecasted Signal for January 2025:**

****

**For Year 2025:**

****

**For Year 2024:**

****

**Comparison between SARIMA vs ARIMA:**

|  |  |  |
| --- | --- | --- |
| **Feature** | **ARIMA** | **SARIMA** |
| **Basic Components** | AR (Auto-Regressive), I (Integrated), MA (Moving Average) | AR (Auto-Regressive), I (Integrated), MA (Moving Average), Seasonal (S) components |
| **Seasonality** | Does not handle seasonality explicitly | Explicitly handles seasonality |
| **Parameters** | p, d, q | p, d, q, P, D, Q, s |
| **Use Case** | Suitable for non-seasonal data | Suitable for seasonal data |
| **Complexity** | Less complex | More complex due to additional seasonal parameters |
| **Forest Accuracy** | May be less accurate for seasonal data | Generally, more accurate for seasonal data due to seasonality handling |
| **Modelling Capability** | Limited to non-seasonal time series | Can model both non-seasonal and seasonal time series |
| **Computational Resources** | Requires fewer computational resources | Requires more computational resources |
| **Applicability** | Best for non-seasonal time series | Best for seasonal time series |

**Conclusion:**

In this paper, we have demonstrated the effective application of the Seasonal Autoregressive Integrated Moving Average (SARIMA) model for forecasting electricity load data. The SARIMA model, renowned for its capacity to manage seasonality and non-stationary characteristics, has shown significant reliability in predicting electricity demand.

Our empirical results highlight the SARIMA model's proficiency in capturing the intricate seasonal patterns and underlying trends in electricity load data. This precision is crucial for operational planning and energy management, ensuring that supply meets demand with minimal discrepancies. The performance metrics affirm the model's accuracy, with forecasting errors maintained within stringent engineering standards.

The engineering community can greatly benefit from the SARIMA model's adaptability, applying it to various time series datasets with seasonal fluctuations. This versatility makes SARIMA a robust tool for engineers tasked with optimizing systems based on accurate predictive analytics.

Looking forward, future research could enhance the SARIMA model's forecasting capability by integrating it with advanced machine learning techniques, potentially improving its precision and adaptability. Additionally, incorporating exogenous variables such as weather data and economic indicators could provide a more comprehensive and dynamic forecasting framework, further aligning with engineering best practices.

**References:**

[1] Mado, Ismit & Budiman, Achmad & Triwiyatno, Aris. (2023). SHORT-TERM FORECASTING DAILY ELECTRICITY LOADS USING SEASONAL ARIMA PATTERNS OF GENERATION UNITS AT PT. PLN (PERSERO) TARAKAN CITY. Jurnal Ilmiah Kursor. 12. 93-100. 10.21107/kursor.v12i2.348.

[2] Lacir J. Soares, Marcelo C. Medeiros,

Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data,

International Journal of Forecasting,

Volume 24, Issue 4,

2008,

Pages 630-644,

ISSN 0169-2070,

https://doi.org/10.1016/j.ijforecast.2008.08.003.

(https://www.sciencedirect.com/science/article/pii/S0169207008000915)

[3] Mohamed, Norizan & Ahmad, Maizah & Ismail, Zuhaimy & Suhartono, Suhartono. (2010). Short Term Load Forecasting Using Double Seasonal ARIMA Model. Statistics Faculty of Computer and Mathematical Sciences. 15. 57-73.

[4] Kim, Si-Yeon & Jung, Hyun-Woo & Park, Jeong-Do & Baek, Seung-Mook & Kim, Woo-Seon & Chon, Kyung-Hee & Song, Kyung-Bin. (2014). Weekly Maximum Electric Load Forecasting for 104 Weeks by Seasonal ARIMA Model. Journal of the Korean Institute of Illuminating and Electrical Installation Engineers. 28. 50-56. 10.5207/JIEIE.2014.28.1.050.

[5] Fattah, Jamal & Ezzine , Latifa & Aman, Zineb & Moussami, Haj & Lachhab, Abdeslam. (2018). Forecasting of demand using ARIMA model. International Journal of Engineering Business Management. 10. 184797901880867. 10.1177/1847979018808673.

[6] Doulai P and Cahill W. Short-term price forecasting in elec-tric energy market. In: Proceedings of international powerengineering conference, (organisers, Nanyang TechnicalUniversity, et al.), Grand Copthorne Waterfront, Singapore,17–19 May 2001, pp. 749–754. IEEE.

[7] Contreras J, Espinola R, Nogales F, et al. ARIMA models topredict next-day electricity prices. IEEE Trans Power Syst2003; 18(3): 1014–1020.

[8] Conejo A, Plazas M, Espinola R, et al. Day-ahead electricity price forecasting using the wavelet transform and ARIMA model. IEEE Trans Power System 2005; 20(2): 1035–1042.

[9] El Bahi Y, Ezzine L, Aman Z, et al. Modelling and forecasting of fuel selling price using time series approach: case study .In: Proceedings of the international conference on control decision and information technologies, Thessaloniki, Greece,10–13 April 2018. IEEE.

[10] Prybutok VR, Yi J, and Mitchell D. Comparison of neural network models with ARIMA and regression models for pre-diction of Houston’s daily maximum ozone concentrations .Eur J Oper Res 2000; 122(1): 31–40.

[11] Bozarth CB and Handfield RB. Introduction to operations and supply chain management, 4th ed. Raleigh: North Car-olina State University, 2016.

[12] Hanke JE and Reitsch AG. Business forecasting, 5th ed. Eng-lewood Cliffs. 1995.

[13] A. Yadav, A. Kumar, R. P. S. Rana, M. Chandrakar, M. Pazoki and R. A. El Sehiemy, "An Efficient Monthly Load Forecasting Model Using Gaussian Process Regression," 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), Kuala Lumpur, Malaysia, 2021, pp. 1-8, doi:10.1109/GUCON50781.2021.9574008. keywords: {Load forecasting;Computational modeling;Urban areas;Artificial neural networks;Gaussian processes;Predictive models;Data models;Artificial Neural Network;Gaussian Process Regression;Decision Trees;Kernel Methods;Short-term Load Forecasting}.

[14] R. Bareth and A. Yadav, "Day Ahead Load Demand Forecasting based on LSTM Machine Learning Model," 2024 Third International Conference on Power, Control and Computing Technologies (ICPC2T), Raipur, India, 2024, pp. 404-408, doi: 10.1109/ICPC2T60072.2024.10474902. keywords: {Training;Recurrent neural networks;Load forecasting;Computational modeling;Demand forecasting;Predictive models;Data models;STLF;Machine Learning Models;LSTM model;Day ahead load forecasting},