International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Comparative Analysis of Load Forecasting Algorithms Using Machine Learning

KARTHIKEYAN S

UG Scholar
Department of Electrical and Electronics Engineering
St. Joseph's College of Engineering
Affiliation of Anna University
Chennai, India
karthickarul254@gmail.com

***Abstract*—Load forecasting is a critical aspect of energy management, enabling utilities to optimize resource allocation and ensure grid stability and reliable power supply. This study proposes a comprehensive approach to enhancing load forecasting accuracy by combining Autoregressive Integrated Moving Average (ARIMA) with Seasonal and Exogenous factors known as Regressor Integrated Moving Average with exogenous factors (SARIMAX) which is categorized under statistical machine learning modeling techniques. The integration of exogenous variables, such as weather conditions and socio-economic factors, enhances the forecasting accuracy by capturing external influences on load demand. The ARIMA/SARIMAX hybrid model is trained on historical load data, exogenous variables, and seasonality patterns, providing a robust framework for capturing complex temporal dependencies. The study evaluates the proposed approach using real-world load datasets, comparing its performance against traditional ARIMA and machine learning models. The results demonstrate superior forecasting accuracy, illustrating the effectiveness of integrating exogenous factors into the ARIMA/SARIMAX model for load forecasting applications. This hybrid approach contributes to the advancement of reliable and precise load forecasting techniques, crucial for effective energy planning and management in modern power systems.**

Keywords—Load Forecasting, Statistical modeling techiques, Exogenous Regressors, Autoregressive Integrated Moving Average (ARIMA), Seasonal and Exogenous Regressors Integrated Moving Average (SARIMAX)

I. INTRODUCTION

Electricity holds immense importance in modern life, serving as the lifeblood of technological, social, and economic progress. From powering our homes and lighting our streets to fueling industries and driving innovations, electricity is integral to nearly every aspect of our daily existence.

Estimating the load is essential for planning energy production in order to reliably supply enough electricity to meet needs at any moment. Due to the increase in electricity demand, the management of the maintenance schedule, the selection of suitable and affordable generators, and plant planning is a difficult procedure for the electricity supply [1],[3],[8].

Electrical load forecasting, a critical component of energy management, which can be broadly classified into three distinct categories based on the forecast period: short-term, medium-term, and long-term [2],[3]. Short-term load forecasting, spanning hours to a few days, focuses on predicting immediate electricity demand fluctuations [5],[7]. Medium-term load forecasting extends the horizon to several weeks, months, or even a year, aiding in mid-range planning and resource allocation. It considers factors like seasonal variations and specific events affecting demand. Long-term load forecasting, covering periods beyond a year, is essential for strategic planning, infrastructure development, and policy formulation.

Forecasting is a systematic process of predicting future outcomes based on historical data, trends, and patterns. It involves the use of various statistical, mathematical, or computational models to analyze past observations and make informed estimates about what is likely to occur in the future. The primary objective of forecasting is to reduce uncertainty and assist in decision-making by providing insights into potential developments, trends, or events. Time series approaches, linear regression, linear autoregressive methods, exponential methods, and the stochastic time series method are among the various techniques that make up statistical methods. Autoregressive Integrated Moving Average (ARIMA) [3],[4], Autoregressive (AR) and Seasonal and Exogenous Auto Regressive Integrated Moving Average (SARIMAX) are also included.

This paper presents the short and medium-term load forecasting using ARIMA and SARIMAX Statistical models by the help of Python programming via Google Colaboratory/Jupiter Notebook platform and the objective for using this two models for an a comprehensive approach to enhancing load forecasting accuracy and considers factors like seasonal variations and specific events affecting demand.

This paper is structured as follows. In part II the models used to obtain short and medium-term load forecasting. Section III presents the step by step approaches in forecasting and results obtained from each of the methods compared in this paper and discussion of their results given in part IV. Finally, conclusions are drawn in section V.

II. Models for Forecasting

A. Autoregressive Integrated Moving Average(ARIMA)

The Autoregressive Integrated Moving Average (ARIMA) model stands as a cornerstone in time series analysis and forecasting, renowned for its effectiveness in capturing and predicting temporal patterns. Comprising three key components—autoregressive (AR), differencing (I), and moving average (MA)—ARIMA is designed to handle nonstationary time series data, transforming it into a stationary form through differencing. The autoregressive component accounts for the correlation between a variable's current value and its past values, the differencing component addresses trends and seasonality, and the moving average component captures the influence of past white noise or random error terms. The power of ARIMA lies in its adaptability to a wide array of time series data, making it particularly valuable in scenarios where underlying patterns are not immediately evident. Its parameterization involves the selection of the order of autoregressive, differencing, and moving average terms (p, d, q), demanding a nuanced understanding of the data's characteristics.

Simplified representations of the formulas used for Autoregressive Integrated Moving Average (ARIMA) model is shown below.

$$Y_t=c+\phi_1 Y_{t-1}+\phi_2 Y_{t-2}+\ldots+\phi_p Y_{t-p}+\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}+\ldots+\theta_q\varepsilon_{t-q}+\varepsilon_t$$

Where:

Y_t are the value of the time series at time t,

C *is* a constant, $\phi_1,\phi_2,\ldots,\phi_p$ are autoregressive parameters, $\theta_1,\theta_2,\ldots,\theta_q$ are moving average parameters, ε_t are white noise at time t, d is the degree of differencing.

B. The Seasonal and Exogenous Auto Regressive Integrated Moving Average with exogenous factors (SARIMAX)

The Seasonal and Exogenous Auto Regressive Integrated Moving Average with exogenous factors (SARIMAX) model represents an advanced and versatile extension of the traditional ARIMA framework, tailored to incorporate both seasonal patterns and external influences. Building upon the autoregressive integrated moving average (ARIMA) structure, SARIMAX introduces exogenous variables, enabling the model to account for factors beyond the inherent time series dynamics. This inclusion of exogenous regressors enhances the model's capability to capture the impact of external events or predictors that may influence the observed time series. The seasonal component addresses recurring patterns that exhibit regularity over fixed intervals, providing a more comprehensive representation of the data's temporal characteristics. SARIMAX, with its adaptable structure, becomes a potent tool for forecasting when external variables, such as economic indicators, weather conditions, or policy changes, play a significant role in shaping the time series behavior. The parameterization of SARIMAX involves the meticulous selection of seasonal periods, autoregressive, differencing, and moving average orders, as well as the identification of relevant exogenous variables. This model has proven valuable in fields such as economics, energy demand forecasting, and epidemiology, where both internal temporal patterns and external factors significantly influence the observed phenomena. SARIMAX's integration of exogenous variables positions it as a robust and sophisticated statistical model, allowing analysts to harness its capabilities for more nuanced and accurate predictions in scenarios characterized by complex temporal dynamics and multifaceted influencing factors.

Simplified representations of the formulas used for The Seasonal and Exogenous Regressors Integrated Moving Average with exogenous factors (SARIMAX) model is shown below.

$$Y_t=c+\phi_1 Y_{t-1}+\phi_2 Y_{t-2}+\ldots+\phi_p Y_{t-p}+\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}+\ldots+\theta_q\varepsilon_{t-q}$$
$$+\sum j=1s\beta_j X_{t,j}+\varepsilon_t$$

Where:

$X_{t,j}$ represents the exogenous variables at time t and seasonal period j, β_j are the corresponding coefficients for the exogenous variables,

Y_t is the value of the time series at time t, c is a constant, $\phi_1,\phi_2,\ldots,\phi_p$ are autoregressive parameters, $\theta_1,\theta_2,\ldots,\theta_q$ are moving average parameters, ε_t is white noise at time t, d is the degree of differencing

III. Approaches and Results

A. Data Set Loading/importing

To compile more thorough information on all the variable factors that impact a location's electrical usage, the entire dataset was gathered from various sources. The period of data collection was July 2014–July 2022. The entire data is recorded hourly.

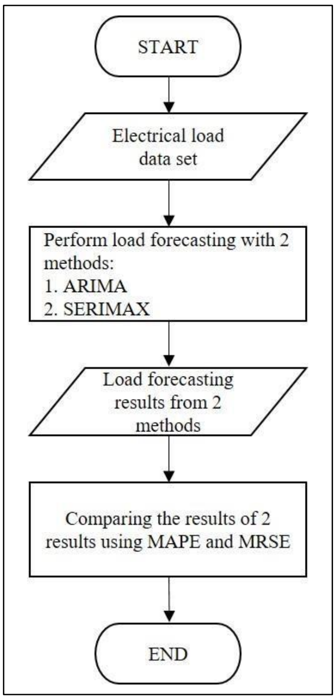
Total number of observations are 48048 rows x 16 columns. Table I provide details of dataset-

Table 1 DATASET DESCRIPTION

Parameters	Type	Description
Date ID	Numerical	-
Date	'yyyy-mm-dd'	2015-01-03 to 2020-06-27
Time	'hh:mm:ss'	01:00:00 – 23:00:00
Day of Week	Categorical	{0,1,2,3,4,5,6}
Holiday	Categorical	{0, 1}
Load	Numerical	Mega Watts

Forecasting procedure follows the steps outlined in the sequence of the flow chart, which is provided below.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930



The figure 1 shows the input dataset plot between the Date ID and Load and graph is plotted for weekly biases by sum the data of every hour to sum of a single each week in a month for clear to make easier prediction.

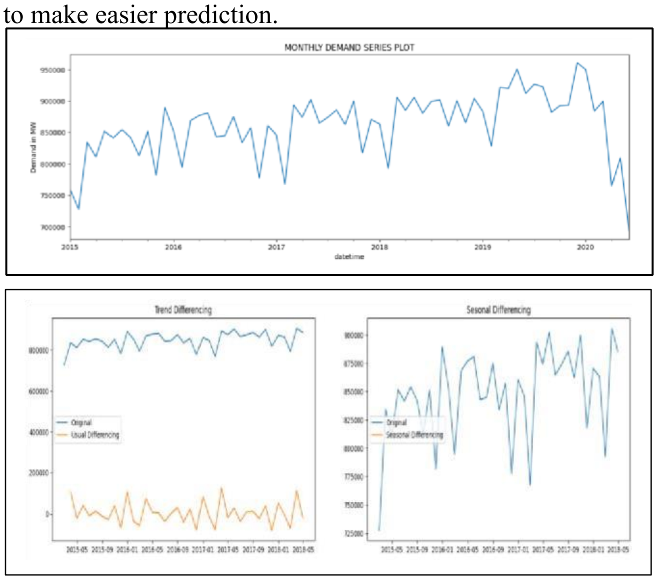


Fig.1 Input Data Sheet Month Series Plot

Checking for the correlation between the all variables in the data set via heat map. The purpose of using heat map in forecasting is to visualize the correlation matrix of variables in a dataset. This helps identify patterns and relationships between variables and its shown in fig. 2.

Making the decomposition of the data sheet to seasonal, load and trends variables for partial and auto-correlation between the decomposed variables and train/test data and the need for decomposition of data sheet where the seasonal component captures recurring patterns, be they daily, weekly, or monthly. Meanwhile, the trend component illuminates the long-term directional movement, indicating if the data is on an upward, downward, or stable trajectory. The residual component captures the random variability that remains after accounting for the seasonal and trend elements. and its given in fig.3,4.

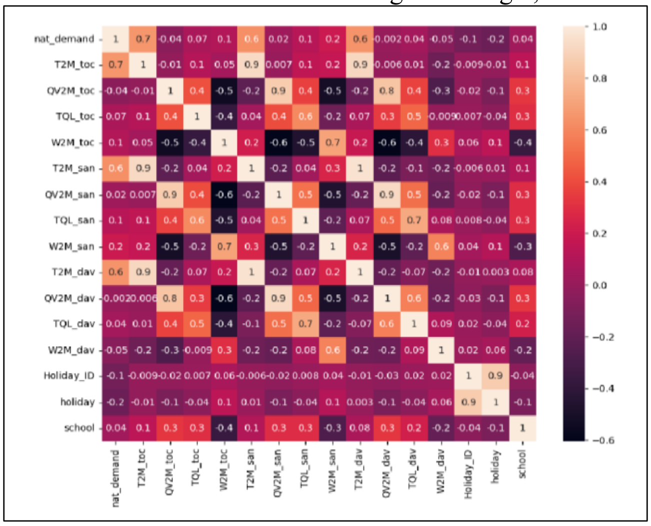


Fig.2 Heat map between all variables in Data Sheet.

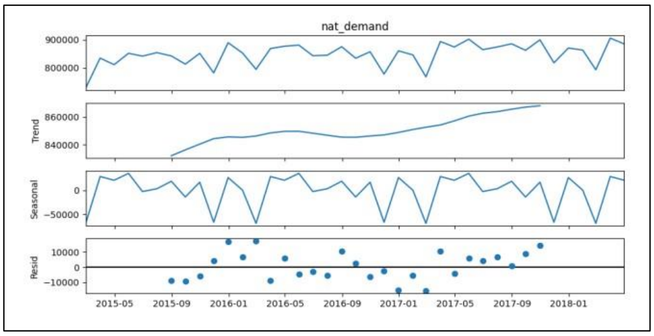


Fig.3 Decomposition of required variables and Seasonal/Trend difference of the main data set.

In ARIMA model, the Augmented Dickey-Fuller test (ADF test) is done for checking the stationarity of the data variables from data set. If the p-value obtained from the test is less than a certain threshold of 0.05, we can reject the null values and therefore the series is stationary.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

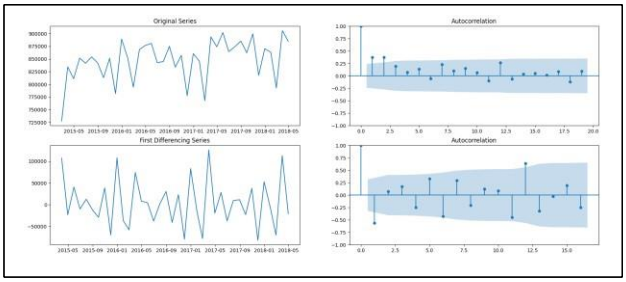


Fig.4 Auto -correlation between date -time and training data set.

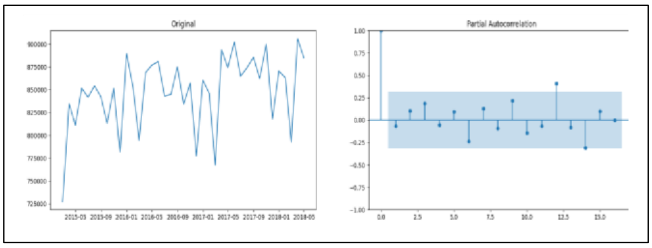


Fig.5 Partial-correlation between date-time and training data set.

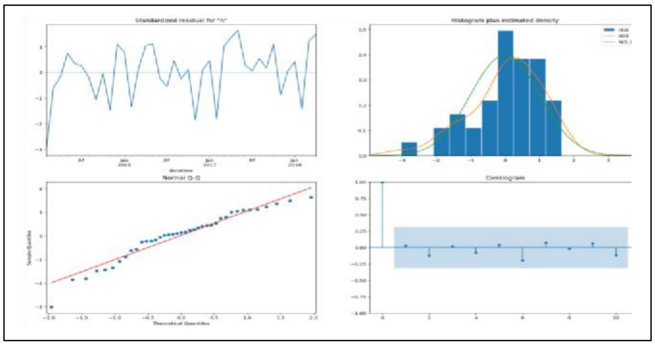


Fig.6 Diagnostics result of ARIMA model.

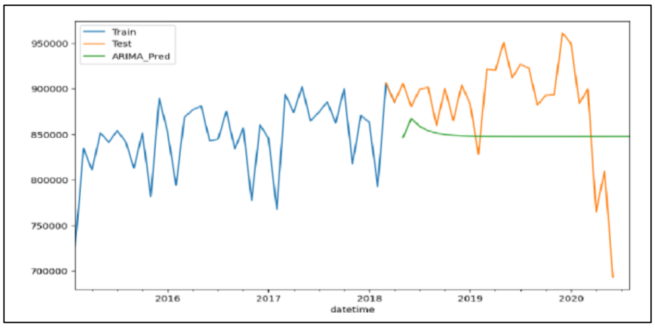


Fig.7 Output from ARIMA model without exogenous variables

After the ARIMA methodology, predicting the same data set with SARIMAX model with exogenous variables.

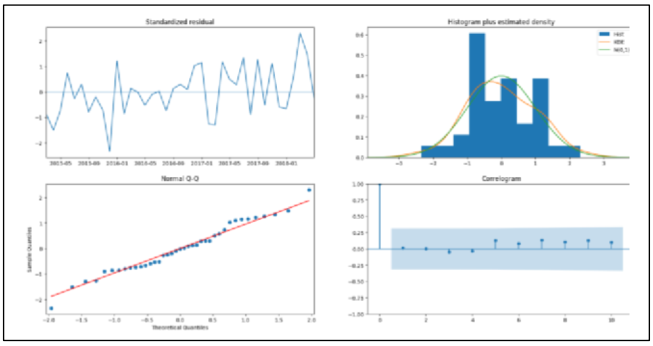


Fig.8 Diagnostics result of SARIMAX model.

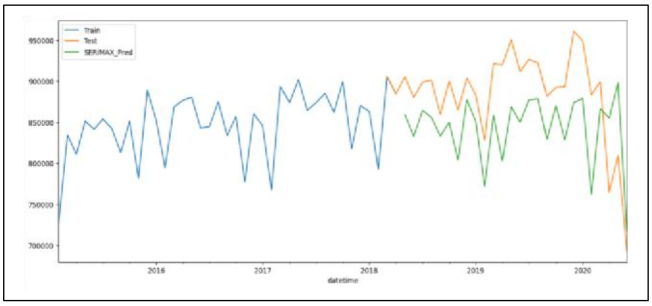


Fig.9 Output from SARIMAX without exogenous variables.

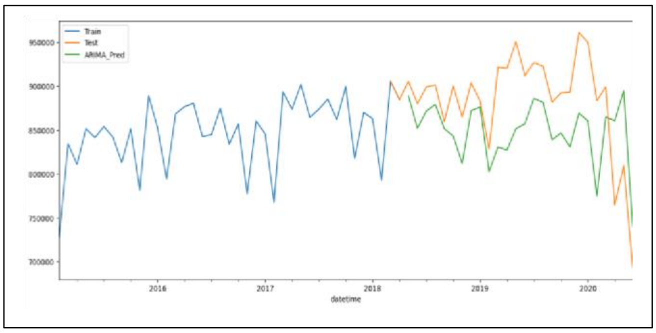


Fig.10 Final Results of the SARIMAX with exogenous variables methodology.

IV. RESULTS

In this paper, the load prediction of the dataset is done in two methodology and comparative output result s has been shown in fig. 7,9,10.

To finding the best results compare the Mean Absolute Percentage Error (MAPE) and Mean Relative Squared Error (MRSE) values. MAPE measures the percentage difference between the predicted values and the actual values and MRSE is a measure of the squared difference between the predicted values and the actual values, normalized by the squared mean of the actual values.

Model	MAPE	MRSE
ARIMA	80.3405	12745.58488
SARIMAX without exogenous variables	90.6356	8515.11278
SARIMAX with exogenous variables	92.3592	8318.69852

This paper findings reveal superior forecasting accuracy of the proposed ARIMA/SARIMAX with exogenous variable model. By integrating exogenous factors, the model excels in capturing the complexities of load demand, outperforming both traditional ARIMA and standalone machine learning models. The study underscores the effectiveness of this hybrid approach in load forecasting applications.

V. Conclusion

The ARIMA/SARIMAX hybrid model, leveraging the strengths of ARIMA and the flexibility of SARIMAX, emerges as a potent tool for load forecasting. The successful integration of exogenous factors significantly contributes to the precision of load forecasts, making this hybrid approach instrumental in advancing reliable load forecasting techniques. This advancement is pivotal for effective energy planning and management in modern power systems.

VI. References

[1]. Agus Setiawan; Zainal Arifin; Budi Sudiarto; Fauzan Hanif Jufri; Qasthalani Haramaini; Iwa Garniwa, "Comparison of Medium-Term Load Forecasting Methods (Splitted Linear Regression and Artificial Neural Networks) in Electricity Systems Located in Tropical Regions" 2022 3rd International Conference on Clean and Green Energy Engineering (CGEE)

[2]. Pande Popovski; Goran Veljanovski; Mitko Kostov; Metodija Atanasovski, "Optimizing Short Term Load Forecast: A study on Machine Learning Model Accuracy and Predictor Selection" 2022 57th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST) [2]. Refer simply to the reference number, as in

[3]. M. Abdullah Al Amin; Md. Ashraful Hoque, "Comparison of ARIMA and SVM for Short-term Load Forecasting" 2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON)

[4]. K. W. Yu; C. H. Hsu; S. M. Yang, "A Model Integrating ARIMA and ANN with Seasonal and Periodic Characteristics for Forecasting Electricity Load Dynamics in a State" 2019 IEEE 6th International Conference on Energy Smart Systems (ESS)

[5]. Rajat Sethi; Jan Kleissl, "Comparison of Short-Term Load Forecasting Techniques" 2020 IEEE Conference on Technologies for Sustainability (SusTech)

[6]. Mansi Bhatnagar; Vivek Dwivedi; Divyanshu Singh; Gregor Rozinaj, "Comprehensive Electric load forecasting using ensemble machine learning methods" 2022 29th International Conference on Systems, Signals and Image Processing (IWSSIP)

[7]. Jinjin Zhang; Tao Wang; Junyong Wu; Hainan Zhu; Dong Lan; Fengshuo Li, "Short-term Load Forecasting Method Based on Artificial Intelligence Highway Neural Network" 2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2)

[8]. Mansi Bhatnagar; Vivek Dwivedi; Divyanshu Singh; Gregor Rozinaj, "Comprehensive Electric load forecasting using ensemble machine learning methods" 2022 29th International Conference on Systems, Signals and Image Processing (IWSSIP)

[9]. Akanksha Jain; S.C. Gupta, "Peak load Forecasting using Machine Learning Algorithms" 2023 IEEE Renewable Energy and Sustainable E-Mobility Conference (RESEM)

[10]. M. Abdullah Al Amin; Md. Ashraful Hoque, "Comparison of ARIMA and SVM for Short-term Load Forecasting" 2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON)

[11]. Rizwan A Khan; C. L. Dewangan; S. C. Srivastava; S. Chakrabarti, "Short Term Load Forecasting using SVM Models" 2018 IEEE 8th Power India International Conference (PIICON)

[12]. Stefan Ungureanu; Vasile Ţopa; Andrei Cziker, "Integrating the industrial consumer into smart grid by load curve forecasting using machine learning" 2019 8th International Conference on Modern Power Systems (MPS)

[13]. Can Wang; Thomas Bäck; Holger H. Hoos; Mitra Baratchi; Steffen Limmer; Markus Olhofer, "Automated Machine Learning for Short-term Electric Load Forecasting" 2019 IEEE Symposium Series on Computational Intelligence (SSCI)

[14]. Denis Sidorov; Qing Tao; Ildar Muftahov; Aleksei Zhukov; Dmitriy Karamov; Aliona Dreglea; Fang Liu, "Energy balancing using charge/discharge storages control and load forecasts in a renewable-energy-based grids" 2019 Chinese Control Conference (CCC)

[15]. Lingxiao Wang; Shiwen Mao; Bogdan Wilamowski, "Short-Term Load Forecasting with LSTM Based Ensemble Learning" 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)

[16]. Ihsan A. S. Abu Amra; Ashraf Y. A. Maghari, "Forecasting Groundwater Production and Rain Amounts Using ARIMA-Hybrid ARIMA: Case Study of Deir El-Balah City in GAZA" 2018 International Conference on Promising Electronic Technologies (ICPET)

[17]. Stylianos I. Vagropoulos; G. I. Chouliaras; E. G. Kardakos; C. K. Simoglou; A. G. Bakirtzis, "Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting" 2016 IEEE International Energy Conference (ENERGYCON)

[18]. Akshita Gupta; Arun Kumar, "Mid Term Daily Load Forecasting using ARIMA, Wavelet-ARIMA and Machine Learning" 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)