A Methodological Review on Time Series Forecasting by using ARIMA

B. Dilip Kumar Reddy[1*], J. Swami Naik[2], S. Vinay Kumar[3], Suresh Kumar[3], G. Haritha[4] and M. Raghavendra Reddy[1]

[1]Assistant Professor, Department of CSE, G. Pulla Reddy Engineering College, Kurnool, India.

[2]Associate Professor, Department of CSE, G. Pulla Reddy Engineering College, Kurnool, India.

[3]Assistant Professor Department of ECS, G. Pulla Reddy Engineering College, Kurnool, India.

[4]Assistant Professor, Department of EEE, G. Pulla Reddy Engineering College, Kurnool, India.

dilipkumarreddy.cse@gprec.ac.in

Abstract. Time series forecasting using ARIMA (AutoRegressive Integrated Moving Average) is well-liked and tested machine learning technique for projecting upcoming values depending on historical data or previous observations. Although ARIMA is a strong and popular technique for time series forecasting, because of its capacity to identify and predict temporal patterns in data, ARIMA models have a broad variety of applications in numerous areas. Energy forecasting, healthcare, supply chain management marketing, customer behaviour analysis, and environmental sciences are a few industries that frequently use ARIMA. These are a handful of the various uses for ARIMA models. In reality, ARIMA is a flexible technique, time series analysis are tailored to a variety of domains and particular forecasting applications. Ultimately, it does have some significant limitations. These drawbacks, which include sensitivity to parameters, seasonality limits, the inability to capture non-linear correlations, and data requirements, make it challenging for companies to employ ARIMA for forecasting. In light of these limitations, my survey suggests using the R language. Modeltime package in conjunction with ensemble learning models (Random Forest, XGBoost) FB Prophet, LSTM for forecasting. When these three models are taken into account for forecasting, the best forecasting values will be determined by computing each model's MSE (Mean Squared Error) value, which will be used for production further.

Keywords: Time Series Forecasting, ARIMA, Ensemble Learning, Random Forest, XGBoost, FBProphet, Modeltime, LSTM

B. S. Babu et al. (eds.), *Proceedings of International Conference on Advanced Materials, Manufacturing and Sustainable Development (ICAMMSD-2024)*, Advances in Engineering Research 257,
https://doi.org/10.2991/978-94-6463-662-8_55

1 Introduction

Over past decades, time series have attracted large communities due to their high prominent in vast applications, like the finance domain, health industry, energy sector, meteorology, engineering sections [1] actual forecast is important in every sectors of governments and enterprises to upskill quality and improve the decision taking strategy [3] actually time series forecasting is crucial for many temporal data challenges. Ensemble modeling was good approach for using numerous predictive methods to boost accuracy, robustness, and performance[2].This research presents a novel way for developing robust ensemble models. Statistical and Machine Learning (ML) models are highly important in time series forecasting. Traditionally employed in the literature are statistical linear methods such as Moving Average (MA), Auto-Regressive (AR), and Auto-Regressive Integrated Moving Average (ARIMA).Time series methods, like the ARMA model, have been the standard model for analysing linear data properties such as autocorrelation in financial time series[4]. And ARIMA (p, d, q) is the representation of the Arima mode. Seasonality is indicated by the notation ARIMA(p,d,q) (PDQ).m. The variables d, q, and m represent the number of non-seasonality differences, moving average times, and periods in each season, respectively, and p is also an auto-regressive number, also referred to as the lag pattern. ARIMA model changes to either AR, I, or MA when any two of the (p,d,q) variables are zero. For instance, ARIMA equals AR(p) it is denoted like (p,0,0). Selecting the appropriate forecasting model is crucial when predicting using the Arima model. This implies that it is necessary to calculate and determine the value of pdq. The model is trained using the fit() method. When computing, it's The technique which has been proposed and popular in recent years and also differed from the traditional image steganography known as coverless image steganography [8]. In this technique the confidential information can be mapped in to the carrier itself based on its characteristics. The communication can be possible secretly by transmitting stego images without alterations.important to keep the knowledge, time series be stationary. One supervised learning technique that aids in data analysis for the regression process and classification process is the SVM. Additionally, it was first presented in 1992 by Vapnik and Boser [7]. SVM has been widely employed in the fields of biological research and health care, and it performed well in text and picture classification [8]. Predicting the outcomes of economic and econometric research has advanced recently. SVM model was laid on linear regression, as was previously mentioned. Three types of analytic kernels—the RBF kernel, linear kernel and poly kernel were introduced here. With b as the bias, $\boldsymbol{\pi}(x)$ as the mapping function, and w as the weight vector, SVM may be expressed mathematically as:

$$f(x) = w\ T\varphi(x) + b\ . \quad (1)$$

Data-driven methods and machine learning are becoming more and more applicable in many different domains. I presented the tree-boosting approach, a popular and very successful ML technique, in my paper. Data scientists may use XGBoost, a scalable top-to-bottom tree-boosting method in this work. Commonly use to get results on a range of machine learning problems.[6]. XGBoost, the best ML framework for boosting, is presented in this paper. An open-source bundle of the system is available. The system's impact on several ML and mining issues is widely recognised.

2 Literature Review

Several approaches have been used in time series prediction to determine the most effective forecasting technique, particularly in econometrics and business.[5] Forecasting is a statistical strategy for predicting the values of a time-dependent variable based on previous data. It is made up of observations that are gathered, noted, or monitored over time; the order in which the observations are made is also significant. This approach has multiple possibilities., including signal processing, meteorology, finance, and economics.

Time Series Data: A time series is an observational sequence gathered at regular periods across time. Each observation is associated with a particular timestamp or timeframe. Monthly sales figures, hourly temperature readings, daily stock prices, and annual GDP growth rates here are a few instances.

Components of Times Series: Time series data typically exhibit a few components:

Trend: The data's directionality, or long-term movement, showing whether it is steadily rising, falling, or staying the same over time.

Seasonality: Repeating cycles or patterns that happen regularly, including weekly, monthly, yearly, or weekly cycles.

Noise: Random irregularities or fluctuations in the data that cannot be accounted for by seasonality or trend

Objective of Forecasting: The primary objective of forecasting is to generate accurate predictions of future values based upon previous infomation. These estimates can help with decision-making, resource allocation, risk management, and strategic planning.

Methods in time series forecasting

- The following are a few traditional time series forecasting techniques:
- Autoregression (AR)
- Moving Average (MA)
- Autoregressive Moving Average (ARMA)
- Autoregressive Integrated Moving average (ARIMA)
- Seasonal Autoregressive Integrated Moving Average (SARIMA)
- Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors (SARIMAX)

2.1 ARIMA

One kind of stochastic process model that uses integration (I) to get rid of non-stationarity is the ARIMA model. [9] They expand on ARMA models as AR, functions like a regression of lag time series, and as MA, functions as an error as the previous error terms, a linear combination of the two components of ARMA models of time series the ARMA typically mentioned as **ARMA (p, q)**.Where, number of lag terms in auto-regressive section, or p, represents the order of the AR portion. The number of lagged terms in the moving average component, or q, is the order of the MA part. Seasonal autocorrelations and trends can be correctly accounted for using the SARIMA model, which incorporates seasonal features of time series [10]. SARIMA (p, d, q) (P, D, Q) s or ARIMA (p, d, q)(P, D, Q)s are two ways to represent the SARIMA model. Seasonal autoregression, seasonal differencing, seasonal moving average, and seasonal cycle are represented by the parameters P, D, Q, and s, respectively [10]. A frequently used time series method is the ARIMA method, which is highly regarded because of its great dependability and mathematical correctness. There were numerous uses for the ARIMA method. The model is used to predict a number of things, including the price of commodities [11]. The environment and ARIMA are closely related. Since these factors directly affect health, forecasting is essential in a number of environmental domains, such as air and noise pollution, fossil fuels, rainfall, and groundwater. In many different industries, forecasting may give forewarning, which can be very helpful. Comprehensive evaluations of the ARIMA's development phases are scarce, despite the fact that they were numerous articles laid on the ARIMA on a variety of sectors, including the environment. Three phases of ARIMA's development are presented in this paper. Highlighting thorough information on the process is crucial. It is equally vital to comprehend evolution and its application.

Table 1. Time Series forecasting articles and sources

Title	***Author***	***Methods***	***Dataset***	***Findings***
Profit prediction based on ARIMA and SARIMA and LSTM	Meena Sirisha et al.	ARIMA model SARIMA method LSTM neural network model	A few features are order, gross, are in thousands A count of 548 lines in dataset	LSTM model performs ARIMA and SARIMA in profit prediction accuracy. Forecasts for the next 5 years were successfully generated.
ARIMA model, in forecasting Solar Radiation at Different Locations	chodakowska et al.	ARIMA model SARIMA model	Publicly available datasets from JRC solar radiation database. Data accessed on 1 January 2023.	ARIMA models effectively forecast solar radiation in various climates. Location-specific models are crucial due

				to solar radiation variability.
Forecasting using a hybrid model, laid on ARIMA	Yihuai Huang et al.	ARIMA model Self-adaptive filtering method	Time Series Data Library (TSDL) test sets. Real cases from maternal health center and child care center	Hybrid improves prediction accuracy by 80–99% over ARIMA. MAPEreduced to 2.79% and 1.25% in actual cases.
Time series, regression models for single variate environmental forecasting: An empirical evaluation	Jamie Hannaford	Time series methods for environmental forecasting	Analysed on forecasting methods conducted. Comparison of time series and regression techniques.	---
Comparing accuracy of LSTM and ARIMA models for time series, with permanent fluctuation	Ghahreman Abdoli et al.	LSTM,ARIMA	Tehran Stock Exchange intraday data used. Data spans over 10 years.	LSTM outperforms ARIMA in forecast accuracy significantly. Long-term prediction accuracy decreases for both models.
Online ARIMA algorithms for time series prediction	Chenghao Liu et al.	Reformulation of ARIMA into task of full information online optimization	All datasets are available on the authors' webpage. Specific datasets used are not mentioned.	Proposed online ARIMA method outperform existing online ARMA model. Regret bounds ensure performance matches best ARIMA model in hindsight.

ARIMA-Time Series Model of Stochastic Wind Power Generation	Peiyuan Chen et al.	Stochastic wind power based on ARIMA process Comparison with first-order transition matrix based discrete Markov model	Wind power measurement from Nysted offshore wind farm. Data collected over one year duration	The ARIMA model outperforms Markov model in correlation and distribution. Monthly variation in wind-based power generation also in model
Online Learning for Time Series Prediction	Oren Anava et al.	ARMA for time series prediction. Online Newton Step (ONS) algorithm for regret minimization.	Experiments demonstrate prediction effectiveness of proposed algorithms. Performance Compared to ARMA-RLS and Yule-Walker estimation methods.	Developed online learning algorithms for ARMA model predictions. Relaxed strict assumptions on noise terms for broader applicability
ARIMA-Time Series Model of Stochastic Wind Power Generation	Peiyuan Chen et al.	Stochastic wind power based on ARIMA process Comparison with first-order transition matrix based discrete Markov model	Wind power measurement from Nysted offshore wind farm. Data collected over one year duration	The ARIMA model outperforms Markov model in correlation and distribution. Monthly variation in wind-based power generation also in model
Use of series analysis, in infectious disease.	R Allard et al.	ARIMA modelling for time series analysis INAR modelling and Mar-	---	ARIMA modelling aids in infectious disease surveillance. Provides estimates of future

		kov chain analysis (less practical)		observation variability.
Time-Series Modelling and Prediction of Global Monthly Absolute Temperature for Environmental Decision Making	Liming Ye et al.	Deterministic-stochastic combined (DSC) approach for modelling. SARIMA models for stochastic processes.	monthly combined data on land, air, and sea surface temperatures worldwide. Data spans January 1880 to December 2011.	Developed a DSC modelling framework for absolute temperatures. Predicted global temperature rise at twice past average rate.
Analysis of ML, AutoML Solutions for forecast, time-series Data	Ahmad Alsharef et al.	Conventional methods for linear modelling Frameworks for AutoML, such as DL models	--	Recall of time series forecasting methods and techniques. Identifies gaps in previous forecasting research studies
Forecasting Methods by using Time Series Data: A Survey	Zhenyu Liu et al.	Method for predicting fuzzy time series Building a fuzzy time series model	Time series data from sensors and equipment. Data often contains noise and missing vales	Classifies various existing time series methods. Identifies challenges and potential research directions in forecasting

Table 1 lists time series forecasting methods that have been examined and their parameters assessed from a variety of sources, including IEEE, Springer, and Scopus.

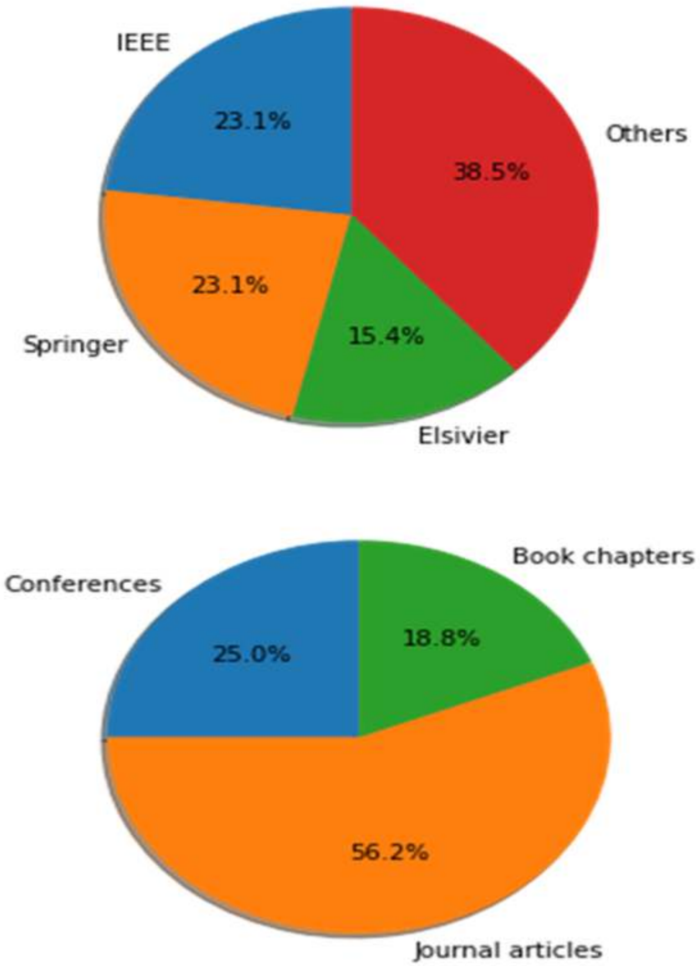

Fig. 1. (a) Clustering by percentage in a small number of databases over the keyword "Time Series Forecasting by ARIMA" and 1(b) clustering by percentage in a small number of databases over the keyword "Time Series Forecasting."

2.2 Bibliometric analysis:

My study examines a number of publications over the past ten years that have been examined utilising state-of-the-art models and databases such as Scopus, Springer, Science Press, Oxford Academic Press, Copernicus, AAS, IEEE, SCI, MDPI, and ASCE library.With the help of keywords we extracted the data from databases are *"Time Series Forecasting by ARIMA"* and *"Time Series Forecasting"* Analysis of each database revealed the following: 60 % in Articles ,13 % in Book Chapters , 8% in Conference Papers , 2% in Short communication , 2% in Editorial, 2% in Abstract , 2% Mini review , 4% in Case report , and 4% in News . Because of this, it is clear that every element in database,was retrieved a certain weight, and Science Direct has the most articles.

3 Methodologies

As the part of literature survey and domain knowledge of Time Series forecasting most of the applications and industries are using ARIMA for forecasting because of a traditional method. But when the data was non-linear the ARIMA model was not working up to mark.

To make better use of forecasting, my survey pointed out to use of ensemble methods with Modeltime and calibrating the accuracy by using different measures like MAE, RSQ, RMSE

3.1 Modeltime:

The Modeltime is a library in R and a framework for Time Series forecasting and Machine Learning modeling. It provides a streamlined interface for fitting, evaluating, and deploying time series models. Modeltime builds upon the tidyverse ecosystem and integrates seamlessly with other popular packages like tidyverse, tidymodels, and time tk. It provides a unified interface for working with various Time Series Models, Making it easier to switch between different algorithms without interpreting code.

It offers a variety of Pre-processing tools to handle tasks such as Missing Values, Feature Engineering, and supports a wide range of models including classical models like ARIMA, Exponential Smoothing, and Prophet as well as Modern It enables ensemble learning techniques to combine predictions from multiple models, potentially improving forecasting accuracy, and includes tools for Visualizing time series data it enhancing easier to understanding and communicating model results.

It Provides functionality for deploying models to a production environment, allowing for real-time forecasting or integration with other systems.

Overall, Modeltime simplifies the process of forecasting with a high accuracy rate, making a valuable tool for Data Scientists and Analysts

3.2 Model time workflow:

The Modeltime workflow demonstrates six easy actions to take:

Gather information and divide it into test and training sets. Construct and Fit a Variety of Models. Fitted models are added to a Model Table, and the models are calibrated to a testing set. Execute Forecast Testing and Accuracy Assessment

4 Conclusion

In this paper, we have covered time series forecasting, briefly described the methodologies employed, compared various approaches to determine which could be superior, and spoke about the benefits and drawbacks of time series forecasting. Depending on the data and the use case, ARIMA and SARIMA produced different findings when compared to determine which was better between AR and MA. Therefore, it is reasonable to state that each model is unique to its use case and the data used to develop the model Seasons, trends, and other natural phenomena are all taken into account in time series forecasting.

References

1. Dimitris Bertsimas Leonard Boussioux,“Ensemble modeling for time series forecasting : An adaptive robust optimization approach”.
2. Domingos S. de O. Santos Junior a,Paulo S.G. de Mattos Neto a, Joao F.L. de Oliveira b, George D.C. Cavalcantia,“A hybrid system based on ensemble learning to model residuals for time series forecasting”.
3. J.F.L. de Oliveira, E.G. Silva, P.S.G. de Mattos Neto, “A hybrid system based on dynamic selection for time series forecasting”, IEEE Trans. Neural Netw. Learn. Syst. (2021) 1–13.
4. Kaijian He 1,Qian Yang 1,Lei Ji 2,Jingcheng Pan 3 and Yingchao Zou 1,“Financial Time Series Forecasting with the Deep Learning Ensemble Model” .5.D Makala, Z Li “Prediction of gold price with ARIMA and SVM”
5. XGBoost:A Scalable Tree Boosting System Tianqi Chen University of Washington tqchen@cs.washington.edu Carlos Guest.
6. Vapnik V N 2000 The nature of statistical learning theory (Springer).
7. Huang S, Nianguang C A I, Penzuti Pacheco P, Narandes S, Wang Y and Wayne X U ,” Applications of support vector machine (SVM) learning in cancer genomics Cancer Genomics and Proteomics”2018.
8. Ewa Chodakowska 1 , Joanicjusz Nazarko 1 , Łukasz Nazarko 1, Hesham S. Rabayah 2 , Raed M. Abendeh 2 and Rami Alawneh 2,”ARIMA Models in Solar Radiation Forecasting in Different Geographic Locations”.
9. Lu Bai1,2,7, Ke Lu3,7, Yongfei Dong1,2, XichaoWang1,2, YaqinGong4 , Yunyu Xia5 , XiaochunWang5 , Lin Chen6 , ShanjunYan6 , ZaixiangTang1,2,8 & Chong Li3,”Predicting monthly hospital outpatient visits based on meteorological environmental factors using the ARIMA model
10. Jatinder Kaur1,2.Kulwinder Singh Parmar 2,Sarbjit Singh,”Autoregressive models in environmental forecasting time series: a theoretical and application review”.
11. M. Khashei and M. Bijari, “An artificial neural network (p, d, q) model for timeseries forecasting,” Expert Systems with Applications”
12. J. Sterba and K. Hilovska, “The implementation of hybrid ARIMA neural network prediction model for aggregate water consumption prediction,” Aplimat—Journal of Applied Mathematics.
13. Y. Chen, B. Yang, J. Dong and A. Abraham, "Time-series forecasting using flexible neural tree model," Inf.Sci., vol. 174, October 2004, pp. 219-235
14. Meyler, G. Kenny, and T. Quinn, “Forecasting Irish Inflation Using ARIMA Models,” Technical Paper, Central Bank of Ireland Research Department, 1998
15. .Luo, L. et al. “Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models” BMC Health Serv.
16. Nelson, B. K. Statistical methodology: V. Time series analysis using autoregressive integrated moving average (ARIMA) models.Acad. Emerg. Med.
17. Wu, X. et al. Time series analysis of air pollution and asthma outpatient visits in Hefei. J. Public Health Prev. Med
18. Hossain, M.M., Abdulla, F., 2016. “Forecasting potato production in Bangladesh by ARIMA model”
19. Wen, Q., Wang, Y., Zhang, H., et al., 2019. Application of ARIMA and SVM mixed model in agricultural management under the background of intellectual agriculture. Cluster Comput. 22 (Suppl 6), 14349–14358. http://dx.doi.org/10.1007/s10586
20. Kawakita, S. et al. Time-series analysis of population dynamics of the common cutworm, Spodoptera litura (Lepidoptera: Noctuidae), using an ARIMAX model. Pest. Manag. Sci.
21. Liu, L. et al. Predicting the incidence of hand, foot and mouth disease in Sichuan

province, China using the ARIMA model.
22. Pimentel, K. et al. Prediction of visceral leishmaniasis incidence using the Seasonal Autoregressive Integrated Moving Average model (SARIMA) in the state of Maranhão, Brazil. Braz.
23. Nelson, B. K. Statistical methodology: V. Time series analysis using autoregressive integrated moving average (ARIMA) models.Acad. Emerg. Med. 5(7), 739–744 (1998).
24. Luo, L. et al. Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. BMC Health Serv. Res. 17(1), 469 (2017)
25. Atique, S.; Noureen, S.; Roy, V.; Subburaj, V.; Bayne, S.; Macfie, J. Forecasting of Total Daily Solar Energy Generation Using ARIMA: A Case Study. In Proceedings of the 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 7–9 January 2019; pp. 0114–0119.
26. Shadab, A.; Ahmad, S.; Said, S. Spatial Forecasting of Solar Radiation Using ARIMA Model. Remote Sens. Appl. Soc. Environ. 2020,20, 100427
27. Vagropoulos, S.I.; Chouliaras, G.I.; Kardakos, E.G.; Simoglou, C.K.; Bakirtzis, A.G. Comparison of SARIMAX, SARIMA,Modified SARIMA and ANN-Based Models for Short-Term PV Generation Forecasting. In Proceedings of the 2016 IEEE International Energy Conference (ENERGYCON), Leuven, Belgium, 4–8 April 2016;pp. 1–6
28. Atique, S.; Noureen, S.; Roy, V.; Bayne, S.; Macfie, J. Time Series Forecasting of To tal Daily Solar Energy Generation: A Comparative Analysis between ARIMA and Ma chine Learning Techniques. In Proceedings of the 2020 IEEE Green Technologies Conference (GreenTech), Oklahoma City, OK, USA, 1 April 2020; pp. 175–180
29. C. Chatfield, Time-series forecasting. Chapman and Hall/CRC, 2000.
30. R. H. Shumway, D. S. Stoffer, and D. S. Stoffer, Time series analysis and its applica tions. Springer, 2000, vol. 3.
31. Takahashi, H. Aida, and T. Saito, “Arima model’s superiority over f-arima model,” in WCC 2000-ICCT 2000. 2000 International Conference on Communication Technolo gy Proceedings (Cat. No. 00EX420), vol. 1.IEEE, 2000, pp. 66–69