**A Comparative Study of Time Series Forecasting Models for Future Value Prediction**

Sreeram B. 1

Student, Department of Computer Science and Engineering, Vellore Institute of Technology Chennai, Tamil Nadu, India

Dr.R.Sujithra @ Kanmani2 Assistant Professor, Department of Computer Science and Engineering, Vellore Institute of Technology,

Chennai, Tamil Nadu, India

Dr. E.Elakiya3

Assistant Professor, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India

Mail: [elakiya.e@vit.ac.in](mailto:elakiya.e@vit.ac.in)

**Abstract**  
Time series forecasting, a fundamental branch of predictive analytics, entails the analysis and prediction of data points collected over time. It underpins decision-making across diverse domains, including finance, economics, weather, and sales. The core objective is to unveil underlying patterns, trends, and seasonality in historical data, enabling the development of robust predictive models.

This field employs a range of methodologies, from traditional statistical techniques to advanced machine learning algorithms. Researchers and practitioners select models that best capture the unique characteristics of the time series data, such as autoregressive integrated moving average (ARIMA), and recurrent neural networks (RNNs).

Time series forecasting facilitates resource allocation, risk assessment, and strategic planning. Financial analysts use it to predict stock prices, businesses rely on it for inventory management, and meteorologists harness it for weather predictions. Accurate forecasts empower stakeholders to optimize operations and make informed decisions. They help mitigate the impact of uncertainties and market fluctuations.

**1 Introduction**

Time series forecasting is a pivotal area of research and practical application in data science, playing a fundamental role in making informed decisions in various fields. In this research project, we delve into the realms of time series forecasting, with a specific focus on comparing three distinct methodologies: ARIMA (Auto-Regressive Integrated Moving Average), SARIMAX (Seasonal Auto-Regressive Integrated Moving Average with Exogenous Factors), LSTM from RNN (Recurrent Neural Networks) and Exponential Smoothing.

The significance of time series forecasting cannot be overstated, given its vital implications in finance, economics, environmental sciences, and beyond. Accurate predictions enable organizations to optimize resource allocation, anticipate market trends, and mitigate risks, which are essential for strategic planning and competitive advantage.

Our project aims to use these models to predict the future values of the dataset taken and to find which method is the most effective. ARIMA, a classical statistical method, offers a solid baseline for comparison. SARIMAX extends this by incorporating seasonality, making it highly suitable for data with repeating patterns. LSTM, a cutting-edge machine learning approach, harnesses the power of recurrent neural networks to capture complex dependencies in sequential data.

By undertaking this comparative analysis, we seek to uncover the strengths and weaknesses of each methodology and provide valuable insights into their effectiveness across different types of time series data, thereby contributing to the ongoing evolution of time series forecasting methodologies.

**2 Literature Survey**  
A literature survey on time series forecasting using ARIMA, SARIMA, exponential smoothing, and LSTM provides an overview of the existing research and developments in this field.

The seminal work done by Box and Jenkins serves as a cornerstone in time series analysis, particularly for ARIMA modeling [1]. Published in 1970, it introduces fundamental concepts such as differencing, autoregressive, and moving average components. The book remains a classic reference for researchers and practitioners in the field of time series forecasting and control.

Hyndman and Athanasopoulos provide a comprehensive and influential online textbook that delves into forecasting principles and practices [2]. Notably, the text covers seasonal decomposition of time series and introduces SARIMA modeling. This resource is widely utilized in both academia and industry, serving as a valuable guide for those seeking a practical understanding of forecasting methodologies.

Gardner's work in 1985 offers an extensive review of exponential smoothing methods, providing insights into various formulations and their applications [3]. This comprehensive overview contributes to the understanding of the state-of-the-art in exponential smoothing, making it a valuable resource for researchers and practitioners exploring forecasting techniques.

The seminal paper by Hochreiter and Schmidhuber published in 1997 introduces Long Short-Term Memory (LSTM) networks, a breakthrough in recurrent neural networks (RNNs) [4]. Overcoming vanishing gradient problems, LSTM finds application in sequence prediction. Widely cited and influential, this paper is pivotal for those interested in advanced neural network architectures for time series analysis.

Focused on the medical domain, Lipton et al.'s study from 2015 showcases the effectiveness of LSTM-RNNs in diagnosing medical conditions [5]. The paper demonstrates the potential of LSTM networks for time series forecasting applications beyond traditional domains, emphasizing the adaptability of these models to diverse datasets.

In a 2018 paper, Medeiros and Mendes explore the synergy between ARIMA models and neural networks for modeling and forecasting short time series [6]. The research contributes to the evolving landscape of hybrid forecasting approaches, providing insights into the complementary strengths of traditional time series methods and modern machine learning techniques.

Harvey's paper is a landmark in the application of structural time series models and the Kalman filter [7]. It presents a flexible framework for modeling time series data by decomposing it into unobserved components such as trend, seasonality, and irregularities. The Kalman filter plays a pivotal role in estimating these components, making the methodology powerful for a wide range of forecasting applications.

Huang et al.'s paper [8] presents Empirical Mode Decomposition (EMD), a novel method for decomposing time series into intrinsic oscillatory modes called Intrinsic Mode Functions (IMFs). EMD is particularly valuable for analyzing nonlinear and non-stationary time series data, where traditional linear methods may fall short. The paper introduces the Hilbert Spectrum as a tool for analyzing the instantaneous frequency and energy distribution of these IMFs, enhancing our understanding of complex time series patterns.

Taylor and Letham's paper [9] introduces Prophet, a forecasting model developed by Facebook for handling time series data with daily observations and multiple seasonal patterns. Prophet is designed to capture various components such as trend, seasonality, and holidays, making it robust for business-related time series forecasting. The paper provides insights into the model's architecture, highlighting its ease of use and scalability for large-scale forecasting applications.

Friedman's tutorial is a comprehensive guide to Gradient Boosting Machines (GBM) [10], including the well-known XGBoost algorithm. While not specifically tailored for time series, the paper discusses the principles behind boosting algorithms, ensemble learning, and the regularization techniques employed by XGBoost. This tutorial has been instrumental in popularizing the use of XGBoost for various machine learning tasks, including time series forecasting.

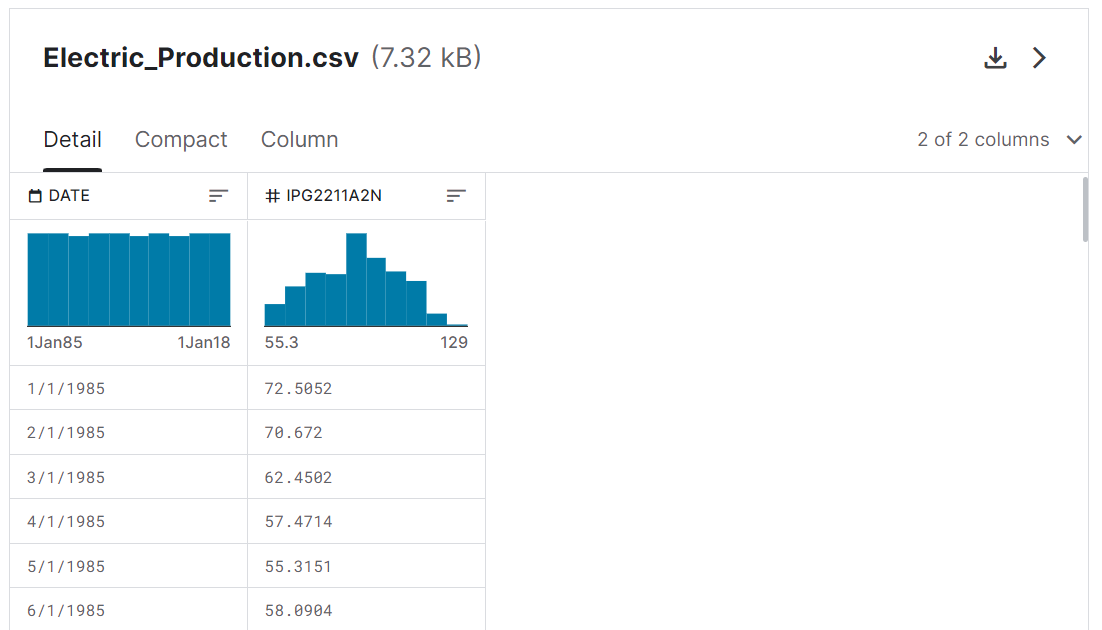
Bengio et al.'s paper [11] introduces a Neural Probabilistic Language Model, emphasizing the use of neural networks for probabilistic language modeling. While not directly focused on time series, the concept of probabilistic modeling is crucial in capturing uncertainty in time series forecasting. This paper contributes to the understanding of generating probabilistic forecasts, a valuable aspect when dealing with uncertain and dynamic time series data

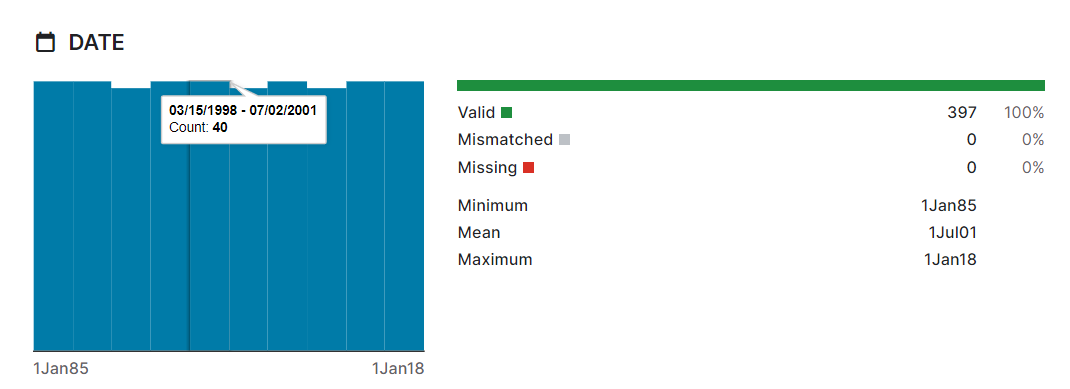
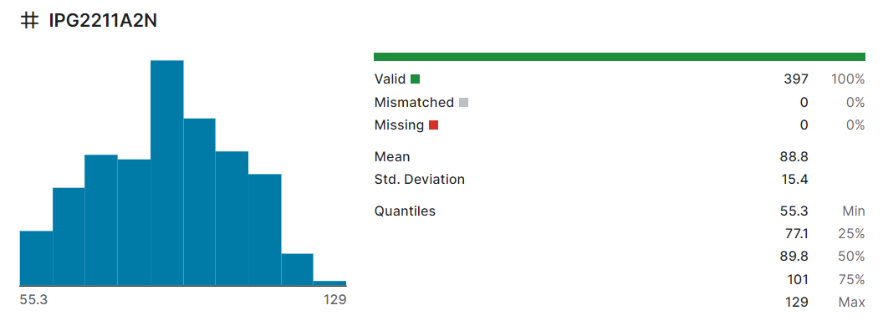
**3 Methodology**

In this research project, we employ four distinct machine learning models to predict future values within a given time series dataset. Our primary objective is to identify the most suitable model for this specific forecasting task based on the Root Mean Square Error (RMSE) metric and the Mean Absolute Error (MAE) metric. This analysis allows us to determine which model consistently provides the most accurate predictions, thus offering a data-driven selection of the optimal forecasting approach.

**3.1 Data Set Description**

The dataset was taken from Kaggle and consisted of 16 years of electricity production data taken on a monthly basis.





**4 Models**

Four machine learning models were used in this project:

1. **ARIMA (AutoRegressive Integrated Moving Average):**

The ARIMA (AutoRegressive Integrated Moving Average)[12] model is a powerful time series forecasting technique that combines three key components: AutoRegressive (AR), Integration (I), and Moving Average (MA). The AR component quantifies the influence of past data points on the present, while the I component handles differencing for data stationarity, and the MA component models short-term noise and fluctuations. ARIMA models are widely used in domains like finance and economics, adept at capturing both short-term and long-term dependencies in time series data. Effective model parameter selection, denoted as 'p,' 'd,' and 'q,' is crucial for optimal forecasting performance, often determined through statistical analysis and model selection methods.

Mathematical Representation:

2. **SARIMAX (Seasonal ARIMA):**

The SARIMAX[13] (Seasonal AutoRegressive Integrated Moving Average) model is an extension of the ARIMA model, designed to handle time series data with significant seasonality. It combines the fundamental ARIMA components (AutoRegressive, Integration, and Moving Average) with additional seasonal counterparts, denoted as (P, D, Q, S). The seasonal components capture periodic patterns in the data,

making SARIMAX particularly effective for forecasting tasks in domains with recurring patterns, such as monthly or quarterly data. SARIMAX models are valuable tools for time series forecasting when seasonality plays a crucial role, enabling analysts to account for both short-term and long-term dependencies in the data while addressing season-specific variations.

3. **Exponential Smoothing:**

Exponential smoothing[14] is a time series forecasting method that relies on weighted averages of past data points to make future predictions. It assigns exponentially decreasing weights to historical observations, with more recent data points receiving higher importance. This approach allows the model to capture short-term fluctuations and trends in the data, making it particularly useful for scenarios where there is no clear seasonality or where data exhibit changing patterns over time. Exponential smoothing methods come in various forms, including Simple Exponential Smoothing (SES), Holt's Linear Exponential Smoothing (Holt's), and Holt-Winters' Exponential Smoothing (HW). In this project Simple Exponential smoothing model is used.

Mathematical Representation:

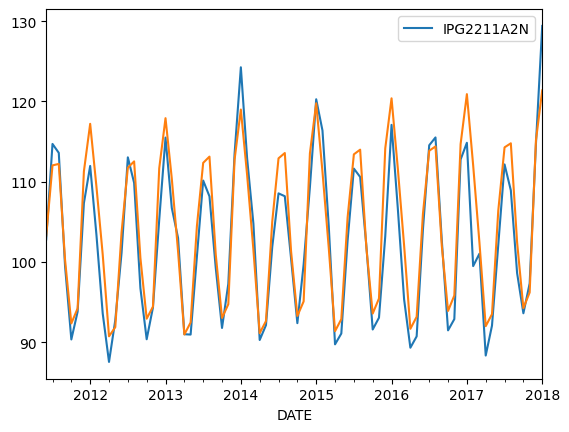
4. **LSTM (Long Short-Term Memory):**

The Long Short-Term Memory (LSTM) model[15] is a specialized recurrent neural network (RNN) architecture designed to excel in capturing and modeling long-range dependencies in sequential data. Its unique architecture features gates that regulate information flow, preventing vanishing gradient problems encountered in traditional RNNs. LSTM has found extensive use in various applications, such as natural language processing, speech recognition, and time series forecasting, thanks to its ability to effectively remember and learn from both short-term and long-term temporal patterns in data, making it a robust and versatile tool in the field of deep learning and sequential data analysis.

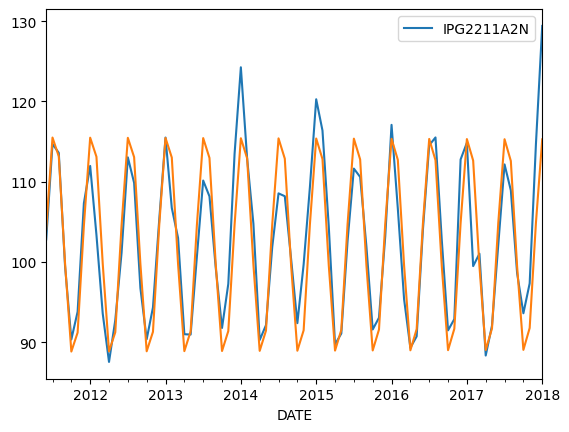
Mathematical Representation:

**5 Evaluation Matrix**

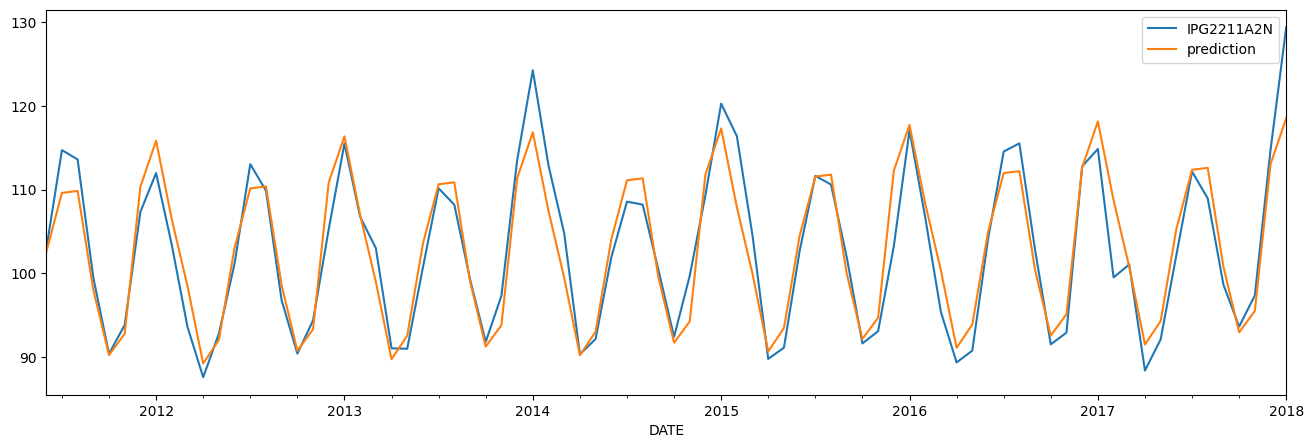
1) Using SARIMAX model:



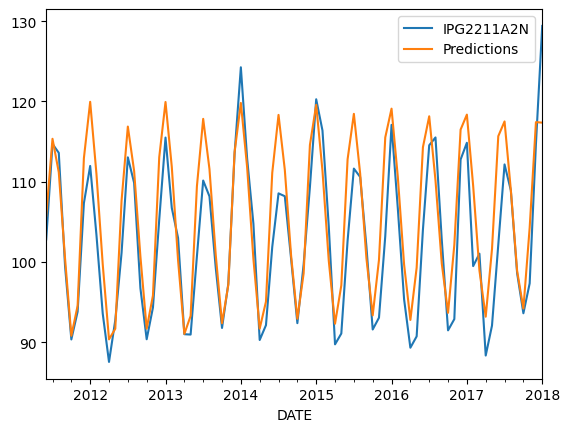
2) Using ARIMA model



3) Exponential Smoothing model:



4) LSTM model:



**ERROR COMPARISON**

:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| MODEL | | MSE | RMSE | MAE |
| ARIMA | 19.357927650037503 | | 4.399764499383746 | 3.305294362821269 |
| SARIMAX | 13.934959254951739 | | 3.7329558335120625 | 2.9224272624082888 |
| SES | 11.414163237160977 | | 3.378485346595568 | 2.5321792816595488 |
| LSTM | 21.86528210671711 | | 4.676032731570333 | 3.3016430048045513 |

**6 Result**

From the values obtained, SES exhibited the lowest RMSE and MAE values among all models, indicating superior predictive accuracy. This suggests that for datasets with a consistent trend, SES can be a suitable choice. SARIMAX followed closely behind SES in terms of performance, indicating its effectiveness in capturing both seasonal and exogenous factors.

ARIMA, while a robust model, demonstrated slightly higher errors compared to SES and SARIMAX, suggesting limitations in capturing certain patterns present in the dataset. LSTM, being a deep learning model, showed the highest RMSE and MAE values, indicating that its complexity might not be necessary for this particular dataset. These findings contribute to the understanding of time series forecasting models and can guide practitioners in choosing the most suitable approach based on their dataset characteristics.

**7 References**

**[1]** Box, George EP, and Gwilym M. Jenkins. "Time series analysis forecasting and control-rev." (1976).

**[2]** Hyndman, Rob J., and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

**[3]** Gardner Jr, Everette S. "Exponential smoothing: The state of the art." *Journal of forecasting* 4.1 (1985): 1-28.

**[4]** Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* 9.8 (1997): 1735-1780.

**[5]** Lipton, Zachary C., et al. "Learning to diagnose with LSTM recurrent neural networks." *arXiv preprint arXiv:1511.03677* (2015).

**[6]** Masini, Ricardo P., Marcelo C. Medeiros, and Eduardo F. Mendes. "Machine learning advances for time series forecasting." *Journal of economic surveys* 37.1 (2023): 76-111.

**[7]** Harvey, Andrew C. "Forecasting, structural time series models and the Kalman filter." (1990).

**[8]** Huang, Norden E., et al. "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis." *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences* 454.1971 (1998): 903-995.

**[9]** Sean, J., and B. Letham. "Prophet: forecasting at scale." *Facebook Research* (2017).

**[10]** Natekin, Alexey, and Alois Knoll. "Gradient boosting machines, a tutorial." *Frontiers in neurorobotics* 7 (2013): 21.

**[11]** Bengio, Yoshua, Réjean Ducharme, and Pascal Vincent. "A neural probabilistic language model." *Advances in neural information processing systems* 13 (2000).

**[12]** Ariyo, Adebiyi A., Adewumi O. Adewumi, and Charles K. Ayo. "Stock price prediction using the ARIMA model." *2014 UKSim-AMSS 16th international conference on computer modelling and simulation*. IEEE, 2014.

**[13]** Arunraj, Nari Sivanandam, Diane Ahrens, and Michael Fernandes. "Application of SARIMAX model to forecast daily sales in food retail industry." *International Journal of Operations Research and Information Systems (IJORIS)* 7.2 (2016): 1-21.

[**14**] Hyndman, Rob, et al. *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media, 2008.

[**15**] Sunny, Md Arif Istiake, Mirza Mohd Shahriar Maswood, and Abdullah G. Alharbi. "Deep learning-based stock price prediction using LSTM and bi-directional LSTM model." *2020 2nd novel intelligent and leading emerging sciences conference (NILES)*. IEEE, 2020.