ISSN: 2583-9292 (ONLINE)
DOI: 10.21917/ijdsml.2025.0170

AN EMPIRICAL COMPARISON OF MACHINE LEARNING MODELS FOR TIME SERIES FORECASTING

Simranjeet Kaur and Jayshree Kundargi
Department of Electronics and Telecommunication Engineering, Somaiya Vidyavihar University, K.J. Somaiya College of Engineering, India

Abstract

Time series data analysis and forecasting stands as a critical information source, shaping future decision-making, strategy formulation, and operational planning across diverse industries. Ranging from marketing and finance to education, healthcare, and robotics, the time series data has become pivotal in guiding effective actions. Time series data analysis plays a pivotal role in understanding sequential trends and patterns present in the data. The Time series forecasting has been used for prediction for effective decision making. The forecasting techniques consist of statistical models and machine learning models. This paper examines different methods, including AR, MA, ARMA, ARIMA, SARIMA, ARIMAX, SARIMAX, Prophet and LSTM. Two meteorological datasets have been analyzed and the above models have been applied and evaluated using various performance metrics.

Keywords:

Time Series Analysis, Forecasting, Statistical Models, Prophet, LSTM

1. INTRODUCTION

Time series analysis is a systematic approach that plays a crucial role in uncovering the underlying patterns present in data sequences indexed over time. Its primary objective is to understand and predict trends, fluctuations, and periodic variations observed in these chronological data points, ranging from long-term yearly trends to short-term minute-by-minute fluctuations.

Time series analysis has a huge number of applications. In the field of meteorology, Time series analysis is of utmost importance for modelling and predicting weather patterns. By scrutinizing historical data, meteorologists can anticipate future climatic shifts and make informed predictions. Similarly, in the empire of finance, particularly in stock markets, Time series analysis, assists in judging price movements and trends. This aids investors and analysts in making well-informed decisions regarding their investments.

This paper presents the empirical comparison of various machine learning models used for the Time Series Forecasting. Two meteorological datasets have been considered; the various performance metrics are applied and appropriate metric is considered for evaluation to obtain the best model for each of the dataset. Thereby, providing the model with best performance for the dataset. This also helps prove the ability of Time Series Analysis to uncover historical trends and predict future trajectories which empowers stakeholders to take firm decisions to promote growth of their respective business.

2. LITERATURE REVIEW

Wibawa et al. [1] emphasized the effectiveness of a Convolutional Neural Network (CNN) with AR and optimal smoothing for time series forecasting, focuses on the utilization of a CNN with smoothed features for the analysis of time-series data. Specifically, the study examines a year-long time series of daily website visitors. The main findings of the study reveal that a CNN with an autoregressive (AR) component outperforms both standalone CNN and Long Short-Term Memory (LSTM) models in the context of time-series forecasting. Furthermore, the paper suggests that incorporating an optimal smoothing factor enhances the performance of the CNN, making it more effective compared to other selected methods for time-series forecasting. To summarize, this research emphasizes the effectiveness of a CNN with AR and optimal smoothing in improving the accuracy of time-series forecasting, particularly when applied to daily website visitor data.

Yunus et al. [2] focused light on a comprehensive examination of open-source tools designed for the analysis of time series data. The primary focus of this paper is to highlight the accessibility of these tools, as they are freely available for use. Rather than conducting specific analyses, the paper serves as a reference point by providing a detailed list and description of these tools based on the existing literature. The models implemented were ARIMA, LSTM, Profet the datasets used were benchmark dataset from University of Irvine (UCI) repository. However, review work did not carry out evaluation of models implemented or results.

Wen et al. [3] emphasized on the representative methods across different categories, highlighting their strengths and limitations through experimental evaluations. Additionally, it sheds light on future research directions in the application of transformers to time series analysis, specifically mentioning the use of benchmark ETTm2 dataset on the various models like ARIMA, RNN and CNN. It was seen that ARIMA model had outperformed the other two models.

Zhou, H. et al. [4] emphasized on predicting the oil temperature of transformers in long sequence time-series forecasting. The study evaluates the performance of five models (ARIMA, LSTMa, LSTnet, Prophet, DeepAR) and introduces the Informer mechanism. The results show that the Informer mechanism effectively reduces the forecasting problem for long sequence time-series and improves model performance in various scenarios, including ETT dataset, ECL dataset, and weather dataset.

Ray et al. [5] digs into various models such as ARIMA, vector AR and a Bayesian nowcasting model. These models were applied to daily stock market data from National Stock Exchange (NSE) and twitter dataset. The research concludes that the Bayesian model, which incorporates sentiment scores and anomaly detection, outperforms the other models in short-term stock price forecasting.

Wijesinghe et al. [6] The researchers conducted a study to investigate various time series forecasting algorithms, including ARIMA and ANN. They utilized the Colombo Stock Exchange (CSE) dataset for their analysis. Upon applying these models, it was observed that the ANN algorithm outperformed the ARIMA model.

Parmar et al. [7] in their study proposed utilization of clustering methods to categorize different regions. Additionally, it employs ARIMA, AR, MA, and ARMA models to forecast values using a dataset of water samples that were examined for minerals. The research findings indicate that the ARIMA model demonstrates higher precision in predicting water quality and identifying the influence of minerals found in the Madiyan-rood River.

After an extensive review of diverse literature in time series analysis, statistical methods stand out for their reliability, interpretability, and proven efficacy across multiple domains. The comprehensive exploration of numerous research papers showcases statistical models like ARIMA, AR, MA, ARMA, and ARIMA consistently delivering accurate forecasts in various applications. These methods offer tangible advantages, such as handling data effectively, incorporating sentiment analysis for stock price prediction, and providing precision in water quality forecasting.

Both the statistical models and machine learning models (especially Deep Learning model) make use of Time series data for modelling and although the Deep Learning models have been deployed for Time Series forecasting, they need complex computation requirement in terms of memory time and processing power whereas the literature review indicates that statistical models offer excellent performance and do not need any special computing requirement. The hybrid combination of both is likely to offer better performance. In this paper the performance of various models has been discussed in detail.

3. IMPLEMENTATION OF MACHINE LEARNING MODELS ON METEOROLOGICAL DATASET

The utilization of statistical models facilitates experiential insights across datasets. In this context, two meteorological datasets have been employed: one of the datasets encompassing the Annual Mean Temperature of India (AMT dataset), and another dataset containing the daily temperature readings specific to Delhi city (DCT dataset).

3.1 ANNUAL MEAN TEMPERATURE DATASET (AMT)

3.1.1 Dataset Description:

The dataset used for the implementation of Statistical Time series models is obtained from the Open Government Data [8], which is a data repository maintained by the Government of India. The dataset contains the annual mean temperature of India from the year 1901 to 2021. The dataset has total 121 data instances. Since univariate forecasting is implemented the variables 'YEAR' and 'ANNUAL' have been considered where 'YEAR' represents the years ranging from 1901 to 2021 and 'ANNUAL' represents mean annual temperature of India in degree Celsius. This dataset will be further referred as 'AMT'. The dataset is split into train data and test data, 1901 to 1994 are in the train data (80%) and rest from the year 1995 to 2021 are in test data (20%). The Fig.1 illustrates the plot for the AMT dataset.

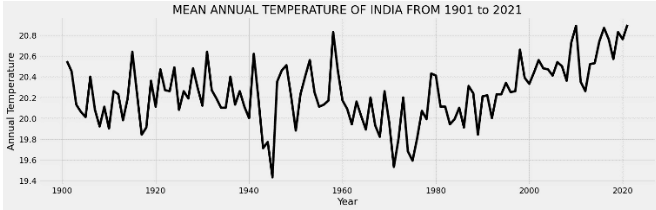

Fig.1. Plot of year v/s the Annual Mean Temperature (°C)

3.1.2 Decomposition of Data:

The decomposition of dataset helps to realize the trend, seasonality and residual. The additive model has been used to decompose the dataset. As shown in Fig.2. the dataset has a very slight upward trend. No seasonal component is present in the data. The empty residual plot indicates all the information of the data has been extracted.

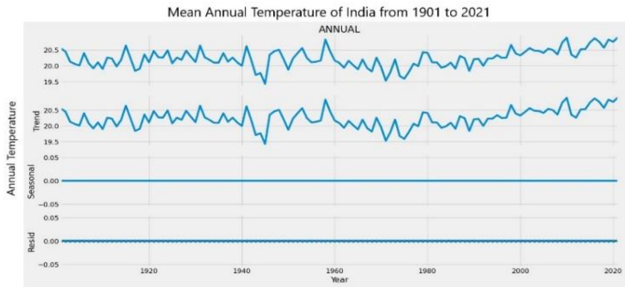

Fig.2. Decomposition of AMT dataset

3.1.3 Stationarity Test:

Since the ADF and KPSS test both indicate that dataset is non stationary it needs to be converted into stationary series by applying the Box-Cox and differencing.

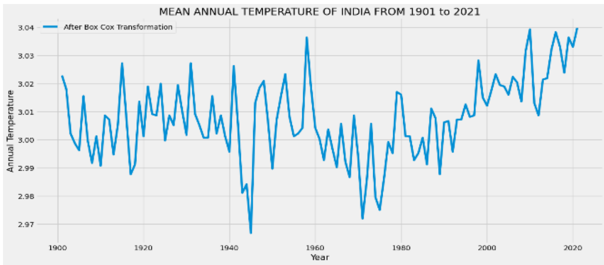

Fig.3.Box-Cox Transformation on AMT dataset

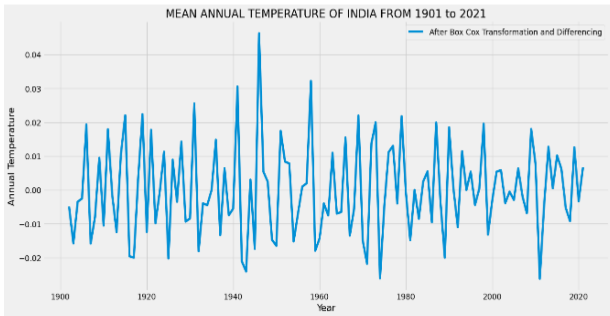

Fig.4. Differencing on AMT dataset

The Fig.3. shows the series after the Box-Cox Transformation and Fig.4. shows that the trend component has been removed and the 'ANNUAL' variable has been made stationary and further models can be applied.

3.1.4 AR Model:

For the implementation of AR model, we need to calculate the lag value which is calculated from the Partial Autocorrelation Function also called as the PACF plot. The value of 'p' is 3 for this series as shown in the Fig.5., so the model can be written as AR (3). The Fig.6. illustrates that the AR model has successfully captured the trend of the model.

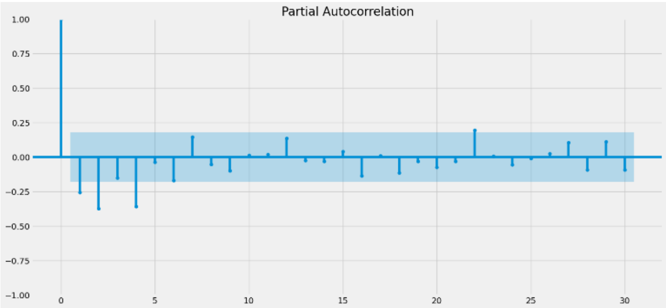

Fig.5. PACF plot for AMT dataset

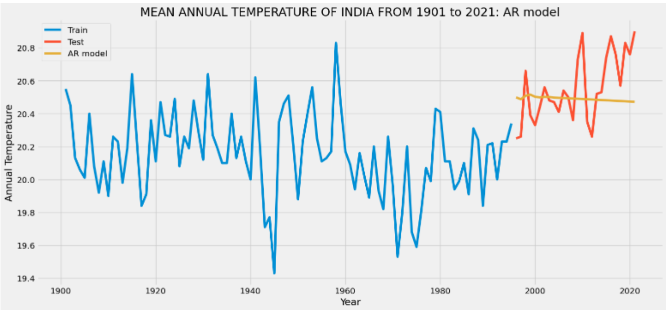

Fig.6. AR model

3.1.5 MA Model:

For the implementation of MA model, we need to calculate the past forecast error value which is calculated from the Autocorrelation Function also called as the ACF plot. Here the value of '*q*' is 2 as shown in the Fig.7, so the model becomes MA (2). The Fig.8. illustrates that the MA model has successfully captured the trend of the model.

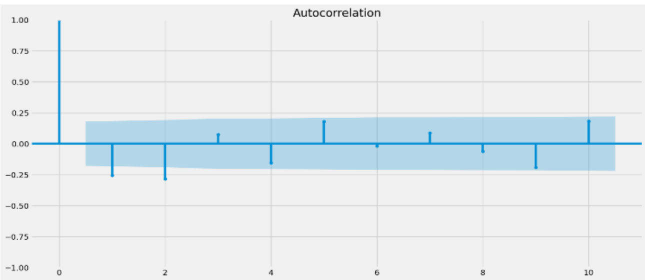

Fig.7. ACF plot

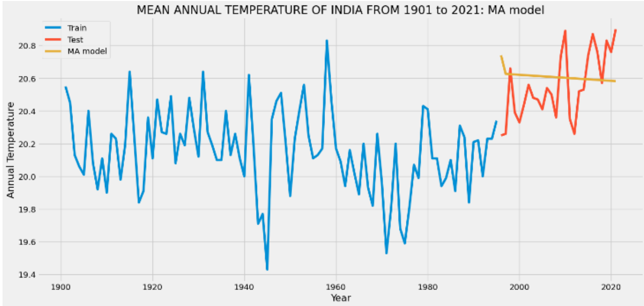

Fig.8. MA model

3.1.6 ARMA and ARIMA Model:

For the ARMA model the combined values of AR and MA are taken into consideration so the ARMA model can be written as ARMA (4,0,2). The forecast of ARMA model is shown in the Fig.9. For the ARIMA model we consider the differencing factor which helped make the variable stationary here on differencing once the data became stationary so 'd' is 1. ARIMA model here is using the autoarima function to find the best parameters for the dataset; the model can be written as ARIMA (1,1,2). The forecasted values of ARIMA model can be seen in the Fig.10.

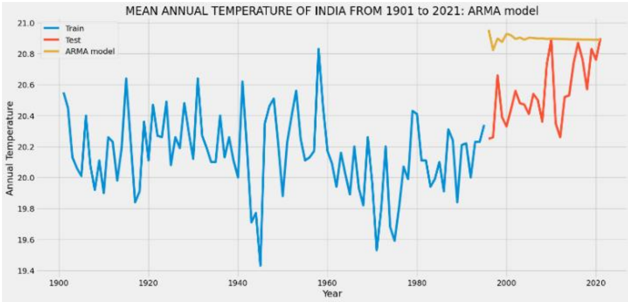

Fig.9. ARMA model

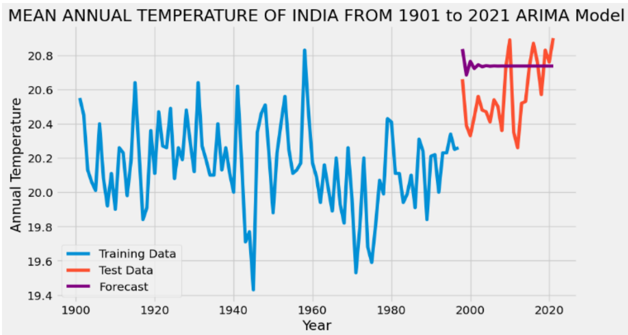

Fig.10. ARIMA model

3.1.7 SARIMA Model:

Making use of the autoarima function to plot the best parameters for SARIMA model we can definitely conclude from the parameters that seasonality is absent in the 'ANNUAL' variable and the model parameters obtained are SARIMA (1,1,2)(0,0,0,12). The Fig.11. shows the plot of SARIMA model.

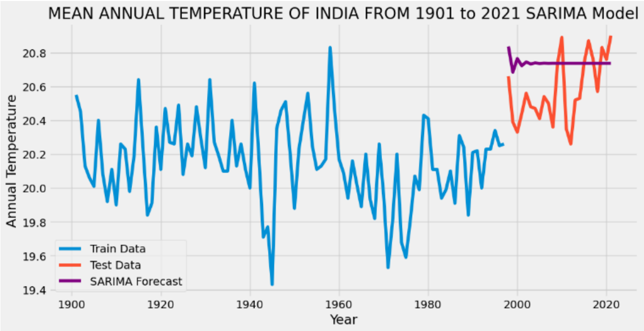

Fig.11. SARIMA model

3.1.8 *LSTM Model:*

Hyper parameter grid search was performed to obtain the select the best parameters for the Prophet model. On performing the exhaustive search, the best values obtained are mentioned below: units: 64, activation: relu, optimizer: rmsprop, loss: mean_squared_error and epochs: 20. The Fig.12. shows the plot of LSTM model.

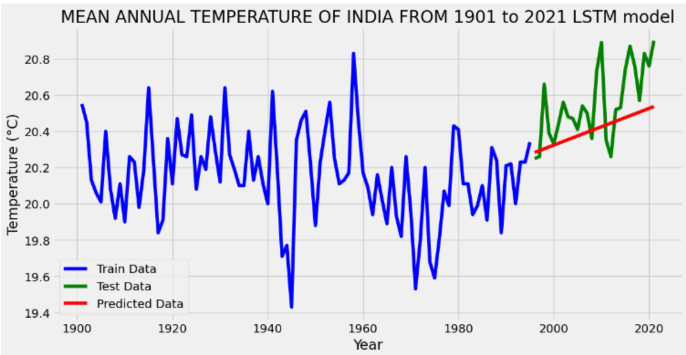

Fig.12. LSTM model

3.1.9 *Prophet Model:*

There are various parameters that are required to be tuned in order to obtain optimum results for the Prophet model. Using the Grid Search method the best values for the parameters are 'changepoint_prior_scale': 0.5, 'holidays_prior_scale': 0.01, 'n_changepoints': 20, 'seasonality_mode': 'additive', 'seasonality_prior_scale': 0.1 these parameters provided the minimum value of MAE. It must be noted that the Prophet model takes care of stationarity of data and outliers and hence can be applied directly. The Fig.13. shows the plot of Prophet model.

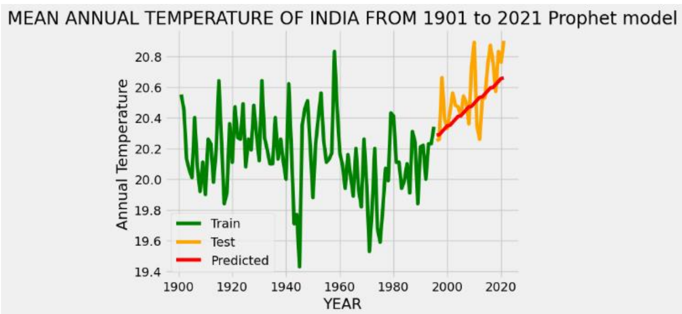

Fig.13. Prophet model

3.2 DAILY CLIMATE TIME SERIES DATASET (DCT)

3.2.1 *Dataset Description:*

The dataset used for the implementation of ARIMAX and SARIMAX is taken from the Kaggle [9], the dataset contains the daily temperature of Delhi city from the year 2013 to 2017. The dataset has total 1463 data instances. The variable 'meantemp' is the endogenous variable and the remaining variables 'humidity', 'meanpressure' and 'wind_speed' represent exogenous variable. The variable 'date' represents the daily dates from 2013 to 2017; 'meantemp' refers as the mean temperature throughout the day in degree Celsius; 'humidity' represents the moisture present on a particular day in grams of water vapor per cubic meter of air; 'meanpressure' refers as the mean of the atmospheric pressure; 'wind_speed' refers to the speed of the winds blowing for the particular day in kilometer per hour. The dataset is split into train data (80%) and test data (20%) the instances 1-1-2013 to 11-11-2015 are used in train data and instances 12-11-2015 to 01-01-2017 are in test data.

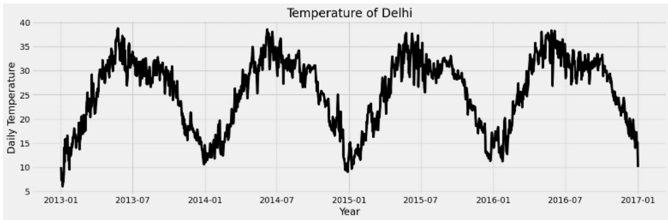

Fig.14. Plot of Year v/s Daily Temperature in (°C)

3.2.2 *Decomposition of DCT Dataset:*

As shown in Fig.15. the dataset has an upward trend. A repetitive seasonal component is present in the data. The residual plot indicates few points present indicating all the information of the data has not been extracted.

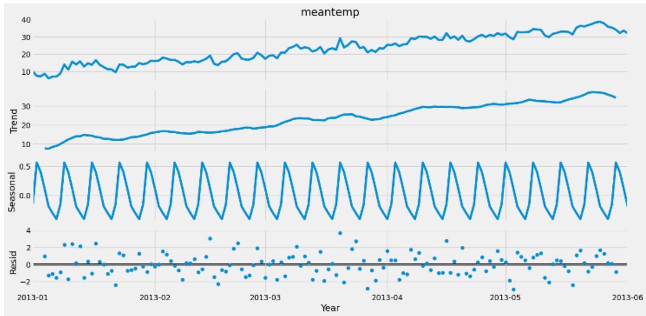

Fig.15. Decomposition of DCT dataset

On applying the ADF test and KPSS test it is observed that the series is stationary so further transformation, and differencing need not be applied.

3.2.3 *ARIMAX Model:*

The ARIMAX model will help understand the effect of various factors on the daily mean temperature by taking into account the past temperature values (AR component), differences between the temperature values (I component), past forecast errors (MA component), and external factors like humidity and wind speed (exogenous variables) to improve the accuracy of temperature forecasts as shown in the Fig.16. and Fig.17. The model parameters are determined using the PACF and ACF. The model can are ARIMAX (7,0,10).

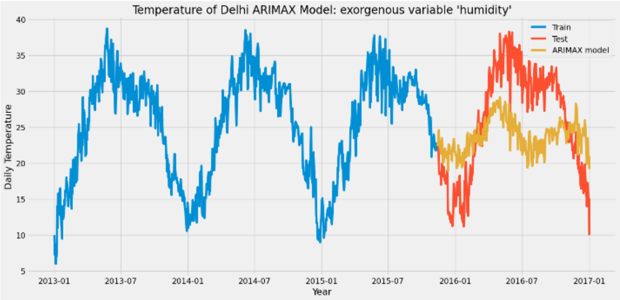

Fig.16. ARIMAX model on DCT dataset with ‘humidity’ as exogeneous variable

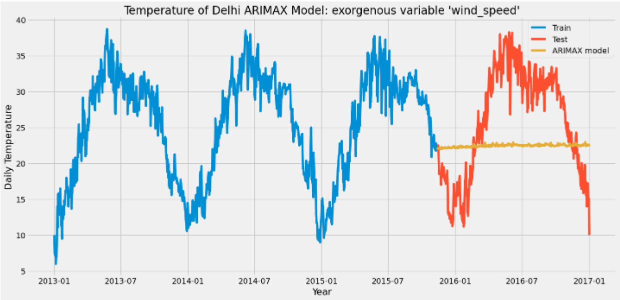

Fig.17. ARIMAX model on DCT dataset with exogenous variable ‘wind_speed’

3.2.4 SARIMAX Model:

The SARIMAX model is similar to the ARIMAX model but here it takes seasonality into account along the other parameters present. The model parameters for the DCT are given as SARIMAX (6,0,10) (2,1,1,12).

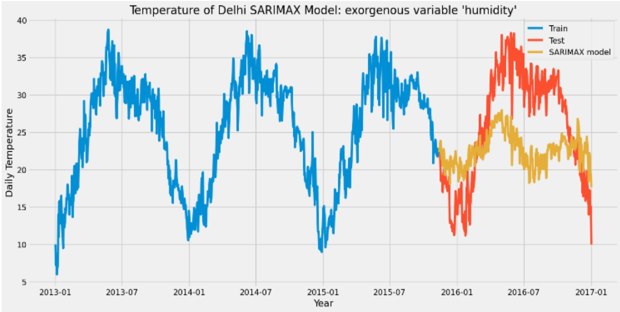

Fig.18. SARIMAX model on DCT dataset with ‘humidity’ as exogeneous variable

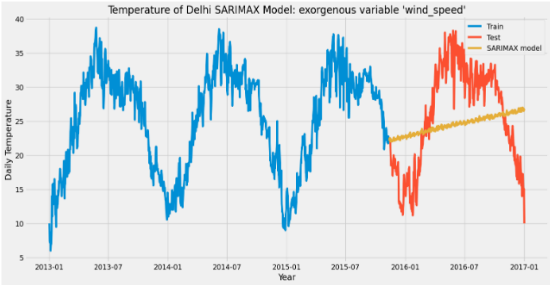

Fig.19. SARIMAX model on DCT dataset with ‘wind_speed’ as exogeneous variable

4. RESULTS

The choice of performance metric depends on the specific characteristics of the problem and the type of errors that needed to be accounted for. For forecasting the AMT dataset and the DCT dataset, using MAE is more interpretable as the metric provide a clearer understanding of the average magnitude of errors without squaring the differences.

Table.1. Comparison of performance evaluation metrics on AMT dataset

	Values for first 20 instances of test data				Values for all instances in test data			
Model	**MSE**	**MAE**	**MAPE**	**SMAPE**	**MSE**	**MAE**	**MAPE**	**SMAPE**
AR	0.17	0.14	0.66	0.66	0.22	0.17	0.85	0.02
MA	0.23	0.19	0.94	0.94	0.22	0.19	0.95	0.02
ARMA	0.45	0.41	2.03	2.00	0.40	0.34	1.69	0.03
ARIMA	0.06	0.20	1.97	2.00	0.20	0.21	1.03	0.02
SARIMA	2.26	2.25	10.98	10.40	5.23	2.28	11.08	0.22
LSTM	0.03	0.14	0.76	15.43	0.04	0.17	84.31	145.75
Prophet	0.04	0.17	0.87	0.87	0.06	0.13	1.01	1.02

Table.2. Comparison of Exogenous variable for ARIMAX model on DCT dataset

	Values for first 200 instances of test data				Values for first all instances of test data			
Exogenous Variable	**MSE**	**MAE**	**MAPE**	**SMAPE**	**MSE**	**MAE**	**MAPE**	**SAMPE**
humidity	6.04	4.69	25.58	21.63	30.05	5.34	24.39	22.05
Meanpressure	163617	32.26	127.66	17.71	289.37	24.64	96.07	17.22
wind_speed	63.20	6.86	32.37	30.10	5.06	7.08	28.70	28.96

Table.3. Comparison of Exogenous variable for SARIMAX model on DCT dataset

	Values for first 200 instances of test data				**Values for first all instances of test data**			
Exogenous variable	**MSE**	**MAE**	**MAPE**	**SMAPE**	**MSE**	**MAE**	**MAPE**	**SAMPE**
humidity	6.49	4.89	25.10	22.13	32.18	5.80	24.62	23.78
meanpressure	25233	16.60	70.02	25.54	113.64	13.25	53.05	22.84
wind_speed	59.50	6.66	31.85	29.24	7.64	6.73	27.99	27.45

They also provide insights into the average magnitude of errors without overemphasizing extreme outliers, enabling a more comprehensive assessment of the model's performance in predicting daily temperature changes.

The Table.1 shows the forecasting on the AMT dataset with different statistical models. It also shows the performance of the model over short duration and long duration. The machine learning model that performed well on AMT dataset is the Prophet model however its performance improves on the long-term duration indicating the requirement of the model to have significant amount of data available for training.

The Table.2 shows the performance of ARIMAX model on DCT dataset with respect to exogenous variables. So, taking into account the MAE values the model performed well exogenous variable 'humidity' the model however has performance dip on the long-term forecasting. Similarly, Table.3 shows the performance of SARIMAX model on DCT dataset with respect to exogenous variables the model performed well on exogenous variable 'humidity'. So, for the DCT dataset on comparing the MAE values of ARIMAX and SARIMAX model for the exogenous it can be concluded that ARIMAX model works well with exogenous variable.

Also, it can be observed from the Tables above that the MAE showed slight drop in performance on long-term forecasting. This proves that the statistical models and LSTM models have slight incompetence to forecast values for longer duration of time and hence for this overcoming this short coming the Prophet model is purposed.

5. CONCLUSION

The AMT dataset was modelled using AR, MA, ARMA, ARIMA, SARIMA, Prophet and LSTM models and the performance of each model was evaluated using various performance evaluation metrics for short term and long-term forecasting. The Prophet model has shown the best performance. The DCT dataset was modelled using ARIMAX and SARIMAX models and the performance of the models was evaluated using various performance evaluation metrics for short term and long-term forecasting The ARIMAX and SARIMAX both performed well on 'humidity' exogenous variable but ARIMAX model provided better performance. For both the datasets MAE metric showed minor drop in performance for long term forecasting.

REFERENCES

[1] A.P. Wibawa, A.B.P. Utama, H. Elmunsyah, U. Pujianto, F.A. Dwiyanto and L. Hernandez, "Time-Series Analysis with Smoothed Convolutional Neural Network", *Journal of Big Data*, Vol. 9, No. 44, pp. 1-18, 2022.

[2] Yunus Parvej Faniband, I. Ishak and S.M. Sait, "A Review of Open Source Software Tools for Time Series Analysis", *Proceedings of International Conference on Artificial Intelligence*, pp. 1-21, 2022.

[3] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan and L. Sun, "Transformers in Time Series: A Survey", *Proceedings of International Conference on Artificial Intelligence*, pp. 1-9, 2023.

[4] H. Zhou, S. Zhang, J. Peng and Z. Shuai, "Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting", *Proceedings of International Conference on Artificial Intelligence*, Vol. 35, No. 12, pp. 11106-11115, 2021.

[5] P. Ray, B. Ganguli and A. Chakrabarti, "A Hybrid Approach of Bayesian Structural Time Series with LSTM to Identify the Influence of News Sentiment on Short-Term Forecasting of Stock Price", *IEEE Transactions on Computational Social Systems*, Vol. 8, No. 5, pp. 1153-1162, 2021.

[6] G.W.R.I. Wijesinghe and R.M.K.T. Rathnayaka, "Stock Market Price Forecasting using ARIMA vs ANN; A Case Study from CSE", *Proceedings of International Conference on Advancements in Computing*, pp. 269-274, 2020.

[7] K.S. Parmar and R. Bhardwaj, "Water Quality Management using Statistical Analysis and Time-Series Prediction Model", *Applied Water Science*, Vol. 4, No. 4, pp. 425-434, 2020.

[8] "Aadhaar Enrolment and Update Data", Available at https://www.data.gov.in/, Accessed in 2025.

[9] "Daily Climate Time Series Data", *Kaggle*, Available at https://www.kaggle.com/datasets/sumanthvrao/daily-climate-time-series-data, Accessed in 2019.