Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 235 (2024) 656–670

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering (ICMLDE 2023)

SARIMA Model: An Efficient Machine Learning Technique for Weather Forecasting

Shabnam Kumari[a], P. Muthulakshmi[b]

[a,b]Department of Computer Science, SRM Institute of Science and Technology, Kattankulathur, Chennai, 603203, Tamilnadu, India

Abstract

Weather forecasting is a critical tool for many different applications, from agriculture and transportation to disaster preparedness and response. While weather forecasts are not always perfect, they provide valuable information that can help people make informed decisions and take appropriate actions to protect themselves and their property from the impacts of extreme weather events. In this paper, to forecast the weather we are using the SARIMA model as RMSE is 1.24 and analyse how it forecast the weather with high accuracy. Note that the SARIMA model is a machine learning (ML) technique used to forecast time series data that has both trend and seasonal components. It is an extension of the ARIMA model, which stands for Autoregressive Integrated Moving Average. To account for periodic changes in the data, the SARIMA model augments the ARIMA model with a seasonal component. The SARIMA model is trained on historical weather data and can be used to predict future weather patterns. It is a powerful tool for weather forecasting as it can accurately predict both short-term and long-term weather trends. The quality of the input data determines how accurate the SARIMA model will be, and the model can be adjusted to perform better. In conclusion, the model known as SARIMA is a ML based approach that has been successfully used to predict weather. It is a powerful tool that can handle non-stationary data and seasonal components and can accurately predict short-term and long-term weather trends.

Peer-review under responsibility of the scientific committee of the International Conference on Machine Learning and Data Engineering

Keywords: *Climate forecasting; Weather prediction; Machine learning, SARIMA model;*

1. Introduction

Making forecasts or projections about future events or conditions based on historical and current data is the process of forecasting. It is an important tool for decision-making in many fields, including business, economics, finance, and weather forecasting, among others [1]. The goal of forecasting is to reduce uncertainty about the future

* Corresponding author.Shabnam Kumari.

E-mail address: sk2581@srmist.edu.in

Peer-review under responsibility of the scientific committee of the International Conference on Machine Learning and Data Engineering
10.1016/j.procs.2024.04.064

and provide meaning information that can guide strategic planning and decision-making. There are various methods of forecasting, ranging from simple extrapolation of trends to complex statistical models that incorporate multiple variables and factors. The type of data and the application's particular requirements will determine the approach to use. Regression analysis, time series analysis, and ML algorithms are a few of the techniques that are often used. Forecasting can be both challenging and rewarding. While accurate predictions can help organizations and individuals make better decisions, inaccurate forecasts can lead to costly errors and missed opportunities. Therefore, it is important to use appropriate methods and tools and to constantly monitor and update the forecasts as new data becomes available.

There are several uses for weather forecasting. First and foremost, it helps individuals and organizations to plan and prepare for weather-related events, such as storms, floods, and extreme temperatures. This can help to reduce the impact of these events and minimize damage to property and infrastructure [2]. Weather forecasting can also be valuable for businesses that are directly impacted by weather conditions, such as agriculture, construction, and transportation. By forecasting weather patterns and conditions, businesses can make informed decisions about planting, harvesting, construction schedules, and transportation routes. This can help to optimize operations and reduce costs. In addition, weather forecasting is important for emergency management agencies, which rely on accurate forecasts to plan and coordinate responses to natural disasters and other emergencies. By anticipating weather-related risks, these agencies can organise resources and respond more effectively to crises. Weather forecasting can also be valuable for individuals in planning their daily activities, such as choosing appropriate clothing and planning outdoor activities. Accurate weather forecasts can help individuals to avoid discomfort or danger due to weather-related conditions, such as heat exhaustion or frostbite.

Furthermore, weather forecasting is essential for climate modelling and scientific research. By collecting and analysing weather data over time, scientists can develop models that help to explain and predict climate patterns and trends. This can be important for understanding the impact of climate change on the environment and for developing policies to address these challenges.

In addition to the practical applications of weather forecasting, it is also valuable for public safety and awareness. Weather forecasts and alerts can help to inform individuals and communities of impending weather-related risks, such as severe thunderstorms, tornadoes, and hurricanes. This can help to reduce the risk of injury and loss of life. Moreover, weather forecasting can also contribute to the growth of new technologies and innovations. For example, advances in weather forecasting have led to the development of more accurate and reliable weather monitoring equipment, as well as new techniques for analysing weather data and predicting future conditions [3].

Finally, weather forecasting is important for global commerce and trade. Accurate weather forecasts can help to inform decisions related to shipping, transportation, and logistics, which are critical components of international trade. This can help to optimize supply chains and reduce costs, benefiting both businesses and consumers. In summary, weather forecasting is a valuable tool that has a wide range of applications. From emergency management to agriculture to individual planning, accurate weather forecasts help individuals and organizations to make informed decisions and prepare for weather-related events (refer Figure 1). They also play a significant role in scientific research and in understanding the impact of climate change.

Fig. 1. Weather forecasting and different sectors.

Organisation of the work: In this paper to forecast the weather we are using the SARIMA model. The remaining paper is structured as follows: Section II discusses about Literature Review, Section III details the problem definition. Section IV provides a proposed methodology using SARIMA model. Section V implies the simulation results in detail followed by Section VI that describes the model parameters. Section VII discusses about Limitations towards weather prediction. In last, conclusion is given in section VIII.

2. Literature Review

Narasimha Murthy et.al [4] proposed that they forecast the rainfall southwest monsoon in North East India. The pattern and amount of rainfall are critical elements influencing agricultural systems. The monsoon season of Southwest lasts four months in India, from June to September. The current study provides an empirical analysis for predicting and forecasting rainfall patterns associated with the Southwest monsoon in North-East India. For this region, model selection, diagnostic analysis, and forecasting have all been done using the Box-Jenkins Seasonal Autoregressive Integrated Moving Average (SARIMA) method. Further, Peng Chen et.al [5] investigates in Time Series Forecasting of Temperatures using SARIMA: An Example from Time series modelling and forecasting, which predicts future values by analysing previous values, is useful in many practical applications. In this research, we evaluate the average monthly temperature in Nanjing, China, from 1951 to 2017 using SARIMA (Seasonal Autoregressive Integrated Moving Average) techniques. The training set consists of data from 1951 to 2014, whereas the testing set consists of data from 2015 to 2017. The choice of a model and the precision of prediction are thoroughly discussed. The findings suggest that proposed research method achieves excellent predicting accuracy. Then, Mehmet Tektaş et.al [6] suggested that the neuro-fuzzy network and statistical models at Göztepe, Istanbul, Turkey. We used data from nine years (2000-2008) to create the models, which included wind speed, air pressure, and daily average temperature (dry-wet). The ARIMA and the Adaptive Network Based Fuzzy Inference System (ANFIS) models were employed. It also examined that the different models uses a separate training and test dataset to confirm the efficiency of ARIMA and ANFIS approaches. Further, Afan Galih Salman et.al [7] proposed the auto-regressive integrated moving average (ARIMA) model in this article uses the grid approach to predict greater clarity for the varied value of the parameters p, d, and q. ARIMA has lowest MSE value which is 0.00029 and the lowest coefficient of variation value which is 0.00315. The MSE value grows as the quantity of estimating data in the ARIMA model increases. Later on, S. I. Vagropoulos et.al [8], in his investigation, contrasts four useful techniques for predicting the power production of grid-connected Photovoltaic (PV) plants: Seasonal Autoregressive Integrated Moving Average (SARIMA) modelling, SARIMAX modelling (SARIMA modelling with exogenous factor), modified SARIMA modelling (a posteriori modification of the SARIMA model), and ANN-based modelling. This comparison yields interesting conclusions on the requirement and benefits via exogenous elements in time series model. Lastly, intra-day prediction updates are used to calculate the SARIMA and SARIMAX models' predicting errors.

Further, Mohammad Valipour et.al [9] suggested about the efficacy of the autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA) models for long-term runoff forecasting in the United States was explored in this work. In the first step, the quantity of runoff for 2011 in each US state is anticipated using data from 1901 to 2010. (mean of all station in each state). The findings demonstrate that SARIMA model outperforms the ARIMA model in terms of accuracy. For all states, the SARIMA model has a relative error of 5%. The runoff is anticipated for 2001 to 2011 in second stage using average yearly runoff data from 1901 to 2000. Flávio Fonseca Nobre et.al [10] proposed This study used data from a national public health surveillance system in the United States to evaluate and compare the outcomes of a seasonal autoregressive integrated moving average (SARIMA) and a dynamic linear model (DLM) in order to estimate the case incidence of two notifiable illnesses. The United States' reported cases of hepatitis A and malaria from January 1980 to June 1995 are used in the comparison. Both predictor models' residuals demonstrate that, they were suitable instruments for usage in epidemiological monitoring. For both models, qualitative factors were taken into account in order to more accurately compare their applicability to public health.

Further, Tsan-Ming Choi et.al [11] investigates that they contrast SW's performance with three forecasting methods: I pure SARIMA, (ii) LESA + CSD, and (iii) evolving neural networks (ENN). It highlights the importance of SW and identifies the circumstances in which it outperforms pure SARIMA and LESA + CSD. It investigates the time series characteristics that affect forecasting accuracy in more detail, and it provides a technique for doing sales prediction based on characteristics of provided sales time series. Further, M. Milenković et.al [12] suggested that the passenger flows on trains are examined, and a workable modelling approach is suggested. It is determined that the time series exhibits a substantial autocorrelation of seasonal features based on previous data assembled from monthly passenger counts obtained on the Serbian railway network. The Seasonal AutoRegressive Integrated Moving Average (SARIMA) approach is used for fitting and predicting the time series that covers the periods from January 2004 to June 2014 in order to cope with seasonal periodicity. The outcomes of the experiments indicate high predicting abilities.

Garima Jain et.al [13] proposed that it was researched independently in the training data and produces a large number of related strategies which are useful in weather forecasting. ARIMA MODEL method is a major approach reviewed in this study for daily meteorology. The method makes use of many criteria to forecast the daily weather, including the amount of precipitation, humidity, temperature, cloud cover, and day's weather. This paper's main contribution is to compare the current meteorological models and to choose the precise model to support their capacity to predict the weather.

Current and Previous Start of Art

SARIMA (Seasonal Autoregressive Integrated Moving Average) is a powerful time series forecasting model used extensively in various domains, including weather forecasting. it extends the traditional ARIMA (Autoregressive Integrated Moving Average) model by incorporating seasonality and trends. Hence, now current and previous start of art for SARIMA model can be discussed as:

Previous State of Art:

- Early Applications: This model has been used in weather forecasting for several decades. They were initially implemented as an improvement over simpler time series models due to their ability to capture both trend and seasonality in historical weather data.
- Manual Tuning: In the previous era, this model required significant manual tuning of hyperparameters, such as the order of differencing, autoregressive, moving average, and seasonal components. This made them somewhat cumbersome to use effectively.
- Data Limitations: In early SARIMA model, we faced difficulties in handling large and complex datasets, including missing data and outliers, which are common in weather data. Hence, this model limited their accuracy and robustness.

Current State of Art:

- Automation and Hyperparameter Tuning: The current state of this model benefits from automation and ML techniques for hyperparameter tuning. Many algorithms like Grid search, genetic algorithms, and Bayesian optimization have been applied to optimize model parameters efficiently.
- Incorporating Exogenous Variables: In current, this model is integrated with exogenous variables, such as satellite data, climate indices, and atmospheric pressure measurements, to enhance forecasting accuracy. This model allows for the consideration of external factors that influence weather patterns.
- High-Performance Computing: This model uses large data and high-resolution datasets with advances in computational power and access to cloud resources, which allows for more accurate and detailed weather forecasting.

- Ensemble Methods: This model is integrated into ensemble forecasting systems, where it works in conjunction with other models, such as neural networks and numerical weather prediction models. This ensemble approach can improve forecasting reliability.
- Probabilistic Forecasting: This model is increasingly used to generate probabilistic weather forecasts, providing not only point predictions but also uncertainty estimates. Hence, this model is more important for risk assessment and decision-making in various applications, such as agriculture and emergency management.
- Integration with Real-Time Data: Real-time data assimilation techniques have been developed to continuously update this model with the recent observations, and improve short-term weather forecasting accuracy.
- Hybrid Models: SARIMA models are often combined with other time series forecasting techniques, such as Seasonal Decomposition of Time Series (STL) or Prophet, to capture complex seasonal and trend patterns effectively.
- Open Source Tools: The availability of open-source libraries and packages (e.g., Python's statsmodels) has made this model more accessible to future researchers and its widespread use in weather forecasting.

In summary, SARIMA model is used for weather forecasting since early days. It now benefits from automation, the integration of exogenous variables, improved handling of large datasets, ensemble techniques, probabilistic forecasting, real-time data integration, and open-source tools. These advancements have made SARIMA a valuable and efficient ML technique in the field of weather forecasting, contributing to more accurate and reliable predictions for various applications.

3. Problem Definition

In the realm of meteorology, weather forecasting is important application of machine learning. Accurate weather forecasts are important for a wide range of sectors, including agriculture, transportation, energy management, and disaster preparedness. One popular ML technique used for weather forecasting is Seasonal Autoregressive Integrated Moving Average (SARIMA) model [14]. The problem definition for using SARIMA model for weather forecasting typically involves the following:

- Time series data: The problem requires a time series dataset, which is a collection of observations recorded at regular time intervals. In the case of weather forecasting, this could be historical data of weather variables such as wind speed, temperature, humidity, precipitation, etc. recorded at hourly, daily, or monthly intervals.
- Seasonal and trend components: Weather data often exhibits seasonal and trend components, such as daily or yearly patterns, which need to be considered in the forecasting model. It specifies the nature of the seasonal component, such as daily, monthly, or yearly, and the expected trend, weather it is linear, exponential, or none.
- Forecast horizon: It specify the forecast horizon, which is the time duration for which the forecasts are needed. It could be short-term forecasts (e.g., next hour, next day) or long-term forecasts (e.g., next week, next month).
- Performance metrics: It specify the performance metrics that will be used to estimate the accuracy of SARIMA model. Common performance metrics for time series forecasting includes: mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and forecast accuracy measures such as forecast error and forecast bias.
- Data preprocessing: It include details on how the time series data will be pre-processed before fitting it to the SARIMA model. It also includes data cleaning, handling missing values, transforming variables if needed, and splitting the dataset into training, validation, and testing sets.

- Model selection and hyperparameter tuning: It specify the process for selecting the appropriate SARIMA model and tuning its hyperparameters. This may involve techniques such as grid search, cross-validation, or model selection based on domain knowledge.
- Deployment and monitoring: It include issues for deploying the trained SARIMA model in a real-world setting and monitoring its performance over time. It also updates the model with new data, evaluating its accuracy, and taking corrective actions if necessary.

In summary, SARIMA model provides its need for predicting the weather accurately, and precisely (refer Figure 2). For that, it needs a clear understanding of the data, the forecasting requirements, the performance evaluation criteria, and the steps involved in model selection, training, deployment, and monitoring. This will help guide the development and implementation of an effective SARIMA-based weather forecasting system [15]. As the problem is clear and refined, it can serve as process of data preparation, model training, evaluation, and deployment.

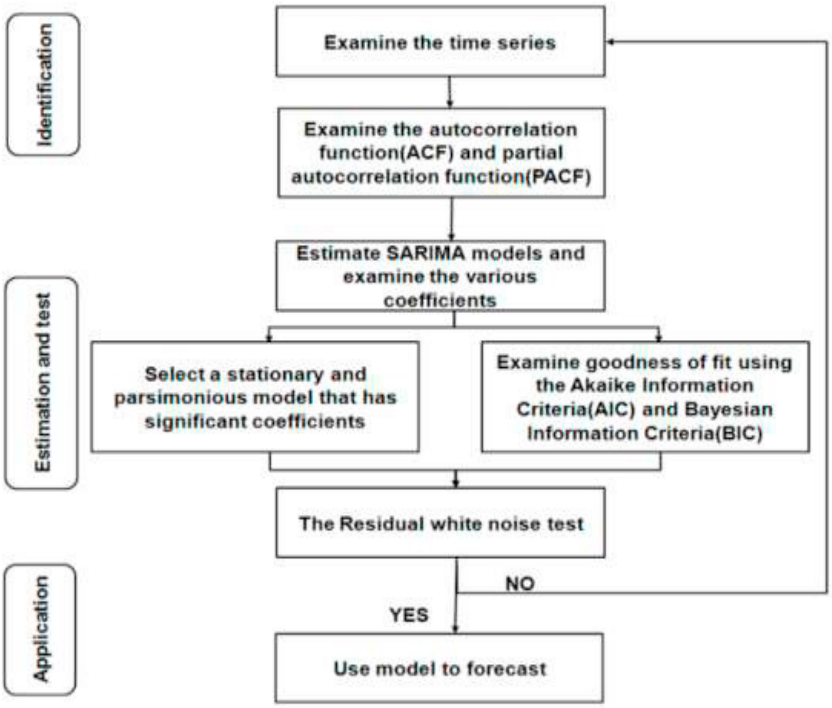

Fig. 2. SARIMA model for forecasting climate change.

4. Proposed Methodology

Weather forecasting is always an important issue (in terms of accuracy) to solve and need efficient solutions from scientific community. Now in this section, we will discuss our proposed work in detail.

4.1. SARIMA MODEL

SARIMA is a standard time series forecasting model which is used to predict future values based on historical data. It is an elongation of the ARIMA model, which is widely used for time series analysis. SARIMA models are useful when the time series being analysed exhibits seasonality, which is a repeating pattern in the data that occurs at fixed intervals. This can include, for example, weekly, monthly, or yearly patterns. SARIMA models consider both the seasonality of the data as well as any trend or irregularities in the time series [16]. To estimate a SARIMA model, the first step is to define the values of p, d, and q, as well as P, D, and Q, by analysing the partial autocorrelation and autocorrelation functions of time series. These functions help identify the lag values that are significant and should be incorporated in the model. Once parameters are determined, the model can be fitted to the data with maximum likelihood estimation. Several statistical tests and metrics, such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), can be used to assess the model's quality of fit.

After the model is fitted, it can be used to generate forecasts of future values. Measures like mean squared error (MSE) or mean absolute error (MAE) can be used to assess the forecasts' accuracy. SARIMA models can be useful in a wide range of applications, such as predicting demand for products, forecasting stock prices, or predicting traffic patterns. However, like any model, SARIMA has limitations and assumptions, and its effectiveness based on quality and relevance of historical data used to estimate the model. It is always important to assess the accurateness and reliability of the model before making any important decisions based on its forecasts [17]. In addition to the basic SARIMA model, there are variations and extensions which can be used to address types of time series data or forecasting problems. For example, the SARIMAX model includes additional exogenous variables that can be used to increase the accurateness of the forecasts. Another variation is the Vector Autoregression (VAR) model, which allows for the modelling of multiple time series that are interrelated. Note that SARIMA models can also be combined with ML techniques, such as neural networks or random forests, to improve forecasting accuracy and incorporate more complex patterns in the data.

As discussed, SARIMA is an effective time series forecast tool that considers the seasonality and pattern aspects of the data. By selecting the appropriate parameters and evaluating the accuracy of the forecasts, SARIMA can help organizations and businesses make more knowledgeable decisions and better plan for the future. However, it is important to use SARIMA in conjunction with other analysis techniques and to carefully evaluate its effectiveness before relying on its forecasts (refer Figure 3 and Figure 4).

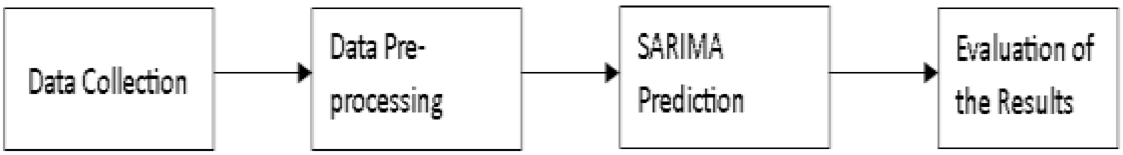

Fig. 3. Steps involved in machine learning techniques.

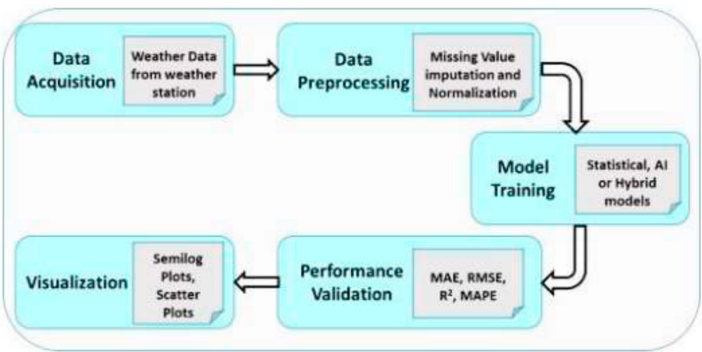

Fig. 4. Machine learning-based approaches for climate change.

5. Implementation Results

From the Figure 5, we can conclude that it is very clear that yearly the data's trend. While there is a high temperature in may month of every year and low temperature during the last months of every year and first months of every year.

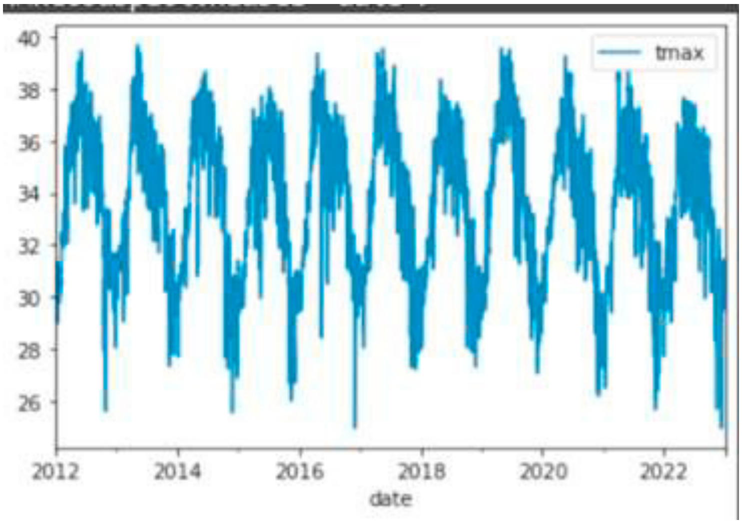

Fig. 5. Plots of residuals: (a) residuals over time

So, we can perform the SARIMA model but we have the condition to apply the SARIMA model that is data should be stationary, then only we can apply SARIMA or forecasting model. For every data to apply the timeseries forecasting model, we need to check the specified data is stationary or not. The data is stationary, data should follow that it is constant mean, constant variance, constant autocorrelation structure and no period component.

In coding to find the stationary of the data we apply augmented Dickey–Fuller test. Augmented Dickey-Fuller Test means if p-value is less than 5%, then we reject the null hypothesis which means given data is stationary if not, then given data is non-stationary. That given data is the stationary data, because the p-value is less than 5% so here we reject the null hypothesis. So, data has no unit root and the data is stationary. So, we apply the SARIMA model to forecast the temperature.

6. SARIMA Model Parameters

As discussed in section 4, SARIMA models are determined by three parameters: p, d, and q. Parameter p represents the number of autoregressive terms, or number of past values used to predict future values. The d parameter indicates how many differences must be made to the time series in order to eliminate trend and seasonality. Parameter q represents the number of moving average terms, or number of past forecasting errors used to predict future values [18]. In addition to these three parameters, SARIMA models also include three seasonal parameters: P, D, and Q. These represent the same concepts as p, d, and q parameters, but are applied to seasonal component of the data.

The seasonal element of the ARIMA model is called SARIMA (Seasonal Auto Regressive Integrated Moving Average). The parameters for these kinds of models are given by SARIMA (p, d, q) x (P, D, Q, s) as follows:

- p and seasonal P: indicates the number of autoregressive terms (lags of stationarized series)

- d and seasonal D: indicates the differencing that must be done to stationarize the series.
- q and seasonal Q: indicates the number of moving average terms (lags of the forecast errors).
- s: indicates the seasonal length in the data.

PACF and ACF plots were used to establish the optimal SARIMA parameters. The optimal values were found using a grid search after narrowing the number of candidate parameters using the PACF and ACF plots as a starting point.

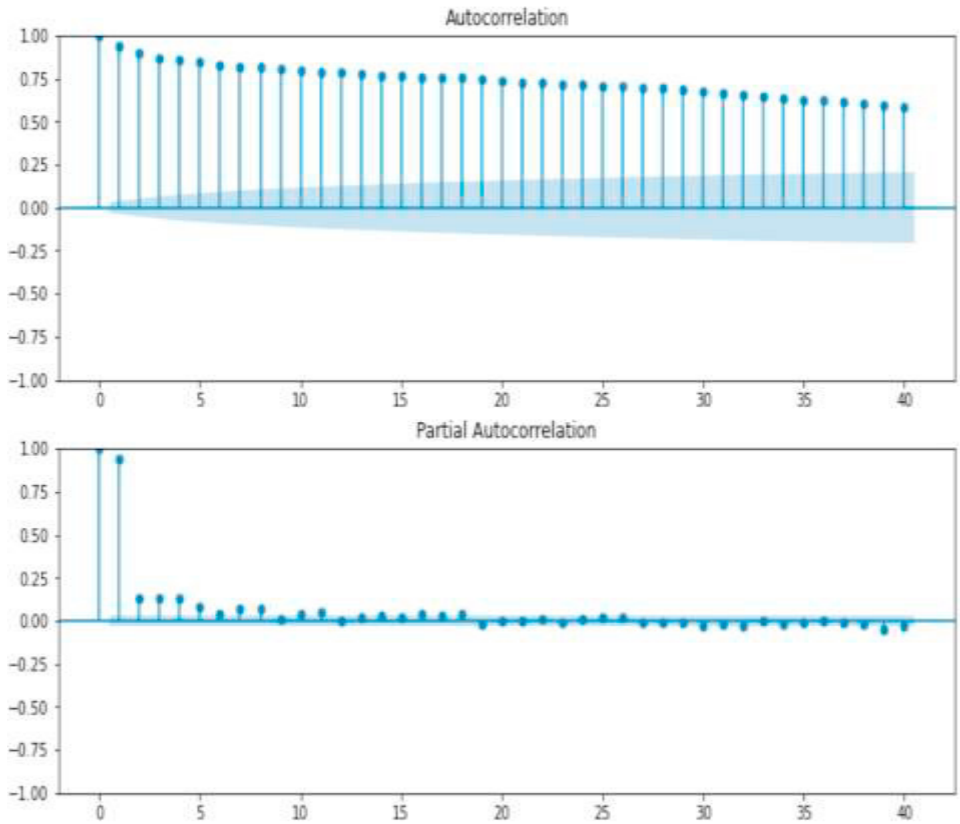

Fig. 6. Autocorrelation, Partial Correlation

Figure 6 shows that there are no negative spikes so we can q and Q as 0, there is no difference but there is seasonal difference so d as 0 and D as 1, in autocorrelation there are all positive spikes so we can take 40 but also, we can any number and there are 6 positive spikes in the partial Autocorrelation so P is 6 and s is the seasonal that is 12 months so we have taken as 12. Then we can apply the SARIMA model with p is 3, d is 0, q is 0, P is 6,D is 1,Q is 0,s is 12.

SARIMA model has skewness as -0.74, kurtosis as 5.82, heteroskedasticity (the condition in which the variance of the residual term in a regression model varies widely) as 1.23, Jarque Beta test (goodness of fit test of weather sample data that have the skewness and kurtosis matching normal distribution.) as 1699.52, Ljung-Box (test to determine whether time series' autocorrelations are different from zero out of a collection of them) as 0.8 and coefficient, standard error, z value of the variables in the SARIMA equation. Next, we can test the data by using the model for the input range 2022-01-01 to 2022-12-31 then predicted the temperature as per the dates and also, we have found the residuals by the difference between actual and predicted values.

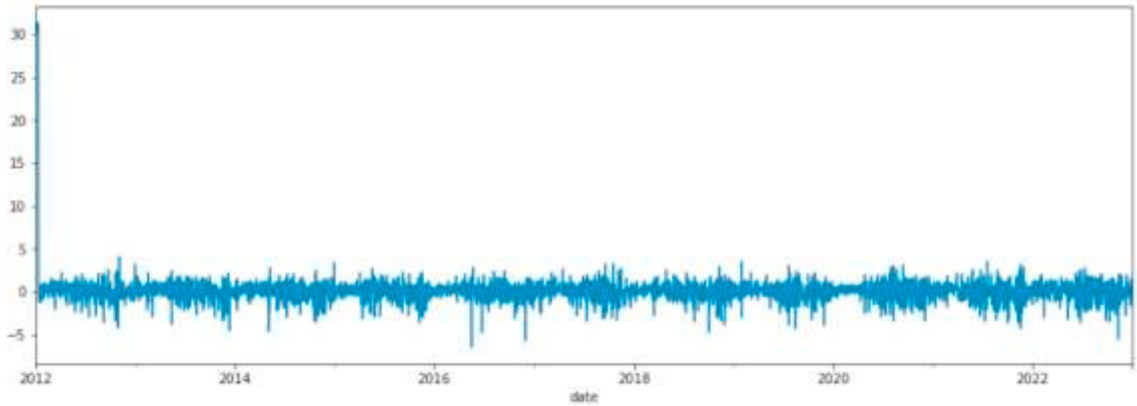

Fig. 7. Autocorrelation

Figure 7 discuss about the error or residuals of the data. We can see from the figure there are errors particularly in the middle of the year and also less errors in the beginning and ending of every year.

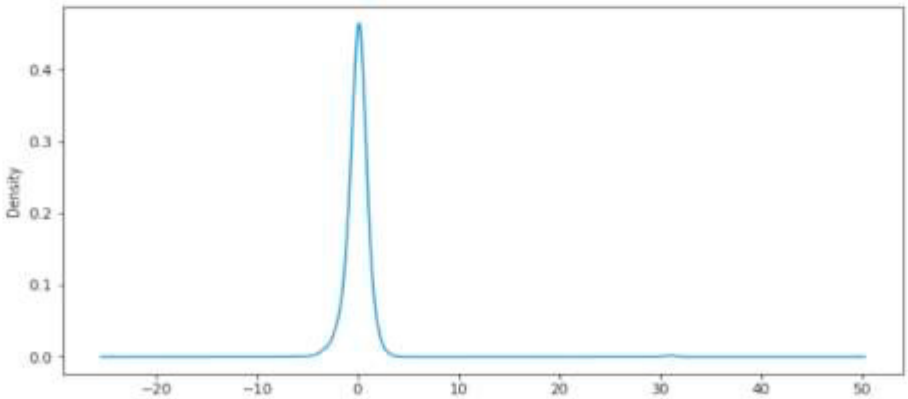

Fig. 8. Residual values plot

The above Figure 8 is residual values plot. By this graph we can say that residual values of the data is normally distributed by mean as 0 because this graph has the peak at the point 0. So, we can easily say that it is normally distributed.

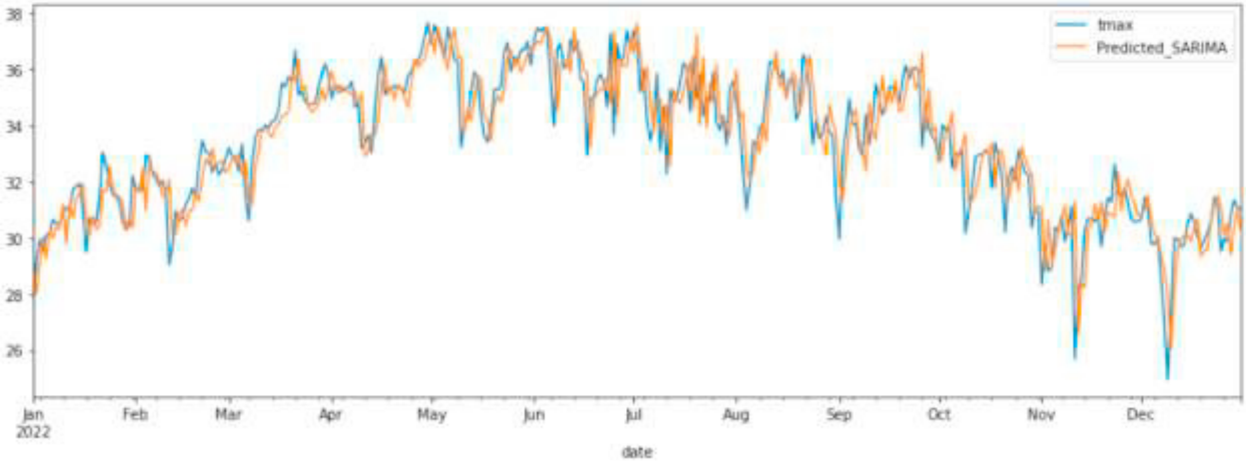

Fig. 9. Actual values of the test data and predicted values of the test data by SARIMA model

Lastly figure 9 discuss about the actual values of the test data and predicted values of the test data by SARIMA model. We can easily say that actual values and the predicted values are approximately same values. We can say error is 1%. So, we can conclude that SARIMA model is the good model to forecast the temperature.

7. Limitations towards Weather prediction

We can discuss limitations about weather forecasting/ predicting weather as**:**

7.1. In General

Weather prediction is a complex and challenging task due to its chaotic and dynamic nature of atmosphere [19]. Some of the limitations towards weather prediction include:

- Chaos Theory: The atmosphere is a complex, non-linear system that is sensitive to initial conditions. Even small errors in initial data can leads to significant differences in predicted weather patterns.
- Incomplete Data: Weather models rely on large amount of data, including humidity, pressure, temperature, wind speed and direction, and more. However, there are many areas of the world where weather data is incomplete or missing entirely.
- Computational Power: Available weather prediction models require large amount of computing power to process this large amount of data and perform complex calculations. While computing power has increased dramatically in recent years, it is still limited in its ability to accurately model complex weather patterns.
- Unforeseen Events: Weather prediction models are based on historical data and patterns, which may not accurately predict extreme weather events such as hurricanes, tornadoes, and thunderstorms.
- Physical Limitations: Weather prediction is limited by the laws of physics, which may not accurately model complex weather patterns such as turbulence, convection, and atmospheric instability.
- Time Constraints: Weather prediction models need to be completed in a timely manner, often within a few hours, to provide accurate forecasts. This time constraint can limit the ability of weather models to accurately predict changes in weather patterns over longer time frames.

In last, weather prediction models can produce large amounts of data, which must be interpreted and analyzed by meteorologists to produce accurate weather forecasts. Note that interpretation errors can lead to inaccurate predictions, which is not a precise way/ approach for weather forecasting.

7.2. Using Machine Learning Techniques

While machine learning has shown its important use/ role in many applications like weather prediction. But there are still some limitations that need to be considered [20 -25], are listed here as:

- Data availability: A lot of high-quality data is needed for ML algorithms to learn from. However, weather data can be limited, especially for certain regions or time periods. Also, weather data can contain errors or missing values, which can affect the performance of ML models.
- Complexity of weather systems: Weather prediction is a complex task, i.e., influenced by many factors, such as pressure, humidity, temperature, wind speed, and more. ML models may struggle to capture all of these factors and their interactions, especially for extreme weather events.
- Limited understanding of physical processes: ML models depend on statistical patterns to make predictions, but they may not fully capture the physical processes that govern weather phenomena, which lead to inaccurate predictions/ poor generalization to new situations.
- Changing weather patterns: Weather patterns can change rapidly, and ML models may not adapt quickly enough to keep up with these changes, which result in outdated or unreliable predictions.
- Interpretability: ML models may be difficult to interpret, which can make it challenging for meteorologists to understand why a particular prediction was made. This can limit their ability to make informed decisions based on the predictions.

In last, this work will be summarised with included few interesting remarks in brief.

8. Limitations of SARIMA Model for Weather Forecasting

As discussed above, SARIMA model is an important and efficient algorithm/ tool for time series forecasting, including weather forecasting, but it also has few limitations. We will understand these limitations, which will be good for using SARIMA effectively and for considering alternative approaches when necessary. Here are some limitations of SARIMA models in the context of weather forecasting:

- Assumption of Linearity: This model assumes that the underlying relationships in the time series data are linear. Weather data often exhibit nonlinear behaviour, especially during extreme weather events or sudden shifts in conditions. SARIMA may struggle to capture such nonlinear patterns.
- Sensitivity to Model Order Selection: This model requires the selection of appropriate orders for autoregressive (p), differencing (d), and moving average (q) components. Hence, selecting the right orders can be challenging, and different choices can lead to significantly different forecasts. Automated methods, like grid search or cross-validation, can help but are not foolproof.
- Limited Handling of Seasonal Patterns: While this model is designed to capture seasonal patterns, they may struggle with irregular or complex seasonality. Seasonal patterns in weather data can be influenced by various factors, and SARIMA model may not always capture these intricacies accurately.
- Data Quality Issues: This model is sensitive to data quality issues such as missing values and outliers. Weather data often have missing observations, especially in remote areas or during extreme weather events. Handling missing data appropriately can be challenging.
- Inability to Incorporate External Factors: This model depends on historical data and do not naturally incorporate external factors that can influence weather patterns, such as atmospheric pressure systems,

oceanic indices, or climate change trends. Incorporating these factors typically requires additional modelling steps.

- Limited Long-Term Forecasting: This model is more suited for short- to medium-term forecasting. When it comes to long-term climate predictions or forecasts beyond the observed historical data, SARIMA models may not perform well.
- Computational Intensity: This model can be computationally intensive, particularly when dealing with high-resolution or large datasets. This can limit their applicability in real-time or resource-constrained forecasting scenarios.
- Complexity of Extreme Events: Weather forecasting often involves predicting extreme events like hurricanes, which have complex dynamics. This model does not capture the intricate details and rapid changes associated with such events, and specialized models may be required.
- Lack of Spatial Information: This model work with time series data from a single location or station. They do not naturally account for spatial dependencies or variations across different geographic areas, which can be important in weather forecasting.
- Model Interpretability: This model is not always easily interpretable. Understanding the model's internal parameters and explaining forecasts to non-technical stakeholders can be challenging.

Hence, this models still remain valuable for many weather forecasting tasks, especially for short to medium time horizons and when seasonality and historical data play a significant role.

9. Future Work towards using SARIMA model for Weather Forecasting

This model has several promising directions for further research and development. Here are some potential future works and areas of improvement:

- Hybrid Models and Ensembles: We need to continue to explore the integration of this model with other forecasting techniques, including ML algorithms like neural networks and deep learning models. Also, hybrid and ensemble approaches can potentially capture complex weather patterns more effectively.
- Incorporation of Spatial Data: We need to extend SARIMA models to incorporate spatial dependencies by considering data from multiple weather stations or grid cells. Note that spatial SARIMA models can provide more accurate localized weather forecasts.
- Multivariate SARIMA: We need to develop multivariate SARIMA models that can handle multiple weather variables simultaneously. It will allow a basic understanding of the interrelationships between different meteorological factors.
- Real-Time Data Assimilation: We need to improve real-time data assimilation techniques to continuously update SARIMA models with the latest observations, satellite data, and other sources. It will enhance the model's ability to provide up-to-date forecasts.
- Extreme Weather Events Prediction: We need to focus on improving the prediction of extreme weather events, such as hurricanes, tornadoes, and heavy rainfall, by developing specialized SARIMA models or incorporating additional features and data sources.
- Long-Term Forecasting: We need to extend SARIMA models to handle long-term weather forecasting, such as seasonal and yearly predictions. It will benefit agriculture, water resource management, and climate research.
- High-Resolution Modeling: We need to develop SARIMA models capable of handling high-resolution data, including sub-hourly observations and fine-grained spatial data. It will be essential for urban and regional weather forecasting.

- User-Friendly Tools: We need to create user-friendly software tools and platforms that enable meteorologists and decision-makers to easily apply SARIMA models to their forecasting needs, with options for customization and real-time updates.
- Climate Change Adaptation: We need to investigate how SARIMA models can contribute to climate change adaptation by forecasting long-term climate trends, extreme weather events associated with climate change, and their impacts on various sectors.

Further research into probabilistic SARIMA models that can provide probabilistic weather forecasts, including quantifying uncertainties at different time horizons and spatial scales. Hence, this model plays an important role in weather forecasting, ongoing research and innovation in these areas which can lead to more accurate, reliable, and actionable weather predictions, ultimately benefiting society in various ways, from disaster preparedness to agriculture and infrastructure planning.

10. Conclusion

In conclusion, SARIMA model is a powerful ML technique for weather forecasting, used as a time-series model which can identify seasonal trends and patterns in weather data and provide precise forecasts for future time periods. Through the analysis of historical weather data and the application of SARIMA model, we can gain valuable information in the patterns and trends in weather patterns. This information can be used to develop effective strategies for managing weather-related risks and enhancing public safety. In terms of future work, there are several directions that could be pursued. Firstly, SARIMA model can be further optimized by incorporating more sophisticated ML methods like deep learning, ensemble models, or hybrid models. Secondly, incorporating more data sources, such as satellite data or weather station data, can lead to even more accurate predictions. Finally, the application of SARIMA model can be extended to other domains beyond weather forecasting, such as finance, marketing, or healthcare.

References

[1] Ardabili, S., Mosavi, A., Dehghani, M., & Várkonyi-Kóczy, A. R. “Deep learning and machine learning in hydrological processes climate change and earth systems a systematic review.” *International conference on global research and education*, (2019).

[2] Cifuentes, J., Marulanda, G., Bello, A., & Reneses, J. J. E. “Air temperature forecasting using machine learning techniques: a review.” 13(16), 4215, (2020).

[3] Derbentsev, V., Datsenko, N., Stepanenko, O., & Bezkorovainyi, V. “Forecasting cryptocurrency prices time series using machine learning approach.” *SHS Web of Conferences*, (2019).

[4] Narasimha Murthy, K.V., Saravana, R. & Vijaya Kumar K. “Modeling and forecasting rainfall patterns of southwest monsoons in North–East India as a SARIMA process.” *Meteorol Atmos Phys* **130**, 99–106 (2018)

[5] Peng Chen, Aichen Niu, Duanyang Liu, Wei Jiang, Bin Ma. “Time Series Forecasting of Temperatures using SARIMA: An Example from Nanjing.” *OP Conference Series: Materials Science and Engineering*, Volume 394, Issue 5, (2021).

[6] Mehmet Tektaş, “Weather Forecasting Using ANFIS and ARIMA MODELS.” Vol. 51 No. 1 (2010).

[7] Afan Galih Salman, Bayu Kanigoro, “Visibility Forecasting Using Autoregressive Integrated Moving Average (ARIMA) Models.” (2021).

[8] S. I. Vagropoulos, G. I. Chouliaras, E. G. Kardakos, C. K. Simoglou and A. G. Bakirtzis, "Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting," *IEEE International Energy Conference (ENERGYCON)*, pp. 1-6, (2016).

[9] Mohammad Valipour, “Long-term runoff study using SARIMA and ARIMA models in the United States.” Volume 22, Issue 3 (2019).

[10] Flávio Fonseca Nobre, Ana Beatriz Soares Monteiro, Paulo Roberto Telles, G. David Williamson, “Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology.” Volume20, Issue20 (2019).

[11] Tsan-Ming Choi, Yong Yu, Kin-Fan Au, “A hybrid SARIMA wavelet transform method for sales forecasting, Decision Support Systems.” Volume 51, Issue 1, (2011).

[12] M. Milenković, L. Švadlenka, V. Melichar, N. Bojović, and Z. Avramović, “SARIMA modelling approach for railway passenger flow forecasting”, *Transport*, vol. 33, no. 5, pp. 1113-1120, (2018).

[13] Garima Jain1, Bhawna Mallick2, “A Review on Weather Forecasting Techniques.” Vol. 5, Issue 12, (2016).

[14] Mohammed Farsi, Doreswamy Hosahalli, B.R. Manjunatha, Ibrahim Gad, El-Sayed Atlam, Althobaiti Ahmed, Ghada Elmarhomy, Mahmoud Elmarhoumy, Osama A. Ghoneim, "Parallel genetic algorithms for optimizing the SARIMA model for better forecasting of the NCDC weather data." *Alexandria Engineering Journal,* Volume 60, Issue 1, Pages 1299-1316, (2021).

[15] Yasaman Ensafi, Saman Hassanzadeh Amin, Guoqing Zhang, Bharat Shah. "Time-series forecasting of seasonal items sales using machine learning – A comparative analysis." *International Journal of Information Management Data Insights,* Volume 2, Issue 1, 100058, (2022).

[16] James H. Stock. "Time Series: Economic Forecasting." *International Encyclopedia of the Social & Behavioral Sciences (Second Edition), Elsevier,* Pages 337-340, (2015).

[17] Tsan-Ming Choi, Yong Yu, Kin-Fan Au. "A hybrid SARIMA wavelet transform method for sales forecasting." *Decision Support Systems,* Volume 51, Issue 1, Pages 130-140, (2011).

[18] Antonios Parasyris, George Alexandrakis,Georgios V. Kozyrakis. "Predicting Meteorological Variables on Local Level with SARIMA, LSTM and Hybrid Techniques." *Atmosphere*, Volume 13, Issue 6, (2022).

[19] Vijay Kumar Didal, Brijbhooshan, Anita Todawat, Kamlesh Choudhary. "Weather Forecasting in India: A Review." *International Journal of Current Microbiology and Applied Sciences*. Volume 6, Number 11, pp. 577-590, (2017).

[20] Chantry M, Christensen H, Dueben P, Palmer T. "Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft AI." *National Center for Biotechnology Information*. 379(2194):20200083, (2021).

[21] D. Goyal, R. Goyal, G.Rekha, S. Malik and A. K. Tyagi, "Emerging Trends and Challenges in Data Science and Big Data Analytics," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1-8, doi: 10.1109/ic-ETITE47903.2020.316.

[22] Akshita Tyagi, Swetta Kukreja, Meghna Manoj Nair, Amit Kumar Tyagi, Machine Learning: Past, Present and Future, Neuroquantology, Volume 20, No 8 (2022), DOI: 10.14704/nq.2022.20.8.NQ44468.

[23] Goyal, Deepti & Tyagi, Amit. (2020). A Look at Top 35 Problems in the Computer Science Field for the Next Decade. 10.1201/9781003052098-40.

[24] Varsha R., Nair S.M., Tyagi A.K., Aswathy S.U., RadhaKrishnan R. (2021) The Future with Advanced Analytics: A Sequential Analysis of the Disruptive Technology's Scope. In: Abraham A., Hanne T., Castillo O., Gandhi N., Nogueira Rios T., Hong TP. (eds) Hybrid Intelligent Systems. HIS 2020. Advances in Intelligent Systems and Computing, vol 1375. Springer, Cham. https://doi.org/10.1007/978-3-030-73050-5_56.

[25] Akshara Pramod, Harsh Sankar Naicker, Amit Kumar Tyagi, Emerging Innovations in the Near Future Using Deep Learning Techniques, Book: Advanced Analytics and Deep Learning Models, Wiley Scrivener, 2022, https://doi.org/10.1002/9781119792437.ch10