Temperature Forecasting and Analysis Using Linear and Timeseries Models

DR. S. K SINGH[1], RISHABH SINGH[2], SHREYASH SHERIGAR[3]

[1] *Head of Department, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India*

[2, 3] *PG Student, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India*

Abstract— Effective weather forecasting is a cornerstone in various sectors, ranging from agriculture and disaster management to climate research. Accurate predictions facilitate proactive decision-making and the mitigation of potential risks. This research paper presents a comprehensive investigation aimed at enhancing weather forecasting through a systematic exploration of linear and time series models. The study commences with meticulous data collection from diverse sources, encompassing weather stations, satellite observations, oceanic sensors, and historical records. Subsequent data preprocessing ensures data quality and integrity. Various climate parameters, including solar radiation, wind direction, and carbon emissions, are examined to refine forecasting accuracy and understand their intricate interactions. Christopher Vu et al [4]. and Mehmet et al. [3] in their research have proven the effectiveness of Linear and Time-series models building upon that, the initial modeling phase employs linear models, featuring linear regression and L1 regularization, known for their simplicity and effectiveness as our focused region is Mumbai metropolitan which does not experience extreme temperature changes. Moving forward, the research delves into time series models, notably ARIMA and its seasonal variation, SARIMA. These models are introduced to capture the temporal dynamics and dependencies inherent in climate data. The research highlights the suitability of specific models for different climate parameters. Linear models prove valuable for stable, less-variable parameters, while time series models excel in capturing intricate temporal dependencies. The shift from linear to time series models results in significantly enhanced accuracy in temperature predictions, emphasizing the need for tailored methodologies in data-driven projects. This paper contributes to the realm of weather forecasting and analysis by offering a structured approach to model selection, dataset preparation, and the interpretation of complex climate dynamics. It underscores the critical role of precise forecasting in promoting environmental sustainability and well-informed decision-making.

Indexed Terms— ARIMA, Linear Regression, Mumbai Metropolitan Region, Time Series, Weather Forecasting

I. INTRODUCTION

In the dynamic realm of weather forecasting, precision is the key factor, and predictions underpin vital decisions in various sectors, ranging from agriculture to disaster preparedness. Traditional forecasting methods, while valuable, often grapple with the intricate temporal dynamics that define climate data, limiting their accuracy and reliability. To enhance the precision of weather and temperature forecasts, our research explores the integration of advanced Artificial Intelligence (AI) and Machine Learning (ML) techniques, specifically time series and linear models. Our research aims towards addressing this challenge head-on by embracing the power of AI and ML. It begins with a rigorous transition from linear models, such as linear regression, to advanced time series models. These models, notably Autoregressive Integrated Moving Average (ARIMA) and Seasonal ARIMA (SARIMA), offer the ability to capture the complex temporal patterns within temperature data. This transition is pivotal in our pursuit of enhanced forecasting precision, where historical data is harnessed to predict future temperature variations.

While our initial explorations with linear models laid a strong foundation, it became evident that embracing the complexities of temperature fluctuations required more nuanced approaches. ARIMA and SARIMA introduce an element of temporal intelligence, enabling our forecasting models to recognize cyclic fluctuations and long-term trends that traditional methods cannot fully grasp. The outcome of this transition is a substantial improvement in temperature forecasting accuracy, as demonstrated by the reduced Mean Squared Error (MSE). These advanced models are not only advancing forecasting precision but also laying the groundwork for a deeper understanding of the intricate temporal dependencies inherent in weather and temperature data. As our research unfolds, we navigate a path to refining temperature forecasting, embracing AI and ML techniques that harness temporal nuances, bridging the gap between past observations and future predictions. This journey is not just about predicting temperatures; it's about ensuring the accuracy and reliability of forecasts that hold profound consequences for our societies and ecosystems. Our commitment to precision reflects our dedication to addressing the challenges posed by extreme weather events and advancing a sustainable, informed future.

II. LITERATURE REVIEW

In the realm of oceanography, ML techniques have also shown promise. Complex and nonlinear relationships within oceanic phenomena, such as El Niño-Southern Oscillation (ENSO), ocean currents, and sea surface temperatures, can be unraveled using algorithms like Support Vector Machines (SVMs), Random Forests, and Gradient Boosting (Chawla et al., 2001). Consequently, accurate ocean forecasting becomes attainable, enabling the prediction of events like harmful algal blooms, ocean acidification, and marine heatwaves.

Furthermore, the integration of real-time data sources through Internet of Things (IoT) devices, weather stations, buoys, and satellites hold great potential for enhancing weather and ocean forecasting systems (Minoli et al., 2017). By assimilating historical data with real-time observations, AI and ML models can adapt to changing conditions, leading to improved accuracy and responsiveness.

Mehmet [3], used ANFIS and ARIMA and Dataset was targeted towards the Göztepe, Istanbul region in turkey and the prime parameters include average temperature, wind-speed and pressure variables. 5% data was used in testing followed by 95% for training. ANFIS had a better yield as compared to ARIMA model in this scenario with a deviation in the range 0.1-0.5 within cost functions

Holmstrom M, et al. [4] incorporated Linear Regression and Variation of Functional Regression Dataset was targeted towards Stanford, CA (2011-2015). Prime parameters include maximum temperature, minimum temperature, mean humidity, mean atmospheric pressure and weather classification. They implemented a classification approach by reducing the features to 4 classification parameters. Linear regression was used to predict and summarize the maximum and minimum temperature. The model would predict a week's max and min temperature given 2 days prior data. Hence, the model failed to outperform professional weather forecasting services. Linear Regression performed slightly better than functional regression

Y. E. Cebeci (2019) [8] presents a study on an RNN model for weather forecasting. The paper emphasizes the challenges associated with traditional forecasting methods and showcases the effectiveness of RNNs in capturing complex weather patterns and improving forecasting accuracy.

Abhishek et al. (2012) [9] discusses a weather forecasting model using ANNs. The paper describes the architecture and training process of the ANN model and demonstrates its capability to capture patterns and make accurate predictions.

D. N. Fente and D. Kumar Singh (2018) [10] present a study on weather forecasting using ANNs. The paper explores the application of ANNs in predicting weather conditions and discusses the architecture and performance evaluation of the ANN model.

S. Kothapalli and S. G. Totad (2017) [11] focus on real-time weather forecasting and analysis. The paper proposes a system that integrates data from multiple sources to provide up-to-date and reliable weather forecasts, emphasizing the importance of real-time information for various sectors.

G. K. Rahul et al. (2020) [12] discuss weather forecasting using ANNs and highlight their potential in predicting weather conditions accurately. The paper presents the architecture and evaluation of an ANN model for weather forecasting.

A growing body of literature highlights the effectiveness of AI and ML in weather and ocean prediction. Deep learning models, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have demonstrated remarkable performance in analyzing large-scale weather data and extracting meaningful patterns (LeCun et al., 2015). These models have excelled in short-term weather prediction tasks such as rainfall estimation, temperature forecasting, and severe weather event detection.

III. PARAMETER ANALYSIS

This research places a strong emphasis on the quality and reliability of data, with a particular focus on weather data. The source of weather data, https://open-meteo.com, is known for its high quality, making it a foundation for trustworthy analysis and forecasting. The project's geographical focus is the Mumbai Metropolitan Region (MMR), a strategic choice to gain localized insights into the unique weather and patterns of this area. This specific focus allows for analysis and forecasting that are tailored to the MMR's needs, ensuring the outcomes are both accurate and relevant.

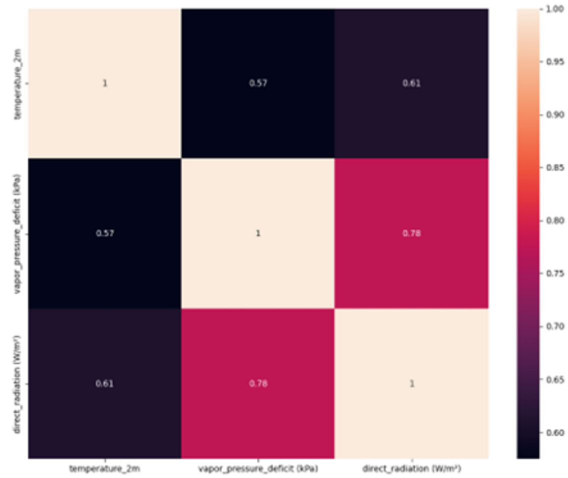

Figure 1. Co-Relation between Temperature, Vapor Pressure, Direct Radiation

The project's analytical rigor involves two crucial components: feature extraction and correlation analysis. Through rigorous investigation, "Vapor Pressure Deficit" and "Direct Radiation" are identified as primary parameters with a significant impact on temperature dynamics. They serve as reliable indicators of temperature range and provide valuable insights into temperature fluctuations. Correlation analysis delves into the relationships between various parameters in the dataset. Notably, it reveals that soil temperature exhibits the highest degree of correlation among the considered parameters. Importantly, it's recognized as an "effect" rather than a "cause," which is crucial for analytical precision. This distinction enables the differentiation between parameters that drive temperature changes and those that are simply correlated, improving the accuracy of the models. One intriguing discovery in the research reveals the weak correlation between greenhouse gas emissions and temperature variations. This finding challenges the initial expectation of a direct link between these factors. It prompts questions about the primary drivers of temperature changes and the influence of other variables in the dataset. This weak correlation serves as a starting point for further exploration.

IV. METHODOLOGY

i. Linear Models

In the temperature analysis project for the Mumbai region, linear regression was initially employed to estimate temperatures based on parameters like solar radiation, vapor pressure, wind direction, and time of day, producing a Mean Squared Error (MSE) of 3.1324. To address uncertainties and enhance forecasting, a novel integration of Monte Carlo simulation with linear regression was introduced, yielding an average temperature range of 26-27°C and an MSE of 13.342. Subsequently, the project incorporated the Random Forest algorithm, resulting in a lower MSE of 2.08 and a Root Mean Squared Error (RMSE) of 1.44, indicating improved forecasting accuracy. Additionally, Lasso regularization was implemented for predicting daily maximum temperatures, effectively mitigating overfitting, selecting relevant features, and improving model accuracy with an MSE of 1.8.

ii. Time Series Models.

In the quest for more accurate temperature predictions, the project transitioned from linear models to advanced time series models. This shift was prompted by the recognition that linear models had limitations in capturing complex temporal patterns and autocorrelation, which are often present in temperature data. The move towards time series modeling marked a crucial step in enhancing forecasting accuracy.

ARIMA for Time Series Forecasting

. The project's initial foray into time series modeling led to the adoption of ARIMA (AutoRegressive Integrated Moving Average), a powerful and versatile method tailored for handling time-dependent data. ARIMA's three key components, AutoRegressive (AR), Integrated (I), and Moving Average (MA), enabled it to capture intricate temporal patterns and improve accuracy. Implementing ARIMA led to a significant reduction in the Mean Squared Error (MSE), enhancing the accuracy of temperature forecasts. The implementation of ARIMA resulted in a substantial reduction in the Mean Squared Error (MSE) to 1.227, demonstrating an improvement in forecast accuracy.

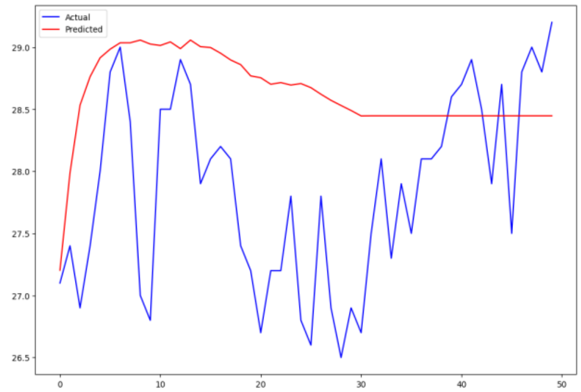

Figure 2. Temperature Prediction using ARIMA

ARIMAX: Incorporating Explanatory Variables

. Continuing its pursuit of improved forecasting, the project explored ARIMAX (AutoRegressive Integrated Moving Average with Explanatory Variables), an extension of ARIMA that integrates external variables. This addition allowed the model to consider factors such as solar radiation, vapor pressure, wind direction, and day/night indicators, leading to a substantial reduction in the MSE and further enhancing forecast accuracy. The introduction of ARIMAX (AutoRegressive Integrated Moving Average with Explanatory Variables) led to a significant reduction in the MSE to 0.3114, emphasizing the power of incorporating external variables to enhance forecast accuracy.

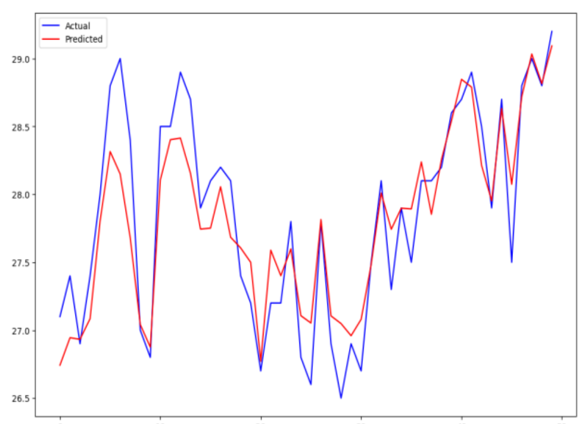

Figure 3. Temperature Prediction using ARIMAX

Seasonal ARIMA (SARIMA) for Seasonal Patterns

. Recognizing the importance of seasonality and recurring patterns in temperature data, the project embraced Seasonal ARIMA (SARIMA). SARIMA, an extension of ARIMA, is explicitly designed to handle time series data with distinct seasonal fluctuations and recurring patterns. This model excels at capturing both auto-correlations and seasonal correlations, resulting in an improved MSE and showcasing the significance of aligning modeling techniques with the data's temporal characteristics. The adoption of Seasonal ARIMA (SARIMA) further improved forecast accuracy, reducing the MSE to 0.5763, showcasing SARIMA's effectiveness in capturing seasonal patterns and temporal dependencies.

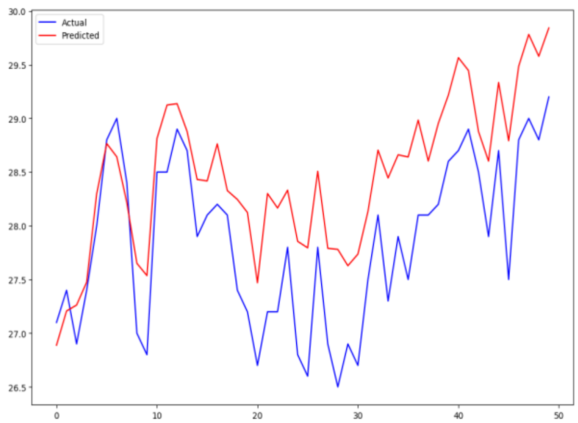

Figure 4. Temperature Prediction using SRIMAX

Ensembling Models for Unparalleled Accuracy
. In the pursuit of unparalleled forecasting accuracy, the project explored the art of ensembling, which combines the strengths of top-performing models, in this case, ARIMAX and SARIMA. By merging these models through a statistical function, the project harnessed their diverse insights to maximize overall forecasting performance. The collaborative efforts of these models in an ensemble approach resulted in a remarkable reduction in the MSE, underscoring the power of ensemble techniques to enhance prediction accuracy. The collaborative efforts of ARIMAX and SARIMA through an ensemble approach resulted in an exceptional reduction in the MSE to 0.223, underlining the power of ensemble techniques to enhance prediction accuracy.

V. RESULT

The project aimed to enhance temperature prediction accuracy in Mumbai by employing a range of models and techniques. Starting with Linear Regression, the project transitioned to advanced approaches including Monte Carlo Simulation, Random Forest, and LASSO regularization to mitigate overfitting. Recognizing the limitations of linear models in capturing complex temporal patterns, the project adopted time series models, beginning with ARIMA and later introducing ARIMAX and Seasonal ARIMA to handle seasonal patterns. The combination of ARIMAX and SARIMA through an ensemble approach yielded exceptional forecasting accuracy. These efforts resulted in a substantial reduction in Mean Squared Error (MSE), with the ensemble approach achieving the lowest MSE of 0.223, showcasing the power of these techniques in improving temperature forecasts tailored to Mumbai's climate and data characteristics.

Model	Description	MSE
Linear Regression	Predicting temperature in Mumbai based on climate parameters	3.1324
Monte Carlo Simulation	Incorporating randomness into forecasting process	13.342
Random Forest	An ensemble learning technique to improve forecasting accuracy	2.08
LASSO (L1 regularization)	To resolve overfitting and enhancing model accuracy	1.8
ARIMA	Capturing temporal patterns and improving temperature forecasting	1.227
ARIMAX	Integrating external variables for enhanced forecasting accuracy	0.3114
Seasonal ARIMA (SARIMA)	Handling seasonal patterns and improving forecast accuracy	0.5763
Ensembling Model	Combining ARIMAX and	0.223

	SARIMA for exceptional accuracy	

Table 1. Performance Comparison of Temperature Prediction Models.

CONCLUSION

The pursuit of enhancing weather and ocean forecasting for sustainable environmental management led us through a comprehensive journey, deploying a diverse array of predictive models, each tailored to specific aspects of the challenging task at hand. As we draw our conclusions, it's imperative to reflect on the cumulative insights and accomplishments gained from our in-depth analyses and the various modeling techniques employed throughout the project. In our quest to elevate temperature predictions, we initially employed linear models. Linear regression, with its foundational principles of relationship estimation, served as a robust starting point. However, we discerned certain limitations in capturing the intricate temporal dynamics inherent in weather and ocean data. These models excelled in scenarios where the temperature variations were relatively linear. Still, when we encountered non-linearity, they proved less effective. Hence, while linear models presented valuable foundations for straightforward forecasting, they found their boundaries in the complex world of climate prediction. Transitioning to time series models marked a pivotal moment in our project. ARIMA (AutoRegressive Integrated Moving Average), with its capacity to capture temporal dependencies, introduced a new dimension to our forecasting arsenal. It provided the power to address cyclic fluctuations and long-term trends that often-eluded linear models. However, ARIMA was not the sole star in our time series exploration. The introduction of seasonal ARIMA (SARIMA) further enriched our forecasting capabilities, offering an enhanced understanding of complex temporal patterns. Together, these models significantly improved temperature forecasting accuracy, demonstrating the importance of aligning modeling techniques with the inherent characteristics of the data.

The pursuit of precision led us to another compelling approach—model ensemble. The synergy of ARIMA and SARIMA, achieved through statistical functions, unveiled an ensemble model capable of providing temperature forecasts with exceptional accuracy. Model ensemble tactics, including mathematical functions for blending predictive outcomes, underpinned our ability to extract greater predictive power from the union of two outstanding time series models.

ACKNOWLEDGMENT

We extend our sincere gratitude to all those who have played an indispensable role in the realization of this project. We are indebted to our academic advisors and mentors, whose wisdom, guidance, and unwavering support have been our guiding light throughout this journey. Our heartfelt appreciation goes to the institutions and organizations that have provided access to data, resources, and collaborative opportunities. We would like to acknowledge Thakur College of Science and Commerce and CSIR – National Institute of Oceanography (NIO), Andheri west, for their invaluable contributions to our research. To our peers and colleagues, your insightful discussions and collaborative spirit have enriched this project and our understanding of the subject matter. We are grateful for the exchange of ideas and the collective pursuit of knowledge. Finally, we extend our profound gratitude to our families and friends, whose encouragement and patience have sustained us during this project. Your unwavering support has been the bedrock of our achievements. This project stands as a testament to the power of collective effort, and it is with sincere appreciation that we acknowledge the contributions of all those who have been a part of this endeavor.

REFERENCES

[1] The World Bank. (2023). World Bank Open Data. Retrieved from [https://data.worldbank.org/]

[2] Open Meteo. (2023). Free Weather API. Retrieved from [https://open-meteo.com]

[3] S. Tektaş, "Weather forecasting using ANFIS and ARIMA models," “Environmental Research, Engineering and Management”, vol. 51, no. 1, pp. 5-10, Apr. 2010.

[4] M. Holmstrom, D. Liu, and C. Vo, "Machine learning applied to weather forecasting,"

Meteorological Applications, vol. 10, pp. 1-5, Dec. 2016.

[5] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "SMOTE: Synthetic minority oversampling technique," “Journal of Artificial Intelligence Research”, vol. 16, pp. 321-357, 2001.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," “Nature”, vol. 521, no. 7553, pp. 436-444, 2015.

[7] D. Minoli, K. Sohraby, and B. Occhiogrosso, "IoT considerations, requirements, and architectures for smart buildings—Energy optimization and next-generation building management systems," “IEEE Internet of Things Journal”, vol. 4, no. 1, pp. 269-283, 2017.

[8] Y. E. Cebeci, "A recurrent neural network model for weather forecasting," in “2019 4th International Conference on Computer Science and Engineering (UBMK)”, Samsun, Turkey, 2019, pp. 591-595.

[9] K. Abhishek, M. P. Singh, S. Ghosh, and A. Anand, "Weather forecasting model using artificial neural network," “Procedia Technology”, vol. 4, pp. 311–318, 2012.

[10] D. N. Fente and D. Kumar Singh, "Weather forecasting using artificial neural network," in “2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)”, Coimbatore, India, 2018, pp. 1757-1761.

[11] S. Kothapalli and S. G. Totad, "A real-time weather forecasting and analysis," in “2017 IEEE International Conference on Power, Control, Signals, and Instrumentation Engineering (ICPCSI)”, Chennai, India, 2017, pp. 1567-1570.

[12] G. K. Rahul, S. Singh, and S. Dubey, "Weather forecasting using artificial neural networks," in “2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)”, Noida, India, 2020, pp. 21-26.