|  |
| --- |
| **Team no –12** |

Submitted by-

|  |  |
| --- | --- |
| 1. Harshal Sanjiv Patil | (50604782) |
| 1. Mrudula Chandrakant Deshmukh | (50605669) |
| 1. Siddhi Sunil Nalawade | (50613176) |

**INTRODUCTION: -**

This project aims to compare two well-known methods for predicting short-term temperature changes: ARIMA, a traditional statistical approach, and LSTM, a type of machine learning model. Predicting temperatures is important for weather forecasting and will offer valuable insights for industries like agriculture, air travel planning, and transportation etc.

This project uses a large collection of daily temperature records from many U.S. cities, covering several decades. This detailed data provides a strong basis for testing how well each method works at predicting short-term temperature changes and understanding their strengths and weaknesses.

By comparing these two approaches, we aim to evaluate their accuracy and reliability under different conditions, explore potential hybrid solutions, and analyse the practical applications of each method in real-world scenarios. Our approach includes processing the data, models’ implementations, comparing their performance, and analysing the results. The results of this comparison will give useful information for industries that need temperature forecasts, like farming, air travel, city planning, and transportation. This can help these industries make better decisions and run more smoothly.

**Proposal: -**

We aim to enhance the accuracy of short-term temperature forecasts by developing and testing two predictive models: ARIMA and LSTM. Our objective is to compare these models to determine which provides more reliable forecasts for short-term weather predictions, thereby improving decision-making in weather-dependent industries (Heitmeier’s 1st).

Current weather forecasting relies on a mix of traditional statistical models and advanced machine learning techniques. ARIMA models are renowned for handling linear trends and seasonal patterns but struggle with capturing non-linear, chaotic weather variations, limiting their effectiveness for localized forecasts [1] In contrast, LSTM models excel at detecting complex, non-linear patterns in data, making them more suitable for chaotic weather predictions. However, LSTM models demand substantial computational resources and large datasets, which can restrict their use in localized settings [2]; [1].

To overcome these limitations, we will explore hybrid models that combine ARIMA’s ability to model linear trends with LSTM's capacity to handle non-linear dependencies. These hybrid approaches have demonstrated improved accuracy over using either model independently [1]; [3](Heitmeier’s 2nd)

Our approach involves directly comparing ARIMA and LSTM models for short-term temperature forecasting using a comprehensive dataset of daily temperature records from various U.S. cities. This allows us to assess the strengths of both traditional statistical methods and modern deep learning techniques. ARIMA is well-suited for capturing linear trends and seasonal patterns in weather data, while LSTM excels in modelling non-linear, complex temporal dependencies [4]; [5]

We believe this approach will be successful due to the extensive historical data available, which will enable thorough training and testing of both models. By comparing both methods on the same dataset, we can determine which model performs best under different conditions [6] Additionally, we plan to investigate a hybrid ARIMA-LSTM model, inspired by the work of [1], which showed that merging these models could significantly improve prediction accuracy compared to using either one alone.

The hybrid approach, leveraging ARIMA’s strength in capturing linear trends and LSTM’s capacity to handle non-linear patterns, has been shown to outperform standalone models in various studies [1]; [3]Combining these methods allows us to exploit their respective strengths, providing a more accurate and robust forecasting solution, particularly for short-term temperature predictions (Heilmeier's 3rd).

This research on short-term temperature forecasting has significant implications for various sectors. While meteorologists and climate scientists directly benefit, industries such as energy, agriculture, transportation, and local government rely heavily on accurate temperature forecasts. For instance, power companies use these forecasts to manage supply and demand, especially during extreme weather, while farmers rely on them for informed decisions about planting and irrigation to enhance crop yields [6]; [1] [5]Ultimately, improving temperature forecasting can lead to better public safety, resource management, and cost savings across various industries [2]; [7]

Moreover, reliable weather predictions are crucial for the aviation industry to ensure flight safety and scheduling, enabling airlines to avoid hazardous conditions [1]. Local governments depend on precise forecasts for emergency preparedness, allowing for effective resource allocation during extreme weather events [4] Additionally, construction and retail sectors benefit from accurate forecasts to optimize operations and reduce financial losses [3]; [8]Overall, enhancing temperature forecasting supports various industries in making informed decisions and improving operational efficiency (Heitmeier’s 4th).

Success in this project would lead to more accurate short-term temperature forecasting, significantly aiding decision-making in agriculture, energy, airport flight management and emergency management. By reducing prediction errors, we can assist these industries in planning better and utilizing their resources more effectively. For example, [6] found that LSTM models can cut prediction errors by 84-87% compared to ARIMA.

To measure our success, we will utilize common metrics such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) to evaluate the performance of ARIMA, LSTM, and hybrid models against standard forecasts. These metrics are widely used in the field, as demonstrated by [1], where RMSE effectively tracked improvements in accuracy for hybrid ARIMA-LSTM models. Improved forecasts can have a significant impact across various sectors. [5]emphasize that accurate weather predictions are essential for agriculture, transportation, and other industries (Heitmeier’s 5th).

The main risks of this project arise from the computational demands of LSTM models and the potential for overfitting. [1] highlight the complexity of running LSTM models, especially with large datasets, which can require significant computational resources. [6] further note that LSTM models may overfit when trained on limited or low-quality datasets, leading to inaccurate predictions. Additionally, there is a risk that our models may not generalize well across different geographic regions or time periods, limiting the broader applicability of the findings [5]

While ARIMA is simpler to implement and computationally less demanding, it struggles with non-linear and complex patterns in weather data [5] [4]. This could lead to limitations in ARIMA’s effectiveness for short-term temperature forecasting, particularly in rapidly changing or chaotic weather conditions.

However, the potential payoffs are substantial. By improving the accuracy of short-term temperature forecasts, this project could benefit industries such as agriculture, where better forecasts assist with planting and harvesting [5], and the energy sector, which can better manage supply and demand during extreme weather [6]; [1]. [1] demonstrated that combining ARIMA and LSTM models can yield more accurate results than using either method alone, suggesting a strong payoff if the hybrid approach is successful.

Furthermore, clarifying the strengths and weaknesses of ARIMA and LSTM will contribute to developing more robust forecasting models [6] found that LSTM significantly reduces forecasting errors compared to ARIMA, indicating that this project could lead to advances in weather prediction techniques and practical applications, especially if we successfully integrate the strengths of both models [3](Heitmeier’s 6th).

The project’s costs will mainly involve computational resources for processing data and training the models, particularly for LSTM, which is computationally demanding [1]. To keep expenses low, we will utilize freely available weather datasets and open-source tools, along with R libraries for implementing the models [6](Heitmeier’s 7th).

We expect the project to take approximately two months, beginning with learning the R programming language and its statistical packages, followed by familiarization with ARIMA and LSTM models for time series forecasting. Subsequent steps will involve data collection and preparation, followed by model development and testing. Finally, we will analyse the results and refine the models as needed, concluding with documentation of findings and project report preparation [9] (Heitmeier’s 8th).

Midterm evaluations will assess the completion of data collection and pre-processing, alongside the implementation of initial versions of the ARIMA and LSTM models. For the final assessment, we will compare the models' accuracy using RMSE and MAE and evaluate the performance of the hybrid model. The goal is to identify the most effective approach for short-term temperature forecasting based on various metrics [1] [6] (Heitmeier’s 9th)

**Plan of Activities :-**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Task** | **Assigned Members** | **Start Date** | **End Date** | **Duration (Days)** |
| Learning R programming language and tools | All Members | 10/20/2024 | 10/22/2024 | 3 |
| Familiarization with ARIMA and LSTM models | All Members | 10/23/2024 | 10/24/2024 | 2 |
| Data collection and preparation | Member 1 & 2 | 10/25/2024 | 10/31/2024 | 7 |
| ARIMA Model development | Member 1 | 11/01/2024 | 11/04/2024 | 4 |
| LSTM Model development | Member 2 | 11/05/2024 | 11/09/2024 | 5 |
| Hybrid Model development | Member 3 | 11/10/2024 | 11/14/2024 | 5 |
| Testing and evaluating models (ARIMA, LSTM, Hybrid) | Member 1 & 3 | 11/15/2024 | 11/21/2024 | 7 |
| Result analysis and model refinement | All Members | 11/22/2024 | 11/30/2024 | 9 |
| Final report and documentation preparation | All Members | 12/01/2024 | 12/05/2024 | 5 |

**References:**

|  |  |
| --- | --- |
| [1] | [2. H. 2. A. a. 3. S. Afan Galih Salman, “Weather Forecasting Using Merged Long Short-Term Memory Model (LSTM) and Autoregressive Integrated Moving Average (ARIMA) Model,” *Journal of Computer Science,* 2018](https://thescipub.com/abstract/10.3844/jcssp.2018.930.938). |
| [2] | [Y. Q. Bangyu Li, “Weather prediction using CNN-LSTM for time series analysis: A case study on Delhi temperature data.,” *International Conference On Computing and Data Science,* 2024.](https://www.arxiv.org/abs/2409.09414) |
| [3] | [N. A. ,. J.-M. M. ,. M. O. P. C. Wissam Abdallah, “A Hybrid Methodology for Short Term Temperature Forecasting,” *International Journal of Intelligence Science ,* 2020.](https://www.scirp.org/journal/paperinformation?paperid=100852) |
| [4] | [I. M. M. G. a. J. K. Małgorzata Murat1, “Forecasting daily meteorological time series using ARIMA and regression models,” *Institute of Agrophysics, Polish Academy of Sciences,* 2018.](http://www.international-agrophysics.org/Forecasting-daily-meteorological-time-series-using-ARIMA-and-regression-models,104317,0,2.html) |
| [5] | [L. R. Eranga De Saa, “Comparison between ARIMA and Deep Learning Models for Temperature Forecasting,” *eprint arXiv:2011.04452,* 2020.](https://arxiv.org/abs/2011.04452) |
| [6] | [S. Siami-Namini, N. Tavakoli and A. S. Namin, “A Comparison of ARIMA and LSTM in Forecasting Time Series,” *IEEE International Conference on Machine Learning and Applications (ICMLA),* 2018.](https://ieeexplore.ieee.org/document/8614252) |
| [7] | [T.-L. Nghiem, V.-D. Le, T.-L. Le, P. Maréchal, D. Delahaye and A. Vidosavljevic, “Applying Bayesian inference in a hybrid CNN-LSTM model for time-series prediction,” *International Conference on Multimedia Analysis and Pattern Recognition (MAPR),* 2022.](https://ieeexplore.ieee.org/document/9924783) |
| [8] | [S. S. ·. L. Y. P. ·. Y. R. K. L. P. ·. S. Jitte, “Enhancing Forecasting Accuracy with a Moving Average-Integrated Hybrid ARIMA-LSTM Model,” *SN Computer Science,* 2024.](https://ieeexplore.ieee.org/document/9924783) |
| [9] | [L. W. A. Z. P. W. X. Y. Jingyan Ye Xuezhe Gao, “Prediction Of Global Temperature Based on ARIMA And GM Model,” *ASIA CONFERENCE ON GEOLOGICAL RESEARCH AND ENVIRONMENTAL TECHNOLOGY,* 2023.](https://drpress.org/ojs/index.php/HSET/article/view/12056) |