|  |
| --- |
| **Team no –12** |

Submitted by-

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

**INTRODUCTION: -**

The goal of this project is to evaluate and compare the performance of two popular time series forecasting methods, ARIMA (a traditional statistical approach) and LSTM (a deep learning-based model), for short-term temperature prediction. The motivation behind this study lies in the importance of accurate weather forecasting for various sectors such as agriculture, energy management, and transportation. We also explore the potential of a hybrid ARIMA-LSTM model to combine the strengths of both approaches.

The dataset used for this project contains historical daily temperature records from multiple U.S. cities spanning several decades. This comprehensive dataset allows us to analyze seasonal patterns, detect anomalies, and implement robust forecasting techniques.

**Literature Survey: -**

Our research builds on the findings from several key papers:

1. ARIMA's Strength in Linear Forecasting: Studies have consistently shown ARIMA's robustness in handling linear trends and seasonality in time series data​​.
2. LSTM for Non-linear Dependencies: LSTM models are highly effective in capturing complex temporal relationships in weather data, outperforming traditional methods in scenarios with non-linear variations​​.
3. Hybrid Approaches: Combining ARIMA and LSTM models has demonstrated improved accuracy over standalone methods, leveraging ARIMA's ability to capture linear patterns and LSTM's strength in handling residual non-linearities​​.

These findings guide our implementation and evaluation strategy.

**Experiment and Evaluation: -**

Progress so Far

1. Data Preparation:

1. The dataset was pre-processed to remove missing values and filtered to focus on individual cities for initial analysis (e.g., Albuquerque).
2. Data was transformed into time series objects with monthly frequency for compatibility with ARIMA modelling.

2. ARIMA Model:

1. Stationarity was checked using the Augmented Dickey-Fuller test.
2. The auto.arima() function was used to select the optimal parameters for the model.
3. Forecasts for the next 12 months were generated and visualized.

3. LSTM Model:

1. Data was scaled to a range of 0–1 to ensure better performance in LSTM training.
2. The data was reshaped into a 3D array format for LSTM input (samples, timesteps, features).
3. A sequential LSTM model was designed with two LSTM layers followed by a dense output layer.
4. The model was trained using the Adam optimizer for 50 epochs with a batch size of 32.

Preliminary Results:

ARIMA Performance:

1. Root Mean Square Error (RMSE): 2.45 (preliminary).
2. ARIMA performed well in capturing seasonal trends but struggled with abrupt non-linear variations.

LSTM Performance:

1. RMSE: 1.90 (preliminary).
2. The LSTM model provided more accurate predictions, especially for non-linear trend

**Plan of Study: -**

Next Steps:

1. Fine-tune LSTM hyperparameters (e.g., number of layers, units, learning rate).
2. Develop and test the hybrid ARIMA-LSTM model by combining ARIMA's predictions for trends with LSTM's predictions for residuals.
3. Evaluate all models using metrics such as RMSE, MAE, and MAPE for a comprehensive comparison.
4. Generate additional visualizations to illustrate model performance and residual patterns.

Timeline:

|  |  |  |  |
| --- | --- | --- | --- |
| Task | Start Date | End Date | Duration(Days) |
| Data Collection and prepressing | Completed | Completed | N/A |
| ARIMA implementation | Completed | Completed | N/A |
| LSTM model training | On – going | Nov 25,2024 | 10 |
| Hybrid model development | Nov 26 ,2024 | Dec 1,2024 | 6 |
| Final evaluation and reporting | Dec 2,2024 | Dec 5,2024 | 4 |

**Discussions: -**

Challenges Faced:

1. Computational demands of LSTM: Training the LSTM model requires significant computational resources, especially with longer time series data.
2. Overfitting in LSTM: Preliminary results indicate overfitting during training, which we plan to address using regularization techniques (e.g., dropout layers).
3. ARIMA's Limitations: The ARIMA model's inability to handle non-linear patterns limits its accuracy for certain temperature variations.

Innovations and Contributions:

1. Combining ARIMA and LSTM models to leverage their respective strengths.
2. Developing a reproducible pipeline for time series analysis that integrates statistical and deep learning methods.
3. Providing actionable insights into the suitability of forecasting models for real-world applications.

**Effort Distribution: -**

All team members have contributed equally to the project:

1. Harshal Sanjiv Patil: ARIMA implementation and analysis.
2. Mrudula Chandrakant Deshmukh: LSTM model development and training.
3. Siddhi Sunil Nalawade: Data pre-processing, hybrid model design, and reporting

**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) |