Time Series Analysis

Short-term temperature forecasting using ARIMA and LSTM models

Team no. 12

Submitted by-

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

INTRODUCTION: -

The aim of the on-going study is to differentiate the raising demand for precise weather forecasts in numerous areas, that includes transportation, management, urban development, energy, and agriculture that motivated this research.

Data preprocessing, individual ARIMA and LSTM model implementation, hybrid ARIMA-LSTM model development, comparison analysis, and testing across several time scales and climatic circumstances are all part of the methodology. This strategy seeks to improve temperature forecasting methods by providing a more complex way to foresee short-term variations with repercussions, for the companies that rely on accurate temperature forecasts.

The technique involves collecting the data, implementing the model, comparing the results, and thoroughly assessing the process. The goal of the study is to focus on forecasting method's precision, dependability, and effectiveness in a range of scenarios. This thorough examination will shed light on the usefulness of various prediction techniques in actual situations. The study looks at the advantages and disadvantages of each strategy to provide useful advice for choosing the best forecasting instruments for various industries and environmental circumstances.

The findings of this research have broad ramifications. By identifying the forecasting systems that perform best under given conditions, we can assist weather-affected businesses in making better judgments. This could lead to faster vitality dispersion, better agrarian edits administration, more successful asset assignment, and safer urban planning and transportation methods. Our goal with this investigation is to see through the advanced machine learning approaches and traditional factual procedures in temperature estimation, that results in more precise and reliable short-term climate forecasts.

Literature Survey: -

Focusing on ARIMA models, LSTM models and hybrid methods, the study focuses on recently developed methods on time series forecasting and weather prediction.

- Linear Forecasting Model of ARIMA:
  ARIMA models properly captured the seasons and regularities in time series data. Murat et al. demonstrated how well the ARIMA forecasted daily weather data, obtaining a mean absolute percentage error (MAPE) of 3.2% for temperature predictions in the 2018 research analysis. Ye et al. (2018), however, emphasized ARIMA's shortcomings when dealing with non-linear data, hence demonstrating the necessity for more sophisticated methods.

- LSTM Networks for Dependencies:
  Nonlinear time patterns in weather forecasts can be well represented by LSTM models (2018) found that using LSTM instead of ARIMA increases the RMSE by 84% to 87% for various time series forecasting operations. De Sa & Ranatunga (2021) further demonstrated the superiority of LSTM, with an overall RMSE 15% lower than ARIMA, especially in situations with rapid temperature fluctuations. Nonlinear time patterns in weather forecasts can be well represented by LSTM models (2018) found that using LSTM instead of ARIMA increases the RMSE by 84% to 87% for various time series forecasting operations. Later, De Sa & Ranatunga (2021) demonstrated the advantages of LSTM, in situations with rapid temperature fluctuations with an overall RMSE 15% lower than ARIMA.

- Hybrid ARIMA-LSTM Approaches:
  Recent studies have shown promising results when combining ARIMA and LSTM models, compared to a standalone model we demonstrated that a hybrid ARIMA-LSTM model diminishes RMSE by 20%. Abdullah et al (2020) enhanced this approach, demonstrating a 25% improvement in forecast accuracy compared to traditional methods across a range of weather parameters compared to conventional methods.

- Research Gaps and Our Contributions:
  Our assessment of the literature revealed numerous notable research gaps, including a lack of research into hybrid models in different geographical regions and a disdain for scalability. Our study intends to solve these gaps by constructing and analysing a hybrid ARIMA-LSTM model in a variety of climates across the United States, using a scalable technique for big datasets, and giving a comprehensive comparison of ARIMA, LSTM, and hybrid models for short-term temperature forecasting.

By these ideas our approach to implementation and evaluation has been developed.

Experiment and Evaluation: -

Progress so Far

1. Data Preparation:

- Loaded and processed the dataset (US_City_Data.csv).
- To start, Albuquerque and other randomly chosen cities were included in the data.
- A time series item for each month between January 1948 and 2022 was constructed. The time column was changed to date format to facilitate precise time series analysis and model training.

2. Exploratory Data Analysis (EDA):

- Analysed the summary data which used the str (), head (), and summary () methods to analyse the structure of the dataset.

- Base R and ggplot2 were used to create visuals that showed the temperature trends of Albuquerque throughout time.
- Clear seasonal patterns and a possible long-term warming trend were revealed be the time series plot.

3. ARIMA Model Implementation:

- Checked stationarity using the Augmented Dickey-Fuller (ADF) test:
  Results: Dickey-Fuller = -11.841, p-value = 0.01, indicating stationarity.
- Applied first-order differencing and re-tested:
  Results: Dickey-Fuller = -38.015, p-value = 0.01, confirming improved stationarity.
- Used auto. Arima () function to select optimal parameters for the ARIMA model.
- The best model was identified as ARIMA (2,0,1) (2,1,0) with drift.

4. Preliminary Results:

ARIMA Model Performance:

- Mean Error (ME): 0.005223061
- Root Mean Square Error (RMSE): 1.464187
- Mean Absolute Error (MAE): 1.115177
- Mean Absolute Scaled Error (MASE): 0.8001062

As indicated by the seasonal components in the ARIMA model, it captured seasonal trend effectively.

5. Work in Progress:

The LSTM model is currently being implemented.

- LSTM and ARIMA comparative study between these models are scheduled.
- The formation of an ARIMA-LSTM hybrid model is being observed.

Working on the experiment and testing a basic ARIMA model for short-term temperature forecasting has been successfully built. The initial results indicate encouraging performance in identifying general temperature patterns as well as seasonal trends. The LSTM implementation will be finished, and thorough performance comparison of the models will be carried out.

Study Plan:

The next steps include:

1. LSTM Model Improvement:
   Performance improves due to tuning of hyperparameter which involves changing the number of units and the learning rate. The layers for dropout are appended to reduce overfitting and new designs resembling bidirectional and stacked LSTMs will be examined as well.

2. Development of Hybrid Models:
   To maximize efficiency, several integration procedures are examined, including LSTM for residuals, weighted normalization and ARIMA.

3. All-inclusive Evaluation Strategy:
   Apart from RMSE other parameters can be utilized to determine performance over various predicted periods, such as one month, four months, and eight months. Used Time series cross-validation to increase the accuracy and robustness of model evaluations.

4. Multiple cities Analysis:
   Broadening the study and including several US cities to compare model performance among regions and their climatic changes. This will judge the model's ability for changing conditions.

5. Advanced visualization:
   Forecast accuracies, residual analyses and model comparisons are displayed using ACF and PACF plots. Additionally, forecast uncertainty will be demonstrated using methods such as fan charts.

6. Statistical Testing:
   Use the Diebold-Mariano test to assess forecast accuracy among models.

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
Multi-city Analysis	Dec 2,2024	Dec 4,2024	3
Final evaluation and reporting	Dec 5,2024	Dec 7,2024	3

Each team members have made equal contributions to the project:

- Harshal Sanjiv Patil: manages data pre-processing, ARIMA implementation, multi-city analysis, and statistical testing.
- Mrudula Chandrakant Deshmukh: focuses in creating and improving LSTM models.
- Siddhi Sunil Nalawade: focuses on data visualization, hybrid model creation, and reporting.

Discussions: -

Challenges Faced:

1. Data Processing and Model Implementation:
   The initial ARIMA implementation struggled with non-linear patterns, as evidenced by performance measures (RMSE: 1.464, MAE: 1.115). Meanwhile, scaling LSTM models to many cities and longer time horizons presents computational hurdles.
2. Technical challenges:
   Include addressing overfitting in the LSTM model with regularization approaches and guaranteeing correct time series cross-validation while preserving temporal relationships in data.

Innovation and Contributions:

1. New Hybrid Architecture:
We developed a hybrid ARIMA-LSTM model that combines ARIMA's capacity to capture linear trends and seasonality with LSTM's ability to predict non-linear patterns.
2. Multi-City Analysis:
Developed a systematic approach to evaluate model performance across various geographical areas and patterns.
3. Improved Methodology:
Metrics like RMSE, MAE, and MASE are used for accurate model evaluation.
4. Practical application:
The study provides information on temperature forecasts that can be applied practically in energy management, transportation, and agriculture. By providing predictions more precise and reliable, the hybrid approach can help in making decisions in real-world circumstances

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.

[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.

[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.

[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.

[5] L. R. Eranga De Saa, "Comparison between ARIMA and Deep Learning Models for Temperature Forecasting," *eprint arXiv:2011.04452,* 2020.

[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.

[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.

[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.

[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.