Prediction of Daily Climate Using Long Short-Term Memory (LSTM) Model

Jinxin Xu[1*], Zhenglin Li[2], Shaojie Li[3], Zichao Li[4], Yuhong Mo[5]

1 Department of Cox Business School, Southern Methodist University, Dallas, TX, USA

2 Department of Software Engineering, Sichuan University, Chengdu, Sichuan, China

3 Huacong Qingjiao Information Technology (Beijing) Co., Ltd., Beijing, China

4 Canoakbit Alliance Inc., Oakville, Canada

5 Department of Electrical and Computer Engineering, Carnegie Mellon University, PA,

*Corresponding author:[Jinxin Xu, E-mail:jensenjxx@gmail.com] USA

Abstract: Climate prediction plays a vital role in various sectors, including agriculture, disaster management, and urban planning. Traditional methods for climate forecasting often rely on complex physical models, which require substantial computational resources and may not accurately capture local weather patterns. This study explores the potential of Long Short-Term Memory (LSTM) networks, a type of recurrent neural network, for predicting daily climate variables such as temperature, precipitation, and humidity. Utilizing historical climate data from the city of Delhi, we developed an LSTM model to forecast short-term climate trends. The model consists of two LSTM layers followed by three Dense layers and is compiled with the Adam optimizer, mean squared error loss, and mean absolute error as a metric. Our results demonstrate the model's capability to capture temporal dependencies in climate data, achieving a satisfactory level of accuracy in temperature forecasting. This research underscores the potential of machine learning techniques, particularly LSTM networks, in enhancing climate prediction and contributing to more informed decision-making in weather-sensitive sectors.

Keyword: machine learning, prediction model, time series forecasting, long short-term memory

I. Introduction

Climate prediction is a critical area of research that has far-reaching implications for agriculture, disaster management, and urban planning. The ability to forecast weather conditions accurately is vital for farmers to plan their crop cycles, for disaster response teams to prepare for extreme weather events, and for city planners to design infrastructure that can withstand climatic changes. Traditional methods for climate forecasting rely heavily on complex physical models that simulate atmospheric and oceanic processes. These models incorporate various factors such as wind patterns, ocean currents, and solar radiation to predict future weather conditions.

While traditional models have been instrumental in advancing our understanding of climate dynamics, they often require substantial computational resources and may struggle to capture the intricacies of local weather patterns. The complexity of these models lies in their attempt to represent the Earth's climate system with a high degree of precision. However, this complexity comes at a cost. High-performance computing environments are necessary to run these simulations, which can be both time-consuming and expensive. Additionally, the granularity of these models might not always be sufficient to provide accurate predictions at a local level, where microclimates can significantly deviate from broader regional patterns.

In recent years, machine learning techniques, particularly deep learning, have emerged as powerful tools for modeling nonlinear and complex relationships in various fields, including climate science. Machine learning models can process vast amounts of data and identify patterns that might be missed by traditional methods. Among these techniques, Long Short-Term Memory (LSTM) networks, a type of recurrent neural network, have shown great promise in capturing temporal dependencies in time series data, making them well-suited for climate prediction tasks. LSTMs are designed to remember long-term dependencies and are particularly effective in scenarios where past events significantly influence future outcomes, such as weather prediction.

LSTM networks offer several advantages over traditional climate models. Firstly, they can handle large datasets efficiently, learning from extensive historical climate data to make accurate predictions. Unlike physical models that require explicit programming of each climate factor, LSTM networks learn directly from the data, identifying complex patterns and relationships autonomously. This data-driven approach allows for more flexible and potentially more accurate climate predictions, especially when dealing with the chaotic and nonlinear nature of weather systems.

This paper aims to explore the potential of LSTM networks in predicting daily climate variables such as temperature, precipitation, and humidity. By leveraging historical climate data, we seek to develop a model that can accurately forecast short-term climate trends. The methodology involves collecting and preprocessing historical weather data, training the LSTM network on this data, and then validating the model's predictions against actual observations. Key factors in this process include selecting appropriate features (such as past temperature, humidity, and atmospheric pressure), tuning the network architecture (including the number of layers and neurons), and optimizing the training process to avoid overfitting.

Accurate climate prediction has significant implications for various weather-sensitive sectors. In agriculture, better forecasts can help farmers optimize planting schedules, irrigation practices, and pest control measures, ultimately leading to higher yields and reduced losses. In disaster management, reliable weather predictions enable authorities to issue timely warnings and prepare for events such as hurricanes, floods, and heatwaves, thereby reducing the potential for human and economic losses. For urban planning, accurate climate models inform the design of resilient infrastructure, ensuring that buildings, roads, and public services can withstand future climatic changes.

The exploration of LSTM networks for climate prediction is a promising frontier in climate science. By harnessing the power of deep learning, we can develop models that provide more accurate and timely weather forecasts, supporting better decision-making across various sectors. Future research will focus on refining these models, incorporating additional data sources, and extending the prediction horizon to provide even more valuable insights into our changing climate.

II. RELATED WORK

These papers collectively provide a strong foundation for understanding the advancements in using LSTM networks for climate prediction. They cover various aspects of model development, optimization, and application, offering valuable insights into the effectiveness of deep learning techniques in this field.

1. Climate Time Series Prediction with Deep Learning and LSTM

This paper presents the application of LSTM networks to climate time series prediction, highlighting their ability to capture complex climate patterns and improve prediction accuracy compared to traditional methods. It provides a detailed comparison of different deep learning approaches and their effectiveness in climate modeling [1].

2. TD-LSTM: Temporal Dependence-Based LSTM Networks for Marine Temperature Prediction

This study proposes a new method for predicting sea surface temperature using Temporal Dependence-Based LSTM Networks (TD-LSTM). The model is evaluated using Argo data, demonstrating its effectiveness in capturing temporal dependencies and improving prediction accuracy across various depths and regions [2].

3. A Sequence-to-Sequence Approach for Remaining Useful Lifetime Estimation Using Attention-Augmented Bidirectional LSTM

This paper [3] proposed a sequence-to-sequence approach for predicting the remaining useful lifetime of equipment using attention-augmented bidirectional LSTM networks. While the focus was not directly on climate prediction, the methods and findings regarding temporal sequence modeling and attention mechanisms are highly relevant to enhancing LSTM-based climate models.

III. Methodology

1. Data Collection and Preprocessing

The study utilizes historical temperature data collected from Daily climate data in the city of Delhi. The dataset comprises daily temperature readings spanning from 2013 to 2017. To prepare the data for LSTM modeling, the following preprocessing steps were undertaken:

- **Data Cleaning:** Missing or erroneous temperature readings were identified and addressed through imputation or removal.
- **Normalization:** The temperature data were normalized to a specific range (e.g., 0 to 1) to enhance the model's convergence during training.
- **Sequence Generation:** The data were transformed into sequences of fixed lengths to capture temporal dependencies. Each sequence consists of a set of consecutive daily temperatures as input and the temperature of the following day as the target output.

2. LSTM Model Architecture

The LSTM model employed in this study is designed to capture the temporal patterns in temperature data. The architecture comprises the following layers:

- **Input Layer:** Accepts sequences of normalized temperature readings. The input shape is determined by the number of time steps and the number of features.
- **LSTM Layer 1:** The first LSTM layer has 50 units and returns sequences to allow the subsequent LSTM layer to receive sequences of inputs. This layer is responsible for capturing the short-term dependencies in the data.
- **LSTM Layer 2:** The second LSTM layer has 64 units and does not return sequences, preparing the data for the dense layers that follow. This layer aims to capture longer-term dependencies.
- **Dense Layer 1:** A fully connected layer with 32 neurons, introducing additional complexity and non-linearity to the model.
- **Dense Layer 2:** Another fully connected layer, this time with 16 neurons, further processing the learned features.

- **Output Layer:** A final dense layer with a number of units, which corresponds to the number of features in the output (in this case, the predicted temperature).

The model is compiled with the Adam optimizer for efficient gradient descent and mean squared error (MSE) as the loss function to quantify the difference between the predicted and actual temperatures [4,5,7]. Additionally, the model's performance is monitored using the mean absolute error (MAE) metric.

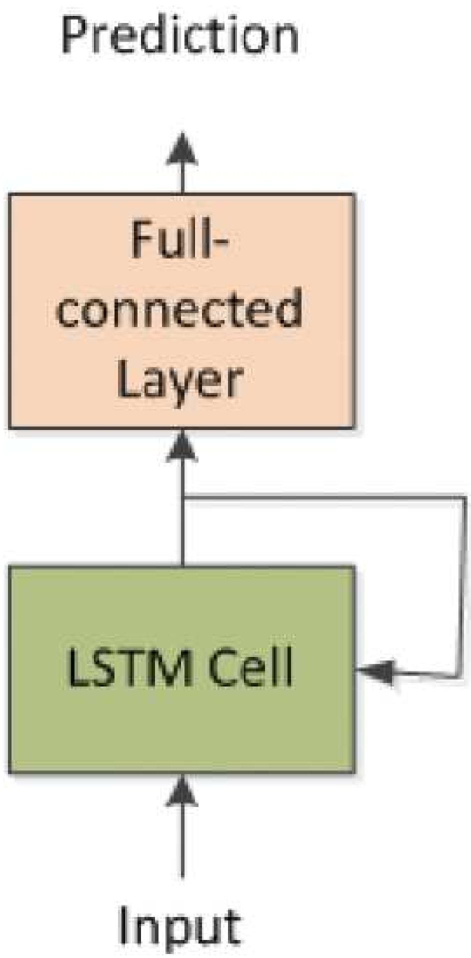


Figure 1 shows the virilization of the architecture in a high-level view. [1]

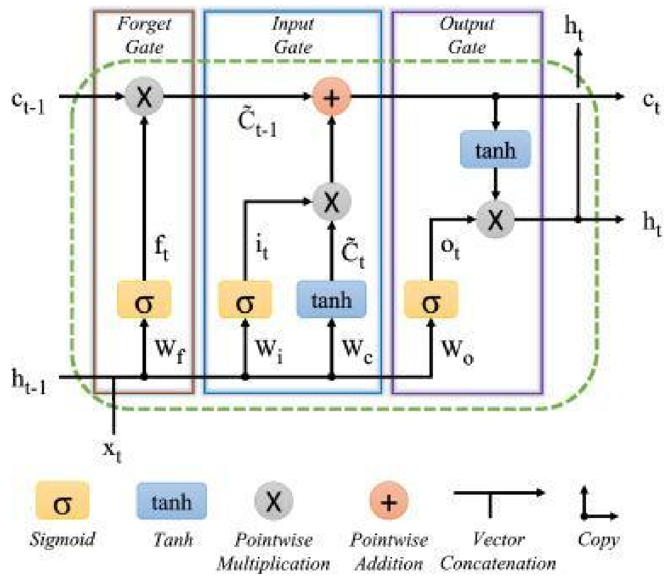


Figure 2 shows the visualization of such LSTM cell. [2]

3. Structure of LSTM cell

A. LSTM Cell State Update

$c_t = f_t \odot c_{t-1} + i_t \odot$ tanh $tanh\left(W_c \cdot \left[h_{t-1}, x_t\right] + b_c\right)$ c_t is the cell state at time t , f_t is the forget gate's output, controlling the extent to which the previous cell state c_{t-1} is retained, i_t is the input gate's output, controlling how much of the new candidate cell state tanh $tanh\left(W_c \cdot \left[h_{t-1}, x_t\right] + b_c\right)$ is added to the current cell state, $\odot$ denotes element-wise multiplication, W_c and b_c are the weights and bias for the candidate cell state.

B. LSTM Hidden State Update

$$h_t = o_t \odot \text{tanh}\ tanh\left(c_t\right)$$

Where h_t is the hidden state at time t , o_t is the output gate's output, controlling the extent to which the cell state c_t is exposed as the hidden state.

C. Gate Activation Functions

Input gate

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right)$$

Forget gate

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right)$$

Output gate

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right)$$

Where σ is the sigmoid activation function, and W_i, W_f, W_o, b_i, b_f, and b_o are the weights and biases for the respective gates.

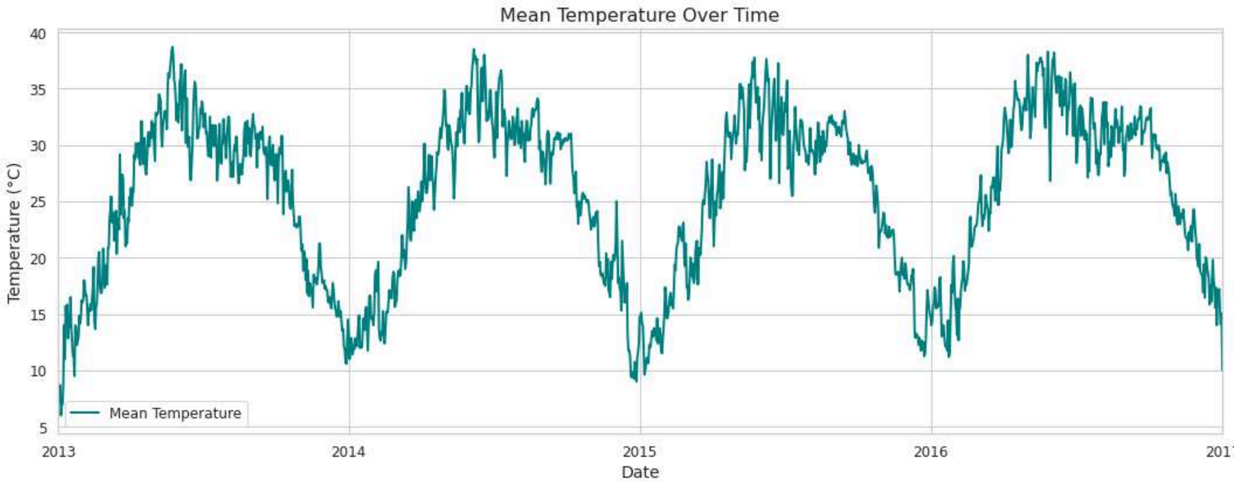
IV. Experiment

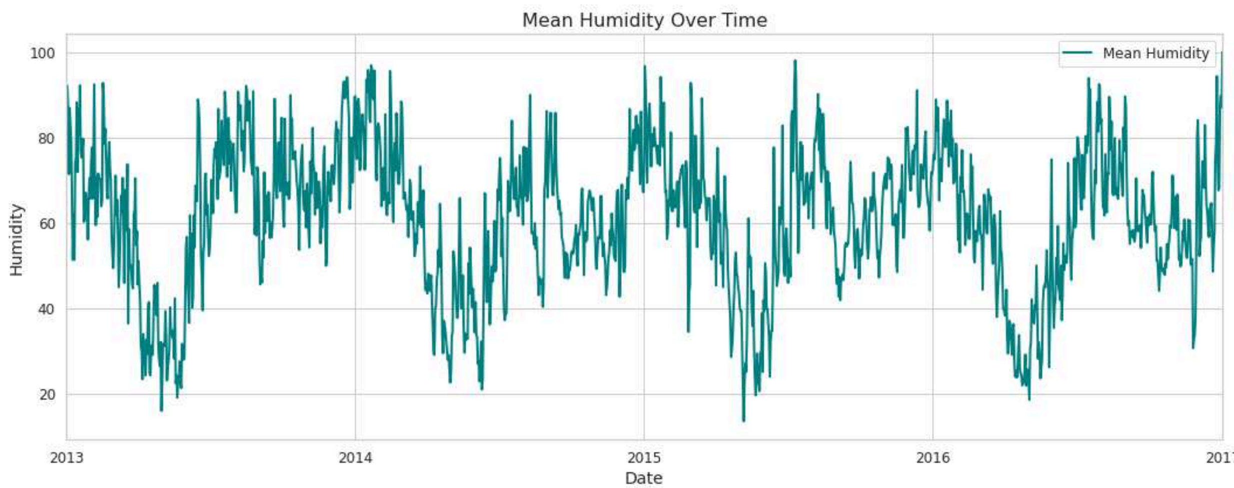
1. Data Collection and Preprocessing

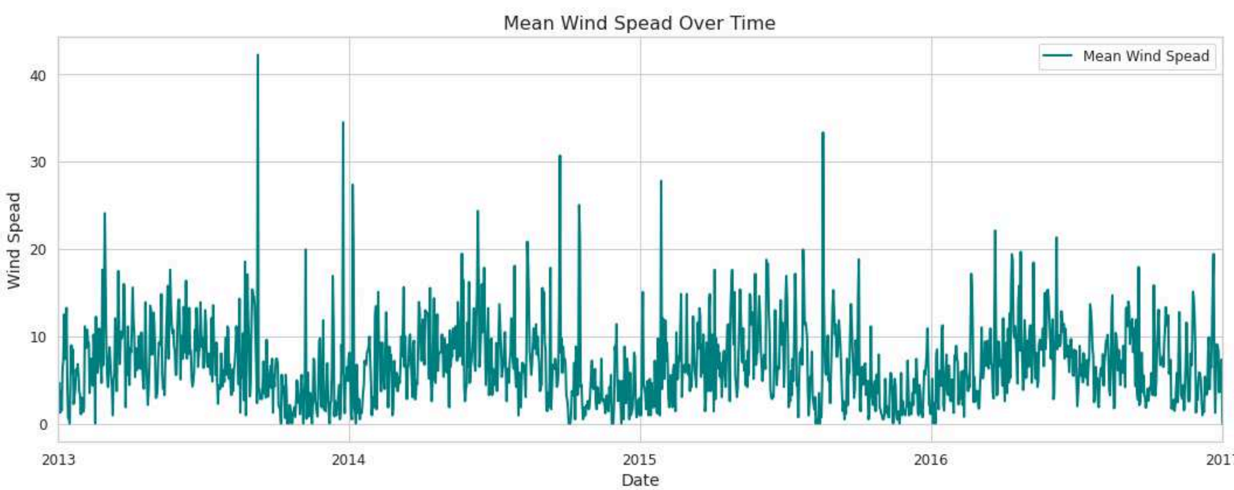
We use daily climate data in the city of Delhi from 2013 to 2017 to conduct the experiment, from https://www.wunderground.com/. [6] This dataset contains weather data from January 1, 2013, to April 24, 2017, for the city of Delhi, India. It includes four parameters: mean temperature, humidity, wind speed, and mean pressure. To prepare the data for the LSTM model, the following preprocessing steps were undertaken:

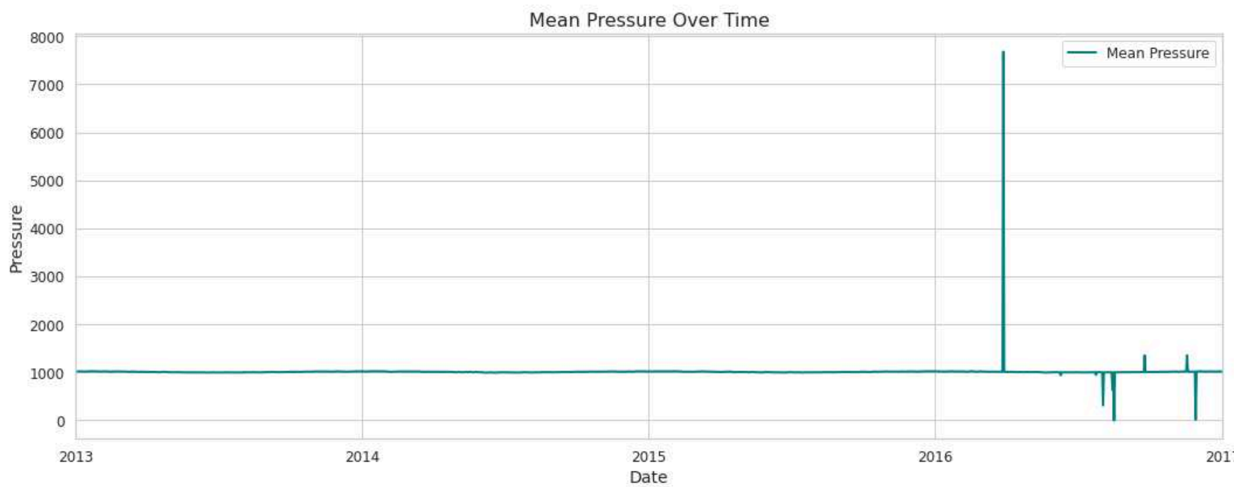
1. Data Cleaning: Missing values were handled by interpolation, and outliers were identified and removed based on statistical thresholds.
2. Feature Selection: Based on correlation analysis and domain knowledge, relevant features influencing daily climate were selected.
3. Normalization: The data was normalized using Min-Max scaling to bring all features to a similar scale, facilitating faster convergence during training.
4. Time Series Transformation: The data was transformed into a time series format.

Figure 3,4,5,6 shows the visualization of the temperature, humidity, wind speed and pressure.


Mean Temperature Over Time
Temperature (°C)
Date
Mean Temperature
40
35
30
25
20
15
10
5
2013
2014
2015
2016
2017


Mean Humidity Over Time
Humidity
Date
Mean Humidity
100
80
60
40
20
2013
2014
2015
2016
2017


Mean Wind Spead Over Time
Wind Spead
Date
Mean Wind Spead
40
30
20
10
0
2013
2014
2015
2016
2017



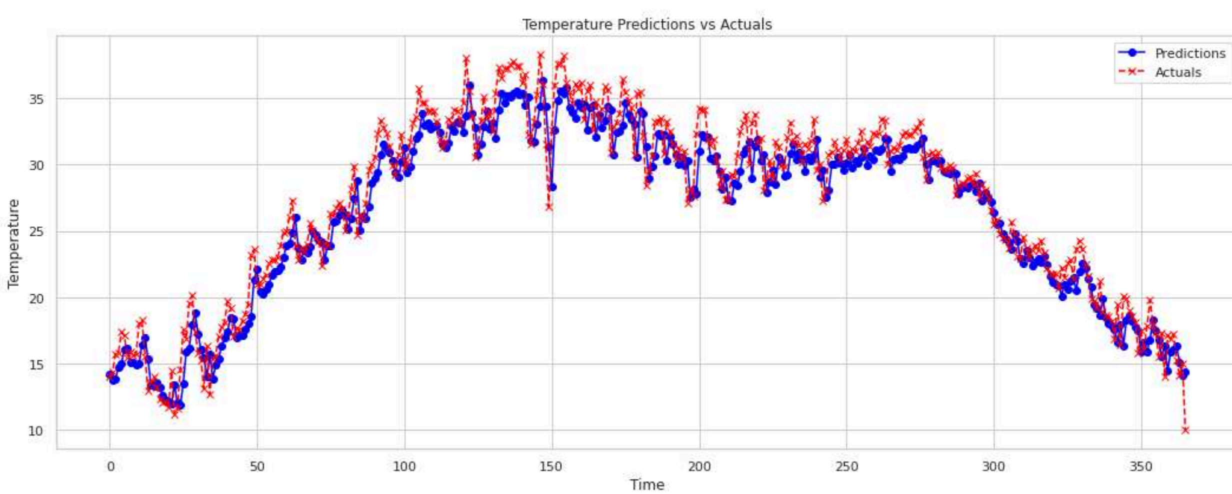
2. Evaluation Metrics

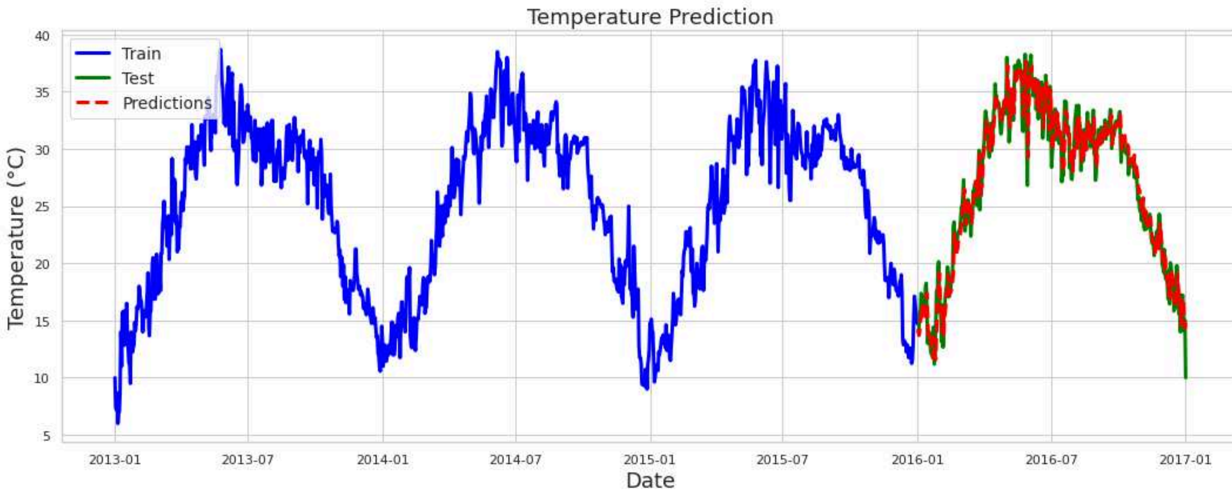
The performance of the LSTM model was evaluated using the following metrics:

- Mean Absolute Error (MAE): Measures the average magnitude of errors in temperature predictions.
- Root Mean Squared Error (RMSE): Provides a measure of the square root of the average squared differences between predicted and actual values.

V. Result and Discussion

Our Root mean square error is 0.78.





The experimental results, evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), indicate that the LSTM model can achieve a satisfactory level of accuracy in forecasting daily temperatures [16-17]. With an RMSE of 0.78, the model demonstrates its capability to capture the underlying patterns in the climate data, making it a valuable tool for short-term climate prediction.

Future work could explore the integration of additional climate variables, such as humidity and wind speed, into the LSTM model to provide a more comprehensive view of climate dynamics. Moreover, the model could be tested on different geographical regions to assess its generalizability and adaptability to various climate patterns [8-10].

Overall, this research result contributes to the growing body of knowledge on the application of machine learning in climate prediction and underscores the potential of LSTM networks in addressing complex time series forecasting challenges in the field of climate science [12-15].

VI. Conclusion

In this study, we have demonstrated the potential of Long Short-Term Memory (LSTM) networks in predicting daily climate variables, with a focus on temperature forecasting. The LSTM model, designed with two layers and additional dense layers, effectively captured the temporal dependencies in the historical climate data from the city of Delhi. The preprocessing steps, including data cleaning, normalization, and sequence generation, played a crucial role in preparing the data for the LSTM model, ensuring accurate and reliable predictions.

The successful application of LSTM networks in this study opens up new avenues for utilizing machine learning techniques in climate science. By leveraging the power of deep learning, we can enhance our ability to predict climate variables, thereby improving our preparedness for weather-related events and informing decision-making in sectors sensitive to climate variability.

VII. REFERENCE

1. Taylor, J., & Feng, M. (2022). A deep learning model for forecasting global monthly mean sea surface temperature anomalies. Frontiers in Climate, 4, 932932.
2. Liu, J., Zhang, T., Han, G., &aUniqueId6689 Gou, Y. (2018). TD-LSTM: Temporal dependence-based LSTM networks for marine temperature prediction. Sensors, 18(11), 3797.
3. Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. Advances in neural information processing systems, 27.
4. Zhang, Q., Wang, H., Dong, J., Zhong, G., & Sun, X. (2017). Prediction of sea surface temperature using long short-term memory. IEEE geoscience and remote sensing letters, 14(10), 1745-1749.
5. Shah, S. R. B., Chadha, G. S., Schwung, A., & Ding, S. X. (2021). A sequence-to-sequence approach for remaining useful lifetime estimation using attention-augmented bidirectional LSTM. Intelligent Systems with Applications, 10, 200049.
6. https://www.wunderground.com/
7. Sun, Y., Cui, Y., Hu, J., & Jia, W. (2018). Relation classification using coarse and fine-grained networks with SDP supervised key words selection. In Knowledge Science, Engineering and Management: 11th International Conference, KSEM 2018, Changchun, China, August 17–19, 2018, Proceedings, Part I 11 (pp. 514-522). Springer International Publishing.
8. Luo, H., Liu, N., & Feng, C. (2021). Question and Answer Classification with Deep Contextualized Transformer. In Advances in Information and Communication: Proceedings of the 2021 Future of Information and Communication Conference (FICC), Volume 2 (pp. 453-461). Springer International Publishing.
9. Bi, Q., Miao, Z., Shen, J., Wang, H., Kang, K., Du, J., ... & Fan, S. (2022). Detecting the research trends and hot spots in external irradiation therapy for rectal cancer. Journal of Cancer, 13(7), 2179.
10. Yang, S., Zhao, Y., & Gao, H. Using Large Language Models in Real Estate Transactions: A Few-shot Learning Approach.
11. Zhong, Z., Yang, S., & Becigneul, G. (2021, January). Environment and speaker related emotion recognition in conversations. In The 2nd International Conference on Computing and Data Science (pp. 1-6).

12. Liu, T., Zhu, F., Yu, H., & Gu, H. (2020). Surface defect detection and root cause analysis. Glob. J. Sci. Front. Res.(GJSFR), 1, 1-8.
13. Han, X., Yu, H., & Gu, H. (2019). Visual inspection with federated learning. In Image Analysis and Recognition: 16th International Conference, ICIAR 2019, Waterloo, ON, Canada, August 27–29, 2019, Proceedings, Part II 16 (pp. 52-64). Springer International Publishing.
14. Zhao, P., Qi, C., & Liu, D. (2017). Resource-constrained Hierarchical Task Network planning under uncontrollable durations for emergency decision-making. Journal of Intelligent & Fuzzy Systems, 33(6), 3819-3834.
15. Liu, Z., Leng, Y., Wang, M., & Lin, C. (2020, September). Named Entity Recognition and Named Entity on Esports Contents. In 2020 15th Conference on Computer Science and Information Systems (FedCSIS) (pp. 189-192). IEEE.
16. Zhao, P., Wang, H., Qi, C., & Liu, D. (2017). HTN planning with uncontrollable durations for emergency decision-making. Journal of Intelligent & Fuzzy Systems, 33(1), 255-267.
17. Wang, H. W., Liu, D., Zhao, P., & Chen, X. (2016). Review on hierarchical task network planning under uncertainty. Acta Autom. Sin, 42, 655-667.