TEMP-LOGISTICS

Detroit's Seasonal Strategy

Submitted by:

Ashwini Awathe(h04121)

Dinesh Arasavalli(hm3731)

Sravanthi Lankalapalli(hp8500)

Submitted to:

Dr. Qingyu Yang

DSA-6100

ABSTRACT

The transportation and logistics sectors are critically impacted by climatic factors, particularly temperature fluctuations which can significantly affect operational efficiency and safety. This research's objective is to improve the robustness and effectiveness of Detroit's logistics and transportation systems by using sophisticated predictive modeling techniques to foresee high heat periods. Random Forest, Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) networks are the three prediction models we used. To assess each model's performance, we performed a comparison analysis. The evaluation criteria for each model were the R^2 Score, Mean Absolute Error, Root Mean Squared Error, and Mean Squared Error. According to our research, the GRU model has the highest explanatory power, the lowest error rates, and the most accurate forecasts. This study explores the useful implications of precise temperature forecasting in enhancing resource management, vehicle maintenance, cargo safety, and routing efficiency in addition to showcasing the GRU model's better capacity to capture and predict temperature changes. The study's findings demonstrate how temperature prediction can significantly impact weather-related disruption mitigation and strategic planning in the logistics and transportation industries.

INTRODUCTION

In the ever-changing global economy, the transportation and logistics industry is critical, but it is also quite vulnerable to disruptions imposed on by climate change. Climate change is making extreme weather events more frequent and severe, which poses a serious threat to logistics networks' operating effectiveness and safety. The City of Detroit, a major hub for logistical and industrial activity, is the subject of this study since it endures notable seasonal temperature variations. Our goal is to improve the logistics and transportation infrastructure's resilience by creating a prediction model that forecasts the upcoming month's extreme temperatures utilizing Long Short-Term Memory (LSTM) networks and Gated Recurrent Unit (GRU).

Our model is intended to deliver precise predictions of peak seasonal temperatures by utilizing twenty years of historical temperature data. These forecasts are to help logistics organizations in their strategic planning and decision-making, resulting in improved cargo safety, proactive vehicle maintenance, and more effective routing. The ultimate objective is to guarantee smooth logistics operations and reduce disruptions during critical weather conditions.

Our approach provides a strategic advantage in risk management related to climatic unpredictability by incorporating insights from forecast temperature into logistical planning. This encourages cost-effective resource management in addition to operational continuity and safety. The knowledge gathered from this research will help Detroit's logistics and transportation organizations adapt to and prosper in the face of escalating climate problems, changing the way they handle weather-related risks.

OBJECTIVES

This project aims to enhance the operational resilience and efficiency of the transportation and logistics sector in Detroit. The specific objectives are outlined as follows:

Development of Predictive Models: Develop and implement prediction models to precisely predict extreme temperatures for the upcoming month by utilizing Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) networks. In order to train and evaluate the models, this project will make use of historical temperature data spanning over twenty years.

Model Comparison and Analysis: Determine which network offers the most accurate and dependable temperature forecasts by conducting a thorough comparison of the LSTM and GRU models. This analysis will involve evaluating various performance metrics to identify the model that best handles the complexity of temperature patterns in Detroit.

Enhancement of Logistics Efficiency: Leverage the insights gained from the temperature forecasts to enhance logistics efficiency. To ensure smoother and more dependable logistical operations, this involves developing strategies to lower the risks of disruptions at peak seasonal extremes.

Strategic Planning for Transportation: Utilize forecast insights for strategic planning in transportation management. This will ensure that the transportation sector can successfully adapt to and reduce the effects of extreme weather events. The focus will be on enhancing resilience and cost management against climate variability.

METHODOLOGY

DATA GATHERING

For the weather prediction project, data collection was meticulously carried out using the Meteostat website. Historical weather data was gathered for each year from the comprehensive archives available on the platform. The data acquisition involved downloading individual yearly datasets, which were then manually combined to create a comprehensive dataset spanning multiple years. This consolidated dataset serves as the foundation for our predictive analysis, allowing for a detailed exploration of weather patterns and trends over time.

DATA PRE-PROCESSING

The comprehensive data preprocessing was conducted on the historical temperature dataset with the following steps outline the meticulous approach taken to prepare the data:

1. Initially, checked for missing values across all the columns in the dataset to identify gaps that could potentially affect the analysis.
2. The columns 'tsun' (sunset time) and 'wdir' (wind direction), which had a high proportion of missing data were removed from the dataset to streamline the analysis as they seemed less relevant for the purposes of temperature prediction.
3. For the 'tavg' (average temperature) column, missing values were filled by calculating the average of 'tmin' (minimum temperature) and 'tmax' (maximum temperature) for each entry. This approach improved the dataset's quality by yielding a credible approximation.
4. The remaining rows that still had missing 'tavg' values after imputation were removed from the dataset to maintain a high standard of data integrity.
5. The missing values for precipitation (prcp), snowfall (snowfall), and wind speed (wspd) were filled only for the concerned season in using the corresponding median

values that were computed. This strategy was used to lessen the effect of outliers and maintain the data's central tendency.

6. The revised dataset was saved to an Excel file called "Final.xlsx" when the preprocessing stages were done, ensuring that the data would be easily accessible for further modeling and analysis.

MODEL BUILDING:

DATA PREPARATION

The Data from an Excel file named 'Final.xlsx' is loaded into a pandas DataFrame.

The 'date' column of this DataFrame is converted to datetime format, ensuring accurate handling of date-specific operations. Subsequently, a new DataFrame is created, consisting only of the 'date' and 'tavg' (average temperature) columns, with the 'date' column set as the index. This structure facilitates easier manipulation and analysis of time series data related to temperature, making it well-suited for tasks such as trend analysis and forecasting.

We created lagged features in the temperature_data DataFrame. These features ('lag1', 'lag2', 'lag3') represent temperatures from one, two, and three days prior, respectively. These lagged variables are essential for forecasting models that predict future values based on past data. After generating these features, any rows containing NaN values are removed to maintain data integrity.

Preparing Data for Model Training:
The data is then organized into features (**X**) and labels (**y**). The features include the lagged temperature values, while the labels are the current day's temperatures. This setup is typical for regression tasks where past values are used to predict current or future outcomes.
Splitting Data into Training and Testing Sets:
The dataset is split into training (80%) and testing (20%) sets using the train_test_split function, with a specified random seed for reproducibility. This step is crucial for training the model on a portion of the data and then evaluating its performance on unseen data.

Random Forest Model

A Random Forest Regressor is initialized with 100 trees and trained using the training dataset. Random Forest is a robust and versatile machine learning algorithm suitable for regression tasks, capable of handling nonlinear relationships without severe overfitting.

Model Evaluation

The model's performance is evaluated using the Mean Squared Error (MSE) metric, which provides a measure of the model's prediction error. The MSE is calculated by comparing the predicted temperatures against the actual temperatures in the testing set.

Mean Squared Error: 13.220
Root Mean Squared Error: 3.636
Mean Absolute Error: 2.694
R^2 Score: 0.883

Predicting Future Temperatures: Using the trained model, the code predicts temperatures for the next month. This is achieved by continually updating the input features with the most recent predictions, allowing the model to forecast temperatures one day at a time for 30 days. This recursive method leverages the model's ability to predict based on recent trends.

LSTM:

We used a Long Short-Term Memory (LSTM) network, a kind of recurrent neural network (RNN) renowned for its capacity to comprehend data sequences, to effectively anticipate severe temperatures. Since precise temperature forecasting requires the ability to capture temporal dependencies and patterns in time-series data, the LSTM model was used.

Data Preparation:

Normalization: The MinMaxScaler function from sklearn.preprocessing is utilized to normalize the average temperature information. By ensuring that all input features are on a comparable scale, this normalization helps to train neural networks more effectively by scaling the temperature data to a range of 0 to 1.

Dataset Creation for LSTM:

Creating Sequences: The temperature series is converted into a format appropriate for LSTM training using the create_dataset function. The function creates sequences with a duration of 30 days, given by look_back, and uses each series to forecast the temperature for the next day.

Data Reshaping: The LSTM model requires that the input data be modified to [samples, time steps, features]. After the sequences are created, they are transformed into the necessary three-dimensional shape.

Model Building and Training:

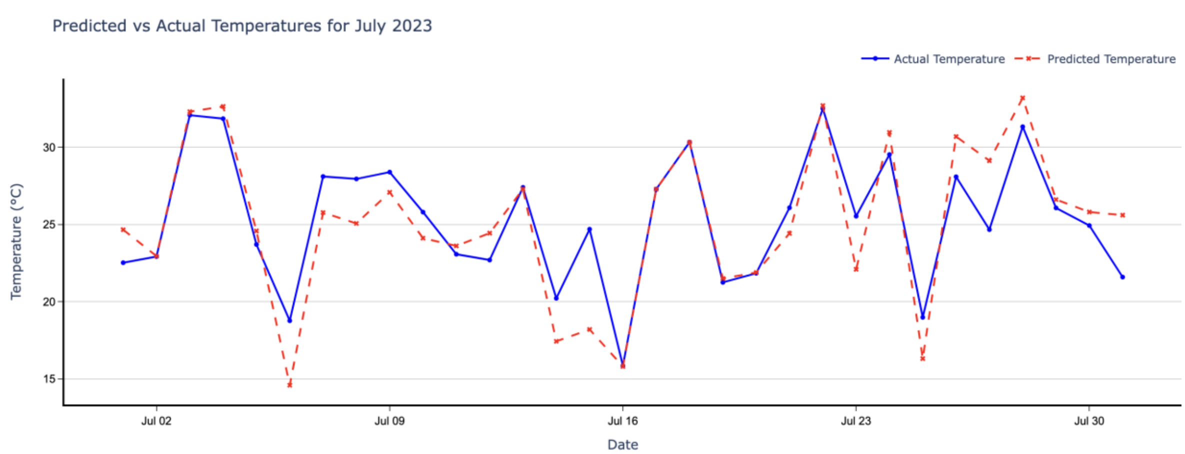
The LSTM model is constructed using Keras with 50 neurons in the LSTM layer and one neuron in the output layer for temperature prediction. The loss function of the model is mean squared error, and the Adam optimizer is used. It is trained with a batch size of 32 across 100 epochs. Using the training data (i.e., 90% of the dataset), the network's weights are adjusted during the training phase to minimize the loss function.

Prediction and Evaluation:

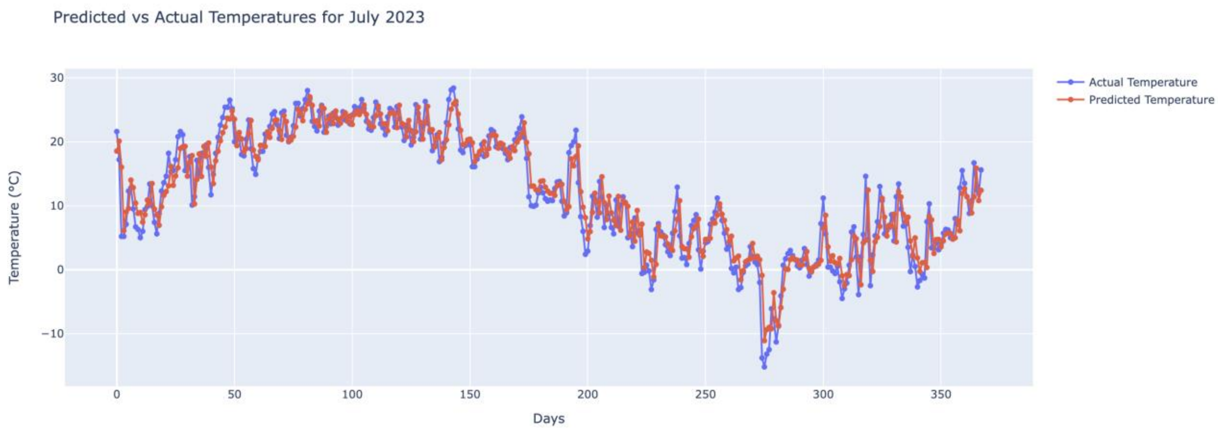
The model was applied to the training and testing datasets to provide predictions after training. Using the MinMaxScaler's inverse transform capability, these scaled forecasts were first converted back to the original temperature scale. Metrics including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the R^2 Score, which offer insights into the model's accuracy and capacity to capture the variance in the temperature data, were used to assess the model's performance.

Mean Squared Error for July 2023: 7.941
Root Mean Squared Error for July 2023: 2.818
R^2 Score for July 2023: 0.910

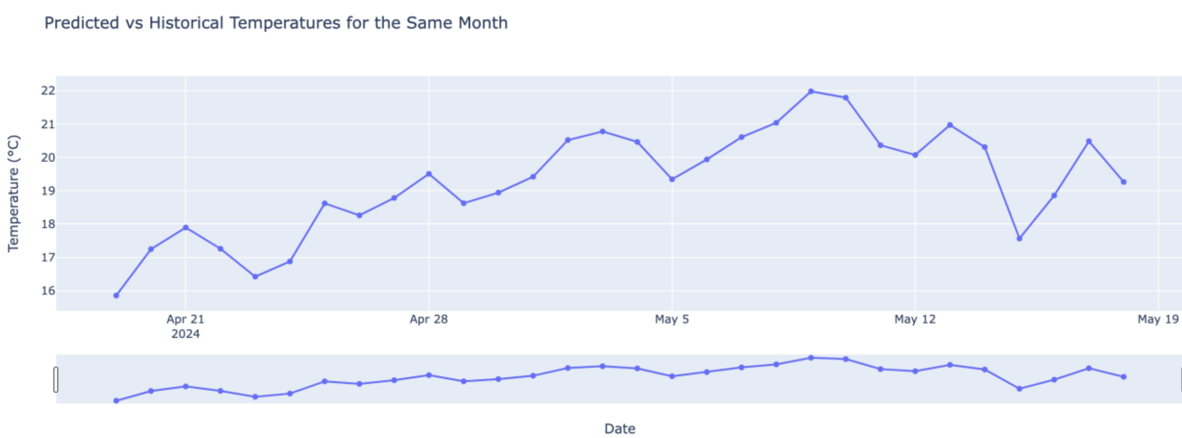
Visualizations:



The graph shows a thorough comparison of the actual and expected temperatures for July 2023. The solid blue line shows the actual recorded temperatures, whereas the dashed red line shows the forecasts made by the LSTM model, which generally follow the solid blue line.



The graph compares the actual and expected temperatures for July 2023, demonstrates how the two datasets closely match throughout the month, indicating that the LSTM model is effective at capturing temperature changes.



The graph illustrates predicted temperatures from April 19th to May 19th, 2024, showing a general upward trend as the month progresses, with temperature variations reflecting typical summer weather patterns.

GRU

GRU Network Architecture and Training

Our GRU model was constructed with two layers of GRU units, each containing 50 units. The first GRU layer returned sequences to provide a comprehensive temporal feature set to the next GRU layer. A dense layer was added after this architecture to get the final output prediction. The Adam optimizer and mean squared error, which are common options for regression tasks, were used in the model's compilation. With a batch size of 32, the model was trained for 30 epochs. A 10% validation split was added during training to keep an eye out for and reduce overfitting.

Model Evaluation

Post-training, the model's performance was assessed on the test set to gauge its predictive accuracy. Predictions were generated for the test set, and the mean squared error (MSE) was calculated, providing a quantitative measure of the model's prediction error.

Mean Squared Error: 0.00297608
R2 Score: 0.91218564

COMPARITIVE ANALYSIS

This study compared three predictive models—Random Forest, LSTM, and GRU—to forecast temperatures. With an MSE of 13.220, RMSE of 3.636, MAE of 2.694, and R^2 of 0.883, the Random Forest model demonstrated dependable but improvable accuracy. With an MSE of 7.941, RMSE of 2.818, and R^2 of 0.910, the LSTM model outperformed Random Forest in terms of accuracy and error measures. With an impressively low MSE of 0.00297608 and a R^2 of 0.91218564, the GRU model, on the other hand, outperformed the others. This indicates that, because of to its effective architecture, it is the most accurate and suitable model for time-series forecasting in crucial applications such as logistics and transportation planning.

BUSINESS PROBLEM

Enhanced Routing Efficiency: When designing a strategic route, accurate temperature forecasts are essential, especially when it comes to adjusting to road conditions caused by excessive summer heat or winter ice. Logistics businesses can improve routes to prevent weather-related delays, guaranteeing timely deliveries and upholding service reliability, by predicting and adapting to these conditions.

Proactive Vehicle Maintenance: Temperature predictions enable transportation companies to implement preventative maintenance schedules that are crucial for vehicle performance and safety during seasonal extremes. This foresight ensures operational efficiency, prevents breakdowns, and helps schedule maintenance when it's most required. **Improved Cargo Safety:** Knowing the weather forecast is essential for businesses shipping temperature-sensitive products. The integrity and quality of fragile goods like medications, perishable foods, and chemicals must be preserved during transit, and accurate temperature forecasts make it possible to modify heating or cooling during this process. **Resource Optimization:** Reducing the amount of resources used is made easier by accurate temperature predictions, especially when it comes to controlling the climate in storage and transit vehicles. Businesses can drastically cut expenses and energy usage by coordinating energy use with both current and predicted temperature conditions. This results in more environmentally friendly operations.

CONCLUSION

The present investigation has effectively exhibited the effectiveness of sophisticated predictive modeling methodologies in predicting temperature, an essential aspect for streamlining operations in the transportation and logistics domains. through a comparison of Random Forest, LSTM, and GRU, three distinct models. It became clear that the GRU model performed better than the others since it produced the most trustworthy and accurate forecasts. These results highlight the potential of GRU models in real-time applications where accuracy is critical. By putting these predictive capabilities to use, logistics and transportation systems may become more reliable and effective by improving vehicle maintenance, cargo safety, resource optimization, and routing efficiency. The insight acquired through this study opens up new avenues for investigating how to incorporate more intricate datasets and improve model architectures to increase forecast accuracy.