Weather Forecasting Using Live API Data: Prophet vs LSTM Models

By Santhosh Srinivasan

1. Introduction

This project focuses on forecasting weather variables, such as temperature, using live dynamic weather data accessed via an API. Accurate weather forecasting is critical for various sectors including agriculture, disaster management, and daily planning. The motivation behind this project is to compare classical time-series models like Prophet against deep learning models like LSTM to evaluate their performance on real-time data.

2. Data Collection and Preprocessing

The data is collected through a live weather API (such as the Metostat API), which provides up-to-date weather observations. The dataset includes timestamps and weather features such as temperature. The data is cleaned and preprocessed to fit the time-series modeling requirements, including scaling and splitting into training and testing subsets.

3. Modeling Approach

3.1 Prophet Model

Prophet is an additive time series forecasting model that explicitly handles trends and seasonality. In this project, daily seasonality was disabled to suit the data specifics. The model was trained on historical data, and forecasts were generated with prediction intervals.

3.2 LSTM Model

Long Short-Term Memory (LSTM) is a deep learning architecture designed to model complex, nonlinear temporal dependencies. The LSTM model in this project was configured with multiple layers, neurons, and trained over several epochs. Input-output sequences were constructed to capture the temporal dynamics.

4. Model Evaluation and Backtesting

Models were evaluated using multiple error metrics including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Symmetric MAPE (sMAPE). Backtesting was performed using rolling forecast splits to simulate realistic forecasting scenarios. The following are the mean backtesting results across 5 splits:

- Prophet metrics: MAE=0.7440, RMSE=0.9635, MAPE=2.59%, sMAPE=2.58%
- LSTM metrics: MAE=0.6199, RMSE=0.7503, MAPE=2.04%, sMAPE=2.04%

LSTM consistently outperformed Prophet in accuracy and error metrics.

5. Statistical Significance Testing

The Diebold-Mariano test was conducted to compare forecast accuracy between Prophet and LSTM. The test yielded a statistic of 15.5737 with a p-value of 0.0001, indicating the difference in forecast accuracy is statistically significant, strongly favoring the LSTM model.

6. Residual Diagnostics

Residual analysis was performed using the Ljung-Box test to check autocorrelation:

- Prophet residuals: lb_stat=868.94, p-value approximately 0 (highly significant autocorrelation, indicating model misfit)
- LSTM residuals: lb_stat=7.35, p-value=0.69 (no significant autocorrelation, indicating good fit)

7. Prediction Interval Coverage

Prediction interval coverage was evaluated to assess uncertainty quantification:

- Prophet coverage: 91.28%
- LSTM coverage: 91.01%

Both models provide reliable prediction intervals.

8. Visual Analysis

Visual comparisons of predicted vs actual values with uncertainty bands across selected time windows showed that the LSTM model's forecasts closely match observed data with tighter error margins than Prophet.

9. Conclusions and Future Work

The deep learning LSTM model demonstrated superior performance for weather forecasting using live, dynamic data by effectively capturing complex temporal patterns. Prophet, while useful as a baseline, showed limitations in handling nonlinearities and dynamic factors present in the data. Future work includes expanding to multivariate forecasting, real-time deployment, anomaly detection, and enhanced interpretability.

Appendix

Detailed code snippets, evaluation metric definitions, and statistical test descriptions are provided in the project repository.