Comparative Analysis of Machine Learning Models for Temperature Forecasting Using the Jena Climate Dataset

Anay Halwasiya
Statistics Department

University of Michigan
Ann Arbor, United States
anayh@umich.edu

***Abstract*—Accurate temperature forecasting is critical for understanding and mitigating climate change impacts. This study evaluates three machine learning models—LSTM, RNN, and XGBoost—on the Jena Climate dataset to predict temperature values. The XGBoost model achieved the best performance with an RMSE of 0.1414, significantly outperforming LSTM (0.2321) and RNN (0.2062). Feature importance analysis highlighted the pivotal roles of T (dew) [°C], T (pot) [K], and date time in prediction accuracy. Future extensions could explore real-time deployment and additional parameters to improve forecasting models further.**

I. Introduction

The prediction of weather parameters, particularly temperature, plays a critical role in understanding climate dynamics and supporting environmental decision-making. Traditional forecasting models often face challenges in handling complex temporal dependencies in datasets. This project involves exploring advanced machine learning models to predict temperature using the Jena climate dataset, which contains hourly weather measurements over several years. The dataset provides a rich source of temporal and multivariate data, offering an opportunity to implement and compare various approaches.

My primary goal was to evaluate the performance of deep learning models like LSTMs and RNNs against a gradient-boosting method, XGBoost. These models were selected for their strengths in capturing temporal patterns and feature importance. Initial experimentation with a pretrained Hugging Face LSTM model yielded promising results, but further custom implementations allowed for performance enhancements. A comprehensive evaluation of these models was conducted using RMSE as the primary metric to ensure robust comparisons.

II. Method

A. Problem Formulation

The goal of this project is to predict hourly temperature values based on historical weather data from the Jena climate dataset. The dataset includes measurements such as temperature, pressure, humidity, wind speed, and solar radiation. The input consists of a sliding window of 720 time steps (30 days of hourly data), while the output is the temperature value 72 time steps (3 days) ahead.

```
LSTM Hyperparameters:
            Parameter   Value
                Units     128
           Activation    tanh
 Recurrent Activation sigmoid
              Dropout     0.2
    Recurrent Dropout     0.2
            Optimizer    Adam
        Learning Rate   0.001
           Batch Size      64
               Epochs       5
```

Fig. 1. LSTM Hyperparameters

B. Dataset Description

The Jena climate dataset comprises hourly weather readings recorded between 2009 and 2016. Key variables include temperature (°C), pressure (hPa), relative humidity (%), wind direction and speed, and solar radiation. The dataset was preprocessed by normalizing features and creating overlapping windows of time-series data. The data was split into training (80%) and testing (20%) subsets to ensure unbiased evaluation.

C. Model Formulation

I implemented three models for comparative analysis:

- Pretrained Hugging Face LSTM: Leveraged a prebuilt model fine-tuned for time-series forecasting.
- Custom RNN: Developed a recurrent neural network model with layers optimized for sequential data.
- XGBoost: Applied a gradient-boosting decision tree model known for its interpretability and efficiency.

D. Methodology

- Preprocessing: All continuous variables were normalized to have zero mean and unit variance, ensuring that features with larger ranges did not dominate the learning process. The raw data was converted into sliding windows for input-output pair generation.

```
XGBoost Hyperparameters:
            Parameter             Value
Number of Estimators                100
        Learning Rate               0.1
            Max Depth                 6
            Subsample               0.8
    Colsample by Tree               0.8
            Objective reg:squarederror
```

Fig. 2. XGBoost Hyperparameters

```
RNN Hyperparameters:
           Parameter   Value
               Units      64
          Activation    relu
Recurrent Activation sigmoid
             Dropout     0.2
   Recurrent Dropout     0.2
           Optimizer    Adam
       Learning Rate   0.001
          Batch Size      64
              Epochs       5
```

Fig. 3. RNN Hyperparameters

- Model implementation: The LSTM model was trained according to the parameters given in Hugging Face but its performance was suboptimal compared to custom-built models.
  The RNN model consisted of multiple recurrent layers designed to process time-series data effectively. Dropout regularization was employed to mitigate overfitting, and the Adam optimizer was chosen for its adaptive learning rate capabilities. Mean squared error (MSE) was used as the loss function, aligning with the goal of minimizing prediction error.
  XGBoost, a gradient-boosting framework, was implemented to exploit its strength in handling structured tabular data. Feature engineering ensured that temporal information was encoded effectively.
- Evaluation: Predictions were made on the test set, and the root mean squared error (RMSE) was computed for comparison.
- Visualization: Predicted vs. actual temperature values were plotted for all models, allowing qualitative assessment of model performance. For the XGBoost model, feature importance was calculated to identify the most influential variables contributing to temperature prediction. This analysis helped interpret the model's decision-making process and highlighted key predictors.

III. Results

The LSTM model achieved an RMSE of 0.2321, providing a strong initial benchmark. While it captured broad temperature trends, its generalization to unseen test data was limited due to potential mismatches between its training dataset and the Jena Climate data.

The RNN model outperformed the pretrained LSTM with an RMSE of 0.2062, showcasing the benefits of a custom architecture tailored to the dataset. The RNN effectively captured long-term dependencies, improving predictive accuracy.

XGBoost delivered the best performance with an RMSE of 0.1414, significantly reducing the error. Its ability to model complex relationships and handle tabular data with feature importance analysis provided a deeper understanding of the variables influencing temperature predictions.

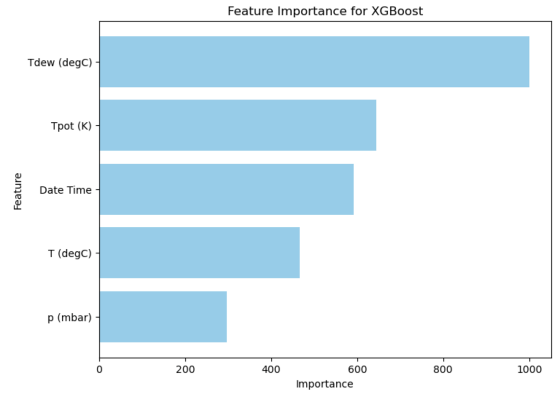

Fig. 4. Feature Importance Values for XGBoost

The feature importance values were also analyzed using the XGBoost model and dew point temperature emerged as the most critical feature, significantly impacting the prediction accuracy. Its high importance aligns with the physical relationship between dew point and air temperature. Other important features include the potential temperature, actual temperature and pressure.

IV. Conclusion

This project investigated the application of advanced machine learning models for temperature forecasting using the Jena Climate dataset. Three distinct models—LSTM, RNN, and XGBoost—were implemented and evaluated based on their predictive performance.

The LSTM model effectively captured sequential dependencies in the dataset, but its relatively high RMSE indicated some limitations in handling complex, nonlinear interactions. The RNN model, while simpler, slightly improved on the LSTM model, suggesting reduced overfitting and improved generalization. However, the XGBoost model emerged as the clear winner due to its ability to model complex feature interactions and handle heterogeneity in the data effectively.

Feature importance analysis further emphasized the role of specific meteorological variables in accurate temperature prediction. The dominant influence of T (dew) [°C], followed by T (pot) [K], date time, T [°C], and p [mbar], underscores the physical relevance of these features. This insight aligns with established meteorological knowledge, demonstrating the model's capacity for meaningful interpretation.

Future work could involve extending this analysis to incorporate additional meteorological parameters or ensemble methods to further enhance predictive accuracy. Additionally, deploying these models for real-time forecasting and evaluating their performance under different temporal and spatial resolutions would provide practical value.