Rainfall Prediction Model Report

Introduction

The purpose of this report is to present a model for predicting daily rainfall using open meteo data. The model was trained on a dataset consisting of daily weather data from 2019-01-14 to 2022-01-14. The model's performance was evaluated using mean squared error, R2 score, and accuracy. This report also provides information on the model architecture and conclusions on the model's performance.

Input Features

The model uses the following features as input:

- time
- weathercode (wmo code)
- temperature_2m_max (°C)
- temperature_2m_min (°C)
- temperature_2m_mean (°C)
- apparent_temperature_max (°C)
- apparent_temperature_min (°C)
- apparent_temperature_mean (°C)
- sunrise (iso8601)
- sunset (iso8601)
- shortwave_radiation_sum (MJ/m²)
- precipitation_sum (mm)
- rain_sum (mm)
- snowfall_sum (cm)
- precipitation_hours (h)
- windspeed_10m_max (km/h)
- windgusts_10m_max (km/h)
- winddirection_10m_dominant (°)
- et0_fao_evapotranspiration (mm)
- Model Performance

After loading and preprocessing the dataset we checked if there were any null values, as there were no null values we processed further to check the correlation of the variables and moved further towards creating a model to predict rainfall

Two models were evaluated on the held-out test set using the following metrics:

Linear Regression Model
Mean Absolute Error (MAE): **0.343**
Root Mean Squared Error (RMSE): **0.522**
Adjusted R2 Score: **0.80**

Decision Tree Regressor Model
Mean Absolute Error (MAE): **0.173**
Root Mean Squared Error (RMSE): **0.375**
Adjusted R2 Score: **0.94**

The mean squared error measures the average squared difference between the predicted and true rainfall values on the test set. A lower mean squared error indicates better performance. The R2 score measures the proportion of variance in the target variable explained by the model. An R2 score of 1.0 indicates perfect fit, while a score of 0.0 indicates that the model is no better than predicting the mean of the target variable. An R2 score close to 0.0 indicates that the model is not a good fit for the data. In our case, the R2 score for the Linear regression model is 0.84, indicating that the model explains a significant proportion of the variance in the rainfall data. For the Decision Tree Regressor model, the R2 score is 0.94, indicating that this model explains even more of the variance in the rainfall data.

The MAE, MSE, and RMSE metrics for the Decision Tree Regressor model also indicate good performance. The MAE measures the absolute difference between the predicted and true rainfall values, while the MSE and RMSE measure the squared difference and square root of the squared difference, respectively. A lower value of these metrics indicates better performance.

Model Architecture

The Linear regression model used in the previous section attempts to find a linear relationship between the input features and the target variable. The Decision Tree Regressor model, on the other hand, builds a decision tree to make predictions. It splits the input data into subsets based on the values of the input features and recursively splits the subsets until a stopping criterion is met. The final predictions are made by averaging the rainfall values of the training examples that reach the same leaf node.

Conclusion

The Decision Tree Regressor model outperformed the Linear regression model in predicting daily rainfall. Further experimentation with different models and input features may be necessary to improve the model's performance even further.