Rainfall Prediction Model Report

This report's objective is to offer a model for hourly rainfall prediction using accessible meteorological data. A dataset of daily weather data from 2019-01-14 to 2022-01-14 was used to train the model. The accuracy, R2 score, and mean squared error were used to assess the model's performance. The model architecture and results of the model's performance are also covered in this study.

Input Features

The model uses the following features as input:

- **time**
- **temperature_2m temperature_2m_max (°C)**
- **precipitation_hours (h)**
- **windspeed_10m**
- **windgusts_10m**
- **rain_sum (mm)**

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

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

Linear Regression Model

Mean Absolute Error (MAE): **0.393**

Root Mean Squared Error (RMSE): **0.578**

R2 Score: 0.779

Adjusted R2 Score: **0.77**

Lasso Regression Model

Mean Squared Error (MSE): **0.6623**

Root Mean Squared Error (RMSE): **0.81377**

R2 Score : 0.788

Adjusted R2 Score: **0.789**

Random forest Model

Mean Squared Error (MSE): **0.5273**

Root Mean Squared Error (RMSE): **0.7260**

R2 Score : **0.8381**

Adjusted R2 Score: **0.9279**

The average squared difference between the test set's actual rainfall and what was expected is measured by the mean squared error. Better performance is indicated by a smaller mean squared error. The R2 rating expresses how much of the variance in the target variable is accounted for by the model. A score of 0.0 means that the model does no better than forecasting the target variable's mean, while a score of 1.0 shows perfect fit. The model is not a good match for the data when the R2 value is close to 0.0. In our situation, the linear regression model's R2 score is 0.779, meaning that a sizable percentage of the variance in the rainfall data is explained by the model. The R2 score of 0.83 for the Random Forest Regressor model shows that this model explains even more variance in the rainfall data.

The Random forest model performs well, according to the MAE, MSE, and RMSE metrics. Whereas the MSE and RMSE quantify the squared difference and square root of the squared difference, respectively, the MAE measures the absolute difference between the predicted and actual rainfall levels. Better performance is indicated by measures with lower values.

Conclusion

When it came to forecasting daily rainfall, the Random Forest model fared better than the Lasso and linear regression models. To further enhance the model's performance, more testing with various models and input features may be required.