Rainfall Prediction Model Report

We trained a Linear regression model to predict daily rainfall using all the input features from open meteo:
'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)'

- 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

The model was trained on a dataset consisting of daily weather data from 2020-02-13 to 2023-02-11.

Model Performance

We evaluated the performance of the model on a held-out test set using mean squared error, R2 score, and accuracy.

Mean Squared Error: 0.272
R2 Score: 0.84
Accuracy: 0.80

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 is 0.84, indicating that the model explains a significant proportion of the variance in the rainfall data. Finally, the accuracy of the model is also reported, which is equal to the R2 score.

Decision Tree Regressor Model Performance

We evaluated the performance of the Decision Tree Regressor model on a held-out test set using mean squared error, root mean squared error, mean absolute error, R2 score, and adjusted R2 score.

Mean Squared Error: 0.138
Root Mean Squared Error: 0.371
Mean Absolute Error: 0.169
R2 Score: 0.941
Adjusted R2 Score: 0.939

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 mean squared error value of 0.138 for the Decision Tree Regressor model is significantly lower than the Linear regression model, suggesting that the Decision Tree Regressor model is better at predicting rainfall.

The root mean squared error and mean absolute error provide additional measures of the model's performance. The root mean squared error of 0.371 and mean absolute error of 0.169 indicate that the Decision Tree Regressor model's predictions are on average about 0.371 and 0.169 units away from the true rainfall values, respectively.

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. A higher R2 score indicates better performance. The R2 score of 0.941 for the Decision Tree Regressor model is significantly higher than the Linear regression model, indicating that the Decision Tree Regressor model is a good fit for the data.

Finally, the adjusted R2 score is a modified version of the R2 score that takes into account the number of input features used by the model. The adjusted R2 score of 0.

Conclusion

In this report, we presented a rainfall prediction model that uses daily temperature data as input features. However, the Linear regression model used did not perform well in predicting rainfall. Further analysis and experimentation with different models and input features may be necessary to improve the model's performance.