Model Selection Report

Project Name: Rainfall Prediction

Team:

Lakshmi Sravya Savaram

Mohammad Shouqat Azeez

N Gokul Chowdary

Nallabotula Vijaya Karthik

Date: 13/10/2025

Objective

The objective of this report is to evaluate different machine learning models for predicting rainfall based on historical climate data and select the best-performing model based on accuracy and other relevant metrics.

Dataset Overview

Feature	Description	Data Type
Date	Date of observation	datetime
Temperature	Daily average temperature (°C)	float
Humidity	Daily average humidity (%)	float
Wind Speed	Wind speed (km/h)	float
Rainfall	Rainfall amount (mm)	float
Weather	Weather description	categorical

Total Records: [Number of rows]
Train-Test Split: [e.g., 80%-20%]

Models Considered

Model	Description	Hyperparameters
Linear Regression	Predicts rainfall using linear relationship with features	Default
Random Forest Regressor	Ensemble model using multiple decision trees	n_estimators=100, max_depth=None
Gradient Boosting Regressor	Boosted ensemble of decision trees	n_estimators=100, learning_rate=0.1, max_depth=3
XGBoost Regressor	Gradient boosting optimized for speed and performance	n_estimators=100, learning_rate=0.1, max_depth=3
Support Vector Regressor (SVR)	Uses kernel-based regression for prediction	kernel='rbf', C=1.0, epsilon=0.1

Evaluation Metrics

Metric	Description	Formula/Explanation
Mean Absolute Error (MAE)	Average absolute difference between predicted and actual values	`MAE = (Σ
Mean Squared Error (MSE)	Average squared difference between predicted and actual values	MSE = (Σ(y_pred - y_actual)^2)/n
Root Mean Squared Error (RMSE)	Square root of MSE, measures prediction error in same units as target	RMSE = sqrt(MSE)
R-squared (R^2)	Proportion of variance in the dependent variable explained by the model	R^2 = 1 - (SS_res/SS_tot)

Model Performance Summary

Model	MAE	MSE	RMSE	R^2	Remarks
Linear Regression	[value]	[value]	[value]	[value]	Baseline model
Random Forest Regressor	[value]	[value]	[value]	[value]	Handles non-linear relationships well

Model	MAE	MSE	RMSE	R^2	Remarks
Gradient Boosting Regressor	[value]	[value]	[value]	[value]	Slightly better than Random Forest
XGBoost Regressor	[value]	[value]	[value]	[value]	Best performance observed
SVR	[value]	[value]	[value]	[value]	Sensitive to feature scaling

Note: Fill in [value] with your experimental results.

Model Selection Criteria

- Best **R^2** and lowest **MAE/RMSE**.
- Robustness to overfitting and generalization capability.
- Computational efficiency and scalability for future predictions.
- Interpretability (optional, if required).

Selected Model

Model	Reason for Selection
[Selected Model Name, e.g., XGBoost Regressor]	Achieved highest R^2 and lowest MAE/RMSE. Handles non-linear relationships in rainfall data effectively and performs well on unseen test data.

Conclusion

The selected model is ready for final training on the complete dataset and deployment for rainfall prediction. Future steps include:

- Hyperparameter tuning (if not already done)
- Feature importance analysis
- Model deployment and monitoring