REPORT ON MODEL EVALUATION RESULTS

-Kaavya Varshitha Raman Shantha

EVALUATION OF MACHINE LEARNING MODELS ON WINE QUALITY DATASET:

The dataset used for this evaluation is the **'Wine Quality'** dataset, which contains various chemical properties of red wine. The goal is to predict the quality of the wine, which is a score between 0 to 10. For this, the input variables are fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol.

SOURCE: Kaggle.

LINK: https://www.kaggle.com/datasets/huseyinelci/wne-qualty-by-uci?select=WineQuality-RedWine.csv

MODEL EVALUATION RESULTS:

The models that were evaluated for this dataset are the Multiple Linear Regression Model, Polynomial Regression Model, Support Vector Regression Model, Decision Tree, and Random Forest.

The evaluation metrics used were R-squared value, mean-squared error, root-mean-squared error, normalized root-mean-squared error, and mean absolute percentage error.

MODEL	R-squared value	Mean-squared error	Root mean-squared error	Normalized root mean-squared error	Absolute percentage error.
Multiple Linear	0.328	0.384	0.620	12.40	8.78
Polynomial	-16721.03	9572.71	97.84	1956.80	433.09
Support Vector	0.348	0.373	0.611	12.22	8.25
Decision Tree	-0.168	0.669	0.818	16.36	8.59
Random Forest	0.369	0.361	0.601	12.02	8.03

CONCLUSION:

Based on the evaluation results, the **Random Forest** model performed the best on the Wine Quality dataset because it had the highest R-squared value indicating that it explained the most variance in the data. It also had the lowest mean-squared error, root-mean-squared error, normalized root-mean-squared error, and mean absolute percentage error indicating that it made the smallest errors in prediction. Therefore, the Random Forest model is the best model for this dataset.