Q1 & Q2 Progress Report: Machine Learning Tasks

1. Classification Task (Datasaurus Dozen)

- **Baseline Performance:** I established initial performance metrics using baseline J48 and Random Forest classifiers. The Decision Tree achieved an accuracy of **39.73%**, while the Random Forest achieved **46.22%**. These scores serve as a benchmark for further improvement.

- **Parameter Optimization:** I employed GridSearchCV and RandomizedSearchCV techniques to identify optimal hyperparameter settings for both classifiers. This process aimed to fine-tune the models and potentially enhance classification accuracy.

- **Evaluation:** After optimization, the Decision Tree reached an accuracy of **43.77%**, while the Random Forest improved to **47.42%**. These results indicate that parameter tuning successfully improved the models' ability to classify data points correctly.

-

2. Regression Task (Used Cars Prices)

- **Baseline Performance:** I initially established a baseline model using a Decision Tree Regressor with default parameters. This model produced a Root Mean Squared Error (RMSE) of **452,514.75** on the testing data. RMSE measures the average magnitude of the difference between predicted and actual values. Lower RMSE indicates better predictive accuracy.

  Parameter Optimization: Similar to the classification task, I utilized GridSearchCV to find the best hyperparameter combination for the Decision Tree Regressor. This could potentially minimize the model's prediction error.

- **Evaluation:** Following hyperparameter tuning, the optimized Decision Tree Regressor achieved a best score of **0.66** (explained variance score). While the report doesn't directly show a new RMSE value, a higher explained variance

score generally translates to a lower RMSE, suggesting improvement in prediction accuracy.