A screenshot of a computer program

Description automatically generated with medium confidence

# Evaluating the Performance of Multiple Regression Models on a Multi-Output Dataset

This code snippet is evaluating the performance of several machine learning models on a dataset with multiple output dimensions. The models being evaluated are Ridge, Lasso, Tweedie Regressor, Lasso Lars, Linear Regression and Random Forest Regressor. These models are all regression models that can be used to predict numerical values.

The code first imports the ‘sqrt’ function from the ‘math’ module. This function will be used later to calculate the root mean squared error (RMSE) of the models.

Next, a list of tuples is created where each tuple contains the name of the model and an instance of the model wrapped in a ‘MultiOutputRegressor’. This is because the code is designed to handle multiple output dimensions such as Sea level (mm), Wheat Yield (Mt), Rice Yield (Mt), Cotton Yield (Mt), and Mortality under 5 years (No deaths/1000 births). The ‘MultiOutputRegressor’ is a meta-estimator that extends regression estimators to multioutput problems.

An empty DataFrame is then created to store the results of the evaluation. The columns of this DataFrame are 'Model', 'MAE', 'MSE', 'RMSE', and 'R^2'. These are common evaluation metrics for regression models. MAE stands for mean absolute error, MSE stands for mean squared error, RMSE stands for root mean squared error and R^2 stands for coefficient of determination.

The code then enters a loop where each model is fit on the training data ‘X’ and ‘y’ and used to make predictions on the test set. The evaluation metrics are calculated for each output dimension and stored in the results DataFrame.

After all models have been evaluated, the results DataFrame is indexed by the model name and printed to the screen. Finally, a bar plot of the results is created using the ‘plot’ method of the DataFrame and displayed using ‘plt.show()’.

This code snippet provides a way to compare the performance of several machine learning models on a given dataset with multiple output dimensions using common evaluation metrics.