Model Development Phase Template

Date	11 JULY 2024
Team ID	SWTID1720000556
Project Title	Predicting Co2 Emission by Countries Using Machine Learning
Maximum Marks	6 Marks

Model Selection Report

In the forthcoming Model Selection Report, various models will be outlined, detailing their descriptions, hyperparameters, and performance metrics, including Accuracy or F1 Score. This comprehensive report will provide insights into the chosen models and their effectiveness.

Model Selection Report:

Model	Description	Hyperparameters	Performance Metric (R2 score)
Linear Regression	Linear regression is a statistical technique that models the relationship between a dependent variable and one or more independent variables using a linear equation. The model predicts the dependent variable as a linear combination of the independent variables, minimizing the sum of the squared differences between observed and predicted values.	NA	0.00026

Ridge Regression	Ridge regression, or Tikhonov regularization, is a linear regression technique that includes a regularization term to penalize large coefficients, thereby reducing model complexity and multicollinearity. The model minimizes the sum of squared residuals plus a penalty proportional to the square of the magnitude of the coefficients.	Ridge Regression
		
		`ridge_reg = Ridge()`
		
		`x=np.logspace(-4, 4, 50)`
		`param_distributions = {'alpha': x}`
		
		`ridge_random = RandomizedSearchCV(ridge_reg, param_distributions, n_iter=10,random_state=42`
		`ridge_random.fit(x_train, y_train)`
		
		RandomizedSearchCV
		best_estimator_: Ridge
		Ridge
			0.00024

Decision Tree Regressor	A Decision Tree Regressor is a non-linear machine learning model that splits the data into subsets based on feature values, forming a tree-like structure. Each internal node represents a decision based on a feature, and each leaf node represents a predicted continuous value.	`tree_reg = DecisionTreeRegressor()`
		`param_distributions = {`
		`  'max_depth': randint(1, 20),`
		`  'min_samples_split': randint(2, 20),`
		`  'min_samples_leaf': randint(1, 20),`
		`  'max_features': ['auto', 'sqrt', 'log2', None]`
		`}`
		
		`tree_random = RandomizedSearchCV(tree_reg, param_distributions, n_iter=100, cv=5, random_state=42, n_jobs=`
		`tree_random.fit(x_train, y_train)`
		
		RandomizedSearchCV
		best_estimator_: DecisionTreeRegressor
		DecisionTreeRegressor
			0.85651

Random Forest Regressor	Random Forest Regressor is an ensemble learning method that combines multiple decision trees to improve predictive accuracy and control overfitting. It builds each tree using a random subset of the data and features, and the final prediction is made by averaging the predictions of all individual trees.	NA	0.324888
XGBoost Regression	XGBoost regression is an advanced ensemble learning technique that uses gradient boosting to build a series of decision trees, where each tree corrects the errors of the previous one. It optimizes model performance by combining predictions from multiple trees and includes regularization to prevent overfitting.	NA	0.983078