

## Model Development Phase Template

Date	15 March 2024
Team ID	SWTID1720007638
Project Title	Predicting CO2 emissions by countries using Machine Learning.
Maximum Marks	6 Marks

### Model Selection Report

In this report, we will select the most accurate model for this project by analyzing the accuracies produced by different model on training and testing them with the appropriate data set and required features. Below are the following models used and their respective accuracies.

#### Model Selection Report:

Model	Description	Hyperparameters	Performance Metric (e.g., Accuracy, F1 Score)
Random Forest Regressor	This model constructs a series of decision trees at training time . The final prediction is obtained by averaging the result of all individual trees.	n_estimators=10, random_state=52, n_jobs=1, verbose=2	Accuracy: The accuracy of the model is approximately 98.97%, indicating that the model correctly predicts about 98.97% of the instances in the training dataset.
Ridge regression model	It's main objective is to remove overfitting. It is a type of Linear Regression that includes a regularization term which adds to the squared magnitude of the coefficients to the loss	alpha = 0.1, random_state =52	Mean Squared Error (MSE): MSE=2.1550023905273144 $\times 10^{27}$ This metric indicates the average of the squares of the errors—that is, the average squared difference between the estimated values ( $\hat{y}$ pred) and the actual value (y test).

	function, penalizing large coefficients.		<p>A lower MSE value indicates better model performance. Score (<math>R^2</math> Score):</p> <p>Score=0.02462738768 This represents the coefficient of determination, which indicates how well the independent variables explain the variability of the dependent variable. A score closer to 1 indicates a better fit, while a score closer to 0 suggests that the model does not explain much of the variability</p>
Polynomial regression model	This is another form of regression model where it represents the relationship between the dependent and independent variables in the form of a nth degree polynomial. This model is effective in capturing non-linear relationships.	Default hyper parameter values.	<p><b><math>R^2</math> score (Coefficient of Determination):</b></p> <p><b>Score=0.1233308059404914</b></p> <p><b>The <math>R^2</math> score indicates the proportion of variance for the dependent variable that is explained by the independent variables in the model. A score close to 0 suggests that the model does not explain much of the variability in the data.</b></p>