Model Development Phase

Date	15 March 2024
Team ID	SWTID1720007638
Project Title	Predicting CO2 emission by countries using Machine Learning.
Maximum Marks	5 Marks

Feature Selection Report:

In this report, We will see all the features present in the dataset and choose specific features which are required for model building. Below are the features available in the data set. Some important features are used while others are not used for training and testing the model to get a better accuracy and the reasoning is also mentioned below.

Feature	Description	Selected (Yes/No)	Reasoning
Country name	This feature contains all the country names in the data set.	Yes	We must use the country name to train and test the model, so that we can predict the emission values for different countries.
Country code	This feature contains the abbreviation for all the country names to identify each country uniquely.	No	We don't use country code because we use the country's name which is more specific than the country code.
Indicator name	This feature contains all the different indicator names present in	Yes	We use indicator name because in the data set there are multiple indicators available.

	the data set. The main indicator that is required for our project is "CO2 emissions (metric tons per capita)".		For this project, we only need CO2 emission indicator. With specifying the indicator name, we can perfectly identify the CO2 emission value.
Indicator code	This feature contains all the abbreviations for the indicator names to identify each of them uniquely.	No	Since we are using Indicator name, we don't have the necessity to use indicator code.
Year	This feature is to indicate from which year we want the prediction of CO2 emission by a country.	Yes	We use the year value because, the model is trained on predicting CO2 emission values for different countries and different years for each country. Thus, given any year, the model will predict the emission value.
Value	This is the predicted CO2 emission value based on the other input values.	Yes	This feature contains all the emission values for different countries in different years. On training the model with this emission value, we expect the model to predict the CO2 emission for different countries in different years.