|  |
| --- |
| **Dataset Description** |
| The dataset comprises of 1,068 rows of data consisting of specific data on different cars. We will be using this dataset to build a model that can effectively predict the carbon dioxide emissions of a car given certain specifications. The data set comes in the form of a .CSV file. This file contains 13,884 individual data points in 13 columns. Following are the columns present in the dataset:   * **MODELYEAR** – Model Year tells us when the specific car model was launched in the market.   *During feature selection this column was dropped because its not contributing to the carbon dioxide emissions of the cars.*   * **MAKE** – This column gives us the name of the company that developed the car. * **MODEL** – This column gives us the specific model name of the car for whose the specifications are written. * **VEHICLECLASS** – This column gives us the class in which the specified car belongs to (E.g. – compact, SUV, etc.). * **TRANSMISSION** – This column tells us about the type of transmission system used in the vehicle (E.g.- Automatic, Manual, Semi-automatic, etc.). * **FUELTYPE** – This column tells us the type of fuel being used by the car (E.g.- Petrol, Diesel, etc.).   *The five aforementioned features were removed during feature selection because of its categorical nature.*   * **ENGINESIZE** – This column tells us the size of the engine (in Liters). * **CYLINDER** – This column tells us how many cylinders are present in the engine of the car. * **FUELCONSUMPTION\_CITY** – This column gives us the mileage of the car in city streets (liters per 100 km). * **FUELCONSUMPTION\_HWY** – This column gives us the mileage of the car in highway (liters per 100 km). * **FUELCONSUMPTION\_COMB –** This column gives us the mileage of the car in both city roads and the highway combined (liters per 100 km). * **FUELCONSUMPTION\_COMB\_MPG** – This column gives us the mileage of the car in both city roads and the highway combined (miles per gallon).   *The aforementioned 6 columns were selected through feature selection to predict CO2EMISSIONS.*   * **CO2EMISSIONS** – This column tells us how much carbon dioxide is emitted by the given car. (in grams per kilometer)   *This is our output variable. The model will help us in finding out this variable when the input variables are provided.* |