**Project Description Document**

**General Information on dataset:**

1. Numerical:

* Name of dataset (Car details v3).
* Labeled by selling\_price.
* Number of samples in dataset 8128. Train to test ratio is 80% to 20% which is 6502 to 1626 and validation is 1084.

**Implementation details:**

1. Numerical:

* Dimension of features is 11 which are (name, year, selling\_price, km\_driven, fuel, seller\_type, transmission, owner, mileage, engine, max\_power, torque, seats).
* Number of fold is 6 and ratio of training to validation is 5 to 1.
* Hyperparameters used are n\_splits in KFold = 6 and n\_neighbors in KNeighborsRegressor = 3

**Results details:**

1. Numerical:

* Linear Regression model: r2.score = 57.9%, rmse = 160637. Polynomial Regression with degree 2: r2.score = 70%, rmse = 137513.98.
* KNN model: r2.score = 77%



