**DETAILS OF PROJECT**

**1.Data preprocessing and Preparation**

1.1 Dataset is loaded.

1.2 The file car\_dataset\_india.csv is read using pandas.

1.3 First 5 row displayed using df.head() function.

1.4 General dataset information are retrieved using df.info() function.

1.5 Missing values count are checked using df.isnull().sum().

1.6 Filled null values using .ffill() function.

1.7 Basic statistical information is displayed using df.describe().

-->Feature transformation and scaling:

1.8 Categorical variables are converted to numerical values using

Label\_Encoder.

1.9 Standardized the dataset using standard\_scaler

**2. Supervised Learning :**

2.1 select the input x value (ie. Independent value) and y value

(ie. dependent or target).

2.2 Data is split into train and test sets.

-->Model Selection:

2.3 select the supervised learning model from sci-kit learn library

(Linear\_Regression,Logistic\_Regression, KNN, SVM)

2.4 train the models using model.fit()

2.5 Next predict the model

2.6 Find the accuracy score using r2 score formula via prediction model

Best accuracy score is above the 0.7 -0.9

But this model is come only -> -.0018...

2.7 so we using hyper\_parameter tuning to improve the model accuracy

- This program we used GridsearchCV method for tuning

2.8 Then visualize the predicted model using matplotlib.pyplot

**3. Unsupervised Learning:**

3.1 For unsupervised learning used K-Mean clustering

3.2 load and read the dataset using pandas

3.3 First we visualize the dataset using seaborn scatterplot

3.4 Select the X value. this case we don't select the Y value because we play

using X data only. in unsupervised learning we don't know the target.

3.5 Normalize the X value using preprocessing

3.6 The optimal value of K in the K-Means algorithm can be found using the

Elbow method

3.7 We use K-Means algorithm from Sci-kit-learn and provide it the K value.

After that we will fit it on our training dataset and get cluster labels

3.8 finally we visualize datasets

3.9 Unsupervised learning are mainly used for preparing the data for

analyzing not for prediction and accuracy.