Machine Learning Project

Dataset: Car-Class

1. The Data given is a Classification data, it has multiple targets.
2. The Dataset has 719 samples.
3. There are 18 numerical features.

Pre-processing

* This Data has no missing values.
* The Data was split-ted into 80% and 20% for training and testing.
* Then the Data I standardized by StandardScaler in sklearn.

Modelling

The Data is fitted in following algorithmic techniques:

* The Logistic Regression gives a train score of 0.8243478260869566, and a test score of 0.7986111112. It fits Justright to the model
* KNN gives a Train score of 0.834347826067, and a Test score of 0.715277778. It fits Justright to the model
* Decision Tree gives a Train score of 1.0, and Test score of 0.71527777728. This is Overfitting the model.
* SVM, it gives a Train score of 0.834782, and a test score of 0.7430555555. This fits Justright to the model
* Gradient Boosting Classifier, it gives a Train score of 0.9982609, and a Test score of 0.75694444

This fits Jusright to the model

* Random Forest Classifier, it gives a Train score of 1.0, and a test score of 0.79169999. It Over fits the model
* GaussianNB Classifier, it gives a Train score of 0.4682607, and a test score of 0.4375. It is Underfitting the model

Final Model

Among These Models, Logistic Regression fits well.

* F1-score of this model is 0.8724377
* Accuracy is 0.81

Confusion Matrix

