**Car Class Multiclass Classification Data**

**Table of Content :**

**Step 1 : Importing Libraries and Importing Data**

In this step we are importing libraries like pandas,matplotlib,seaborn to upload the cars\_class.csv data to the workbook and tried to explore the data using pandas.

**Step 2 : Data Cleaning**

In this step I done exploratory Data Analysis to check missing values,null values are other misleanous thing in data after that I tried to remove the unwanted columns and also splitting the data to x and y to further the process.

Step 3 : Training and Testing Data

In this step we importing train\_test\_split library and then we are training the model x\_train,x\_test,y\_train,y\_test.

**Step 4 : Data Processing**

In this step we first check the outliers in data because if we have outliers our end result will be affected to clear then and any inconsistency between the data I use Standard Scaler method to standardised the data and then there is a drastic change in the data.

**Step 5 : Model Building**

After Standardizing the data I created a 8 Models like Logistic regression,SVM,KNN,Decision Tree,Enemble Learning,Nayes I created all this model on default value which this libraries are have then out of all this model I select the best model which produces the best result out by checking their Accuracy and F1-Score,confusion matrix,precision,recall Score.

**Step 6 : Hypertuning the Model**

From the above Model Building we selected Random Forest Classifier Method because it has best values among them then with the help of Randomised and Gridsearch CV I decided the parameters that we going to use it on the Model.

**Step 7 : Feature Selection**

In this Method I just check the score of all the features present in the data as a end result we will get the least scored features based on that and also with the help of visualisation I even plot the graph to see importance of the features and then i removed the least score features.

**Step 8 : Final Model**

* In this project, I build a Random Forest Classifier to predict the Class of the cars. I build a models with 100 decision-trees.
* The model accuracy score with Orginal Features the 100 decisiontrees is **0.7569** but the same with 100 decisiontrees with 13 features after reduction of the least score features the score is **0.8194**. So, as expected accuracy increases with number of decision-trees and Important Feature Selection in the model.
* I have used the Random Forest model to find only the important features, build the model using these features and see its effect on accuracy. The most important feature is **Max.L.Ra** and least important feature is **Kurt.maxis,Ra.Gyr,Kurt.Maxis,Pr.Axis.Rect,Circ**.
* I have removed the **['Kurt.maxis','Ra.Gyr','Kurt.Maxis','Pr.Axis.Rect','Circ']** variable from the model, rebuild it and checked its accuracy. The accuracy of the model **with ['Kurt.maxis','Ra.Gyr','Kurt.Maxis','Pr.Axis.Rect','Circ'] variable removed is 0.8194**. The accuracy of the model with all the variables taken into account is **0.7569**. So, we can see that the model accuracy has been improved with **['Kurt.maxis','Ra.Gyr','Kurt.Maxis','Pr.Axis.Rect','Circ'] variable removed from the model**.
* Confusion matrix and classification report are another tool to visualize the model performance. They yield good performance.