**Car\_Class Classification**

* The csv file contains multi-class classification data set with 719 samples and 18 numerical features.
* The target variable is the class of the car which may be one of: 0 –bus, 1 – Opel Manta, 2 –Saab, 3 – Van.

* After importing the required libraries and the csv file, the data-frame is checked for missing values and unique values.
* The “ID” column is dropped as its not giving any information to the model.
* Each features are visualized using box plot for better understanding.
* The data-set contains outliers. Outliers are data Points which are far away from the nearest point.
* There are four ways for handling outliers.
  + Sorting method
  + Using visualization
  + Using Z-score
  + Using inter-quartile range
* The Heat-map is plotted for the given data-set for visualizing the correlation between the features.
* Observing the heat-map found that the feature “elong” has no correlation between the other features. Hence dropping the feature.
* The data-set is splitted into train and test for building the model.
* The models used are K-nearest neighbour,Decision Tree,support vector machine,logistic regression,Random Forest.
* The Scores obtained in each models are

|  |  |  |
| --- | --- | --- |
| **Learning Algorithm** | **Train Score** | **Test Score** |
| **Decision Tree** | 0.8044692737430168 | 0.674074074074074 |
| ****K**-Nearest Neighbors** | 0.8305400372439479 | 0.6222222222222222 |
| **Support Vector Classification** | 0..8584729981378026 | 0.7407407407407407 |
| **Random Forest**: | 0.8640595903165735 | 0.7333333333333333 |
| **Logistic Regression**:**Train Score** | 0.7355679702048417 | 0.6814814814814815 |

* Considered the random forest as my final model.
* The scores obtained by the model are Train Score=0.8640595903165735,Test Score=0.7333333333333333
* Accuracy:0.7407407407407407
* F1 score: 0.7037037037037037
* Confusion matrix for the final model are obtained

