

ML Term Assignment

19BCE248- Shah Dhruvil
19BCE286- Jurin Vachhani



Dataset

Car Evaluation

These particular dataset consist of 1728 rows and 7 columns which are as follows:

- 1) Buying
- 2) Maintainance
- 3) Doors
- 4) Persons
- 5) Luggage Boot
- 6) Safety
- 7) Acceptance





Data Preprocessing



Data Analysis



Model Prediction without any transformation with
different ML models



Testing



Examining problem of overfitting



Applying Model prediction after oversampling data
with various techniques



Final Testing



Conclusion



Preprocessing Dataset

Here as the data provided was in the textual format so we have converted into numeric so that model training becomes easy.

Data Analysis

Here as the dataset was so small dropping any column doesn't seem feasible for classification model. But though we have analysed effect of each of the column with the output parameter.



**Random Forest
Classifier**

67%

Naive Bayes

68%

KNN

90%

SVC

91%

Decision Tree

94%

Different Result with various sampling technique

Manual Approach

97%

BorderLine SMOTE

95%

RandomOverSampler

99%

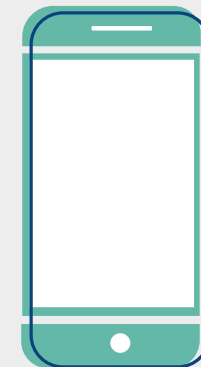
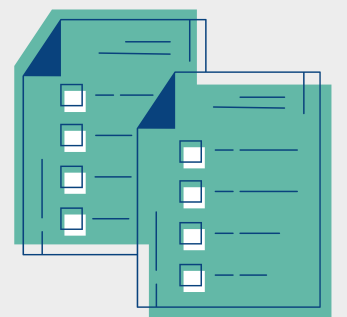
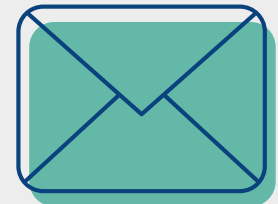
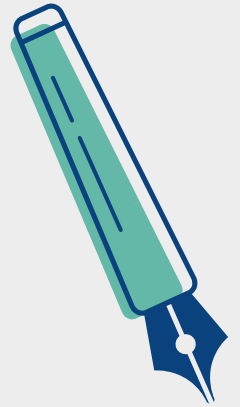
SVMSmote

93%

Testing

Testing can be done in the following two ways:

- 1) Prediction of a whole new instance from the user for a single instance.
- 2) By replacing the test.csv file with your input file.



Conclusion

From performing above ML classification we learned about different classifiers. Tackling of different problems like the Imbalance dataset, Non-numerical value, outliers were all explored during these courses of action. The most difficult problem we found was a limited amount of data which is not feasible in real-world scenarios but to deal with it was great exposure. Lastly by exploring various models, we got more clarity about which sets of algorithms to be used in various scenarios.



Thank You

