SUMMARY

ABSTRACT

This project is designed to create a classification model that classifies seeds into 7 Registered bean groups that will be used with a computer vision system in the crop processing line to get a uniform seed variety.

DESIGN

- Data cleaning
- Exploratory data analysis
- Classification models

DATA From UCI machine learning repository.
Data points : 13,611
Features : 16, 14 were used(
Classes: 7

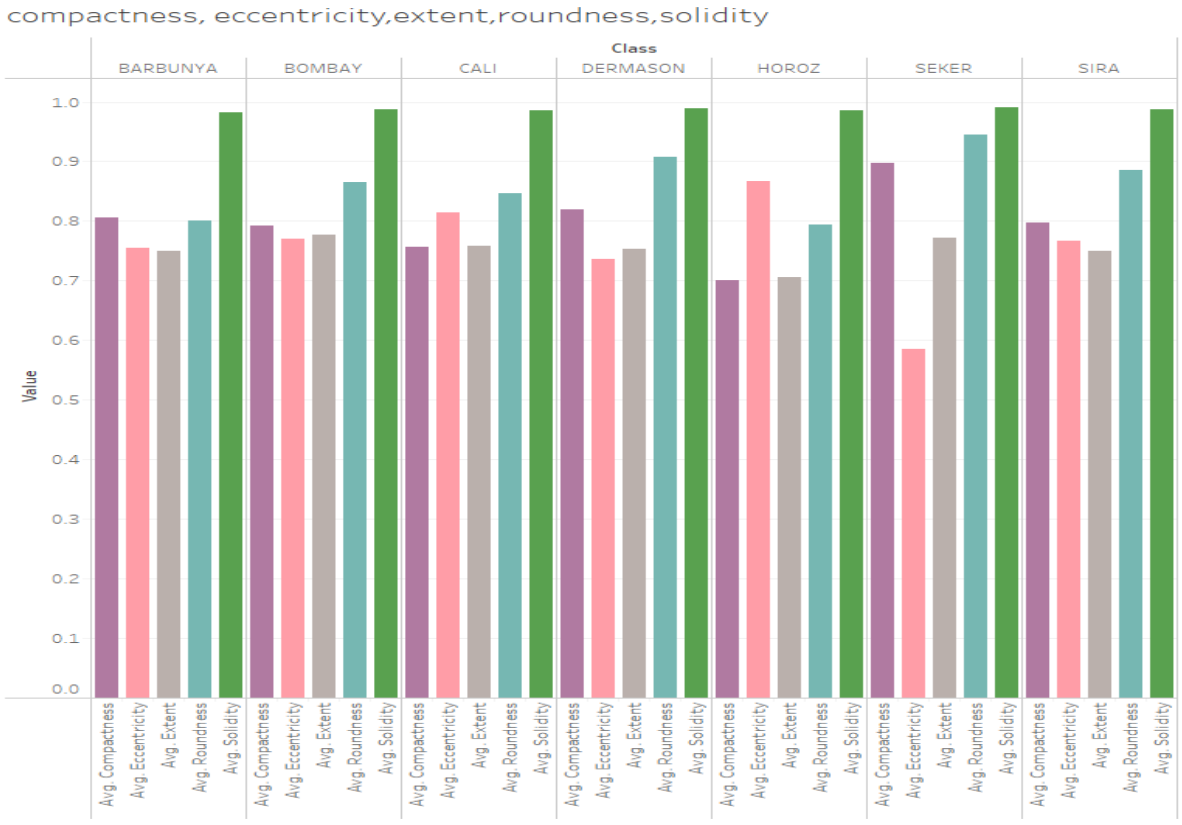
METHODOLOGY

- Check for n/a values
- Use pair plots and Bar charts to see the relationship between features and classes
- Remove the two features that are not predictive ; Solidity and shape factor 4
- Split the dataset into 80% train & 20 % test sets
- Build a baseline model using KNN algorithm
- Check for accuracy, precision, recall and confusion matrix
- Build a second model using random Forest
- Improve random forest model by feature scaling and parameter tuning
- Crid_search for hyperparameter tuning
- Build third model ,Naive bayes
- Choose best model

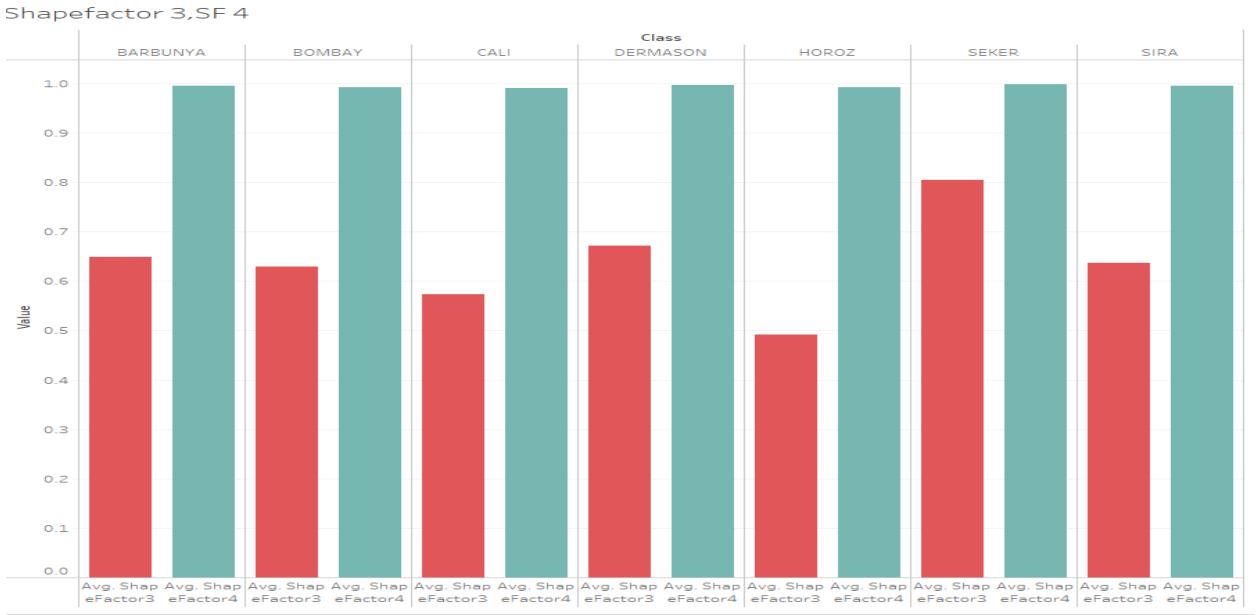
TOOLS

TABLEAU - Data analysis
Python - Data analysis, model building and testing

COMMUNICATION : Tableau Sheets



*All Classes have almost the same



solidity value.

*Shape Factor 4 is also almost the same for all classes.