DRUG CONSUMPTION ANALYSIS

By

Sarika Kshatriya

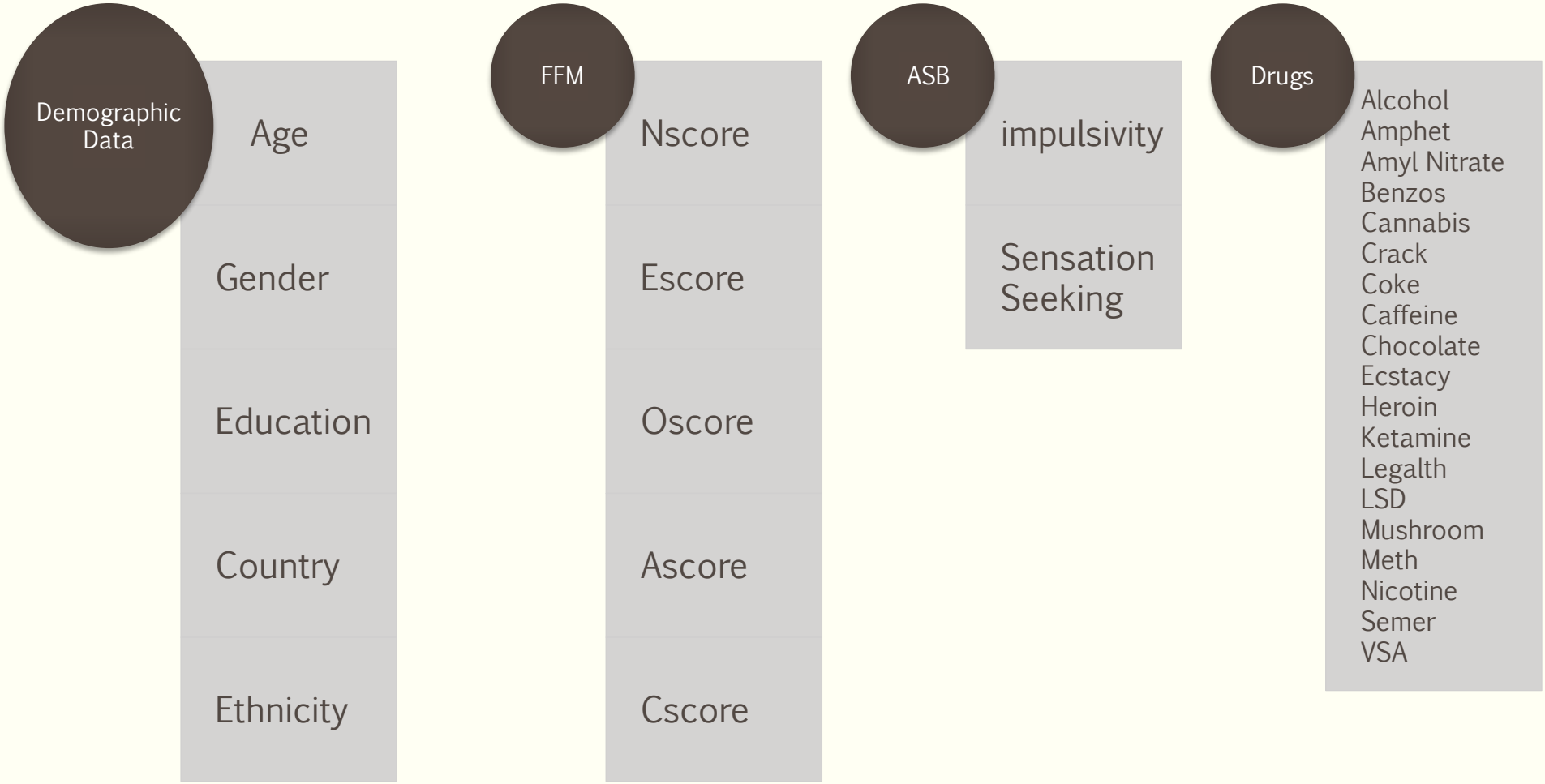
What is a Drug Abuse?

- An abusable psychoactive drug is a ‘drug whose mental effects are sufficiently pleasant or interesting or helpful that some people choose to take it for a reason other than to relieve a specific malady.
- Serious Global problem.
- Risk Factors.
- Associated with a number of personality traits.

OBJECTIVE

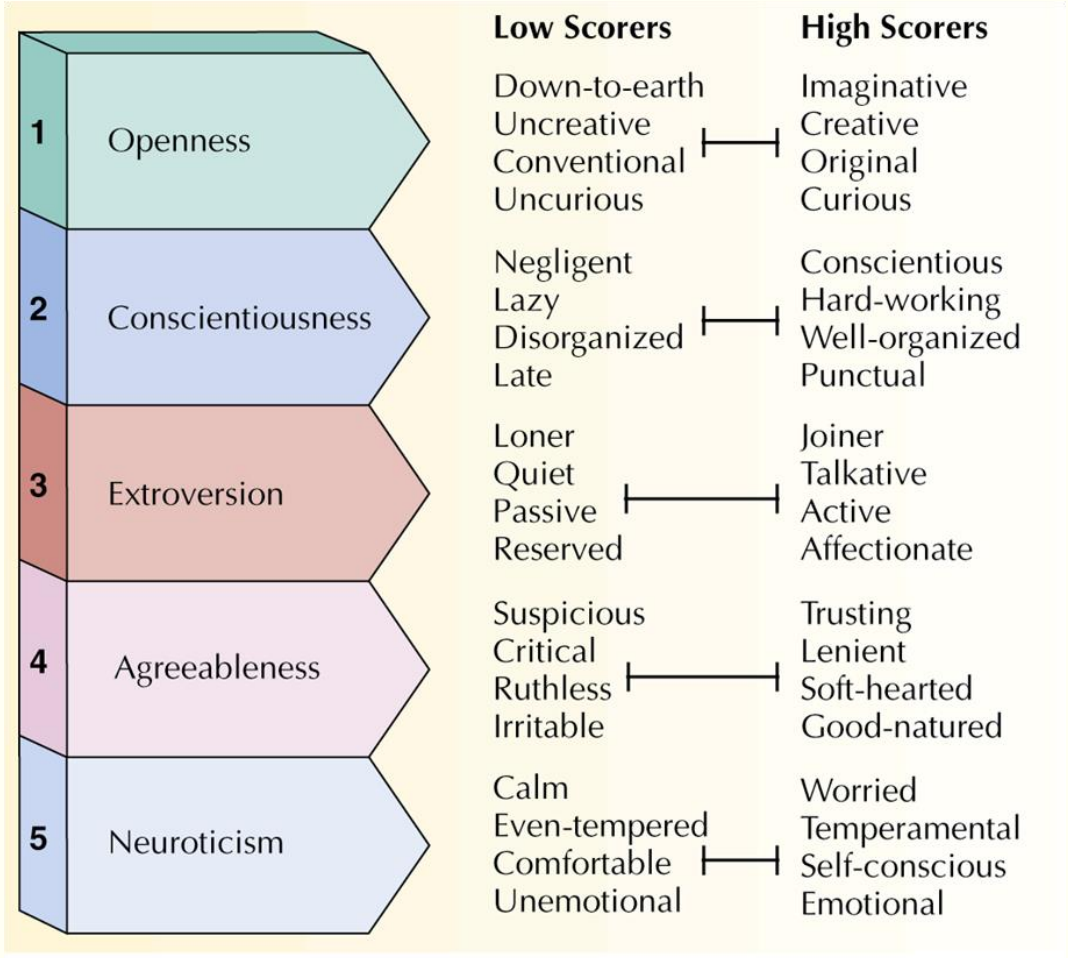
- Assess the potential effect of the data on different drug consumptions.
- Predict the risk of drug consumption for each individual.

Dataset



link: https://archive.ics.uci.edu/ml/datasets/Drug+consumption+%28quantified%29

Personality Traits



Five Factor Model

Anti Social Behavior

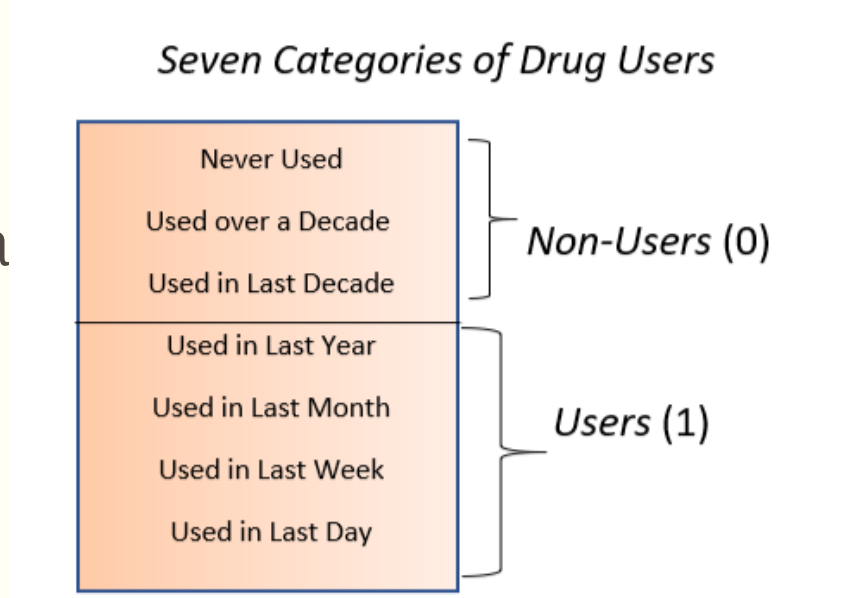
Impulsiveness:

- Motor impulsiveness: reflects acting without thinking.
- Attentional impulsiveness: poor concentration and thought intrusions.
- Non-planning. a lack of consideration for consequences.

Sensation-Seeking (**SS**):
Measure of high-risk behavioral such as, substance misuse.

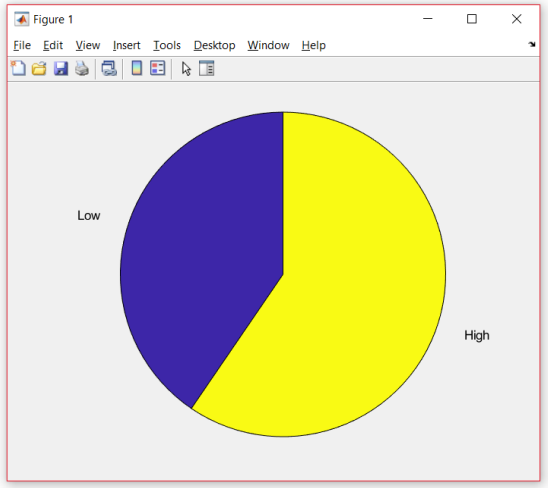
DATA PREPROCESSING

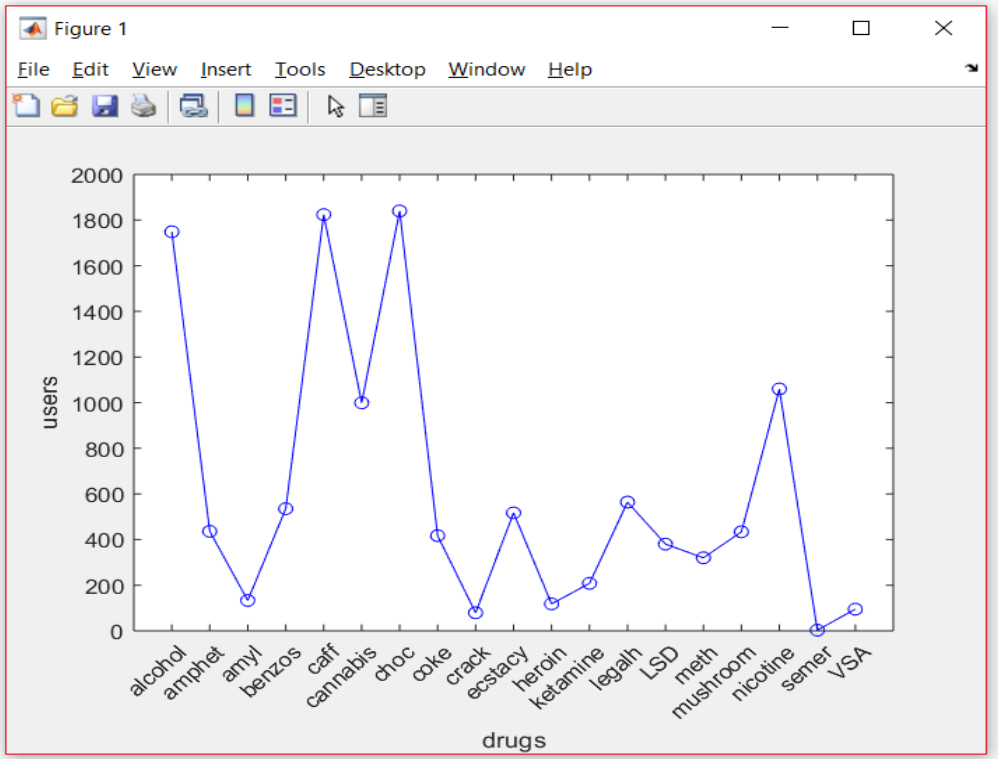
- Age : divided into 6 categories 18 to 24 (18), 25-34(25),35-44 (35), 45-54 (45), 55-64 (55) a above (65)
- Binary Classification ‘decade-based’.



Some Important Findings

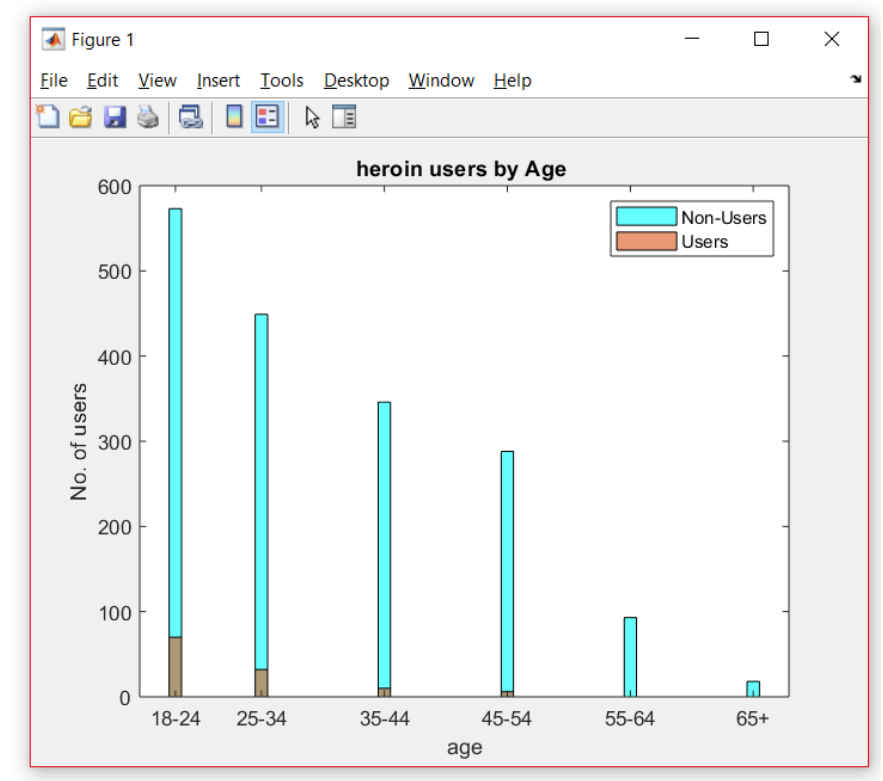
- Highest drug users are teenagers
- Highly educated people tend towards drug abuse.
- Male Female Equality in drug abuse.
- Legal drugs are very popular.



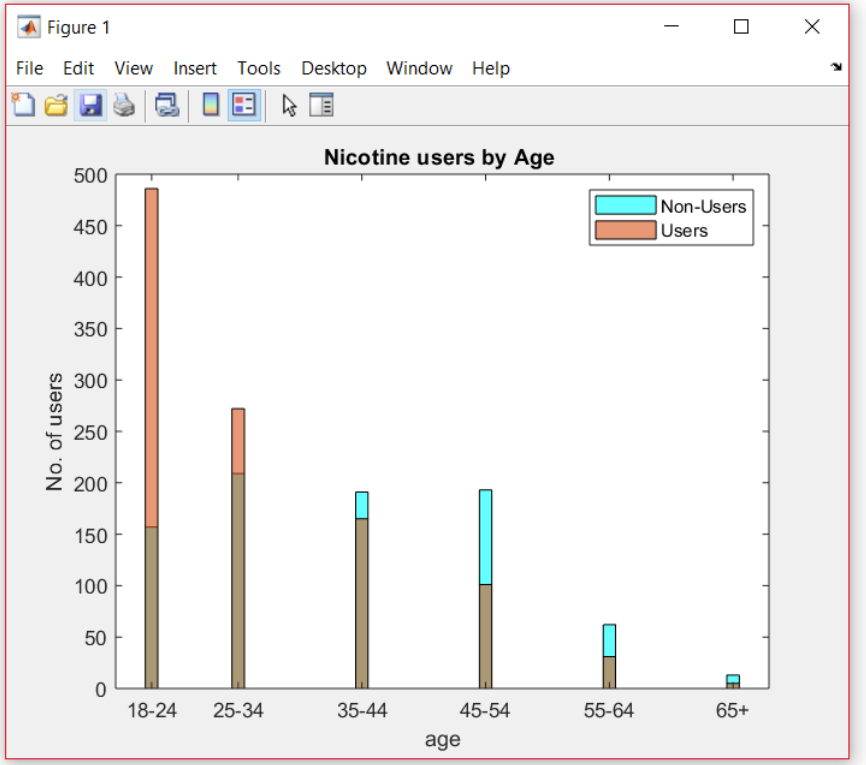


Focus on youth

Illegal…



- Legal, Safer, Branded
- Main ingredient in tobacco



Classification Models

- Naive Bayes Model
- K Nearest Neighbors with k = 3, 5, 7
- Decision Tree
- Linear Discriminant Analysis
- Ensemble Model: Adaptive Boosting
- Support Vector Machine: Linear Support Vector Machine, Standardize Feature Matrix, Nonlinear SVM with RBF Kernel, SVM Auto Kernel Scale
- Random Forest

Best classifier for each drug

Best K fold cross validation : 10 fold

Best Accuracy above 0.90

Poorest Accuracy below 0.75

Drugs with Best Classifier		
Drug Name	**Classification Model**	**Accuracy**
Alcohol	Adaboost Ensemble	0.936
Amphet	KNN Classifier k=7	0.824
Amyl nitrite	Nonlinear SVM with RBF Kernel	0.936
benzos	KNN Classifier k=7	0.782
Caffeine	Naive Bayes Model	0.973
Cannabis	Linear Discriminant Analysis	0.878
Chocolate	KNN Classifiers	0.979
Cocaine	Naive Bayes Model	0.798
Crack	KNN Classifier k=5	0.962
Ecstacy	Naive Bayes Model	0.787
Heroin	AdaBoost Ensemble Model	0.941
Ketamine	KNN Classifier k=3	0.894
Legalth	Naive Bayes Model,SVM	0.824
LSD	Linear Discriminant Analysis	0.888
Meth	KNN Classifier k=7	0.856
Mushroom	Linear Discriminant Analysis	0.846
Nicotine	Linear Discriminant Analysis	0.744
VSA	Linear Discriminant Analysis	0.952

Conclusion

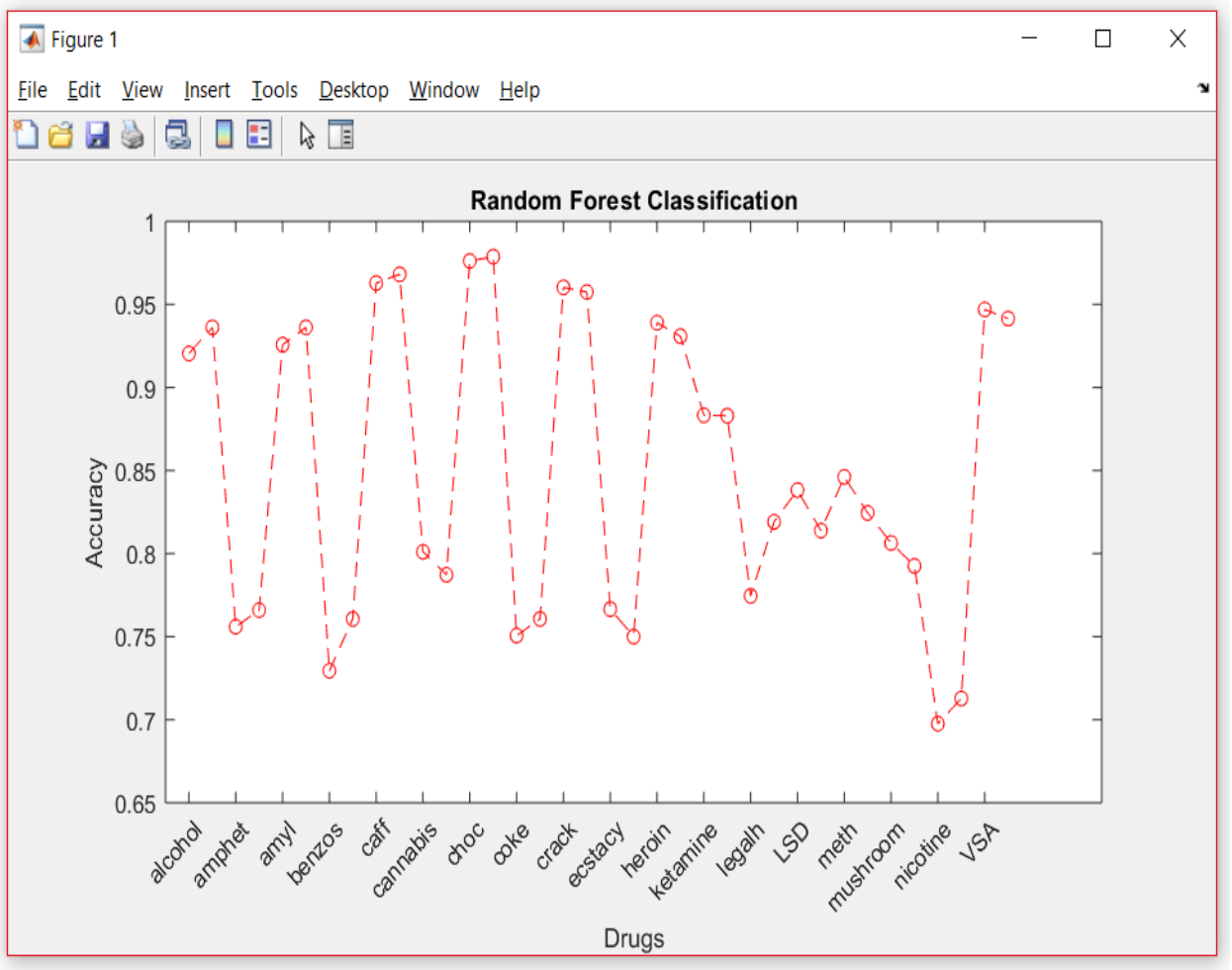
- To predict the usage of the most drugs K Nearest Neighbor's is the best classifier.
- Decision Tree is classifier with lowest accuracy for almost all drugs.
- All legal drugs has good accuracy. Highest Accuracy: 0.97 by '**Chocolate**'.
- LDA is the best classifier for Cannabis, LSD, Mushroom, Nicotine and VSA users.

Cont..

- KNN is the best classifier for Amphet, benzos, Chocolate, Crack, Ketamine and meth users.
- Naïve Bayes is the best classifier for Caffeine, Coke and Legalth users.
- For Alcohol and Heroin users Adaboost Ensemble is best classifiers.
- SVM worked best for Amyl nitrite.

Random Forest

- Works fine for all drugs but doesn't give highest accuracy for any drug.
- No significance change in both folds for any drug.

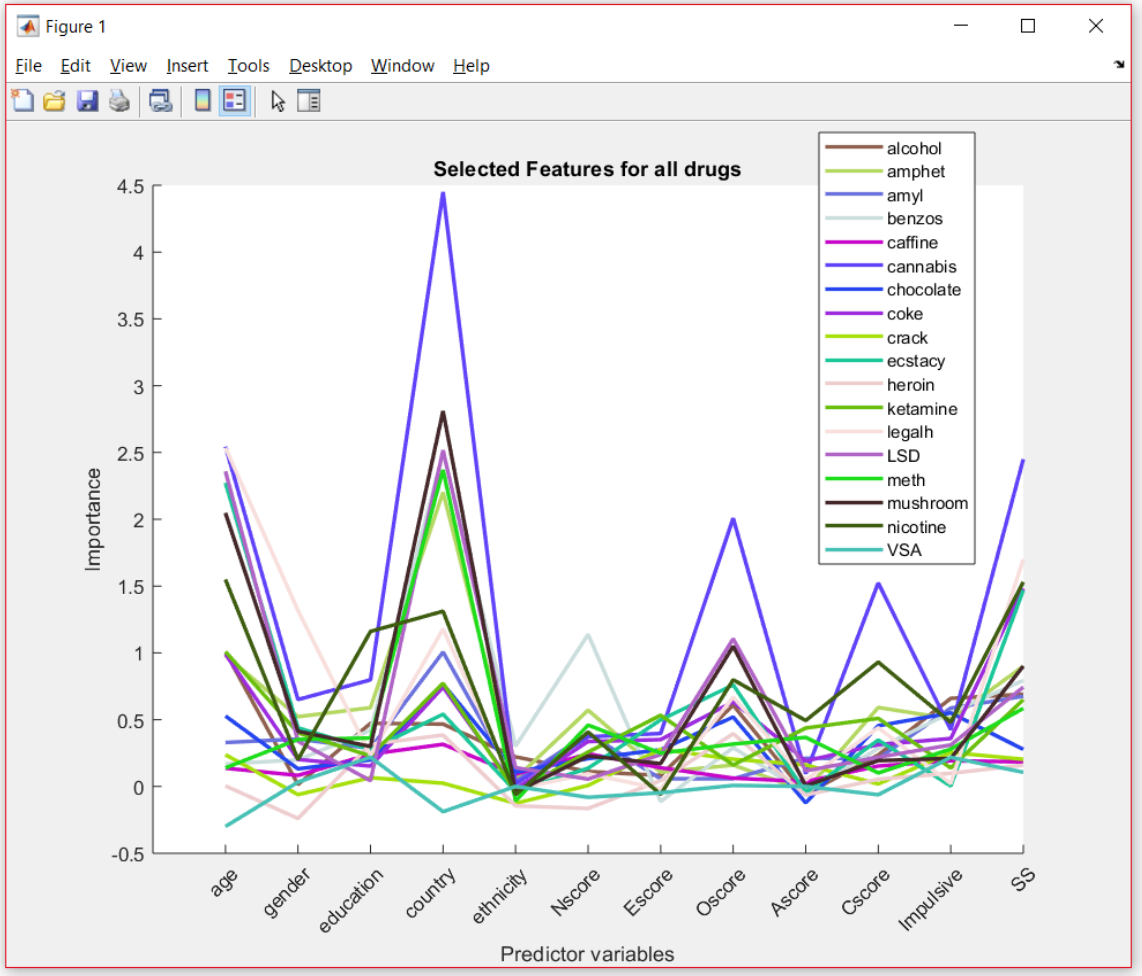


FEATURE SELECTION

- Dimensionality reduction technique.
- Useful when dealing with very high-dimensional data or when modeling with all features is undesirable.
- Used boosted and bagged decision trees approach.

Features Importance

- Age and Country as a highest priority or neutral position.
- Gender and Ethnicity Least important feature.
- Higher N and O importance.
- Higher Sensation seeking is also higher for users of recreational drugs.



Thank you !!!