Drug Consumption

Risk Assessment

What do we want to do?

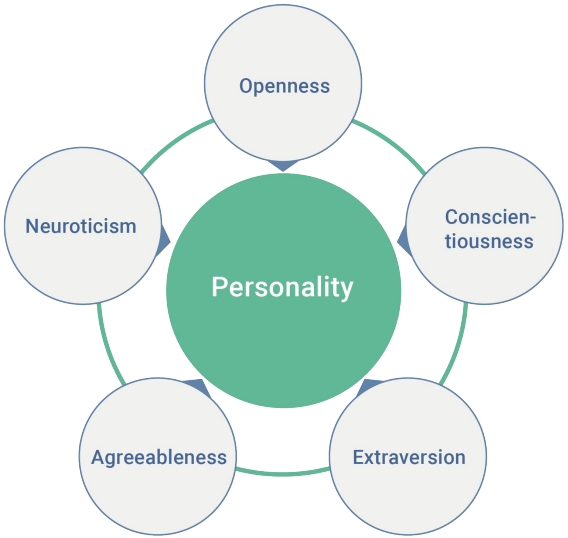
- ➔ Tool for insurance company
- ➔ Take better individual preventive measures
- ➔ Trade-off between short-term and long-term costs
- ➔ We want to have both high recall and precision

	Health Care
Tobacco[1,2]	$168 billion
Alcohol[3]	$27 billion
Illicit Drugs[4,5]	$11 billion
Prescription Opioids[6]	$26 billion

source: https://www.drugabuse.gov/drug-topics/trends-statistics/costs-substance-abuse

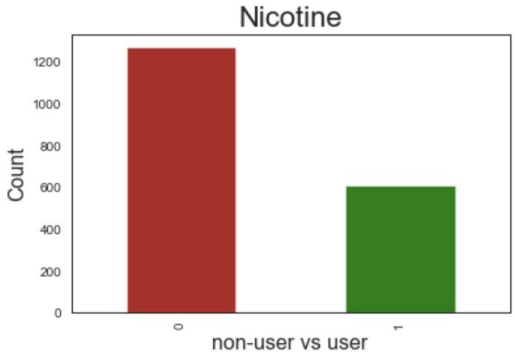
Overview of the Dataset

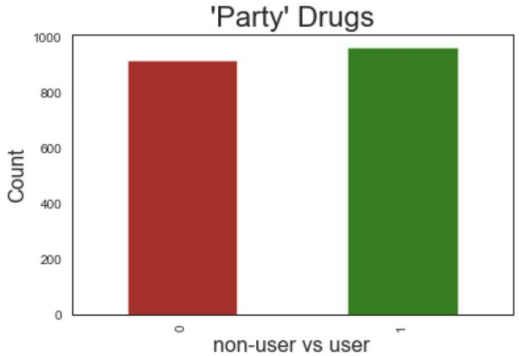
- data collected by online survey 2011-2012
- 1885 study participants
- 12 attributes
- 18 legal and illegal Drugs and one fictitious drug (to identify over-claimers)



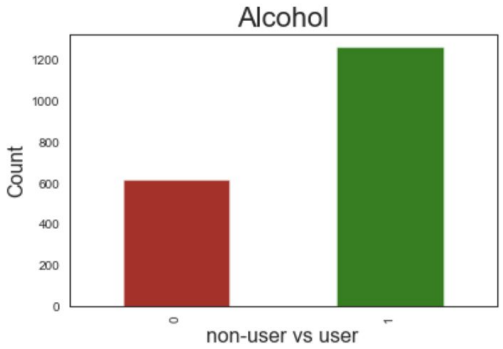
Drug						
"Never Used"	"Used over a Decade Ago"	"Used in Last Decade"	"Used in Last Year"	"Used in Last Month"	"Used in Last Week"	"Used in Last Day"

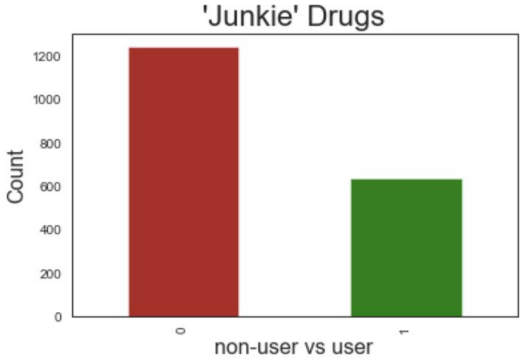
Clustering of the different drugs





-Cannabis
-Legalhighs
-Ketamine
-Shrooms
-Amphet
-Ecstasy
- LSD
-Coke
-Amyl

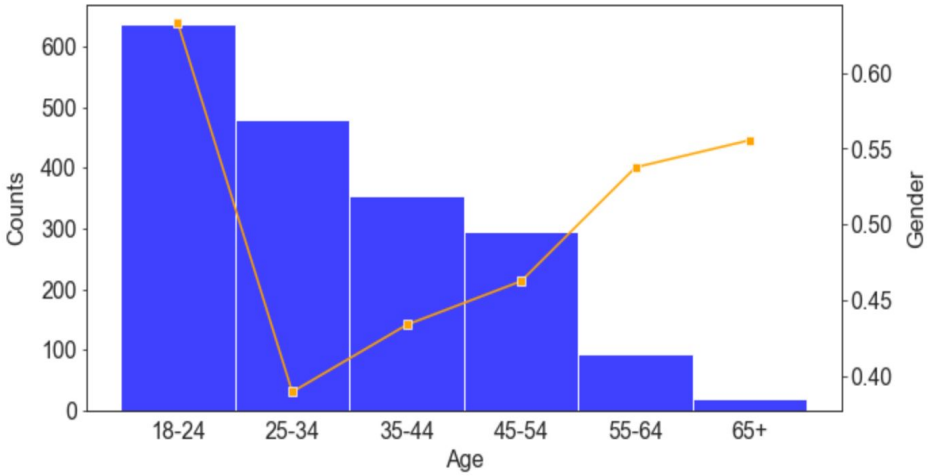




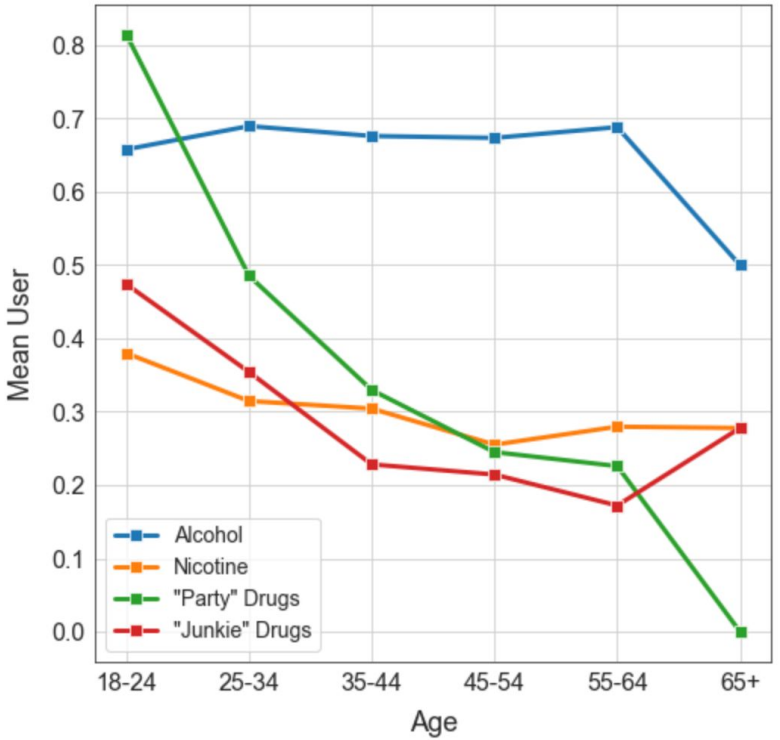
-Meth
-Benzos
-Heroin
-Crack

Gender and Age Distribution

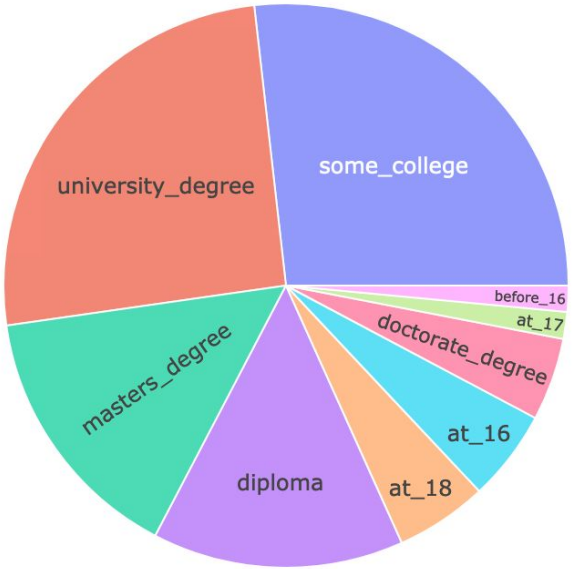
male=1, female=0

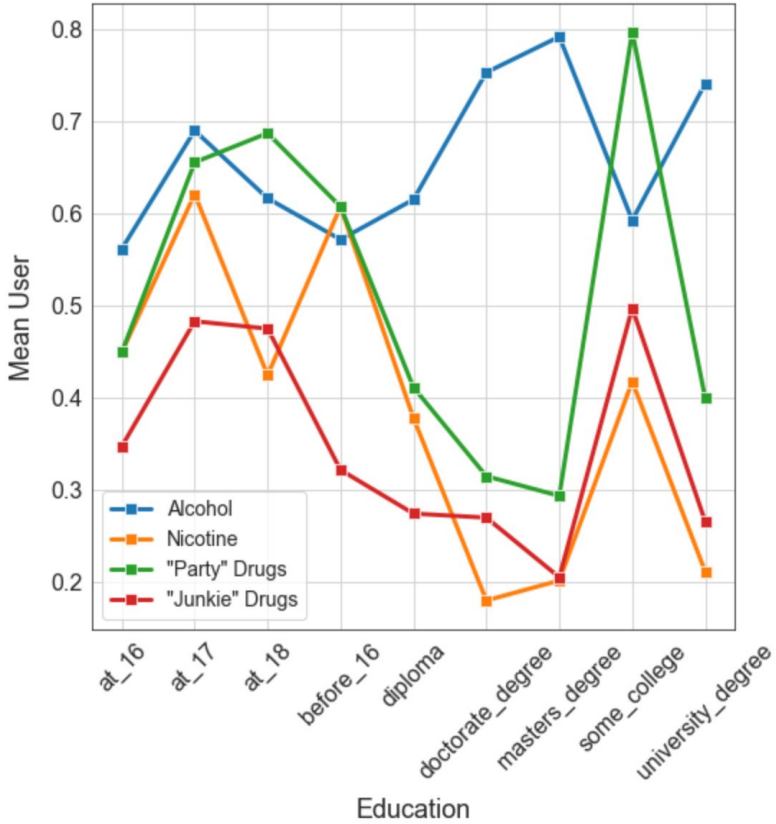


→ Overall drug usage declines with age

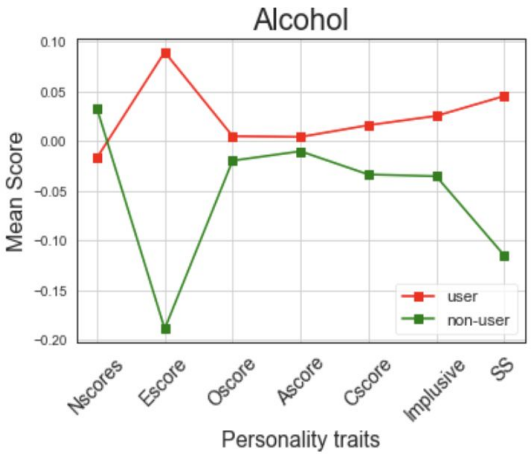


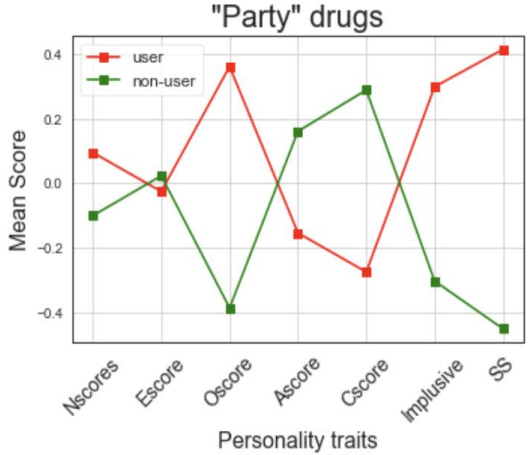
Educational Background and Drug Usage

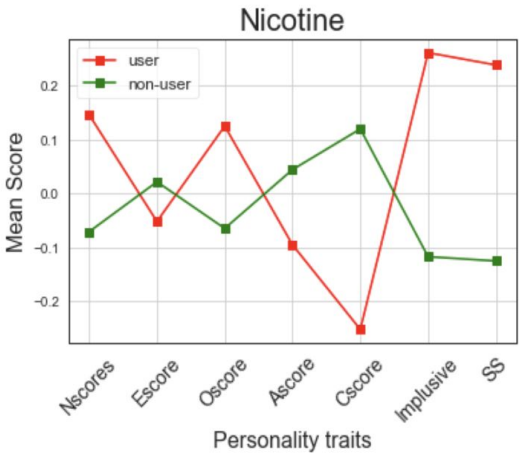


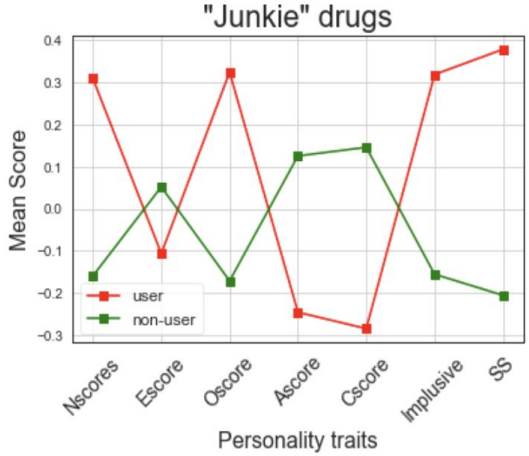


Personality traits for each drug cluster







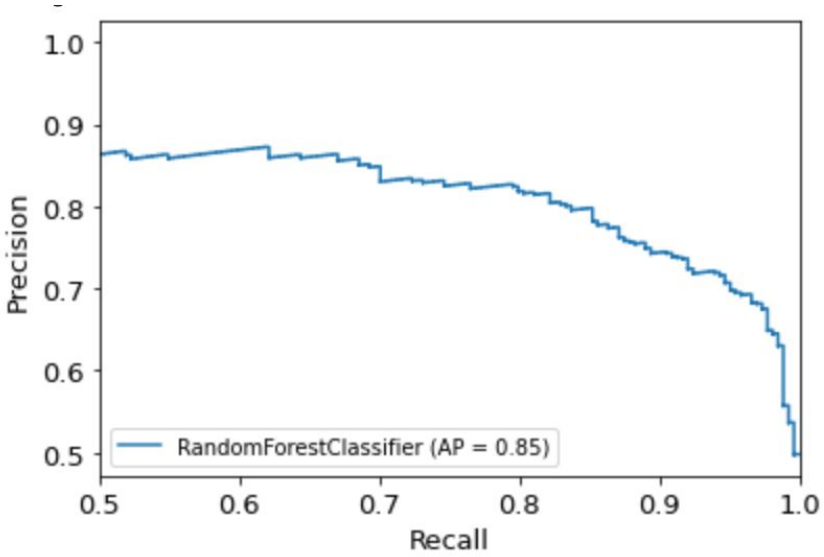


Modelling of the different Drug Cluster

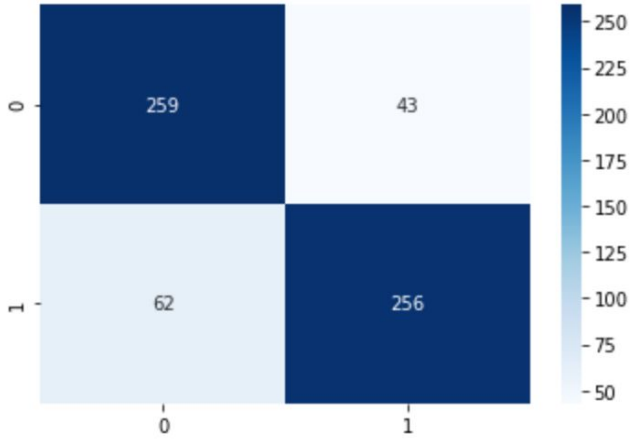
- Focus on high precision and recall → optimized on f1-score
- Started with basic models
- Different Classifiers for the Drug cluster

Modelling “Party” Drugs Cluster

- best model with Random Forest



	precision	recall	f1-score
0	0.93	0.74	0.82
1	0.72	0.92	0.81

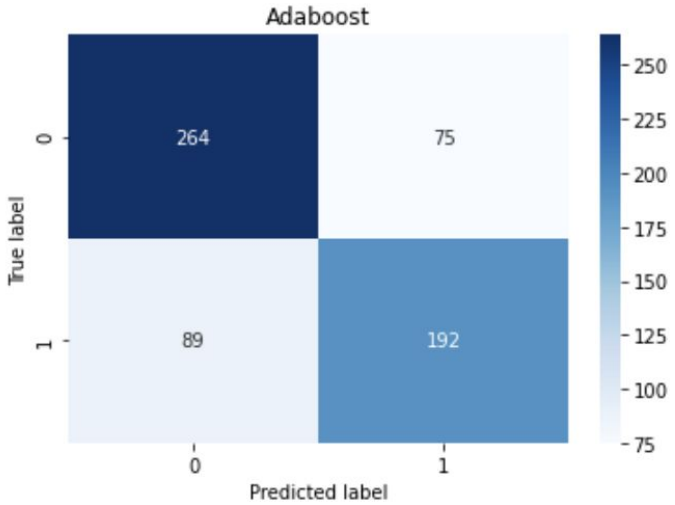


Modelling “Junkie” Drug Cluster

-Best Model: AdaBoost
-Good to predict not “Junkie” Drug user

	precision	recall	f1-score
0	0.75	0.78	0.76
1	0.72	0.68	0.70

feature	importance
Nscore	0.191667
Country	0.150000
Education	0.116667
Oscore	0.108333
Impulsive	0.083333

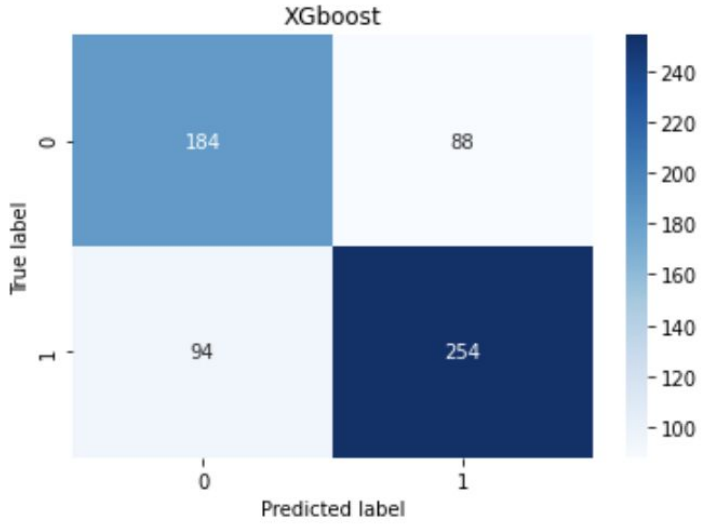


Modelling Nicotine

-Best Model: XGboost
-Good to predict smokers, bad for non-smokers

	precision	recall	f1-score
0	0.66	0.68	0.67
1	0.74	0.73	0.74

feature	importance
Country	0.189594
Age	0.165033
SS	0.109168
Education	0.082662
Cscore	0.078560

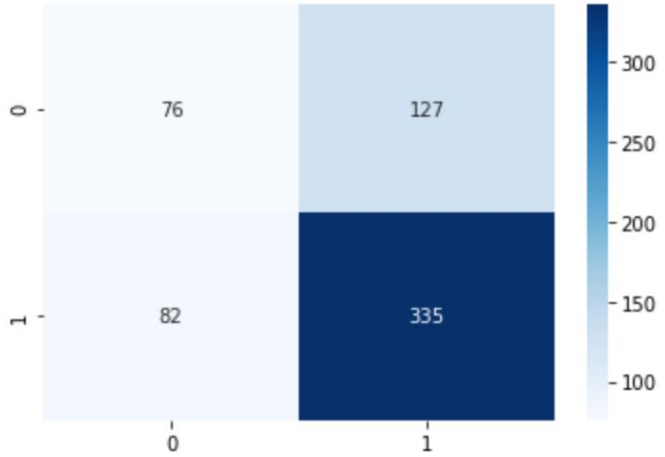


Modelling Alcohol

- good for predicting moderate to heavy drinkers, pretty bad for predicting casual and non-drinkers

feature	importance
Education	0.170616
Escore	0.126103
SS	0.106070
Nscore	0.090500
Cscore	0.088550

	precision	recall	f1-score
0	0.48	0.37	0.42
1	0.73	0.80	0.76



Summary

→ Robust Model to separate user and non-user

‘Junkie’ Drugs	Nicotine	Alcohol	‘Party’ Drugs
Bonus for non-user user: Preventive measures	Higher insurance Rate for smoker / reject	Higher insurance Rate for drinker	user: Higher insurance Rate non-user: Bonus

Future Work

- Other classification algorithms (SVM)
- Different clustering of the drugs
- Classification for individual drugs
- Collect more data to balance the Dataset

Thank you!