

High Times or Drug Crimes?



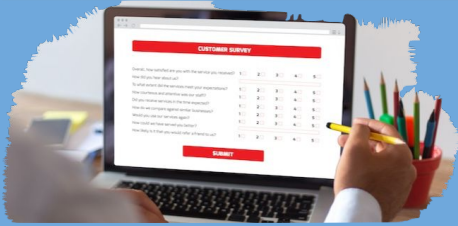
The Goal

We want to develop a system that can be used by **parents, teachers, social workers et.**

which can assess whether somebody getting caught using an **illegal drug for the first** time is at the risk of starting a **drug career**

or if the person really **only just likes to try out new stuff**

The Data



an online survey undertaken 2011 with ~
1900 respondents*

from 7 different countries



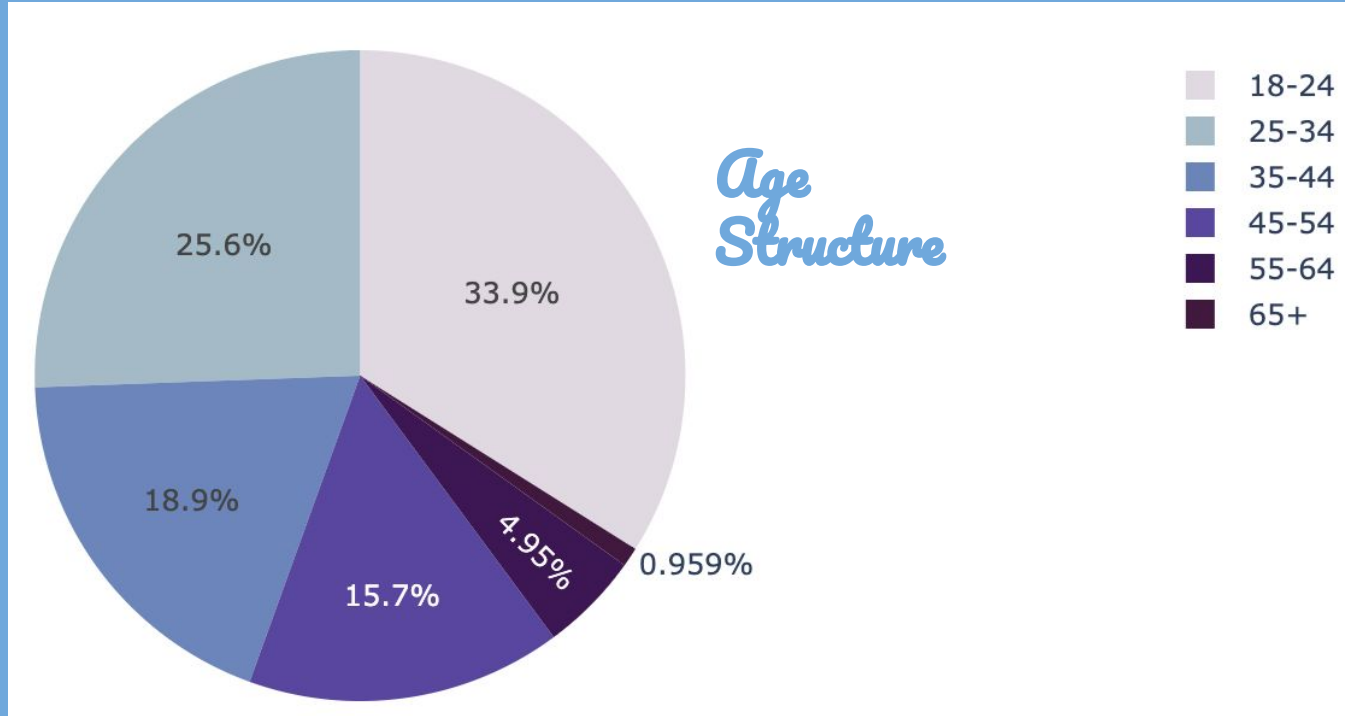
with information on 7
personality traits

and on the consumption of 18
legal and illegal drugs

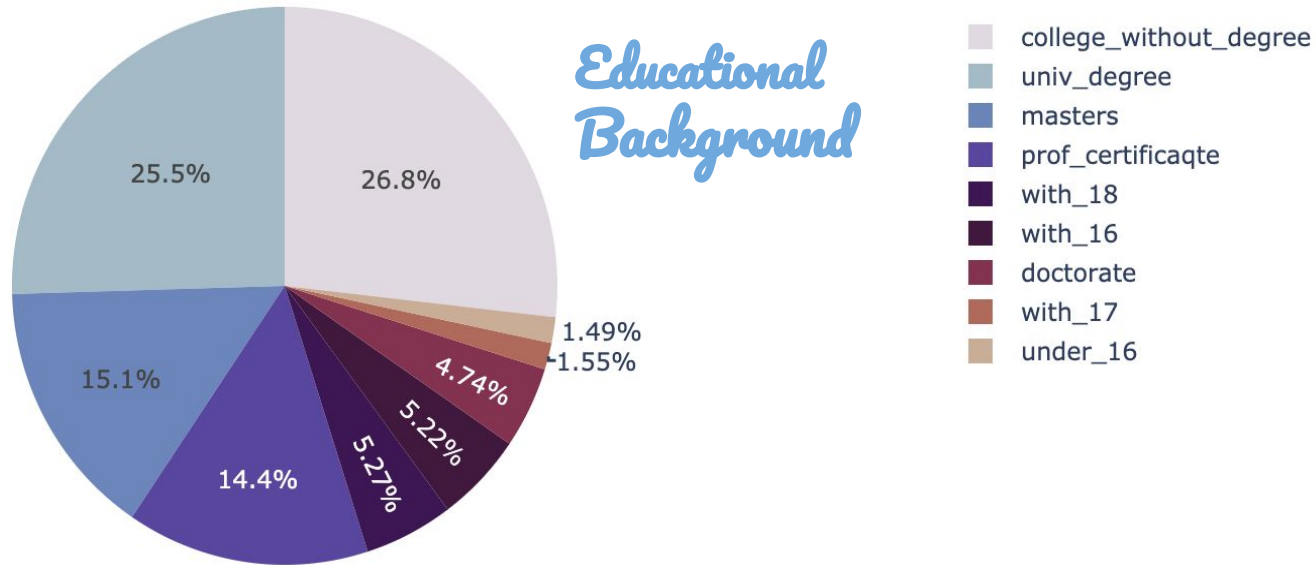


*<https://archive.ics.uci.edu/ml/datasets/Drug+consumption+%28quantified%29>

High Times or Drug Crimes?

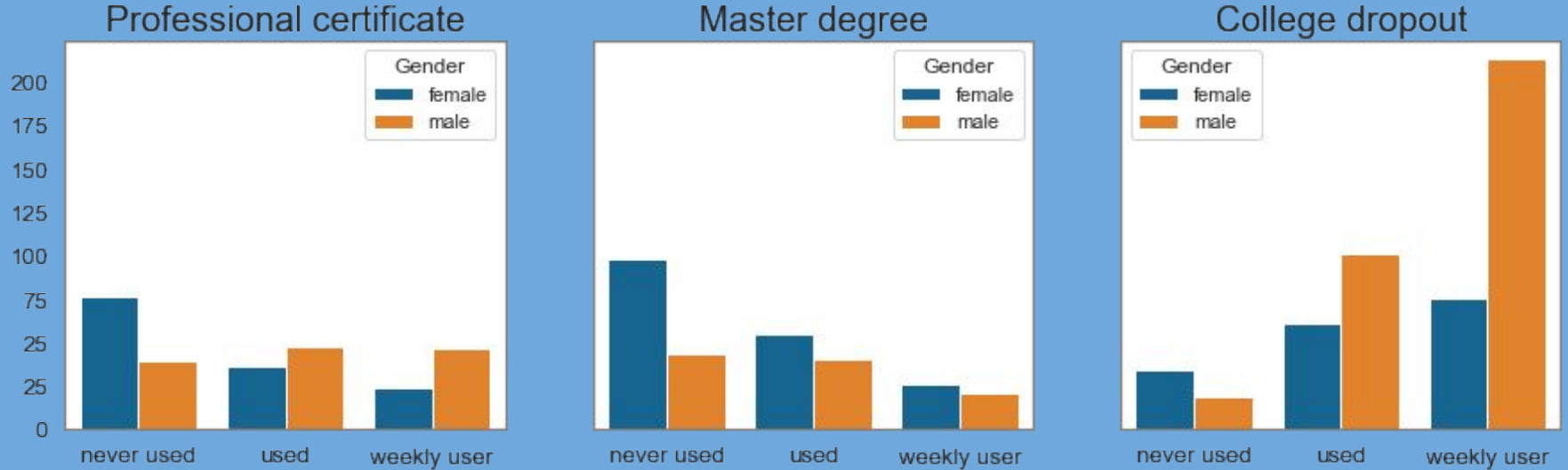


High Times or Drug Crimes?



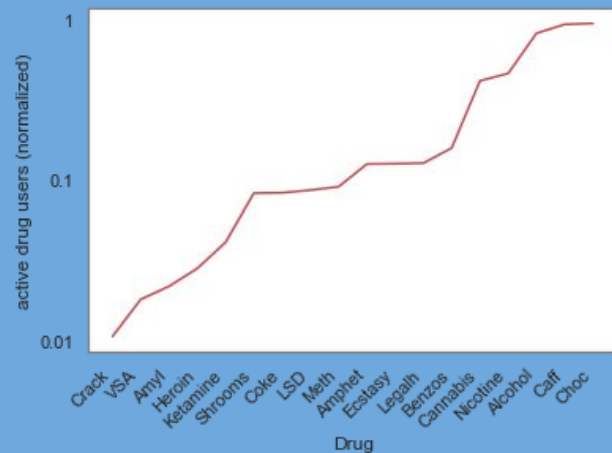
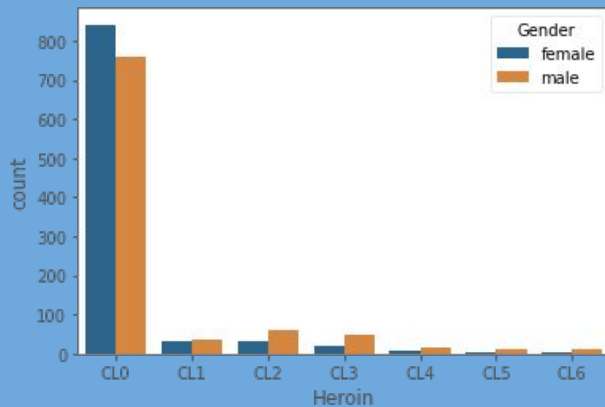
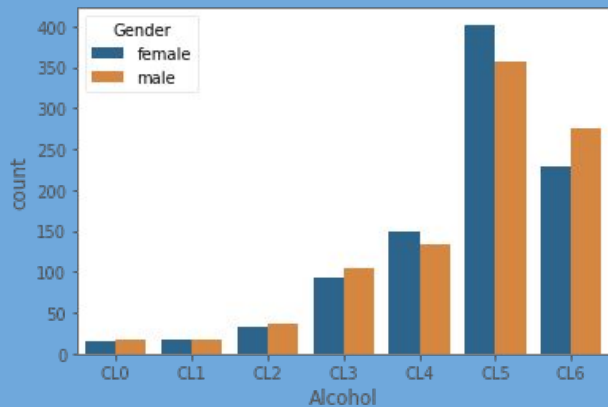
High Times or Drug Crimes?

cannabis consumption by education



High Times or Drug Crimes?

The Imbalanced data Problem?



High Times or Drug Crimes?

The Onset of illegal Drug Addiction?

For classification purposes, we had to find a **subset** of respondents who can be thought of **having an illegal drug addiction**

we only considered people that **consumed at least one illegal substance in the last week**



The Onset of illegal Drug Addiction?

thus, we identify **855** respondents that most likely exhibit a **substance addiction** vs 1022 that do not

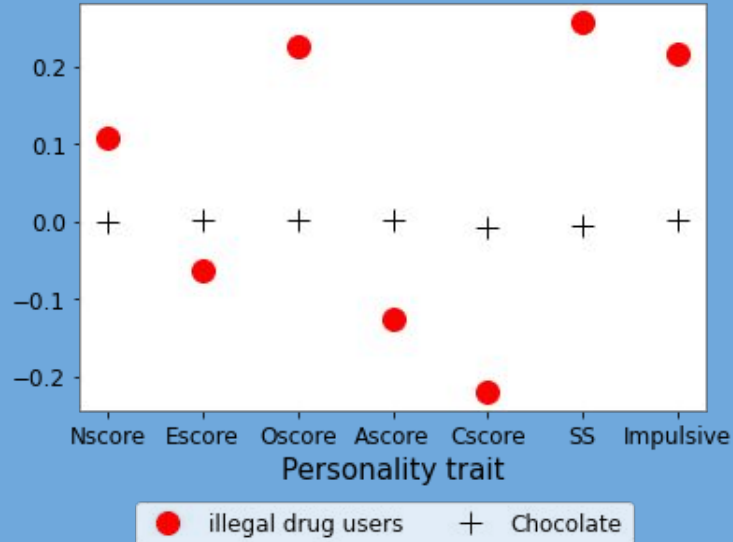
our best classification model for the test data is a **tuned Extreme-Gradient-Boost-Classifer**, with a **recall of 81%**



High Times or Drug Crimes?

The Onset of illegal Drug Addiction?

Mean personality traits of illegal drug users



Drug Cultures

since our model is not exactly fine grained when it comes to mapping **specific substance abuse** to sociodemographic background and personality traits, we introduce **three different drug culture-clusters**

The Party Cluster

for the party drug cluster, we only considered survey participants who used **ecstasy, ketamine, amphetamine and/or cocaine in the last month**

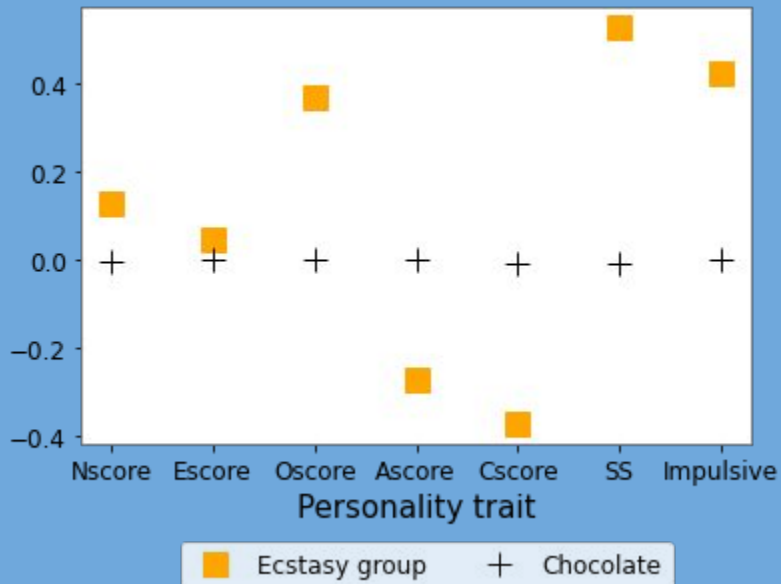
given that this classification was very unbalanced (462 vs. 1415), we had to use **synthetic minority oversampling** for our training data



High Times or Drug Crimes?

The Party Cluster

Mean personality traits of Ecstasy group members



High Times or Drug Crimes?

The Party Cluster

our best classifier was a **tuned random forest model**, which yields a **recall of 70 %**



The Stoner Cluster

for the stoner cluster, we only considered couch potatoes who used **cannabis, magic mushrooms and/or lsd in the last week**

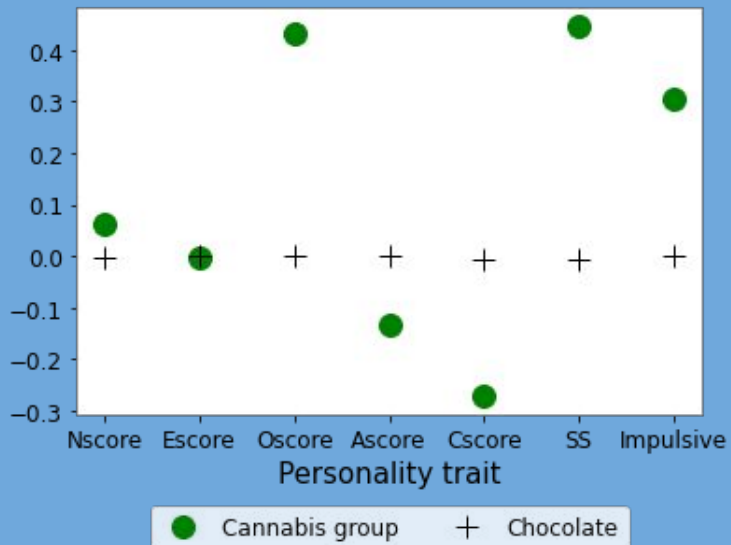
given that this classification was very unbalanced as well (670 vs.1207), we again used **synthetic minority oversampling**



High Times or Drug Crimes?

The Stoner Cluster

Mean personality traits of Cannabis group members



High Times or Drug Crimes?

The Stoner Cluster

with a **tuned random forest classifier**, we were
able to get a **recall** of **78%**

wicked!



The Junkie Cluster

here, we grouped people that used **heroin, crack, meth and/or benzos** in the last week

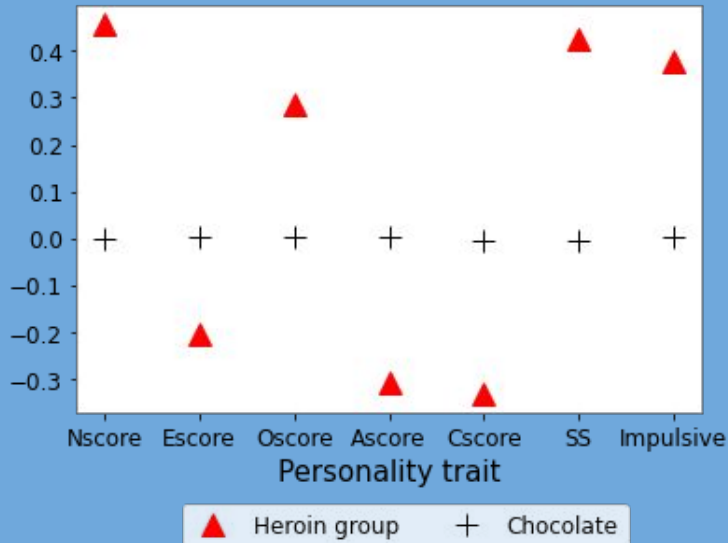
luckily, these people were a **clear minority** (259 vs. 1618), so we constructed some artificial junkies for training purposes



High Times or Drug Crimes?

The Junkie Cluster

Mean personality traits of Heroin group members



The Junkie Cluster

our best model was a **tuned random forest classifier** with a poor **recall of 35%** (27 out of 78 cases in the training data classified correctly)



High Times or Drug Crimes?

Modelling Single Drugs

turned out to be very difficult

we invested quite some time and energy
(computer and brain wise) in **modelling**
heroin and alcohol

but our best models **sacrificed precision**
for recall way too much



The Conclusion

- our general model works pretty well:
we should be able to evaluate if a situation calls for intervention measures
- **but:** personal assessment is still extremely important
- much more data concerning this topic should be collected

High Times or Drug Crimes?

**Thank
You!**