

Drug Consumption

Explained and Predicted (Preliminary Version by Jasmin)

Definitions

- Drug: A chemical that influences biological function
- Legal substances sugar, alcohol, and tobacco
- Substance Use Disorder (DSM)
- Five Factor Model for describing a personality
- Other Psychometrics: Sensation Seeking and Impulsiveness

The Data

- Comprised of 1885 respondents and their usage of 18 drugs
- Differentiation the substance users on the basis of recency of use
- No identification of depth of substance dependence
- Five Factor Model Personality Traits
 - Neuroticism
 - Extraversion
 - Openness to Experience
 - Agreeableness
 - Conscientiousness
- Impulsivity and Sensation Seeking
- Demographics: Age, Gender, Country, Ethnicity
- All data were centered and scaled to Standard Deviation of approximately 1

Goal of Data Exploration and Modeling

Finding the link between personality traits, gender, education, nationality, age and the risk of substance use

This Dataset can tell many stories

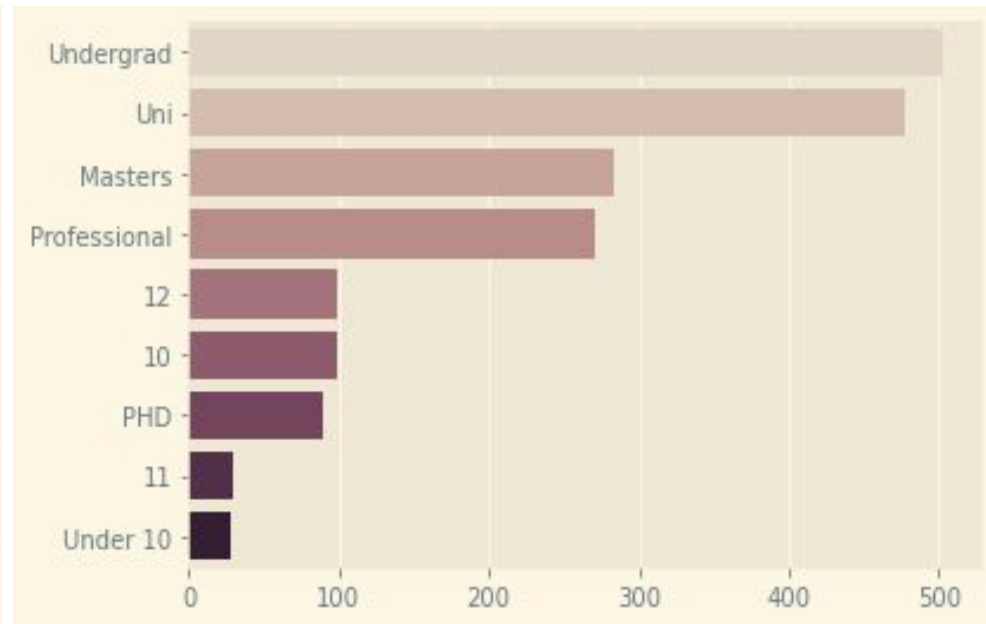
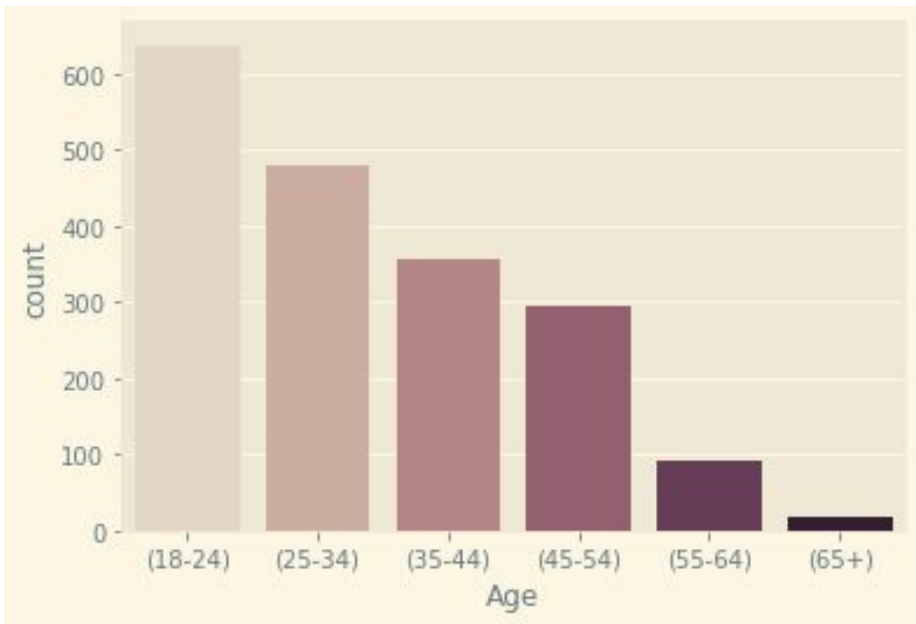
We looked at special case for Modeling: Prediction of Cannabis Use for a person

Reason: Most instances for this drug (except the 'everyday' stimulants like chocolate)

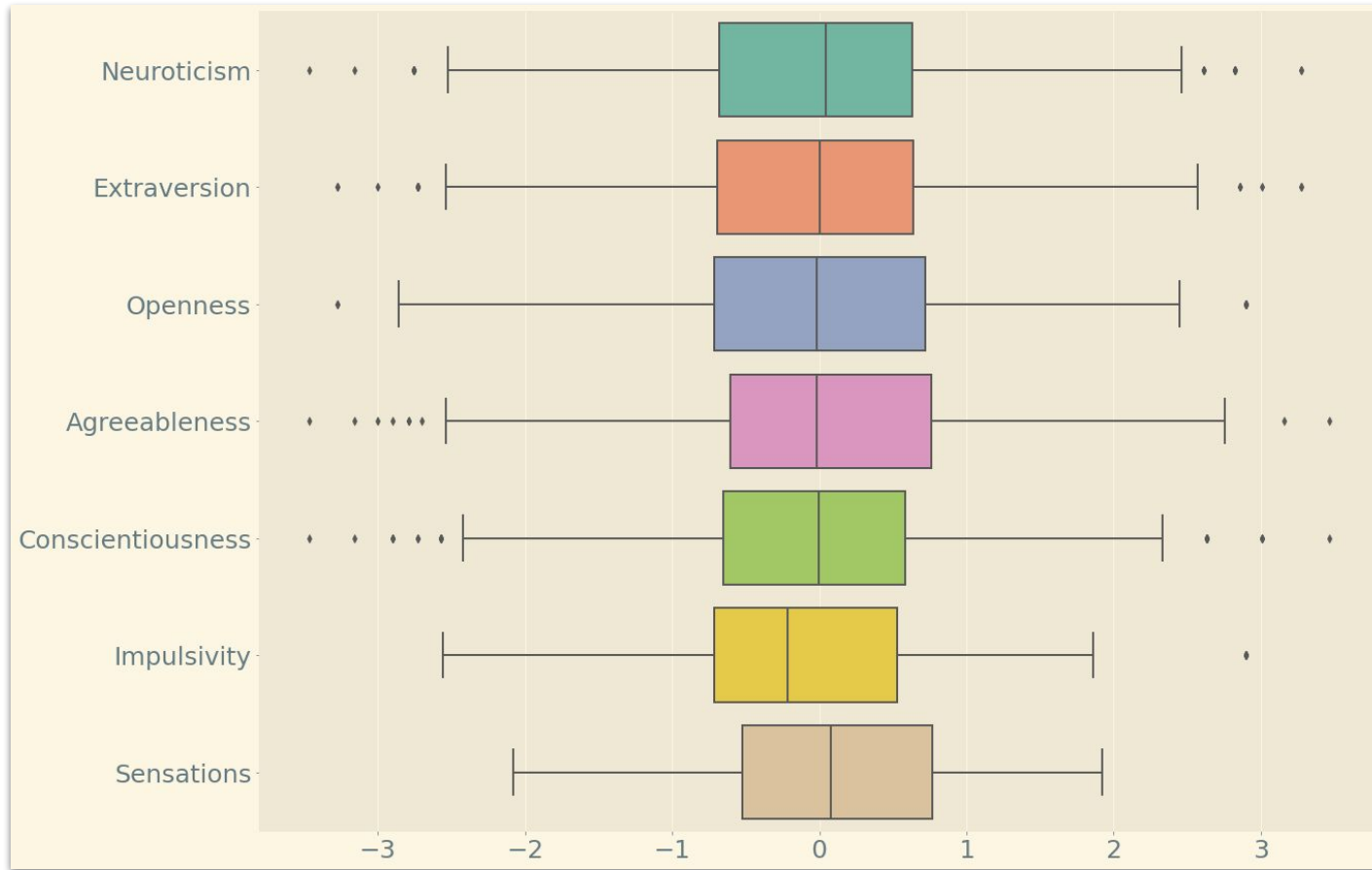
Method

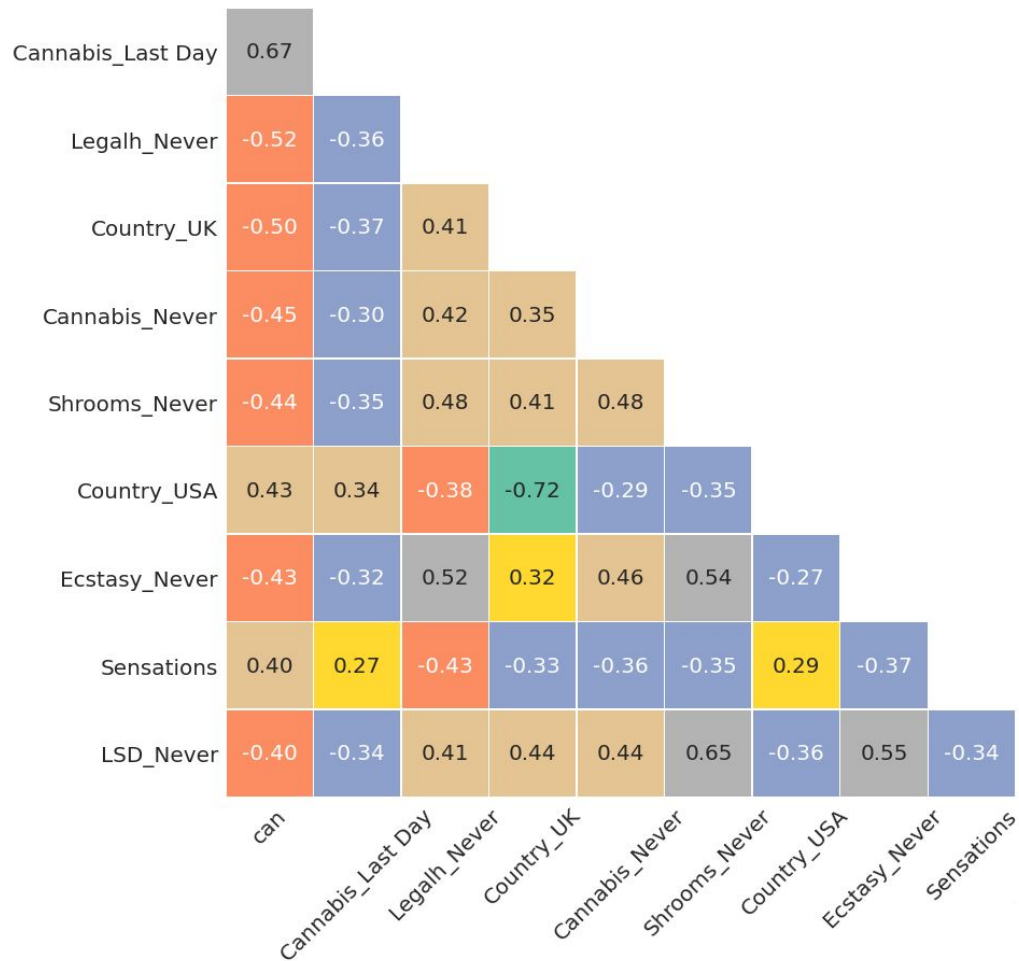
1. Clean Data, Change Names and Categories for descriptive Analysis
2. Start with EDA for all interesting characteristics of Dataset
3. Pick a target for Prediction
4. Prepare Dataset for task
5. Select Algorithm
6. Tune Hyperparameters
7. Interpretation
8. Visualization of Findings

Distribution of Age and Education Level



Overview of Personality Traits

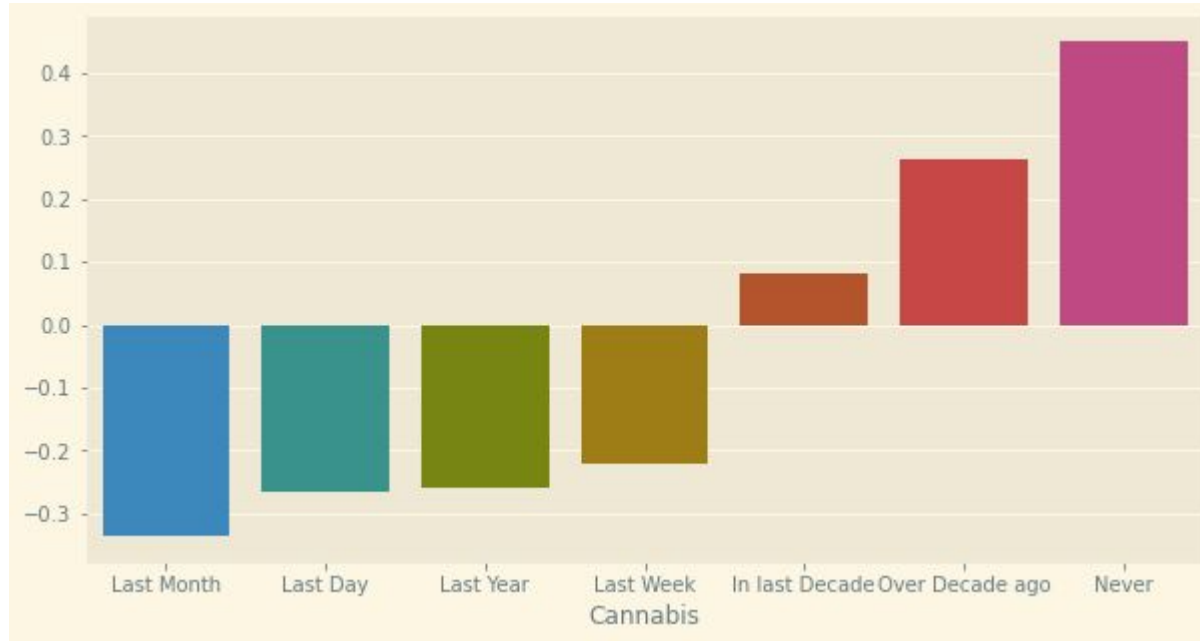




Most interesting
Variables for
Prediction of
Cannabis Use (can)

Personality Traits and Cannabis Use

Conscientiousness

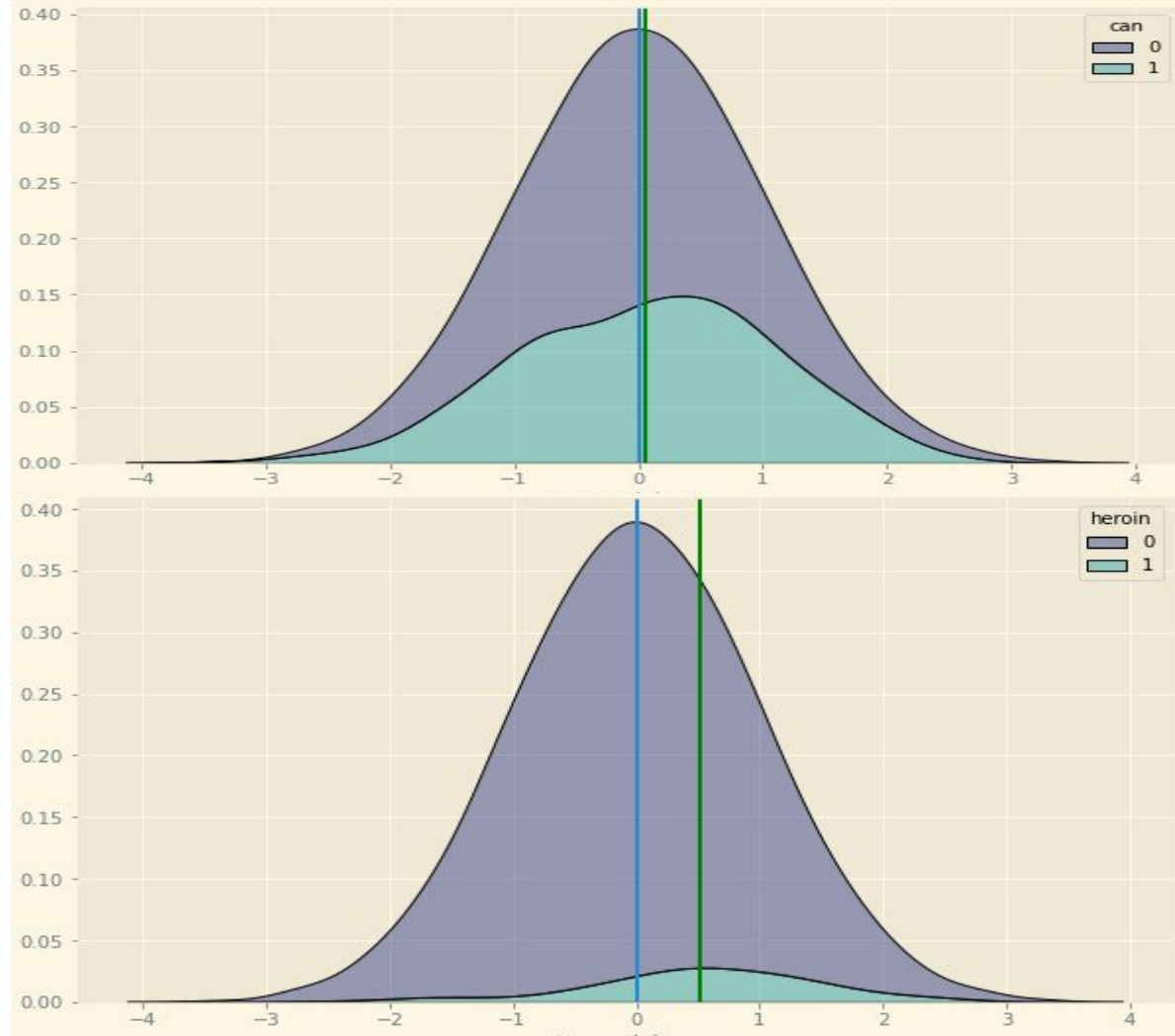


Personality Traits and Cannabis Use

Openness



Neuroticism Score of Cannabis and Heroin Users



Less well represented variables

Country

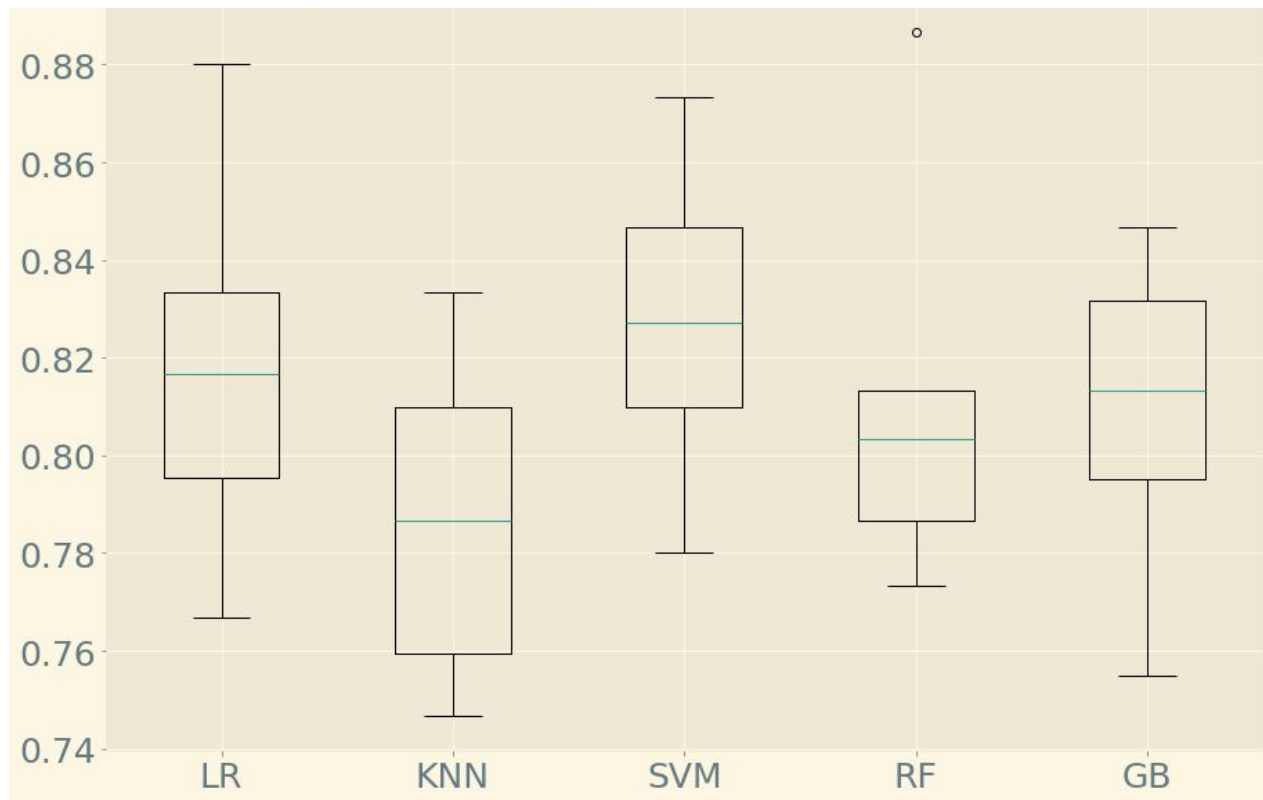
Ethnicity

Age: Mostly young people (?) between 18 and 24

Receiver Operating Characteristic

Interpretation of Effects on Target with Feature Importance

Algorithm Selection on Accuracy Score



Conclusion and further research

It could be shown, that the prediction of a propensity to cannabis use could be linked to personality traits and predicted quite well with an Accuracy Score of 0.82.

The best Classifier was the Support Vector Classifier.

Ideas for further research

- Association between personality traits and other drugs
- Effect of other demographics on drug usage
- Interaction Effects between drugs (People who like MJ don't drink)