Drug Use Prediction: A Classification Study

Analysis by Ben Geissel

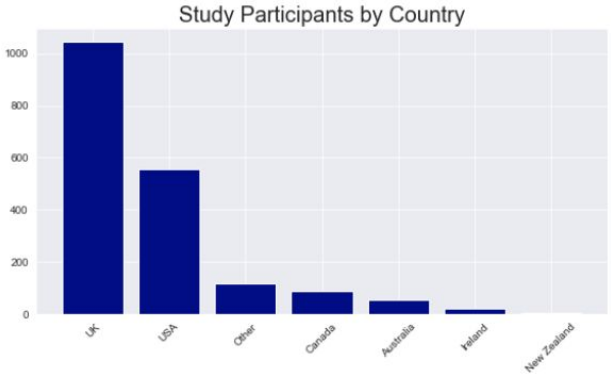
How can we identify those susceptible to drug use?

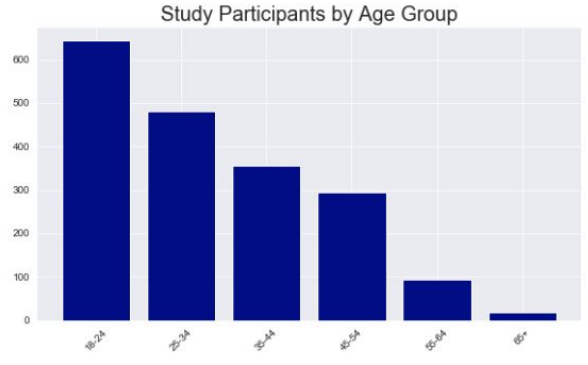
- Demographic Traits: Age, Country, Education Level, etc…
- Personality Traits: Neuroticism, Extraversion, Agreeableness, etc…
- With this information rehab clinics, hospitals, and others are able to identify who may be at risk for certain types of drug use and abuse
- Case studies:
  - Cannabis
  - Heroin

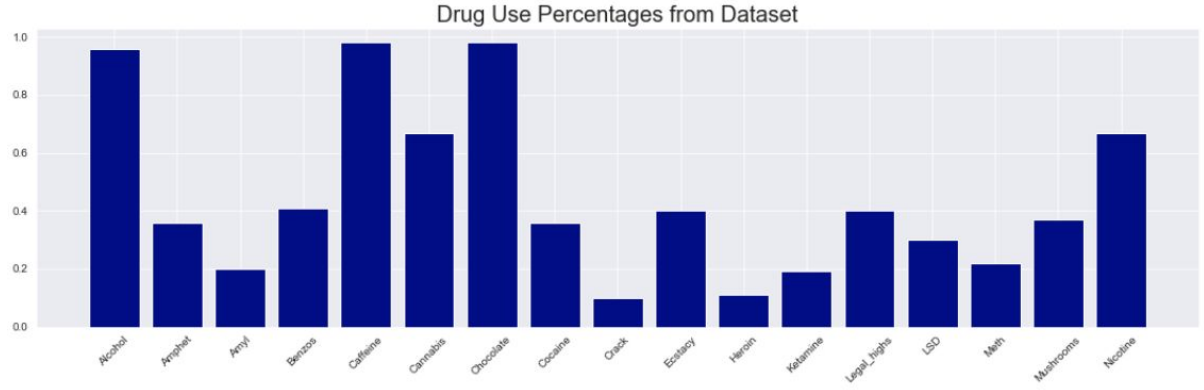
Data Source

- 1885 data points
- 13 features:
  - Demographic Traits
  - Personality Traits
  - Deception Test
- Drug use data for 17 different drugs
  - Made binary (Never Used & Used Over a Decade Ago → Non-User)
- Source:
  - Data comes from UCI Machine Learning Repository
  - Originally collected by Elaine Fehrman, Vincent Egan, and Evgeny M. Mirkes during a 2015 study

Data Insights







50/50 Gender split

0.3% Fail Deception Test

Model Evaluation

- Classifier models can be evaluated by many metrics
  - AUC
  - Precision
  - Recall
  - Accuracy
  - F1 Score
- Focus on Accuracy in this study
  - Ability to pinpoint user vs non-user
- Limit False Positive results
  - Don’t want to say someone is a user when they are not

Case Studies

Cannabis

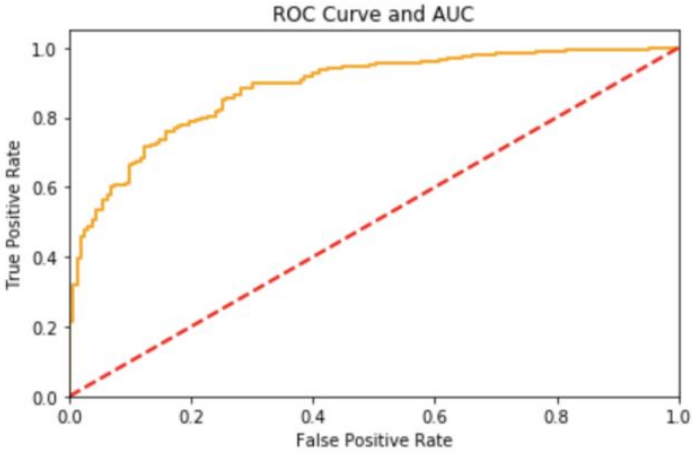
- 67% of study participants have used Cannabis
- Legalization in many US States
- New research will be conducted on addictions and treatments

Heroin

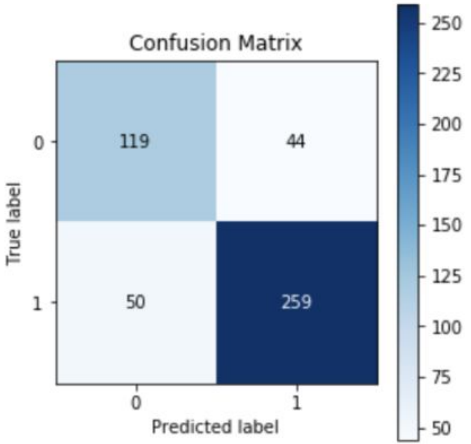
- 11 % of study participants have used Heroin
- Opioid Epidemic
- More treatment sought out for opiates than for any other drug (including Alcohol)

Case Study: Cannabis

Metric	Best Model	Score
Precision	SVM	91.5%
Recall	Random Forest	83.8%
Accuracy	Random Forest	80.1%
F1 Score	Random Forest	84.6%
AUC	Logistic Regression	88.3%



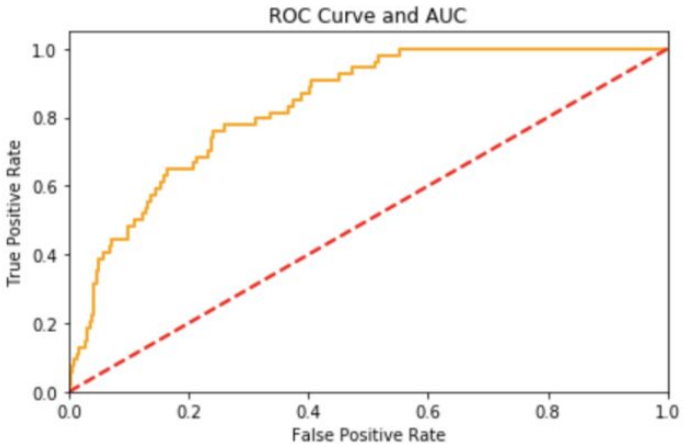
Logistic Regression



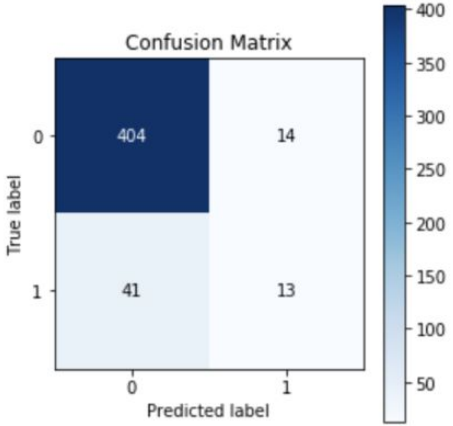
Random Forest

Case Study: Heroin

Metric	Best Model	Score
Precision	Random Forest	48.2%
Recall	Naive Bayes	90.7%
Accuracy	Random Forest	88.4%
F1 Score	Logistic Regression	41.6%
AUC	Logistic Regression	83.4%



Logistic Regression



Random Forest

Model Insights - Feature Importance

Cannabis

Best Predictors from Random Forest Model:

1. Sensation Seeking Score
2. Living in the UK
3. Openness to New Experiences Score
4. Conscientiousness Score
5. Neuroticism Score

Heroin

Best Predictors from Random Forest Model:

1. Living in the US
2. Living in the UK
3. Impulsiveness Score
4. Openness to New Experiences Score
5. Sensation Seeking Score

Conclusions

- Personality and demographic traits can help determine susceptibility to using a certain drug
  - Demographics: Location
  - Personality: Sensation Seeking and Openness to New Experiences
- Having the ability to pick a model based on drug of interest and performance metric helps generate quick reliable results
- Rehab clinics, hospitals, and more can benefit from these model insights to create effective tailored treatment plans for individuals