Interim Report: **Detecting Drug Consumption Risk**

A machine learning project by Mirko Knoche and Nina Notman, by request of the *American Psychology Association (APA)*

REQUEST BY THE APA

REQUEST BY THE APA

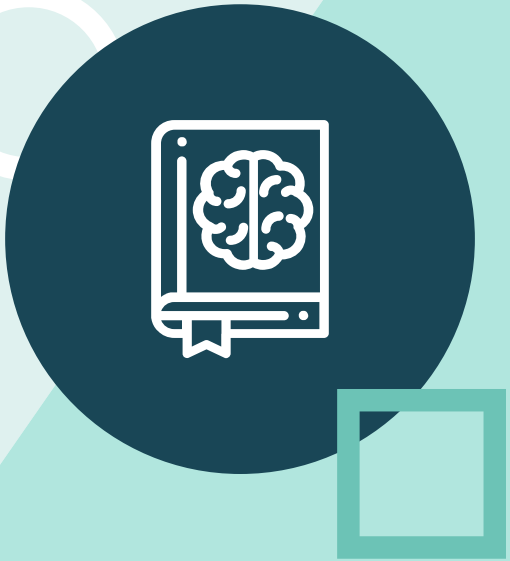
> I wish I could identify patients at risk for cannabis consumption so it won't influence therapy at a future time point.

REQUEST BY THE APA

DATASET BY ELAINE FEHRMAN

> I wish I could identify patients at risk for cannabis consumption so it won't influence therapy at a future time point.

THE DATASET



FEATURES

DEMOGRAPHICS

Age, Gender, Education, Ethnicity, Country

NEO-FF-R

Neuroticism, Extraversion, Openness, Agreeableness, Conscientiousness

BIS-11 & ImpSS

Impulsiveness, Impulsiveness-Sensation-Seeking

LABELS

PSYCHOTIC DRUGS

Alcohol, Amphetamines, Amyl Nitrite, Benzodiazepines, Cannabis, Chocolate, Cocaine, Caffeine, Crack, Ecstasy, Heroin, Ketamine, Legal Highs, LSD, Methadone, Mushrooms, Nicotine, VSA

PREDICTION OF CANNABIS CONSUMPTION RISK

> I wish I could identify patients at risk for cannabis consumption so it won't influence therapy at a future time point.

PREDICTION OF CANNABIS CONSUMPTION RISK

MODEL SELECTION

- K-nearest Neighbors
- Logistic Regression
- Random Forest Classifier
- Support Vector Classifier
- XGBoost

PREDICTION OF CANNABIS CONSUMPTION RISK

MODEL SELECTION

- K-nearest Neighbors
- Logistic Regression
- Random Forest Classifier
- Support Vector Classifier
- XGBoost

MODEL EVALUATION

- Classification Report
- Confusion Matrix
- ROC Curve & AUC Score

PREDICTION OF CANNABIS CONSUMPTION RISK

MODEL SELECTION

- K-nearest Neighbors
- Logistic Regression
- Random Forest Classifier
- Support Vector Classifier
- XGBoost

MODEL EVALUATION

- Classification Report
- Confusion Matrix
- ROC Curve & AUC Score

MODEL TUNING

- Grid Search (scorer = F1)
- Feature Importance

“

I wish I could identify patients at risk for cannabis consumption so it won’t influence therapy at a feature time point.

“

I wish I could identify patients at risk for cannabis consumption so it won’t influence therapy at a feature time point.

BEST MODEL

XGBoost after GridSearch

> I wish I could identify patients at risk for cannabis consumption so it won't influence therapy at a feature time point.

BEST MODEL

XGBoost after GridSearch

> I wish I could identify patients at risk for cannabis consumption so it won't influence therapy at a feature time point.

BEST MODEL

XGBoost after GridSearch

- Identify 4 out of 5 correctly as cannabis consumer.
- Fine. But let's not leave the one hanging. We can do better!

PREDICTION OF HIGHLY ADDICTIVE DRUG CONSUMPTION RISK

PREDICTION OF HIGHLY ADDICTIVE DRUG CONSUMPTION RISK

CHANGE LABEL

HIGHLY ADDICTIVE PSYCHOTIC DRUGS

Ecstasy, Heroin, LSD, Meth, Cocaine

PREDICTION OF HIGHLY ADDICTIVE DRUG CONSUMPTION RISK

CHANGE LABEL

HIGHLY ADDICTIVE

PSYCHOTIC DRUGS

Ecstasy, Heroin, LSD, Meth, Cocaine

SAME FUNCTION

MODEL SELECTION

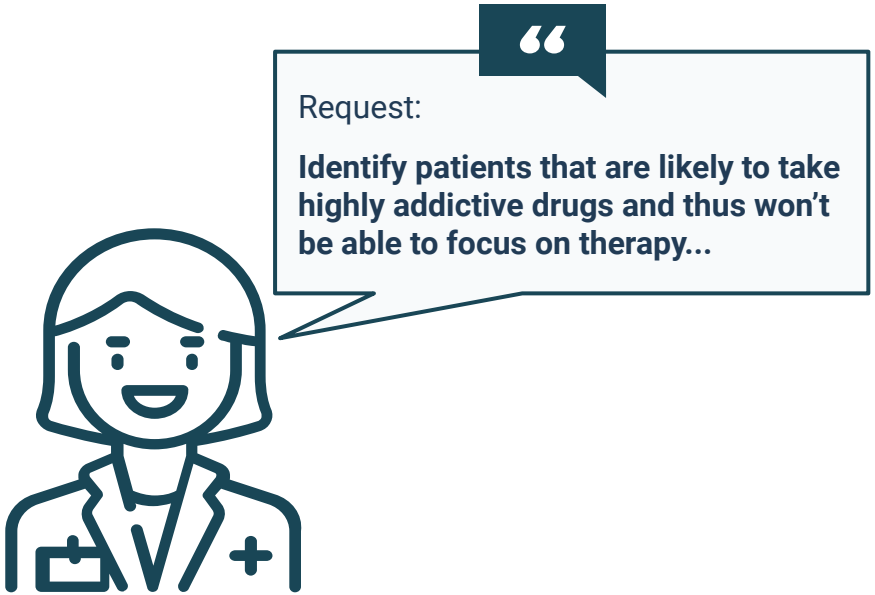
- K-nearest Neighbors
- Logistic Regression
- Random Forest Classifier
- Support Vector Classifier
- XGBoost

MODEL EVALUATION

- Classification Report
- Confusion Matrix
- ROC Curve & AUC Score

MODEL TUNING

- Grid Search (scorer = F1)
- Feature Importance



BEST MODEL

SVM after GridSearch

Only 75% accuracy on predicting consumption of highly addictive drugs.

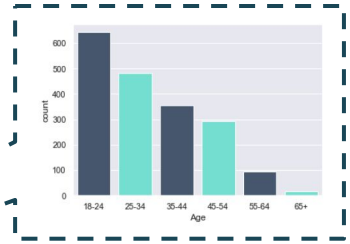
Future Work: Feature Engineering

WATCH OUT!

WATCH OUT!

ONLY FOR WHITE, ENGLISH SPEAKING PEOPLE!

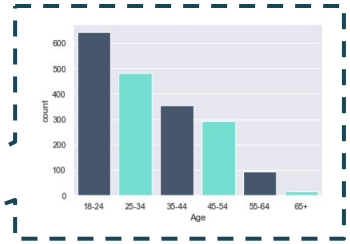
FUTURE WORK



DISTRIBUTION

Age / Country / Ethnicity

FUTURE WORK



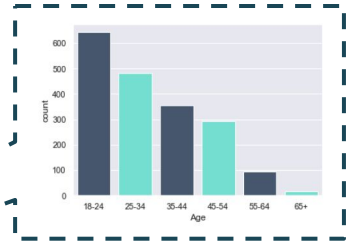
DISTRIBUTION

Age / Country / Ethnicity

COLLECT MORE DATA

Expand data collection to optimize and generalize model

FUTURE WORK



DISTRIBUTION

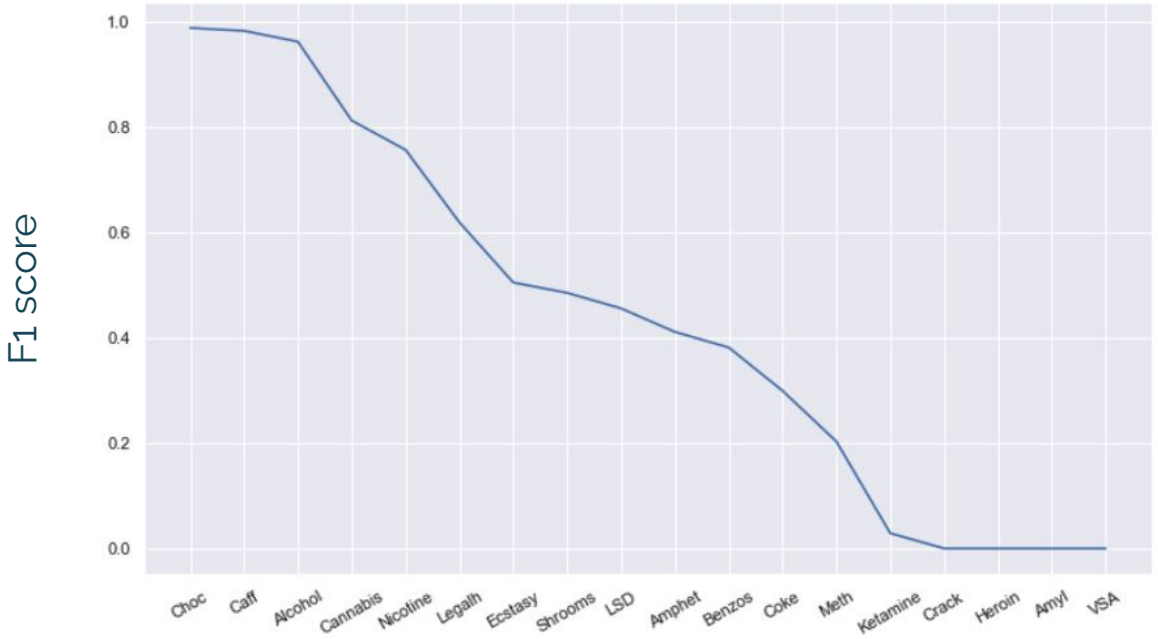
Age / Country / Ethnicity

COLLECT MORE DATA

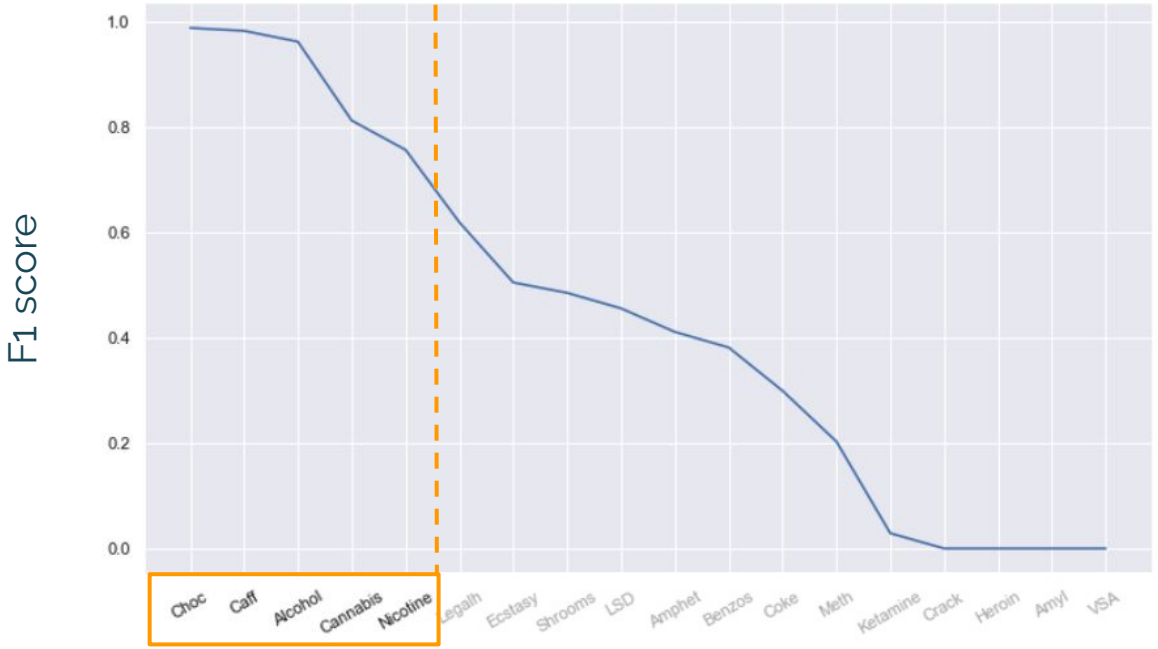
Expand data collection to optimize and generalize model

FEATURE ENGINEERING

GROUPING OF DRUGS



GROUPING OF DRUGS



LET'S TALK!

Nina Notman
nina.notman@gmx.net

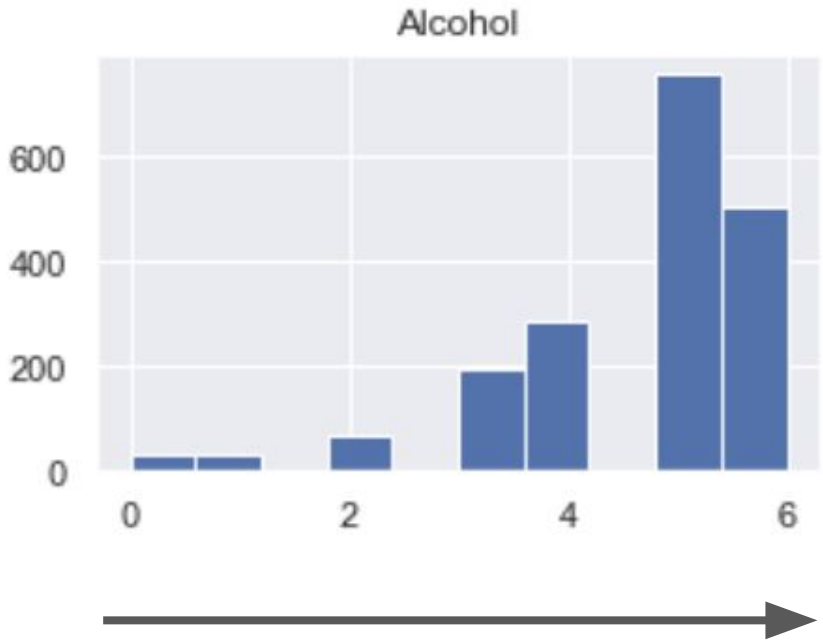
Mirko Knoche
mirko.knoche@posteo.de

THANK YOU!

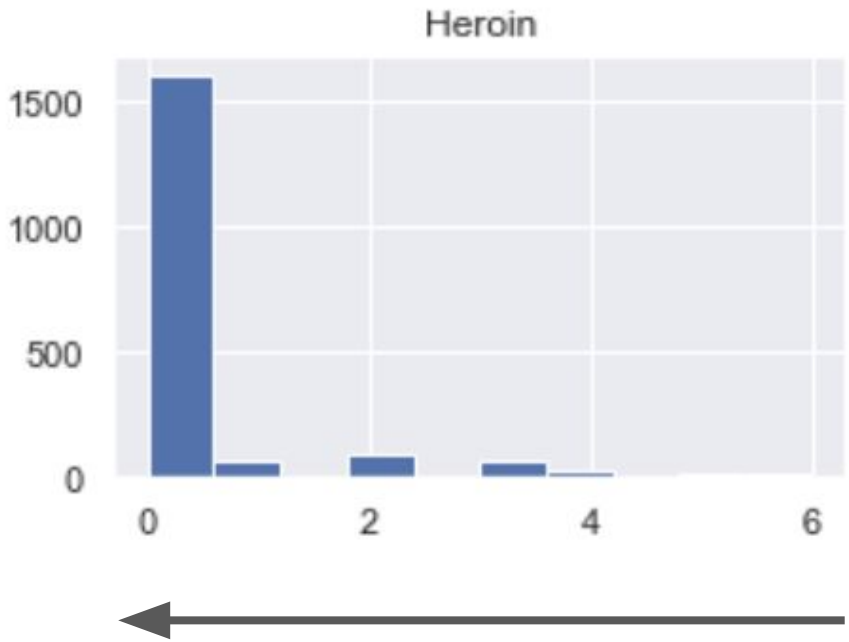
Questions?

Appendix

Distribution of drug use



High use



Low use