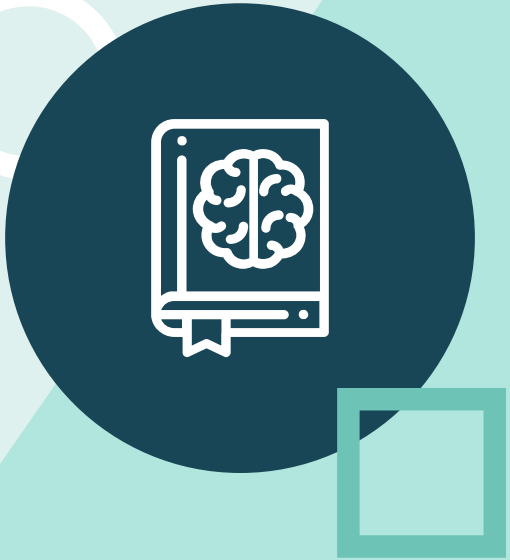
Predicting Drug Consumption Risk

A presentation by Nina Notman and Mirko Knoche

TABLE OF CONTENTS

INTRODUCTION TO 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

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

PREDICTION

XGBoost after GridSearch

- Identify 4 out of 5 correctly as cannabis consumer
- Not reliable enough (yet!)

FEATURE IMPORTANCE

For preventive reasons, it would be very helpful to **identify patients that are likely to take highly addictive drugs at some point in their life**.

We need to adjust therapy before it's too late!"

— FAMOUS THERAPIST

Let's do it!

PREDICTION OF HIGHLY ADDICTIVE DRUG CONSUMPTION RISK

CHANGE LABEL

HIGHLY ADDICTIVE PSYCHOTIC DRUGS

Ecstasy, Heroin, LSD, Meth, Cocaine

ALL IN ONE 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

PREDICTION

> For preventive reasons, it would be very helpful to **identify patients that are likely to take highly addictive drugs at some point in their life**.

SVM after GridSearch

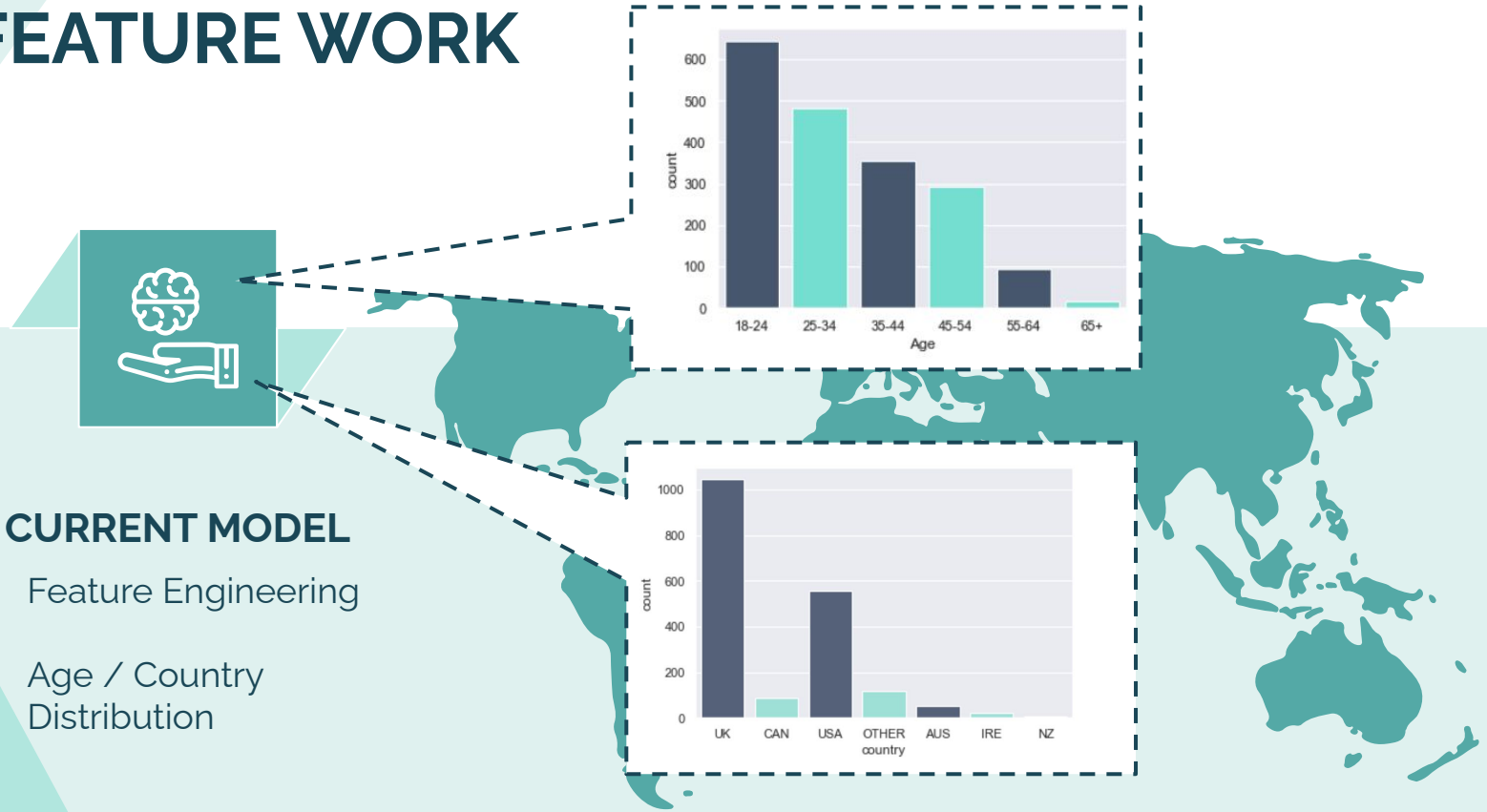
Only 75% accuracy on predicting consumption of highly addictive drugs.

Feature Engineering necessary!

WATCH OUT!

ONLY FOR WHITE, ENGLISH SPEAKING PEOPLE!

FEATURE WORK

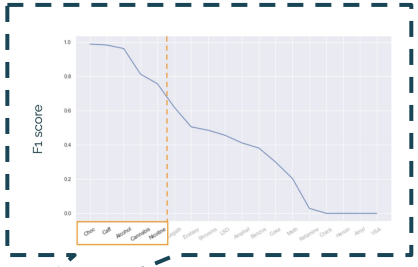


CURRENT MODEL

Feature Engineering

Age / Country Distribution

FEATURE WORK



CURRENT MODEL

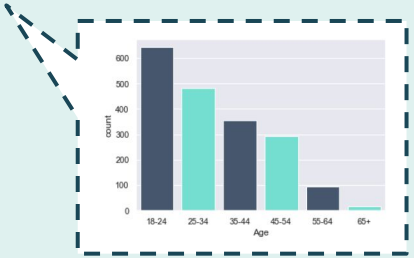
Feature Engineering
Age Distribution

NEW MODELS

Grouping of Drug Consumers

NEW DATA

Expand data collection to generalize model



LET'S TALK!

Nina Notman
nina.notman@gmx.net

Mirko Knoche
mirko.knoche@posteo.de

THANK YOU!

Questions?