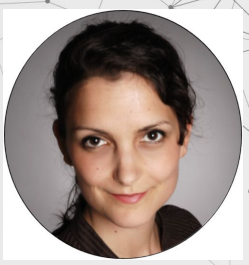
WELCOME!

Sebastian Siegler

Economist
Work experience in logistics sector

Dr. Karima Chakroun

Cognitive Neuroscientist
Psychologist
Biochemist



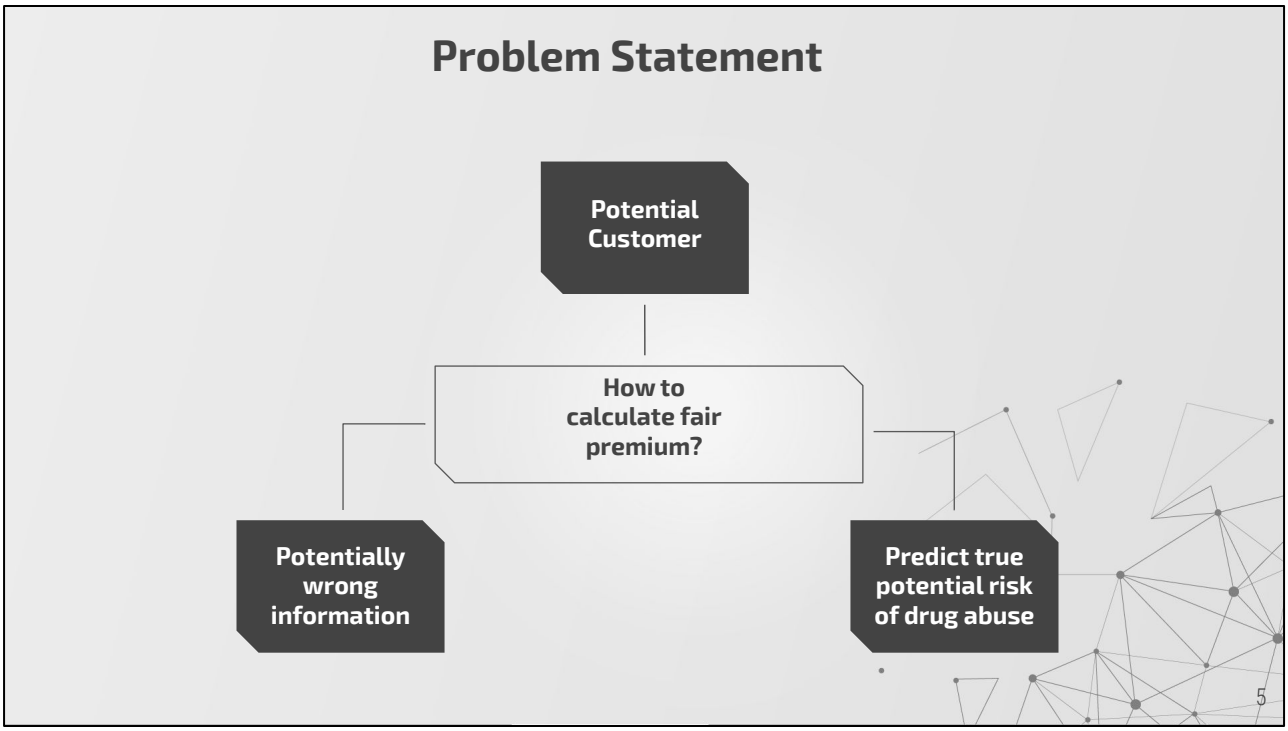
Today: Consulting Data Scientist at FutureMinds

Health Insurance Premium

Calculations based on psychological predictors for drug abuse risk

TABLE OF CONTENTS


01
Problem Statement



Health insurance companies have to calculate a premium for their products that make them competitive at the market while ensuring appropriate profits

Drug abuse causes immense costs for the healthcare sector

Potential customers are dishonest when it comes to admitting previous drug use

It is of great interest for us to predict a potential drug abuse risk for potential customers of private health insurances in order to calculate a fair premium


02
Business Value

Predict potential costs caused by customers with drug abuse problems (treatment, sick leaves …) in order to

- Reject customers who potentially consume heavy drugs like cocaine
- Charge customers who potentially consume moderately heavy drugs like speed or cannabis a risk premium
- Accept customers who do not consume any drugs with the basic insurance premium


03
Methodology

Methodology

Dataset

- Online survey of 2051 people
- 12 numeric features
  - Demographic
  - Personality Traits
- 18 Drugs (we selected three)
  - Categorized into different temporal consumption patterns

Exploratory Data Analysis
Application of Machine Learning

Data source:

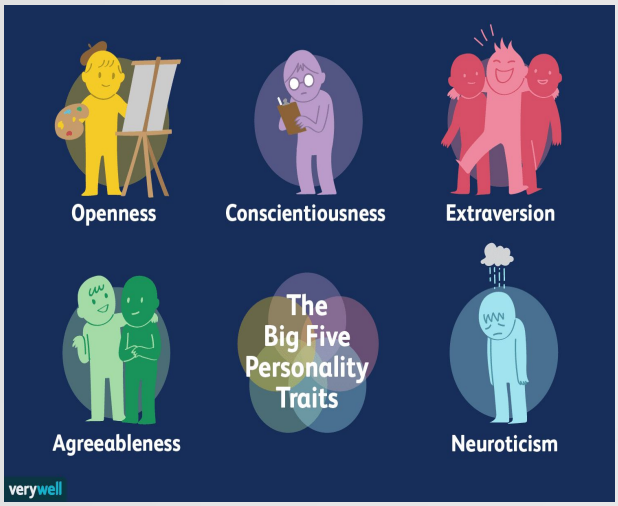
- we were able to use data collected by research facilities
  - 'The Five Factor Model of personality and evaluation of drug consumption risk' (Fehrman et al., 2017)
  - Cornell University, UK
  - link to research article: https://arxiv.org/abs/1506.06297
  - link to dataset and data dictionary: https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/36536
- anonymous online survey, ran over 12 months, snowball sampling method (biased, see Data Exploration)
- 2051 took part in the survey, 166 excluded due to incorrect answers or inattentiveness

Dataset:

- 12 features: all numeric based on numerical/categorical feature quantification as described in research article
  - demographic: age, gender, education, country, ethnicity
  - personality:
    - Impulsivity
    - Sensation Seeking
    - Openness
    - Conscientiousness
    - Extraversion
    - Agreeableness
    - Neuroticism
- 18 drugs: categorized into different temporal consumption patterns
  - CL0 Never Used

- CL1 Used over a Decade Ago
- CL2 Used in Last Decade
- CL3 Used in Last Year
- CL4 Used in Last Month
- CL5 Used in Last Week
- CL6 Used in Last Day

Methodology



- Easy to obtain
- Different personalities are connected to certain drug use patterns
- Drug abuse connected with high N and low A and C
- More patterns can be found

Possible questions
Openness:
- I have excellent ideas
- I am quick to understand things

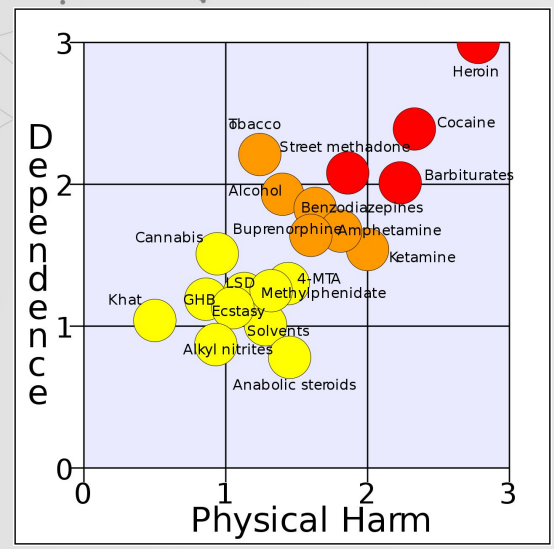
Conscientiousness
- I am always prepared
- I pay attention to details

Extraversion
- I don't mind being the center of attention.
- I feel comfortable around people.

Agreeableness
- I am interested in people.
- I sympathize with others' feelings.

Neuroticism
- I am relaxed most of the time. (*reversed*)
- I seldom feel blue. (*reversed*)[1]

Methodology



- Picked one drug form every physical harm category
- Used different evaluation metrics to find the best prediction model for each drug
- Min FN for class 2 and 3
- Min FP for class 1

11

For our business case, we have different requirements concerning the evaluation metrics of our models, depending on the physical damage potential of the drug

1) moderate (amphetamine) and very high (heroin)

- we want to correctly identify customers with a high probability of taking very harmful drugs due to the high expected treatment costs associated with that drug usage
  - in case of heroin, we want to completely reject the customer because the insurance fees will not cover potential treatment costs
  - in case of amphetamine, we want to charge the customer an additional risk premium in order to minimize the potential loss due to costs associated with drug abuse
- in both cases we therefore want to avoid classifying a person as non-user, who has actually already consumed the drug or has a high probability to do so in the future
- we want to **minimize false negatives (FN)**, hence we want to **maximize recall**

2) low (cannabis)

- in this case, the risk premium we charge will cover the potential treatment costs associated with the consumption of the drug, hence we are indifferent when it comes to accepting these customers
- but we want to maximize our client pool and profits, therefore we don't want to charge an additional premium to potential customers falling into this category because this might result in loosing the customer to the competitors who classifies the customer correctly and doesn't charge a risk premium
- we therefore want to avoid classifying a person as user, who has actually never consumed the drug or has no risk of doing so in the future
- we want to **minimize false positives (FP)**, hence we want to **maximize precision**

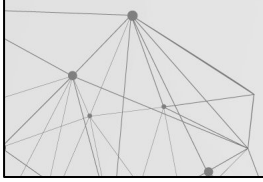
Our Model

Customer — Applies for private insurance

Ml Model — Calculates Drug abuse risk

Basic Premium — Multiplied by model parameter

Final Premium — Based on business needs



04

Conclusion

Conclusion

- Importance of drug abuse for healthcare sector
- Drug abuse risk prediction based on questionnaire data
- Presented different evaluation metrics to find the best prediction model for each drug
- Have shown how to deal with different drug abuse patterns
- Presentation of business model


05
Future Work

Future Work

- Dataset is imbalanced
- Only analyzed consumption patterns of three drugs
- Data does not give insights in frequency patterns

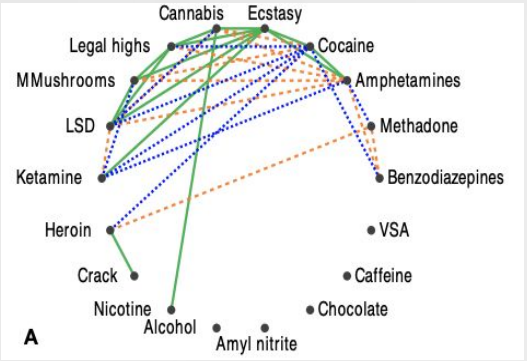


Image: https://arxiv.org/pdf/1506.06297.pdf

THANKS

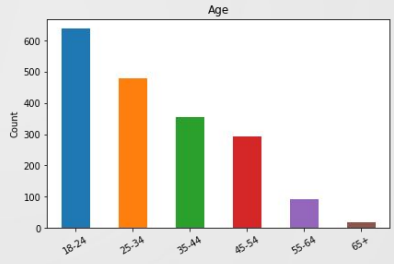
Does anyone have any questions?

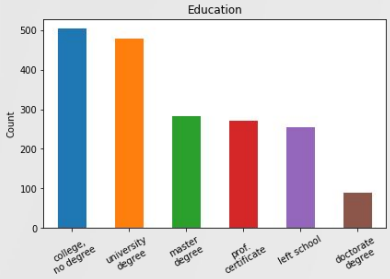
CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

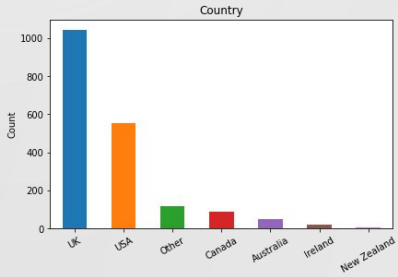
Please keep this slide for attribution.

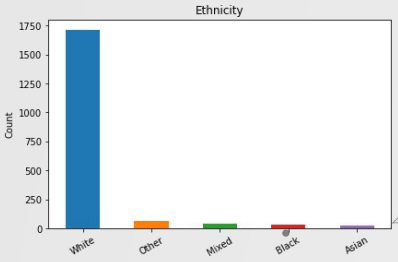

A
Appendix

Demographics: distribution

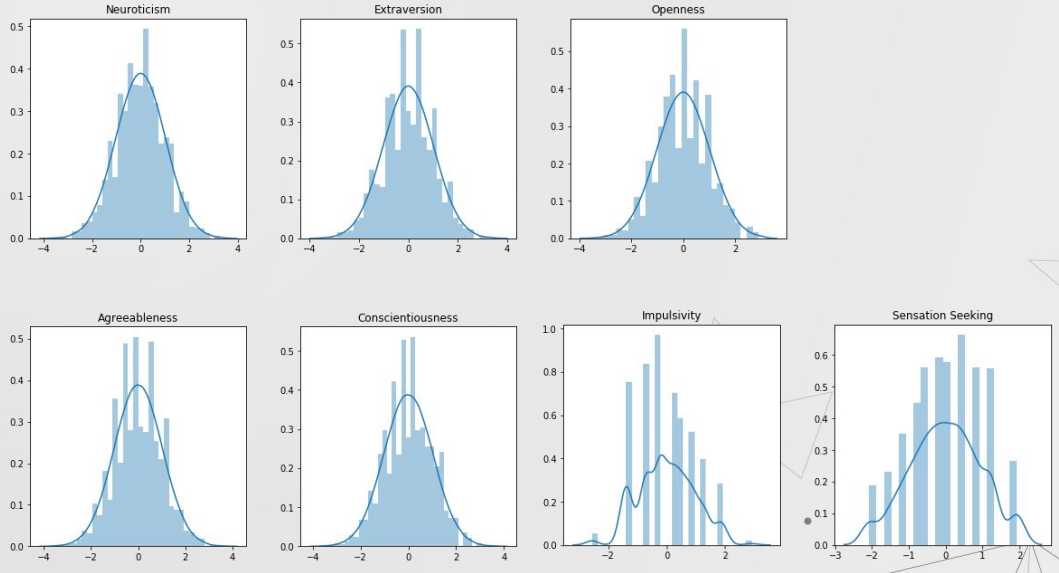




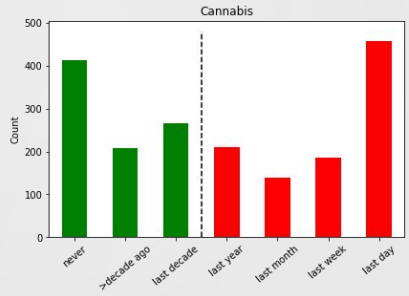


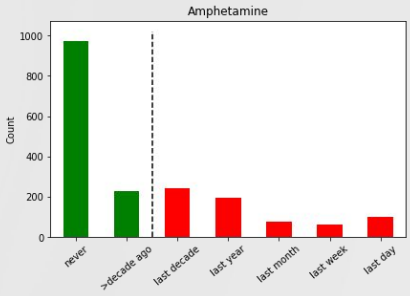


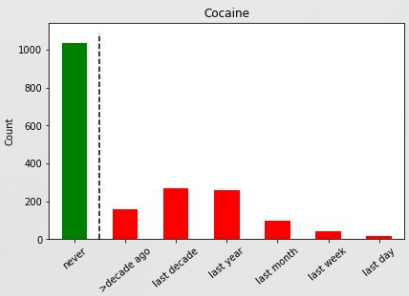
Personality traits: distribution



Drug consumption patterns







non-user
user