THE FIVE FACTOR MODEL OF PERSONALITY AND EVALUATION OF DRUG CONSUMPTION RISK

Elaine Fehrman / Rampton Hospital

Evgeny M. Mirkes / University of Leicester

Awaz K. Muhammad / University of Leicester

Vincent Egan / University of Nottingham

Alexander N. Gorban / University of Leicester

Conference of the International Federation of Classification Societies

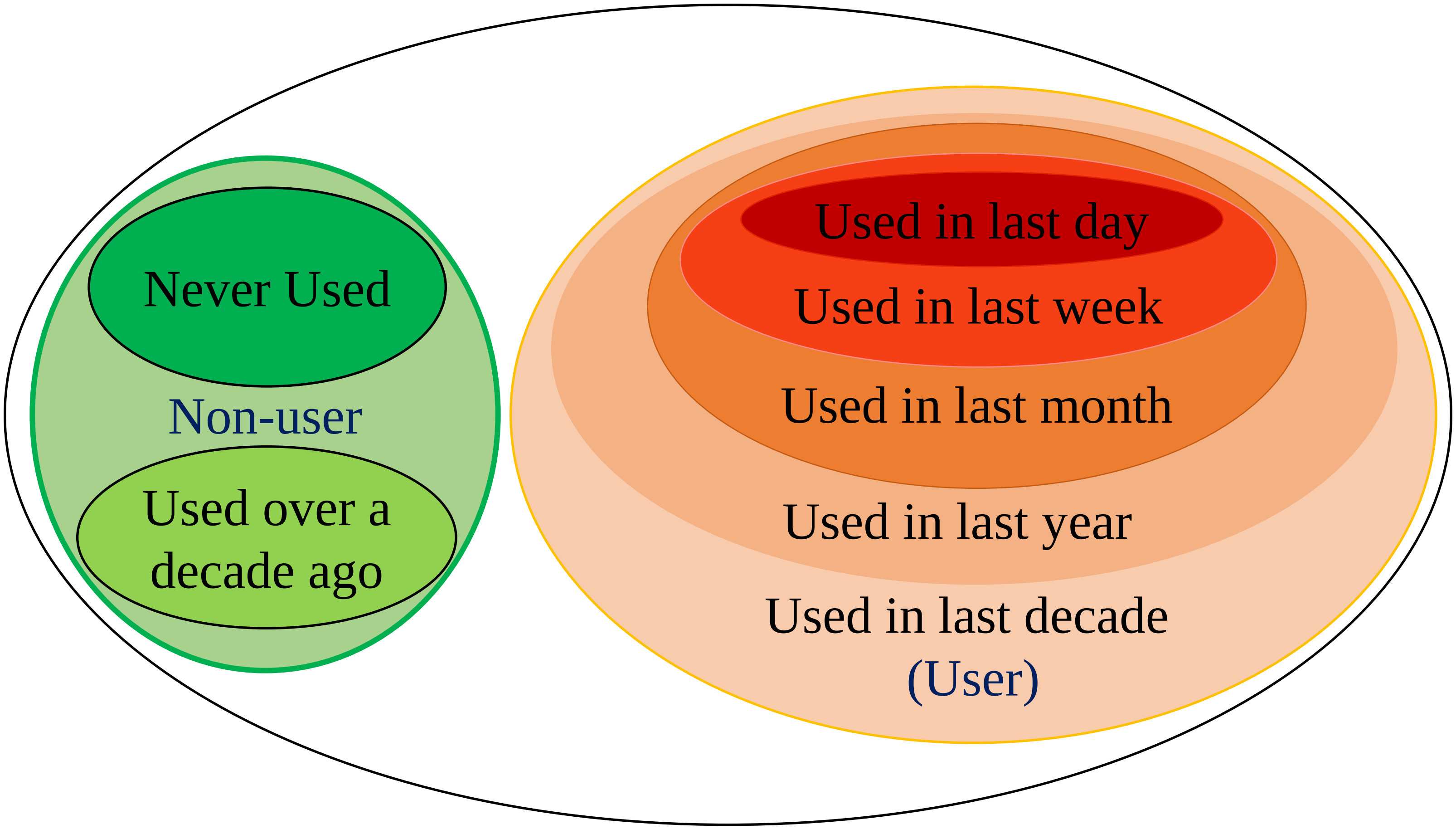
University of Bologna

8th July 2015

OUTLINE

- Introduction.
- The problem of risk evaluation for drug usage.
- Classification methods and results.
- Risk evaluation.

THE PROBLEM OF RISK EVALUATION FOR DRUG USAGE



Categories of drug users

INPUT FEATURE TYPES

- Personality traits:
  - Revised NEO-Five Factor Inventory (NEO-FFI-R) (McCrae & Costa, 2004): Neuroticism (**N**), Extraversion (**E**), Openness to Experience (**O**), Agreeableness (**A**), Conscientiousness (**C**).
  - Impulsivity (BIS-11) (Stanford et al., 2009) (**Imp**).
  - Sensation-seeking (ImpSS) (Zuckerman, 1994) (**SS**).
- Demographic data:
  - Age.
  - Gender.
  - Education level (**Edu**).

PSYCHOLOGICAL HYPOTHESIS

- We expect that drug usage is associated with high N, and low A and C.
- It is known that the 'dark dimension' of personality can be described in terms of low A (Jakobwitz & Egan, 2006).
- Much of the antisocial behaviour in normal persons appears underpinned by high N and low C.
- The 'negative urgency' is the tendency to act rashly when distressed; it is characterised by high N, low C, and low A (Settles et al., 2012).
- The 'negative urgency' is partially proved by us for users of the majority of illegal drugs.
- In addition, we demonstrate that O is higher for drug users.

INPUT FEATURES AND DRUGS

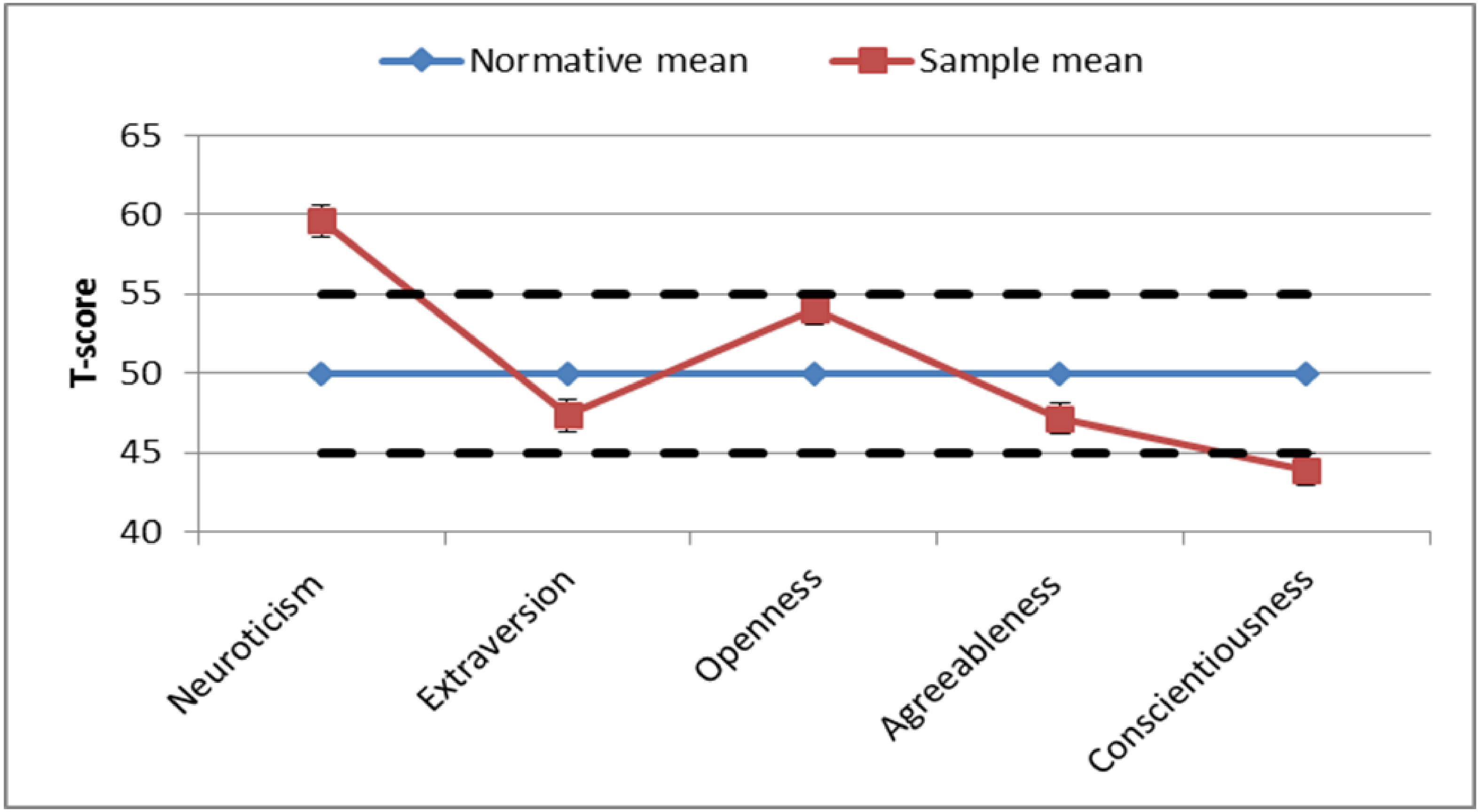
Input
Neuroticism
Extraversion
Openness
Agreeableness
Conscientiousness
Impulsiveness
Sensation-seeking
Age
Gender
Education

Drugs for risk evaluation	
Alcohol	Ecstasy
Amphetamines	Heroin
Amyl nitrite	Ketamine
Benzodiazepines	Legal highs
Cannabis	LSD
Chocolate	Methadone
Cocaine	Magic mushrooms
Caffeine	Nicotine
Crack	VSA

THE SAMPLE

- *N* = 2,051; 1,885 useable cases.
- Gender: Male *(n* = 943*)*, female (*n* = 942).
- Age: 18 – 24 years (*n* = 643; 34.1%), 25 – 34 years (*n* = 481; 25.5%), 35 – 44 years (*n* = 356; 18.9%), 45 – 54 years (*n* = 294; 15.6%), 55 – 64 (*n* = 93; 4.9%), and over 65 years (*n* = 18; 1%).
- Education: Professional certificate or diploma (*n* = 271; 14.4%), undergraduate degree (*n* = 481; 25.5%), master's (*n* = 284; 15%), doctorate (*n* = 89; 4.7%), some college / university (*n* = 506; 26.8%), left school ≤ 18 years (*n* = 257; 13.6%).
- Country of origin: UK (*n* = 1,044; 55.4%), USA (*n* = 557; 29.5%), Canada (*n* = 87; 4.6%), Australia (*n* = 54; 2.9%), New Zealand (*n* = 5; 0.3%), Ireland (*n* = 20; 1.1%), and 'Other' (*n* = 118; 6.3%).
- Ethnicity: White (*n* = 1,720; 91.2%), Black (*n* = 33; 1.8%), Asian (*n* = 26; 1.4%), and 'Other / Mixed'(*n* = 106; 5.6%).

THE SAMPLE VS. POPULATION NORM



COMPARISON OF AVERAGE PERSONALITY TRAITS FOR DRUG USERS AND NON-USERS

The relationship between personality and risk of drug consumption:

- High risk of drug use is correlated with High N and O.
- High risk of drug use is correlated with Low A and C.
- The influence of E is drug specific.

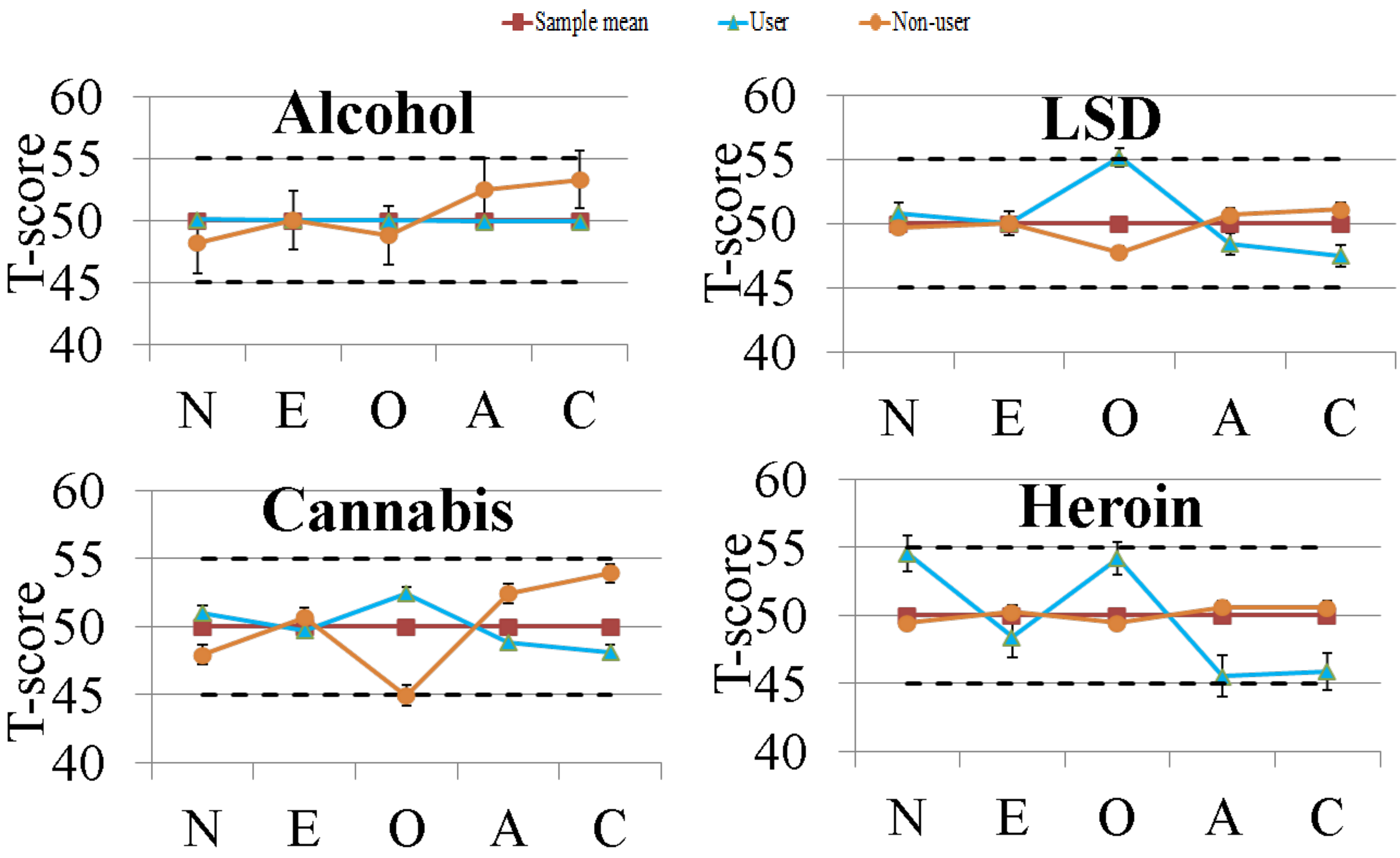
For each drug, drug users scored higher on Neuroticism and Openness, and lower on Agreeableness and Conscientiousness when compared to non-users.

Moderate subcategories of T-score$_{sample}$ with respect to the sample mean for group of users. The white background corresponds to neutral score (0), the green background corresponds to high score (+), and the pink background corresponds to low score (−).

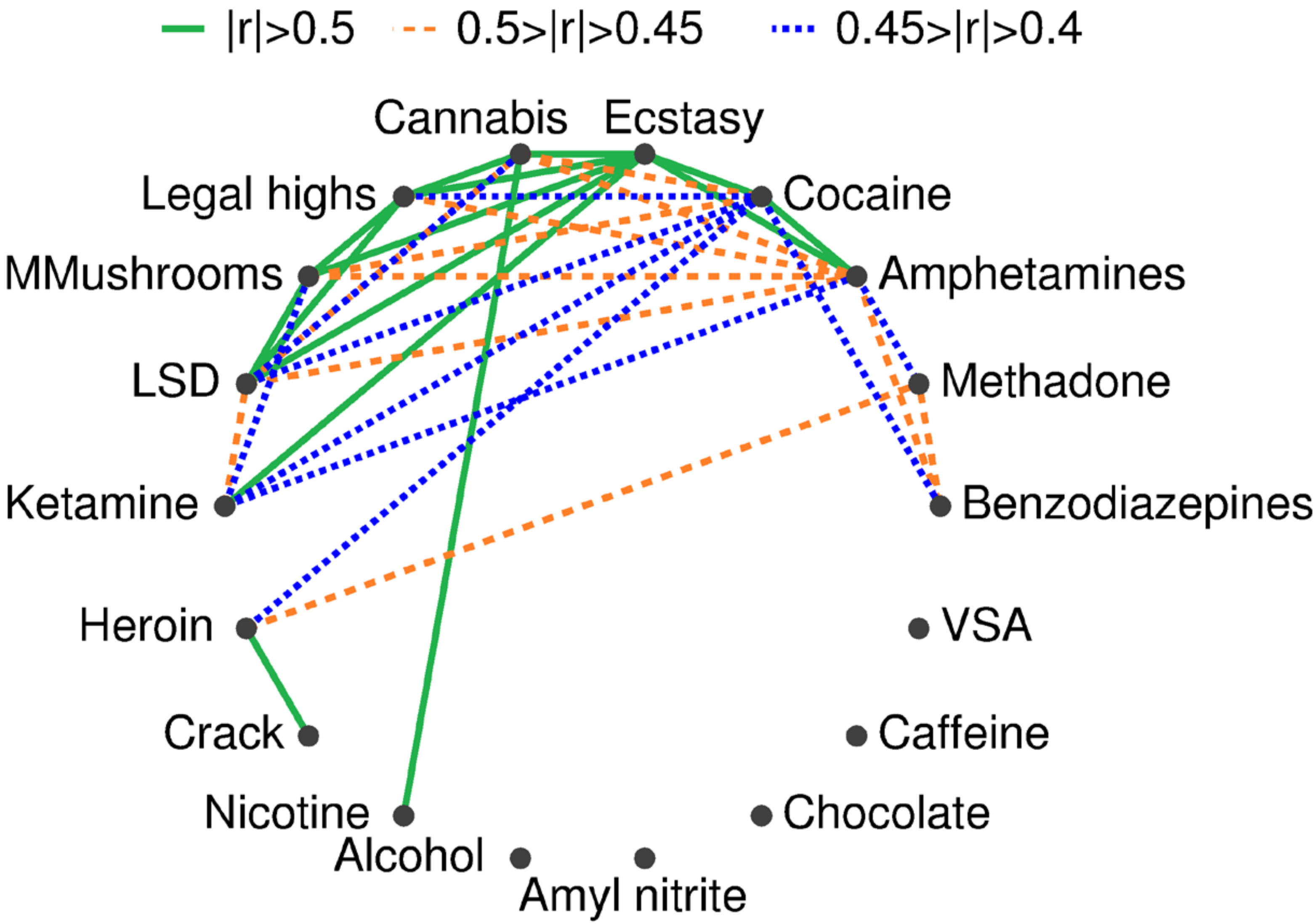
N	E	O	A	C
Alcohol, Chocolate, and Caffeine				
0	0	0	0	0
Nicotine				
0	0	+	0	−
Amyl nitrite, LSD, and Magic mushrooms				
0	0	+	−	−
Amphetamines, Benzodiazepines, Cannabis, Cocaine, Ecstasy, Ketamine, and Legal highs				
+	0	+	−	−
Crack, Heroin, VSA, and Methadone				
+	−	+	−	−

N=Neuroticism, E= Extraversion, O= Openness to experience, A=Agreeableness, C=Conscientiousness

AVERAGE PERSONALITY PROFILES FOR DRUG USERS AND NON-USERS



CORRELATIONS BETWEEN USAGE OF DIFFERENT DRUGS



INFORMATION GAIN

RIG of the drug X usage with respect to the drug Y usage is defined as:

$$\text{RIG}(X|Y) = \frac{Entropy(X) - Entropy(X|Y)}{Entropy(X)},$$

where $Entropy(X)$ is the entropy of drug X usage:

$$Entropy(X) = -\mu \ln \mu - (1 - \mu) \ln(1 - \mu),$$

where μ is the fraction of drug X users among all participants, $Entropy(X|Y)$ is the relative entropy:

$$\begin{aligned} &Entropy(X|Y) \\ &= \nu\, Entropy(X|y = \text{User}) + (1 - \nu)\, Entropy(X|y = \text{Non} - \text{user}), \end{aligned}$$

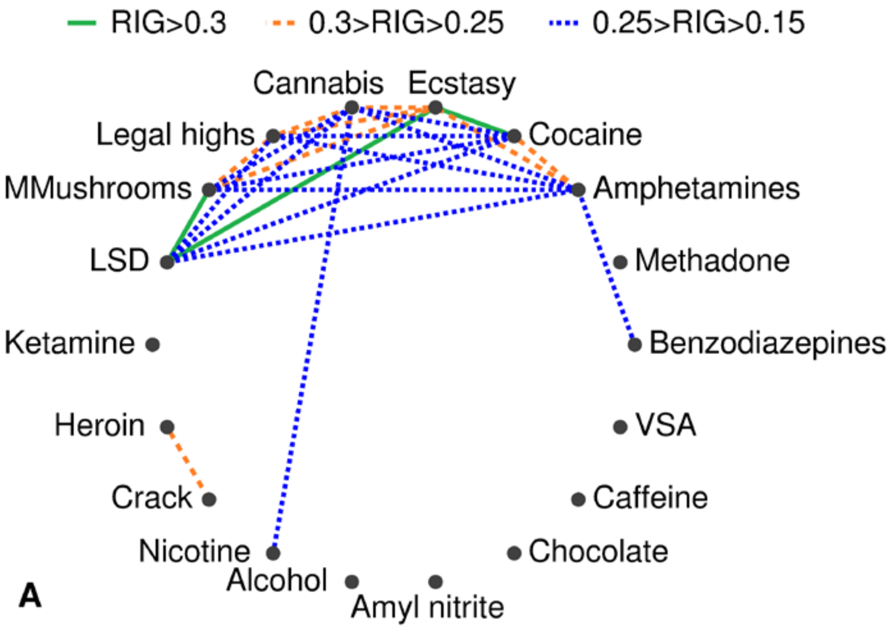
where ν is the fraction of drug Y users among all participants, $Entropy(X|y = \text{User})$ and $Entropy(X|y = \text{Non–user})$ are the specific conditional entropies:

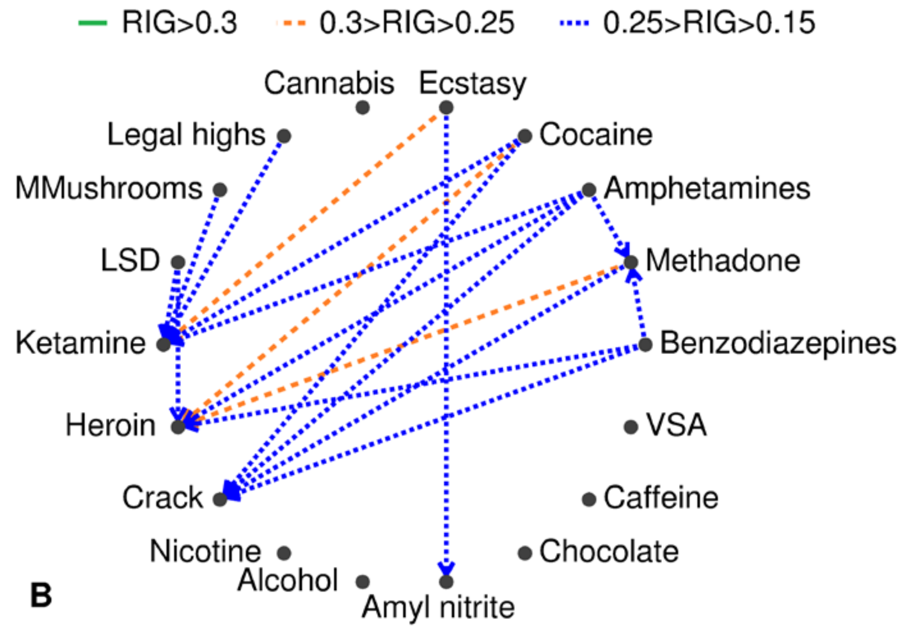
$$\begin{aligned} &Entropy(X|y = \text{User}) \\ &= -\mu_{y=\text{User}} \ln \mu_{y=\text{User}} - \left(1 - \mu_{y=\text{User}}\right) \ln\left(1 - \mu_{y=\text{User}}\right), \end{aligned}$$

$$\begin{aligned} &Entropy(X|y = \text{Non–user}) \\ &= -\mu_{y=\text{Non–user}} \ln \mu_{y=\text{Non–user}} - \left(1 - \mu_{y=\text{Non–user}}\right) \ln\left(1 - \mu_{y=\text{Non–user}}\right), \end{aligned}$$

where $\mu_{y=\text{User}}$ is the fraction drug X user among all drug Y users and $\mu_{y=\text{Non–user}}$ is the fraction of drug X users among all drug Y non-users.

PAIRS OF DRUG USAGES WITH HIGH RELATIVE INFORMATION GAIN





More or less symmetric RIG

Essentially asymmetric RIG

CLASSIFICATION METHODS

- Decision Tree (**DT**).
- K-Nearest Neighbours (**KNN).**
- Random Forest (**RF**).
- Linear Discriminant Analysis (**LDA**).
- Gaussian Mixture (**GM**).
- Probability Density Function Estimation (**PDFE**).
- Logistic Regression (**LR**).
- Naïve Bayes (**NB**).

CLASSIFICATION METHODS

- Decision Tree (**DT**)**:** 166M models per drug
  - Split criterion : information gain, Gini gain or DKM gain.
  - Linearly combined or separately used input features.
  - The set of the input features.
  - Minimal number of cases in the leaf is varied between 3 and 30.
  - Weight of class “User” is varied between 0.01 and 5.0.
- K-Nearest Neighbours (**KNN**)**:** 1,683M models per drug
  - k is varied between 1 and 20.
  - The set of input features.
  - Distance: Euclidean, adaptive, and Fisher’s.
  - The kernel function for adaptive distance transformation.
  - The kernel functions for voting.
  - Weight of class “User” is varied between 0.01 and 5.0.

CLASSIFICATION METHODS

- Random Forest (**RF):** 2,048 models per drug
  - The set of the input features.
- Linear Discriminant Analysis (**LDA):** 8,192 model per drug
  - The set of the input features.
  - RIG, Gini gain, DKM gain, or accuracy as criterion for threshold defining.
- Gaussian Mixture (**GM**)**:** 1.024M models per drug
  - The set of the input features.
  - Weight of class “User” is varied between 0.01 and 5.0.

CLASSIFICATION METHODS

- Probability Density Function Estimation (**PDFE): ** 426K models per drug
    - The number of the NN is varied between 5 and 30.
    - The set of the input features.
    - The kernel function which was placed in each data points.
- Logistic Regression (**LR**)**:** 2,048 models per drug
    - The set of the input features.
- Naïve Bayes (**NB**)**:** 2,048 models per drug
    - The set of the input features.

THE BEST CLASSIFIER SELECTION

- Sens+Spec is the distance from ‘completely random guess’ classifier.
- *Balanced classifier* is the classifier with Sens=Spec.
- *Measure of classifiers balance* is min{Sens,Spec}.
- *The best classifier* (in this case study) is the balanced classifier with Sens+Spec $\to$max.

THE BEST RESULTS OF THE ***LEGAL DRUG*** USERS CLASSIFIERS

Target feature	Meth	Age	Gen	Edu	N	E	O	A	C	Imp	SS	Sens. %	Spec. %
Alcohol	LDA	X	X	X	X						X	75.34	63.24
Chocolate	KNN	X	X			X			X			72.43	71.43
Caffeine	KNN	X		X			X	X		X		70.51	72.97
Nicotine	DT		X		X	X			X			71.28	79.07

'X' means used input feature. LOOCV test results.

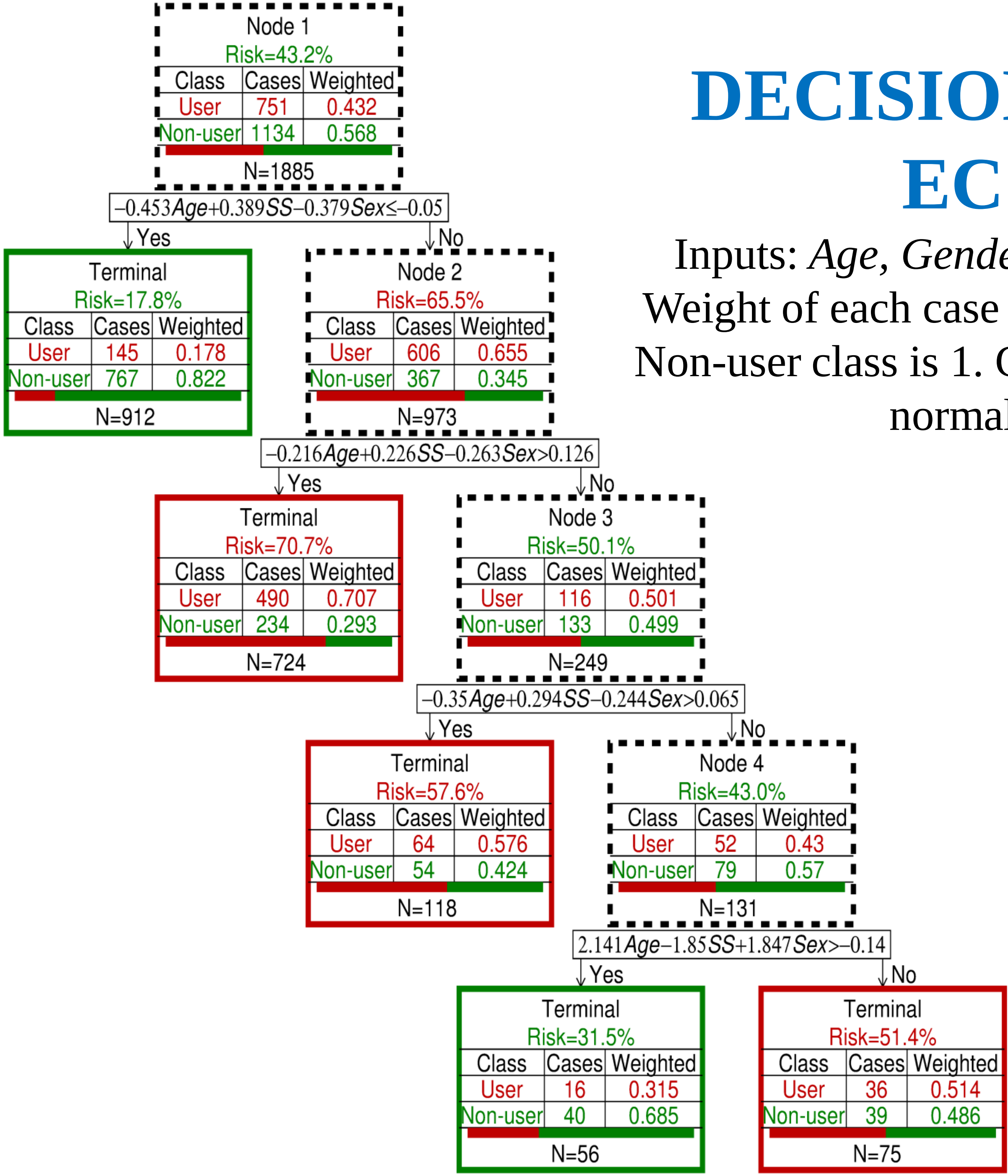
LDA = Linear Discriminant Analysis
KNN = K-Nearest Neighbours
DT = Decision Tree

THE BEST RESULTS OF THE *ILLEGAL DRUG* USERS CLASSIFIERS

Target feature	Age	Gen	Edu	N	E	O	A	C	Imp	SS	Sens. %	Spec. %
Amphetamines	X			X		X		X	X	X	81.30	71.48
Amyl nitrite				X		X		X		X	73.51	87.86
Benzodiazepines	X	X		X	X				X	X	70.87	71.51
Cannabis	X		X			X	X	X	X		79.29	80.00
Cocaine	X					X	X		X	X	68.27	83.06
Crack					X			X			80.63	78.57
Ecstasy	X	X								X	76.17	77.16
Heroin	X	X							X		82.55	72.98
Ketamine	X				X		X		X	X	72.29	80.98
Legal highs	X	X				X	X	X		X	79.53	82.37
LSD	X	X		X	X	X			X		85.46	77.56
Methadone	X	X	X		X	X					79.14	72.48
MMushrooms		X			X						65.56	94.79
VSA	X		X		X		X	X		X	83.48	77.64

DECISION TREE FOR ECSTASY

Inputs: *Age, Gender, and Sensation-seeking*. Weight of each case of User class is 1.15 and of Non-user class is 1. Columns ‘Weighted’ present normalised weights.



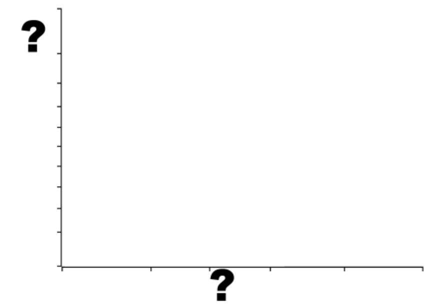
RISK MAP CREATION

DB

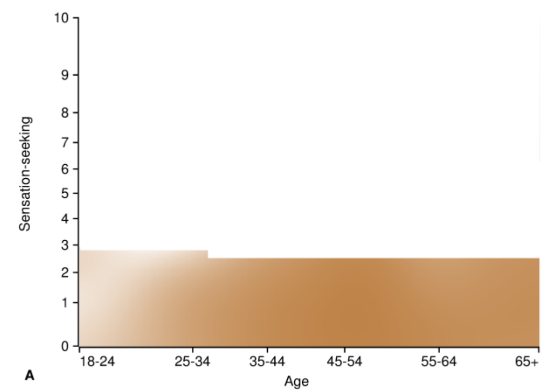
Select the best classifier

Pool of classifiers

Select coordinate system for map

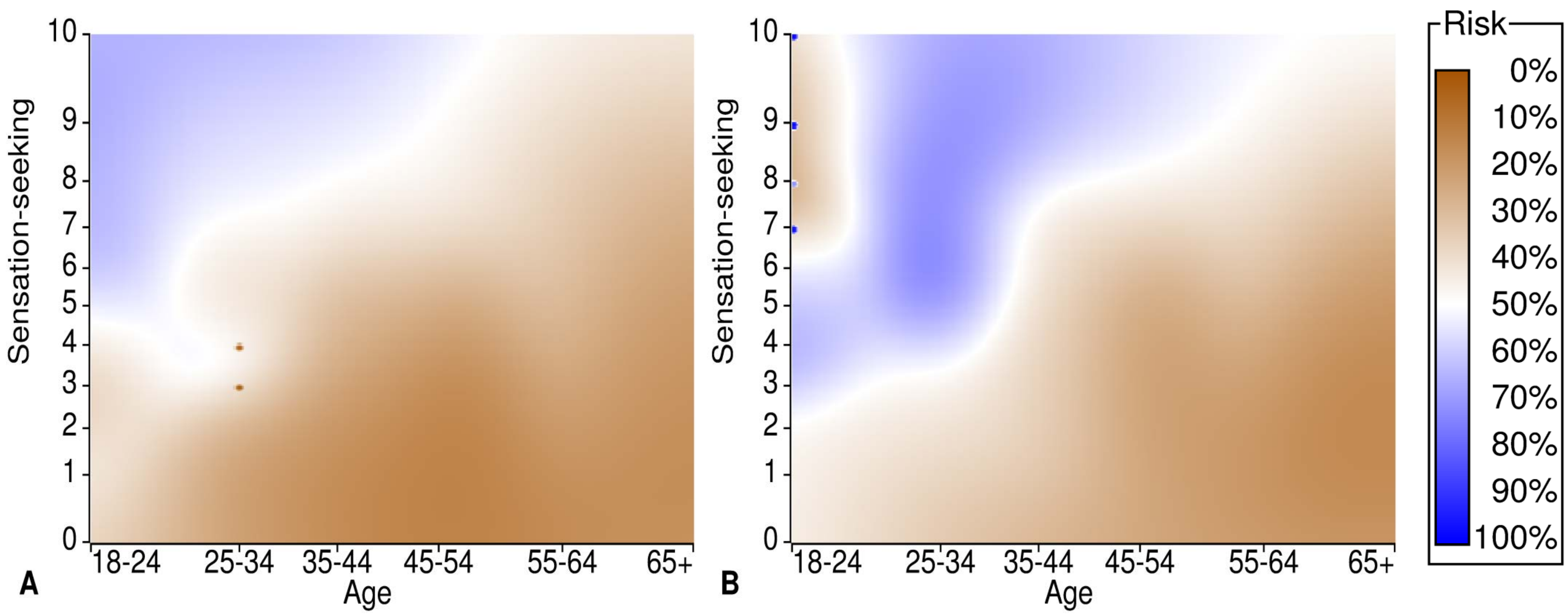


Fix values of other input features and calculate risk by scanning of area



THE RISK MAP FOR ECSTASY

Inputs: *Age, Gender, and Sensation-seeking*
(**PDFE – kernel radial basis functions**)



A female

B male

THANK YOU FOR YOUR ATTENTION!

- Questions?

- *Detailed e-print:*
  Fehrman, E., Muhammad, A.K., Mirkes, E.M., Egan, V., & Gorban, A.N. (2015). The Five Factor Model of personality and evaluation of drug consumption risk, arXiv:1506.06297 [stat.AP]. http://arxiv.org/abs/1506.06297

REFERENCES

Psychology

- Jakobwitz, S., & Egan, V. (2006). The dark triad and normal personality traits. *Personality and Individual Differences*, *40*, 331-339.
- McCrae, R. R., & Costa, P. T. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences*, *36*, 587-596.
- Settles, R.E., Fischer, S., Cyders, M.A., Combs, J.L., Gunn, R.L., & Smith, G.T. (2012). Negative urgency: A personality predictor of externalizing behavior characterized by neuroticism, low conscientiousness, and disagreeableness. *Journal of Abnormal Psychology*, *121*, 160-172.
- Stanford, M.S., Mathias, C.W., Dougherty, D.M., Lake, S.L., Anderson, N.E., & Patton, J.H. (2009). Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personality and Individual Differences*, *47*, 385-395.
- Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking*. New York: Cambridge University Press.

REFERENCES

Data Mining

- Dinov, I.D. (2008). Expectation maximization and mixture modelling tutorial. ExStatistics Online Computational Resource.
- Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*, 179-188.
- Hastie, T., & Tibshirani, R. (1996). Discriminant adaptive nearest neighbor classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18*, 607–616.
- Rokach, L., & Maimon, O. (2010). Decision trees. In O. Maimon, & L. Rokach (Eds.), *Data mining and knowledge discovery handbook* (pp. 165-192). Berlin: Springer.
- Scott, D.W. (1992). *Multivariate density estimation: Theory, practice, and visualization* (1st ed.). New York: Wiley.