Columbia
Business
School

Universiteit
Antwerpen

SaverLife®

Explainable AI for psychological profiling from behavioral data

Yanou Ramon, Sandra Matz,
Robert Farrokhnia, David Martens

Joint seminar adm+adrem, Dec 15, 12:30 pm

https://doi.org/10.3390/info12120518

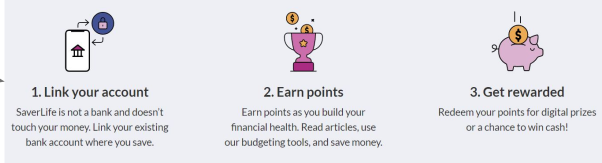
Context

Psychological profiling = “the automated assessment of psychological traits from digital footprints” (Matz, 2020)

Case study: personality profiling from consumer spending

Case study: personality profiling from consumer spending

SaverLife



1. Link your account

SaverLife is not a bank and doesn't touch your money. Link your existing bank account where you save.

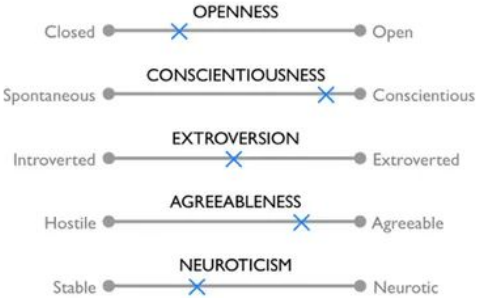
2. Earn points

Earn points as you build your financial health. Read articles, use our budgeting tools, and save money.

3. Get rewarded

Redeem your points for digital prizes or a chance to win cash!

Big Five Personality Traits



OPENNESS

Closed — Open

CONSCIENTIOUSNESS

Spontaneous — Conscientious

EXTROVERSION

Introverted — Extroverted

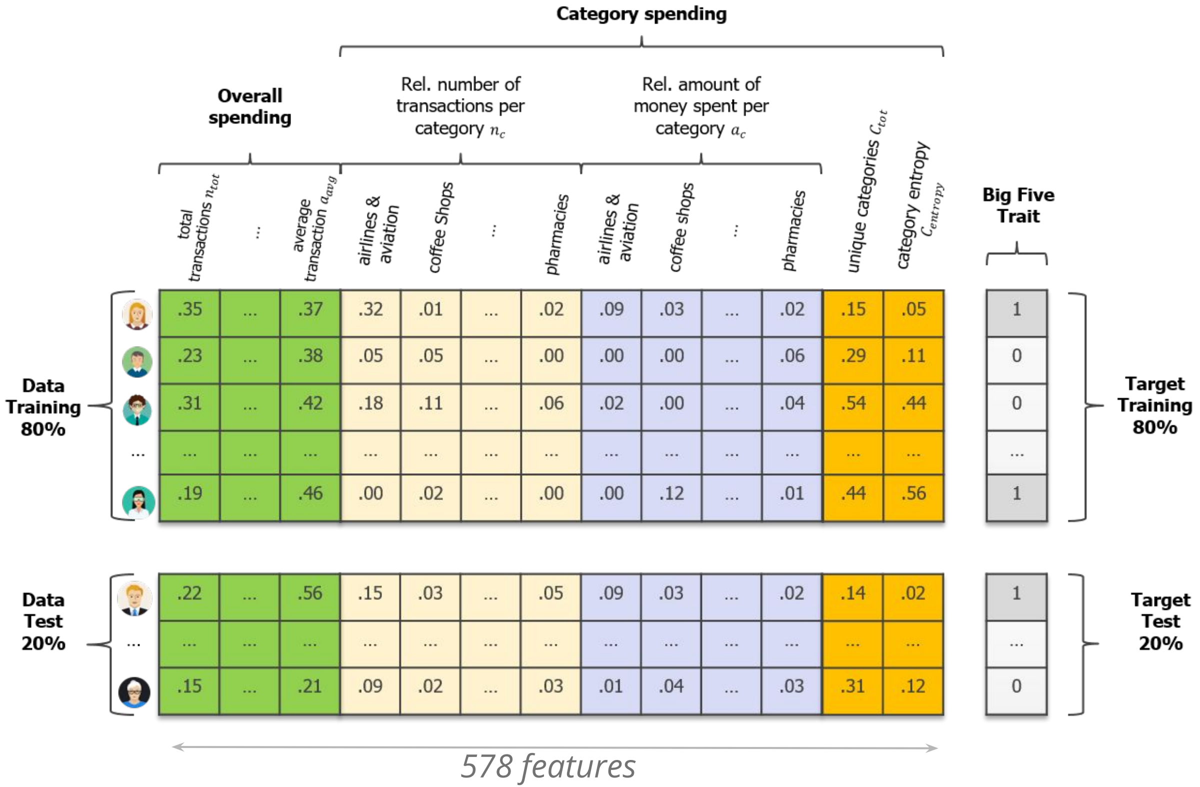
AGREEABLENESS

Hostile — Agreeable

NEUROTICISM

Stable — Neurotic

Case study: sample data (N=6,408)



				Category spending											
	Overall spending			Rel. number of transactions per category n_c				Rel. amount of money spent per category a_c						Big Five Trait	
	total transactions n_{tot}	…	average transaction a_{avg}	airlines & aviation	coffee Shops	…	pharmacies	airlines & aviation	coffee shops	…	pharmacies	unique categories C_{tot}	category entropy $C_{entropy}$		
Data Training 80%	.35	…	.37	.32	.01	…	.02	.09	.03	…	.02	.15	.05	1	Target Training 80%
	.23	…	.38	.05	.05	…	.00	.00	.00	…	.06	.29	.11	0	
	.31	…	.42	.18	.11	…	.06	.02	.00	…	.04	.54	.44	0	
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	
	.19	…	.46	.00	.02	…	.00	.00	.12	…	.01	.44	.56	1	
Data Test 20%	.22	…	.56	.15	.03	…	.05	.09	.03	…	.02	.14	.02	1	Target Test 20%
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	
	.15	…	.21	.09	.02	…	.03	.01	.04	…	.03	.31	.12	0	

578 features

Predictability of personality

- Random Forest models work best
- Decent performance: min=53.4%, max=61.8%
- Best accuracy for Neuroticism
- Conscientiousness & Neuroticism easier to predict than Agreeableness & Openness

Complication

Black box models → Why?

(1) High dimensionality

(2) Sparsity

(3) Non-redundancy

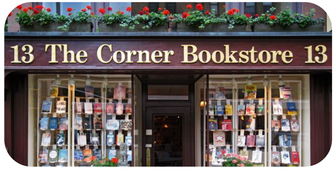
(e.g., De Cnudde et al., 2019; Clark & Provost, 2019;
Junqué de Fortuny et al., 2013)

Complication

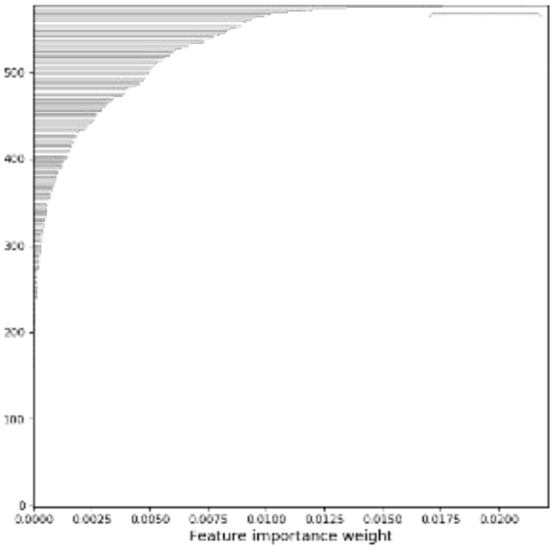
Black box models → Why?

(1) High dimensionality
(2) Sparsity
(3) Non-redundancy

(e.g., De Cnudde et al., 2019; Clark & Provost, 2019; Junqué de Fortuny et al., 2013)



Y = Extraversion



Main claim

Global and local rule-based Explainable AI (XAI) methods are important to gain insight into models for psychological profiling & particularly suitable for digital footprints data

Global XAI: rule-extraction

Extract if-then-else rules using data and predictions of the model:

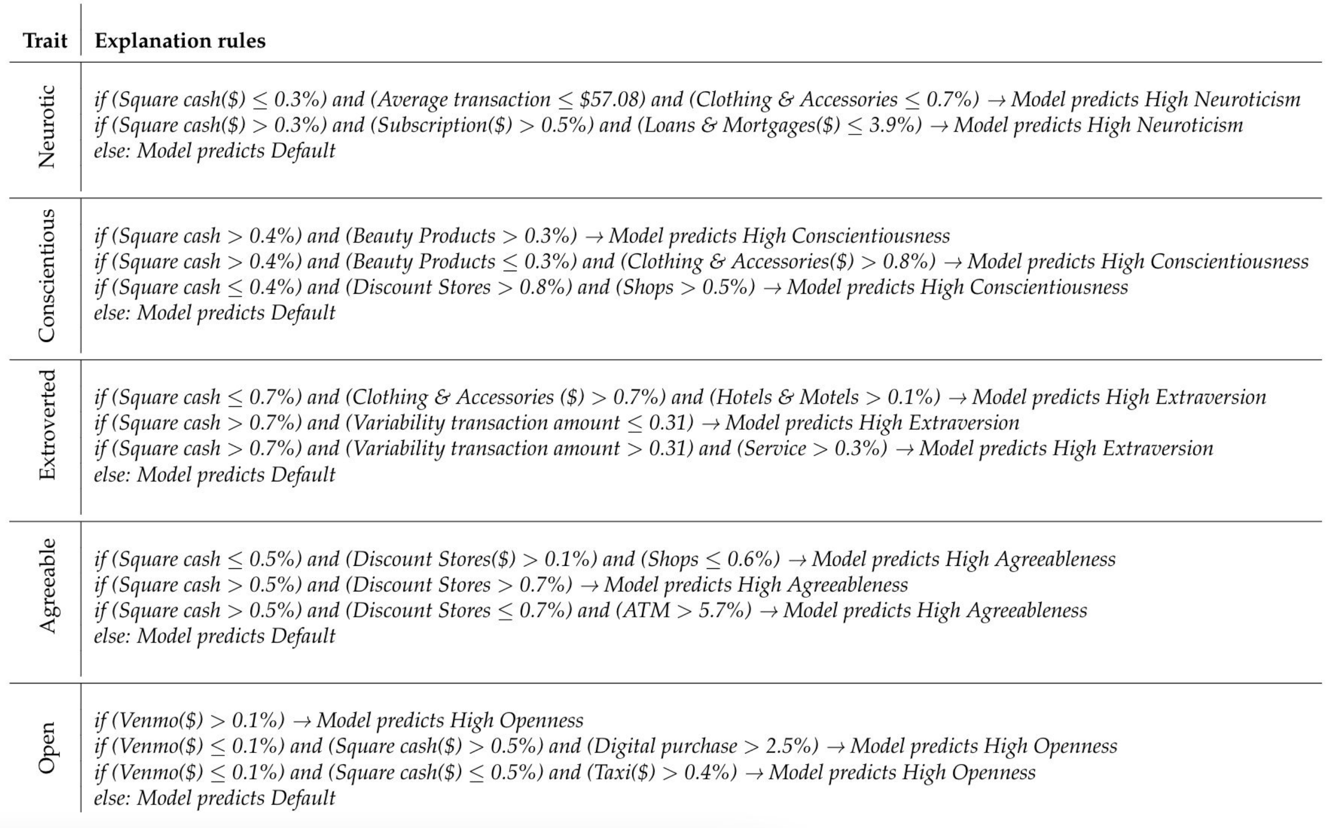
if **<condition1>** and **<condition2>** and … ⇒ **class 1**

elif **<condition3>** ⇒ **class 1**

else **class 2**

Fidelity (%) → measures overlap between predictions of the model and predictions of the explanation rules

Global XAI: results (max. features per rule = 3)



Trait	Explanation rules

Neurotic	*if (Square cash($) ≤ 0.3%) and (Average transaction ≤ $57.08) and (Clothing & Accessories ≤ 0.7%) → Model predicts High Neuroticism*
	if (Square cash($) > 0.3%) and (Subscription($) > 0.5%) and (Loans & Mortgages($) ≤ 3.9%) → Model predicts High Neuroticism
	else: Model predicts Default

Conscientious	*if (Square cash > 0.4%) and (Beauty Products > 0.3%) → Model predicts High Conscientiousness*
	if (Square cash > 0.4%) and (Beauty Products ≤ 0.3%) and (Clothing & Accessories($) > 0.8%) → Model predicts High Conscientiousness
	if (Square cash ≤ 0.4%) and (Discount Stores > 0.8%) and (Shops > 0.5%) → Model predicts High Conscientiousness
	else: Model predicts Default

Extroverted	*if (Square cash ≤ 0.7%) and (Clothing & Accessories ($) > 0.7%) and (Hotels & Motels > 0.1%) → Model predicts High Extraversion*
	if (Square cash > 0.7%) and (Variability transaction amount ≤ 0.31) → Model predicts High Extraversion
	if (Square cash > 0.7%) and (Variability transaction amount > 0.31) and (Service > 0.3%) → Model predicts High Extraversion
	else: Model predicts Default

Agreeable	*if (Square cash ≤ 0.5%) and (Discount Stores($) > 0.1%) and (Shops ≤ 0.6%) → Model predicts High Agreeableness*
	if (Square cash > 0.5%) and (Discount Stores > 0.7%) → Model predicts High Agreeableness
	if (Square cash > 0.5%) and (Discount Stores ≤ 0.7%) and (ATM > 5.7%) → Model predicts High Agreeableness
	else: Model predicts Default

Open	*if (Venmo($) > 0.1%) → Model predicts High Openness*
	if (Venmo($) ≤ 0.1%) and (Square cash($) > 0.5%) and (Digital purchase > 2.5%) → Model predicts High Openness
	if (Venmo($) ≤ 0.1%) and (Square cash($) ≤ 0.5%) and (Taxi($) > 0.4%) → Model predicts High Openness
	else: Model predicts Default

Global XAI: results for Conscientiousness

if (Square Cash > 0.4%) and (Beauty Products > 0.3%) ⇒ Model predicts **High C**

elif (Square Cash > 0.4%) and (Beauty Products <= 0.3%) and (Clothing & Accessories($) > 0.8%)
⇒ Model predicts **High C**

elif (Square Cash <= 0.4%) and (Discount Stores > 0.8%) and (Shops > 0.5%) ⇒ Model predicts **High C**

else: Model predicts **Default**

Fidelity: 75.8%

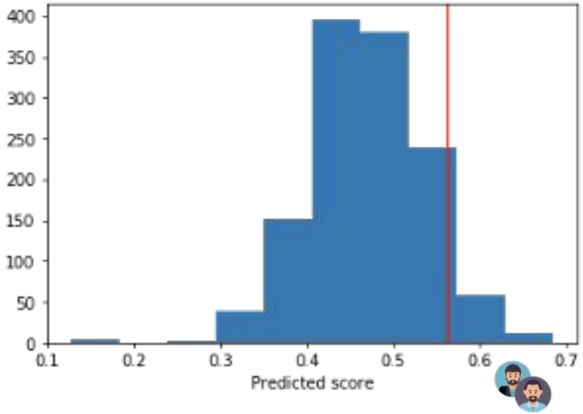
Local XAI: counterfactual explanations

Extract if-then-else rules using instance **x** and scoring function:

if **<condition1>** and **<condition2>** and ...

⇒ class *changes* from **class 1** to **class 2**

Local XAI: results for Neuroticism



Person A:

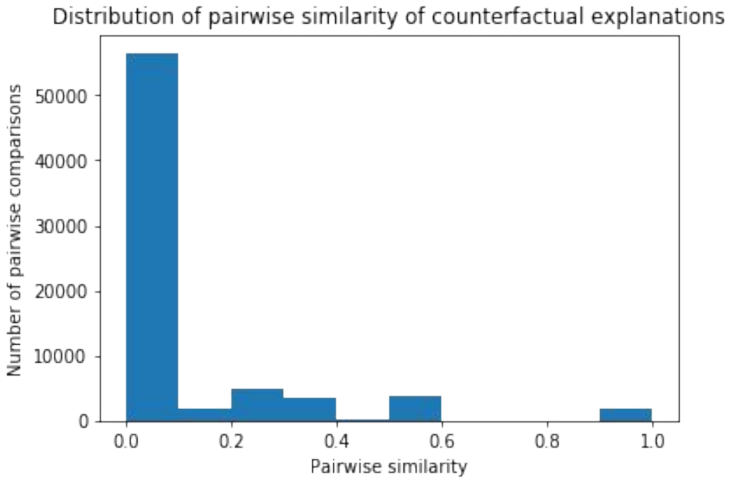
IF person A had spent *less frequently* in Computers & Electronics, Insurance and Shops, and *more frequently* in Clothing & Accessories and Restaurants ⇒ THEN he would not have been predicted as Neurotic

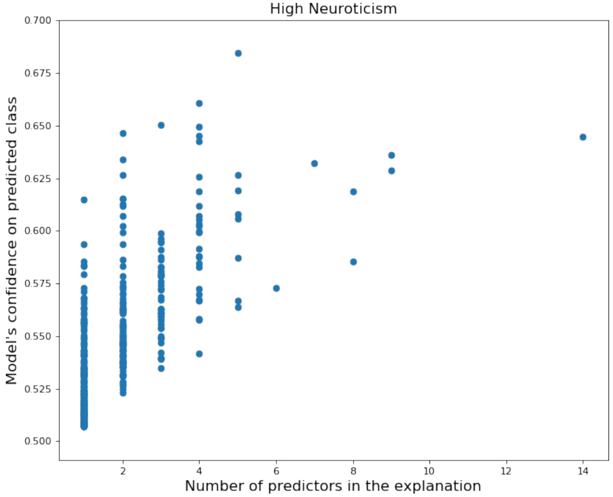
Person B:

IF person B had spent *less frequently* in Shops and Tobacco, and *less money* on Subscription and Tobacco ⇒ THEN he would not have been predicted as Neurotic

Local XAI: results for Neuroticism

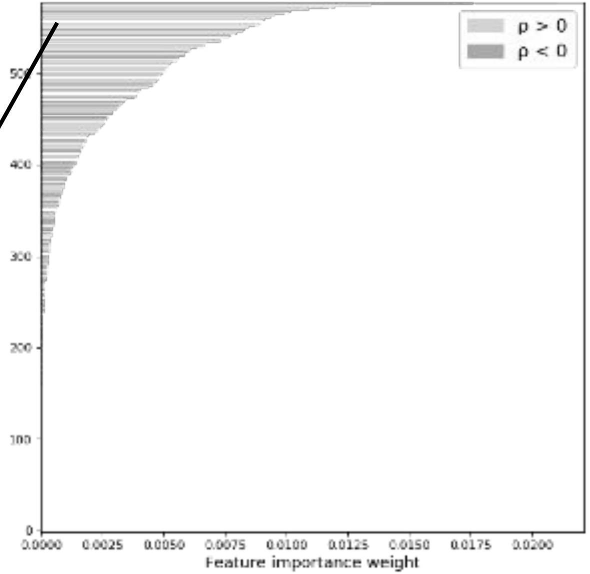
(1) **Uniqueness**: variety of features in explanations
(2) **Concise**: on average, 0.3% of the features in the explanation
(3) **Comply** with regulatory requirements (e.g., GDPR)





Local XAI (vs. global)

‘Tobacco’ ranked 73rd out of 578 features in feature relevance list of *‘Neuroticism*’ model



Person B:

IF person B had spent *less frequently* in Shops and Tobacco, and *less money* on Subscription and Tobacco ⇒ THEN he would not have been predicted as Neurotic

Conclusions

- Both global & local XAI methods are important to open black box, especially when modeling digital footprints data
- Different use cases:

  → **Global:** (i) trust & validation, (ii) audit functionality, (iii) insights, (iv) improve

  → **Local:** (i) provide unique & personalized insight into how data is used, (ii) validate individual predictions

Thank you!