



COS10022 – DATA SCIENCE PRINCIPLES

Dr Pei-Wei Tsai (Lecturer, Unit Convenor)
ptsai@swin.edu.au, EN508d

SWIN
BUR
NE

SWINBURNE
UNIVERSITY OF
TECHNOLOGY



HOW ARE YOU CONTROLLED BY BIG DATA + PSYCHOLOGY?

What Data Can Do for You?

The base for recommendation applications

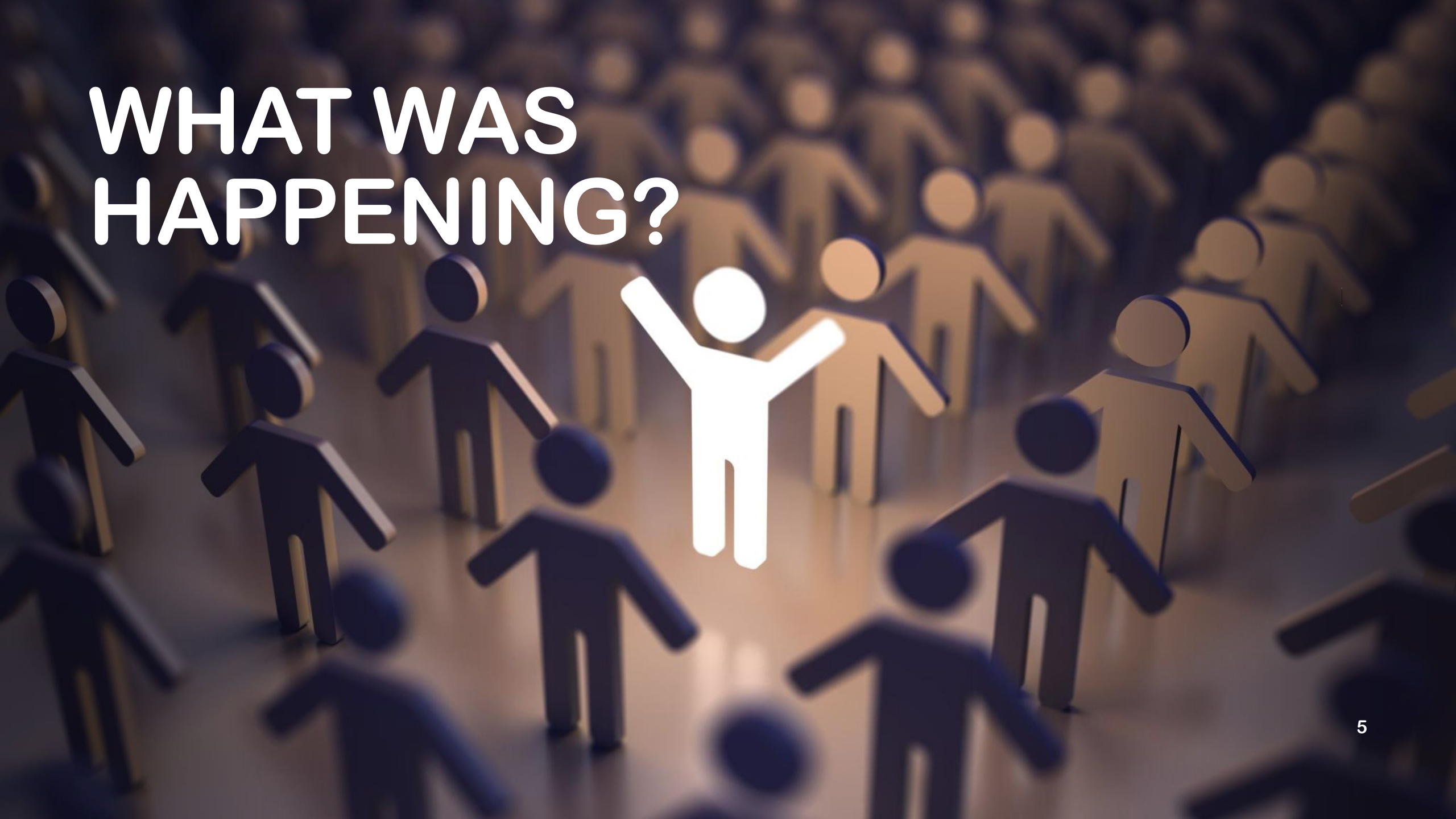
Driving peoples' behaviours without being noticed

Impact Example

**Facebook & Cambridge Analytica
Scandal**

**Behavioural Psychology + Big Data
can change human behaviour**

WHAT WAS HAPPENING?





Cambridge
Analytica

Main Character

Strategic Communication Laboratories (SCL) Group

- Was a private British behavioural research and strategic communication company.
- SCL's subsidiaries: Cambridge Analytica and Crow Business Solutions MENA are involved in the Facebook-Cambridge Analytica data scandal.

Elections that SCL was Spotted

Thailand, 1997



Trinidad & Tobago, 2009



Italy, 2012



Argentina, 2015



Kenya, 2013



Ukraine, 2004



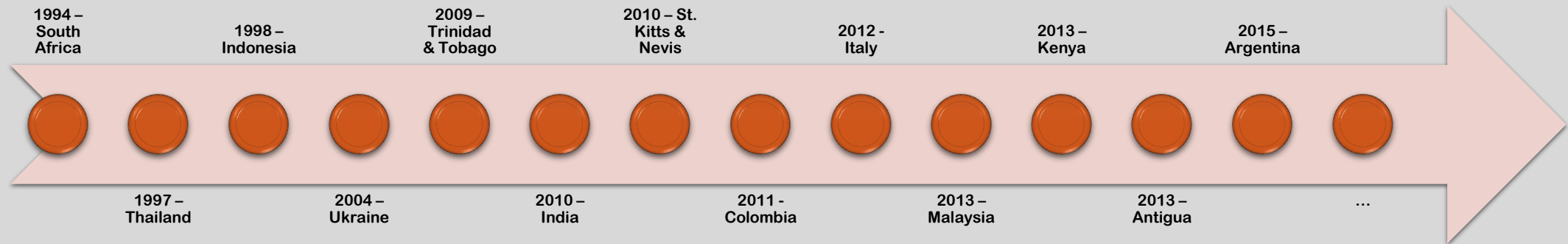
India, 2010



Malaysia, 2013



Elections that SCL Involved



RESEARCH ARTICLE | SOCIAL SCIENCES | OPEN ACCESS



Private traits and attributes are predictable from digital records of human behavior

Michal Kosinski , David Stillwell, and Thore Graepel [Authors Info & Affiliations](#)

March 11, 2013 | 110 (15) | <https://doi.org/10.1073/pnas.1218772110>



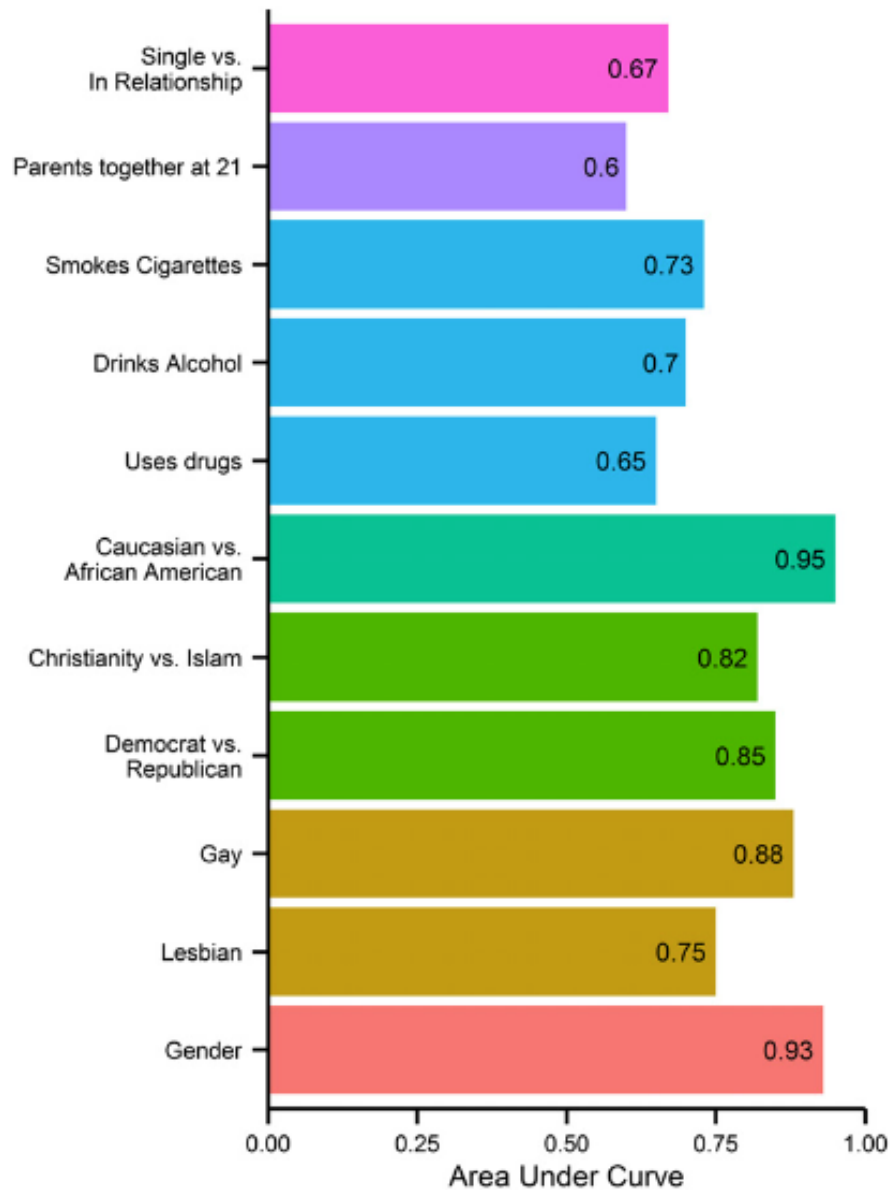


Fig. 2. Prediction accuracy of classification for dichotomous/dichotomized attributes expressed by the AUC.

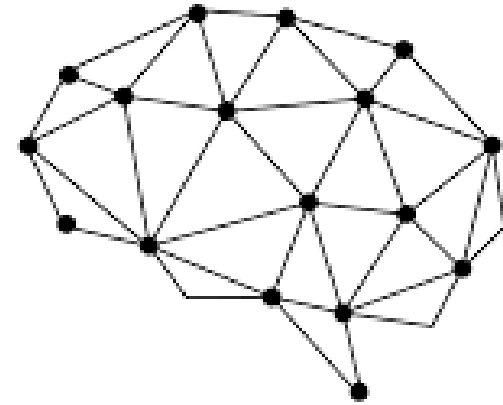
Prediction Accuracy Analysis

- By simply utilise the comments, forward, and the click on the like button can reveal a high accuracy prediction results on some dichotomous/dichotomized attributes.

ESTABLISHING THE SUBSIDIARY

Since this is a good business, SCL creates a subsidiary called “Cambridge Analytica.”

Let’s call it **CA** for simplification.



Cambridge
Analytica

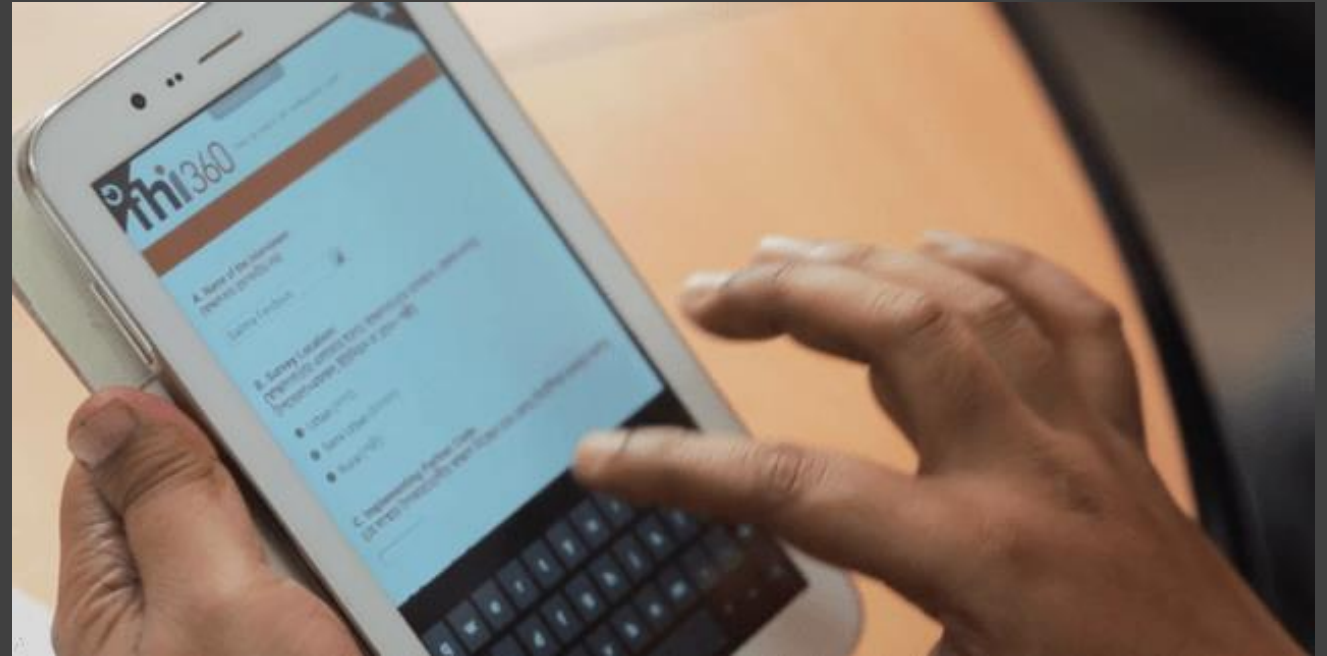
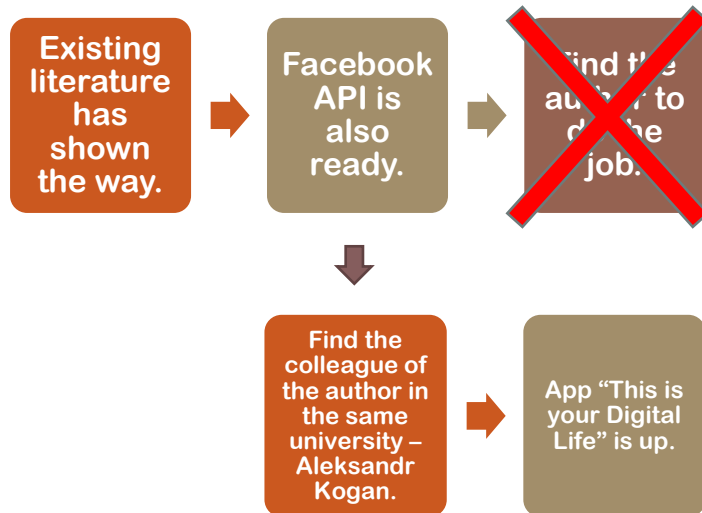
Goal of CA

CA is a profit-driven company. The goal is specific and thus the operation can be clearly defined.

What CA does include:

- **Aggregating Data.**
- **Analysing Data.**
- **Changing Crowd Behaviours.**

DATA AGGREGATION



This Is Your Digital Life
will receive your public profile.



Security Breach

“This is your digital life” is an upgraded version of the original questionnaire created by the original authors.

- It has 120 questions, which is not that attractive. But, they offer money (around \$2 - \$4) for the participant who completed the questionnaire.

When retrieving the data of the users who completed the questionnaire, they found a security breach of Facebook resulting that they can also gather data of the users' friends.

- This means that any of your friend on Facebook filled in the questionnaire, which is totally not related to you, results in all your data is collected without your permission.

Finding this “exciting” fact, Kogen speeded up the issuing of the questionnaire. In a few months, he has issued 270,000 questionnaires. As the return, he collected 50,000,000 users' data. (some reports says 70,000,000 or 80,000,000)

CA confessed that they had collected the data of all people in US who is eligible to vote.



Data Analysis

CSL Research Toolkit

Data Analysis

The big data source is collected. Now CA can move into the data analysis phase.

With building models to find relations between attributes, it is easy to conclude what is the personality of a user.

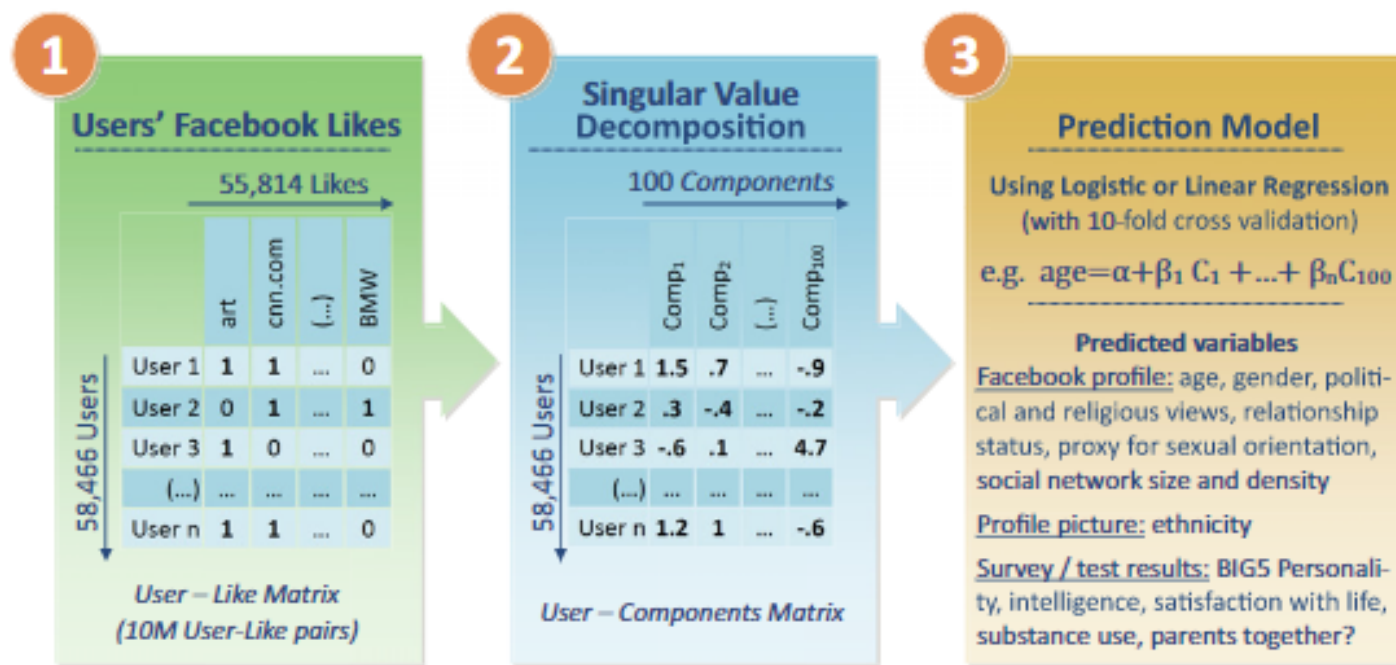
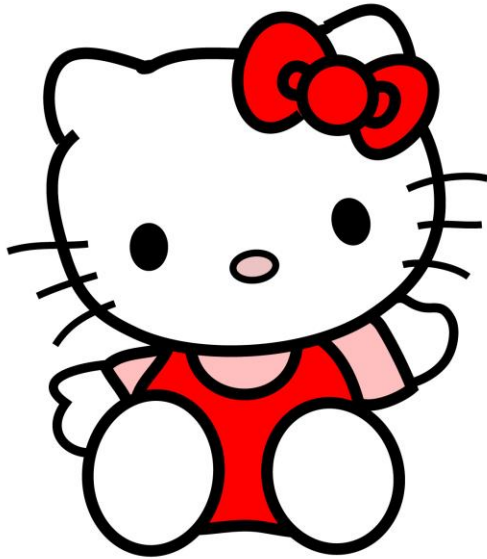


Fig. 1. The study is based on a sample of 58,466 volunteers from the United States, obtained through the myPersonality Facebook application (www.mypersonality.org/wiki), which included their Facebook profile information, a list of their Likes ($n = 170$ Likes per person on average), psychometric test scores, and survey information. Users and their Likes were represented as a sparse user–Like matrix, the entries of which were set to 1 if there existed an association between a user and a Like and 0 otherwise. The dimensionality of the user–Like matrix was reduced using singular-value decomposition (SVD) (24). Numeric variables such as age or intelligence were predicted using a linear regression model, whereas dichotomous variables such as gender or sexual orientation were predicted using logistic regression. In both cases, we applied 10-fold cross-validation and used the $k = 100$ top SVD components. For sexual orientation, parents' relationship status, and drug consumption only $k = 30$ top SVD components were used because of the smaller number of users for which this information was available.



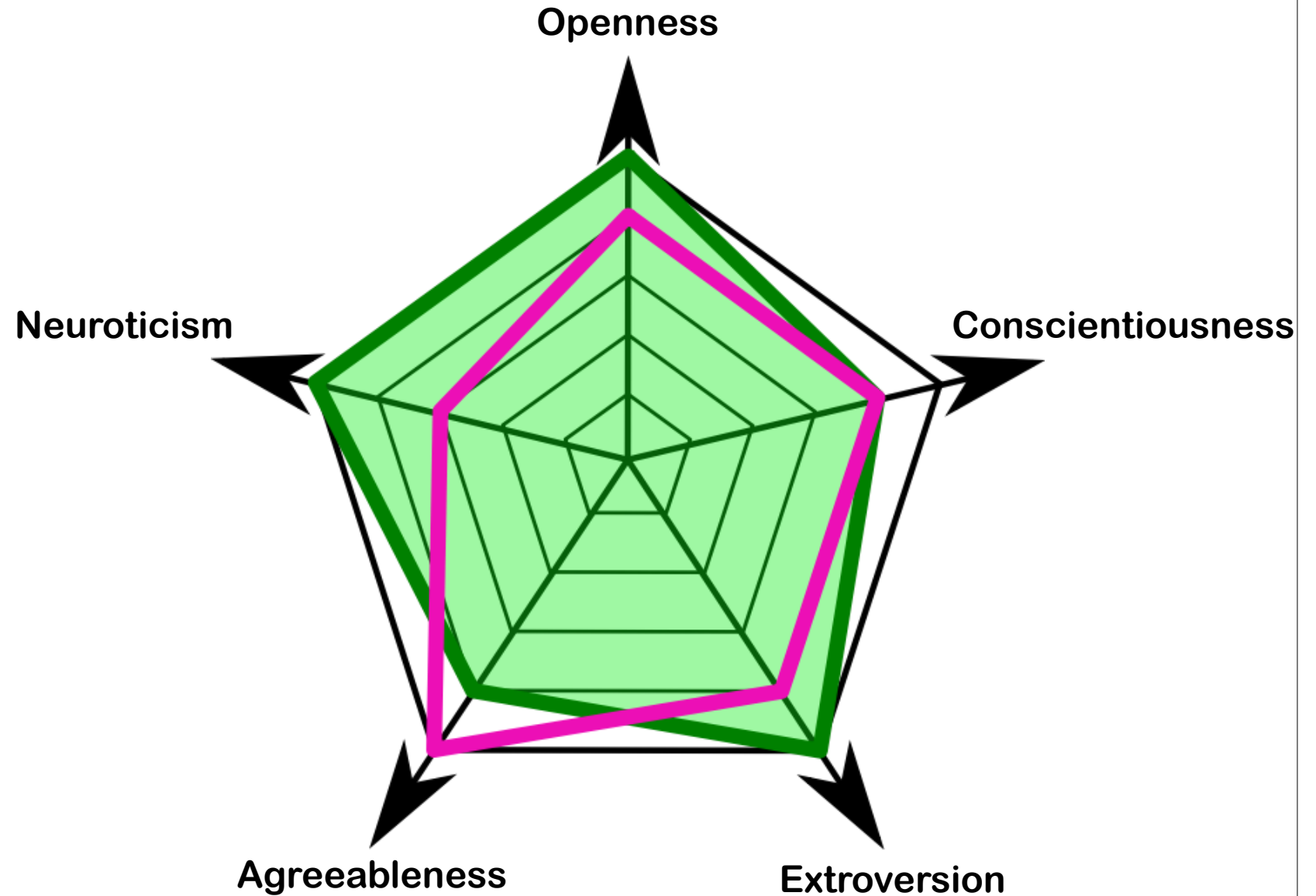
Example Analysis Outcome

If a user who liked the “Hello Kitty” brand tended to be **high on Openness** and **low on “Conscientiousness,” “Agreeableness,” and “Emotional Stability.”**

They were also more likely to have **Democratic political views** and to be of **African-American origin**, **predominantly Christian**, and **slightly below average age**.

CSL Proposed OCEAN Analysis

- OCEAN is a scoring system with 5 attributes.
- The result can be used to conclude the characteristic of a user.



CROWD BEHAVIOUR CHANGING

- Since the data aggregation and analysis are nailed, CA start to involve in the US election and joined the team of Trump.
- The last work is what CA most capable of: affect the election result by changing crowd behaviours.



of consistent policy and variegated communications contributed to the People's Partnership Coalition's landslide victory. The victory likewise saw Trinidad & Tobago's first-ever female Prime Minister. SCL subsequently supported the governing coalition with ongoing advice.



How PSYOP Affects the Election Result?

In 2009-2010, SCL was involved in the election in Trinidad and Tobago.

The population in this country can be roughly divided into two races (Let's call them A and B).

Team A is more traditional, and the family education is stricter while team B is a bit loose on family education.

SCL is involved as the consultant in team A.

After a series of analysis, they decide to focus on the youth people.

The corresponding strategy is proposing an activity called "Do So."





2016 US ELECTION

Involved?

Cambridge Analytica

US Election

Brexit



**DO PEOPLE REALLY
NOT KNOW THEIR
DATA IS BEING
COLLECTED?**



RELATED NEWS

<https://about.fb.com/news/2018/03/suspending-cambridge-analytica/>

Meta

Suspending Cambridge Analytica and SCL Group From Facebook

March 16, 2018

“
Protecting people’s information is at the heart of everything we do, and we require the same from people who operate apps on Facebook.
”

By [Paul Grewal](#), VP & Deputy General Counsel

Update on March 17, 2018, 9:50 AM PT: The claim that this is a data breach is completely false. Aleksandr Kogan requested and gained access to information from users who chose to sign up to his app, and everyone involved gave their consent. People knowingly provided their information, no systems were infiltrated, and no passwords or sensitive pieces of information were stolen or hacked.

Originally published on March 16, 2018:

We are suspending Strategic Communication Laboratories (SCL),