university of groningen / faculty of economics and business / customer insights center

Turning rich data into insights in a privacy-concerned world

IMU spring conference
Mannheim March 18, 2021

Prof. dr. Jaap Wieringa
Department of Marketing
University of Groningen
J.E.Wieringa@rug.nl

Joint work with Gilian Ponte

Prof. dr. Jaap E. Wieringa

- Current position:
  - full professor of Research Methods in Business at the Department of Marketing, University of Groningen; research RUGCIC
- Experience:
  - 20 years of academic and applied research and teaching at the bachelor, master, PhD, and excutive level
  - business consulting (many member companies of Customer Insights Center: Philips, GE, DE, Paccard, B&D)
- Education: MSc & PhD in econometrics
- Expertise: data science, marketing analytics, econometrics, statistics, pharmaceutical marketing
- Hobbies: theater, running

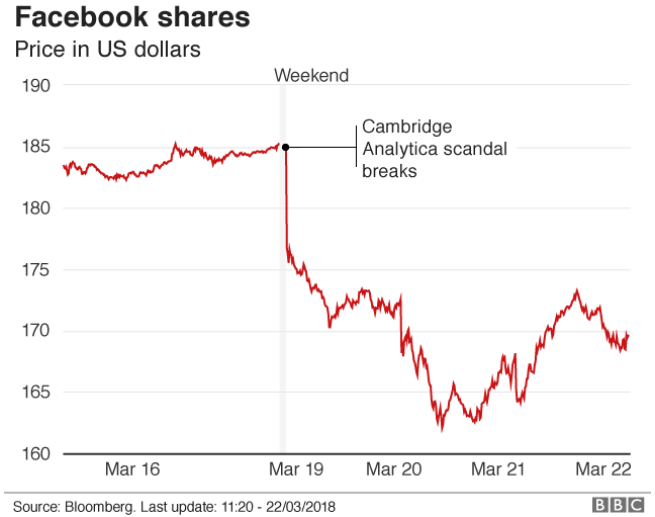
Privacy concerns…

- 92% of the US consumers are concerned about their privacy (TrustE 2016)
  - Collection/storage/use

- Do consumers accept that firms collect and use all this information?

Example: Facebook - Cambridge Analytica

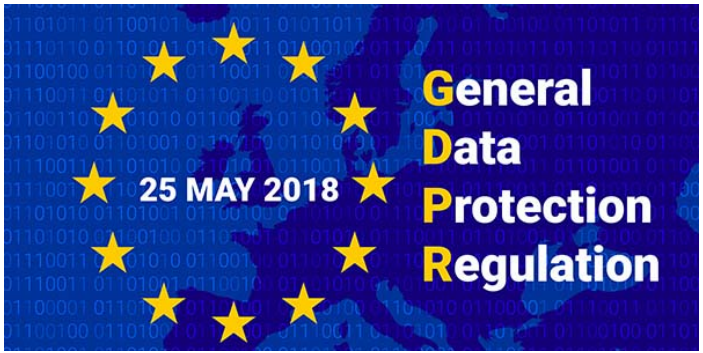
- Shared data of 87 million users
- Risks of such endeavors are:
  - Potential privacy costs
  - Losses in brand value
  - Legal fines or customer trust



(Reuters, 2019; Schneider, Jagpal, Gupta, Li & Yu, 2017)

Legislation…

- European Union (GDPR) and the United States adopted legislation to regulate and protect the individual data of customers.
  - Consent
  - Fair processing notices
  - Data subject rights
  - Personal data breach



(Columbus, 2014; European Parliament, 2013; PCAST, 2014, The California Consumer Privacy Act of 2018)

Legislation…

› As of 1-1-2020, CCPA became effective

The intentions of the Act are to provide California residents with the right to:

1. Know what personal data is being collected about them
2. Know whether their personal data is sold or disclosed and to whom
3. Say no to the sale of personal data
4. Access their personal data
5. Request a business to delete any personal information about a consumer
6. Not be discriminated against for exercising their privacy rights

Consequences for marketing analytics?

- Question:
  - how to guarantee privacy
  - without sacrificing power in applications of analytics?

- Possible lines of development:
  - ‘Data minimization’
  - ‘Data anonymization’

university of groningen / faculty of economics and business / customer insights center

Data minimization

Data minimization

- Data should be stored only if necessary
- When required for and in line with goals, determined a priori

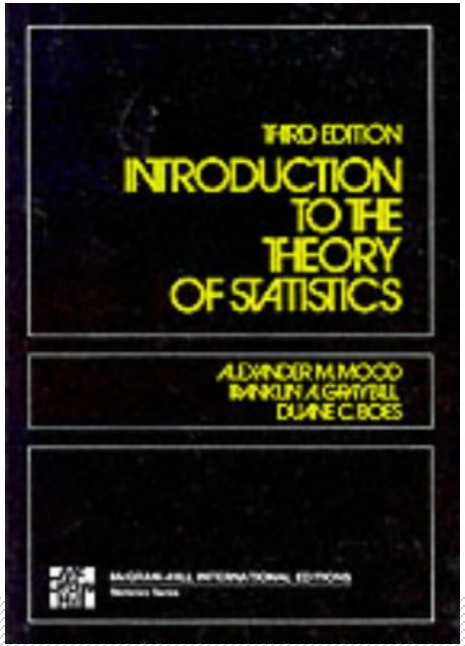
AUTORITEIT PERSOONSGEGEVENS

BfDI Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit

Store or destroy

Has the agreed term of personal information passed? Or is the information no longer required. Organizations should then delete the observations.

How?



State Space modeling

Professor Rudolf Emil Kalman
(1930-2016)

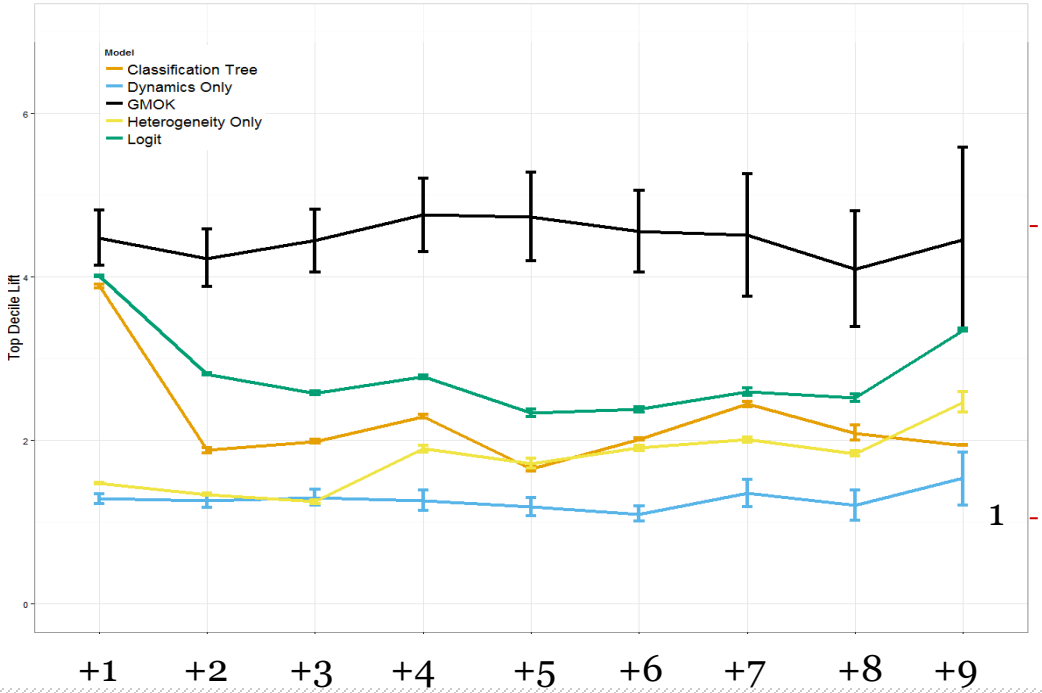
Kalman filter


?
State ↔ Stimuli

Application: predicting customer churn

- The number one problem for telcos around the world (Mattison 2005)
- Churn is bad…
- Churn is expensive
- Churn is difficult to manage
  - Big surprise
  - Difficult to explain
  - Difficult to predict
  - Difficult to defend against

Insurance case: Top decile lift



≈ 4.5 times better than random selection

1

Random selection

Quality of predictions does not deteriorate with time!

Paper:

ELSEVIER

International Journal of Research in Marketing

Volume 34, Issue 1, March 2017, Pages 154-172

No future without the past? Predicting churn in the face of customer privacy ☆

Niels Holtrop ✉, Jaap E. Wieringa, Maarten J. Gijsenberg, Peter C. Verhoef

Show more

https://doi.org/10.1016/j.ijresmar.2016.06.001

Abstract

university of groningen / faculty of economics and business / customer insights center

Data anonymization

Data anonymization

- Basic idea: can we generate ‘fake data’ from a real-life marketing data set, such that
  - anonymity is realized
  - the fake data retains the same marketing insights as the original data?


Generative adversarial networks (GANs).

GANs…

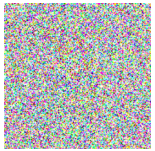
- Developed in 2014 by Ian Goodfellow
- Basic idea: make two neural networks compete (or collaborate, it's a matter of perspective) with each other
- "the most interesting idea in the last 10 years in Machine Learning" (Yann LeCun, co-recipient of the 2018 ACM A.M. Turing Award for his work in deep learning)

Create Anime characters

- GAN can auto-generate and colorize Anime characters.

How do GANs work?

$z \sim \text{Normal}(0,1)$



Generator (G)

Discriminator (D)

real/fake

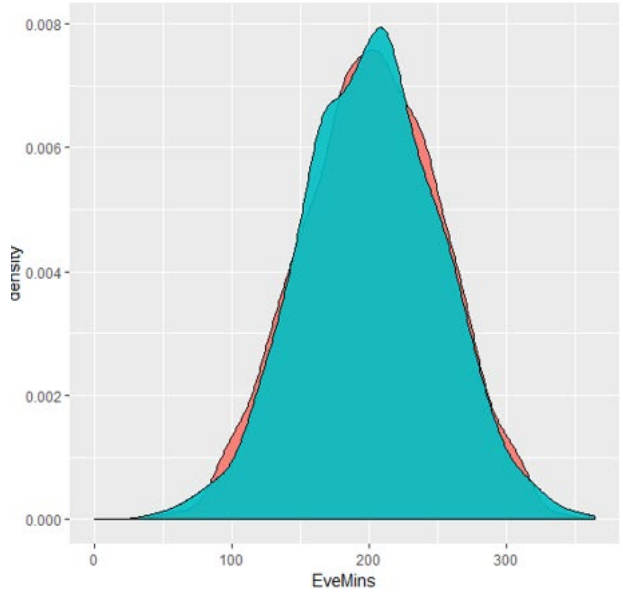
Data set

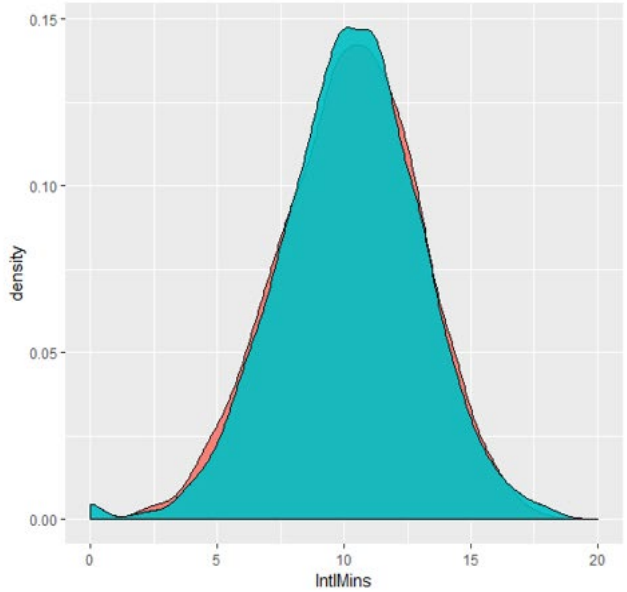
So we can generate fake data that resembles real data?

Three data sets

1. Synthetic churn data – 3333 observations, individual customer level, Telco, 17 explanatory variables, cross-sectional
2. Real churn data – 1,262,423 observations, individual customer level, insurance, 25 explanatory variables, cross-sectional
3. Sales data – 4858 observations, supermarket chain-level, FMCG, 15 explanatory variables, panel

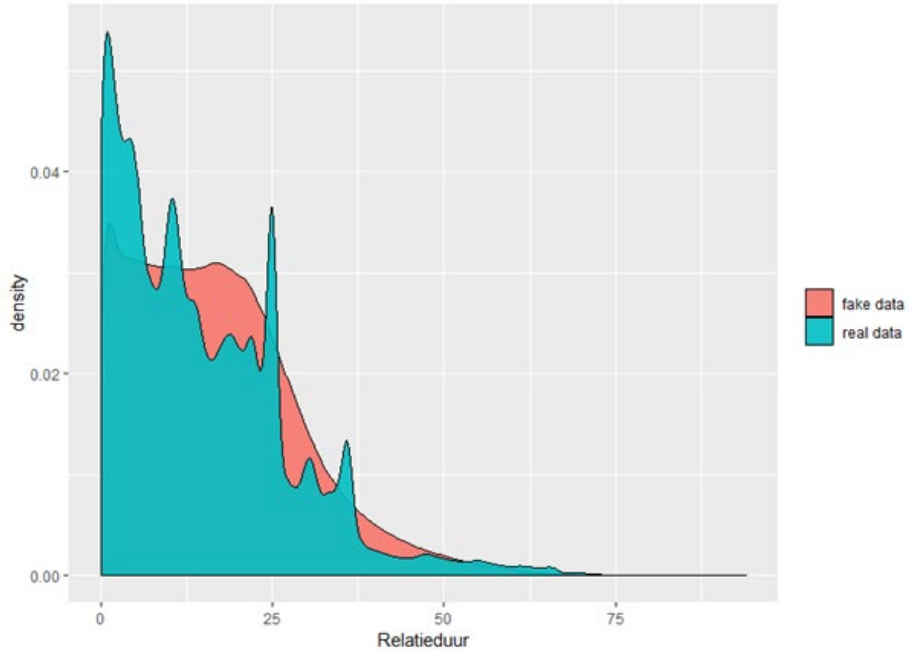
‘Model-free’ evidence



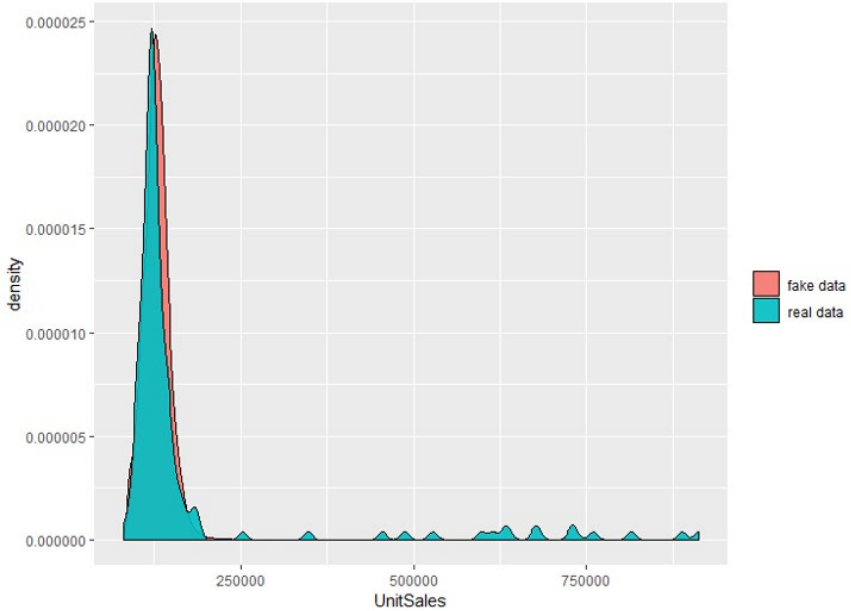


*Red = fake data, green = real data.

‘Model-free’ evidence

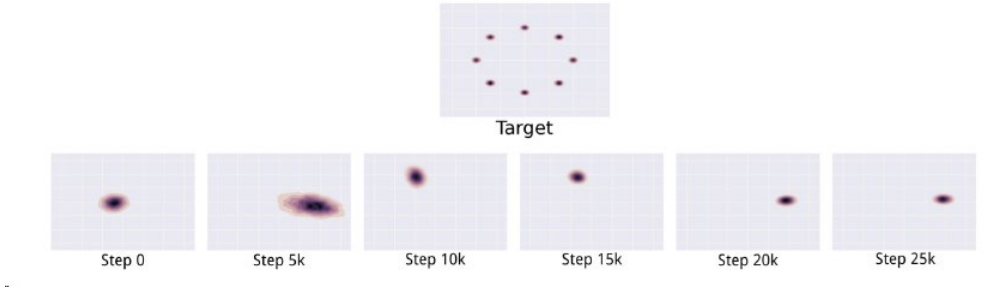


‘Model-free’ evidence



Issue with GANs - Mode collapse

- Mode collapse is a problem that occurs when the generator learns to map several different input z values to the same output point.
- Problems:
  - Counting
  - Perspective
  - Global structure



(Goodfellow, 2016)

So, who cares?

- Companies can use fake data to estimate models while accounting for privacy.
- GANs can help academics in obtaining data
- GANs are *"cool"* and have **great potential** in many fields

university of groningen / faculty of economics and business / customer insights center

Conclusion

Effective analytics can be GDPR-friendly!

- There are ways to be compliant with two important GDPR requirements
  - Data minimization
  - Data anonymization

- ánd generate powerful customer insights from data using analytics

- We discussed two relatively new examples today
  - Promising results
  - Should be further developed

Data minimization

- By employing and implementing existing statistical concepts in state-of-the-art recursive methodology, data minimization can be realized
- Data is only used to update relevant parts of ‘the system’
- After all relevant information is extracted, the raw data can be deleted
- Predictive power is comparable (and in some respects better) than full-data methods.

Data anonymization

- GANs are very new and can be useful in preserving privacy through anonymity
- Are able to retain marketing insights
- Can be trained by relatively small networks
- Are difficult to train
- Need to be developed further

university of groningen / faculty of economics and business / customer insights center

Thank you for your attention!