Measuring patterns of human behaviour through large-scale mobile phone data

Defence of Dr.philos, 22.02.2017

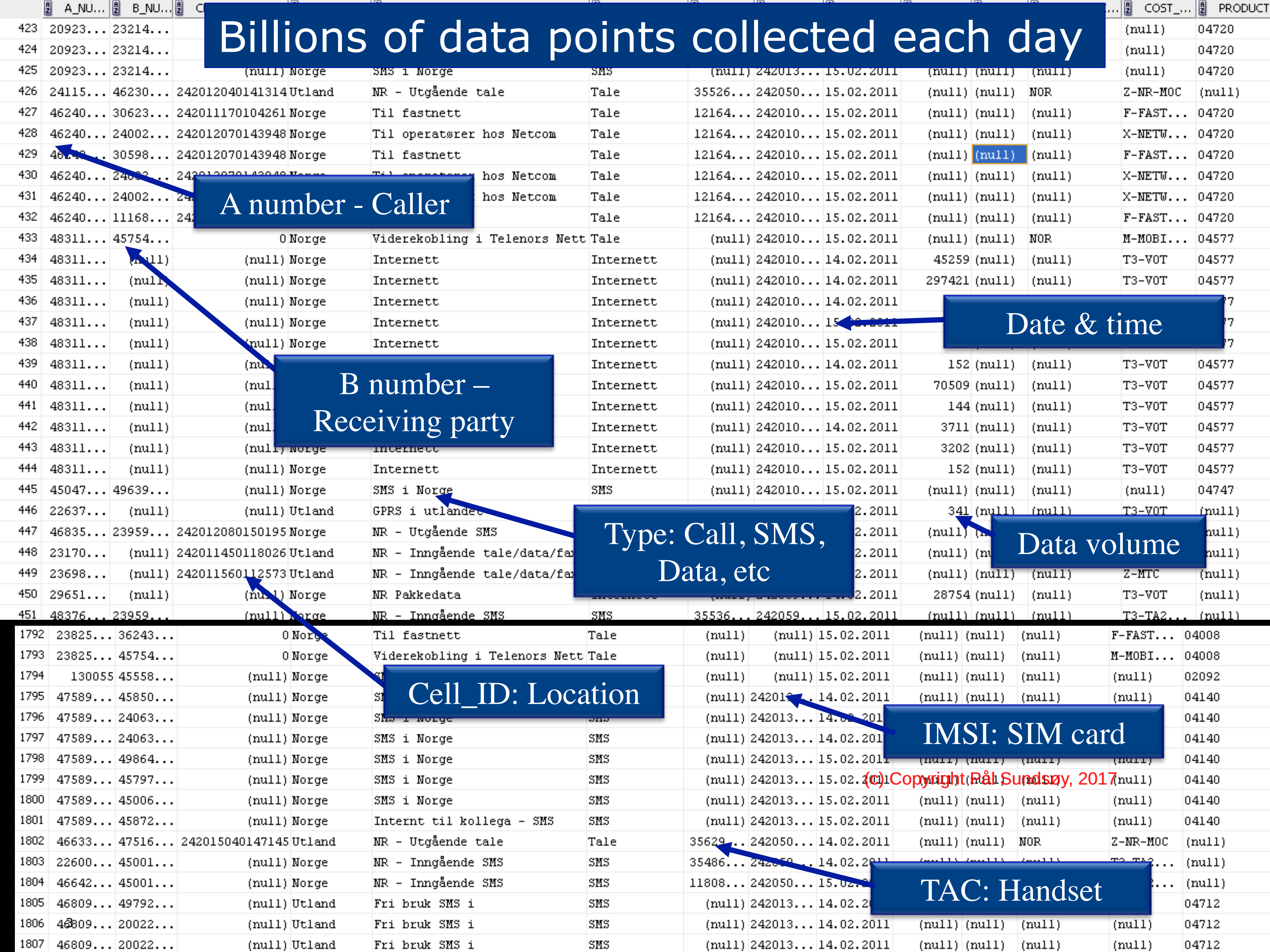
Pål Sundsøy

Members of commitee: Prof. Kåre Synnes, Prof. Zbigniew Smoreda, Prof. Petter Nielsen

Motivation

- One of the most promising rich Big Data sources is mobile phone data
- Mobile phone data can give us new insight into human sociology
- Traditionally mobile phone data has mostly been used for billing the customers and network maintenance.
- Untapped potential

Lazer, D. et al (2009). Computational social science. *Science, 323*, 721-723.


Billions of data points collected each day
A number - Caller
B number – Receiving party
Date & time
Type: Call, SMS, Data, etc
Data volume
Cell_ID: Location
IMSI: SIM card
TAC: Handset
(c) Copyright Pål Sundsøy, 2017

Key contributions

Socioeconomics


Illiteracy


Income


Poverty

Disaseters


Terror attack


Cyclone disaster

Product uptake


Product spreading


Data-driven marketing

Research objective

Apart from providing basic communication services, what kinds of positive impacts can we create for society or individuals using large-scale mobile phone datasets?

Main methodology

- Descriptive
- Prediction

Socioeconomics

Illiteracy

Income

Poverty

Research challenges

- Lacking official statistics in developing countries
- Evaluate if mobile phone data can complement official statistics
- Evaluate different metrics and methods

Main methodology

Prediction

Predicting illiteracy

Mobile phone input features

Approach

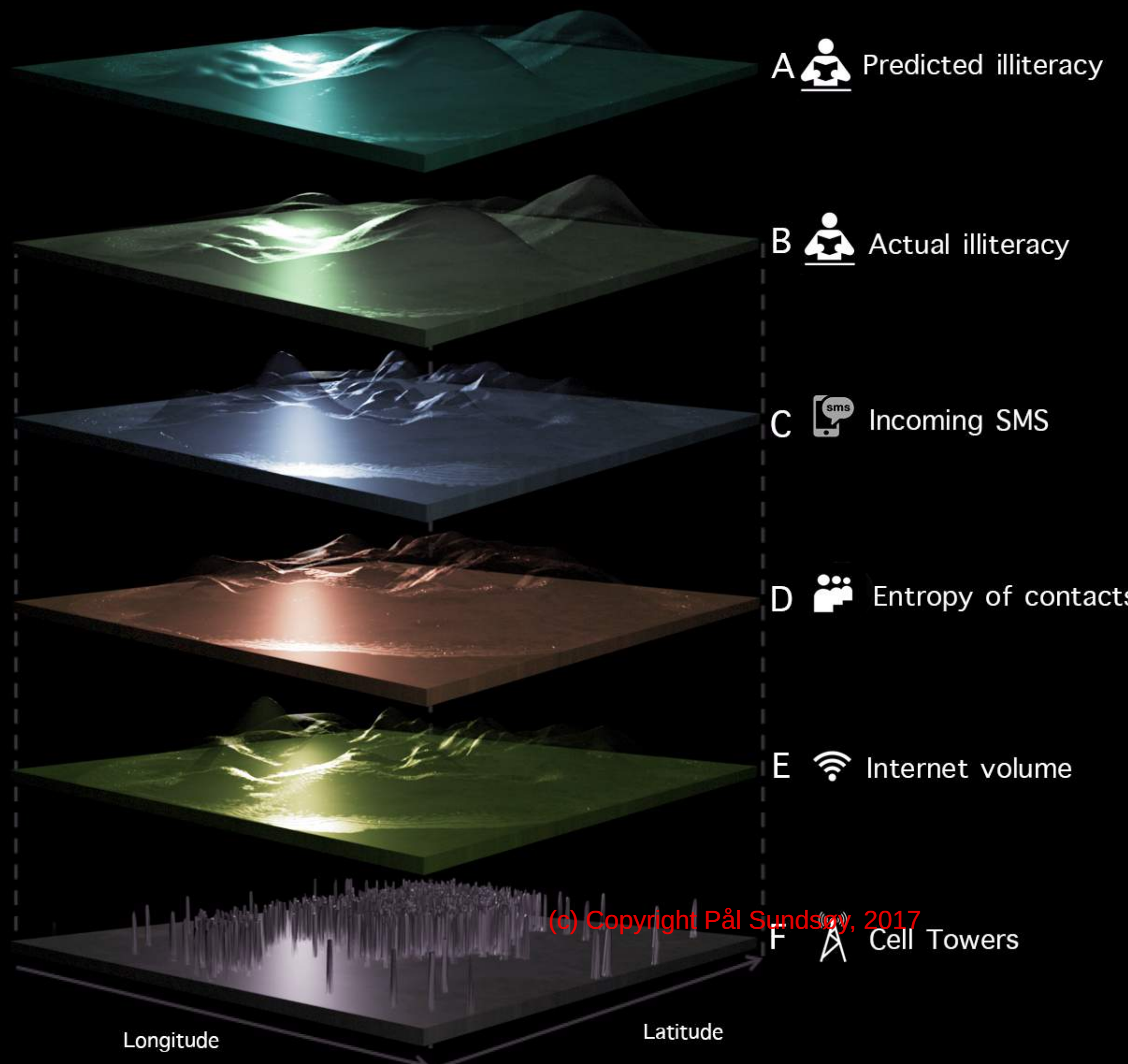
70.1%
Accuracy

Algorithm

Gradient Boosted Trees

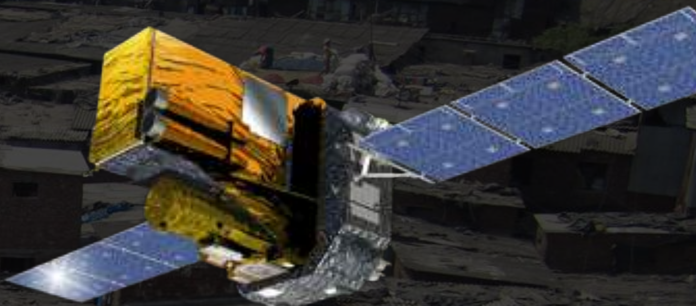
Top illiteracy predictors

1. Location
2. Incoming SMS
3. Entropy of contacts
4. Internet volume
5. Number of places
6. Interactions per contact
7. Recharge amount per transaction



Predicting poverty

Survey data
- Income survey
- DHS
- PPI

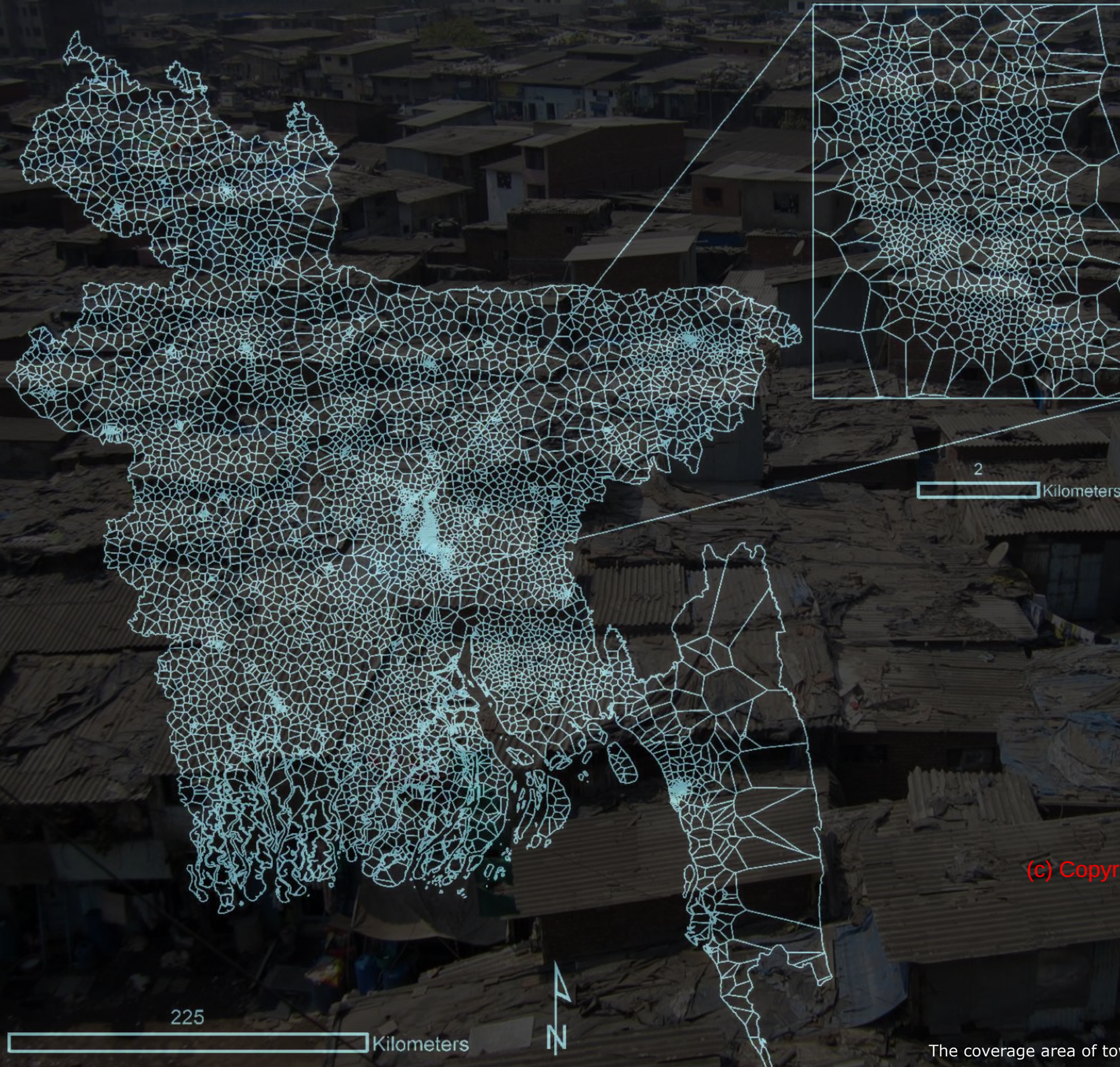


Satellite layers
- Population
- Aridity index
- Evapotranspiration
- Various animal densities
- Night time lights
- Elevation
- Vegetation
- Distance to roads/waterways
- Urban/Rural
- Land cover
- Pregnancy data
- Births
- Ethnicity
- Precipitation
- Annual temperature
- Global human settlement layer

Mobile phone data
- Aggregated anonymized non-personal information
- E.g. average recharge amount per tower

PREDICTION

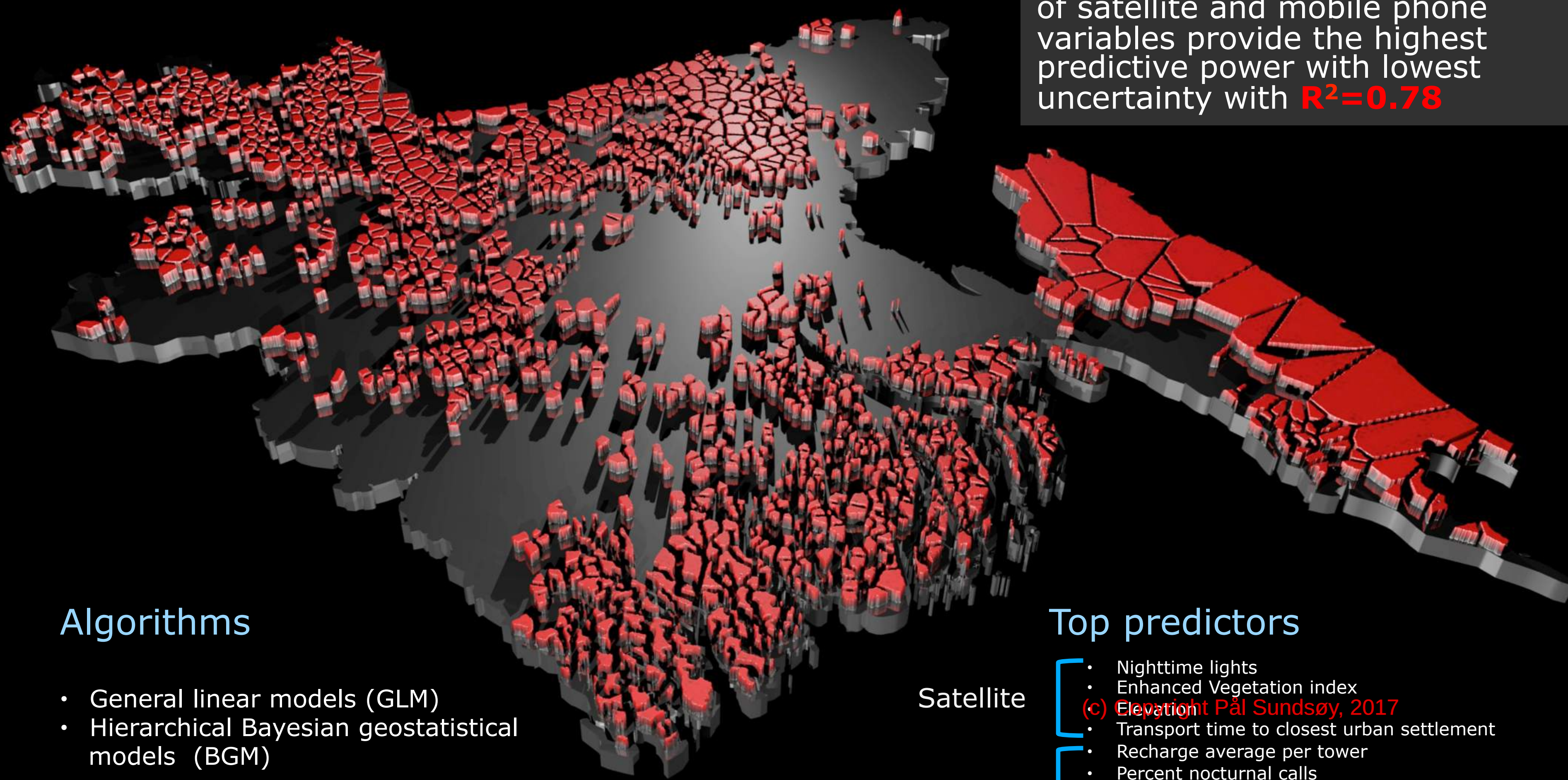
- **Poverty levels**
- **Prediction maps**



The coverage area of towers are approximated with Voronoi-like tessellation

Okabe, A., Boots, B., Sugihara, K. and Chiu, S.N., 2009. *Spatial tessellations: concepts and applications of Voronoi diagrams* (Vol. 501). John Wiley & Sons.

= Poorest areas (Wealth index)



Models employing a combination of satellite and mobile phone variables provide the highest predictive power with lowest uncertainty with $R^2=0.78$

Algorithms

- General linear models (GLM)
- Hierarchical Bayesian geostatistical models (BGM)

Top predictors

Satellite
- Nighttime lights
- Enhanced Vegetation index
- Elevation
- Transport time to closest urban settlement

Mobile phone
- Recharge average per tower
- Percent nocturnal calls
- Outgoing internet sessions
- count incoming VAS
- Recharge amount per transaction
- Count incoming texts
- Weekly recharge amount

Disasters

Research challenges

- Evaluate if mobile phone data can give better insight into social patterns during disasters
- Evaluate if behavioral signals may provide insights into damages and where the vulnerable population is located

Terror attack

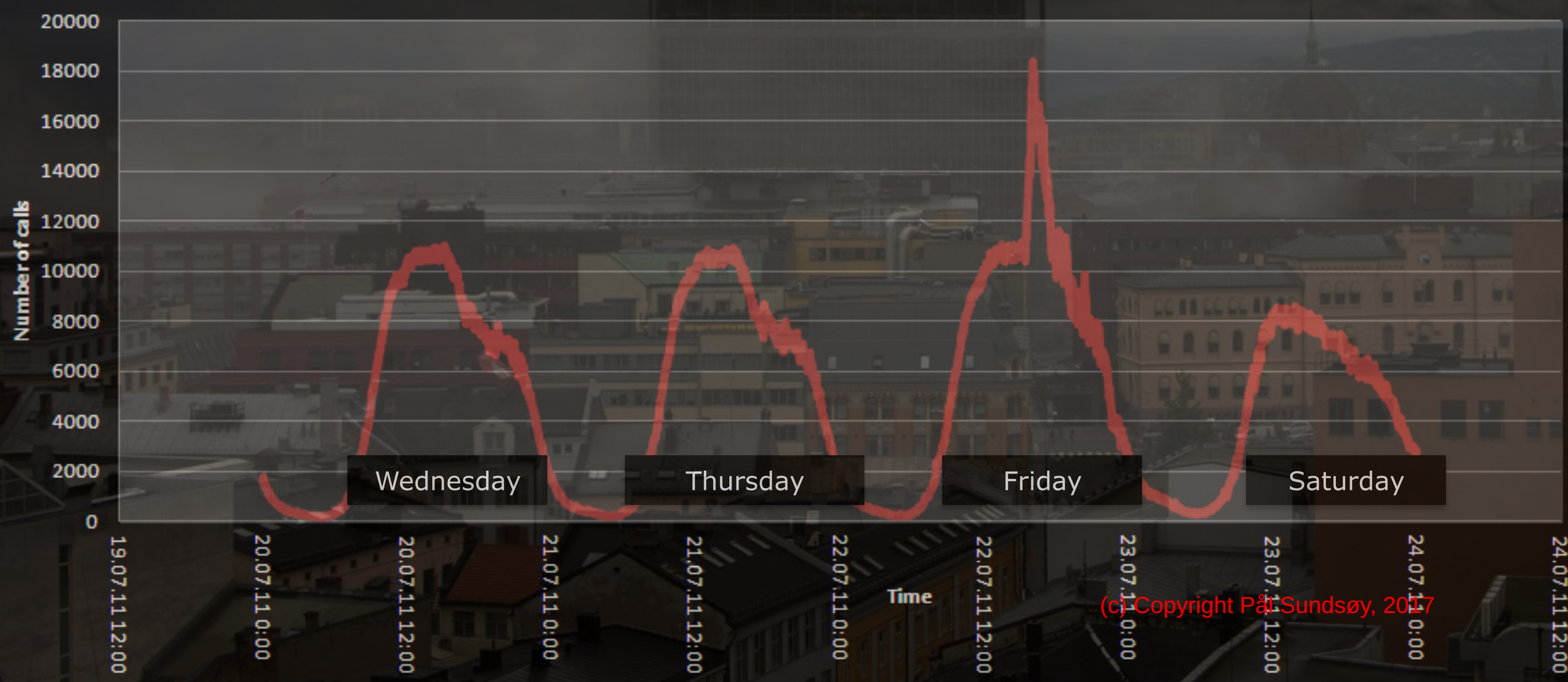
Cyclone disaster

Main methodology

Descriptive

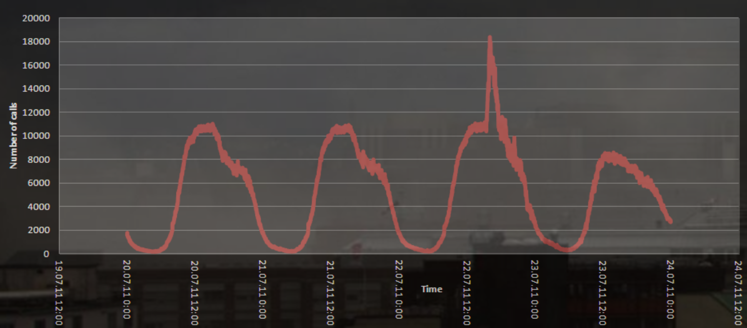
Voice calls minute by minute

Oslo terror attack, 22nd July 2011



Voice calls minute by minute

Oslo terror attack, 22nd July 2011



Socioeconomics | Disasters | Product uptake

Illiteracy | Income | Poverty | Terror attack | Cyclone | Product spreading | Data-driven marketing

= Normal top-up activity

Product uptake

Product spreading

Data-driven marketing

Research challenges

- Evaluate if mobile phone data can be used to understand how products spread over large-scale social networks
- Evaluate if product uptake can be increased by incorporating social effects
- Evaluate how data-driven marketing benchmark against marketers' gut-feeling

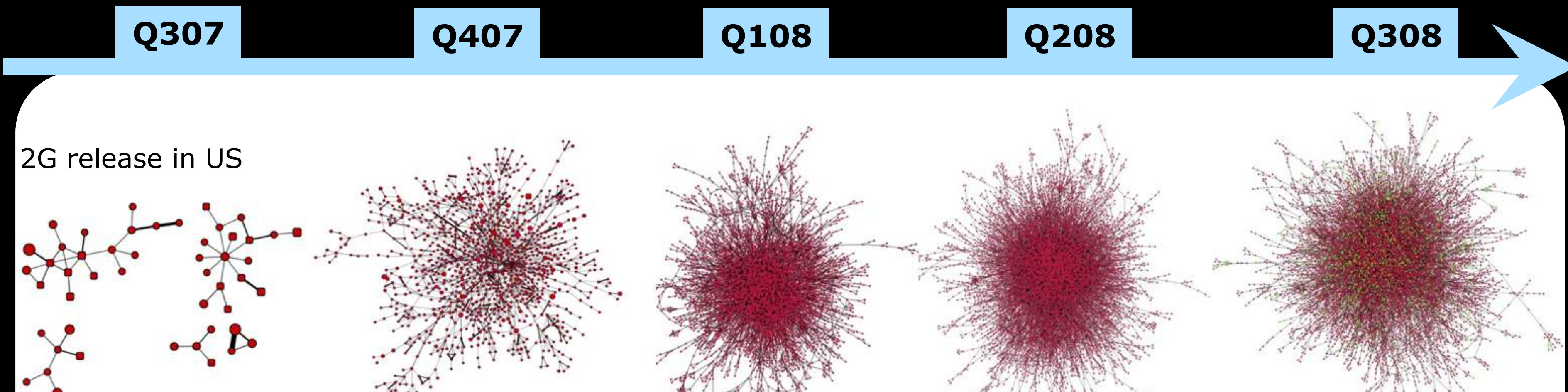
Main methodology

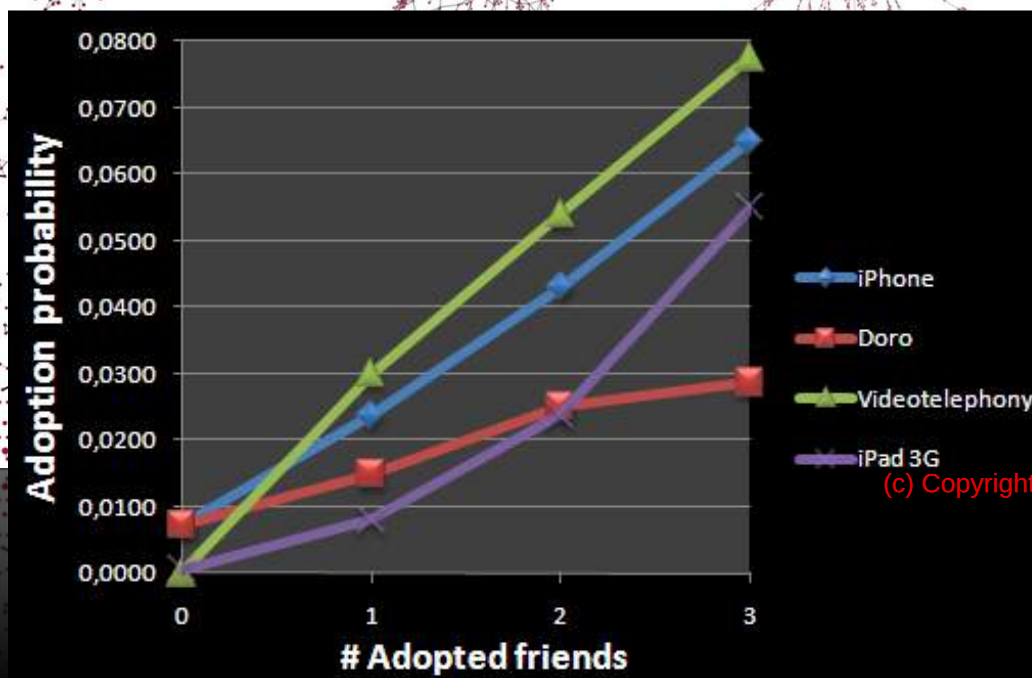
Descriptive

Prediction

Socioeconomics
Disasters
Product uptake
Illiteracy
Income
Poverty
Terror attack
Cyclone
Product spreading
Data-driven marketing
(c) Copyright Pål Sundsøy, 2017
(built from traffic data)
Research on human interactions: By analyzing anonymized CDR-data we can map out a proxy for the social network among our customers

Socioeconomics
Illiteracy
Income
Poverty
Disasters
Terror attack
Cyclone
Product uptake
Product spreading
Data-driven marketing
(c) Copyright Pål Sundsøy, 2017
(built from traffic data)
Research on human interactions: By analyzing anonymized CDR-data we can map out a proxy for the social network among our customers


Q307
Q407
Q108
Q208
Q308
2G release in US


Adoption probability
0,0800
0,0700
0,0600
0,0500
0,0400
0,0300
0,0200
0,0100
0,0000
0
1
2
3
Adopted friends
iPhone
Doro
Videotelephony
iPad 3G

Data-driven approach: Who are most profitable targets for SMS campaign?

Non-convertors 'negatives'

Natural Data Convertors 'positives'

Create model

Find patterns identifying the data convertors based on historic data

300 variables

40M customers

2-6 months back: Use Historical data

Today: Present time data

Non Data Customers today

Model deployment

Use the patterns to identify likely adopters

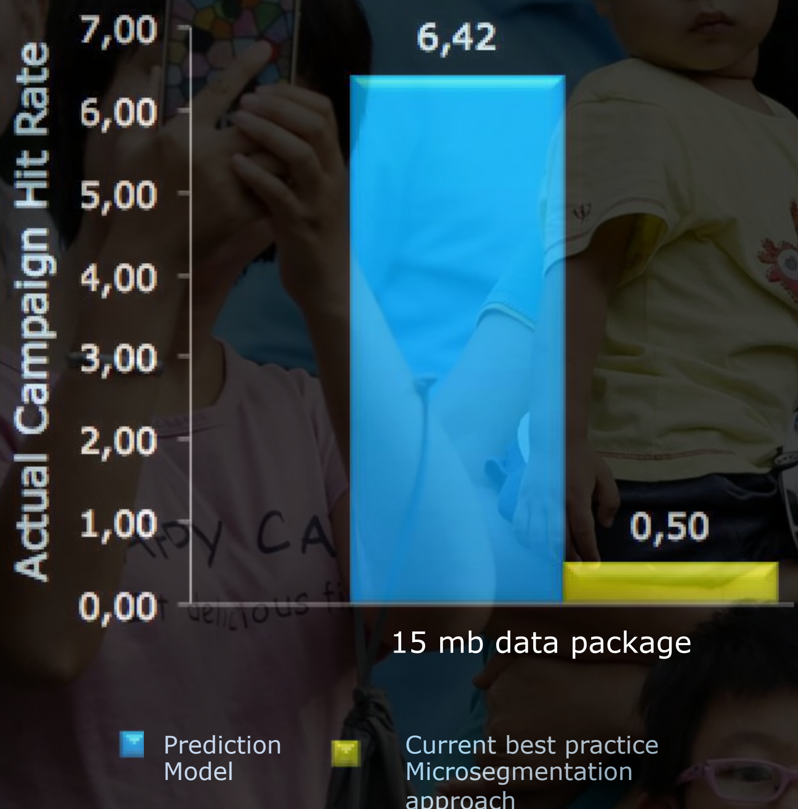
Identify and run campaign on 200k most likely adopters

The predictive model learns from existing cases of data conversion

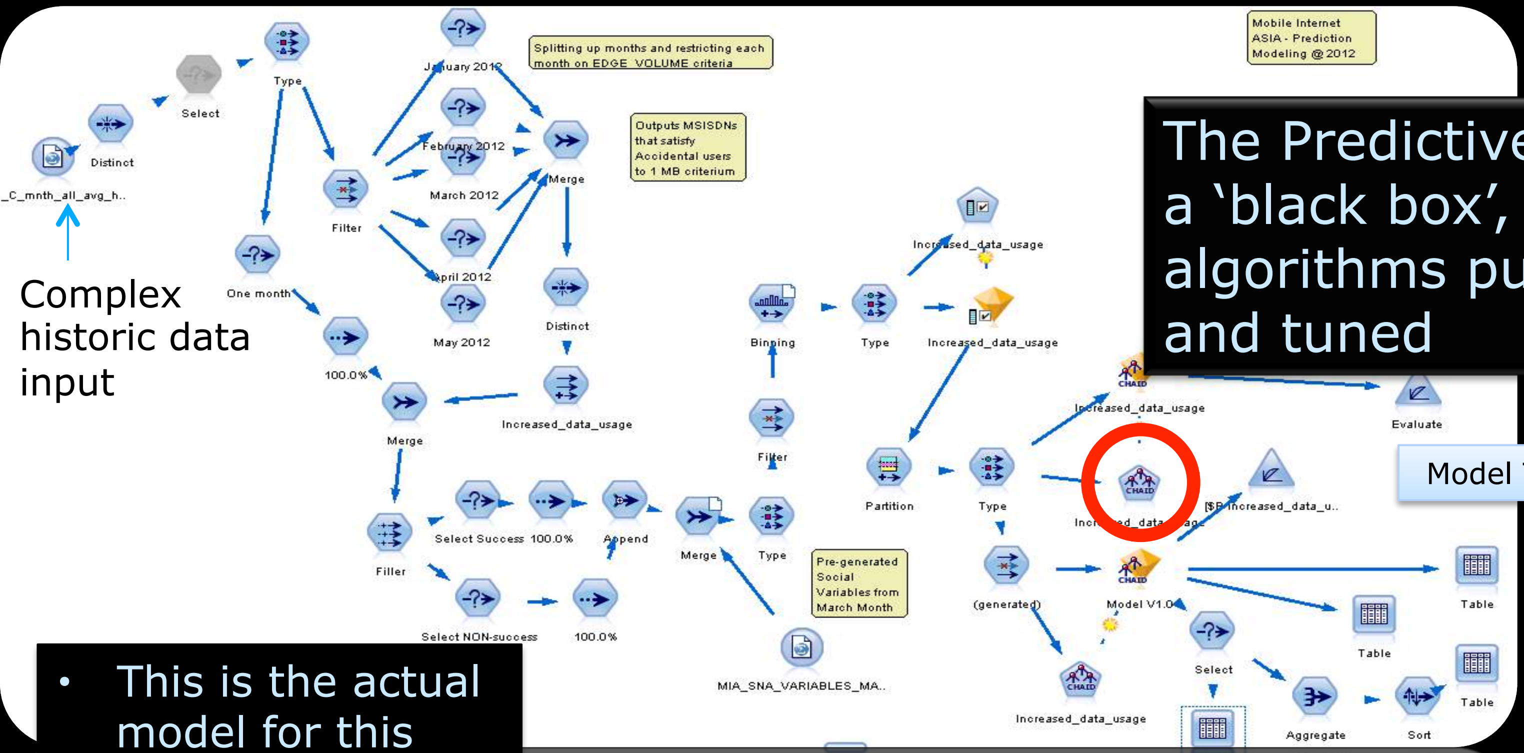
Offers are 15 MB & 99 MB data packages offered for half-price

Algorithm used is Bagging Trees

The prediction model outperforms existing best practice approach with **13 times** better performance



99% Renewal– the algorithm is optimized to avoid 'freeriders'

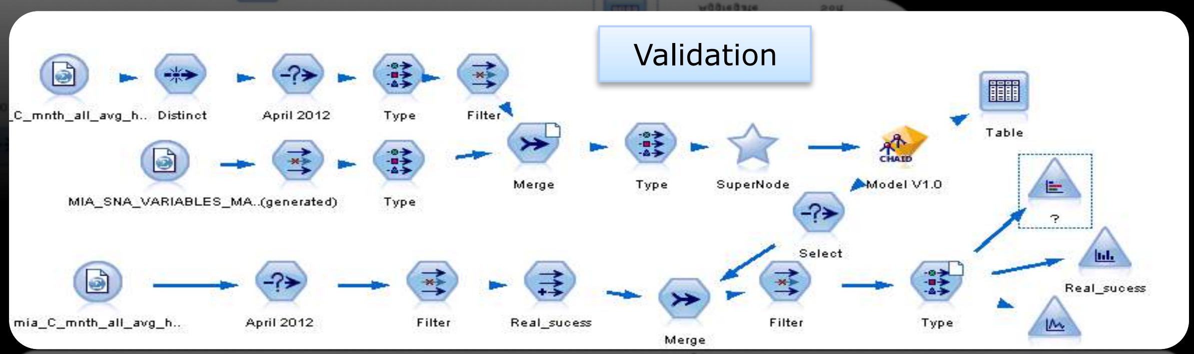


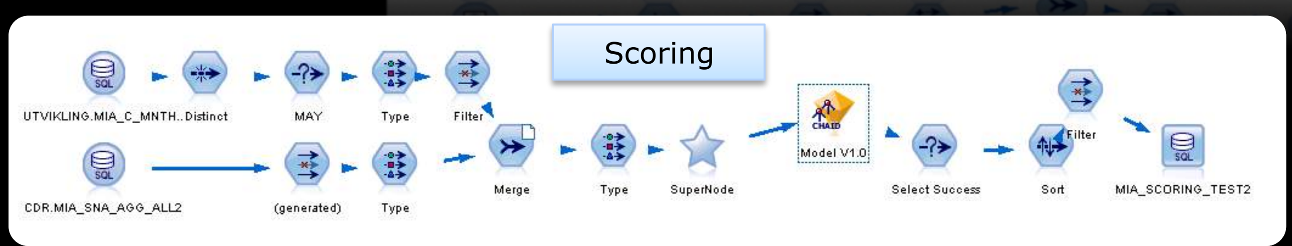
Complex historic data input

The Predictive Model is not a 'black box', but algorithms put together and tuned

Model Training

- This is the actual model for this pilot
- All the boxes are model interaction points
- 80% of the work is data preparation

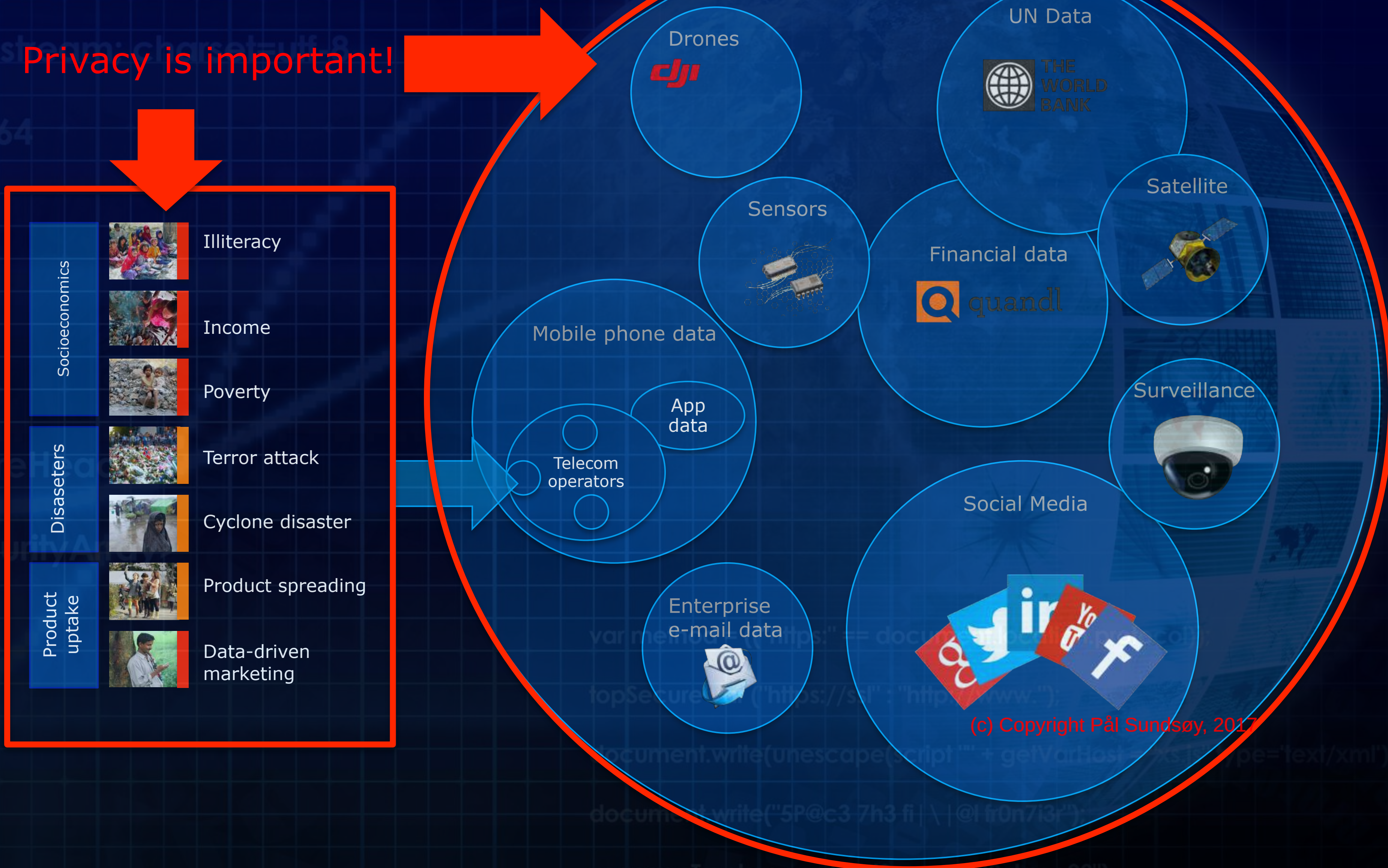




Final Output

The greater 'Big Data' perspective

Sources of behavioral data



1. Lazer, D. et al (2009). Computational social science. Science, 323, 721-723.
2. Golder, S. and Macy, M., 2012. Social science with social media. *ASA footnotes*, *40*(1), p.7
3. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A.H., 2011. Big data: The next frontier for innovation, competition, and productivity

Conclusion

Mobile phone data is useful to :

1. Inform socially beneficial policies

2. Provide insights into human behavior, with the aim of gaining:

    I. A better understanding of human behavior and interactions

    II. Better insights into human behavior to improve marketing

Thank you

1. Can mobile usage predict illiteracy in a developing country?
Preprint available at *arXiv:1607.01337* [cs.AI]. 2016.

2. Deep learning applied to mobile phone data for Individual income classification
Joint work with Bjelland, J., Reme B.A., Iqbal A. and Jahani, E.
Published in *International conference on Artificial Intelligence: Technologies and Applications* (ICAITA). Atlantic Press. 2016.

3. Mapping Poverty using mobile phone and satellite data
Joint work with Steele, J.E., Pezzulo, C., Alegana, V., Bird, T., Blumenstock, J., Bjelland J., Engø-Monsen, K., de Montjoye, Y.A., Iqbal, A., Hadiuzzaman, K., Lu, X., Wetter, E., Tatem, A. and Bengtsson, L.
Published in *Journal of The Royal Society Interface 17*. 2017

4. The activation of core social networks in the wake of the 22 July Oslo bombing
Joint work with Ling, R., Engø-Monsen, K., Bjelland, J. and Canright, G.
Published in *Social Networks Analysis and Mining ASONAM* (pp. 586-590). 2012.

5. Detecting climate adaptation with mobile network data: Anomalies in communication, mobillity and consumption patterns during Cyclone Mahasen
Joint work with Lu, X., Wrathall, D., Nadiruzzaman, M., Wetter, E., Iqbal, A., Qureshi, T., Tatem, A., Canright, G., Engø-Monsen, K. and Bengtsson, L.
Published in *Climatic Change*, 138(3-4), pp.505-519. 2016.

6. Comparing and visualizing the social spreading of products on a large-scale social network
Joint work with Bjelland, J., Engø-Monsen, K., Canright, G. and Ling, R.
Published in *Influence on Technology on Social Network Analysis and Mining, Tanzel Ozyer et. al.* Springer International Publishing. 2012.

7. Big Data-Driven Marketing: How Machine Learning outperforms marketers' gut-feeling
Joint work with Bjelland, J., Iqbal, A., Pentland, A. and de Montjoye, Y.A. Published in *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction* (pp. 367-374). Springer International Publishing. 2014.