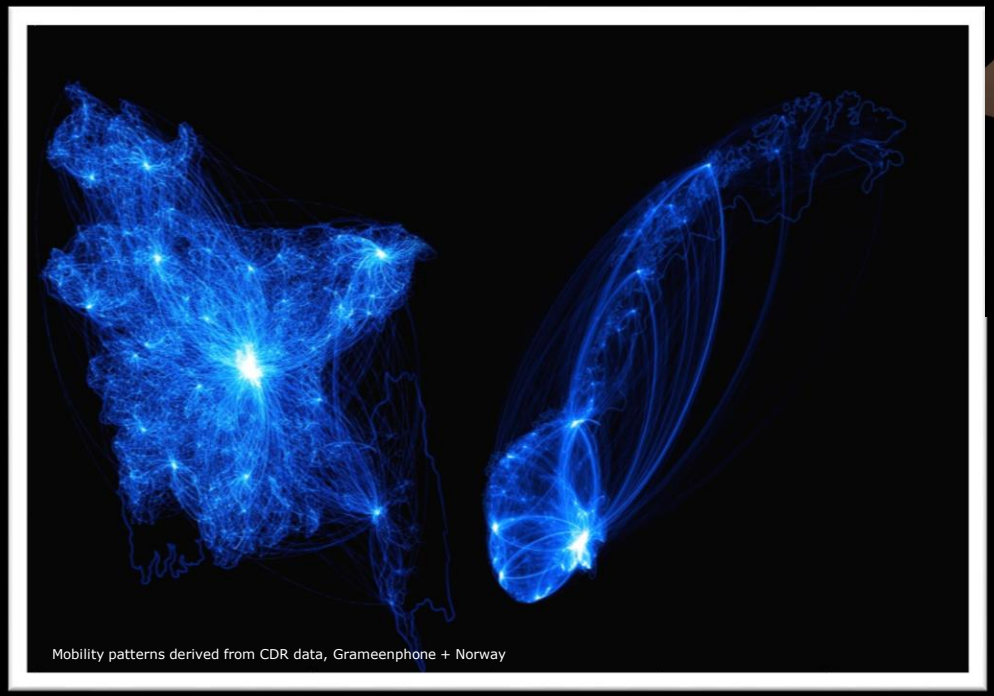
From Viral product spreading to Poverty prediction

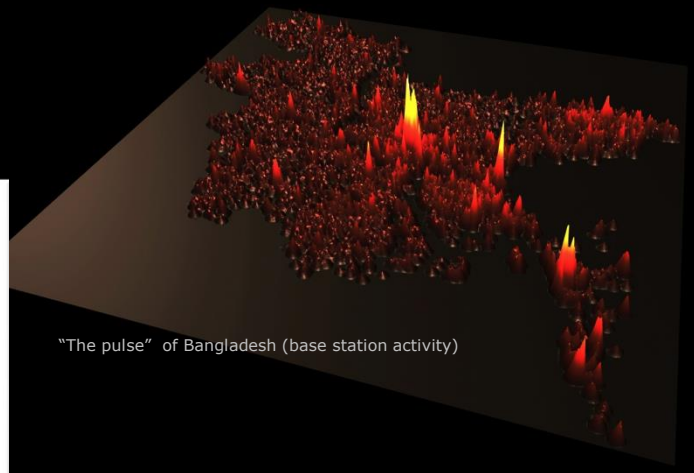
Data Science from a telecom perspective

Pål Sundsøy, Telenor Group Research

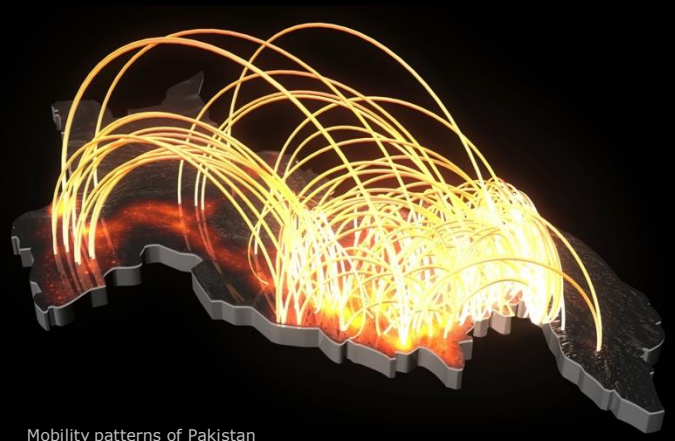
Oct 22nd 2015 International Conference on Big Data for Official Statistics, Abu, Dhabi



Mobility patterns derived from CDR data, Grameenphone + Norway



"The pulse" of Bangladesh (base station activity)

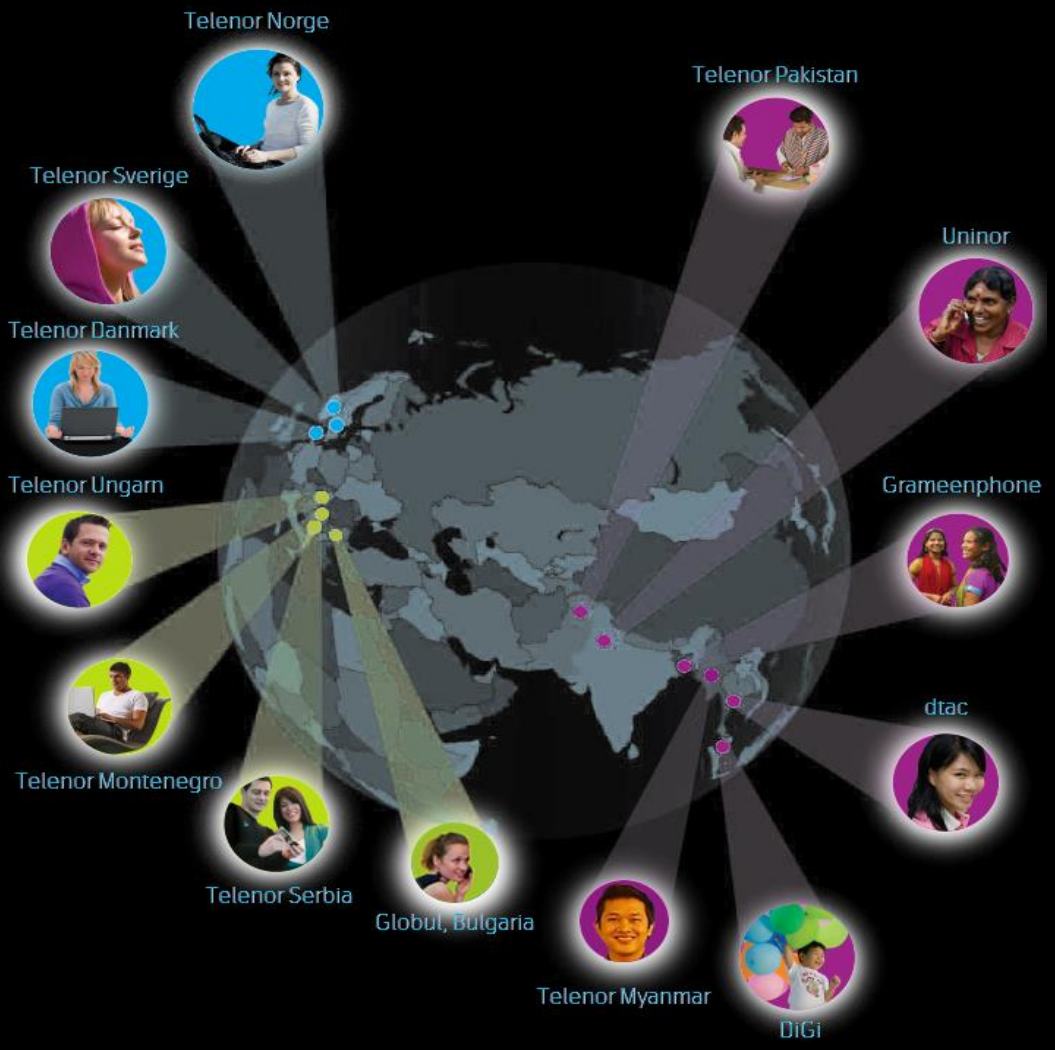


Mobility patterns of Pakistan

Telenor Research

Among the major mobile operators in the world
Approaching 200 million mobile subscriptions

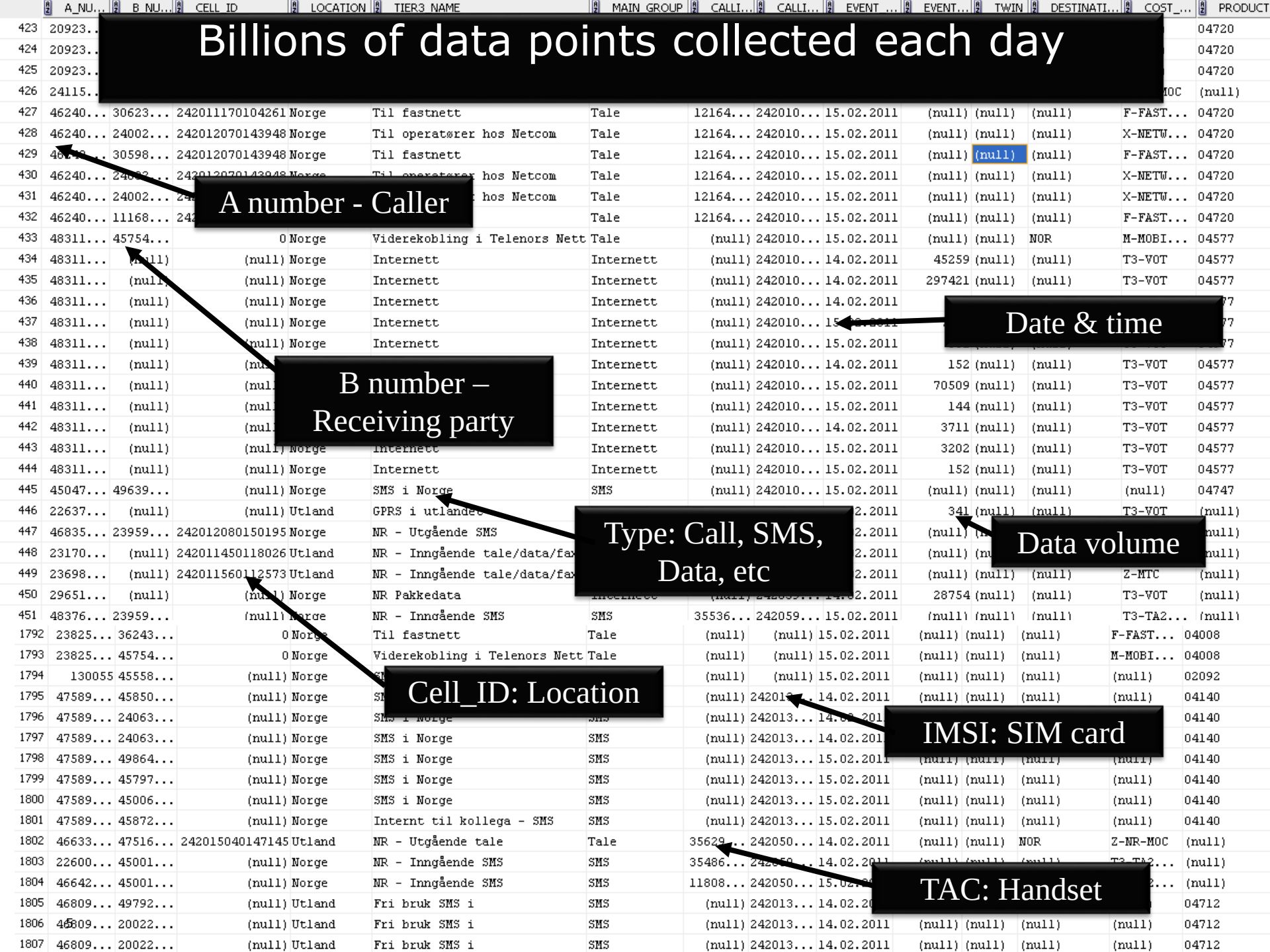
33 000 employees
Present in markets with 1.6 billion people



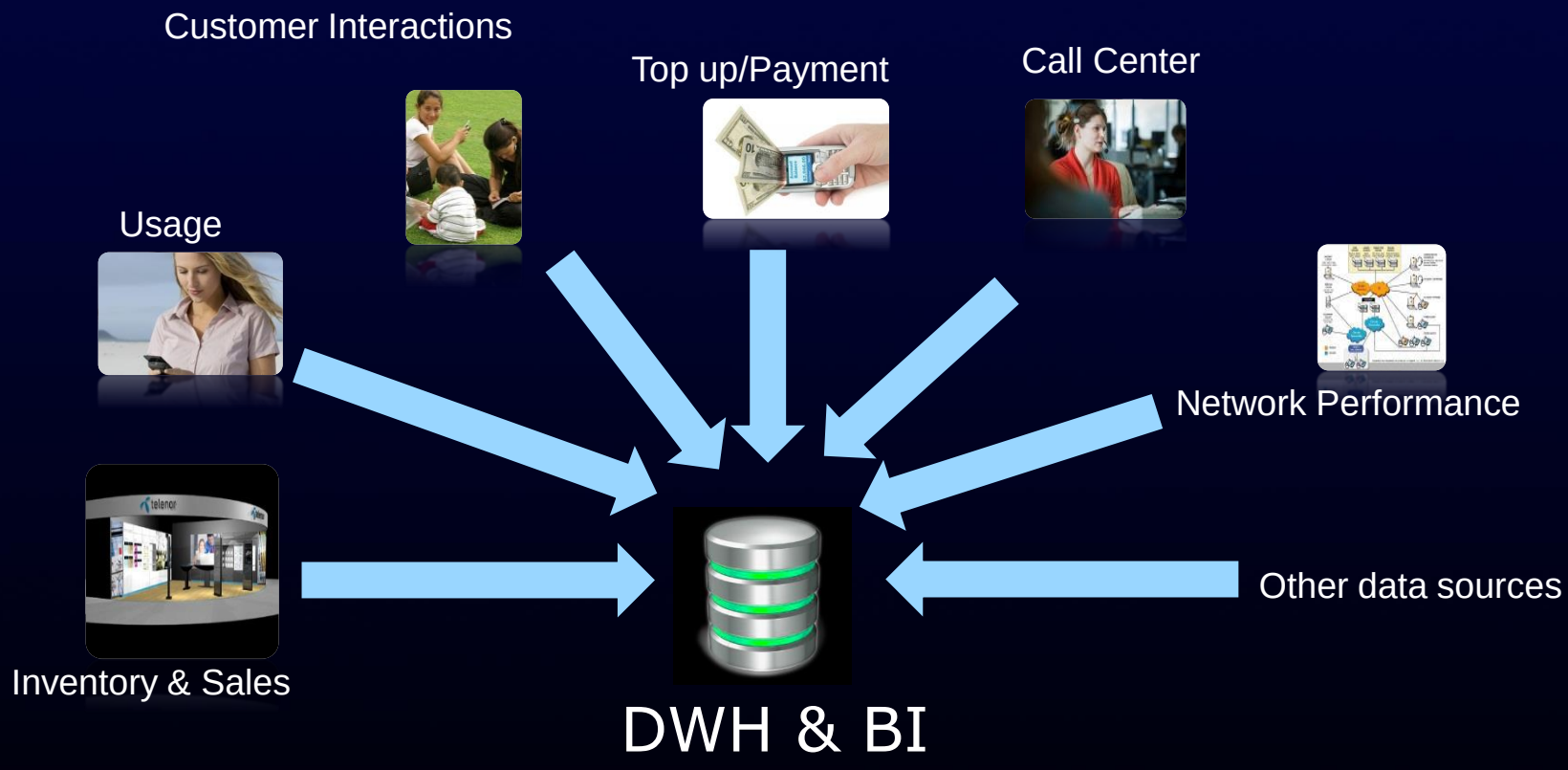
The Big Data Analytics team at Telenor Group Research work across all markets

- A team of 9 Data scientists
- Collaboration partners at leading academic research institutions
- Bridge between academic research and all business units
- Explore and develop new ways to utilize customer data across markets


Our customers generate an increasing amount of information in our systems
What's in it for Telenor?


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

Each BU generates huge amounts of data

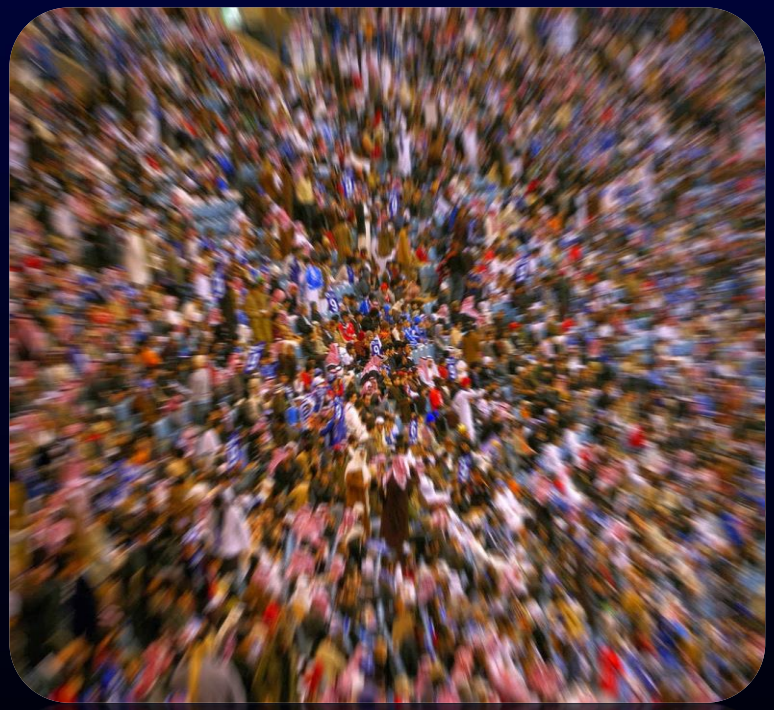


The data enables us to make better decisions across markets

Data Science: Learning from data

Study mobile phone data can give us new insight into human sociology

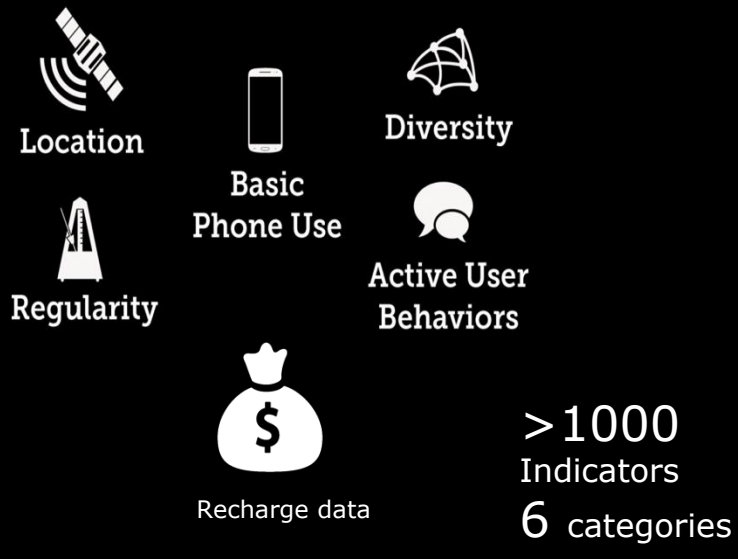
Telenor Research collaborates with MiT, Harvard and NorthEastern University, Flowminder ++



We us a large database with de-identified data for research

Testing advanced behavioral indicators

- Or partners at MiT have successfully predicted mobile phone users' personality class ('Big 5 personal traits') using CDR data
- Together with MiT we explore the use of advanced derived variables to predict telco relevant behaviour
- Example variables
  - Percentage of initiated contacts
  - Variations in response time
  - Entropy in contacts
  - Entropy of visitited places



>1000 Indicators

6 categories

Gender prediction

Support vector machine (SVM) and Random Forest (RF) are the best candidates
Sampling (test+train): South Asia 50K, Europe 500K

Total Accuracy	Europe	South Asia
Linear SVM	72.2%	73.1%
Kernel SVM	73.5%	74.1%
RF	73.1%	74.8%

The top 20% segment gives around **85-90%** accuracy

Top 6 features (SVM)

1. Interactions_per_contact__text__median__sem
2. Number_of_interactions__call__mean
3. Interactions_per_contact__text__median__mean
4. Duration_of_calls__call__mean__mean
5. Interactions_per_contact__call__std__mean
6. Number_of_interactions__call__sem

1. Duration_of_calls__call__mean__sem
2. Interactions_per_contact__call__mean__mean
3. Entropy_of_contacts__call__mean
4. Duration_of_calls__call__median__mean
5. Interactions_per_contact__call__std__mean
6. Number_of_interactions__call__mean

Note: Radius of gyration among top predictors in RF ranking

A cross country study of gender prediction
E.Jahani, P. Sundsøy, Y.Montjoye, A.Pentland, J.Bjelland
Netmob 2015, MIT

”The greatest value of a picture is when it forces us to notice what we never expected to see.”

John Tukey, American mathematician

Data-Driven Development

Using data for social good

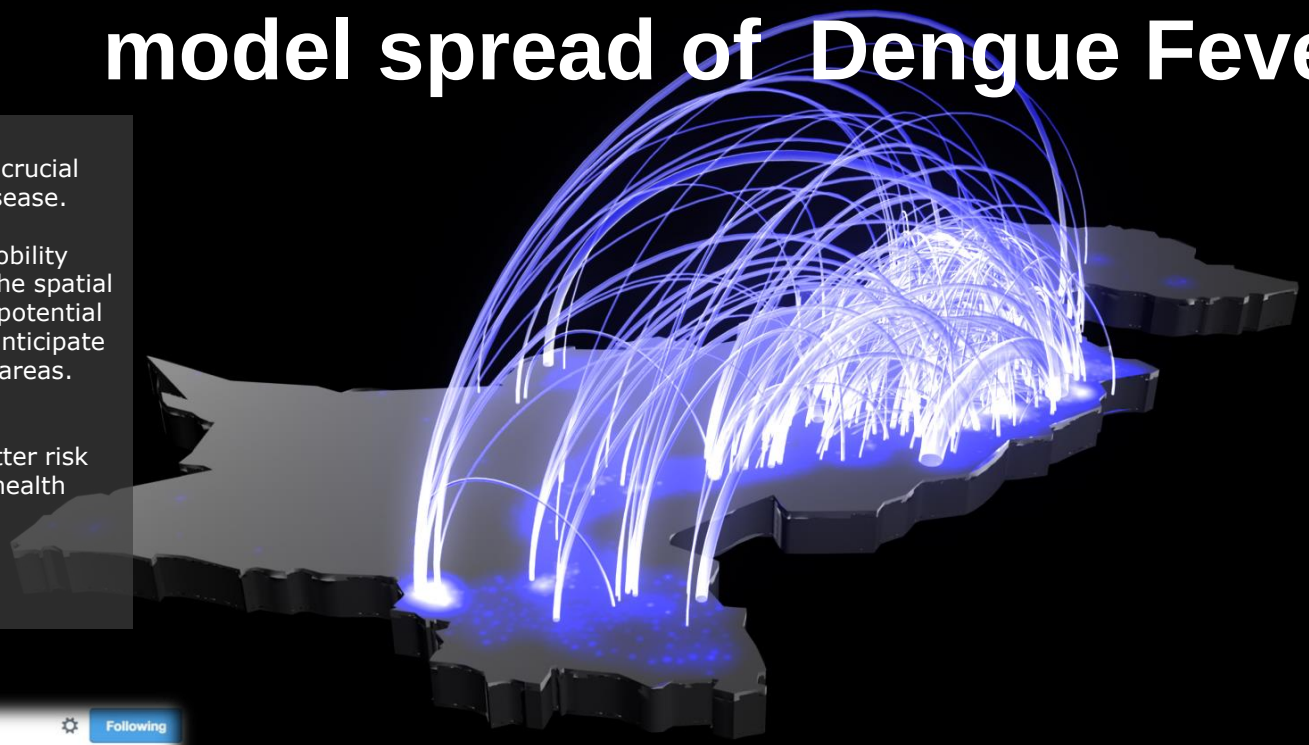
- **Disaster mobility and behavior**
  - Collaboration between Grameenphone, ICCCAD, Flowminder & Telenor Group (Research + Corporate Responsibility)
  - Assessing mobility patterns and changes in economic behavior during the Cyclone Mahasen (May 2013).
- **Improve models for Infectious disease spread**
  - Use mobile phone data to understand the spread of Dengue fever in Pakistan
  - Collaboration with epidemiologists from Harvard T.H. Chan School of Public Health
- **Poverty Prediction**
  - Using telecom data in assessing poverty (Flowminder, MIT)
- **Privacy**
  - Collaboration with UN Global Pulse

Heartbeat of Bangladesh

Assessing mobility patterns and changes in economic behavior during the Cyclone Mahasen

Using power of mobile data to model spread of Dengue Fever

- Human mobility plays a crucial role in the spread of the disease.
- Incorporating human mobility data from CDRs improves the spatial and temporal prediction of potential outbreaks, and it also will anticipate potential outbreaks in new areas.
- The method can be operationalized through better risk maps for government and health practitioners



Privacy and mobile data for the dengue project

- We follow the guidelines in the Telenor Privacy Toolbox developed by the Group Privacy Officer
- For Pakistan there were additional considerations, coming from national considerations:
  - All involved under Non-Disclosure Agreements
  - Personal information (CDRs) processed within the data warehouse and on location in Islamabad, Pakistan, and no personal information was exported
  - The results of the processing resulted in anonymous summations/aggregations
- Only the aggregate data was used in the study and the epidemic modeling of dengue

Article published in PNAS (2015): Impact of human mobility on the emergence of dengue epidemics in Pakistan, A.Wesolowski, T. Qureshi, M.Bonid, P. Sundsøy, M.Johansson, S.Rasheed, K. Engø-Monsen, Caroline O. Buckee

Introducing mobile phone data in **Poverty prediction**

Current Research: Improving Accuracy, Resolution and Regularity

- CDRs can produce rapidly updateable and spatially detailed metrics on consumption, social network structure and mobility – all shown to be related to poverty
- CDRs can improve resolution in urban areas where satellite only-based approaches lack detail

Spatial resolution: Phone data has only be used for coarse mapping (large admin areas)

Technology comparison: Not evaluated what mobile data adds vs remote sensing/other spatial layers. E.g. urban areas problematic using remote sensing

Metric comparison: Different economic welfare measures have not been included (income vs wealth vs consumption in mapping)

Introducing mobile phone data in **Poverty prediction**

Survey data
- Telco surveys
- 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
- Basic phone usage
- Advanced phone usage
- Social Network
- Mobility
- Top-up
- Revenue
- Handset

PREDICTION

- **# poor per km²**
- **Prediction maps**

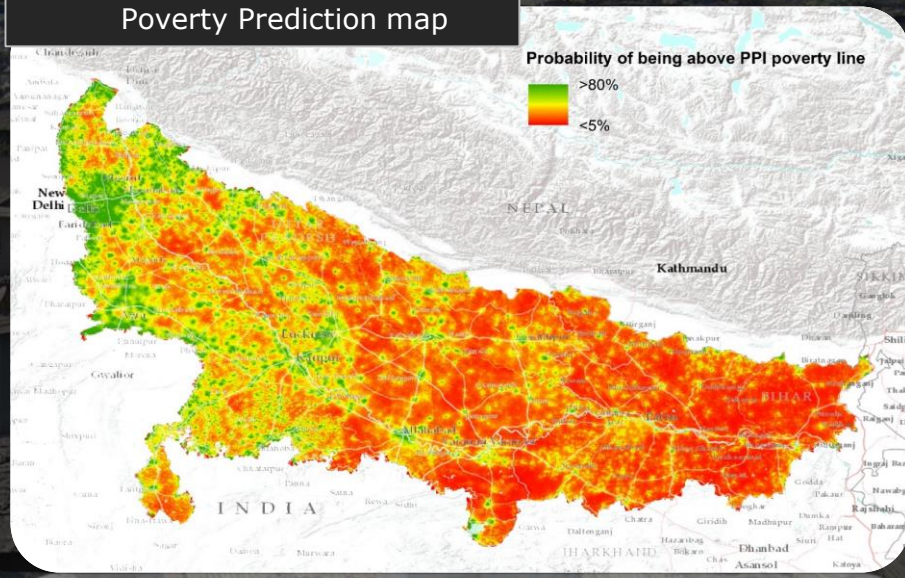
Introducing mobile phone data in **Poverty prediction**

Methods

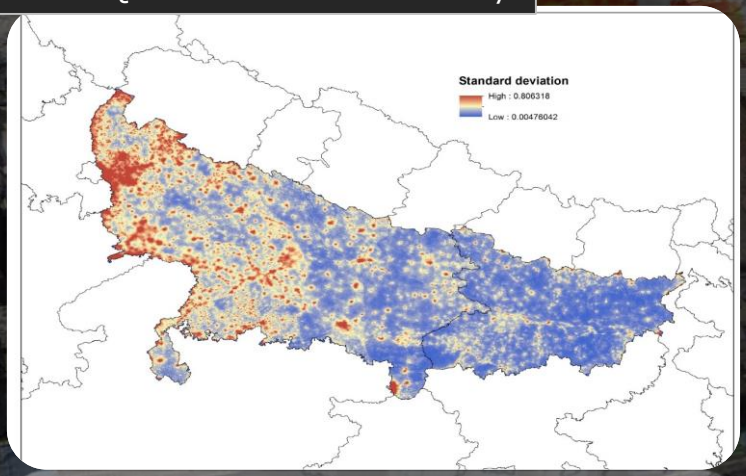
1. **Spatial prediction**
   - Bayesian geostatistical modelling
   - Prediction maps

2. **Individual classification using machine learning methods**
   - RF
   - GBM
   - SVM
   - Deep learning

Poverty Prediction map



Quantification of uncertainty



worldpop

FLOWMINDER.ORG

Partnerships and Principles on Big data for social good

BIG impact & shared value is our goal

Our **DATA** & **EXPERTISE** is our contribution

Handling data **RESPONSIBLY** is key

COLLABORATION is the name of the game

Kenth Engø-Monsen kenth.engo-monsen@telenor.com
Geoffrey Canright Geoffrey.canright@telenor.com
Pål Sundsøy pal-roe.sundsoy@telenor.com