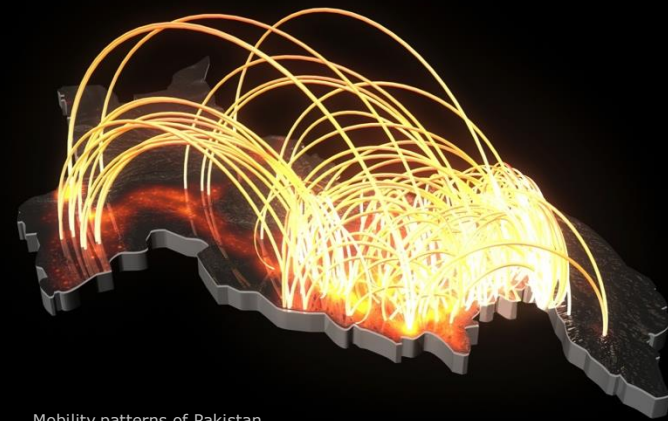
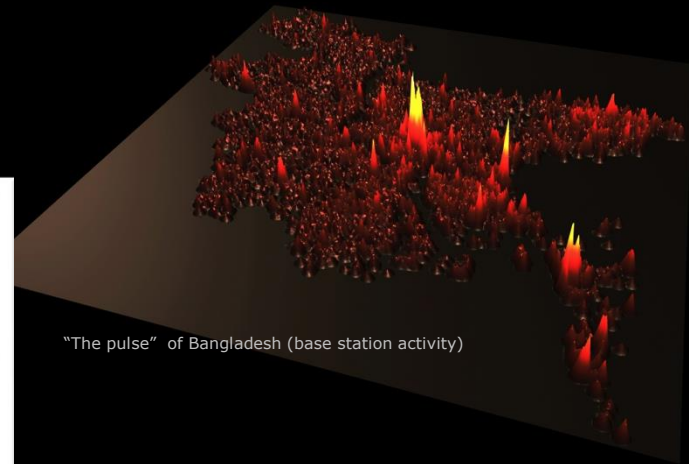
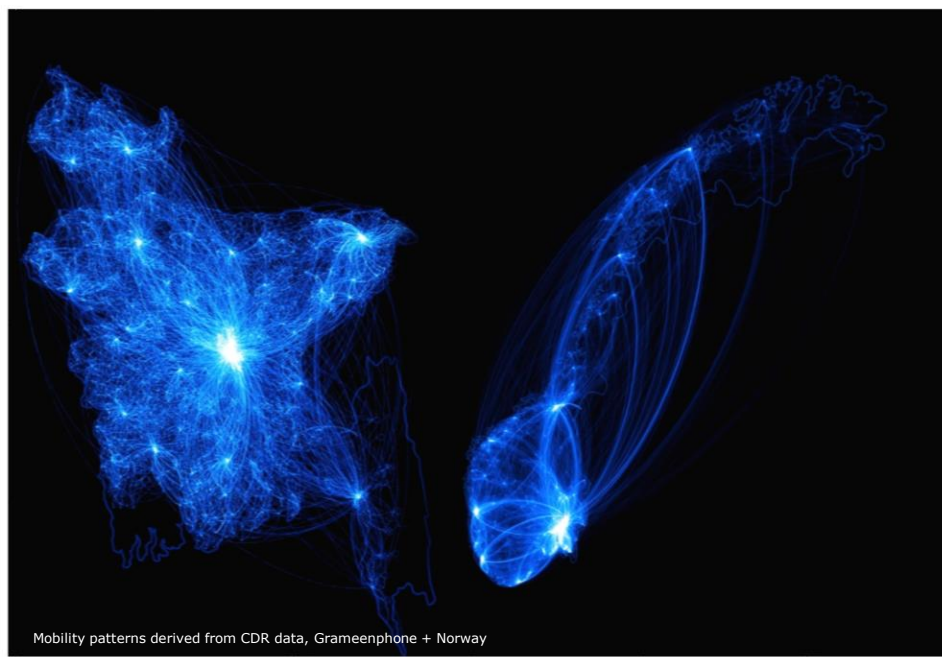


From Viral product spreading to Poverty prediction

Data Science from a telecom perspective

Pål Sundsøy, Telenor Group Research

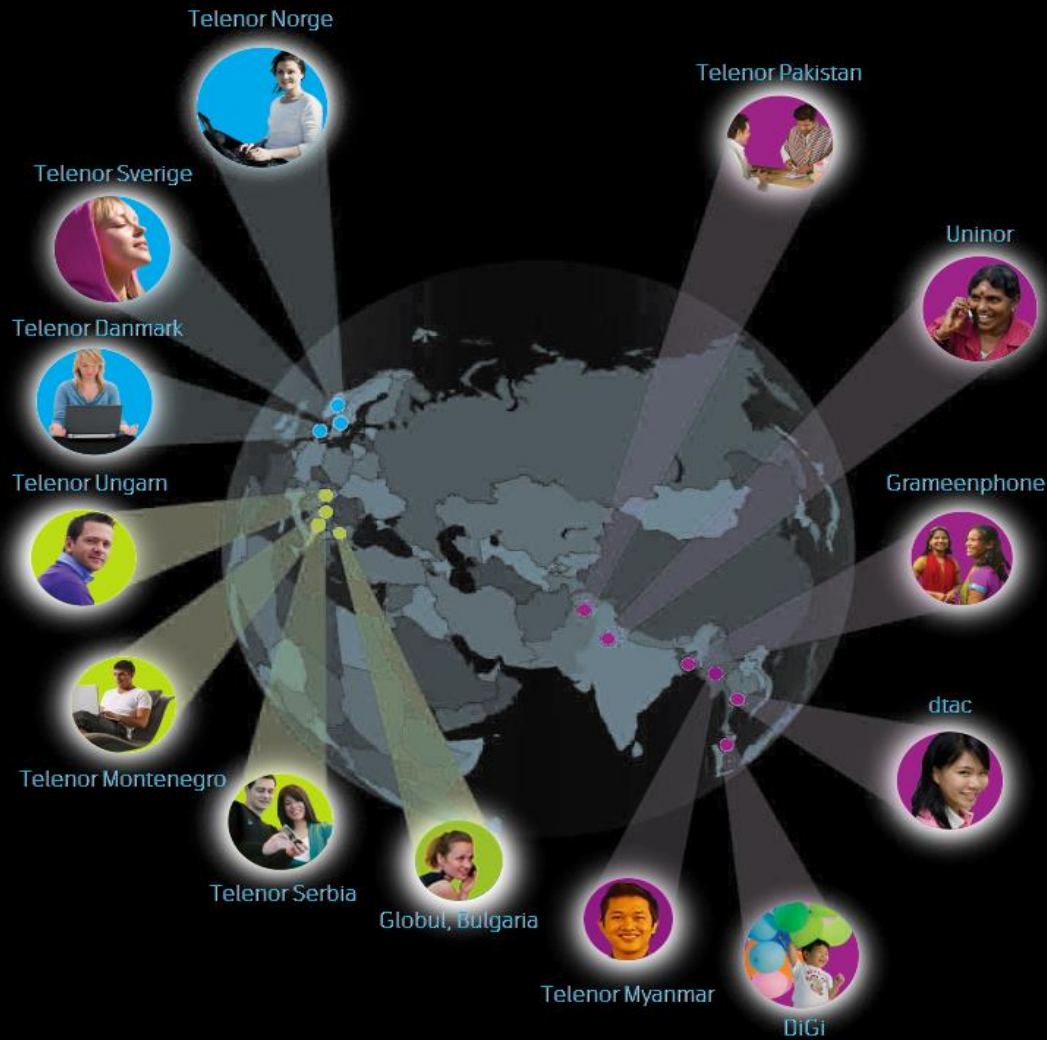
Oct 22nd 2015 International Conference on
Big Data for Official Statistics, Abu, Dhabi



**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

Type: Call, SMS,
Data, etc

Cell_ID: Location

Date & time

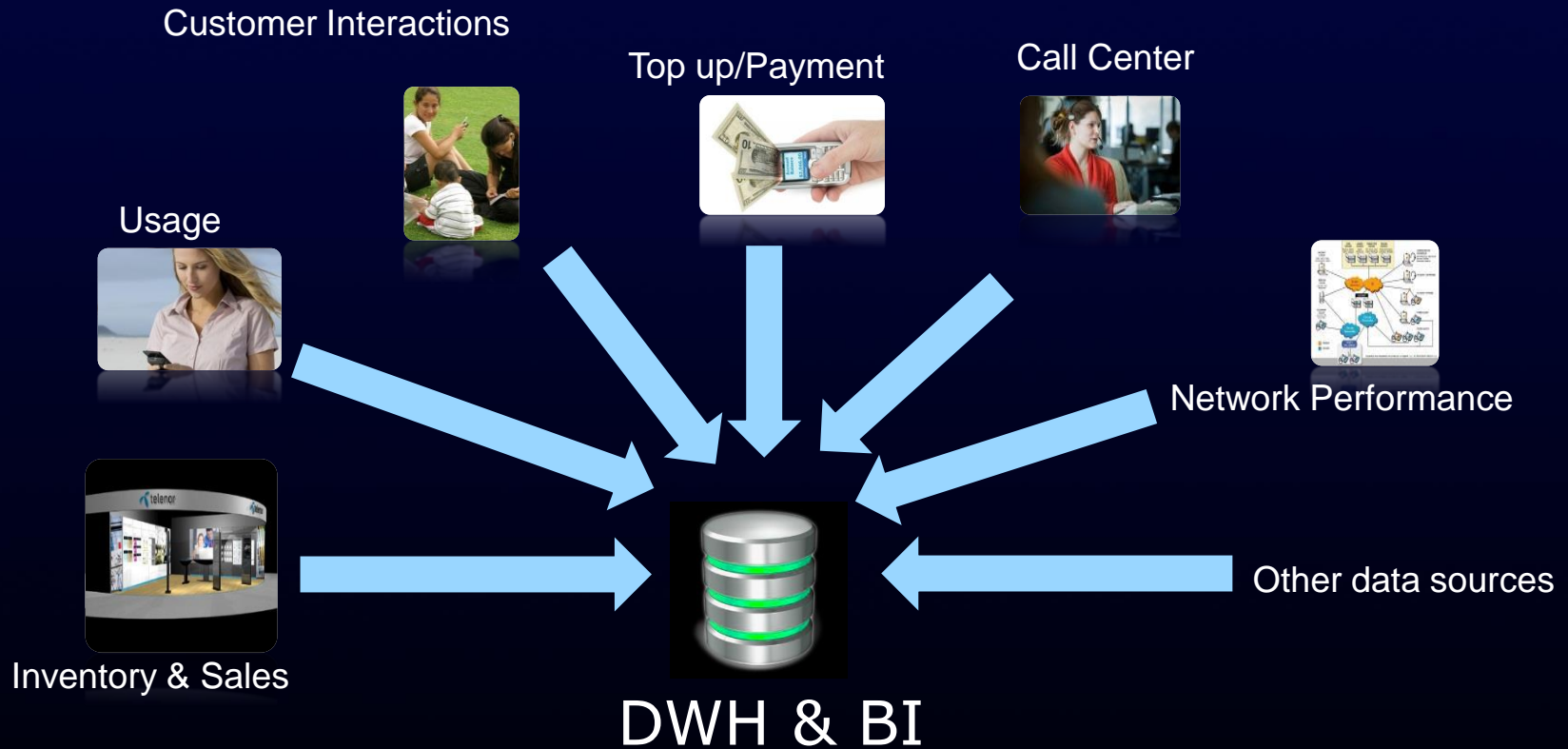
Data volume

IMSI: SIM card

TAC: Handset

A NU...	B NU...	CELL ID	LOCATION	TIER3 NAME	MAIN GROUP	CALLI...	EVENT...	EVENT...	TWIN	DESTINATI...	COST...	PRODUCT
423	20923...											04720
424	20923...											04720
425	20923...											04720
426	24115...											04720
427	46240...	30623...	242011170104261	Norge	Til fastnett	Tale	12164...	242010...	15.02.2011	(null) (null) (null)	F-FAST...	04720
428	46240...	24002...	242012070143948	Norge	Til operatører hos Netcom	Tale	12164...	242010...	15.02.2011	(null) (null) (null)	X-NETW...	04720
429	46240...	30598...	242012070143948	Norge	Til fastnett	Tale	12164...	242010...	15.02.2011	(null) (null) (null)	F-FAST...	04720
430	46240...	24002...	242012070143948	Norge	Til operatører hos Netcom	Tale	12164...	242010...	15.02.2011	(null) (null) (null)	X-NETW...	04720
431	46240...	24002...	242012070143948	Norge	Til operatører hos Netcom	Tale	12164...	242010...	15.02.2011	(null) (null) (null)	X-NETW...	04720
432	46240...	11168...	242012070143948	Norge	Til operatører hos Netcom	Tale	12164...	242010...	15.02.2011	(null) (null) (null)	F-FAST...	04720
433	48311...	45754...		Norge	Viderekobling i Telenors Nett	Tale	(null)	242010...	15.02.2011	(null) (null) NOR	M-MOBI...	04577
434	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	14.02.2011	45259 (null) (null)	T3-VOT	04577
435	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	14.02.2011	297421 (null) (null)	T3-VOT	04577
436	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	14.02.2011			04577
437	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	15.02.2011			04577
438	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	15.02.2011			04577
439	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	14.02.2011	152 (null) (null)	T3-VOT	04577
440	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	15.02.2011	70509 (null) (null)	T3-VOT	04577
441	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	15.02.2011	144 (null) (null)	T3-VOT	04577
442	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	14.02.2011	3711 (null) (null)	T3-VOT	04577
443	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	15.02.2011	3202 (null) (null)	T3-VOT	04577
444	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	15.02.2011	152 (null) (null)	T3-VOT	04577
445	45047...	49639...	(null)	Norge	SMS i Norge	SMS	(null)	242010...	15.02.2011	(null) (null) (null)	(null)	04747
446	22637...	(null)	(null)	Utland	GPRS i utlandet	GPRS	(null)	242010...	15.02.2011	341 (null) (null)	T3-VOT	(null)
447	46835...	23959...	242012080150195	Norge	NR - Utgående SMS	SMS	(null)	(null)	15.02.2011	(null) (null) (null)	(null)	(null)
448	23170...	(null)	242011450118026	Utland	NR - Inngående tale/data/fax	Tale	(null)	(null)	15.02.2011	(null) (null) (null)	(null)	(null)
449	23698...	(null)	242011560112573	Utland	NR - Inngående tale/data/fax	Tale	(null)	(null)	15.02.2011	(null) (null) (null)	Z-MTC	(null)
450	29651...	(null)	(null)	Norge	NR Pakkedata	Tale	(null)	(null)	15.02.2011	28754 (null) (null)	T3-VOT	(null)
451	48376...	23959...	(null)	Norge	NR - Inngående SMS	SMS	35536...	242059...	15.02.2011	(null) (null) (null)	T3-TA2...	(null)
1792	23825...	36243...		Norge	Til fastnett	Tale	(null)	(null)	15.02.2011	(null) (null) (null)	F-FAST...	04008
1793	23825...	45754...		Norge	Viderekobling i Telenors Nett	Tale	(null)	(null)	15.02.2011	(null) (null) (null)	M-MOBI...	04008
1794	130055	45558...	(null)	Norge	SMS i Norge	SMS	(null)	(null)	15.02.2011	(null) (null) (null)	(null)	02092
1795	47589...	45850...	(null)	Norge	SMS i Norge	SMS	(null)	242013...	14.02.2011	(null) (null) (null)	(null)	04140
1796	47589...	24063...	(null)	Norge	SMS i Norge	SMS	(null)	242013...	14.02.2011	(null) (null) (null)	(null)	04140
1797	47589...	24063...	(null)	Norge	SMS i Norge	SMS	(null)	242013...	14.02.2011	(null) (null) (null)	(null)	04140
1798	47589...	49864...	(null)	Norge	SMS i Norge	SMS	(null)	242013...	15.02.2011	(null) (null) (null)	(null)	04140
1799	47589...	45797...	(null)	Norge	SMS i Norge	SMS	(null)	242013...	15.02.2011	(null) (null) (null)	(null)	04140
1800	47589...	45006...	(null)	Norge	SMS i Norge	SMS	(null)	242013...	15.02.2011	(null) (null) (null)	(null)	04140
1801	47589...	45872...	(null)	Norge	Internt til kollega - SMS	SMS	(null)	242013...	15.02.2011	(null) (null) (null)	(null)	04140
1802	46633...	47516...	242015040147145	Utland	NR - Utgående tale	Tale	35629...	242050...	14.02.2011	(null) (null) NOR	Z-NR-MOC	(null)
1803	22600...	45001...	(null)	Norge	NR - Inngående SMS	SMS	35486...	242050...	14.02.2011	(null) (null) (null)	T3-TA2...	(null)
1804	46642...	45001...	(null)	Norge	NR - Inngående SMS	SMS	11808...	242050...	15.02.2011	(null) (null) (null)	(null)	(null)
1805	46809...	49792...	(null)	Utland	Fri bruk SMS i	SMS	(null)	242013...	14.02.2011	(null) (null) (null)	(null)	04712
1806	46809...	20022...	(null)	Utland	Fri bruk SMS i	SMS	(null)	242013...	14.02.2011	(null) (null) (null)	(null)	04712
1807	46809...	20022...	(null)	Utland	Fri bruk SMS i	SMS	(null)	242013...	14.02.2011	(null) (null) (null)	(null)	04712

Each BU generates huge amounts of data

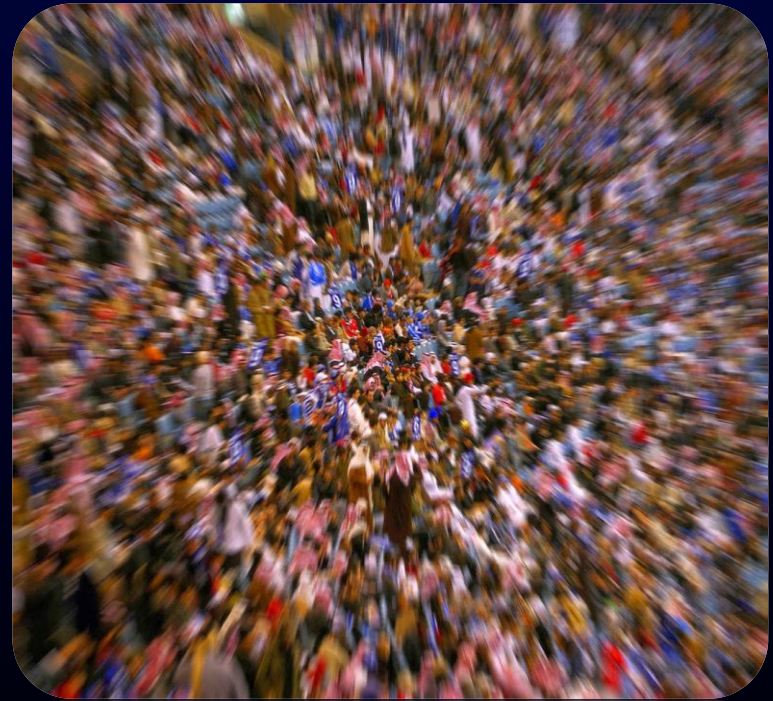


The data enables us to make better decisions across markets

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 use 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 visited places



Location



Basic
Phone Use



Diversity



Regularity



Active User
Behaviors



Recharge data

> 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 | 1. Duration_of_calls_call_mean_sem |
| 2. Number_of_interactions_call_mean | 2. Interactions_per_contact_call_mean_mean |
| 3. Interactions_per_contact_text_median_mean | 3. Entropy_of_contacts_call_mean |
| 4. Duration_of_calls_call_mean_mean | 4. Duration_of_calls_call_median_mean |
| 5. Interactions_per_contact_call_std_mean | 5. Interactions_per_contact_call_std_mean |
| 6. Number_of_interactions_call_sem | 6. Number_of_interactions_call_mean |

Note: Radius of gyration among top predictors in RF ranking

"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

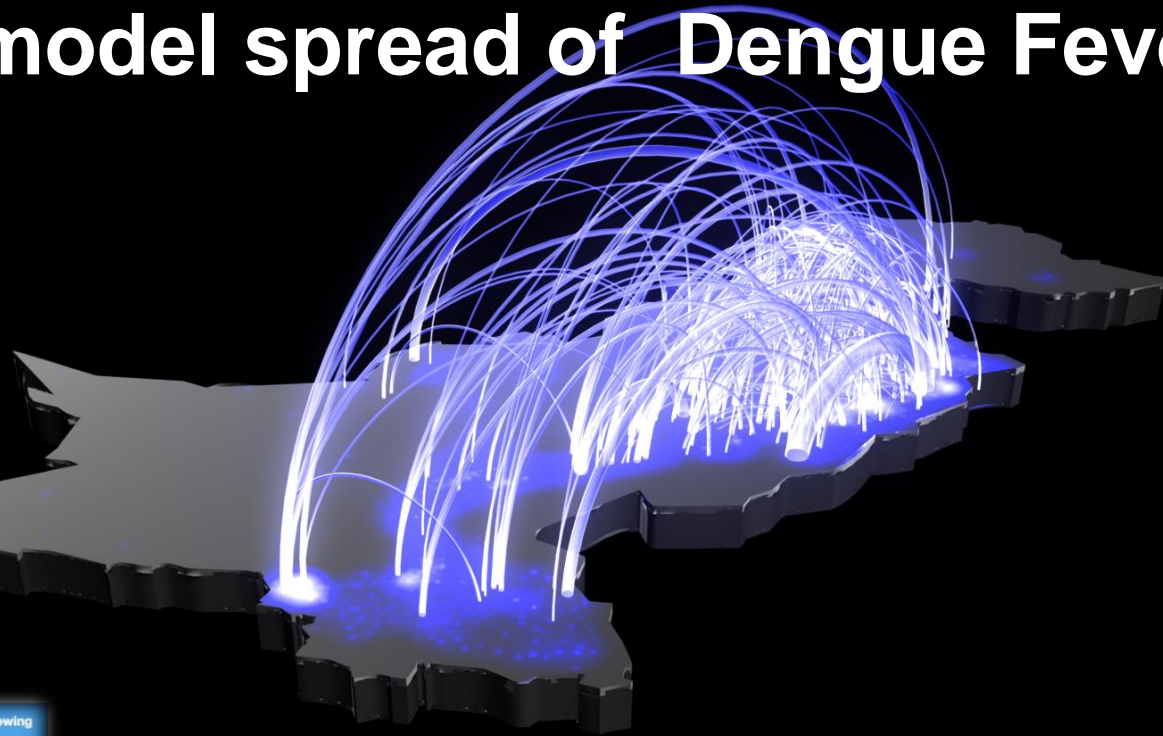
Heartbeat of Bangladesh





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



Using power of mobile data to model spread of Dengue Fever

- o Human mobility plays a crucial role in the spread of the disease.
- o 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.
- o The method can be operationalized through better risk maps for government and health practitioners



 **Bill Gates**   

Cellphone records could help doctors predict which places might be hit by dengue: [b-gat.es/1Obnehb](https://twitter.com/b-gat.es/1Obnehb)



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

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

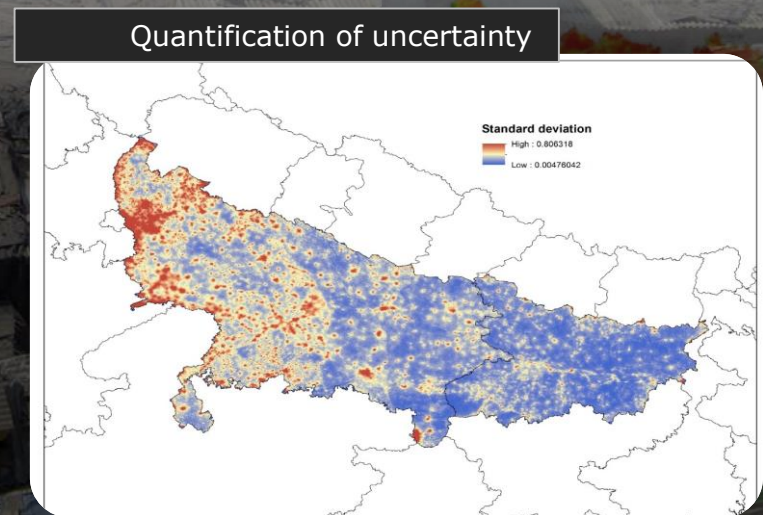
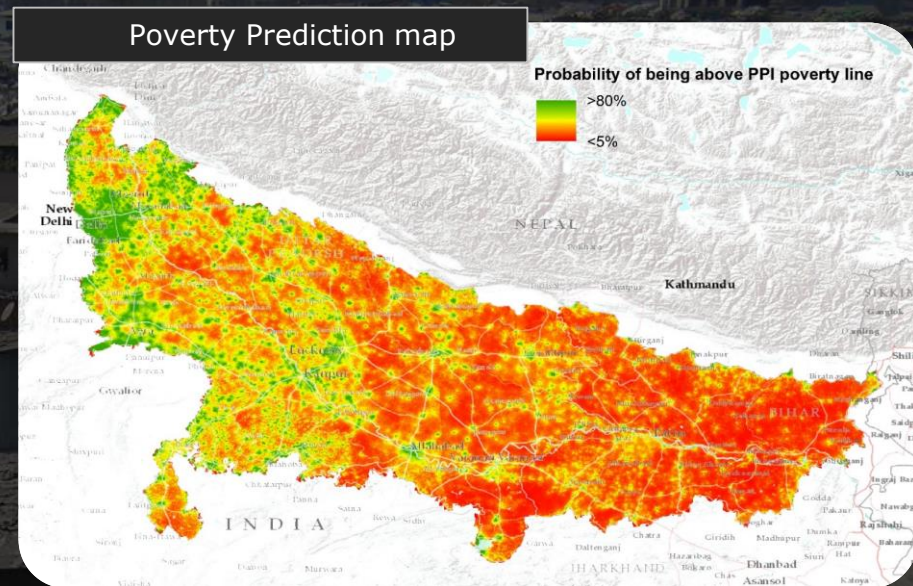
Methods

1. Spatial prediction

- Bayesian geostatistical modelling
- Prediction maps

2. Individual classification using machine learning methods

- RF
- GBM
- SVM
- Deep learning



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
Geoffrey Canright
Pål Sundsøy

kenth.engo-monsen@telenor.com
Geoffrey.canright@telenor.com
pal-roe.sundsoy@telenor.com