

# **Tapping on unconventional data sources to obtain actionable intelligence on the connections between gender and environment**

**Rajius Idzalika - Junior Data Scientist**

## UN Global Pulse

Global Pulse is an innovation initiative of the United Nations Secretary-General on big data and AI. Our vision is a future in which big data is harnessed safely and responsibly as a public good. Our mission is to accelerate discovery, development and scaled adoption of big data innovation for sustainable development and humanitarian action.



Pulse Lab Jakarta combines data science and social research to help make sense of our interconnected, interdependent, and complex world. The Lab is a joint initiative of the United Nations and the Government of Indonesia, via United Nations Global Pulse and the Ministry of National Development and Planning (Bappenas) respectively.

## OUR SERVICES



Drive **exploratory research** on new insights that can be gleaned from unconventional data sources



Help UN agencies, governments and development partners make **better use of their data**

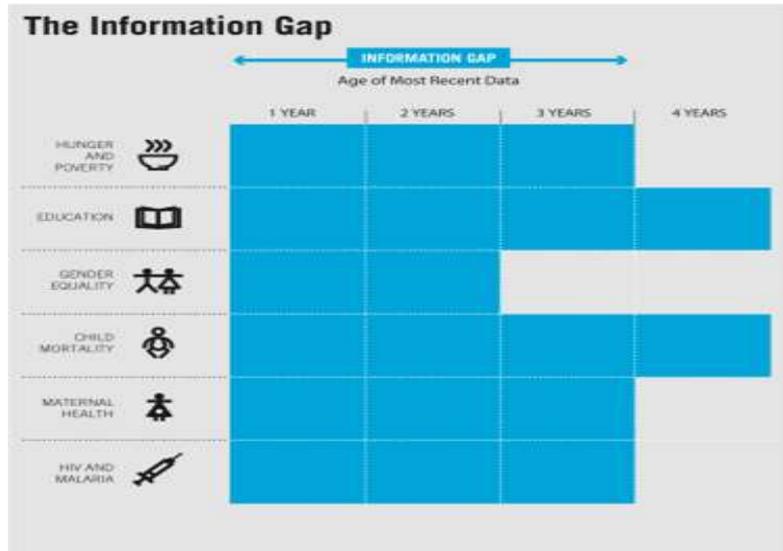


Advocate for the **ethical use of data** and technological platforms in line with the protection of individual privacy

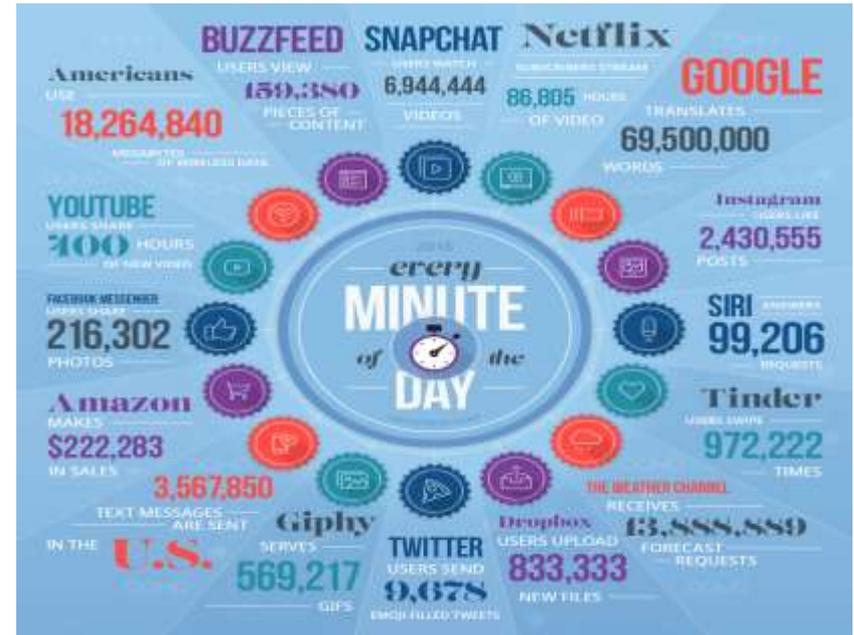
# How do you decide to cross the street? Considering the current traffic or the traffic an hour ago?



# There is an information gap between conventional data source and decision making.



Household-level data is challenging to collect on a real-time basis, making development progress difficult to track. (Source: Millennium Development Goals Report, 2011)



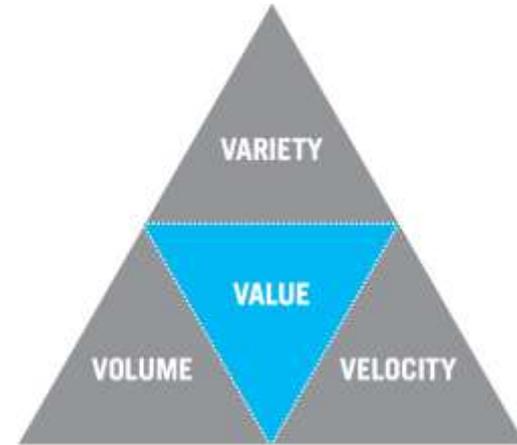
Sumber: <https://www.domo.com/learn/data-never-sleeps-4-0>

# Big data is a new data source

*The basic idea behind the phrase **Big Data** is that everything we do is increasingly leaving a digital trace (or data), which we (and others) can use and analyse*

*“**Big Data** therefore refers to our **ability** to **make use** of the ever-increasing **volumes of data.**”*

***BUT...Big data is not intended to replace conventional data, instead they complement each other to generate richer insights.***





## ...What people have said

- Social media (content focused)
- Online ads
- Community complaints management system
- Radio



## ...What people have done

- Social media (location focused)
- Utilities information (electricity, clean water, etc.)
- Postal data
- Transportation data
- Keywords search
- Online/offline retail data
- Remote sensing
- Financial service data
- Call Data Record (CDR)

Obtaining big data is not easy, but there are ways to get through it.

Two working strategies:



1. Public private partnership



1. Public generated data (citizen science and crowdsourcing)

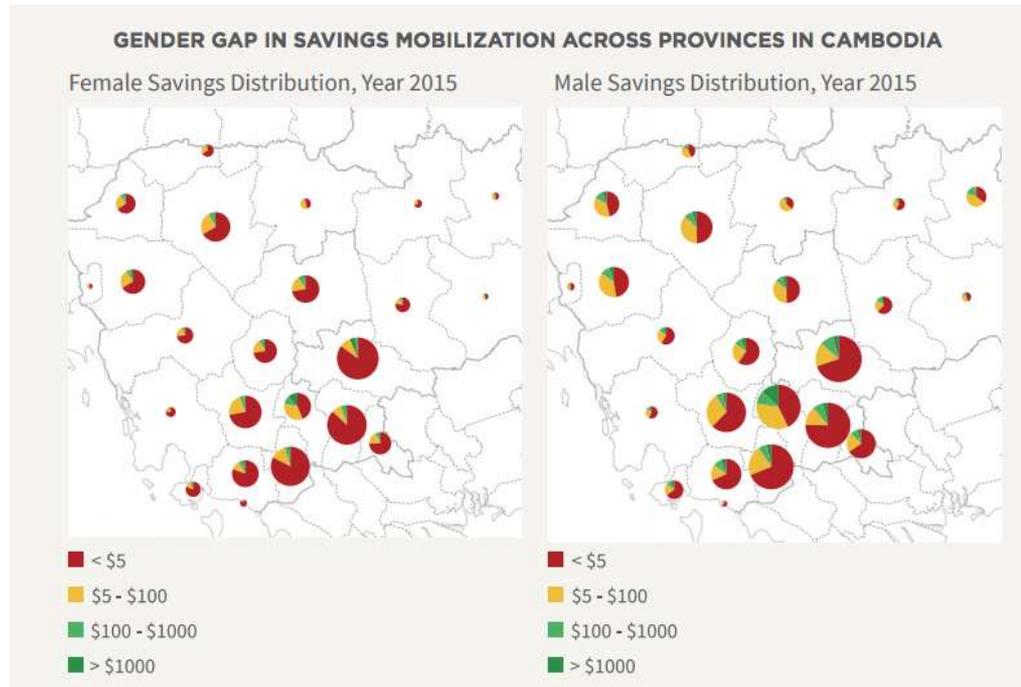
## Examining customers journey at financial institution in Cambodia

UNCDF SHIFT and UN Pulse Lab Jakarta are pleased to launch their new report 'Examining Customer Journeys at Financial Institutions in Cambodia'. This study encourages a shift in focus from examining access to finance to understanding actual usage of financial products. **The study demonstrates the potential of Big Data analytics to generate granular sex- and youth-disaggregated information** on the use of financial services, and to apply insights to inform product development and policy making.

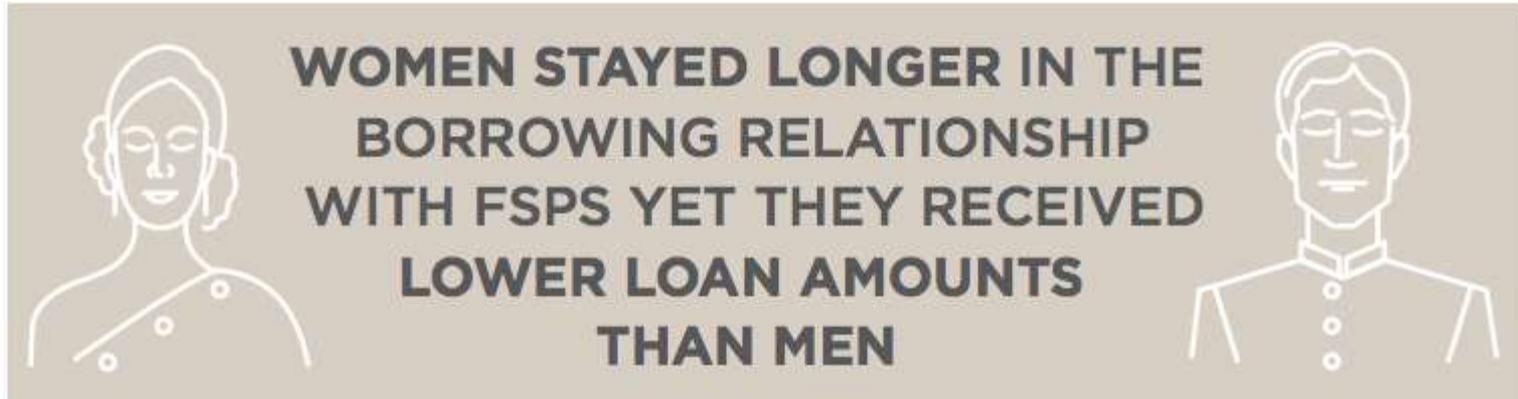
## Finding 1: Different customer profile



## Finding 2: Gender gap on saving mobilization



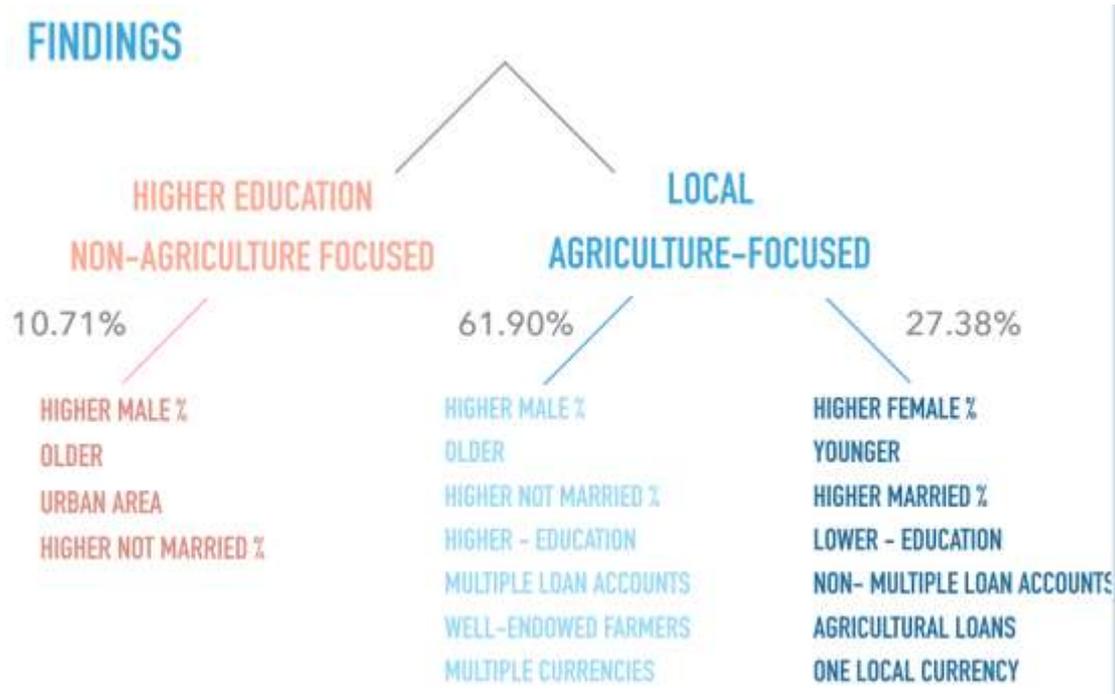
## Finding 3: Gender gap on customer journey



## Works in the pipeline ...

- Measure resilience by adaptive capacity index.
- Find proxy indicators for poverty.
- Understanding the relationship between loan and climate change and deforestation.

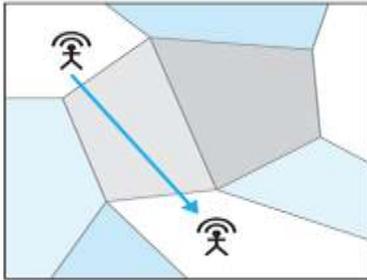
An illustration to repurpose loan data with cluster analysis.



# Data Call Detail Record (CDR)

Call data records could be harnessed to learn human behavior.

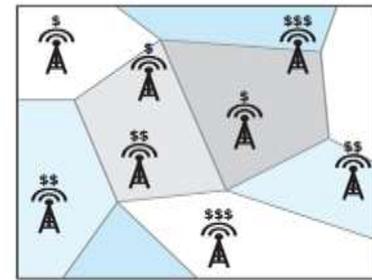
CALLER ID	CALLER CELL TOWER LOCATION	RECIPIENT PHONE NUMBER	RECIPIENT CELL TOWER LOCATION	CALL TIME	CALL DURATION
X76VG588RLPQ	2°24' 22.14", 35°49' 56.54"	A81UTC93KK52	3°26' 30.47", 31°12' 18.01"	2013-11-07T15:15:00	01:12:02



Mobility



Social interaction

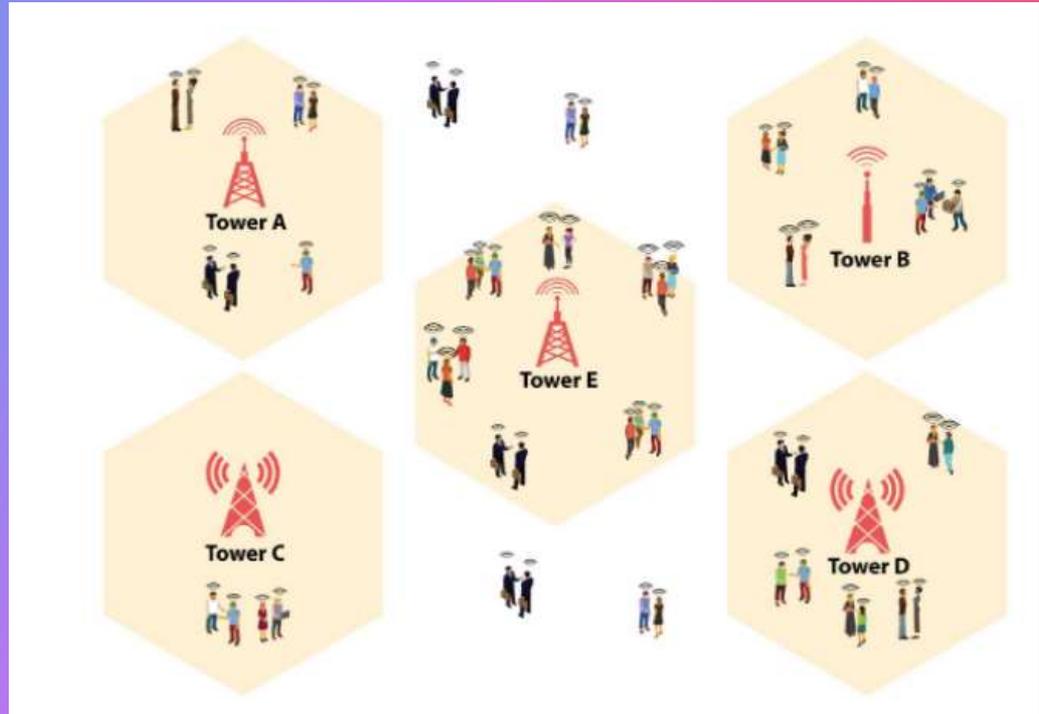


Economic activity

## RURAL TO URBAN MIGRATION

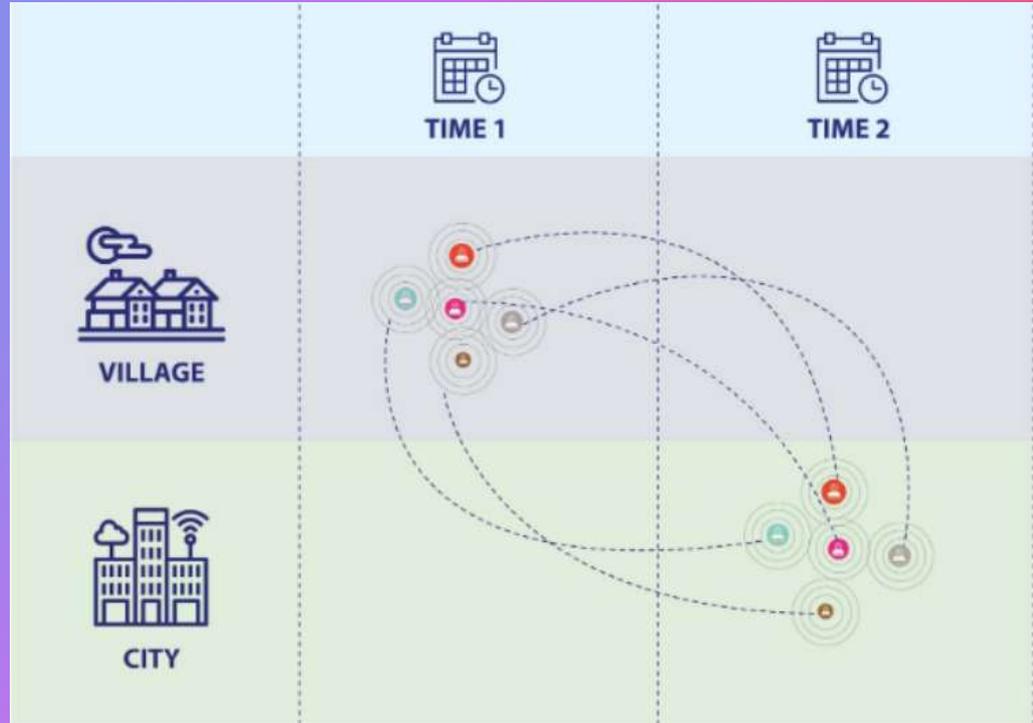
Commissioned by the World Bank, PLJ and Empatika conducted research into the experiences of rural to urban migrants.

PLJ led the quantitative component of the project which used mobile network data to develop statistics on the magnitude of short term migration and the source communities of migrants to seven major cities within Indonesia.

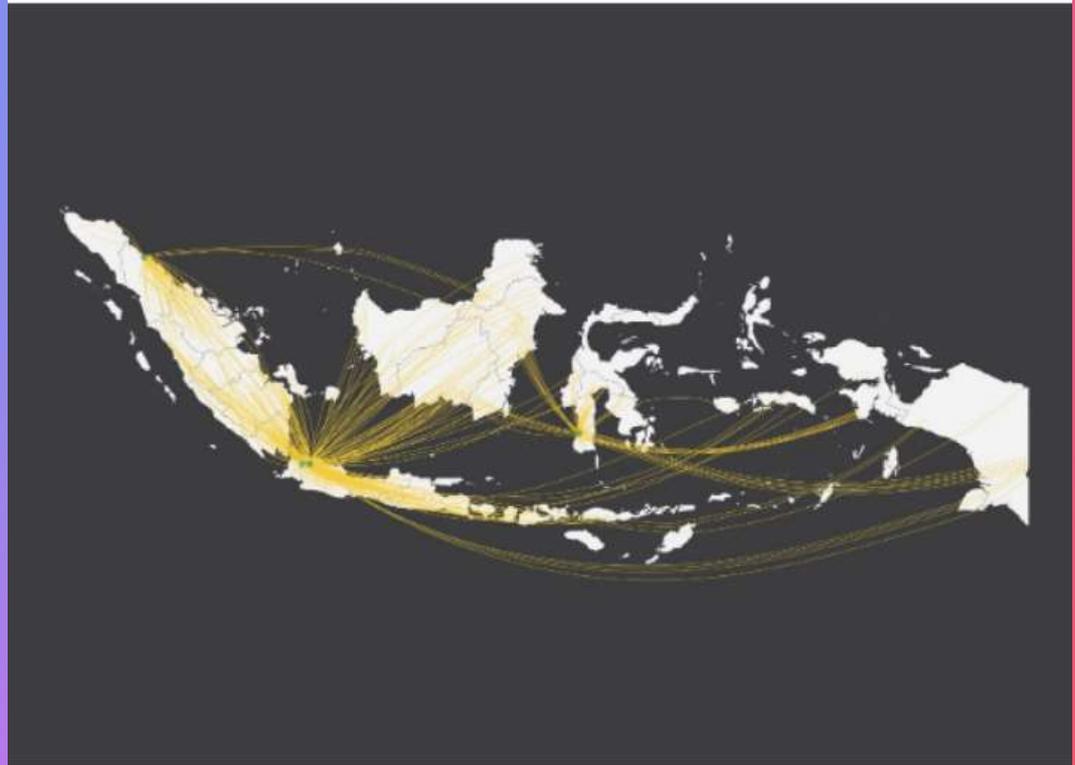


# Data Call Detail Record (CDR)

## RURAL TO URBAN MIGRATION

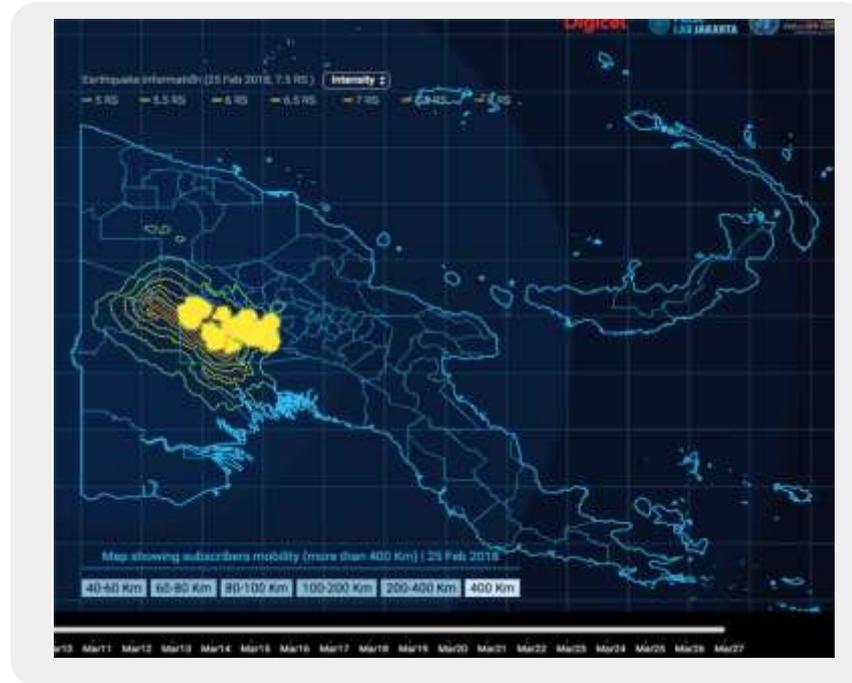


## RURAL TO URBAN MIGRATION



# Data Call Detail Record (CDR)

One output is visualization with high resolution.

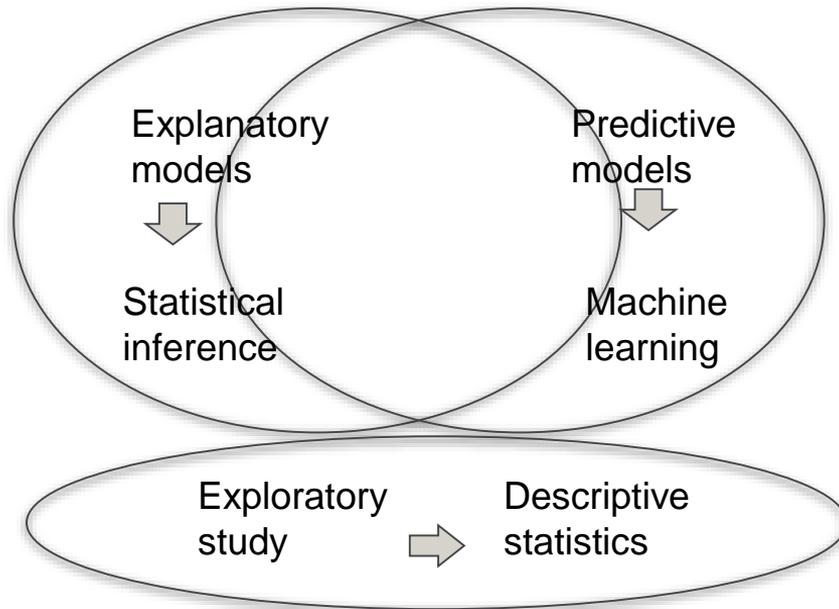


The challenge related to gender statistics is no gender information.

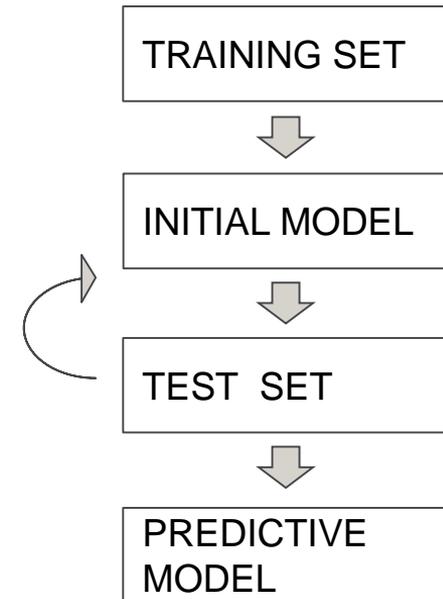
- We need to conduct foundational research to predict gender from the call or text behavior.
- It is known that machine learning is really good for a classification task. The ground truth is determined by conducting a telesurvey.
- Our recent research shows the the accuracy is 0.88 or higher for the prediction of selected household assets.

# Exploratory model versus predictive model

Predictive model is getting more popular for timely decision making.



## MACHINE LEARNING



It is possible to predict gender of mobile user with high accuracy.

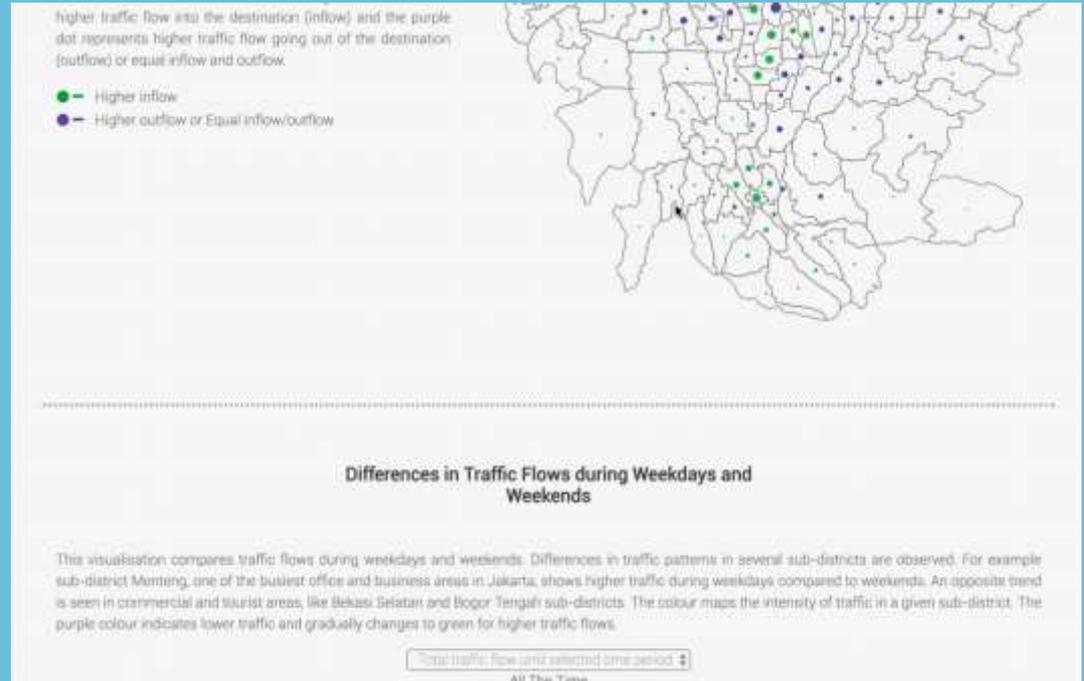
TABLE IV: Household prediction results

Type	Method	None	FCBF	Boruta	XGBoost
Electricity	SVM	0.77	0.58	0.71	0.74
	NB	0.67	0.6	0.71	0.74
	Elastic-Net	0.76	0.6	0.8	0.83
	NN	0.67	0.42	0.73	0.77
	Decision tree	0.85	0.66	0.85	0.88
Bank Account	SVM	0.78	0.59	0.77	0.8
	NB	0.61	0.57	0.68	0.66
	Elastic-Net	0.76	0.55	0.77	0.79
	NN	0.67	0.47	0.69	0.7
	Decision tree	0.89	0.69	0.88	0.9
Television	SVM	0.75	0.62	0.81	0.84
	NB	0.66	0.61	0.73	0.76
	Elastic-Net	0.77	0.6	0.85	0.85
	NN	0.7	0.57	0.8	0.82
	Decision tree	0.87	0.74	0.86	0.88
Refrigerator	SVM	0.77	0.6	0.79	0.78
	NB	0.7	0.61	0.72	0.73
	Elastic-Net	0.77	0.6	0.78	0.77
	NN	0.7	0.51	0.71	0.71
	Decision tree	0.87	0.68	0.89	0.89
Radio	SVM	0.77	0.61	0.78	0.8
	NB	0.55	0.64	0.57	0.58
	Elastic-Net	0.75	0.59	0.76	0.77
	NN	0.62	0.47	0.66	0.65
	Decision tree	0.86	0.7	0.87	0.9

## Inferring Greater Jakarta's Traffic Patterns

Pulse Lab Jakarta, in partnership with Grab, has been investigating how ride-hailing data can be leveraged to better understand Greater Jakarta's traffic flows at a macroscopic level.

This visualization shows traffic patterns (inflows and outflows) in Greater Jakarta.

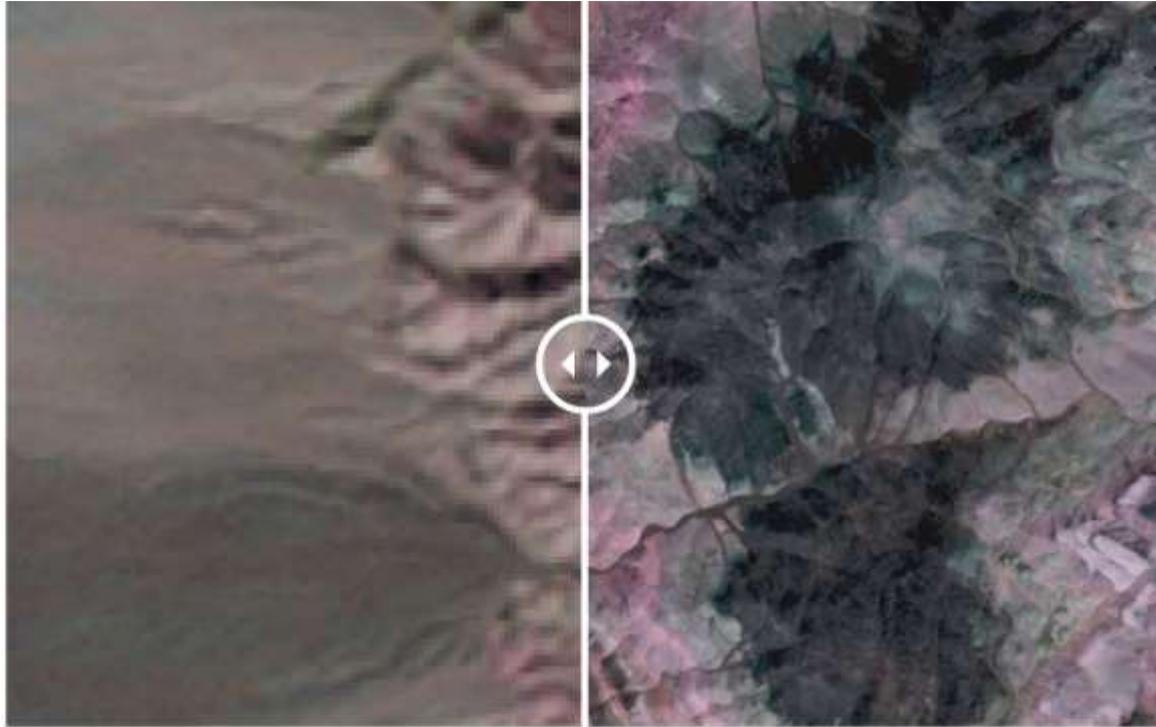


There is gender information that can be used.

- To get the gender (and other) information, the challenge is data partnership. Two modalities: data sharing or insight sharing.
- Homework: build trust and define the shared values.

# Satellite images

DigitalGlobe provides 30 cm resolution imagery.



## Fight climate changes with machine learning and ground truth.

- Better estimates on how much energy we are consuming
- Improve deforestation tracking

Gender disaggregated?

Overlay with other (big) data with disaggregated by gender for real time information.



Harnessing data for  
development.  
Translating insights  
for social innovation.

-  [plj@un.or.id](mailto:plj@un.or.id)
-  [@PulseLabJakarta](https://twitter.com/PulseLabJakarta)
-  [@PulseLabJakarta](https://www.instagram.com/PulseLabJakarta)
-  [@PulseLabJakarta](https://www.facebook.com/PulseLabJakarta)
-  [Pulselabjakarta.org](http://Pulselabjakarta.org)