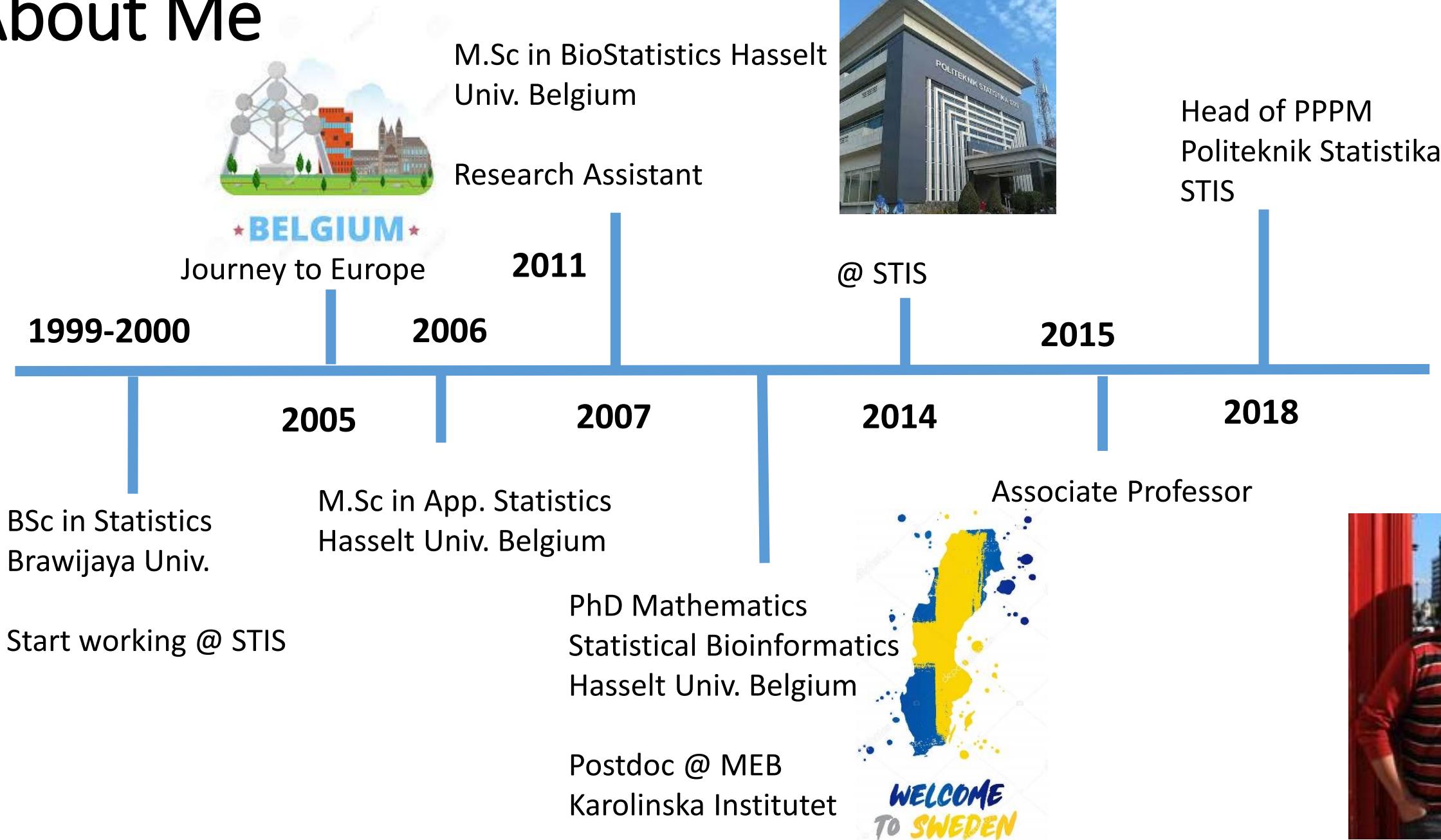




Data Science, the Key for Competing in Data-driven World

Setia Pramana

About Me



Board Member

- UN Global Working Group Big Data for Official Statistics
- Asosiasi Ilmuwan Data Indonesia
- Ikatan Statistisi Indonesia
- Forum Pendidikan Tinggi Statistika
- Masyarakat Biodiversiti dan Bioinformatika Indonesia

“Data! Data! Data!” he cried impatiently. “I can’t make bricks without clay!”
– **Sherlock Holmes**, Conan Doyle’s “The Adventure of the Copper Beaches”

The World is Changing

Big Data



90%

of the data created in the last two years alone.

Mobile



1 billion (plus)
(plus) smart devices shipped in 2013 alone.

Social



81%

of customers depend on social sites for purchasing advice.

Cloud



62%

of total workloads will be in the cloud by 2016.

Internet of Things



50 billion

devices connected to the internet by 2020.

API Economy

Global m-commerce sales were



85 billion

in 2013 and forecast to rise to \$120 billion by 2015 and an estimated \$1 trillion by 2017



We can do everything Online!

YESBASS

traveloka[®]

Qraved

www.BerryKitchen.com
Fresh • Tasty • Delivered

Spotify

OPINI.id

HijUp.com

uangteman

waze
OUTSMARTING TRAFFIC. TOGETHER

Grab

Mobil123.com
PORTAL OTOMOTIF NO. 1

GOJEK

livingsocial Quipper SCHOOL

hlukum online.com

airbnb

danabijak

R

tunaiku
PINJAM CEPAT MUDAH

klikDOKTER[®]
MENUJU INDONESIA SEHAT

LewatMana.com

foodpanda

HappyFresh
JOY DELIVERED!

telunjuk.com
Sahabat Kita Saat Belanja

TunaiKita
RESPONSIBLE CREDIT. GOOD LIVING.

okulaku

UANGTEMAN

pakdok

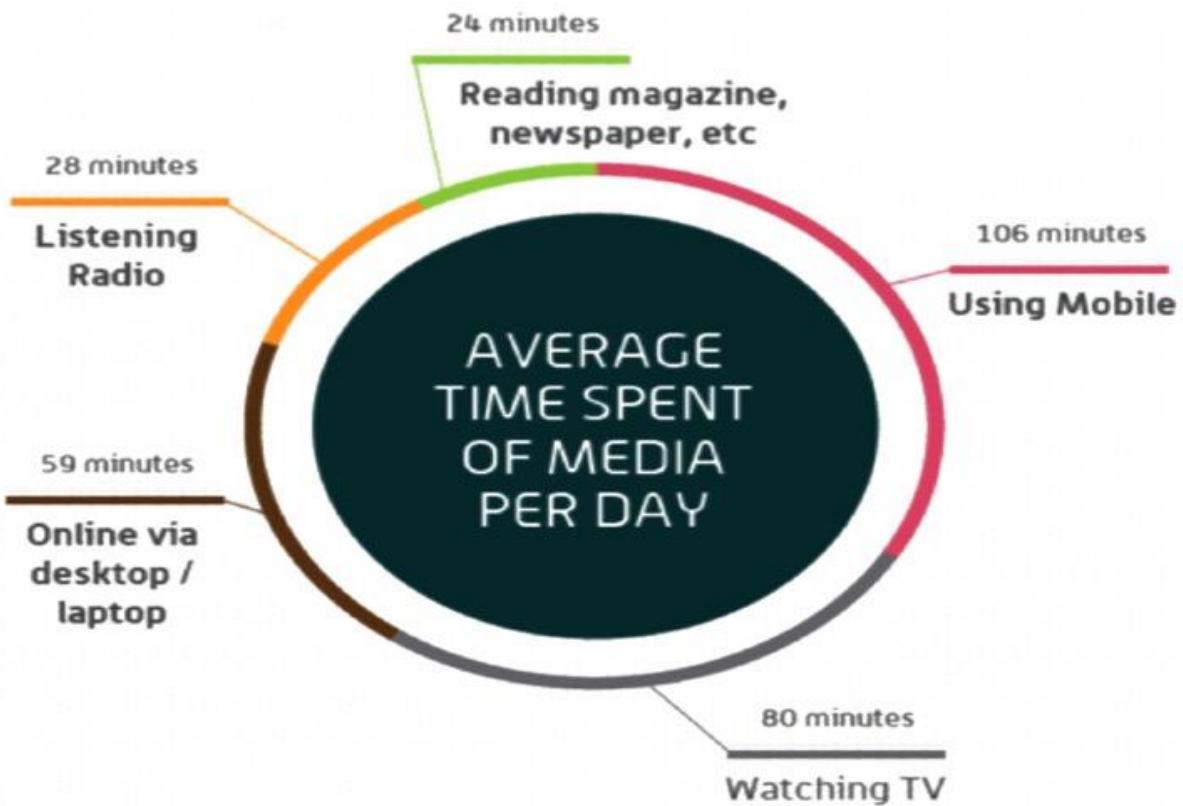
PROSEHAT
Aplikasi Kesehatan Indonesia



JobStreet.com

GUDANG LISTRIK
PIONEER

What we “produce”?



2018 This Is What Happens In An Internet Minute



How about Indonesia?



JAN
2018

TIME SPENT WITH MEDIA

SURVEY-BASED DATA: FIGURES REPRESENT RESPONDENTS' SELF-REPORTED ACTIVITY

AVERAGE DAILY TIME
SPENT USING THE
INTERNET VIA ANY DEVICE



AVERAGE DAILY TIME
SPENT USING SOCIAL
MEDIA VIA ANY DEVICE



AVERAGE DAILY TV VIEWING TIME
(BROADCAST, STREAMING
AND VIDEO ON DEMAND)



AVERAGE DAILY TIME
SPENT LISTENING TO
STREAMING MUSIC



8H 51M

3H 23M

2H 45M

1H 19M

JAN
2018

SMARTPHONE LIFE MANAGEMENT ACTIVITIES

PERCENTAGE OF THE TOTAL POPULATION USING A SMARTPHONE TO PERFORM EACH TASK [SURVEY-BASED]

USE THE ALARM
CLOCK FUNCTION



17%

MANAGE DIARY
OR APPOINTMENTS



6%

CHECK THE
WEATHER



4%

TRACK HEALTH, DIET,
OR ACTIVITY LEVELS



we
are.
social

4%

TAKE PHOTOS
OR VIDEOS



44%

CHECK
THE NEWS



18%

READ E-BOOKS
OR E-MAGAZINES



3%

MANAGE LISTS
(E.G. SHOPPING, TASKS)



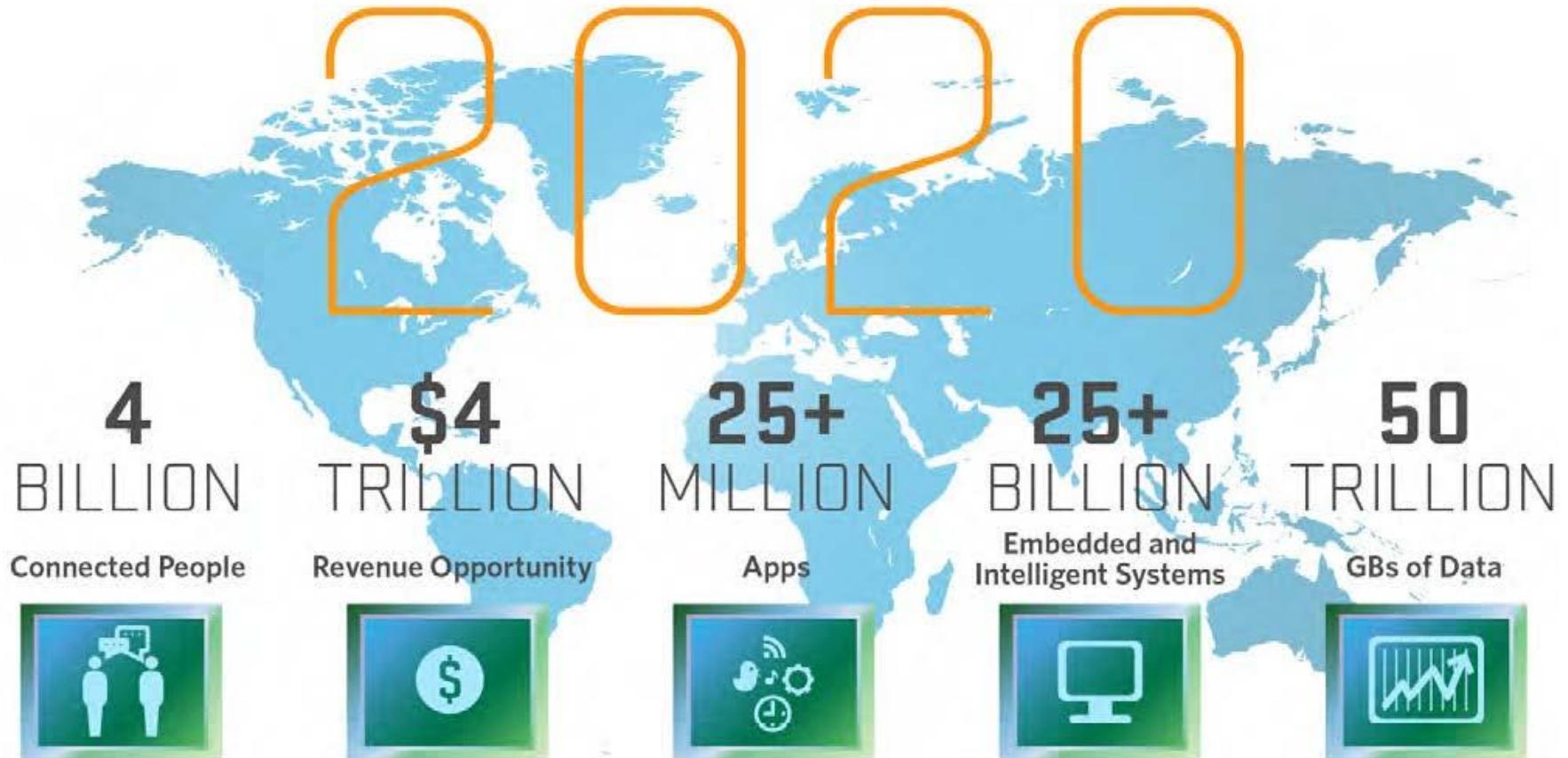
4%

62

SOURCE: GOOGLE CONSUMER BAROMETER, JANUARY 2018. FIGURES BASED ON RESPONSES TO A SURVEY. *NOTES: DATA BASED ON SURVEY RESPONSES FROM ADULT INTERNET USERS ONLY; PLEASE SEE THE NOTES AT THE END OF THIS REPORT FOR MORE INFORMATION ON GOOGLE'S METHODOLOGY AND THEIR AUDIENCE DEFINITIONS. DATA HAS BEEN REBASED TO SHOW TOTAL NATIONAL PENETRATION, REGARDLESS OF AGE.

Hootsuite™ we
are.
social

Data Explosion



Source: Mario Morales, IDC

BIG DATA

with 8 V's

01

01

05



06



VELOCITY

Information gains momentum and crises & opportunities evolve in real time. How is outlook for today?

07



VISCOSITY
Does it stick with you? Does it call for action?

08



VIRALITY
AHA to-go?
Does it convey a message that can be pasted into a presentation or Instagrammed?

04



VISUALISATION

Can you make sense at a glance? Does it trigger a decision?

02



VALUE
Can you find it when you most need it?

03



VERACITY
Are you dealing with information or disinformation?

By 2014, it's anticipated there will be
420 MILLION WEARABLE, WIRELESS HEALTH MONITORS

4 BILLION+
HOURS OF VIDEO
are watched on YouTube each month

400 MILLION TWEETS
are sent per day by about 200 million monthly active users

Poor data quality costs the US economy around
\$3.1 TRILLION A YEAR



IBM.

40 ZETTABYTES

[43 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005

2020

6 BILLION PEOPLE
have cell phones



WORLD POPULATION: 7 BILLION

The New York Stock Exchange captures
1 TB OF TRADE INFORMATION during each trading session



By 2016, it is projected there will be
18.9 BILLION NETWORK CONNECTIONS

– almost 2.5 connections per person on earth



Sources: McKinsey Global Institute, Twitter, Cisco, Gartner



Taxonomy Big Data Sources

- **Exhaust data:** Passively collected data from people's use of digital services such as mobile phones, financial transactions or web searches.
- **Sensing data:** Actively collected data from sensors, e.g. in smart cities or from wearables and also through remote sensing and satellite images.
- **Digital content:** Open web content actively produced by people such as social media interactions, news articles, blogs or job postings. Unlike exhaust and sensing data is digital content intentionally edited by somebody, i.e. subjective or even deceptive, depending on the intentions of the author.

Letouzé (Data-Pop Alliance, 2015)

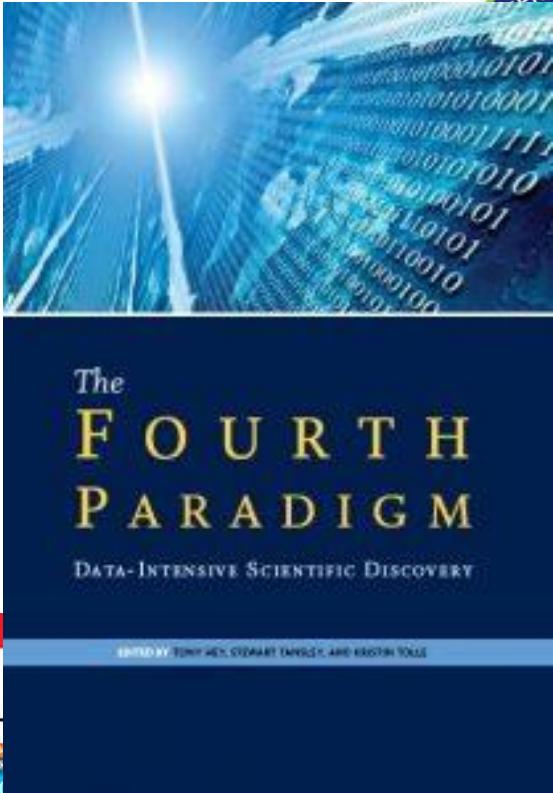
Taxonomy Big Data Sources

Exhaust data	Mobile phone data Financial transactions Online search and access logs Citizen card Postal data
Sensing data	Satellite and UAV imagery Sensors in cities, transport and homes Sensors in nature, agriculture and water Wearable technology Biometric data Internet of Things (IoT)
Digital Content	Social media data Web scraping Participatory sensing / crowdsourcing Health records Radio content

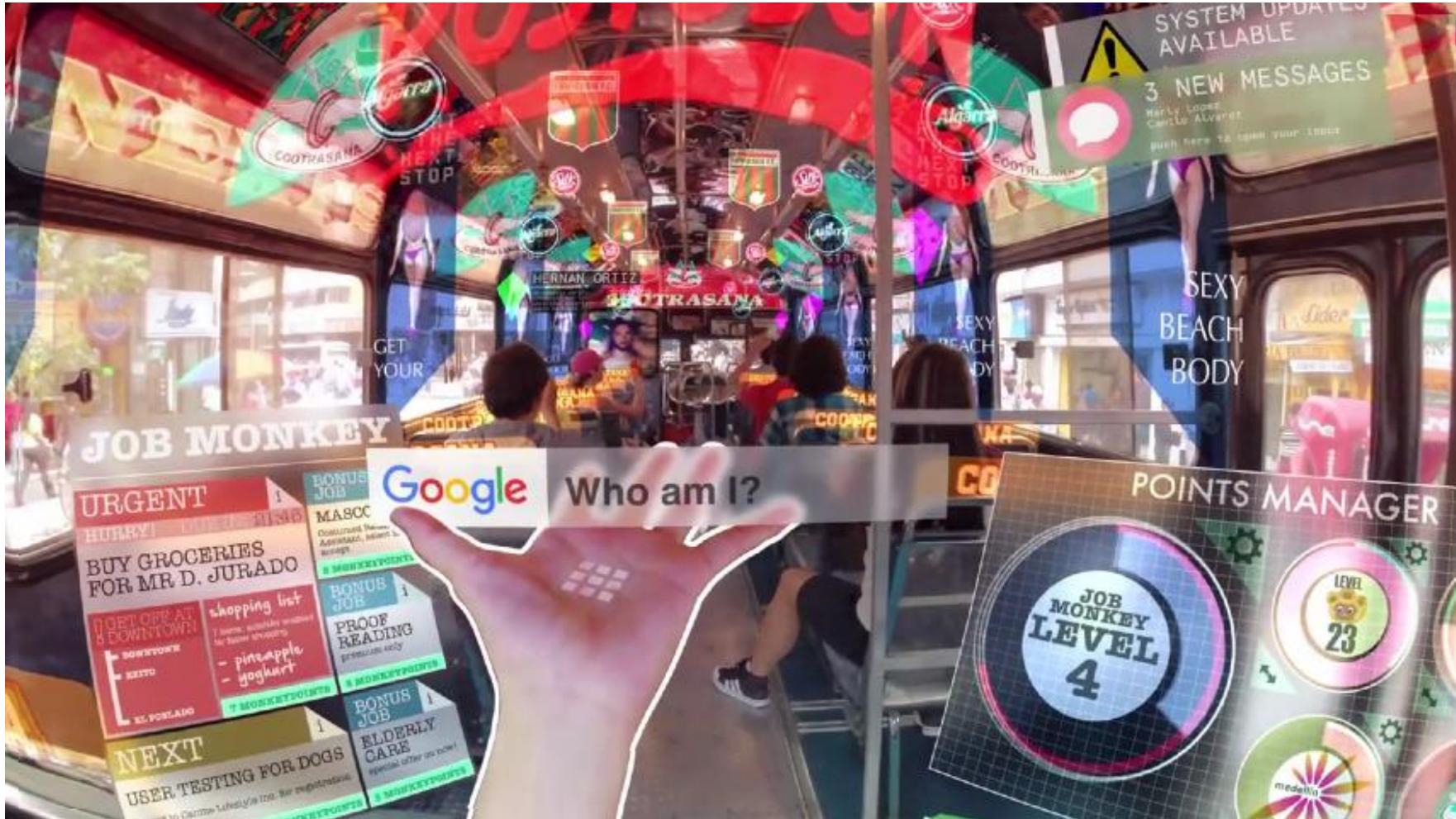
Measurement Revolution

What People Do

What People Say

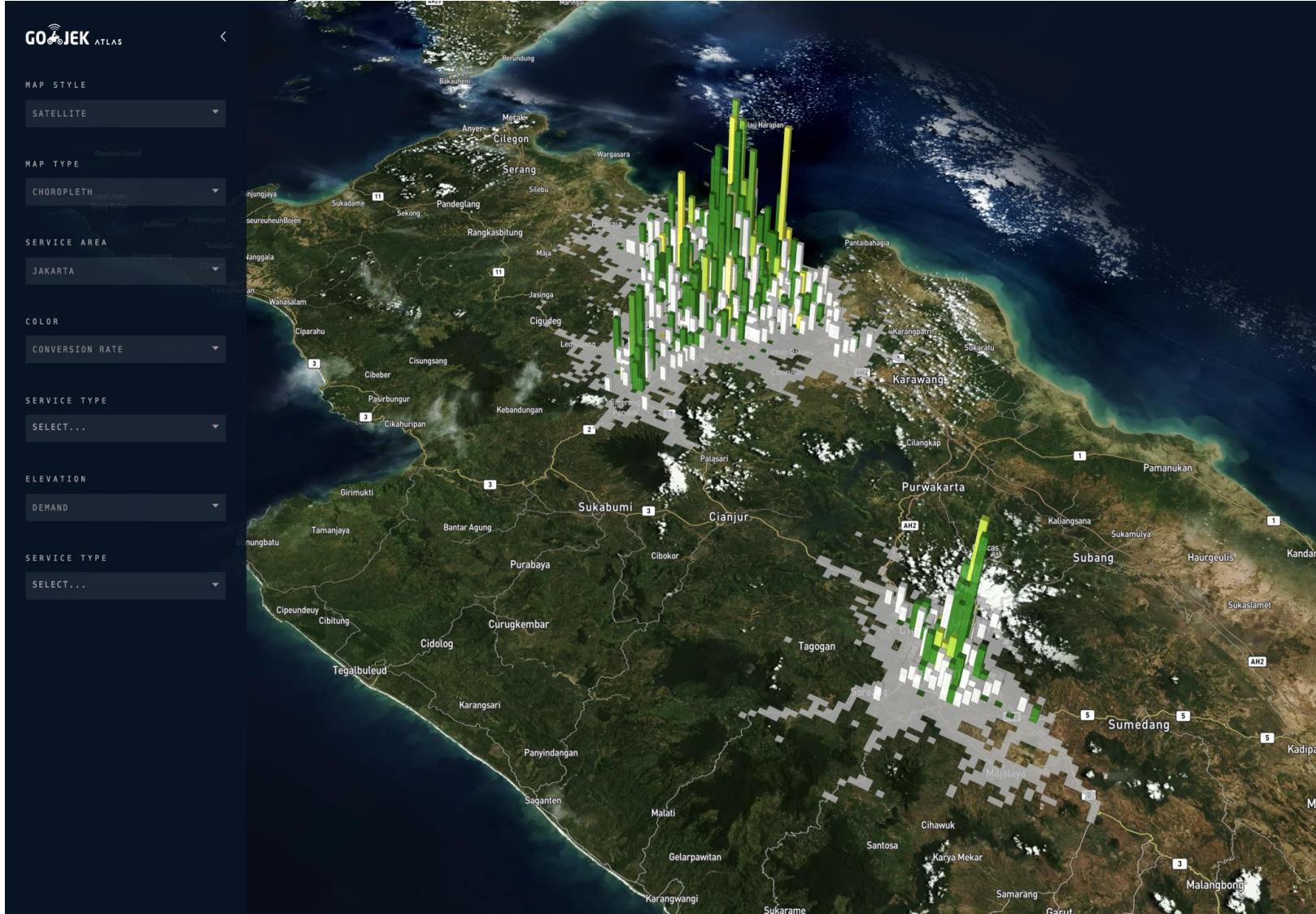


Data Driven World



Keiichi Matsuda's "Hyper-Reality."

ATLAS: GO-JEK's real-time geospatial visualization platform



Ben Wellington:

How we found the worst place to park in New York City – using big data

TEDxNewYork · 11:48 · Filmed Nov 2014

20 subtitle languages ⓘ

View interactive transcript



- Watch later
- Favorite
- Download
- Rate

Big Data

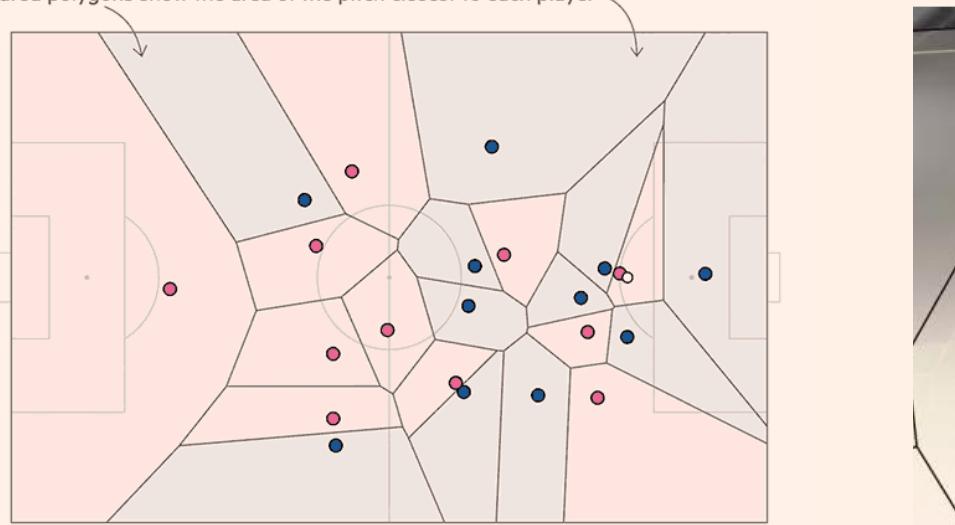
+ Add to myFT

How data analysis helps football clubs make better signings

Machine-learning tools track player movements in the build-up to a goal

Algorithms calculate metrics such as how much space each player creates

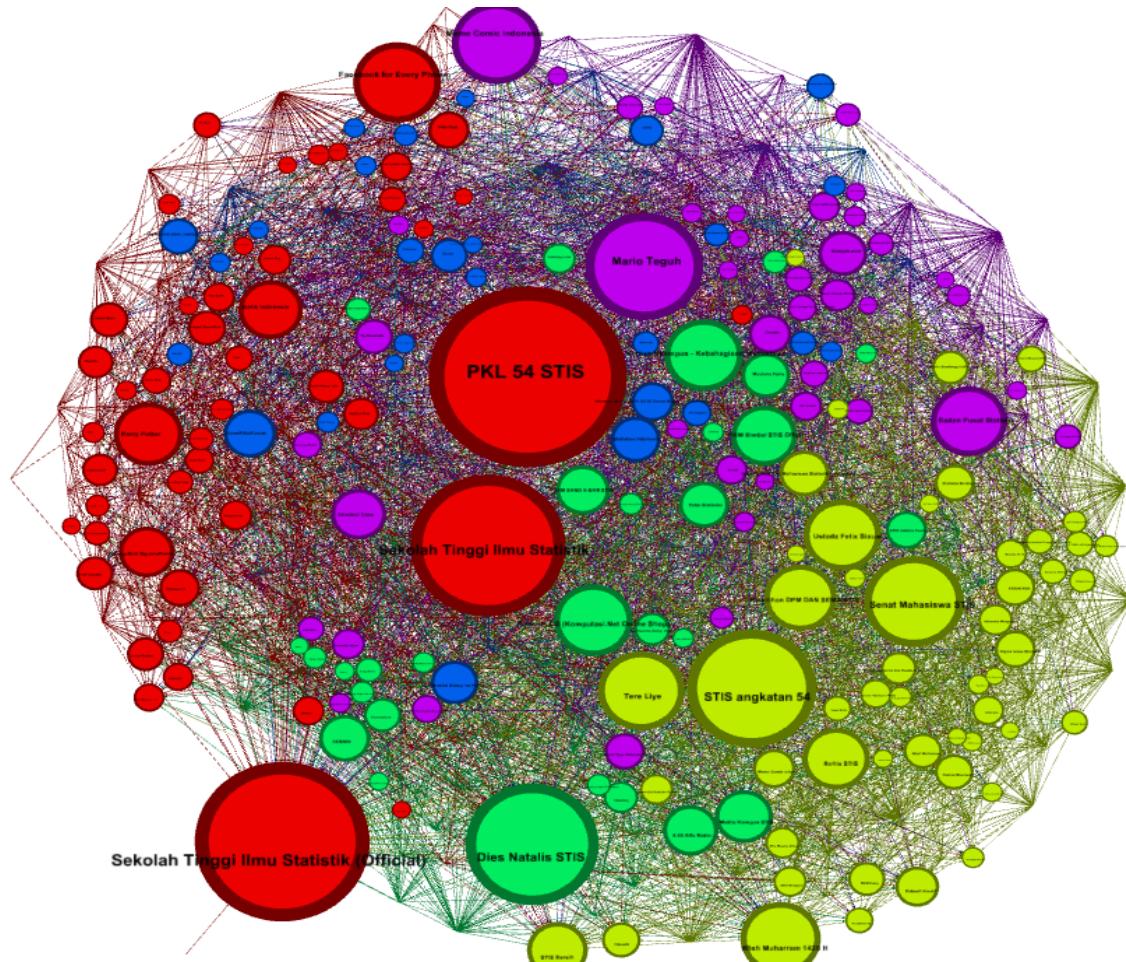
Coloured polygons show the area of the pitch closest to each player



The amount of data collected by 1 million metering devices in a year.

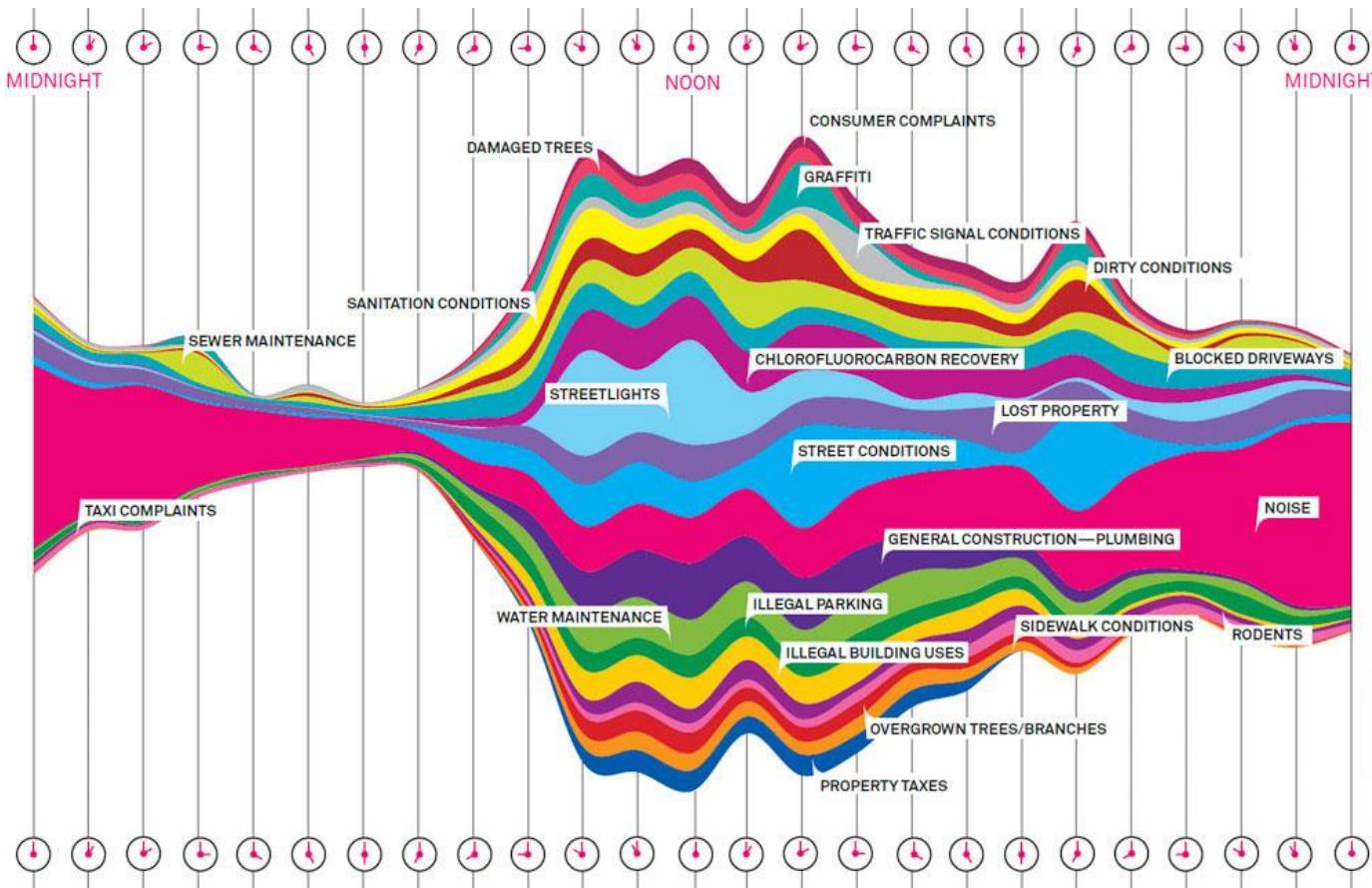
Collection frequency	1/day	1/hour	1/30 min	1/15 min
Records (billion)	0.37	8.75	17.52	35.04

Capturing People Behaviour



What can open data tell us?

- What a Hundred Million Calls to 311 Reveal About New York...



From *Wired* magazine.
“There were 34,522
complaints called in to
311 between September
8 and September 15,
2010. Here are the most
common, plotted by time
of day.
Illustration: Pitch
Interactive”

UK | World | Politics | Budget 2017 | Science | Education | Health | Brexit | Royals

◆ PREMIUM

News

Census questions could be replaced with data from mobile phones to track where people live and work



17



POVERTY IS THE #1 UN SUSTAINABLE DEVELOPMENT GOAL

The elimination of poverty worldwide is the first of 17 UN Sustainable Development Goals for the year 2030. To track progress towards this goal, we require more frequent and more reliable data on the distribution of poverty than traditional data collection methods can provide.

In this project, we propose an approach that combines machine learning with high-resolution satellite imagery to provide new data on socioeconomic indicators of poverty and wealth. Check out the short video below for a quick overview and then read the paper for a more detailed explanation of how it all works.

Combining satellite imagery and machine learning to predict poverty



Inferring Commuting Statistics in Greater Jakarta from Social Media Locational Information from Mobile Devices

Imaduddin Amin, Ni Luh Putu Satyaning P.P, Yulistina Riyadi, and Jong Gun Lee

Pulse Lab Jakarta - United Nations Global Pulse

Email: {imaduddin.amin, ni.paramita, yulistina.riyadi, jonggun.lee}@un.or.id

Setia Pramana and Robert Kurniawan

Institute of Statistics, Jakarta, Indonesia

Email: {setia.pramana, robert}@stis.ac.id

ABSTRACT

Jakarta is the biggest city in Indonesia with a population of more than 10 million and a megacity with 1.38 million people commuting around Greater Jakarta [1]. In order for the Government of Indonesia to understand the commuting behaviors of citizens and better plan the transportation system in Greater Jakarta, the Indonesian Central Bureau of Statistics first conducted its commuting survey in 2014. In this paper, we produce commuting statistics from locational information on social media and show that the information from mobile phone apps is a promising source of data. It implies that the commuting behaviors from big data can allow the public sector more frequent statistics.

people, specifically as about 88.1 million people use the Internet and among them about 79 million people use the Internet for accessing social media¹. Jakarta is often named as the Twitter capital of the world with 10 million tweets every day. As another example, in Greater Jakarta, between January and May 2014 for five months, about 40 million tweets were posted with GPS information.

In this paper, we test how the locational information from social media on mobile devices can reveal commuting patterns in Greater Jakarta. First we produce Origin-Destination statistics given 10 cities in Greater Jakarta from the entire GPS-stamped tweets posted by mobile devices, by identifying a set of people who commute in these areas.



The main conference on the scientific analysis
of mobile phone datasets
5-7 April 2017 Vodafone Theatre, Milan, Italy

commuting-statistics-twitter

Inferring Jakarta Commuting Statistics from Twitter

by Pulse Lab Jakarta

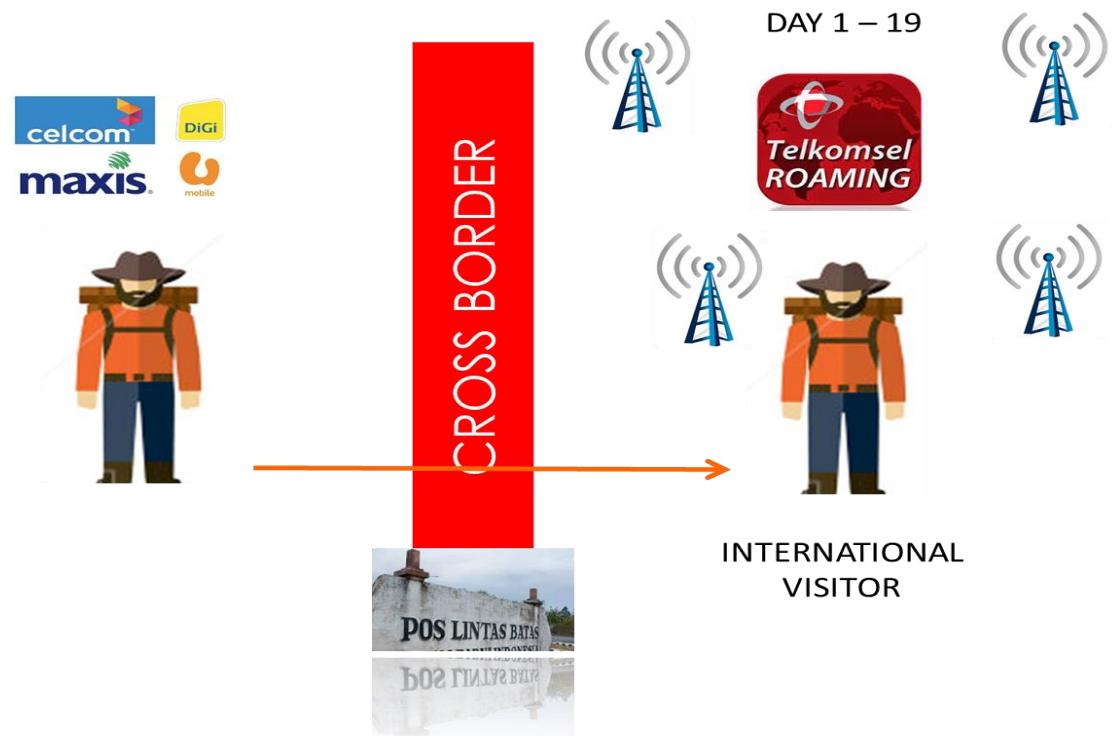
May 30, 2017



Some estimates for Greater Jakarta put the population at over 30 million. Within the boundaries of the city itself the transport system has to handle 1.38 million daily commuters. Policy makers need regular updates to track the rhythm of the city and best optimise public transport. Pulse Lab Jakarta teamed up with the Indonesian Institute of Statistics to look at whether data from social media could help.

Transport policy in a megacity

The Use of Mobile Positioning Data in Tourism Statistics



- Applications:
- International Visitors
 - Domestic Tourists
 - Mobility between regions

Measurement Configuration via Method of Roaming in Cross Border Area

Data Driven Marketing

Collate the Data Sources



Micro-Segmentation

Devise Micro-segments based on combining multiple factors:

- Age
- Location
- Spending History
- Channel Preferences
- Content Preferences
- Apps Usage
- Social Influence
- Churn Score
- Lifetime Value
- Usage Patterns
- Data Usage



Drive Personalized Campaigns

Drive Personalized Campaigns for specific micro-segments

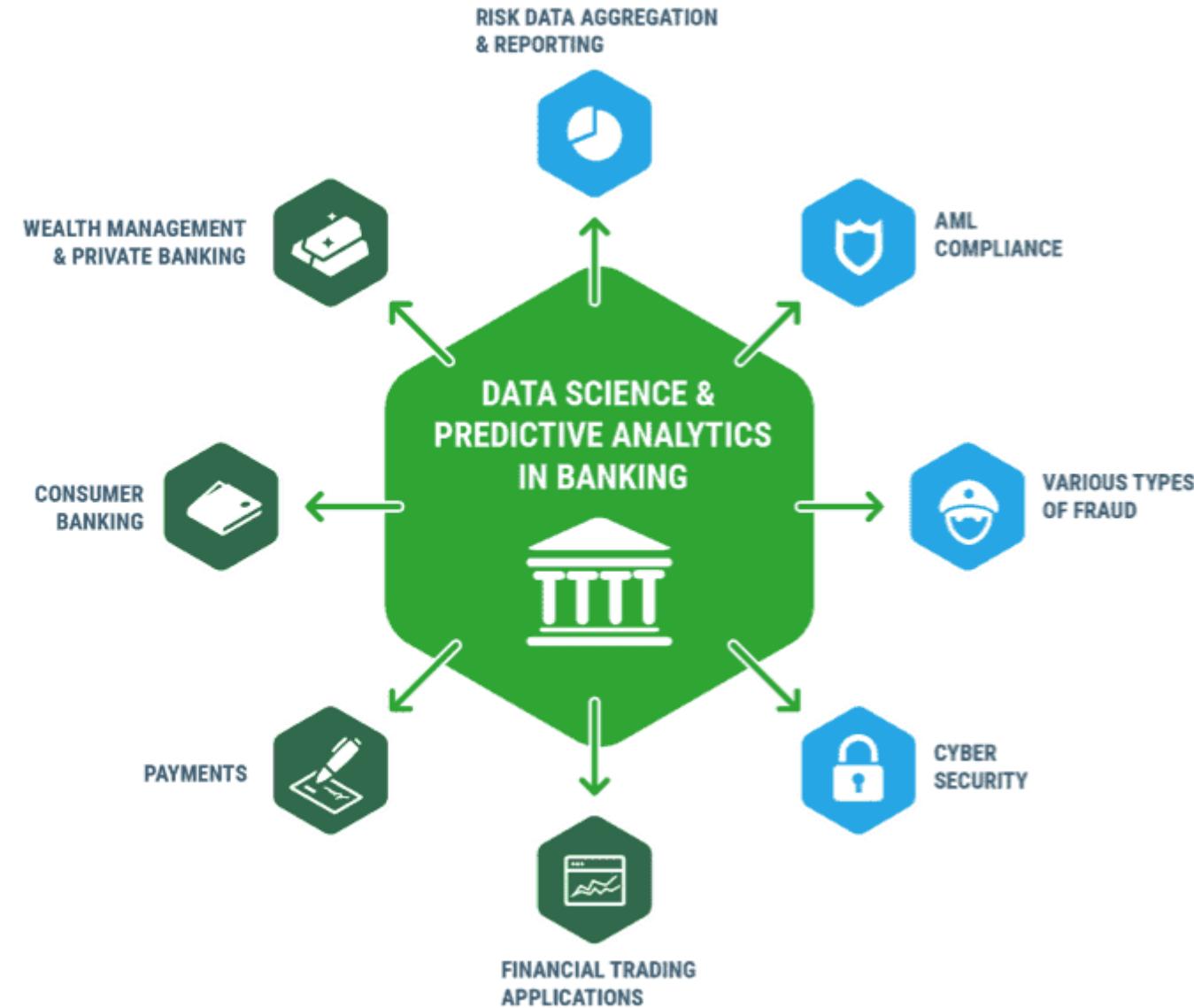


PREDICTIVE ANALYTICS ACROSS BANKING

COLOR KEY

 **DEFENSIVE**
SAVE THE BANK

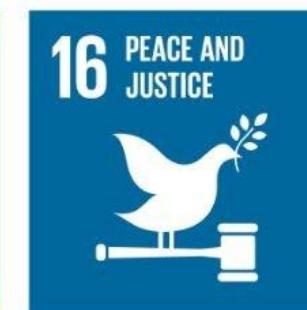
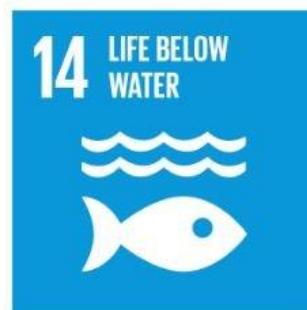
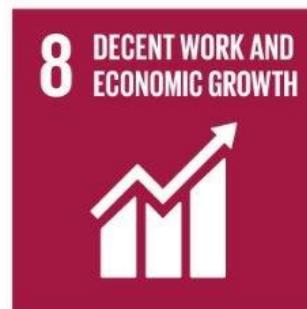
 **OFFENSIVE**
DRIVE PROFITS & COMPETITIVE ADVANTAGE



Business Insight in Banking

- Customer account data and demographics
- Core banking data
- Transactional data at every level of detail
- Wire and payment data
- Trade and position data
- General ledger data including accounts payable, accounts receivable, cash management, purchasing information
- Support data from banking reporting

The World's To-Do List by 2030





How data science and analytics can contribute to sustainable development



GLOBAL
PULSE

www.unglobalpulse.org
©UNGGlobalPulse 2017

- | | | | |
|---|--|--|---|
| 1 NO POVERTY
Spending patterns on mobile phone services can provide proxy indicators of income levels | 6 CLEAN WATER AND SANITATION
Sensors connected to water pumps can track access to clean water | 10 REDUCED INEQUALITY
Speech-to-text analytics on local radio content can reveal discrimination concerns and support policy response | 14 LIFE BELOW WATER
Maritime vessel tracking data can reveal illegal, unregulated and unreported fishing activities |
| 2 ZERO HUNGER
Crowdsourcing or tracking of food prices listed online can help monitor food security in near real-time | 7 AFFORDABLE AND CLEAN ENERGY
Smart metering allows utility companies to increase or restrict the flow of electricity, gas or water to reduce waste and ensure adequate supply at peak periods | 11 SUSTAINABLE CITIES AND COMMUNITIES
Satellite remote sensing can track encroachment on public land or spaces such as parks and forests | 15 LIFE ON LAND
Social media monitoring can support disaster management with real-time information on victim location, effects and strength of forest fires or haze |
| 3 GOOD HEALTH AND WELL-BEING
Mapping the movement of mobile phone users can help predict the spread of infectious diseases | 8 DECENT WORK AND ECONOMIC GROWTH
Patterns in global postal traffic can provide indicators such as economic growth, remittances, trade and GDP | 12 RESPONSIBLE CONSUMPTION AND PRODUCTION
Online search patterns or e-commerce transactions can reveal the pace of transition to energy efficient products | 16 PEACE, JUSTICE AND STRONG INSTITUTIONS
Sentiment analysis of social media can reveal public opinion on effective governance, public service delivery or human rights |
| 4 QUALITY EDUCATION
Citizen reporting can reveal reasons for student drop-out rates | 9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
Data from GPS devices can be used for traffic control and to improve public transport | 13 CLIMATE ACTION
Combining satellite imagery, crowd-sourced witness accounts and open data can help track deforestation | 17 PARTNERSHIPS FOR THE GOALS
Partnerships to enable the combining of statistics, mobile and internet data can provide a better and real-time understanding of today's |
| 5 GENDER EQUALITY
Analysis of financial transactions can reveal the spending patterns and different impacts of economic shocks on men and women | | | |

Big Data for Government Policy: Potential Implementations of BigData for Official Statistics in Indonesia

Setia Pramana, Budi Yuniarto, Robert Kurniawan, and
Ricky Yordani

Center for Computational Statistics Studies
Institute of Statistics
Jakarta Indonesia
setia.pramana@stis.ac.id

Jonggun Lee, Imaduddin Amin, Ni Luh Putu
Satyaning P.P, and Yulistina Riyadi
United Nations Global Pulse Lab, Jakarta
Indonesia

Atika Nashirah Hasyyati and Rina Indriani
BPS-Statistics Indonesia

How about your department?

Data, data and data everywhere.....

Big Data is affecting people everywhere.



Big Data is changing business



Big data industry estimated to be worth
\$100 BILLION+

IDC believes that within five years.... companies will finally be able to run a real-time enterprise that simultaneously transacts, analyzes and acts on big data.



80% of the most competitive organizations capitalize on data for decision-making.
58% of the least competitive organizations capitalize on data for decision-making.

42% of Asia Pacific organisations expect customer service analytics to benefit most from an in-memory data management and analysis technology

7.9 ZETTABYTES (ZB)
ESTIMATED AMOUNT OF DIGITAL DATA WORLDWIDE BY 2015
If one dollar bill represented one byte, a zettabyte would stretch from the Sun to Pluto 18,000 times over

SAP is helping customers get real value from Big Data

MKI performs genome analysis with SAP HANA
“...with this (SAP HANA) we've found a way to shorten the genomic analysis time from several days down to only 20 minutes.”

YUKIHIKO KATO, CTO AND DIRECTOR OF MKI

eBay uses predictive analytics to gain new insights
“With the speed of HANA great people become exceptional at what they do because of the speed that they can interact with the data. That is truly awesome.”

DANIEL SCHWARZBACH,
VP & CFO EBAY NORTH AMERICA AT EBAY INC.

Bigpoint solves big data challenges with SAP HANA
Our expectation – and it actually seems to be coming true – is that the use of this technology and the methods behind it helps us realize sales growth spurts of 10 – 30%.

MICHAEL GUTSMANN, CFO BIGPOINT



Find out what opportunities Big Data holds for you.

Follow us on Twitter @asiarunbetter Visit www.sap.com/bigdata Contact SAP





Analytics Approaches

- Descriptive: What happened or what is happening now?
- Diagnostic: Why did it happen or Why is it happening now?
- Predictive: What will happen next? What will happen under various conditions?
- Prescriptive: What are the options to create the most optimal/high value result/outcome?

ANATOMY OF A DATA SCIENTIST

SALARY

Average salary of data scientists is
\$120,000/year

BENEFITS

- Harvard Business Review called data science the **"Sexiest Job of the 21st Century"**
- One of the fastest growing careers in the United States
- 94%** of data science graduates have found jobs since 2011

RESPONSIBILITIES

- Conduct research
- Extract, clean, and analyze data from varied sources
- Solve problems
- Build automation tools
- Communicate findings to management

EDUCATION

- 88%** of all data scientists have at least a Master's degree
- 46%** of data scientists have a PhD

SKILLS

- Programming languages (R, Python, SQL, Hive, etc.)
- Statistics
- Multivariable calculus and linear algebra
- Machine learning
- Software engineering
- Wrangle, visualize, and communicate data to management

CAREER POSSIBILITIES

- The majority of data scientists work in the **technology industry**.
- Other options include marketing, consulting, healthcare and pharmaceuticals, finance, government, gaming, and many more.



MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21th century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- Machine learning
- Statistical modeling
- Experiment design
- Bayesian inference
- Supervised learning: decision trees, random forests, logistic regression
- Unsupervised learning: clustering, dimensionality reduction
- Optimization: gradient descent and variants



PROGRAMMING & DATABASE

- Computer science fundamentals
- Scripting language e.g. Python
- Statistical computing package e.g. R
- Databases SQL and NoSQL
- Relational algebra
- Parallel databases and parallel query processing
- MapReduce concepts
- Hadoop and Hive/Pig
- Custom reducers
- Experience with xaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- Passionate about the business
- Curious about data
- Influence without authority
- Hacker mindset
- Problem solver
- Strategic, proactive, creative, innovative and collaborative



COMMUNICATION & VISUALIZATION

- Able to engage with senior management
- Story telling skills
- Translate data-driven insights into decisions and actions
- Visual art design
- R packages like ggplot or lattice
- Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of e-commerce marketing. Our fields of expertise include: marketing strategy and optimization; customer tracking and on-site analytics; predictive analytics and econometrics; data warehousing and big data systems; marketing channel insights in Paid Search, SEO, Social, CRM and brand.

Our Road Map

- Introduction to Data Science
- Introduction to R
- Introduction to Python
- Data Management and Wrangling
- Data Presentation and Visualization
- Data Mining
- Supervised Learning (classification)
- Unsupervised Learning (cluster Analysis)
- Forecasting
- Text Mining
- Case Studies
- Coaching

Thank you!

setia.pramana@stis.ac.id

