Hype and Reality: Big Data

Dr. David Bergvinson, Director General, International Crops Research Institute for the Semi-arid Tropics (ICRISAT)

Dr. Usha Zehr, Chief Technology Officer, Maharashtra Hybrid Seeds Company Private Limited (Mahyco)

Dr. Ranjan Patnaik, Director, Dupont Knowledge Center, Technology Director SA & ASEAN, Dupont

Frederic Pivetta, Co-founder & Managing Partner, Dalberg Data Insights
Hype and Reality: Big Data

• Panel Discussion

• Moderator: Dr David Bergvinson – Director General, ICRISAT

• Questions: #ICT4D2017
Introducing our Panelists

• Using data from the Global Food Security Index to Assess SDG 2 - Eliminating Hunger
  – Dr. Ranjan Patnaik, Director – DuPont Knowledge Center

• An overview of using Big Data to measure impact and how it can improve targeting and impact assessment and highlight challenges.
  – Frederic Pivetta, Co-Founder and Managing Partner, Dalberg Data Insights

• An overview on continuum of crop improvement, precision agriculture and adapting to climate change through analytics
  – Usha Zehr, Chief Technology Officer, MAHYCO
The Global Food Security Index

Hyderabad, May 16, 2017

Dr. Ranjan Patnaik
DuPont Knowledge Center
Dalberg Data Insights combines Strategy Advisory with Big Data skills

DDI builds on a global footprint:

**Dalberg**

*world class advisory for social and economic development impact*

- 21 offices, ~400 people focusing on advising in the context of economic development and social impact
- Key practices around agriculture and food security, energy, cities and urban development, financial inclusion, health and nutrition, design, national statistics, education and employment, environment, gender empowerment, humanitarian assistance, infrastructure, water and sanitation,
- Key focuses around global strategy, evaluation, human centred design, inclusive business growth, inclusive economic development, leadership and talent, organizational effectiveness, policy and advocacy

**RIA**

*clients globally with real time data on over 400 million people*

- 3 locations, 140 people focusing on Big Data for telecom operators
- Commercial approach – operate relations with 5 out of the top 10 telecom operators in the world through about 30 countries and develop global product portfolio in Big Data
- Social approach – work with the largest and most innovative donors in the world and have secured access to telecom data in 22 emerging countries, mostly in data-poor environments
Dalberg Data Insights is part of the Dalberg Group

Combining Big Data capabilities with strategy can lead to transformational social impact and business value
Working with governments, private businesses and ecosystem players

<table>
<thead>
<tr>
<th>Governments</th>
<th>Private businesses</th>
<th>Ecosystem players</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brazil - City of Sao Paulo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uganda - Kampala capital city authority and National Statistics office</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haiti – Ministry of Public transport</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Digicel</td>
<td>Bill &amp; Melinda Gates foundation</td>
</tr>
<tr>
<td></td>
<td>MTN</td>
<td>World Bank</td>
</tr>
<tr>
<td></td>
<td>Airtel</td>
<td>United Nations</td>
</tr>
<tr>
<td></td>
<td>Telefonica</td>
<td>USAID</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IDB</td>
</tr>
</tbody>
</table>

Have framework agreement to access data or specific contract for a specific topic

Align ecosystem player agenda with end-user / operational partner needs
DDI creates platforms of algorithms and secure privacy

- All individual data remain within the premises of the data providers
- All individual data are anonymized
- All individual data are aggregated
- All algorithms are open and available
- Pushing algorithms to the data

### Big Data smart city platform

<table>
<thead>
<tr>
<th>Module 1 – Telecom data module</th>
<th>Module 2 – Survey data module</th>
<th>Module 3 – Administrative data module</th>
<th>Module 4 – Retailers’ data module</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module 5 – Satellite data module</td>
<td>Module 6 – Public transport data module</td>
<td>Module 7 – Social media data module</td>
<td>Module 8 – Basic technical layers</td>
</tr>
<tr>
<td>Module 9 – Mobility monitoring module</td>
<td>Module 10 – Public transport module</td>
<td>Module 11 – Road network module</td>
<td>…</td>
</tr>
</tbody>
</table>

…
DDI focuses on 4 areas in social impact

1. Financial Inclusion
   - Where and how to push digital payments?
   - Where to further develop BC network?

2. Smart cities
   - What are the traffic patterns?
   - Where to further develop urban infrastructures?
   - How to optimize public transport?

3. Public health
   - Where to prioritize disease control and eradication?
   - Are the quarantine zones enforced?

4. National Statistics and SDGs
   - Where do people work and live?
   - Where are the poor communities?
   - What are the female communities?

---

Uganda
Haiti

Uganda
Mobility mapping and assessment in the Greater Kampala Area, Uganda

Using telecom data, we mapped mobility in Kampala in real-time and assessed commuting patterns across the city.

- Understanding origin / destination flows between neighbourhoods, prioritising commuting routes and peak travel times

We answer key questions:

- Which places attract the most people at what times?
- What are the preferred routes between neighbourhoods?
- How many trips are taking place and how is traffic evolving over time?
- What is the speed and travel time?

These insights can be used for:

- Traffic monitoring
- Better planning of public transport networks
- Prioritise infrastructure investments
Impact Assessment of Road Works in Kampala, Uganda

Using telecom data, we measured mobility performance before and after the building of a new junction in Kampala to understand its impact.

Using our algorithms we mapped the origin/destination of daily commuters, estimating the flow of people and their travel time over various time periods.

These insights can be used for:
- Infrastructure planning and decision-making
- Assessing future investments in infrastructure projects

SHOW ME
- Impact Evaluation: Situation in May 2015, Situation in Feb 2017

TRAFFIC DIRECTION
- All Traffic going through Fairway Junction
- Only Traffic entering Fairway Junction
- Only Traffic exiting Fairway Junction

DAY TYPE
- Business Day
- Saturday
- Sunday

TIME PERIOD
- 00:00 to 23:59

Impact Analysis per Zone
- Nakasero: +52% (May 2015: 1,110, Feb 2017: 1,760)
- Wandegeya: +11% (May 2015: 1,100, Feb 2017: 1,213)
- Kibuli: +11% (May 2015: 947, Feb 2017: 1,018)
- Kalaba: -5% (May 2015: 847, Feb 2017: 802)

Traffic over time (ppl/h)
- 12 AM to 6 AM:
  - May 2015: 0, Feb 2017: 0
  - 6 AM to 12 PM:
    - May 2015: 0, Feb 2017: 0
  - 12 PM to 6 PM:
    - May 2015: 0, Feb 2017: 0
  - 6 PM to 12 AM:
    - May 2015: 0, Feb 2017: 0

Average Speed over time (km/h)
- 12 AM to 6 AM:
  - May 2015: 0, Feb 2017: 0
  - 6 AM to 12 PM:
    - May 2015: 0, Feb 2017: 0
  - 12 PM to 6 PM:
    - May 2015: 0, Feb 2017: 0
  - 6 PM to 12 AM:
    - May 2015: 0, Feb 2017: 0

Total Time Lost over time (hours)
- 12 AM to 6 AM:
  - May 2015: 0, Feb 2017: 0
  - 6 AM to 12 PM:
    - May 2015: 0, Feb 2017: 0
  - 12 PM to 6 PM:
    - May 2015: 0, Feb 2017: 0
  - 6 PM to 12 AM:
    - May 2015: 0, Feb 2017: 0
Using telecom and public health data, we mapped mobility and incidence rates in São Paulo to develop real-time insights on Zika flows.

We answer key questions:

- Which regions are the biggest exporters of Zika?
- Which regions should elimination programmes be targeted in?
WHAT ARE THE KEY APPLICATION FEATURES?

Identification of priority target areas

We can identify areas in which eliminating disease would have the greatest impact on disease in the overall region.

Areas with the highest target effectiveness

The areas on the map are coloured based on how effective elimination of malaria in that area would be on reducing malaria import to other regions. Click on an area to see which other areas are currently importing malaria from there.

- High export areas can be identified by the number of areas they distribute disease to
- We can model the decrease in risk of disease by eliminating disease from the high export area
Identifying female communities and address the gender biases/gaps

Women are reliable and efficient users of microcredit loans, but how do you target them to increase financial inclusion?

➢ We can identify women through their phone usage patterns

Data
- Airtel Uganda (39% women): CDRs, Top-ups + CRM data for the whole customer base
- Dataset A (28% women): outgoing CDRs, Top-ups + CRM data for ca 160,000 users
- Dataset B (42% women): CDRs, TOP-UPS + CRM for ca 160,000 users

Methodology
- Over 150 features summarizing usage patterns, social network, mobility and top-ups
- Trained random forests and support vector machines based on a labelled training sample

Results
- 3 key variables determining female gender: call duration, number of outgoing calls and number of contacts for incoming calls
- Predicted gender accuracy of 70-75% and option to better target one gender group
- Lowering the coverage increases the accuracy: for dataset A we are 90% sure of the gender for 30% of our sample
Mapping granular poverty pockets

Distribution of telecom spending and population density based on Telecom usage

Mapping of poverty pockets and evolution of granular slum areas
Create a gradual digital strategy around key building blocks

<table>
<thead>
<tr>
<th>Topics / Sectors</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pilot use cases</strong> using aggregated public open data and some private data sources to show value and opportunities / ecosystem</td>
<td>Platforms of mostly open algorithms with a network of technical partners accessing and integrating multiple public and private data sources to address scalable topics</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regulation</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Research environment</strong></td>
<td>Supportive set of laws and regulations</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data providers</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Research partner</strong></td>
<td>Market for data, where data providers see data as a commodity, including economic / financial flows</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>End-users</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Co-developing third parties</strong></td>
<td>Ecosystem of end-users of operational tools, involving specific processes, e.g. resilience officer for smart cities</td>
<td></td>
</tr>
</tbody>
</table>
PLANT BREEDING
Seed

Markers

Genotype

Phenotype

Germplasm

Prebreeding

Breeding

Trials

Production
BIG DATA
MOLECULAR FINGERPRINTS OF PLANTS
CLIMATE SMART BREEDING

EYE IN THE SKY
Farmers app

तपासणी साधी वाण निखाडा

○ कॉटन-27BG-FS

१५ दिवसात पाहणी करा

१५ दिवसात पाहणी करा

आपली टिऩ्याणी

Receipt request

Product: Bittergourd-231

<table>
<thead>
<tr>
<th>Product</th>
<th>Bittergourd-231</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company</td>
<td>Maharashtra Hybrid Seeds Co. Pvt. Ltd.</td>
</tr>
<tr>
<td>Grower Code</td>
<td>0040030874</td>
</tr>
<tr>
<td>Grower Name</td>
<td>RAOASAHEB RAGHUNATH</td>
</tr>
<tr>
<td>Village</td>
<td>CHANDAI</td>
</tr>
<tr>
<td>P.O. NO.</td>
<td>7062083</td>
</tr>
<tr>
<td>Year-Season</td>
<td>2016-17</td>
</tr>
<tr>
<td>Batch</td>
<td>FKG20266</td>
</tr>
<tr>
<td>Rept. DL.</td>
<td>15.10.2016</td>
</tr>
<tr>
<td>Un-Process Qty (Kg.)</td>
<td>53.6</td>
</tr>
<tr>
<td>Bags</td>
<td>1</td>
</tr>
</tbody>
</table>

Field visit request

या संकेतस्थायीतः दर्शवलेली महत्त्वपूर्ण नाही, किंवा कोणतेही शासकीय अथवा कायदेद्वीर बाबतीत वापरणा गेला नाही.
Data challenges

- Data filtering
- Data analytics
- Managing data
- Affordability of data