Title: Big data research at VTT
Author(s): Södergård, Caj
Citation: Vetenskapsademiens årsmöte 28.3.2017, Helsinki, Finland (2017), pages 13.
Rights: This presentation may be downloaded for personal use only.
Big Data Research at VTT

Svenska Tekniska Vetenskapsademiens årsmöte 28.3.2017

Prof. Caj Södergård
"Data is the world´s great new natural resource. What steam power was to the 18th century, electromagnetism was to the 19th and fossil fuels to the 20th... data will be to the 21st."

Ginni Rometty, President and CEO of IBM
What is Big Data?

3V

▪ Volume
▪ Velocity
▪ Variety

(Veracity, Value)

Digital data grows at an annual growth rate (CAGR) of 42% until 2020 (source IDC)

Source: Patrick Cheesman
HOW CAN BIG DATA MAKE A DIFFERENCE?

DO YOU KNOW?
Decoding the human genome took
in 2003
today
near future
10 Years
less than a WEEK
a few HOURS

REAL LIFE APPLICATIONS

ON THE FARM
More efficient use of natural resources

IN SHOPS & FACTORIES
Improve efficiency and productivity

IN THE HOSPITAL
Better diagnosis & clinical decision

IN THE HOME
Reduce home energy consumption

ON THE MOVE
Management of traffic flows

ALL OF THIS IS POSSIBLE BECAUSE OF

High-speed broadband
Cloud services
High-performance computing
Data analytics tools and methods

Source: EU
Data-Driven Bioeconomy

- **Bioeconomy**: The utilization of raw materials from agriculture, forestry and fishery for food, energy and biomaterials with responsibility & sustainability

- Large quantities of data from
  - farm & forestry machines & fishing vessels
  - satellite & air imagery and local sensors

  About
  - soil, microclimate, trees, plant diseases, sunlight
  - marine environment

- **DataBio** shows the benefits of Big Data in the raw material production for the bioeconomy industry

- EU H2020 Lighthouse. Duration 2017 - 2019, Volume 16,2 M€, 48 partners, 71 associate partners

- VTT biggest partner and Technical Manager

Estimating forest cuttings from satellite images

Cuttings in Finland from Landsat-8 2014 & Sentinel-2 2015

Autochange method (VTT). Area size 15 km by 15 km

<table>
<thead>
<tr>
<th></th>
<th>N=108</th>
<th>Keskipituus (m)</th>
<th>Keskiläpi-mitta (cm)</th>
<th>Runkoluku (/ha)</th>
<th>Keski-tilavuus (m³/ha)</th>
<th>Pohjapinta-ala (m²/ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ennusteen keskiarvo</td>
<td>17</td>
<td>21</td>
<td>1666</td>
<td>216</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>-6</td>
<td>-7</td>
<td>1207</td>
<td>-51</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Bias %</td>
<td>-32</td>
<td>-31</td>
<td>72</td>
<td>-23</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>7</td>
<td>9</td>
<td>1424</td>
<td>93</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>RMSE %</td>
<td>41</td>
<td>41</td>
<td>85</td>
<td>43</td>
<td>26</td>
<td></td>
</tr>
</tbody>
</table>

Source: F-TEP (forestry-tep.eo.esa.int) & Tuomas Häme, VTT
Map of broadleaved shrubs from Sentinel-2

Source: F-TEP and Tuomas Häme, VTT
**Example of VTT´s research**

**Mental Concentration Sensor**

- **Need**
  - Mental concentration is central in all learning and work
  - Individuals strive to find optimal conditions – sweet spot –, where their focus is as sharpest

- **Our result**
  - By gathering biosignals from students and using machine learning, we predicted concentration with 75% confidence

Online Quality Monitoring Tool - Case SSAB

Over 100 000 measurement points/
industrial site

Measurements made 24/7

Monitored parameters
thousands per process point

Source: Heli Helaakoski, VTT
Case SSAB - Quality monitoring for liquid steel making process

Benefits

- Detection of process deviation from normal state
- Quick analysis how a melting succeeded
- Give overall understanding of whole liquid steel making quality
- Tool to improve steel cleanness for strengthened customer demands e.g. automotive industry
- Help and support product development and operators in processing quality deviations
- Fasten development work
- Monitoring tool can be used as learning tool

Contact
Heli Helaakoski, VTT
Agne Bogdanoff, SSAB

04/04/2017
Customer data analysis – case media users
Clustering users based on their reading habits

0: Most active (17%)

1: Iltalehti only (59%)

2: Kauppalehti readers (4%)

3: Kotikokki readers (20%)
Most users read daytime

Percentage of use

- Early morning (5 %, 29 x)
- Morning centric (11 %)
- During day (49 %, 82 x)
- Evening centric (26 %)
- Only evenings (9 %, 14 x)
- Night

Cluster 0
Cluster 1
Cluster 2
Cluster 3
Cluster 4

5 %
11 %
26 %
9 %
49 %
TECHNOLOGY FOR BUSINESS