DATA SCIENCE IN INDUSTRY 4.0 ERA

SANI M. ISA
MASTER OF COMPUTER SCIENCE
BINUS GRADUATE PROGRAM
OUTLINE

- What is Industry 4.0?
- Why Industry 4.0?
- What is Data Science?
- Challenges
WHAT IS INDUSTRY 4.0?

The next phase in the digitization of the manufacturing sector, driven by four disruptions:
1. the astonishing rise in data volumes, computational power, and connectivity, especially new low-power wide-area networks;
2. the emergence of analytics and business-intelligence capabilities;
3. new forms of human-machine interaction such as touch interfaces and augmented-reality systems;
4. and improvements in transferring digital instructions to the physical world, such as advanced robotics and 3-D printing.
WHY INDUSTRY 4.0?
WHY INDUSTRY 4.0?

- Faster production
- Better quality
- Increased productivity
- Mass customization
- Sharing economy
- Improved decision making
WHY INDUSTRY 4.0?

- Faster production
- Better quality
- Increased productivity
- Mass customization
- Sharing economy
- Improved decision making
FASTER PRODUCTION

Speed up the manufacturing process:

- **120%** in time needed to deliver orders
- **70%** in time to get products to market
WHY INDUSTRY 4.0?

- Faster production
- Better quality
- Increased productivity
- Mass customization
- Sharing economy
- Improved decision making
Top 100 European manufacturers could save an estimated €160 billion in the costs of scrapping or reworking defective products if they could eliminate all defects.
WHY INDUSTRY 4.0?

- Faster production
- Better quality
- Increased productivity
- Mass customization
- Sharing economy
- Improved decision making
INCREASED PRODUCTIVITY

Top 100 European manufacturers could save an estimated €160 billion in the costs of scrapping or reworking defective products if they could eliminate all defects.
WHY INDUSTRY 4.0?

- Faster production
- Better quality
- Increased productivity
- Mass customization
- Sharing economy
- Improved decision making
MASS CUSTOMIZATION

Increased flexibility in production and manufacturing on demand
WHY INDUSTRY 4.0?

- Faster production
- Better quality
- Increased productivity
- Mass customization
- Sharing economy
- Improved decision making
SHARING ECONOMY

Custom manufactured products at low-cost prices and delivered home on time

“We will buy device usage”
WHY INDUSTRY 4.0?

- Faster production
- Better quality
- Increased productivity
- Mass customization
- Sharing economy
- Improved decision making
**Real-time Monitoring**
Process production sensor data in real-time from machines to the global reliability portal.

**Predictive Quality**
Increase product quality and reduce waste and costs by analyzing sensor data in the production process.

**Process Mining**
Analyze process steps, throughput time and its main drivers to derive process optimization measures.

**Location-Based Solutions**
Record, analyse and visualize position and movements of any resources in real time to improve shop floor and logistics efficiency.
WHAT IS DATA SCIENCE?
WHAT IS DATA SCIENCE?
DATA SCIENCE IS TEAM WORK
DATA SCIENCE STEPS

Data Engineering

Computational Data Science

ACQUIRE  PREPARE  ANALYZE  REPORT  ACT
Step 1: Acquire Data

- Identify data sets
- Retrieve data
- Query data
Data comes from many places…

…with many ways to access it
Step 2: Prepare Data

- Step 2-A: Explore
- Step 2-B: Pre-process
Step 2A: Explore Data

Understand nature of data

Preliminary analysis
VISUALIZE YOUR DATA

- Histogram
- Line graphs
- Heat maps
- Scatter plots
- Boxplots
Step 2B: Pre-Process Data

ACQUIRE  PREPARE  ANALYZE  REPORT  ACT

Clean  Integrate  Package
Real-world data is messy!

Data Quality Issues

- Inconsistent values
- Duplicate records
- Missing values
- Invalid data
- Outliers
Step 3: Analyze Data

Select analytical techniques
Build models
SUPERVISED VS UNSUPERVISED

Supervised Learning

Unsupervised Learning
Select technique

Classification
Regression
Clustering
Association
Analysis
Graph Analytics

Build model

Evaluate
Classification

Goal: Predict category

Sunny
Windy
Rainy
Cloudy
PCB AUTOMATIC VISUAL INSPECTION SYSTEM

- **Solder Paste Inspection**: Machine vision checks for slumped or scavenged print, bridging, and peaking.

- **Surface Mount Device Inspection**: Machine vision inspects lead length, width, pitch, bent, lead absence, chip size, and ball position, size, and pitch.

- **Automated Optical Inspection (AOI)**: A visual test of the populated board inspects placement of parts and detects missing, reversed, or incorrect components.
Regression

Goal: Predict numeric value
PREDICTIVE MAINTENANCE

1# COLLECT
Gather near real-time data e.g. via an IoT CAN bus logger and transfer data to the cloud via a WiFi hotspot.

2# PREDICT
Process the data in the cloud and use it in the a model (e.g. regression) to predict future failure.

3# REACT
Use auto-reactions to ensure that necessary action is taken by e.g. informing & guiding maintenance teams.
Clustering

Goal: Organize similar items into groups

- Seniors
- Adults
- Teenagers
Traditional approach

Cluster resources by its capability and parts by process step

Cellular Layout

Resources to produce similar products placed together
Association Analysis

Goal: Find rules to capture associations between items
ASSOCIATION ANALYSIS IN PCB PRODUCTION

- Association of failures with root cause.
- Association of failures with suppliers.
- Identifying failures occurring in sequence.
- Association of failures with the location of the root cause on the board.

<table>
<thead>
<tr>
<th>Board Type</th>
<th>serial</th>
<th>supplier</th>
<th>Failure</th>
<th>reason-of-failure</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOODBOARD</td>
<td>2459</td>
<td>TATCHUN-GIA TZOONG</td>
<td>display error</td>
<td>no solder</td>
<td>U45.6 PIN</td>
</tr>
<tr>
<td>736</td>
<td></td>
<td>AUX1 error</td>
<td>short circuit</td>
<td></td>
<td>U8.2 PIN</td>
</tr>
<tr>
<td>GIA TZOONG</td>
<td>990</td>
<td>device-not-work</td>
<td>sw</td>
<td></td>
<td>L71</td>
</tr>
<tr>
<td>TATCHUN-GIA TZOONG</td>
<td>700</td>
<td>display error</td>
<td>short circuit</td>
<td></td>
<td>R407</td>
</tr>
<tr>
<td>UNAL ELEKTRONIK</td>
<td>712</td>
<td>rgb-cvbs error</td>
<td>flash error</td>
<td></td>
<td>R412</td>
</tr>
<tr>
<td>GOODBOARD</td>
<td>1411</td>
<td>sw error</td>
<td>sw</td>
<td></td>
<td>K23</td>
</tr>
<tr>
<td>GOODBOARD-TATCHUN</td>
<td>663</td>
<td>AUX1 error</td>
<td>no solder</td>
<td></td>
<td>C130</td>
</tr>
<tr>
<td>UNIWELL ELECTRONIC</td>
<td>627</td>
<td>audio error</td>
<td>upside-down</td>
<td></td>
<td>B353</td>
</tr>
<tr>
<td>GOODBOARD</td>
<td>1169</td>
<td>sw error</td>
<td>sw</td>
<td></td>
<td>U6</td>
</tr>
</tbody>
</table>
Graph Analytics

**Goal:** Use graph structures to find connections between entities
To improve an existing complex production system, which involves many operations occurring in series and parallel, the dominating path needs to be analyzed. The path that takes the maximum time to complete the process is defined as the critical path for that process.
MACHINE LEARNING MODELS

- **Clustering**
  - K-means
  - Hierarchical Clustering
  - DBSCAN
  - KNN
  - One Class SVM
  - Neural Networks

- **Anomaly Detection**
  - DBSCAN
  - KNN
  - One Class SVM
  - Neural Networks

- **Regression**
  - Linear Regression
  - GLM
  - ARIMA
  - Regression Trees
  - Neural Networks

- **Classification**
  - Logistic Regression
  - Decision Trees
  - Random Forest
  - Support Vector Machine
  - Neural Networks
  - Dimensionality Reduction (LDA, PCA)
  - Ensembling (Boosting, Bagging, Stacking)

- **Association & Sequence**
  - Apriori
  - Markov Chains

- **Dimensionality Reduction (LDA, PCA)**
- **Ensembling (Boosting, Bagging, Stacking)**
Step 4: Communicate Results
What?

How?

Tools?

Tableau

Google Developers
Step 5: Apply Results
Determine Next Steps

Action

Evaluation

- Favorable Results?
- Revisit?
- Further Opportunities?
CHALLENGES
## Challenges

### Technical Challenges
- Communication reliability and QoS
- Cyber security
- Maturity of machine intelligence
- Handling big data

### Social Impacts
- General reluctance to change by stakeholders
- Lack of adequate skill-sets
- Unemployment
BIG DATA

Autonomous Vehicle

4 TB Per Day

RADAR: ~10-100 KB per second

Camera: ~20-40 MB per second

SONAR: ~10-100KB per second

LIDAR: ~10-70 MB per second

GPS: ~50 KB per second
SOCIAL IMPACT

- We can expect a wave of structural unemployment to spring from the technology in the medium term.

- A study out of Oxford University in 2014 found that in the near future artificially intelligent technology could take over nearly half of all U.S. jobs.
CHALLENGES

Machine
- Faster
- More precise
- Able to work in harsh environments
- Higher initial cost and lower running cost
- Fit for 4D tasks: dull, dangerous, dirt, dumb

Human
- Creative
- Lower initial cost and higher running cost
Thank you

sani.m.isa@binus.ac.id