Big Data, AI & Analytics

Use case from Novo Nordisk

Dan Holst Jakobsen
Business Analyst
Novo Nordisk A/S
Agenda

- Background
  - Having realistic expectations
  - Being ready and mature
  - Creating the right team

- Use case
  - Getting the data
  - Making it into information
  - Gaining the knowledge
Background
We have also been struck by the hype!
We had to ask ourselves some questions

- Can the systems provide the data
  - Structure
  - Availability
  - Quality
  - Integrity
- Are we ready to use the data
  - Trust
  - Accept
  - Maturity
Developing the readiness and maturity

- You need to climb the stairs one step at a time

Baseline: What happened?

Step 1: Descriptive
What is happening?

Step 2: Predictive
What will happen?

Step 3: Prescriptive
How should we act?
Competencies and resources

- A lot of focus is on the data scientist role
  ...but you also need data engineers and domain experts
  ...and business solution architects/digital translators to make the wheels turn
- The balance between the roles will change throughout the different phases
Use case
Approach for implementing MI on production line

- **Define & Measure**
  - Pinpoint which business pains to be targeted
  - Make draft of expected visualisation goal
  - Define which data sources are needed for visualisation, **and verify data sources for data structure, availability, quality, integrity etc.**
  - Feasibility study – iterate on benefits vs. efforts.

- **Analyse & Improve**

- **Control (& Iterate)**
FlexTouch® packaging line 47 in Hillerød

- Needed OEE improvement to meet production target
- Had several data sources available
- Geographical location

IT Systems supplying data
- Performance management system
- Manufacturing Execution System
- Material Control System
- Serialisation Line Controller
- Building Management System
Approach for implementing MI on production line

**Define & Measure**

- Make a detailed mapping of data and create the needed master data (e.g. mapping of root cause code)

**Analyse & Improve**

- Document the required data (interface agreement) and request and document needed infrastructure changes (e.g. firewall openings) to get access to data
- Set up data acquisition from data sources (enabling real-time data availability), and if necessary, normalise data either on the edge or in data platform.
- Structure and prepare data for use in visualisations and begin building visualisations and models in the platform

**Control (& Iterate)**
Data analysis

- Manually extracting historical data from the various different systems
- Matching each dataset on batch level
  $\approx 4500$ batches
- Removing batches missing in one or more systems
  $\approx 1200$ batches
- Removing outliers (test-batches, aborted batches etc.)
  $\approx 650$ batches
  ...with 184 variables

Data analysis
Data analysis

PCA: Significant variables
(Principal component analysis)

AI model

Temperature

Pens/carton

Day of year

Batch size

Predicted pens/min.
Batch progress

Classical OEE target

Output prediction from AI model

Actual output
Approach for implementing MI on production line

Define & Measure

Ensure sustained use of solution, dashboards and related tools through effective user training and support

Analyse & Improve

Continuously improve solution via enhancements (e.g. ideas from users) or bug fixes

Evaluate maturity and readiness to move to the next step from descriptive analytics to predictive analytics or from predictive analytics to prescriptive analytics

Control (& Iterate)
Many more iterations to make

- We have now been running with the model for six months and have both the input and output data for this period. Should we re-train the model using the new data? ...or do we risk introducing bias when not having data for a full season?
- Get feedback from production and adjust the visualisations
- Build a standard solution and roll out to other lines
Unexplored areas

Correlation with PENS_PER_MINUTE

- We seem to have two groups of SKUs behaving differently
- Is there something we have not yet included in the analysis?

...or is it just a correlation linking back to one of the already identified causes?
Learnings and take-aways
Understanding data processing consumption in the Amazon cloud

- What is it that we do, to cause these spikes in data processing?
- Move processing from the cloud to the edge devices

![Graph showing data processing consumption](image)
- Total per range: 1.50 PB
- Avg per month: 249.38 TB
- Avg last week: 13.16 TB
- Avg per day: 8.31 TB
Data Architecture Principles

Guiding Principles for when connecting data between data source and data lake

**Principle 1:** Compute at lowest possible level

**Principle 2:** Cleaned, elevated and fit for purpose

**Principle 3:** Entity values are time stamped when they happen

**Principle 4:** Consider sample frequency carefully

**Principle 5:** Apply relational tagging of data entities

**Principle 6:** Associate data points to physical assets and group them

**Principle 7:** Define master processes

**Principle 8:** Storing is faster than calculating

**Principle 9:** Store JSON strings for multiple value entries at same timestamp

**Principle 10:** Remove correlating sensors
Take-aways

• One maturity step at a time

• Learn from pilots before scaling up

• Create Data Architecture Principles