Taking the leap: Expanding your Reliability engineering workfield

Dorien Lutgendorf

Reliability Seminar 2021

holland innovative

Why take the leap?
Data scientists need you

Data engineers, analysts, and scientists

Data engineer

- Databases and data warehousing systems
- Data pipelines (ETL)
- Big Data storage and computing solutions

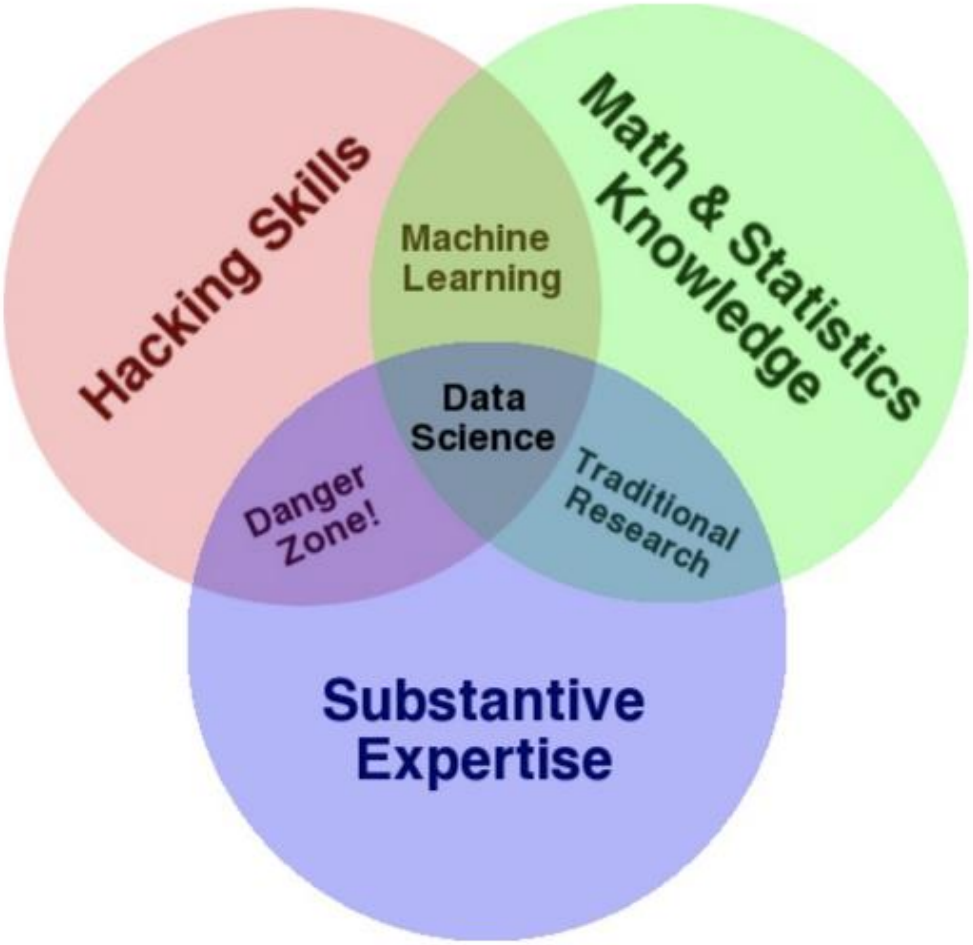

Data analyst

- Explore data (data mining)
- Build models and algorithms
- Visualizations (dashboards)

Data scientist

- Translate business problems into DS problems
- Translate data and analytics into business value
- Manage DS-project & stakeholders, deliver and deploy results, get things done in large and complex organizations

Data science and reliability



Venn diagram by Drew Conway

To predict or to explain?

$Y = f(X)$

Causes → Effects

	Batch no	Date	Origin	Raw Caf	Treatm	Moisture	Heattime	Soaktime	ExTime	Temperature	Cycle time	Caffeine%
		X	X	X	X	X	X	X	X	X	Y	Y
1	C160032	05-Apr-16	Uga	3.47	0	12.6%	75	40	490	59.7	605	0.126
2	C160065	05-Apr-16	Bra	3.89	0	10.2%	75	40	388	56.7	503	0.102
3	C160006	05-Apr-16	Uga	3.51						60.2	551	0.117
4	C160123	05-Apr-16	Ind	2.59						60.5	466	0.105
5	C160062	05-Apr-16	Ind	2.81	0	11.8%	75	40	489	61.4	604	0.118

Causal / explanatory model

$X \rightarrow \square \rightarrow Y$

Predictors → Response

	Batch no	Date	Origin	Raw Caf	Treatm	Moisture	Heattime	Soaktime	ExTime	Temperature	Cycle time	Caffeine%
		X	X	X	X	X	X	X	X	X	Y	Y
1	C160032	05-Apr-16	Uga	3.47	0	12.6%	75	40	490	59.7	605	0.126
2	C160065	05-Apr-16	Bra	3.89	0	10.2%	75	40	388	56.7	503	0.102
3	C160006	05-Apr-16	Uga	3.51						60.2	551	0.117
4	C160123	05-Apr-16	Ind	2.59						60.5	466	0.105
5	C160062	05-Apr-16	Ind	2.81	0	11.8%	75	40	489	61.4	604	0.118

Predictive model

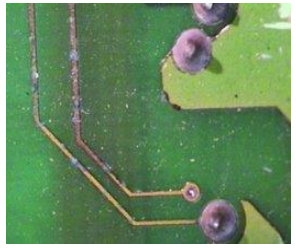
Join forces

Smart Maintenance (predictive algorithms)

Traditional reliability methods (physics-of-failure modelling)

Two main advantages:

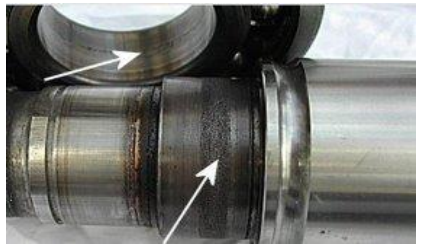
- By understanding the failure mechanism the failure may be reduced or degradation delayed by redesigning.
- Some types of failures are inherently unpredictable and predictive maintenance (based on preventive replacement) is unlikely to be an effective maintenance strategy



Corrosion

Bolted Joint

Welded Joint



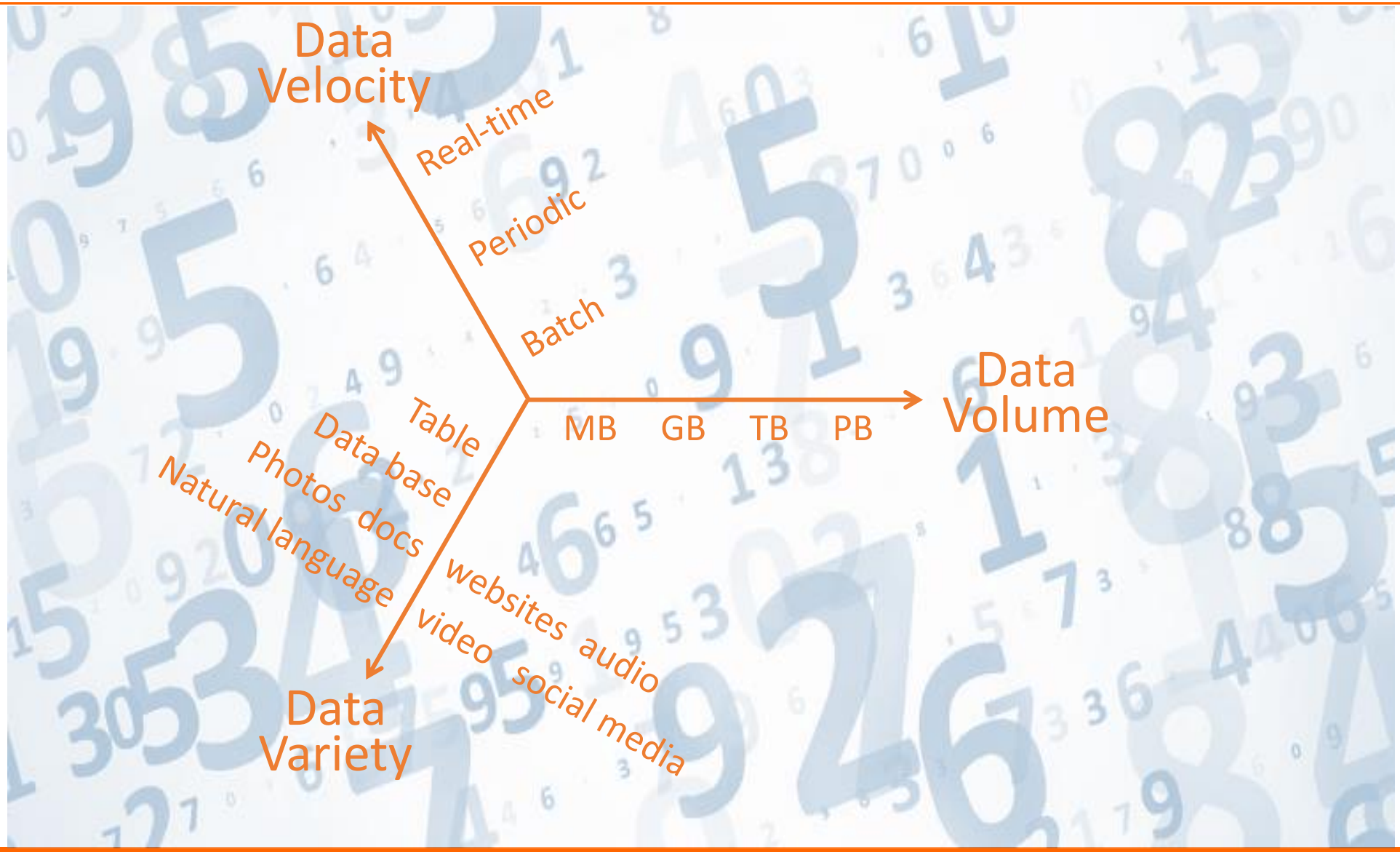
Fretting Wear

Why take the leap?

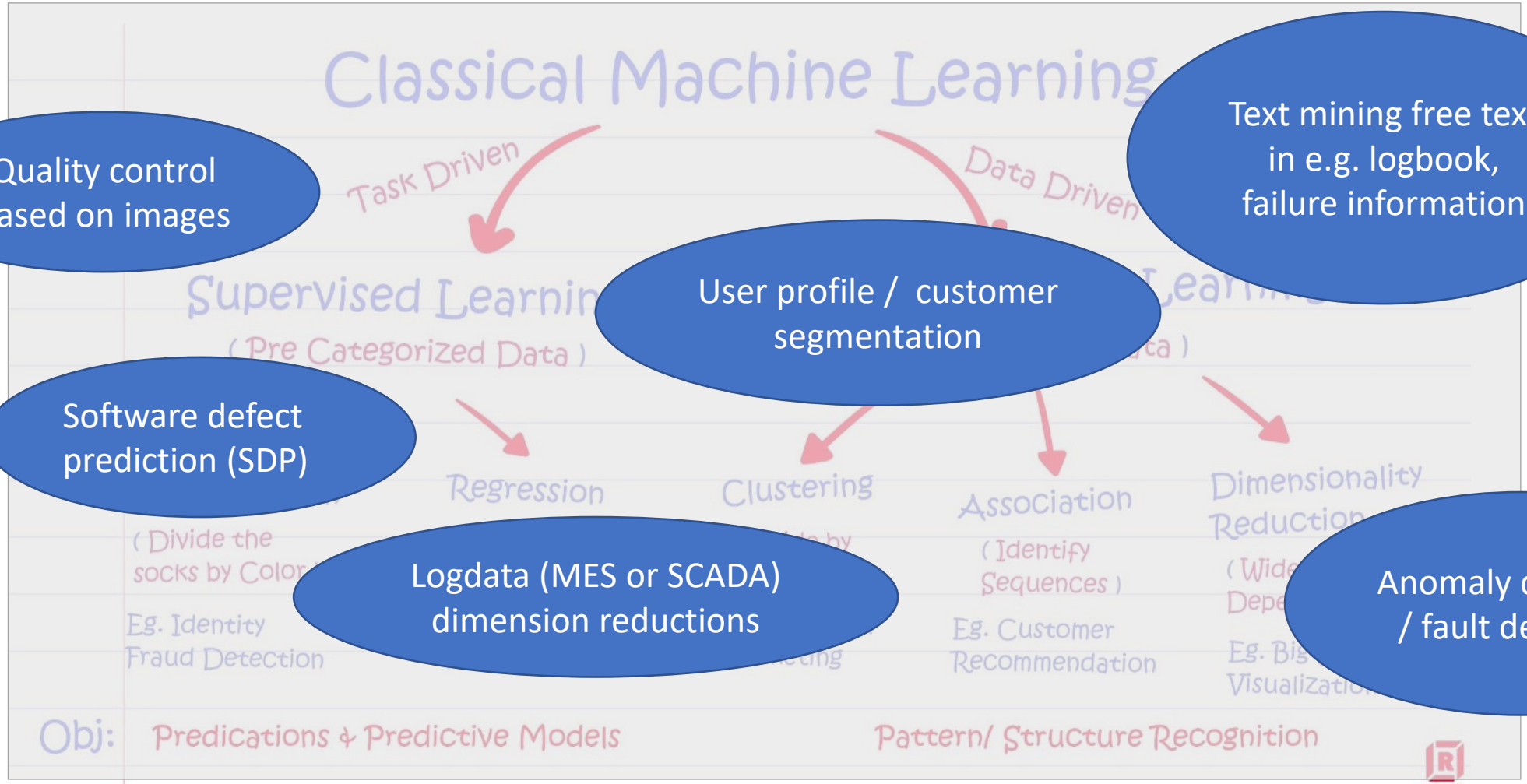
World of algorithms

New forms of data …

Order	Items	Pick-time
1	18	11.6
2	11	9.5
3	7	6.0
4	7	4.7
5	22	15.0
6	5	7.1
7	9	5.4
8	5	9.3
9	11	9.0
10	24	17.6
11	12	10.9
12	8	8.8
13	15	12.3
14	20	16.6
15	7	8.9
16	21	17.2
17	20	15.0
18	13	10.6
19	22	13.2
20	10	10.1
21	19	10.8
22	9	10.2
23	10	9.9
24	2	6.1
25	5	3.3
26	2	5.1
27	4	7.0
28	24	18.5
29	3	7.1
30	2	3.5



Model types



Quality control based on images

Text mining free text in e.g. logbook, failure information

User profile / customer segmentation

Software defect prediction (SDP)

Logdata (MES or SCADA) dimension reductions

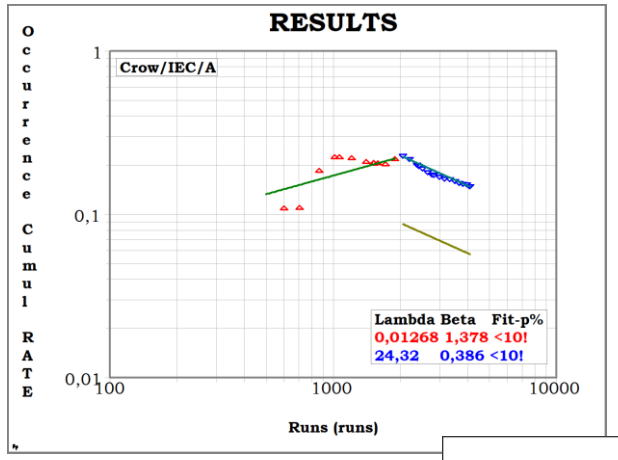
Anomaly detection / fault detection

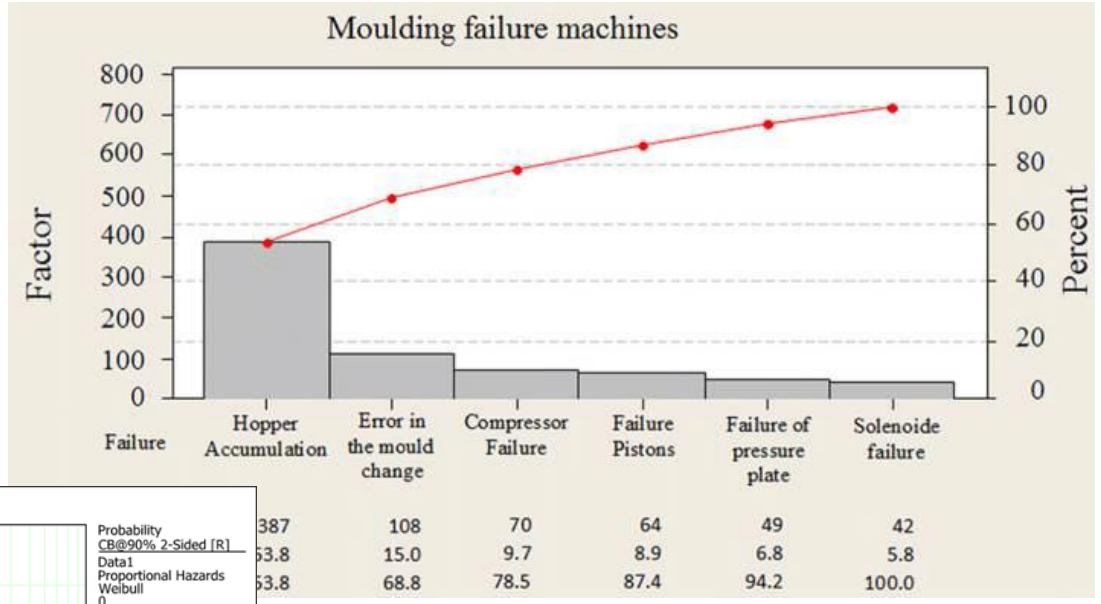
- https://medium.com/@recrosoft.io/supervised-vs-unsupervised-learning-key-differences-cdd46206cdcb

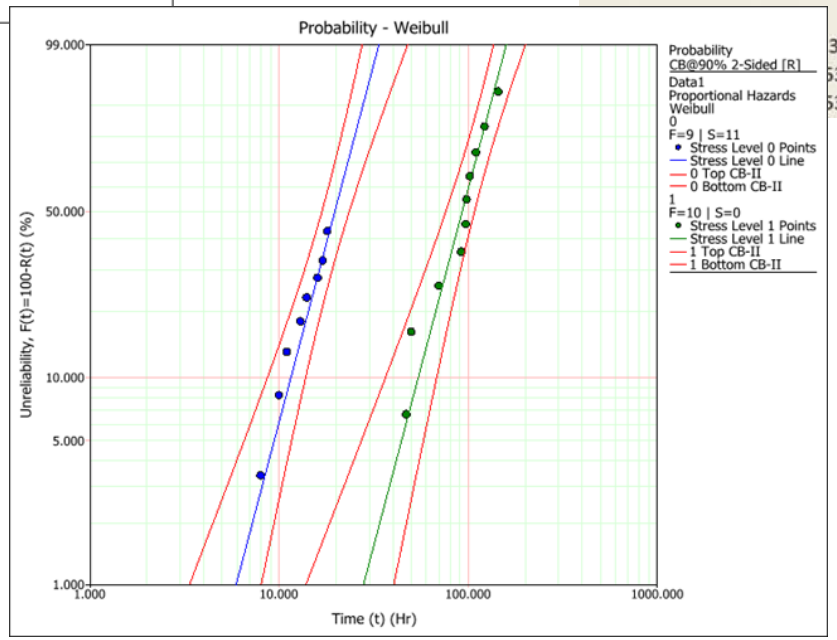
Why take the leap?

Work efficiency

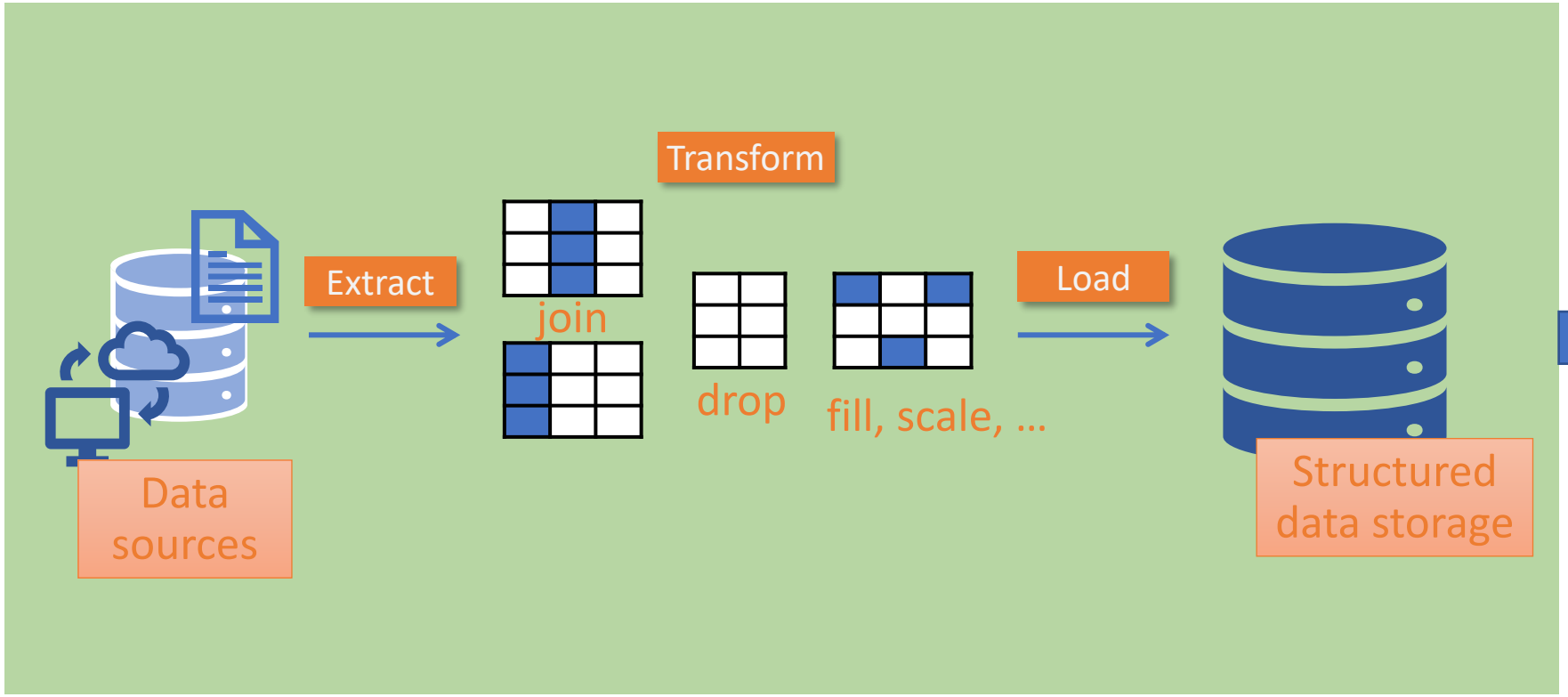
ETL Pipeline

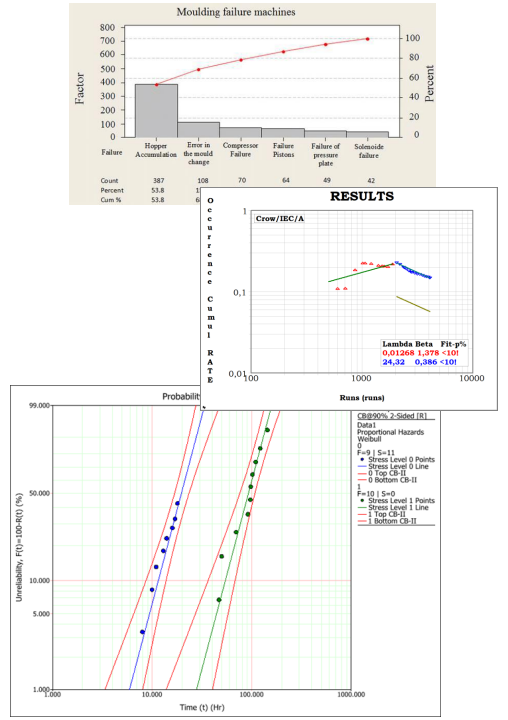




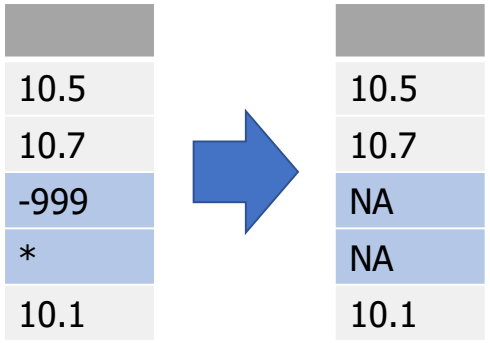


ETL Pipeline

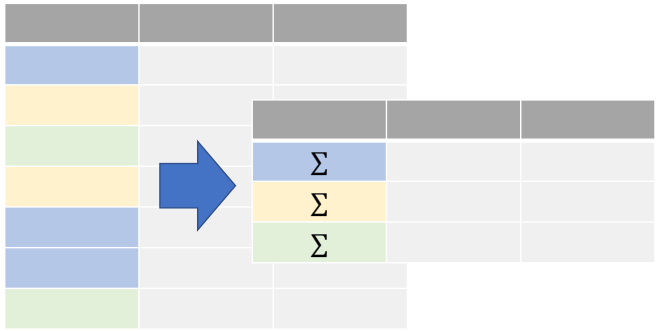




Data wrangling …



Validate and recode values, handle missing values



Group and aggregate data

X_1	X_2
A	1
B	2
C	3

X_1	X_2
A	T
B	F
D	T

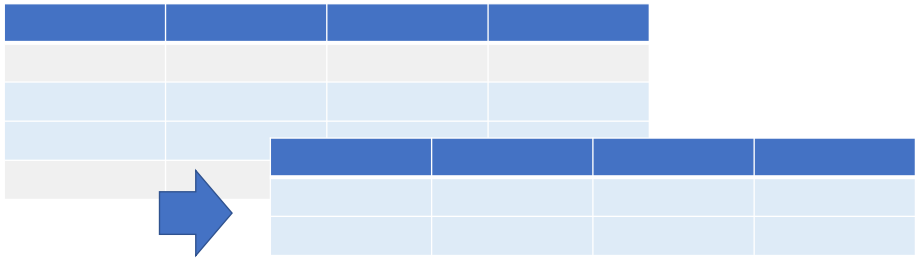
X_1	X_2	X_3
A	1	T
B	2	F
C	3	NA
D	NA	T

Combine data from multiple sources

Stack columns

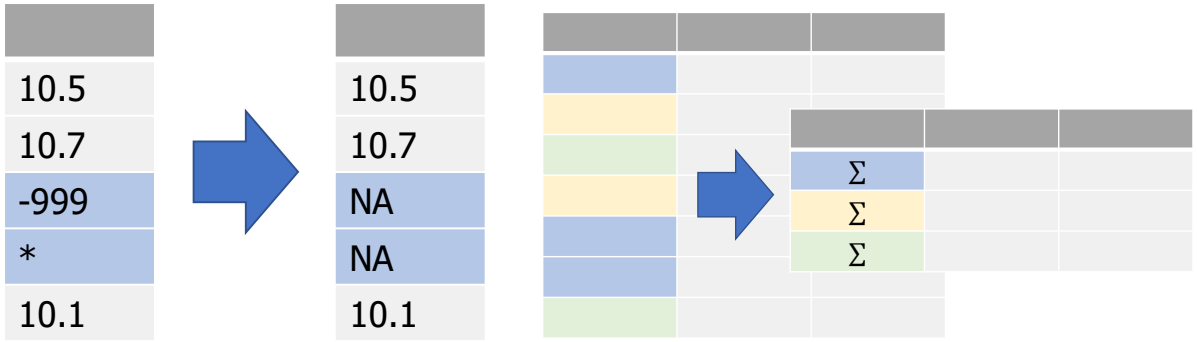
x_1	y_1	$f(x_1, y_1)$
x_2	y_2	$f(x_2, y_2)$
x_3	y_3	$f(x_3, y_3)$

Compute variables from raw data



Filter data (subsets)

Data wrangling …



X_1	X_2
A	1
B	2
C	3

X_1	X_2
A	T
B	F
D	T

X_1	X_2	X_3
A	1	T
B	2	F
C	3	NA
D	NA	T

Validate and recode values, handle missing values

Group and aggregate data

Combine data from multiple sources

```
data.replace('*', 'NaN', inplace=True)
data.replace(-999, 'NaN', inplace=True)

data.dropna(axis=0, how='any', inplace=True)
```

```
installs = pd.read_excel('Sales_data.xls')
failures = pd.read_excel('Failure_data.xls')

data = failures.join(installs, on='Serial number', how='left')
```

```
data_agg = data.groupby(['Product ID'])['Errors'].sum()
```

Data wrangling …

```
stacked = data.stack()
```

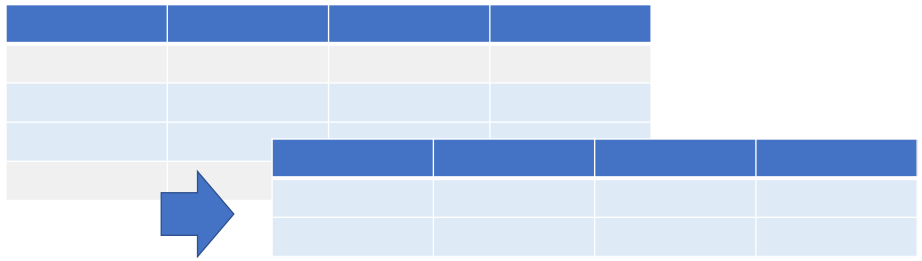
```
product 'type A'
date = '1-1-2021'

datafiltered = data[(data['Product']==product]) & (data['Install date']>date])
```

```
data['ttf [hrs]'] = (data['Fail date'] - data['Start date']) * 24
```

Stack columns

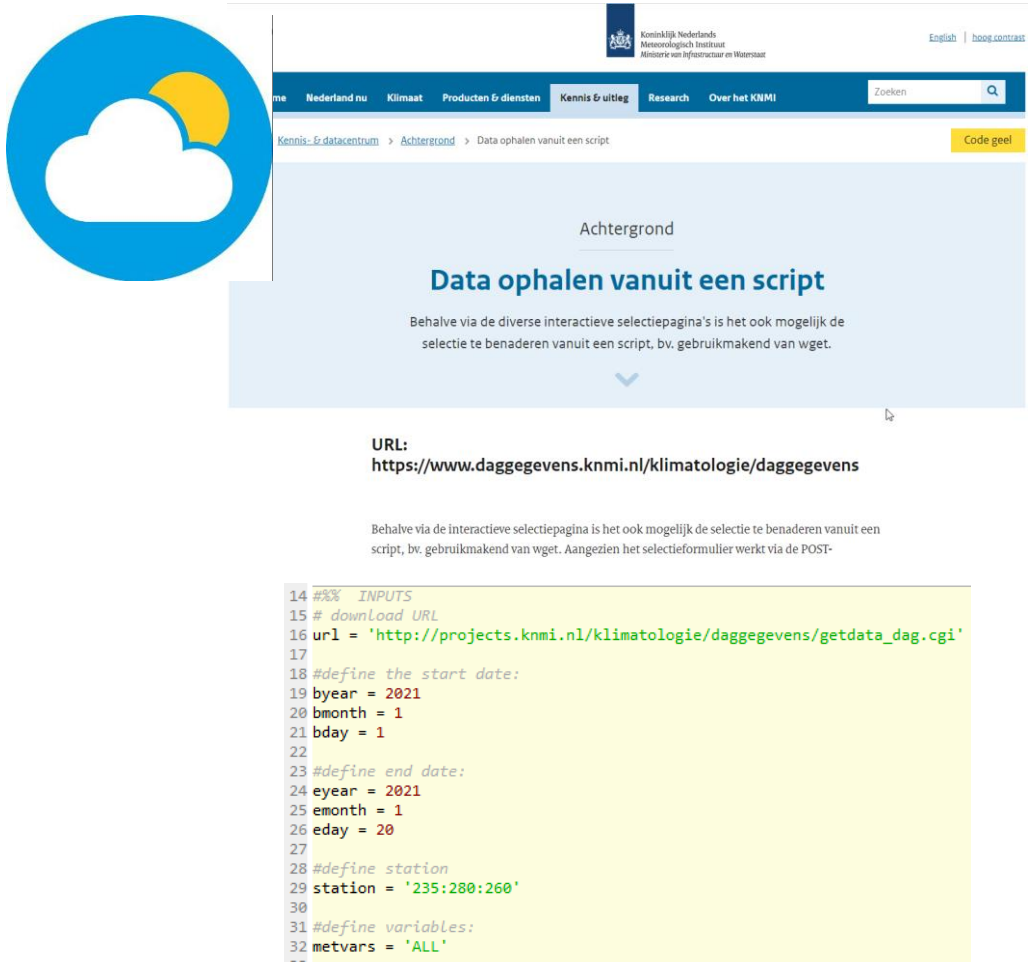
x_1	y_1	$f(x_1, y_1)$
x_2	y_2	$f(x_2, y_2)$
x_3	y_3	$f(x_3, y_3)$

Compute variables from raw data



Filter data (subsets)

Access data via API or from database




```
#%%  INPUTS
# download URL
url = 'http://projects.knmi.nl/klimatologie/daggegevens/getdata_dag.cgi'

#define the start date:
byear = 2021
bmonth = 1
bday = 1

#define end date:
eyear = 2021
emonth = 1
eday = 20

#define station
station = '235:280:260'

#define variables:
metvars = 'ALL'
```

- https://www.knmi.nl/kennis-en-datacentrum/achtergrond/data-ophalen-vanuit-een-script

Example how to connect to AWS database

```
In [ ]: import pandas as pd
        import psycopg2
```

Login details

```
In [ ]: plant_code = "plantcode"
        process_name = "assembly"
        aurora_host = "hostname"
        db_name = "{0}_{1}".format(plant_code, process_name)
        username = Test123
        password = Test123
        port = "1234"
        hydra_table_name = "table_name"
```

Connect to database

```
In [ ]: aurora_conn = psycopg2.connect(host=aurora_host, dbname=db_name, user=user_name, password=password, port=port)
```

Run SQL query on database

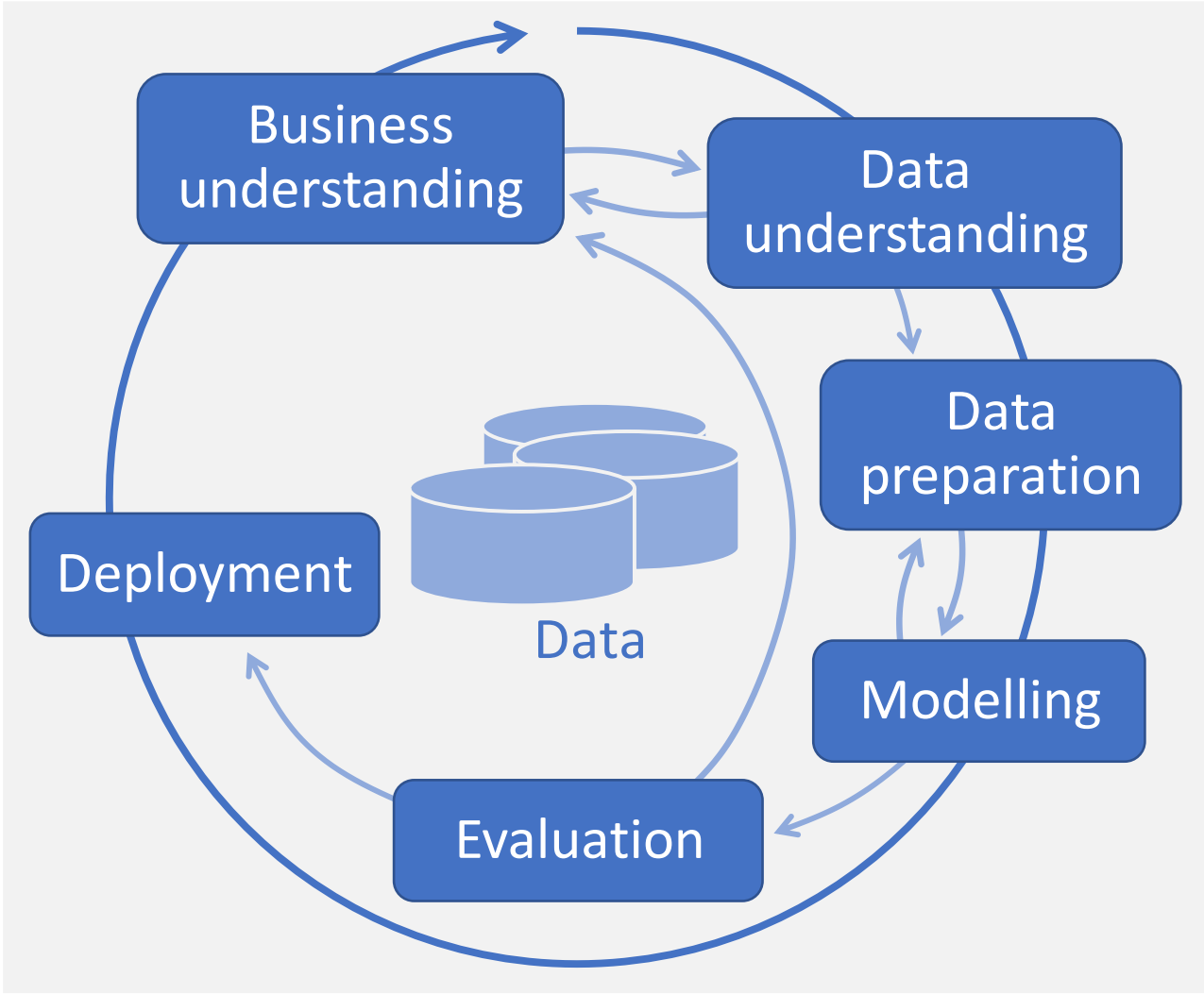
```
In [ ]: df = pd.read_sql("SELECT * FROM {}".format(hydra_table_name)
                        + " WHERE line = \'6\'"
                        + " AND lt_date BETWEEN \'{}\' AND \'{}\'".format(start_time, end_time) , aurora_conn)

In [ ]: df.head()
```

- https://aws.amazon.com/rds/aurora/?aurora-whats-new.sort-by=item.additionalFields.postDateTime&aurora-whats-new.sort-order=desc
- https://towardsdatascience.com/amazon-rds-step-by-step-guide-14f9f3087d28

How to take the leap?

Cultivate your domain knowledge

CRISP-DM Process for Data Mining



Learn (a bit of) scripting

Understand the vocabulary

Tires …

- Pressure
- Temperature
- Wear

X

Impending blowout or flat

Y

$X \rightarrow \rightarrow Y$

Independent variables

Explanatory variables

Physical properties

Attributes

Predictor variables

Features

$X \rightarrow$ Algorithm / Classifier / Model $\rightarrow Y$

Model fitting
Training
Supervised learning

Dependent variable

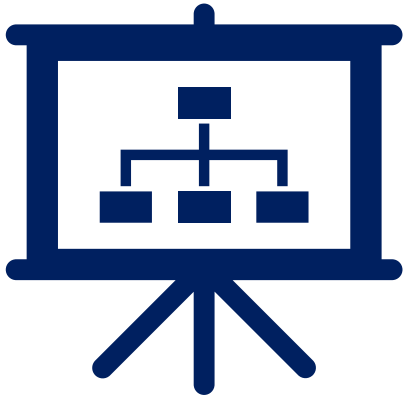
Response variable

Probability of failure

Output variable

Labels

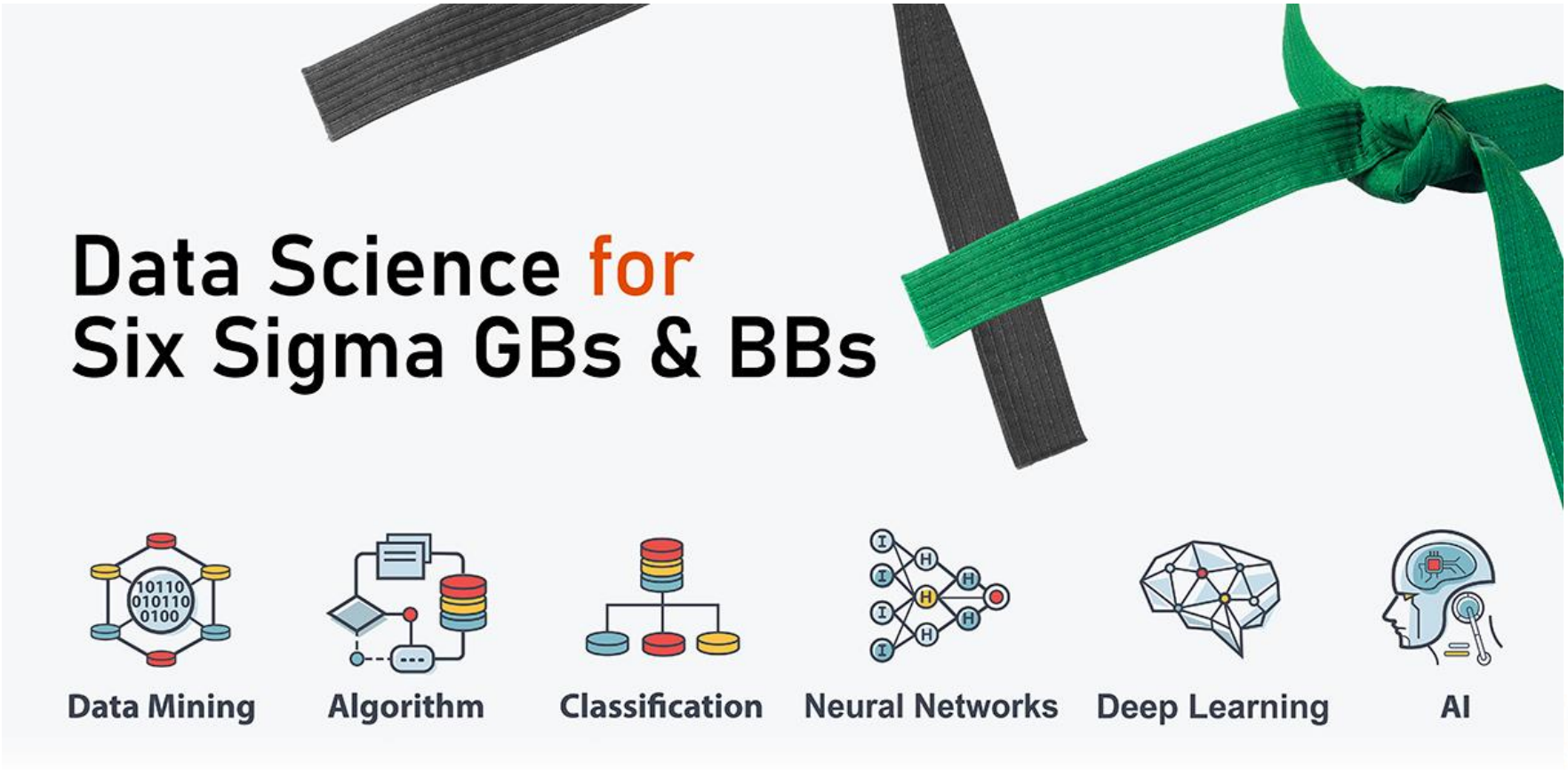
Initiate and join teams



&

...

Jump and enjoy!


Data Science for
Six Sigma GBs & BBs
10110
010110
0100
Data Mining
Algorithm
Classification
Neural Networks
Deep Learning
AI

DS4GBB course

Data science and six sigma have much in common. Both operate on data-driven analysis techniques, offer structures for framing and solving problems, and follow a project-based approach. But also: six sigma is at least 25 years old … data science takes it to the next level. With new analytics, new forms of data, and new opportunities!

The last decades have seen the emergence of a totally new brand of analytics from statistical learning, machine learning and AI. While Six Sigma focuses on optimizing business processes and current product lines (“Horizon 1 innovation”), current industry recognizes data and analytics as valuable assets in themselves, and explores data-driven business models and strategies (“Horizon 3 innovation”).

https://www.holland-innovative.nl/academy/data-science-six-sigma-green-belts-black-belts

Data Science for Six Sigma Greenbelts and Blackbelts

holland innovative

Data science… the next step in Six Sigma? Data science and Six Sigma have much in common. Both operate through data-driven analysis techniques, offer structures for framing and solving problems, and follow a project-based approach. But also: Six Sigma is at least 25 years old…

Data science takes Six Sigma to the next level!

Register at: www.holland-innovative.nl

New analytics, new forms of data, new opportunities …

Six Sigma represents the first generation of computer-aided analysis techniques, such as regression, design of experiments and control charts. Driven by discoveries in mathematics and the tremendous power of modern computers, the last decades have seen the emergence of a totally new brand of analytics from statistical learning, machine learning and AI.

Also, where Six Sigma uses powerful techniques to get the most out of small datasets (say, N=20 to 100 or so), modern IT infrastructures and the IoT, cheap storage and computing capacity, and the resulting huge streams of data enable totally new applications of analytics, where data could also be images, audio or natural language.

While Six Sigma focuses on optimizing business processes and current product lines (“Horizon 1 innovation”), current industry recognizes data and analytics as valuable assets in themselves, and explores data-driven business models and strategies (“Horizon 3 innovation”).

Instructors

Besides his affiliation with HI, **Jeroen de Mast** is a professor at the University of Waterloo and Academic Director at the Jheronimus Academy of Data Science. **Jörg Bewerunge** is a lead data scientist and project manager.

For whom?

Six Sigma, DfSS and Lean Six Sigma green belts and black belts eager to enrich their expertise with machine learning and data science.

Course duration and number of participants: 4 days, minimum 8 participants

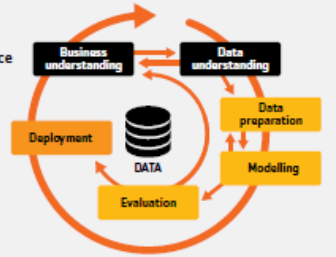
Instructors: Jeroen de Mast, Jörg Bewerunge, Eric Drost

Location and costs: Holland Innovative, High Tech Campus 29, Eindhoven. The costs are € 2.950,- (ex. VAT) per participant, including course materials, drinks and lunch.

Dates, registration and more information: See www.holland-innovative.nl under Academy, where you can also sign up. A registration form can be requested via academy@holland-innovative.nl.

Contact: HI Team Academy, tel. +31 40 85 14 611, academy@holland-innovative.nl

CRISP-DM model for data-science projects



Headquarters
High Tech Campus 29
NL - 5656 AE Eindhoven

T +31 40 85 14 611
E academy@holland-innovative.nl
W www.holland-innovative.nl

Focus on complex business processes

DS4GBB Course content (4 days)

1. Understand data science and machine learning (4u)
2. Learn to work in a data-science analytics environment (4u)
3. Data Engineering 101 (2u)
4. Visualization (6u)
5. Machine learning I - *Regression techniques (6u)*
6. Machine learning II - *Classification techniques (6u)*
7. Text mining and natural language processing (NLP) (3u)
8. Data science, machine learning & Six Sigma (1u)