Semantic Data Mining based decision support for quality assessment in steel industry

Artificial Intelligence in Research and Applications Seminar

Maciej Szelążek

Motivation

Semantic Data Mining based decision support for quality assessment in steel industry

Integration of machine learning approach with ruled-based quality management systems

Semantic compatibility with current support decision systems

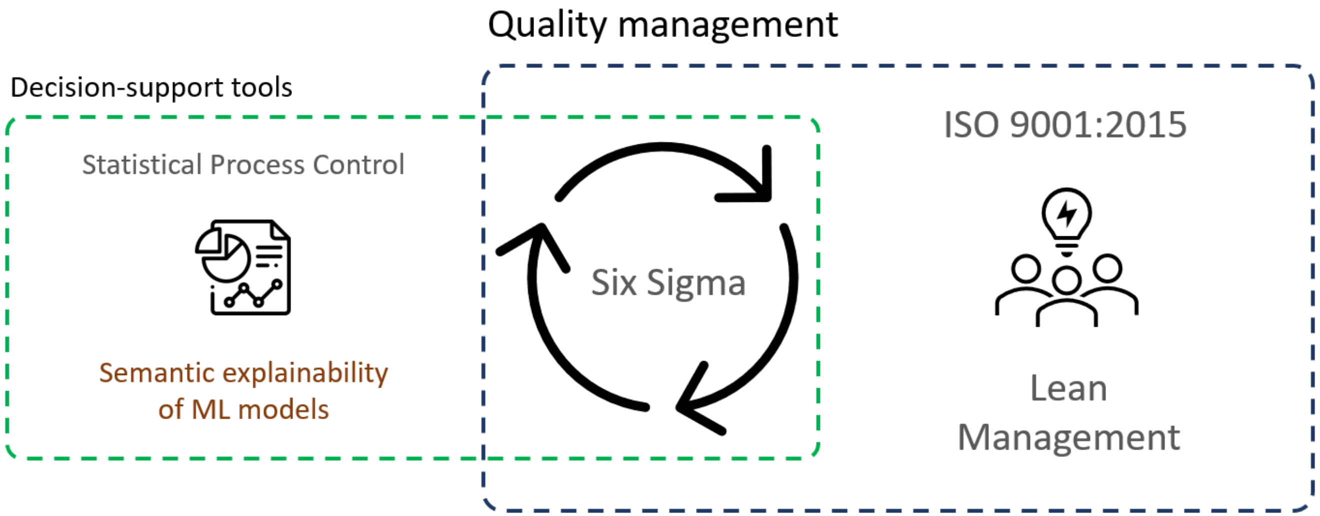
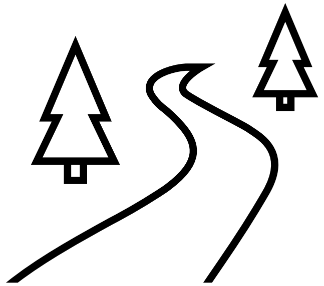


Table of content

Semantic Data Mining based decision support

for quality assessment in steel industry



Quality

ISO 9001:2015

- The totality of features and characteristics of product or service that bear on its ability to satisfy stated or implied needs revolving around the customer.

British Defence Industries Quality Assurance Panel

- Quality is conformance to specifications.

Degree of preference

- It is the degree to which a particular product is preferred over competing products of similar grade, supported comparative test by customers, normally called as customer's preference.

Manufacturing-based definition

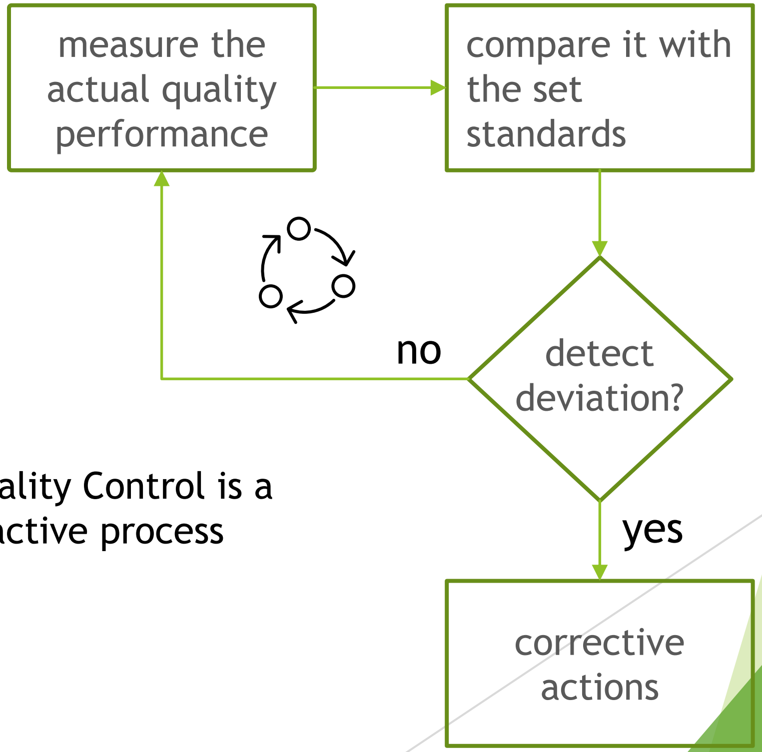
- Quality of a product, means conformance to customer's requirements

Degree of Excellence

- It is a measure of a degree of excellence at a suitable price and control of variability at a suitable cost. this is often a Value-based definition.

Quality Control is a set of activities for ensuring quality starting from raw material to end product.

- devices
- tools
- skills

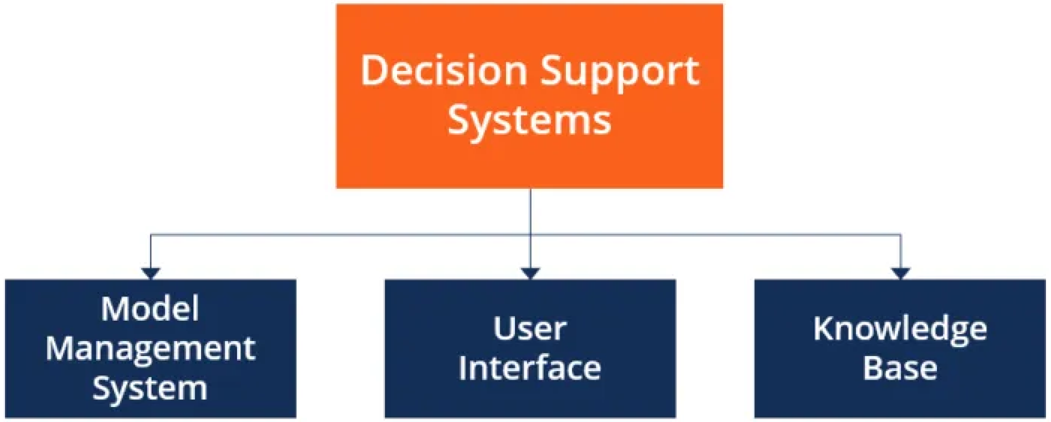


Quality Control is a reactive process

https://techqualitypedia.com/quality/

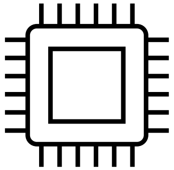
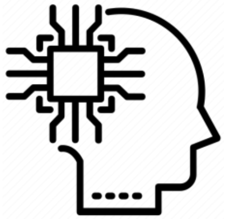
Decision support system

A decision support system produces detailed information reports by gathering and analyzing data.



A **decision support system** (DSS) information system that aids a business in decision-making activities that require judgment, determination, and a sequence of actions.

A DSS is either human-powered, automated, or a combination of both.

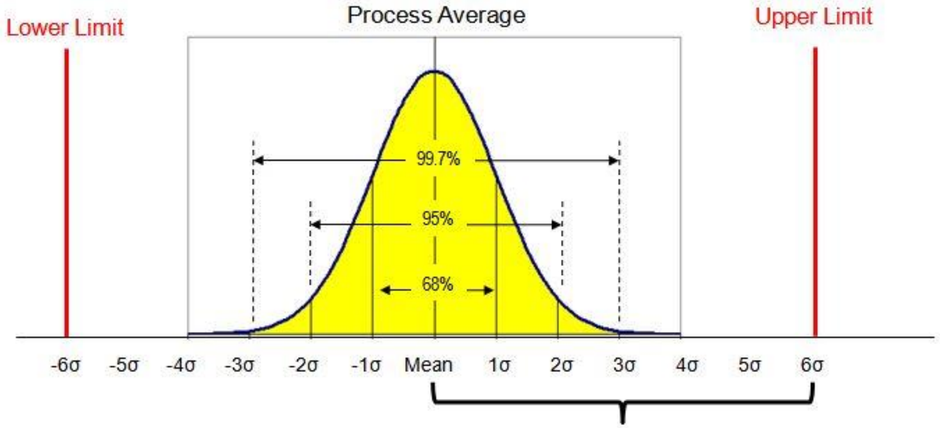
https://corporatefinanceinstitute.com/resources/knowledge/other/decision-support-s

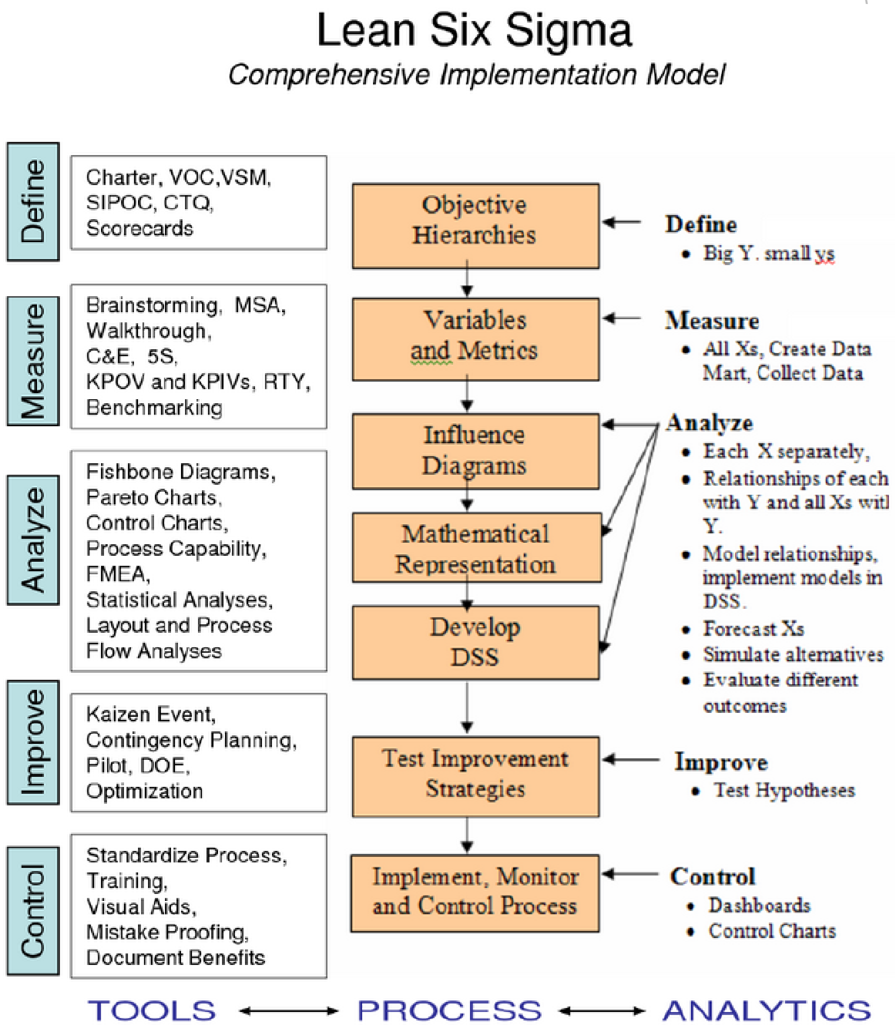
ISO 9001:2015 norm

International standard that specifies the requirements for companies quality management systems

Six Sigma

Well-tested set of instruments and techniques aimed at reducing the variability (tolerance) and defectiveness of a product/process

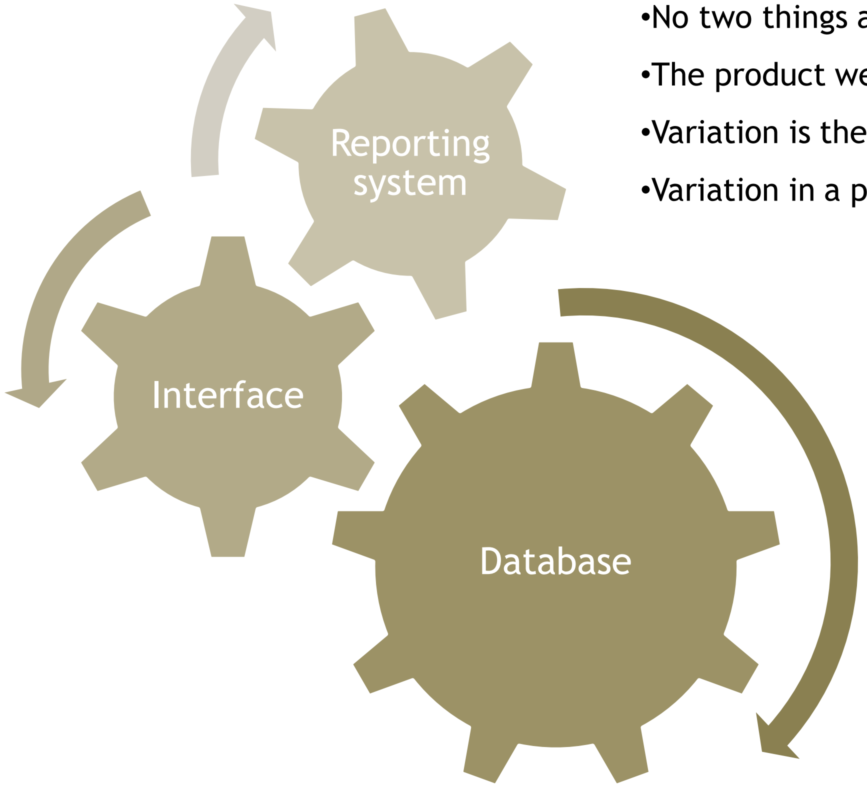




https://www.researchgate.net/publication/333648011_Quality_a_Key_Value_Driver_in_Value_Based_Management
https://lablean.blogspot.com/2015/03/dmaic-applied-to-biopmanufacturing.html

Statistical Process Control

Statistical Process Control is a process of finding and measuring **variability** in the manufacturing process.



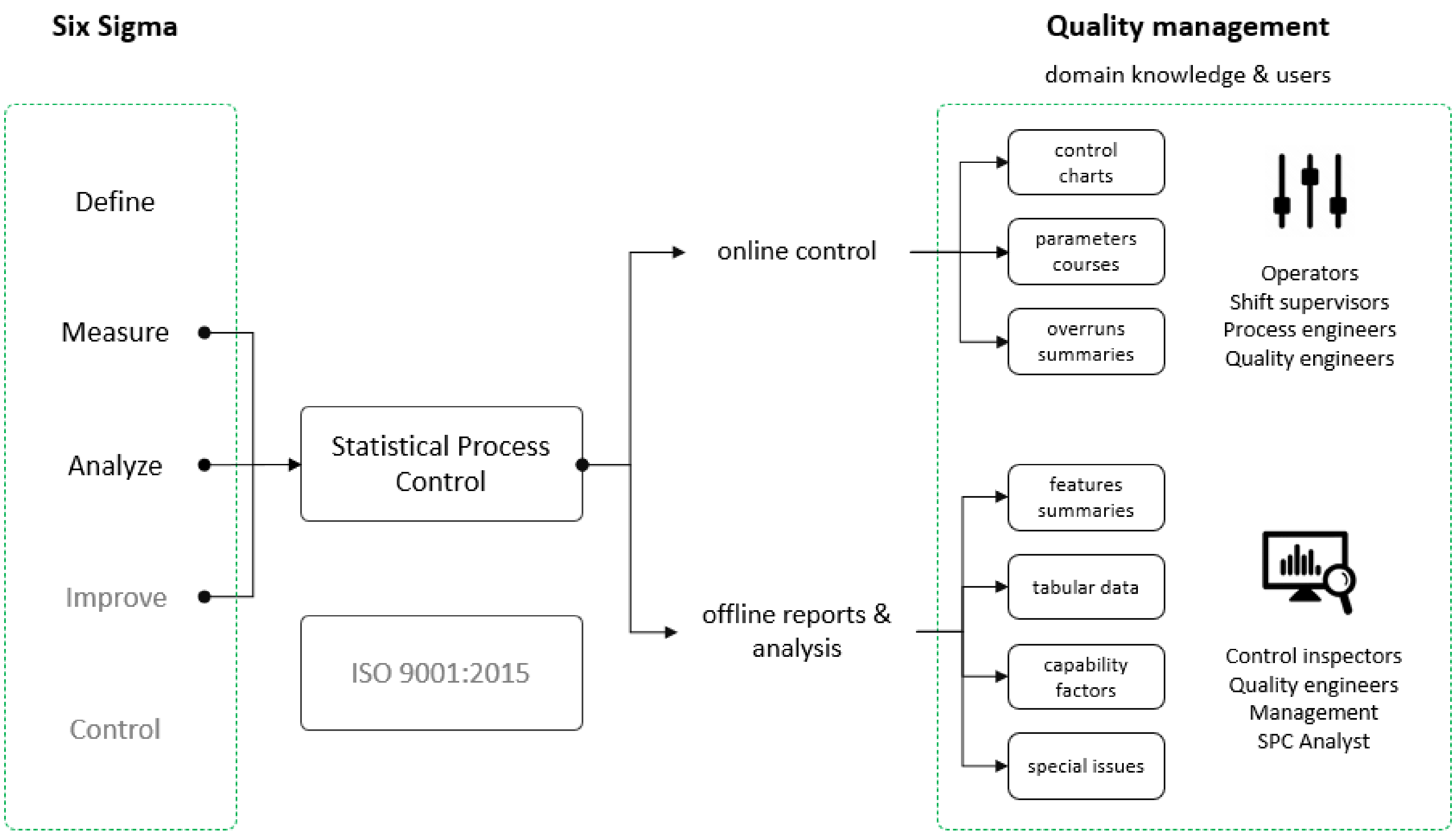
Principles:

- No two things are alike due to variability.
- The product we produce changes continuously.
- Variation is the law of nature and exists in all processes.
- Variation in a product and process can be always measured.

Sources of knowledge:

- Product specification
- Customers order
- Technology constraints

SPC vs Six Sigma

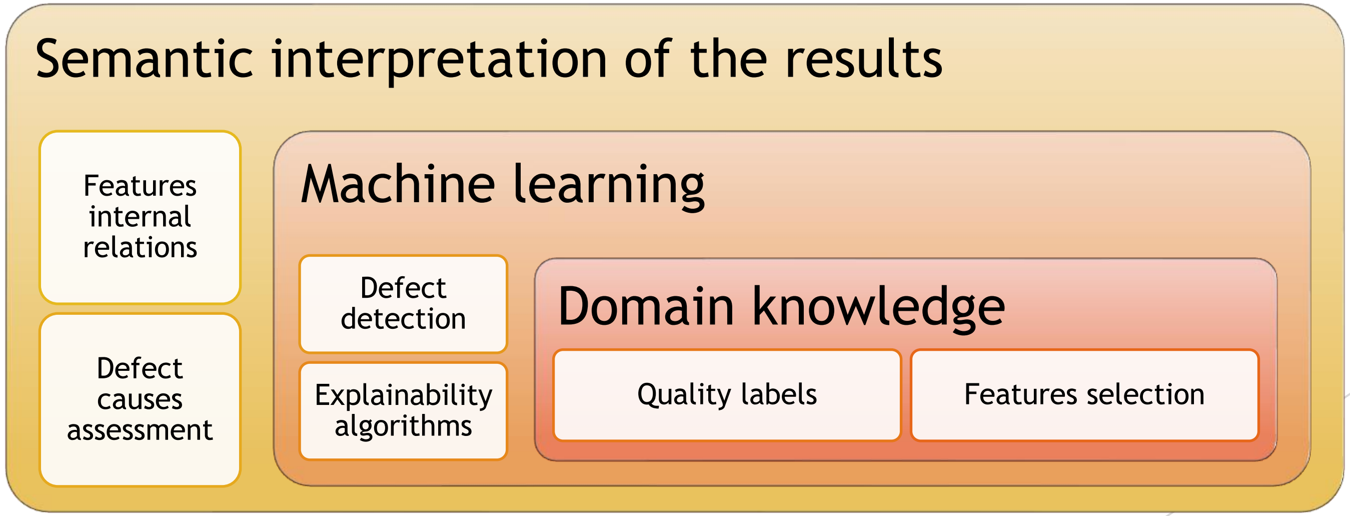


Semantic Data Mining

Data mining is the proces of uncover novel, interesting and understandable patterns related to valuable, useful, and implicit knowledge.

Semantic Data Mining is a knowledge-based analysis approach.

Focused on exploiting the formalized information in order to enhance interpretability of the applied data mining models



Hot Rolling Mill process

Start date - 2019-08-01 00:00:56
End date -   2019-08-31 23:57:18
Products amount - 9551 coils
Parameters amount - 394
Process mesurements amount - 463 401 627

Process nodes:

Walking beam furnace

Single pass per product
Product - hot slab
Single value measurements
Main parameters - charge/discharge temperatures, heating time

⇩

Roughing mill (RM)

5 or 7 passes per product
Product - transfer bar
Continuous measurements + single value
Main parameters - temperature, speed, forces, width, thickness
~9000 records for all passes per product

⇩

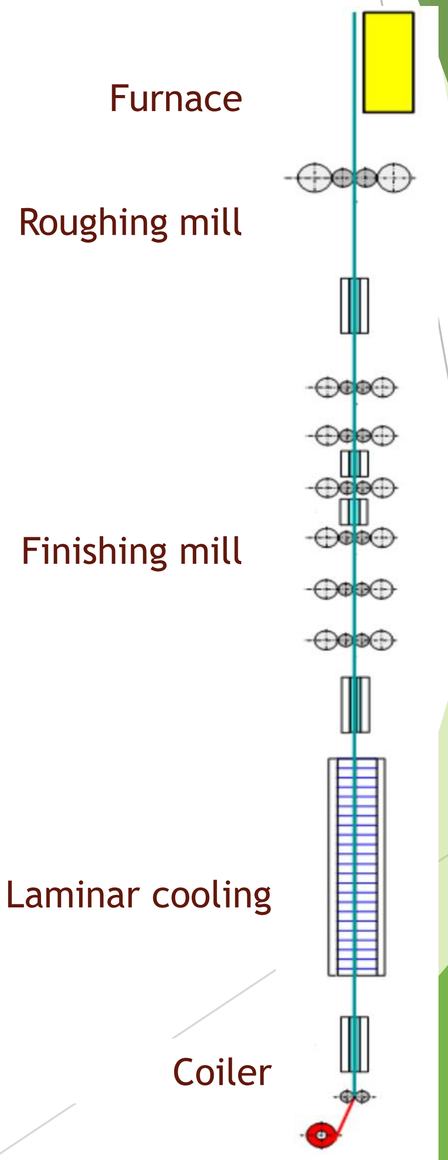
Finishing mill (FM)

Single pass per product
Product - strip
Continuous measurements + single value
Main parameters - temperature, dimensions, forces, rolling mills setups
~31 200 records per product

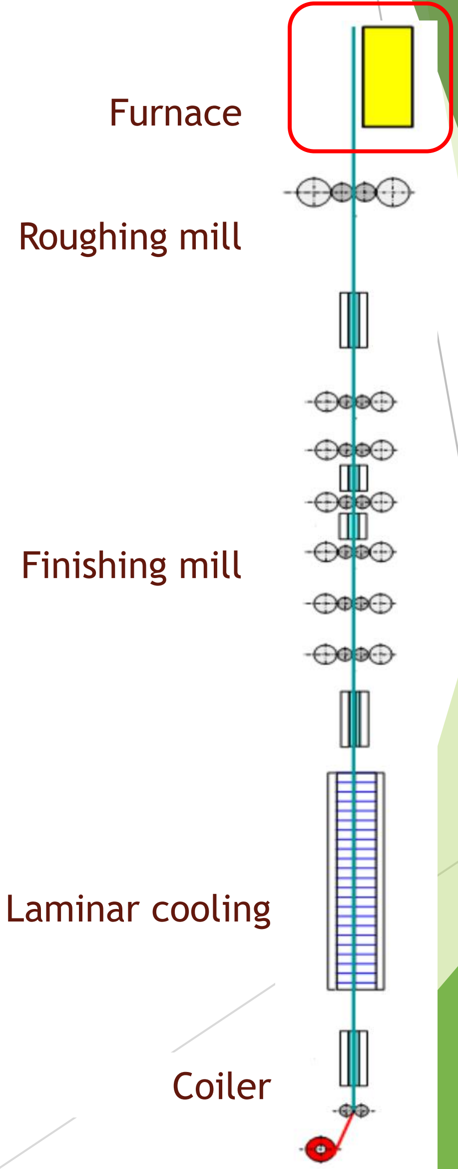
⇩

Laminar cooling + Coiler

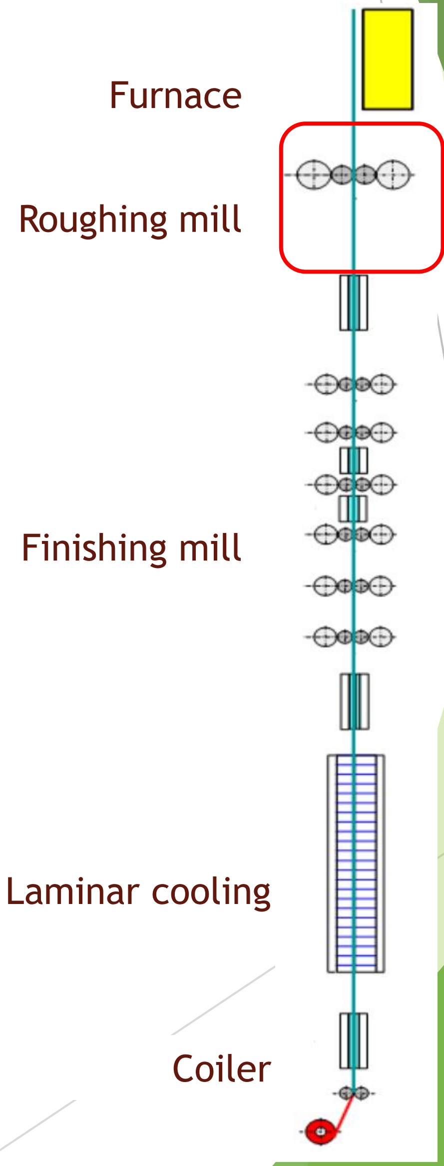
Single pass per product
Product - coil
Continuous measurements + single value
Main parameter - temperatures, water amount



Furnace

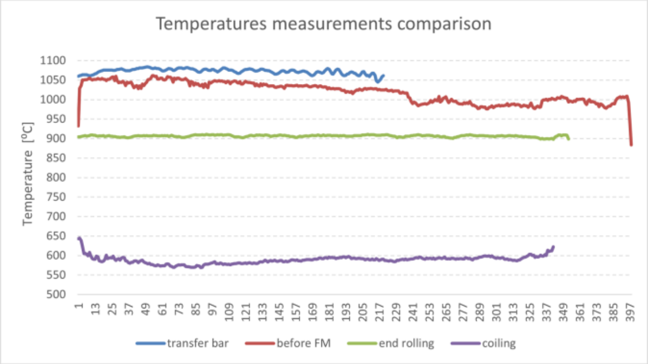

Furnace
Roughing mill
Finishing mill
Laminar cooling
Coiler

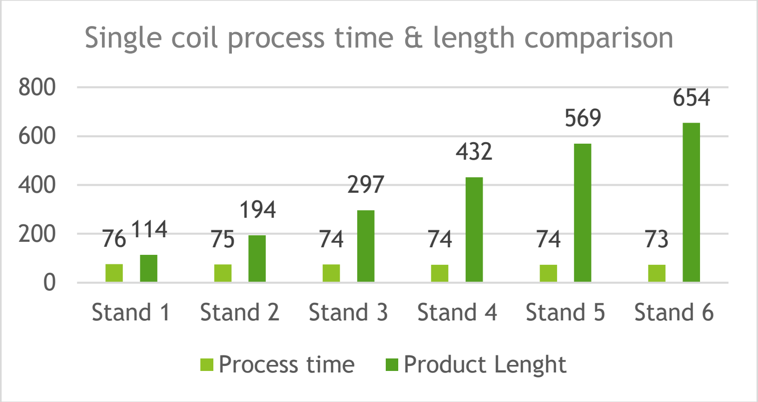
Roughing mill

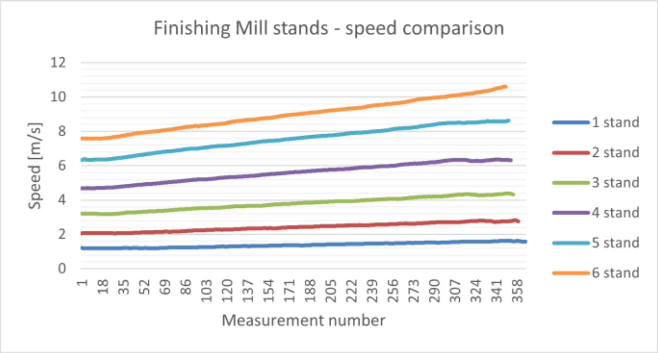


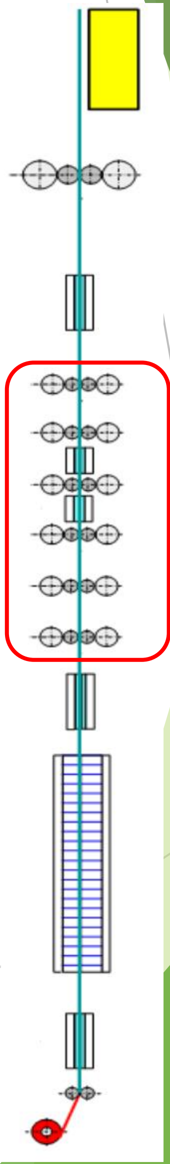
Finishing Mill device

- 100 s of rolling
- Over 10 times thickness reduction
- 3.5 min whole process
- 4 mln tonnes/year production capability



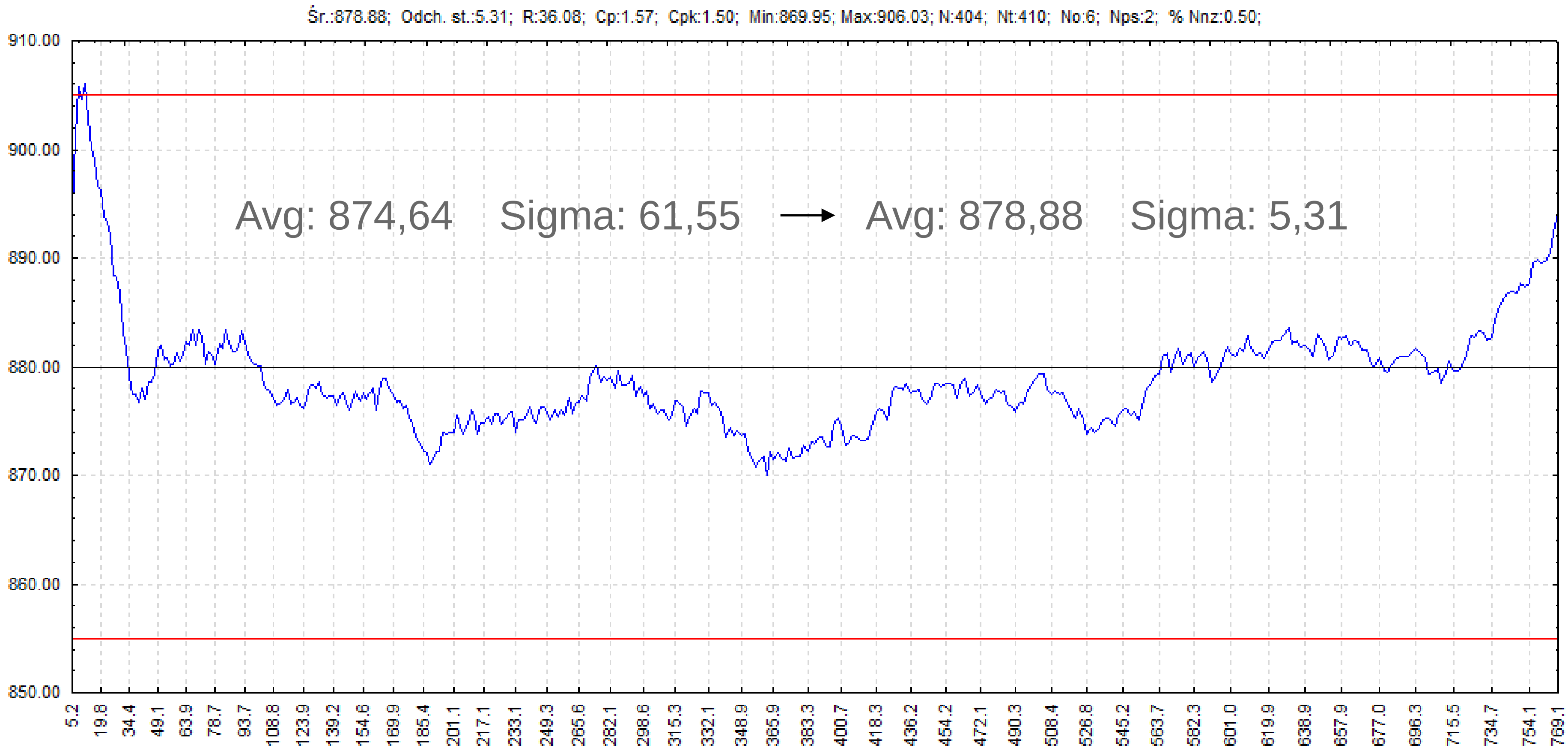






Capability factors

Data cleaning:
- Sensor response time flag
- Sensor default value flag



Distribution after data cleaning

Capability factors

Process capability - Cp : tolerance width divided by the total spread of process (6 Sigma).

Calculation of Process Capability (Cp) :

$$Cp = \frac{\text{Design Tolerance}}{6\sigma} = \frac{USL - LSL}{6\sigma}$$

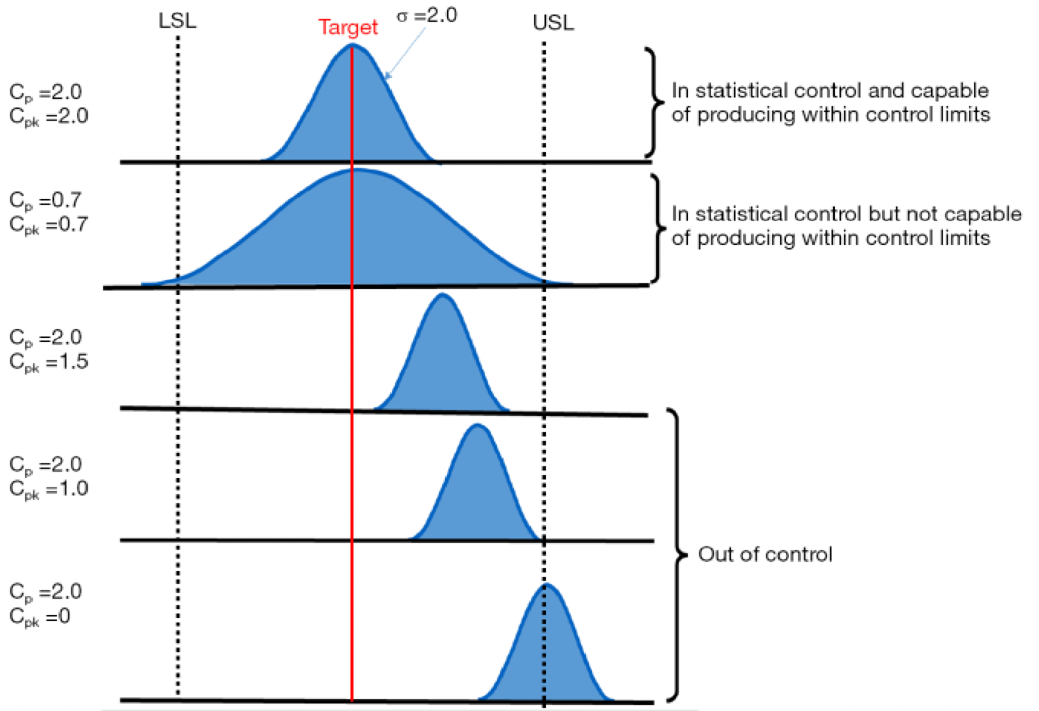
USL = Upper Specification Limit, **LSL** = Lower Specification Limit

Cp	Defects amount
1	2700 ppm
1,33	63 ppm
1,67	0,57 ppm
2	0,002 ppm

Process Capability Index - Cpk: indicates shifting of the process, the minimum of Cpk upper and Cpk lower.

Calculation of Process Capability Index (Cpk) :

$$Cpk_U = \frac{USL - \bar{\bar{X}}}{3\sigma} \quad \text{and} \quad Cpk_L = \frac{\bar{\bar{X}} - LSL}{3\sigma}$$



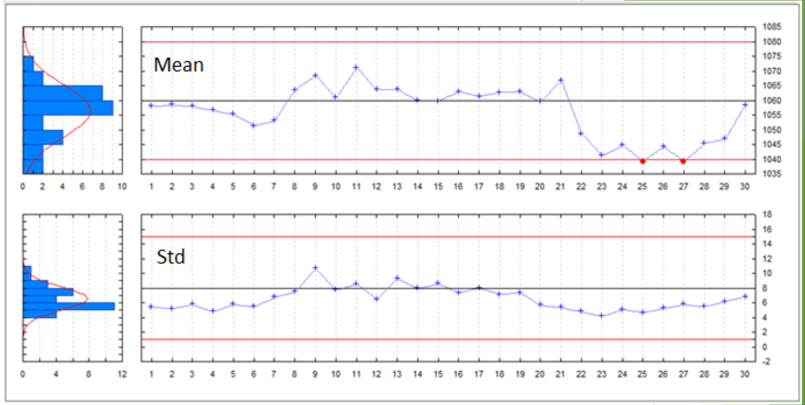
https://www.doriane-copar.com/process-capability-index-cpk/

https://techqualitypedia.com/cp-and-cpk/

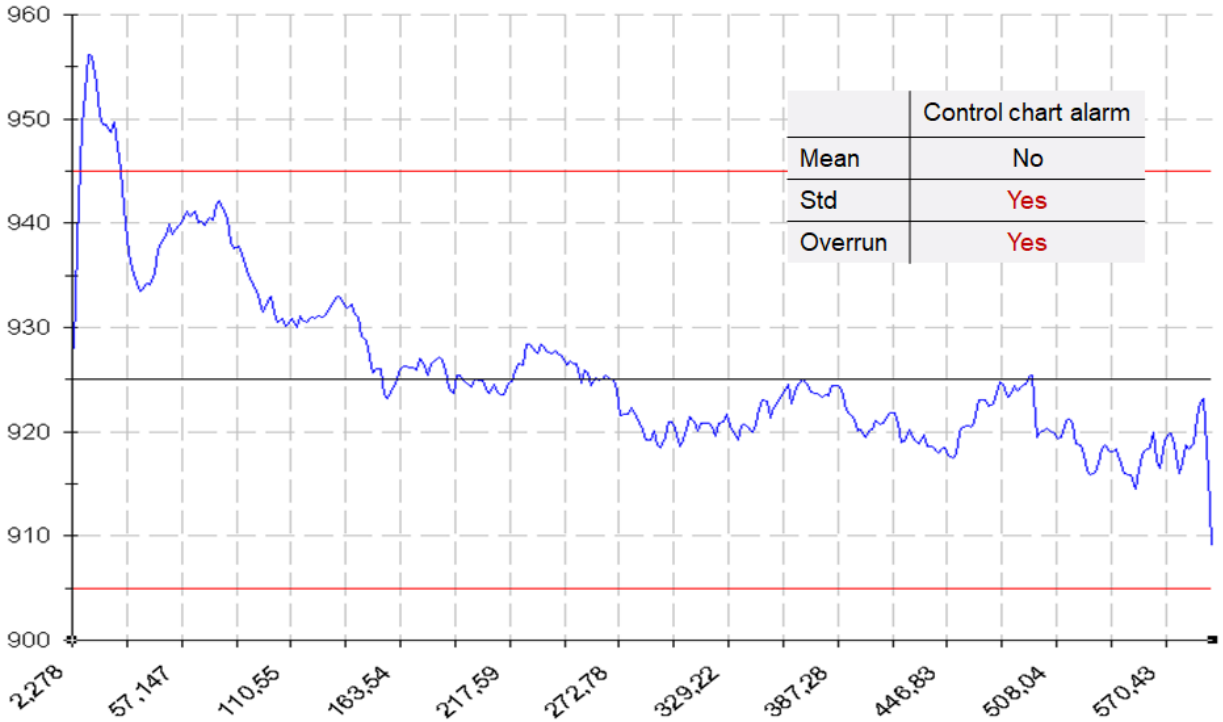
SPC decision support

3 scopes of information
- Single feature, single product, all measurements (data series)
- Single feature, two statistical factors, many products (single value)
- Many features, many products, single statistical factor

Control chart



Single coil temperature distribution



	Control chart alarm
Mean	No
Std	Yes
Overrun	Yes

Summary matrix

KRĘGA	N168041	N168042	N168043	N168044	N168045	N168046	N168047	N168048	N168049	N168050	N168051	N168052	N168053	N168054	N168055
pasma (za FM)	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
pasma (za FM)	9,6	7,3	9,8	18,4	12,0	31,4	46,6	70,9	26,4	52,1	12,3	9,8	40,1	0,0	0,0
pasma (za FM)	0,0	0,0	0,0	0,0	2,5	0,0	24,4	2,5	1,7	0,0	0,0	0,0	0,0	0,0	0,0
pasma (za FM)	0,2	0,0	0,0	0,0	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,2	4,3	0,5
ica walcowania	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,0	0,0
eratura zwijania	7,1	3,8	6,7	7,6	5,0	8,7	3,3	9,9	1,4	4,3	4,3	8,2	1,2	7,9	2,1
pasma (za FM)	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

KRĘGA	N168056	N168057	N168058	N168059	N168060	N168061	N168062	N168063	N168064	N168065	N168066	N168067	N168068	N168069	N168070
pasma (za FM)	0,0	0,0	7,3	1,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
pasma (za FM)	2,4	0,0	100,0	3,9	3,7	21,5	16,8	59,5	33,0	25,3	14,0	10,8	67,6	79,6	30,9
pasma (za FM)	0,0	0,0	92,4	8,5	0,0	0,0	0,0	100,0	91,1	2,5	0,0	0,0	5,4	13,1	34,1
pasma (za FM)	0,2	0,0	0,4	0,0	0,0	0,0	0,0	0,5	0,0	0,0	0,0	0,8	1,0	0,5	0,6
ica walcowania	0,0	0,2	0,0	0,0	0,0	0,0	0,0	2,8	2,2	1,4	0,0	0,5	0,0	0,0	0,0
eratura zwijania	9,5	5,5	6,1	3,1	2,3	5,8	1,9	8,8	3,8	0,3	0,3	4,0	0,3	0,3	0,3
pasma (za FM)	0,0	0,0	0,0	0,0	0,0	0,0	0,0	5,3	0,3	0,0	0,0	0,0	0,0	0,5	0,0

Constraint overrun

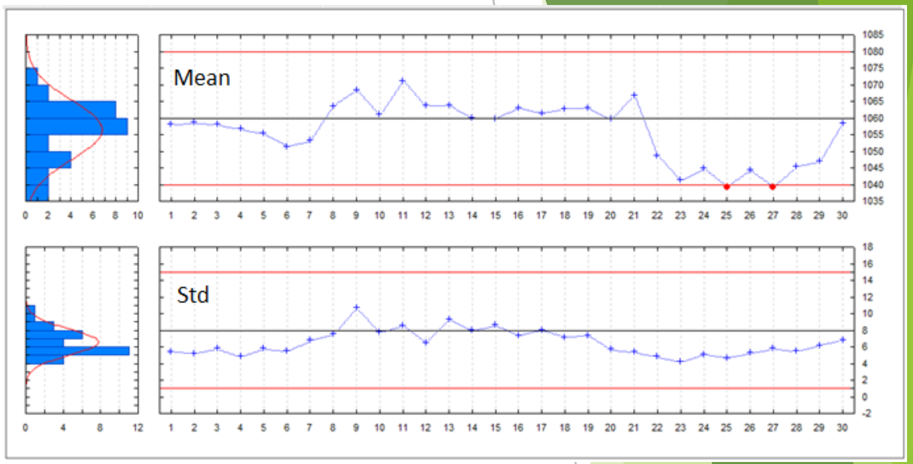
0.3	[0%, 1%)
1.1	[1%, 5%)
5.9	[5%, 100%)

SPC decision support

Single coil temperature distribution

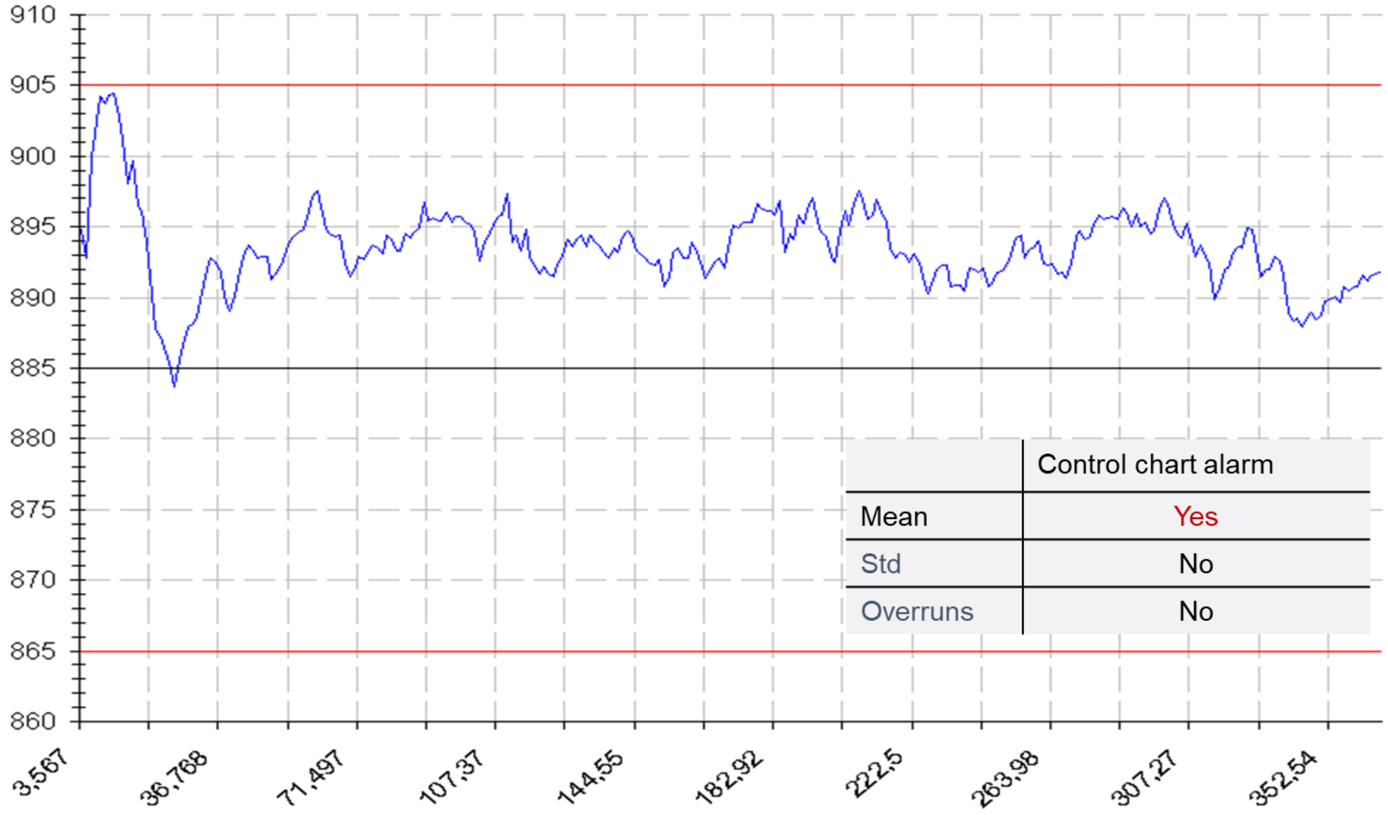
	Control chart alarm
Mean	No
Std	Yes
Overrun	No

Control chart



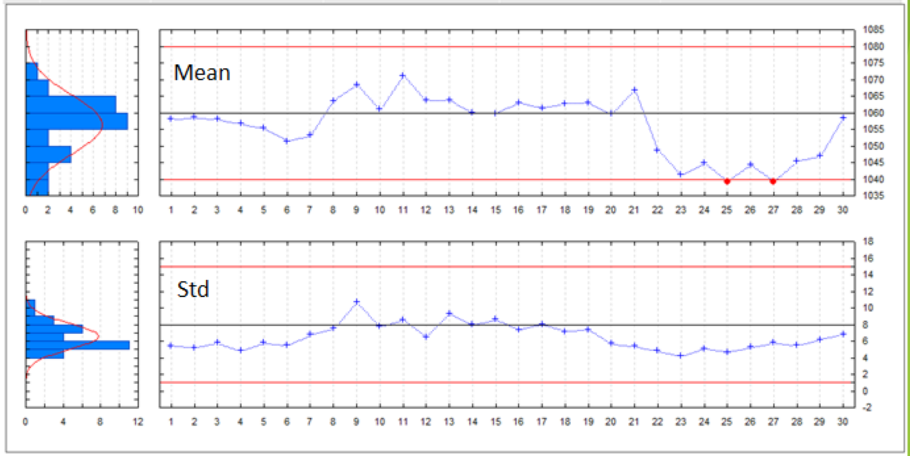
SPC decision support

Single coil temperature distribution



	Control chart alarm
Mean	Yes
Std	No
Overruns	No

Control chart

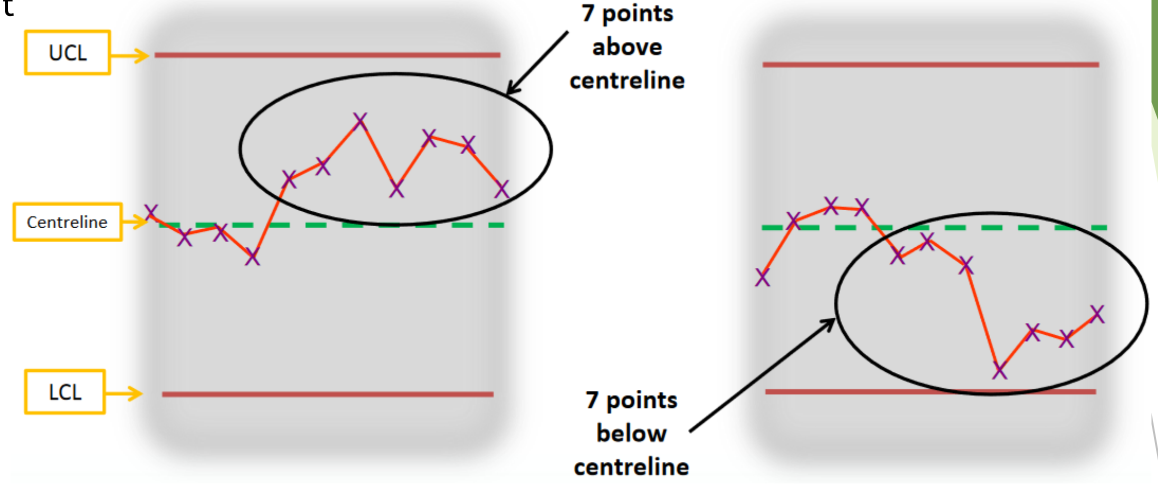


SPC decision support - trends control

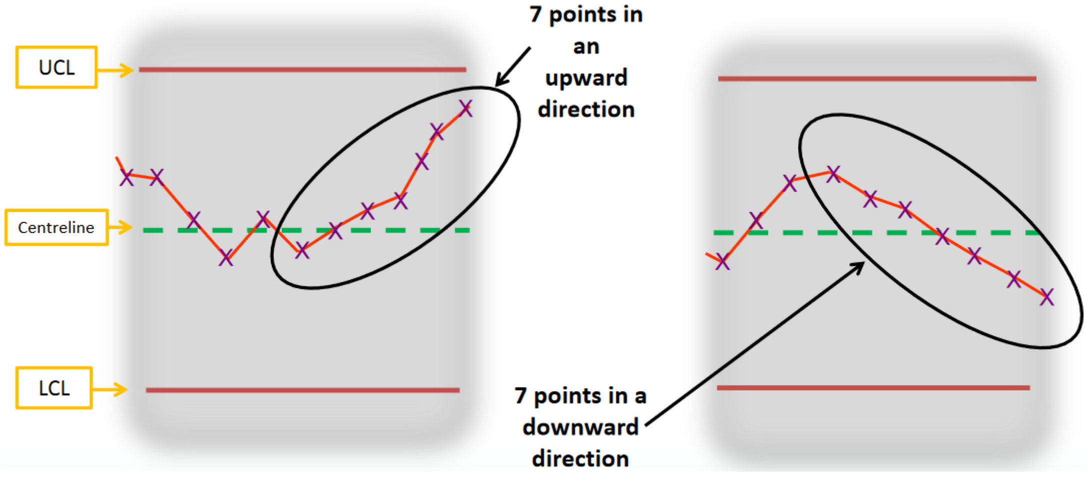
Trend tracking:
Predictive maintenance
Proces capability control

Rules for run charts and control charts

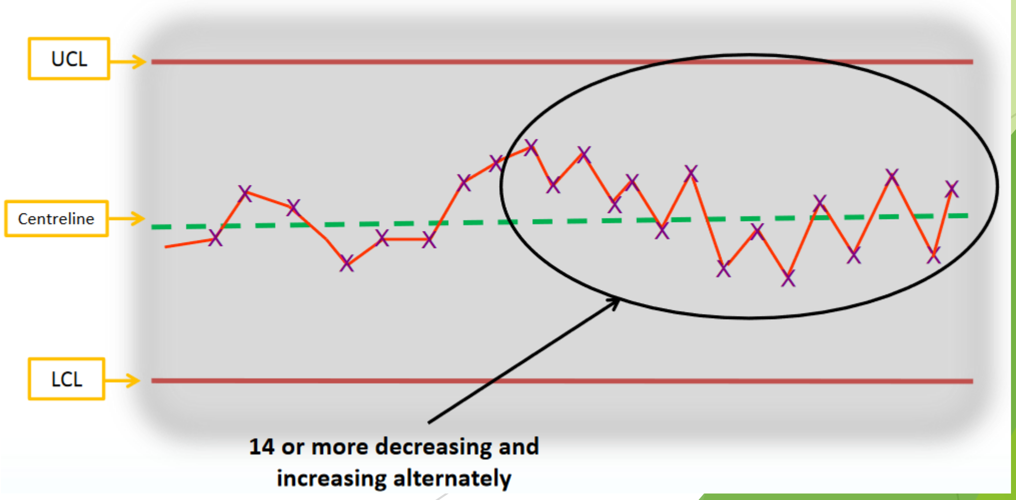
Shift



Indication of trend



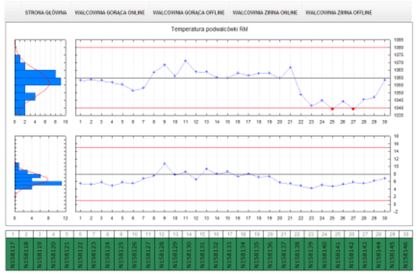
Fluctuations



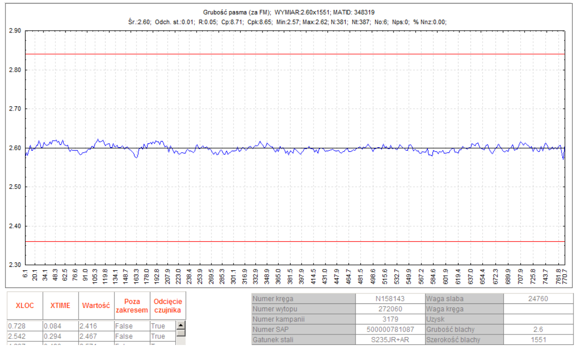
https://www.skillsdevelopmentnetwork.com/infoskills/assets/documents/IS/Understanding%20SPC%20Charts.pdf

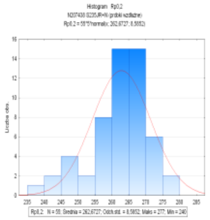
SPC reporting system on Hot Rolling Mill

Control charts



Parameters courses



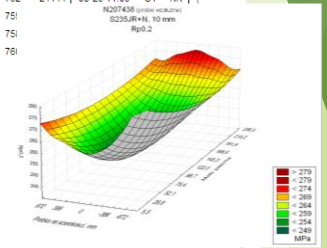


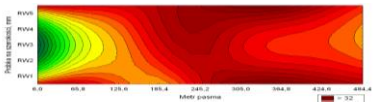
Reports & shift or campagne summary

Lp	SAP	Kampania	ID slaba	Nr kręga	MATID	Gat. COS	Gat. KLIENT	Slab Szer	Slab Gru	Slab Dł	Slab Waga	Krąg Szer	Krąg Grub	Krąg Dł	Krąg Waga	Czas wyład.	Kier	Typ
1	500000781086	3179	K113037104	N158154	348240	F1-COS	F1	1289	220	10992	24560	1270	2	1197	-23814	08-23 11:42	B2	KR
2	500000781086	3179	K113037103	N158153	348239	F1-COS	F1	1281	220	11003	24500	1270	2	1192	-23314	08-23 11:37	B2	KR
3	500000781088	3179	K113055201	N158152	348238	HC180AM-1	HC180AM	1283	220	10737	23980	1275	3	931	-22819	08-23 11:32	B2	KR
4	500000781088	3179	K113055106	N158151	348237	HC180AM-1	HC180AM	1289	220	10720	23780	1275	3	926	-23058	08-23 11:10	B2	KR
5	500000781088	3179	K113055101	N158150	348236	HC180AM-1	HC180AM	1285	220	10747	23800	1275	3	928	-22648	08-23 11:07	B2	KR
6	500000781087	3179	C362264801	N158149	348311	S235-c-COS	S235JR+AR	1611	220	9138	24900	1551	3	762	-24144	08-23 11:03	G1	KR
7	500000781087	3179									00	1551	3	75				
8	500000781087	3179									60	1551	3	75				
9	500000781087	3179									40	1551	3	76				

Podsumowanie statystyczne:

	WG	WZ	Total
Uzysk:	95.96%	98.01%	96.91%
Średnia długość slaba:	8508	9763	9102
Średni ciężar slaba:	22679	21404	22070
Średnia długość pasma:	515	865	681
Średni ciężar kręga:	21756	20979	21384
Średnia szerokość slaba:	1590	1273	1439
Średnia szerokość pasma:	1568	1255	1419





Strip reports

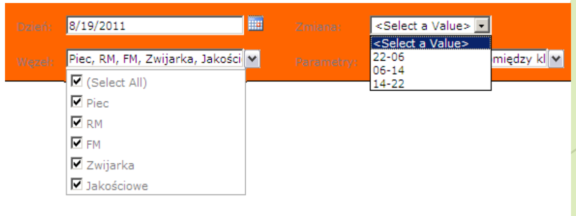
Raport przekroczeń parametrów

PARAMETRY / NR SLABA	N158152
Klin pasma (za FM)	|||
Temperatura końca walcowania	0.0
Temperatura zwijania	0.8
Profil pasma (za FM)	4.3
Szerokość pasma (za FM)	0.0
Grubość pasma (za FM)	0.0
Naprężenie pasma pomiędzy klatką 5. a zwijarką	|||
Temperatura podwalcówki (Encopanel)	|||
Temperatura podwalcówki przed FM	|||
Siła nacisku walców - Strona operatora - klatka 0	|||
Siła nacisku walców - Strona napędu - klatka 0	|||
Przebieganie walców Bending (FM 0)	|||
Prędkość walcowania (FM 0)	|||
Nastawa klatki 0	|||

A clear overruns summaries

PARAMETRY / NR KRĘGA	N158123	N158124	N158125	N158126	N158127	N158128	N158129	N158130	N158131	N158132	N158133	N158134	N158135	N158136	N158137
Klin pasma (za FM)	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡
Temperatura końca walcowania	0,3	0,0	0,0	1,0	1,0	0,0	1,1	0,3	0,0	1,1	1,1	0,0	0,0	0,0	0,5
Temperatura zwijania	1,7	3,5	1,8	1,7	0,8	1,1	0,5	1,4	0,5	1,3	1,1	0,3	1,0	1,5	0,5
Profil pasma (za FM)	69,8	100,0	4,6	7,3	1,9	4,8	2,2	28,6	1,9	28,3	4,3	7,5	2,1	1,8	3,5
Szerokość pasma (za FM)	0,8	1,0	1,1	0,7	0,5	0,5	0,5	1,1	0,0	0,5	0,5	0,5	0,5	0,5	0,5
Grubość pasma (za FM)	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

PARAMETRY / NR KRĘGA	N158138	N158139	N158140	N158141	N158142	N158143	N158144	N158145	N158146	N158147	N158148	N158149	N158150	N158151	N158152
Klin pasma (za FM)	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡
Temperatura końca walcowania	0,0	0,0	0,3	0,0	0,0	0,5	0,5	0,5	0,2	1,0	0,0	0,7	0,2	1,8	0,0
Temperatura zwijania	0,8	0,5	1,5	0,5	1,3	0,8	0,8	1,0	1,2	1,0	0,2	0,5	7,8	1,8	0,8
Profil pasma (za FM)	7,3	6,3	7,3	1,9	4,7	2,9	4,2	7,1	11,1	4,1	3,7	4,2	90,1	5,5	4,3
Szerokość pasma (za FM)	0,8	0,5	0,0	0,0	0,3	0,0	0,5	0,5	0,5	0,5	0,5	0,5	0,2	0,0	0,0
Grubość pasma (za FM)	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

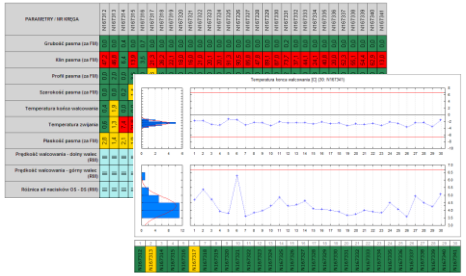


Interactive access to historical data

Users of quality decision support system

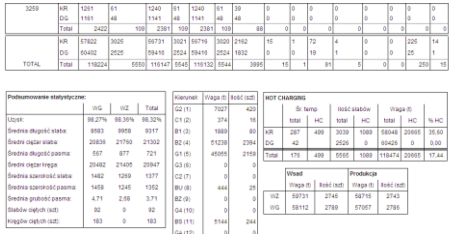
Specialist users

Control charts



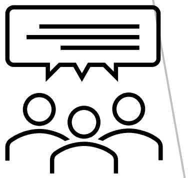
- Operators
- Shift supervisors
- Process engineers
- Quality engineers

Reports



- Control inspectors
- Quality engineers
- Directors

Project teams



R&D – technology optimisation,

Quality department - complaint evaluation

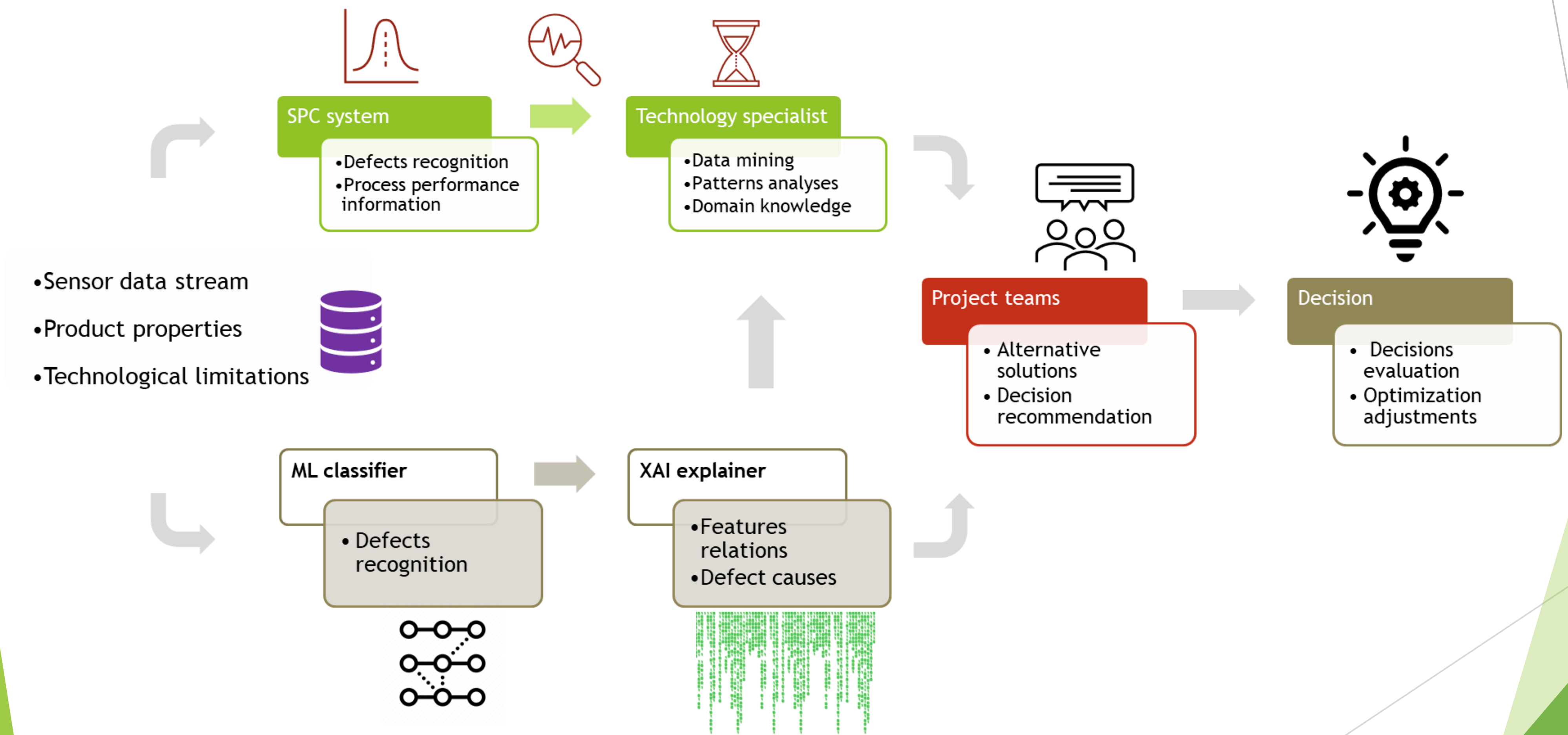
Production departments - logistics

IT systems

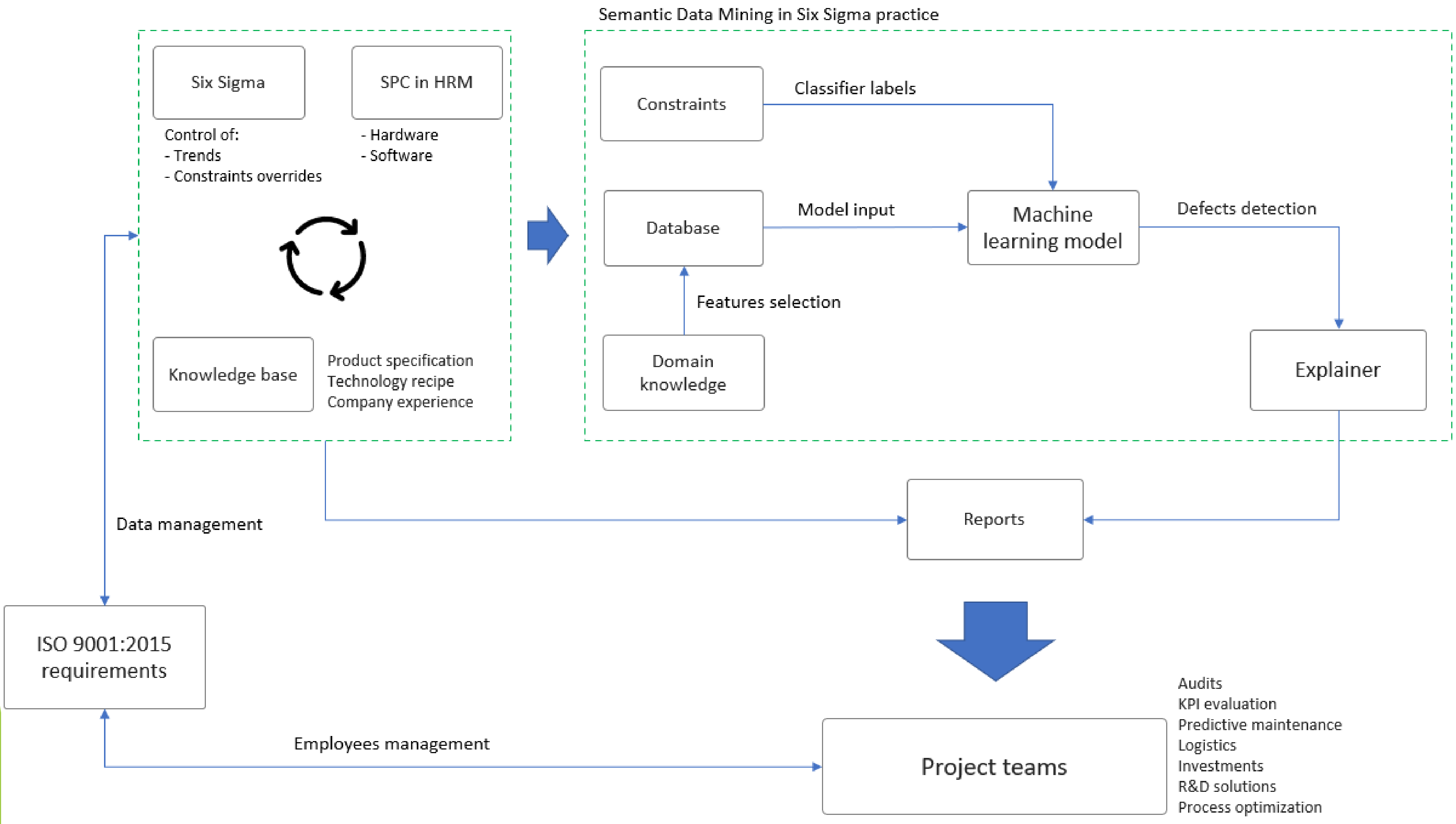
- Enterprise Resource Planning (ERP)
- Logistics systems
- Physical models

Decision support in practise

Goal: find causes of the defect



Semantic Data Mining application



Model input - Feature selection

Single and continuous measured parameters groups

Process parameters data series	**Product quality parameters** data series	**Metadata** single data
Speed	Thickness	Product IDs
Gap	Width	Targets and constraints
Left side pressure force	Profile	Products dimensions in all nodes
Right side pressure force	Wedge	Rollers dimensions
Bending	FM exit temp.	Thickness reduction settings
Shifting	Coiling temp.	Chemical composition

Profile

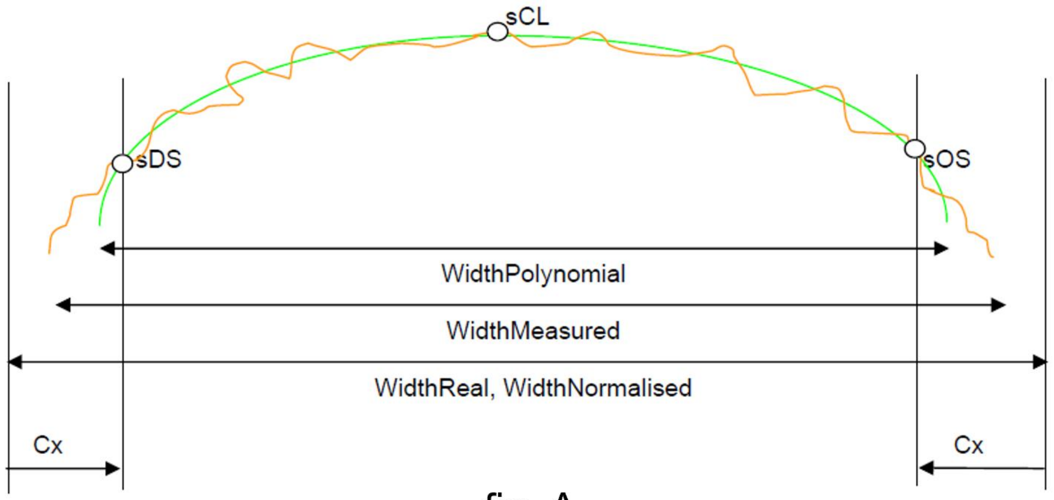


fig. A

Rollers bending

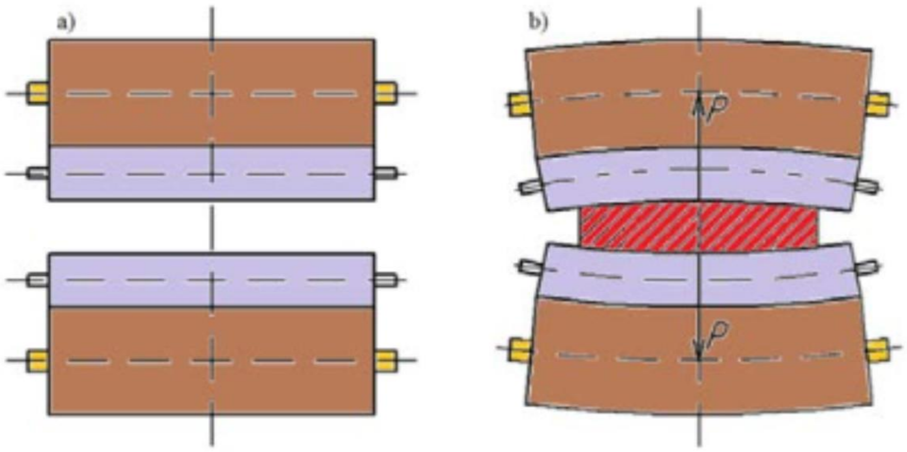


fig. B

Model design

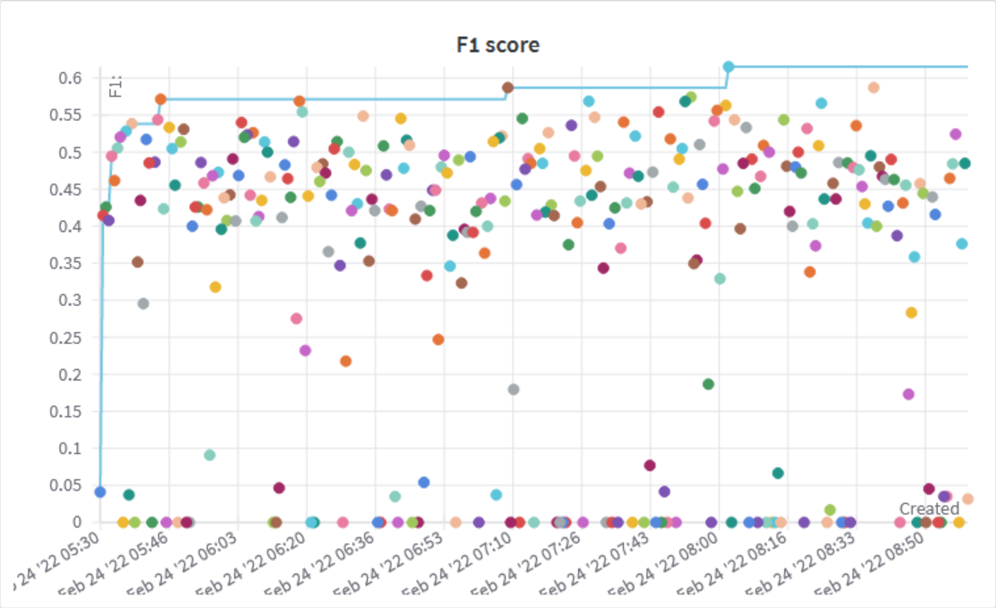
Table 2. Comparison of model performance due to their best F1 score metric

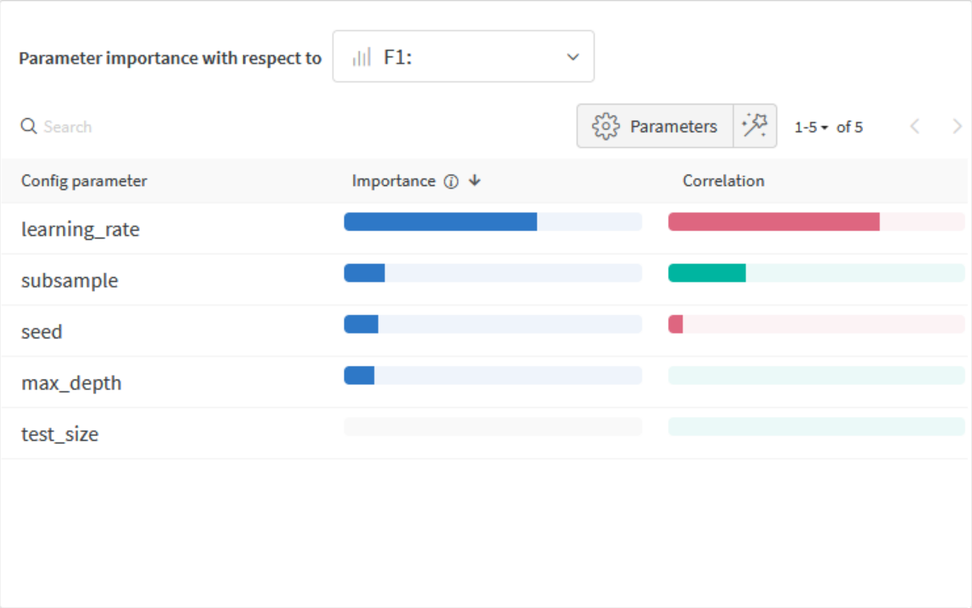
Metrics	XGBoost	Decision Tree	K–Nearest Neighbor
F1	0,61	0,58	0,5
Accuracy	0,986	0,983	0,981
Precision	0,77	0,61	0,66
Recall	0,51	0,55	0,41

Input dataset – Product quality parameters

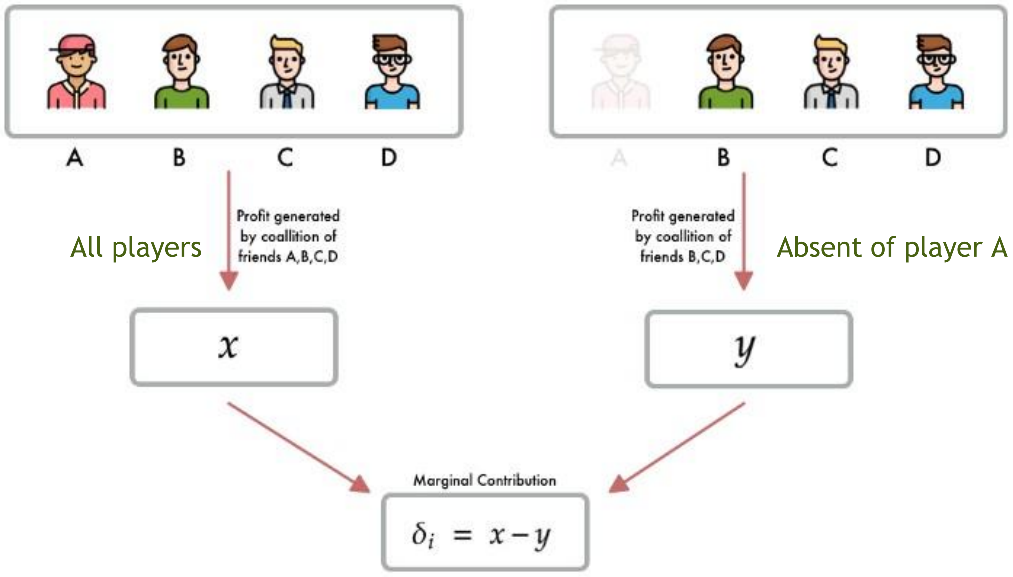
ML classifier – XGBoost

Model explanation – SHAP values



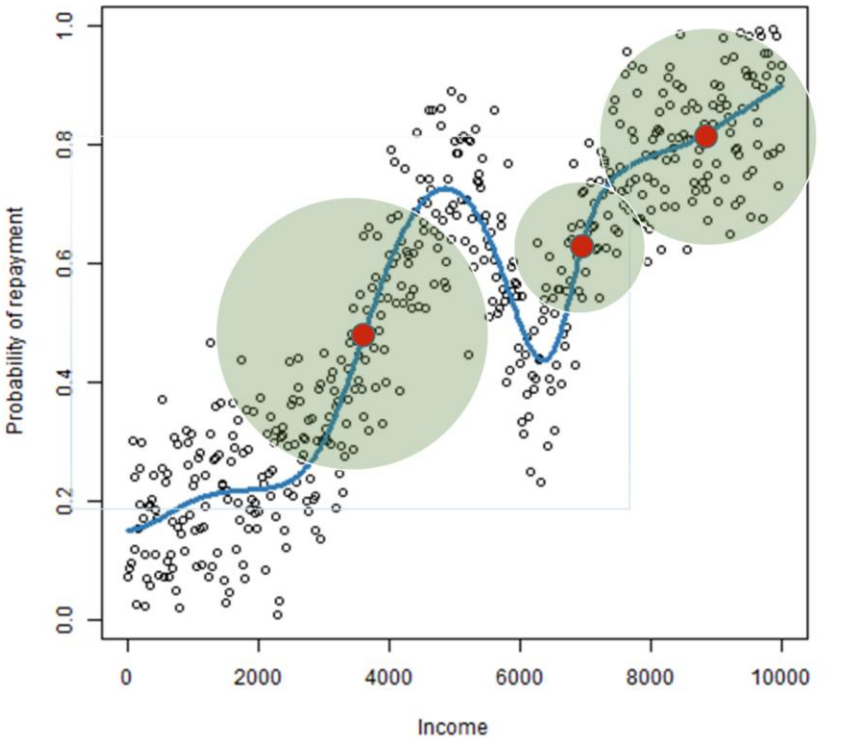


Explainer module - SHAP vs Lime



Marginal contribution of
A for single coalition

SHAP - SHapley Additive exPlanations

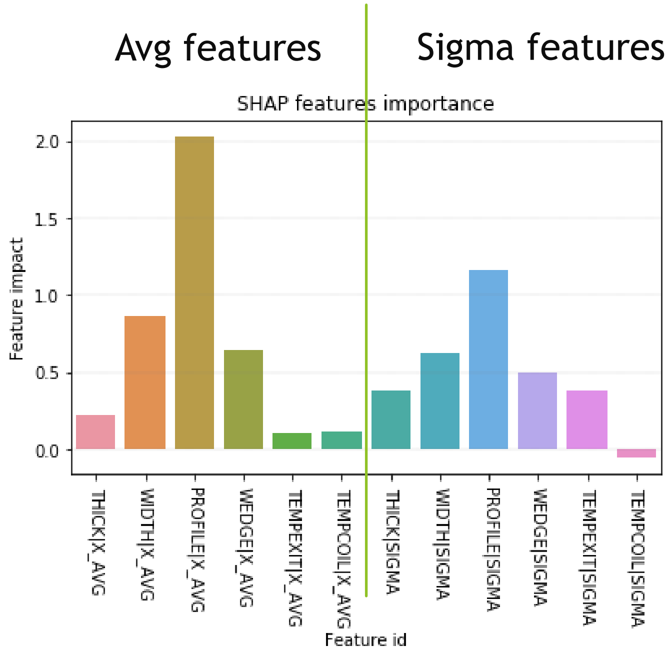


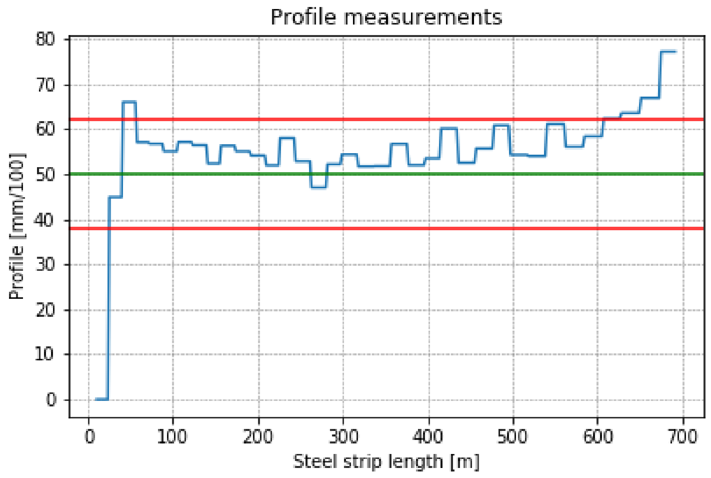
LIME - Local Interpretable Model-agnostic Explanations

https://towardsdatascience.com/shap-shapley-additive-explanations-5a2a271ed9c3
https://towardsdatascience.com/lime-explain-machine-learning-predictions-af8f18189bfe
https://dl.acm.org/doi/abs/10.1145/2939672.2939778

Single prediction results

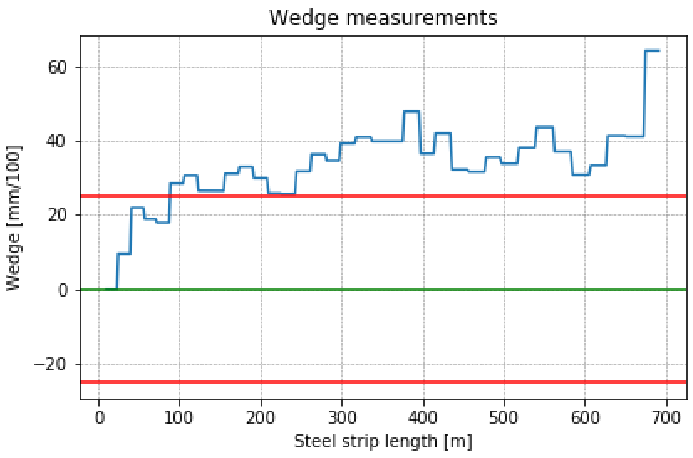
Feature importance

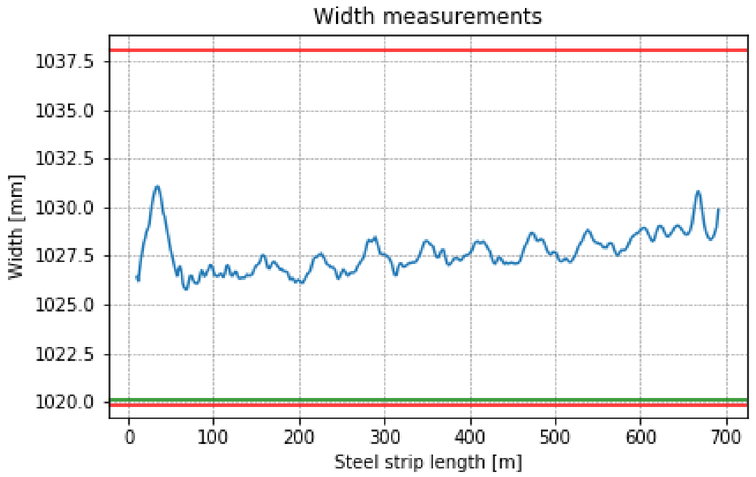




Profile measurements

Wedge measurements



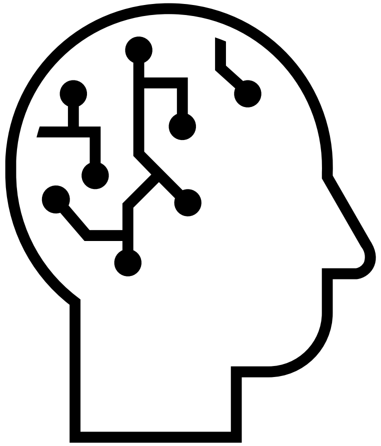


Width measurements

Future plans

Predictive maintanance
Assesment of oscilating signals like forces or vibrations

Introduct **additional types of machine learning models** like LSTM to analysie performance of Finishing Mill device in subsequent stands



Use of machine learning **model to create labels** for subsequent model used for semantic explanations

New forms of the **results visualization** to show summarized impact of the individual parameter

Related works

1. Nalepa, G.J.; Bobek, S.; Kutt, K.; Atzmueller, M. Semantic Data Mining in Ubiquitous Sensing: A Survey. *Sensors* **2021**, *21*. doi:10.3390/s21134322.
2. Yadav, N.; Shankar, R.; Singh, S.P. Impact of Industry4.0/ICTs, Lean Six Sigma and quality management systems on organisational performance. *The TQM Journal* **2020**, *32*, 815–835. doi:10.1108/TQM-10-2019-0251.
3. Gil-Vilda, F.; Yagüe-Fabra, J.A.; Sunyer, A. From Lean Production to Lean 4.0: A Systematic Literature Review with a Historical Perspective. *Applied Sciences* **2021**, *11*. doi: 10.3390/app112110318.
4. Yang, H.; Rao, P.; Simpson, T.; Lu, Y.; Witherell, P.; Nassar, A.; Reutzel, E.; Kumara, S. Six-Sigma Quality Management of Additive Manufacturing. *Proceedings of the IEEE* **2020**, *PP*. doi:10.1109/JPROC.2020.3034519.
5. Tran, K.P. Artificial Intelligence for Smart Manufacturing: Methods and Applications. *Sensors* **2021**, *21*. doi:10.3390/s21165584.
6. Huang, P.M.; Lee, C.H. Estimation of Tool Wear and Surface Roughness Development Using Deep Learning and Sensors Fusion. *Sensors* **2021**, *21*. doi:10.3390/s21165338.
7. He, Z.; Tran, K.P.; Thomassey, S.; Zeng, X.; Xu, J.; Yi, C. Multi-objective optimization of the textile manufacturing process using deep-Q-network based multi-agent reinforcement learning. *Journal of Manufacturing Systems* **2021**. doi:https://doi.org/10.1016/j.jmsy.2021.03.017.
8. Fazelpour, S.; Lipton, Z.C. Algorithmic Fairness from a Non-ideal Perspective. *CoRR* **2020**, *abs/2001.09773*, [2001.09773].
9. Bezobrazov.; Yuriy.; Naumov.; Anton.; Kolbasnikov.; Nikolay. High strength dual-phase steel structure evolution during hot rolling. *Materials Science and Technology Conference and Exhibition 2012, MS and T 2012* **2012**, *2*.
10. Visani, G. LIME: explain Machine Learning predictions, 2020.
11. Biewald, L. Experiment Tracking with Weights and Biases, 2020. Software available from wandb.com.