DOCTORAL DISSERTATION DEFENSE

Semantic Data Mining methods

for decision support in smart manufacturing.

Supervisor: *prof. dr hab. inż. Grzegorz J. Nalepa*

Co-supervisor: *dr inż. Szymon Bobek*

Maciej Szelążek

Agenda

- Motivations
- Research problems and tasks
- Main results of the dissertation
- Further directions of research

Motivations

How to extend the capabilities of statistical-based practices for Quality Control?

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.

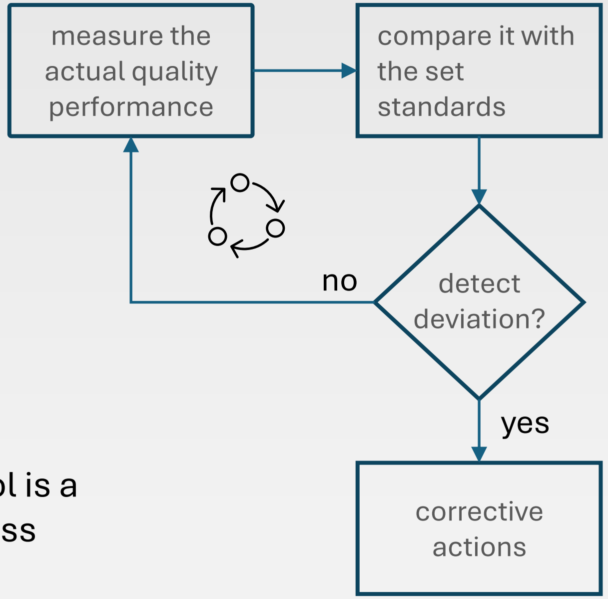
Manufacturing-based definition

- 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.



Quality Control is a reactive process

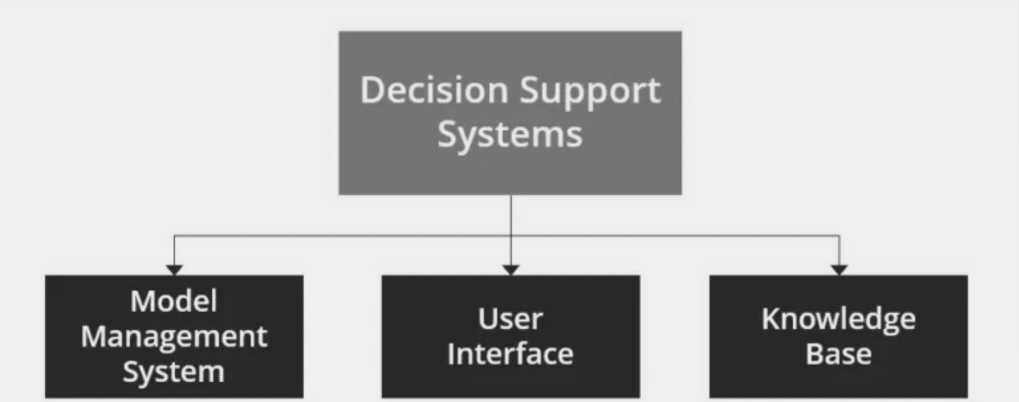
Motivations

How to translate the reasoning of the ML model into actionable knowledge?

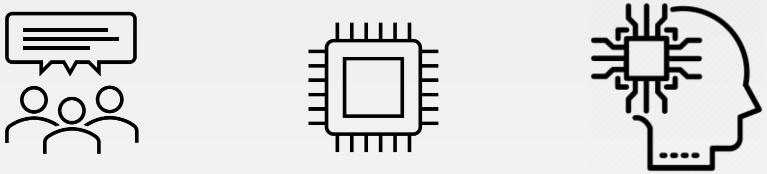
explainable AI as decision support enhancement

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

DSS produces detailed information reports by gathering and analyzing data.



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



Research problems

RP 1:

How to link the results from explainable AI as additional set of information in industrial QM systems and preserve meaning for end users?

RP 2:

What are the key aspects of visualisation methods to fuse information from ML–based indicators and quality control principles?

RP 3:

What is the most effective way to apply Semantic Data Mining to provide decision support in the area of quality control for smart manufacturing?

Research tasks

1. Conduct a comprehensive study of QM and decision support methods in the industrial practice.
2. Analysis of industrial data characteristics focusing on prospective use as input to the SDM procedure.
3. Investigate technology indicators and forms of information presentation used in industrial quality control.
4. Developing a new approach on interpreting explanatory results to provide domain context for end users.
5. Propose a SDM component that will enable practical application in industrial quality control for smart manufacturing.

Main results of the dissertation

[P1] Maciej Szelążek, Szymon Bobek, Antonio Gonzalez--Pardo, Grzegorz J. Nalepa *„Towards the modeling of the hot rolling industrial process : preliminary results.”*; In: Analide, C., Novais, P., Camacho, D., Yin, H.(eds) Intelligent Data Engineering and Automated Learning– IDEAL 2020. IDEAL 2020. Lecture Notes in Computer Science(), vol 12489. Springer, Cham.

[P2] Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa *„Semantic Data Mining Based Decision Support for Quality Assessment in Steel Industry.”*; Expert Systems 41, 2 (2024), e13319.

[P3] Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa *„Improving understandability of explanations with a usage of expert knowledge.”*; Polish Conference on Artificial Intelligence (PP–RAI), (2023.04)

[P4] Iwona Grabska-Gradzińska, Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa *„Visual patterns in an interactive app for analysis based on control charts and SHAP values.”*; In: Nowaczyk, S., et al. Artificial Intelligence. ECAI 2023 International Workshops. ECAI 2023. Communications in Computer and Information Science, vol 1948. Springer, Cham.

[P5] Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa *„Poster: Application of knowledge transfer to ML–based Quality Decision Support practice in the steel manufacturing process.”*; In Proceedings of the 15th Biannual Conference of the Italian SIGCHI Chapter (CHItaly ’23). Association for Computing Machinery, New York, NY, USA, Article 50, 1–2.

[P6] Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa *„Improving understandability of explanations with a usage of expert knowledge.”*; Artificial Intelligence. ECAI 2023 International Workshops (2024), 36–47.

[P7] Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa *„Enhanced Explanations for Knowledge–Augmented Clustering using Subgroup Discovery.”* Maciej Szelążek, Daniel Hudson, Szymon Bobek, Grzegorz J. Nalepa and Martin Atzmueller; 2023 IEEE 10th International Conference On Data Science And Advanced Analytics (DSAA). (2023.10),

[P8] Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa *„Method for using explanations for machine learning model as actionable decision support for data series analysis.”*; Submitted

Main results of the dissertation

Article	Published	Date	Contribution [%]
P1	IDEAL	2020.10	70
P2	Expert Systems	2022.05	80
P3	PP-RAI	2023.04	90
P4	SEDAMI @ ECAI	2023.09	40
P5	CHItaly	2023.09	90
P6	SEDAMI @ ECAI	2023.10	80
P7	DSAA	2023.10	40
P8	Journal of Industrial Information Integration	submitted	80

Paper	T1	T2	T3	T4	T5
P1	✓	✓			
P2	✓				✓
P3			✓	✓	
P4			✓		
P5		✓	✓		
P6		✓			
P7				✓	
P8			✓	✓	✓

Other activities during the doctorate

Publications not included in the dissertation

Published	Date
ICAISC	2020.10
Human-AI Interaction Workshop @ ECAI	2020.11
Scientific Data	2022.06
xAI 2023	2023.07

Performing 17 reviews for IEEE Access

Participation in 7 national and international conferences

Participation in 4 international research projects

Conclusions of research task I

Conduct a comprehensive study of QM and decision support methods in the industrial practice.

Process Approach and Risk-Based Thinking

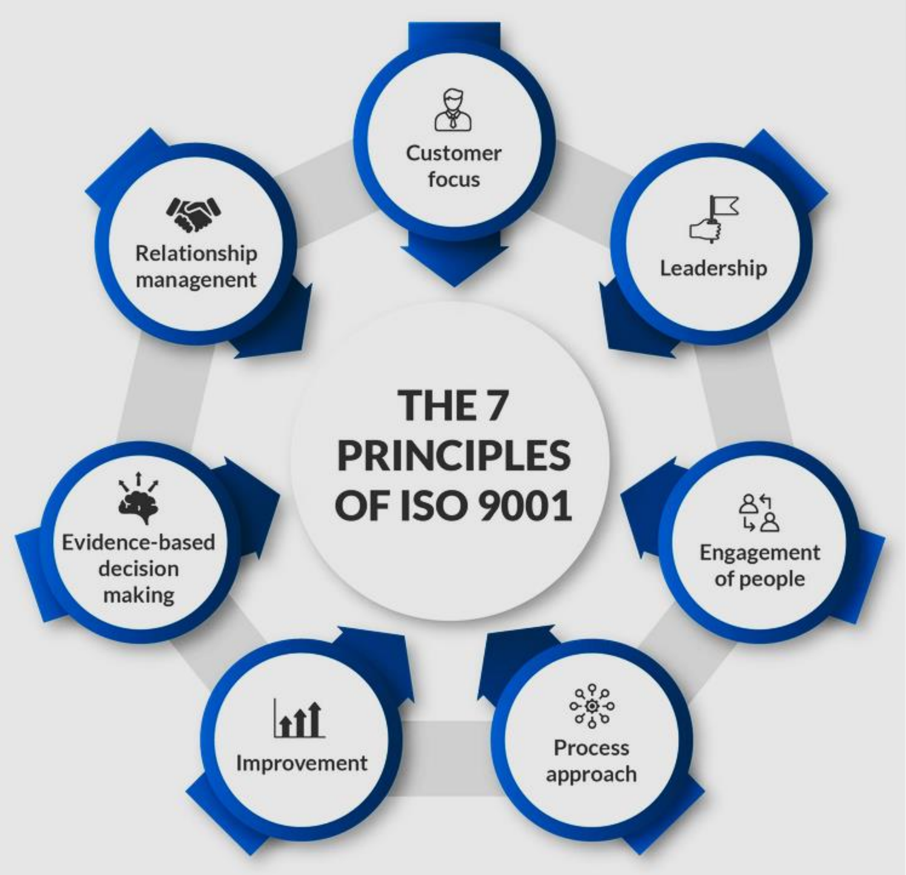
- cross-functional teams work within clearly defined workflows
- anticipate potential failures

Evidence-Based Decision Making

- use of data-driven analysis and performance metrics to make informed decisions

Continuous Improvement and Customer Focus

- project teams constantly refine their methods and solutions
- develop innovative and high-quality technological solutions



Conclusions of research task I

Well-tested set of techniques aimed at reducing the variability (tolerance)

and defectiveness of a product-process

Data-Driven Decision Making

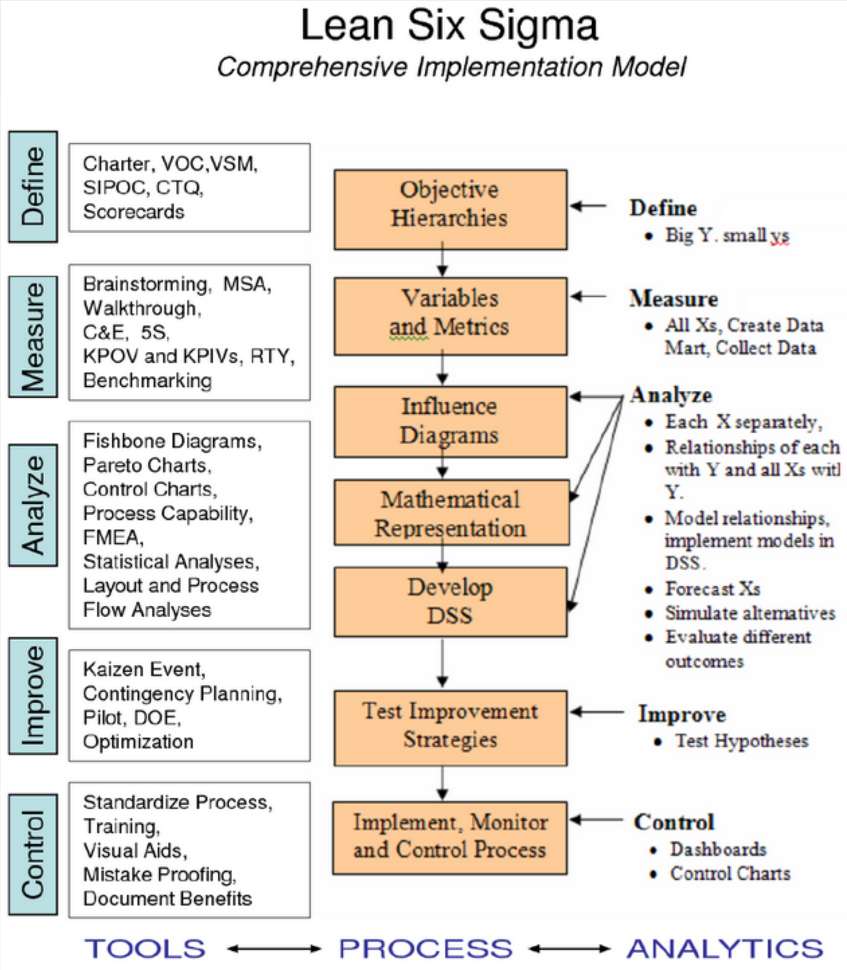
- DMAIC Approach
- Define, Measure, Analyze, Improve, and Control

Continuous Improvement (Kaizen)

- iterative refinements to develop and sustain high-performance, technology-driven solutions

Defect Reduction and Process Optimization

- reduce variability
- increase operational efficiency



https://www.researchgate.net/publication/333648011_Quality_a_Key_Value_Driver_in_Value_Based_Management

[P1] *„Towards the modeling of the hot rolling industrial process : preliminary results."* Maciej Szelążek, Szymon Bobek, Antonio Gonzalez--Pardo, Grzegorz J. Nalepa

[P2] *„Semantic Data Mining Based Decision Support for Quality Assessment in Steel Industry."* Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa

Conclusions of research task II

Analysis of industrial data characteristics focusing on prospective use as input to the SDM procedure.

- Analysis of process data streams
- Selection of the most reliable information and a set of characteristics

Measurement location	Stand 1	Stand 6
Beginning of the strip	0.25m	1.6m
End of the strip	0.35m	2.2m
Length of the strip	114m	654m
Subprocess duration	76s	72s

Furnace

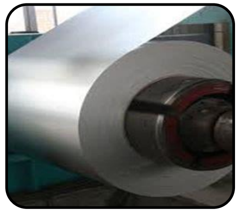
- Heating
- Product - hot slab
- Main parameters - *charge/discharge temperatures, heating time*

Roughing Mill

- Preliminary iterative rolling
- Product - transfer bar
- Main parameters - *forces, temperatures, dimensions*

Finishing Mill

- Primary single pass rolling
- Product – strip
- Main parameters - *temperatures, dimensions, forces, mills setups*



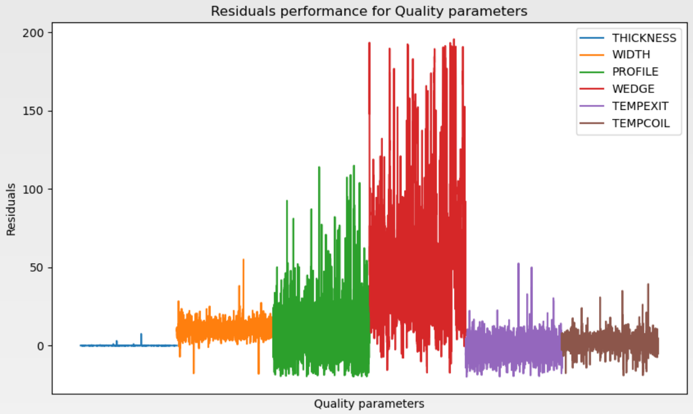
Cooling and coiling

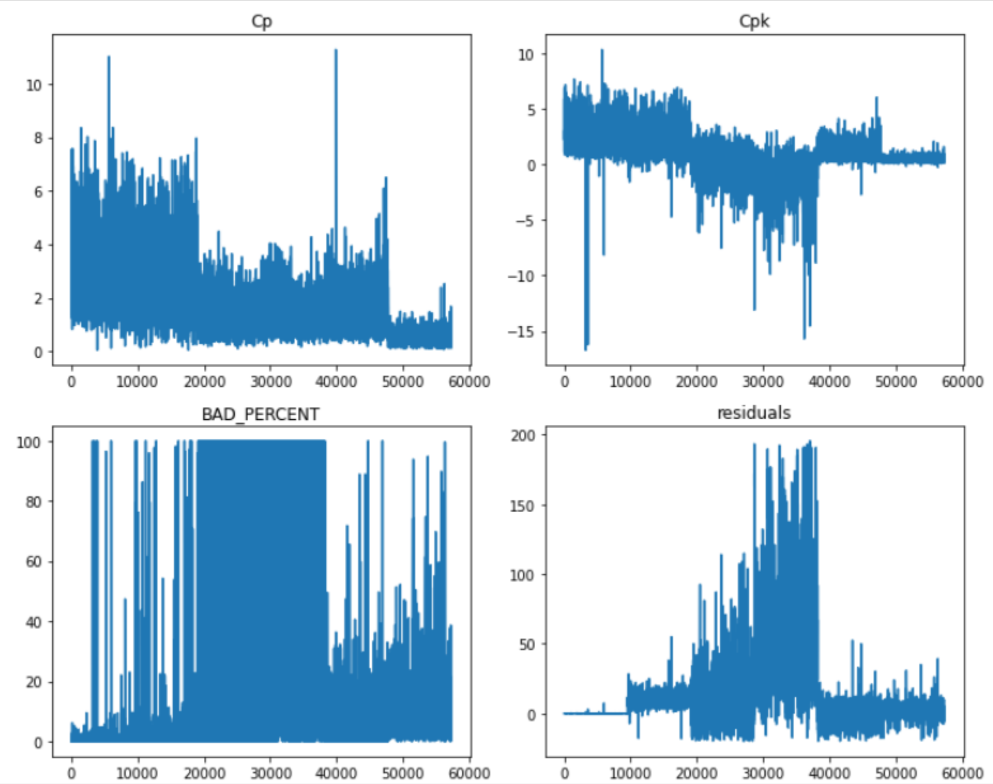
- Temperature reduction
- Product - coil
- Main parameters - *temperatures, water amount*

Conclusions of research task II

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

Measurement location	Stand 1	Stand 6
Beginning of the strip	0.25m	1.6m
End of the strip	0.35m	2.2m
Length of the strip	114m	654m
Subprocess duration	76s	72s





[P1] „*Towards the modeling of the hot rolling industrial process : preliminary results.*" Maciej Szelążek, Szymon Bobek, Antonio Gonzalez--Pardo, Grzegorz J. Nalepa

[P5] „*Poster: Application of knowledge transfer to ML–based Quality Decision Support practice in the steel manufacturing process.*" Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa

[P6] „*Improving understandability of explanations with a usage of expert knowledge.*" Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa

Conclusions of research task III

Investigate technology indicators and forms of information presentation used in industrial quality control.

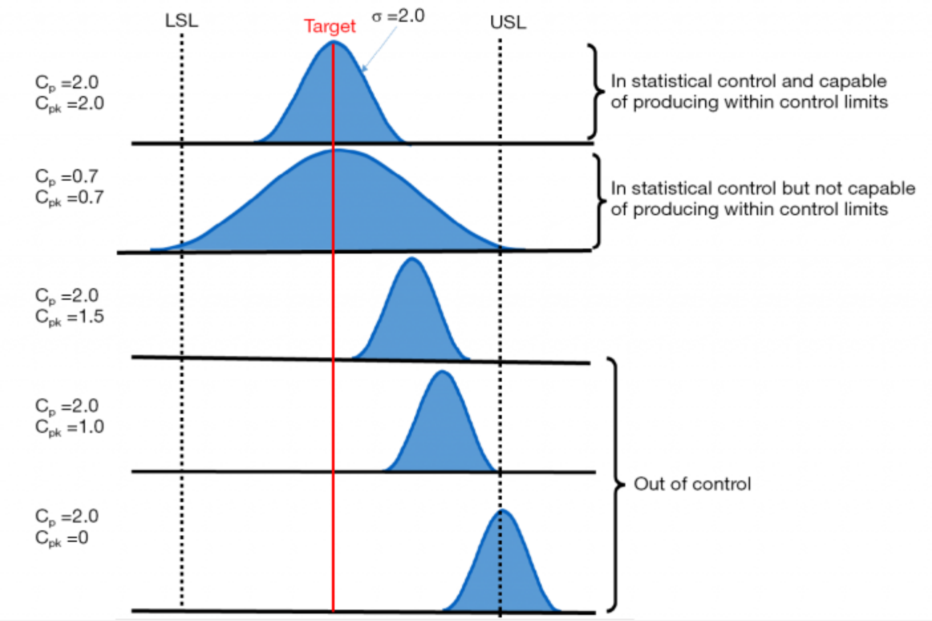
Process capability – Cp : tolerance width divided by the total spread of process.

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

Calculation of Process Capability (Cp) :

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

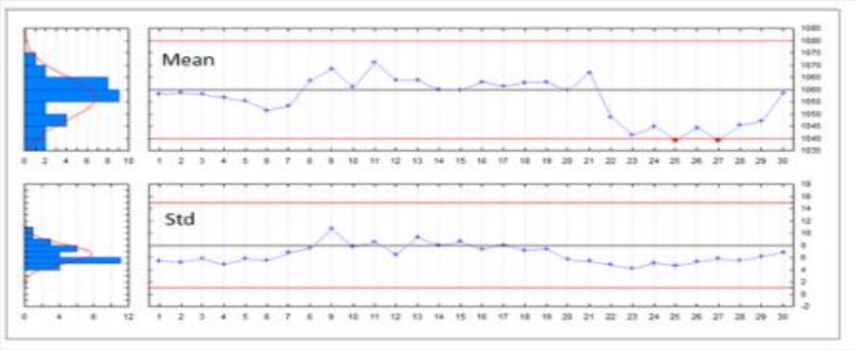
USL = Upper Specification Limit, **LSL** = Lower Specification Limit



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

Conclusions of research task III

Consistency with quality control methods based on commonly used statistical evaluation of sensory process measurements

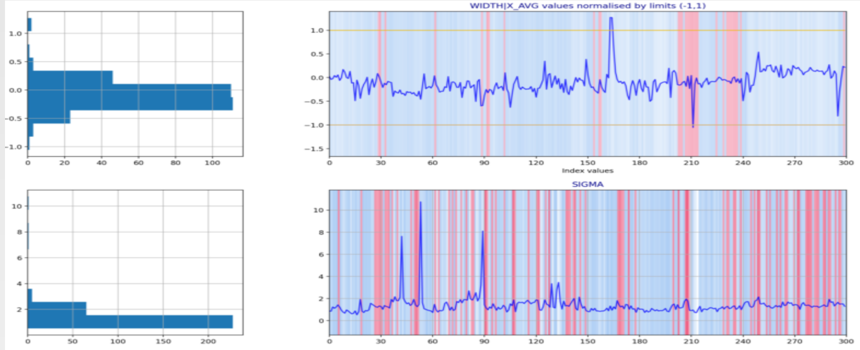


Quality KPIs

- Capability factors
- Bad percent
- Residuals

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



[P3] *„Improving understandability of explanations with a usage of expert knowledge."* Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa

[P4] *„Visual patterns in an interactive app for analysis based on control charts and SHAP values."* Iwona Grabska-Gradzińska, Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa

[P5] *„Poster: Application of knowledge transfer to ML–based Quality Decision Support practice in the steel manufacturing process."* Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa

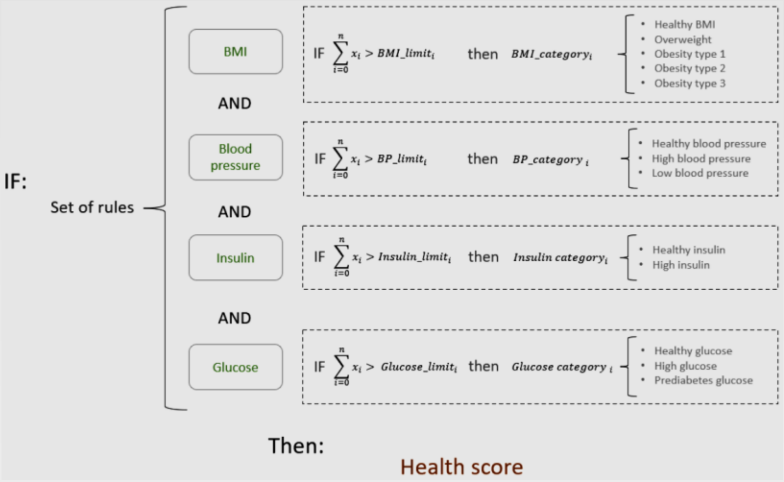
[P8] *„Method for using explanations for machine learning model as actionable decision support for data series analysis."* Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa

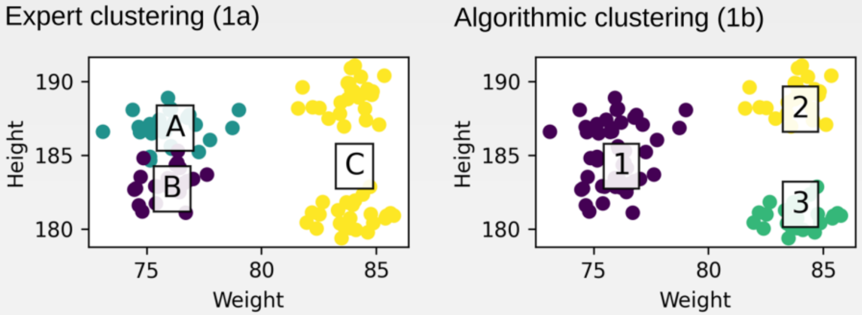
Conclusions of research task IV

Developing a new approach on interpreting explanatory results to provide domain context for end users.

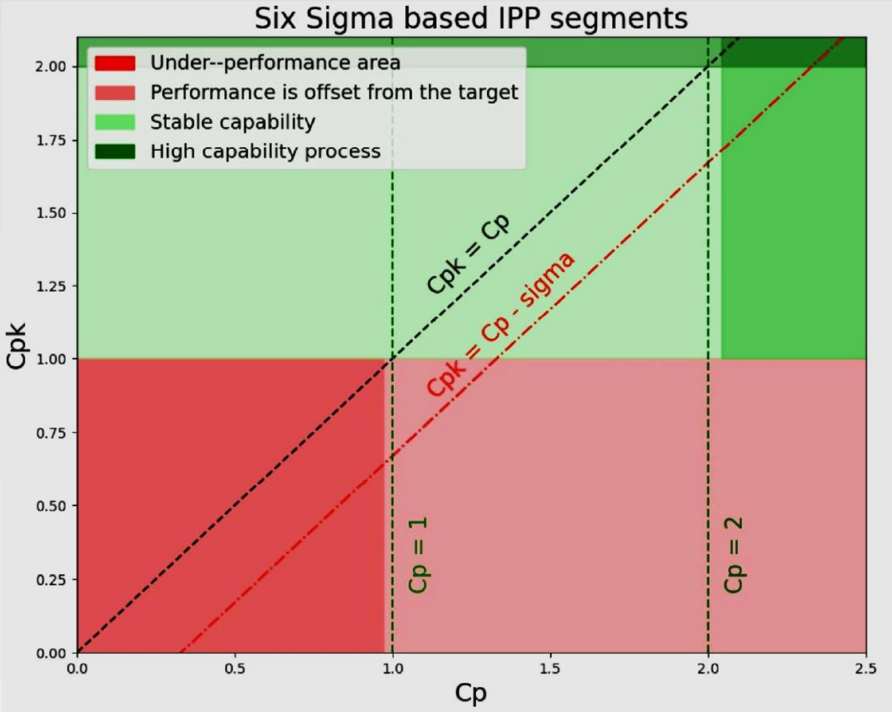
Integration of domain knowledge into the analytical process:

- Knowledge–Augmented Clusters (KnAC)
- Subgroup Discovery
- Set of rules based on domain knowledge
- Quality Management standards





Conclusions of research task IV



Incorporation of domain knowledge into Quality Management analysis:

- Additional set of rules or conditions
- Compliance with company practice
- Evolution not revolution. Extension of QM standards, maintaining compatibility with previous solutions

Model agnostic: regression models and classifiers, regardless of the ML model used

XAI agnostic: flexibility of using any explainability-based indicator

[P3] *„Improving understandability of explanations with a usage of expert knowledge."* Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa

[P7] *„Enhanced Explanations for Knowledge–Augmented Clustering using Subgroup Discovery."* Maciej Szelążek, Daniel Hudson, Szymon Bobek, Grzegorz J. Nalepa and Martin Atzmueller

[P8] *„Method for using explanations for machine learning model as actionable decision support for data series analysis."* Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa

Conclusions of research task V

Propose a SDM component that will enable practical application in industrial quality control for smart manufacturing.

Actionability

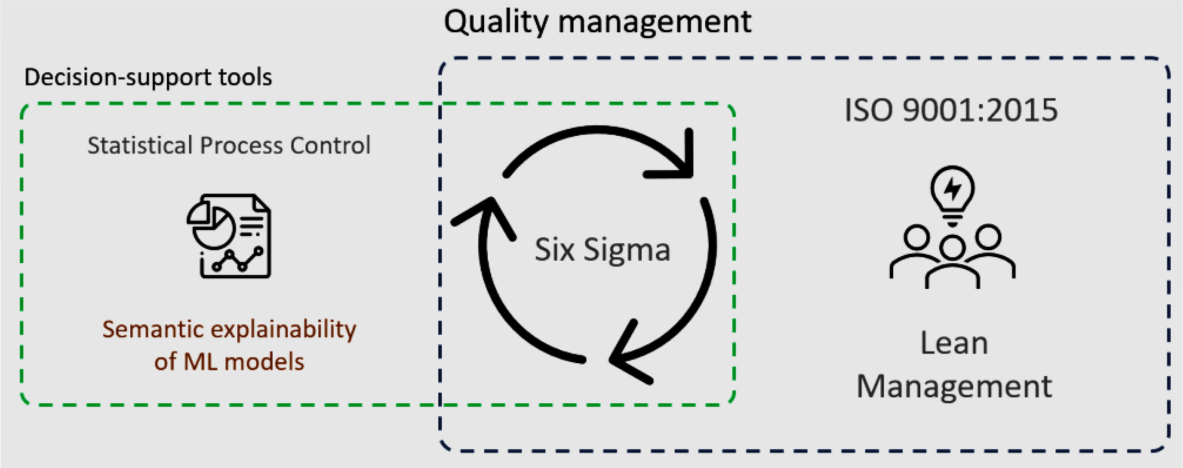
ML methods enhance traditional statistical approaches, improving process control, deviation detection, and cost optimization within QM systems.

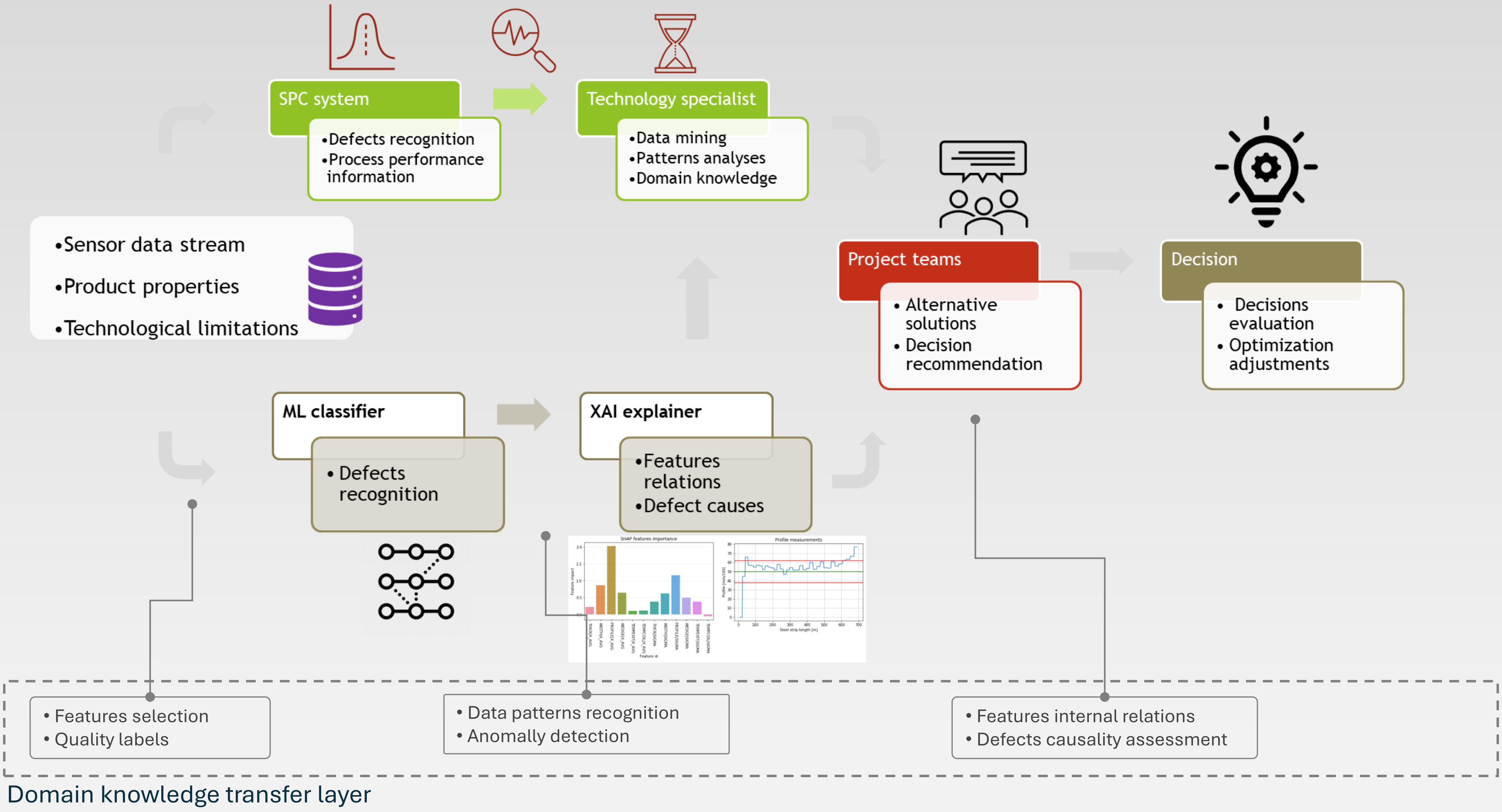
Visualization

A standardized format enables effective comparison of independent metrics, improving the presentation of explainability outcomes in IPA analysis

Contextualization

Combining classic SPC indicators with XAI-based metrics ensures interpretability, making ML-generated insights more accessible for decision-making.



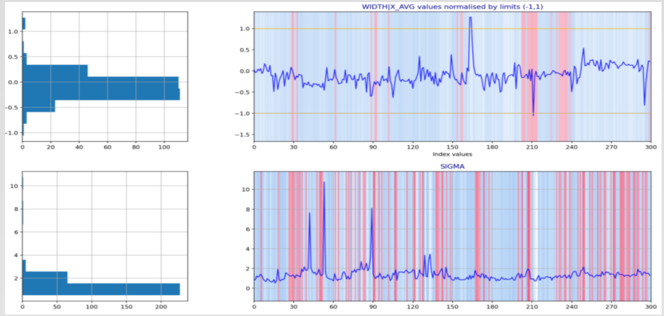

SPC system
•Defects recognition
•Process performance information
Technology specialist
•Data mining
•Patterns analyses
•Domain knowledge
•Sensor data stream
•Product properties
•Technological limitations
Project teams
• Alternative solutions
• Decision recommendation
Decision
• Decisions evaluation
• Optimization adjustments
ML classifier
• Defects recognition
XAI explainer
•Features relations
•Defect causes
SHAP features importance
Profile measurements
• Features selection
• Quality labels
• Data patterns recognition
• Anomally detection
• Features internal relations
• Defects causality assessment
Domain knowledge transfer layer

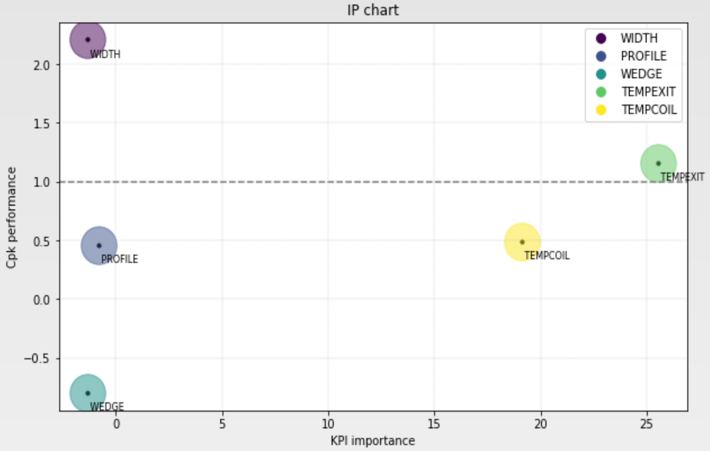
Conclusions of research task V

- The possibility **of full integration with existing systems** in terms of contextualization and interpretation of the obtained results.

- **Universality of applications**, possibility of creating multidimensional links between data streams and business KPI's.

  - Consistent visualization scheme that will allow interpretation of results regardless of ML methods, XAI and the set of information presented.





Feature	Cpk	residuals shap
WIDTH	2.21	-1.35
PROFILE	0.46	-0.83
WEDGE	-0.79	-1.33
TEMPEXIT	1.16	25.54
TEMPCOIL	0.49	19.10

[P2] „*Semantic Data Mining Based Decision Support for Quality Assessment in Steel Industry.*" Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa

[P8] „*Method for using explanations for machine learning model as actionable decision support for data series analysis.*" Maciej Szelążek, Szymon Bobek, Grzegorz J. Nalepa

Further directions of research

- Recognizing the impact of data disparity between classes. An important aspect of analysis related to high-precision manufacturing processes.
- Assessing the differences between XAI methods in terms of response variability for a similar dataset.
- Utilization of developed methods in areas other than smart manufacturing.

Dziękuję za uwagę