UNIVERSITY OF WARSAW

SOFTWARE

Model-agnostic Explanations of Black-box Prediction Models using Rough Sets – the case of post-competition analytics at KnowledgePit.ai

Andrzej Janusz

European Funds Smart Growth

Republic of Poland

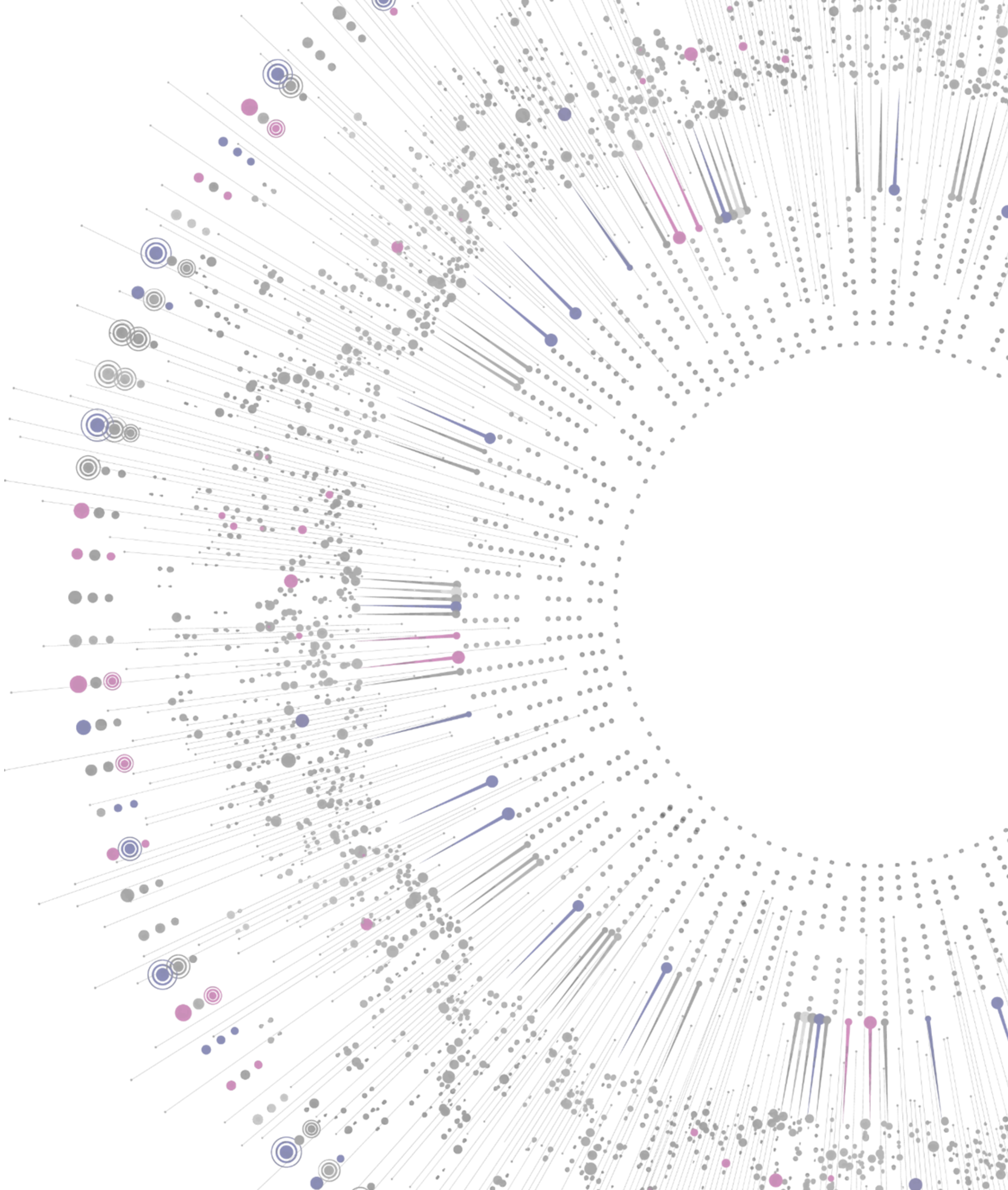
The National Centre for Research and Development

European Union European Regional Development Fund

This research was co-funded by Smart Growth Operational Programme 2014-2020, financed by European Regional Development Fund, in frame of project POIR.01.01.01-00-1070/21, operated by National Centre for Research and Development in Poland.

Agenda

1. Data science competitions:
   - What are they?
   - How they work?
   - An exemplary competition.
2. Post-competition data analysis:
   - What can we learn about solutions?
   - How rough sets can help?
   - What can we learn about data?



Data science competitions

For organizers/companies:

- an easy way of outsourcing work to the community
- a reliable feasibility study
- significant reduction of research costs
- an opportunity to acquire new or evaluate data science specialists

For contestants:

- experience
- a publication opportunity
- fame and glory 😊
- rewards
- exposure to novel research topics and job opportunities

A historical perspective

Late 90's
Early competitions organized at ML and KDD conferences (e.g. the KDD Cup)

2006
The Netflix Prize ($1 million)

2010
Kaggle was founded

2012
Heritage Health Prize organized at Kaggle ($3 millions)

2013
KnowledgePit was founded

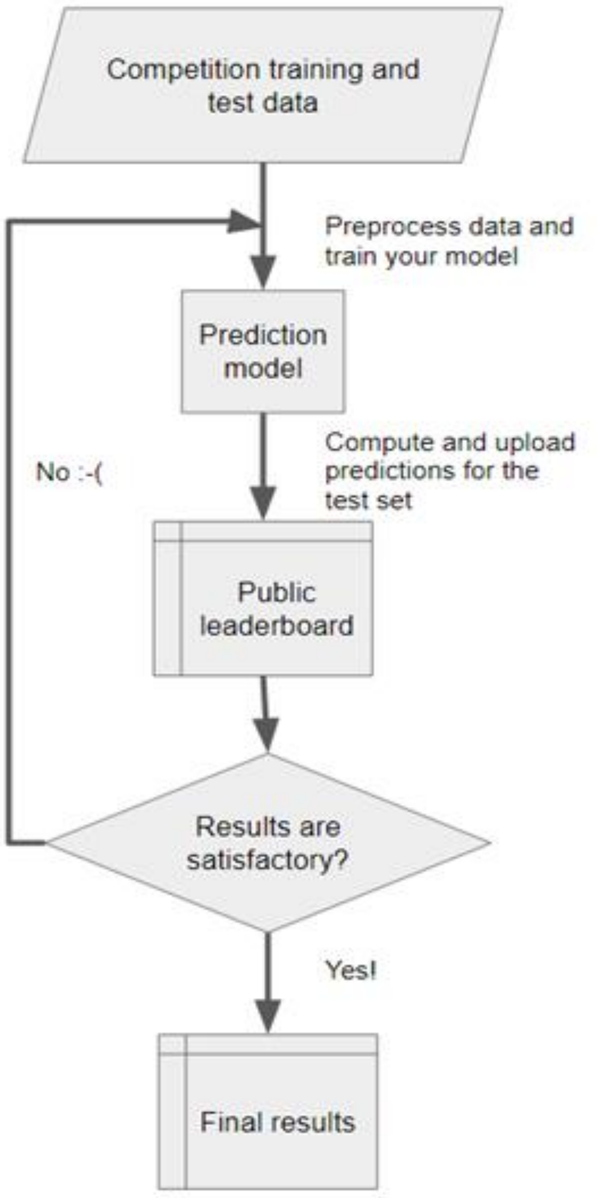
2017
Google buys Kaggle

2023
QED FORCE
Launch of the **QED Force** initiative

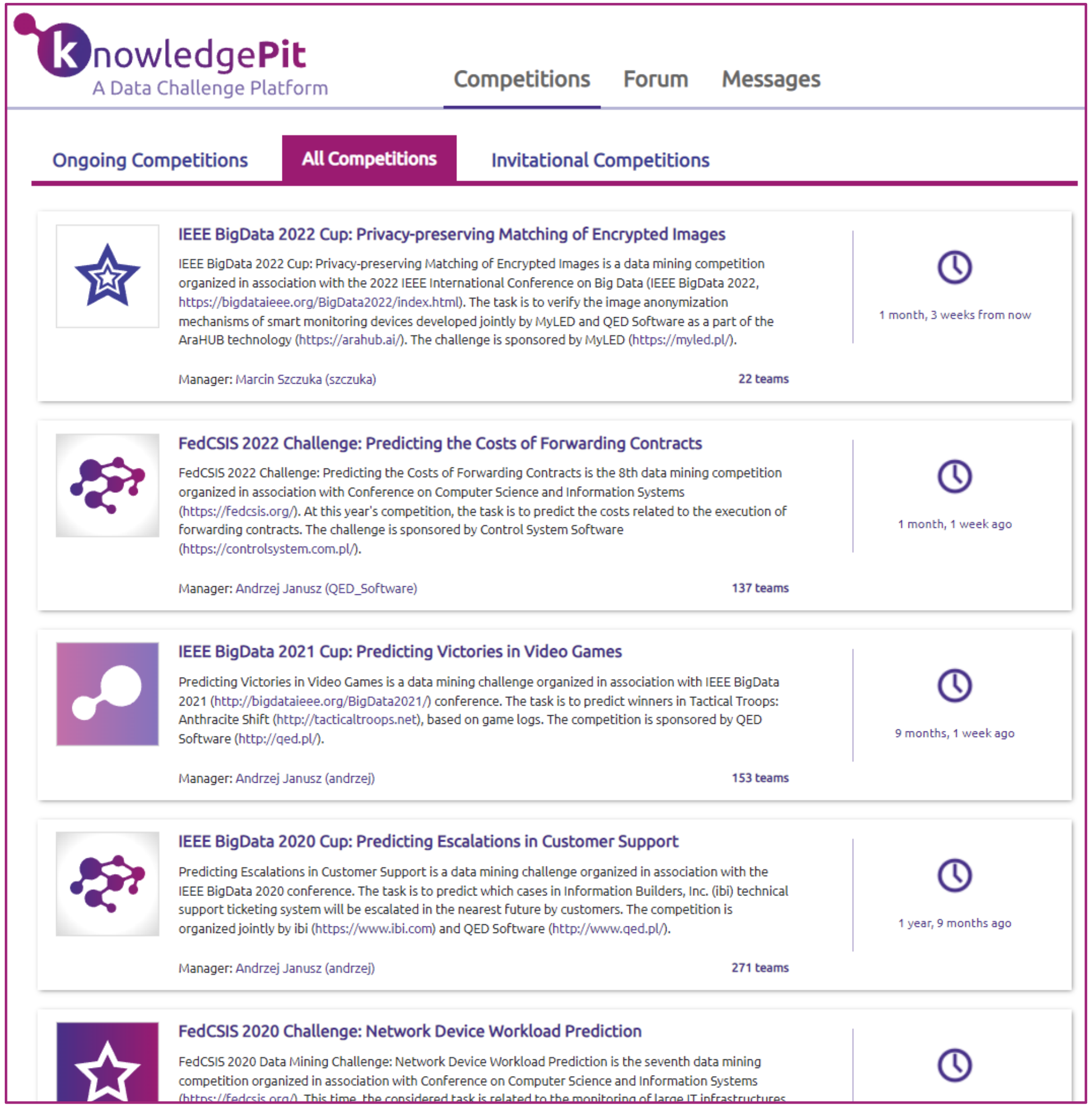
How does it (usually) work?

A typical competition schema

1. The available data set is divided into the training and test parts.
2. Target values (e.g., labels) for the test set are hidden from participants – they need to be predicted.
3. Participants submit solutions which are assessed on a sample from the test set.
4. Participants select their most reliable models and write short reports.
5. The final solutions are evaluated on the remaining test data.



KnowledgePit.ai

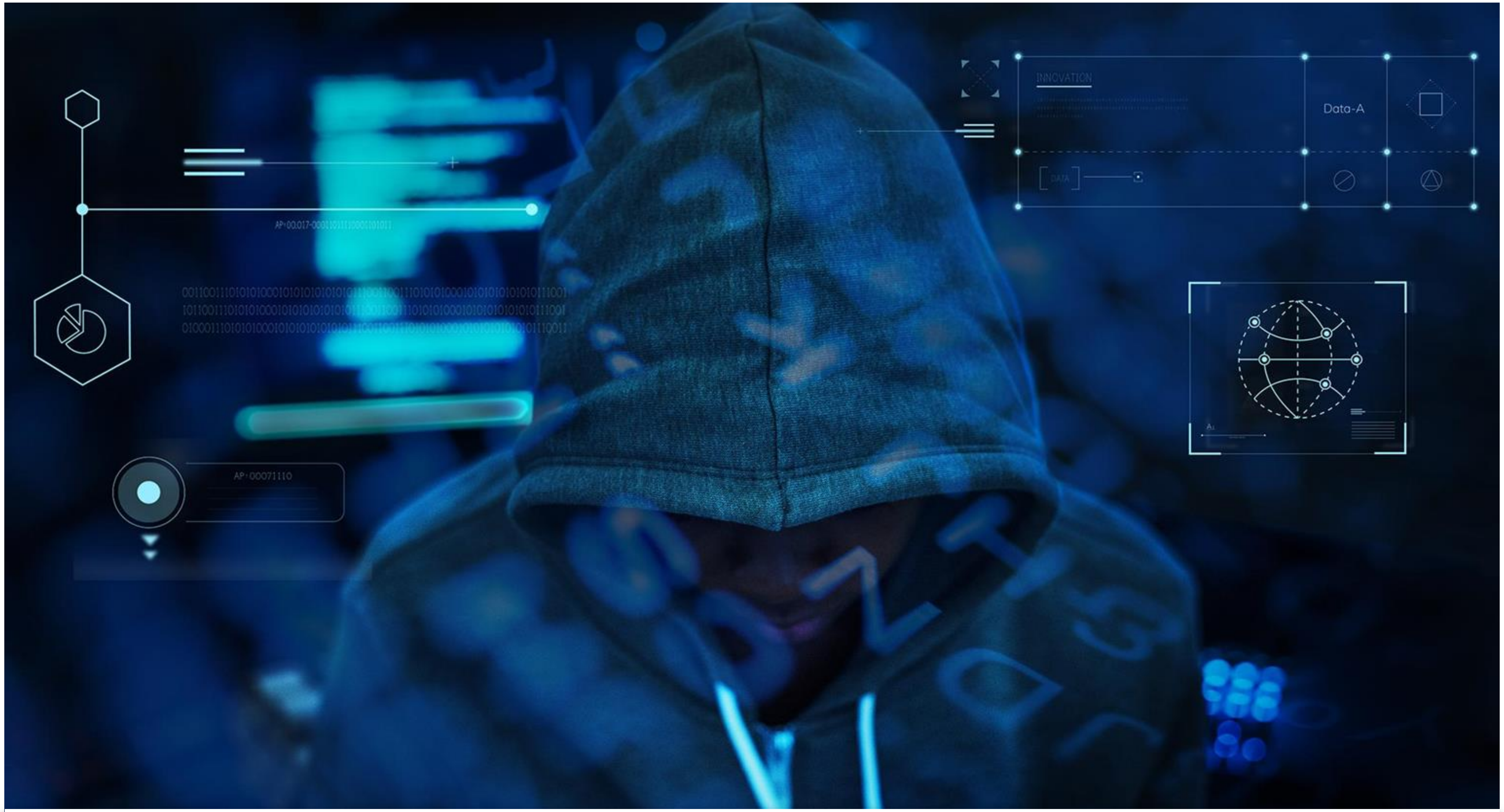


- Stimulates data science research and knowledge exchange in the community.
- Provides interesting data sets and research topics.
- Establishes connections between industry and academia.
- Promotes interesting events and conferences.
- Provides commercial services to companies that seek state-of-the-art in ML.
- Includes modern XAI functionalities to support competition participants, sponsors, and organizers.

Main competition topics (so far)

- Firefighting – 2014, 2015
- Financial/retail industry – 2015, 2017
- Coal mining – 2015, 2016
- Video games – 2017, 2018, 2019, 2021, 2023 (we like video games 😊)
- Customer service – 2020
- Privacy preservation in video analysis – 2022
- Transportation and logistics – 2022
- Cyber-security and hardware monitoring – 2019, 2020, 2023
- and many other, smaller challenges for my students at MIM UW…

Suspicious Network Event Recognition



IEEE BIG DATA 2019
Los Angeles, CA, USA • 9–12 December

SECURITY ON-DEMAND

https://knowledgepit.ai/suspicious-network-event-recognition/

Scope of the challenge

Learn to identify ThreatWatch alerts that ought to be reported to SOD's clients as truly suspicious.

Raw data provided by the SOD company:

- ThreatWatch alerts investigated by analysts at SOD's Security Operations Center (~$6*10^4$ records, ~15MB of data).
- Localized alerts corresponding to the investigated ThreatWatch alerts (~$8.7*10^6$ records, ~2GB of data).
- History of log events for each of the investigated ThreatWatch alerts (~$9*10^9$ records, ~4.5TB of data).

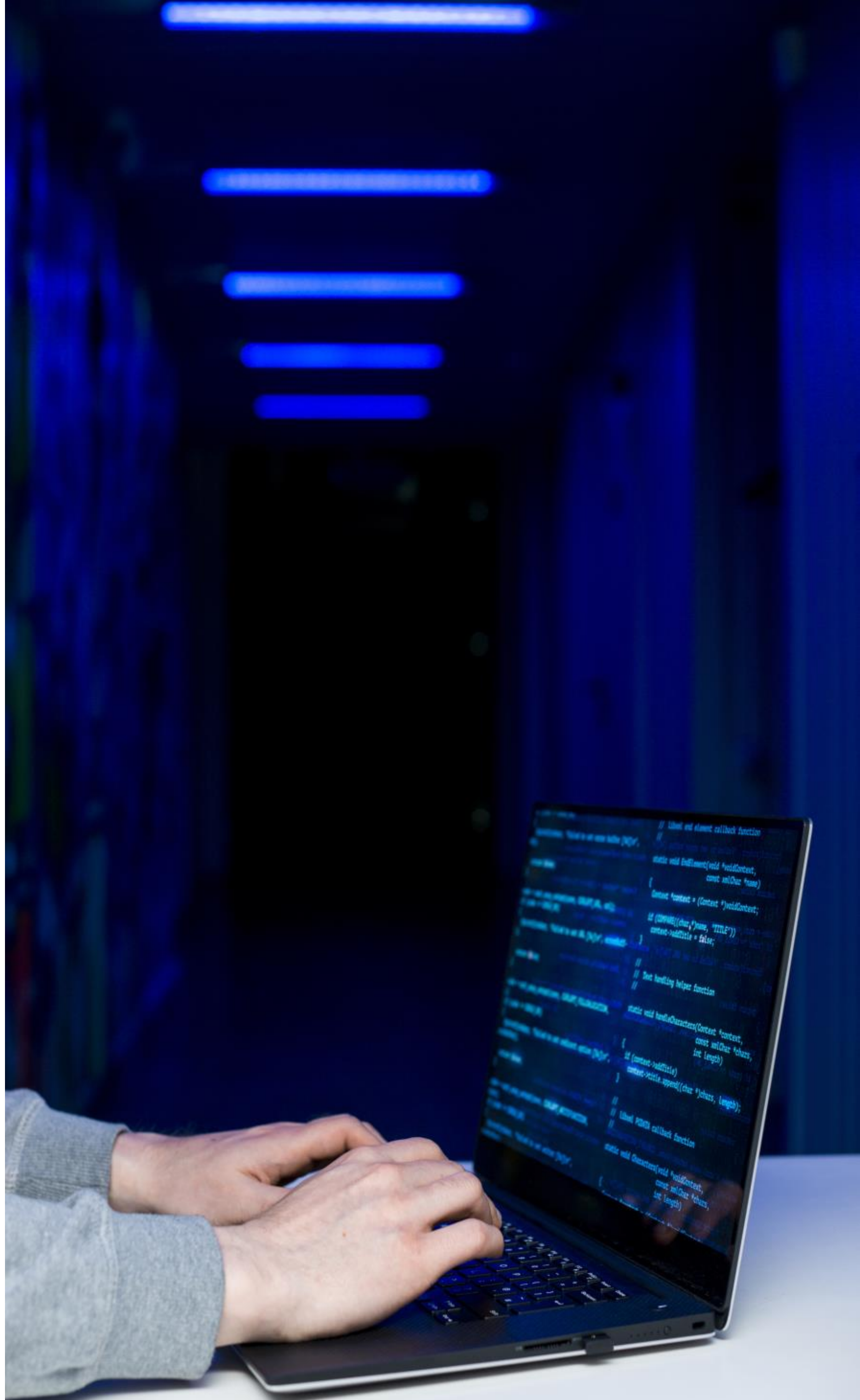


Image: freepik.com

Available data

- ThreatWatch alerts (61 features):
  - ➢ imbalanced decision classes (only ~5.7% of ‘1’ class),
  - ➢ hidden ordering of alerts to avoid data leaks.
- Localized alerts identified by SOD’s rule-based heuristics:
  - ➢ series of events ordered in time,
  - ➢ each alert described by 20 features.
- Log events – all logs gathered for individual ThreatWatch alerts
  - ➢ 29 data chunks, each between 10-20GB
  - ➢ records ordered in time, described by 26 features

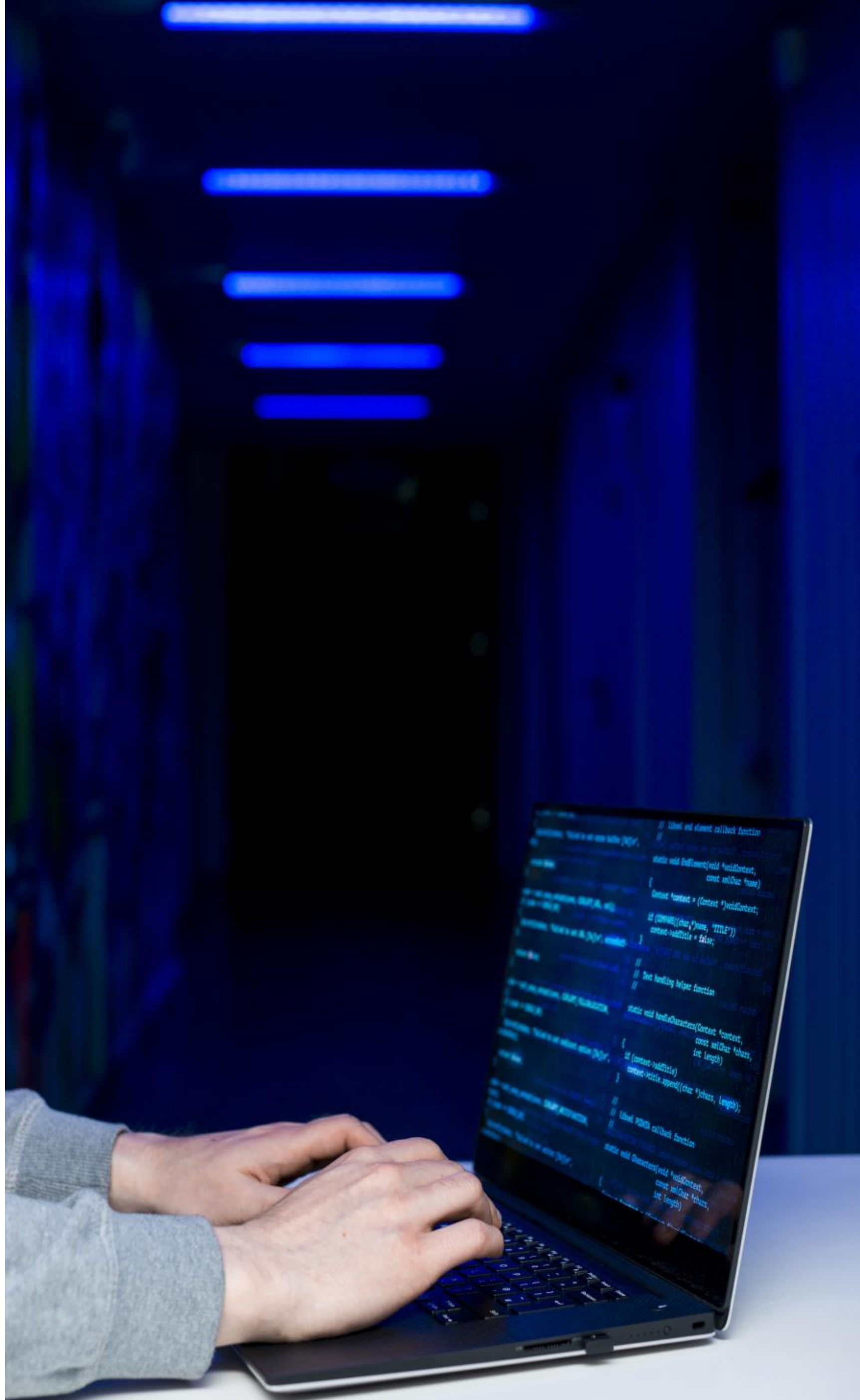


Image: freepik.com

Results – what have we learned?

- Top solutions dominated by the tree-based gradient boosting machines.
    - ➢ Out of nearly 250 participating teams only 10 beat our baseline model.
    - ➢ Feature engineering was pivotal to success.

- Prioritizing alerts using top solutions showed a great potential to optimize SoC operations:
    - ➢ 86% of suspicious events identified in only 20% of alerts with the highest predicted scores.
    - ➢ 25% reduction in the number of alerts that had to be manually investigated to detect all positive cases.

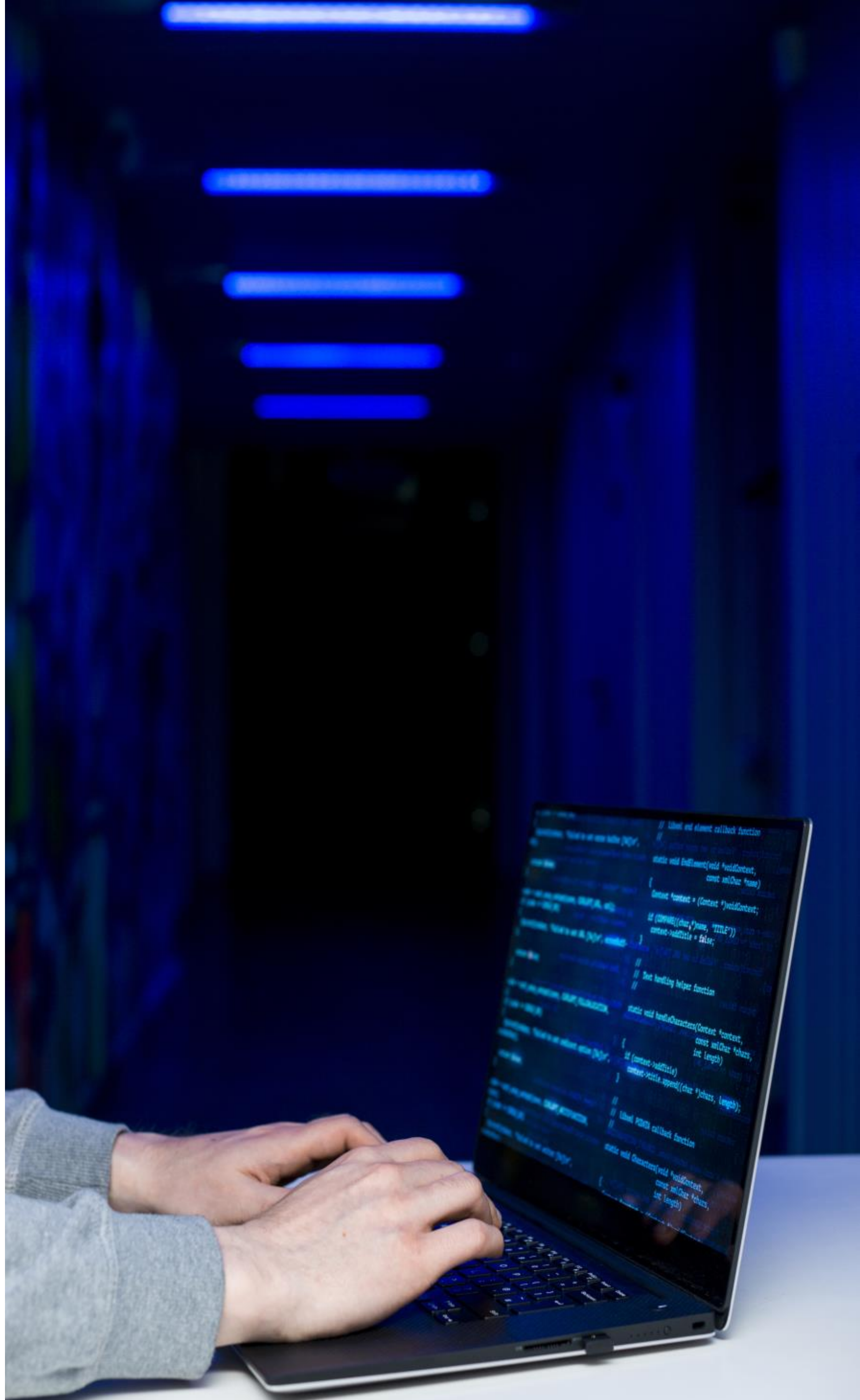


Image: freepik.com

XAI for the competition results?

1

Can we explain how a model works using only its predictions and competition data?

2

Can we analyze and diagnose errors in a submitted solution?

3

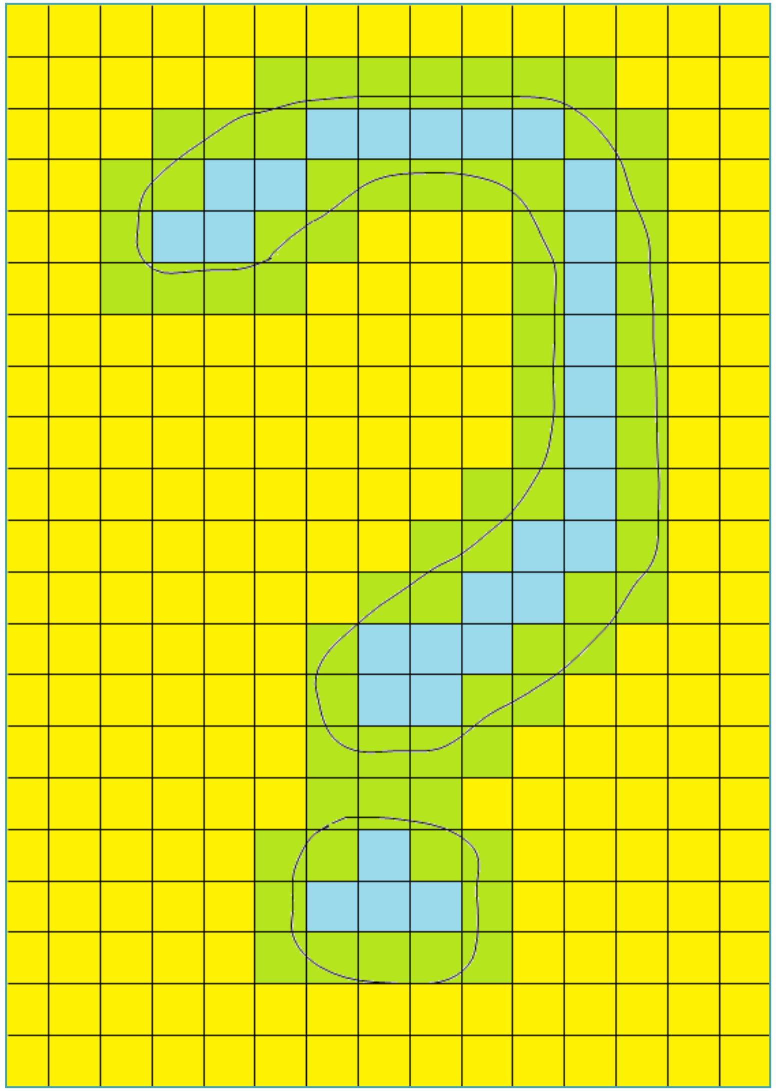
Can we measure a similarity of instances with regard to a particular solution?

4

Can we learn which errors are more important and use this information to create a better model?

Approximations of solutions – rough set theory basics

- In the rough set theory, concepts in data are described by their lower and upper approximations.
- Decision reducts – the base for RS approximations.
- Each decision reduct is an intrinsically interpretable prediction model – it corresponds to a set of rules.
- A collection of reducts can approximate predictions of an arbitrary ML model.
- Such a surrogate model can be interpreted using common XAI techniques.

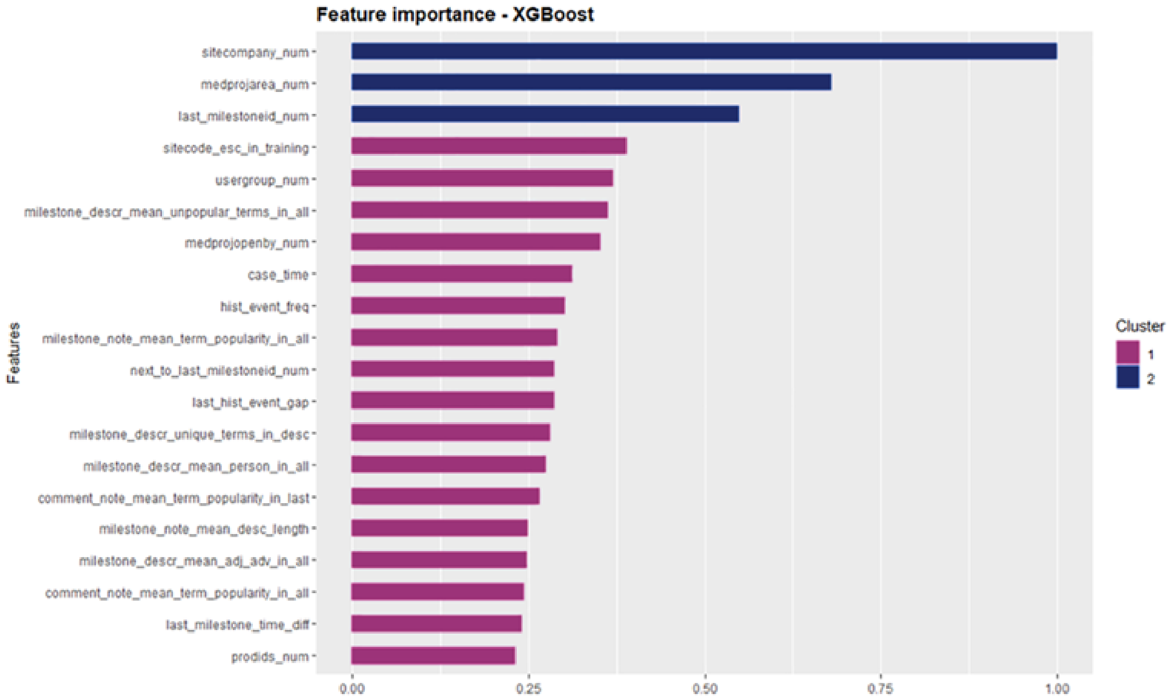


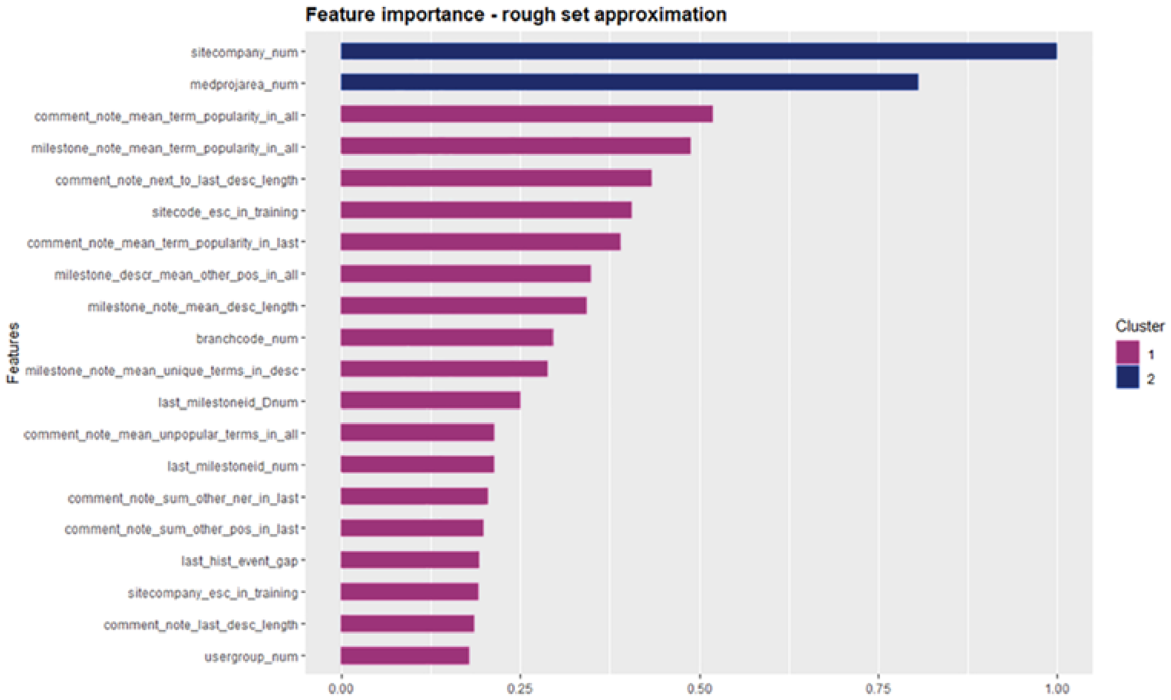
A reduct is a minimal subset of attributes (in the sense of inclusion) that holds the same amount of information about the decision attribute as the whole attribute set.

Reduct-based similarities and neighborhoods

- Reducts used to construct the surrogate model can be used to find data instances that are similar in the context of the investigated solution...
- and facilitate the computation of neighborhoods.
- We used the neighborhoods for a detailed diagnostic of prediction errors.
- We compute, so called, diagnostic attributes and used them to provide insightful information about the diagnosed data instance and the model.
    - For instance: *the model made an error, but the label of the instance is inconsistent with labels of similar data cases from the training set*.
- We constructed expert rules using selected diagnostic attribute values to indicate possible causes of errors made by investigated models.

Quality of reduct-based approximations



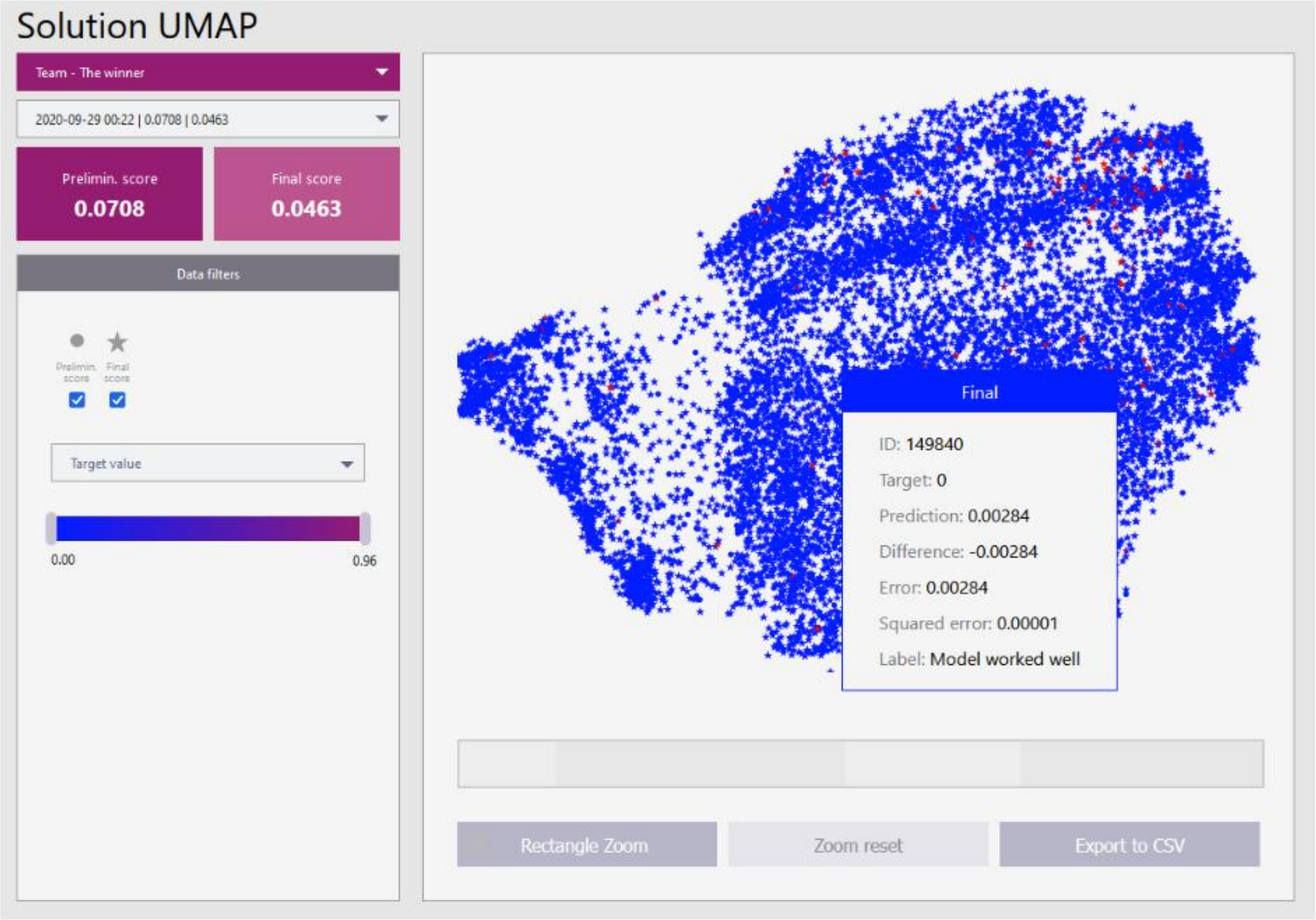


Correlation of importance coefficients ≈ 0.7

R2(approximations, predictions) ≥ 0.9

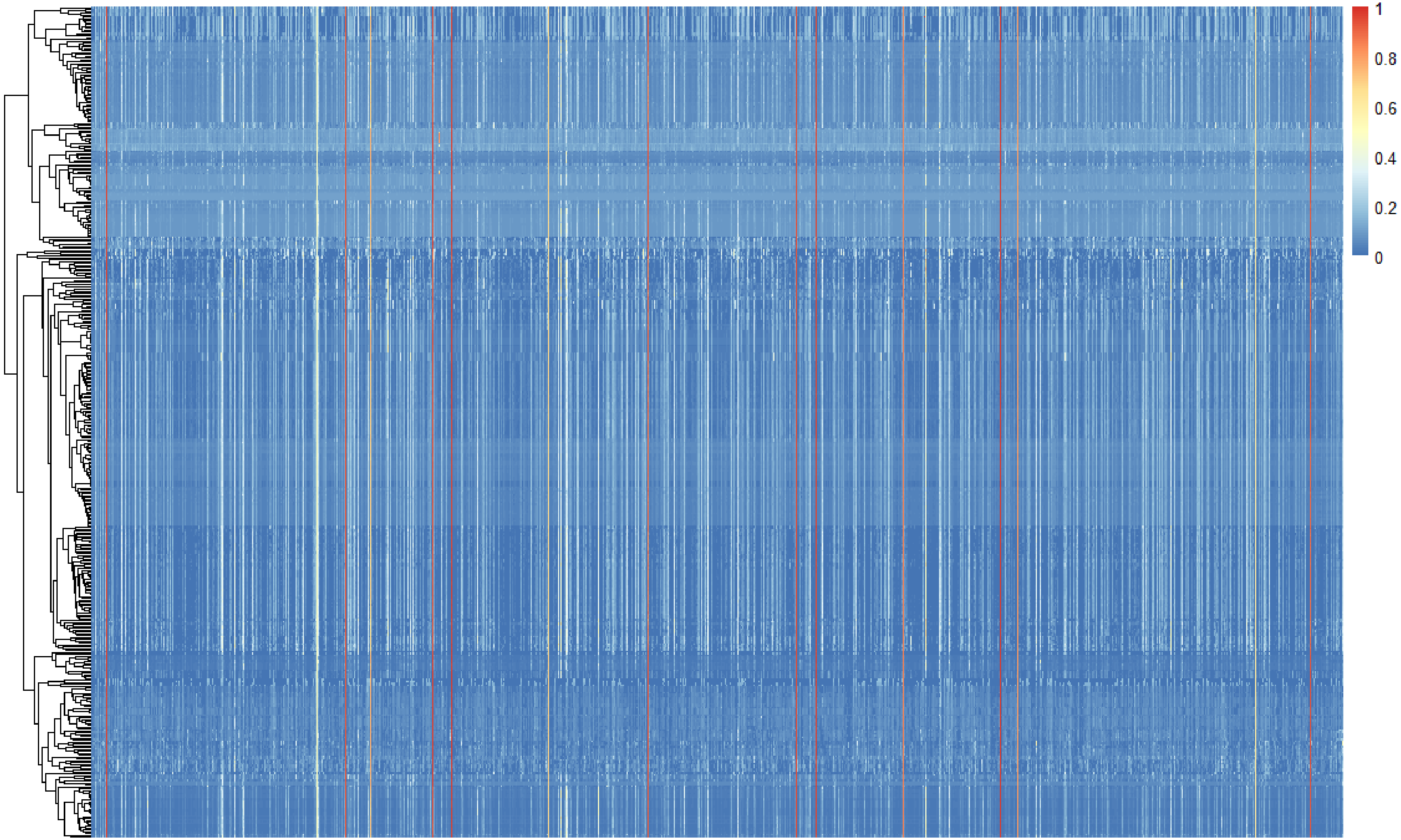
Similarity and neighborhoods

- We can use reducts to find similar instances with regard to a particular solution and compute neighborhoods.
- The similarity between neighborhoods can be used to visualize the data (e.g., using UMAP).
- We may visualize prediction errors and allow users for a more detailed investigation of selected instances.
- The neighborhoods can also be used in a more detailed error diagnostics.



Errors overview

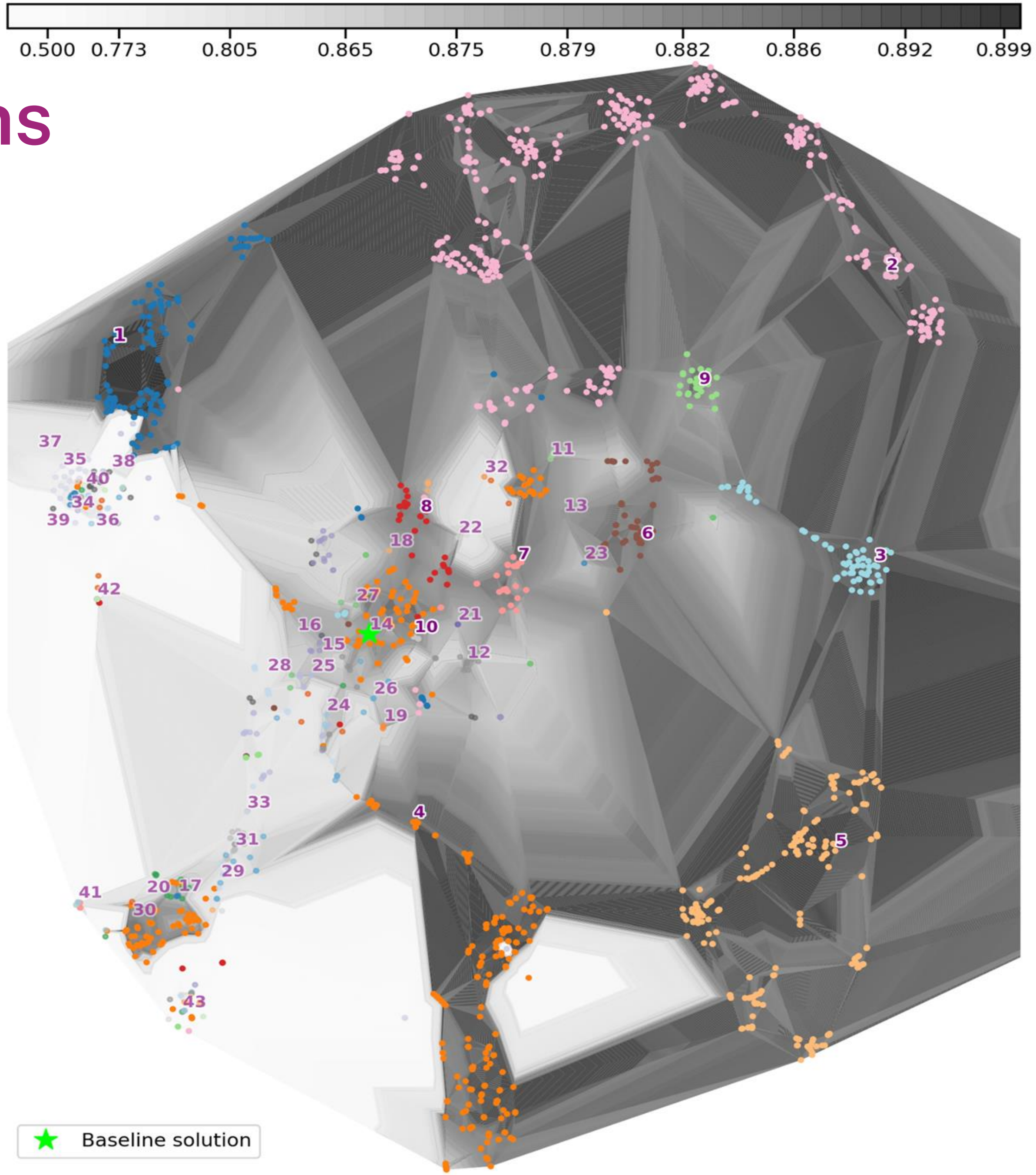
- Error heatmaps help us in finding the most difficult or problematic cases.
- We can identify potential issues in the test data.
- We can also cluster the solutions with regard to type of errors they make.



Visualizations of submissions

UMAP plot of submitted solutions:

- Embeddings of submissions represented as ranking vectors.
- Spearman's correlation used as the similarity metric.
- Shades in the background represent the estimated score of a submission from a given spot.



Conclusions

- Data mining competitions can help you in data science research:
  - You may outsource your research to ML community.
  - You may use results in post-competition research.
  - You may start cooperation with ML experts.
  - They are an objective benchmark for ML algorithms and a source of topics for publications.
- Model-agnostic analysis of prediction outcomes helps:
  - It makes the post-competition analysis of results more insightful.
  - .It enables discovering potential issues with the models and the used datasets.

KnowledgePit.ai as a recruitment support tool

- Data science challenges are an objective way of verifying practical data science skills.
- Recruiters may use it to make the initial assessments of candidates for related positions.
- We plan to launch a service dedicated to recruiters – which will allow to configure “quizzes” composed of data science tasks.
- We will use the BrightBox technology to provide insightful reports to recruiters.

Image: freepik.com

Thank you!

Andrzej Janusz
andrzej.janusz@qed.pl