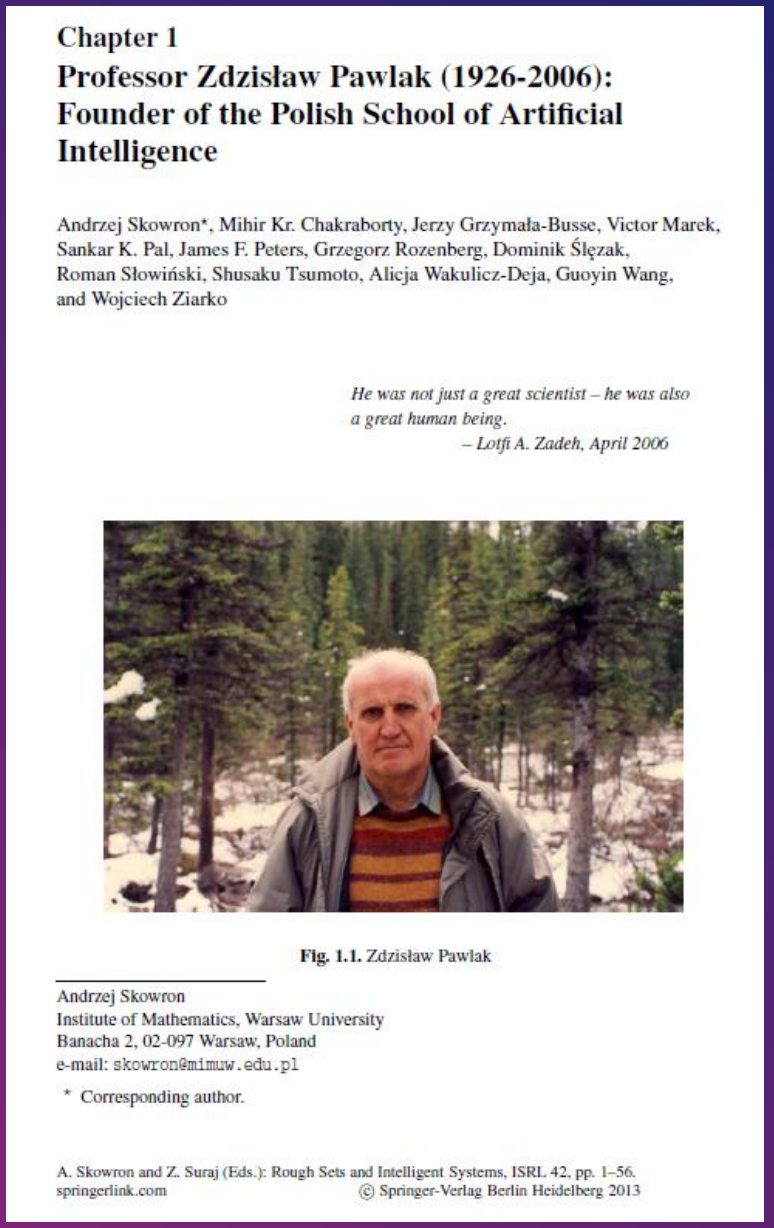
Theory of Rough Sets

- Rough Set Approximations
- Simple data and information representations
- Approximating complex phenomena with compacted and intuitive models

Publications authored by Professor Pawlak have still the highest number of citations when considering all scientists affiliated in Poland (https://data.mendeley.com/datasets/btchxktzyw/2)



Chapter 1

Professor Zdzisław Pawlak (1926-2006): Founder of the Polish School of Artificial Intelligence

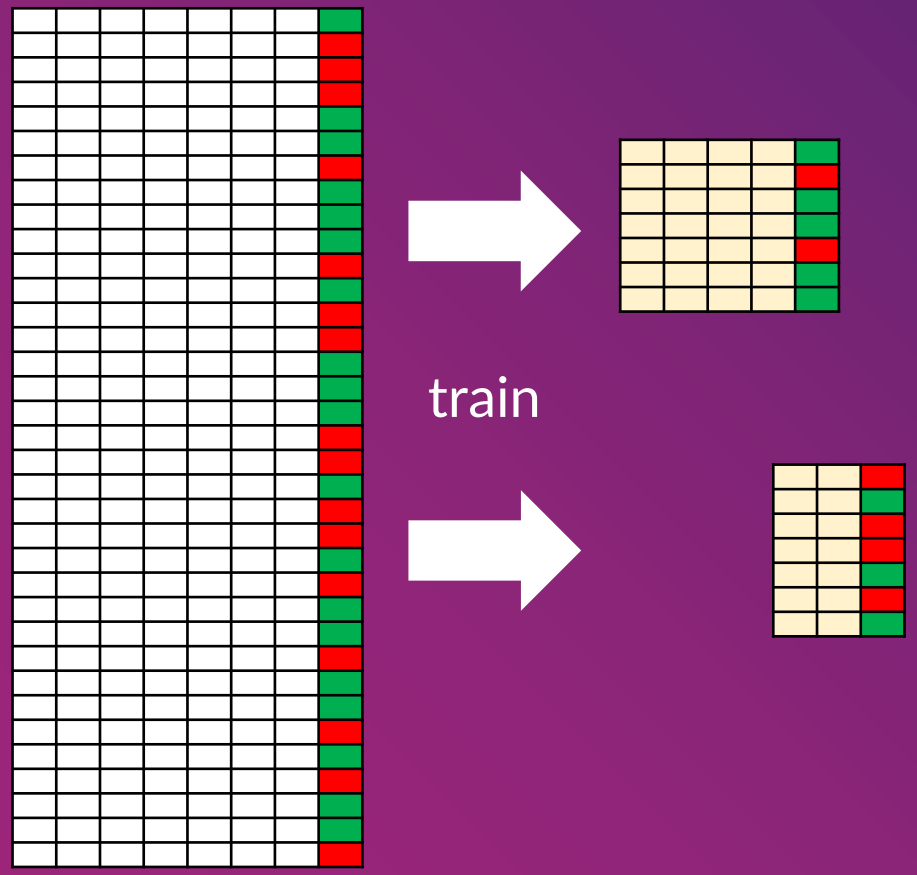
Andrzej Skowron*, Mihir Kr. Chakraborty, Jerzy Grzymała-Busse, Victor Marek, Sankar K. Pal, James F. Peters, Grzegorz Rozenberg, Dominik Ślęzak, Roman Słowiński, Shusaku Tsumoto, Alicja Wakulicz-Deja, Guoyin Wang, and Wojciech Ziarko

He was not just a great scientist – he was also a great human being.
– Lotfi A. Zadeh, April 2006

Andrzej Skowron
Institute of Mathematics, Warsaw University
Banacha 2, 02-097 Warsaw, Poland
e-mail: skowron@mimuw.edu.pl
* Corresponding author.

A. Skowron and Z. Suraj (Eds.): Rough Sets and Intelligent Systems, ISRL 42, pp. 1–56.
springerlink.com

Ensembles of Reducts



Information Sciences 451–452 (2018) 112–133

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A framework for learning and embedding multi-sensor forecasting models into a decision support system: A case study of methane concentration in coal mines

Dominik Ślęzak[a,*], Marek Grzegorowski[a], Andrzej Janusz[a], Michał Kozielski[b], Sinh Hoa Nguyen[c], Marek Sikora[b,d], Sebastian Stawicki[a], Łukasz Wróbel[b,d]

[a] Institute of Informatics, University of Warsaw, ul. Banacha 2, Warsaw 02-097, Poland
[b] Institute of Informatics, Silesian University of Technology, ul. Akademicka 16, Gliwice 44-100, Poland
[c] Polish-Japanese Academy of Information Technology, ul. Koszykowa 86, Warsaw 02-008, Poland
[d] Institute of Innovative Technologies EMAG, ul. Leopolda 31, Katowice 40-189, Poland

ARTICLE INFO

Article history:
Received 31 May 2017
Revised 18 February 2018
Accepted 3 April 2018
Available online 4 April 2018

Keywords:
Sensor data processing
Methane concentration forecasting
Sliding-window feature engineering
Feature subset ensemble selection

ABSTRACT

We introduce a new approach for learning forecasting models over large multi-sensor data sets, including the steps of sliding-window-based feature extraction and rough-set-inspired feature subset ensemble selection. We show how to integrate this approach with the major data-processing-related components of DISESOR – a decision support system which is a coherent and complete framework for exploring streams of sensor readings registered in underground coal mines. As a case study, we report our experiments related to the task of methane concentration forecasting. The contributions in this paper refer to both the analysis how the nature of sensor readings influenced the architecture of the developed system and the empirical proof that the designed methods for data processing and analytics turned out to be efficient in practice.

1. Introduction

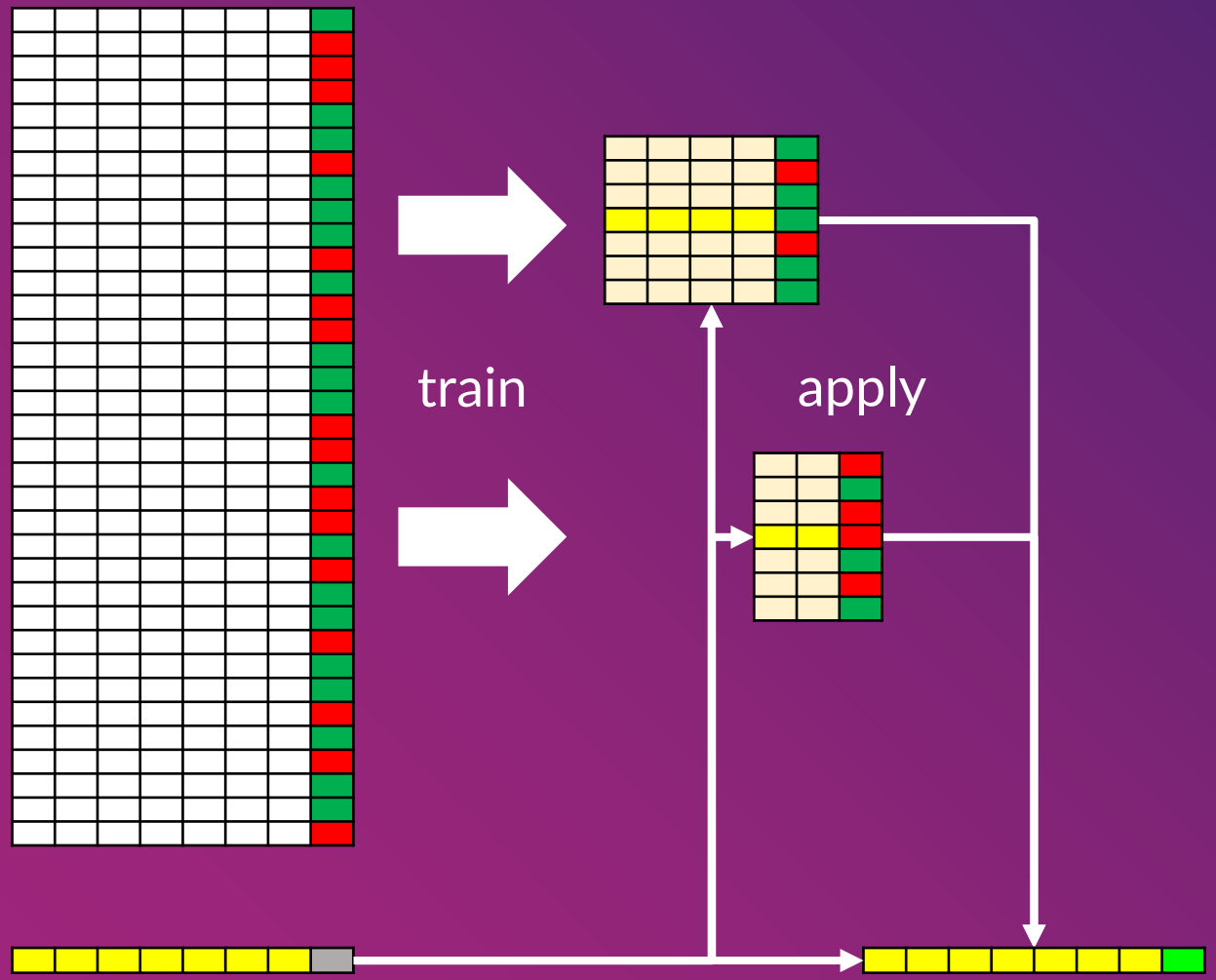
In the last decade, intensive growth in capabilities and popularity of analytical environments containing data mining solutions has been observed. Marketing, insurance, banking and finance, trade (especially e-commerce) and health care are the most popular applications. Less frequently, data mining methods are used to analyze and supervise industrial processes. The industrial monitoring systems usually produce multivariate streams of sensor readings for which performing standard preprocessing steps (such as data integration, data cleaning, feature extraction, etc.) is quite challenging. It is also difficult to construct and maintain forecasting models that should be used in an on-line fashion in industrial decision support systems. Nevertheless, potential benefits coming from intelligent utilization of this data source are truly huge.

In this paper, we investigate new approaches for learning forecasting models from multi-sensor data, for the purposes of monitoring natural hazards and industrial processes. We discuss them from the viewpoint of our decision support system – called DISESOR – which comprises of the expert system shell with the knowledge base that can be used together with the data incoming on-line, the feature engineering module that can derive the most meaningful statistics describing multivariate

* Corresponding author.
E-mail address: slezak@mimuw.edu.pl (D. Ślęzak).

https://doi.org/10.1016/j.ins.2018.04.026
0020-0255/

Ensembles of Reducts



Information Sciences 451–452 (2018) 112–133

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A framework for learning and embedding multi-sensor forecasting models into a decision support system: A case study of methane concentration in coal mines

Dominik Ślęzak[a,*], Marek Grzegorowski[a], Andrzej Janusz[a], Michał Kozielski[b], Sinh Hoa Nguyen[c], Marek Sikora[b,d], Sebastian Stawicki[a], Łukasz Wróbel[b,d]

[a] Institute of Informatics, University of Warsaw, ul. Banacha 2, Warsaw 02-097, Poland
[b] Institute of Informatics, Silesian University of Technology, ul. Akademicka 16, Gliwice 44-100, Poland
[c] Polish-Japanese Academy of Information Technology, ul. Koszykowa 86, Warsaw 02-008, Poland
[d] Institute of Innovative Technologies EMAG, ul. Leopolda 31, Katowice 40-189, Poland

ARTICLE INFO

Article history:
Received 31 May 2017
Revised 18 February 2018
Accepted 3 April 2018
Available online 4 April 2018

Keywords:
Sensor data processing
Methane concentration forecasting
Sliding-window feature engineering
Feature subset ensemble selection

ABSTRACT

We introduce a new approach for learning forecasting models over large multi-sensor data sets, including the steps of sliding-window-based feature extraction and rough-set-inspired feature subset ensemble selection. We show how to integrate this approach with the major data-processing-related components of DISESOR – a decision support system which is a coherent and complete framework for exploring streams of sensor readings registered in underground coal mines. As a case study, we report our experiments related to the task of methane concentration forecasting. The contributions in this paper refer to both the analysis how the nature of sensor readings influenced the architecture of the developed system and the empirical proof that the designed methods for data processing and analytics turned out to be efficient in practice.

1. Introduction

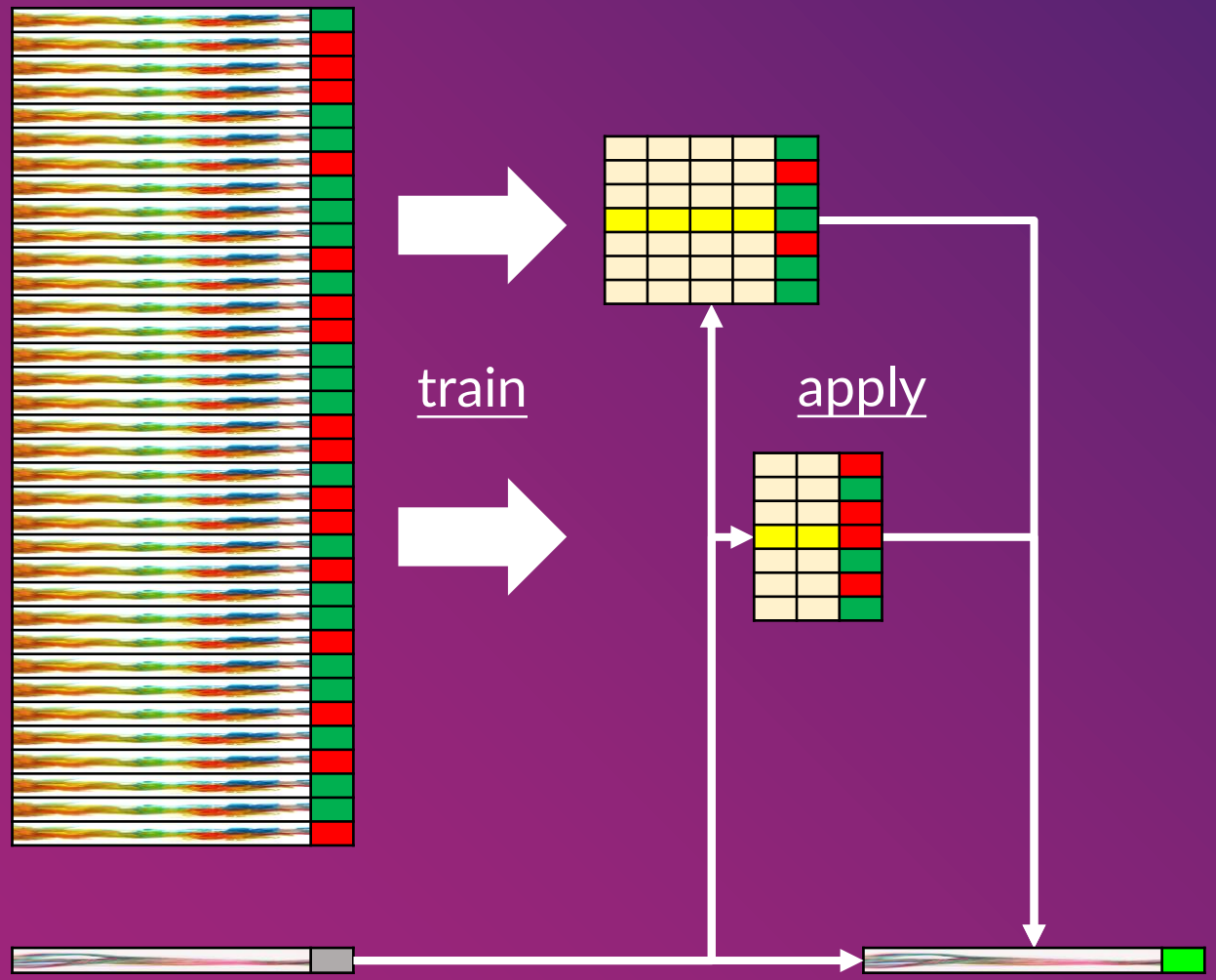
In the last decade, intensive growth in capabilities and popularity of analytical environments containing data mining solutions has been observed. Marketing, insurance, banking and finance, trade (especially e-commerce) and health care are the most popular applications. Less frequently, data mining methods are used to analyze and supervise industrial processes. The industrial monitoring systems usually produce multivariate streams of sensor readings for which performing standard preprocessing steps (such as data integration, data cleaning, feature extraction, etc.) is quite challenging. It is also difficult to construct and maintain forecasting models that should be used in an on-line fashion in industrial decision support systems. Nevertheless, potential benefits coming from intelligent utilization of this data source are truly huge.

In this paper, we investigate new approaches for learning forecasting models from multi-sensor data, for the purposes of monitoring natural hazards and industrial processes. We discuss them from the viewpoint of our decision support system – called DISESOR – which comprises of the expert system shell with the knowledge base that can be used together with the data incoming on-line, the feature engineering module that can derive the most meaningful statistics describing multivariate

* Corresponding author.
E-mail address: slezak@mimuw.edu.pl (D. Ślęzak).

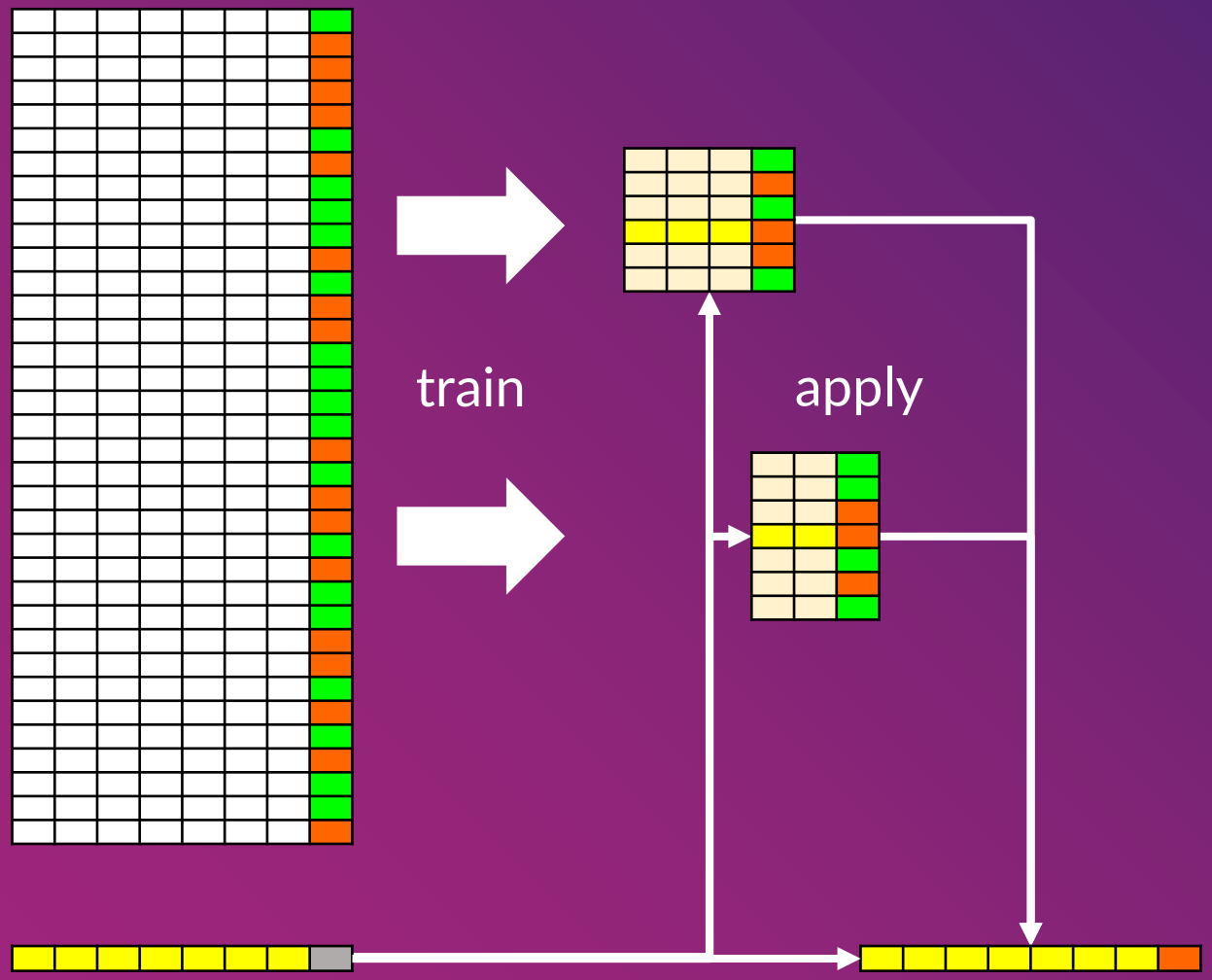
https://doi.org/10.1016/j.ins.2018.04.026
0020-0255/

Feature Magazine



- At its growth phase, the reduct creation algorithm repeats the following steps:
  1. Take a random subset of features / attributes
  2. Evaluate its elements
  3. Add the best element to a reduct candidate that is being constructed
- Step 1 can draw random subsets from a space of possible features, instead of a closed set of columns

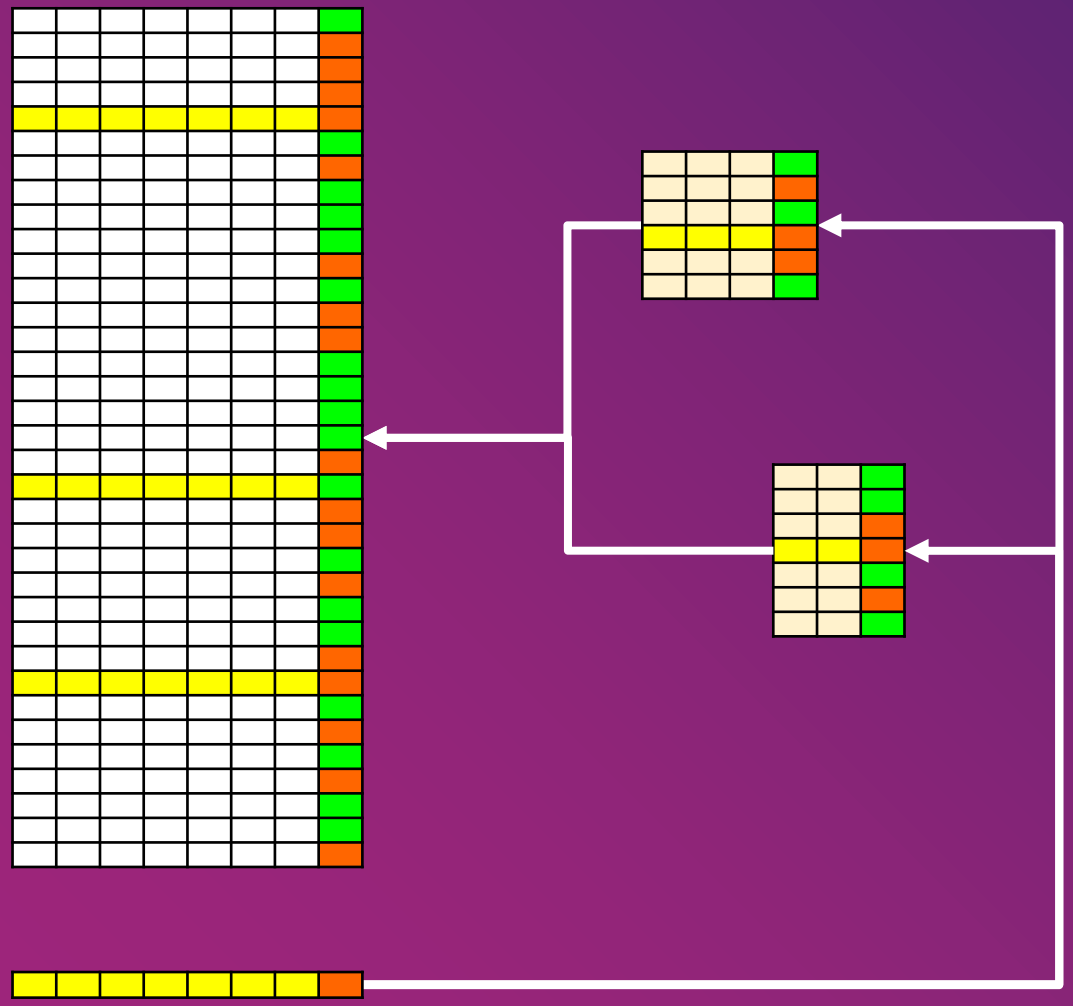
Reduct Approximators



BrightBox

The **BrightBox Technology** leverages our proprietary set of **diagnostic attributes** and **error taxonomy** to explain the **underlying reasons of errors** and recommend **corrective actions**.

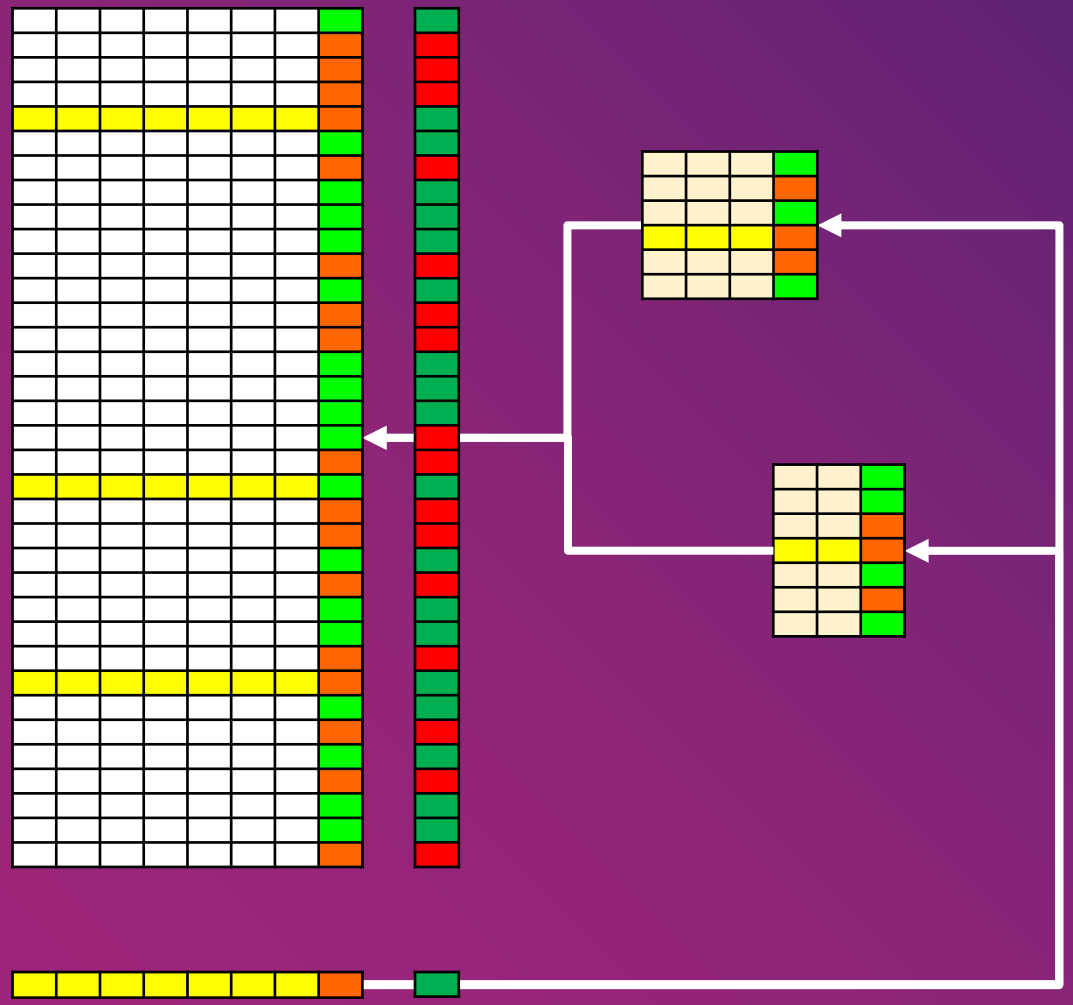
Reduct Neighborhoods



BrightBox

The **BrightBox Technology** leverages our proprietary set of **diagnostic attributes** and **error taxonomy** to explain the **underlying reasons of errors** and recommend **corrective actions**.

BrightBox Diagnostics



IEEE BigData 2020 Cup: Predicting Escalations in Customer Support

Predicting Escalations in Customer Support is a data mining challenge organized in association with the IEEE BigData 2020 conference. The task is to predict which cases in Information Builders, Inc. (ibi) technical support ticketing system will be escalated in the nearest future by customers. The competition is organized jointly by ibi (https://www.ibi.com) and QED Software (http://www.qed.pl/).

Manager: Andrzej Janusz (andrzej) — 259 teams — 2 months, 2 weeks ago

FedCSIS 2020 Challenge: Network Device Workload Prediction

FedCSIS 2020 Data Mining Challenge: Network Device Workload Prediction is the seventh data mining competition organized in association with Conference on Computer Science and Information Systems (https://fedcsis.org/). This time, the considered task is related to the monitoring of large IT infrastructures and the estimation of their resource allocation. The challenge is sponsored by EMCA Software and Polish Information Processing Society (PTI).

Manager: Andrzej Janusz (andrzej) — 162 teams — 6 months, 1 week ago

IEEE BigData 2019 Cup: Suspicious Network Event Recognition

Suspicious Network Event Recognition is a data mining challenge organized in association with IEEE BigData 2019 conference. The task is to decide which alerts should be regarded as suspicious based on information extracted from network traffic logs. The competition is kindly sponsored by Security On-Demand (https://www.securityondemand.com/) and QED Software (http://qed.pl/).

Manager: Andrzej Janusz (andrzej) — 293 teams — 1 year, 2 months ago

Clash Royale Challenge: How to Select Training Decks for Win-rate Prediction

Clash Royale Challenge is the sixth data mining competition organized in association with the Federated Conference on Computer Science and Information Systems (https://fedcsis.org/). This year, the task is related to the problem of selecting an optimal training data subset for learning how to predict win-rates of the most popular Clash Royale decks. The competition is kindly sponsored by eSensei, QED Software and Polish Information Processing Society (PTI).

Manager: Andrzej Janusz (andrzej) — 117 teams — 1 year, 6 months ago

ESENSEI Challenge: Marking Hair Follicles on Microscopic Images

ESENSEI Challenge is a data mining competition, whereby the task is to design an algorithm for accurate marking of follicle positions on microscopic images.

Manager: Andrzej Janusz (andrzej) — 49 teams — 2 years, 6 months ago

knowledgePit
A Data Challenge Platform

XAI

Thank you!

A Skowron, D Ślęzak: Rough Sets Turn 40:
- From Information Systems to Intelligent Systems. FedCSIS 2022: 23-34

A Janusz, D Ślęzak: KnowledgePit Meets BrightBox: A Step Toward Insightful Investigation of the Results of Data Science Competitions. FedCSIS 2022: 393-398