SEG 2018
Metals, Minerals, and Society
Keystone, Colorado, USA
September 22-25, 2018

NSERC-CMIC FOOTPRINTS

Assembling Machine Learning Workflows to Assist Mineral Exploration

[1,2] L. FELTRIN, [1] M. BERTELLI, [3] J. MCGAUGHEY, [4] W. A. MORRIS

[1] WESTERN UNIVERSITY, LONDON, ONTARIO
[2] LAURENTIAN UNIVERSITY, SUDBURY, ONTARIO
[3] MIRA GEOSCIENCE, MONTRÉAL, QUÉBEC
[4] MCMASTER UNIVERSITY, HAMILTON, ONTARIO

LaurentianUniversity UniversitéLaurentienne
HARQUAIL EARTH SCIENCES
MERC Mineral Exploration Research Centre AT THE HARQUAIL SCHOOL OF EARTH SCIENCES
NSERC-CMIC FOOTPRINTS
canadamining innovationcouncil
NSERC CRSNG

Machine learning

NSERC-CMIC FOOTPRINTS

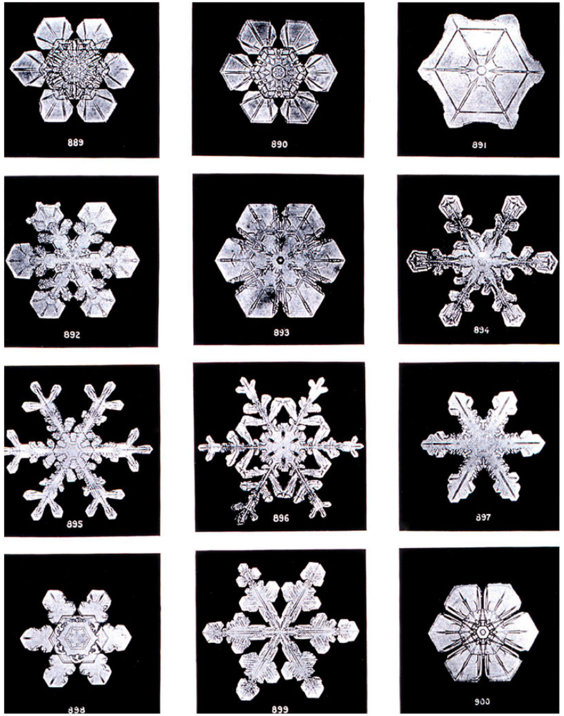
- Role and purpose of machine learning in modern society?
- Foreseeable impact that ML methodologies can lead to?
- Machine learning to automate the recognition of specific geological processes?

- Machine learning accelerates the pace of AUTOMATION reducing human labor.
- It's a general purpose technology like the invention of the steam engine or electricity.
- Offers better decisions if compared to discipline experts.
- Increased Usability.

The challenge of automation

NSERC-CMIC FOOTPRINTS

The problem of natural complexity



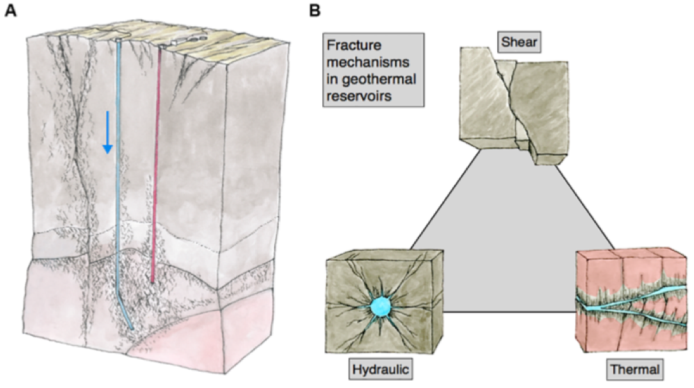
Snowflakes "Emergence"

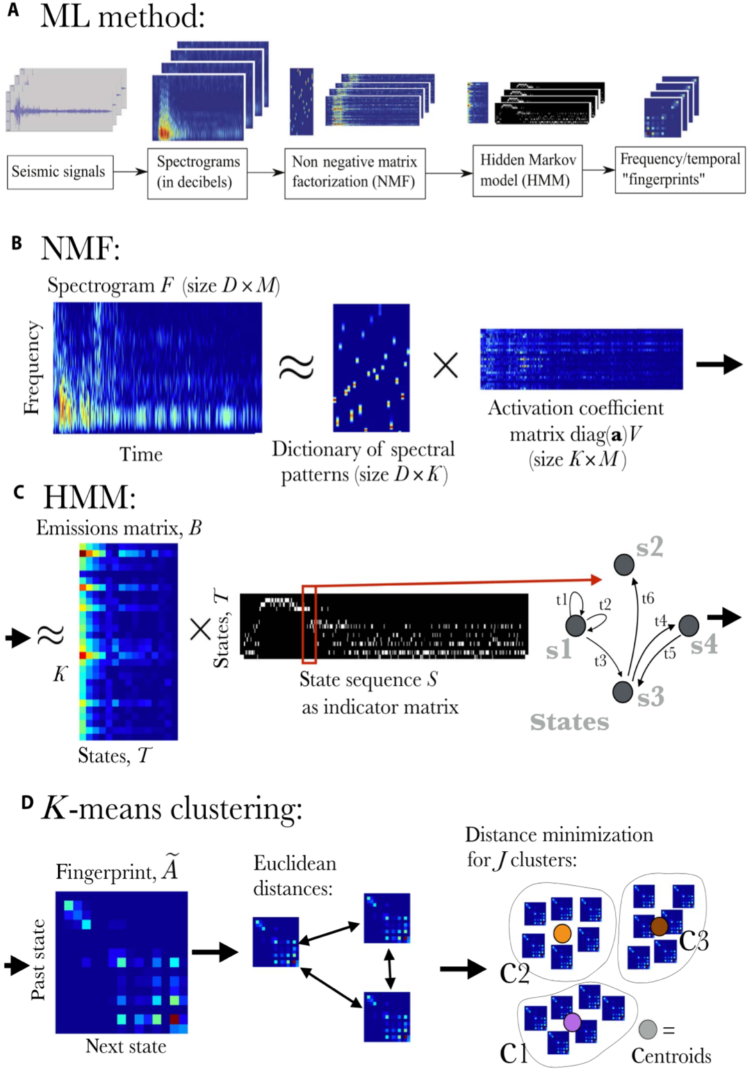
Flakes 2000 – the computational beauty of nature

- **Development of new breeds of more "organic" ML applications is necessary to achieve a more advanced level of automation.**
- **Example: training simple neural network architectures is a superseded paradigm. Trend to more complex and specialized networks (e.g. Deep Learning).**
- **A modular array of specialised algorithms can be pipelined to organize and drive automation.**
- **Complex systems are required to obtain more automations of automation paradigms and increase the applicability of ML technology leading to more flexible/usable mathematical models superior to traditional methods.**

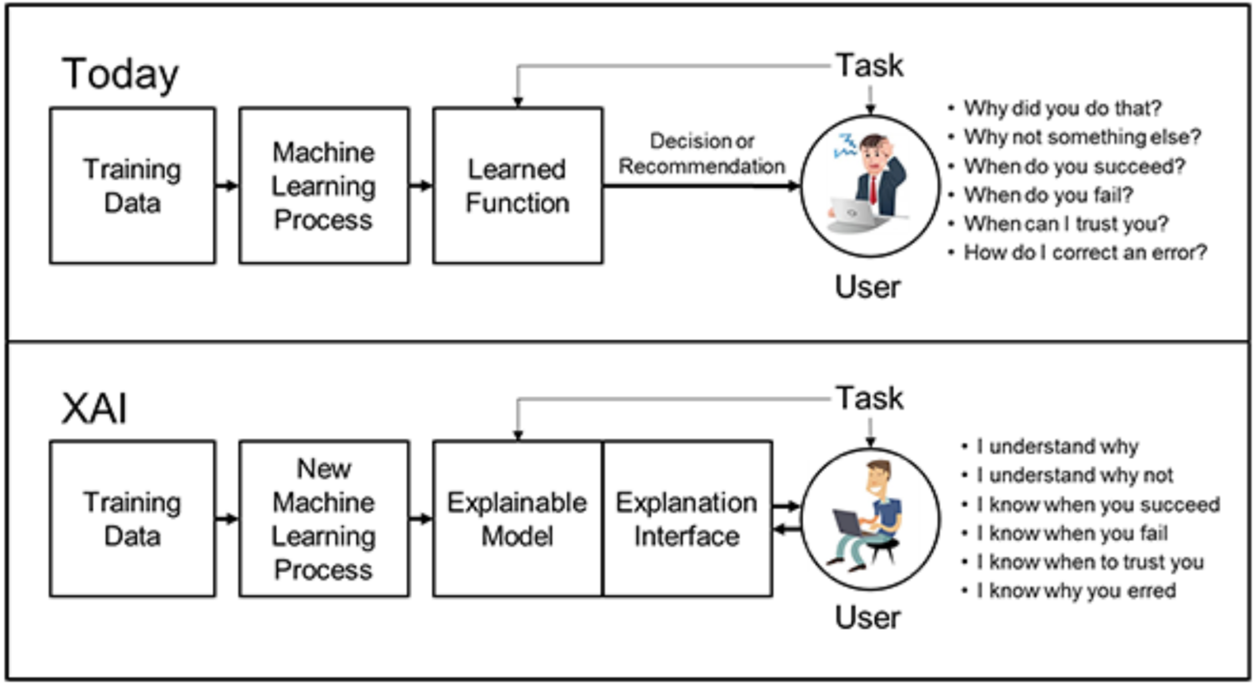
Automation requires adaptable & interactive workflows

- One powerful algorithm is insufficient. More complex architectures accommodating the input variability and processing needs are becoming a standard in geoscience [automation of the automation concept].
- Example of complex but relatively successful workflow (see Holtzman et al., 2018, Sci. Adv.)





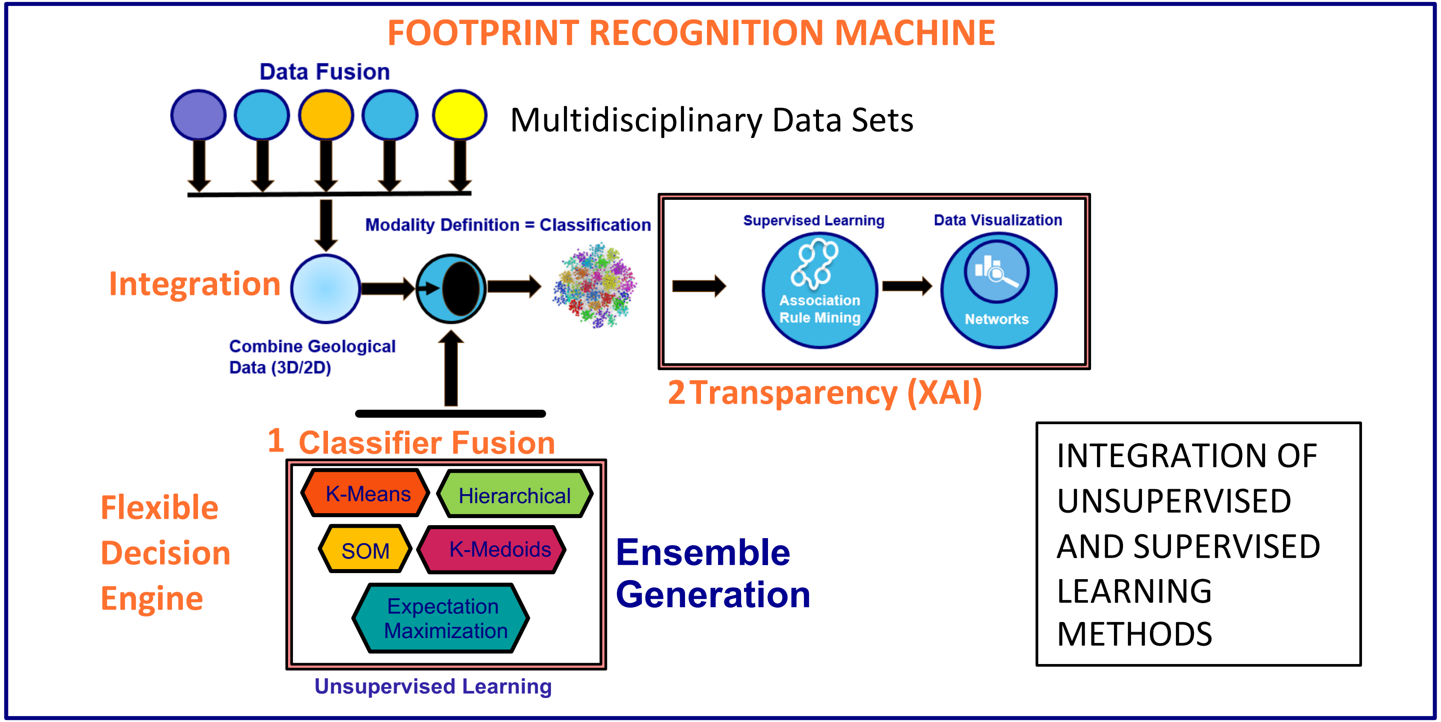
Automation may lead to a relatively "blind" approach reducing the capacity of the scientist to derive insight from the results and understand the machine reasoning and decisions



Advanced Data Visualization Solutions

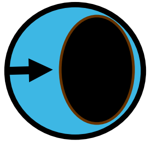
Explainable Artificial Intelligence (XAI) Gunning (2017) DARPA

If an increased complexity of workflows allows more powerful applications, then advancement should consider this type of approaches also in a mineral exploration application



- **Our proposed model attempted this operation by looking at constructing a data pipeline that uses a variety of algorithms to solve specialized tasks**

Research Problem: Geospatial Footprint Segmentation (Zone Attribution)



Progress domains in cluster analysis

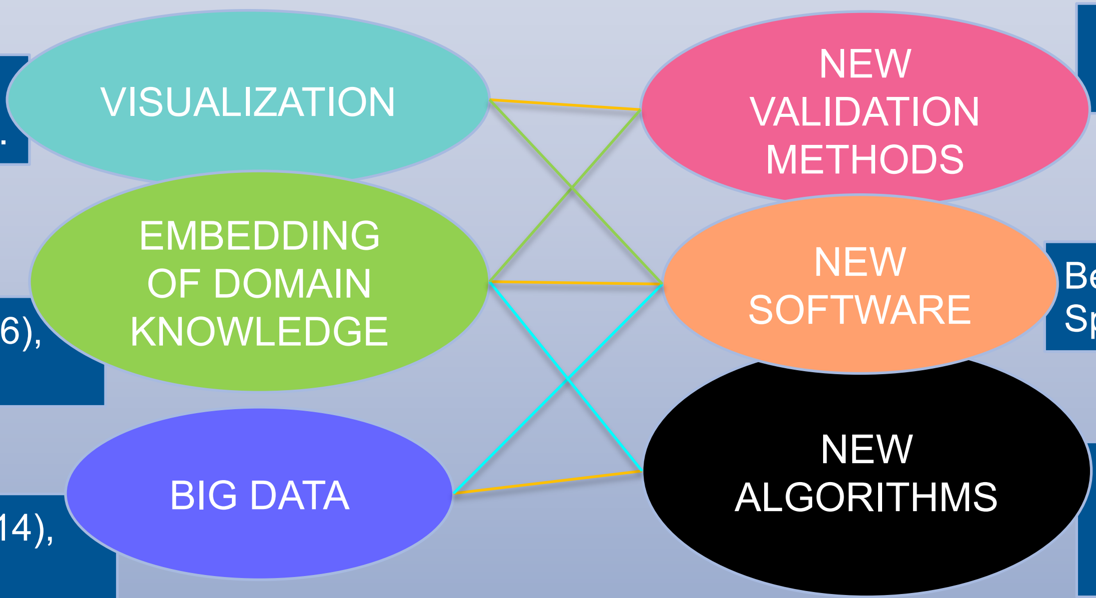
Areas with most significant progress

Areas with significant progress

Viks et al., (2003), Bioinf.

Datta and Datta (2006), BMC Bioinf.

Kurasova et al. (2014), ICTAI, IEEE



Wu et al. (2013), Exp. Sys.

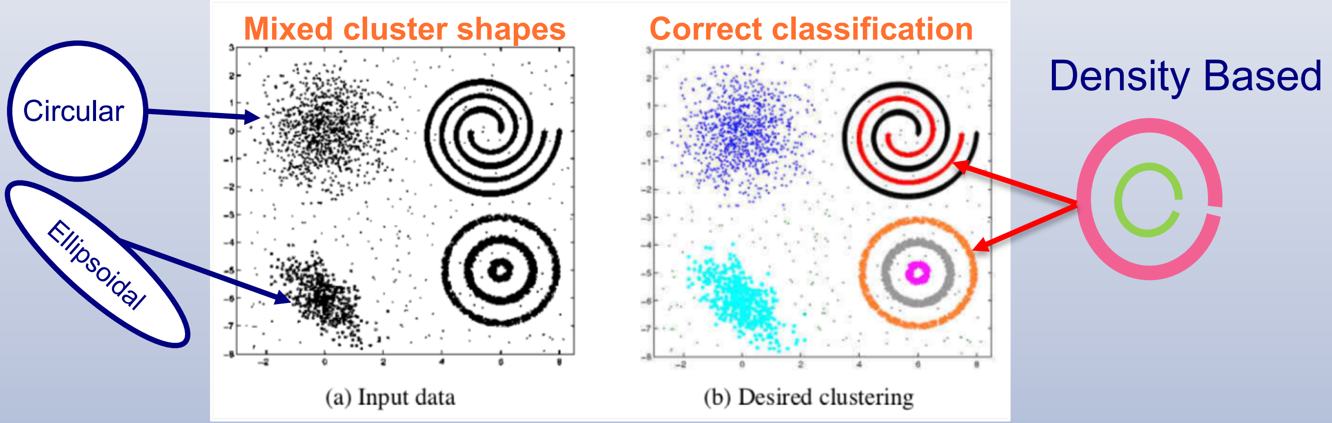
Berthold et al. (2010), Springer

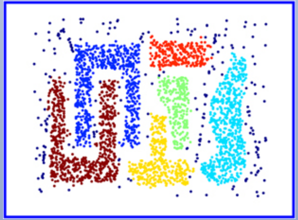
Narmadha and Rathipriya (2016) IEEE Expl.

Develop flexible architectures that can accommodate for rapid changes in the clustering arena

The problem of optimization: shape of the clusters non-homogeneous, non-separated

- **Labelling of different cluster shapes. Each algorithm is good at classifying specific topologies**





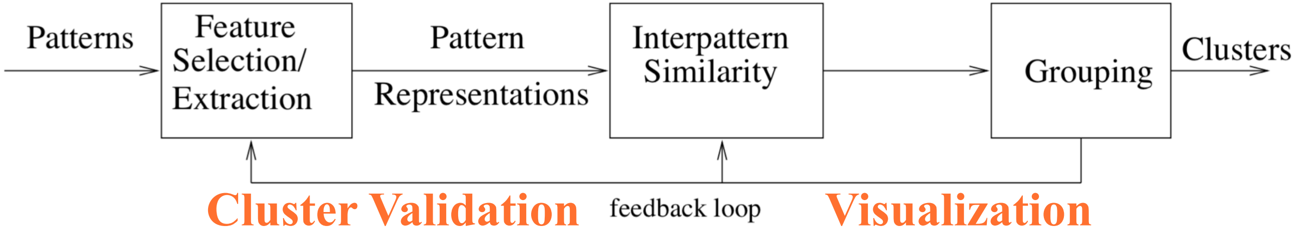
Types of clusters with different morphology in bi-variate space. It is very difficult to obtain the classification reported in (b) with a single clustering algorithm. Visualization is critical.

Jain, 2008, ICPR

Cluster analysis iterative workflow

NSERC-CMIC FOOTPRINTS

Stages in cluster analysis & flexibility in terms of parametrization



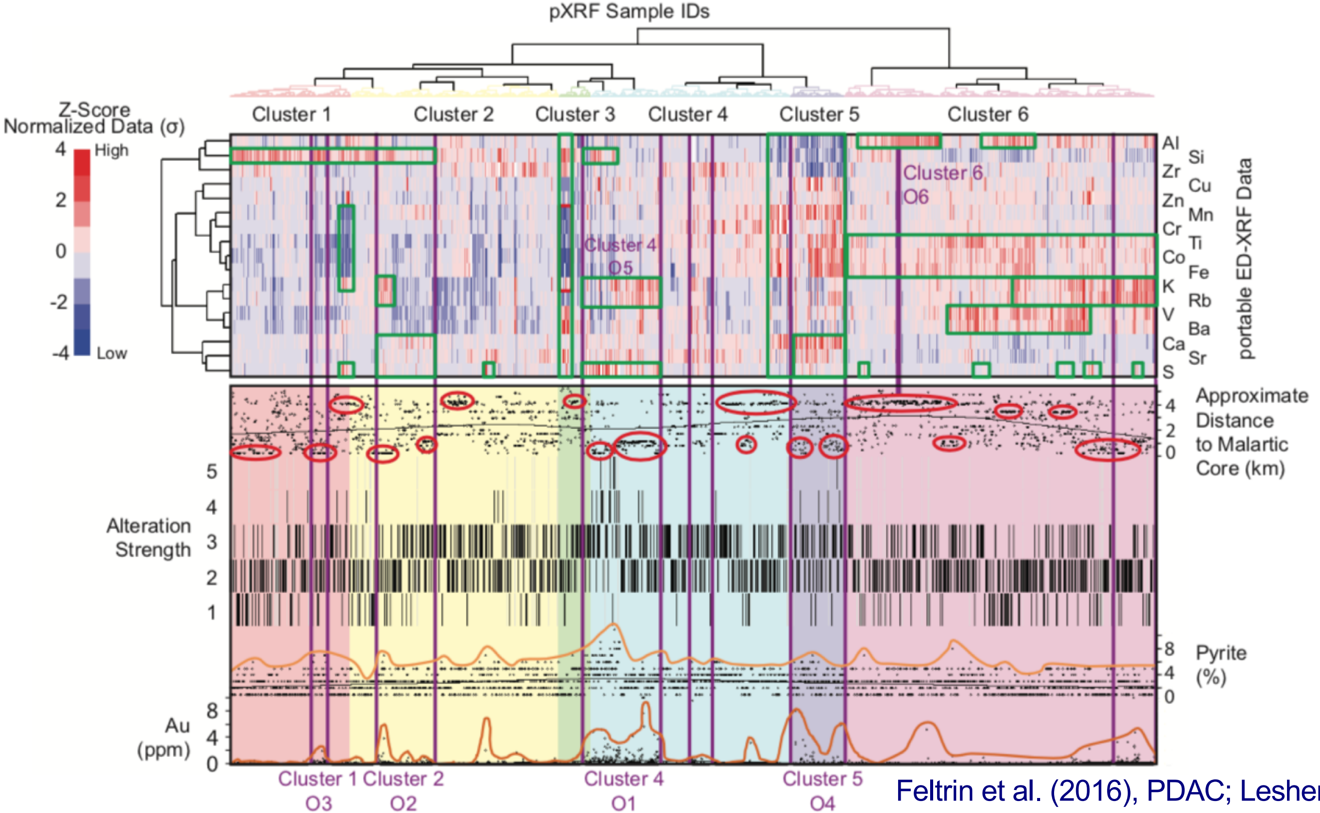
Modified from Jain et al., 1999, ACM

- Feedback loop allows the generation of multiple clustering outputs (useful for comparative purpose)
- Validation can be used to rank the quality of each clustering output
- Common changes involve either the clustering algorithm or its parameters:
  - (1) A transformed (different) input data set
  - (2) A different similarity metric (Euclidean, Manhattan, Minkowski, etc.)
  - (3) A different clustering algorithm (K-Means, K-Medoids, Model Based, etc.)

Malartic Au deposit, Quebec (p-xrf) footprint

NSERC-CMIC FOOTPRINTS

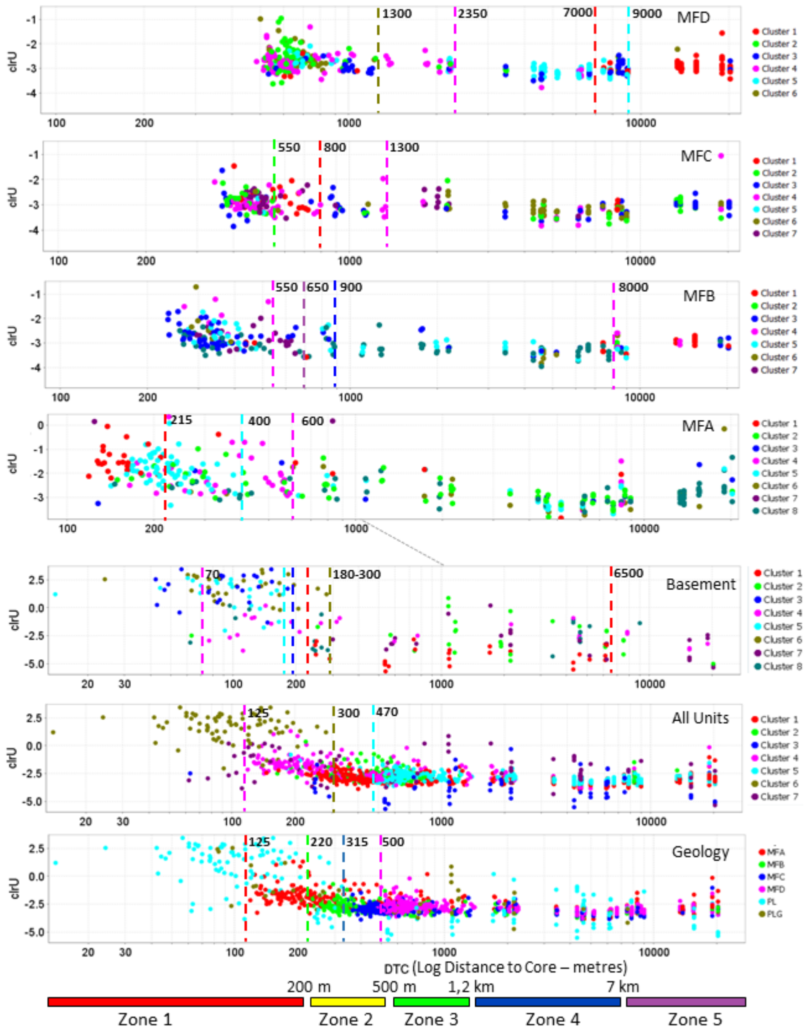
Coupling statistical visualization with external domain knowledge



Feltrin et al. (2016), PDAC; Lesher et al. (2017), DMEC

K-Means groups – Millennium deposit

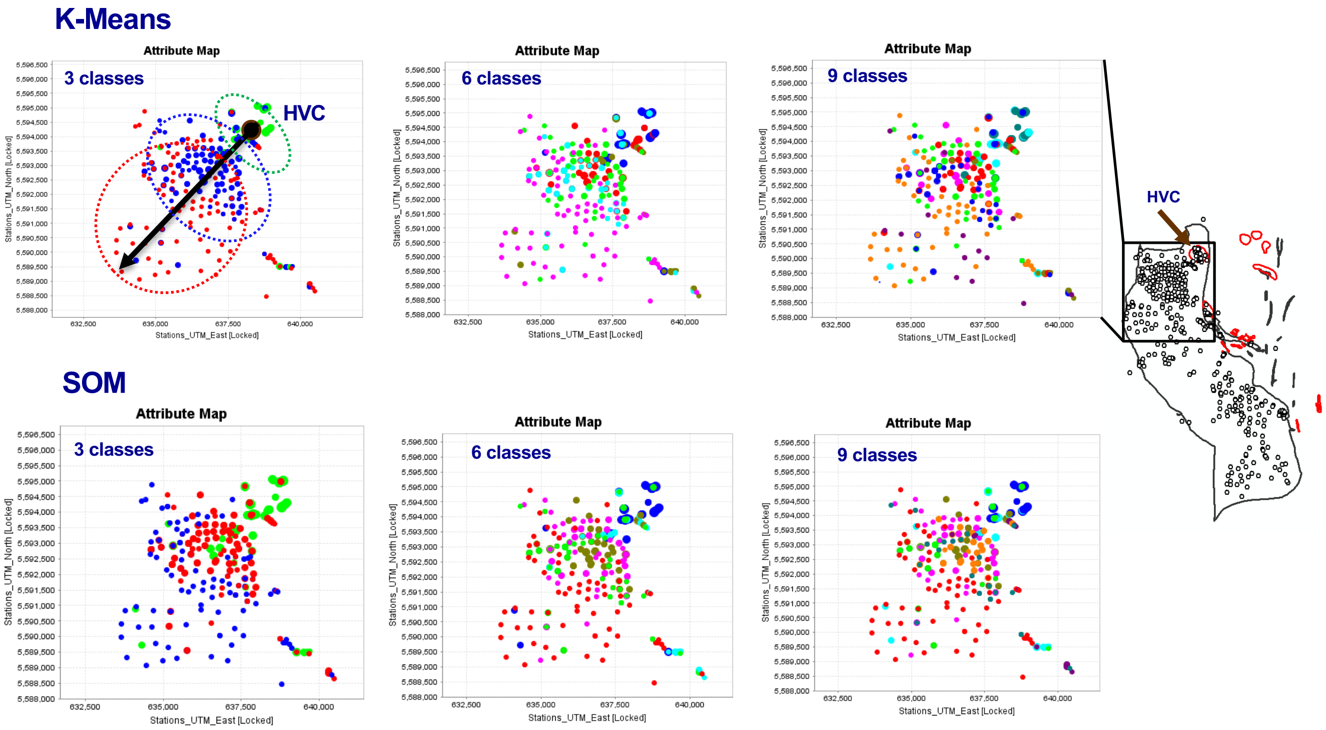
- K-Means clustering of selected data was carried out to define semi-automatically discrete zone boundaries.
- K-Means cluster analysis was used on distinct lithological sub-categories to limit the effect of lithology on the results of cluster analysis.
- 5 Zones were identified by comparing boundaries defined in each sub-unit.



Feltrin et al. 2016, GEOSTATS

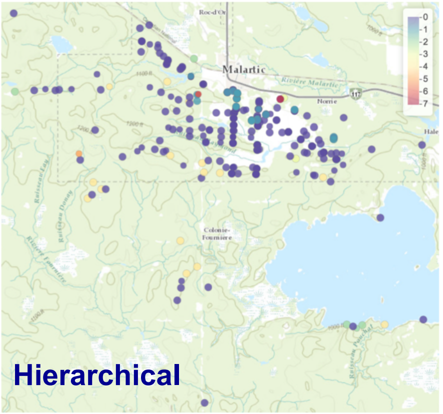
K-Means groups – HVC deposit

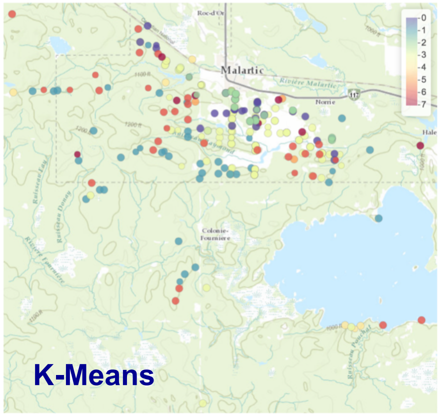
Comparison of classifiers with different K=[3,6,9]. The ternary classification is the easiest to interpret and shows differences in the sub-clusters at the 100-250 m scale (SOM vs K-Means). Zones show minimal population mixing in the ternary clustering. Generalization is similar for the two classifiers in this case. An increase of K causes a reduction of the number of samples per cluster reducing the classification performance even if 5 fold cross-validation was used.

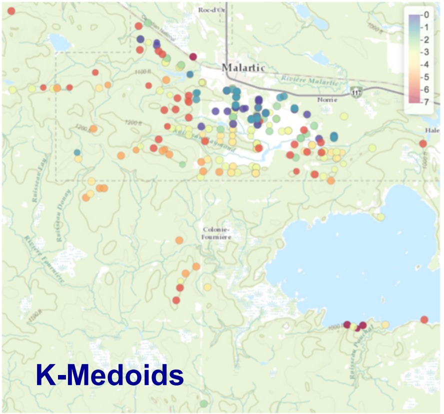


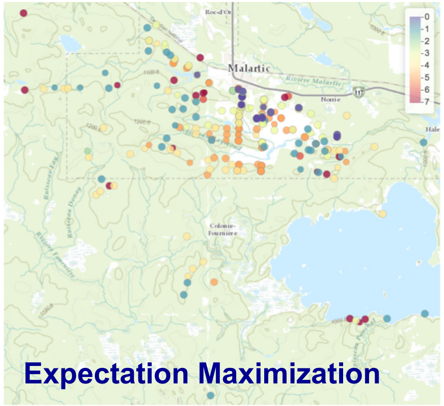
Classifier 2D mapping – Malartic Au deposit

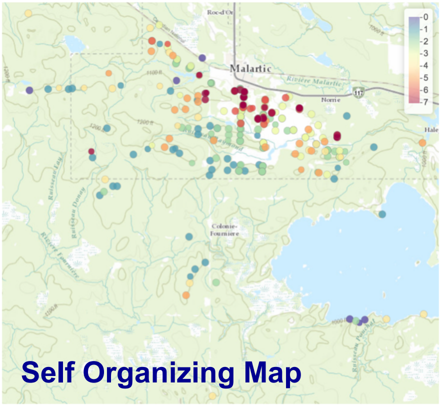
Although it is relatively easy to conduct a classification with modern analytics software, one outstanding problem is the <u>variability of classifier performance,</u> which is internally dependent on the classification algorithm selected to accomplish a discretization and also externally dependent on the input data supplied to the algorithm (*free lunch theorem*).









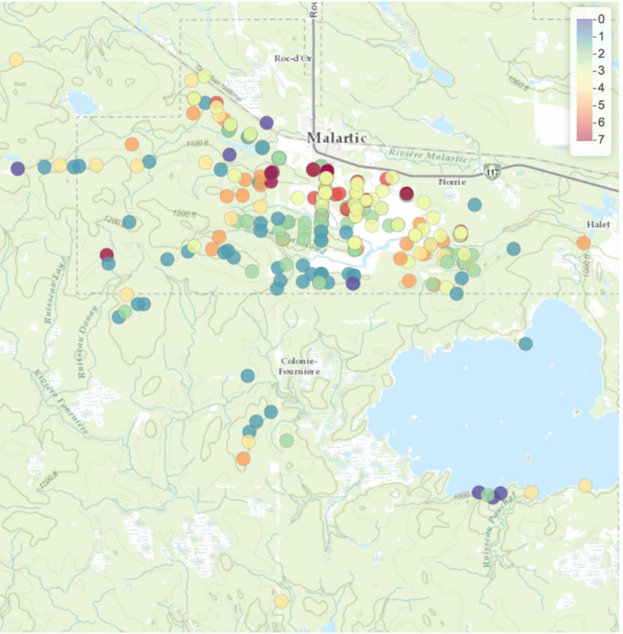


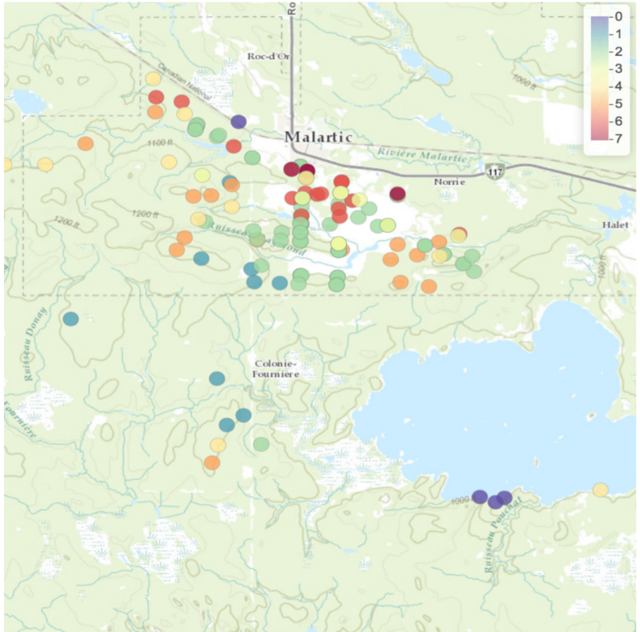
- **Visualization of classification results as 2D GIS Map.**
- **Results show some of the problems encountered: mainly heterogeneous cluster size and labelling.**
- **A certain level of agreement is also observed (despite labelling).**

Which one is right?

Solution filtering

- **Classifier Integration was implemented to mitigate classifier dependence.**
- **Effect of filtering of samples using combined scoring of sureness and agreement weight (not meeting the 50% majority vote). Samples colored using the SOM classifier.**

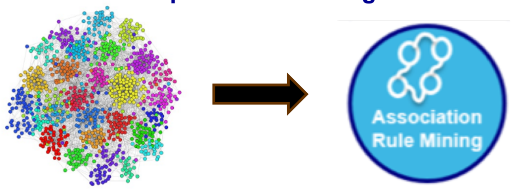




- **Overall outlier reduction process**

Research Problem: Zone Investigation Applying Data Mining with Advanced Machine Learning Tools

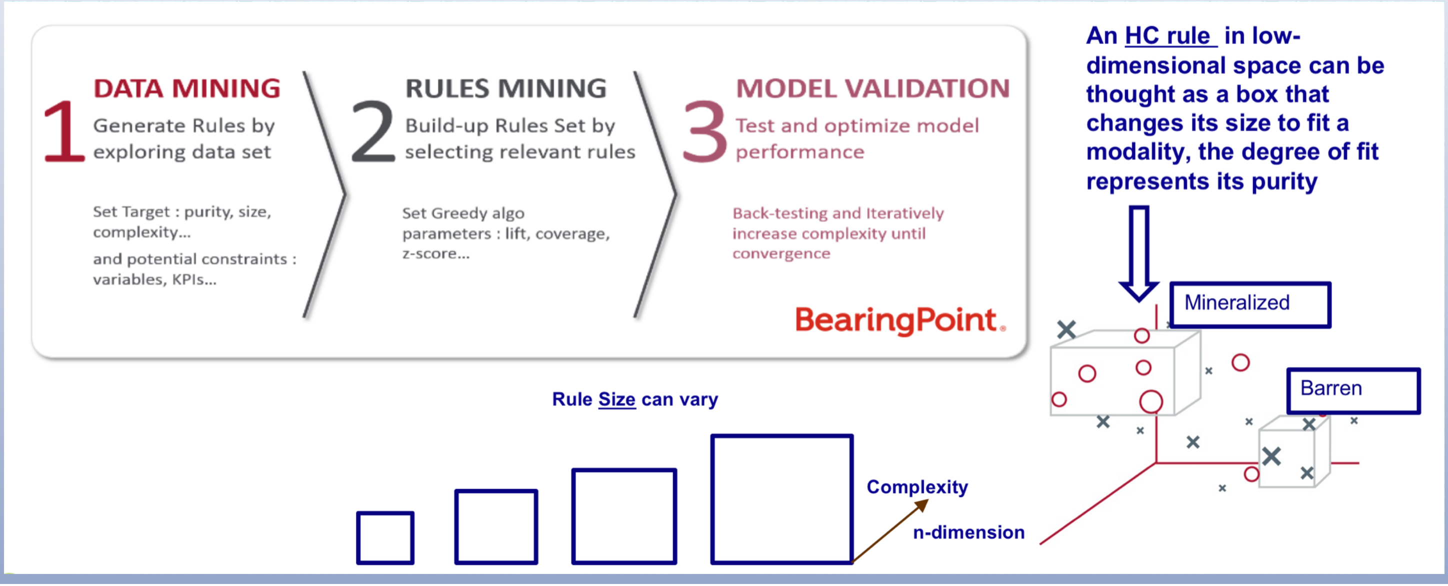
Supervised Learning



Supervised learning

NSERC-CMIC FOOTPRINTS

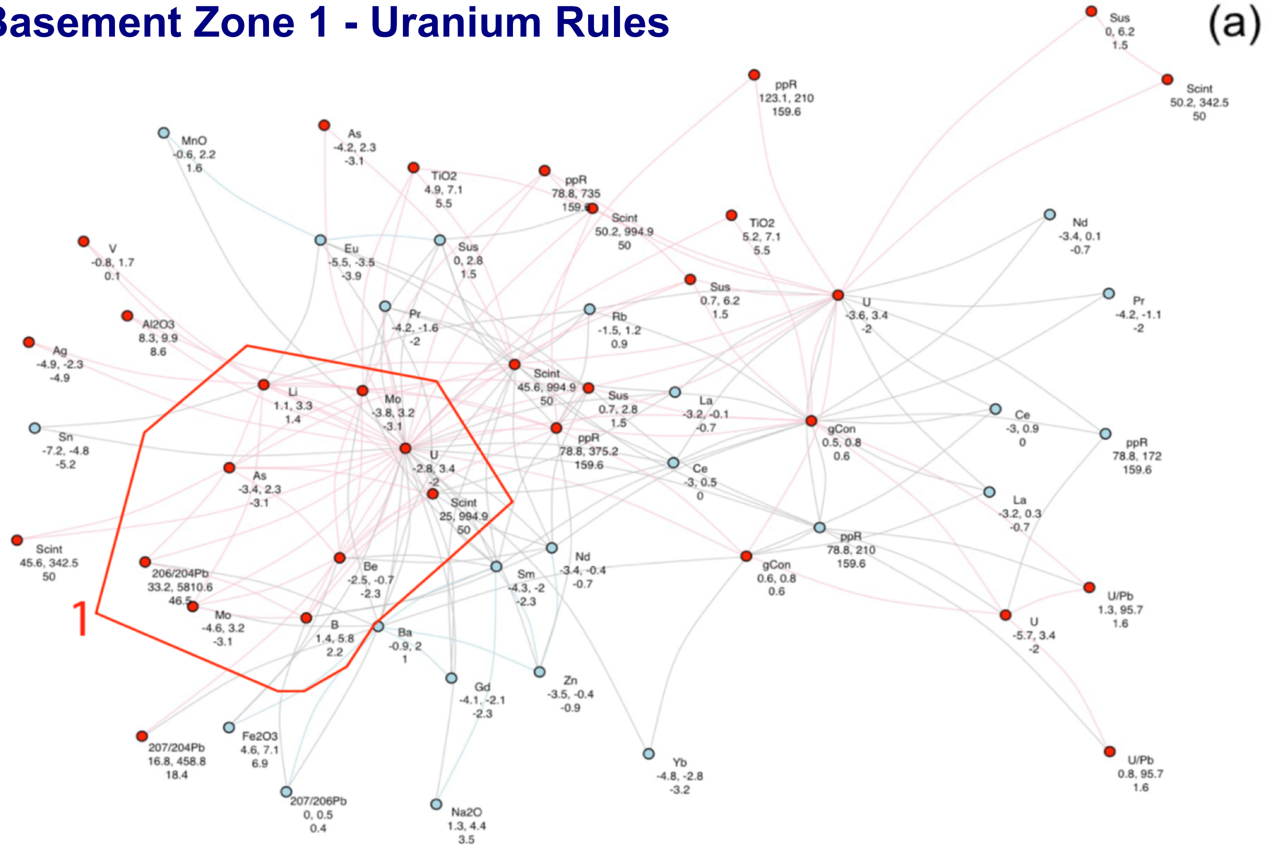
- Understanding what is controlling the clustering provides support for the interpretation of the results and represents a key aspect of footprint recognition.
- To accomplish this task we used Data Mining to define the characteristics that typify each group/class.
- Hypercube a French software with a history in solving healthcare related problems (Loucoubar 2012) was applied to test the formulated hypothesis that association rule learning can be a fruitful solution to more explicitly expose characteristics that determined the clustering observed. DISTINCTIVE PATTERNS



Zone investigation using association rule mining

Inference concerning mineralogy is possible

Basement Zone 1 - Uranium Rules



Feltrin et al., in preparation

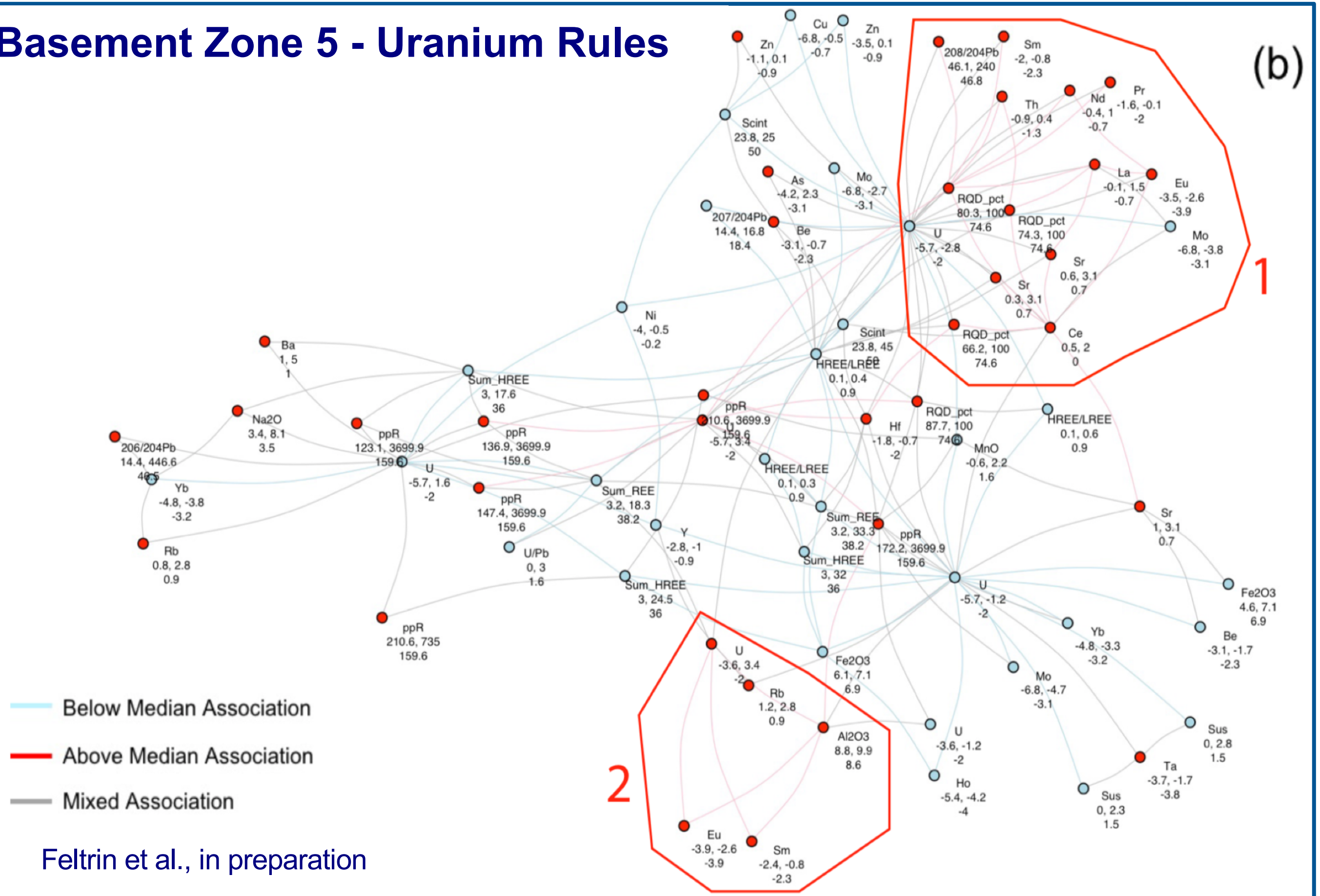
- Geochemical: Ag, V, Al, 206/204 Pb, 207/204Pb, As, Li, B, Be, Mo, (U), Ti, U/Pb
  Geophysical: Scint, ppR, Sus, gCon
- Geochemical: Sn, Mn, Fe, 207/206 Pb, Eu, Pr, Ba, Na, Gd, Sm, Zn, Nd, Rb, Yb, Ce, La, Nd, Pr
  Geophysical: Sus, ppR

canadamining innovationcouncil

Zonal statistics and footprint recognition

- Confirm known transitions
- Discover unknown transitions

Basement Zone 5 - Uranium Rules



Feltrin et al., in preparation

Geochemical: 206/204Pb, Rb, Na, Ba, Eu, Sm, U, Al, Be, As, Zn, Hf, 208/204Pb, Th, Sr, Ce, Nd, Pr, La, Ta
Geophysical: ppR, RQD

Geochemical: Yb, U, HREE, U/Pb, Ni, Y, Fe, HREE/LREE, 207/204Pb, Cu, Zn, Mo, Ho, Mn, Be
Geophysical: ppR, Scint, Sus

Concluding remarks

- We provided a series of snapshots of how our machine learning workflow operates that emphasize its organic and transparent nature.
- We demonstrated plausible ways of increasing the embedding of domain knowledge to improve solution optimization and showed how interpretations can be carried out efficiently using association rule learning to expose mineralogical transitions from the center to the periphery of an ore system.
- Mining companies need to make best use of their data. The experience suggests that obtaining quantitative representations that are visual and automate classification is essential to improve our use of information for mineral discovery.

Sponsors/Collaborators

NSERC-CMIC FOOTPRINTS

ABITIBI GEOPHYSICS, Actlabs, AGNICO EAGLE, ALS, ANGLOGOLD ASHANTI, BARRICK, BearingPoint, Cameco, CGG Passion for Geoscience, Denison Mines, DGI Geoscience Inc., Franklin Geosciences Ltd., GEOVIA, GEOSOFT, GOLD FIELDS, HUDBAY MINERALS, IAMGOLD CORPORATION, KINROSS, Mira Geoscience ...modelling the earth, orano, Paradigm Deeper Insight into the Subsurface, Paterson, Grant & Watson Limited Consulting Geophysicists, Pitney Bowes Software, REFLEX intelligence on demand, SGS, SRC Saskatchewan Research Council, srk consulting, Teck, YAMANAGOLD

Collaborators: GSC TGI4 Program
MRNQ
Saskatchewan Geol Survey
BC Geological Survey

Supporters: Fullagar Geophysics
Rekasa Rocks
UBC Geophysical Inversion Facility