PGO
PROFESSIONAL GEOSCIENTISTS ONTARIO

2025 Virtual Symposium

April 15, 17, 22, & 24, 2025

2025 Virtual Symposium Chair

Hannah Chessell, P.Geo.
Geosyntec Consultants
International Inc.

2025 Symposium Planning Committee


Chair - Hannah Chessell, P.Geo.


Vice-Chair - James Whyte, P.Geo., FGC


Sheila Ballantyne, P.Geo.


Simon Gautrey, P.Geo., FGC


Robert Hearst, P.Geo.


Andrea Waldie, P.Geo., FGC


Craig Waldie, P.Geo., FGC


Marilen Miguel, CAE

Land Acknowledgement

2025 Virtual Symposium Program

- **Today's session (April 15):** *Rocks & Robots - The AI Revolution in Geoscience*
- **April 17:** *Fieldwork Unleashed - Exploring the Power of Practical Experience, Inclusion, and Respect in Geoscience*
- **April 22:** *Geoscience Journeys - Insights and Advice for Aspiring Professionals*
- **April 24:** *Professional Conduct and Complaints - Navigating Ethics, Accountability, and Trauma-Informed Regulation*

Group Insurance Program Partner

TD Assurance

Associate Sponsors

Manulife

HUB

Panel Session A

Rocks & Robots: The AI Revolution in Geoscience

Panel Session A: Speakers

David Slade, P.Eng.
Practice Advisor, Professional Practice Standards & Development
Engineers and Geoscientists British Columbia (EGBC)

Chris Gerrits, M.Sc., P.Eng.
Director, Land Development
CROZIER

Shishi Chen, Ph.D.
Principal – Data Science, Exploration Excellence and Innovation
BHP

Disclaimer:

The information, views and statements presented by speakers at PGO symposium are solely those of the speakers and do not reflect the views of PGO nor do they represent explicit or implied endorsement by PGO.

Presentation 1

Use of Artificial Intelligence (AI) in Professional Practice

David Slade, P.Eng.
Practice Advisor
Professional Practice Standards & Development
Engineers and Geoscientists British Columbia

USE OF ARTIFICIAL INTELLIGENCE (AI) IN PROFESSIONAL PRACTICE

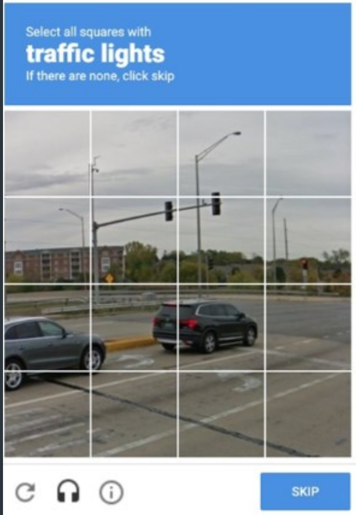
ENGINEERS &
GEOSCIENTISTS
BRITISH COLUMBIA

INTRODUCTION

- Practice Advisory – Use of Artificial Intelligence in Professional Practice
- Guidance on how to use AI appropriately
- Professionals remain responsible
- Considerations when using AI
- How to maintain regulatory compliance

AI OVERVIEW & BACKGROUND

- What is AI
- Types of AI
  - Expert Systems
  - Machine Learning
  - Generative



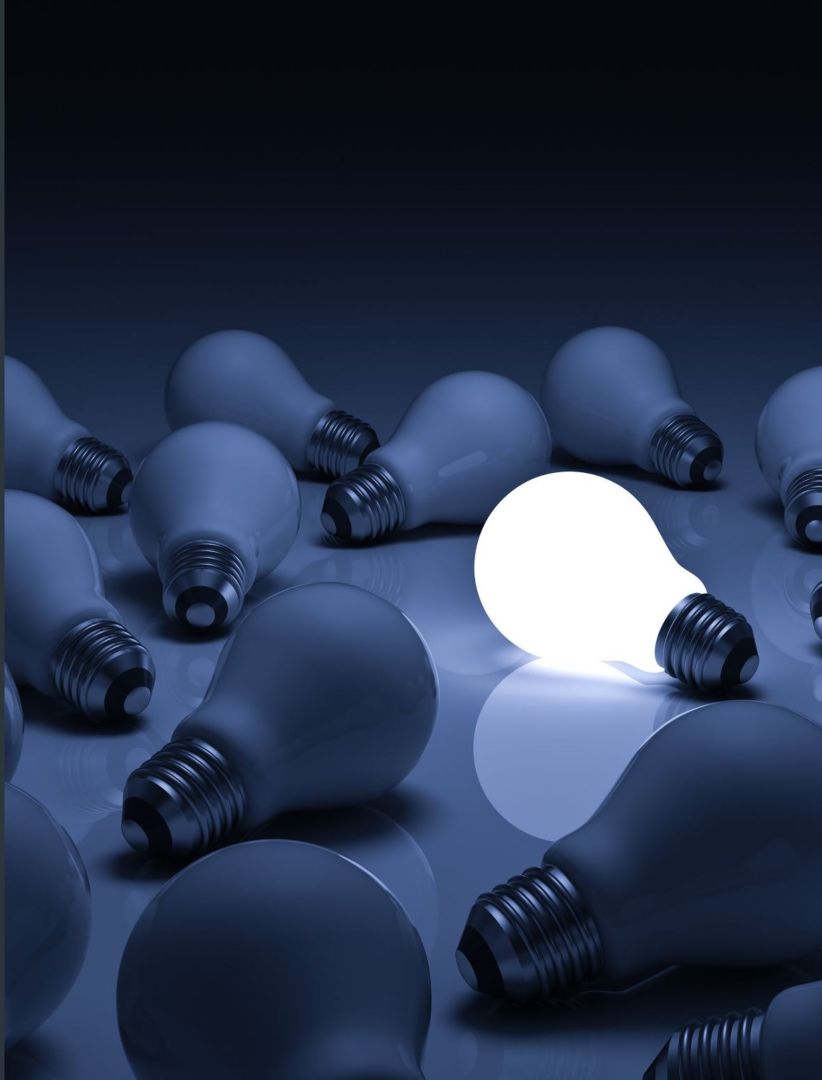
Machine Learning

Expert System

Generative

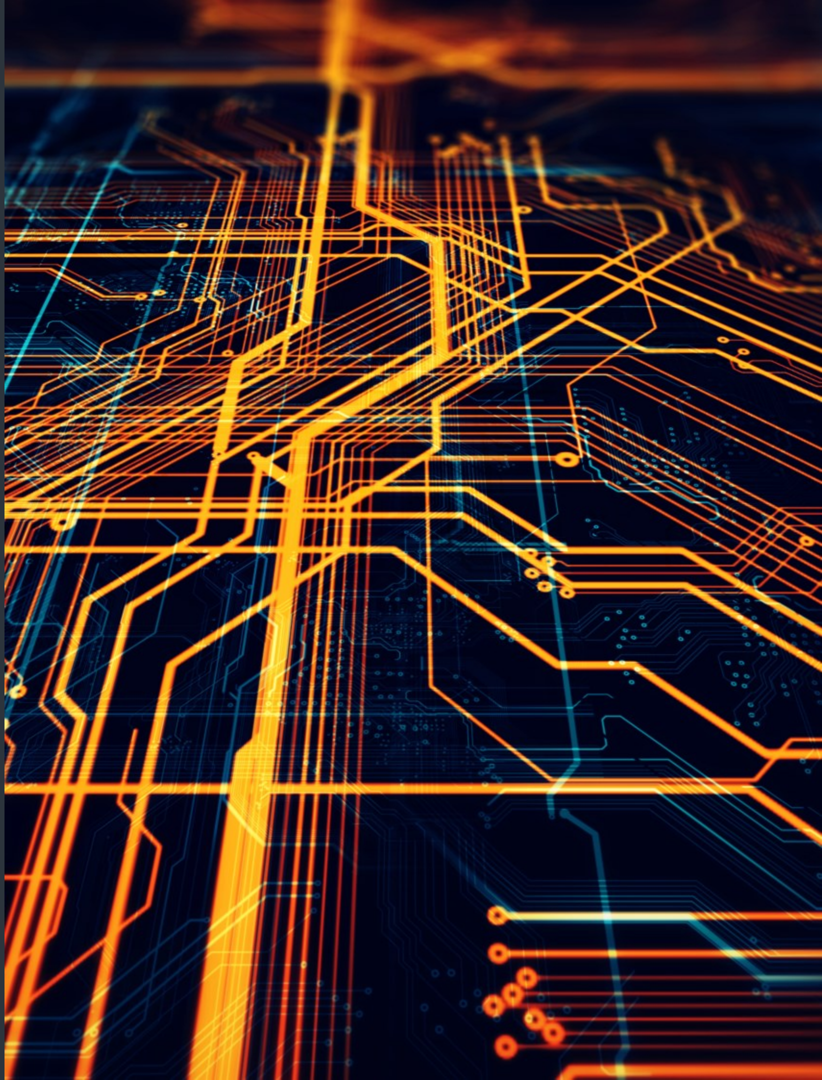
AI OVERVIEW & BACKGROUND

- Static AI
  - Remains unchanged
- Dynamic AI
  - Changes over time
  - Learns, evolves



EXAMPLES OF AI IN PROFESSIONAL PRACTICE

- Text and Image Generation
- Design and Modelling
- Predictive Maintenance
- Quality Control and Inspection
- Process Optimization
- Summarization



RISK MANAGEMENT

- AI risks unique from traditional software
- Maintain familiarity and understanding of system functionality
- Practice only in areas where professionally competent
- Assess, understand and manage/mitigate risks
- Risk Assessments

KEY RISK FACTORS

- Bias
- Trustworthiness
- Transparency, explainability, interpretability
- Repeatability
- Confidentiality and Privacy
- Hallucinations

RISK ASSESSMENTS

Questions to Consider:

- How will the AI output be used?
- What are the potential risks of a flawed output?
- Are you comfortable taking full professional responsibility?
- Does the system exhibit trustworthy characteristics?
- What data was it trained on?
- Are professionals involved competent?
- Can the system be validated?

FIRM CONSIDERATIONS

- Internal Policies & Procedures
- Maintain familiarity
- Legal considerations
- Provision of training

SUMMARY

- Protection of the public & environment paramount
- Professionals remain responsible
- Maintain competence, understand the risks
- Perform risk assessments

THANK YOU

ENGINEERS & GEOSCIENTISTS
BRITISH COLUMBIA

Presentation 2

Data Does Not Lie… or Does It? The Hidden Pitfalls of Machine Learning/AI Workflows in Mineral Exploration

Shishi Chen,
Principal – Data Science, Exploration Excellence and Innovation
BHP

PROFESSIONAL GEOSCIENTISTS ONTARIO

Data does not lie... or does it? The hidden pitfalls of machine learning/AI workflows in mineral exploration

Shishi Chen
Principal Data Science, Exploration, BHP
April 15, 2025

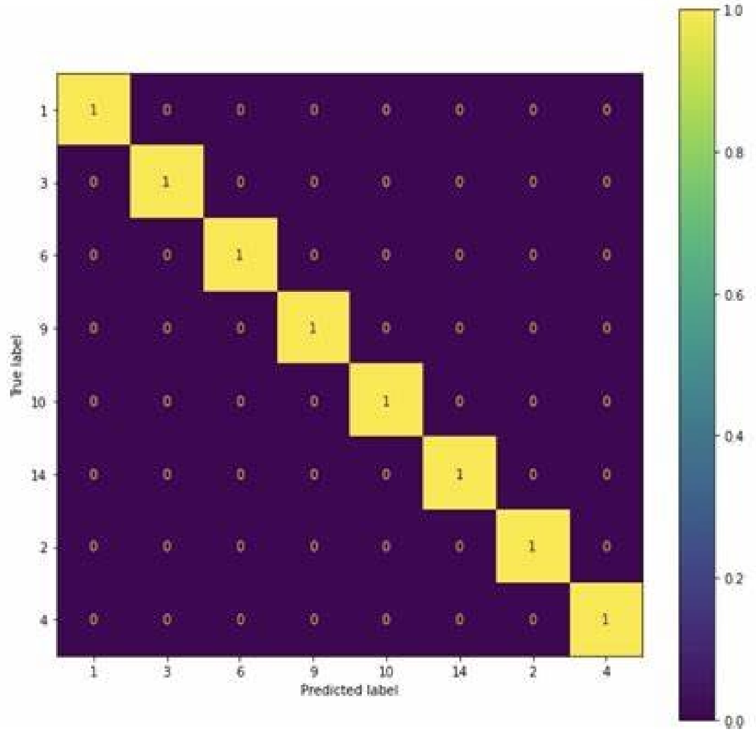
Disclaimer

The presenter maintains an unbiased stance and refrains from endorsing or criticizing any person or data science service provider;

The maps and data have been altered (modified, tilted, or reprojected), and therefore, they do not represent any actual real-world information.

What are the major differences between ML and AI?

- Definition (broader concept vs. narrower)?
- Approach (Mimics human intelligence vs. use algorithms to learn data pattern)?
- Techniques (knowledge representation reasoning, planning, natural language understanding… vs. supervised learning/unsupervised learning)?



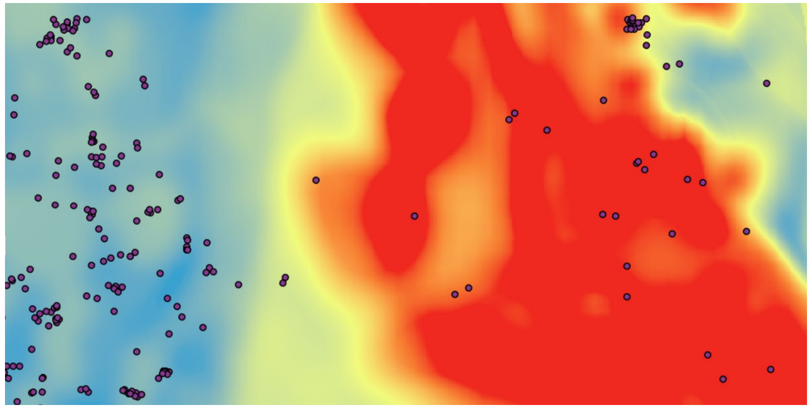
Example: This model has 100% accuracy.. But is it useful? Are we fooled? Why?

The BIGGEST difference is the focus: AI is to **reduce risks and improve success rates** in decision-making, while ML focuses on **enhancing performance by optimizing learning from data**.

Risk assessment matters!
AI system needs human to teach the **potential pitfalls** to improve success rate.

Pitfalls in Data

Exploration data are geospatial data



Source: Lovelace R et al., Statistical Learning: Spatial CV, Geocomputation with R.

When random partitioning is employed in the training process, the accuracy can appear to be **exceptionally high** (>90%) as a result of spatial correlation and data leakage*. This can lead to potentially misleading final predictions.

What you can do:

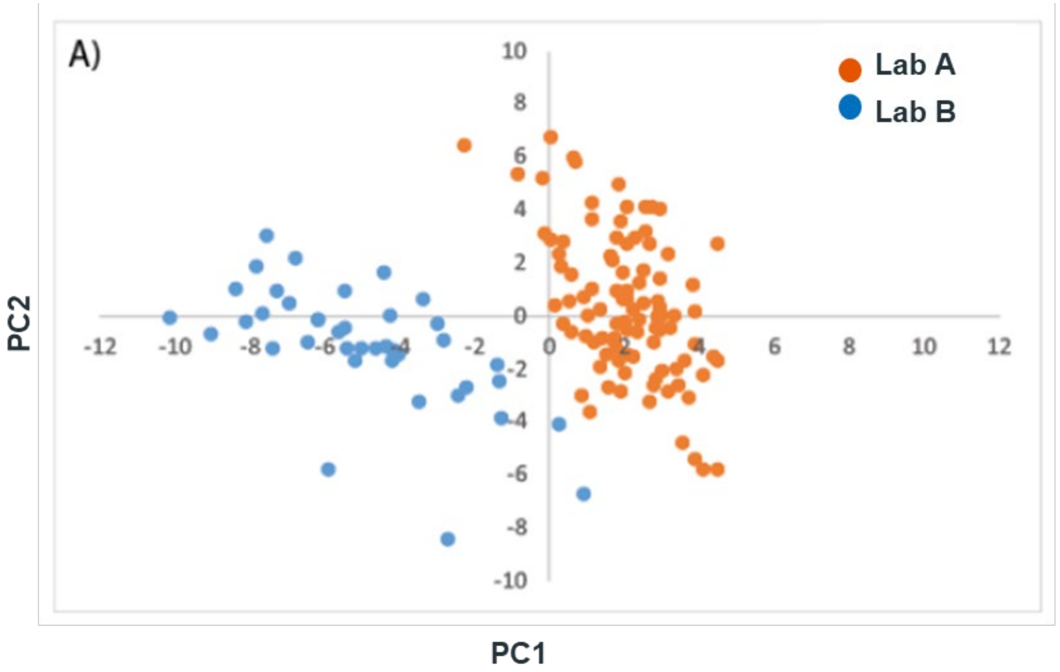
- Use spatial cross validation.

* Data leakage: due to improper test/train data splitting or data processing, the model has learned information of test/validation data , which lead to overly optimistic performance metrics.

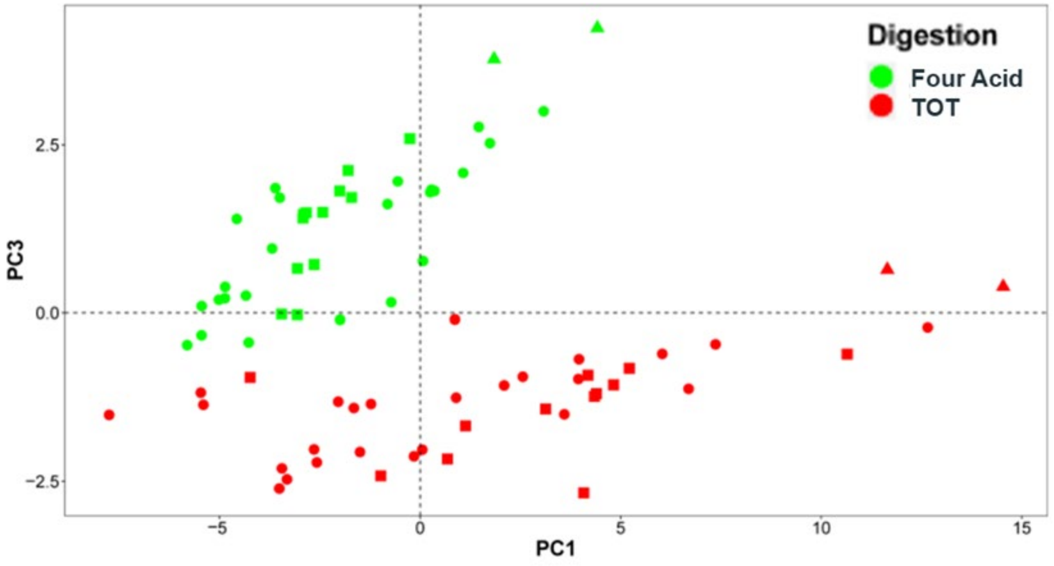
Pitfalls in Data

Data Inconsistency

Example 1



Example 2



Models based on data from different sources, labs, digestion methods, analytical technique or sample medium may yield misleading results.
Need to compare apples with apples.

Pitfalls in Data

Geochemical data are always problematic

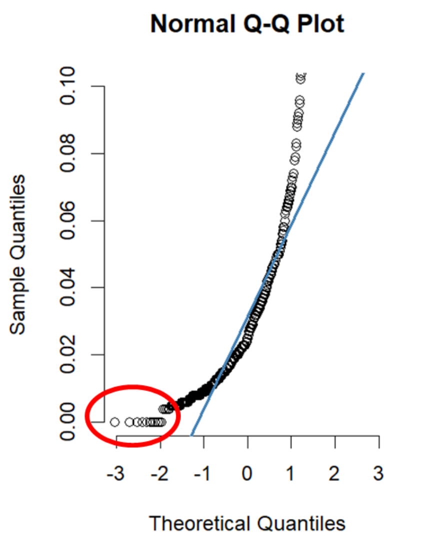
Data quality check:

- Too few data (n of data < n of features * 5, and no less than 100);
- Missing values (NaN or blanks);
- Below detection limit values (e.g. <x, -x,);
- "0" values (likely missing values)
- Constant values (likely BDLs)

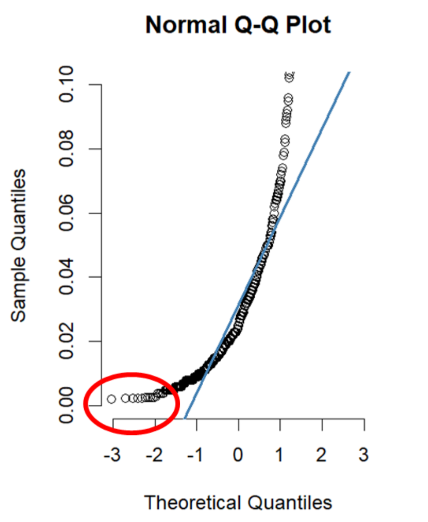
Au_MEMS41L_ppm	Hg_MEMS41L_ppm	Na_MEMS41L_pct
string	string	string
Decimal	Decimal	Decimal
<0.0002	0.053	<0.001
<0.0002	<0.004	0.02

Au_MEMS41L_ppm	Hg_MEMS41L_ppm	Na_MEMS41L_pct
double	double	double
Decimal	Decimal	Decimal
0.000118217919451	0.053	0.000818515812254
0.000106617539635	0.0021841265536061	0.02

Before imputation



After imputation



What you can do:

- Remove data with missing values;
- Replace missing values with mean/median values;
- Replace BDLs with ½ or 1/3 of DL;
- Imputation (but make sure the imputed data is only small portion (<30%))
- *"Thank you for reminding me, but my fancy ML can handle missing values" – you sure?*

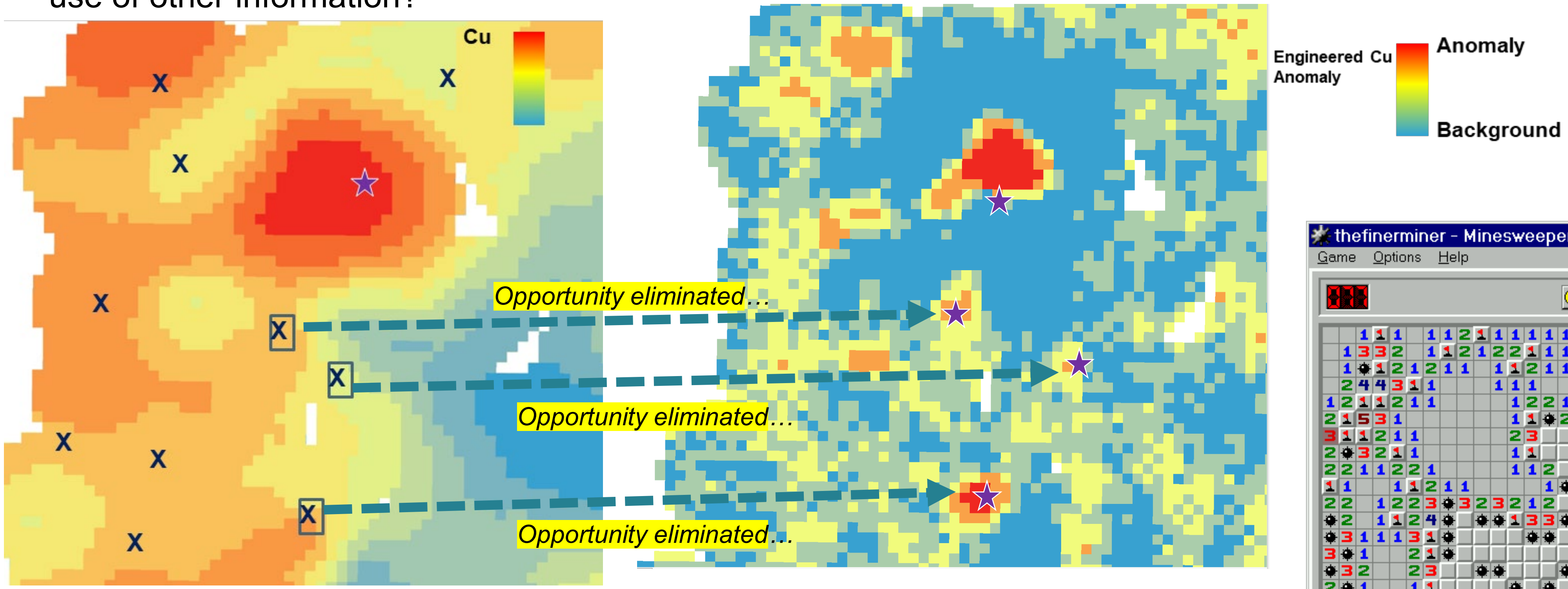
Pitfalls in Data

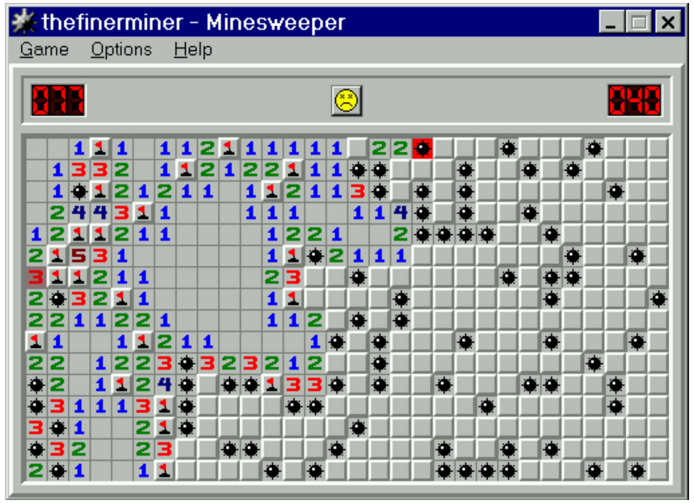
What data can be selected as background?

You *never* know before you drill!

Most of the data analysis in mineral exploration involves comparison of known mineralization and background (non-mineralization). But…

How to define non-mineralization data? Randomly select data from map? SME to select? Make data driven use of other information?



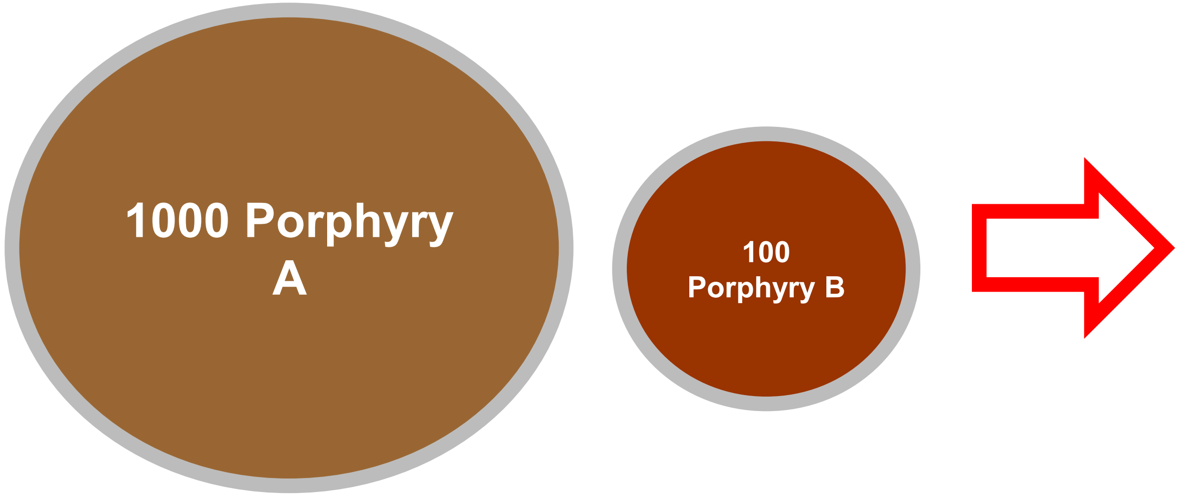


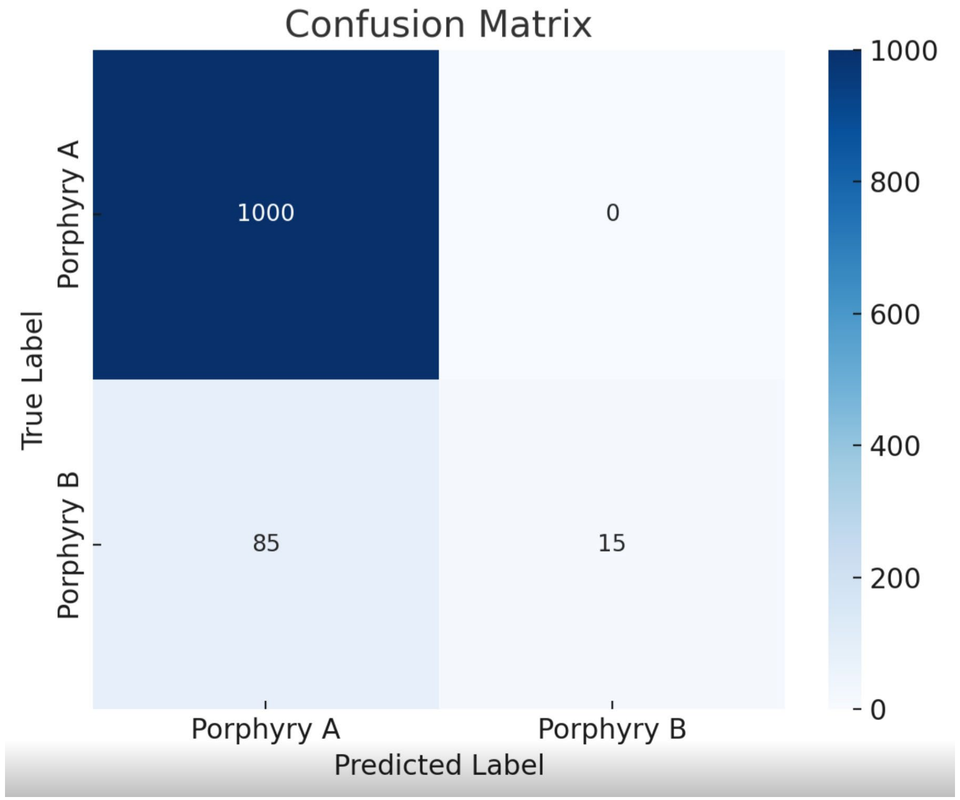
What you can do:

- Try one class and positive unlabeled machine learning

Pitfalls in Data

Imbalanced dataset





Total accuracy = 0.92

Generally speaking, this model has very nice total accuracy, but is not very useful.
The model tends to predict everything Porphyry A to achieve best accuracy and may not have enough data to learn Porphyry B.

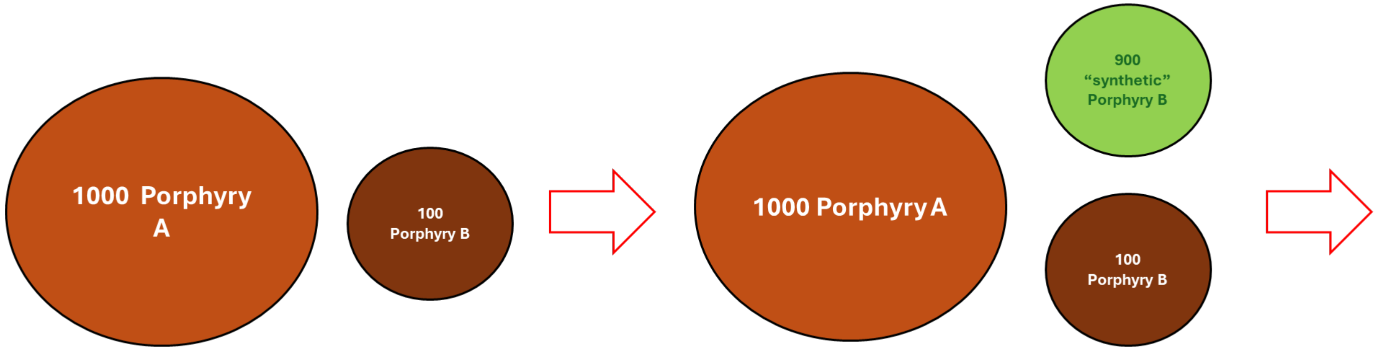
What we can do:

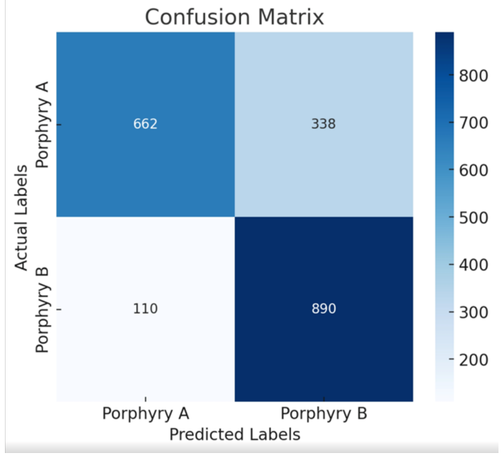
- Look at performance metrics for each class;
- Look at other performance metrics that are robust to imbalanced dataset (e.g. F1)
- Perform data augmentation?

Pitfalls in Data

Data augmentation

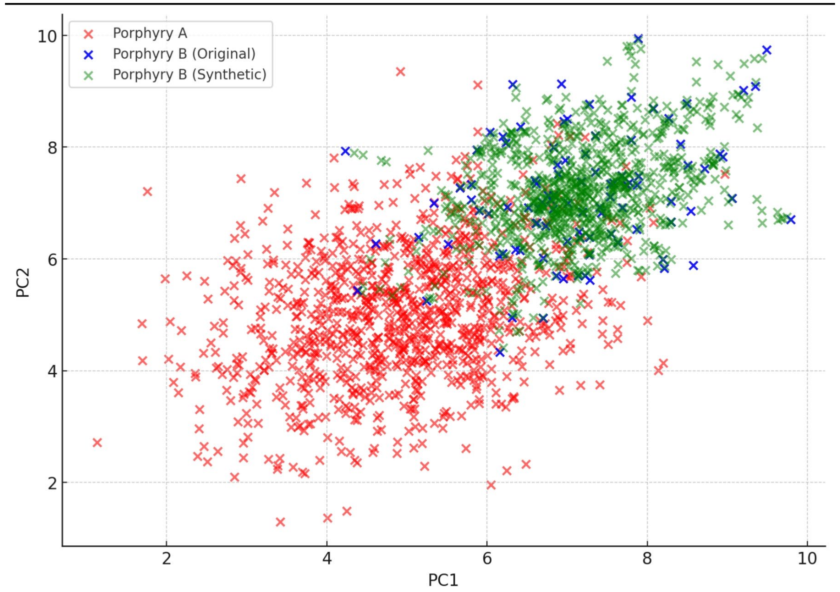
Data augmentation works by interpolating between existing minority-class samples to create new synthetic examples, which helps improve the performance and reliability of machine learning models.





The accuracy for Porphyry B increase to 89% - seems like significant improvement.

Is the improvement real? Or we are fooled again?



Pitfalls in Data

Data augmentation

Food for thought:
We know Prominent Hill, Olympic Dam and Oak Dam are giant, and **unique** IOCG deposits.

We don't have sufficient IOCG data to training model.

Can we use data augmentation to create another 1000 fake Prominent Hill, Olympic Dam and Oak Dam that never exist in real world, for training?

Pitfalls in Algorithms and Performance Metrics

Algorithms:
Select a proper algorithm that fit your data and business question.
Linear? Outliers? Interpretability? Training time?

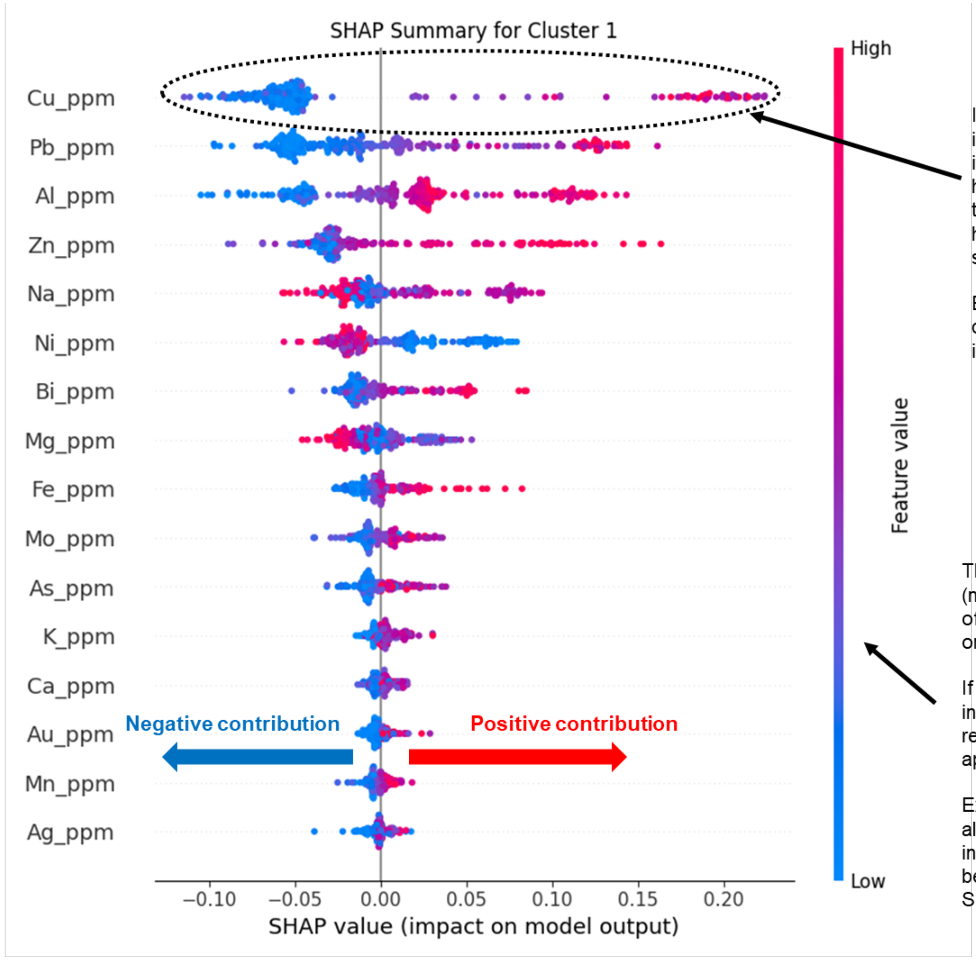
Metrics:
Accuracy: could be misleading with imbalanced data;
Precision: Only focuses on relevance of positive predictions. Ignores false negatives.
Recall: Ignores false positives. A model with high recall but low precision may generate too many positives.
F1 score: harmonic mean of precision and recall. Works better than accuracy for imbalanced data
Mean Absolute Error (MAE): Not sensitive to outliers;
Mean Squared Error (MSE): sensitive to outliers. Better if you are interested in outliers.
R square (R2): Sensitive to outliers. Not good for non-linear models.

Know what your data look like, what you want to optimize and use a proper metric to assess your model!

Interpretable ML/AI

Use interpretation techniques, such as Shapley analysis and LIME to open model black box and interpret predictions.
This will greatly reduce risk and improve confidence of decision-making.

These are the input variables, ranked from top to bottom by their mean absolute SHAP values for the entire dataset.



In a beeswarm plot, for each variable, every instance (i.e. row) of the dataset appears as it's own point . The points are distributed horizontally along the x-axis according to their SHAP value. In places where there is a high density of SHAP values, the points are stacked vertically.

Examining how the SHAP values are distributed reveals how a variable may influence the model's predictions.

The color bar corresponds to the raw values (not to be confused with the SHAP values) of the variables for each instance (i.e. point) on the graph.

If the value of a variable for a particular instance is relatively high, it appears as a red dot. Relatively low variable values appear as blue dots.

Examining the color distribution horizontally along the x-axis for each variable provides insights into the general relationship between a variables' raw values and its SHAP values.

An example of SHAP analysis to interpret geochemical anomalies generated by ML.

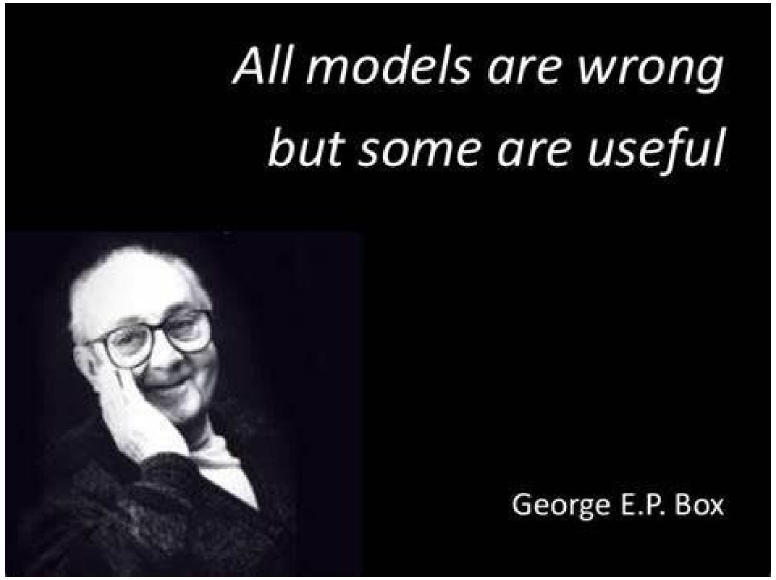
Summary

- Understand and evaluate your data;
- Select the algorithm and performance metrics according to your business problem to avoid hidden pitfalls.
- Risk assessment is essential, requiring data scientists to collaborate with SMEs for data understanding and model interpretation.
- Due to time constraints, many topics were not covered in this presentation.

e.g. bias introduced by feature engineering, geochemical data closure etc.

Let’s go back the question at the very beginning…

Does data lie? Not exactly. Data is neutral, but results can be misleading. Models often prioritize accuracy over risks. Without experts to guide AI on risks, AI remains ineffective and fails to aid exploration strategies.



BHP

Thank you

PGO
PROFESSIONAL GEOSCIENTISTS ONTARIO

Presentation 3

Using Artificial Intelligence and Machine Learning in Canada's Water

Chris Gerrits, M.Sc. P.Eng.
Director, Land Development
CROZIER

PGO
PROFESSIONAL GEOSCIENTISTS ONTARIO

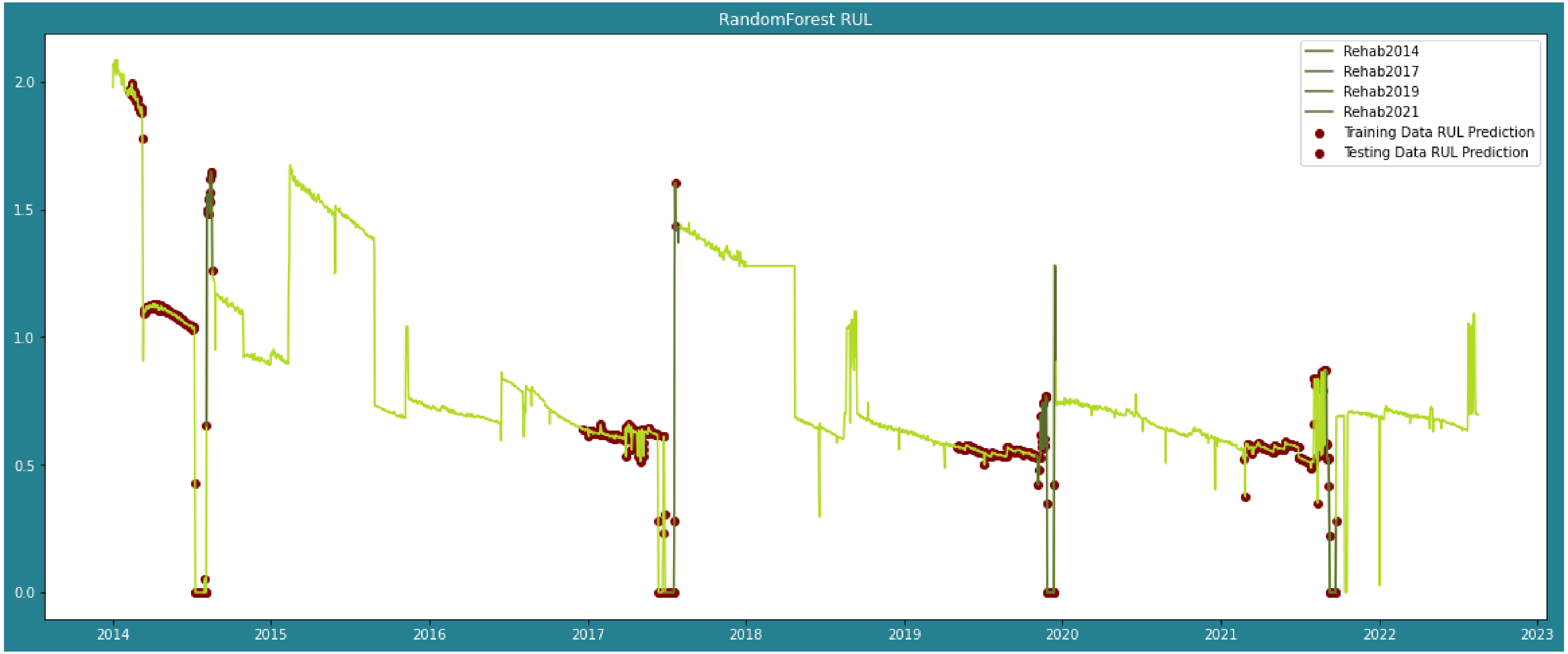
Using Artificial Intelligence and Machine Learning in Canada's Water Sector

Presented by: Chris Gerrits, M.Sc., P.Eng.

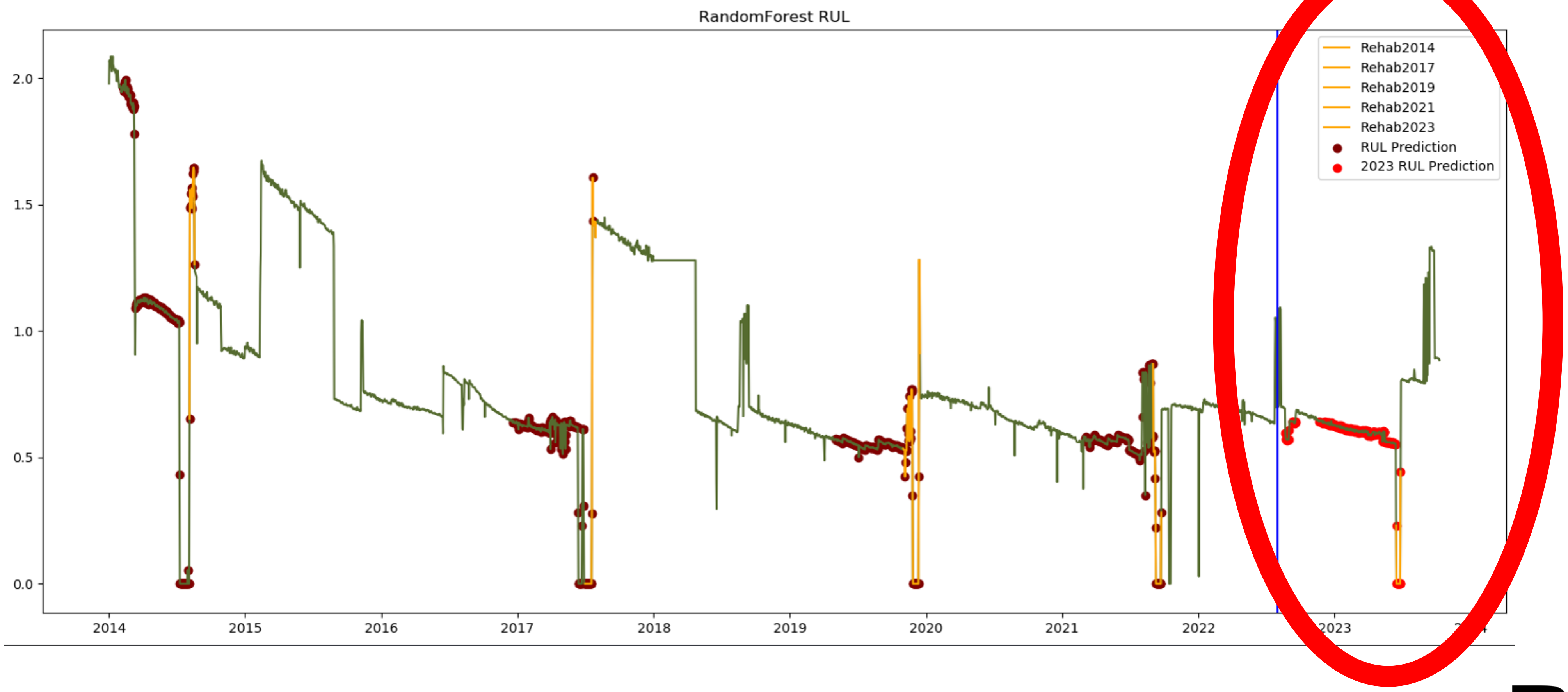
April 15, 2025

CROZIER

Nov 2022 - Prediction of Rehab Before 180 Days

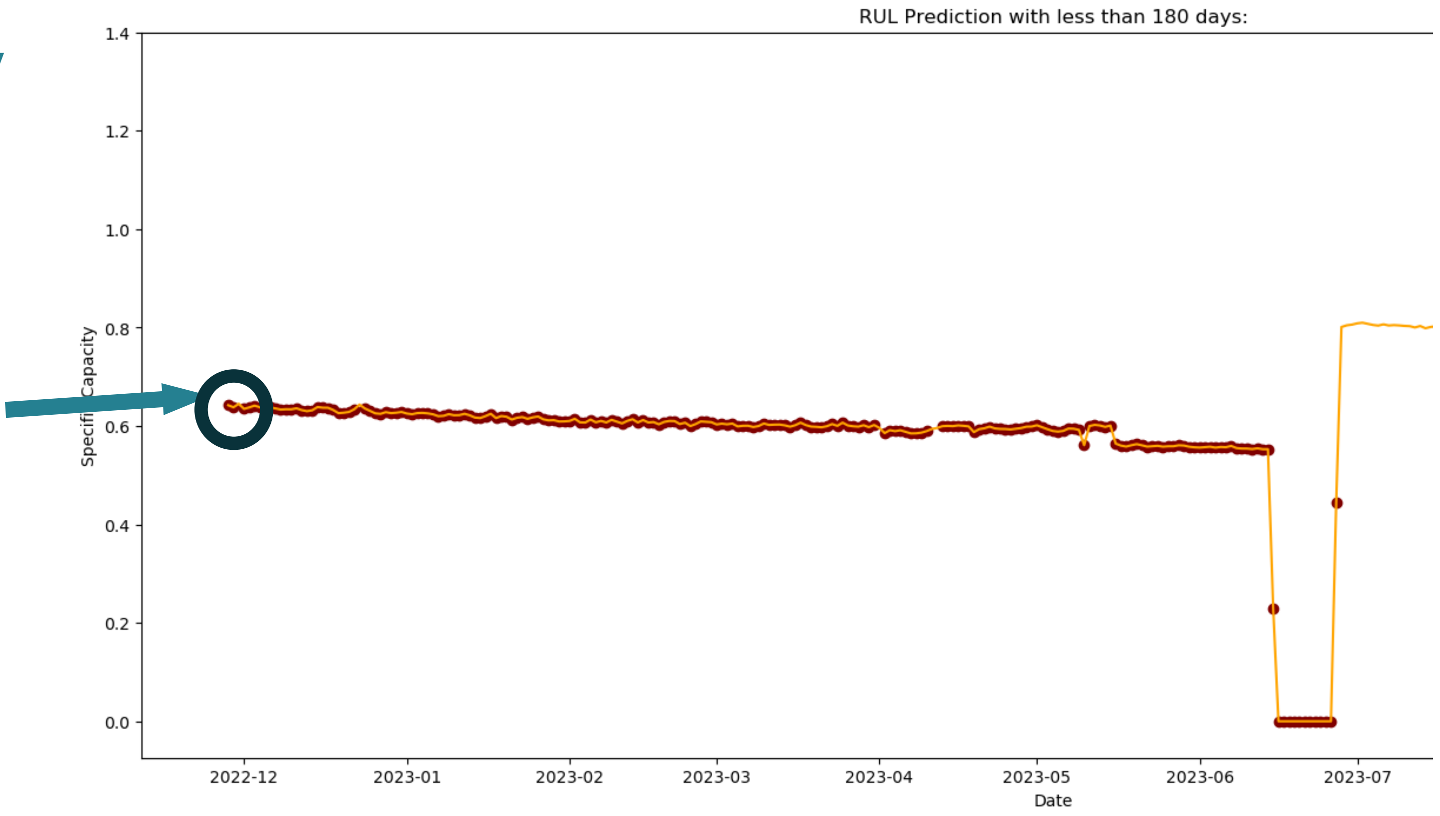


Nov 2023 - Prediction of Rehab Before 180 Days



Nov 2022 - Prediction of Rehab Before 180 Days

- The model consistently indicated commending November 28, 2022 that well rehabilitation would be required within 180 days from that date.
- RUL < 180 days
- Therefore model predicts rehab on/ around May 26, 2023.



What Does This All Mean?

- Rehabilitation has been required 5 times over the 13 years of the life of the system with increasing frequency.
- Assuming rehabilitation is required every 2 years and there is a two-month gap between identification of the need and completion of the work.
- 0.0044 L/s/m per month lost (0.23%).
- In 2023 there were two months between when the rehabilitation was identified as being required and the commencement of the rehabilitation effort.
- Additional avoidable losses of 0.008 L/s/m over a two-month period.
- **Effective prediction and proactive maintenance (rehabilitation) can add over 5% to the total life of the asset in this situation.**
- Each situation is different and in some cases the savings could be significantly greater, some cases negligible.

Case Study: Using Artificial Intelligence Object Detection in Hydrogeology

Problem Statement

- Identification of threats on a large scale requires a significant amount of effort.
- Threats need to be constantly updated.
- Training a model to identify potential threats through object detection could save considerable time and effort.

Gas Tank

SWM Facility

Salt Storage

Object Detection Model

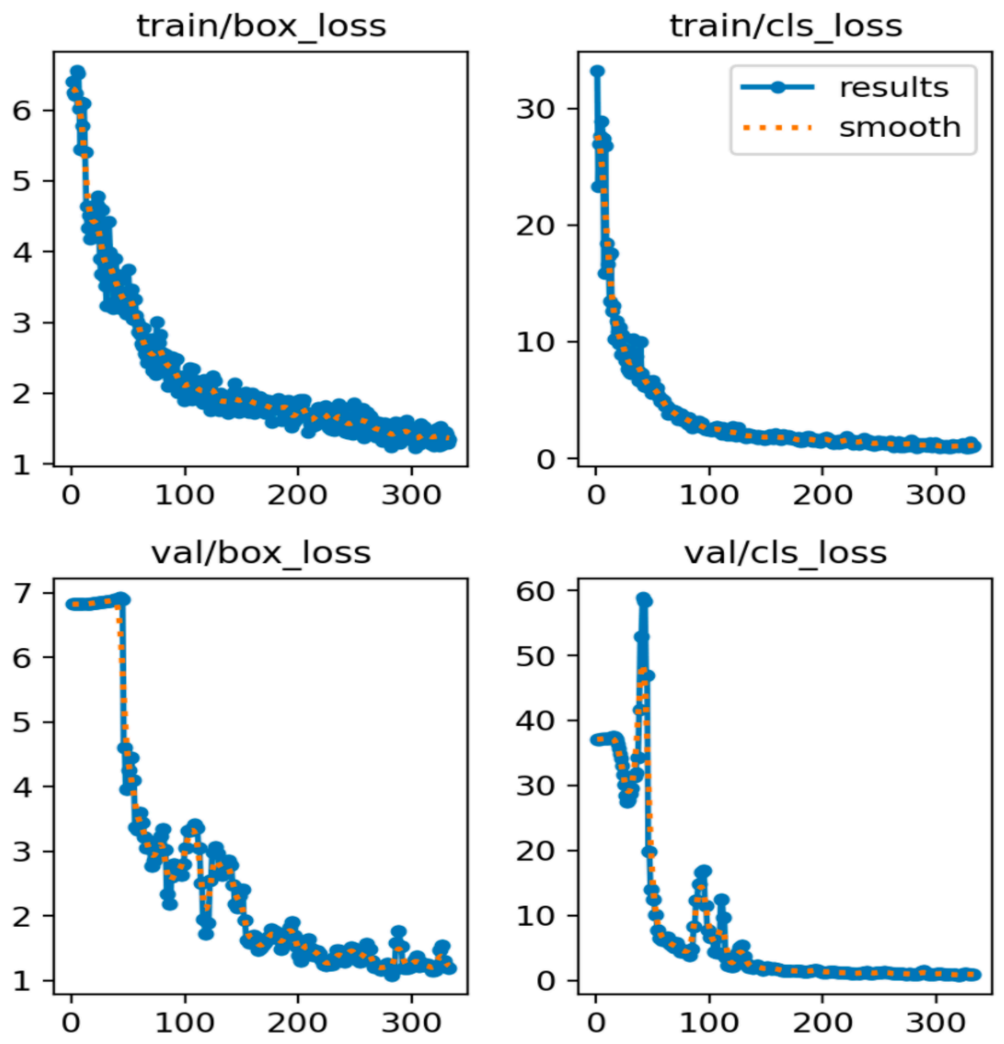
- Image (or object) detection is a task in computer vision that involves identifying the presence, location and type of one or more objects in a given image.
- It is a challenging problem that involves building upon methods for object recognition (e.g. where are they), object localization (e.g. what are their extent), and object classification (e.g. what are they).
- A series of images of known storage facilities are used to train the model (training set) and the model is validated with a separate set of images (validation set).
- The selected images are on different scales and in different settings to ensure a variety of circumstances with which to train the model.

Model Development

- First step in the development is to use Data Annotation (with CVAT Tools).
- Images are converted with annotation into YOLOv8 training data.
- Manual process that relies on the model developer to input/annotate the known features.

Training

- Once the model has been developed and trained with the training data the model is validated.
- The validation data is a separate data set used to validate the results. Data has not previously been "seen" by the model.
- The model runs through enough epochs until the losses are sufficiently low as shown in the graphs.



PROFESSIONAL GEOSCIENTISTS ONTARIO

Prediction

- The model is then used to locate the feature along with a confidence score (from 0 to 1).
- In this case, the model correctly identified twelve (12) manure storage tanks with a confidence from 0.3 – 0.8.
- The model also failed to locate five (5) manure storage tank.
- These results are from a single model run.
    - It is expected that with further refinement the number of successful features located as well as the confidence score would both increase.


manure_tank 0.5
Google

Potential Uses

Regional or local scale

- Invasive species detection – identifying areas that meet the conditions for giant hogweed to grow.
- Sinkhole floodplain area detection in Kentucky.
- Asset location – can be used to map bus stops, backyard pools, SWM ponds.

Site scale

- Identifying and automating weed spraying.
- Road (asphalt) condition assessment.
- Detection of Potential Vernal Pools on the Canadian Shield (Ontario) Using Object-Based Image Analysis in Combination with Machine Learning (Luymes & Patricia Chow-Fraser, Canadian Journal of Remote Sensing, May 2021).

Thank you

PROFESSIONAL GEOSCIENTISTS ONTARIO

Symposium contact information

Marilen Miguel
mmiguel@pgo.ca

PROFESSIONAL GEOSCIENTISTS ONTARIO

Panel Session B

April 17, 2025 from 12 noon to 1:30 p.m. ET

Fieldwork Unleashed: Exploring the Power of Practical Experience, Inclusion, and Respect in Geoscience

Panel Session B: Co-chairs

Andrea Waldie, P.Geo., FGC

Sheila Ballantyne, P.Geo.

Thank you

PROFESSIONAL GEOSCIENTISTS ONTARIO