An Automated Conformalized Causal Learning System for Enhanced Mineral Prospectivity Mapping

Evelyn Jessica Jaya[1], Xinbing Wang[1], Chenghu Zhou[1], and Nanyang Ye[1]

[1]*Shanghai Jiao Tong University, Shanghai, China (evelyn321jaya@gmail.com)*

EGU General Assembly #EGU24 14 -19 April 2024 · Sharing is encouraged

Session GI2.4 EGU24-8771 · Outstanding Student & PhD candidate Presentation contest

1. Introduction

- Mineral Prospectivity Mapping (MPM) is pivotal in economic geology, enabling strategic identification of mineral deposits through sophisticated analysis of geological data.
- Recently, MPM has navigated a shift from expert-driven interpretations to data-driven approaches, leveraging advanced machine learning to enhance predictive accuracy.
- Despite these advancements, a critical challenge remains: accurately predicting in Out of Distribution (OOD) scenarios, essential for reliable mineral exploration.

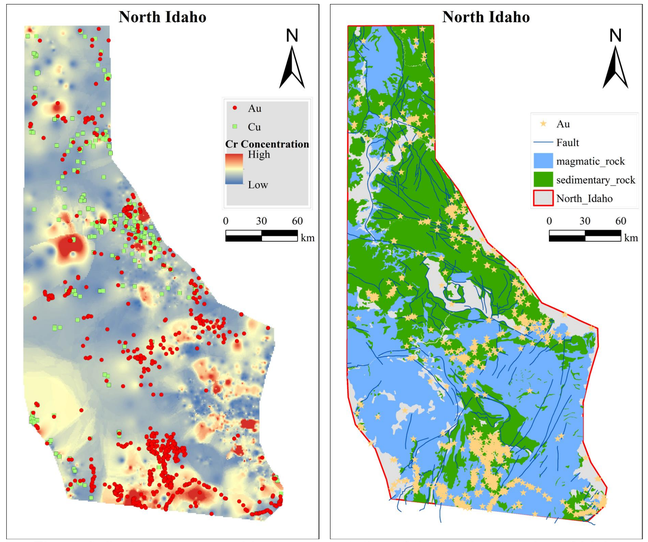
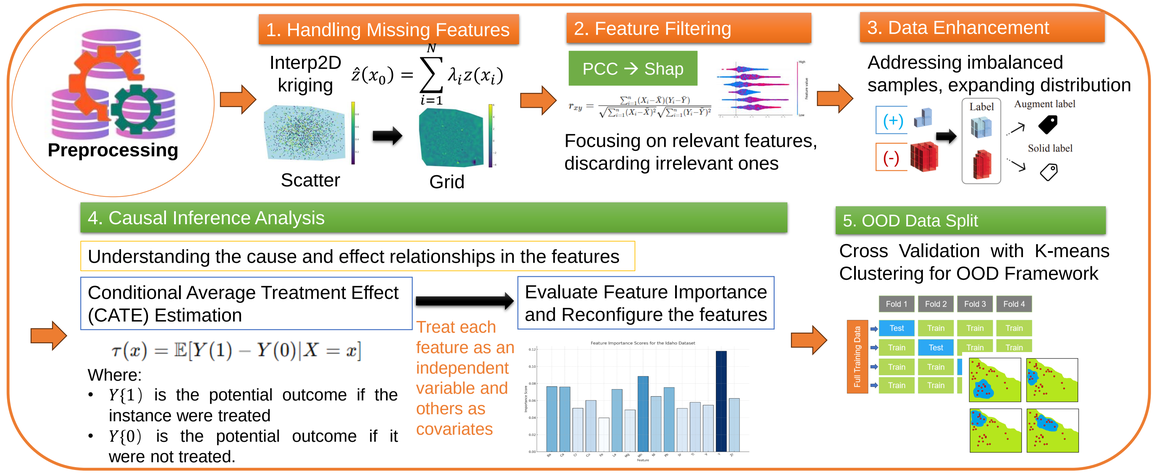


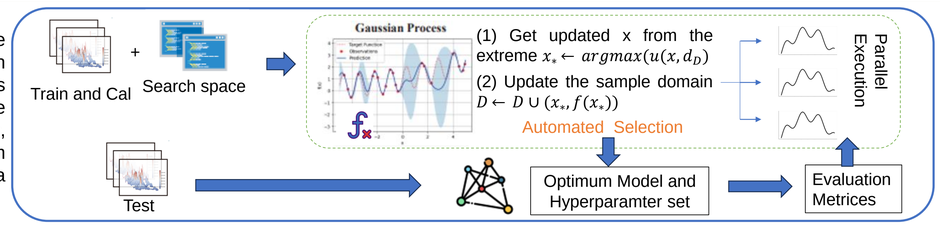
Figure 1 displays a geographical representation of the northern region of Idaho state, collected by the Idaho Geological Survey and the US Geological Survey.

2. Preprocessing with Causal Inference

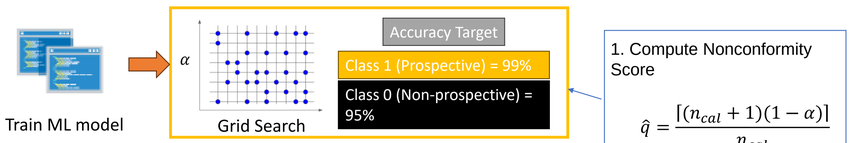


3. Bayesian Optimization

- In each iteration, the true objective function (a surrogate model) is constructed using the Gaussian Process, consisting of random variables following a Gaussian distribution.



4. Automated Alpha Parameter Tuning



Conformal Prediction

$$p_i = \frac{|\{j \in X_{cal}: \alpha_j \geq \alpha_i\}|}{n_{cal} + 1}$$

- Implement conformal prediction to provide an "uncertainty region" for each prediction, quantifying the confidence and reliability of the machine learning model's outputs.
- It operationalizes the notion of nonconformity, assigning a scalar α to new observations, signifying their deviation from the established data distribution.
- The generated p-values are consequently interpreted as confidence measures, encapsulating the likelihood of prospective predictions.

1. Compute Nonconformity Score

$$\hat{q} = \frac{\lceil (n_{cal} + 1)(1 - \alpha) \rceil}{n_{cal}}$$

2. Predict with $1 - \alpha$

3. Construct *prediction set*

$$C(X_{cal}) \subset \{0, 0.5, 1\}$$

5. Results

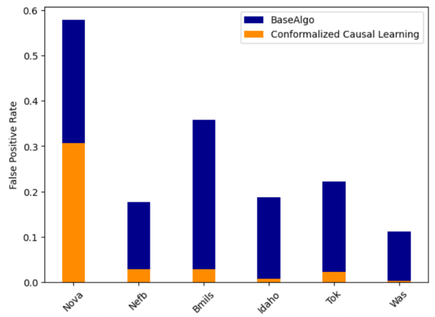


Figure 2. A Comparative Evaluation of FPR between the Proposed Methodology and the Baseline Algorithm (RF)

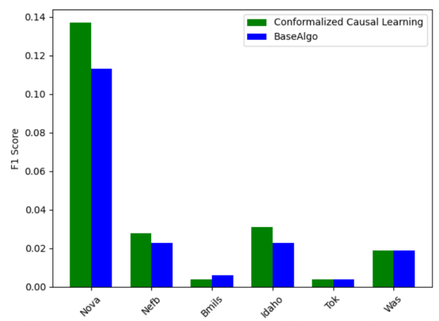


Figure 3. Comparative Analysis of F1 Scores Across Six Diverse Datasets.

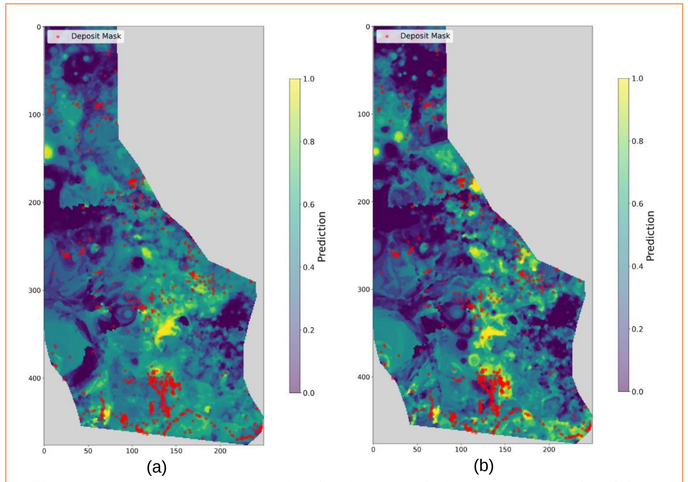


Figure 4 presents a comparative visualization of prediction outcomes for the Idaho dataset in an OOD context:(a) Baseline Algorithm (RF) (b) Conformalized Causal Learning Algorithm. Upon visual inspection, it is evident that our method delineates a more precise region of anticipated mineral presence when contrasted with the baseline algorithm.

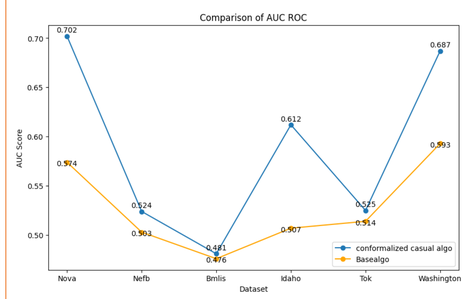


Figure 5. Comparison of AUC ROC Scores for Conformalized Casual Algorithm and Basealgo Across Six Datasets

- We evaluated the performance of our methodology on six datasets, where the AUC ROC of our methodology surpassed the baseline method by an overall **17.84%,**
- The false positive rate (FPR) was reduced by an overall **84.31%**

6. Conclusions

- Conformalized causal learning system represents a monumental leap forward in MPM by deftly tackling the challenges of OOD scenarios and subjective analysis.
- The system advanced data preprocessing and Bayesian Optimization, paired with the novel application of conformal prediction and causal learning, empower geologists with unprecedented precision and confidence.
- The remarkable advancements demonstrated by our methodology, marked by a substantial improvement in AUC ROC scores and a drastic reduction in the false positive rate, solidify the promise and potency of data-driven approaches in economic geology.