Predicting Magma Fertility for Porphyry Copper Exploration Using Machine Learning

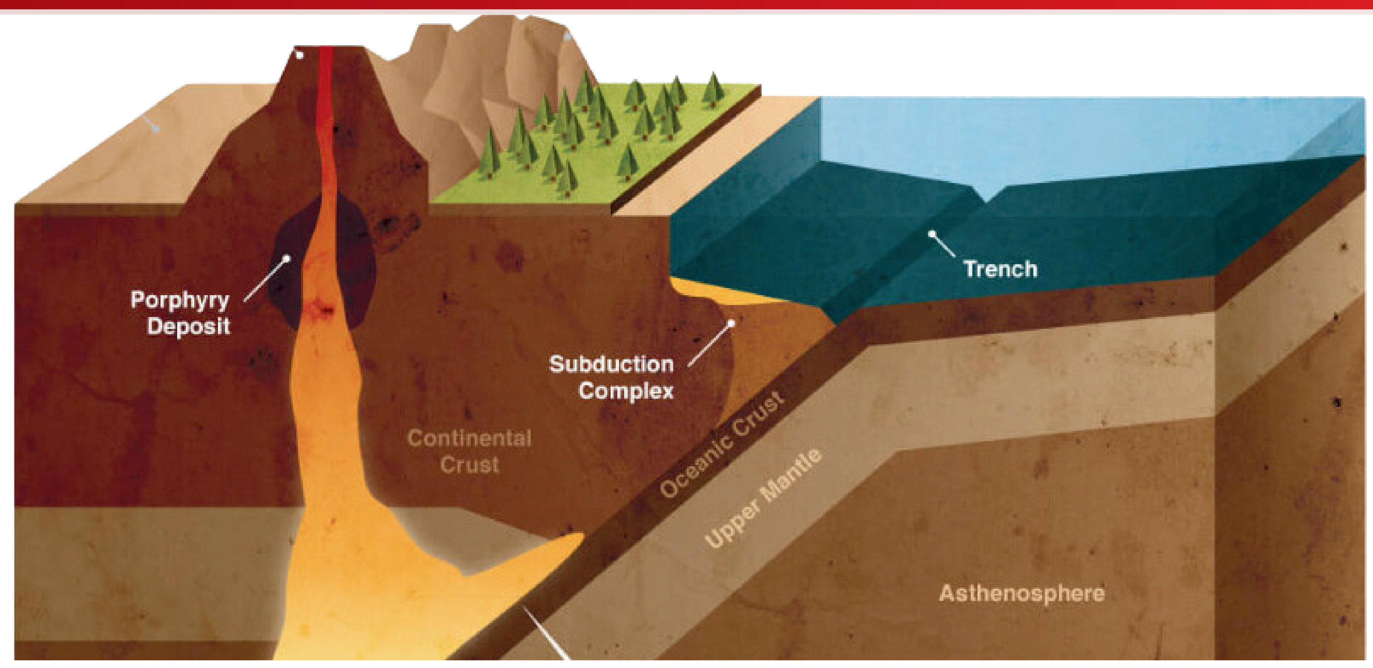
Data Science for Energy Transition

Dr. Jiajia Sun[1]
Professor of Geophysics

Sai Phani Ram Popuri[1], Kaethleen Siguenza[2], Hemanth Kumar Nichenametla[1], Ryan Nguyen[1]
University of Houston[1], University of Houston - Downtown[2]

Dr. Yingcai Zheng[1]
Professor of Seismology

INTRODUCTION



- *Scan the "Explore" QR code to discover the geology of porphyry.*
- **Magma fertility is a crucial indicator for porphyry copper deposits; its accurate assessment requires analyzing complex geological data, a task that machine learning significantly streamlines to enhance exploration efficiency.**

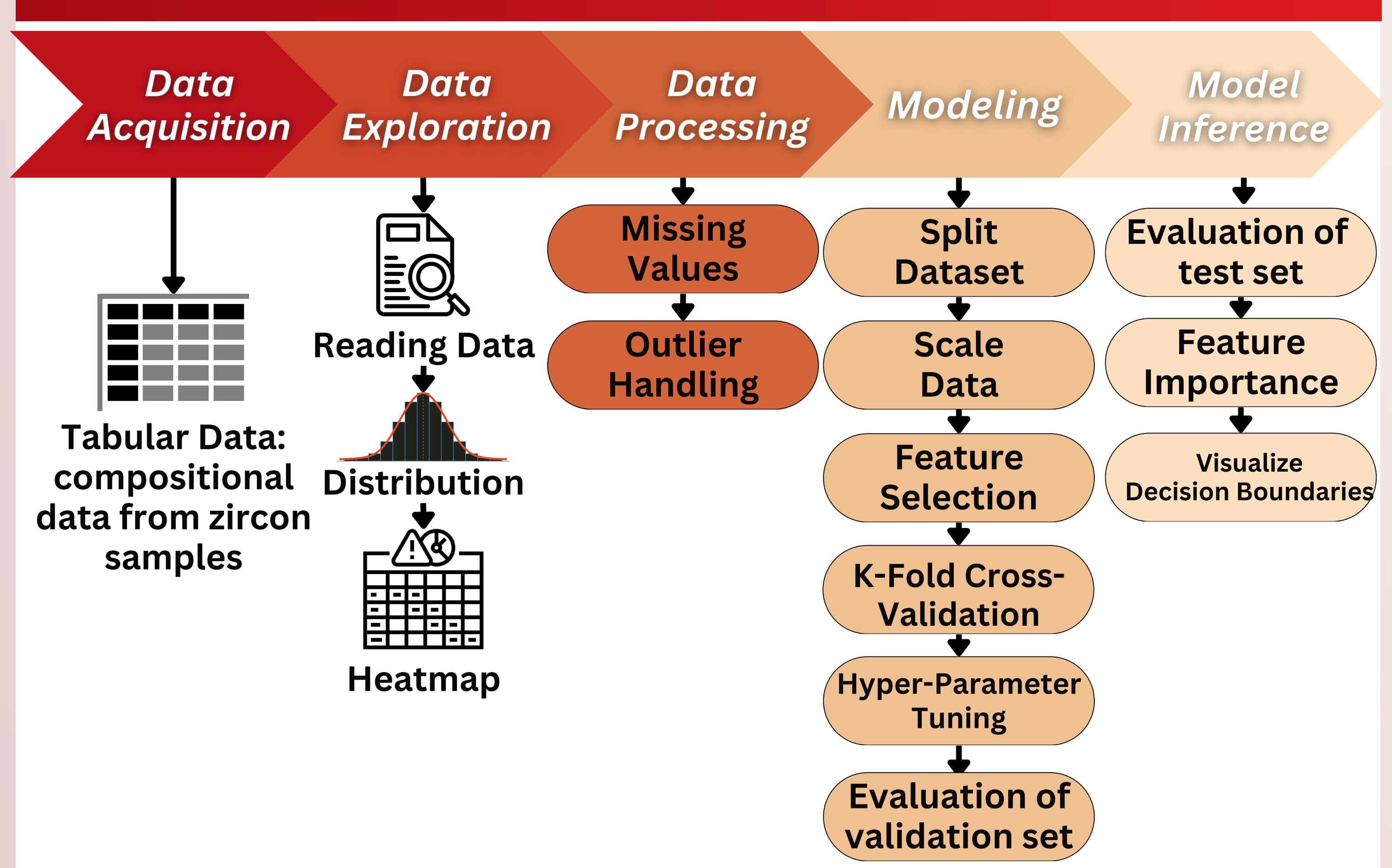
OBJECTIVE

To evaluate and compare the efficacy of various machine learning models in predicting magma fertility for enhanced porphyry copper deposit identification.

DATA DESCRIPTION

- **2988 Zircon Samples - 80:20 train test split**
- **28 Total Features |** *Excluded 8 highly correlated features*
- **Elements:** Neodymium (Nd), Samarium (Sm), Europium (Eu), Hafnium (Hf), Dysprosium (Dy), Uranium (U), Praseodymium (Pr), Thorium (Th), Cerium (Ce), etc.

METHODOLOGY



EVALUATING THE IMPACT OF IMPUTATION TECHNIQUES AND OUTLIERS ON LOGISTIC REGRESSION'S (BASELINE) PERFORMANCE

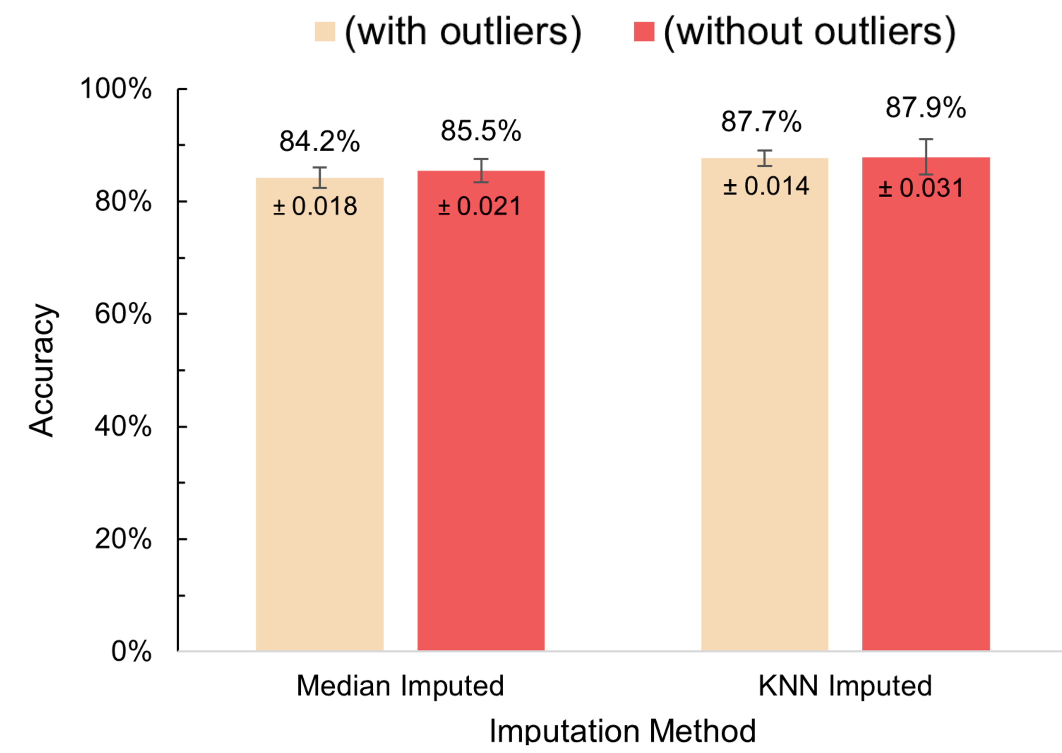


Figure 1: Accuracy Comparison of Imputation Methods

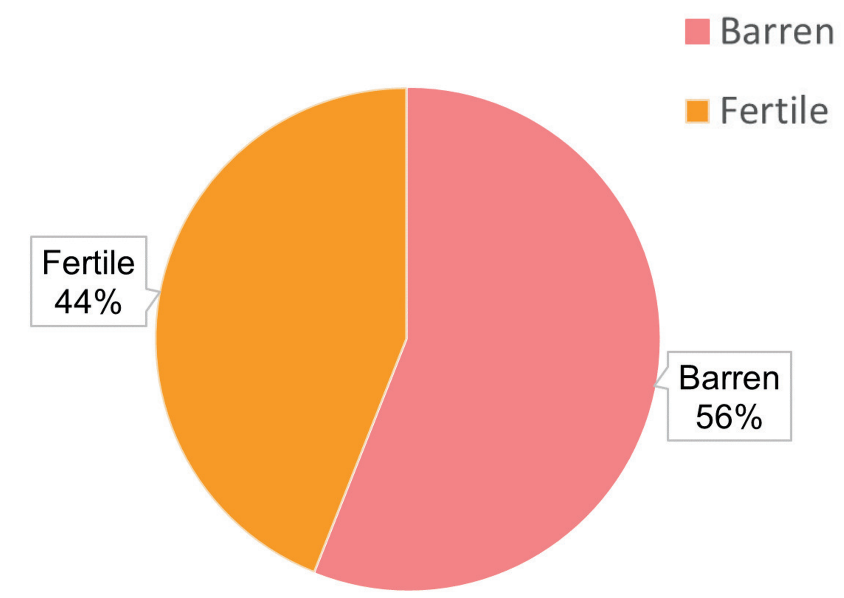


Figure 2: Distribution of Magma Fertility in Dataset

PERFORMANCE OF TEST SET OVER HYPER-PARAMETER TUNED MODELS

Model	Precision	Recall	F1-Score	Overall Accuracy
Logistic Regression	0.81	0.93	0.86	0.873
Support Vector Machines	0.89	0.95	0.92	0.928
Decision Tree	0.90	0.87	0.88	0.898
Random Forest	0.96	0.95	0.96	0.962

Table 1: Evaluation Metrics for Minority Class (Fertile) across test data against 10-Fold Cross Validated Models

ANALYZING DECISION BOUNDARIES OF PCA-TRANSFORMED FEATURES ACROSS TEST SET

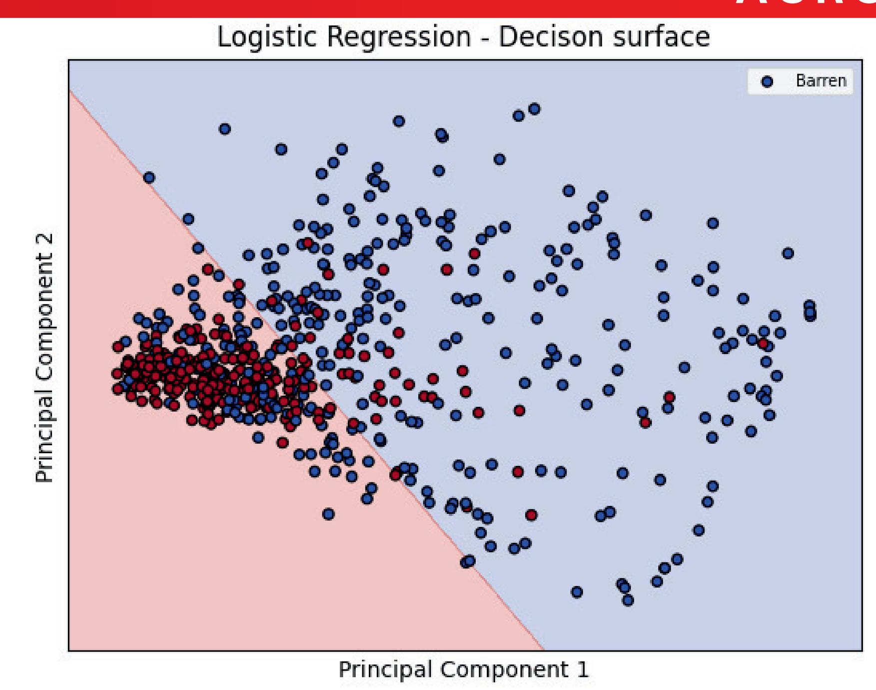


Figure 3: Logistic Regression - Linear Decision Boundary

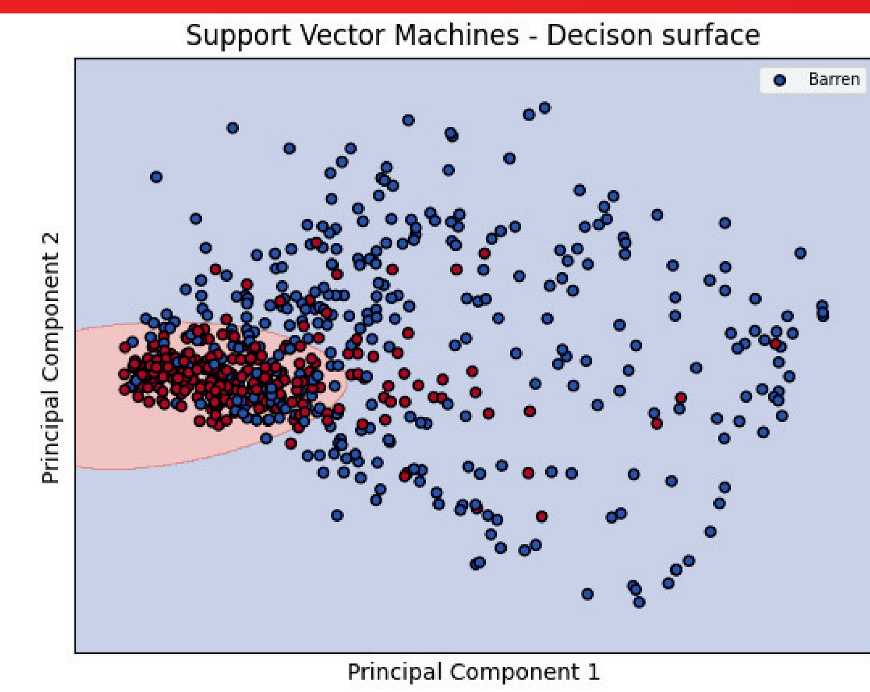


Figure 4: SVM - Radial Basis Function

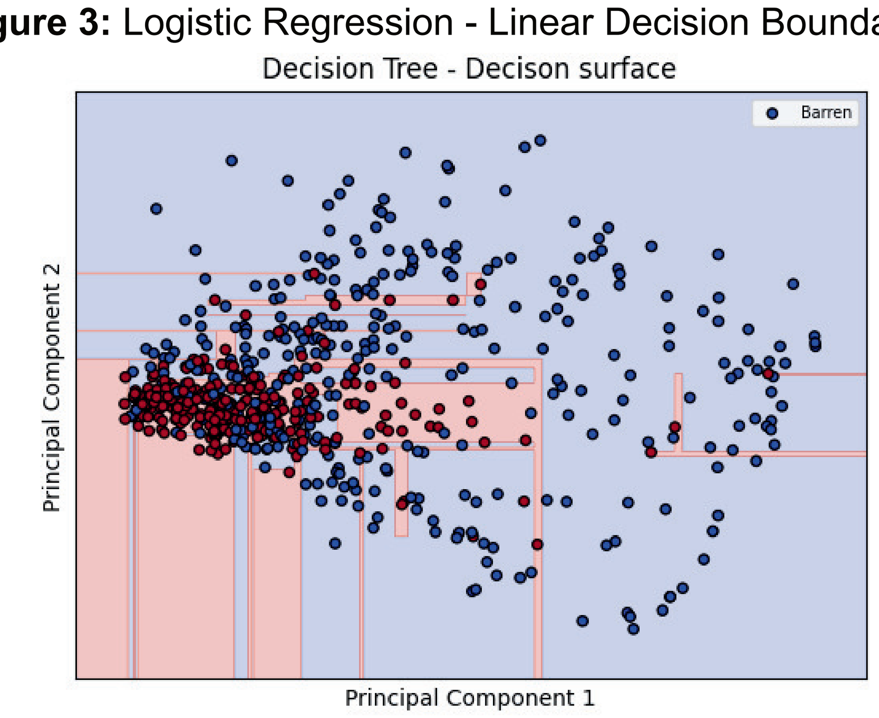


Figure 5: Decision Tree - Segmented Decision Boundary

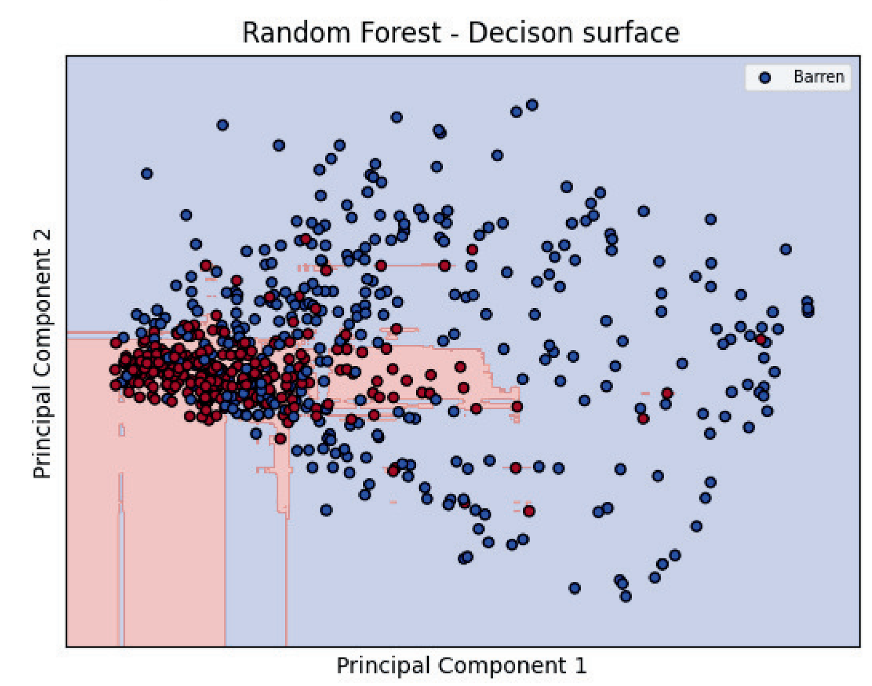


Figure 6: Random Forest - Segmented Decision Boundary

FEATURE IMPORTANCE - RANDOM FOREST

Top 10 Features Based On Random Forest (~ 77% contribution)

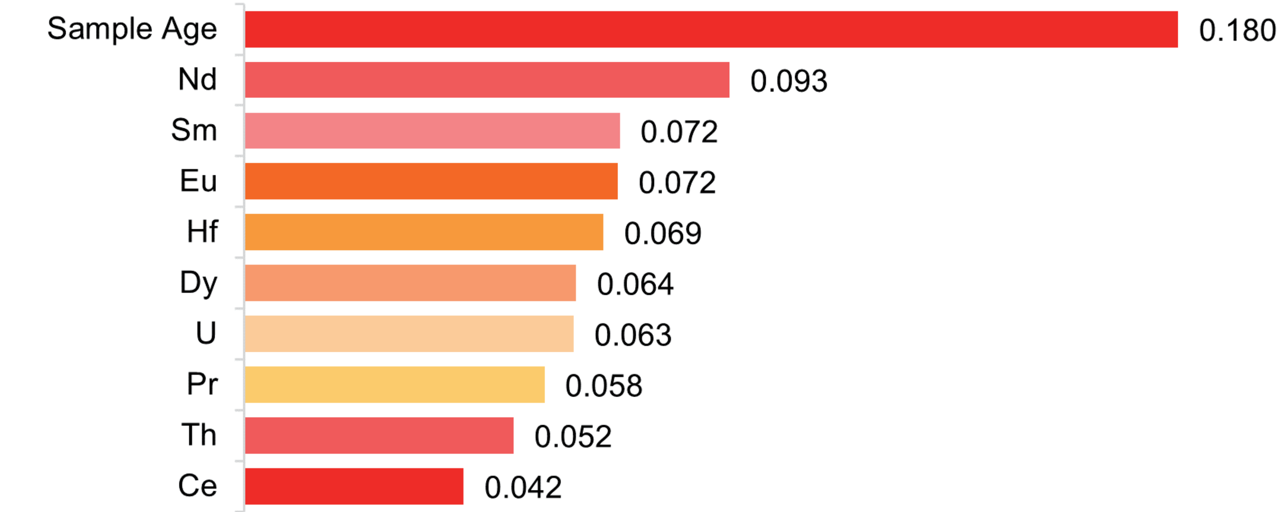


Figure 7: Feature importance scores quantify the contribution of each feature to the model's predictive performance, indicating how much each feature impacts the model's decisions.

BEST PERFORMING MODEL: RANDOM FOREST

Confusion Matrix for Random Forest Classifier Over Test Set

Figure 8: This confusion matrix indicates high accuracy in predicting 'Barren' and 'Fertile' classes, with correct predictions of 327 and 248, respectively, and only 23 misclassifications."

INFERENCES & FUTURE DIRECTION

- After evaluating several machine learning models, the Random Forest model demonstrates superior performance in classifying the dataset, achieving high accuracy as well as balanced precision and recall.
- In the future, feature selection should be enhanced by incorporating a broader range of geochemical data and work closely with geochemists to identify key trace elements and isotopic ratios which would help filter the data leading to more accurate and scientifically grounded Random Forest model.

Links for more information:

Explore **Github** **Link to Poster**