12th International Kimberlite Conference Extended Abstract No. 12IKC-59, 2024

Predicting geochemical and isotopic compositions as well as lithosphere-asthenosphere boundary depth and diamond grade in kimberlites and lamproite using Artificial Intelligence

Luc S. Doucet[1] and Andrea Giuliani[2]
[1]Curtin University, Perth, Australia
[2] ETH, Zurich, Switzerland

Introduction

Machine learning is a subset of artificial intelligence (AI) that enables computer systems to learn from data, thereby improving their performance over time. This allows them to make predictions or decisions without being explicitly programmed. For instance, machine learning is widely used to anticipate stock market trends and has been employed to predict diamond prices (e.g., Kigo, et al. 2023). Machine learning is also starting to be used to predict tectonic settings using geochemical and isotopic data (Doucet, et al. 2022; Takaew, et al. 2024). In this study, we evaluate the capability of machine learning in predicting major and trace element compositions, isotopic ratios, lithosphere-asthenosphere boundary depths and diamond grades in diamond-bearing volcanic rocks. Our approach utilizes available compilations of major-, trace-element and radiogenic isotope compositions, including their ages and diamond grades for kimberlites and olivine lamproites, combined with independent constraints on lithospheric thickness based on xenolith and xenocryst thermobarometry Giuliani (2021, 2023a,b)

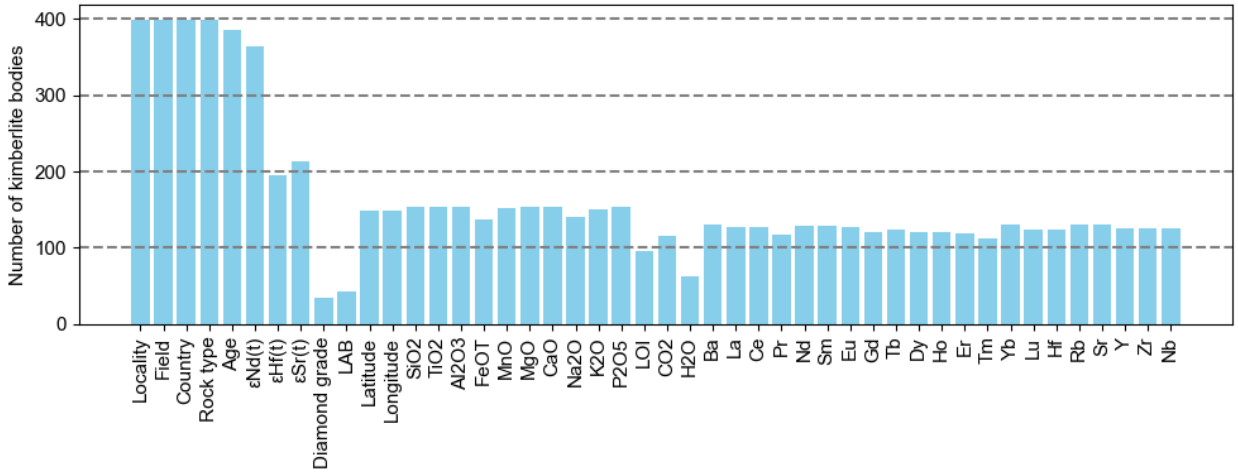

Figure 1: Histogram showing the number of kimberlite bodies (n = 399), and showing availability of data for each bodie in our dataset.

Method

The machine learning was performed using opensource program language Python 3.1 using Scikit-learn library. We used the `IterativeImputer', which employs a Bayesian Ridge estimator. This advanced imputation technique iteratively models each feature with missing values as a dependent function of other features, allowing for a more nuanced and statistically informed estimation of the missing data. Through

this method, we sought to preserve the inherent structure of the dataset, minimizing information loss and maximizing the integrity of subsequent analyses. This approach is particularly apt for datasets where the pattern of missingness is complex and not well-suited to simpler imputation methods. To evaluate prediction capabilities of our model, we intentionally removed known values of interest, such as diamond grade or lithosphere-asthenosphere boundary depth estimates, from the dataset. Then, we use the imputer to predict these values based on the rest of the dataset. This method allows us to assess how accurately our model can estimate missing information under controlled conditions, providing insights into its effectiveness in real-world scenarios where similar data might be absent or incomplete.

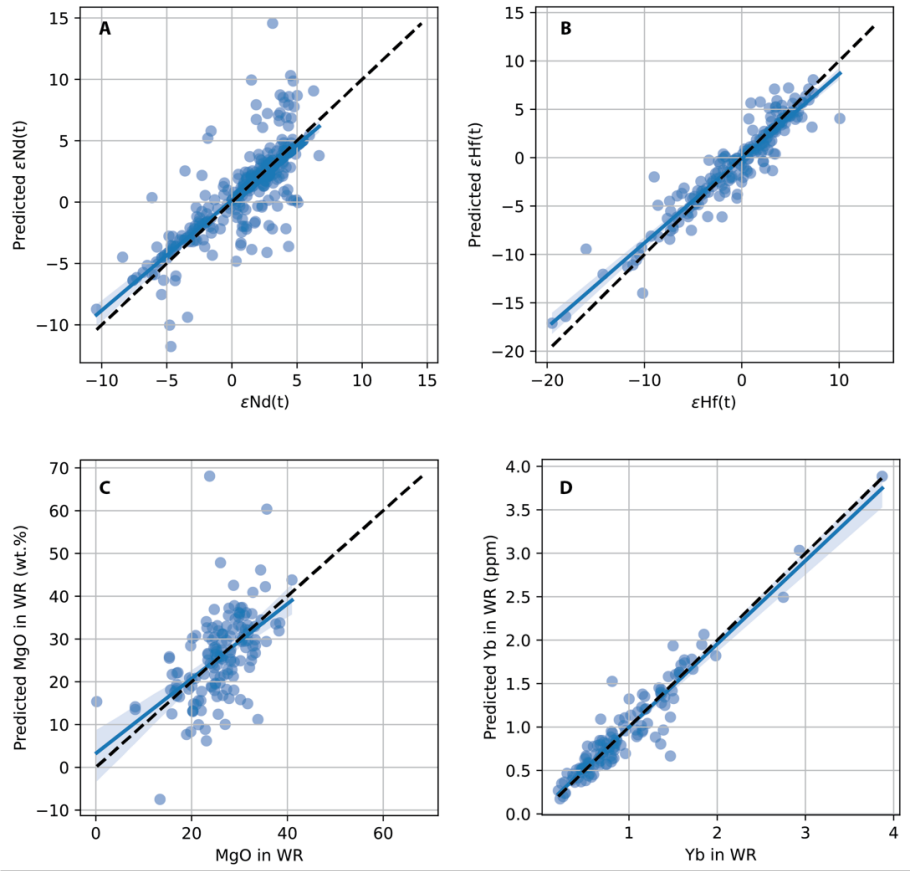

Figure 2: Whole-rock data vs predicted values for (A)εNd(t), (B) εHf(t), (C) MgO and (D) Yb.

Results

Our results show that we are able to accurately predict both Nd-Hf isotopic ratios and trace element compositions (Fig. 2A, B and D). However, the IterativeImputer did not perform as well in predicting major element compositions (Fig. 2C), which aligns with current understanding that the major element composition of kimberlites and lamproites is a complex blend of melt, mantle, and hydrothermal components. The IterativeImputer shows some capability in predicting the depth of the lithosphere-asthenosphere boundary using geochemical and isotopic compositions (Fig. 3A). This accuracy improves considerably when focusing solely on trace elements and isotopic ratios (Fig. 3C), highlighting the interplay between lithosphere thickness and melt composition. The predictions for diamond grade using the complete dataset (major-, trace-element and radiogenic isotope compositions, plus ages and lithospheric thickness) is less definitive as this stage, mainly due to limited amount of available data (Fig. 1 and Fig. 3B). We observe a systematic underestimation of diamond grade, but an overall better diamond

grade prediction when major element composition are used alone (Fig. 3D), which is promising for diamond exploration.

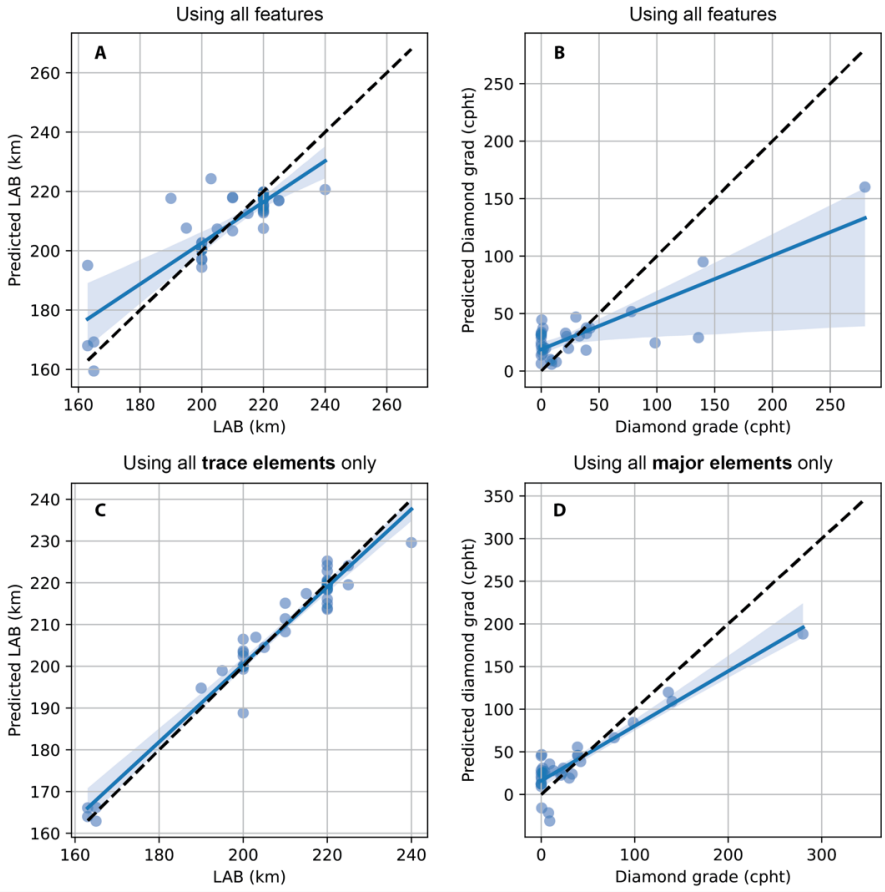

Figure 3: Lithosphere-asthenosphere boundary (LAB) and diamond grade (cpht) for kimberlites compared to predictions using all features (Fig. 1) (A and B, respectively). (C) LAB versus prediction using only trace elements, and (D) diamond grade versus prediction using only major elements.

References

Doucet LS, Tetley MG, Li Z-X, Liu Y, Gamaleldien H (2022) Geochemical fingerprinting of continental and oceanic basalts: A machine learning approach. Earth-Science Reviews 233:104192-104192 doi:10.1016/j.earscirev.2022.104192

Giuliani A, Jackson MG, Fitzpayne A, Dalton H (2021) Remnants of early Earth differentiation in the deepest mantle-derived lavas. Proceedings of the National Academy of Sciences 118(1):e2015211118

Giuliani A, Phillips D, Pearson DG, Sarkar S, Müller AA, Weiss Y, Preston R, Seller M, Spetsius Z (2023a) Diamond preservation in the lithospheric mantle recorded by olivine in kimberlites. Nature Communications 14(1):6999

Giuliani A, Schmidt MW, Torsvik TH, Fedortchouk Y (2023b) Genesis and evolution of kimberlites. Nature Reviews Earth & Environment 4(11):738-753

Kigo SN, Omondi EO, Omolo BO (2023) Assessing predictive performance of supervised machine learning algorithms for a diamond pricing model. Scientific Reports 13(1):17315

Takaew P, Xia JC, Doucet LS (2024) Machine learning and tectonic setting determination: Bridging the gap between Earth scientists and data scientists. Geoscience Frontiers 15(1):101726