

Machine-Learning Approaches to Predict Undetectable pXRF Elements: A Case Study of Na Prediction From the Northern Bushveld Complex, South Africa

Mpho Ndebele¹, Iain McDonald¹, Huw Davies¹, Hannah Hughes², Julian M. Ortiz²

¹Cardiff University, Cardiff, United Kingdom, ²University Of Exeter (Camborne School Of Mines), Penryn, United Kingdom

The use of Portable X-Ray Fluorescence (pXRF) equipment in mineral exploration provides, cheap, rapid, real-time geochemical analysis of soil and rock samples, enabling swift on-site decision-making. Despite these advantages, pXRF has limitations; particularly its inability to detect light elements (atomic numbers below Mg) and its shallow penetration depth. This limits the capability to accurately predict Na-bearing minerals when calculating a mineral balance using norms or other approaches. In the Northern Limb of the Bushveld Complex, Na is an essential component of the plagioclase–orthopyroxene–clinopyroxene dominated Platreef Ni-Cu-PGE ore system at Mogalakwena mine. This study explores the use of tree-based Machine-Learning (ML) algorithms to predict Na concentrations using pXRF analyses on diamond drill core. Four tree-based models were applied: one bagging method (Random Forest, RF) and three boosting methods (Gradient Boosting; Extreme Gradient Boosting; and Adaptive Boosting). The pXRF dataset consisted of 18 elements and four elemental ratios (Sr/Cr, Sr/Ti, Sr/Al, and Sr/Y). Three boreholes representative of the Zwartfontein mining area, with accompanying 10-cm resolution pXRF analyses and 1-m resolution 72 element assay data, were selected as the training dataset. For training, Na₂O concentrations were “borrowed” from the assay data and inserting them into the pXRF dataset. Both datasets were composited to 1-m intervals to preserve underlying stratigraphic relationships. The test dataset consisted of 2 more Zwartfontein area boreholes and 6 from exploration drilling along strike. The Gradient Boosting model performed best with predicted Na concentrations matching the observed Na and picking out the most important trends and breaks, demonstrating the value of ML techniques in predicting undetectable elements like Na. The model performed better on other Zwartfontein boreholes compared to those farther away, illustrating the impact of out-of-distribution (OOD) prediction, and the importance of selecting training data that represent the same geological domain as the target prediction area.