Determining the Lowest Quantity of Variables Required for Geometallurgical Machine Learning-Based Modelling: A Review

Ethan E. Mack, Bjorn von der Heyden, Margreth Tadie, Tobias Louw
Stellenbosch University, Stellenbosch, South Africa

Geometallurgy involves predicting the metallurgical performance of a mineral deposit, including its extraction and processing characteristics. A crucial component of geometallurgy is the prediction of extracted mineral content such as manganese, being a critical element in battery production. Machine learning has emerged as the leading industry standard for geometallurgical prediction, with random forest being one of the most universally utilised algorithms. However, using all available variables for prediction can lead to overfitting, poor generalisation performance, and increased computational cost. Therefore, determining the minimum quantity of variables required for accurate prediction is important. In this study, random forest machine learning was applied with SHAP (SHapley Additive exPlanations) analysis to identify the most important variables for manganese geometallurgy prediction modelling. SHAP analysis is an innovative technique that provides explanations for individual predictions made by machine learning models. It calculates the contribution of each variable to the prediction outcome, providing a ranking of feature importance. In this study, we used SHAP analysis to determine the minimum number of variables required for accurate prediction. Data was collected from Black Rock manganese mine located in South Africa. A random forest model was trained using all variables, followed by SHAP analysis to determine the feature importance ranking. Based on the SHAP analysis results, influential variables were retained to determine accuracy variance for the secondary model. The secondary model's results indicate that dimensionality reduction using random forest machine learning paired with SHAP analysis produces acceptable accuracy in the modelling of manganese geometallurgical processing while decreasing computational cost. In conclusion, this study highlights the usefulness of SHAP analysis in determining the most important variables for prediction modelling. By using this technique, reduction in the number of variables required for accurate prediction can be achieved, resulting in more efficient and effective manganese geometallurgy prediction models.