
The Application of Unsupervised Machine Learning to Identify Geochemical Associations with Geological Domains of Nickel Laterite Deposit in Sorowako, Indonesia

Alya Sevy Setiawan[1], Fadlin Fadlin[1], Reza Febri Agus Setyawan[2]
[1]Jenderal Soedirman University, Banyumas, Indonesia, [2]PT Vale Indonesia Tbk, East Luwu, Indonesia

Unsupervised machine learning, including Principal Component Analysis (PCA) and K-Means Clustering was employed to identify complex multi-chemistry relationships and trends within geological domains by reducing data dimensionality and performing multivariate grouping. The study utilized drill hole geochemical data from two prospect areas: “AS” and “AL”. This study used 1120 samples from Prospect AS and 1053 samples from Prospect AL with different geological settings, dunite, and lherzolite bedrock, resulting in more diverse geochemical characteristics. Elements used include Ni, Co, Fe, SiO_2, MgO, Mn, Cr, Al, and Ca. The results of PCA and clustering were interpreted based on their spatial distribution and geochemical characteristics of nickel laterite deposits, with geological domains serving as validation references to assess the objectivity and applicability of the method for exploration purposes. Robust PCA with Box-Cox (No Trim) transformation revealed that one principal component captured 96.7% of the total data variance in Hill AS, while two principal components accounted for 85.82% in Hill AL. K-Means clustering, performed using an unsupervised method, grouped data based on laterite geochemical characteristics. PCA and clustering results showed that element distributions varied by geological domain. Spatial analysis of PCA and clustering identified multi-chemistry relationships and geochemical trends in laterite deposits, with optimal cluster numbers at k = 3 and k = 4. The three-group model identified Limonite, Saprolite, and Bedrock domains, while the four-group model revealed an additional transitional domain between saprolite and bedrock, referred to as “saprock” formed through lateritization processes. The method also identified high MgO and SiO_2 contents in dunite bedrock, attributed to olivine ($(Mg,Fe)_2SiO_4$) dominance forming Mg-rich clay minerals, whereas higher Al content in lherzolite bedrock dominated by pyroxene, generated Al-rich clay minerals. Cross-matrix validation showed average accuracies of 87.9% in Hill AS and 80.13% in Hill AL, indicating strong model performance.