K-NEAREST NEIGHBORS ALGORITHM APPLIED TO LITHOLOGY PREDICTION BASED ON HANDHELD GEOCHEMISTRY AND PETROPHYSICS ANALYSIS – A CASE STUDY OF METASEDIMENTARY AND INTRUSIVE ROCKS FROM THE JACOBINA RANGE

Guilherme Ferreira da Silva[1]; Anderson Dourado Rodrigues Silva[2]; Carina Graciniana Lopes[3]; Evandro Luiz Klein[4]; Adalene Moreira Silva[5]; Catarina Labouré Bemfica Toledo[6].

[1] Geólogo, M.Sc., CPRM - Serviço Geológico do Brasil, Porto Velho, Rondônia, Brasil.
[2] Geólogo, CPRM - Serviço Geológico do Brasil, Brasília, DF, Brasil.
[3] Geóloga, M.Sc., CPRM - Serviço Geológico do Brasil, Salvador, Bahia, Brasil.
[4] Geólogo, D.Sc., CPRM - Serviço Geológico do Brasil, Brasília, DF, Brasil.
[5] Eng. Geóloga, D.Sc., Universidade de Brasília, Brasília, DF, Brasil.
[6] Geóloga, D.Sc., Universidade de Brasília, Brasília, DF, Brasil.

Integration of multi-source data for geological mapping or mineral exploration is one of the most challenging branches of earth science. The assistance of computers in recognition of hidden patterns in structured data for this purpose has been developed for the last ten years but remains underestimated in the mineral industry. This work presents a database built from petrophysics and lithochemistry assays acquired in halved core samples by portable X-ray fluorescence (pXRF), magnetic susceptibility meter and gamma-ray spectrometer. The work aims to discuss the automatic classification of rock type using the machine learning algorithm K-Nearest Neighbors (KNN) in a dataset obtained by handheld equipment. The KNN is one of the simplest machine learning algorithms and is an example of instance-based learning. In this method, new data are classified based on labeled instances. The samples originate from the Jacobina Group, a sequence of Archaean rocks, including gold-bearing metaconglomerate interspersed by other metasedimentary rocks, locally intruded by ultramafic dikes and sills. The rocks were classified as pyrite-bearing metaconglomerates with fuchsite and/or iron oxide occurring in the matrix, fuchsite-bearing quartzite, tremolitite (meta-ultramafic rock) and ordinary quartzite (without fuchsite, pyrite or iron oxide in the matrix). In the Exploratory Data Analysis step, it was possible to distinguish between sulfide-bearing samples and the others, because as the main sulfide found consists of pyrite, some samples define a linear trend in the Fe vs. S bivariate plot. Similarly, the fuchsite-bearing samples contributed to a moderate to a strong positive correlation between Al, K, and Cr (all present in the fuchsite green mica). From an initial number of 120 analysis, an equal number of samples of the four types of rocks were selected in a total of 40 samples, aiming to have a homogeneous database. Then, the database was evaluated to choose the number of variables that have statistical meaning or achieve at least 50% of the values above the detection limit. Thus, 16 variables were selected: Total Count (gamma-ray), Magnetic Susceptibility, Density and the elements Al, Si, S, Ti, V, Cr, Mn, Fe, Ni, Zr, Cd, Sn, and Sb, which were selected from the original list of 25 elements resultants from the pXRF assay. In a Principal Component Analysis, it is possible to notice some structure on data, clearly putting apart the tremolitites from the metasedimentary rocks. It is also possible to say that these metasedimentary rocks stand as a poorly distinguished cloud of points, hindering an ordinary classification. In preparation for the KNN model, the database was split into two parts aiming to compose the Training and the Testing population. Three sizes of Training Population were tested: 33% (Model 1), 50% (Model 2) and

67% (Model 3) of data, and the complementary data in each case was used as Testing population. The parameter K was fixed in odd numbers by 1 to 11, given the database size limitation. The Model 1 resulted in very poor predictions, hitting a maximum success rate of 48% with the parameter N fixed in a high value. This was expected because this model had the lowest training population. The Model 2 resulted in predictions that varied from 75% of success rate (with K fixed to 1) to 17.85% (with N fixed to 9), what gives a good but not stable solution. At last, the Model 3 resulted into a stable solution with a rate of 75% of correct predictions, in most values of K, classifying correctly in all scenarios tremolitites and ordinary quartzite, with a reasonable rate of success for metaconglomerates, but not always classifying fuchsite-bearing quartzites correctly. It may imply into one interpretation that these two last groups of rocks do not have enough contrast into physical or chemical properties evaluated in this work, or some other issue into the analytical procedures applied, as where the spot analysis of the equipment was pointed to (matrix or clast). In the other hand, some different ML algorithms must be tested to compare results and more samples could be collected to compose more significant Training and Testing Populations. Nevertheless, these preliminary results show that Applied Machine Learning is a useful tool to help in some general problems faced daily into mineral exploration.

Acknowledgment: The authors would like to thank the Exploration Staff of JMC/Yamana Gold Inc. for the donation of the drill core samples and the cooperation in this work, especially to the geologists Juliano de Souza and Wagner Pocay. Undergraduate student Matheus Ferreira da Silva and the geologists Fernando Rossi and Daniel Miranda participated in the acquisition of part of data. The local manager of the Salvador Unit of the Geological Survey of Brazil/CPRM, Valter Sobrinho, coordinated the logistics to the selection of the samples. João H. Larizzatti collaborated in the initial conception of this work.