Lithological Layer Prediction (LLP) through Machine Learning in Mineral Exploration at Paragominas Bauxite Province, Brazil

Acácio Nunes de Pina Neto[1], Jefferson Klister[2], Andras Fulop[3], Tamas Petz[3], Gustavo Loureiro[1], Dayane do Nascimento Coelho[1], Ricardo Radtke[1], Bruno Lima Gomes[4] and Helcio José Prazeres Filho[5]

1. Exploration Geologist
2. System Analyst
3. Solutions Architect, Datapao
3. Senior Data Scientist, Datapao
4. Mineral Exploration Manager
5. Mineral Exploration Senior Manager

Hydro Paragominas, Paragominas, Brazil

Corresponding author: acacio.pina.neto@hydro.com

Abstract

An accurate database is essential for geological interpretation, geological modeling, and decision-making in mineral exploration and mining activities. To ensure database reliability, validation of chemical samples is necessary to identify and correct inconsistencies as well as to minimize classification errors during data input. After a brief geological description, sampling, and lithological classification, a final validation based on laboratory results is required. Which can be a tedious and time-consuming process consisting of a set of conditions and filters applied using Excel to validate initial classification and identify outliers. This paper investigates a means to improve this process by applying a Deep Neural Network (DNN) classifier running in Azure Databricks environment (a Microsoft® platform) to make lithology predictions and database validation from chemical samples sourced from two different targets located in the Paragominas Bauxite Province (PBP), northeastern Pará State (Brazil). The application of Lithological Layer Prediction (LLP) using machine learning algorithms was found to significantly improve database validation. By analyzing and interpreting a vast amount of geological data using machine learning techniques, the accuracy and speed of lithology prediction was significantly improved. This technology has proven to be a valuable tool in identifying and characterizing bauxite deposits, allowing for more efficient and targeted exploration efforts.

Keywords: Lithology prediction, Machine learning, Bauxite, Paragominas Bauxite Province.

1. Introduction

Robust review and database validation are essential to provide valuable inputs for subsequent geological modelling and mineral resources estimates, as they ensure accuracy, completeness, and consistency of the data used in the analysis. Complexities involving historical data, many drilling campaigns as well as copious amounts of time for manual data interpretation have unique challenges and can introduce potential errors. Without proper validation, the results of the analysis can be unreliable, and decisions made based on the analysis could be incorrect.

In standard mineral exploration procedures, after a brief geological description (logging), sampling, and lithological classification, a final validation based on laboratory results is required (Figure 1). Although the latter is highly necessary, it is a time-consuming process consisting of a set of conditions and filters applied using Excel to validate initial classification and identify inconsistencies/outliers.

Figure 1. Mineral exploration workflow.

Due to these challenges, the prediction of lithological layers using machine learning (ML) models has been extensively explored [1][2][3][4]. While various machine learning techniques have been applied to lithological layer prediction, the utilization of deep learning (DL) models in conjunction with chemical components as features in the context of bauxite mines remains relatively new.

Many studies have incorporated traditional supervised techniques, such as ensemble tree models [5] or support vector machines (SVMs) [6] for lithological layer prediction [7]. Sebtosheikh *et al.* (2015) [8] demonstrated the effectiveness of SVMs in lithology prediction and provided insights on selecting optimal kernel functions and parameters for small datasets. Dev *et al*. (2019) [9] proposed the use of Extreme Gradient Boosting Trees for prediction using seismic logs, while Martin *et al.* (2021) [10] compared Extreme Gradient Boosting Trees with Convolutional Neural Networks [11] (CNNs) to improve prediction accuracy.

DL algorithms achieve high performance for classification, regression, clustering, and other applications [12][13][14]. The multiple hidden layers of DNN, the activation functions commonly employed, and methods of regularization to prevent overfitting, among other benefits allow DL algorithms to learn hierarchical feature representations of data with multiple levels of abstraction [13]. These methods create a powerful tool to help identify anomalies and reveal patterns in large datasets with less manual feature engineering than traditional ML methods.

2. Methodology

The description below presents a DNN-based method from chemical samples analysis for validation and lithological classification in geochemical data. This approach is entirely data-driven and, once the network is trained, delivers the results in real time by predicting the samples classification from input data in a single step.

2.1 Study Area

The study area is located at Paragominas Bauxite Province (PBP) central domain, in the northeast region of the state of Pará, in the Eastern Amazon, Brazil. The PBP represents one of the most important, extensive, and dense groupings of bauxite deposits in Brazil, with a potential of more than 3 billion tonnes of metallurgical ore, about 70% of Brazil's total bauxite reserves [15]. PBP is characterized by a relief of plateaus covered by a thick layer of clays (Belterra clay) and ferro-aluminous crusts. The formation of these deposits was originated by the lateritic alteration of siliciclastic deposits from the Cretaceous, in this case, sediments from the Itapecuru and the Ipixuna Formation, during the Paleogene [15] (Figure 2).

5. References

1. Dramsch, J. S. (2020). 70 years of machine learning in geoscience in review. *Advances in geophysics*, *61*, 1-55.
2. Vladimir Puzyrev, Mario Zelic, Paul Duuring. Applying neural networks-based modelling to the prediction of mineralization: A case-study using the Western Australian Geochemistry (WACHEM) database, Ore Geology Reviews, Volume 152, 2023.
3. Kirkwood, 2016 - A machine learning approach to geochemical mapping https://doi.org/10.1016/j.gexplo.2016.05.003
4. Mohamed, 2019 - Formation Lithology Classification: Insights into Machine Learning Methods https://doi.org/10.2118/196096-MS
5. Kuhn, 2018 - Lithologic mapping using Random Forests applied to geophysical and remote-sensing data: A demonstration study from the Eastern Goldfields of Australia https://doi.org/10.1190/geo2017-0590.1
6. Hearst, 1998 - Support vector machines https://doi.org/10.1109/5254.708428
7. Otchere, 2021 - Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models https://doi.org/10.1016/j.petrol.2020.108182
8. Sebtosheikh, 2015 - Lithology prediction by support vector classifiers using inverted seismic attributes data and petrophysical logs as a new approach and investigation of training data set size effect on its performance in a heterogeneous carbonate reservoir https://doi.org/10.1016/j.petrol.2015.08.001
9. Dev, 2019 - Gradient Boosted Decision Trees for Lithology Classification https://doi.org/10.1016/B978-0-12-818597-1.50019-9
10. Martin, 2021 - Centimeter-Scale Lithology and Facies Prediction in Cored Wells Using Machine Learning https://doi.org/10.3389/feart.2021.659611
11. O'Shea, 2015 - An Introduction to Convolutional Neural Networks https://doi.org/10.48550/arXiv.1511.08458
12. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
13. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.
14. Liu, W.B., Wang, Z.D., Liu, X.H., Zengb, N.Y., Liu, Y.R., Alsaadi, F.E., 2017. A survey of deep neural network architectures and their applications. Neurocomputing 234, 11–26.
15. Basile Kotschoubey et al., Caracterização e Gênese dos depósitos de bauxita da Província Bauxitífera de Paragominas, Noroeste da Bacia de Grajaú, Nordeste do Pará/Oeste do Maranhão. In: Marini Onildo João et al. (orgs.). Caracterização de depósitos minerais em distritos mineiros da Amazônia. Brasília, DF, DNPM-CT Mineral, ADIMB, 2005, 613-698.
16. Boslaugh, Sarah and Paul Andrew Watters. 2008. Statistics in a Nutshell: A Desktop Quick Reference, ch. 7. Sebastopol, CA: O'Reilly Media.
17. Pedregosa, 2011 - Scikit-learn: Machine Learning in Python https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html
18. Chen, 2016 - XGBoost: A Scalable Tree Boosting System https://doi.org/10.48550/arXiv.1603.02754
19. Breiman, L. Bagging predictors. Machine Learning 24, 123–140 (1996). https://doi.org/10.1007/BF00058655
20. He, 2015 - Deep Residual Learning for Image Recognition https://doi.org/10.48550/arXiv.1512.03385
21. Hochreiter, 1997 - Long Short-Term Memory https://doi.org/10.1162/neco.1997.9.8.1735
22. Dwarampudi, M. & Reddy, N. V. S. Effects of padding on LSTMs and CNNs (2019). https://arxiv.org/pdf/1903.07288.
23. Zaharia, 2018 - Accelerating the Machine Learning Lifecycle with MLflow https://cs.stanford.edu/~matei/papers/2018/ieee_mlflow.pdf