Development of a Model to Predict Average Moisture of Bauxite Shipments from Moisture of the Stockpile, Residence Time and Rainfall

Newton Flávio dos Santos Oliveira[1], Octávio Guimarães Rosa[2] and Gustavo Ferreira Correia[3]
1. Mining Development Analyst
2. Mine Planning Engineer Msc.
3. Mining Development Engineer
Alcoa World Alumina, Poços de Caldas, Brazil
Corresponding author: newton.oliveira2@alcoa.com

Abstract

The moisture content of bauxite is one of the main characteristics of the ore that needs constant monitoring and control, due to operational, commercial and financial impacts throughout the production chain. It is important especially in cases where ore transportation takes place in by vessels, since there are intensive material handling considerations and high costs involved in this stage of production. The Amazon Rainforest biome, where the operation of bauxite mining in the north of Brazil is located, is characterized by its high temperature and intense rainfall, directly influencing the moisture content of the ore shipped, due to storage in stockpiles, which exposes the material to the climatic conditions. The objective of this paper is to demonstrate the existence of the relationship between the average moisture of bauxite shipments, the moisture content of the stockpile, the residence time and the rainfall during the pre-shipment storage period, through a model using these variables that could predict the average moisture of bauxite shipments. In this study, different methods of data analysis were used, among them the Machine Learning, a method of data analysis that automates the construction of analytical models, providing advanced statistical analyzes from different regression methods. The following machine learning methods were directly tested: Random Forest Regression, Support Vector Regression, Gradient Boosting Regression, and Polynomial Regression, to find out the best model for the prediction of the average moisture content. For the database used for the analysis considers the period from April 2016 to March 2018 as a reference. It was possible to figure out the best parameters for each method of predicting the average moisture content of bauxite shipments, resulting in a variation between methods of at most 10 %. The Random forest regression produced the greatest results with a correlation of 0.7141 between the variables.

Keywords: Bauxite Moisture Content, Average Moisture Content; Machine Learning Regression Analysis.

1. Introduction

1.1. Amazon Bauxite Moisture Content

The bauxite deposits from the north of Brazil are located in the Amazon Rainforest with a considerable high moisture content, that may increase during the washing stage, the process step needed for the enrichment of the product, is required in order to reduce the grade(s) of reactive silica and enriching the available alumina content.

Ore shipping to refineries, when done by shipment or railroad, can be impacted by the moisture of the material since the costs in material handling is related to the quantity of mass moved. So, moving ore with high moisture content can result in greater production costs [4]. In addition, the moisture content can also create operational difficulties and delays, making civil maintenance services needed for the continuity of production [5].

Also, the refinery processes the ore by the Bayer Process - in which bauxite is treated with a hot solution of sodium hydroxide, and the high moisture content can dilute the solution, resulting in a greater consumption of sodium hydroxide, one of the main costs of refinement [4].

Alcoa's Juruti Mine is located in the north region of Brazil which is subject to high rainfall, a dry season ranging from July to December and a rainy season from January to June) [9]. It is possible to observe the effect of the climate on the moisture content of the shipped ore, having an average variation of 1 % in the moisture content of ore shipped between seasons.

The last stage in the bauxite production at the Juruti Mine is the stockpiling of the material for homogenization and drainage before the material is shipped to the refineries. The process of storing the ore in stockpiles allows the drainage of excess moisture, and it must be stored for a determined time. For iron ore the indicated period is from days to weeks, while for coal it can last from 6 – 8 weeks [8]. During the stockpiling stage, the ore is exposed to the climatic conditions of the environment, the moisture content being affected by rainfall during the time of residence.

Considering all the exposed variables and the importance of moisture content control and monitoring, the present work aimed to demonstrate the relationship between residence time of the stockpiles, the rainfall in the pre-shipment storage period, and formation moisture content with the shipped moisture content.

1.2. Multivariate Regression Analysis

The exposed problem presents 3 independent variables for the prediction of the shipped moisture content model, the stockpile formation moisture content, rainfall and time of residence; making it necessary to use multivariate analysis techniques to study the interrelationship of variables [7].

Multivariate statistics started to be cited in the early twentieth century, but it was only with the advancement of technology that it became possible to conduct studies quickly and clearly [7]. Learning the parameters of a prediction function and testing it on the same data would just repeat the labels of the samples that it has just seen but would fail to predict new unseen data. Machine learning algorithms allow us to learn the trend of the database and to generalize it as accurately as possible, thus improving the quality of the model developed [1].

Figure 1 shows the relationship between the accuracy of the model with its complexity and the point at which the model reaches the best fit of reality. The line corresponding to Training are the results obtained with the data that allows the generation of the model, and the line corresponding to generalization are the results obtained when generalizing the results of the model for the resultant simulation in the training [1].

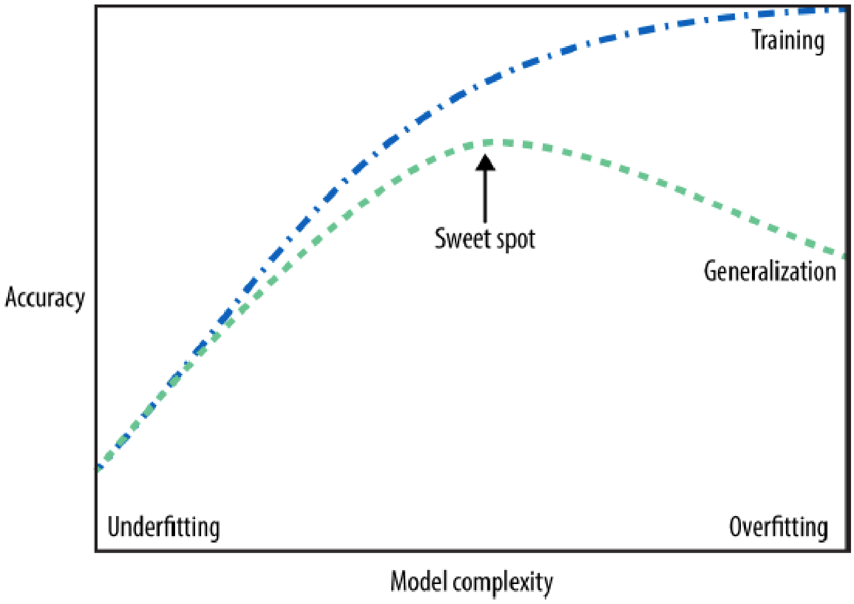

Figure 1. Relationship between the accuracy of the model with its complexity [1].

2. Materials and Methods

2.1. Data Base

The database was collected from operational control worksheets in the Juruti mine.
Assumptions:

- The stockpiles considered for the study were the final product stockpiles,
- The residence time is the time between the end of the stacking and the beginning of its recovery,
- The rainfall considered was the rains in which the material was exposed during the beginning of stacking until the end of its recovery,
- The information was collected in the meteorological station of the port of Juruti,
- The period from April 2016 to March 2018 was used as the reference.

2.2. Multivariate Regressions – Machine Learning

Many algorithms were developed for multivariate regressions using machine learning methodology. In order to obtain the best model, four different techniques were applied:

- Multivariate Polynomial Regression – Polynomial Regression is used when the relation between the variables is non-linear. [1]
- Random Forest Regression – This method adds a "*layer*" of randomness to bagging, changing how the regression trees are constructed, resulting in a subset of predictor randomly chose at each node [2].
- Support Vector Regression – This method implements the risk minimization inductive principle, in order to obtain the best generalization of the model, involving simultaneous simulations to minimize the empirical risk [3].
- Gradient Boosting Regression – This method is based on constructing regression models by adjusting the parameterized function to the current "pseudo"-residuals by the least square of each interaction [6].

2.3. Data Grouping

In order to obtain a better correlation between the variables, the monthly data grouping was performed. A weighted average of each variable was performed as a function of the shipped mass of each ship to make the model more representative.

3. Results and Discussion

Aiming to demonstrate the correlation between the variables and to find the best method to develop a model to predict the average moisture of bauxite shipments it was applied four different methodologies of multivariate regressions using machine learning. Also, different scenarios were simulated considering different features applying in the different methods of regression, evaluating the impact of each variable in the model.

A preliminary statistical analysis of the database was performed in order to achieve a better level of knowledge of the database used in the simulations. Table 1 shows the results:

Table 1. Database exploration statistics.

	Rainfall	Residence Time	Stockpile Moisture Content	Average Shipment Moisture
mean	23.7	1.8	13.2	13.2
std	26.3	1.9	0.5	0.5
min	0.8	0.0	12.3	12.2
25%	6.1	0.2	12.8	12.9
50%	15.2	0.8	13.2	13.2
75%	35.0	3.2	13.6	13.5
max	99.3	6.9	14.0	14.2

Figure 2, below, shows the average R-Squared obtained in training and tests simulations and the standard deviation of the R-Squared obtained. Multivariate Polynomial Regression (MPR) and Support Vector Regression (SVR) showed the worst results between the tools applied. Despite the good results with just one feature, Gradient Boosting Regression (GBR) showed a slightly smaller average R-Squared with two (2) and three (3) features. The table 2, show the values resulted in the simulations.

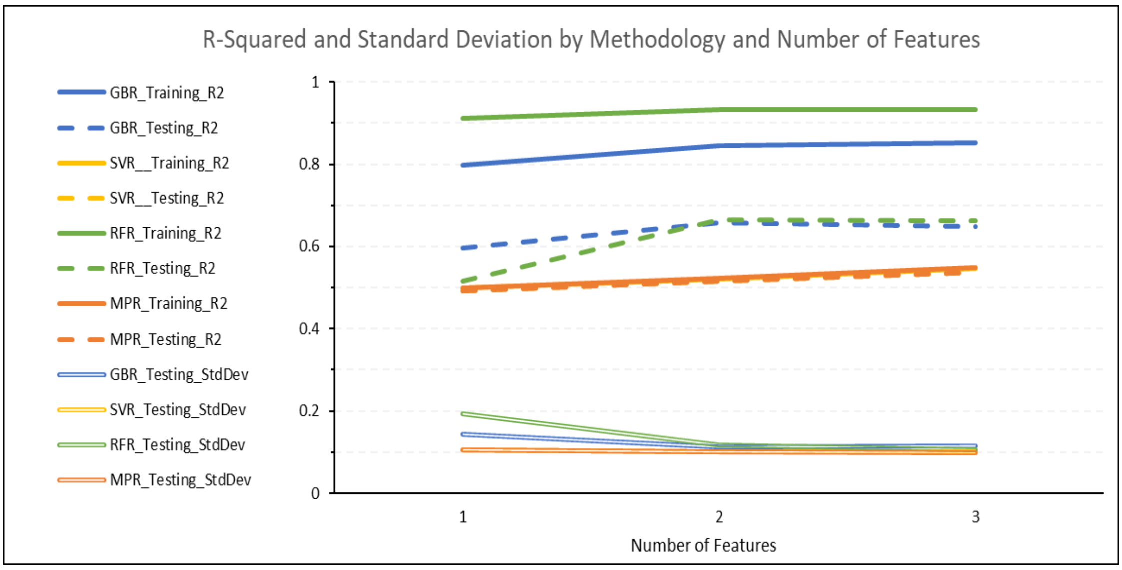

Figure 2. R-Squared and standard deviation of the different regression methodologies.

Table 2. R-Squared resulted from simulations of the different methods of multivariate regressions.

	Gradient Boosting Regression			Random Forrest Regression		
Feature	Training R-Squared Average	Testing R-Squared Average	Standard Deviation of R-Squared	Training R-Squared Average	Testing R-Squared Average	Standard Deviation of R-Squared
1	0.80	0.60	0.14	0.91	0.52	0.19
2	0.85	0.66	0.11	0.93	0.67	0.12
3	0.85	0.65	0.11	0.93	0.66	0.10
	Support Vector Regression			Multivariate Linear Regression		
Feature	Training R-Squared Average	Testing R-Squared Average	Standard Deviation of R-Squared	Training R-Squared Average	Testing R-Squared Average	Standard Deviation of R-Squared
1	0.50	0.49	0.11	0.50	0.49	0.11
2	0.52	0.52	0.10	0.52	0.51	0.10
3	0.55	0.54	0.10	0.55	0.54	0.10

Since the greatest average R-Squared obtained in the simulations was via Random Forrest Regression, this method was chosen as the method to obtain the model to predict average moisture of Bauxite Shipments. Aiming to finding the best methodology, simulations varying the features and parameters of the regression was performed. The figure 3 and table 3 shows the results.

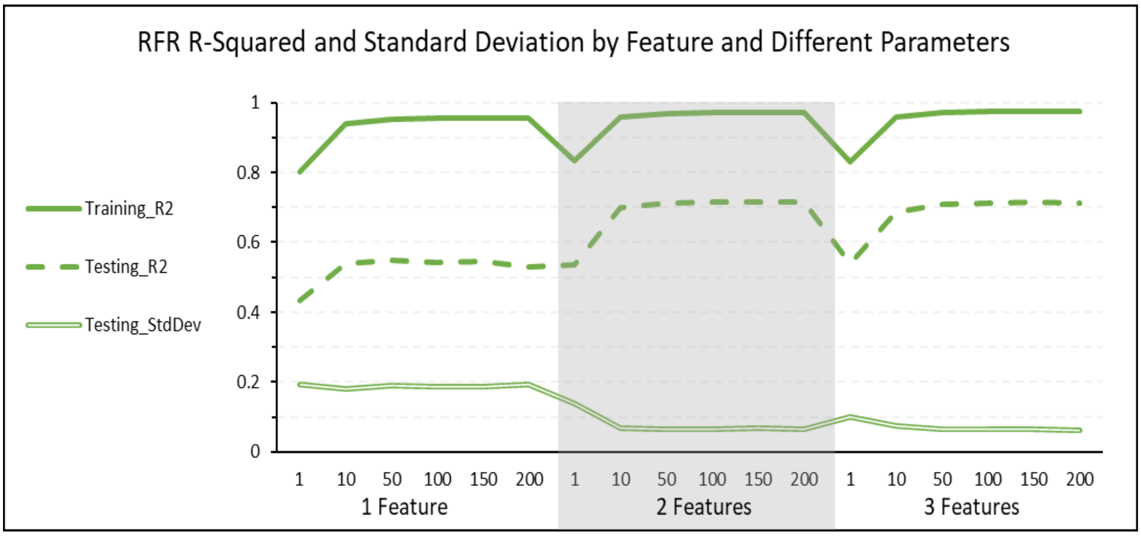

Figure 3. R-Squared and standard deviation obtained from different simulations of Random Forrest Regression using 1, 2 or 3 features.

Table 3. R-Squared resulted from Random Forest Regression simulations varying the features applied and the parameters of the simulation.

	1 Feature			2 Features			3 Features		
Parameters	Training R-Squared Average	Testing R-Squared Average	Standard Deviation of R-Squared	Training R-Squared Average	Testing R-Squared Average	Standard Deviation of R-Squared	Training R-Squared Average	Testing R-Squared Average	Standard Deviation of R-Squared
1	0.8016	0.4325	0.1924	0.8335	0.5359	0.1396	0.8308	0.5407	0.1006
10	0.9392	0.5399	0.1813	0.9572	0.7004	0.0671	0.9572	0.6850	0.0737
50	0.9535	0.5501	0.1898	0.9688	0.7121	0.0661	0.9712	0.7074	0.0665
100	0.9546	0.5409	0.1875	0.9705	0.7143	0.0656	0.9730	0.7131	0.0642
150	0.9555	0.5440	0.1884	0.9712	0.7138	0.0673	0.9737	0.7147	0.0641
200	0.9549	0.5296	0.1944	0.9713	0.7142	0.0669	0.9740	0.7134	0.0636

As discussed previously, when applying regressions to develop a model to predict a variable it's important to generalize the data assuring the most accurate prediction in the unseen data. Therefore, this paper aimed to find the greatest correlation in the testing simulations considering different parameters and features, showed by Random Forest Regression with a parameter of 150 and all the 3 features.

The best method showed an averaged R-Squared of 0.7147, with a standard deviation of 0.0641.

4. Conclusion

There was a good correlation among the average moisture of bauxite shipments, the moisture content of the stockpile, the residence time and the rainfall during the pre-shipment storage. Evaluating different method of regression, it was determined that the Rain Forest Regression as the best method to develop a model to predict average moisture of bauxite shipments from moisture of the stockpile, residence time and rainfall.

The best parameter to be used in the Rain Forest Regression is 150, with the 3 features, allowing to predict the average moisture of bauxite shipments in a monthly basis, enabling a forecast for operational decision-making and foresee the moisture impact over the production.

5. References

1. Andres C. Müller; Sarah GUIDO. *Introduction to Machine Learning with Python***.** Sebastopol: O'riley Midia, Inc., 2017.
2. Andy Liaw; Matthew Wiener. Classification and Regression by Random Forest. *R News.*, Austria, v. 2/3, p.18-22, dec. 2002. <: https://www.researchgate.net/publication/228451484>. Accessed by: 15 Aug 2018.
3. Debasish Basak; Srimanta Pal; Dipak Chandra Patranabis. Support vector regression. *Neural Information Processing* – Letters and Reviews, Switzerland, v. 11, n. 10, p.203-224, oct. 2007.
4. Everton de Melo Dias. Alternativas para a redução de umidade em minério de alumínio. Estudo de Caso: Votorantim Metais/ Poços de Caldas / MG. 2014. *Universidade Federal de Alfenas*, Poços de Caldas, 2014.
5. Jaqueline do Carmo Ferreira. Estudo sobre drenagem e redução de umidade do minério de ferro de Carajás com a utilização de Geossintéticos. 2009. *Universidade de Brasília, Brasília,* Brasília, 2009.
6. Jerome H. Friedman. Stochastic gradient boosting. *Computational Statistics & Data Analysis*, Rome, v. 38, n. 4, p.367-378, feb. 2002.
7. Lorena Vicini. Análise Multivariada da Teoria À Prática. Santa Maria*: UFSM*, 2005.
8. S. McElwain; F. Fulford; “Moisture movement in bulk stockpiles”, *Proceedings of the Mathematics-in-Industry Study Group* (John Hewitt, editor, 1995), 48-65.
9. Thadeu Keller Filho; Eduardo Delgado Lima and Paulo Roberto Schubnell De Rezende Assad. Regiões pluviometricamente homogêneas no Brasil. *Pesq. agropec. bras.* [online]. 2005, vol.40, n.4, pp.311-322. ISSN 0100-204X. <http://dx.doi.org/10.1590/S0100-204X2005000400001>.