Use of Advanced Analytics to Forecast Low Efficiency Pots

Walisson Alves Aguiar[1], Michel Adolfo dos Santos[2], Jean Carlos Pardo[3], Diego Cota Marinho[4], Nilton Freixo Nagem[5], Kenji U. Junior[6], Hugo T. Teixeira[7], Luiz Joaquim Dias de Lima Nunes[8] and Yugo Watari[9].

1. Reduction Process Coordinator
2. Integrated System Consultant
3. Reduction Process Manager
4. Reduction Production Manager
5. Reduction Principal Process Consultant

Companhia Brasileira de Alumínio (CBA), Aluminio, Brazil

6. Data Analyst
7.Data Consultant
8. Data Consultant
9. Position of Analytics Manager

Votorantim, São Paulo, Brazil

Corresponding author: walisson.aguiar@cba.com.br

Abstract

CBA is a Brazilian aluminum company with VSS Søderberg technology and is moving towards Industry 4.0 to improve potroom operations and make more consistent and predictable performance. One of several initiatives is the use of advanced data analytics in a cloud environment to forecast pot conditions in advance to avoid pot problems and maintain high productivity. The project started in 2020 and after several attempts with different numbers of variables the neural networking configuration was selected as the best candidate. The training data set has two years of history and has more than ten variables in a weekly average to check pot conditions as a supporting tool for the technical team. The test was conducted in our six potrooms with two groups: the control group and the test group (calculated by the neural network). The algorithm forecasts one week ahead the probability to lose production by pot. After one year of test the results showed 12 % decrease in production loss due to early diagnosis and action. The results were only achieved through combined software and human effort. This work changes the potroom operation from a reactive and manual interference in low production pots to a proactive and automated early detection.

Keywords: Aluminum production, Advanced data analytics, Neural networks, Industry 4.0.

1. Introduction

CBA is a Brazilian aluminum company with vertical stud Søderberg (VSS) technology and is moving towards the 4.0 industry to improve potroom operations making it with more consistent and predictable performance. From Cyber-Physical Systems, Internet of Things and Internet of Services, production and processes tend to become more efficient, autonomous and customizable. The CBA Industry 4.0 program focuses on several initiatives such as: artificial intelligence (AI), business intelligence and others, as well as a pipeline of R&D projects, which are expected to enhance the sustainability of processes and products.

This paper explores one of the fields of Industry 4.0 Big Data Analytics. These are very extensive and complex data structures that use new approaches to collect, analyze and manage information. Applied to Industry 4.0, Big Data technology consists in handling relevant information: Connection (to industrial network, sensors and PLCs), Cloud (cloud/data on demand), Cyber (model and memory), Content, Community (sharing of information) and Customization

(customization and values). From Big Data technology emerges the Advanced Analytics that is a methodology of data analysis that uses predictive modeling, machine and deep learning algorithms, process automation and other statistical methods to analyze information from a different data source.

Nowadays one of the best performance algorithms that ensemble classifiers/predictors is the extreme gradient boosting (XGBoost). It is an optimized distributed gradient boosting algorithm designed to be highly efficient, flexible and portable. It implements machine learning algorithms under the Gradient Boosting framework. The algorithm category is based on Decision Trees with Gradient Boosting. Gradient means that the algorithm uses the Gradient Descent algorithm to minimize loss. It has many hyperparameters that can be tuned to adjust the algorithm to best suit to the scenario. Decision Trees are methods where there is a function that takes a value as a vector of attributes as input and returns a decision (output). For a decision tree to arrive at the output value, it performs a series of steps, creating various branches throughout the process. Each node in this tree represents a single decision. The more times an attribute is used for decision making, the greater its relative importance in the model. Gradient boosting is a machine learning technique, which produces a prediction model in the form of an ensemble prediction models, typically decision trees [1, 2]. When a decision tree is the weak learner, the resulting algorithm is called gradient boosted trees, which usually outperforms random forest [2, 3].

2. Experimental

The first step was to build the data architecture that consists in the acquisition of data from our source of historic process data (PIMS), where the tags are sequenced and queued for consumption based on the frequency of capture by the sensors. The Data Ingestion process uses the extraction, transformation and loading (ETL) resources combined between the Data Factory and the Event Hub, forming the first stage, where the data presents a raw and semi-structured format.

In the next step, we internalize the information in our Data Lake, producing an accurate version, enriched with other information from auxiliary and complementary spreadsheets, as well as a source that is fed back by Tableau's own visualization, which allows an evaluation of the actions taken later.

In the intelligence stage, we apply machine learning algorithms in order to obtain the evaluated insights from the period in question, acting in the prescriptive way of analyzing the acquired variables.

Some of the parameters used in the algorithm were: DEPTH, SHRINKAGE, SUBSAMPLE, COLSAMPLE_BYTREE, GAMMA, MIN_CHILD_WEIGHT. Where DEPTH means maximum depth of a tree. Increasing this value will make the model more complex and more likely to overfit. SHRINKAGE which means in modifying the update rule. It can increase the computational time both during training and querying. SUBSAMPLE is the ratio of the training instances. COLSAMPLE_BYTREE is the subsample ratio of columns when constructing each tree. GAMMA is the minimum loss reduction required to make a further partition on a leaf node of the tree. MIN_CHILD_WEIGHT is the minimum sum of instance weight needed in a child. If the tree partition step results in a leaf node with the sum of instance weight less than min_child_weight, then the building process will give up further partitioning.

The general objective function of the gradient tree boosting is presented by Equation (1)

$$\gamma_{jm} = \arg min_{\gamma} \sum_{x_i \epsilon R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma) \tag{1}$$

5. Conclusion

Data analytics became a new reality as a tool for potroom decisions. This project is still improving the results as the overall knowledge increases. Several optimization algorithms are planned to be tested to improve the results. The current project decreases the loss of production by early pot diagnosis but it requires the potroom engagement not only by using the tool but by putting in the system the correct root cause to support the weights of the model.

6. References

1. Tianqi Chen and Carlos Guestrin, Xgboost: A scalable tree boosting system, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, 785–794.
2. Yoav Freund and Robert E. Schapire, Experiments with a new boosting algorithm, *Proceedings of the Thirteenth International Conference on Machine Learning*, 1996, 148–156.
3. Jerome H. Friedman, Stochastic gradient boosting, *Computational Statistics & Data Analysis, Elsevier*, 38(4), 2002, 367–378.