ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Estimate the Percentage of Silica in Iron Ore Using Machine Learning

K.Srinija[1], A.Nikita[2], G.Gopichand[3], B.Pushpa[4]
Aditya Institute Of Technology and Management
Department of Computer Science And Engineering

Abstract: A study of silica and iron recovery from the iron ore concentration tailing is presented. The residue is composed of 40.1% Fe, 33.4% SiO_2, 8.31% Al_2O_3, 0.08% P, and 0.34% MnO2. Iron is the most used material in the world. It is used in architecture, bearings, surgical instruments and jewellery. We get iron from iron ore. Almost 98% of iron ore is used in steelmaking. The iron ore has an impurity called silica. The impurity such as silica adversely affect the production of iron and steel. In this system , we predict the percentage of silica concentrate in iron ore using a random forest. Random forest is a one type of supervised learning method used for classification and regression. First we loaded the data set into the system and then we preprocessed the data by handling missing values, removing outliers and correlated variables . We Split the data into 70% as a training set and 30 % as a testing set. With the help of a training data set, we train our model using a random forest algorithm. After that we tested our model with our test data set. we got an accuracy of 91% within a short period of time which will be very useful in the real world without any delays .Then we have deployed our model.

I. INTRODUCTION

The investigation of primordial drivers regarding iron ore quality recovery in froth flotation processing plant has been of great interest lately. As seen in the Financial Times, ore iron, the large raw material for steel production are more integral to the global economy than any other commodity, except perhaps soil [1]. It has been proven that approximately 2.5-3.0 tons of iron ore tailings are discharged for every one ton of iron ore produced. Moreover, statistics show that there is more than 130 million tons of iron ore produced annually. This indicates that if for example the mine tailing dams contain an average of approximately 12% iron ore, there would be approximately 1.52 million tons of iron wasted each fiscal year [2]. It takes hours to ascertain the two variables of interest, which are the percentage of iron ore and silica concentrate. Such practice, however has demonstrated to be a non-novel technique to monitor and control the global unit circuit of the froth flotation system. In addition, the concomitant variation in the ore feed coupled with output stipulation changes make it cumbersome to put the plant in a steady state as a result of lengthy delays of the laboratory test. This includes reduction of ore feed rate or increasing or decreasing the reagents flow or air flow.

Cite this article as: K.Srinija, A.Nikita, G.Gopichand & B.Pushpa., " Estimate the Percentage of Silica in Iron Ore Using Machine Learning", International Journal & Magazine of Engineering, Technology, Management and Research (IJMETMR), ISSN 2348-4845, Volume 9 Issue 6, June 2022, Page 42-46.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

II. EXISTING SYSTEM

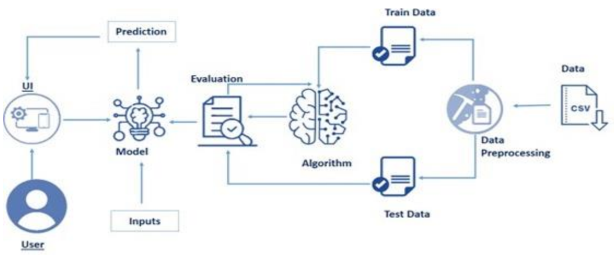
Forth Flotation is defined as a flotation process in which the minerals floated gather in and on the surface of bubbles of air or gas driven into or generated in the liquid in some convenient manner. The power input and floaatation technique required for optimum flotation condition is also determined by agitation. However, the forth is collected and dried under controlled sunlight. Most importantly, the percentage of silica concentrate and iron ore concentrate are ascertained at this crucial phases of the plant processes. Enterprises often make concerted efforts to collect empirical data at each processing phase and store in their repository database for consumption.

Drawbacks:

1. Time consumption process
2. Involves high cost
3. Sampling period takes more than 2 hours

III. PROPOSED SYSTEM

The main goal of this project is to build a Machine Learning model to predict the silica concentrate present in Iron ore. The user interacts with the UI (User Interface) to enter the data. Then, entered data is analyzed by the model which is integrated. Once the model analyzes the input, the prediction is showcased on the UI.



IV. Algorithms Used

Random Forest: Random Forest is a one type of supervised learning methods used for classification, regression, and outliers detection. All of these are common tasks in machine learning. The random forest algorithm is an extension of the bagging method as it utilizes both bagging and feature randomness to create an uncorrelated forest of decision trees. Feature randomness to create an uncorrelated forest of decision trees also known as feature bagging, genetares a random sunset of features which ensures low correlation among decision trees .While decision are common supervisied learning algorithms, they can be prome to problems, such as bias and over fitting. However, when multiple decision trees form an ensemble in random forest algorithm they predict more accurate results, particularly when a individual trees are uncorrelated with each other.

How an random forest works?

Random forest algorithms have three main hyper parameters, which need to be set before training. These include node size, the number of trees, and the number of features sampled. From there, the random forest classifier can be used to solve regression or classification problems

The random forest algorithm is made up of a collection of decision trees, and each tree in the ensemble is comprised of a data sample drawn from a training set with replacement, called the bootstrap sample. Of that training sample, one-third of it is set aside as test data, known as the out-of-bag (oob) sample, which we'll come back to later. Another instance of randomness is then injected through feature bagging, adding more diversity to the dataset and reducing the correlation among decision trees.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

The advantages of random forest are:

1. Random Forest can be used to solve bothclassification as well as regression problems.
2. Random Forest works well with both categorical and continuous variables.

3. No feature scaling required: No feature scaling (standardization and normalization) required in case of Random Forest as it uses rule-based approach instead of distance calculation.

4. Random Forest can automatically handle missing values.

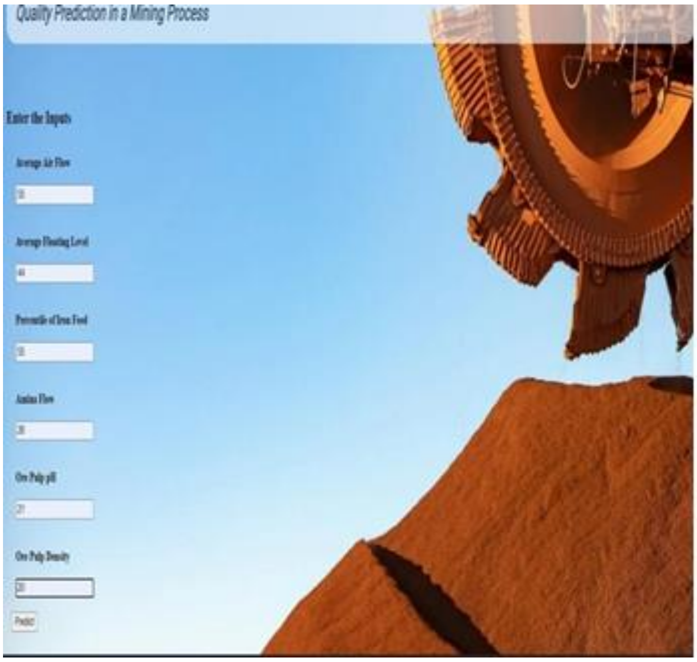
IV. DATA SET

For our model , we obtained the dataset called "Quality Prediction in a Mining Process" obtained from Kaggle is used . Since it is a mining plant data, it is the product of a manufacturing process . The data were collected from March of 2017 until September of 2017 during 6 months .

The instances are the observed measurement values which were recorded in every hour and in every 20 seconds .

Structure of a Data Set:

The dataset contains 24 columns representing the measurements, 737,453 samples exist. The 24 columns include the date and time of the measurement, which will not be used as an input feature. The last columns of the dataset represent the targets of this prediction task: the percentages of iron ore and silica concentrate, which are highly inversely correlated. Our goal is to predict silica concentrate without the use of iron concentrate. The other 21 columns will be used as features for predicting the target value.



COLUMN PROCESS	DESCRIPTION OF VARIABLES IN FORTH PLANT
Date	date of the measurement
% Iron Feed	% of Iron that comes from the iron ore that is being fed into the flotation cells
% Silica Feed	% of silica (impurity) that comes from the iron ore that is being fed into the flotation cells
Starch Flow	Starch (reagent) Flow measured in m3/h
Amina Flow	Amina (reagent) Flow measured in m3/h
Ore Pulp Flow	t/h
Ore Pulp pH	pH scale from 0 to 14
Ore Pulp Density	Density scale from 1 to 3 kg/cm^3
Flotation Column 01 Air Flow	Air flow that goes into the flotation cell measured in Nm^3/h
Flotation Column 02 Air Flow	Air flow that goes into the flotation cell measured in Nm^3/h
Flotation Column 03 Air Flow	Air flow that goes into the flotation cell measured in Nm^3/h
Flotation Column 04 Air Flow	Air flow that goes into the flotation cell measured in Nm^3/h
Flotation Column 05 Air Flow	Air flow that goes into the flotation cell measured in Nm^3/h
Flotation Column 06 Air Flow	Air flow that goes into the flotation cell measured in Nm^3/h
Flotation Column 07 Air Flow	Air flow that goes into the flotation cell measured in Nm^3/h
Flotation Column 01 Level	Froth level in the flotation cell measured in mm (millimeters)
Flotation Column 02 Level	Froth level in the flotation cell measured in mm (millimeters)
Flotation Column 03 Level	Froth level in the flotation cell measured in mm (millimeters)
Flotation Column 04 Level	Froth level in the flotation cell measured in mm (millimeters)
Flotation Column 05 Level	Froth level in the flotation cell measured in mm (millimeters)
Flotation Column 06 Level	Froth level in the flotation cell measured in mm (millimeters)
Flotation Column 07 Level	Froth level in the flotation cell measured in mm (millimeters)
%Iron Concentrate	% of Iron which represents how much iron is presented in the end of the flotation process (0-100%, lab measurement)
% Silica Concentrate	% of silica which represents how much iron is presented in the end of the flotation process (0-100%, lab measurement)

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Dataset	MAE	RMSE
Data1	0.4938	0.9081
Data2	0.5141	0.9093

RStudio generated detail statistics of the dataset such as the mean, median, min and so forth. The aim was to assess skewness of each variable and detecting outliers. It was realized that all variables were stored as numeric with the exception of date variable which was stored as a factor. Kaggle is an online community for descriptive analysis and predictive modelling. It collects variety of research fields' dataset from data analytic practitioners. Data scientists compete to build the best model for both descriptive and predictive analytic. It however allows individual to access their dataset in order create models and also work with other data scientist to solve various real world analytics problems.The input dataset used in developing this model has been downloaded from Kaggle . The dataset contains design characteristics of iron ore froth flotation processing plant which were put together within three (3) months. This is nicely organized using common format and a standardized set of associate features of iron ore froth flotation system.

VI. Result and Analysis

After running the python application, we get an output screen like the image below. Then we have to fill in the values of features in their respective fields. After filling the values, we have to click the predict button which is present at the end of the page.

After clicking the predict button, we get the percentage of silica in the iron ore. We can see the output in the below image. By using a Random Forest Regressor , we got an accuracy of 91% within a short period of time which will be very useful in the real world without any delays

VII. Conclusion

In this study, the empirical evidence of building machine learning algorithm using the past 3 months of iron ore froth flotation processing plant dataset is examined. This study aims to investigate data driven model to estimate the percentage of silica concentrate in iron ore froth flotation processing plants in real-time. It has been shown that Random Forest improved the execution time, enabling a real-time silica prediction of 7 seconds in best case scenario running on a platform-based intel i7. we got an accuracy of 91% within a short period of time which will be very useful in the real world without any delays. This study has not only selected highly significant features for the stochastic nature inherent in the iron ore froth flotation dataset, but also presented a general model that should work well for other froth flotation plants which have similar attributes in the data sample. Future study for further development of the method will be in two-fold. On the one hand, a validation study should be conducted in order to evaluate the predictive power of the proposed model using a dataset of more parameters and factors like froth physical characteristics such as speed, size distribution, color, bubbles shape and size from a different study.

VIII. References

1. Galit Shmueli. (2018). Data mining for business analytics : Concepts, techniques, and applications in R. Hoboken, New Jersey: John Wiley & Sons, Inc.

2. Awson, P., & Koorts, R. (2014). Flotation Control Incorporating Fuzzy Logic and Image Analysis: IFAC Proceedings Volumes, 47(3), 352—357.

3. Nakhaei, F., & Irannajad, M. (2015). Application and comparison of RNN, RBFNN and MNLR approaches on prediction of flotation column performance. International Journal of Mining Science and Technology, 25(6), 983—990.

4. Luo, X., Feng, B., Wong, C., Miao, J., Ma, B., & Zhou, H. (2016). The critical importance ofpulp concentration on the flotation of galena from a low grade lead—zinc ore. Journal ofmaterials Research and Technology, 5(2), 131—135.

5. Horn, Z. C., Auret, L., mccoy, J. T., Aldrich, C., & Herbst, B. M. (2017). Performance of Convolutional Neural Networks for Feature Extraction in Froth Flotation Sensing. Ifacpapersonline, 50(2), 13—18.

6. Horn, Z. C., Auret, L., mccoy, J. T., Aldrich, C., & Herbst, B. M. (2017). Performance of Convolutional Neural Networks for Feature Extraction in Froth Flotation Sensing. Ifacpapersonline, 50(2), 13–18.

7. Bergh, L. G., & Yianatos, J. B. (2011). The long way toward multivariate predictive control offlotation processes. Journal of Process Control.

8. Eduardo Magalhães Oliveira. (2017). Quality Prediction in a Mining Process. Retrieved September 7, 2019, from Kaggle.com