International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 07 | July - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Survey paper on Iron Ore Impurity Prediction Using Machine Learning

Dr M Prasad[1], Chandrashekar C[2]

[1]Associate Professor, Department of Master of Computer Application, BIET, Davangere
[2]Student, Department of Master of Computer Application, BIET, Davangere

ABSTRACT:

The main goal of this project is to predict how much impurity is in the ore concentrate. The% of Silica is measured in a lab experiment it takes at least one hour for the process engineers to have this value. As this impurity is measured every hour, if we can predict how much silica (impurity) is in the ore concentrate, we can help the engineers, giving them early information to take actions (empowering!). Hence, they will be able to take corrective actions in advance (reduce impurity, if it is the case) and also help the environment (reducing the amount of ore that goes to tailings as you reduce silica in the ore concentrate).

1.Introduction

The approach is simple. It aims whether we can predict the silica concentrate without iron concentrate and approached with simple way of developing the model with concentrate and model without concentrate and compare the performance of model using various regression metric like R^2 or MAE and drawing conclusion based on the results.

When multiple dependent variables exist in a regression model, this task is called as multi-target regression. In this case, a multi-output regressor is employed to learn the mapping from input features to output variables jointly. In this study, multi-target regression technique is implemented for quality prediction in a mining process to estimate the amount of silica and iron concentrates in the ore at the end of the process. In the experimental studies, different regressors that use Random Forest, AdaBoost, k-Nearest Neighbours and Decision Tree algorithms separately in the background were compared to determine the best model. Coefficient of determination (R^2) measure was used as the evaluation metric. There are some studies that predict iron concentrate and silica concentrate separately. However, this Model provides a new contribution to the field by calculating these two values jointly since they have a great correlation.

Our Approaches is whether
1. % Iron Concentrate is correlated with % Silica Concentrate
2. Predict the % silica concentrate without using % iron concentrate.
3. If it is correlated and we can predict both % Iron and Silica concentrate at same time using power of ML and DL .

Research Objectives

1. To evaluate the feasibility of using machine learning algorithms to predict in real-time the percentage of silica concentrate of froth flotation processing plant.

2. Model selection: The project finds out which variable associated with iron ore extraction is statistically significant.

3. Estimate: The project will propose a model to predict percentage of silica concentrate in froth flotation

2. LITERATURE REVIEW

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/cm³
Flotation Column 01 Air Flow	Air flow that goes into the flotation cell measured in Nm³/h

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 07 | July - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Flotation Column 02 Air Flow	Air flow that goes into the flotation cell measured in Nm³/h
Flotation Column 03 Air Flow	Air flow that goes into the flotation cell measured in Nm³/h
Flotation Column 04 Air Flow	Air flow that goes into the flotation cell measured in Nm³/h
Flotation Column 05 Air Flow	Air flow that goes into the flotation cell measured in Nm³/h
Flotation Column 06 Air Flow	Air flow that goes into the flotation cell measured in Nm³/h
Flotation Column 07 Air Flow	Air flow that goes into the flotation cell measured in Nm³/h
Flotation Column 01 Level	Froth level in the flotation cell measured in mm (millimetres)
Flotation Column 02 Level	Froth level in the flotation cell measured in mm (millimetres)
Flotation Column 03 Level	Froth level in the flotation cell measured in mm (millimetres)
Flotation Column 04 Level	Froth level in the flotation cell measured in mm (millimetres)
Flotation Column 05 Level	Froth level in the flotation cell measured in mm (millimetres)
Flotation Column 06 Level	Froth level in the flotation cell measured in mm (millimetres)
Flotation Column 07 Level	Froth level in the flotation cell measured in mm
%Iron Concentrate	% of Iron which represents how much iron is presented in the end of the flotation process
% Silica Concentrate	% of silica which represents how much iron is presented in the end of the flotation process

2.1 Source of Data

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 months. This is nicely organized using common format and a standardized set of associate features of iron ore froth flotation system.

Structure of Dataset

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. Description of each feature can be found in Table above

2.2 Proposed Solution

Over the past two decades, there has been an upsurge of academic research work within froth flotation process fraternity. Though, a significant number of the plant processing problems are being successfully modelled using machine learning algorithms but other unresolved issues and impediment still remain.

Random Forest Regressor

This method basically trains a number of classifying decision trees on various different subsamples. It benefits from averaging mechanism to improve the predictive accuracy and to control over-fitting. Training samples are randomly selected with replacement. The size of each new training set is the same as the original dataset. That is to say, a chosen instance is likely to be chosen again and again as an element of distinct subsets. As input parameters, the number of trees in the algorithm and maximum depth should be determined initially. The change in their values may affect the performance and predictive power of the algorithm. Therefore, all possible parameters in the range for the size of the dataset are given to the method and tested. The parameters leading to best results become candidates to be used. This method performs efficiently without causing too much computational cost.

3. THEORETICAL ANALYSIS

When multiple dependent variables exist in a regression model, this task is called as multi-target regression. In this case, a multi-output regressor is employed to learn the mapping from input features to output variables jointly. In this study, multi-target regression technique is implemented for quality prediction in a mining process to estimate the amount of silica and iron concentrates in the ore at the end of the process.

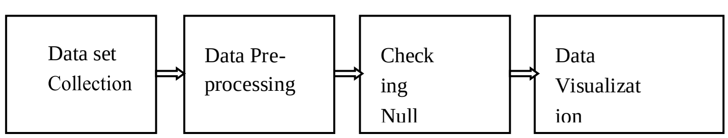
In this study, two inter-dependent single target regression tasks are transformed into a multiple output regression problem for quality prediction in a mining process.

In the previous models have been conducted to estimate silica concentrate with or without taking iron concentrate as input parameter. In this aspect, the problem is a single-target regression problem. However, this study that focuses on the estimation of both iron and silica concentrates simultaneously as output variables. We compared different multi-target regressors that use Random Forest, AdaBoost, XGBOOST ,RIDGE and Decision Tree algorithms separately in the background. Coefficient of determination (R^2) metric and MSE was used to evaluate predictive performance of the regression methods for the mentioned data.

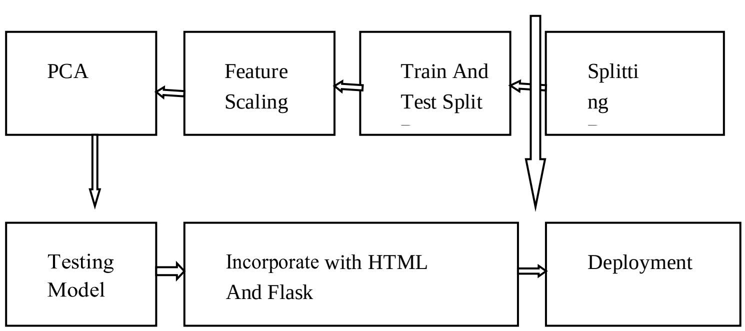
The prediction error is defined as the difference between its actual outcome value and its predicted outcome value. In this study, two metrics were used to compare models: - RMSE and MAE. RMSE (root mean squared error) is calculated. This is computed by taking the differences between the target and the actual algorithm outputs, squaring them and averaging over all classes and internal validation samples.

MAE (mean absolute error/deviation) is calculated as MAE This gives the magnitude of the average absolute error.

3.1Block Diagram



3.2 Software Designing



Jupyter Notebook Environment

SpyderIde

Machine Learning Algorithms

Python(pandas, numpy, matplotlib, seaborn, sklearn)

HTML

Flask

We developed this loan status prediction by using the Python language which is a interpreted and high level programming language and using the Machine Learning algorithms. for coding we used the Jupyter Notebook environment of the Anaconda distributions and the Spyder, it is an integrated scientific programming in the python language.

For creating an user interface for the prediction we used the Flask. It is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions, and a scripting language to create a webpage is HTML by creating the templates to use in the functions of the Flask and HTML.

4.EXPERIMENTALINVESTIGATION

Below is the image of data set and it has totally 737453 data points and 24 attributes. In this blog, we used the first 21 attributes as independent variables and the last two attributes (% iron and % silica concentrate) as target variables.

5. RESULT

Dataset: -

In this analysis, we evaluate the predictive performance of the aforementioned ML models. The values for RMSE, MSE, and R2 for all models are reported in Figure below. Low values for RMSE and MSE, and high values for R2 indicate better model performance, respectively. For simplicity of discussion, RMSE is used as the primary metric of performance. In addition, both the testing and validation performance is reported, which facilitates the discussion on overfitting in the models. These performance measures are plotted as boxplots to illustrate the range and variance of the error.

	r2_score	mse	rmse
% silica concentrate without using % iron concentrate.	0.99417	0.002450	0.049504
% silica concentrate with using % iron concentrate.	0.99391	0.001925	0.043886

6. ADVANTAGES AND DISADVANTAGES

ADVANTAGES

1. Silica is basically impurity in iron ore and by predicting the impurity in ore we can help the engineers in the plant to take measurements in early stages of manufacturing. To help the environment by reducing the amount of ore that goes to tailing as you reduce silica in the ore concentrate.
2. silica fume is a kind of neutral inorganic filler with very stable physical and chemical properties. It does not contain crystalline water, does not participate in the curing reaction, and does not affect the reaction mechanism.
3. good infiltration for various kinds of resin, good adsorption performance, easy to mix, no agglomeration phenomenon.
4. it can increase the thermal conductivity, change the adhesive viscosity and increase the flame retardancy.
5. due to the fine grain size and reasonable distribution of silica fume, it can effectively reduce and eliminate precipitation and stratification.
6. pure silicon powder, low content of impurities, stable physical and chemical properties, so that the curing material has good insulation properties and arc resistance.

DISADVANTAGES

1. dry shrinkage.
2. it is easy to produce temperature cracks.
3. Silica fume requires a high amount of water and needs to be used with a superplasticizer.
4. The price of silica fume is relatively high compared to cement and fly ash.
5. Silica fume will increase the autogenous shrinkage of the cement slurry, and the amount of inclusion will exceed 5%, which may increase the risk of cracking. It is easy to cause cracks in mortar and concrete and need concrete maintenance.

7.CONCLUSION

- This Project presents a simple mathematical model to predict the quality prediction in a mining process from the early time test results. In this study, the silica concentrate characteristic with date is modelled by a Random Forest regression mathematical equation. Early age test data are being used in this case to get reliable values of the 20 seconds silica prediction. Herein, a simple and practical approach has been described for prediction of quality prediction in a mining process and the proposed technique can be used as a reliable tool for assessing the mining process from quite early test results. This will help in making quick decision at site and reduce delay in the execution time of large construction projects.
- To predict the silica(impurity) % in the ore concentrate in a less time we are building a predictive analytics system in that we are applying various machine learning algorithms and find the best accurate model. Here web application will be used to display the prediction. The web application is built by using flask framework and it is integrated with trained ML model.

8. REFERENCES

- Breiman, L., &Schapire, R. (2001). Random Forests. IEEE, 45, 5–32. Retrieved from https://www4.stat.ncsu.edu/~lu/ST7901/reading%20 materials/Breiman2001_Article_Random Forests.pdf.
- https://ieeexplore.ieee.org/abstract/document/8907120 .
- Eduardo Magalhães Oliveira. (2017). Quality Prediction in a Mining Process. Retrieved September 7, 2019, from Kaggle.com website: https://www.kaggle.com/edumagalhaes/quality-prediction-in-a-mining-process
- Li, C., Sun, H., Bai, J., & Li, L. (2010). Innovative methodology for comprehensive utilization of iron ore tailings. Journal of Hazardous Materials