Predictive Models For Estimating Impurities In Iron Ore Using Machine Learning

Submitted by,

Manasi S Pillai

Raimah Rahmath

1.INTRODUCTION

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).

1.1 OVERVIEW

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 Neighbors 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 .

1.2 PURPOSE

The main objective of the project is to predict the impurities that are present in the iron ore .The process of detecting silica impurities in lab or by experiment causes time and is very expensive. The project which includes machine learning helps us to detect the impurities present in the iron ore without any delay or waiting.

2.LITERATURE SURVEY

2.1 Existing problem

ML depends heavily on data, without data, it is impossible for an "AI" to learn. It is the most crucial aspect that makes algorithm training possible. In Machine Learning projects, we need a training data set. It is the actual data set used to train the model for performing various actions.

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	Air flow that goes into
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

2.2 Proposed solution

Iron ores are rocks and minerals from which metallic iron can be economically extracted. Iron is usually found in the form of Magnetite, Hematite, Goethite, Limonite, or Siderite. Usually, Magnetite Iron ore concentrate contains an impurity of 3–7% of silica. Estimation of silica involves a lot of chemical analysis which is time-consuming and involves high operational cost. In order to cut down the operational cost and also to help engineers by predicting at a faster rate, we make use of Machine Learning (ML). So the main goal of this project is to build a Machine Learning model to predict the impurities present in an Iron ore.

3.THEORITICAL 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, multitarget 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 .

3.1 HARDWARE / SOFTWARE DESIGNING

The hardware required for the development of this project is:

Processor	: Intel CoreTM i5-9300H
Processor speed	: 2.4GHz
RAM Size	: 8 GB DDR
System Type	: X64-based processor

SOFTWARE DESIGNING:

The software required for the development of this project is:

Desktop GUI	: Anaconda Navigator

Operating system : Windows 10
Front end : HTML, CSS,JAVASCRIPT
Programming : PYTHON
Cloud Computing Service : IBM Cloud Services

4.EXPERIMENTAL INVESTIGATION

IMPORTING AND READING THE DATASET

Importing the Libraries
First step is usually importing the libraries that will be needed in the program.
Pandas: It is a python library mainly used for data manipulation.
NumPy: This python library is used for numerical analysis.
Matplotlib and Seaborn: Both are the data visualization library used for plotting graph which will help us for understanding the data.
csr_matrix() :A dense matrix stored in a NumPy array can be converted into a sparse matrix using the CSR representation by calling the csr_matrix() function.
Train_test_split: used for splitting data arrays into training data and for testing data.
Pickle: to serialize your machine learning algorithms and save the serialized format to a file.

Reading the Dataset
For this project, we make use of three different datasets (Books_Ratings, Books, Users). We will be selecting the important features from these datasets that will help us in recommending the best results.
The next step is to read the dataset into a data structure that's compatible with pandas. Let's load a .csv data file into pandas. There is a function for it, called **read_csv().**We will need to locate the directory of the CSV file at first (it's more efficient to keep the dataset in the same directory as your program).If the dataset in same directory of your program, you can directly read it, without any path. After the next Steps we made following bellow:

1.Data visualization
2.Collabrative and filtering
3.Creating the Model
4.Test and save the model
5.Buil Python Code
6.Build HTML Code
7.Run the Application
We are the following above sections we did and investigate it.

5.FLOWCHART

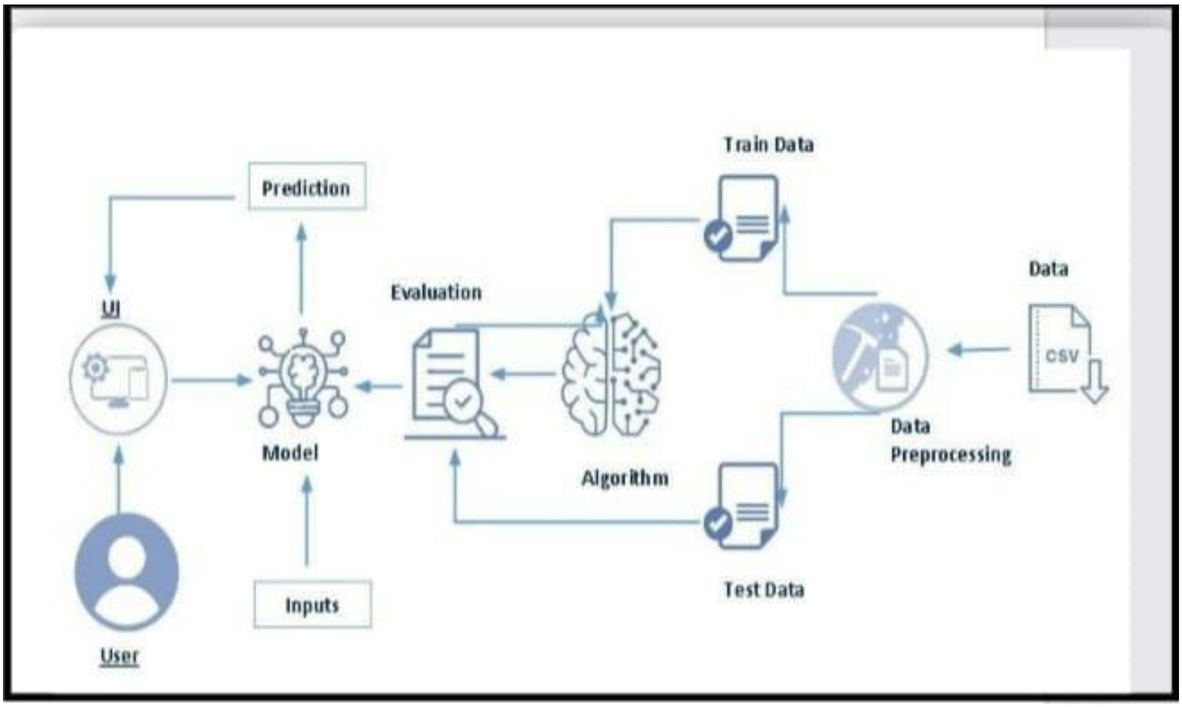

Fig 5.1 Flowchart of the project

Project Flow:

- User interacts with the UI (User Interface) to upload the input features.
- Uploaded features/input is analysed by the model which is integrated.

Once a model analyses the uploaded inputs, the prediction is showcased on the UI.

1. Data Collection.

- Collect the dataset or Create the dataset

2. Data Pre- processing.

- Import the Libraries.
- Importing the dataset.
- Exploratory Data Analysis
- Data Visualization

3. Collaborating Filtering

- Merging datasets
- Creating the Model
- Predicting the results

- Saving our model and dataset

4. **Application Building**

- Create an HTML file
- Build a Python Code

6.RESULT

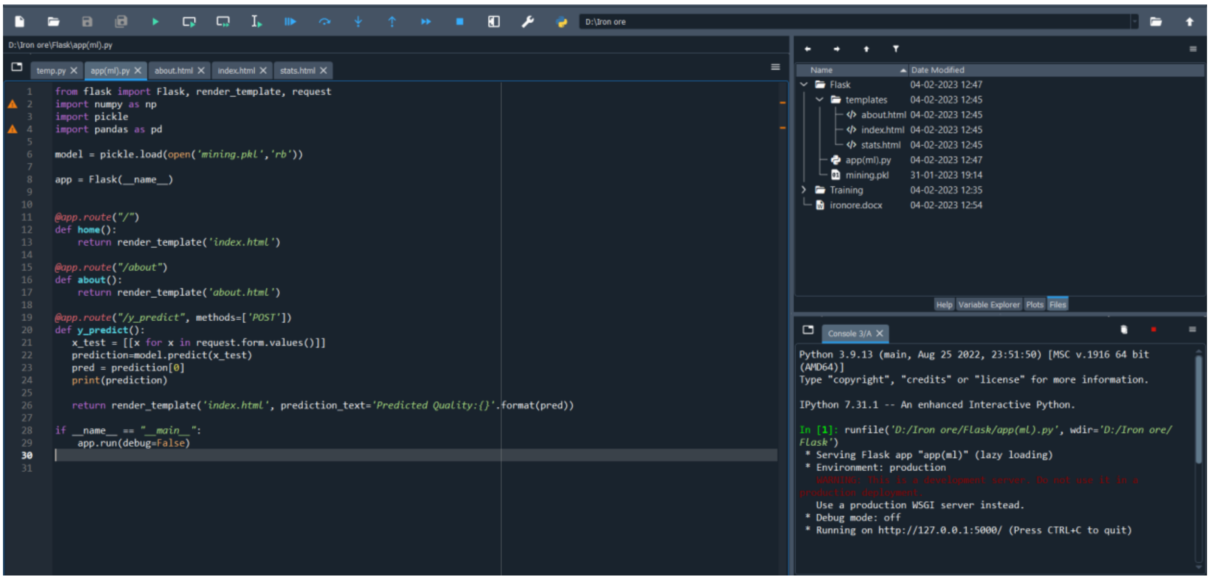

Fig 6.1 Flask code on Spyder

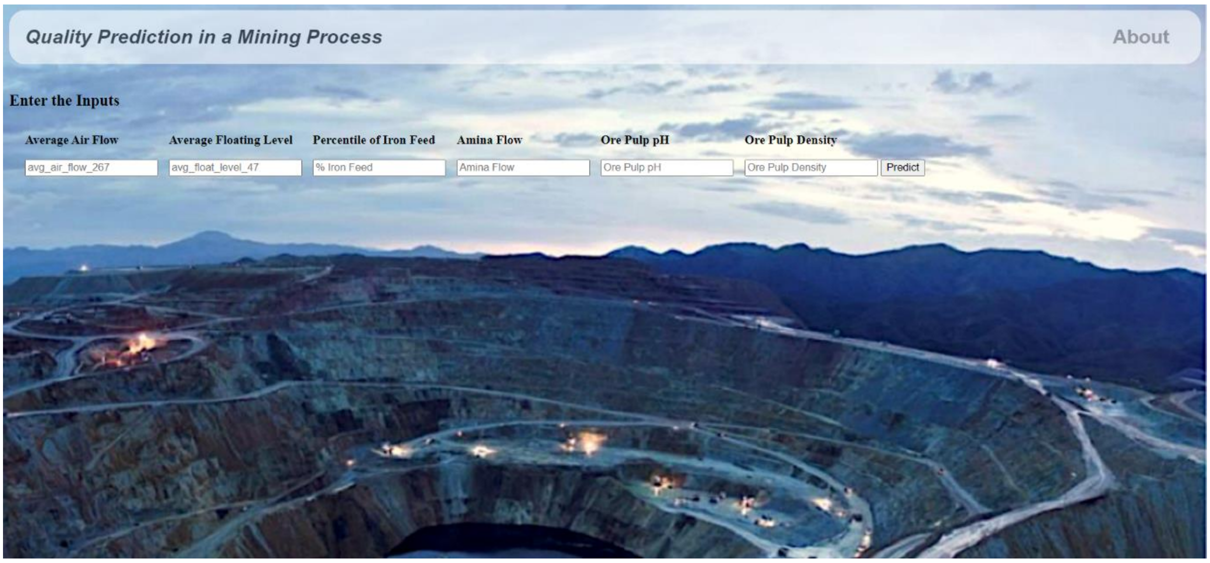

Fig 6.2

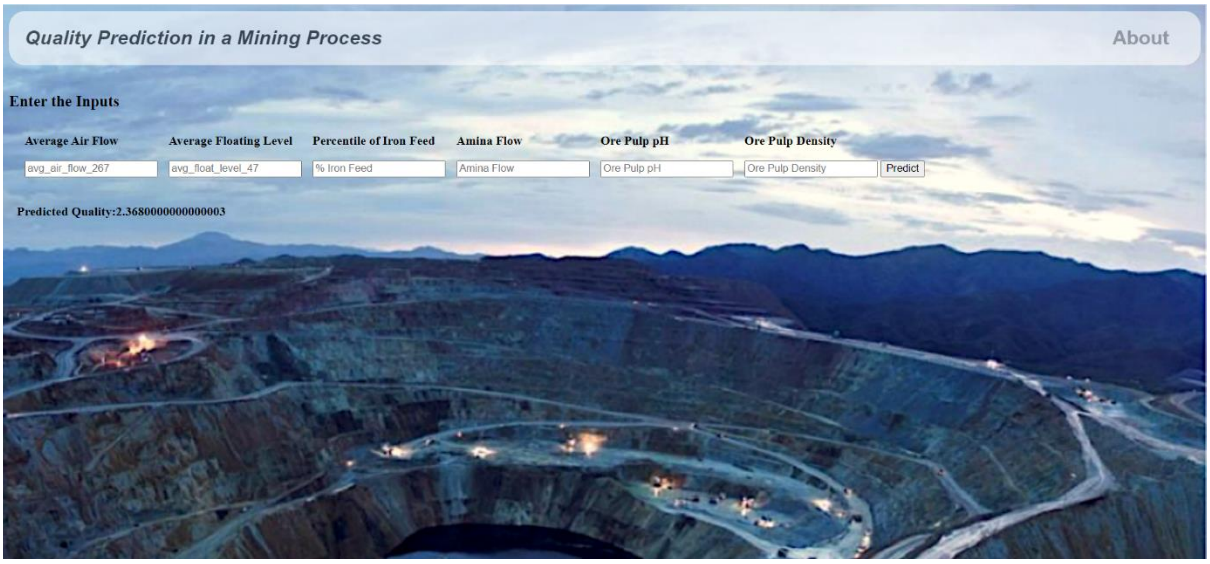

Fig 6.3

7.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 radience

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..

8.APPLICATIONS

- The main application of this project is at the minig field,as we can use this project for detecting the impurities present in the iron ore.
- We can also use this project at the iron industries.

9.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 slica concentrate characteristic with date is modeled 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.

10.FUTURESCOPE

Enhancements that can be made in the future:

- Iron ores are rocks and minerals from which metallic iron can be economically extracted. Iron is usually found in the form of Magnetite, Hematite, Goethite, Limonite, or Siderite.

- Usually, Magnetite Iron ore concentrate contains an impurity of 3–7% of silica. Estimation of silica involves a lot of chemical analysis which is time-consuming and involves high operational cost.
- In order to cut down the operational cost and also to help engineers by predicting at a faster rate, we make use of Machine Learning (ML).
- So the main goal of this project is to build a Machine Learning model to predict the impurities present in an Iron ore.

11.BIBILOGRAPHY

- Hastie, Friedman, and Tibshirani, *The Elements of Statistical Learning*, 2001
- Bishop, *Pattern Recognition and Machine Learning*, 2006
- Ripley, *Pattern Recognition and Neural Networks*, 1996
- Duda, Hart, and Stork, *Pattern Classification*, 2nd Ed., 2002
- Tan, Steinbach, and Kumar, Introduction to Data Mining, Addison-Wesley, 2005.
- Scholkopf and Smola, *Learning with Kernels*, 2002

APPENDIX

A Source Code of Flask:

```
from flask import Flask, render_template, request
import numpy as np
import pickle
import pandas as pd

model = pickle.load(open('mining.pkl','rb'))

app = Flask(__name__)


@app.route("/")
def home():
```

```python
    return render_template('index.html')


@app.route("/about")
def about():
    return render_template('about.html')


@app.route("/y_predict", methods=['POST'])
def y_predict():
   x_test = [[x for x in request.form.values()]]
   prediction=model.predict(x_test)
   pred = prediction[0]
   print(prediction)

   return render_template('index.html', prediction_text='Predicted Quality:{}'.format(pred))


if __name__ == "__main__":
    app.run(debug=False)
```