Machine Learning-based Quality Prediction in the Froth Flotation Process of Mining

Master's Degree Thesis in Microdata Analysis

Author: Eric Kwame Osei
Supervisor: Professor William Song
Examiner: Siril Yella
Subject/main field of study: Microdata Analysis
Course code: MI4002
Credits: 15 ECTS
Date of examination: December 18,2019

Abstract

In the iron ore mining fraternity, in order to achieve the desired quality in the froth flotation processing plant, stakeholders rely on conventional laboratory test technique which usually takes more than two hours to ascertain the two variables of interest. Such a substantial dead time makes it difficult to put the inherent stochastic nature of the plant system in steady-state. Thus, the present study aims to evaluate the feasibility of using machine learning algorithms to predict the percentage of silica concentrate (SiO2) in the froth flotation processing plant in real-time. The predictive model has been constructed using iron ore mining froth flotation system dataset obtain from Kaggle. Different feature selection methods including Random Forest and backward elimination technique were applied to the dataset to extract significant features. The selected features were then used in Multiple Linear Regression, Random Forest and Artificial Neural Network models and the prediction accuracy of all the models have been evaluated and compared with each other. The results show that Artificial Neural Network has the ability to generalize better and predictions were off by 0.38% mean square error (mse) on average, which is significant considering that the SiO2 range from 0.77%-5.53% -(mse 1.1%) . These results have been obtained within real-time processing of 12s in the worst case scenario on an Inter i7 hardware. The experimental results also suggest that reagents variables have the most significant influence in SiO2 prediction and less important variable is the Flotation Column.02.air.Flow. The experiments results have also indicated a promising prospect for both the Multiple Linear Regression and Random Forest models in the field of SiO2 prediction in iron ore mining froth flotation system in general. Meanwhile, this study provides management, metallurgists and operators with a better choice for SiO2 prediction in real-time per the accuracy demand as opposed to the long dead time laboratory test analysis causing incessant loss of iron ore discharged to tailings.

Keywords: Froth Flotation; Machine Learning; Random Forest; Multiple Linear Regression; Artificial Neural Network.

Acknowledgements

It would be impossible to acknowledge all the various people whose positive influence to my life at various stages have shaped me to be the man I am today. I would like to express my profound gratitude to Almighty God for His unfailing love and mercies toward me through His Son Jesus Christ and the Holy Spirit for teaching me all things and humbling me through the word.

I wish to acknowledge my lovely wife Mrs. Priscilla Amma Osei for her enormous support, love, prayers and for standing with me in the years I had spent on this master's program. I could never have been able to accomplish this without you, I love you dearly.

My gratitude goes to my parents, Mr. Lawrence Anang, Mr Samuel Boama-Wiafe and Mrs. Rosina Eshun for their support and prayers.

I am also indebted to my supervisors: Professor William Song, Dr.Qingshang Dengand Dr. Xiangdong Peng for been helpful and patient with me right from the beginning to the end of this research work. I am truly grateful and promise to make you proud when the time comes.

Also not forgetting my good friend Nejib BenSlah Bouzaabia here in Sweden for your support and candid advices throughout my study period.

Last but not least, I wish to thank all lecturers of Dalarna University especially Professor Moudud Alam, Prof. Hasan Fleyeh, Dr.Ilias Thomas, Dr. Reuben Laryea, Murshid Saqlain (PhD) for their support and cooperation in helping me finish my master's. God bless you all.

Table of Contents

Figures and Tables

FIGURES

TABLES

CHAPTER 1. INTRODUCTION

In this section, I introduce the reader to the thesis background. I discuss the research problem in detail. I further describe the objectives, scope and limitation of this project. The section concludes with how the proceeding chapters are organized.

1.1 Background

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].

However, the recent boom in machine learning techniques offers an alternative paradigm for illustrating the relationships between real-time percentage of iron ore or silica concentrate and its superfluous features, thereby necessitating diverse models compared to the conventional laboratory approaches. It is also established that by employing powerful Artificial Intelligence model classes, such as Fuzzy Logic [3], Deep Neural Network [4], and PCA [5] etc. the control and prediction efficiency quality iron recovery are significantly ameliorated.

Generally, froth flotation processing plant has a vital role to play in the portfolio management.

It is by far one of the novel invention and widely used technique employed in conventional mineral processing fraternity over the century [6]. The plant remains one of the gigantic tonnage separation processing mechanisms used in a number of industries processing such as paper mills, and water purification, aside mineral processing [7]. Despite its significant impact in the mining ecosystem, the inherent stochastic nature of the plant processing makes it difficult to comprehend and effectively control [19]. The reasons include but not limited to, inherent complexity and unpredictability of the response of most flotation circuits to upset conditions, unclear expectations of what can be achieved by a control system, unrealistic objectives for control systems and excessive complexity of the actual control strategies [9]. As a result, real-time quality grade recovery and percentage of impurity estimation have been

approached from diverse perspectives and still an open research area in which new models are sought to optimal prediction.

Froth 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” [10].Metal sulfide ore such as Zinc, Phosphate, Cobalt, Copper, and Iron ore etc. are often extracted from gangue, consisting of quartz, feldspar and similar materials. These gangues often encapsulate the ore body which are mined from the earth crust and crushed into a finely particulate form. Usually, the density of the discharged slurry from the mill is approximately between 30% and 70 % solids [17]. The crushed feed is fed to the froth flotation processing plant as the input disturbances.

Depending on the mineral of interest, all individual mineral in the rock body are made to liberate either by natural size reduction or artificial means of grinding. The motivation is to liberate gangue from precious mineral, given a raw material which takes place in the froth flotation processing plant [11]. This vital mineral liberation process is effectuated in a form of slurry, often referred to as flotation. The slurry is mixed with frothers which aid in the formation of the bubbles under agitation [12]. Effective compressed air is usually discharged into the solution which causes turbulence by the rising bubbles. It therefore enables the hydrophobic particles to stick to the air bubbles whilst the hydrophilic particles stay inundated in the pulp or tank. The bubbles are then carried above the chemical treated solution, thereby forming froth on top of the slurry.

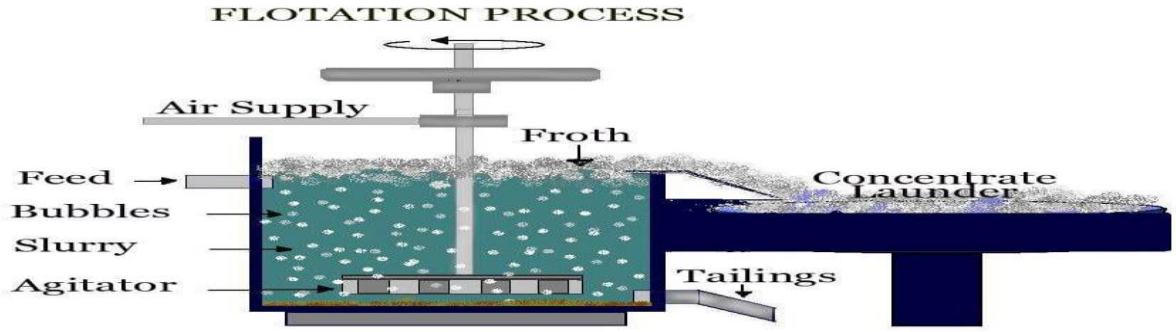


Figure 1. Schematic micro-cell flotation column [13]

Moreover, retention time and agitation are two most significant parameters required to stabilize the froth concentration in addition to the reagents added to the slurry. The volume and the number of flotation cells required for efficient separation process to happen is normally determined by the retention time [17]. On the other hand, the power input and

flotation technique required for optimum flotation condition is also determine by agitation as shown in the schematic microcell flotation column in Figure 1.

However, the froth is collected and dried under controlled sunlight. Most importantly, the percentage of silica concentrate and iron ore concentrate are ascertained at this crucial phase of the plant processes. Parameters such as pulp, operation conditions, appropriate dosage of reagent, mineralogy determined the ore recovery and the fastidious of the froth flotation process [14]. Therefore, enterprises often make concerted efforts to collect empirical data at each processing phase and store in their repository database for consumption.

1.2 Problem Definition

In practice, management, metallurgists and control operators, bank on laboratory test to take ad-hoc decision for corrective action in order to achieve optimal quality of ore recovery in the froth flotation system. Usually, the laboratory analysis takes two or more 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.

Although researchers have studied numerous different techniques on how to find a lasting solution to circumvent this relative efficient way of liberating gangue particles given a measureable input but little efforts have been made to estimate percentage of silica concentrate in real-time. More importantly, the silica concentrate is often quite susceptible to estimation error especially when it would end up in the tailings. Another problem worth mentioning is that company's losses colossal amount of money each fiscal year as chucks of quality ore recovery end up in the tailings. There is therefore the need to derive models for efficient predictions that would incorporate all the stochastic nature of a flotation plant so as to build a reliable platform for decision making in real-time.

1.3 Research Objectives

The study is in the field of micro-data analysis jurisdiction, a normative science which connotes that the objective and results are decision-support oriented. Furthermore, this project seeks to appropriately address the following goals:

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

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

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

1.4 Scope and limitations

In this study, different methods of supervised machine learning predictive models performance would be evaluated using froth flotation processing plant dataset. R statistical capacities for modeling would be applied to the data and therefore no attempt is being made to propound any novel predictive algorithm. However, froth physical characteristics such as bubble shape, speed, size distribution and colors were not considered as model input parameters.

1.5 Overview of the report

The work in this thesis has been respectively structured into five chapters.

Chapter 1 presents the work study and chapter 2 dilates on a review of previous studies conducted regarding froth flotation processing plant. Also chapter 3 provides the methodologies reflections of this thesis whilst chapter 4 presents the results of the models evaluation. Finally, chapter 5 focuses on discussion and concluding remarks.

CHAPTER 2. RELATED WORK

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.

Notably amongst papers which have presented similar work on using Artificial Intelligence in the froth flotation process plant worth mentioning is the work of Dawson [3] in which the researcher explored the use of Fuzzy Logic to control the flotation stochastic process, resulting in increased grade and recoveries. The author further investigated the use of image analyser coupled with Fuzzy logic and reagents control to provide a starting point for expert knowledge to be utilized in order to monitor, evaluate and control grade and recovery. He however averred that, despite the success, there are still opportunities for further enhancements within the control jurisdiction of employing image analysing in the froth concentrate. The research concluded that availability of measurable features would be the basis for machine vision model to improve the froth flotation system problem.

Another author in the paradigm of froth flotation system has explored and discovered the most significant parameters that influence the flotation performance of lead mineral. Seemingly, the new model suggested that grinding time, flotation pH, for comparable collector, solid-in-pulp concentration and the increase of solid-in-pulp concentration have the most significant effect on the ore recovery and selective separation of lead mineral. He concluded that solid-in-pulp concentration was the most important parameter that influences the flotation of lead mineral [20].

In a recent study, advanced imaging systems based on Convolutional Neural Networks were employed to extract features from forth flotation plant, a case study of platinum flotation images at four distinct-grades. The extracted features were trained and compare with traditional texture feature extraction method. It was found that the results were competitive and nearly comparable [21]. Another study examined the used of several Neural Networks architecture models to predict metallurgical performance of the flotation column at Sarcheshmeh copper complex pilot plant. Apparently, 8 parameters were used namely: The chemical reagents dosage, froth height, air, wash water flow rates, gas holdup, Cu grades in the rougher feed, column feed and final concentrate streams concentrate streams were used for the simulation. The authors proposed Artificial Neural Networks (ANN) and Multivariate

Nonlinear Regression (MNLR) as the most robust model for predicting copper ore recovery and grade [4].

However, Bergh el [5] all conducted a pilot and industrial research to establish how the characteristics of flotation processes, the quality of measurements of key variables and general lack of robust models are thwarting the appropriate use of predictive control. The authors proposed a multivariate statistics model such as PCA to explain the relationship between operation data for on-line diagnosis and fault detection. It was found that statistical methods seem to provide a general framework to build models in latent variables related to froth characteristics at short sampling internal. The authors further discovered a supervisory and stabilizing control is a form of sub-optimal expert systems. The only challenge found was the difficulties in replicating a particular solution from one plant to another.

Furthermore, another study proposed artificial neural networks as most robust predictive model to estimate the percentage of silica concentration in iron ore mining plant with the aid of virtual sensor. In his work, 700K observations and 120 parameters including desliming variables which correspond to 10 seconds sampling of the plant process and laboratory variables [22].

The above review highlights the myriad of issues informing and contextualizing the novel approach in the froth flotation system. From this overview, it is apparent that predicting real-time percentage of silica concentrate would be an essential component of the roll-out process.

The chief objective, therefore in this thesis is to focus on using machine learning to predict online percentage of silica concentrate in the froth flotation process plant. More importantly, how to incorporate lagged values of silica concentration into the model.

CHAPTER 3. METHODOLOGY

In this section, I shall describe the methodologies used, the source of data, structure of dataset, data pre-processing, correlation analysis, methods and the statistics used for the predicting error evaluation.

3.1Source 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 [23]. 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.

3.1.1 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 A in Appendix 1.

3.1.2 Data Exploration and Pre-processing

Preliminary analysis was carried out to understand the global landscape of the dataset. By summarizing, 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. This however did not pose any challenge since it was further eliminated from the analysis as non-predictive feature. All distribution of variables was examined. A typical example is Figure 2 which shows a snapshot of the distributions of pair variables in the dataset explored.

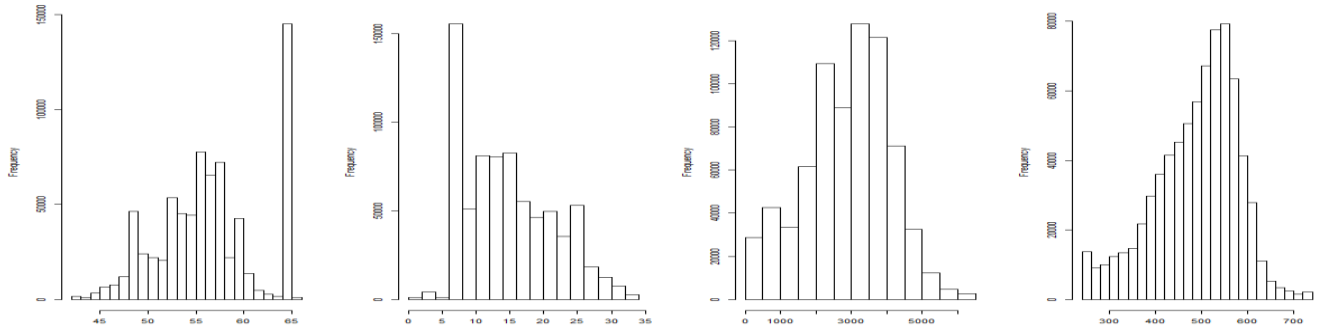


Figure 2.Pair distribution charts for numerical variables-histogram: left-to-right: iron feed, X…silica Feed, Starch Flow and Amina Flow.

As you can see, there is variation in the distribution of each variable as expected. I also established other interesting observations in the outcome variable of interest. The distribution of the silica concentrate in Figure 3 depicts randomness which however buttresses the hypothesis that the froth flotation processing plant is indeed stochastic in nature and difficult to operate.

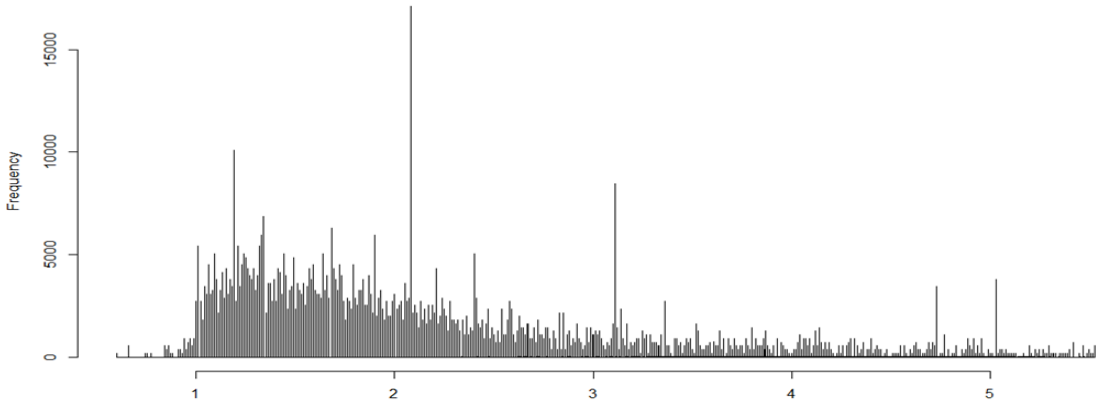


Figure 3. Distribution of Silica concentrate

Moreover, boxplot and threshold adjustment were used to detect outliers and strange numbers. Outliers were observed in Ore-Pulp Flow and Ore-Pulp.pH variables. Excerpt is in Figure 4 which shows the distribution of Ore-Pulp Flow variable's outliers. The outliers were eliminated as can be seen on the same snapshot for further analysis in order to make the analysis trustworthy. A more refined dataset was however, obtained totalling approximately 521188 observations.

Figure 4. Distribution of Ore-Pulp Flow variable to detect outliers.

Moreover, it is worth mentioning the sampling technique, by which the percentage of silica and iron concentrate was measured in this study. Sample of the slurry known as froth as indicated earlier on is drawn at twelve (12) minutes interlude. The laboratory result usually takes at least 2 hours to be available and only pertains to the amalgamation of such partition. This does not include the process sequence of events at that specific instant. However, lag-transformed was generated on the silica concentrate variable to incorporate vital information that might have lost as the process changes during the 2 hours' dead time.

3.1.3 Correlation Analysis

A correlation analysis was employed in this study to examine if explanatory variables share the same linear relationship with the outcome variable in order to detect duplications of variables in the dataset. Amongst other things, highly correlations between variables were observed in the dataset. Table B in Appendix 1 shows the correlation table heatmap. The following clusters of correlated features were found using Pearson correlation coefficient r cutoff of [-0.5, 0.5]. The Pearson correlation coefficient r, takes a range of values between +1 to -1. A value of 0 indicates that there of no relationship between the two variables. A value less than zero indicate a negative relationship and a value greater than zero connotes a positive association: that is as one unit of variable increases, so does the value of the other variable.

The correlation matrix of Table 1 shows a summary of high level of correlation between the following variables:

Table 1. Correlated variables

Flotation.Conlumn.02.Level	Flotation.Conlumn.02.Level
Flotation.Conlumn.03.Level	Flotation.Conlumn.01.Level
Amina. Flow	Ore.Pulp.density
Flotation.Conlumn.06.Level	Flotation.Conlumn.07.Level
Flotation.Conlumn.05.Level	Flotation.Conlumn.04.Level
Flotation.Conlumn.04.Level	Flotation.Conlumn.06.Level
Amina. Flow	Ore.Pulp.density
Flotation.Conlumn.02.Air.Flow	Flotation.Conlumn.01.Air.Flow
Flotation.Conlumn.03.Air.Flow	Flotation.Conlumn.01.Air.Flow
Flotation.Conlumn.06.Air.Flow	Flotation.Conlumn.01.Air.Flow
Flotation.Conlumn.07.Air.Flow	Flotation.Conlumn.01.Air.Flow
Flotation.Conlumn.07.Air.Flow	Flotation.Conlumn.03.Air.Flow
Flotation.Conlumn.07.Air.Flow	Flotation.Conlumn.06.Air.Flow
Flotation.Conlumn.07.Air.Flow	Flotation.Conlumn.04.Air.Flow
Flotation.Conlumn.02.Air.Flow	Flotation.Conlumn.03.Air.Flow

3.1.4 Normalization

For the neural network, the dataset was pre-processed in a format they could ingest. Since the data is already numerical, it was not prudent to vectorise the dataset. But each variable in the dataset was stored on a different scale (for example Ore-pulp pH is typically between 8.753 and 10.808 pH, and Anima. Flow, measured around 500 m3/h. However, each variable was normalized independently in order to put them on the same scale. That is using (Z-score normalization) - [(observed-mean)/ (standard deviation)] formula. This was done by subtracting the mean of each observed variable and dividing by the standard deviation. More also, in order for the model to stay comparable to both the baseline model and other models, the target vector was not normalized.

Z-score = $\mathbf{(x-\mu)/\sigma}$

3.1.5 Variable

In this study, features selection methods have been used to extract significant features to be included in the model for the purpose of ensuing parsimony and manageability. However, Backward Elimination and Random Forest were used to curtail the dimension of the dataset.

Linear Regression. The backward elimination technique arrives at best 18 features with the exception of Flotation.Column.02.Air.Flow, Flotation.Column.07.Air.Flow, and Flotation.Column.03.Air.Flow variables. All variables were used in the initial model and at each step one predictor was removed at the conventional significant p-value of 0.05. The procedure truncated when all variables in the model were statistically significant.

Random Forest. Random Forest algorithm returns important scores for each of the predictors. Figure 9 in Chapter 4 shows the importance of each variable and suggests that 19 variables, except Flotation.Column.02.Air.Flow were selected for the predictive task at hand. I further investigated the significant of a feature by comparing a model with all features verse a model with this feature eliminated from training. It helped to derive most accurate feature significant. The only drawback was that high computation cost encountered.

3.1.6 Partition of the Data

The dataset was partitioned into two parts for training and testing purpose: 70% of the entire dataset for training the selected models using 10-fold cross validation method and 30% for testing purpose. Most importantly, the respective training and validation dataset were randomly sampled to circumvent sampling biasness.

Particularly, 10-fold cross validation is a technique to evaluate a predictive model by partitioning the original dataset into a training set to train the model, and a validation/test dataset to assess it. In this study, the training set was partitioned into 10 equal size subsamples. A single subsample is held as validation set for testing the model, while the remaining 9 subsamples are used to train the data until all the subsamples have been used as validation set.

3.1.7 Evaluation Metrics.

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 as RMSE = $\sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}$. 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 [24].

MAE (mean absolute error/deviation) is calculated as MAE = $\frac{1}{n}\sum_{i=1}^{n} |e_i|$. This gives the magnitude of the average absolute error [24].

3.2 Naïve benchmark

The performance of all models using the different feature set was assessed in a number of ways and compared with each other. However, in order to comprehend how our models performance, a naïve benchmark was established and compared against our models. It served as sanity check. This was calculated as the average outcome percentage of silica concentrate thereby ignoring all predictors' information. In other words, the prediction for a new record is simply the average across the outcome variable (silica concentrate) of the records in the training set. More often than not a good predictive model outperforms the benchmark criterion in terms of predictive accuracy. The RMSE and MAE are respectively given in Table 2 in Chapter 4.

3.3 Machine learning

The steady progress of machine learning has been quite phenomenon over a decade where data is now seen as an important asset to be used judiciously by all companies. Machine learning, in simple term referred to algorithms that learned from data in an iterative manner to identify trends, patterns and correlations. Moreover, it is more applicable especially to datasets where past observations are potent predictor of the future. In this study, among wide variety of existing methods, we employed three supervised machine learning algorithms to a historical dataset observed in the froth floatation plant system.

Supervised machine learning is a process of furnishing the algorithm with observations in which the variable outcome of interest is known beforehand and the algorithm learn from the observed data to make prediction of future values [25].

3.4 Selected Methods

Multiple Linear Regressions. Multiple Linear Regression is one of the most popular and simple machine learning algorithms employed in numerical predictive task. It is mainly used to model a relationship between a numerical outcome variable and a set of explanatory variables. In other words, the model is expected to fit a relationship between a numerical outcome variable and a set of predictors. Y (called the response, target, or dependent variable) and a set of predictors X1, X2, X3…….Xn. (also referred to as independent variables, input variables or covariates). The assumption is that the function in Equation 1 approximates the relationship between the predictors and the outcome variable:

$Y=\beta_o + \beta_1 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots\ldots\ldots + \beta_p \ \ x_p + \epsilon$,---------- (1)

$\beta_o \ldots\ldots, \beta_p \ are \ cofficients$, and ϵ is the noise or unexplained part.

In predictive modeling, data are often used to estimate the coefficient and the unexplained noise that could better lead to prediction of individual records. Regression modeling means not only estimating the coefficients but also choosing which predictor can be included in the model. And the performance of the model is evaluated using a holdout dataset. In R, I use the lm () function to fit the model.

Moreover, Multiple Linear Regression model does exceptionally well with linearly separability phenomena and has a faster computational capability. On the other hand, the limitation is that it does not perform well in non-linearly-separable cases [25].

Random Forest. Random Forest (RF) is an ensemble machine learning proposed by Leo Breiman that combines predictors tree for predictive or classification task based on independent random samples of observations. In order to grow a tree, multiple random samples are drawn with replacement from the training dataset. This connotes that each tree would be grown with its version of the training dataset. Moreover, a subset of the explanatory variables is also randomly selected at each node during the learning process. After training, prediction scores are however determined by averaging the predictions from all individual

regression trees. This helps to attenuate overfitting and variance. Though, there are many methods that perhaps most likely to have impact on model performance. In this study, I focused on two main tuning hyperparameters that have significant effect on the prediction outcome of Random Forest. I employed repeatedcv method to split the dataset into 10 folds cross-validation in [Caret]. For the implementation of the Random Forest in this project, the hyperparameters namely number of trees (*ntree*) and the number of randomly selected features (*mtry*) were experimentally tuned as shown in Figure 7 &8 in Chapter 4.

- **mtry**- Number of variable is randomly collected to be sampled at each split time.
- **ntree**- Number of trees will grow after each time split

For tuning the hyperparameters, *ntree* and *mtry* of the Random Forest, I used the function "randomForest" available within the R-Package. For tuning *ntree*, I tried values (100, 200, 300, 500, and 1000) as shown in Figure 8.

The advantage of using this algorithm is that it does not over fit the data and does exceptionally well with non-linear relationship task. One challenge worth mentioning here is that the algorithm is computationally expensive [26].

Artificial Neural Networks.Artificial Neural Networks are biologically motivated computer algorithms that mimic the way the human brain processes information. This can be comprehended as a random weighted sum of the input parameters employed in an activation function to compute an output as shown in Figure 5. It is a supervised and powerful method for modeling complex, linear and non-linear relationships between a set of predictors and one or more outcome variables. Artificial Neural Networks are often employed to predict outcome variables, provided they have been previously trained on. The network is adjusted using back propagation technique, based on the comparison between the output and the target variable, until the network output matches the target. A weakness of this algorithm is that it can easily overfit the data, causing the error rate on validation data to be too large. For this prediction task, I choose to use a Neural Networks Architecture implemented in the R library Keras [10], with the backend TensorFlow. The choice of employing Keras is based on the fact that it enables fast experimentation and has user- friendly API which makes it easy to quickly prototype deep learning models [18].

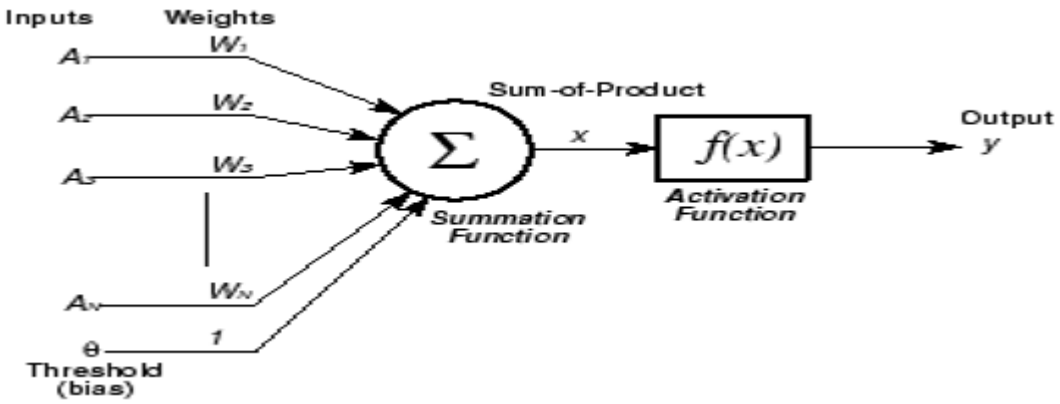


Figure 5. More general neural model [27].

Optimizing the Artificial Neural Network

In this study, different hyperparameters for the implementations are trained and tested with a subset of the iron ore froth flotation processing plant dataset. This aided to find the optimal parameters for the network architecture in order to solve the predictive problem at hand. However, the dataset was divided into three chronologically order set, namely training dataset-validation dataset using 10 fold-cross validation and test dataset.

Loss Function

This is the most important parameter to measure the performance of the network during training. In order to measure the network performance, two different network architectures metrics scores were observed. Namely, the loss function which tells whether the network is estimating the data in appreciable manner and the loss function for optimizing a regression task is mean squared error. It is obtained by taking the differences between the target and the actual algorithm output and squaring them. However, fast convergence rate coupled with the potential to achieve the least error score is often the most significant characteristics in Neural Network training. In contrast, the mean square error metric was also observed in parallel with the loss function as the network was trained.

Network Architecture

In this study, the network architecture was designed in manner tailored to solve the problem at hand. A typical example is a network with few neurons in the hidden layer perhaps has the capacity to generalize from the trained data. On the other hand, sophisticated network with relative much layers and neurons would be subjected to computational cost and overfitting and thereby failing to generalize unseen data. Therefore, considerably quality time was invested to circumvent the challenge of finding the optimal network architecture capable to solve the predictive problem.

Input and Output Layers. The input layer usually has the same number of nodes or units as the input features of the dataset. For example, if the number of input features is 10, then the input layer would be furnished with 10 units. On the other hand, in this experiment, the outcome variable was a numerical value. Hence the output layer was provided with only one unit.

Hidden Neurons. A single layer network with different neurons was used with the default optimizer RMSProp and hyperbolic tangent activation function. As shown in Table 6, in Chapter 4, different neurons were tested to obtain the optimal neurons for further optimization performances.

Hidden Layers. An experiment was carried out with the aim of seeking for an adequate number of hidden layers for the intended network topology. Four different layer namely, 1,2,3,4 were experimented with 100 neurons each in the hidden layer, default RMSProp optimizer and hyperbolic tangent activation function. The simulation results are shown in Table 7 in Chapter 4.

Optimization Algorithm. I employed three optimizer algorithms such as RMSProp, Adam and Stochastic Gradient Descent to establish which amongst them has the ability to converge relatively fast and considerably avoiding be stuck in local optima. The aforementioned optimizers were simulated with hyperbolic tangent activation function in each layer with 500 training iterations as can be seen in Table 9 in Chapter 4.

Activation Functions. Three different activation functions were attempted in order to choose the best one for the final network architecture. The generic activation function such as sigmoid, hyperbolic tangent and ReLU were used in the experiment. Table 10 shows the result of loss scores for each activation function.

Overfitting

In Neural Network modeling, it's often common to see from the training and validation curves that the model is overfitting. That is the training and validation loss begin to diverge considerably after a few iterations or epochs. In this study, a regularization technique, drop out was used to attenuate overfitting. It is the where the input units of the layer are randomly zero in order to break the happenstance correlations in the training dataset that the layer are exposed to. However, I used three-layered network with 500 neurons each in each hidden layer with hyperbolic tangent activation function, RMSProp optimizer and the input features for 2000 epoch training. Yarin Gal [16] proposed the appropriate way to apply dropout regularization parameter and as such this was done in three variations. 30% dropout in the first hidden layer, 30% dropout in the second layer and 10 % in the third hidden layer.

3.5 Sensitivity Analysis

In this study, a Deterministic Sensitivity Analysis (DSA) method was used to examine the sensitivity of the results from the proposed model analysis to variations in a specific input feature or set of features. For example, in the mathematical equation defining a simple machine learning regression.

$Y=\beta_o + \beta_1 x + \epsilon$,---------- (2)

The slope parameter in Equation 2 can be directly interpreted as sensitivity. That is a unit change in x is associated with a unit change in the outcome variable Y. Nonlinear models as opposed to linear model parameters cannot be interpreted in his manner. Therefore, a simple deterministic sensitivity analysis approach has been proposed as an indirect means of analysing causal relationship between model outputs and inputs. Moreover, the silica concentrate was assessed in order to test the robustness of the model results in the presence of uncertainty [15].

A deterministic function, SensitivityAnalysCaret ( ) was generated which was implemented in caret package for the spider plot in Chapter 4.

CHAPTER 4. RESULTS

This section presents various experiment scenarios each consisting of different features and hyperparameters evaluation. The results of the main task of this study, finding the optimal parameters for predicting percentage of silica concentrate in iron ore froth flotation processing plant. In each of the scenarios, different dimension of dataset is considered per the earlier dimension technique employed. Consequently, the two evaluation metrics scores are obtained where the lowest metric errors based on the test set indicates superiority in terms of predictive power. For this, I consider a naïve benchmark which provides a healthy sanity check for comparison purposes.

Table 2. Naïve Baseline Model Errors

MAE	RMSE
0.9106	1.1173

4.1. Multiple Linear Regression

The result of the Multiple Linear Regression model experimentusing different feature set is presented in Table 3.

Table 3. Multiple Linear Regression Model Errors

Dataset	Variable selection method	MAE	RMSE
Data1	Variable Eliminated via Backward elimination	0.8121	1.0315
Data2	Random Forest	0.8130	1.0323

It is realized from Table 3 that the error scores for Data1 model is significantly lower than Data2 dataset model. On that basis, the results of the regression analysis are displayed in Table 4.

The research question of determining which of the froth flotation plant process features are statistically significant in predicting the percentage of silica concentrate was answered by obtaining the regression weights and p-values for these weights. In the Table 4, the estimated regression weights, standard errors and p-values for all the features are given.

Table 4. Regression weights

Predictors	*Estimates*	*CI*	*p*
X..Iron.Feed	-0.00	-0.01 – 0.00	**<0.003**
X..Silica.Feed	0.01	0.00 – 0.01	**<0.001**
Starch.Flow	-0.00	-0.00 – -0.00	**<0.001**
Amina.Flow	0.00	0.00 – 0.00	**<0.001**
Ore.Pulp.Flow	-0.00	-0.00 – -0.00	**<0.001**
Ore.Pulp.pH	-0.19	-0.20 – -0.18	**<0.001**
Ore.Pulp.Density	-1.34	-1.41 – -1.27	**<0.001**
Flotation.Column.01.Air.Flow	-0.01	-0.01 – -0.01	**<0.001**
Flotation.Column.02.Air.Flow	0.00	0.00 – 0.00	**0.008**
Flotation.Column.03.Air.Flow	-0.01	-0.01 – -0.00	**<0.001**
Flotation.Column.04.Air.Flow	0.08	0.08 – 0.08	**<0.001**
Flotation.Column.05.Air.Flow	0.01	0.01 – 0.02	**<0.001**
Flotation.Column.06.Air.Flow	-0.00	-0.00 – -0.00	**<0.001**
Flotation.Column.07.Air.Flow	-0.00	-0.00 – 0.00	0.663
Flotation.Column.01.Level	-0.00	-0.00 – -0.00	**<0.001**
Flotation.Column.02.Level	-0.00	-0.00 – -0.00	**0.046**
Flotation.Column.03.Level	0.00	0.00 – 0.00	**<0.001**
Flotation.Column.04.Level	-0.00	-0.00 – -0.00	**<0.001**
Flotation.Column.05.Level	-0.00	-0.00 – -0.00	**<0.001**
Flotation.Column.06.Level	-0.00	-0.00 – -0.00	**<0.001**
Flotation.Column.07.Level	-0.00	-0.00 – 0.00	0.490
Observations	364894		
R^2 / R^2 adjusted	0.149 / 0.149		

Flotation.Column.07.Air.Flow was not statistically significant at p-value of 0.05 in predicting silica concentrate (β=-0.0, p=0.663), so were Flotation.Column.02.Air.Flow (β=-0.0,

p=0.046) and Flotation.Column.07.Level (β=-0.0, p=0.490). However,...Iron.Feed (β= -0.00, p=0.003),so were Flotation.Column.01.Air.Flow (β=-0.01, p=0.001) and Flotation.Column.01.Level (β=-0.00, p=0.001), Flotation.Column.02.Air.Flow (β=0.00, p=0.008),so were Flotation.Column.03.Air.Flow (β=-0.01, p=0.001) and Flotation.Column.04.Level(β=0.08,p=0.001), Flotation.Column.05.Level(β=0.01, p=0.001), so were Flotation.Column.06.Air.Flow (β=-0.00, p=0.001) and so were Flotation.Column.03.Level (β=0.00, p=0.001) and Flotation.Column.07.Level (β=0.04, p=0.45), so were Flotation.Column.06.Air.Flow (β=-0.00, p=0.001) and X..Silica.Feed (β=0.01, p=0.001), Starch.Flow (β=-0.00, p=0.001), so were Ore.Pulp.Flow, Ore.Pulp.pH were all statistically significant features to predict real time percentage of silica concentrate in the froth flotation system.

Moreover, a significant regression equation was found (F (18, 364875) = 1.031, P< 2.2e-16) with an r-square of 0.1486. Anima flow was the most significant variable as by its estimated standardised weight and the least significant was Flotation Column.07. Level. Figure 6 shows the relative importance of each variable in the Multiple Linear Regression model. Collectively, the model suggested that out of 22 features, 18 were statistical significant at p-value of 0.05 in predicting the percentage of silica concentrate. The R-square is 14.86 % and considerably lower. This means that 14.86% of the variance in the silica concentration can only be explained by the selected predictor which is relatively low to make prediction.

Notwithstanding, the multiple linear regression model significantly outperformed the naïve benchmark model as shown in Table 2.

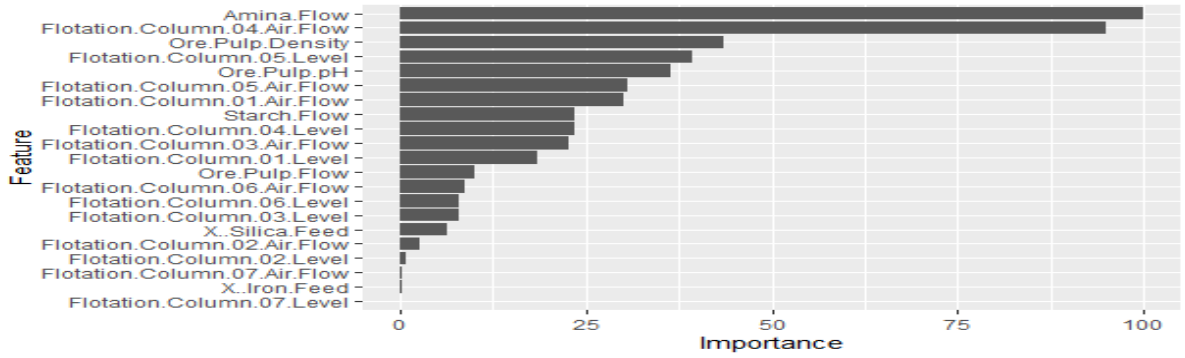


Figure 6. Relative importance of variables

As we can see in Figure 6, the first five most important variables for multiple linear regression predictive model are Anima Flow, Flotation.Column.04.Air.Flow, Ore.Pulp.Density, and Flotation.Column.05. Level and Ore.Pulp.pH, and least are Flotation.Column.07. Level variable.

4.2 Random Forest

In this section, the result of the Random Forest model applied to data1 set is described. I further visualized and analyzed the obtained predictive model and investigated the tuning hyperparameters in order to get best performance from them. The importance of each feature scores is explored as such.

Mtry

The number of predictors or features which is randomly collected to be sampled at each split epoch is described. 15 randomly scores of mtry at each time tuning was generated using a Package caret in R [8].

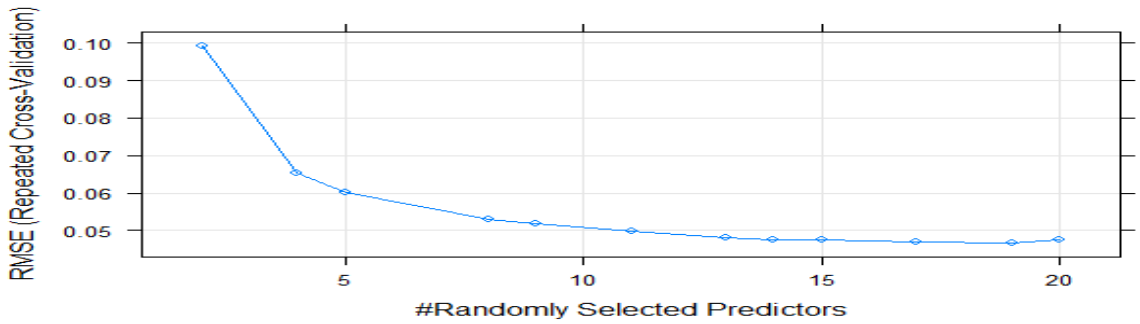


Figure 7.Tune Random Forest Parameters Using Random Search

RMSE was used to select the optimal model using the smallest value. It can be seen that 19 mtry has the lowest scores. Therefore, the mtry hyperparameter search algorithm suggested that 19 random selected predictors are the optimal value for the random forest model.

Ntree

A careful investigation into the number of trees to create a forest as an important hyper parameter in Random Forest was carried out. An objective function for each forest was estimated in order to find the optimal number of trees to build our forest. Figure 8 establishes

various scores of the number of trees with different behavior of the Random Forest method used.

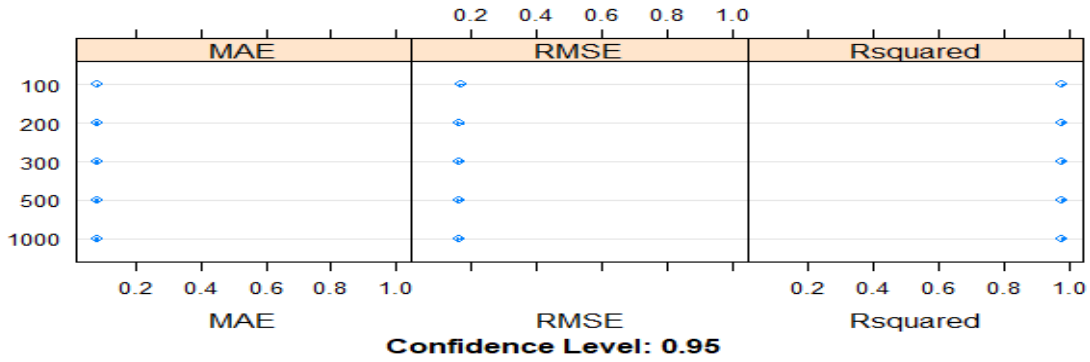


Figure 8. Manually Tune of Random Forest Parameters

As you can see the lowest errors score for ntree was perhaps 500 with a mean absolute error and root mean squared error of 0.0823 and 0.1644 respectively. The results suggest that an optimal score for ntree is between 300 and 500. It is worth noting that, I held mtry =19 per the random search algorithm results in Figure 6 and repeated the experiment with combination of ntree and mtry in order to establish any interactive effects. A Random Forest of 500 trees achieved % variance explained =96.22% during the training phase and lowest error scores of test set.

Variable of Importance

The fundamental features could lead to satisfactory accurate prediction and produce a steady strategy outperformance for this predictive task. It is, therefore, prudent to analyse the feature relevance in this dataset. Figure 9 shows the feature weight distribution for the 21 proprietarily designed factors, where the variable importance is determined by computing the relative influence of each variable.

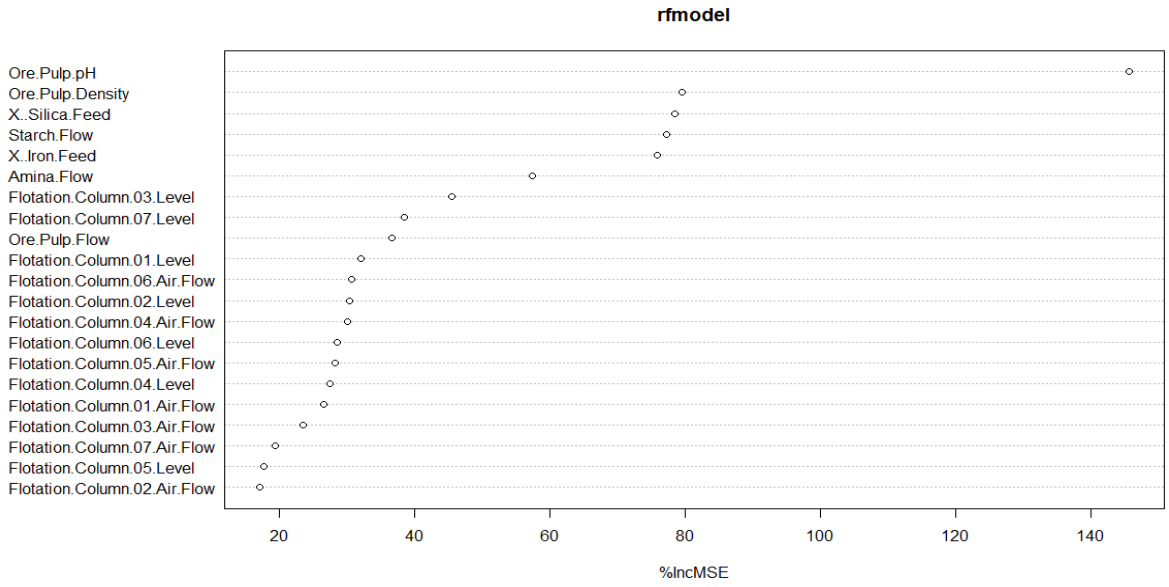


Figure 9. Variable Importance from Random Forest

We can see that Ore.Pulp.ph has the highest scores, with Ore.Pulp.Density, X..Silica. Feed, Starch. Flow being the second, third and fourth respectively. On the other hand, importance scores for the other variables like Flotation.Column.02.Air.Flowand Flotation.Column.05.evel were considerably lower.

However, the Random Forest was applied to both dataset and the results is shown in Table 5 below.

Table 5. Random Forest Model Errors

Dataset	MAE	RMSE
Data1	0.4938	0.9081
Data2	0.5141	0.9093

4.3 Artificial Neural Networks

In this section, the anatomy and architecture of the Artificial Neural Networks are explicitly evaluated.

4.3.1 Hidden Neurons

This section presents the experiment results with different number of hidden neurons for single layer network architecture.

Table 6. Different numbers of neurons for a single layered network

Neurons	MSE	MAE	RMSE
20	1.0006	0.76089	1.0003
30	0.9338	0.7339	0.9663
100	0.8549	0.6896	0.9246
200	0.8525	0.6745	0.9238
500	0.8684	0.6961	0.9300

The hyperbolic tangent (tanh) is used as activation function and trained using the default RMSProp optimizer until no improvement on loss is made for 20 epochs.

As you can see from Table 6, the optimal number of hidden neurons for the single layer in this predictive task is when the neurons exceed 200. Nevertheless, there seems to be diminishing return when the number of neurons moved from 200 to 500. This indicates that a layer units of 100 is relative adequate for this Artificial Neural Networks architecture as further adjustment would be explored. The tradeoff between computational cost and predictive performance suggest that 100 neurons is better as opposed to 200 in order to attenuate the unrealistic computational complexity that networks are often exposed to.

4. 3.2 Hidden Layer

The results of different number of hidden layer (s) of the network architecture are presented.

Table 7. Different number of hidden layers with 100 neurons each

LAYER	MSE	MAE	RMSE
1	0.8549	0.6896	0.9246
2	0.7517	0.6216	0.8670
3	0.6979	0.4902	0.4902
4	0.6558	0.5671	0.8099

The hyperbolic tangent (tanh) is used as activation function and trained using the default RMSProp optimizer until no improvement on loss is made for 20 epochs.

It is realized from Table 7, that there is slightly significant improvement in term of the network learning from the data when the number of hidden layer goes beyond 1. The evaluation metrics scores however, suggested that the benefit seems to diminish with hidden layer is above 3.Moreover, comparing Table 7, metrics scores with baseline model, Random

Forest model and Multiple Linear Regression model, you can observe that the network with 3 hidden layers outperformed these three models.

Hence, it is strongly believed that the network architecture with three hidden layers is considerably sufficient to solve the silica concentrate predictive task at hand.

It is also worth noting that, I once again compared the network with three hidden layer to the respectively number of neurons, 100,200 and 500 above. This aided the justification of any additional neurons that were added to optimise the network performance.

Table 8. Different number of neurons.

NEURONS	LOSS	MAE	RMSE
100	0.6098	0.5135	0.7819
200	0.5534	0.4899	0.7439
500	0.5361	0.4929	0.7322

The network with three hidden layers employing hyperbolic tangent (tanh) activation function and RMSProp optimizer until no improvement on loss is made for 10 epochs.

As you can see, a significant decrease is exhibited when the number of neurons in the hidden layers increased as shown in Table 8. The evaluation metrics scores suggest that 500 neurons each of the three hidden layer is adequate for the final network architecture.

4.3.3 Optimization

In this section, the experiment results for choosing the best optimizer for the artificial neural network training is described.

Table 9. Optimizer parameter.

OPTIMIZER	LOSS	MAE	RMSE
RMSPROP	0.4190	0.3838	0.6473
ADAM	0.4852	0.4517	0.6965
RMSPROP_SGD	0.5662	0.4334	0.7525

Different optimizers are employed to ascertain the neural network performance with 3 hidden layers with 500 neurons each, hyperbolic tangent (tanh) as activation function for 500 training epochs

It is realized that the optimizer from Table 9, suggests the default RMSProp optimizer as the best choice for the optimization algorithm.

4.3.4 Activation Function

In this section, the results of the experiment with different activation functions is presented for the purpose of choosing the optimal activation function of the network architecture.

Table 10. An artificial neural network with three hidden layer of 500 neurons each.

Activation Function	MSE	MAE	RMSE
Tanh	0.5619	0.5180	0.7496
Relu	0.5891	0.5130	0.7675
Sigmoid	0.7034	0.6128	0.8387

The neural net is trained using RMSProp optimizer until MSE has been achieved for 10 training epoch.

We can observe from Table 10 that sigmoid activation function convergence rate is relative low as opposed to both tanh and relu activation function. Tanh activation function seems to have better convergent rate scores than relu activation function. Therefore, tanh activation function was chosen for all three hidden layers for further optimization.

4.3. 5 Overfitting

1. **Regularization**-dropout Regularization technique was also employed for the purpose of attenuating or combating overfitting and achieving a generalization (low test error) as described in Chapter 3.

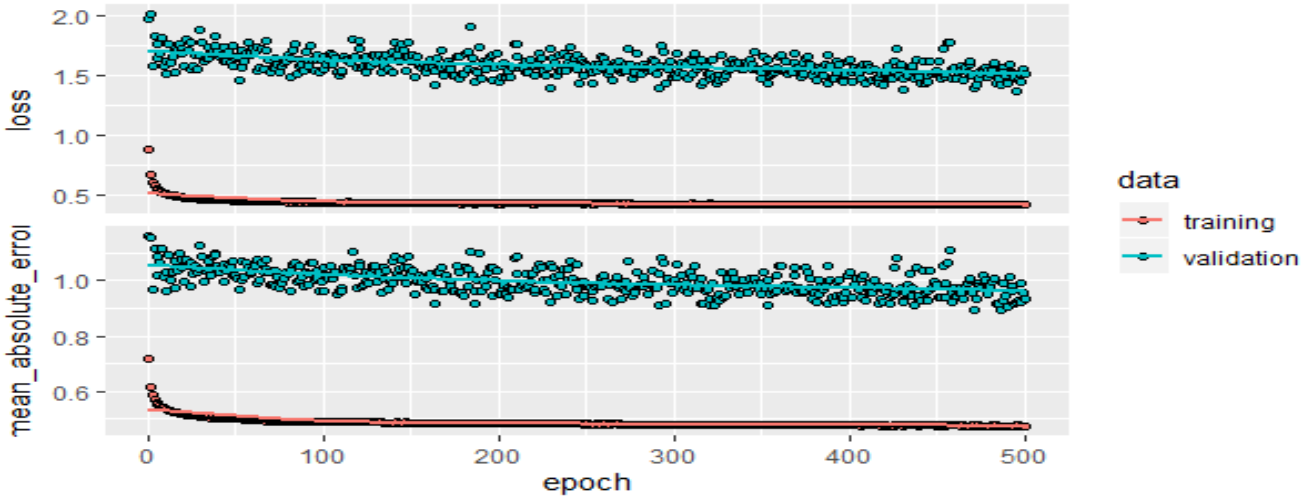


Figure 10. Training and validation metrics.

As you can see, Figure 10, graph shows little improvement in the model after about 500 epochs. Moreover, the loss curves of both training and validation sets are considerably wide which indicate that the Neural Network is overfitting.

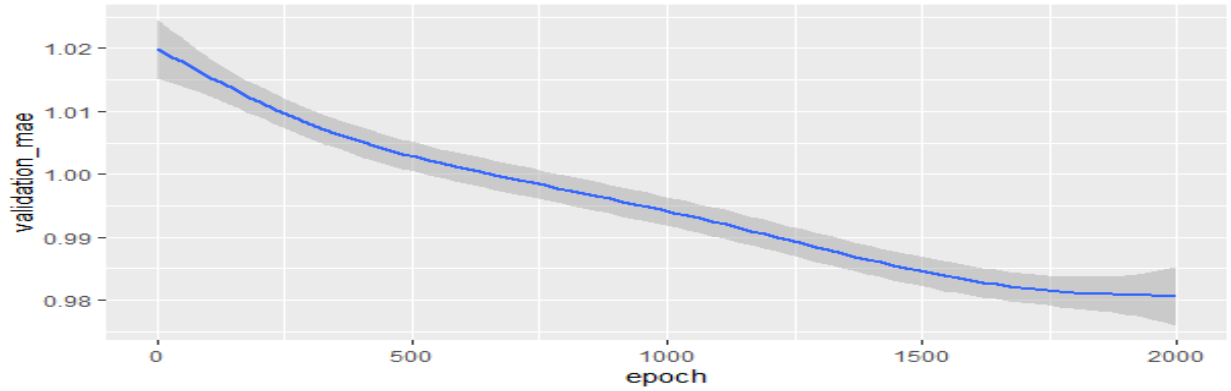


Figure 11. Validation MAE by epoch: smoothed

It is realized from Figure 11, that after 2000 training epochs, the model suggests that I could still tune the hyperparameters to optimise the network results.

4.3. 6. The results for both dataset for the neural network is shown in Table 11

Table 11. Neural network Model Errors

Dataset	MAE	RMSE
Data1	0.4101	0.5765
Data2	0.3810	0.5538

4.4 Final Network Implementation Parameters

In this section that final architecture network is chosen for further optimization.

Table 12. Specifications of the final neural network topology.

No.	Parameters	Value
1.	Number of epochs	2000
2.	Activation function	Tanh
3.	Number of neurons in hidden layer	500
4.	Number of Hidden layer	3
5.	Optimizer	RMSProp
6.	Loss function	mse

4.5 Summary of results

In regard to the performance comparison between different methods, Artificial Neural Network and Random Forest give the best performance, showing strong evidence for predicting real- time percentage of silica concentration in froth flotation plant. The strategy performance for Multiple Linear Regression is relatively weak, probably indicating that the relationship between the process variables and silica concentration is not linear.

Table 13. Prediction Evaluation Metric of the Learning Algorithms in a Test Dataset
Prediction Evaluation

MODEL	MAE	RMSE
BASELINE	0.9107	1.1173
MULTIPLE LINEAR REGRESSION	0.8121	1.0315
ARTIFICIAL NEURAL NETWORK	0.3810	0.5538
RANDOM FOREST	0.4938	0.9081

4.6 Sensitivity Analysis

In this section, the impact of uncertainty in each feature on the percentage of silica concentrate is presented. A spider plot, Figure 12, shows a functional response of the percentage of silica concentrate to each parameter on the same scale.

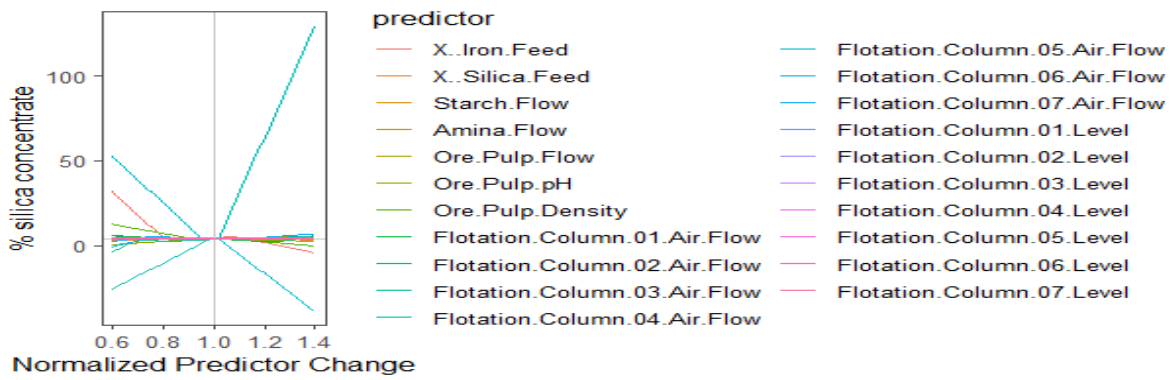


Figure 12. A spider Plot

Illustrates the relationships between model output elucidating percentage of silica concentrations and variations in each of the feature sets, expressed as a percentage deviation from their normalized values. And the relative sensitivity of the performance measure to changes (percentage) in each variable.

4.7 Real-time prediction

This section describes the experiments carried to ascertain the prediction time in real-time.

Table 14. Best-And Worst-And average – Case Execution Times in seconds

Model	Minimum Time	Average Time	Maximum Time
ARTIFICIAL NEURAL NETWORKS	8	9	12

CHAPTER 5. DISCUSSION AND CONCLUSION

5.1 Discussion

The findings from this study suggest that machine learning algorithms have the predictive power to predict percentage of silica concentrate in iron ore froth flotation processing plant in real-time as opposed to 2 hours laboratory analysis. However, after the 3 months' observations of iron ore froth flotation processing plant dataset was analysed, on average, the silica concentrate predictions will be off by 0.38% with a standard deviation of approximately 0.12%, which is significant considering the fact that silica concentrate ranges from 0.77% to 5.53% (mae:1.1% or rmse 0.92%). This result should be interpreted with caution because the silica concentrate variable used in the analysis was lagged 2 hours and could be further explored with diverse residence time

Selecting the significant features is typically arduous, and past literature informed variable selection. In order to judge the robustness and efficacy of my approach, with each different machine learning model, a dimension reduction technique was used to extract significant features for the predictive task. Backward elimination technique was employed of which 18 out of 22 features were statistically significant at a conventional p-value of 0.05. In contrast, 19 features were selected by the Random Forest technique. The experiment results have clearly suggested the usefulness of the feature extraction.

Moreover, when examining which features were the most significant, it could be asserted that the reagents that contributed the most to the construction of the froth flotation processing predictive model were OreAmina. Flow,Ore.Pulp. pH and Ore.Pulp. Density. I observe an interesting contrast in the Multiple Linear Regression model with regard to significant features. For example, the importance of variable graph in Figure 6, experiments results suggest Column.Flotation.07. Level has less influence in the context of silica concentrate prediction. This is quite fascinating, the said cell column happens to be the last phase of the plant system prior to the laboratory analysis. On the other hand, it is prudent to however note that, all the reagents were amongst the most significant variables, which is in accord with a mining perspective

When taking all the results into account in Table 13 from the performance evaluation metrics of the machine learning algorithm, the appraisal was that Artificial Neural Network model had the promising method to generalise unseen test data. Similarly, both multiple linear

regression and random forest models did exhibit some promising predictive capacity, albeit to a moderate percentage.

Although these results are statistically significant, it is possible that a better network topology could have given more definitive results. Also, the reliability and robustness of the model predictions were investigated based on the Artificial Neural Network model to variations in a specific input parameter or set of parameters. Thus, a sensitivity analysis was carried out to analyse the extent of change on the proposed output model if one or more parameters are manually changed. The deterministic sensitivity analysis results in Figure 12 demonstrated a gentle non-linear relationship among silica concentrate and reagents such as ore pulp pH, ore pulp density and iron feed. This means that any change in the outputs of the proposed froth flotation predictive model is not proportional to the change in the aforementioned parameters. In fact, that is one of the reasons why accurate long-time predictions are impossible with most predictive models of current technology, especially in froth flotation plant system. In general, the relationship between the silica concentrate and its associated parameter is nonlinear. On the other hand, the related column cell level and air flow seem to have kept the range very small, given a limited uncertainty associated with those parameters on the model output. It can also be seen that the last column of the column cell level 7 is more sensitivity to the silica concentrate.

Furthermore, the mae values of the Artificial Neural Network model were very similar to the mae values found by Eduardo [22].The evaluation metric value was relatively high as the dimension of the dataset in his study was significantly large.

The real-time prediction experiments have been performed using the Artificial Neural Network model. All experiments were run on a laptop with the following specification: Intel Core i7 CPU with a 2.9 GHz clock rate, 16 GB of RAM and a 64-bit Windows 10 operating system. The experiments were however run on the test dataset and the execution time required was also measured as well.Sys.time function ( ) was used to measure the execution time for each task with a higher degree of accuracy. All experiments were executed 10 times, in order to combat the interference of the operating system routines on the experiments results. The model takes the last 5 minutes observations as input. Table 14 shows the summary results obtained in terms of computation time required to run the experiments. It represents the best-and worst- case and average execution time of the predictive model. The results show an enormous advantage of the Neural Network model over conventional

laboratory test, since its execution time is smaller than the 2 hour laboratory test time. One can also see from Table 14 that the worst execution of time was 12 seconds. This means that real-time silica concentrate could be achieved on the aforementioned laptop specification platform within 12seconds.

However, it is worth noting that each observation in the froth flotation plant can be estimated with respect to silica concentrate as fast as possible. This further connotes that not only the execution time is significant but also the precision of the predictive task. Thus, when both the prediction accuracy and execution time are significant features of an automating the froth flotation plant system, the best option is artificial neural network. This provides in effective predictive in real-time.

More importantly, the accuracy of machine learning model would depend on many factors, such as the number of observations in the training dataset, complexity of the problem domain, number of weights in the model, the error goal and function approximation. The proposed Artificial Neural Network in this study exhibited a superior performance in terms of predictive power and was the simplest to be implemented.

The main limitation in this study was computing time during training. The runtime for training the Artificial Neural Network was quite long prior to converging to achieve optimal minimum. Moreover, finding the optimal topology and hyperparameters were challenging. The whole process is more of an art than science to come up with optimal hyperparameters to solve the problem at hand.

The results of the experiments presented in this study is an indication that Artificial Neural Network model can provide metallurgists, operators and management a better choice according to predictive power and computational cost in practical implementation. This will help them in their decision making in order to achieve quality recovery of iron ore in froth flotation plant, thereby curtailing the incessant discharge of quality iron ore to the tailings.

5.2 Conclusion

In this study, the empirical evidence of building machine learning algorithm using 3 months' iron ore froth flotation processing plant dataset is examined. This study aims to investigate data driven model to estimate percentage of silica concentrate in iron ore froth flotation processing plant in real-time. In addition, the study seeks to find the significant variables that influence silica concentrate in iron ore quality recovery. To elaborate, I examine whether all

features are significant for the predictive task. This is done using feature extraction technique via 10-fold cross validation for the purpose of ensuing parsimony and manageability. Then based on the feature extraction, significant variables dataset is constructed based on 10-fold cross validation performances. Ore. pulp.ph and Amina Flow, and ore. pulp. Density were selected as the most significant reagents parameters that influence the percentage of silica concentrate of iron ore quality recovery in froth flotation plant, while flotation air flow 6 was treated as non-significant variable. In addition to these features, all froth flotation related air flow and cell column level were also involved in the constructing of the Artificial Neural Network model. Finally all the three machine learning algorithms namely Multiple Linear Regression, Random Forest and Artificial Neural network were applied on the feature extraction dataset to train and construct the final model. Consequently, several models were compared and chosen from each other. The experiments were compared with the naïve based benchmark model. The model experiments results have performed consistently better than the naïve benchmark model.

When taking all the results into account from the performance metric evaluation (MAE, RMSE) of the machine learning models (10 fold cross evaluation/test set), the appraisal was that Artificial Neural Networks had the ability to generalize better in estimating real-time percentage of silica concentrate, but the prediction performances of the Random Forest model was found to be better than both baseline model and Multiple Linear Regressions with regard to both error metrics. However, for froth flotation plants with observations beyond this project or different rock formation, it is imperative to used different Neural Networks topology. Conversely, the proposed Artificial Neural Network can have a wider application in select cases when feed parameters with wider range of input values.

A deterministic sensitive analysis was carried out to examine the sensitivity of the results from the proposed model to variations in a specific input parameter or set of parameters.

With regard to the real-time prediction, it has been shown that Artificial Neural Network model can improve the execution time, enabling a real-time silica prediction of 12 seconds in worst case scenario running on a platform based inter i7.

In the light of the results, 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 dataset of more parameters and factors like froth physical characteristics such as speed, size distribution, colour, bubbles shape and size from a different study. On the other hand, plans are in far advance to extend the application of the methodology for different froth flotation processing plants preferably paper mills industry while considering different operating conditions and cell numbers and sizes.

References:

[1] Hume, N. (2014). End of the Iron Age. Financial Time. London, UK.

[2] Li, C., Sun, H., Bai, J., & Li, L. (2010). Innovative methodology for comprehensive utilization of iron ore tailings. Journal of Hazardous Materials, 174(1–3), 71–77.

[3] Dawson, P., & Koorts, R. (2014). Flotation Control Incorporating Fuzzy Logic and Image Analysis: IFAC Proceedings Volumes, 47(3), 352–357.

[4]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.

[5] Bergh, L. G., & Yianatos, J. B. (2011). The long way toward multivariate predictive control of flotation processes. Journal of Process Control, 21(2), 226–234.

[6] Jovanović, I., Miljanović, I., & Jovanović, T. (2015). Soft computing-based modeling of flotation processes – A review. Minerals Engineering, 84, 34–63.

[7] Parada, B. (2012). Study of liquid drainage in flotation foam. London, UK.

[8] Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. Journal of Statistical Software, 28(5).

[9] McKee, D. J. (1991). Automatic flotation control- a review of 20 years of effort. Minerals Engineering, 4(7–11), 653–666.

[10] Pabst, A. (1970). A Dictionary of Mining, Mineral, and Related Terms. Mineralogical Magazine, 37(289), 635–636.

[11] Hodouin, D., Jämsä-Jounela, S.-L., Carvalho, M. ., & Bergh, L. (2001). State of the art and challenges in mineral processing control. Control Engineering Practice, 9(9), 995–1005.

[12] Ghislain Bournival, Seher Ata & Graeme J. Jameson. (2017). Bubble and Froth Stabilizing Agents in Froth Flotation, Mineral Processing and Extractive Metallurgy Review, 38(6), pp. 366-387

[13] Michaud, D. (2017). Froth Flotation Process. Retrieved September 23, 2019, from Mineral Processing & Metallurgy website: https://www.911metallurgist.com/blog/froth-flotation-process

[14] Benzaazoua, M., & Kongolo, M. (2003). Physico-chemical properties of tailing slurries during environmental desulphurization by froth flotation. International Journal of Mineral Processing, 69(1–4), 221–234.

[15] Deterministic Sensitivity Analysis - YHEC - York Health Economics Consortium. (2018). Retrieved December 1, 2019, from YHEC - York Health Economics Consortium website: https://yhec.co.uk/glossary/deterministic-sensitivity-analysis/

[16] Jinzhen Li, Yarin Gal. (2017). PHD Thesis Dropout Inference in Bayesian Neural Networks with Alpha-divergences (PhD Thesis), in ICML.

[17] Metso global website - Metso. (2018). Retrieved December 7, 2019, from Metso website: https://www.metso.com/

[18] Chollet, F., & Allaire, J. J. (2018). Deep learning with R. Shelter Island, Ny: Manning Publications Co.

[19] Bergh, L. G., & Yianatos, J. B. (2011). The long way toward multivariate predictive control of flotation processes. Journal of Process Control, 21(2), 226–234.

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

[21] 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. IFAC-PapersOnLine, 50(2), 13–18.

[22] Eduardo Oliveira, Pablo Drummond, Daniele Kappes, Cássio de Moraes and Mariana Teixeira.(2018). Traditional Machine Learning or Deep Learning, in Conference 22° Seminário de Automação e TI, , São Paulo, Brazil.

[23] 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

[24] Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30, 79–82. https://doi.org/10.3354/cr030079

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

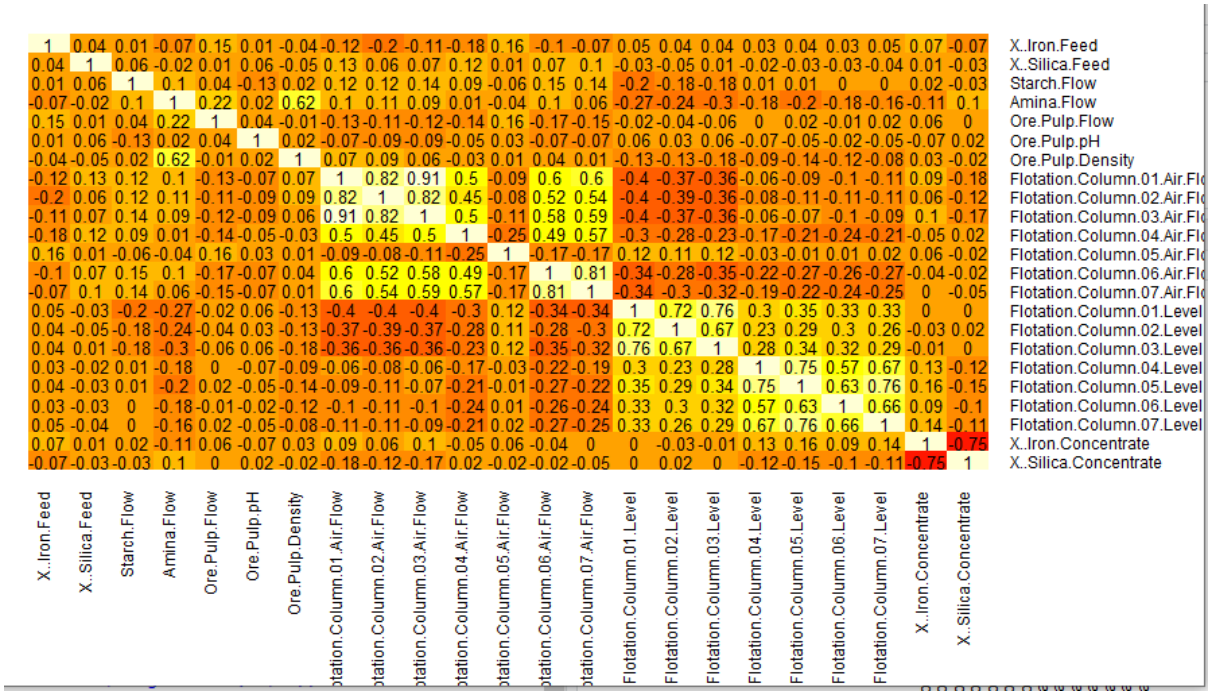
[26] Breiman, L., & Schapire, R. (2001). Random Forests. IEEE, 45, 5–32. Retrieved from https://www4.stat.ncsu.edu/~lu/ST7901/reading%20materials/Breiman2001_Article_RandomForests.pdf

Appendices

TABLE A.

COLUMN	DESCRIPTION OF VARIABLES IN FORTH PLANT PROCESS
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
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 (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)

Table B


X..Iron.Feed
X..Silica.Feed
Starch.Flow
Amina.Flow
Ore.Pulp.Flow
Ore.Pulp.pH
Ore.Pulp.Density
Flotation.Column.01.Air.Flow
Flotation.Column.02.Air.Flow
Flotation.Column.03.Air.Flow
Flotation.Column.04.Air.Flow
Flotation.Column.05.Air.Flow
Flotation.Column.06.Air.Flow
Flotation.Column.07.Air.Flow
Flotation.Column.01.Level
Flotation.Column.02.Level
Flotation.Column.03.Level
Flotation.Column.04.Level
Flotation.Column.05.Level
Flotation.Column.06.Level
Flotation.Column.07.Level
X..Iron.Concentrate
X..Silica.Concentrate