**Industrial Internship Report on**

**”Quality Prediction in a Mining Process”**

**Prepared by**

**Rekha Mathew**

|  |
| --- |
| *Executive Summary* |
| This report provides details of the Industrial Internship provided by Upskill Campus and The IoT Academy in collaboration with Industrial Partner UniConverge Technologies Pvt Ltd (UCT).  This internship was focused on a project/problem statement provided by UCT. We had to finish the project, including the report, in 6 weeks’ time.  My project was ”Quality Prediction in a Mining Process,” aiming to use machine learning to predict the percentage of impurity (silica) in the iron ore concentrate. By predicting the amount of silica in the ore concentrate, engineers can take corrective actions in advance, reducing impurity and helping the environment by reducing the number of ore that goes to tailings.  This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solutions for them. It was an overall great experience to have this internship. |

**TABLE OF CONTENTS**

[1 Preface 3](#_Toc139702806)

[2 Introduction 5](#_Toc139702807)

[2.1 About UniConverge Technologies Pvt Ltd 5](#_Toc139702808)

[2.2 About upskill Campus 9](#_Toc139702809)

[2.3 Objective 11](#_Toc139702810)

[2.4 Reference 11](#_Toc139702811)

[2.5 Glossary 11](#_Toc139702812)

[3 Problem Statement 12](#_Toc139702813)

[4 Existing and Proposed solution 13](#_Toc139702814)

[5 Proposed Design/ Model 14](#_Toc139702815)

[5.1 Model Comparison Diagram 14](#_Toc139702816)

[5.2 Interfaces 16](#_Toc139702818)

[6 Performance Test 18](#_Toc139702819)

[6.1 Test Plan and procedure 18](#_Toc139702820)

[6.2 Performance Outcome 18](#_Toc139702822)

[7 My learnings 19](#_Toc139702823)

[8 Future work scope 20](#_Toc139702824)

# Preface

In the internship with UniConverge Technologies, I was given an industry-grade project as an Data Science and Machine Learning Intern. A time of 6 weeks was provided to complete the project and submit to industrial mentors. Throughout the time of 6 weeks, I worked on a plan provided and reported my weekly progress timely. I also got opportunity to interact and learn from fellow interns and mentors.

As a student of data science and machine learning student, I believe that these internships are very important for career development. Currently, I am interning as a Data Science and Machine Learning intern, which aligns perfectly with my field of study. This internship provides me with hands-on experience, allowing me to apply the theoretical knowledge I have gained in real-world scenarios. It also helps me develop essential skills, build a professional network, and gain industry exposure, all of which are invaluable for my future career in this field.

The project aims to use machine learning to predict the percentage of impurity (silica) in the iron ore concentrate. By predicting the amount of silica in the ore concentrate, engineers can take corrective actions in advance, reducing impurity and helping the environment by reducing the amount of ore that goes to tailings. The project enquires about 3 specific questions:

1. Is it possible to predict % Silica Concentrate every minute?
2. How many steps (hours) ahead can we predict % Silica in Concentrate? This would help engineers to act in predictive and optimized way, mitigatin the % of iron that could have gone to tailings.
3. Is it possible to predict % Silica in Concentrate whitout using % Iron Concentrate column (as they are highly correlated)?

I extend my thanks and gratitude to the UniConverge Technologies, UpSkill Campus and The IoT Academy team for this opportunity. This internship is really helpful for me to gain experience in the relevant industry.

This internship was executed in a very professional and planned manner in which we studied about UniConverge Technologies and explored the project in the first week. From week two to week five were dedicated to solution exploration, analysis, evaluation, performance, and design to the problem statement in the project.  
Here we started with exploratory data analysis, followed by preparing the design and model, then working on implementation. I also evaluated the model to test its quality and general use case.

The following flowchart is the 6 weeks project execution plan of the internship:



This internship has been very helpful for me in terms that I got the chance to understand the mining industry as well as work on a real case problem and worked on it. This opportunity gave me a vivid idea on how machine learning models can be oriented towards problem solving in industries. I also got to connect and network with a lot of professionals in the IoT and Robotics domain.

Thank you all and specially Kaushlendra Sir, Apurv sir, Ankit Sir, and Archana ma’am to be a great guide and beholder to me as an intern at UCT.

I would like to say to my juniors and peers that this internship is definitely going to upskill your career. It is uniquely designed and planned in a manner to provide maximum learning to the interns.

# Introduction

## About UniConverge Technologies Pvt Ltd

A company established in 2013 and working in Digital Transformation domain and providing Industrial solutions with prime focus on sustainability and RoI.

For developing its products and solutions it is leveraging various**Cutting Edge Technologies e.g. Internet of Things (IoT), Cyber Security, Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, Front end**etc.



1. UCT IoT Platform ()

**UCT Insight** is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable “insight” for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

* It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA
* It supports both cloud and on-premises deployments.

It has features to  
• Build Your own dashboard  
• Analytics and Reporting  
• Alert and Notification  
• Integration with third party application(Power BI, SAP, ERP)  
• Rule Engine

1. **Smart Factory Platform (****)**

Factory watch is a platform for smart factory needs.

It provides Users/ Factory

* with a scalable solution for their Production and asset monitoring
* OEE and predictive maintenance solution scaling up to digital twin for your assets.
* to unleased the true potential of the data that their machines are generating and helps to identify the KPIs and also improve them.
* A modular architecture that allows users to choose the service that they what to start and then can scale to more complex solutions as per their demands.

Its unique SaaS model helps users to save time, cost and money.

1.  based Solution

UCT is one of the early adopters of LoRAWAN teschnology and providing solution in Agritech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

1. Predictive Maintenance

UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



## About upskill Campus (USC)

upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process.

USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

<https://www.upskillcampus.com/>

upSkill Campus aiming to upskill 1 million learners in next 5 year



## The IoT Academy

The IoT academy is EdTech Division of UCT that is running long executive certification programs in collaboration with EICT Academy, IITK, IITR and IITG in multiple domains.

## Objectives of this Internship program

The objective for this internship program was to

 ☛ get practical experience of working in the industry.

 ☛ to solve real world problems.

 ☛ to have improved job prospects.

 ☛ to have Improved understanding of our field and its applications.

 ☛ to have Personal growth like better communication and problem solving.

## Reference

[1] Dataset Link: [quality-prediction-in-a-mining-process](https://www.kaggle.com/datasets/edumagalhaes/quality-prediction-in-a-mining-process%09)

[2] Existing solutions: [(Solution 1)](https://www.kaggle.com/code/aydinaktar/concentrate-prediction-from-processing-engineer) [(Solution 2)](https://www.kaggle.com/code/adarshsambare/pca-on-iron-ore)

[3] Froth floatation process in mining industry: [article explaining process with diagram](https://www.911metallurgist.com/blog/froth-flotation-process)

[4]

## Glossary

|  |  |
| --- | --- |
| Terms | Acronym |
| DS-ML | Data Science and Machine Learning |
|  |  |
|  |  |
|  |  |
|  |  |

# Problem Statement

The assigned problem statement in the project is about a froth flotation plant, a process used to concentrate iron ore. This process is very common in a mining plant.

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 a 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. 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 number of ore that goes to tailings as you reduce silica in the ore concentrate).

**Expectation from this project:**  
‘The target is to predict the % of Silica at the end of the process, which is the concentrate of iron ore and its impurity (the % of Silica).’ Although the % of Silica is measured (last column), it is a lab measurement, which means it takes at least one hour for the process engineers to have this value, so it is possible to predict the amount of impurity in the process.

The predictive model was constructed using the iron ore mining froth flotation system dataset 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 was evaluated and compared with each other.

# Existing and Proposed solution

On Keggle, I came across a few existing solutions which proposed multiple linear regression models, random forest model, decision tree, dummy regressor, ridge regression model, etc. The existing solutions were not thorough and detailed in explanation as well as the approach used in some solutions was found to be inappropriate or extrapolatory in nature.

One major limitation was the missing data for a comprehensive study and more accurate prediction. The missing data that should be a minimum for analysis was found to be recovery, feed tons, minerology, feed & tail grades for every column, liberation degree, grinding performance, pumps amperes, pressures etc.   
Another problem in existing solution, the random forest model was applied separately over iron and silica, which would generate good R2 score but this would be fundamentally wrong to do prediction.

In my proposed solution, 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 was evaluated and compared with each other.   
 In addition, I have checked the correlation between airflow and level parameters using heat maps. And it was found that there were high correlations between floatation air flows and floatation column levels. Along with this, I did some feature engineering for percent iron concentrate. Normalization was also done independently for each variable to put them on the same scale.

So overall, the comparison between Linear Regression with regularization term(Lasso), Linear Regression with regularization term(ridge model), and Random Forest Tree model was done, and the Random Forest Tree model performed best on RMSE and R2 score tests.

## Code submission (Github link): https://github.com/rekha-mathew-2023/mining-process-data/blob/main/Mining%20dataset%20project%20(2).ipynb

## Report submission (Github link):

https://github.com/rekha-mathew-2023/mining-process-data/blob/main/Final%20Internship%20report-UCT.docx

# Proposed Model

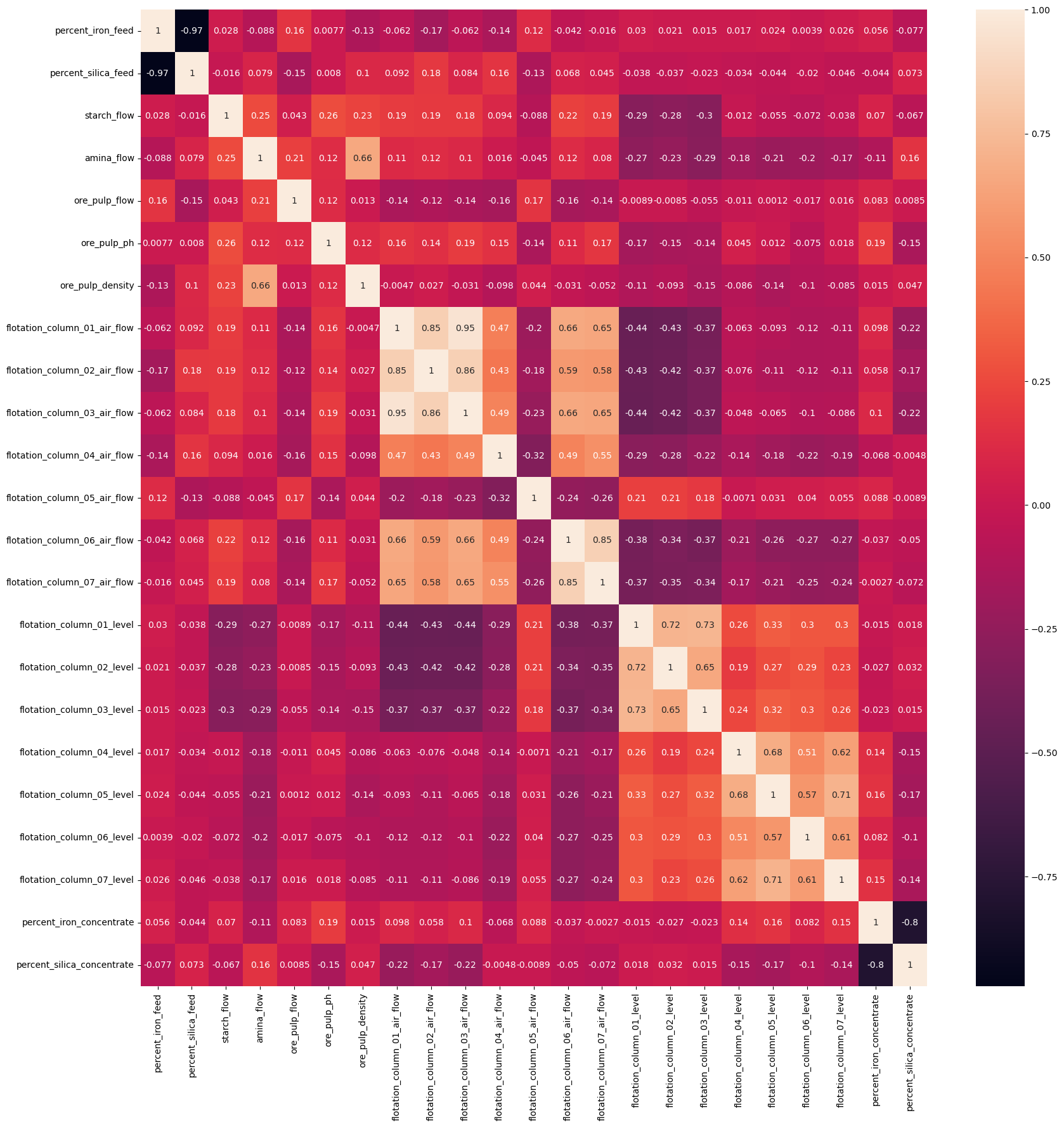
In this study, the empirical evidence of building machine learning algorithm using 3 months’ iron ore froth flotation processing plant dataset is examined and a Random Forest Model was found to be most appropriate prediction model. Along this process, the study seeks to find the significant variables that influence silica concentrate in iron ore quality recovery.

To proceed, I selected only significant features for the predictive task after a thorough study of correlation and feature extraction. 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 random forest model.  
 The scikit-learn library provides additional functionalities for tuning the Random Forest model, such as adjusting the number of trees, maximum depth, and minimum samples required to split a node. These parameters can be optimized to improve the model's performance. Using the Random Forest model for regression allowed me to predict continuous numerical values based on input features.

## Model comparison diagram

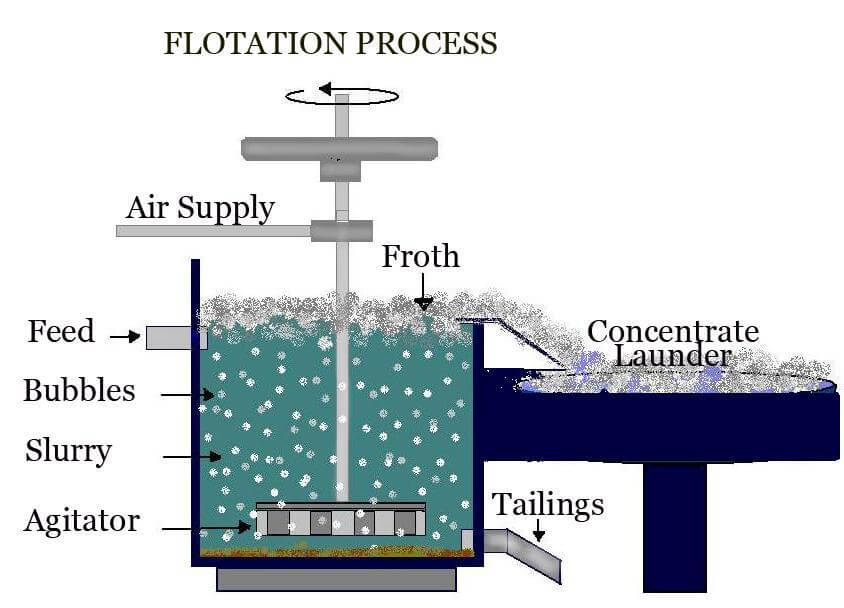


Figure 1: Plot showing comparison of selected models on performance



**Figure [02]: Correlation examination using heatmap**

## Interfaces



**Figure [03]: Schematic micro-cell flotation column**

**5.3 Table about dataset fields**

|  |  |  |
| --- | --- | --- |
| Sr. no. | Field name | Field meaning |
|  | **Date** | Date and Timestamp |
|  | **% 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** | Airflow that goes into the flotation cell is measured in Nm³/h |
|  | **Flotation Column 02 Air Flow** | Airflow that goes into the flotation cell is measured in Nm³/h |
|  | **Flotation Column 03 Air Flow** | Airflow that goes into the flotation cell is measured in Nm³/h |
|  | **Flotation Column 04 Air Flow** | Airflow that goes into the flotation cell is measured in Nm³/h |
|  | **Flotation Column 05 Air Flow** | Airflow that goes into the flotation cell is measured in Nm³/h |
|  | **Flotation Column 06 Air Flow** | Airflow that goes into the flotation cell is measured in Nm³/h |
|  | **Flotation Column 07 Air Flow** | Airflow that goes into the flotation cell is 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) |

# Performance Test

The evaluation metrics to check the prediction error, I used two metrics as RMSE(Root Mean Squared Error) and R Square.

## Test Plan and procedure

RMSE (root mean squared error) is calculated as . 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.

R-squared, also known as the coefficient of determination, is an evaluation metric used to assess the performance of a model. It measures the proportion of the variance in the dependent variable that can be explained by the independent variables. The R-squared value ranges from 0 to 1, where a higher value indicates a better fit of the model to the data. It is calculated as 1 minus the ratio of the sum of squared residuals to the total sum of squares.

The procedure to deploy these tests is using ‘sklearn.metrics’ library and importing their individual program as r2\_square, mean\_squared\_error, etc.

## Performance Outcome

RMSE of test result : 0.18441173179947729

R2 of test result : 0.8540841930721229

Thus the result shows that R2 on test data (R2=0.8539 ) and R2 on CV data (R2=0.8672) is not difference significantly so the model should be just fit.

Comparing performance between Cross-Validation data set and test set. I found that the performance it does not drop significantly, so the model should be used properly in the deployment step.

# My learnings

The internship with UniConverge Technologies and Upskill Campus has helped in a great way that I gained valuable learning and experiences. During the starting period of internship, I got to learn about the IoT industry and how UniConverge Technologies work. The understanding about the products and solutions offered by UCT gave me an exposure to the industrial applications of the technology and use case.  
 Later a project on ‘Quality prediction in mining process’ was assigned to me, which had the application level use case, it being a industrial problem I understood the problem statement in the project. It was followed by the data exploration and processing, which made me learn some new concepts on EDA(Exploratory Data Analysis).

During the project, I laid my hands on various types of machine learning models that can be implemented on the dataset to do prediction.   
In conclusion, I can say that this internship has given me learning on industrial exposure, problem solving, adaptability, professional networking and remote work.

# Future work scope

The dataset provided has the scope of application of more kinds of machine learning models like ANN(Artificial Neural Networks) etc.   
Apart from this the sensitivity analysis can also be done to examine the sensitivity of the results from the proposed model analysis to variations in a specific input feature or set of features. This a Deterministic Sensitivity Analysis(DSA) method can be used.

In case of performance and evaluation tests, some other metrices like p-value, MAE, accuracy, etc. can also be used along with the RMSE and R-squared metrices used in proposed model.

**Use case of this project for UniConverge Technologies:**This project can be relevant to the UniConverge Technologies Pvt Ltd, particularly in the use case of mining industry. IoT sensors can be used to collect data on the quality measures of the iron ore pulp and the process data, which can then be used for machine learning to predict the percentage of silica in the ore concentrate. Wireless communication can be used to transmit the data from the sensors to the machine learning model, allowing for real-time predictions and early information for engineers to take corrective actions.