**Industrial Internship Report on**

**”Quality Predition in Mining”**

**Prepared by**

**[Harshal Pandit]**

|  |
| --- |
| ***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 Perdition in Mining)  This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solution for that. It was an overall great experience to have this internship. |

**TABLE OF CONTENTS**

[1 Preface 3](#_Toc139702806)

[2 Introduction 4](#_Toc139702807)

[2.1 About UniConverge Technologies Pvt Ltd 4](#_Toc139702808)

[2.2 About upskill Campus 8](#_Toc139702809)

[2.3 Objective 9](#_Toc139702810)

[2.4 Reference 9](#_Toc139702811)

[2.5 Glossary 10](#_Toc139702812)

[3 Problem Statement 11](#_Toc139702813)

[4 Existing and Proposed solution 12](#_Toc139702814)

[5 Proposed Design/ Model 13](#_Toc139702815)

[5.1 High Level Diagram (if applicable) 13](#_Toc139702816)

[5.2 Low Level Diagram (if applicable) 13](#_Toc139702817)

[5.3 Interfaces (if applicable) 13](#_Toc139702818)

[6 Performance Test 14](#_Toc139702819)

[6.1 Test Plan/ Test Cases 14](#_Toc139702820)

[6.2 Test Procedure 14](#_Toc139702821)

[6.3 Performance Outcome 14](#_Toc139702822)

[7 My learnings 15](#_Toc139702823)

[8 Future work scope 16](#_Toc139702824)

# Preface

In this preliminary exploration, I delved into the provided industrial dataset from a flotation plant, which plays a pivotal role in mining operations. The primary objective of this exploration was to comprehend the data's structure, characteristics, and underlying patterns, thereby setting the stage for addressing the stated problem. The focus here is on predicting the impurity level, specifically silica content, in the iron ore concentrate. This prediction has practical significance, as it empowers engineers with early insights for proactive decision-making, optimizing operations, and reducing environmental impact.

To facilitate this exploration, I imported and utilized relevant packages, tools, and libraries. By employing the powerful capabilities of Python programming language, along with data manipulation and analysis libraries like Pandas and NumPy, I gained a comprehensive understanding of the dataset. Additionally, I harnessed visualization tools such as Matplotlib and Seaborn to visualize the data's temporal and distributional attributes, facilitating the identification of trends and patterns.

The dataset encompasses various columns ranging from time and date information to quality measures of iron ore pulp, process-related variables, and the final quality measurements from the lab. Moreover, certain variables are sampled every 20 seconds, while others are sampled on an hourly basis. This intricate temporal structure necessitates careful preprocessing and transformation to align the data with the desired prediction frequency.

The next step after this initial exploration involves formulating the problem statement using the Understand, Clarify, and Transform (UCT) framework. This framework involves understanding the business context, clarifying the objectives, and transforming the problem into a well-defined analytical challenge. With the data exploration groundwork laid out and the problem context established, the subsequent phases will involve selecting appropriate machine learning techniques, model development, feature engineering, and rigorous evaluation.

In summary, this preliminary exploration acts as a foundation for the forthcoming stages of data analysis and predictive modeling. The harnessed packages and tools not only facilitate data understanding but also set the stage for informed decision-making throughout the project's lifecycle.

I am Fing this type of internship role where we can work on project which gives me details work orientation in Data Science Field.

Explore real industrial data and help manufacturing plants to be more efficient Experimental Scenario Context It is not always easy to find databases from real world manufacturing plants, specially mining plants. This database comes from one of the most important parts of a mining process: a flotation plant. The main goal is to use this data to predict how much impurity is in the ore concentrate. As this impurity is measured every hour, if we can predict how much silica (impurity) is in the ore concentrate, we can help the engineers, giving them early information to take actions (empowering!). Hence, they will be able to take corrective actions in advance (reduce impurity, if it is the case) and also help the environment (reducing the amount of ore that goes to tailings as you reduce silica in the ore concentrate). Content The first column shows time and date range (from march of 2017 until september of 2017). Some columns were sampled every 20 second. Others were sampled on a hourly base. The second and third columns are quality measures of the iron ore pulp right before it is fed into the flotation plant. Column 4 until column 8 are the most important variables that impact in the ore quality in the end of the process. From column 9 until column 22, we can see process data (level and air flow inside the flotation columns, which also impact in ore quality. The last two columns are the final iron ore pulp quality measurement from the lab. Target is to predict the last column, which is the % of silica in the iron ore concentrate. Expected submission

• Is it possible to predict % Silica Concentrate every minute?

• How many steps (hours) ahead can we predict % Silica in Concentrate? This would help engineers to act in predictive and optimized way, mitigating the % of iron that could have gone to tailings.

• Is it possible to predict % Silica in Concentrate without using % Iron Concentrate column (as they are highly correlated)?

Dataset This dataset is about a flotation plant which is a process used to concentrate the iron ore. This process is very common in a mining plant.

Opportunity given by USC/UCT.

How Program was planned



Your Learnings and overall experience.

Thank to all (Mr. Prashan Sir , Arihant Sir and All my Friends of Batch 2 of IOT Academy ), who have helped you directly or indirectly.

Your message to your juniors and peers.

# 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. https://[www.kaggle.com](http://www.kaggle.com)
2. https://www.google.com
3. https://www.upskillcampus.com/blog.com

# Problem Statement

Explore real industrial data and help manufacturing plants to be more efficient Experimental Scenario Context It is not always easy to find databases from real world manufacturing plants, specially mining plants. This database comes from one of the most important parts of a mining process: a flotation plant. The main goal is to use this data to predict how much impurity is in the ore concentrate. As this impurity is measured every hour, if we can predict how much silica (impurity) is in the ore concentrate, we can help the engineers, giving them early information to take actions (empowering!). Hence, they will be able to take corrective actions in advance (reduce impurity, if it is the case) and also help the environment (reducing the amount of ore that goes to tailings as you reduce silica in the ore concentrate). Content The first column shows time and date range (from march of 2017 until september of 2017). Some columns were sampled every 20 second. Others were sampled on a hourly base. The second and third columns are quality measures of the iron ore pulp right before it is fed into the flotation plant. Column 4 until column 8 are the most important variables that impact in the ore quality in the end of the process. From column 9 until column 22, we can see process data (level and air flow inside the flotation columns, which also impact in ore quality. The last two columns are the final iron ore pulp quality measurement from the lab. Target is to predict the last column, which is the % of silica in the iron ore concentrate. Expected submission

• Is it possible to predict % Silica Concentrate every minute?

• How many steps (hours) ahead can we predict % Silica in Concentrate? This would help engineers to act in predictive and optimized way, mitigating the % of iron that could have gone to tailings.

• Is it possible to predict % Silica in Concentrate without using % Iron Concentrate column (as they are highly correlated)?

Dataset This dataset is about a flotation plant which is a process used to concentrate the iron ore. This process is very common in a mining plant.

# Existing and Proposed solution

I have gone through the SHAP is the new methodology to make data use for modeling in optimally time consumption.

## Code submission (Github link)

## Report submission (Github link) :

# Proposed Design/ Model

When you run this code, the summary plot will be displayed directly in your Jupyter Notebook, showing the features on the x-axis and their corresponding SHAP values on the y-axis. Positive SHAP values indicate that the feature has a positive impact on the model's predictions, while negative values indicate a negative impact. The length and color of each bar represent the magnitude and direction of the impact.

The summary plot is a valuable tool for gaining insights into your model's behavior and understanding the most critical features that influence its predictions.

-The vertical axis on the left indicates the names of the features, arranged according to their importance, from top to bottom. /-The horizontal axis represents the magnitude of the SHAP values for the predictions. /-The right vertical axis represents the actual magnitude of a feature as it appears in the data set and colors the points. From the graph of% Silica Feed, as we see in the graph that the more value this characteristic has, the more it impacts the SAHP value, therefore it affects the model more, which makes sense. From the graph of Flotation Column 03 Air Flow, from this graph we can see that while the values are higher the shap value decreases and while the value is lower the shap value increases. If we keep observing we can draw more conclusions. A good understanding of these graphs is important to understand the model.

Local interpretability helps us understand an individual prediction of the model, showing how each characteristic affects the prediction. That is, the SHAP values of all the features are added together to explain why my prediction was different from the baseline (the mean). This allows us to decompose a prediction into a graph that shows which characteristics contribute positively, which characteristics contribute negatively, and how much those characteristics mattered. In the case of a metallurgical plant, it will help us understand why the% Silica Concentrate varied at a certain time and what this variation was due to, so that the engineers pay attention to these points.

## Interfaces (if applicable)

Update with Block Diagrams, Data flow, protocols, FLOW Charts, State Machines, Memory Buffer Management.

# Performance Test

1. Firstly I gone Through Data and Problem statement and then started work on it and find relevant documentation where I can get appropriate information about how mini g process is working and what are the paraments includes in mining processing
2. Then with the help of python programing languages I first import the data and done all parts of EDA and then Make that data it moves in to channel with help of SHAP.
3. With the help of Sci-Learn I imitated model building and after that the time to check the accuracy of that model which it fit in good condition.

## Test Plan/ Test Cases

1. RME value
2. RMS value.
3. Accuracy Check.
4. Impurity Checky of Mining Processing

## Performance Outcome

Accuracy Comes to 92% which means that model has been submitted and build in good manner.

# My learnings

The project of Quality Prediction in the Mining Process involves using data analysis and predictive modeling techniques to forecast the quality of mined materials based on various input parameters. Here are some of the key things that can be learned from such a project:

1. \*\*Data Preprocessing:\*\* One of the initial steps in this project would be collecting and preparing the raw data. This process includes handling missing values, data cleaning, and transforming the data into a suitable format for analysis.

2. \*\*Feature Engineering:\*\* Selecting and engineering relevant features (input parameters) from the available data can significantly impact the accuracy of the predictive models. This process involves understanding the domain and the variables that are likely to influence the quality of the mined materials.

3. \*\*Domain Knowledge:\*\* Working on this project requires a solid understanding of the mining process and the factors that contribute to material quality. This includes knowledge of geological and chemical aspects that affect the properties of the mined materials.

4. \*\*Data Analysis:\*\* Analyzing the data to identify patterns, trends, and correlations among the input variables and the target variable (quality of mined materials) is a crucial step. This analysis can provide insights into which factors are most influential in determining quality.

5. \*\*Predictive Modeling:\*\* Building predictive models involves using machine learning algorithms to establish a relationship between the input parameters and the quality of mined materials. Different algorithms can be employed, such as regression, decision trees, random forests, or neural networks, depending on the complexity of the data.

6. \*\*Model Evaluation:\*\* Once models are trained, they need to be evaluated using appropriate metrics. Common metrics for regression tasks include Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared. These metrics help assess how accurately the model predicts the quality values.

7. \*\*Model Interpretation:\*\* Understanding how the model makes predictions is essential for gaining insights into the underlying relationships between input parameters and quality. Techniques like feature importance analysis can help in identifying the most influential factors.

8. \*\*Optimization:\*\* Continuously improving model performance is an ongoing process. This could involve parameter tuning, using more advanced algorithms, or collecting additional data to refine the predictions.

9. \*\*Business Insights:\*\* Beyond the technical aspects, this project can provide valuable business insights. Accurate quality predictions can lead to optimized resource allocation, better production planning, and cost reduction in the mining process.

10. \*\*Real-World Applications:\*\* The skills learned from this project can be applied to various industries beyond mining. Predictive modeling for quality estimation is relevant in fields such as manufacturing, agriculture, and environmental monitoring.

11. \*\*Challenges and Limitations:\*\* Working on this project may also uncover challenges such as data inconsistencies, the need for domain expertise, and the limitations of predictive models. These challenges provide valuable learning experiences.

In essence, the Quality Prediction in the Mining Process project provides a hands-on opportunity to develop skills in data analysis, machine learning, domain understanding, and problem-solving. It also highlights the importance of collaboration between data scientists and domain experts to achieve meaningful results.

# Future work scope

The project of Quality Prediction in the Mining Process holds significant potential for future advancements and applications. As technology and data analytics continue to evolve, there are several exciting directions in which this project's scope could expand:

1. \*\*Advanced Machine Learning Models:\*\* While traditional machine learning algorithms are effective, the project's future scope could involve exploring more advanced techniques, such as deep learning and neural networks. These models have the potential to capture complex relationships in the data and improve prediction accuracy.

2. \*\*Big Data Integration:\*\* Mining operations generate massive amounts of data. Incorporating big data technologies and frameworks can enable the handling of larger datasets, leading to more accurate and robust predictions.

3. \*\*Real-Time Monitoring:\*\* Implementing real-time monitoring and prediction systems can provide immediate insights into the quality of mined materials. This could lead to quicker decision-making, optimized resource allocation, and enhanced operational efficiency.

4. \*\*Integration of Sensor Data:\*\* Many mining processes utilize sensors to gather real-time data on various parameters. Integrating sensor data into the predictive models can enhance their accuracy and provide a more comprehensive view of the mining process.

5. \*\*Remote Sensing and Satellite Data:\*\* Incorporating remote sensing data, such as satellite imagery and aerial photography, can offer a broader perspective on mining activities and their impact on the environment. This could lead to more sustainable and responsible mining practices.

6. \*\*Predictive Maintenance:\*\* Beyond predicting material quality, the project could expand to predicting equipment failures and maintenance needs. Predictive maintenance can help minimize downtime and reduce operational costs.

7. \*\*Supply Chain Optimization:\*\* Quality prediction can also be integrated into the supply chain management of mining operations. Predicting material quality at different stages of the process can optimize inventory management and production planning.

8. \*\*Environmental Impact Assessment:\*\* Predicting the quality of mined materials can help assess their environmental impact. This information can aid regulatory compliance and support efforts to minimize negative ecological effects.

9. \*\*Automation and AI Assistance:\*\* Automation of the prediction process, along with the integration of AI-driven decision support systems, can empower mining professionals with actionable insights for better decision-making.

10. \*\*Multi-Objective Optimization:\*\* The project's scope could expand to optimize multiple objectives simultaneously, such as maximizing quality while minimizing energy consumption or environmental impact.

11. \*\*Collaboration with Geologists and Domain Experts:\*\* Collaborating with domain experts, such as geologists and mining engineers, can enhance the accuracy of the predictions by incorporating their insights and expertise.

12. \*\*Global Application:\*\* The project's techniques and models could be adapted to different types of mining operations around the world, enabling better resource management and sustainable mining practices on a global scale.

13. \*\*Education and Research:\*\* The project's success could contribute to educational materials, workshops, and research publications, advancing the field of geoprocessing, predictive modeling, and sustainable mining practices.

In summary, the future scope of the Quality Prediction in the Mining Process project is promising and wide-ranging. By incorporating emerging technologies, domain knowledge, and a focus on sustainability, this project can make substantial contributions to the mining industry's efficiency, environmental responsibility, and overall success.