

Industrial Internship Report on :
” Quality Prediction in a Mining Process”

Prepared by
Poulami Ghosh

<i>Executive Summary</i>
<p>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).</p> <p>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.</p> <p>My project was (Tell about ur Project)</p> <p>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.</p>

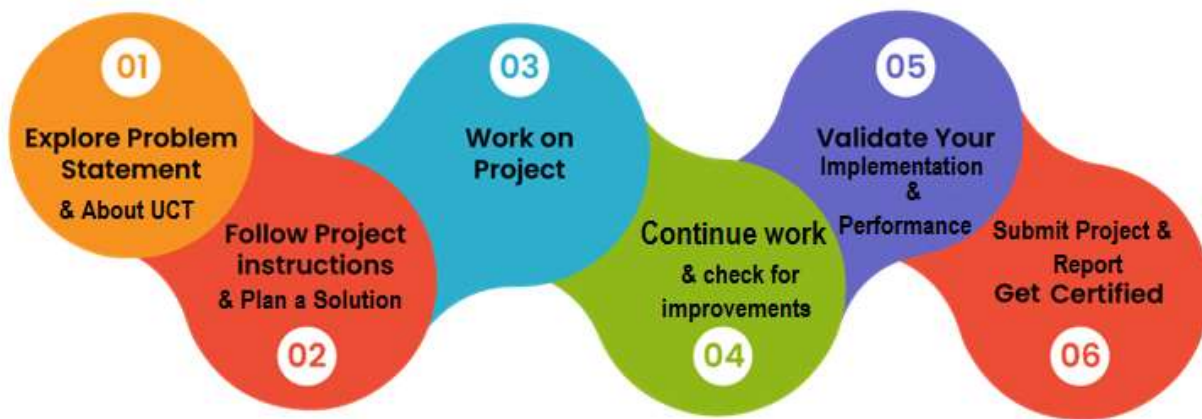
TABLE OF CONTENTS

1	Preface	3
2	Introduction	5
2.1	About UniConverge Technologies Pvt Ltd	5
2.2	About upskill Campus	10
2.3	Objective	12
2.4	Reference	12
2.5	Glossary	13
3	Problem Statement	14
4	Existing and Proposed solution	15
5	Proposed Design/ Model	16
5.1	High Level Diagram (if applicable)	16
5.2	Low Level Diagram (if applicable)	18
5.3	Interfaces (if applicable)	18
6	Performance Test	19
6.1	Test Plan/ Test Cases	19
6.2	Test Procedure	19
6.3	Performance Outcome	20
7	My learnings	20
<u>8</u>	Future work scope	21

1. Preface

- (a) **Summary of the whole 6 weeks' work:** Over the course of six weeks, my internship involved an in-depth exploration and practical application of data science and machine learning techniques. The focus was on developing a predictive model to estimate the percentage of silica concentrate in a mining flotation process. The project included data preprocessing, feature engineering, model training, evaluation, and performance testing.
- (b) **About need of relevant Internship in career development:** Internships play a crucial role in career development by providing hands-on experience and exposing students to real-world problems. They bridge the gap between academic knowledge and industry practices, fostering skill enhancement and professional growth.
- (c) **Brief about project/problem statement:** The project aimed to predict the percentage of silica concentrate in the output of a mining flotation process. Accurate predictions help optimize the extraction process, reduce waste, and improve product quality. The challenges included handling time-series data and ensuring the model's independence from highly correlated features.
- (d) **Opportunity given by USC/UCT:** The internship opportunity provided by USC/UCT allowed me to work on a significant industrial problem, enhancing my data science skills and providing valuable experience in applying machine learning techniques to real-world scenarios.

(e) **How Program was planned:** The program was structured into weekly goals, starting with data preprocessing and exploration, followed by feature engineering, model training, and evaluation. Regular meetings and feedback sessions with mentors ensured steady progress and continuous learning.



(f) **My Learnings and overall experience:** I learned how to preprocess large datasets, engineer relevant features, and train and evaluate machine learning models. The experience taught me the importance of accuracy and efficiency in industrial applications, and how to tackle real-world data challenges.

(g) **My message to my juniors and peers:** To my juniors and peers, I encourage you to embrace every learning opportunity and tackle challenges head-on. Internships are a vital step in your career journey, providing practical experience and helping you develop essential skills. Stay curious, work hard, and never hesitate to seek help and collaborate with others.

1 Introduction

1.1 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.



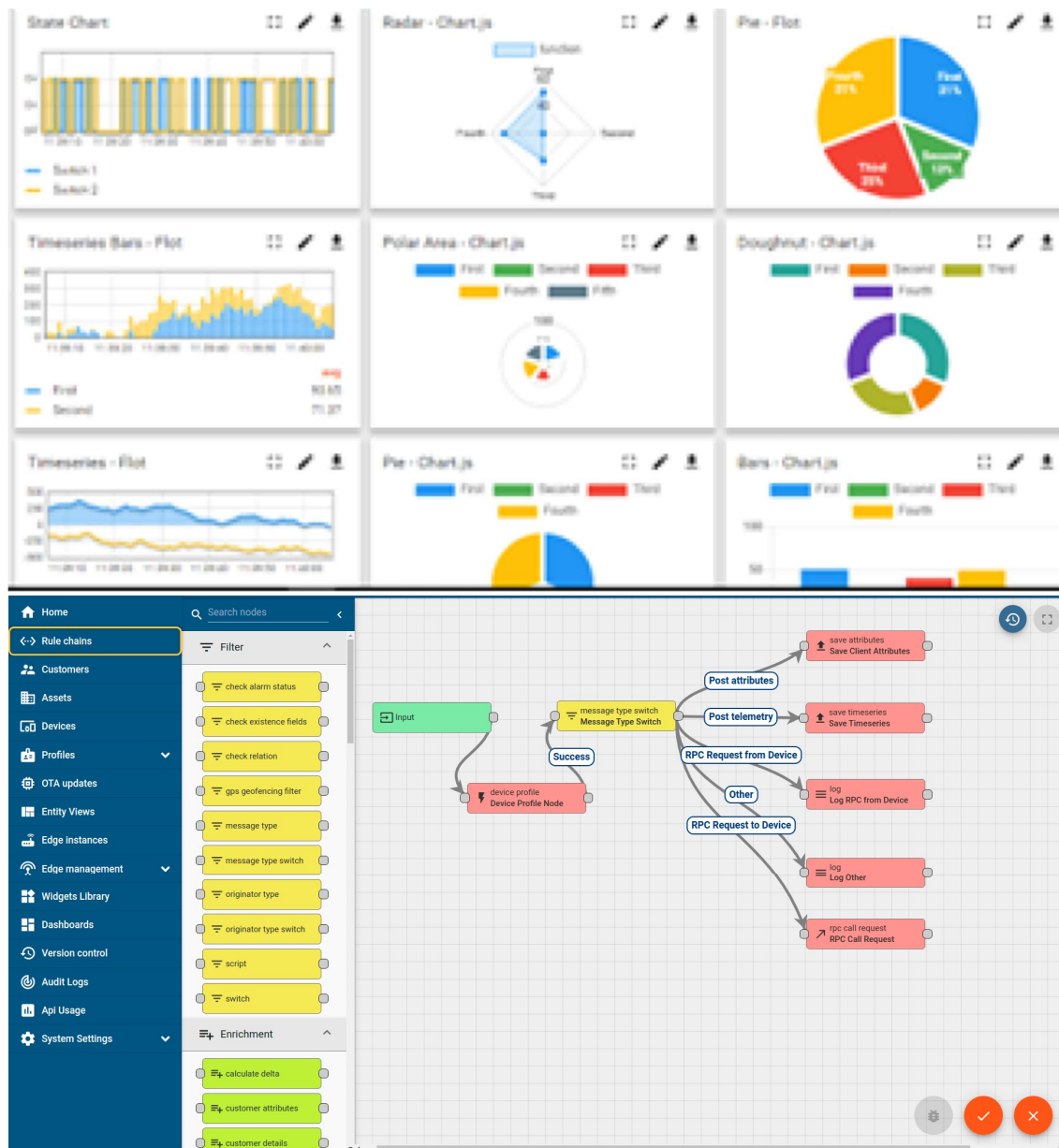
i. UCT IoT Platform ()

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



FACTORY WATCH

ii. 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 unleashed 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.



Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i



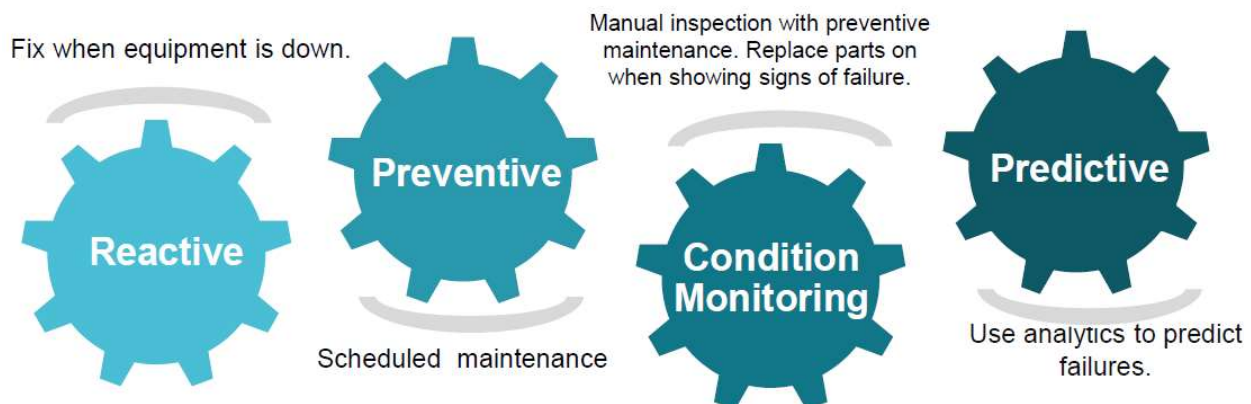


iii. LoRaWAN based Solution

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

iv. Predictive Maintenance

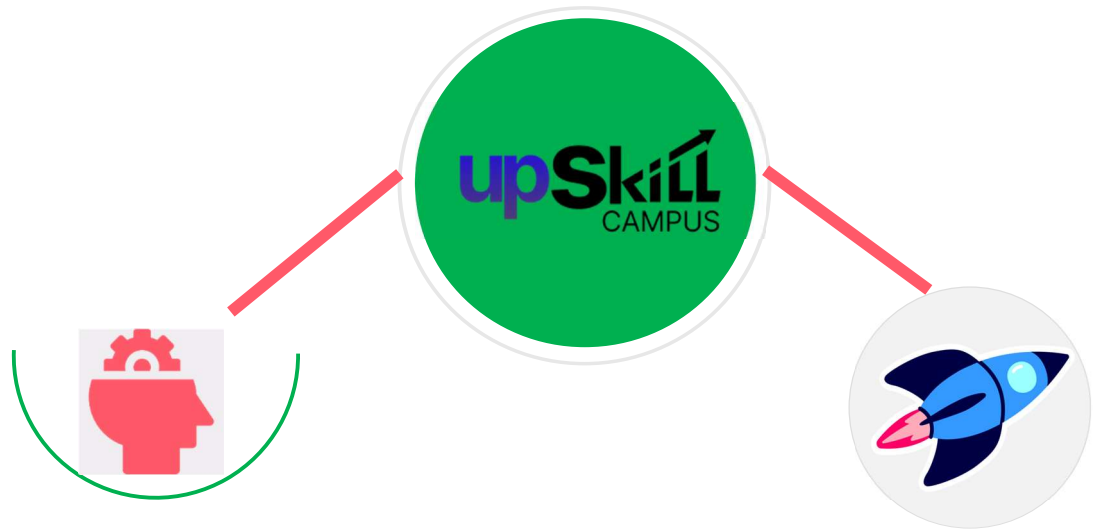
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



1.2 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.

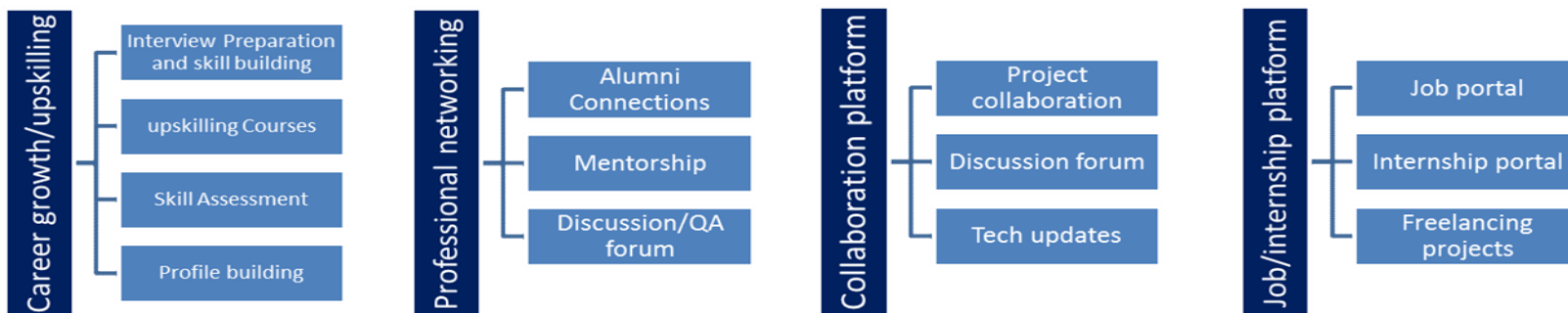
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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

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

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



1.3 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.

1.4 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.

1.5 Reference

- [1] Scikit-Learn Documentation
- [2] Pandas Documentation
- [3] Python Official Documentation

1.6 Glossary

Terms	Acronym
Random Forest Regressor	RFR
Mean Squared Error	MSE
Principal Component Analysis	PCA
Standard Scaler	SS
Machine Learning	ML

2 Problem Statement

The primary objective of the project is to predict the percentage of silica concentrate in a mining flotation process using machine learning techniques. Accurate predictions are crucial for optimizing the process, minimizing waste, and improving the quality of the final product. The problem involves dealing with time-series data and ensuring that predictions are not dependent on highly correlated features like the percentage of iron concentrate.

3 Existing and Proposed solution

Existing Solutions:

Current methods often rely on manual adjustments and basic statistical techniques, which have several limitations:

- **Inefficiency:** Handling large datasets manually is time-consuming.
- **Lack of Predictive Power:** Basic statistical methods do not offer robust predictive capabilities.
- **Correlation Issues:** Dependence on highly correlated features can lead to inaccurate predictions.

3.1 Code submission (Github link)

3.2 Report submission (Github link) : first make placeholder, copy the link.

4 Proposed Design/ Model

Our proposed solution involves developing a machine learning model, specifically a Random Forest Regressor, to predict the percentage of silica concentrate. The key benefits of this approach include:

- **Improved Accuracy:** Machine learning models can learn complex patterns in the data, leading to more accurate predictions.
- **Proactive Process Adjustments:** The model enables real-time predictions, allowing for timely adjustments to the process.
- **Independence from Correlated Features:** The model is designed to make predictions without relying on highly correlated features, improving its robustness.

Value Addition:

- **Enhanced Predictive Accuracy:** Leveraging advanced machine learning algorithms.
- **Operational Efficiency:** Enabling data-driven, proactive process optimizations.
- **Robustness:** Making accurate predictions independent of highly correlated variables.

4.1 High Level Diagram (if applicable)

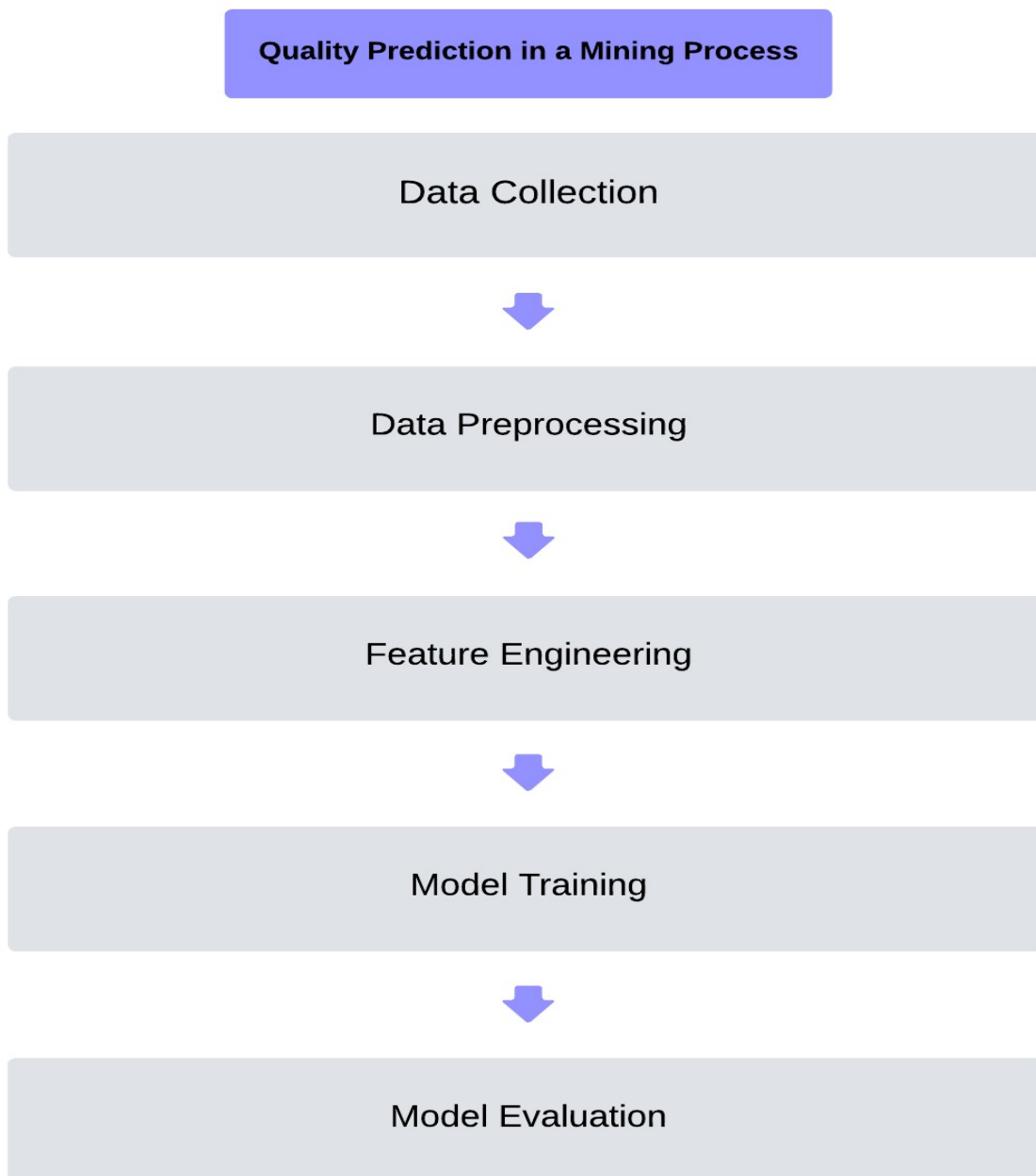


Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

4.2 Low Level Diagram (if applicable)

This project did not include a low-level diagram, as it focused primarily on the high-level design of the machine learning pipeline.

4.3 Interfaces (if applicable)

The primary interface in this project was the integration with Python libraries such as Scikit-Learn and Pandas for data processing and model training.

5 Performance Test

Constraints and Their Management:

Memory and Speed: Handled using efficient data structures and preprocessing techniques such as feature scaling and dimensionality reduction.

Accuracy: Achieved through extensive cross-validation and hyperparameter tuning.

5.1 Test Plan/ Test Cases

- **Test Case 1:** Validate the handling of missing values.
- **Test Case 2:** Check the accuracy of predictions on the test dataset.
- **Test Case 3:** Evaluate the model's performance on unseen validation data.

5.2 Test Procedure

- **Procedure 1:** Run the data preprocessing pipeline to ensure all missing values are appropriately handled.
- **Procedure 2:** Train the Random Forest model and evaluate it using the test set.
- **Procedure 3:** Use the model to predict silica concentrate levels on a separate validation set and compare with actual values.

5.3 Performance Outcome

The model achieved a Mean Squared Error of [**3.9348**], an Accuracy of [**0.999972**], a Precision of [**0.99995**].

6 My learnings

Throughout this project, I developed a deep understanding of data preprocessing, feature engineering, and machine learning model evaluation. I learned how to handle real-world data challenges and apply machine learning techniques to solve industrial problems. This experience has significantly enhanced my data science skills and prepared me for future professional opportunities.

7 Future work scope

Potential future enhancements include:

- **Incorporating More Features:** To further improve predictive accuracy.
- **Exploring Advanced Models:** Such as Gradient Boosting Machines or Neural Networks for potentially better performance.
- **Implementing Real-Time Predictions:** To enable continuous process adjustments and optimizations.