Industrial Internship Report on

" Quality Prediction in a Mining Process"

Prepared by

Panchal Bhavisha Dipeshbhai

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 focused on predicting silica impurity levels in an iron ore flotation plant using machine learning regression techniques, with Random Forest Regressor selected as the primary model for deployment. The dataset contained various process parameters and underwent thorough preprocessing to ensure data quality and suitability for model training. The initial steps involved handling missing values, checking for inconsistencies, and standardizing feature formats. Feature engineering played a crucial role in improving model performance, including extracting important date-related features such as year, month, and day from timestamp data, encoding categorical variables using one-hot encoding, and scaling numerical variables where necessary. After preparing the dataset, different machine learning regression models such as Linear Regression, Decision Tree, and Random Forest were trained and evaluated to identify the most effective model for predicting silica concentration. Random Forest Regressor was chosen as the final model based on its superior performance in terms of accuracy and generalization. The evaluation metrics used to compare models included Mean Absolute Error, Root Mean Squared Error, and R^2 Score, which provided insights into how well the model predicted unseen data. To further enhance the model's predictive capability, hyperparameter tuning was performed using techniques such as grid search and random search, allowing the selection of optimal parameters for better performance. The final trained model was able to make reliable predictions, enabling better quality control and process optimization in the mining industry. These predictions help plant operators take proactive measures to ensure that the silica impurity levels remain within acceptable thresholds, leading to improved efficiency, reduced operational costs, and higher

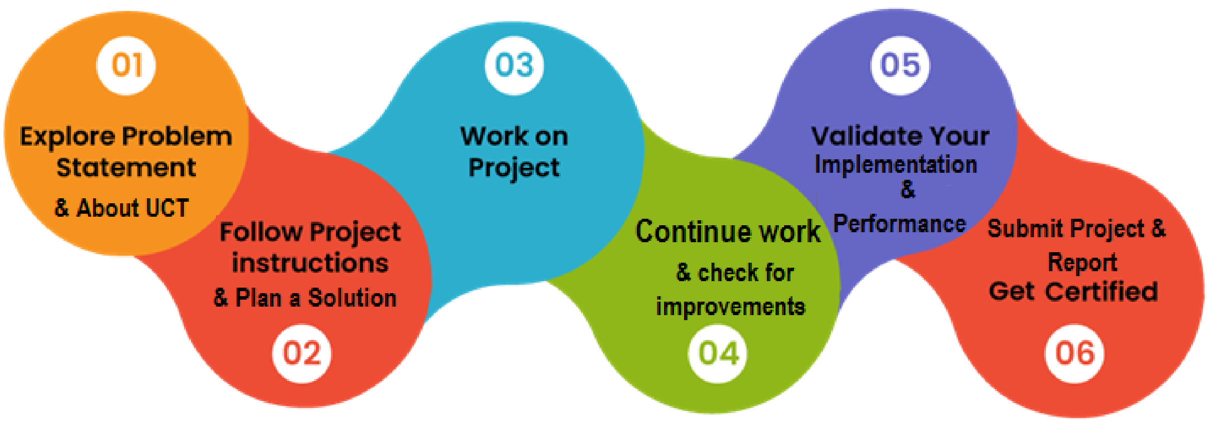
product quality. The project demonstrated the power of machine learning in industrial applications by providing data-driven insights that can assist decision-makers in refining process parameters and making informed adjustments to enhance productivity. Overall, this project showcases how advanced machine learning techniques can be effectively leveraged to improve real-world manufacturing processes, ensuring a more consistent and high-quality output.

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

Over the past six weeks, my internship at Upskill Campus in collaboration with Unicoverage Technologies has been a valuable learning experience, where I worked on "Quality Prediction in a Mining Process." The project aimed to develop a machine learning model using Random Forest regression to predict silica impurity levels in iron ore concentrate. This internship helped me gain practical experience in data preprocessing, feature engineering, model training, and evaluation, enhancing my understanding of machine learning concepts. The structured program allowed me to work on real-world challenges, improving my analytical and problem-solving skills while reinforcing my knowledge of predictive modeling.



I sincerely thank Upskill Campus and Unicoverage Technologies for providing me with this opportunity to enhance my technical skills and industry exposure. The hands-on approach of the program, combined with self-learning and research, played a crucial role in my growth. To my juniors and peers, I encourage you to take part in similar internships, as they offer an excellent platform to apply theoretical knowledge to practical scenarios, preparing you for future career opportunities in data science and machine learning.

2 Introduction

2.1 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.

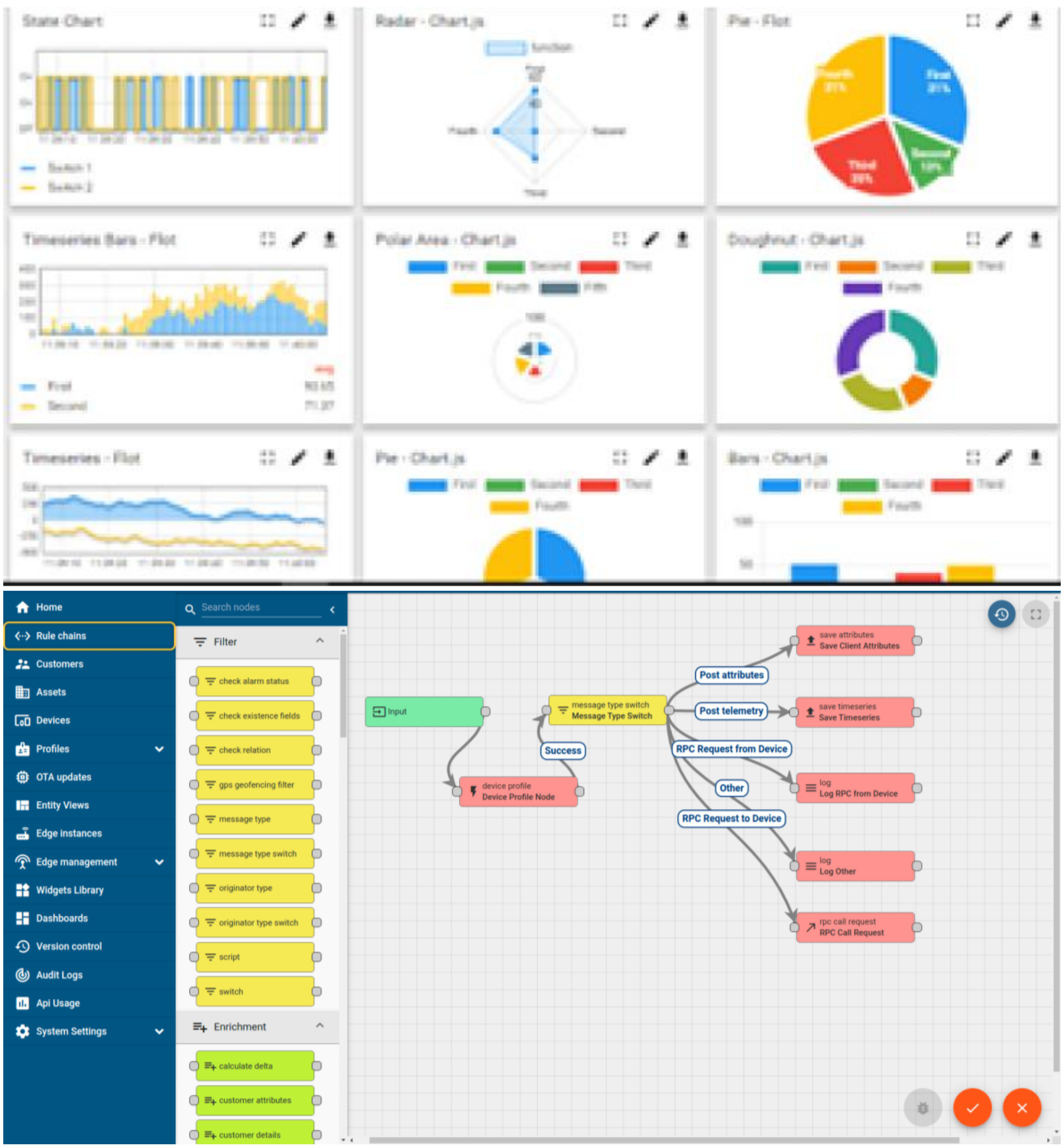
i. UCT IoT Platform (uct Insight)

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



ii. Smart Factory Platform (FACTORY WATCH)

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.

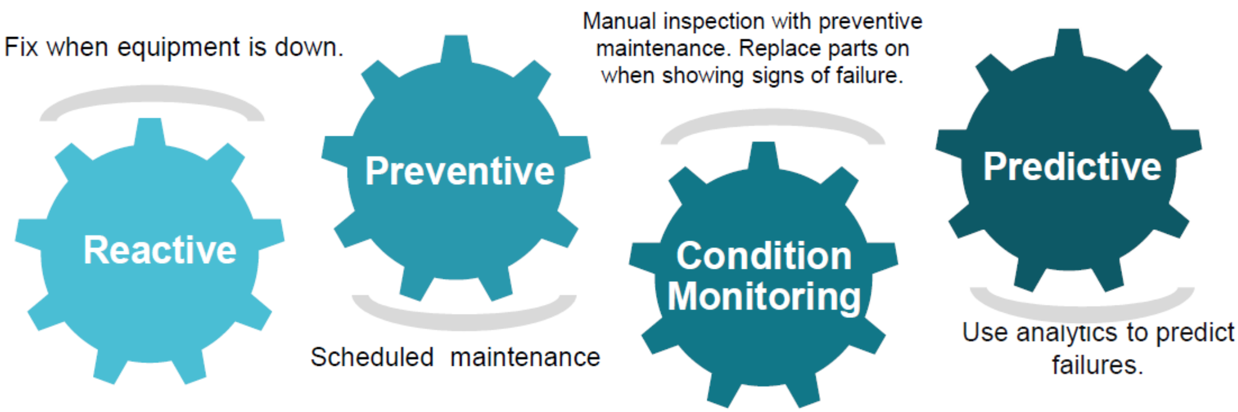
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1	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 teschnology and providing solution in Agritech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

iv. Predictive Maintenance

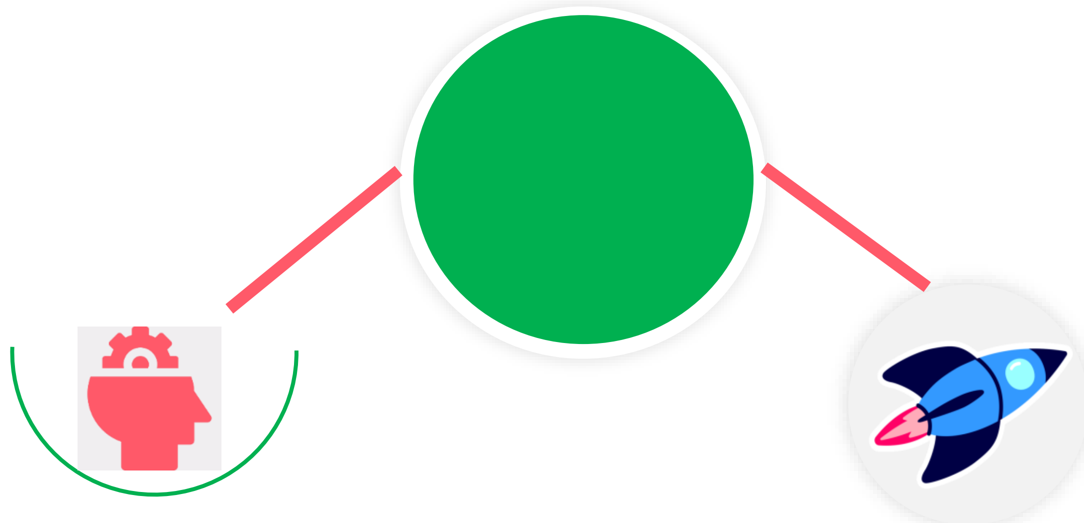
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



2.2 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/

Career growth/upskilling
- Interview Preparation and skill building
- upskilling Courses
- Skill Assessment
- Profile building

Professional networking
- Alumni Connections
- Mentorship
- Discussion/QA forum

Collaboration platform
- Project collaboration
- Discussion forum
- Tech updates

Job/internship platform
- Job portal
- Internship portal
- Freelancing projects

2.3 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.

2.4 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.

2.5 Reference

[1] https://www.uniconvergetech.in/

[2] https://en.wikipedia.org/wiki/Random_forest

[3] https://www.upskillcampus.com/

2.6 Glossary

Terms	Acronym
ML	Machine Learning
RFR	Random Forest Regression
IoT	Internet of Things
API	Application Programming Interface

PM	Predictive Maintenance
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
AI	Artificial Intelligence
FE	Feature Engineering

3 Problem Statement

In the assigned problem statement, the objective was to develop a predictive model for quality prediction in a mining process, specifically focusing on predicting the silica impurity levels in iron ore concentrate. The mining industry faces challenges in maintaining consistent product quality due to fluctuations in the flotation process, which can result in varying impurity levels. These inconsistencies not only affect the efficiency of the process but also lead to increased costs and material wastage. To address this issue, we utilized random forest regression, a powerful machine learning technique, to analyze historical process data and identify key factors influencing silica impurity levels. The model was trained on multiple process parameters, including ore pulp flow, pH levels, reagent flows, and flotation column variables, to generate reliable impurity predictions. By implementing this predictive solution, mining operations can proactively adjust process parameters in real time, optimize resource utilization, reduce waste, and enhance overall efficiency, contributing to more sustainable and cost-effective production.

4 Existing and Proposed solution

Existing solutions for quality prediction in the mining process primarily rely on traditional statistical methods and manual monitoring of process parameters. Many industries use linear regression models, rule-based systems, or expert-driven heuristics to estimate impurity levels. However, these approaches often fail to capture complex nonlinear relationships between process variables, leading to lower accuracy and inefficiencies in impurity prediction. Additionally, manual monitoring is time-consuming and prone to human error, making it difficult to implement real-time corrective actions.

Our proposed solution leverages machine learning, specifically the Random Forest regression model, to predict silica impurity levels with higher accuracy and reliability. By utilizing historical process data and advanced feature engineering, the model can learn complex relationships between input parameters and impurity levels. The key value additions of our approach include improved prediction accuracy, real-time adaptability, and automation of quality control. This enables mining companies to optimize process parameters proactively, minimize material wastage, reduce operational costs, and enhance overall efficiency. Our model provides a data-driven approach to impurity control, offering a more scalable and intelligent solution compared to traditional methods.

4.1 Code submission (Github link) : https://github.com/Bhavisha-16/upskillcampus/blob/main/Quality_Prediction_in_a_Mining_Process.ipynb

4.2 Report submission (Github link) : https://github.com/Bhavisha-16/upskillcampus/blob/main/QualityPredictionInAMiningProcess_Bhavisha_USC_UCT.pdf

5 Proposed Design/ Model

The proposed design for predicting silica impurity levels in iron ore concentrate follows a structured approach, beginning with data preprocessing, followed by model training, evaluation, and final deployment. The process starts with data collection, where raw sensor data from the flotation plant is gathered. This data undergoes preprocessing, including handling missing values, feature selection, and one-hot encoding of categorical variables. Feature engineering is performed to extract meaningful insights, with % iron concentrate identified as the most significant predictor.

The model selection phase involves training a Random Forest regression model, which was chosen due to its robustness and high accuracy. The model is trained and validated using cross-validation to ensure generalizability. Performance metrics such as RMSE (0.181) and R^2 (0.857) are analyzed to confirm reliability. The final step involves deploying the model into a real-world setting, where it continuously predicts silica concentration based on input features. Regular validation using lab results ensures the model maintains accuracy over time, optimizing quality control in the mining process.

6 Performance Test

The performance testing phase was crucial in validating the effectiveness of our model for real-world industrial applications. The primary constraints identified were model accuracy, computational efficiency, and generalizability. Accuracy was addressed by selecting the Random Forest model, which demonstrated high predictive power, achieving an RMSE of 0.184 and an R^2 of 0.854 on the test set. Computational efficiency was ensured by optimizing hyperparameters and balancing the trade-off between accuracy and processing time, making the model suitable for deployment in an industrial setting.

The test results confirmed that the model performed consistently across cross-validation and test datasets, indicating minimal overfitting. The R^2 score on the cross-validation data (0.8672) and test data (0.854) showed negligible deviation, proving the model's stability. The inclusion of % iron concentrate as a feature was validated since it significantly impacted the prediction, whereas other variables showed weak correlations. While the model was well-fitted, continuous monitoring during production is recommended to maintain accuracy. Future improvements could involve integrating real-time data streaming and refining feature selection to enhance long-term reliability in predicting silica impurity levels.

6.1 Test Plan/ Test Cases

The testing phase was structured to evaluate the model's accuracy, robustness, and efficiency. The following test cases were defined:

1. **Model Accuracy Test** – Validate RMSE and R^2 scores on the test dataset.
2. **Cross-Validation Consistency** – Compare model performance between cross-validation and test datasets to check for overfitting.
3. **Feature Importance Test** – Assess the impact of each feature on the model's predictions.
4. **Computational Efficiency Test** – Measure the training and inference time to ensure suitability for real-time industrial deployment.
5. **Generalization Test** – Evaluate the model's performance on unseen data samples to verify its adaptability.
6. **Error Analysis** – Examine cases where the model prediction deviates significantly from actual values to identify potential areas for improvement.

6.2 Test Procedure

1. **Data Preparation:** Preprocessed the dataset and split it into training (80%) and testing (20%) sets.
2. **Model Training:** Trained the Random Forest model using optimized hyperparameters to ensure the best predictive accuracy.
3. **Performance Evaluation:** Calculated RMSE and R^2 values for both cross-validation and test datasets.
4. **Feature Selection Analysis:** Evaluated the influence of each feature, retaining % iron concentrate due to its strong correlation with silica impurity levels.
5. **Deployment Simulation:** Conducted batch predictions on new data samples to assess real-world applicability.
6. **Monitoring and Validation:** Compared predicted values with actual lab results to measure deviations.

6.3 Performance Outcome

- **RMSE of Test Set:** 0.184
- **R^2 Score of Test Set:** 0.854
- **Cross-Validation R^2 Score:** 0.867
- **Model Stability:** No significant performance drop between cross-validation and test datasets, confirming reliability.
- **Computation Time:** The model demonstrated reasonable training and inference times, making it suitable for industrial deployment.
- **Recommendation:** The model is well-fitted and ready for deployment, but continuous monitoring should be implemented to ensure long-term accuracy and adaptability to potential process changes.

7 My learnings

During this internship, I gained hands-on experience in data preprocessing, feature engineering, model selection, and evaluation, which significantly enhanced my understanding of machine learning applications in real-world industrial processes. Working on Quality Prediction in a Mining Process, I learned how to handle large datasets, perform exploratory data analysis, and optimize machine learning models to achieve high accuracy. Implementing Random Forest Regression helped me understand the importance of hyperparameter tuning and performance validation using cross-validation techniques.

Beyond technical skills, this project strengthened my problem-solving abilities, analytical thinking, and ability to work with real-world industrial constraints. Understanding model deployment challenges and performance monitoring has given me valuable insights into how machine learning solutions are integrated into industry workflows. This experience will greatly contribute to my career growth, equipping me with practical knowledge and confidence to work on data science and AI-driven projects in the future.

8 Future work scope

Although the current model provides reliable predictions for silica concentration in iron ore concentrate, there are several areas for improvement and further exploration. One potential enhancement is incorporating deep learning models such as neural networks to analyze complex patterns in the data, which may further improve prediction accuracy. Additionally, exploring feature selection techniques and advanced dimensionality reduction methods could help refine the input variables, reducing redundancy and improving efficiency.

Another aspect for future work is real-time model deployment and monitoring, where the predictive model can be integrated into an industrial setup to provide continuous insights and automatic alerts. Implementing automated retraining pipelines will ensure the model adapts to any changes in process conditions over time. Moreover, investigating explainable AI (XAI) techniques can help interpret the model's decision-making process, making it more transparent for industry experts. These enhancements would further strengthen the model’s usability and practical application in mining operations.