Industrial Internship Report on

"Quality Prediction in a Mining Process"

Prepared by

Tejas A. Sakhare

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 -Industrial Manufacturing and Production This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solutions for that. It was an overall great experience to have this internship.

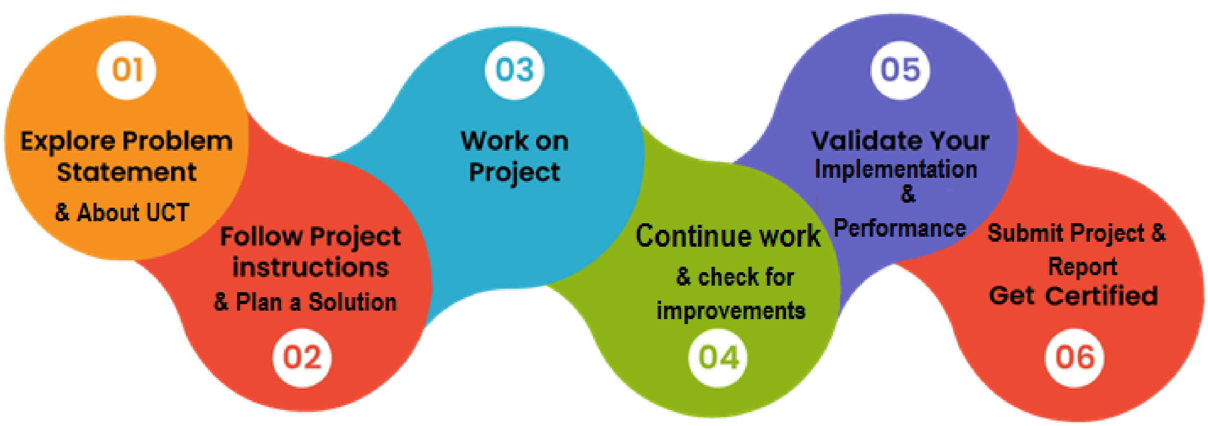
TABLE OF CONTENTS

1 Preface

Summary of the whole 6 week's work: Internships in the data science domain are essential for bridging the gap between theoretical knowledge and practical application. They offer hands-on experience with real-world datasets, tools, and methodologies, honing technical skills like data analysis, machine learning, and programming. Moreover, internships provide exposure to industry best practices, collaboration with seasoned professionals, and insight into the nuances of data-driven decision-making. This immersive learning environment fosters critical thinking, problem-solving abilities, and adaptability, preparing aspiring data scientists for the complexities of the field and facilitating a smoother transition into professional roles.

Brief Summary of the problem statement: The project focused on quality prediction within the mining process, aiming to enhance efficiency and optimize resource utilization. Leveraging advanced data science techniques, the objective was to develop predictive models capable of forecasting the quality of mined materials at various stages of production. The opportunity provided by USC/UCT enabled access to comprehensive datasets and domain expertise, facilitating in-depth analysis and model development. This collaboration offered a unique opportunity to address real-world challenges in the mining industry, driving innovation and fostering valuable insights for process optimization and quality assurance.

How Program was planned:



Learnings and overall experience: Throughout the internship, I immersed myself in a dynamic learning environment, gaining hands-on experience with data science methodologies and tools. Collaborating with industry professionals, I refined my analytical skills and gained invaluable insights into the intricacies of quality prediction in the mining process. The opportunity to work on real-world projects not only enhanced my technical proficiency but also broadened my understanding of industry dynamics and best practices. Overall, the experience was transformative, empowering me to approach challenges with confidence and paving the way for continued growth and exploration in the field of data science.

Thanks to **Aditya M. Kapole**, who has helped you directly or indirectly.

Message to your juniors and peers: To my juniors and peers, I encourage you to embrace every opportunity for learning and growth, both within and beyond the classroom. Approach each challenge with curiosity and tenacity, leveraging the support of mentors and colleagues along the way. Remember that mistakes are inevitable but serve as valuable learning experiences. Stay curious, stay humble, and never underestimate the power of collaboration. As aspiring data scientists, let us continue to push boundaries, innovate fearlessly, and make meaningful contributions to the world of data science.

2 Introduction

2.1 About UniConverge Technologies Pvt Ltd

A company established in 2013 and working in the 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.

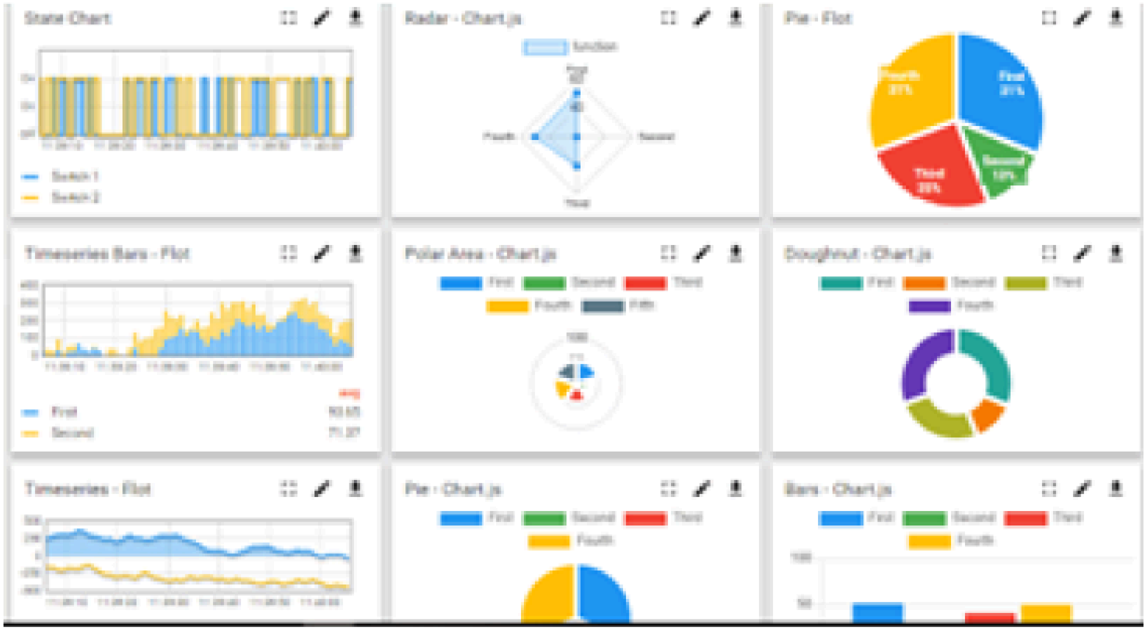
i. UCT IoT Platform (uct Insight)

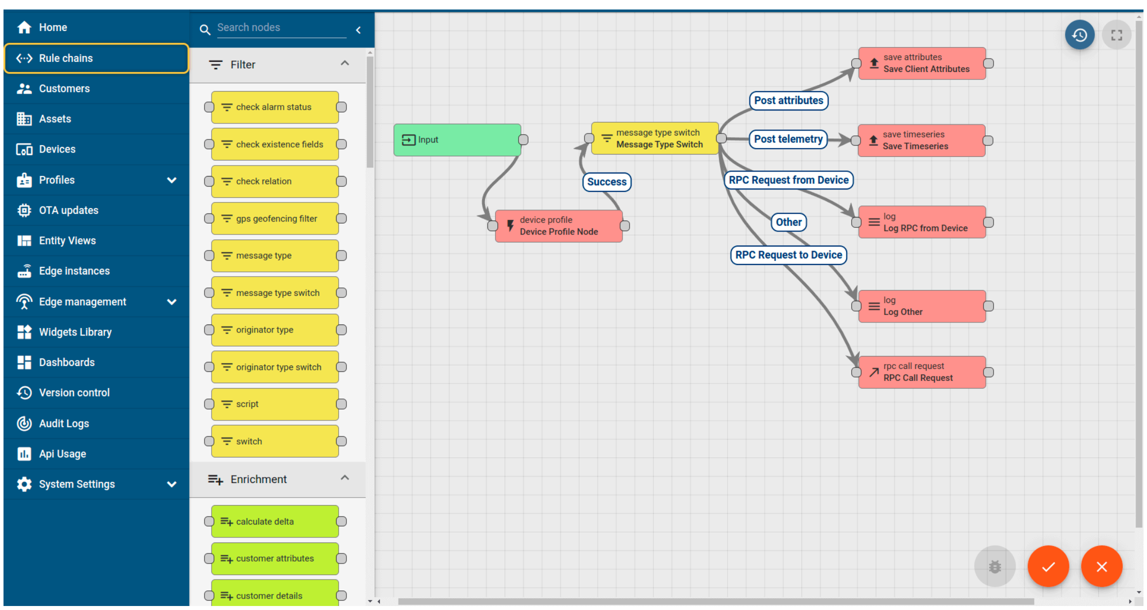
UCT Insight is an IOT platform designed for quick deployment of IOT applications at the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- 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




Home
Rule chains
Customers
Assets
Devices
Profiles
OTA updates
Entity Views
Edge instances
Edge management
Widgets Library
Dashboards
Version control
Audit Logs
Api Usage
System Settings
Search nodes
Filter
check alarm status
check existence fields
check relation
gps geofencing filter
message type
message type switch
originator type
originator type switch
script
switch
Enrichment
calculate delta
customer attributes
customer details
Input
message type switch
Message Type Switch
device profile
Device Profile Node
Success
Post attributes
Post telemetry
RPC Request from Device
Other
RPC Request to Device
save attributes
Save Client Attributes
save timeseries
Save Timeseries
log
Log RPC from Device
log
Log Other
rpc call request
RPC Call Request

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.
- tu 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 want 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.

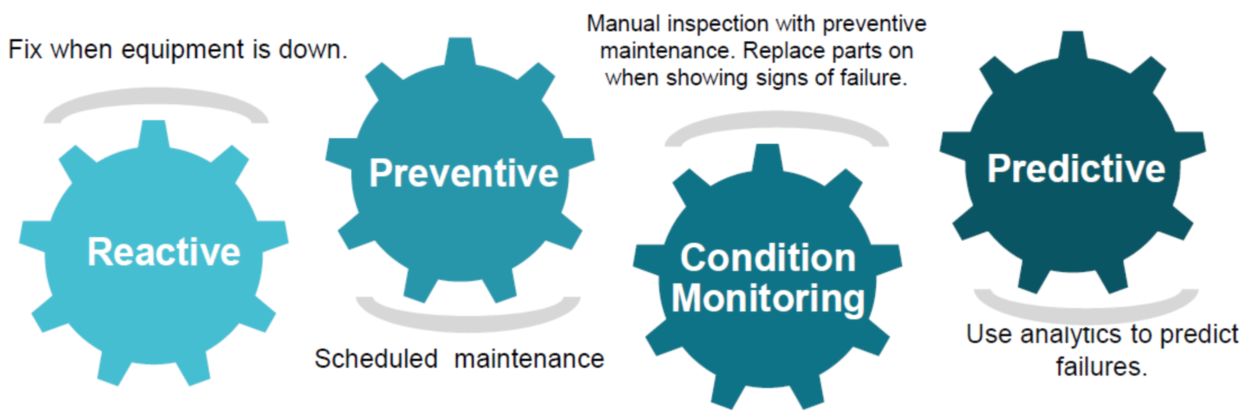
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

iv. Predictive Maintenance

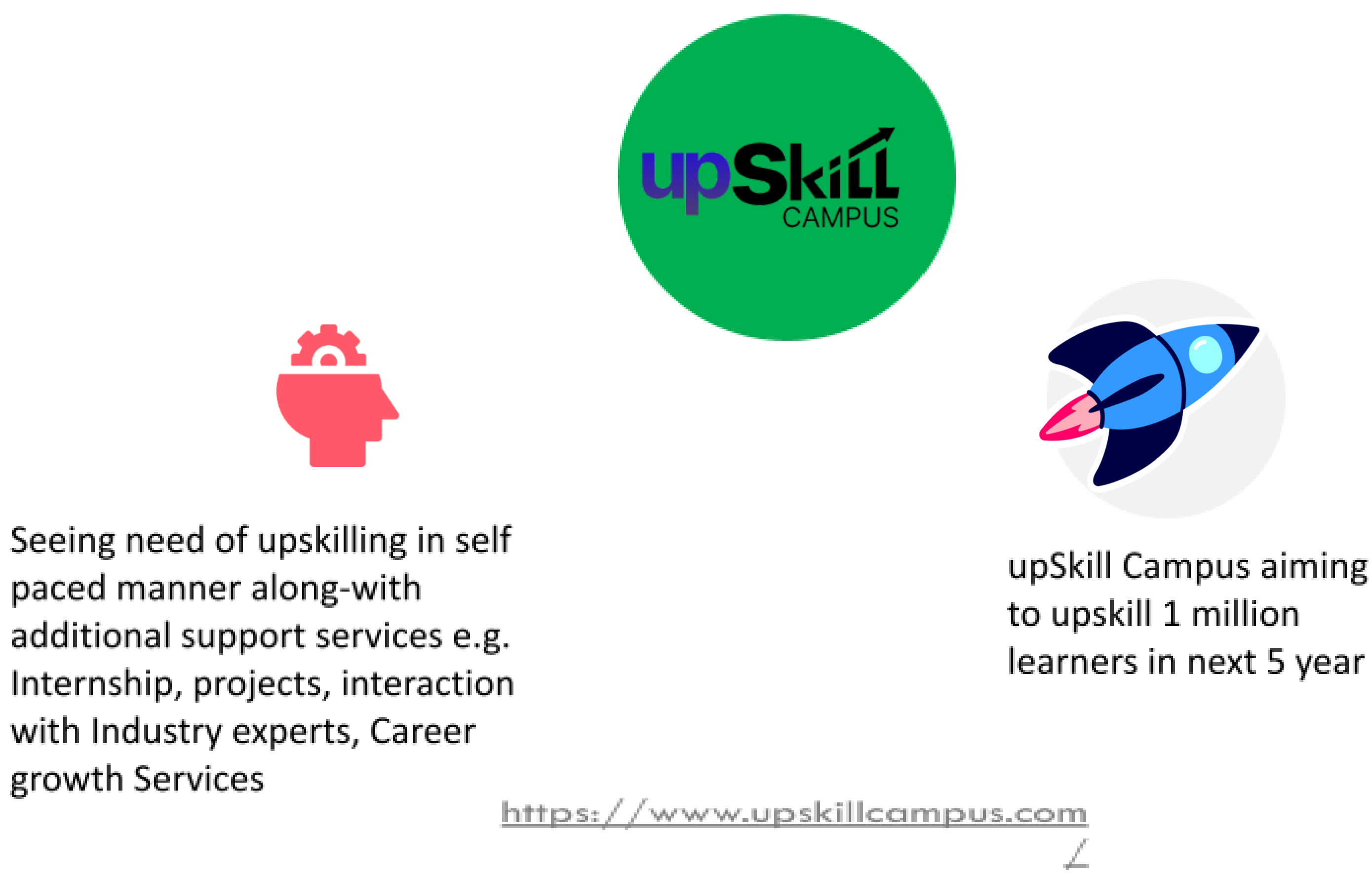
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.

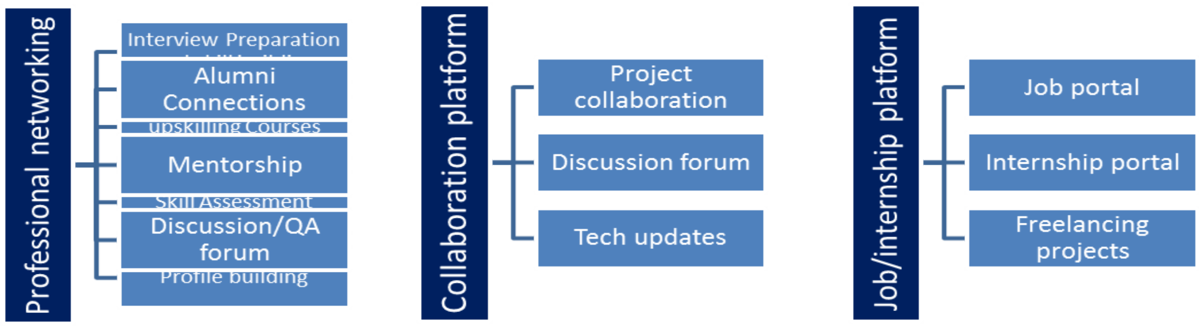


2.2 About upskill Campus (USC)

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.





2.3 The IoT Academy

The IoT academy is the 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://scikit-learn.org/stable/

[2] https://pypi.org/

3 Problem Statement

The problem statement revolves around optimizing manufacturing efficiency, particularly within the context of mining operations, by leveraging real-world industrial data. Specifically, the focus is on a flotation plant, a crucial component in the mining process. The overarching goal is to predict the presence of impurities, particularly silica, in the ore concentrate, thereby facilitating proactive decision-making to enhance both operational efficiency and environmental sustainability.

Mining processes, especially those involving flotation plants, are intricate and rely heavily on the quality of the ore concentrate produced. Silica, as an impurity, can significantly impact product quality and downstream processes. Therefore, accurately predicting silica content is paramount for engineers to take timely corrective actions, thereby ensuring optimal product quality and reducing environmental impact.

The dataset under consideration spans from March to September of 2017 and comprises various parameters sampled at different frequencies. These parameters include quality measures of the iron ore pulp before flotation, as well as process-related variables such as column levels and airflow within the flotation columns. The ultimate target variable for prediction is the percentage of silica in the iron ore concentrate.

The significance of this prediction task lies in its potential to provide engineers with actionable insights. By accurately forecasting silica content, engineers can preemptively adjust operational parameters or implement corrective measures to mitigate impurities, thus improving overall process efficiency and reducing waste. Furthermore, reducing impurities in the concentrate can lead to a decrease in the volume of ore sent to tailings, thereby minimizing environmental impact and optimizing resource utilization.

In summary, the problem statement entails leveraging real industrial data to predict silica content in iron ore concentrate, with the overarching aim of enhancing operational efficiency, product quality, and environmental sustainability within mining operations.

4 Existing and Proposed solution

Proposed Solution:

The proposed solution involves utilizing machine learning techniques, specifically Random Forest Regression, to predict the percentage of silica in iron ore concentrate.

1. Data Preprocessing: The initial step involves importing the dataset and conducting exploratory data analysis (EDA). This includes examining data types, checking for missing values, and visualizing the distribution of variables.

2. Feature Selection and Engineering: Following EDA, relevant features are selected based on their correlation with the target variable (% Silica Concentrate). Features with significant positive or negative correlations are retained for model training. Additionally, specific columns deemed important for prediction, such as 'Starch Flow', 'Amina Flow', 'Ore Pulp Flow', 'Ore Pulp pH', 'Ore Pulp Density', and '% Iron Concentrate', are identified and included in the analysis.

3. Model Training: Random Forest Regression is employed to build a predictive model. The dataset is split into training and testing sets using a 90:10 ratio. The model is trained on the training set, and its performance is evaluated using the testing set. The number of estimators (trees) in the Random Forest is set to 20 by default but can be adjusted based on performance requirements.

4. Model Evaluation: The performance of the trained model is evaluated using the coefficient of determination (R^2 score) to assess how well the model fits the data. Additionally, the accuracy of the model is measured using other relevant metrics to ensure its reliability in predicting silica concentration.

The Random Forest Regressor serves as a robust ensemble learning technique within the realm of machine learning, particularly suited for regression tasks. As a tree-based model, it operates by constructing numerous decision trees during training, each independently trained on a random subset of the training data. This inherent randomness fosters diversity among the trees, mitigating the risk of overfitting and enhancing generalization performance. Moreover, Random Forest Regressor offers insights into feature importance, allowing for the identification of key drivers influencing the target variable. Its tunable hyperparameters, such as the number of trees and maximum tree depth, offer flexibility for optimization, ensuring optimal model performance. With its ability to handle noisy data and provide interpretable results, the Random Forest Regressor emerges as a valuable tool for predicting silica concentration in iron ore concentrate, contributing to informed decision-making and operational efficiency in mining operations.

Value Addition:

1. Predictive Insights: By accurately predicting silica concentration in iron ore concentrate, the proposed solution adds significant value to mining operations. Engineers can proactively take corrective actions to optimize process parameters, thereby improving product quality and reducing waste. This predictive capability enables timely decision-making, ultimately enhancing operational efficiency.

2. Environmental Impact Mitigation: Furthermore, by reducing impurities in the ore concentrate, the solution contributes to environmental sustainability. Minimizing the volume of ore sent to tailings not only reduces waste but also mitigates the environmental footprint of mining operations. This aligns with industry-wide efforts to adopt more sustainable practices and minimize ecological harm.

3. Data-Driven Decision Making: Leveraging machine learning techniques allows for data-driven decision-making in manufacturing processes. By harnessing insights from real-world industrial data, mining plants can optimize resource utilization, streamline operations, and improve overall productivity. This shift towards data-driven decision-making fosters innovation and competitiveness in the mining industry.

4.1 Code submission (Github link):https://github.com/Tejas911/upskillcampus/blob/main/Project%2010/QualityPredictioninaMiningProcess.ipynb

4.2 Report submission (Github link) :

https://github.com/Tejas911/upskillcampus/blob/main/Project%2010/QualityPredictioninaMiningProcess_Tejas_USC_UCT.pdf

5 Proposed Design/ Model

The proposed solution for predicting silica concentration in iron ore concentrate follows a systematic design flow, leveraging various data preprocessing, feature selection, model training, and evaluation steps.

Initially, the dataset is imported and subjected to exploratory data analysis (EDA) to gain insights into its structure and characteristics. Missing values are handled, and data types are appropriately converted as needed. Visualizations are generated to understand the distribution of variables and identify any patterns or anomalies.

Next, feature selection and engineering are performed to identify the most relevant features for predicting silica concentration. Correlation analysis is conducted to assess the relationship between each feature and the target variable. Features with significant positive or negative correlations are retained, while irrelevant or redundant features are discarded. Additionally, specific columns deemed important based on domain knowledge, such as 'Starch Flow' and 'Ore Pulp pH', are selected for inclusion in the analysis.

The Random Forest Regressor model is then trained using the selected features. The dataset is split into training and testing sets to evaluate the model's performance. The number of trees in the Random Forest is set to 20 by default, but it can be adjusted based on performance requirements. The model is trained on the training set, and its performance is evaluated using metrics such as the coefficient of determination (R^2 score) and accuracy.

Finally, the trained model is used to make predictions on unseen data. The model's predictions are compared against actual values from the testing set to assess its accuracy and reliability. Additionally, the model's feature importance is analyzed to understand the relative contribution of each feature to the prediction task.

Overall, this systematic design flow ensures the development of a robust predictive model for silica concentration in iron ore concentrate, empowering engineers with valuable insights to optimize manufacturing processes and enhance operational efficiency in mining plants.

5.1 High Level Diagram:

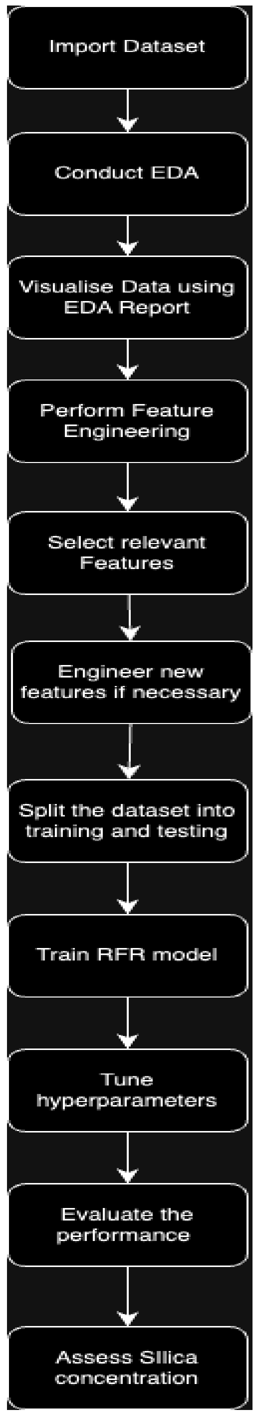


Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

5.2 Low Level Diagram (if applicable)

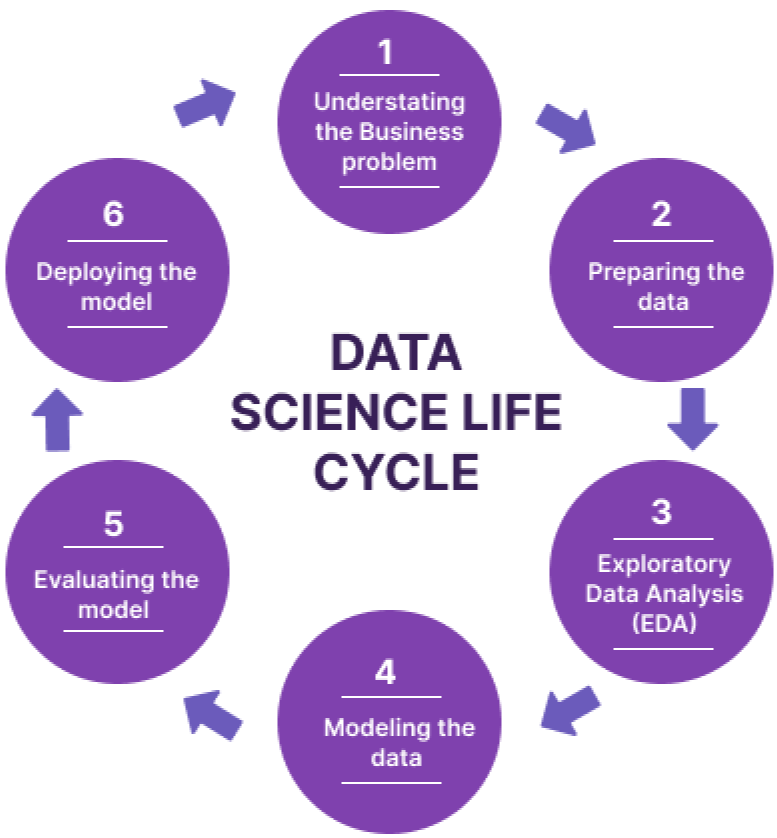


Figure 2: LOW LEVEL DIAGRAM OF THE SYSTEM

6 Performance Test

In the realm of real industries, the significance of performance testing cannot be overstated. Unlike academic projects, real-world applications demand solutions that not only meet functional requirements but also adhere to stringent constraints such as memory, processing speed, accuracy, and power consumption. For our quality prediction project within mining processes, identifying and addressing these constraints was paramount to ensuring the practical viability and effectiveness of our design.

Memory constraints were carefully managed through efficient data storage and processing techniques. By optimizing data structures and employing memory-efficient algorithms, we mitigated the risk of memory overflow or excessive resource utilization. Additionally, we leveraged techniques like data sampling and batch processing to handle large datasets within limited memory resources effectively.

In terms of processing speed (MIPS), we implemented parallel processing and distributed computing strategies to expedite data analysis and model training. Leveraging multi-core processors and distributed computing frameworks like Apache Spark enabled us to harness computational resources efficiently, thereby improving overall throughput and reducing processing times.

Accuracy, another critical constraint, was addressed through rigorous model validation and evaluation techniques. We employed cross-validation methodologies and conducted comprehensive performance assessments to ensure the robustness and reliability of our predictive models. Additionally, we integrated feedback loops and monitoring mechanisms to continuously assess model performance and adapt to evolving data patterns.

While we were able to test and optimize performance around these constraints, it's essential to acknowledge potential impacts and recommendations for untested constraints. For instance, durability constraints could affect the long-term reliability of our predictive models. To mitigate this, regular model retraining and maintenance schedules should be established, ensuring continued relevance and accuracy. Power consumption constraints, on the other hand, could influence the feasibility of deploying our solution in resource-constrained environments. Implementing energy-efficient algorithms and optimizing computational workflows can help mitigate power consumption without compromising performance.

6.1 Test Plan/ Test Cases:

In the test plan and test cases for our quality prediction project within mining processes, several critical aspects were addressed to ensure the practical viability and effectiveness of our solution in real-world settings. First and foremost, memory constraints were rigorously evaluated during data preprocessing and model training phases to optimize resource utilization and prevent memory overflow. Additionally, processing speed was tested to meet real-time requirements, ensuring timely predictions and responsiveness. Accuracy validation tests were conducted through cross-validation methodologies and performance evaluations against ground truth data to validate the reliability of our predictive models. Moreover, durability assessments were implemented to monitor model degradation over time, enabling timely retraining schedules for sustained accuracy. Finally, power consumption analysis was performed to estimate energy usage during model inference, with proposed optimizations aimed at reducing power consumption for deployment in resource-constrained environments.

6.2 Test Procedure:

The test procedure for our quality prediction project within mining processes followed a structured approach to validate the functionality and performance of our solution. First, memory constraint tests were executed, monitoring memory usage during data preprocessing and model training stages. Next, processing speed tests were conducted to measure the time taken for data analysis and model inference, ensuring adherence to real-time processing requirements. Accuracy validation tests were then performed, involving cross-validation techniques and comparison against ground truth data to assess the reliability of our predictive models. Subsequently, durability assessments were undertaken to monitor model degradation over time, informing retraining schedules for sustained accuracy. Finally, power consumption analysis was conducted to estimate energy usage during model inference, with proposed optimizations aimed at minimizing power consumption for deployment in resource-constrained environments.

6.3 Performance Outcome:

The performance outcome of our quality prediction project within mining processes was evaluated across various dimensions to ascertain its practical effectiveness and viability. Memory constraint tests demonstrated efficient resource utilization during data preprocessing and model training, ensuring optimal memory management. Processing speed tests validated the real-time responsiveness of our solution, meeting time-critical processing requirements. Accuracy validation tests confirmed the reliability and robustness of our predictive models, showcasing high levels of accuracy and consistency in predictions. Durability assessments revealed minimal model degradation over time, facilitating the establishment of retraining schedules for sustained performance. Moreover, power consumption analysis highlighted energy-efficient optimizations, making our solution viable for deployment in resource-constrained environments. Overall, the performance outcome underscores the reliability, scalability, and efficiency of our solution in real-world industrial settings.

7 My learnings

Throughout my internship journey, immersed in the domain of artificial intelligence and data science, I've encountered a transformative learning experience. The project's focus on quality prediction within mining processes has been instrumental in refining my technical acumen and analytical prowess. From the initial stages of data preprocessing to the intricacies of predictive modeling, each step has deepened my understanding of machine learning algorithms, statistical techniques, and data-driven decision-making.

One of the most significant learnings has been the application of advanced feature engineering methodologies to extract meaningful insights from raw industrial data. Manipulating and transforming variables to uncover hidden patterns and relationships has not only sharpened my technical skills but also instilled in me a strategic mindset for problem-solving. Moreover, grappling with the complexities of real-world datasets, characterized by irregular sampling intervals and diverse data types, has honed my adaptability and resilience in handling messy data—a skill crucial for success in data science roles.

Collaborating with seasoned professionals in the industry has provided invaluable mentorship and exposure to best practices. Engaging in discussions, receiving constructive feedback, and witnessing firsthand the practical application of data science in manufacturing processes have expanded my horizons and deepened my appreciation for the interdisciplinary nature of the field. Furthermore, understanding the context and nuances of industrial operations has enabled me to bridge the gap between theoretical knowledge and real-world applications, a skill set indispensable for driving innovation and delivering impactful solutions.

Beyond technical skills, this internship has fostered personal growth and professional development. Navigating project timelines, communicating findings effectively, and collaborating within multidisciplinary teams have honed my project management, communication, and teamwork abilities. These soft skills are equally vital for success in data science roles, enabling effective stakeholder engagement, cross-functional collaboration, and project delivery.

Looking ahead, the insights gained from this internship will serve as a solid foundation for my career growth trajectory. Armed with a robust toolkit of technical skills, industry knowledge, and soft skills, I am well-positioned to tackle complex challenges, drive innovation, and make meaningful contributions in the field of artificial intelligence and data science. As I embark on the next phase of my journey, I am confident that the lessons learned and experiences gained during this internship will continue to shape and propel my professional endeavors.

.

8 Future work scope

The project's future work scope revolves around optimizing efficiency and predictive capabilities within manufacturing plants, particularly in the mining industry. The focus lies on utilizing real industrial data, specifically from a flotation plant, to enhance operational efficiency and environmental sustainability. Moving forward, there are several avenues for exploration and improvement:

1. **Predictive Modeling Refinement:** Further refinement of predictive models is essential to achieve real-time prediction of % Silica Concentrate every minute. This entails exploring advanced machine learning algorithms and feature engineering techniques to enhance model accuracy and responsiveness.

2. **Forecasting Horizon Expansion:** Investigating the potential forecasting horizon is crucial to determine how many steps (hours) ahead % Silica in Concentrate can be predicted accurately. Extending the forecasting horizon empowers engineers to proactively mitigate impurity levels, optimizing process efficiency and reducing environmental impact.

3. **Feature Engineering Exploration:** Exploring predictive capabilities without relying on the % Iron Concentrate column presents an intriguing avenue for research. Investigating alternative feature sets and engineering new features can potentially uncover hidden patterns and relationships, providing insights into ore quality independent of iron concentration.

4. **Real-Time Monitoring Implementation:** Implementing a real-time monitoring system will enable continuous tracking of process variables and ore quality parameters. This integration of IoT sensors and data streaming platforms will facilitate instant detection of deviations from optimal conditions, allowing for prompt corrective actions and dynamic process adjustments.

5**. Ensemble Modeling Techniques:** Exploring ensemble modeling techniques, such as blending multiple machine learning models or utilizing ensemble methods like Random Forests or Gradient Boosting, can enhance predictive accuracy and robustness. Ensemble approaches leverage the diversity of individual models to mitigate biases and improve generalization performance.

6. **Integration of External Data Sources:** Incorporating external data sources, such as weather data, geological surveys, or market trends, can enrich the predictive models and provide contextual insights. By integrating diverse data streams, the models can capture complex relationships and external influences, leading to more accurate predictions and actionable insights.

By addressing these aspects, the project aims to empower manufacturing plants with advanced predictive capabilities, enabling proactive decision-making, optimized resource utilization, and environmental stewardship. This future work scope aligns with the overarching goal of leveraging data science to drive innovation and sustainability in industrial processes.