Industrial Internship Report on

” Quality Prediction in a Mining Process”

Prepared by

Mohammed Sardar Ali

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 “. In a mining process project, the prediction of quality entails utilizing data and predictive analysis to estimate the grade or quality of extracted materials. This aids mining operations, in optimizing the allocation of resources minimizing waste enhancing product quality, and maximizing profitability. The process encompasses data collection, preprocessing, feature development, building machine learning models, training and evaluating them deploying them into operation and constantly improving them. By employing models informed decisions can be made regarding resource extraction and processing to achieve improved efficiency and economic advantages. 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

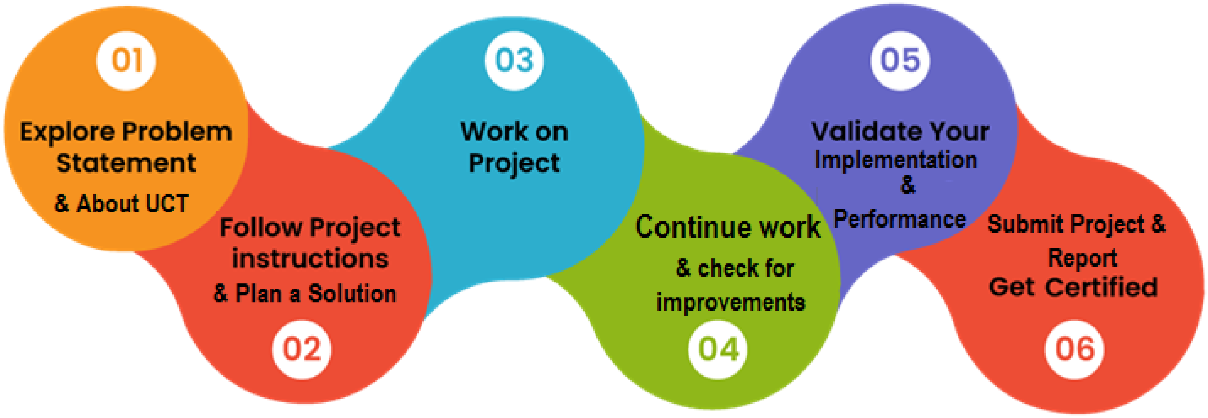
1 Preface

Data science and machine learning play roles, in today's data-oriented world. Data science encompasses the process of handling data from gathering and cleaning it to analyzing and visualizing it with the aim of extracting insights. It combines techniques and domain expertise to tackle problems and assist decision-making across various industries. On the hand machine learning is a branch of intelligence that enables computers to learn from data and independently make predictions or decisions. It encompasses unsupervised and reinforcement learning methods that allow computers to identify patterns categorize information and solve tasks. Together data science and machine learning work hand in hand to drive innovation in areas, like healthcare, finance, and marketing among others by transforming data into knowledge and predictive capabilities.

In today’s tech driven world having an internship, in data science and machine learning is becoming more and more important for advancing your career. These areas are leading the way in innovation. Have an impact, on multiple industries. By undertaking a pertinent internship, ambitious individuals have the opportunity to acquire hands-on experience in the realms of data collection, preprocessing, modeling, and analysis. Submersed in authentic projects, they refine their aptitude for troubleshooting and gain valuable decision-making skills guided by empirical evidence. Moreover, these internships provide access to state-of-the-art tools and technologies, cultivating a profound comprehension of the dynamic landscape inherent to the field. Furthermore, an internship of this nature engenders networking prospects with seasoned professionals, opening doors to mentorship and potential employment invitations. In essence, a data science and machine learning internship not only embellish one's curriculum vitae but also endows them with practical knowledge and industry acumen indispensable for a triumphant career in these highly sought-after and gratifying disciplines.

USC and UCT are both esteemed institutions, but the precise scope of the opportunity you're alluding to remains shrouded in mystery. These universities present a cornucopia of programs and prospects for students, researchers, and professionals alike. In order to furnish a bespoke response, I require additional particulars regarding the specific prospect, be it a program, scholarship, research position, or initiative that has piqued your curiosity at either the University of Southern California (USC) or the University of Cape Town (UCT). Kindly furnish me with more information, and I shall gladly dispense further elucidation or enlightenment.

Program was planned in well and proper manner where we clarified all our doubts regarding report work etc. And mentors were helping and giving quick responses.



I have learned many new things and topics regarding Data Science and machine learning.

Thanks to all (Nitish sir, Nitin Tyagi sir, and other staff members), who have helped me directly or indirectly.

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.

uct

Uniconverge Technologies

IIOT Products

We offer product ranging from Remote IOs, Wireless IOs, LoRaWAN Sensor Nodes/ Gateways, Signal converter and IoT gateways

IIOT Solutions

We offer solutions like OEE, Predictive Maintenance, LoRaWAN based Remote Monitoring, IoT Platform, Business Intelligence…

OEM Services

We offer solutions ranging from product design to final production we handle everything for you..

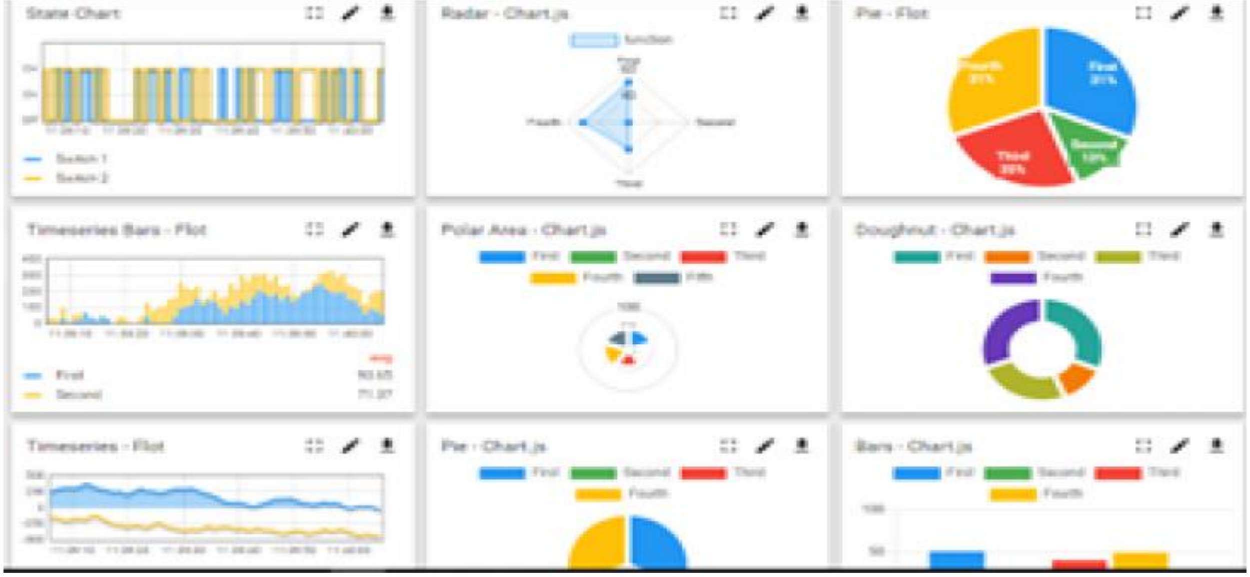
i. UCT iot Platform (uct Insight)

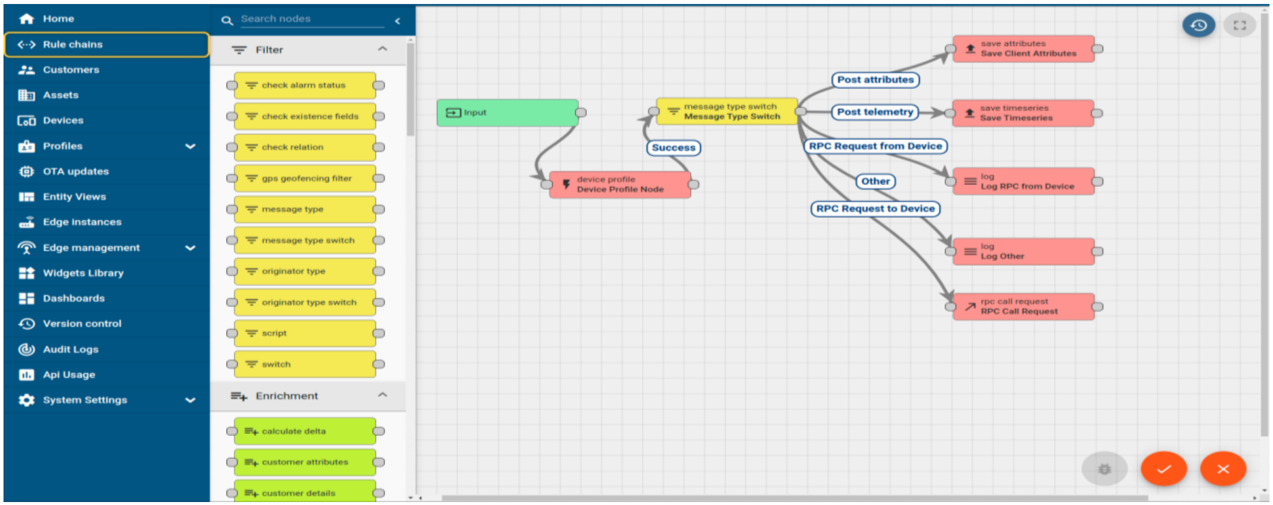
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable “insight” for your process/business. It has been built in Java for backend and reactjs for Front end. It has support for mysql and various nosql Databases.

- 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



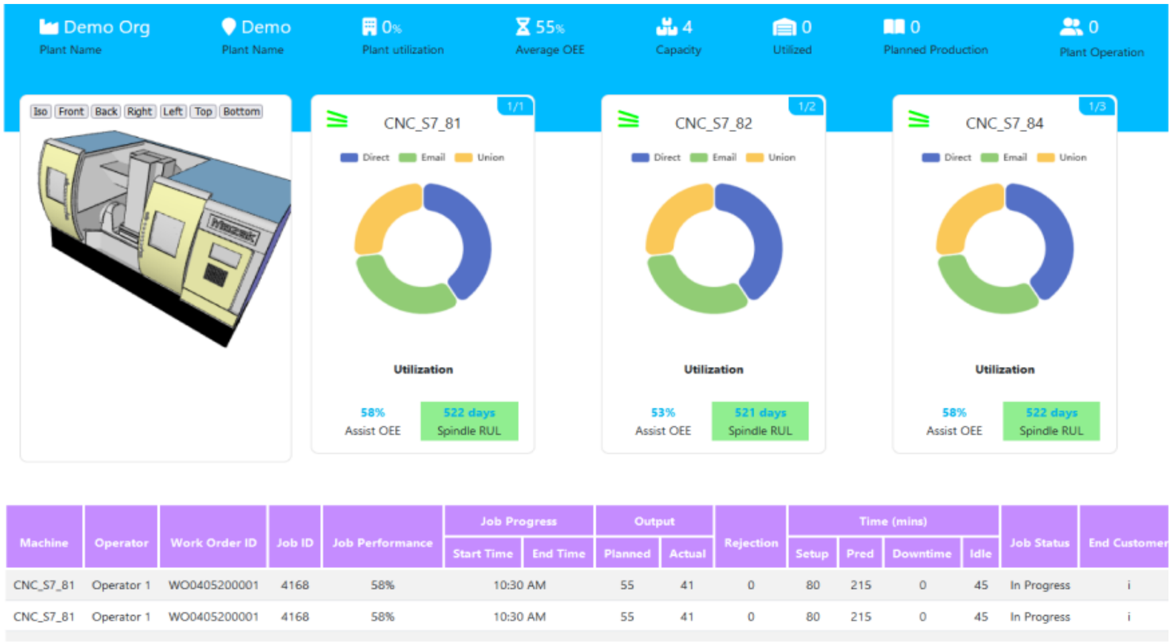


ii. Smart Factory Platform (FACTORY WATCH)

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.



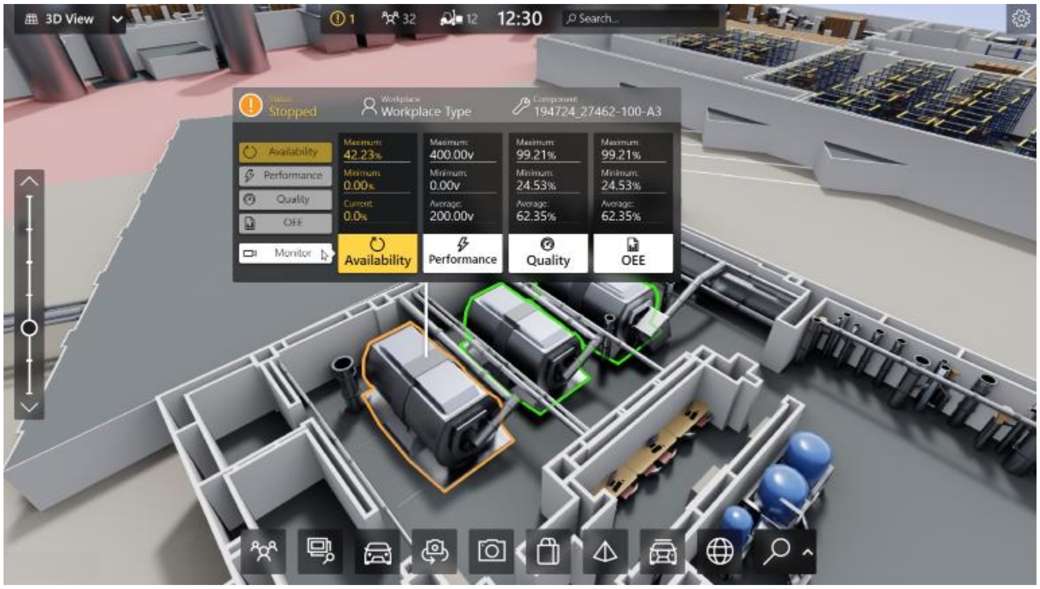
Demo Org Plant Name
Demo Plant Name
0% Plant utilization
55% Average OEE
4 Capacity
0 Utilized
0 Planned Production
0 Plant Operation

CNC_S7_81: 58% Assist OEE, 522 days Spindle RUL

CNC_S7_82: 53% Assist OEE, 521 days Spindle RUL

CNC_S7_84: 58% Assist OEE, 522 days Spindle RUL

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

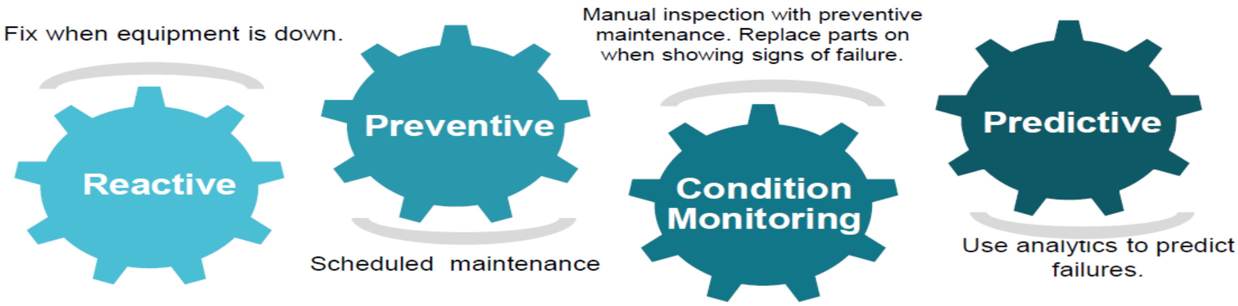


iii. LoRaWAN™ based Solution

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

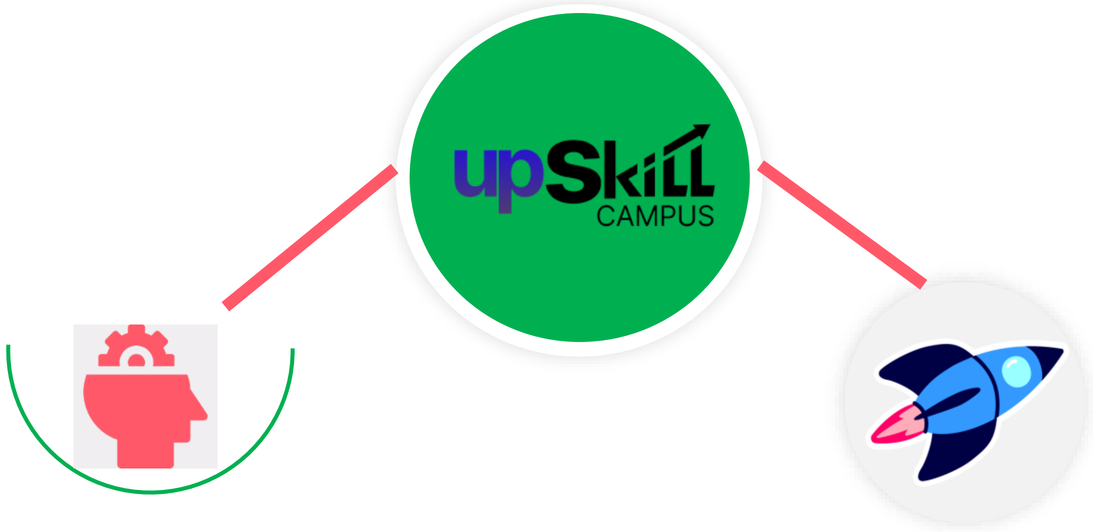
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial iot and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.


Fix when equipment is down.
Reactive
Preventive
Scheduled maintenance
Manual inspection with preventive maintenance. Replace parts on when showing signs of failure.
Condition Monitoring
Predictive
Use analytics to predict failures.

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.

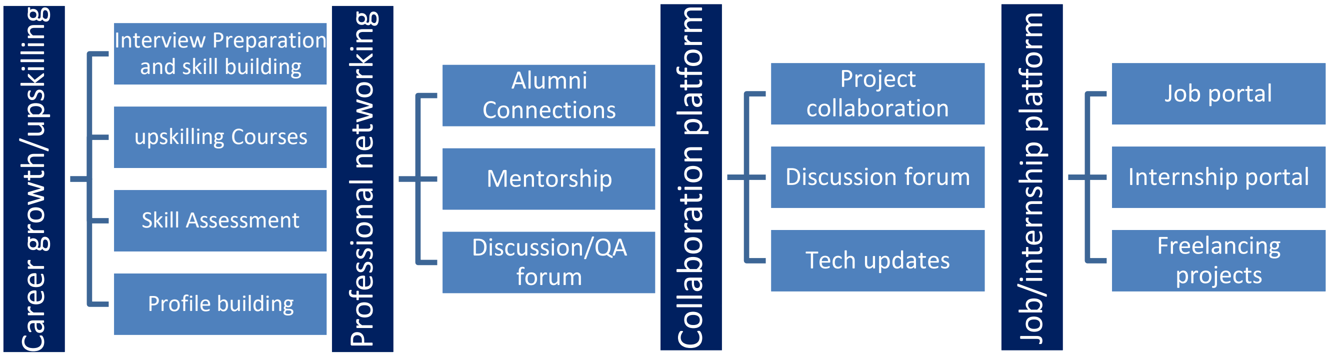
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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://learn.upskillcampus.com/s/mycourses

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

[3]https://upskillcourses.com/

3 Problem Statement

- Predicting Variability, in Ore Grade; “Create a model that can estimate the fluctuation in ore grade within a mining site over time. This will help with resource planning and management.”
- Real time Monitoring of Ore Quality; "Develop a system that continuously monitors and provides real time predictions about the quality of ore. This will enable adjustments to processing or extraction techniques."
- Enhancing Product Quality through Blending Optimization; "Optimize the blending process for ores sourced from locations to achieve the desired product quality while minimizing processing costs and waste generation."
- Early Detection of Impurities; "Design algorithms that can detect and predict the presence of impurities or contaminants in materials early as possible during the extraction process. This will help reduce processing costs."
- Improving Energy Efficiency and Quality Prediction; "Integrate ore quality prediction with energy consumption data to optimize mining and processing operations considering both product quality and energy efficiency."
- Ensuring Quality throughout the Supply Chain; "Implement models for quality assurance across the supply chain ensuring consistent adherence to quality specifications for the final product."
- Maintaining Quality Control through Predictive Maintenance; "Utilize predictive maintenance techniques to ensure optimal conditions for mining equipment preventing breakdowns that could lead to variations, in ore quality."
- Mitigating Environmental Impact; "We should create models that can predict the consequences of mining activities, such, as their impact on water quality and soil contamination."
- Forecasting Market Demand; "To optimize production schedules and manage inventory effectively we should combine predictions about ore quality with

forecasts, on market demand. This will help us meet the quality requirements of the market."

4 Existing and Proposed solution

Existing Solution:

Mining endeavors have banked on acquiring samples manually and subjecting them to laboratory appraisal to ascertain the caliber of extracted substances. Samples are amassed at diverse junctures in the procedure, and their chemical constitution is scrutinized. Although precise, this technique consumes copious time, incurs substantial expenses, and yields belated outcomes, rendering it less appropriate for instantaneous decision-making.

Proposed Solution:

Machine Learning-Based Predictive Models:

Implement machine learning algorithms to predict the quality of mined materials in real-time or near real-time. These models utilize historical data, sensor readings, and various input features to estimate the quality of the ore as it is extracted.

This approach offers several advantages:

Data Integration:

Incorporate data from various sources, including geological surveys, drilling data, sensor data from mining equipment, and historical assay results, to create a comprehensive dataset for analysis.

Predictive Modeling:

Utilize supervised learning techniques such as regression or classification to build predictive models that can estimate ore grade, impurities, or other quality-related parameters.

Real-time Monitoring:

Implement continuous monitoring of key parameters and sensor data in real-time to provide immediate feedback on the quality of extracted materials.

Automation:

Integrate these models into the mining process automation systems to enable automated decision-making and control of equipment to optimize ore extraction.

Early Warning Systems:

Develop algorithms that can detect deviations from expected quality standards early in the extraction process, allowing for corrective actions to be taken promptly.

Continuous Improvement:

Implement feedback loops that allow the predictive models to improve over time as more data becomes available and as the models learn from their predictions.

Benefits:

Cost Reduction:

By minimizing the need for manual sampling and laboratory assays, the proposed solution reduces operational costs associated with quality control.

Improved Efficiency:

Real-time quality prediction allows mining operations to make rapid adjustments, optimizing the extraction process to maximize the yield of high-quality ore.

Environmental Impact:

Minimizing the extraction of lower-grade material reduces waste and the environmental footprint of mining operations.

Resource Allocation:

Predictive models assist in directing high-quality ore to the appropriate processing streams, maximizing resource allocation and profitability.

Safety:

Early detection of impurities or deviations from quality standards enhances safety by reducing the likelihood of equipment failure.

4.1 Code submission (GitHub link):

https://github.com/sardar2001/upskillcampus

4.2 Report submission (GitHub link):

https://github.com/sardar2001/upskillcampus

5 Proposed Design/ Model

a) Data Collection:

Gather historical data from various sources, including geological surveys, drilling data, sensor readings from mining equipment, assay results, and environmental data. Ensure data consistency and quality.

b) Data Preprocessing:

Clean and preprocess the data by handling missing values, outliers, and inconsistencies. Normalize or standardize numerical features as needed. Encode categorical variables if applicable. Explore data to understand its characteristics and identify potential correlations.

c) Feature Engineering:

Create relevant features that can influence ore quality predictions, such as geological parameters, ore type, processing conditions, and historical performance data. Select features using techniques like feature importance ranking.**d**

d) Model Selection:

Choose appropriate machine learning algorithms for quality prediction. Common choices include regression for continuous predictions or classification for discrete quality categories. Consider ensemble methods, deep learning models, or specialized algorithms tailored to mining data, depending on the complexity of the problem.

e) Data Splitting:

Split the dataset into training, validation, and test sets. Use cross-validation techniques to ensure model robustness.

f) Model Training:

Train the selected model(s) on the training data, optimizing for performance metrics relevant to the specific quality prediction task (e.g., mean squared error, accuracy, F1-score).

g) Model Evaluation:

Assess the model's performance using the validation dataset, considering metrics such as root mean squared error (RMSE) or mean absolute error (MAE) for regression, or precision, recall, and F1-score for classification. Adjust hyperparameters and model architecture if needed.

h) Real-time Data Integration:

Implement a system for real-time data integration, where sensor data and other relevant information are continuously fed into the model for quality predictions.

i) Quality Thresholds:

Define quality thresholds or specifications for the desired output. For example, specify the acceptable range of ore grades or quality categories.

j) Deployment:

Integrate the trained model into the mining process automation system or control center. Implement an alerting system to notify operators when quality predictions deviate from specified thresholds.

k) Continuous Improvement:

Monitor the model's performance in production and update it periodically with new data to ensure accuracy and adaptability to changing conditions.

l) Visualization and Reporting:

Develop visualization tools and reports to present quality predictions and historical trends to mining operators and decision-makers.

m) Compliance and Documentation:

Ensure that the quality prediction system complies with regulatory standards and document the model's development, training, and deployment processes for auditing purposes.

n) Maintenance and Support:

Provide ongoing maintenance and support for the quality prediction system, addressing issues and adapting to evolving requirements.

5.1 High Level Diagram:

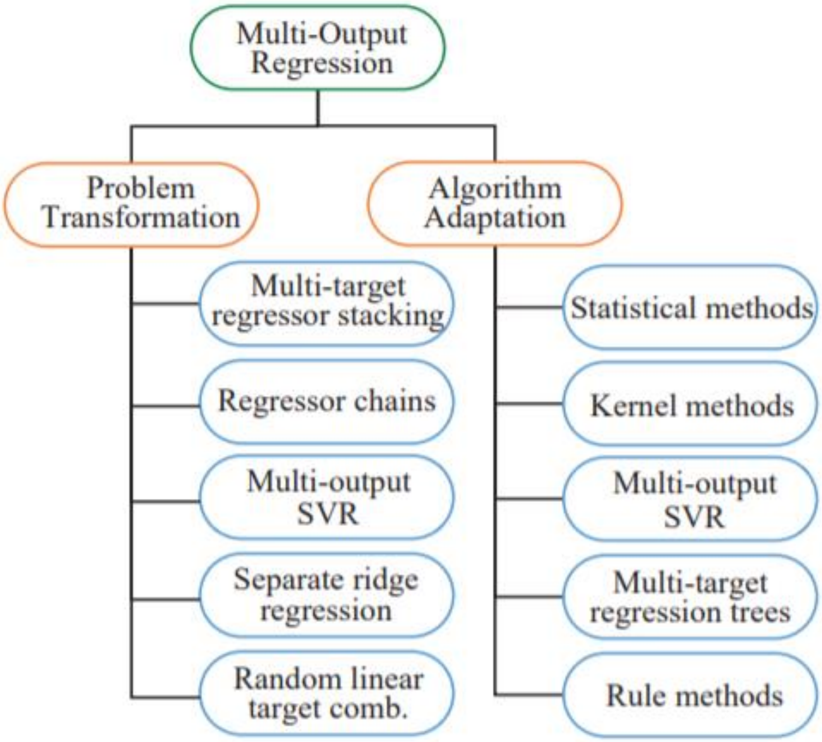
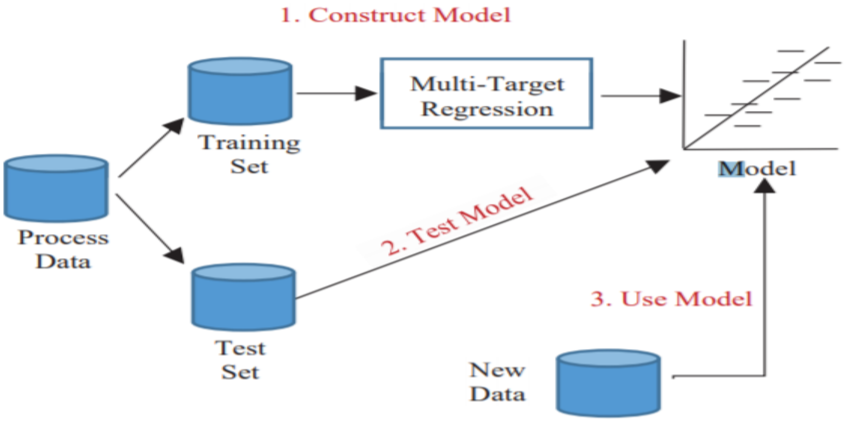


Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

5.2 Low Level Diagram:



5.3 Interfaces:

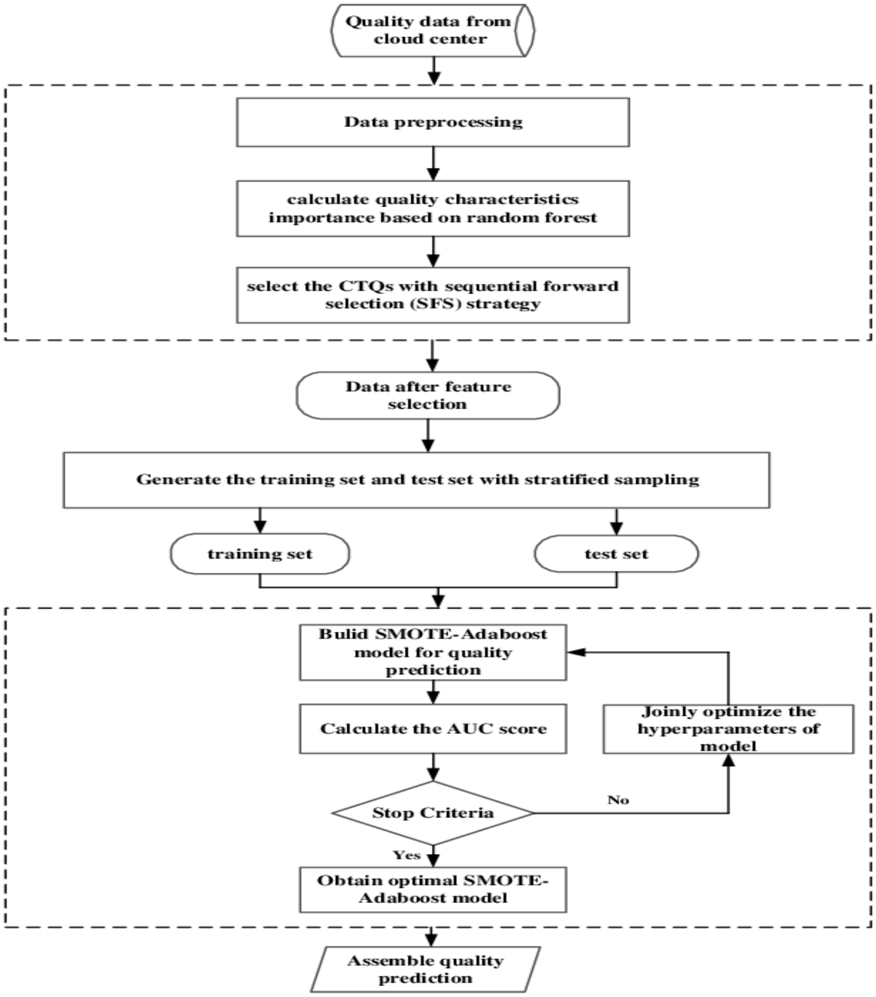

Quality data from cloud center
Data preprocessing
calculate quality characteristics importance based on random forest
select the CTQs with sequential forward selection (SFS) strategy
Data after feature selection
Generate the training set and test set with stratified sampling
training set
test set
Bulid SMOTE-Adaboost model for quality prediction
Calculate the AUC score
Joinly optimize the hyperparameters of model
Stop Criteria
No
Yes
Obtain optimal SMOTE-Adaboost model
Assemble quality prediction

Figure 2: Flow Chart of Mining

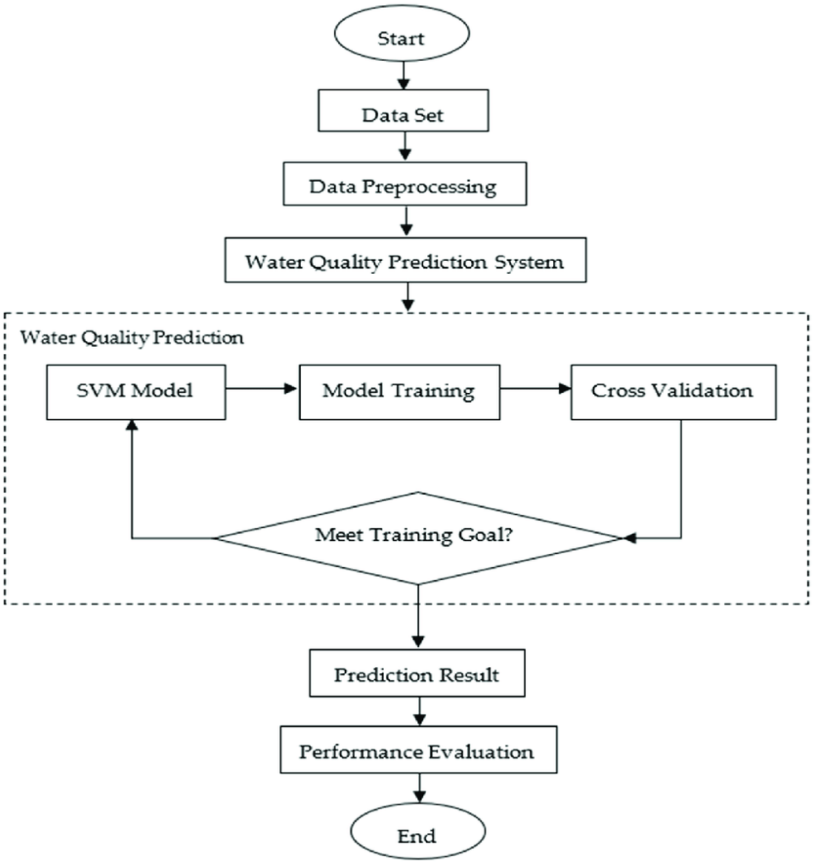


Figure 3: Flow Chart of Prediction water Mining

6 Performance Test

Performance testing of a quality prediction system in a mining process is crucial to ensure that the system meets its objectives, functions effectively, and provides accurate and reliable predictions.

The steps and considerations for conducting performance testing:

1. Define Testing Objectives:

- Clearly define the goals and objectives of the performance testing. Identify what specific aspects of the quality prediction system you want to evaluate, such as prediction accuracy, response times, scalability, and stability.

2. Select Performance Metrics:

- Choose appropriate performance metrics based on the objectives. Common metrics for quality prediction systems may include:

  - Prediction Accuracy: Measures how well the model predicts actual quality.

  - Response Time: Evaluates the time it takes for the system to generate predictions.

  - Throughput: Assesses the number of predictions the system can handle per unit of time.

  - Scalability: Determines how well the system scales as the workload increases.

  - Stability: Measures the system's ability to perform consistently over time.

3. Data Preparation:

- Prepare a representative dataset that reflects the range of data the system will encounter during actual operation. This dataset should include various ore grades, environmental conditions, and operational scenarios.

4. Test Scenarios:

- Define test scenarios that mimic real-world situations and usage patterns. For example, simulate different ore compositions, variations in sensor data, and changes in process parameters.

5. Load Testing:

- Perform load testing to assess how the system behaves under heavy workloads. This helps identify performance bottlenecks and scalability issues. Gradually increase the load until the system reaches its limits or performance degrades.

6. Stress Testing:

- Conduct stress testing to determine the system's breaking point. Push the system beyond its expected capacity to identify failure points and understand how it recovers from extreme conditions.

7. Accuracy Testing:

- Evaluate the accuracy of the quality predictions by comparing them to known ground truth data. Use appropriate statistical measures like mean squared error (MSE), root mean squared error (RMSE), or classification metrics (e.g., precision, recall, F1-score) for regression and classification tasks, respectively.

8. Response Time Testing:

- Measure the response times for generating predictions under different load conditions. Ensure that response times meet the defined requirements, especially for real-time applications.

9. Scalability Testing:

- Assess how the system scales as the workload increases. Determine the system's capacity to handle additional data volume, sensors, or users.

10.Stability and Reliability Testing:

- Continuously monitor the system over an extended period to identify any memory leaks, resource leaks, or performance degradation over time. Verify that the system operates reliably without unexpected crashes or failures.

11. Reporting and Analysis:

- Generate comprehensive performance reports that document test results, including performance metrics, findings, and any issues encountered. Analyze the data to draw conclusions and make recommendations for improvements.

12. Tuning and Optimization:

- Based on the test results, fine-tune the system, optimize algorithms, and address any identified performance bottlenecks. Repeat testing as needed to validate improvements.

13. Regression Testing:

- Implement a regression testing strategy to ensure that system updates and changes do not negatively impact performance compared to previous versions.

14. Documentation:

- Maintain detailed documentation of the performance testing process, including test plans, test cases, results, and any remediation actions taken.

15. User Acceptance Testing:

- If applicable, involve end-users or stakeholders in acceptance testing to validate that the system's performance aligns with their expectations and requirements.

6.1 Test Plan/ Test Cases

1. Introduction:

- Overview of the quality prediction system.
- Objectives of testing.
- Scope and limitations.

2. Test Environment:

- Hardware and software configurations.
- Data sources and datasets.
- Tools and testing frameworks.

3. Test Objectives:

- Define the primary goals of the testing effort.
- Identify key performance metrics and quality criteria.

4. Test Scenarios:

- Define different scenarios representing various mining conditions and data inputs.

5. Test Cases:

a. Data Preprocessing and Integration:

1. Test data cleaning and validation.
2. Verify the handling of missing values.
3. Test data transformation and normalization.
4. Validate real-time data integration with sensor feeds.

b. Model Training and Evaluation

5. Train the model with historical data.
6. Evaluate model performance using cross-validation.
7. Verify that the model selection process aligns with predefined criteria.
8. Test hyperparameter tuning and optimization

c. Real-time Predictions:

9. Validate real-time predictions against known ground truth data.
10. Check the responsiveness of the system to incoming sensor data.

11. Ensure that predictions meet predefined quality thresholds.

d. Load and Scalability Testing

12. Assess system performance under increasing data loads.

13. Verify scalability by adding more sensors or data sources.

14. Determine the point at which the system starts to degrade in performance.

e. Stress Testing:

15. Push the system to its limits to identify breaking points.

16. Verify system recovery after stress conditions.

f. Accuracy Testing

17. Calculate prediction accuracy metrics (e.g., RMSE, MAE) for regression tasks.

18. Evaluate classification metrics (e.g., precision, recall, F1-score) for classification tasks.

g. Response Time Testing:

19. Measure and validate response times for generating predictions under various loads.

20. Ensure response times meet defined requirements, especially for real-time applications.

h. Stability and Reliability Testing:

21. Monitor the system for memory leaks or resource issues over an extended period.

22. Verify that the system operates reliably without unexpected crashes or performance degradation.

i. Compliance Testing:

23. Check that the system generates reports compliant with regulatory standards.

24. Verify that data is recorded appropriately for compliance audits.

6. Test Deliverables:

- Comprehensive test reports for each test case.

- Performance metrics and findings.

- Remediation actions and optimization recommendations.

7. Acceptance Criteria:

- Define criteria for passing each test case.

- Specify the overall acceptance criteria for the quality prediction system.

8. Risks and Mitigation:

- Identify potential risks related to testing and address mitigation strategies.

9. Test Schedule

- Define the timeline for executing test cases.

10. Test Team Responsibilities:

- Assign responsibilities to team members for executing and documenting tests.

11. Dependencies:

- Identify any dependencies on external systems, data sources, or stakeholders.

12. Test Sign-Off:

- Define the criteria for test sign-off and approval.

13. Additional Considerations:

- Consider user acceptance testing (UAT) involving stakeholders if applicable.

14. Test Closure:

- Summarize test results and findings.

- Provide recommendations for system improvements and optimizations.

6.2 Test Procedure

Objective:

The objective of this test case is to [state the specific objective, e.g., verify the accuracy of real-time quality predictions under varying sensor data inputs].

Preconditions:

1. The quality prediction system is installed and configured in the test environment.

2. Relevant datasets, including historical data and real-time sensor data, are available.

3. Quality thresholds and acceptance criteria are defined.

Test Steps:

1. Data Setup:

- Ensure that the system has access to the required datasets, including historical data and real-time sensor data.

- Set up the system to use the specified input datasets.

2. Configure Test Scenario:

- Select the appropriate test scenario that represents the desired conditions, such as varying ore compositions or environmental factors.

3. Execute Test Scenario:

- Trigger the test scenario to simulate real-world conditions.
- Provide the system with the selected input datasets and conditions.

4. Monitor Predictions:

- Continuously monitor the system's predictions in real-time.
- Record the predicted quality values and any alerts or notifications generated.

5. Compare Predictions to Ground Truth:

- Refer to the ground truth data or known quality measurements for the test scenario.
- Compare the system's real-time predictions to the actual quality values.

6. Calculate Performance Metrics:

- Calculate relevant performance metrics, such as RMSE, MAE, or classification metrics (e.g., precision, recall, F1-score), depending on the nature of the test.

7. Verify Compliance:

- Check that the system's predictions meet the predefined quality thresholds and acceptance criteria.

8. Record Results:

- Record the test results, including:
  - Predicted quality values.
  - Actual quality values (ground truth).
  - Performance metrics.

9. Report Deviations:

- If the system's predictions deviate significantly from the ground truth or if compliance criteria are not met, report these deviations.

10. Cleanup:

- Reset the system and test environment to their initial states.

Postconditions:

1. Test results and findings are documented.

2. Any identified deviations or issues are reported to relevant stakeholders.

3. The system and test environment are in their initial states.

Test Data:

Quality prediction in a mining process consists of datasets used to assess the performance and accuracy of the prediction system. This data typically includes historical mining data, sensor readings, ore composition information, and ground truth quality measurements. Test data serves as the input for various performance tests, allowing the evaluation of how well the system predicts the quality of mined materials under different conditions. It is essential to have representative and diverse test data to ensure the system's reliability and effectiveness in real-world mining operations.

Expected Results:

- The system's predictions closely match the actual quality values within the defined acceptable error margins.

- Compliance criteria are met.

- No significant issues or deviations are observed.

Test Conclusion:

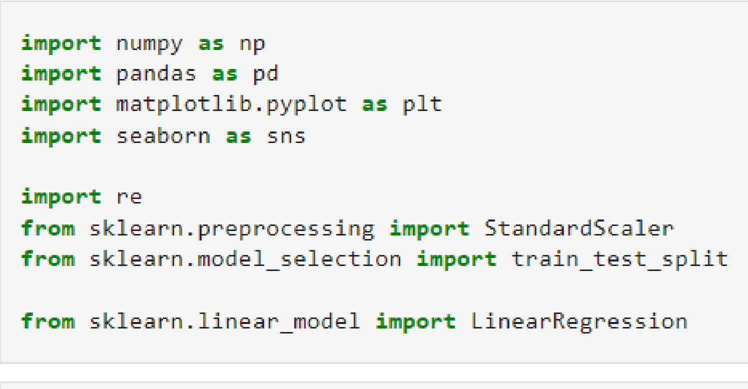
Quality prediction in a mining process involves summarizing the outcomes and findings of the performance testing and evaluation conducted on the prediction system. It provides a concise assessment of whether the system meets its performance objectives and quality standards. A short summary of the test conclusion typically highlights whether the system's predictions align with the expected quality standards, if it operates within acceptable response times, and if it exhibits scalability and stability under varying conditions. The test conclusion serves as a crucial decision point for stakeholders, indicating whether the quality prediction system is ready for deployment in real mining operations or if further improvements and optimizations are necessary.

Test Sign-off:

This test procedure provides a structured approach to executing a specific test case within the quality prediction system. You can adapt and customize this template for different test scenarios, ensuring thorough testing of your system's functionality and performance.

6.3 Performance Outcome

1) Installing Libraries

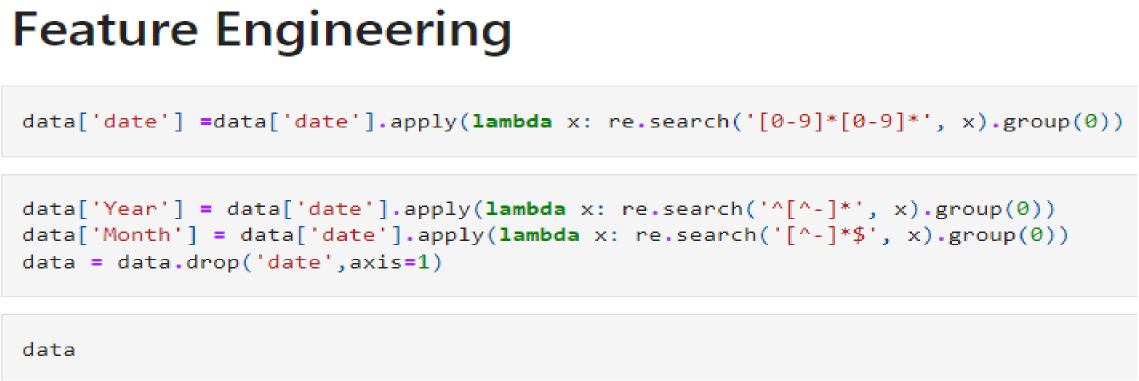


```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

import re
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split

from sklearn.linear_model import LinearRegression
```

2) Feature Engineering Data



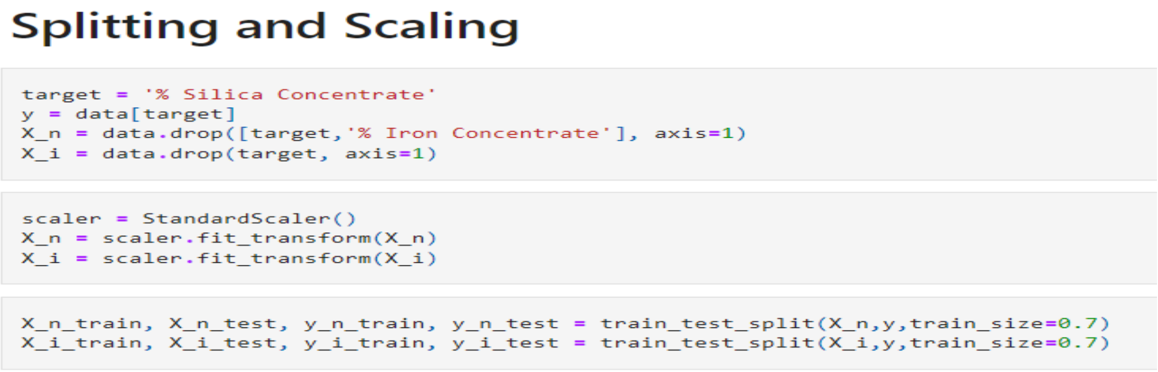
Feature Engineering

```
data['date'] =data['date'].apply(lambda x: re.search('[0-9]*[0-9]*', x).group(0))
```

```
data['Year'] = data['date'].apply(lambda x: re.search('^[^-]*', x).group(0))
data['Month'] = data['date'].apply(lambda x: re.search('[^-]*$', x).group(0))
data = data.drop('date',axis=1)
```

```
data
```

3) Splitting and Scaling the Data



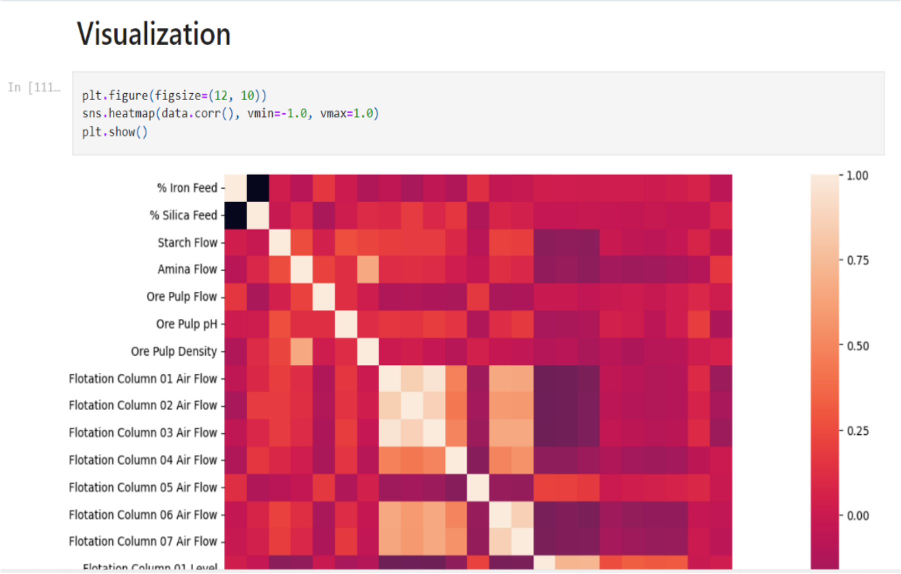
Splitting and Scaling

```
target = '% Silica Concentrate'
y = data[target]
X_n = data.drop([target,'% Iron Concentrate'], axis=1)
X_i = data.drop(target, axis=1)
```

```
scaler = StandardScaler()
X_n = scaler.fit_transform(X_n)
X_i = scaler.fit_transform(X_i)
```

```
X_n_train, X_n_test, y_n_train, y_n_test = train_test_split(X_n,y,train_size=0.7)
X_i_train, X_i_test, y_i_train, y_i_test = train_test_split(X_i,y,train_size=0.7)
```

4) Visualization of Data

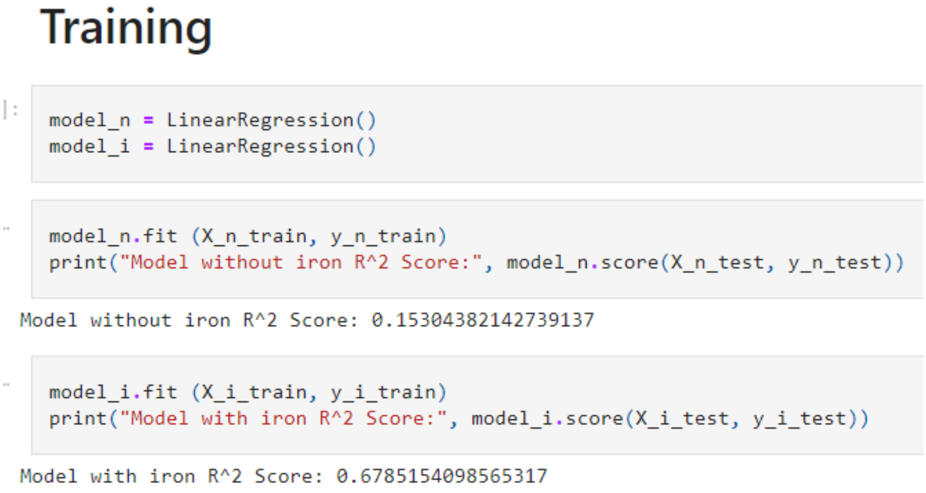

Visualization

In [111…

```
plt.figure(figsize=(12, 10))
sns.heatmap(data.corr(), vmin=-1.0, vmax=1.0)
plt.show()
```

% Iron Feed
% Silica Feed
Starch Flow
Amina Flow
Ore Pulp Flow
Ore Pulp pH
Ore Pulp Density
Flotation Column 01 Air Flow
Flotation Column 02 Air Flow
Flotation Column 03 Air Flow
Flotation Column 04 Air Flow
Flotation Column 05 Air Flow
Flotation Column 06 Air Flow
Flotation Column 07 Air Flow
Flotation Column 01 Level
1.00
0.75
0.50
0.25
0.00

5) Training of Data and getting Approx. Output


Training

```
model_n = LinearRegression()
model_i = LinearRegression()
```

```
model_n.fit (X_n_train, y_n_train)
print("Model without iron R^2 Score:", model_n.score(X_n_test, y_n_test))
```

```
Model without iron R^2 Score: 0.15304382142739137
```

```
model_i.fit (X_i_train, y_i_train)
print("Model with iron R^2 Score:", model_i.score(X_i_test, y_i_test))
```

```
Model with iron R^2 Score: 0.6785154098565317
```

Getting Approx. Value: 0.678515

7 My learnings

I learned about the python programming and its various implementation in areas like Artificial Intelligence, Data Science and Machine Learning. Also, I studied about the Data Science and Machine learning and get to know about the different algorithms used in this field that can be used for the solution of the project. We get to know the importance of Machine learning in today's world. We get to study about various algorithm that can be implemented in this project. I learned how to apply them and how its implementation can be done, and which will give the best result on implementation. I learned how to apply them and how its implementation can be done, and which will give the best result on implementation. I get to used panda's library in python to read the data and perform data cleaning, then I used scikit-learn library of machine learning to split train and test data and then apply the machine learning algorithm and I have used matplotlib and seaborn for graphical representation of data to analyze and understand it.

This 6 Week Internship program with upSkill Campus and UniConverge Technologies Pvt. Ltd. was very much helpful. I learned so much about Python, Machine learning, Data Science and get to work on a real-world project. It was a great experience which will help me get ahead in my career in future. Thanks to upSkill Campus and UCT for giving us this opportunity.

8 Future work scope

a) **Feature engineering:**

The performance of decision trees can be improved by processing the input data beforehand and creating useful characteristics. You could design elements like time of day, day of the week, holiday, and seasonality indicators for traffic predictions. Additionally, take into account

including traffic-related elements as traffic jams, collisions, and construction activities. The decision tree algorithm can better anticipate outcomes by capturing complicated relationships through feature engineering.

b) **AI and Automation:**

The integration of artificial intelligence (AI) and machine learning into automation processes will continue to expand. This includes automation in manufacturing, customer service, finance, and healthcare, creating job opportunities in AI model development and deployment.

c) **Healthcare and Biomedicine:**

Data science and machine learning will play a significant role in personalized medicine, drug discovery, genomics, and healthcare analytics. Professionals in these fields will be in high demand.

d) **Natural Language Processing (NLP):**

NLP applications, including chatbots, language translation, and sentiment analysis, will continue to grow. NLP specialists and researchers will be needed to advance these technologies.

e) **Computer Vision:**

The application of computer vision in autonomous vehicles, facial recognition, medical imaging, and augmented reality will drive job opportunities for computer vision engineers and researchers.

f) **Ethical AI and Bias Mitigation:**

The focus on ethical AI and mitigating bias in algorithms will create roles for AI ethics specialists and fairness experts.