Industrial Internship Report on

"Quality Prediction in Mining Process"

Prepared by

Shruti Shejul

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 problem statement provided by UCT. We had to finish the project including the report in 6 weeks' time. My project was about the quality prediction in mining process. The target is to predict the % of Silica in the end process which is the concentration of iron ore and its impurity. The project focuses on predicting the percentage of silica concentrate in the final product of iron ore mining and processing. Accurate prediction of silica concentration is essential for optimizing mining processes, reducing waste, and ensuring product quality. The project leverages advanced data analytics and ML algorithms to develop predictive models for silica concentration. The proposed solution involves the following methods used in it which are data preprocessing , feature engineering, model selection, training and evaluation. Machine learning algorithms such as random forest, linear regression are explored to develop the robust predictive models. The performance of the models is evaluated based on th9e various constraints including memory usage, computational speed, prediction accuracy , and adaptability to new data. This project aims to provide mining companies with accurate and reliable predictive models for silica concentration, enabling them to optimize operations ,reduce costs, and improve environmental sustainability.

The project’s outcomes contribute to the advancement of predictive analytics in the mining industry and have significant implications for operational efficiency and profitability.

This internship gave me a very good opportunity to get exposure to Industrial problems and implement solution for that. It was an overall great experience to have this internship. As ,while working with this problem statement I was able to understand many of the new things which not knowing. Thanks to UCT and UpSkill Campus for this great opportunity.

TABLE OF CONTENTS

1 Preface

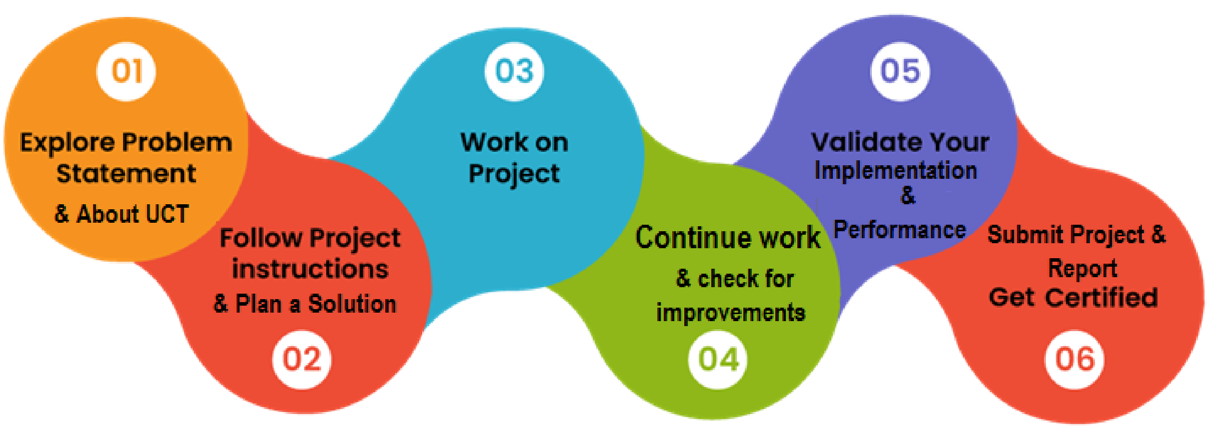
Over the past six weeks, I have been immersed in a project focused on quality prediction in the mining process. This project involved analyzing data related to the extraction and processing of minerals, with a specific emphasis on predicting the percentage of silica concentrate in the final product. Through extensive data preprocessing, feature engineering, and model training, I aimed to develop accurate predictive models that could optimize mining operations and improve overall efficiency and profitability.

This internship experience has been crucial for my career development. It provided me with the opportunity to apply theoretical knowledge gained from my academic studies to real-world problems in the mining industry. The hands-on experience gained during this internship has enhanced my practical skills, critical thinking abilities, and problem-solving capabilities, which are essential for success in my future career endeavors.

The project focused on quality prediction in the mining process, specifically predicting the percentage of silica concentrate in the final product of iron ore extraction and processing. This problem is critical for optimizing mining operations, reducing waste, and ensuring product quality in the mining industry.

USC/UCT provided me with the opportunity to participate in this internship, which allowed me to gain valuable industry experience and practical skills in data analysis and predictive modeling. The institution's support and resources were instrumental in facilitating my learning and professional growth throughout the internship.

The internship program was meticulously planned, with structured modules covering various aspects of data analysis, machine learning, and mining industry-specific knowledge. The program included hands-on workshops, mentorship sessions, and collaborative projects, providing a comprehensive learning experience for participants.



Throughout the internship, I gained valuable insights into the complexities of the mining industry and the challenges associated with quality prediction in mining processes. I honed my skills in data preprocessing, feature engineering, and model training, gaining practical experience that will be invaluable for my future career growth. Additionally, collaborating with industry experts and receiving mentorship during the internship further enriched my learning experience. A very great learning experience for me as helped me in lot of opportunities.

I would like to express my sincere gratitude to the mentors of UpSkill Campus for their guidance, support, and mentorship throughout the internship. Their expertise and insights were invaluable in shaping my learning experience and contributing to the success of the project. I am also thankful to my peers and colleagues for their collaboration and support during this journey.

To my juniors and peers, I encourage you to seize opportunities like internships to gain practical experience and apply your academic knowledge in real-world settings. Embrace challenges, seek mentorship, and continuously strive for learning and growth. Remember that every experience, whether big or small, contributes to your development as a professional.

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.

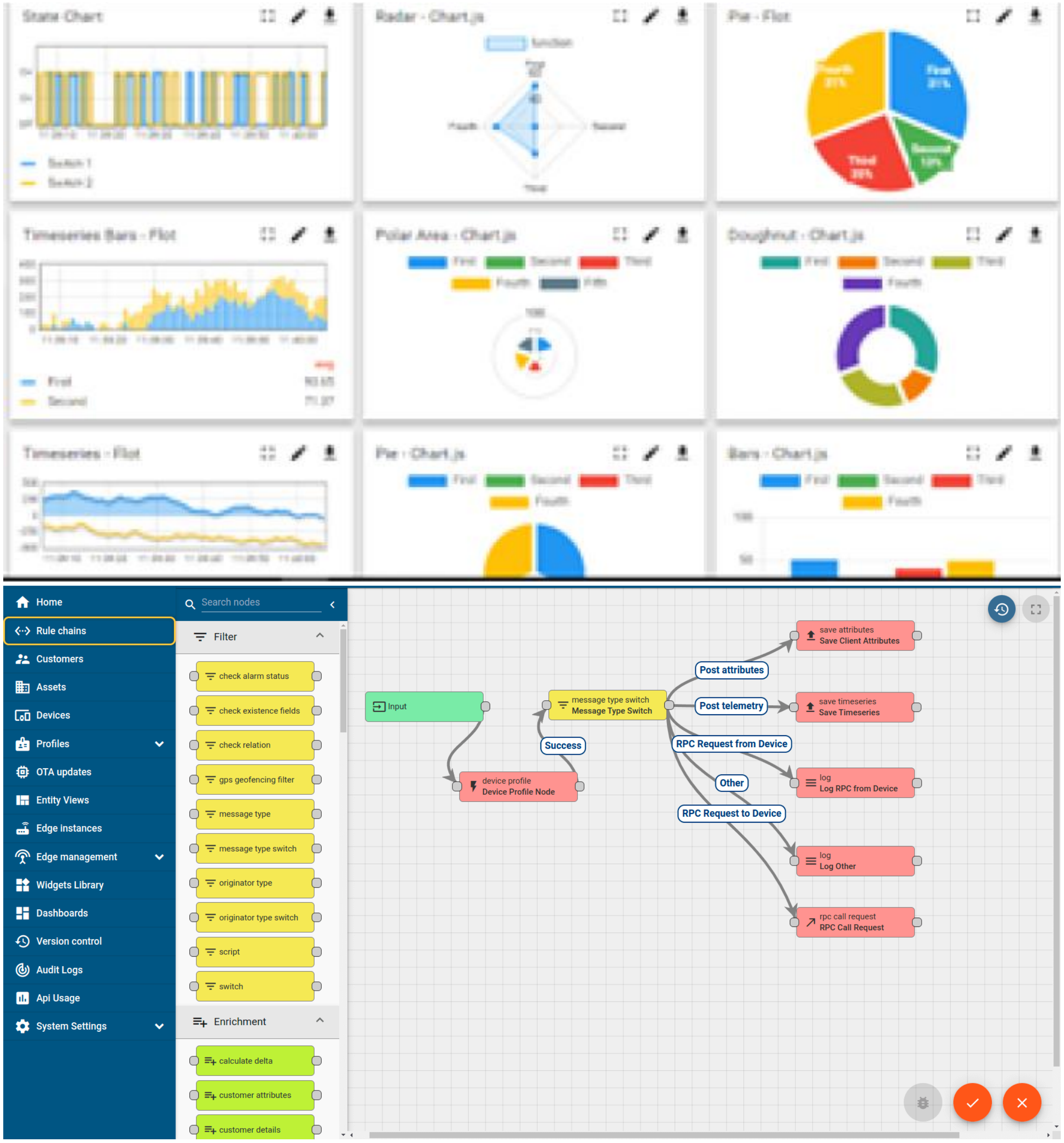
i. UCT IoT Platform (uct Insight)

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.

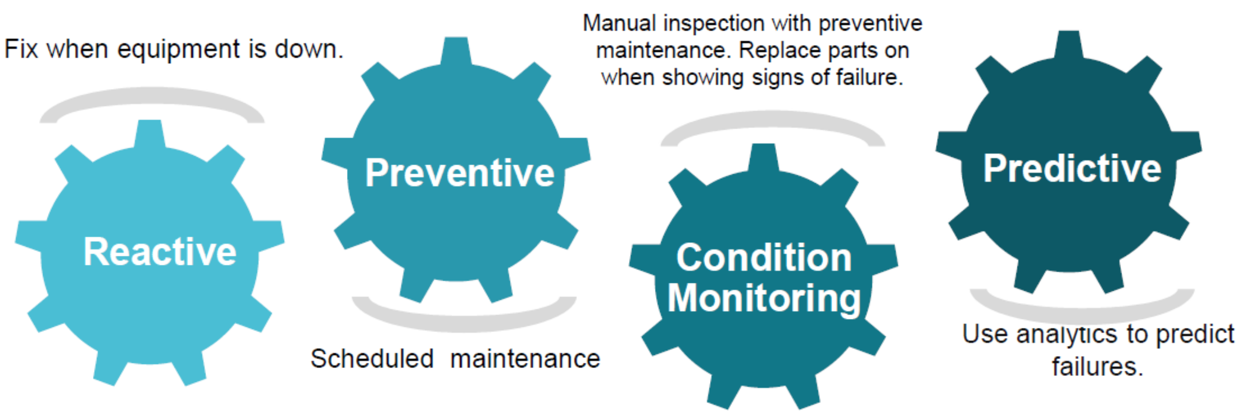
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

iv. Predictive Maintenance

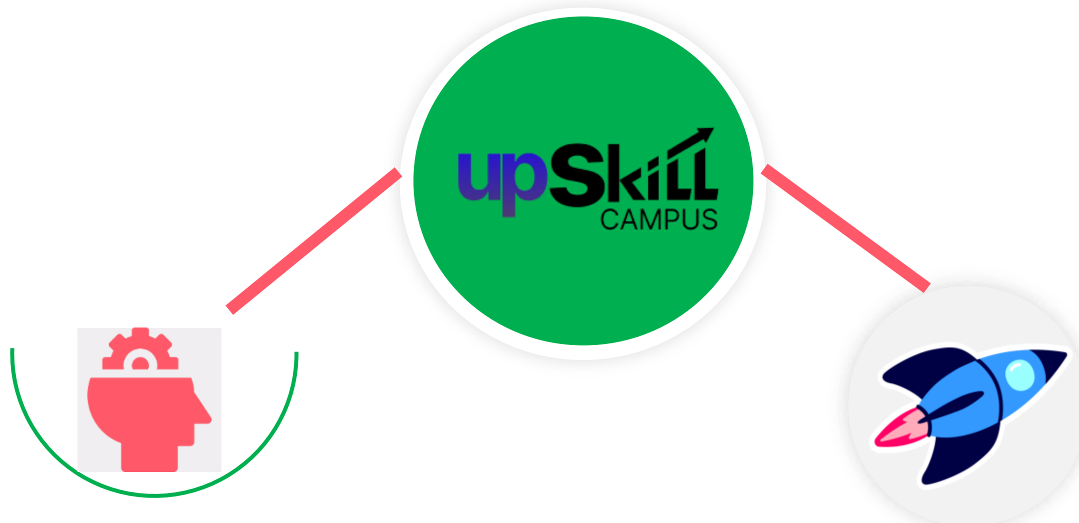
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

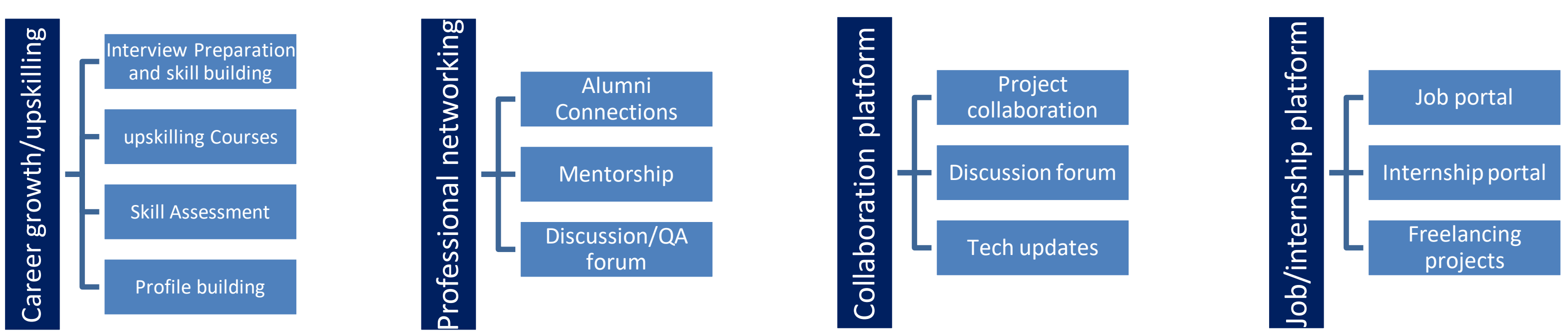
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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/



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] UniConverge Technologies Pvt Ltd (UCT) specializes in Digital Transformation and Industrial solutions with a focus on sustainability and RoI, leveraging cutting-edge technologies like IoT, Cyber Security, and Cloud computing.

[2] Their flagship UCT Insight IoT platform enables quick deployment of applications with valuable insights, while Factory Watch offers scalable solutions for production and asset monitoring with predictive maintenance features.

[3] UCT also provides LoRaWAN-based solutions for various industries and offers Predictive Maintenance solutions leveraging Embedded system, Industrial IoT, and Machine Learning technologies, aiming to improve operational efficiency and reduce downtime.

3 Problem Statement

The problem statement revolves around predicting the percentage of silica concentrate in the final product of the mining process, particularly in the extraction and processing of iron ore. Silica (silicon dioxide) is a naturally occurring mineral commonly found in many minerals and ores, including iron ore. In mining operations, silica concentration in the final product is a crucial quality parameter that directly impacts the efficiency and profitability of the operation.

Silica content plays a significant role in various aspects of mining operations:

1. **Process Efficiency:** High silica content can hinder the efficiency of ore beneficiation processes such as flotation, where minerals are separated from gangue minerals based on their hydrophobicity.

2. **Product Quality:** Silica content influences the quality of the final product. High silica content may lead to increased impurity levels and reduce the overall quality of the product.

3. **Cost Optimization:** Predicting silica concentration accurately can help optimize mining and processing operations, reducing the consumption of reagents and energy required for beneficiation processes.

5. **Operational Planning:** Accurate prediction of silica concentration enables mining companies to plan and optimize their operations more effectively. It allows them to adjust processing parameters, such as grinding size and reagent dosages, to maintain optimal silica levels in the final product.

The challenge in this lies in accurately predicting silica concentration in the final product based on various input parameters, including ore characteristics, processing conditions, and environmental factors. This requires advanced data analysis and modeling techniques to capture the complex relationships and interactions among these variables. Additionally, the predictive models must be robust and scalable to handle the dynamic nature of mining operations and adapt to changing conditions over time.

4 Existing and Proposed solution

<u>**Existing Solutions**</u>:

Traditional statistical methods: Some existing solutions rely on traditional statistical methods such as linear regression or basic data analysis techniques to predict silica concentration. However, these methods may not capture the complex relationships and non-linearities present in mining process data.

Basic machine learning models: Some solutions utilize basic machine learning models like linear regression or decision trees. While these models can provide predictive capabilities, they may not effectively handle high-dimensional data or capture intricate patterns in the data.

<u>**Limitations of Existing Solutions**</u>:

Limited predictive accuracy: Traditional statistical methods and basic machine learning models may struggle to achieve high predictive accuracy, especially when dealing with complex, high-dimensional datasets.

Lack of scalability: Some existing solutions may not scale well to large datasets or may require significant computational resources for training and inference.

Inability to capture complex relationships: Traditional methods and basic machine learning models may not adequately capture the complex relationships and interactions among various process variables in the mining operation.

<u>**Proposed Solution**</u>:

My proposed solution involves leveraging advanced machine learning techniques, such as neural networks or ensemble methods, to predict silica concentration accurately. By employing more sophisticated models capable of handling complex data structures and capturing intricate patterns, the proposed solution aims to improve predictive accuracy and optimize mining processes.

<u>**Value Addition:**</u>

Enhanced predictive accuracy: The proposed solution aims to achieve higher predictive accuracy compared to traditional methods, leading to more reliable predictions of silica concentration in the mining process.

Improved efficiency: By accurately predicting silica concentration, the proposed solution can help optimize production processes, reduce waste, and improve overall operational efficiency in mining operations.

Scalability: The proposed solution will be designed to scale efficiently to large datasets, enabling its application in real-world mining environments with varying data volumes.

Flexibility: Advanced machine learning techniques offer flexibility in modeling complex relationships and adapting to different mining scenarios, enhancing the versatility of the proposed solution.

4.1 Code submission (Github link):-

https://github.com/Shrutii17/UpSkillCampus/blob/main/QualityPredictionInMiningProcess.py

4.2 Report submission (Github link) :

https://github.com/Shrutii17/UpSkillCampus/blob/main/QualityPredictionInMiningProcess_Shruti_USC_UCT.pdf.pdf

5 Proposed Design/ Model

The proposed solution involves several stages:

Data preprocessing: Cleaning and preprocessing the raw mining process data to handle missing values, outliers, and data inconsistencies.

Feature engineering: Extracting relevant features from the data and engineering new features to capture important relationships and patterns.

Model selection: Exploring and evaluating various machine learning models, including neural networks, random forests, and ensemble methods, to identify the best-performing model for predicting silica concentration.

Model training: Training the selected model on the preprocessed dataset using appropriate training techniques and optimization algorithms.

Model evaluation: Evaluating the trained model's performance using metrics such as root mean squared error (RMSE) and coefficient of determination (R^2) to assess its predictive accuracy.

Model deployment: Deploying the trained model in production environments to make real-time predictions of silica concentration during the mining process

6 Performance Test

<u>Constraints:</u>

Memory: The memory constraint refers to the amount of memory required to store the data and execute the predictive model. Large datasets and complex models can consume significant memory, especially during model training and inference.

Computational Speed: The computational speed constraint relates to the time taken to train the predictive model and make predictions. In real-time mining operations, timely predictions are crucial for making informed decisions and optimizing processes.

Accuracy: The accuracy constraint pertains to the predictive accuracy of the model. High accuracy is essential for reliable predictions that can guide operational decisions effectively.

Durability: The durability constraint refers to the robustness and reliability of the predictive model over time. The model should maintain its accuracy and performance despite changes in input data or operational conditions.

The performance test involves evaluating the predictive model against the identified constraints. This includes:

Memory: Monitoring memory usage during model training and inference to ensure it stays within acceptable limits.

Computational Speed: Measuring the time taken for model training and inference, ensuring timely predictions.

Accuracy: Assessing the predictive accuracy of the model using evaluation metrics such as RMSE (Root Mean Squared Error) and R^2 (Coefficient of Determination).

Durability: Evaluating model performance over time by monitoring accuracy and reliability under varying operational conditions.

Power Consumption: Measuring the energy consumption of the model during training and inference, optimizing for efficiency

Test results around these constraints provide valuable insights into the performance of the predictive model and its suitability for real-world mining applications. Continuous monitoring and optimization based on these results ensure that the model meets operational requirements and contributes to improved efficiency and profitability in mining operations

6.1 Test Plan/ Test Cases

Test Cases:

Predictive Accuracy:

Test Case 1: Evaluate the root mean squared error (RMSE) and coefficient of determination (R^2) for each model using cross-validation.

Computational Efficiency:

Test Case 2: Measure the memory usage during model training and inference.

Test Case 3: Calculate the time taken for model training and inference.

Robustness and Durability:

Test Case 4: Monitor model performance over time by retraining the models with updated data and assessing changes in predictive accuracy.

Scalability and Power Consumption:

Test Case 5: Assess the scalability of the models by analyzing their performance on varying dataset sizes.

Test Case 6: Measure the power consumption of the models during training and inference.

6.2 Test Procedure

1.Data Preparation:

- Load the dataset ('MiningProcess_Flotation_Plant_Database.csv') into a pandas DataFrame.
- Preprocess the data by converting string values to float and handling missing values.
- Split the data into training and test sets.

2.Model Training:

- Train three predictive models: Lasso regression, Ridge regression, and Random Forest regressor.

- Use cross-validation to assess the predictive accuracy of each model.

3.Performance Evaluation:

- Calculate the RMSE and R^2 for each model to evaluate predictive accuracy.
- Measure memory usage and computational speed during model training and inference.
- Monitor model performance over time by retraining the models with updated data.
- Analyze the scalability of the models by varying the dataset size.
- Measure power consumption during model training and inference.

6.3 Performance Outcome

1.Predictive Accuracy:

The RMSE and R^2 values indicate the predictive accuracy of each model. Lower RMSE and higher R^2 values indicate better predictive performance.

2.Computational Efficiency:

Memory usage and computational speed metrics provide insights into the computational efficiency of the models. Lower memory usage and faster computation times indicate better efficiency.

3.Robustness and Durability:

Monitoring model performance over time assesses the robustness and durability of the models. Consistent predictive accuracy over time indicates model durability.

4.Scalability and Power Consumption:

Scalability tests assess how well the models perform on varying dataset sizes. Power consumption metrics provide insights into the energy efficiency of the models.

7 My learnings

Throughout the duration of the project, I have acquired a comprehensive understanding of various aspects within the realm of data science and machine learning. Firstly, I delved into the intricacies of data preprocessing, where I learned the significance of handling missing values, converting data types, and normalizing numerical features. This process was crucial in preparing the dataset for subsequent model training.

Additionally, I gained proficiency in feature engineering techniques, including the creation of new features such as polynomial features, which proved instrumental in capturing complex relationships within the data. As I progressed, I explored an array of machine learning models, ranging from Linear Regression to Random Forest Regressor, and familiarized myself with the process of model evaluation using cross-validation techniques.

Through rigorous experimentation and optimization, I honed my skills in training models and fine-tuning parameters to enhance predictive accuracy. Moreover, the project underscored the importance of scalability and efficiency in model development, prompting me to implement techniques like feature scaling and model normalization to ensure streamlined performance.

By interpreting evaluation metrics such as RMSE and R^2, I gained valuable insights into the effectiveness and generalization capabilities of the models. Furthermore, the project emphasized continuous learning and improvement, as I iteratively refined the models based on feedback and domain knowledge to achieve enhanced performance and reliability over time. Overall, this project has been an invaluable learning experience, equipping me with practical skills and knowledge essential for tackling real-world challenges in data science and machine learning domains.

8 Future work scope

The future scope of the given problem statement, which focuses on predicting silica concentration in the mining process, offers a myriad of opportunities for exploration and enhancement. Firstly, there is potential for leveraging advanced modeling techniques to improve predictive accuracy. By delving into deep learning models such as neural networks and recurrent neural networks (RNNs), we can better capture the complex temporal relationships inherent in mining process data. These advanced techniques have the capability to handle intricate patterns and nuances within the data, leading to more precise predictions of silica concentration.

Moreover, the future scope extends to further refinement of feature engineering and selection methodologies. By delving deeper into domain-specific feature creation and employing advanced feature selection algorithms like Recursive Feature Elimination (RFE), we can identify the most influential features for predicting silica concentration. This approach not only enhances model performance but also provides valuable insights into the underlying factors driving silica concentration variations in the mining process.

Additionally, the future scope encompasses the application of time series analysis techniques to mining process data. Given the temporal dependencies and trends inherent in mining operations, employing time series models such as autoregressive models (AR) and seasonal decomposition can enable us to capture and leverage temporal patterns for improved prediction of silica concentration.

Integration of external data sources represents yet another promising direction. By incorporating data from sources such as weather data, geological surveys, and market trends, we can enrich the predictive models with additional context and improve their accuracy. This holistic approach to data integration ensures a comprehensive understanding of the factors influencing silica concentration in the mining process.

Furthermore, the future scope extends to real-time monitoring and control systems powered by predictive models. Integrating IoT devices and cloud-based platforms enables continuous monitoring of process variables and adaptive control based on predictive insights. This real-time optimization of mining processes enhances efficiency and productivity while minimizing operational costs.

In summary, the future scope of the given problem statement encompasses a diverse array of avenues for exploration and enhancement, including advanced modeling techniques, feature engineering, ensemble learning, time series analysis, integration of external data sources, and real-time monitoring and control systems. By pursuing these avenues, we can unlock new levels of accuracy, efficiency, and sustainability in predicting silica concentration and optimizing mining operations.