Project Report: Mining Process Quality Prediction

1. Introduction

This project focuses on predicting the quality of a mining process using real-world data containing operational parameters and output quality indicators. The goal is to model and predict the percentage of silica in the concentrate, which is a key metric in the efficiency of mineral processing plants.

2. Dataset Description

The dataset contains hourly records of a flotation process from a mineral processing plant. Each row represents one hour of operation and includes values such as ore feed properties, reagent flows, pulp characteristics, air flows in flotation columns, and the resulting % of Iron and Silica in the concentrate.

Total Features: 24 raw columns + engineered features (lags, rolling means)
Target Variable: % Silica Concentrate

3. Methodology

1. Loaded dataset and corrected European-style number formatting.
2. Converted all numeric columns and parsed the timestamp column.
3. Performed forward-fill imputation and removed invalid rows.
4. Engineered lag and rolling mean features for air flow and level variables.
5. Visualized correlations to guide feature selection.
6. Standardized the feature set and trained an XGBoost regression model.
7. Evaluated the model using RMSE and R^2 on the test set.
8. Interpreted results using SHAP values to identify feature influence.

4. Results and Evaluation

Model performance on test set:

- Test RMSE: 0.24

- Test R^2 Score: 0.96

These results indicate that the model can reliably predict the silica content in concentrate with low error and high explanatory power. The correlation analysis and SHAP

interpretation confirmed that air flow and level variables were highly influential in prediction.

5. Business Value and Insights

By deploying this model in real-time, plant operators can receive alerts when silica content is likely to increase, allowing for preemptive adjustments to process parameters. This helps in maximizing product quality, reducing waste, and maintaining compliance with operational standards.

6. Conclusion

The model effectively captures quality degradation patterns in the flotation process and provides reliable predictions of silica concentration. With strong performance and interpretability, this project demonstrates a practical use-case for machine learning in industrial process optimization.

7. Future Work

- Integrate model into a live monitoring dashboard (e.g., using Power BI).
- Extend prediction to both Iron and Silica simultaneously (multi-output).
- Experiment with LSTM for sequential modeling.
- Evaluate performance on additional time splits and other processing sites.