# Lead Data Scientist Assignment: Silicon (SI) Prediction in Hot Metal

## Problem Statement

Silicon (SI) content in hot metal is a critical indicator of blast furnace thermal stability. Accurate SI prediction enables early detection of process deviations, allowing operators to make proactive adjustments to maintain efficiency. This assignment focuses on developing an advanced predictive model leveraging raw material chemistry, burden composition, and operating parameters to ensure high accuracy, robustness, and real-world applicability.

## Data Availability

* Dataset Size: 5,704 records, 25 features
* Target Variable: SI (Silicon content in hot metal)
* Feature Categories:
* Operating Parameters: Blast temperature, top pressure, oxygen enrichment, steam flow, blast momentum, permeability index, theoretical combustion temperature.
* Timestamp Available: Allows for time-series analysis and trend identification.

## Expected Deliverables

* High-Performance Predictive Model for SI, optimized for accuracy and robustness.
* Provision Real-time Anomaly Detection to flag unexpected SI deviations and recommend corrective actions.
* Preemptive Root Cause Analysis to identify key parameters affecting SI and guide process control decisions.
* Explainability & Actionability: SHAP analysis, feature importance, and business-relevant insights for furnace optimization.
* Comprehensive Model Evaluation:
  + Performance Metrics: R², RMSE, MAPE.
  + Uncertainty Quantification for confidence estimation.
  + Robustness Testing to validate performance under varying process conditions.
* Optimization Algorithms: Implementation of Reinforcement Learning (RL), Genetic Algorithms (GA), or Bayesian Optimization to dynamically adjust furnace parameters for optimal SI levels.
* Scalable Deployment Strategy\*:
* Real-time inference with low-latency processing.
* Model Drift Detection & Retraining Mechanism for sustained performance.
* Cloud or Edge Deployment Considerations based on operational constraints.

\*Candidate should provide a high-level approach to deployment strategy, outlining key infrastructure and integration aspects.

## Submission Requirements

Candidates must submit the assignment in the following formats:

* Jupyter Notebook/Python Script/Code Repository with a structured approach.
* PowerPoint Presentation summarizing methodology, key findings, insights, and proposed deployment plan.