Hackathon Report

Electrical Conductivity Prediction in Metallurgical Alloys

BY: Sonia (23118074), Swaroop (23118077) & Aahana (23118001)

---

1. Problem Statement

The objective of this hackathon is to predict the **electrical conductivity (%IACS)** of metallurgical alloy samples based on their **chemical composition**, **processing parameters**, and **mechanical properties**. Accurate prediction of conductivity aids in designing high-performance materials in aerospace, automotive, and electronics industries.

---

2. Dataset Description

The dataset comprises detailed records of metallurgical alloys, including:

Features

- **Alloy Formula**: Chemical formula (dropped in model input due to high complexity and sparsity).
- **Alloy Class**: Categorical label for alloy type (also dropped).

Elemental Composition (Weight fractions)

- Cu, Al, Ag, B, Be, Ca, Co, Ce, Cr, Fe, Hf, La, Mg, Mn, Mo, Nb, Nd, Ni, P, Pb, Pr, Si, Sn, Ti, V, Zn, Zr

Processing Conditions

- **Tss (K)**: Solution treatment temperature
- **tss (h)**: Solution treatment time
- **CR reduction (%)**: Cold rolling reduction
- **Aging**: Binary feature indicating aging process
- **Tag (K)**: Aging temperature
- **tag (h)**: Aging time
- **Secondary thermo-mechanical process**: Categorical

Mechanical Properties

- **Hardness (HV)**
- **Yield Strength (MPa)** (Dropped)
- **Ultimate Tensile Strength (MPa)** (Dropped)

Target Variable

- **Electrical Conductivity (%IACS)**: The prediction target

---

3. Evaluation Metric

Submissions are evaluated using **Mean Absolute Error (MAE)**:

Lower MAE indicates better model performance.

---

4. Data Preprocessing

- **Dropped** irrelevant columns: 'ID', 'Alloy Formula', 'Alloy Class', 'Yield Strength', 'Ultimate Tensile Strength'
- **Label encoding** of categorical variables
- **Missing values** filled using mean/mode for numeric fields and a placeholder for categoricals
- **Feature scaling** performed before training deep learning models

---

5. Model Development

Traditional ML Models

- Random Forest
- XGBoost
- LightGBM
- Ridge Regression

Best Performing: Random Forest (post-hyperparameter tuning)

Deep Learning Models

- **Regression MLP**: 2 hidden layers with BatchNorm, Dropout
- **Wide & Deep Network**: Combines linear and deep features
- **Residual Network (ResNet-style)**: Deep model with skip connections
- **TabNet**: Attention-based deep learning for tabular data

Best Performing: Wide & Deep (after Optuna tuning)

Ensemble Learning

- Combined predictions of **Regression MLP** and **Wide & Deep Network**
- Simple average ensemble reduced MAE further

---

6. Hyperparameter Tuning

- Performed using **Optuna** for deep learning models
- **RandomizedSearchCV** for ML models
- Optimization focused on MAE minimization

---

7. Final Submission & Output

- Predictions were made using the best Random Forest model

Output file format:
ID,Electrical conductivity (%IACS)
2,45.5
5,99.2

- 6,94.5
- Submitted `predictions.csv` for leaderboard evaluation

---

8. Conclusion

- **Wide & Deep models** demonstrated strong performance, but Random Forest remained a robust and fast baseline.
- **Ensembling** yielded slight MAE reduction.
- Further improvements may involve:
    - Feature engineering (e.g., interaction terms)
    - Incorporating domain knowledge in pre-processing
    - Advanced architectures (e.g., GNN for structural data)

Goal Achieved: Predictive model with low MAE for alloy conductivity, aiding materials design.