

MetES Kaggle Competition Report

Introduction

The MetES Kaggle Competition focuses on predicting the electrical conductivity (%IACS) of metallurgical alloys based on their composition and processing parameters. This competition aims to develop a robust predictive model to help in material selection and process optimization.

Dataset Overview

The dataset consists of metallurgical alloy samples with features related to:

- **Elemental Composition** (weight fractions of elements like Cu, Al, Ag, etc.)
- **Processing Conditions** (solution treatment temperature & time, cold rolling reduction, aging conditions)
- **Mechanical Properties** (Hardness (HV), Yield Strength (MPa), Ultimate Tensile Strength (MPa))
- **Target Variable:** Electrical Conductivity (%IACS)

Feature Importance Analysis

Using feature importance analysis from Random Forest, the most influential features for predicting electrical conductivity were:

- **Composition** (individual element weight fractions)
- **Solution Treatment Temperature (Tss K)**
- **Solution Treatment Time (tss h)**
- **Cold Rolling Reduction (%)**
- **Aging Temperature (Tag K)**
- **Aging Time (tag h)**
- **Hardness (HV)**

These features were prioritized during model development to enhance predictive performance.

Model Selection and Performance Comparison

Multiple machine learning models were tested to find the most accurate approach. The results were as follows:

Model	Mean Absolute Error (MAE)
Random Forest (Best Model)	13.61438
Linear Regression	13.69008
XGBoost	13.66300
RandomizedSearchCV (Optimized RF)	13.73077
Neural Network	19.68904

Despite experimenting with hyperparameter tuning and feature selection, **neural networks performed the worst**, while **Random Forest outperformed all models**.

Best Model: Random Forest Regression

The best-performing model was **Random Forest Regression**, with the following hyperparameters:

- **n_estimators**: 325
- **max_depth**: None
- **min_samples_split**: 7
- **min_samples_leaf**: 1
- **max_features**: 'sqrt'
- **Mean Absolute Error (MAE)**: 13.61438

This model effectively captured non-linear relationships in the data while maintaining generalizability.

Conclusion & Future Work

Key Takeaways:

- **Random Forest performed the best**, providing the lowest MAE.
- **Linear Regression, XGBoost, and RandomizedSearchCV** delivered similar performance but were slightly less effective.
- **Neural Networks did not perform well**, likely due to limited training data or the need for extensive hyperparameter tuning.
- **Important features like composition, thermal processing parameters, and hardness played a crucial role** in prediction accuracy.

Future Improvements:

- **Feature Engineering:** Experiment with derived features such as interaction terms or domain-specific transformations.
- **Ensemble Methods:** Blend multiple models to further reduce MAE.
- **Advanced Hyperparameter Tuning:** Use Bayesian Optimization for improved tuning.
- **Polynomial Regression:** Investigate if polynomial terms help improve performance.

This report summarizes the approach taken to achieve the best predictive model. Further improvements could refine predictions and enhance industrial applications of the findings.

Prepared by: [Amarjeet]

Date: [05-04-2025]