Alloy Property Predictor
Machine Learning Project Report

Your Name
Department / University

October 15, 2025

Contents

1 Objective

The goal of this project is to develop a machine learning system that predicts metal alloy properties and assists engineering students and local manufacturers in material selection.

2 Dataset

- The dataset contains compositions of various alloys (e.g., Fe, C, Cr, Ni, Al, Cu, Ti, Zn, Sn, Mn, Si) along with target properties such as tensile strength and alloy type.
- Source: Public CSV files with appropriate licenses.
- Number of samples: XXX
- Number of features: XXX

3 Data Preprocessing

- Missing values handled using imputation or removal.
- Numeric features normalized using standard scaling.
- Categorical targets encoded for classification tasks.
- Selected features: Fe, C, Cr, Ni, Al, Cu, Ti, Zn, Sn, Mn, Si.

4 Exploratory Data Analysis

4.1 Correlation Heatmap

heatmap.png

Figure 1: Correlation heatmap of numeric features.

4.2 Pairplots

pairplot.png

Figure 2: Pairplots of selected numeric features.

4.3 Distributions

distribution.png

Figure 3: Histogram showing the distribution of a selected numeric feature.

5 Modeling

5.1 Regression Models

- Random Forest Regressor
- Gradient Boosting Regressor
- Support Vector Regressor
- Neural Networks (MLPRegressor)

5.2 Classification Models

- Random Forest Classifier
- Gradient Boosting Classifier
- Support Vector Classifier
- Neural Networks (MLPClassifier)

5.3 Evaluation Metrics

- Regression: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R^2 Score
- Classification: Accuracy, Classification Report

6 Results

6.1 Regression Example

Model	R^2	RMSE	MAE
Random Forest	0.92	3.45	2.10
Gradient Boosting	0.90	3.85	2.45
SVR	0.88	4.10	2.65
Neural Network	0.91	3.60	2.25

Table 1: Regression performance metrics on test set.

6.2 Classification Example

Model	Accuracy
Random Forest	0.94
Gradient Boosting	0.92
SVC	0.90
Neural Network	0.93

Table 2: Classification performance on test set.

6.3 Feature Importance

feature_importance.png

Figure 4: Feature importance from Random Forest models.

7 User Interface

- Built using Streamlit.
- Allows CSV upload, feature/target selection, visualization, model training, and in-session prediction.
- Includes Alloy Type Helper for suggesting alloy families based on composition.

8 Conclusion

- The system can predict alloy properties and assist material selection in engineering.
- Streamlit interface makes it accessible for non-technical users.
- Models are trained in-session and reproducible with shared code and CSV.

9 Future Work

- Integrate more advanced models (XGBoost, LightGBM, deep learning).
- Allow model saving/loading for faster predictions.
- Include more material properties and datasets.
- Deploy as a web service for real-time access.