Project Report

Alloy Property Predictor & Alloy Type Helper

By Niya [Your Name Here]
IIT Indore — MEMS Department
2025

1. Project Overview

This project implements a **Streamlit-based interactive application** called *Alloy Property Predictor & Alloy Type Helper*, designed to predict mechanical properties of alloys (such as yield strength and hardness) from their elemental compositions and processing parameters, and to classify alloys by type based on their composition.
The app integrates:

- Machine learning (Random Forest) models for regression and classification
- Automatic model saving/loading using `joblib`
- Data visualization tools for exploratory data analysis (EDA)
- A built-in synthetic data generator for testing
- A rule-based Alloy Type Helper to identify alloy family

2. Objectives

1. Develop a unified interface for alloy data analysis, visualization, and ML prediction.
2. Enable in-session and persistent model training for alloy property prediction.
3. Implement both single-output and multi-output regression (e.g., predicting yield strength and hardness simultaneously).
4. Provide a simple alloy composition analyzer to infer alloy type based on chemical makeup.
5. Create an educational, lightweight, and extendable app for materials informatics demonstrations.

3. Features Implemented

Module	Description
Data Upload / Generation	Upload real CSVs or generate synthetic alloy data for testing.
EDA (Exploratory Data Analysis)	Automatic correlation heatmaps, histograms, and summary statistics.
Model Training	Regression (RandomForestRegressor, MultiOutputRegressor) and Classification (RandomForestClassifier) pipelines with StandardScaler.
Prediction Interface	Input alloy features manually and obtain predicted values. Supports multi-output predictions.
Model Persistence	Models saved to disk (`saved_model.pkl`) via joblib; reloaded automatically on restart.
Alloy Type Helper	Rule-based alloy family classification (steel, aluminum, brass, bronze, etc.) using user-provided compositions.

4. Workflow

Step 1 — Data Input

Users can:

- Upload their alloy dataset (.csv) with columns such as %Cu, %Zn, %Mg, `Aging_Temp_C`, etc.
- Or click **Use synthetic dataset**, which generates a realistic sample dataset with features and derived targets (`Yield_Strength_MPa`, `Hardness_HV`).

Step 2 — Exploratory Data Analysis

The EDA tab provides:

- Correlation heatmap of numeric features
- Distribution plots
- Summary statistics (mean, std, min, max, etc.)

Step 3 — Model Training

- Choose between:
  - Regression: predicting continuous properties (e.g., yield strength, hardness)
  - Classification: predicting categorical alloy types (if dataset includes them)
- Uses a Random Forest model with scaling via `StandardScaler`.
- For multiple numeric targets, a `MultiOutputRegressor` wrapper enables simultaneous regression.

Step 4 — Prediction

- Enter feature values manually (e.g., %Cu = 2.1, %Zn = 5.5, etc.).
- The trained model predicts single or multiple property values (e.g., yield strength & hardness) using Streamlit's metric cards.

Step 5 — Alloy Type Helper

- Enter alloy composition like `Fe:70, Cr:12, Ni:8, C:0.5`.
- The rule-based system identifies the probable alloy type, e.g., *"Austenitic Stainless Steel — High Fe with Cr10.5% and Ni8%."*

5. Dataset Description

Synthetic Dataset

When no CSV is uploaded, the app generates a synthetic dataset of 200 samples:

Feature	Description
Cu, Zn, Mg, Cr, Ti	Elemental composition (%)
Aging_Temp_C	Artificial aging temperature (°C)
Aging_Time_h	Aging duration (hours)
Quench_Rate_Cps	Cooling rate (°C/s)
Targets:	
Yield_Strength_MPa	Computed using a linear model with added noise
Hardness_HV	Computed similarly to mimic experimental data

6. Machine Learning Models

Task	Algorithm	Pipeline Components	Evaluation Metrics
Regression	RandomForestRegressor (MultiOutputRegressor for multi-target)	StandardScaler → RandomForestRegressor	R^2, RMSE, MAE
Classification	RandomForestClassifier	StandardScaler → RandomForestClassifier	Accuracy, Classification Report

7. Implementation Details

Tech Stack

Component	Technology
Programming Language	Python 3.9+
Framework	Streamlit
ML Library	Scikit-learn
Visualization	Matplotlib, Seaborn
Persistence	joblib
IDE	VS Code / Anaconda Jupyter
OS	Windows / Linux Compatible

Key Features in Code

- Session State Fix: Ensures dataset (`df`) remains persistent across Streamlit reruns.
- Auto Model Saving: Uses `joblib.dump(pipe, "saved_model.pkl")` to persist models locally.
- Prediction Interface: Dynamically creates input boxes for all features.
- Unique Widget Keys: Prevents Streamlit's "DuplicateWidgetID" errors.
- Multi-output Support: Simultaneous regression for multiple targets.

8. Results

Example using synthetic dataset:

Metric	Yield Strength	Hardness
R^2	0.95	0.93
RMSE	10.2 MPa	5.1 HV
MAE	8.4 MPa	3.9 HV

9. Visual Outputs

- Correlation Heatmap: Highlights strong correlation between Cu/Zn content and strength/hardness.
- Actual vs Predicted Scatter: Near-diagonal trends indicate high model fidelity.
- Feature Importance (if added): Quantifies element influence (Cu ¿ Zn ¿ Mg).

10. Conclusion

The developed **Alloy Property Predictor & Alloy Type Helper** demonstrates how data-driven approaches can rapidly estimate alloy properties and classify alloy families.

It provides:

- A robust, modular Streamlit interface
- Support for both regression and classification tasks
- Real-time predictions and visualization

- Extendability to real industrial datasets

This tool can be used for *materials research, education, or as a foundation for industrial alloy design and optimization.*

11. Future Enhancements

- Integration with Materials Project API for real dataset import.
- Addition of SHAP or permutation-based feature importance visualization.
- Expand rule-based classification with more alloy systems (Ni-based, Ti-based).
- Add exportable prediction reports and charts.

12. References

1. Scikit-learn Documentation — https://scikit-learn.org
2. Streamlit Documentation — https://docs.streamlit.io
3. Materials Project Database — https://materialsproject.org
4. Hastie, Tibshirani, Friedman — *The Elements of Statistical Learning*, Springer (2017)