# Clustering Steel Grades Based on Microstructure and Yield Strength

## Objective

The objective of this project is to classify steel grades using unsupervised learning techniques by analyzing:  
- Microstructural parameters (e.g., grain size, phase content)  
- Mechanical property: Yield Strength (MPa)

## 1. Data Description

The dataset includes the following:  
- Microstructural features (such as ferrite, pearlite, bainite, martensite fractions, grain size, etc.)  
- Yield Strength (MPa), which is the target property for mechanical performance.

## 2. Methodology

### a. Data Preprocessing

- Removed the `Yield\_Strength\_MPa` column from the feature set to use it later for interpretation.  
- Applied StandardScaler to normalize all features to zero mean and unit variance.

### b. Dimensionality Reduction

- Used Principal Component Analysis (PCA) to reduce dimensions from high-dimensional microstructure space to 2 principal components.  
- The first two principal components explained ~34% of the variance in the dataset.  
- This enabled clear 2D visualization of clusters.

### c. Clustering

- Used the Elbow Method on K-Means clustering to determine the optimal number of clusters (k).  
- Based on the elbow plot, we selected k = 4 as the optimal number.  
- Applied KMeans clustering and assigned each data point to a cluster.

### d. Visualization

- Plotted the clusters in the 2D PCA space.  
- Different clusters show distinct groupings of steel grades, likely due to variation in microstructural parameters.

## 3. Insights

- Clustering successfully grouped steels with similar microstructural profiles.  
- Post-clustering analysis (optional next step) can include:  
 - Comparing average Yield Strength across clusters.  
 - Analyzing which microstructural features dominate in each cluster.  
 - Identifying optimal microstructures for high-yield strength grades.

## 4. Future Scope

- Incorporate hierarchical clustering or DBSCAN to validate cluster consistency.  
- Use feature importance analysis (e.g., via decision trees on clusters) to understand what drives clustering.  
- Visualize Yield Strength distributions across clusters to uncover mechanical property trends.