Prediction of Mechanical Properties of Low-alloy Steels

Presented by
Artem Ramus

Background

Data set: "Mechanical properties of low alloy steels" from Kaggle, Contains alloy composition, temperature and mechanical properties

Context: currently there are no precise theoretical methods to predict mechanical properties of steels. All the methods available are by backed by statistics and extensive physical testing of the materials. Since testing each material with different composition is a highly tedious task (imagine the number of possibilities!), let's apply our knowledge of machine learning and statistics to solve this problem.

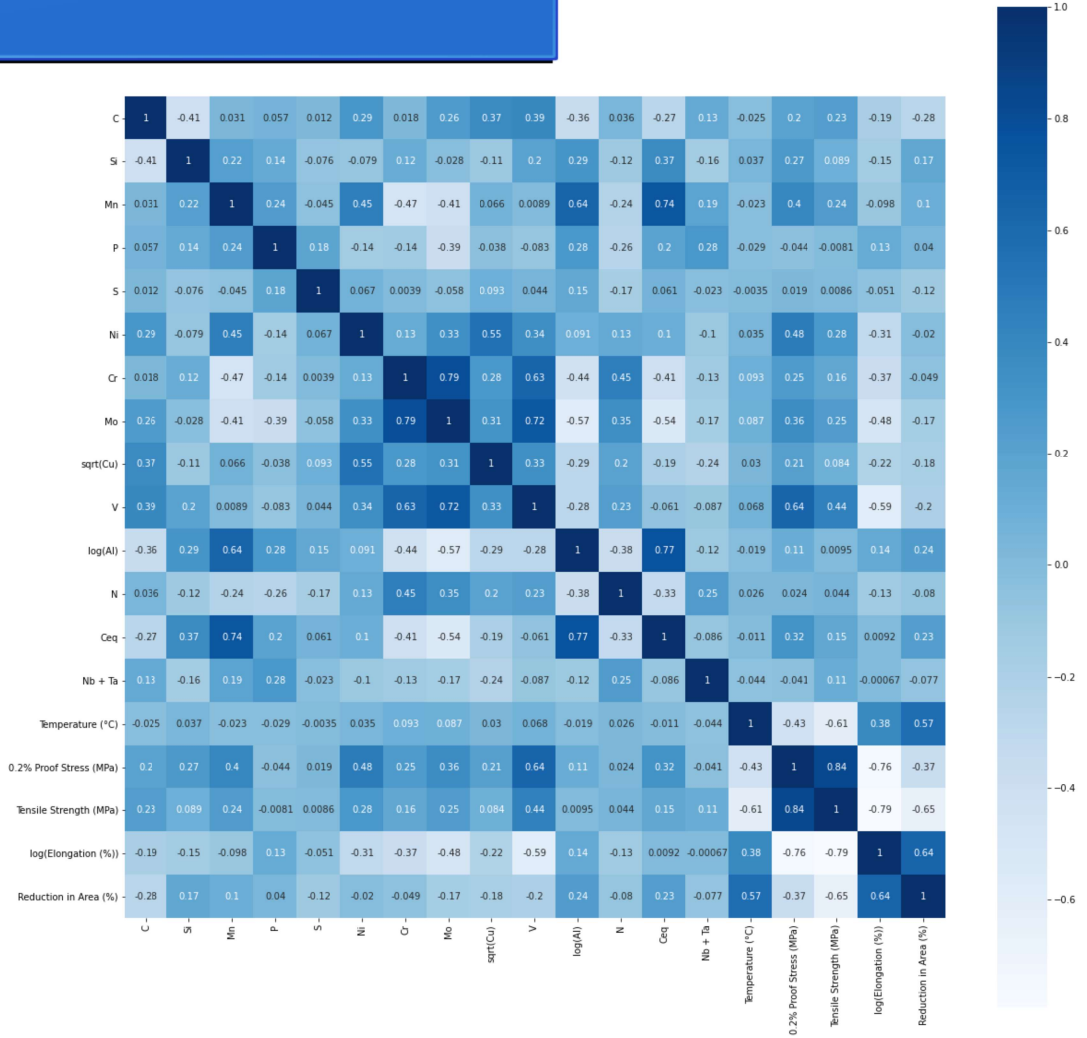
Link to the data set:
https://www.kaggle.com/rohannemade/mechanical-properties-of-low-alloy-steels

Introduction

Content: this dataset contains compositions by weight percentages of low-alloy steels along with the temperatures at which the steels were tested and the values mechanical properties observed during the tests. The alloy code is a string unique to each alloy. Weight percentages of alloying metals and impurities like Aluminum, copper, manganese, nitrogen, nickel, cobalt, carbon, etc are given in columns. The temperature in celsius for each test is mentioned in a column. Lastly mechanical properties including tensile strength, yield strength, elongation and reduction in area are given in separate columns. The dataset contains 915 rows.

Methodology

- The data set is checked for duplicates, null values and homogeneous features.
- Data distribution is checked with histogram, distribution and box plots for skewness and outliers.
- Correlation between features was examined with a correlation matrix.



Performance of the Model

- Performance of the following models was estimated based on Pearson coefficient:
  - Linear regression
  - Decision tree
  - Random forest
  - Artificial neural network

Summary and Conclusions

Random forest regression has showed best performance for this task.

ANN performs similarly to random forest regression for all the four dependent variables. R square score is about 2% less for the ANN.

Data scaling, outliers removal and skewness amendment have sensibly improved the performance.

The end

Thank you for your attention!