International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 666-670 www.ijaem.net ISSN: 2395-5252

Predicting Mechanical Properties of Low-Alloy Steels using Machine Learning

Arnab Adhikary
Student, Department of Mechanical Engineering, Jadavpur University
Corresponding Author:Arnab Adhikary

Submitted: 01-08-2021 Revised: 10-08-2021 Accepted: 13-08-2021

ABSTRACT: This paper introduces a Machine Learning and Neural Network oriented approach to predict the tensile strength of low alloy steel. In this era of data-driven decision making, the use of data science and machine learning has become prevalent not only in the field of management and analytics but also in science. This paper suggests the use of this method, to estimate the mechanical properties of the desired material without having to, physically, test/work with the specimen to do so. It uses the data collected earlier by physical experiments and relates the feature variables with the mechanical property using correlations and various ML algorithms. This method uses the composition(wt%) of alloys, the temperature at which the experiment was being conducted to predict the ultimate tensile strength of steel.

KEYWORDS: Data Science, Machine Learning, Statistical Analysis, Deep Learning, Low alloy steels.

I. INTRODUCTION

In today's world, knowing the mechanical property of a certain object has become mandatory to design, manufacture, test and innovate. With changing demand and purpose, innovation of new materials with the desired properties material and mechanical properties is needed. To do that, the engineers, scientists or concerned individual has to physically test and go through a vast and hectic process to develop such a material and test its sanctity. One such example is that of Steel. Steel is one of the world's most important engineering and construction materials that is used in every aspect of our lives. To name a few of them, in cars and construction products, refrigeration, transports and medicinal purposes. Steel is an alloy made up of iron with usually a few tenths of a percentage of carbon with a few other elements. This paper in particular deals with low-alloy steel that belongs to the low-medium carbon steels. Therefore, it possesses precise chemical composition that provides better mechanical properties to improve strength,hardness ,anti-corrosion properties etc. The main goal of using data science to predict the properties is to reduce the cost, time and labour of the individual/organisation consumed in doing so physically, to improve and explore mechanical properties of steel. Traditional techniques involve using UTM, Heat treatment, trial and error etc but as mentioned before, all this could be done faster and more efficiently using nothing but data. This topic is considered a hot topic for the very reason. Statistics, Machine Learning have immensely evolved to replace human judgement and have often been proved right. The field of data science is widely classified into machine learning, deep learning, image processing, time series forecasting, Anomaly Detection, Clustering and Classification, Regression etc. In this article regression algorithms have been used to rightly predict the mechanical properties of steel using its composition and experimental environment.Algorithms used are namely:

- Decision Tree Regression
- Random Forest Regression
- XGB Regression (Boosting technique)
- SVM Regression
- Artificial Neural Network

II. MATERIAL AND PRE-PROCESSING

Steel is an alloy of iron with typically a few tenths of a percent of carbon to improve its strength and fracture resistance compared to iron. Many other elements may be present or added. Because of its high tensile strength and low cost, steel is used in buildings, infrastructure, tools, ships, trains, cars, machines, electrical appliances, and weapons. Steel is a subsidiary of Ferrous Materials that havea carbon percentage of less than 2.1% in them.Further classifications have been shown in FIGURE 1.

Numerous mechanical properties of steel can be predicted for example yield strength, elongation%,manufacturing process,manufacturing process, heat treatment etc. But in this paper,

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 666-670 www.ijaem.net ISSN: 2395-5252

ultimate tensile strength will be predicted and the same procedure can be implemented for other mechanical properties as well.

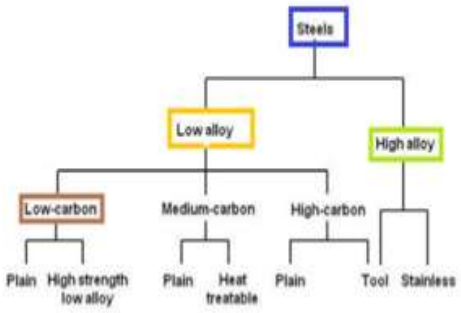

Fig1. Classification of steel

The data that we are going to use include these features:

- Alloy code
- C(Carbon)
- Si (Silicon)
- Mn (Manganese)
- P (Phosphorus
- S(Sulphur)
- Ni(Nickel)
- Cr(Chromium)
- Mo(Molybdenum)
- Cu(Copper)
- V(Vanadium)
- Al(Aluminium)
- N(Nitrogen)
- Ceq,
- Nb + Ta
- Temperature (°C)
- Tensile Strength (MPa)

2.1 Feature Engineering

In the dataset that is being used, the features are tested for their respective correlation with the target variable i.e. Tensile Strength(Mpa) using a popular statistical test, the chi-squared test.A chi-squared test, also written as the $\chi 2$ test, is a statistical test that is used to determine whether there is a statistically significant difference between the expected frequencies and the observed frequencies in one or more categories of a contingency table. In other words, it lets us knowhow good the variable is in terms of affecting the outcome of the target variable. **Figure 2** shows the results of this test. It clearly shows that temperature plays a big role in predictingTensile Strength. One possible explanation for this is that as temperature increases, the atoms become free and the movement of dislocations along slip planes become easier. As a result, the tensile strength may decrease. The second most important seems to be chromium, the possible explanation for this maybe that of solid solution hardening that is chromium atoms block the passage of slip planesaffecting the Tensile strength. Further pre-processing involved removing the skewness in the data. Skewness is the measure of the asymmetry of a probability distribution of a random variable. To remove skewness, log transformations and square root transformations have been applied to the relevant data columns.

Another problem that the data had, was that of outliers that may have generated abnormal predictions. So they have been removed by appropriate methods.

Features	Score
Temperature (°C)	64588.039755
Cr	234.966525
Mo	172.437950
Ceq	155.668642
Ni	101.139152
V	88.788434
Mn	70.954273
Cu	17.008691
Si	11.578424
C	9.200508
Al	6.317302
P	0.771312
Nb + Ta	0.724526
S	0.654711
N	0.309697

Fig2. Feature Importance

III. METHODOLOGY

A comparison between the 5 models have been made to undertake the regression and predict the tensile strength of steel based on the parameters and variables.The regression-based algorithms that have been chosen are Decision Tree, Random Forest, XGBoost, SVM, ANN.

3.1 Decision Tree:

A Decision Tree is a supervised machine learning algorithm that builds both regression and classification models in the form of a tree structure. It disintegrates a dataset into smaller and smaller subsets simultaneously building a tree by training..The final result is a tree with decision nodes and leaf nodes.It creates the tree based on the training data and uses the same to predict the required target variable. The dataset has been separated into training and testing subsets using Train Test Split with 80% being the training dataset and 20% being the testing dataset. The test dataset is then used for testing the model's accuracy.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 666-670 www.ijaem.net ISSN: 2395-5252

3.2 Random Forest Regressor:

A Random Forest is an ensemble technique that is capable of both regression and classification with the use of multiple decision trees. It uses a technique called bootstrap and aggregation,also known as bagging.It uses averaging to improve prediction and control over-fitting. To improve accuracy, RandomSearchCV has been used to tune the hyperparameters of the model which tests the model with certain parameters on cross-validation datasets. The dataset has been separated into training and testing subsets using Train Test Split with 80% being the training dataset and 20% being the testing dataset. The test dataset is then used for testing the model's accuracy.

3.3 XGBoost:

XGBoost refers to eXtreme Gradient Boosting. It implements machine learning algorithms under the Gradient Boosting framework. This is an ensemble method. The boosting technique the trees are built sequentially such that each subsequent tree aims to reduce the errors of the previous tree. It makes use of the tree with fewer splits. Some other boosting techniques are ADAboost and CATBoost. To improve accuracy, RandomSearchCV has been used to tune the model which tests the model with certain parameters on cross-validation datasets. The dataset has been separated into training and testing subsets using Train Test Split with 80% being the training dataset and 20% being the testing dataset. The test dataset is then used for testing the model's accuracy.

3.4 SVM (SUPPORT VECTOR MACHINE):

SVM performs classification and regression techniques under supervised learning. The basic idea behind SVR(Support Vector Regression) is to find the best fit line.In SVR,the best fit line is the hyperplane that contains the maximum number of data points. The dataset has been separated into training and testing subsets using Train Test Split with 80% being the training dataset and 20% being the testing dataset. The test dataset is then used for testing the model's accuracy.

3.5 ANN(Artificial Neural Network):

An artificial neural network is the component of artificial intelligence that imitates the function of thehuman brain. The brain has a hive of networks that takes an input and performs the output through an armature. Processing units are built-in ANNs which in turn have inputs and outputs. Learning is done based on inputs and predicts the output. Backpropagation is the principle used to guide ANNs. There are different layers(input,hidden and output) in a neural network that help in taking input, calculating output and feed the error back to the neural network. Nodes are interconnected among the layers and have three components: weight,bias, and activation function. The dataset has been separated into training and testing subsets using Train Test Split with 80% being the training dataset and 20% being the testing dataset. The test dataset is then used for testing the model's accuracy.

IV. RESULTS

To check the accuracy of the model we used the R^2 metric which one of the best scoring techniques for a regression model.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Where,
SSres is the sum of squares of the residual errors.
SStot is the total sum of the errors.

The corresponding Models and Accuracies(R^2*100) have been shown in the table below:

Model	**Accuracy**
Decision Tree	94.81
Random Forest	98.63
XGBoost	98.61
SVR	93.43
ANN	94.54

Now, the graphs of actual values vs predicted values of each model have been shown:

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 666-670 www.ijaem.net ISSN: 2395-5252

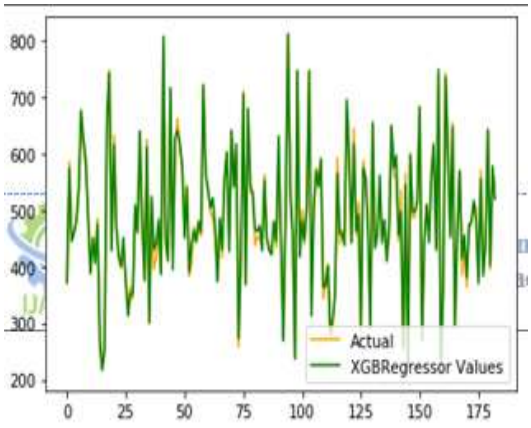

Fig3. Actual vs XGBRegressor Values

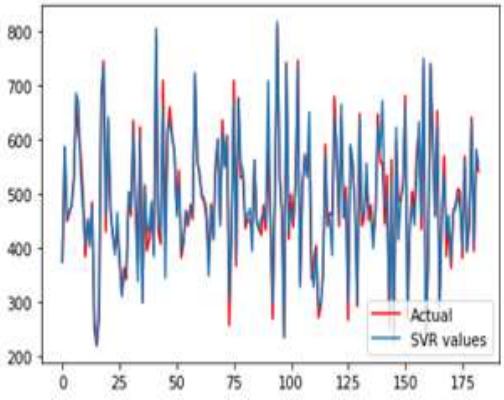

Fig 4. Actual vs SVR Values

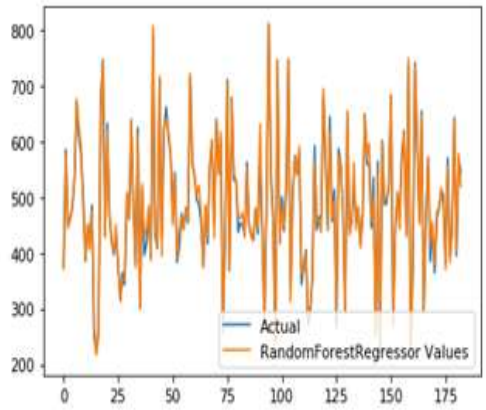

Fig 5. Actual vs RandomForestRegressorValues

Theanalysis shows us that the Random Forest Regression technique is best suited for this data. Fig 8 shows comparison of model accuracies.

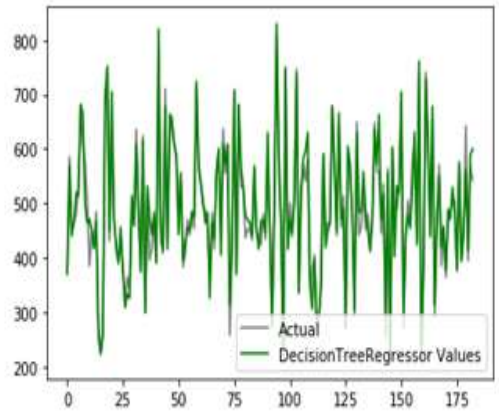

Fig6Actual vs DecisionTreeRegressor Values

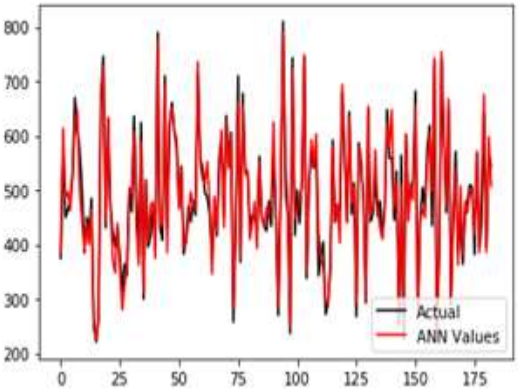

Fig 7. Actual vs ANN Values

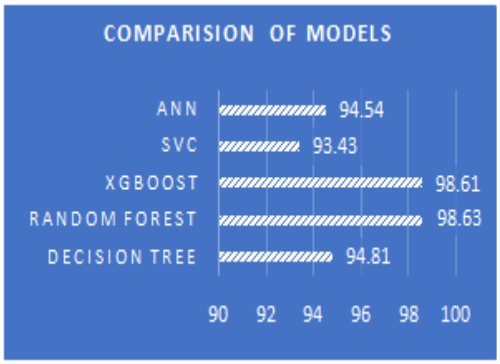

Fig 8 Comparision of models

This suggests that although Random Forest has performed well to predict the tensile strength of steel, XGBoost can also be used to deploy the model.

Accuracy as high as 98.63 suggests that these models can be relied upon if need be.

V. CONCLUSION AND FUTURE SCOPE

This research shows us that the models being built can be useful when time mandates to predict the mechanical properties of steel without actually testing or processing physically. This

process can be improved by using a considerably large amount of data to properly train our models and learn better. The accuracies of 98.63,98.61 is appreciable in the field of material science. Further, these models

can be deployed in real-time, where data could be fed live and the model would re-train. An application can also be made for a better and organised use. Although this paper focuses on the tensile strength of steel, properties like yield strength, elongation %, Recrystallisation Temperature, etc can be predicted. Moreover, since the composition of different elements and temperature were the variables used in the project, features like manufacturing process, cold working/hot working may also be used to train our models better.

REFERENCES

[1]. Wang, Baoji &Kaiqiang, Wu & Li, Yin-Sheng & Chen, Xie& Liu, Yi & Zhang, Yong & Xu, Longfei. (2020). Prediction and Analysis of Tensile Properties of Austenitic Stainless Steel Using Artificial Neural Network. Metals. 10. 234. 10.3390/met10020234.

[2]. Agrawal, Ankit & Deshpande, Parijat&Cecen, Ahmet & Gautham, Bp & Choudhary, Alok &Kalidindi, Surya. (2014). Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters. Integrating Materials and Manufacturing Innovation. 3. 10.1186/2193-9772-3-8.

[3]. Guo, Shun & Yu, Jinxin& Liu, Xingjun& Wang, Cuiping& Jiang, Qingshan. (2019). A predicting model for properties of steel using the industrial big data based on machine learning. Computational Materials Science. 160. 95-104. 10.1016/j.commatsci.2018.12.056.

[4]. Xiong, Jie& Zhang, Tong-Yi & Shi, San-Qiang. (2020). Machine Learning of Mechanical Properties of Steels.

[5]. N. Sandhya, Sowmya, V., Bandaru, C. R., & Babu, R. (2019). Prediction of mechanical properties of steel using data science techniques. International Journal of Recent Technology and Engineering,8(3),235–241.https://doi.org/10.35940/ijrte.c3952.098319

[6]. HnBhadeshia, H.K.D., 1999. Neural networks in materials science. ISIJ International, 39(10): 966-979