Machine Learning for predicting the Fatigue strength of steels

ME 793- 2022
Stage III Presentation

Team ID: 09
Latika Jitendrakumar Patel
Moitreyee Sarkar
Polamreddy Mithun Balram

CONTENT

Background

Where did we find it?

We wanted to apply Data-driven and Machine Learning techniques to predict material property. One with great significance and challenge in materials science was accurate prediction of fatigue strength of steels. We found only a few papers in this domain which predicted fatigue strength of steel.

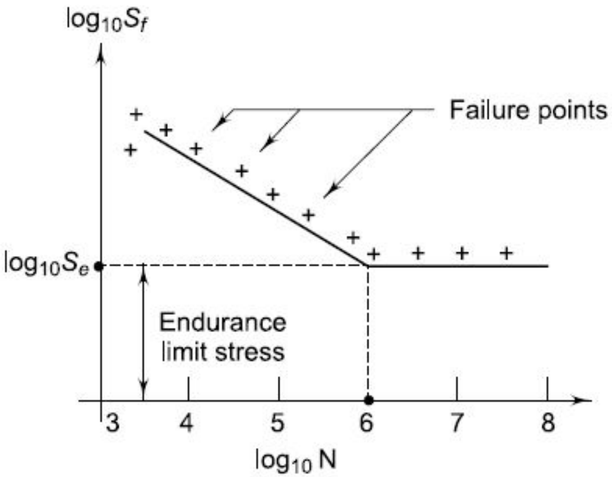
Background

Why it's important?

- *Fatigue* strength is the most important and basic data required for design and failure analysis of a material. It is reported that fatigue accounts for over 90% of all mechanical failures of structural components.
- Here, we have selected *Steel* as the subject as it is a crucial metal from construction and transportation point of view.
- Accurate *prediction* of fatigue strength of steels is of particular significance in materials science because of the extremely high cost and time of fatigue testing and often debilitating consequences of fatigue failures.

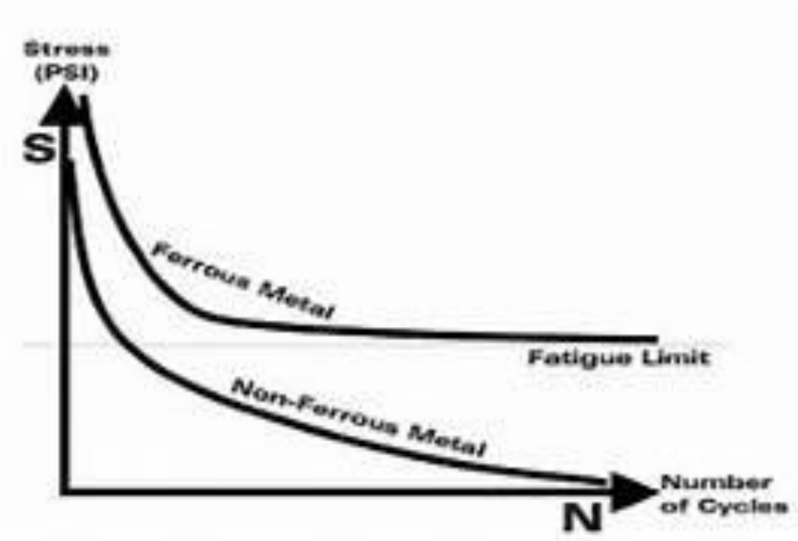
OBJECTIVES

1) Predicting the fatigue endurance limit from auxiliary data
2) Estimating the S-N graph from the endurance limit



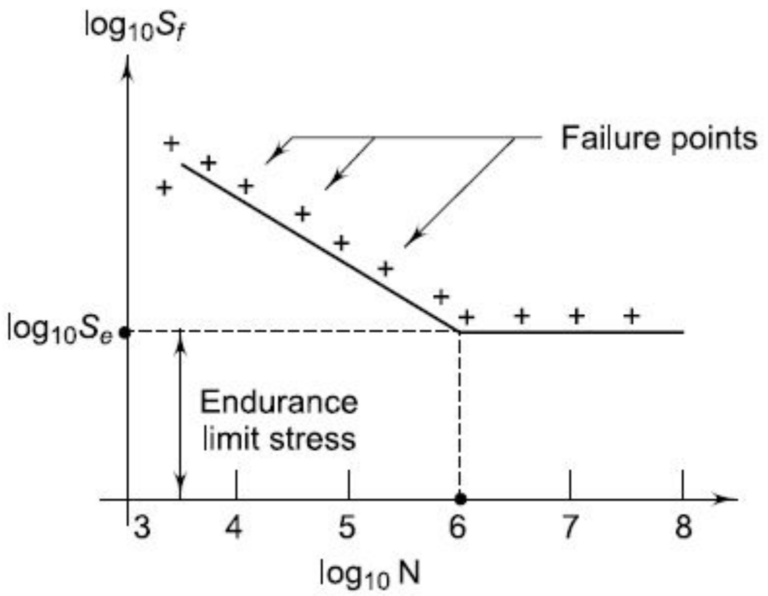
Predicting the Fatigue Endurance limit from auxiliary data

1) The data includes chemical composition of steel, processing, heat treatment conditions & some mechanical properties like hardness, charpy impact value etc.
2) We will train a regression to predict the endurance limit i.e. the threshold stress below which the material will have infinite life.



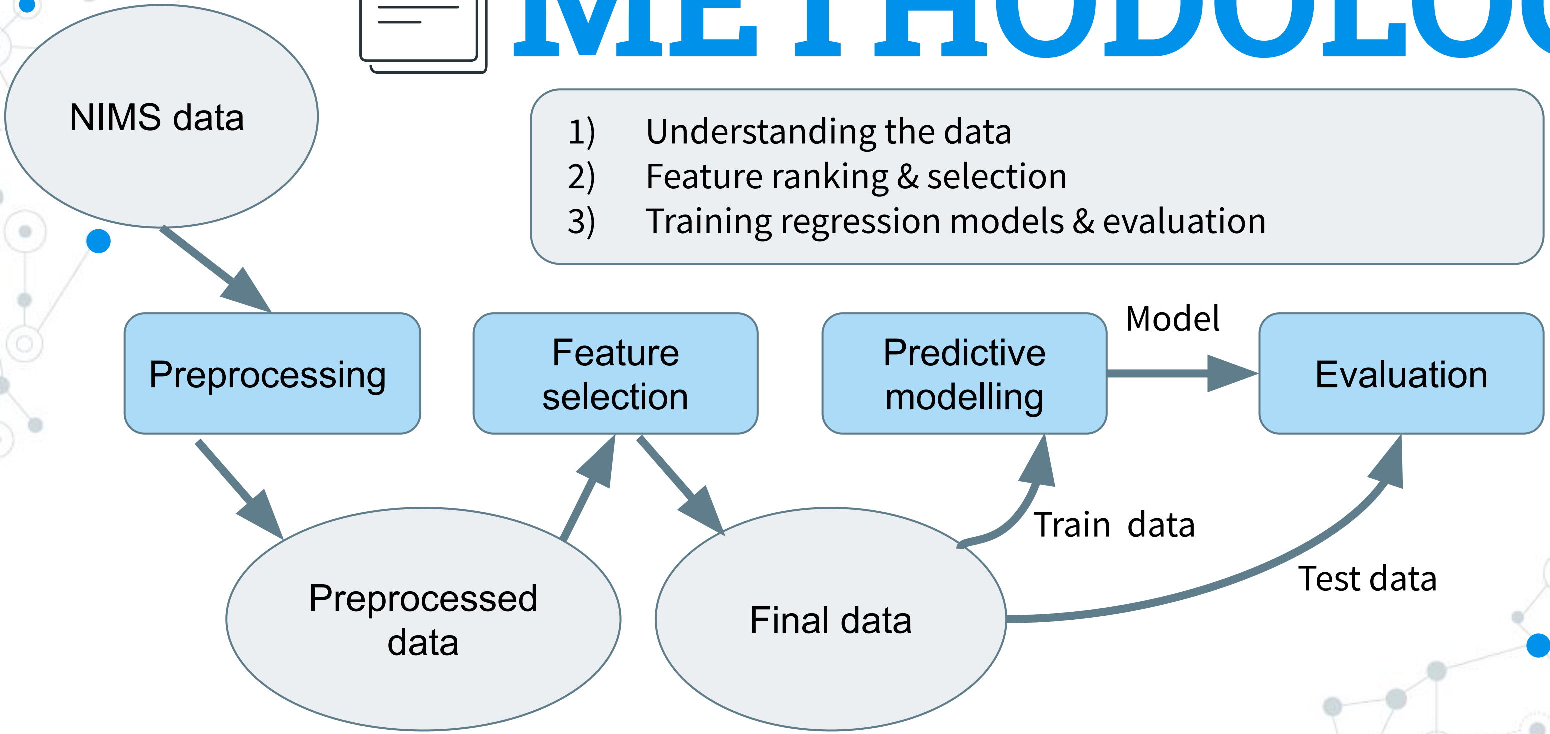
Estimating the S-N curve from the Endurance limit

1) The S-N curve of steel can be modelled using 2 straight lines
2) We directly get the second line from the endurance limit.
3) For the first line, we can get its y-intercept i.e. yield strength by performing a uniaxial tensile test & intersect it with the endurance limit at 10^6 cycles.
4) This is much easier than performing the much more complex fatigue tests for different values of N



METHODOLOGY

1) Understanding the data
2) Feature ranking & selection
3) Training regression models & evaluation



WORK DONE

1. Load and Check Data

The data used in this work has

- ➢ 437 instances/rows
- ➢ 25 features/columns (composition and processing parameters)
- ➢ 1 target property (fatigue strength)
- ➢ The 437 data instances include 371 carbon and low alloy steels, 48 carburizing steels, and 18 spring steels. This data pertains to various heats of each grade of steel and different processing conditions
- ➢ 3 grade Steels

1. Load and Check Data

Fatigue Dataset for Steel from National Institute of Material Science (NIMS)

- ➢ Chemical composition - %C, %Si, %Mn, %P, %S, %Ni, %Cr, %Cu, %Mo (all in wt. %)
- ➢ Upstream processing details - ingot size, reduction ratio, non-metallic inclusions
- ➢ Heat treatment conditions - temperature, time and other process conditions for normalizing, through-hardening, carburizing-quenching and tempering processes
- ➢ Mechanical properties - YS, UTS, %EL, %RA, hardness, Charpy impact value (J/cm2), fatigue strength
- ➢ Corresponding endurance limit stress

1. Load and Check Data

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
Sl. No.	NT	THT	THt	THQCr	CT	Ct	DT	Dt	QmT	TT	Tt	TCr	C	Si	Mn
1	885	30	0	0	30	0	30	0	30	30	0	0	0.26	0.21	
2	885	30	0	0	30	0	30	0	30	30	0	0	0.25	0.18	
3	885	30	0	0	30	0	30	0	30	30	0	0	0.26	0.27	
4	885	30	0	0	30	0	30	0	30	30	0	0	0.26	0.23	
5	885	30	0	0	30	0	30	0	30	30	0	0	0.22	0.19	
6	885	30	0	0	30	0	30	0	30	30	0	0	0.26	0.22	
7	885	30	0	0	30	0	30	0	30	30	0	0	0.28	0.27	
8	885	30	0	0	30	0	30	0	30	30	0	0	0.27	0.25	
9	885	30	0	0	30	0	30	0	30	30	0	0	0.27	0.23	
10	885	30	0	0	30	0	30	0	30	30	0	0	0.22	0.26	
11	885	30	0	0	30	0	30	0	30	30	0	0	0.23	0.26	
12	865	865	30	24	30	0	30	0	30	550	60	24	0.35	0.21	
13	865	865	30	24	30	0	30	0	30	550	60	24	0.35	0.3	
14	865	865	30	24	30	0	30	0	30	550	60	24	0.34	0.26	
15	865	865	30	24	30	0	30	0	30	550	60	24	0.36	0.26	
16	865	865	30	24	30	0	30	0	30	550	60	24	0.33	0.21	
17	865	865	30	24	30	0	30	0	30	550	60	24	0.38	0.29	
18	865	865	30	24	30	0	30	0	30	550	60	24	0.33	0.26	
19	865	865	30	24	30	0	30	0	30	550	60	24	0.32	0.26	
20	865	865	30	24	30	0	30	0	30	550	60	24	0.37	0.27	
21	865	865	30	24	30	0	30	0	30	550	60	24	0.35	0.25	
22	865	865	30	24	30	0	30	0	30	550	60	24	0.37	0.24	
23	865	865	30	24	30	0	30	0	30	550	60	24	0.32	0.21	
24	865	865	30	24	30	0	30	0	30	600	60	24	0.35	0.21	

WORK DONE

2. Visualization

Overview Alerts 95 Reproduction

Dataset statistics

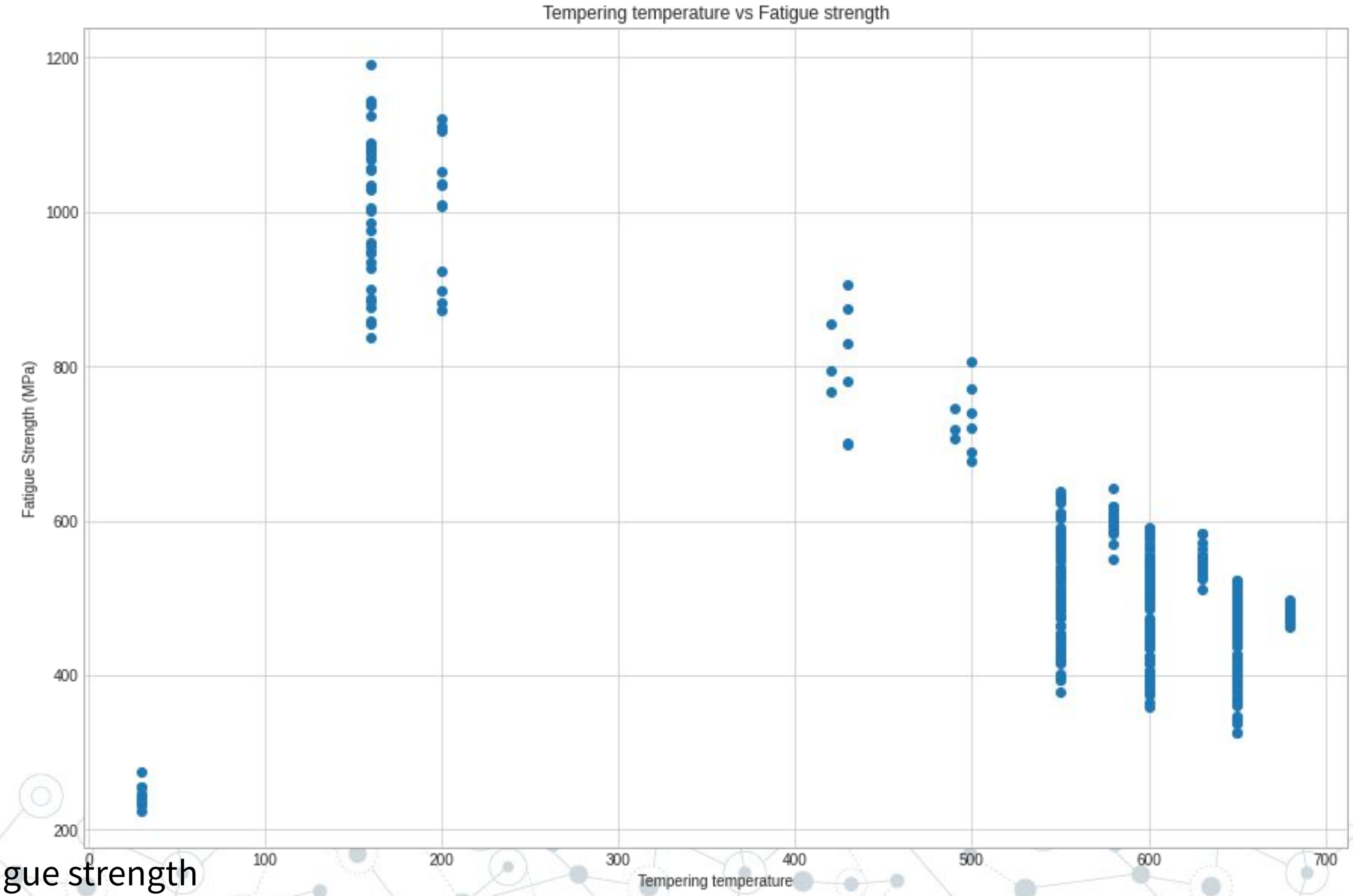
Number of variables	27
Number of observations	437
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	92.3 KiB
Average record size in memory	216.3 B

Variable types

Numeric	20
Categorical	7

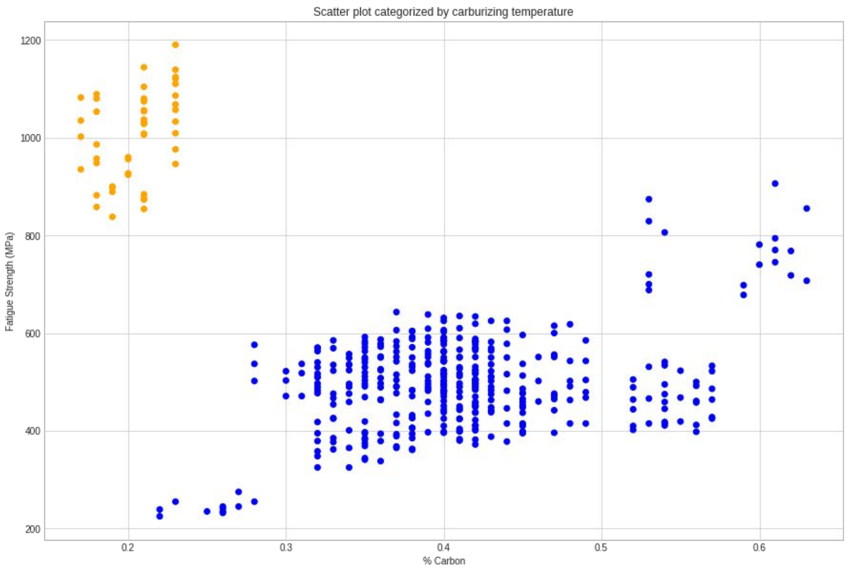
Profile Report of Dataset

2. Visualization

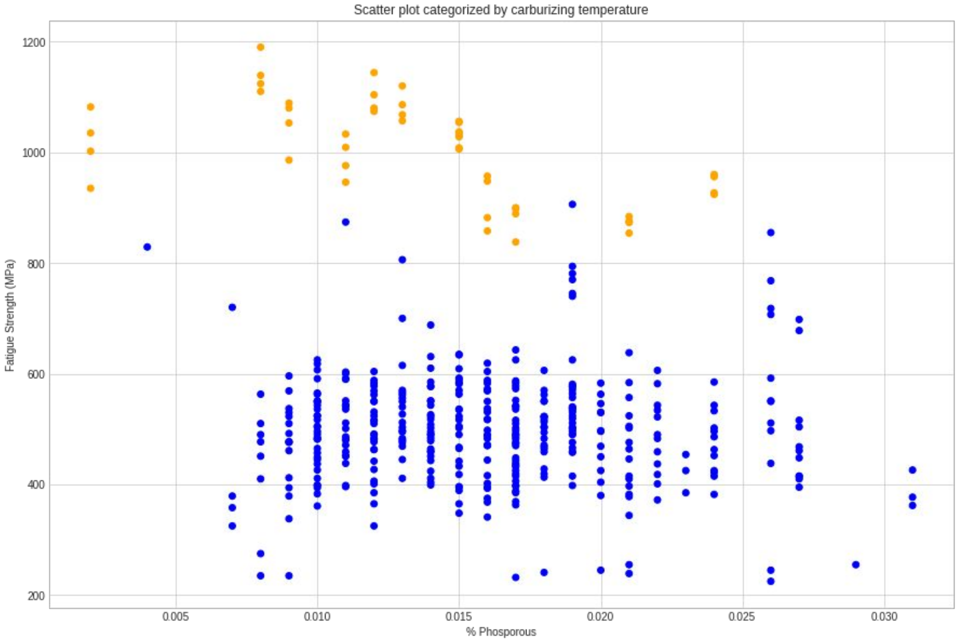


Tempering temperature vs Fatigue strength

2. Visualization



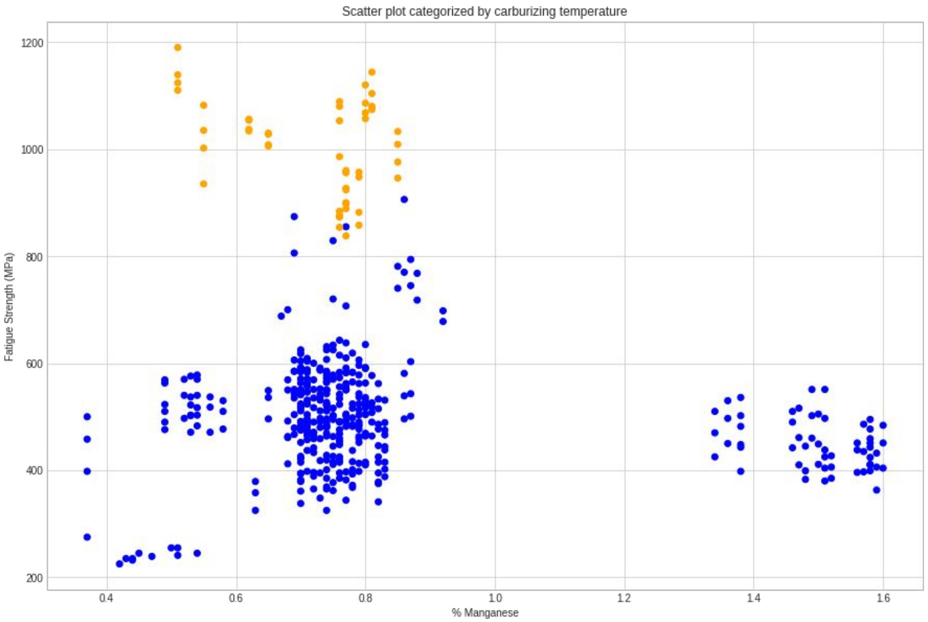
% Carbon



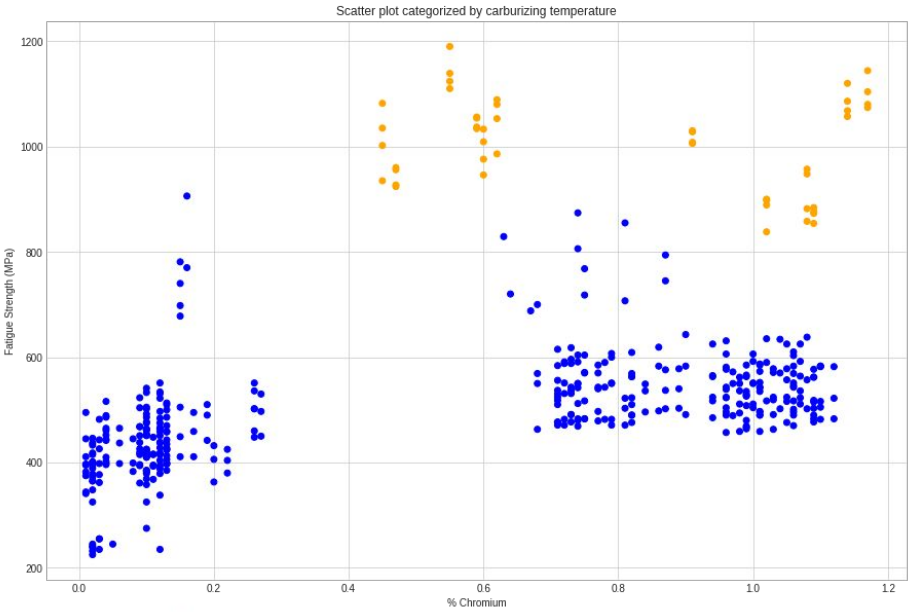
% Phosphorus

Scatter plot categorized by carburizing temperature

2. Visualization



% Manganese

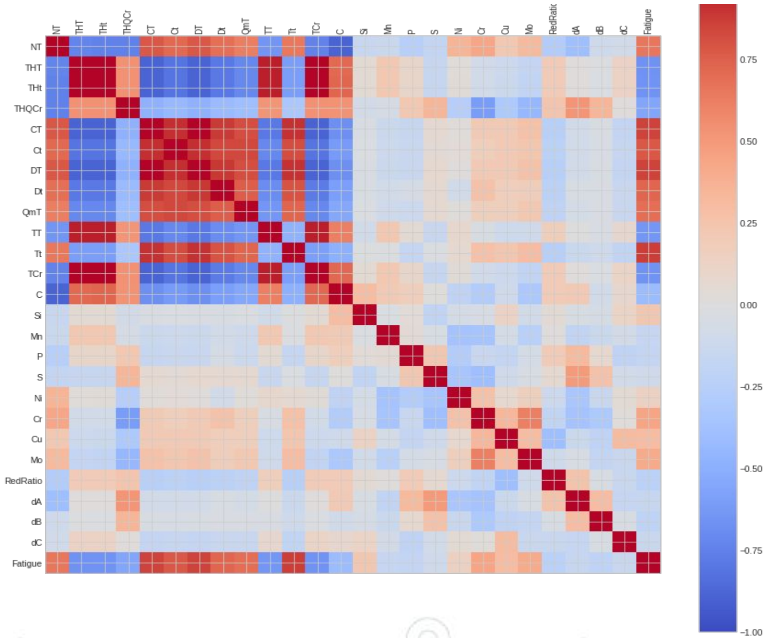


% Chromium

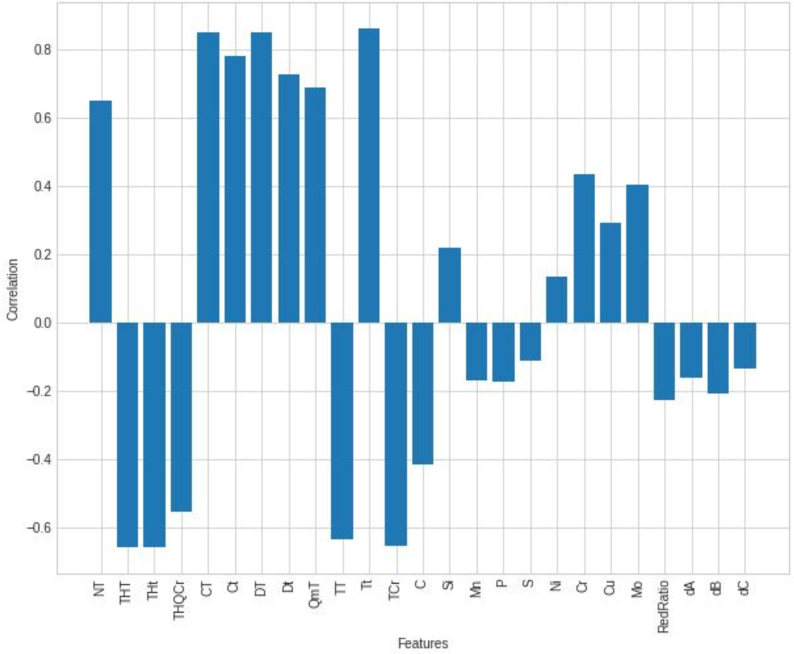
Scatter plot categorized by carburizing temperature

2. Visualization

Correlation Matrix

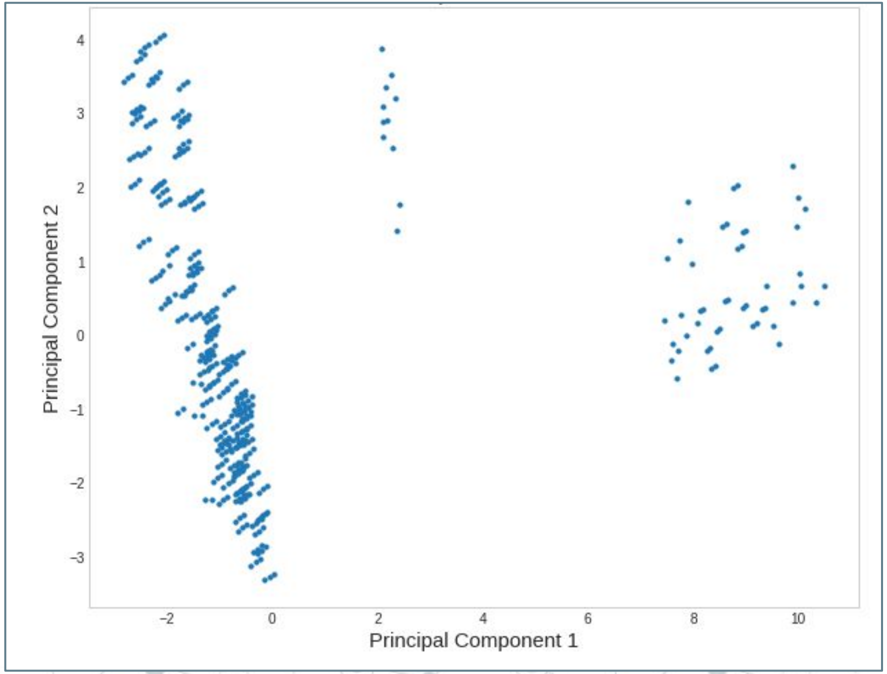


Correlation of Features with Fatigue Strength

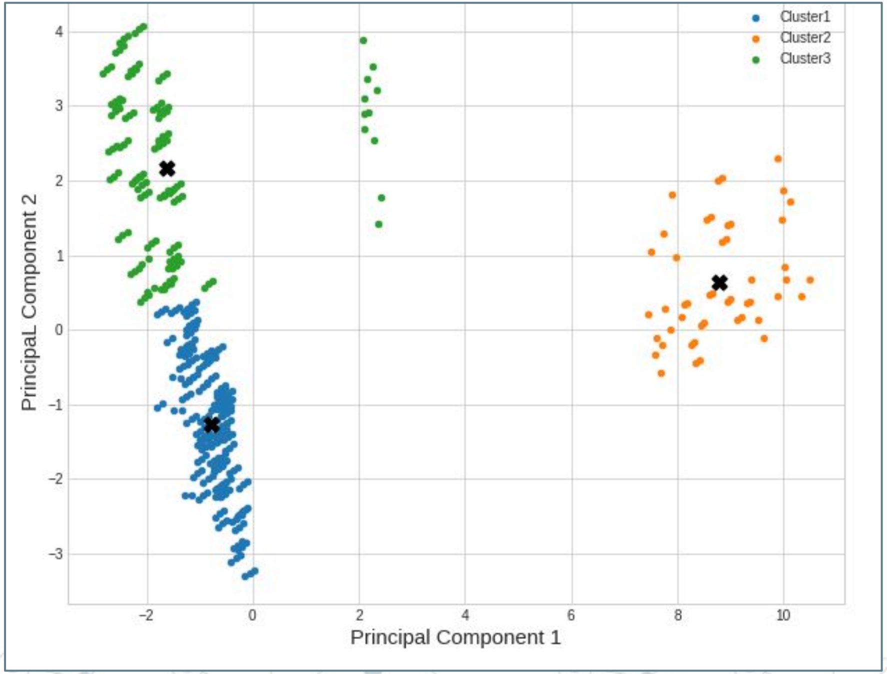


2. Visualization

PCA
n =2



K-means clustering
k=3



WORK DONE

3. Feature Selection

There are two ways for dimensionality reduction-

1. ***Feature Extraction***
   - Transforms the existing features into new ones based on combinations of the raw features.
   - These higher level features are difficult to understand as can't be directly linked to original

2. ***Feature Selection***
   - We can compute information gain for each feature individually with respect to the fatigue strength & rank the available features.
   - We can discard some features with very less information gain to avoid overfitting the data.
   - Finding the best features helps extract valuable information and discover new knowledge.

3. Feature Selection

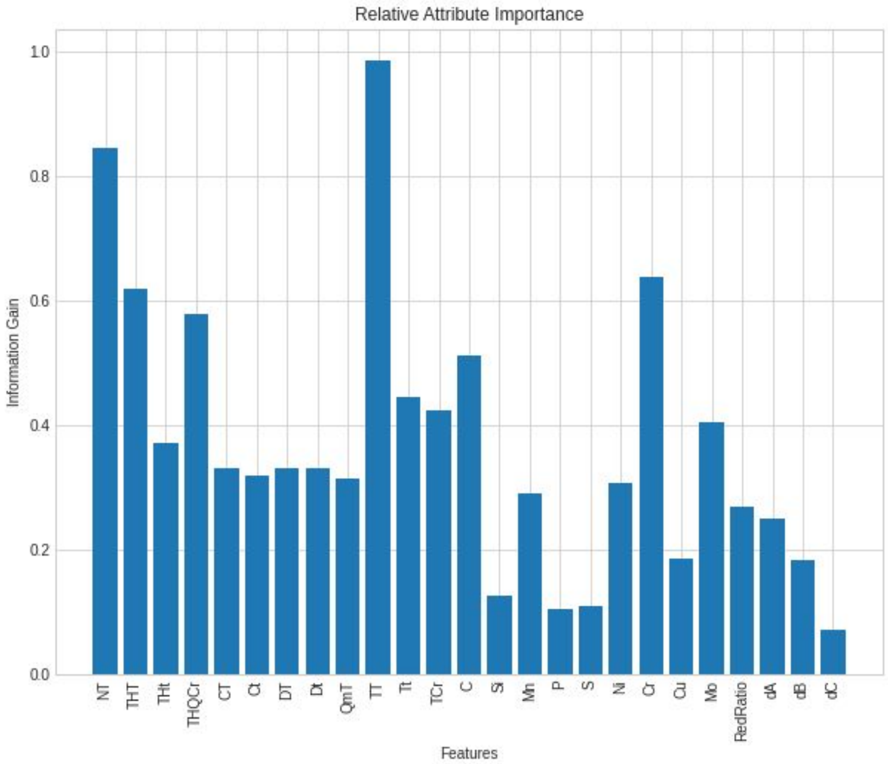
Features with high predictive potential

- TT - Tempering temperature
- NT - Normalization temperature
- THT - through Hardening temperature

Features with relatively lower predictive potential

- dC- Area proportion of Isolated Inclusions
- P - % Phosphorus
- S - % Sulphur
- Si - % Silicon

21 Best Features selected for modelling



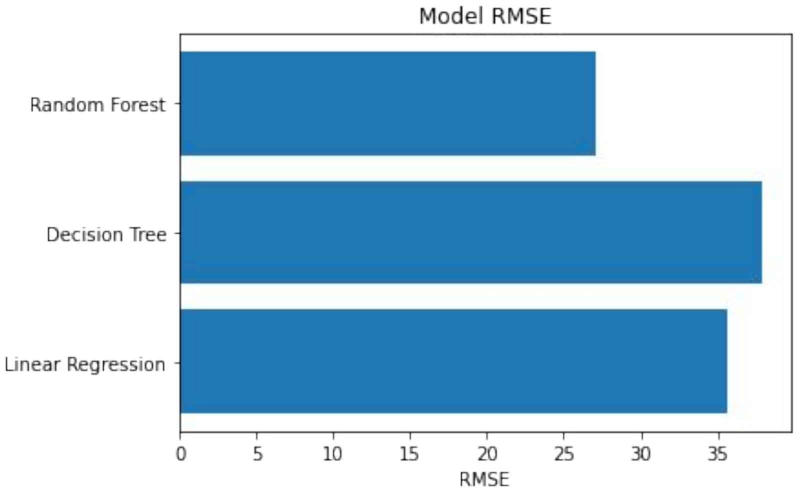
WORK DONE

4. Modelling

We have currently trained 7 ML models including:

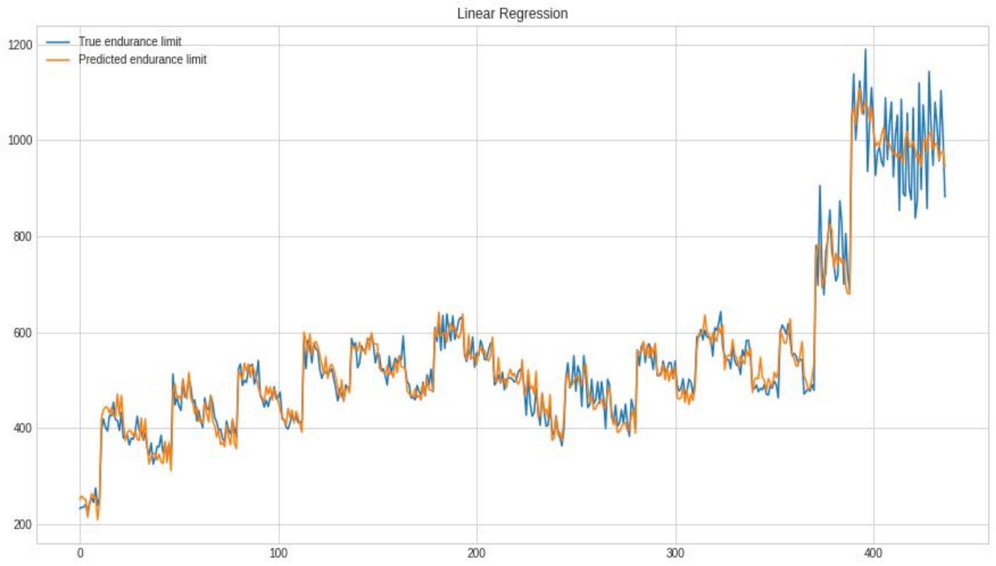
1) Multilinear regression
2) Decision Trees
3) Random Forests

We have evaluated the models using Leave one out cross-validation (LOOCV) since, the dataset is pretty small. The RMSE scores of the models indicate that random forest is the best model.



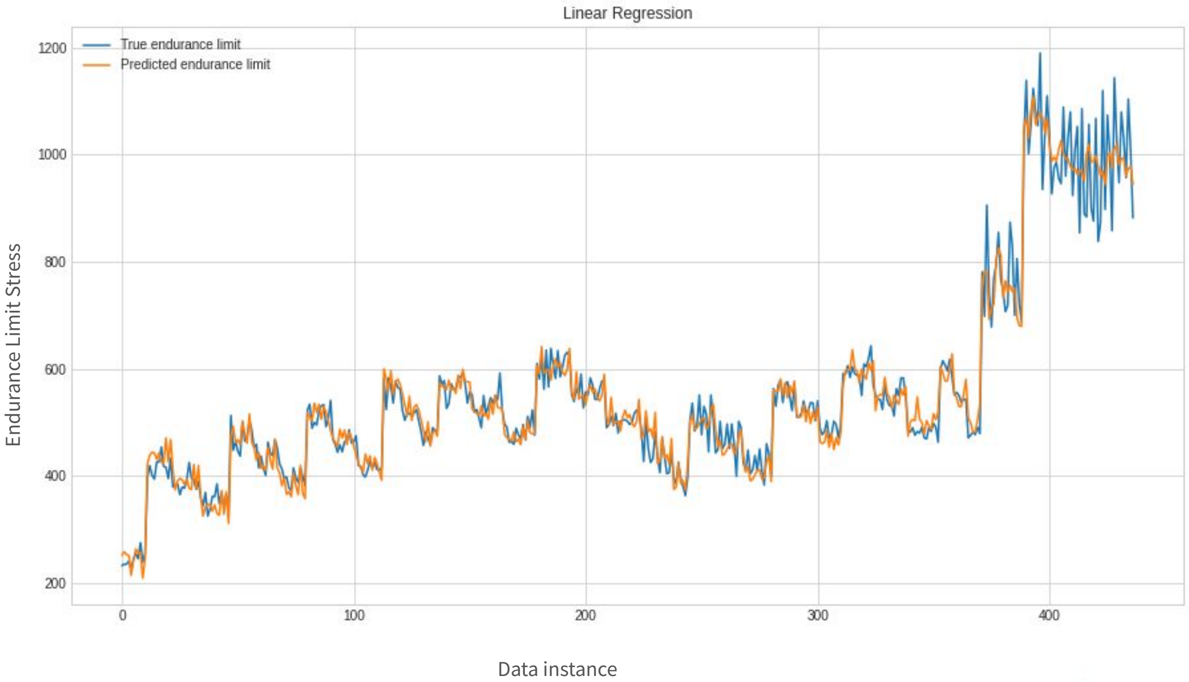
4. Modelling

Now let's look at the model prediction & true values & analyze any overfitting or underfitting the models we have trained



Baseline Linear regression model

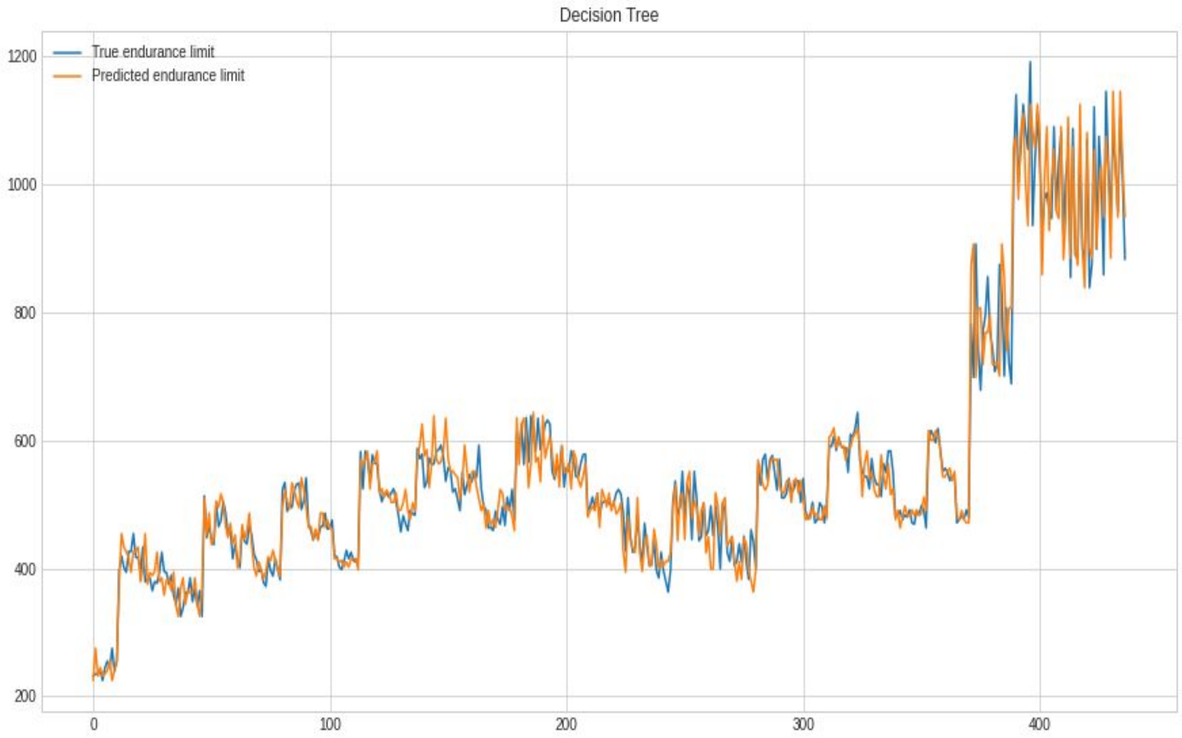
4. Modelling



Linear Regression

We can clearly see that the linear regression model somewhat captures a simple model & ignores some fluctuations and just make them smooth. This is evident in the right part where the predicted values are more or less the mean.

4. Modelling



Decision Tree

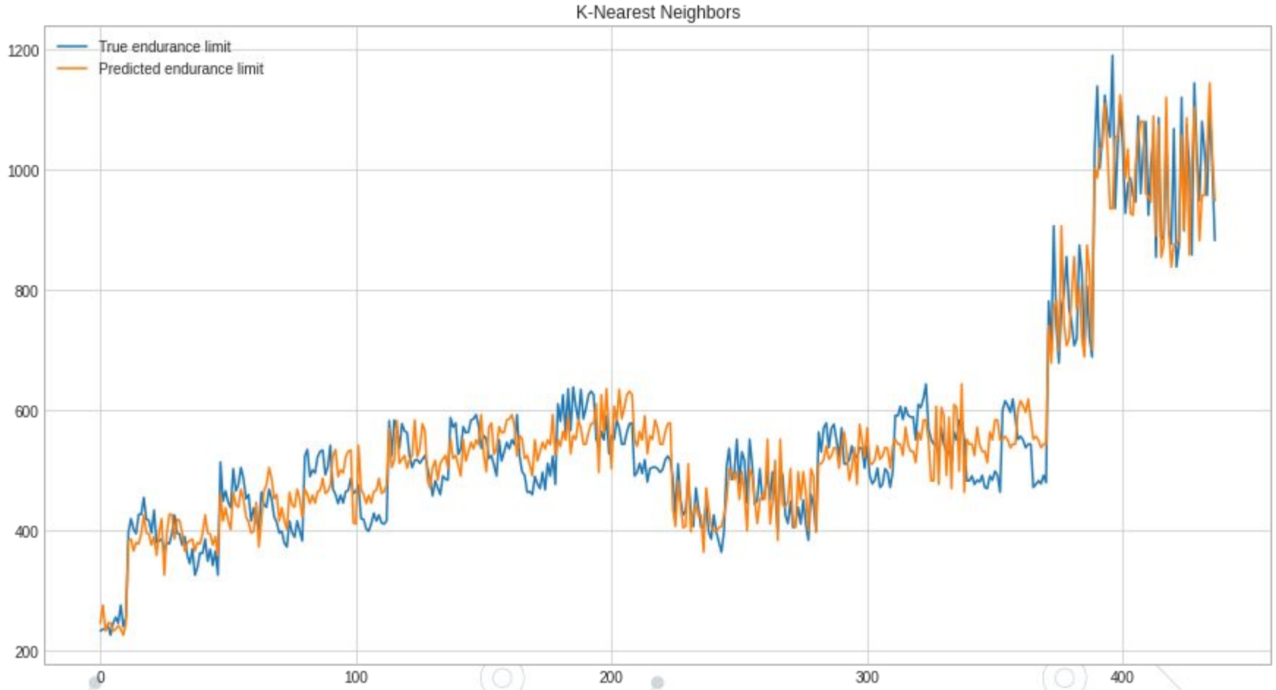
The decision tree model fluctuates more than the true values indicating high variance which is observed when overfitting to the training data. This is clearly visible on the right part where the fluctuations are much more than the true values

4. Modelling

The random forest model is more complex than linear regression & more simpler than decision tree hence, it has a lower RMSE compared to other models.

Random Forest

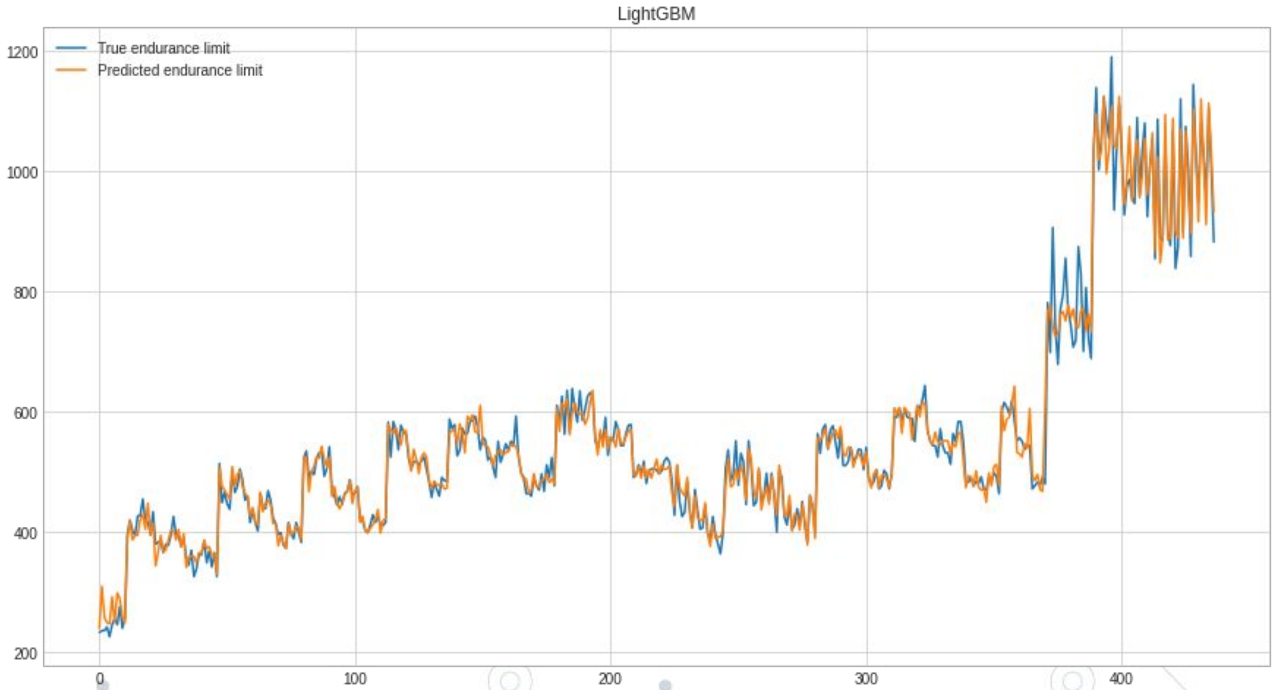
4. Modelling



KNN

KNN is underfitting for low carbon steels and underfitting for high carbon steels. Hence, it has a high RMSE and a low R2 score.

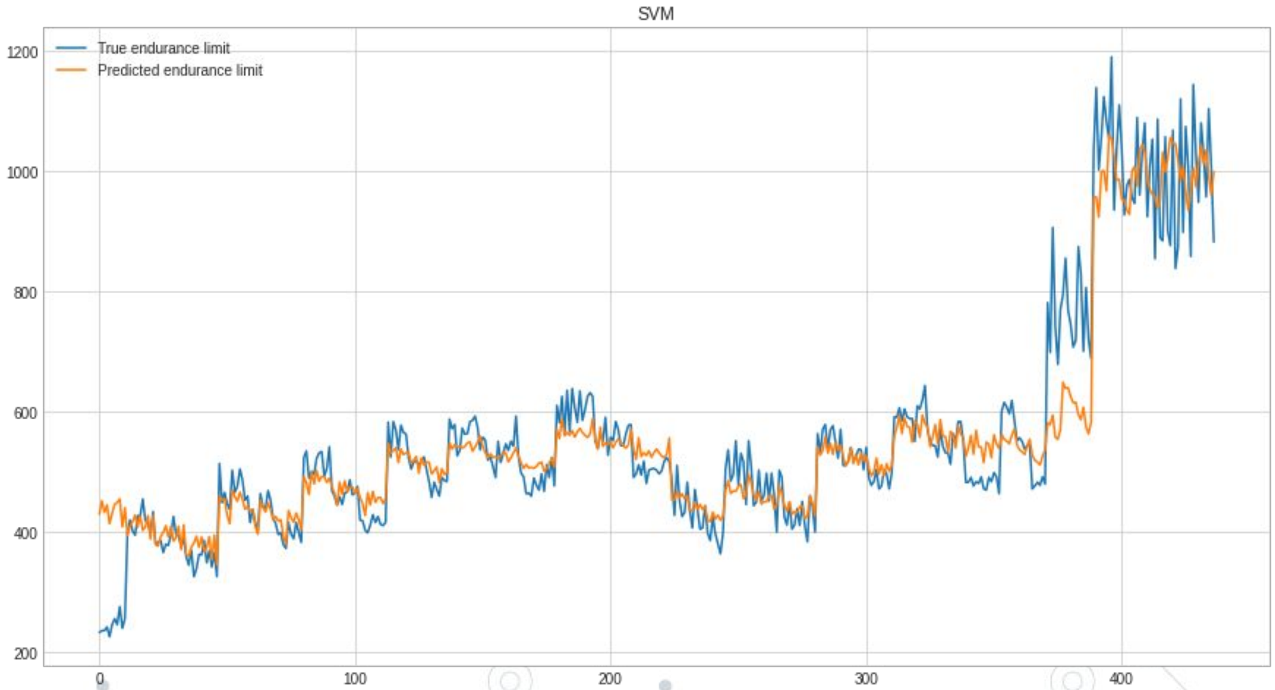
4. Modelling



LightGBM

LightGBM is the best model in terms of RMSE, it fits almost perfectly to data and also generalises well on unseen data.

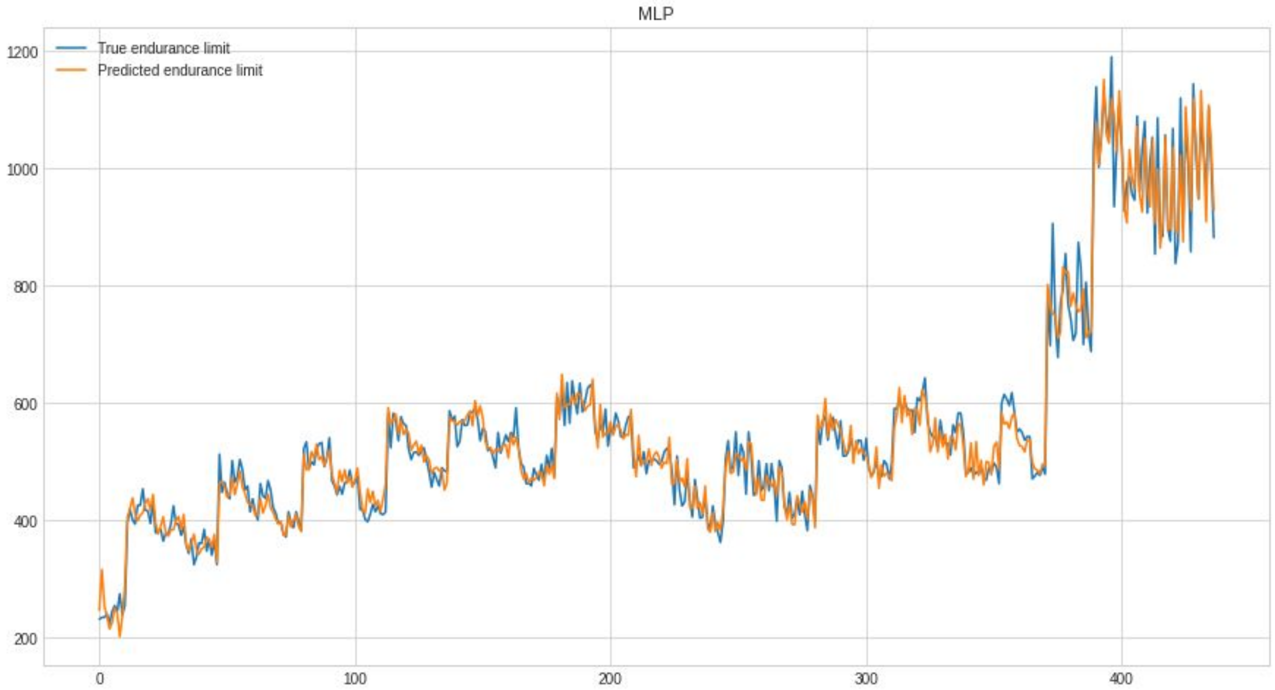
4. Modelling



SVM

SVM had the highest value of RMSE among all the other models due to underfitting. So adding more data points would help increase the accuracy of SVM and also, the variables are highly correlated.

4. Modelling

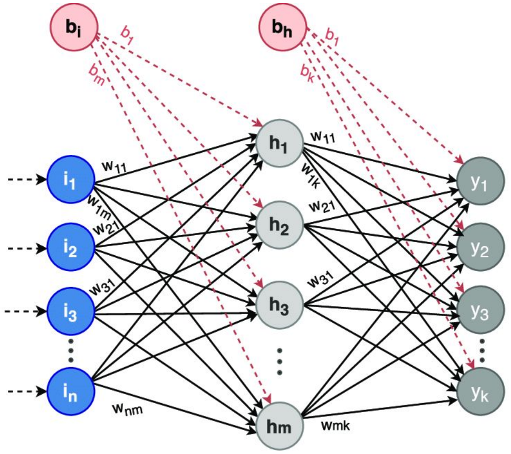


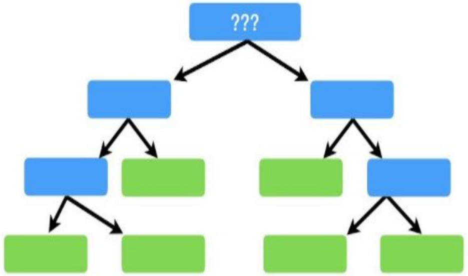
While the model is able to capture most of the values shown by the low RMSE, it still shows some amount of fluctuation in the right side. However, with some hyperparameter tuning, the RMSE can be improved as Neural networks can capture complex non-linear relationships.

Neural Network - MLP (35,35)

CONCLUSION

Model	RMSE Value
Linear Regression	35.69
Decision Tree Regression	37.89
Random Forest Regression	27.14
K Nearest Neighbors	52.29
LightGBM	25.93
SVM	64.31
MLP	29.04





References

1. Lei He, ZhiLei Wang, Hiroyuki Akebono, Atsushi Sugeta, Machine learning-based predictions of fatigue life and fatigue limit for steels, Journal of Materials Science & Technology, Volume 90, 2021, Pages 9-19, ISSN 1005-0302, https://doi.org/10.1016/j.jmst.2021.02.021
2. Agrawal, A., Deshpande, P.D., Cecen, A. et al. Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters. Integr Mater Manuf Innov 3, 90–108 (2014). https://doi.org/10.1186/2193-9772-3-8
3. Shibanuma, K., Ueda, K., Ito, H., Nemoto, Y., Kinefuchi, M., Suzuki, K., & Enoki, M. (2018). Model for predicting fatigue life and limit of steels based on micromechanics of small crack growth. *Materials & Design*, *139*, 269-282. https://www.sciencedirect.com/science/article/pii/S0264127517310122
4. Furuya, Y., Nishikawa, H., Hirukawa, H., Nagashima, N., & Takeuchi, E. (2019). Catalogue of NIMS fatigue data sheets. *Science and Technology of Advanced Materials*, *20*(1), 1055-1072.https://www.tandfonline.com/doi/full/10.1080/14686996.2019.1680574

Thank You