Indian Institute Of Technology Roorkee

THE UNIVERSITY OF BRITISH COLUMBIA

Machine Learning Application in Optimization of woven two-layered hydraulic pipes

Presented By-

Dushyant Singh Udawat, Undergrad, IIT - Roorkee

Table of Contents

Product Description

- Two-layered Woven Composite Hydraulic Pipe
  - Expected Application
    - to transport working fluids (e.g., water for oil and gas firefighting, water sand slurries in mining activities, among others
- Snaking Issue
  - axial buckling of the hose structure under pressurized, causing interference with other personnel or equipment on the worksite

  - Implications on finance
  - Challenges
  - That’s why the use of Machine Learning

Overall Objectives

- Reduce the Number of Snaking, Snaking Length, Snaking Amplitude in the hose in such a way as to decrease the cost to the company.
- Get Insights on the data provided.
- Make prediction models to predict the snaking in a hose.
- Provide the manufacturer with an AI-based decision-making tool.

Data Summary - Inputs

- We have four Parameters:
  - Inner Modulus
  - Outer Modulus
  - Axial Load
  - Friction factor between the hose and ground.

Parameter Name	Units	Input Variables Min-Max (Levels)
Inner Modulus	GPa	0.10-0.18 (5)
Outer Modulus	GPa	0.07-0.15 (5)
Axial Load	kN	6-14(5)
Friction Coefficient	-	0.01 – 45 (10)

These Parameters are changeable either during operations or during design process. Simulations gives us the data which is then used in Machine Learning methods to give predictions and get insights.

The data is generated for five values of each parameter. Therefore, giving a total of 5*5*5*10 1250 samples.

Data Summary - Outputs

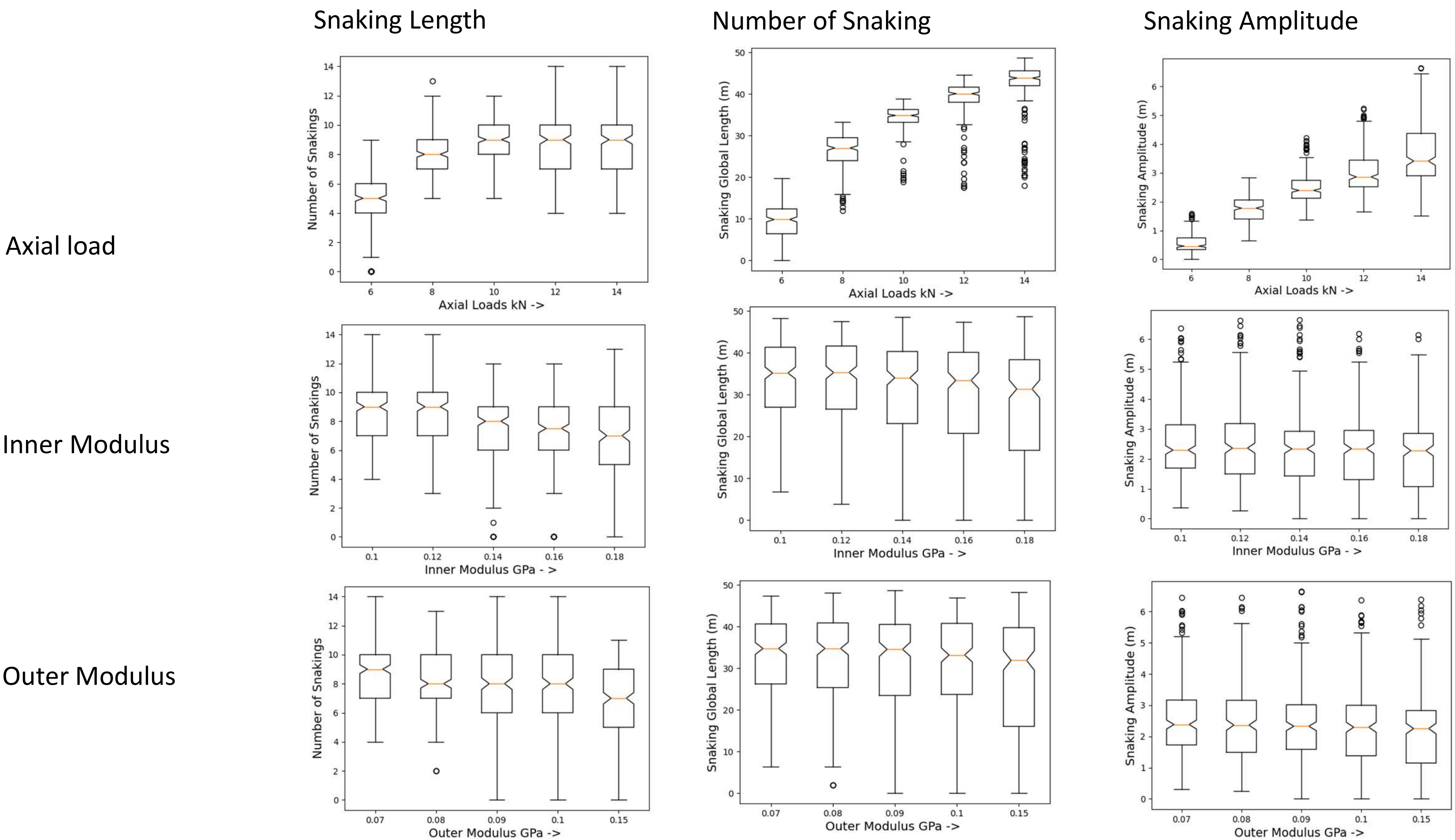
- Output Data Description
  - We have three output variables regarding snaking behavior
    - Snaking Length,
    - No. of Snaking,
    - Snaking Amplitude.
  - Their Description is as shown in the table.
  - Their Correlation with input variables can be seen in the adjoining figure.

Pearson Correlation Coefficient –

Input Variable	Snaking Length	No. of Snaking	Snaking Amplitude
Inner Modulus	-0.321153	-0.152724	-0.112964
Outer Modulus	-0.256012	-0.122518	-0.10633
Axial Load	0.442826	0.87367	0.832798
Friction Coefficient	0.0126775	0.0119381	-0.00499214

Parameter Name	**Units**	**Input Variables Min-Max (Levels)**
Number of Snaking	-	0-14 (7.72)
Snaking Length	m	0-49 (31)
Snaking Amplitude	m	0-6.64 (2.31)

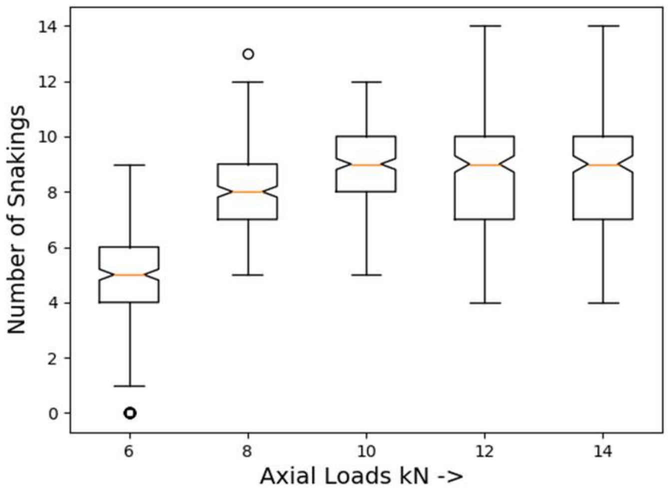
STAGE I
Pre-Analysis of Data

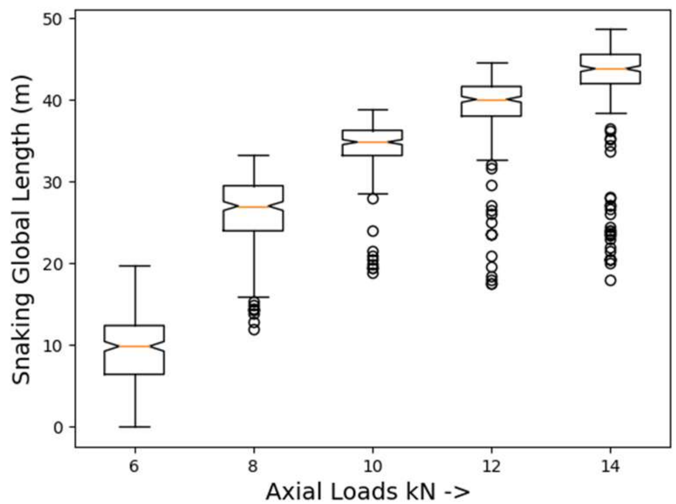


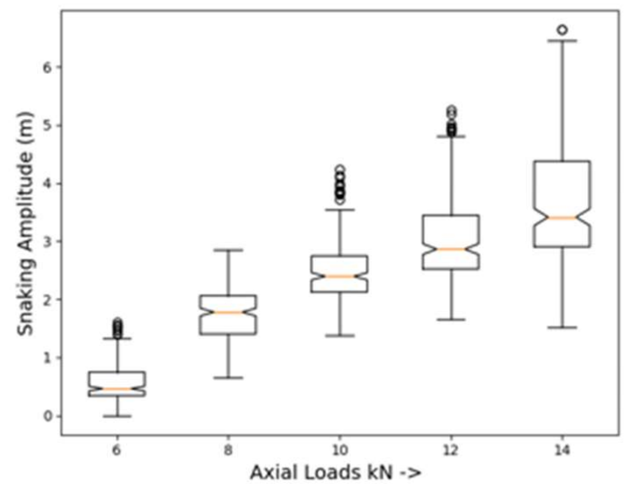
Snaking Length
Number of Snaking
Snaking Amplitude
Axial load
Inner Modulus
Outer Modulus
Number of Snakings
Snaking Global Length (m)
Snaking Amplitude (m)
Axial Loads kN ->
Inner Modulus GPa - >
Outer Modulus GPa ->

Insights

- The single most important factor is the **axial Load**, the values changes a lot as we increase the loads from 6kN to 8kN.
- After load>6kN, the variation in output parameters from modulus is less. This is corroborated by Linear Regression on separate Data, when Axial Load = 6 kN, and axial load > 6kN.







- **After load>6kN, the variation in output parameters from modulus is lesser. (Linear Regression** on separate Data)
- After Axial Load > 6kN, Role of Inner Modulus is more important than Outer Modulus.
- Generally, as inner and outer modulus increases, the tendency for snaking reduces slightly (Correlation)
- For zero Snaking Condition –

    Axial load < 7kN, Inner modulus > 0.13, Outer modulus> 0.13.
- No dependence on the friction coefficient

    (Correlation, Explainable AI, Models testing)

ANOVA

ANOVA Analysis Shows – (Show fig. for Number of Snaking)

1. Significant Variables – Inner , Outer and Axial load.
2. Insignificant = Friction and all its interactions.
3. Interaction between inner and outer modulus is Insignificant. (for Numbers of Snaking)
4. Interaction of Inner and Outer Modulus with Axial load are significant.
5. Interactions are similar in Length and Number of Snaking.
6. Interactions in case of amplitude, are not significant at all.
7. Interactions of Inner and outer modulus are not significant for NUM and AMP but it is significant for LEN.

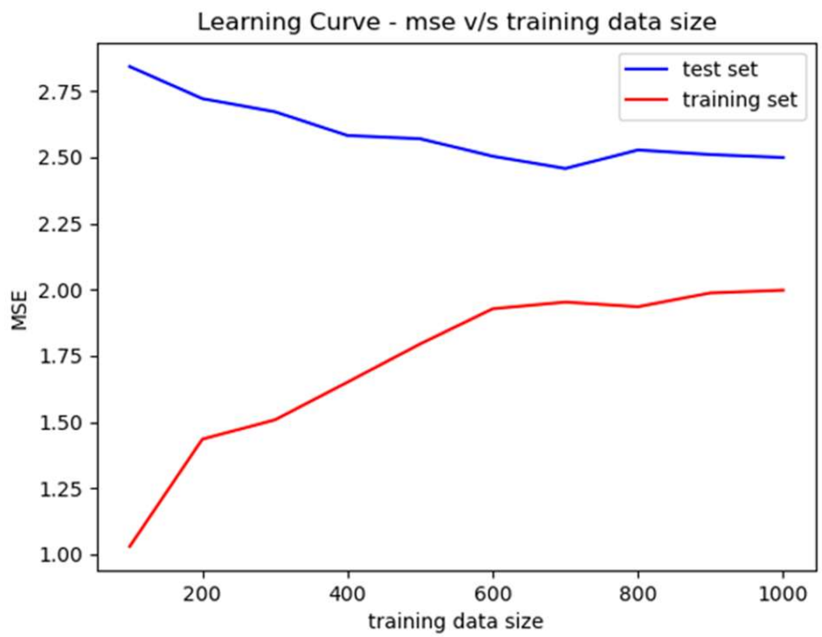
	df	sum_sq	mean_sq	F	PR(>F)
C(INNER)	4.0	792.2432	198.060800	101.760551	1.663331e-65
C(OUTER)	4.0	497.2672	124.316800	63.872033	1.083724e-44
C(AXIAL)	4.0	3027.5232	756.880800	388.873553	4.626749e-162
C(FRIC)	9.0	14.7592	1.639911	0.842561	5.770418e-01
C(INNER):C(OUTER)	16.0	22.5648	1.410300	0.724590	7.698833e-01
C(INNER):C(AXIAL)	16.0	195.9488	12.246800	6.292215	4.767914e-13
C(INNER):C(FRIC)	36.0	75.7728	2.104800	1.081413	3.457641e-01
C(OUTER):C(AXIAL)	16.0	286.0048	17.875300	9.184050	2.155595e-20
C(OUTER):C(FRIC)	36.0	80.5888	2.238578	1.150146	2.552195e-01
C(AXIAL):C(FRIC)	36.0	91.4528	2.540356	1.305195	1.131908e-01
C(INNER):C(OUTER):C(AXIAL)	64.0	268.5232	4.195675	2.155672	2.036038e-06
C(FRIC):C(OUTER):C(AXIAL)	144.0	386.7792	2.685967	1.380008	5.445599e-03
C(INNER):C(FRIC):C(AXIAL)	144.0	334.6752	2.324133	1.194103	8.147473e-02
C(INNER):C(OUTER):C(FRIC)	144.0	308.7792	2.144300	1.101708	2.212602e-01
Residual	576.0	1121.0928	1.946342	NaN	NaN

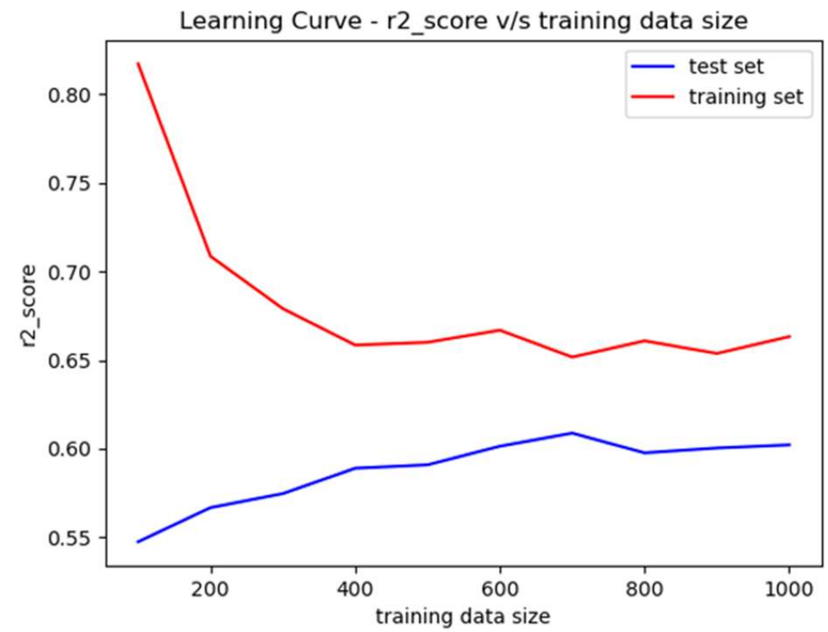
STAGE – II
Machine Learning Prediction Models

Procedure Employed to select Machine Learning Model -

1. Design the model and train it.
2. Optimize hyper-parameters to get the best cross-validation accuracy based on r2_score (Coefficient of Determination) and MSE (Mean Squared Error).
3. Consider the Importance of Accuracy, Speed, Interpretability and choose a Machine Learning model for each of the different output.

Learning Curve for XGBoost Predictor for Number of Snaking (12 such graphs are there for various models and Output variables)





SELECTION OF BEST MODELS

- We have considered four factors that might help us to choose our model.
  - Interpretability
  - Training Speed
  - Accuracy (Both of them)
- Since dataset is small, training speed is not an issue for us.
- Next slide will go through results of selection of best models.

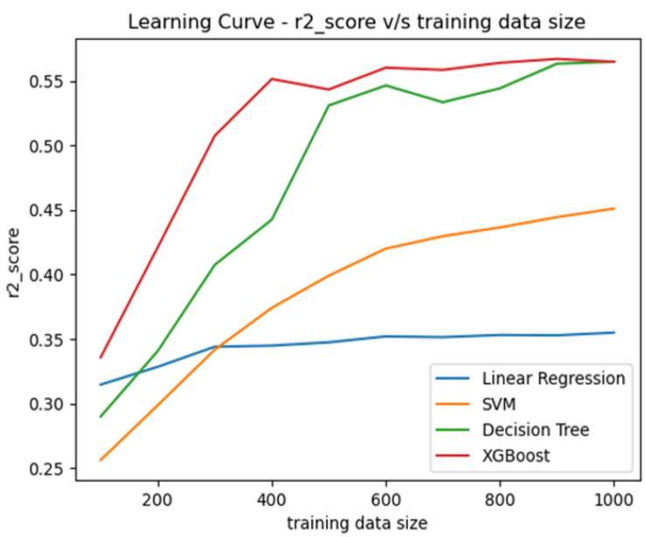
SELECTION OF MODELS -

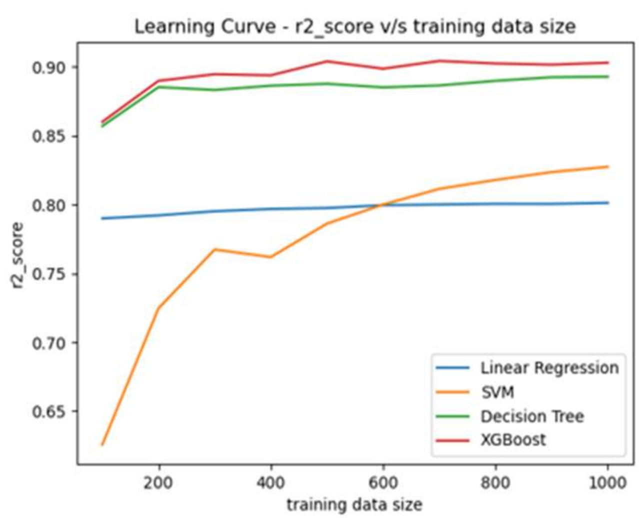
Model	Interpret-ability	R^2 Coefficient of Determination (Highest)	MSE	Training Speed
Number of Snaking				
Linear Regression	High	0.321	4.268	High
SVR	Low	0.545	2.856	Low
Decision Tree(*)	High	0.600	2.508	High
XGBoost(**)	Low	0.602	2.500	Low
Global Length of Snaking				
Linear Regression	High	0.764	41.78	High
SVR	Low	0.873	23.32	Low
Decision Tree (*)	High	0.880	21.24	High
XGBoost (**)	Low	0.881	21.22	Low
Snaking Amplitude				
Linear Regression(*)	High	0.708	0.495	High
SVR	Low	0.712	0.483	Low
Decision Tree	High	0.711	0.490	High
XGBoost(**)	Low	0.723	0.460	Low

Learning Curve Comparison

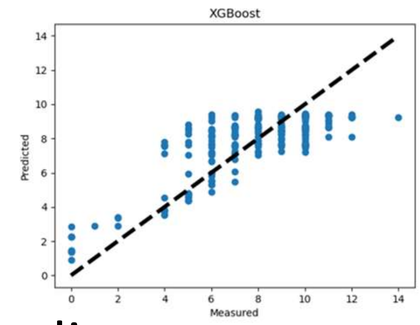
- Below are three figure showing the learning curves of various ML Models.

(Left to Right : For Number of Snaking predictor, Snaking Length, Snaking amplitude.)

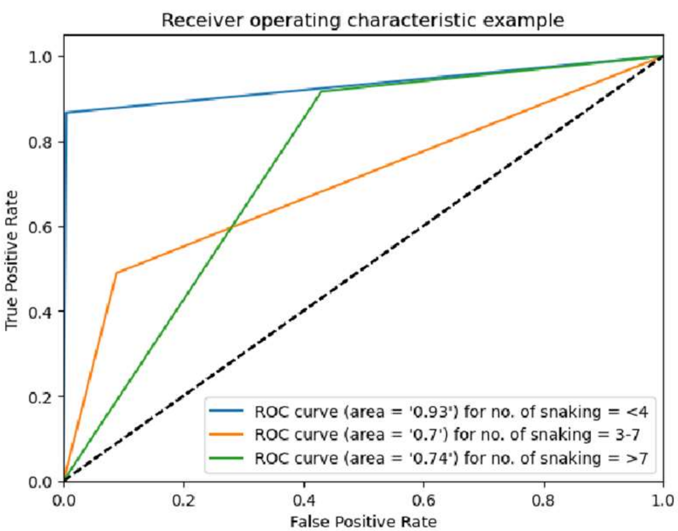


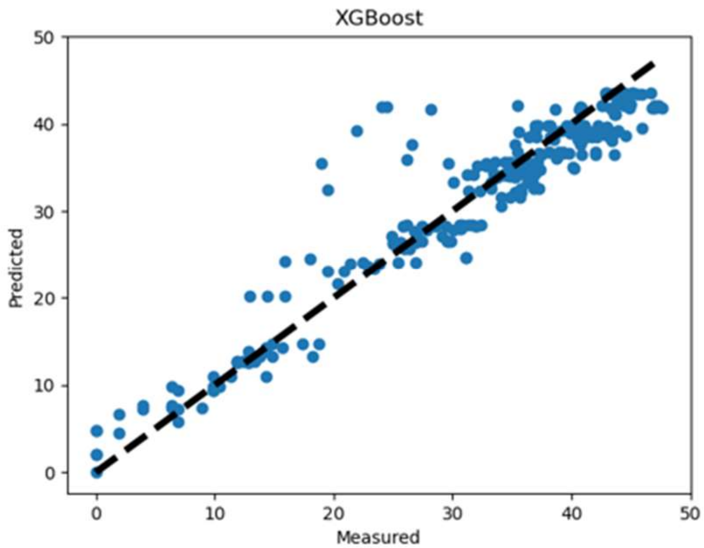


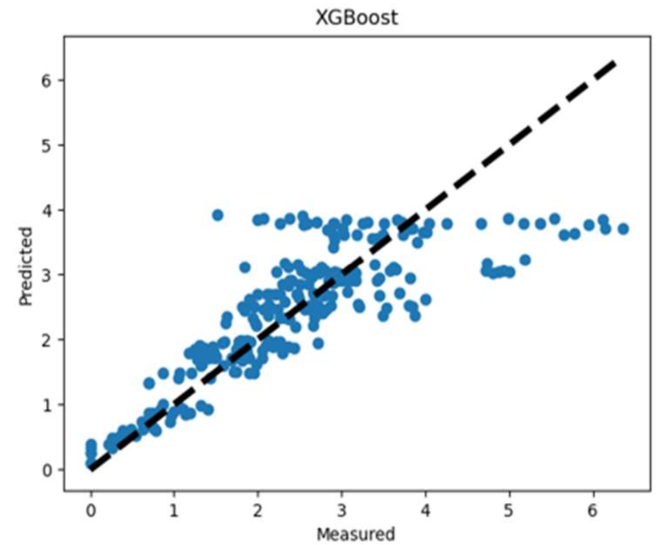
Pareto Charts for Best Models



- Below are three figure showing the Pareto charts of best Predictors (Left to Right : For Number of Snaking predictor, Snaking Length, Snaking amplitude.)
- Used to measure visually see how well a model is performing by plotting the predicted values against true values.







Stage – III
Explainable AI

Interpretability - LIME v/s SHAP

- **Better Interpretability Leads to Better Adoption**
- Initially, **SHAP and LIME** explainable AI methods. SHAP was considered more appropriate – Data Size.
- LIME performs a Linear Regression in vicinity of the input data sample. - Contributions
- SHAP computes the contribution of each feature to the final output of any sample
- Important: Only gives info about Model structure.

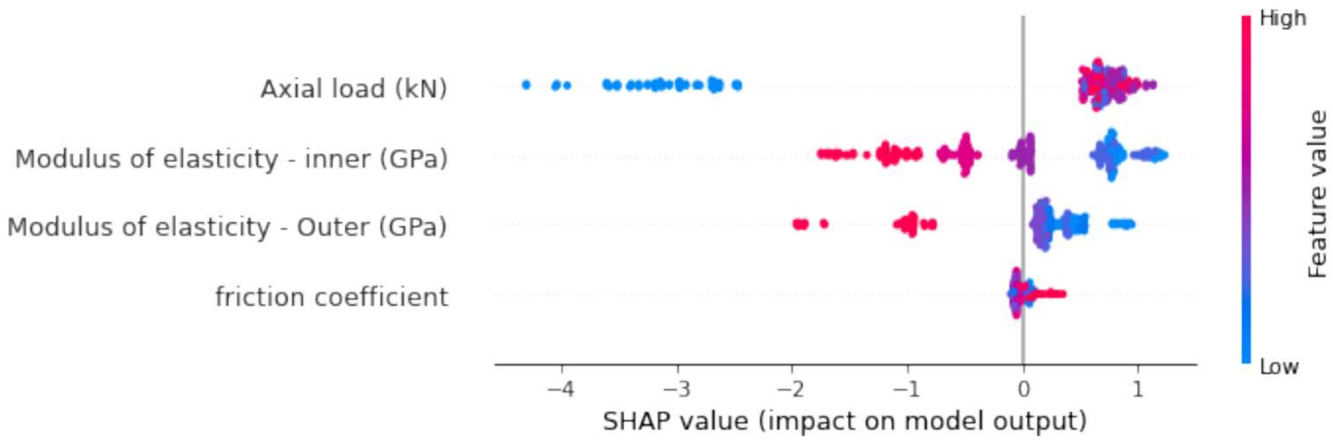
SHAP (SHapley Additive exPlanations)

- The goal of SHAP is to explain the prediction of an instance x by computing the contribution of each feature to the prediction.
- What is the Shapley Value?
  - Contributions
- Example
  - Two people are to break stones. (A- 10, B-15 , together – 30 stones)
  - But how much of these 30 stones were broken by A ? – (12.5 stones)
- Order matters

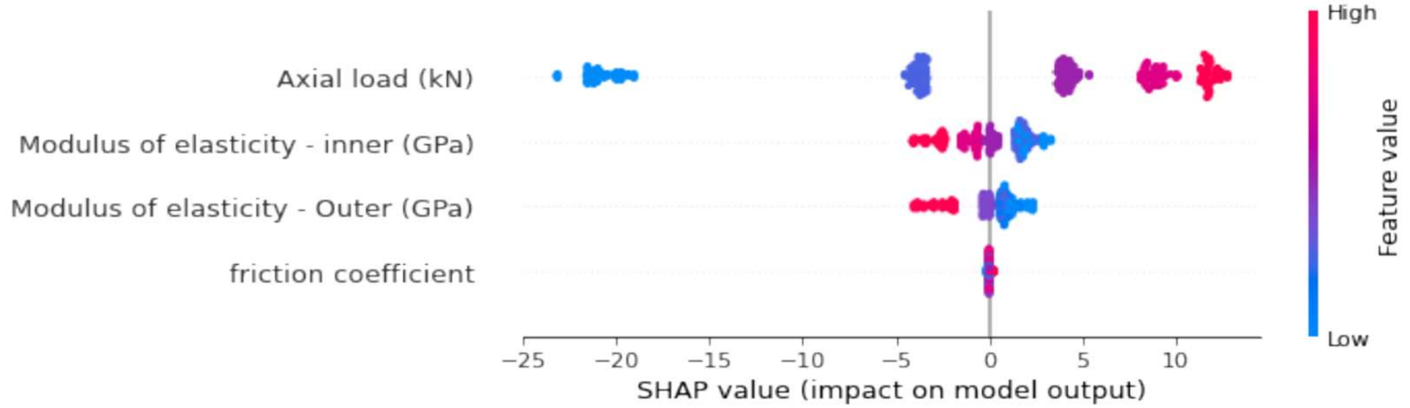
$$\varphi_i(v) = \frac{1}{\text{number of players}} \sum_{\text{coalitions excluding } i} \frac{\text{marginal contribution of } i \text{ to coalition}}{\text{number of coalitions excluding } i \text{ of this size}}$$

SHAP Results - Understanding

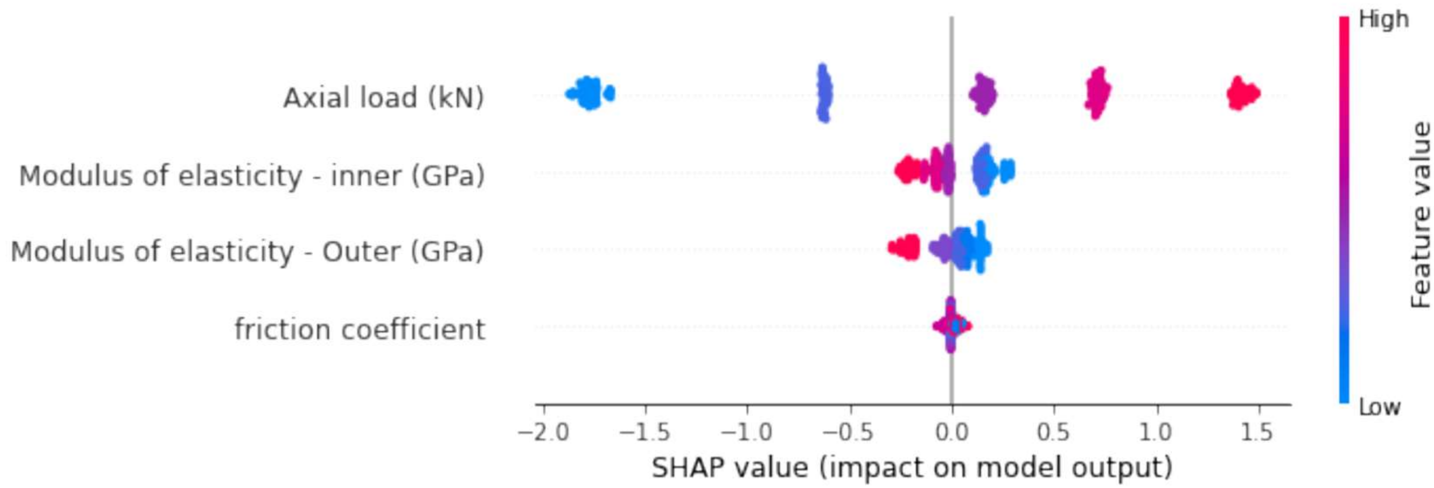
SHAP Summary Plot of Number of Snaking



SHAP Summary Plot of Snaking length



SHAP Summary Plot of Snaking Amplitude

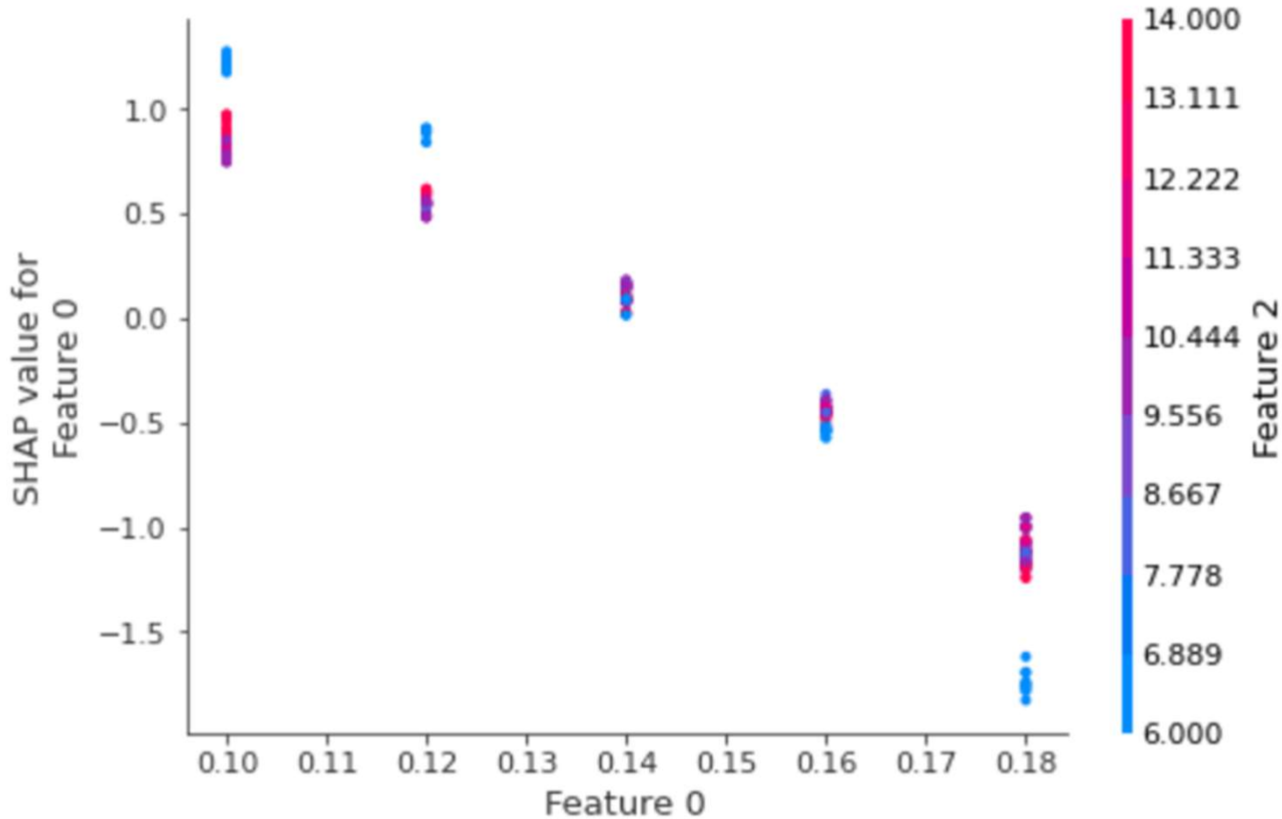


Insights from SHAP -

The insights from SHAP mostly corroborates the pre-analysis but are easy to interpret.

1. Existence of threshold.
2. The impact is almost constant for all the other values of axial loads on number of snaking.
3. Friction factor behavior is as expected, extremely low contact for almost all the values.

- As shown in the dependence plot in figure, the impact of inner and outer modulus is significantly amplified for Axial Load = 6kN. Feature 0 is Inner Modulus in MPa, Feature 2 is Axial Load in kN.



Other Activities -

- Other models tested –
  - Neural Network.
  - Poisson Classification.
  - Classification in case of number of snaking
    - XGBoost and Decision tree.
  - New data creation
    - Randomized.

NEXT STEPS

- Optimization.
  - We want to see the application of Bayesian Optimization in this setting.

THE END.