Investigation of Corrosion detection & Severity Level Prediction Using NDT and ML Techniques

Presented by:

SYED SANAULLAH SHAH

Herbert

Jasmine

Supervised By:
Data Glacier

Introduction

- Million of kilometers of pipelines carrying oil and gas worldwide.
- Integrity of these pipelines with time is of utmost importance for economy.
- Corrosion identified as main challenge affecting the **sustainability & efficiency** by compromising the mechanical strength of these pipelines leading to **environmental** and **economical catastrophe.**

Introduction Cont.

- Hazardous liquid pipelines accidents caused by internal corrosion **34%** & External corrosion **65%**

- Gas transmission pipelines accidents caused by internal corrosion **63%** & External corrosion **36%**

- Mild steel is majorly used for the O&G pipelines.

- These pipelines exposed to harsh pressure and chemical attacks from inside & outside provides a medium to corrosion to wreck havoc eventually like a malignant cancer.

Problem Statement

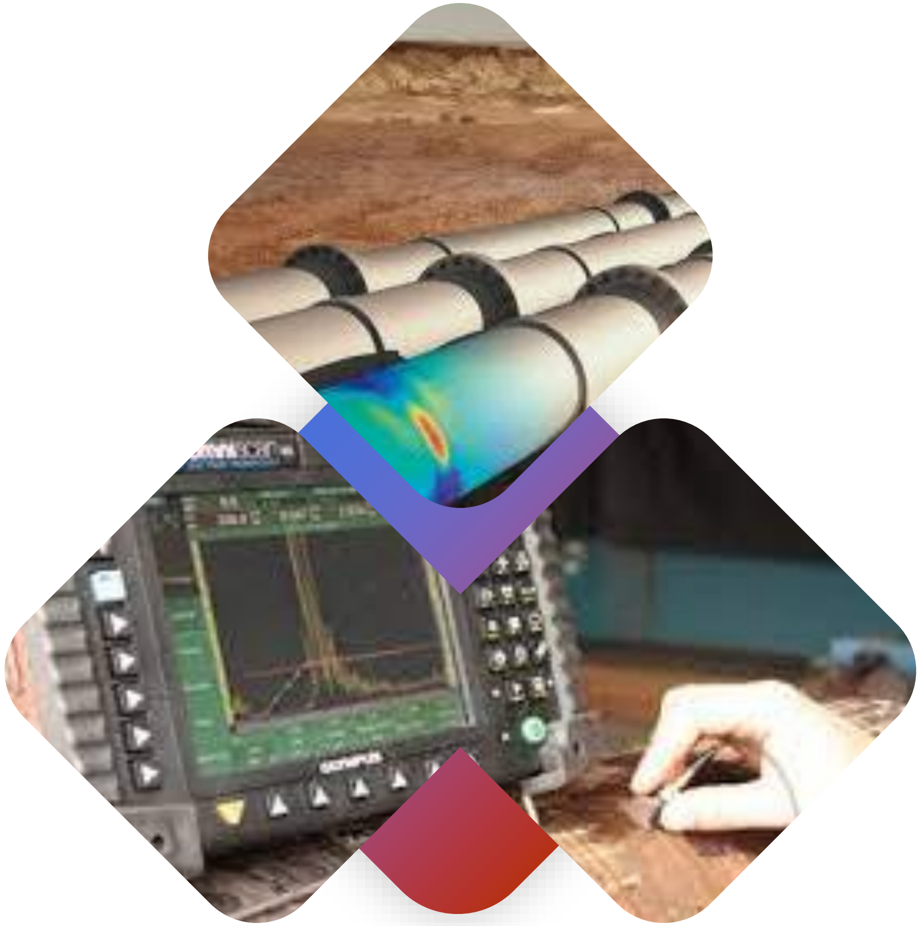
Pipeline Corrosion is a distinct problem in industries.

Problem

Corrosion contributes to about 25% of failure experienced in O&G production industry. While more than 50% of these failures are associated with sweet and Sour Corrosion in pipelines.

Causes

Humidity, Acid rain, Chemical salts, Oxygen, CO_2, High temperatures affect the corrosion pace



Solution

Our project explores risk assessment methodologies by classifying corrosion severity levels using Machine Learning.

Monitoring

NDT based Ultrasonic testing to acquire data

Inspection

Corrosion Severity classified and risk analysis is given as output

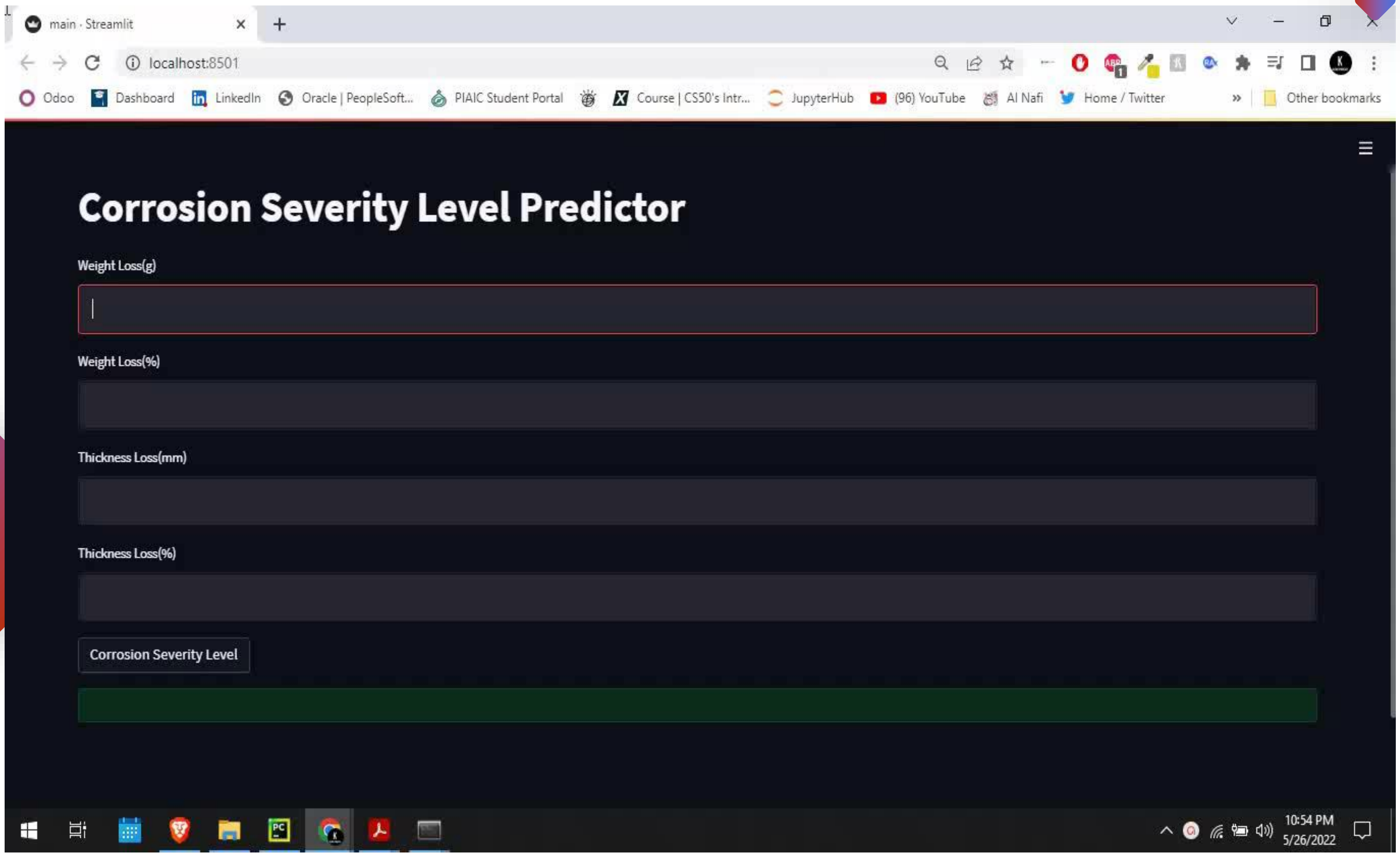
Processing

Sending Corrosion data to Machine learning platform for classification

Digital Product

App collects data & Predicts the failure class based on trained failure data & improves itself with more data

Product Prototype



Methodology

- ✓ Advancements in inspection and monitoring techniques are aiding corrosion expert in decision concerning the "when" & "how", managing to optimize cost and pipelines performance.
- ✓ Data Driven Insights. to surface new key performance Indicators.
- ✓ We utilized Corrosion growth rates estimation from laboratory experiments to generate dataset and did a comparative study of Machine Learning Algorithms, to predict risk based on severity levels.

Percentage Losses

✓ **Percentage Thickness Loss**

The Percentage loss in the thickness of the sample before and after corrosion

$(\textit{Final Thickness} \,/\, \textit{Initial Thickness})$ *100

✓ **Percentage Mass Loss**

The Percentage loss in the mass of the sample before and after corrosion

$(\textit{Final Mass} \,/\, \textit{Initial Mass})$ *100

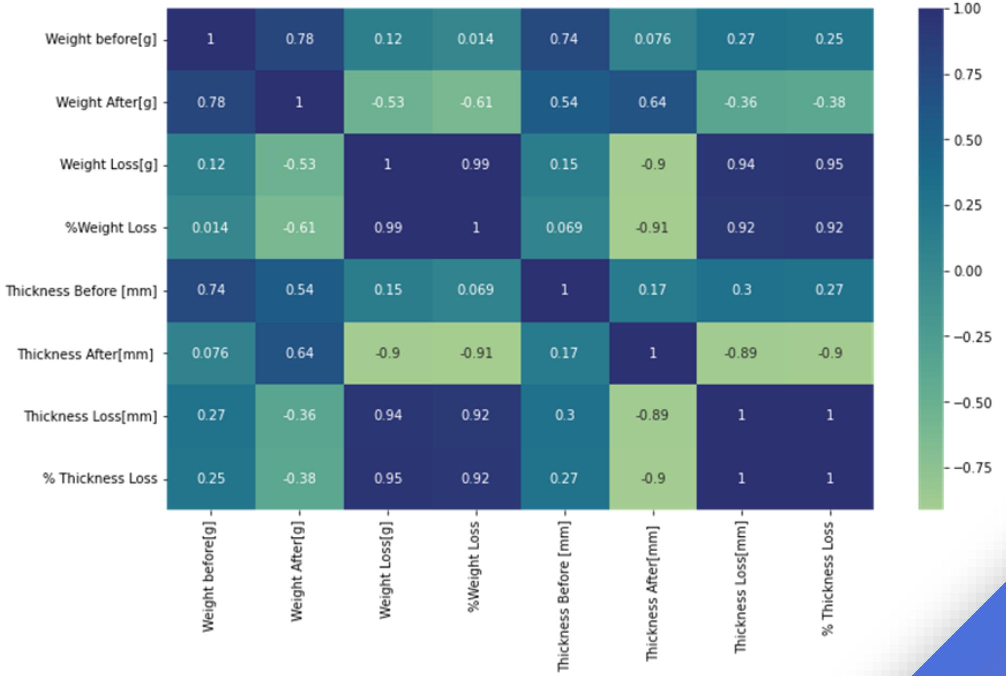
Implementation on Machine Learning Methods

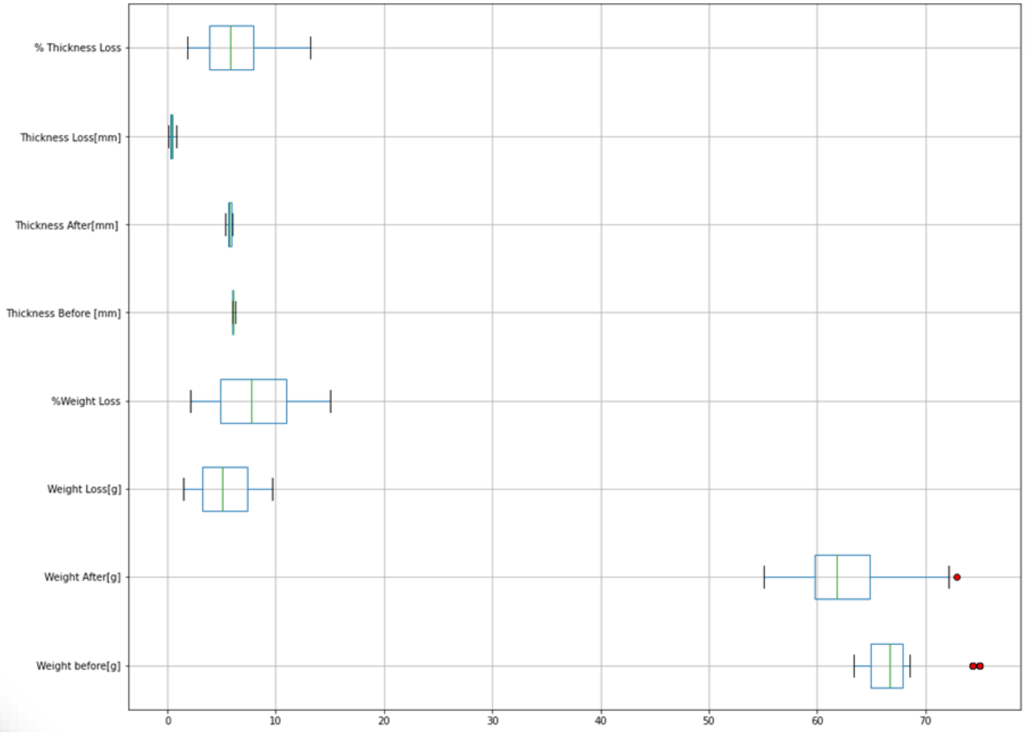
1. Data Loading from Experiments and Pre-Processing
2. Feature Selection (Sensitivity of inputs variables on Corrosion Level)
3. Implementation on Machine Learning(ML) Methods
4. Comparison & Performance of ML Methods

   (Predict Corrosion Level Accurately)
5. Graphical Representation

Data Pre-Processing

(Experimental Data Processed to start Machine Learning Methods)





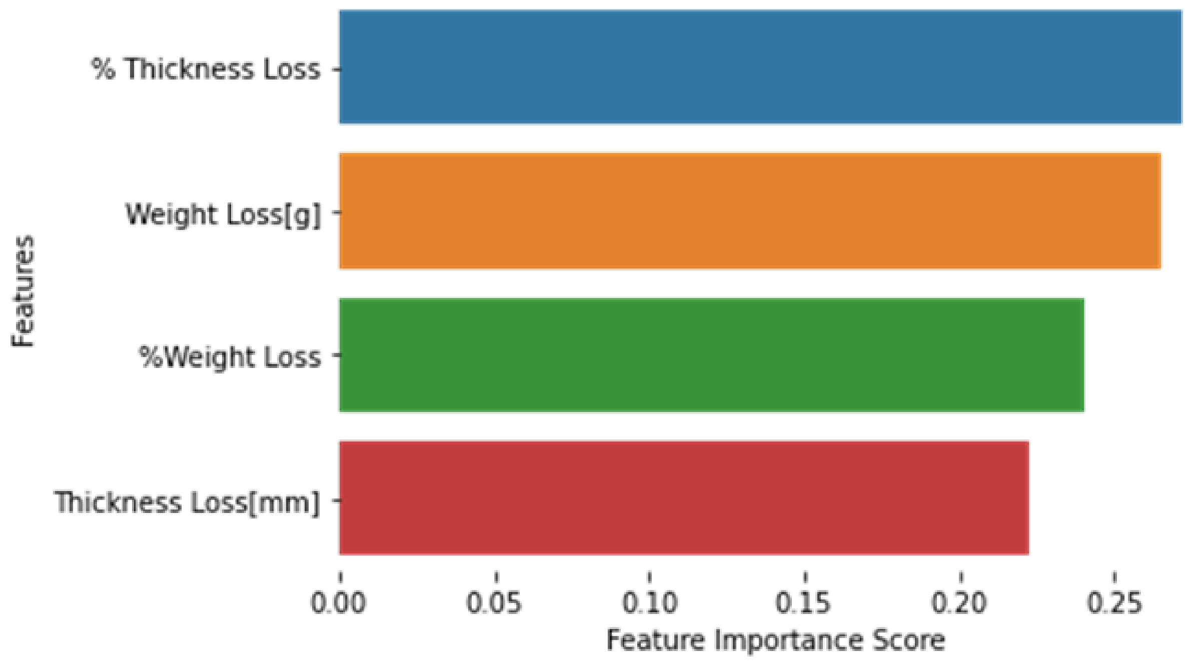
Correlation Heatmap showing Weight losses are strongly correlated with Thickness Losses

Boxplot showing the bigger picture of dataset by showing outliers in weight losses instances i.e Dataset quality check

Feature Selection

(Quantification of Parameters that impact Corrosion Level)

Embedded Method precisely random forests was used to explore key features. The relevance scores that tell us the %thickness loss & weight loss impact the target the most followed by %weight loss and thickness loss whereas other features only created noise.



Machine Learning MODEL Performance

HYPER PARAMETERS

GridSearchCV & RandomSearchCV

Functions to give best parameters to use.

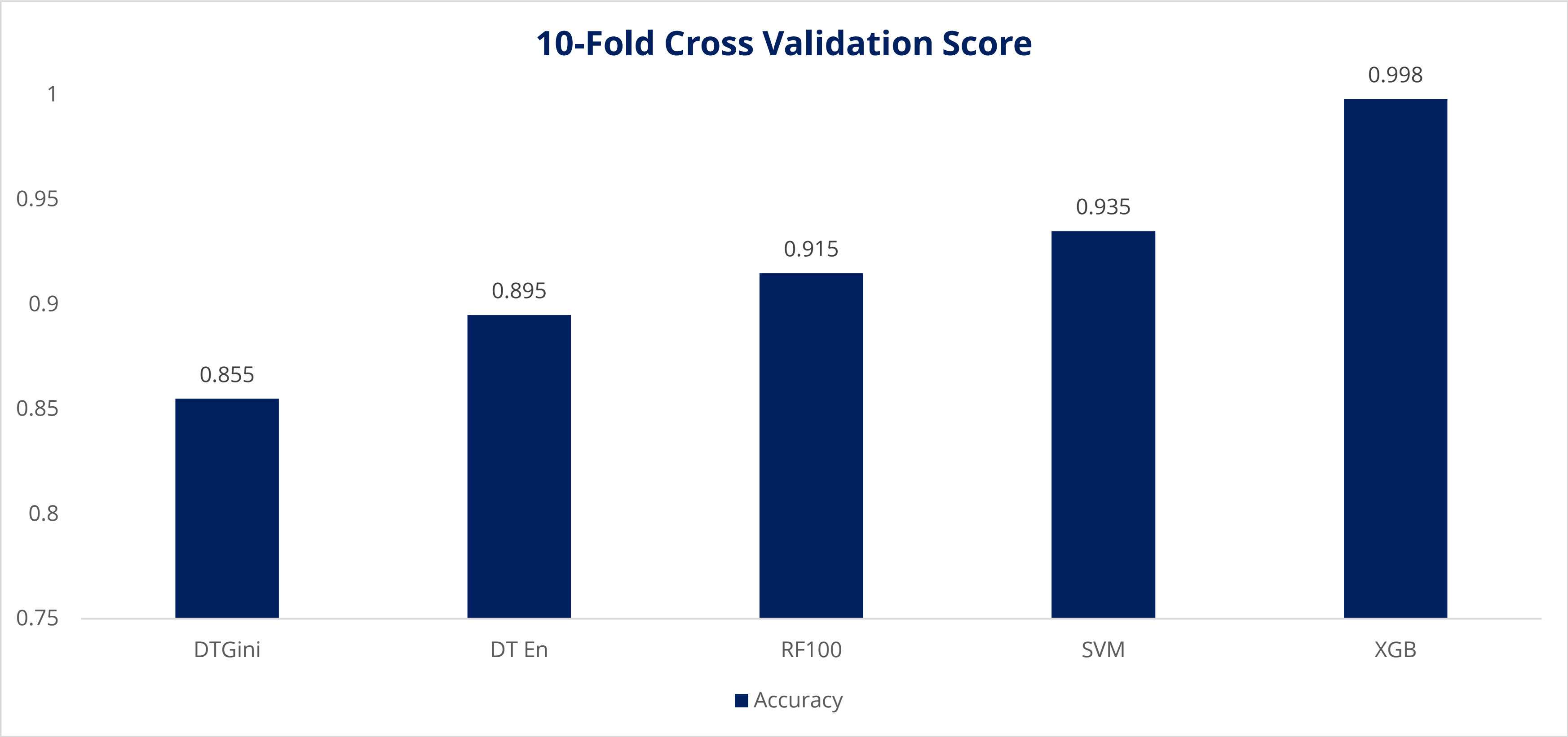
10 fold cross validation mean score to

counter overfitting concerns.

ML Methods	Parameters	Ranges	Optimized Values	Cross Val Score(10F)
Decision Trees	Criterion Max-Depth	GINI 3-5	GINI 3	0.855
Decision Trees	Criterion Max-Depth	Entropy 3-5	Entropy 3	0.895
SVM	‘C’ ‘Gamma’ ‘Kernel’ ‘CV’	0.1-100 ‘scale’-0.001 ‘RBF’, ’POLY’ 3-10	1 ‘scale’ ‘RBF’ 10	0.935
Random Forests	Decision Trees	10-100	10	0.915
XGB	Max-Depth Learning Rate ‘n-estimators’ ‘Gamma;	3-15 0.01-.3 10-500 0-0.4	4 0.25 400 0.3	0.998

Cross Validation Score

Extreme Gradient Boosting outperforming every other technique, proving why it's the most accurate & fast classifier for small to medium size tabular datasets.

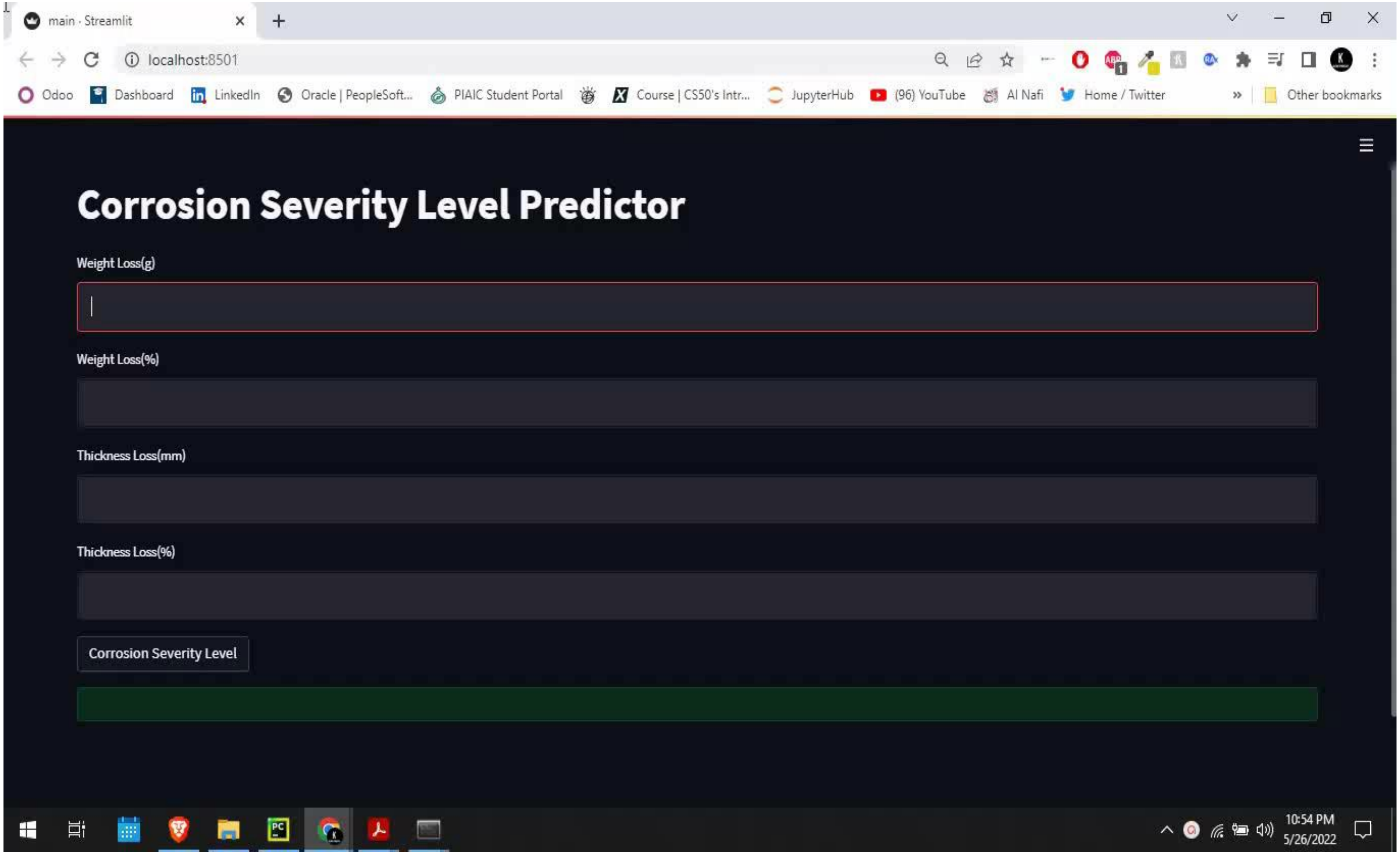


Comparison of ML Methods
(Best Method to Predict Corrosion Level)

Common Classification Evaluation metrics like precision, recall, F1,Accuracy, R^2 & RMSE results are show in table below.

ML Methods	Precision	Recall	F1	Accuracy	R Squared	RMSE
Decision Trees(GINI)	0.75	0.75	0.75	**0.8**	0.852	0.447
Decision Trees(Entropy)	0.92	0.88	0.87	**0.9**	0.926	0.316
SVM	0.88	0.81	0.77	**0.70**	0.663	0.548
Random Forests	0,94	0.88	0.88	**0.9**	0.926	0.316
XGB	1	1	1	**1**	1	0

Streamlit Graphical User Interface



Limitations

- In a mid- to long-term perspective, AI could fundamentally alter the way in which research is conducted, structured and understood
- AI will always answer, But not Same Answer every time
- Over fitting
- Inclusion of new variable could every time Improve accuracy of underperformed algorithm Lack of data as research depends
- Data Collection is costly, difficult task & time consuming
- Conventional Beliefs dulls the vision of Future Works of Artificial Intelligence & Internet of things.

Future Prospects

- AI can be integrated with other technologies. like IOT to amplify efficiency of data gathering and mitigating potential risks
- With more data, model can be transferred to deep learning that can complement forecasts to detect patterns, Classify and Correct response for mitigation by anticipating the future.
- Smart contract protocols embedded in the block chain technology can be used to execute a sensitive decision making.

Conclusion

- Risk based corrosion severity level multi-classification was successfully predicted by our machine learning classifiers using thickness and mass loss derived features
    - Extreme gradient boosting outperform every other classifier according to our evaluation matrix
        - Decision trees, SVM & RF even after hyper parameter optimization show low accuracy scores because they are prone to outliers
            - The future work will focus on deep learning techniques with bigger database

THANK-YOU

- ANY Questions