Data Science Intern at Data Glacier

Project: Corrosion Detection & Severity Level Prediction Using Machine Learning

Week 7: Deliverables

Name: Syed Sanaullah Shah

University: NUST, Islamabad

Email: iamkadhimi@gmail.com

Country: Pakistan

Specialization: Data Science

Batch Code: LISUM11

Date: 30 October 2022

Submitted to: Data Glacier

Table of Contents:

Project Deadlines

Weeks	Date	plan
Weeks 07	September 23, 2022	Problem Statement, Data Collection, Data Report
Weeks 08	September 30, 2022	Data Preprocessing
Weeks 09	October 7, 2022	Feature Extraction
Weeks 10	October 14, 2022	Building the Model
Weeks 11	October 21, 2022	Model Result Evaluation
Weeks 12	October 28, 2022	Webapp Streamlit Deployment
Weeks 13	October 30, 2022	Final Submission

Problem Statement

Corrosion of pipelines has been recognized as a major predicament because it undermines the mechanical strength of pipelines, which can pose a disaster for the environment and the economy. Most of the time, the O&G industry uses mild steels. Ultrasonic testing during accelerated corrosion testing is a way to detect corrosion. AI is already using the scientific approach by popping up with hypotheses and doing simple tests to see if they are true. In the mid to long term, AI could have a big impact on how research is conducted, constructed, and acknowledged. However, using machine learning techniques to classify corrosion based on parameters like weight and thickness loss is still a new idea.

Our framework suggests using a new method that looks for corrosion through thickness and mass losses by NDT. With the help of machine learning methods and ultrasonic testing from accelerated corrosion testing, corrosion severity levels can be predicted based on data generated in the lab. Samples of mild steel could be evaluated for accelerated corrosion over four different time periods in a lab. Samples' thickness and mass loss were noted down in the dataset. Ultrasonic testing readings were done with an Epoch LT Ultrasonic testing

machine with a straight beam configuration, because the relationship between thickness and mass losses and the corrosion process is linear. For multi-class problem, four Corrosion severity levels have been created based on thickness & mass loss occurred during accelerated corrosion testing for which XGB, SVM and Random Forests showed cross validation accuracy score of 99.8%, 93.5%, and 91.5% respectively.

Business Understanding

Globally, the cost of corrosion is in the billions of dollars for every single economy on the planet. Corrosion failures have caused more than $2 trillion dollars in losses around the world. Steel pipes have surpassed all other modes of oil and gas transportation in the previous 50 years. Between 1975 and 2004, the overall length of the gas transmission system in Europe rose from around 30,000 Km to more than 120,000 Km. As a result of transporting corrosive substances and being positioned in severe settings (particularly near-sea and some subsea pipelines), the life expectancy of these pipes has been reduced more rapidly than anticipated, owing to internal and exterior corrosion or erosion. Apart from the fact that carbon steels are susceptible to corrosion and therefore require frequent inspection, they are good and cost-effective materials for use in the oil and gas industry. Carbon steel pipelines corrode both internally and outside. The rate of steel corrosion is determined by the temperature, pressure, composition, and shear of the produced multiphase fluids moving through the pipeline. The rate of corrosion in pipelines can be used to determine the parameters that are causing problems.

A classification model derives some valid mapping functions from the training dataset and predicts the class label for a new data entry using the mapping function that was determined from the training dataset. An attribute or feature is a parameter discovered in the provided problem set that may be used to develop a predictive model that is accurate enough to predict the future.