AMERICAN UNIVERSITY OF BEIRUT

SERVICE LIFE PREDICTION FOR PIPELINES USING MACHINE LEARNING TECHNIQUES

by
MICHAEL ANTOINE TAWK

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Engineering
to the Department of Civil and Environmental Engineering
of the Maroun Semaan Faculty of Engineering and Architecture
at the American University of Beirut

Beirut, Lebanon
August 2023

AMERICAN UNIVERSITY OF BEIRUT

SERVICE LIFE PREDICTION FOR PIPELINES USING MACHINE LEARNING TECHNIQUES

by

MICHAEL ANTOINE TAWK

Approved by:

Prof. Samir Mustapha, Associate Professor Mechanical Engineering Department	Advisor
Prof. George Saad, Associate Professor Civil and Environmental Engineering Department	Member of Committee
Prof. Darine Salam, Associate Professor Civil and Environmental Engineering Department	Member of Committee

Date of thesis defense: 18/08/2023

AMERICAN UNIVERSITY OF BEIRUT

THESIS RELEASE FORM

Student Name:_____TAWK__________MICHAEL_________ANTOINE__

Last First Middle

I authorize the American University of Beirut, to: (a) reproduce hard or electronic copies of my thesis; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes:

☐ As of the date of submission

☐ One year from the date of submission of my thesis.

☒ Two years from the date of submission of my thesis.

☐ Three years from the date of submission of my thesis.

_______________________ 30/08/2023 ____________

Signature Date

ACKNOWLEDGMENTS

First and foremost, I would like to extend my deepest gratitude to my thesis advisor, Dr. Samir Mustapha, without his support, patience, and encouragement, this work would not have been possible. His expertise, guidance, and dedication were instrumental in shaping the direction of this research and enhancing the quality of my work.

I would like to extend my thanks to the thesis committee, Dr. George Saad and Dr. Darine Salam. The assistance provided by Dr. Saad was greatly helpful.

I would like to thank Fouad Trad for his insightful comments, constructive criticism, and encouragement, which significantly contributed to the improvement of my research.

My sincere appreciation goes out to my father, mother, brother, and friends for their unwavering love and support. Their constant encouragement and belief in my abilities kept me motivated during challenging times.

ABSTRACT

OF THE THESIS OF

Michael Antoine Tawk for Master of Engineering
Major: Civil Engineering

Title: Service Life Prediction For Pipelines Using Machine Learning Techniques

Pipelines are a vital part of the global economy as they transport large volumes of fluids across thousands of miles. Pipelines mainly pass through rural areas and are most of the time buried underground or laid down on the seabed. They are prone to various types of failure due to the harsh environmental conditions they serve in, where corrosion is considered one of the major causes.

This study focused on predicting the remaining service life of pipes, before they fails due to corrosion, using a big dataset published by the Pipeline Hazardous Material Safety Administration (PHMSA). The PHMSA dataset includes a large number of predictive fields, along with additional weather data such as temperature and precipitation that were extracted from the National Centers for Environmental Information. A regression model was built to predict the remaining time till failure and classification models to predict an interval of the remaining time till failure. Top performance was achieved using the Extremely Randomized Trees (Extra Trees) algorithm with an R-Squared score of 90.35% for regression and an f1 score of 85% for classification. The importance of each feature used to build the model was assessed using SHapley Additive exPlanations (SHAP) to explain the outputs of the models and identify the most contributing factors responsible for accelerated pipe failure. It was concluded that weather conditions like temperature and precipitation play a major role in pipe failure.

Future work may include implementing predictive maintenance using the precise predictions of the time left before failure, as well as considering other datasets from various types of structures.

TABLE OF CONTENTS

ILLUSTRATIONS

Figure

TABLES

Table

CHAPTER 1

INTRODUCTION

The global total length of crude oil pipelines is 379,000 km while natural gas pipelines constitute 1,295,563 km [1]. The total cumulative length of oil and gas pipelines as of December 2020 is enough to wrap the earth 30 times with pipelines [2], they surpass any other mode of transporting fluids [3]. Pipeline incidents have catastrophic repercussions on the economy as they interrupt the supply to customers, especially a supply of extreme importance on which whole production lines and industries depend. Disruptions to the supply chain don't merely result in increased costs and shortages but also create instability which can foster inter- and instar-national conflicts [4]. The recent discoveries of oil and gas reserves in harsh and remote environments [5] made the pipeline integrity even more critical; a spillage accident due to failures of any type can cause complete ecosystems to be wiped out in these untouched places, which may cause irredeemable damage to the global biodiversity. The environmental risk is not restrained to remote and harsh environments; for instance, a failure caused by corrosion contributed to a spillage of 140,000 gallons of crude oil, the spill polluted miles of coastline, and the oil's tar balls washed ashore on Los Angeles beaches more than 100 miles away [6].

Corrosion which is a natural phenomenon in metallic structures [7] can be a primary factor in pipeline failures [8], leading to large-scale water and soil contamination, toxic emissions from explosions, and fires associated with these failures [9].

Sources, forms, and rate of corrosion – In metals, corrosion can happen if the metal comes into touch with an electrolyte that carries the electric current; that is electrochemical corrosion, and if there's a direct chemical reaction between the metal and the medium (molten salts, gaseous environments), that is dry corrosion [10]. Different factors inducing corrosion, added together with different combinations and intensities can produce different forms of corrosion. The most common forms of corrosion are uniform corrosion, pitting corrosion, crevice corrosion, and under deposit corrosion. Uniform corrosion happens when the interior walls or the exterior walls of the pipe are reduced in thickness uniformly [11]. Pitting corrosion is risky because the depth of corrosion is very high while the superficial area is low, thus, making it more difficult to visually localize, and they develop on the inside as well as the outside walls of the pipe. Crevice corrosion, is generally observed in restricted spaces to which the fluid at work does not have access e.g., between flanges, nuts, and bolts [10]. Under deposit corrosion causes deep, aggressive, and localized penetration of the walls [12]. This type of corrosion develops under or around deposits on the surface, these deposits could be organic or inorganic.

Another harmful aspect of corrosion is its rate. The rate of corrosion depends on the surface of the interface between the metal and the soil and water [13], on the temperature, humidity, pH levels [14], the purity of the metal [14], and on applied outside stress which causes cracks [15].

Many Nondestructive Testing (NDT) techniques have been developed over time to enhance the detection of damages in pipeline networks and mitigate the nefarious consequences of failures including corrosion. More recently machine learning (ML) models have also been used in detecting damages due to their prominence in finding intricate connections between the pipe failures and different variables.

Related Work: Due to the importance of pipeline integrity in providing necessary goods for economic progress and the advancement of the quality of life, much research was done on the subject. NDT techniques include but are not limited to, visual inspection, magnetic flux, ultrasound, and uncompensated volume balance.

Visusal inspection is the most basic form of NDT, it relies on human knowledge and expertise, and gives information about the surface of the material only. Magnetic flux is a state-of-the-art method for damage detection, the method consists in detecting the change in flux. Most of the time homogeneous materials will allow magnetic flux to travel in continuous lines in a fixed direction. Damage, like corrosion, can be detected by the deformation of the magnetic flux traveling through the material [16]. Ultrasonic testing uses a pitch-catch or pulse-echo method to measure the thickness of the pipe walls. Since corrosion causes thinning of the pipe walls, this method is very effective in identifying corrosion locations. A transducer emits waves that travel through the material, the difference in arrival time in the signal arriving to the receiving transducer, helps to deduce the distance traveled knowing the speed of propagation of the wave in the material [17]. Last but not least, the uncompensated volume balance method is based on the flow entering and exiting a pipeline network. Ideally, the sum of flows entering the pipeline and the sum of flows exiting the pipeline should be equal, an unusual difference indicates a leakage.

NDT techniques require the installation of sensors and measuring devices on structures which has been hindered by the restrictions on sensor measurement capabilities, and power and data connection requirements, fortunately, ML models have emerged as an adaptable and appealing solution in detecting damage [18].

The major factors for this recent interest in ML-based solutions can be attributed to (1) big data and cloud-based compute advancements, (2) hardware and software advancements in computers, (3) data science advancements, and (4) transfer learning advancements [19]. These advancements facilitated the collection, storage, and manipulation of sensor data as well as manual aggregation of data.

Multiple works made use of these advancements like Fan, et al. [20] that applied ML models trained on the data from the Cleveland Water Department (CWD) to predict the failure of a water pipe (Boolean output: fail/no fail). To further enrich the dataset, the authors added soil, topographical, climate, and demographic data to the CWD dataset. They then compared the performance of five ML algorithms: LightGBM, artificial neural networks, logistic regression, k-nearest neighbors, and support vector classification. LightGBM achieved the best performance.

Further, Kumari, et al. [21] aimed to implement a model that predicts the risk (a consequence of an incident multiplied by its probability) of corrosion-induced pipeline incidents. Using the PHMSA dataset they trained an artificial neural network to predict the consequences: total cost (US Dollars), the net loss (Barrels), and the type of release using an artificial neural network, and they developed a Bayesian analysis model to predict the probability of an incident happening in the next seven days.

Similar research was conducted by Wang and Li [22] to assess the risk level of an urban pipeline using clustering. They conducted a case study on a dataset from a local gas company consisting of more than 1700 km of pipelines, and the accident reports from the database of the maintenance division. The proposed method identified and separated into clusters the various risk levels.

Yang, et al. [23] went even further by using a graph convolutional network (GCN) to extract topological features of the pipe network. Their conclusion was that the GCN is helpful to have an overall understanding of the risk level in a certain region, while the traditional method (based on the Kent index method and the analytic hierarchy process) is more helpful in a specific pipe section.

Senouci, et al. [24] have used data from the report of the Conservation of Clean Air and Water in Europe (CONCAWE) organization, established in 1963 by leading European oil companies [25], to predict the failure type of oil pipelines. This report aggregates data from oil pipeline accidents in Europe. In their study, Senouci, et al. [24] used reported spillages in

2008 and since 1971 and compared ANNs to the Linear Regression model. They used five inputs (Type of Product, Pipe Location, Pipe Age, Land Use, Pipe Diameter) and trained on 351 data points to predict several failure causes including mechanical, operational, corrosion, natural hazards, and third party. ANN proved to be superior to LR, but practically some causes could not be predicted from these attributes such natural hazards or third party. Kumari, et al. [26] went a step further to predict the sub-category of failure. For example, for an incident with natural force as the cause, they predicted whether the natural force is lightning or heavy rain. They used the incident data collected by PHMSA for 2010–2019 to implement the sub-category 2 prediction models. The first model predicts the main category of the incident (mechanical, operational, corrosion, natural hazard, third party), and based on this prediction a second model predicts the sub-category of the incident (in the case of internal or external corrosion). The models were trained on 3105 data points and they used 16 attributes for the first model, as for the second the number of attributes depends on the category type (e.g., 8 attributes for corrosion).

Liu, et al. [27] exploited the 3587 records of incident narratives of the 'comment' section in the incident database of the PHMSA using Natural Language Processing (NLP). NLP helps concentrate on the failure's underlying causes and contributing elements, such as organizational, managerial, or personnel problems, as these are not always reported by operators and can only be mentioned in free-text incident descriptions if they apply. They mined hidden knowledge beyond the cause assigned by the PHMSA thus offering a creative way to uncover hidden knowledge and improve learning from prior mistakes. The procedure can create an incident's cause-and-effect narrative, and the information made available enables a rapid comprehension of the knowledge amassed in incident narratives.
Similar research was conducted by Ahadh, et al. [28], where a semi-supervised approach was used to analyze reports. The primary advantage of this strategy is that it requires little manual

labor to produce domain-specific predictive models. Keywords were extracted and grouped and used for a variety of data mining tasks, including classification. The suggested method was used to determine the reason for the pipeline failure.

Previous research was limited to predicting the type of failure of the pipeline, or to mining useful data from texts that need to be assessed by humans. Practically the need arises to predict how long a pipeline can still serve before undergoing maintenance or failing; this can include both predicting an interval in years within which the pipe will fail or predicting the exact number of years left till failure, and this is the main objective of this work.

This is the first research to propose an approach to predict the time a pipe still has until it fails due to corrosion. The approach is based on training ML models on the PHMSA [29] dataset and adding to it the temperature and precipitation properties (extracted from the National Centers for Environmental Information [30]) of the state where the failure occurred as well as during the time frame of the failure.

The main contributions of this work are (1) addition of weather data (temperature and precipitation) to the PHMSA dataset after selecting the best features (based on domain knowledge), (2) generation of the time left till failure, and augmentation of the dataset, (3) investigation of the best-performing learning models for regression and classification, and (4) develop a good understanding of the contribution of different factors in the remaining time left.

The remainder of this paper is organized as follows: Section 2 describes the methodology, Section 0 expands on the ML implementation, Section 0 presents results and discussion, finally Section 5 concludes with remarks and future work.

CHAPTER 2

METHODOLOGY

2.1.Dataset used

In the USA, the pipeline incidents database is maintained by the PHMSA [29]. The PHMSA pipeline infrastructure has 347,020 km of pipeline for crude oil, refined products, and natural gas liquids, 513,070 km of pipeline for gathering and transmitting natural gas, 3.5 million km for distributing gas to homes, businesses, and other industrial sites, and a relatively smaller LNG pipeline network [31]. For each incident in the pipeline, operators submit a report, the PHMSA reviews the reports and reorganizes it into a structured table to maintain consistency over the years. PHMSA started collecting data in 1970 and the reporting went through many changes, the most recent reporting system started in 2010. Currently, the HL (Hazardous Liquids) dataset contains 654 fields (e.g., date of the incident, operator, state, pipe properties, etc.) and 4901 incident records. This dataset will be used in the present work, and incident dates range from 2000 until 2023. Most of the incidents due to corrosion occurred in the state of Texas, and the material transported in the pipelines was mainly crude oil, the pipe diameters ranged from 1 inch to 42 inches, and the thickness of the pipe walls ranged from 0.05 inches to 0.75 inches.

Features were selected and filtered based on domain knowledge. The 654 features have been examined one by one; only incidents caused by corrosion were taken into consideration, and features with irrelevant information like the report ID or the operator’s address were disregarded. In addition, features that are a consequence of the incidents (e.g., fatalities, injuries) were also put aside because these variables cannot be known before the failure occurs, as well as highly detailed information about the incident (e.g., whether the incident location is a road, crossing, or bridge) and about the operator (e.g., postal code, latitude, longitude) which is documentation of the incident and is not informative about incident causation.

2.2. Data Processing and Aggregation

After a preliminary examination, the remaining features were analyzed more thoroughly. Incidents, where the pipe diameter, the pipe wall thickness, and installation year are not available, were not selected. Further, some features have a small number of occurrences, for example, offshore water depth and puddle weld indicator, are included in a very small number of incidents, and therefore do not add any positive contribution to the dataset. There were some outliers/mistakes that needed to be dealt with manually, for example, a value of 250 inches for the pipe wall thickness is highly unlikely, hence this was changed to 0.25 inches which lies within the typical range of the wall thickness, additionally some features had missing values, these values were replaced by the mode (the most occurring value) value of the feature. All data manipulation was done using the NumPy [32] and Pandas [33] libraries. With all the filtering and cleaning the initial version of the dataset included 851 incidents and eight features were selected as reported in Table 1.

Table 1 Selected features to be used in models for regression and classification.

Feature	Type	Explanation
COMMODITY_RELEASED_TYPE	String	Fluid inside pipe
SYSTEM_PART_INVOLVED	String	Pump piping, tank piping…
MONITORING_SYSTEM	Boolean	Monitoring system present
PIPE_DIAMETER	Float	Diameter of pipe (in)
PIPE_WALL_THICKNESS	Float	The thickness of pipe wall (in)
PIPE_SMYS	Integer	Specified Minimum Yield Strength (PSI)
MOP_PSIG	Float	Maximum Operating Pressure (PSI)
pipe_age	Integer	Age of pipe (years)

2.3.Generation of Remaining Service Life, Normalization, and Encoding

The database used contains information about the pipes when they failed, and since the focus of this study is to predict the remaining life, it is important to adjust the age of the pipe and then try to predict the failure using the developed model.

To generate the time left until the pipe fails in years a random number between the installation year and the incident year is chosen, this will be called the threshold year. The threshold year is subtracted from the incident year to get the remaining time till failure and from the installation year to get the pipe age at the threshold year. For example, if the installation year of the pipe is 2002, and the incident year is 2016, it is known that the pipe age is 14 years. A random number (threshold year) between 2002 (installation year) and 2016 (incident year) is randomly chosen, in this case, this random number is 2014 (threshold year). After the corresponding subtractions, at the threshold year, the pipe age will be 12 years and the time left till failure will be 2 years. The other features are not affected by the time, which is why they stay the same. The same procedure is applied for all incidents.

Normalization and Encoding: Continuous features can vary a lot in scale; the pipe wall thickness ranges between 0.05 in and 0.5 in while the maximum operating pressure ranges between 10 PSI and 990 PSI. To avoid giving more importance to features with higher ranges and to improve the performance and the stability of the model, normalizing is a necessity. Therefore, the features were standardized by removing the mean and scaling to unit variance. ML models can have as inputs only numbers, this is why classes of categorical data like 'Commodity Released' had to be encoded with integers. For instance, 'Commodity Release' which is the fluid being transported by the pipe has different classes like crude oil and carbon dioxide.

2.4. Precipitation and Temperature

The precipitation and temperature data were extracted from the National Centers for Environmental Information [30]. For each incident, the monthly average, of the last 120 months from the threshold year, of the precipitation and temperature data was added depending on the state where the incident occurred. For example, if the threshold year of a pipe is 2014 and the pipe failed in Texas, the monthly average of temperatures and precipitation is extracted for the interval between 2004 and 2014 for the state of Texas and added to the dataset. Statistical properties like the average, minimum, maximum, and median are calculated for each precipitation and temperature time series and appended to the original dataset. Table 2 shows two examples of data after processing and preparation. The dataset contains 16 features in addition to the column being predicted which is the time left till failure. Temperature keeps varying underground until a soil depth of around 4 m [34], while humidity due to precipitation keeps varying until a depth of around 3 m [35]. Incidents in buried pipes on average occurred at a depth of around 1.25 m so they were still prone to temperature and precipitation variations.

Table 2 Two examples of pipes that will fail with added statistical properties of temperature and precipitation.

Feature	Example 1	Example 2
PIPE_DIAMETER	8	26
PIPE_WALL_THICKNESS	0.375	0.438
PIPE_SMYS	24000	60000
MOP_PSIG	285	285
pipe_age	1	24
COMMODITY_RELEASED_TYPE	CRUDE OIL	CRUDE OIL
SYSTEM_PART_INVOLVED	TANK PIPING	PUMP PIPING
MONITORING_SYSTEM	YES	YES
temp_mean	18.573148	-2.796759
temp_min	5.777778	-24.388889
temp_max	30.222222	13.777778
temp_median	18.833333	-2.944444
prec_mean	5.880735	7.779385
prec_min	0.5842	2.5908
prec_max	13.3096	16.9164
prec_median	5.2832	7.493
time left	2	16

2.5.Linear Correlation of Features

Correlation in ML is the statistical association between two or more variables. It enables us to comprehend how changes in one variable affect changes in another. The intensity and direction of the association between two variables in a dataset are frequently measured using correlation. If the value is positive, there is a positive correlation, which means that if one variable rises, the other one tends to rise as well. If the value is negative, there is a negative correlation, which means that when one variable rises, the other tends to fall. Highly correlated features can make the model unstable and increase the difficulty in interpreting the individual contributions of the correlated variables. On this basis, highly correlated features were disregarded and only the most relevant feature was kept. According to Figure 1, it seems that the highest correlation other than the temperature and precipitation are between the pipe diameter and the wall thickness (positive correlation), and that is expected because in general the bigger the diameter of the pipe the larger the wall thicknesses should be.

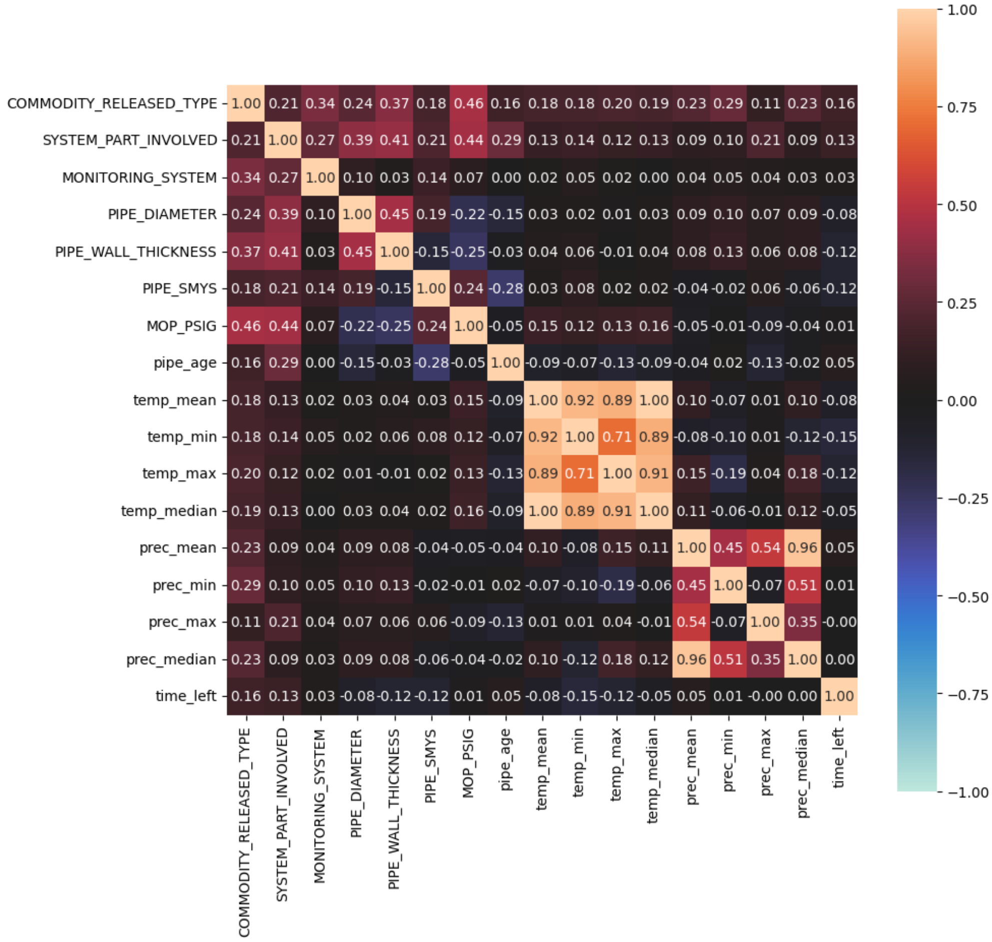

Figure 1 Heatmap of the linear correlation of features.

2.6.Data Augmentation and Binning

The size and diversity of the data dramatically affect the performance of the models. The dataset available, for corrosion, is relatively small, hence an increase in the size of the dataset is needed. Initially, when generating the time left for the pipe to fail, one random number was chosen (threshold year) between the installation year and the incident year. To augment the number of samples, more than one random number was chosen. The same process of generating the time left till failure explained in section 0 is implemented but instead of choosing only one random year three random years are chosen. The example in section 0 will explain this further; for the same pipe whose installation year is 2002 and failure year is 2016, not only 2014 was chosen as a threshold year, but also 2005 and 2015. As such, the same pipe is represented by three examples instead of one. The code was designed in a way to avoid the generation of identical values for threshold years. Using this method, the dataset size increased from 851 data points to 2051. Figure 2 shows the distribution before and after augmentation. The data is more uniformly distributed along the intervals. It is worth mentioning that all examples are unique, there are no duplicates, and only values of "time left till failure" less than 32 years were kept. This goes back to the logical limitation that it is preposterous to try to predict that a pipe will fail after half a century for example.

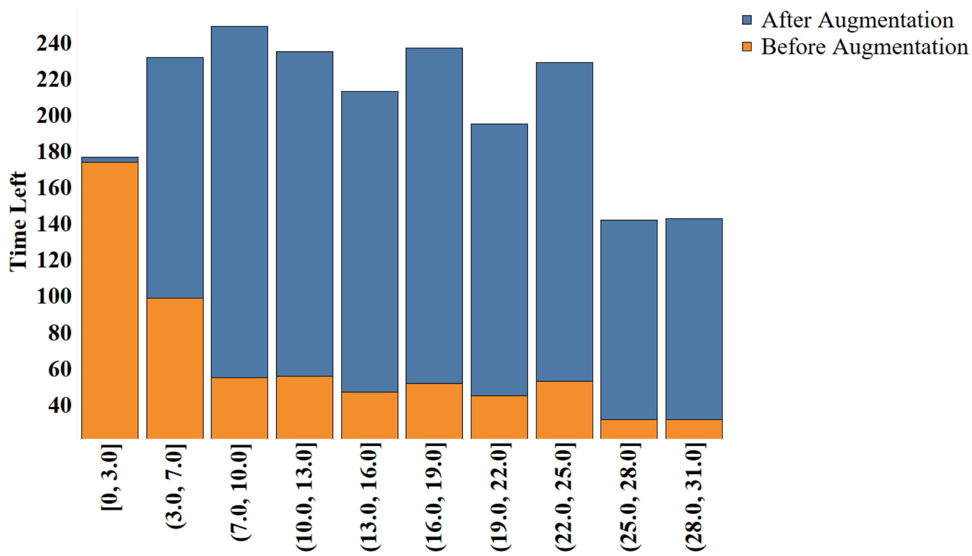

Figure 2 Data distribution along intervals before and after augmentation.

Binning of Time Left till Failure: For training classification models, continuous prediction variables must be transformed into classes or bins. As such, the time left till failure was transformed into a categorical variable by using equal-frequency binning (also referred to as quantile binning or quantization), which is a technique that transforms continuous variables into categorical variables by segmenting the data into bins that contain an equal count of observations. This method guarantees that each category encompasses approximately the same number of data points, allowing for the creation of balanced groups, and mitigating the influence of imbalance. The number of bins will be discussed later in the coming sections.

CHAPTER 3

MACHINE LEARNING IMPLEMENTATION

3.1.Algorithm Selection

Different ML models for regression and classification were trained and assessed using different performance metrics. These models have been trained with the default hyperparameters of their respective libraries; XGBoost [36], LightGBM [37], and Scikit-learn [38]. Once the best algorithm is selected, the hyperparameters are then tuned. For regression, multiple algorithms were investigated including Extra Trees Regressor (XTR), XGB Regressor, Gaussian Process Regressor, Hist Gradient Boosting Regressor, LGBM Regressor, and many more. The best-performing model with the default hyperparameters was the XTR [39] achieving an R-Squared of 88%. A summary of the investigated algorithms along with their achieved performance can be found in Table 3.

For classification, multiple algorithms were used including Extra Trees Classifier (XTC), Random Forest Classifier, XGB Classifier, LGBM Classifier, Bagging Classifier, and many others with a multiple number of classes to predict (notably 4, 8, 12, and 16 classes). The best-performing model with the default hyperparameters was the XTC [39] achieving an accuracy of 80%, 61%, 54%, and 33% for bins of 4, 8, 12, and 16 respectively. A summary of the investigated algorithms along with their achieved performance can be found in Table 4.

Figure 3 summarizes the performance of different classification models on different bin sizes, XTC followed by RandomForestClassifier achieved the best accuracy in total. It can be seen that all models perform best on 4 bins, followed by 8 bins, 12 bins, and finally 16 bins. It can also be noticed that the accuracies of all the models for 16 bins are very similar and do not vary much.

Table 3 Evaluation metrics of best-performing ensemble models for regression.

Model	Adjusted R-Squared	R-Squared	RMSE
ExtraTreesRegressor	0.86	0.88	3.05
XGBRegressor	0.86	0.88	3.05
GaussianProcessRegressor	0.84	0.86	3.26
HistGradientBoostingRegressor	0.83	0.85	3.4
LGBMRegressor	0.83	0.85	3.41

Table 4 Evaluation metrics of best-performing ensemble models for classification for different numbers of bins.

Bins	Model	Accuracy	F1 Score
4	ExtraTreesClassifier	0.8	0.8
	RandomForestClassifier	0.8	0.8
	XGBClassifier	0.78	0.78
	BaggingClassifier	0.77	0.77
	LGBMClassifier	0.77	0.77
8	ExtraTreesClassifier	0.61	0.6
	RandomForestClassifier	0.61	0.59
	XGBClassifier	0.58	0.57
	LGBMClassifier	0.57	0.56
	BaggingClassifier	0.56	0.55
12	ExtraTreesClassifier	0.54	0.55
	LGBMClassifier	0.5	0.52
	RandomForestClassifier	0.5	0.51
	BaggingClassifier	0.48	0.49
	XGBClassifier	0.48	0.5
16	RandomForestClassifier	0.33	0.32
	LGBMClassifier	0.3	0.31
	ExtraTreesClassifier	0.31	0.31
	XGBClassifier	0.3	0.31
	KneighborsClassifier	0.29	0.26

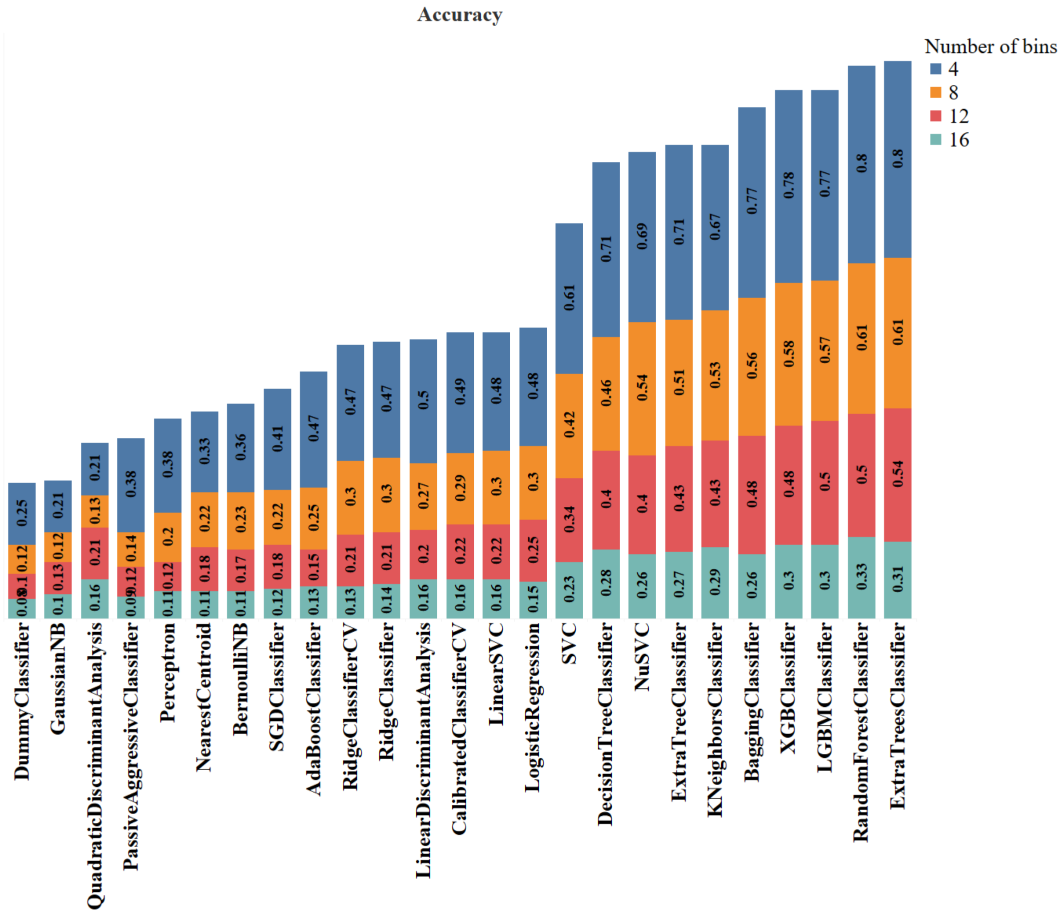

Figure 3 Accuracy of models for different number of bins.

ExtraTrees provide good computational complexity and improved accuracy [39], and like ensemble models can handle different types of data and lessen the variance and therefore increase accuracy [40]. ExtraTrees is an ensemble ML approach, similar to Random Forests (RF), that trains a large number of decision trees and combines the output of the group of decision trees to produce a forecast. To ensure that the decision trees are sufficiently varied, RF employs bagging to choose various versions of the training data. Extra Trees, on the other hand, trains decision trees using the complete dataset. Also, it randomly selects the values at which to split a feature and build child nodes to ensure that there are enough variances between individual decision trees. In a RF, however, a greedy search technique is employed to choose the value at which to split a feature. XTR and the XTC performed the best, and are preferred over XGBoost, LGBM, or RF because they take less time to train and consequently are easier to tune.

3.2.Effect of Bin Size

In classification problems, the number of classes the model needs to predict affect greatly the performance of the model. If a model has to predict one out of two classes, it will be right 50% of the time even if it does not have any predicting power. However, if it has to predict one class out of 30 classes, then being right 50% of the time needs a good predictive power. Using the concept of decision boundaries is another helpful method to approach this. Considering a low dimensional data set that can be linearly separated in the locations where data is located, the decision border and the area around it will occupy a relatively low amount of space. As the dimensionality, curvature, and data range increase, the decision boundary becomes more complex, occupying a larger volume in the data space. Typically, whichever algorithm is employed to build the decision border, all that's provided is an approximation of where it is. The probability that the true border—or the best boundary given the population of data, of which the training data is simply a sample—is on the incorrect side of data points. XTC was trained on a multiple number of classes starting from 2 (binary problem) to 20, and for each case, the accuracy and the f1-score were reported.

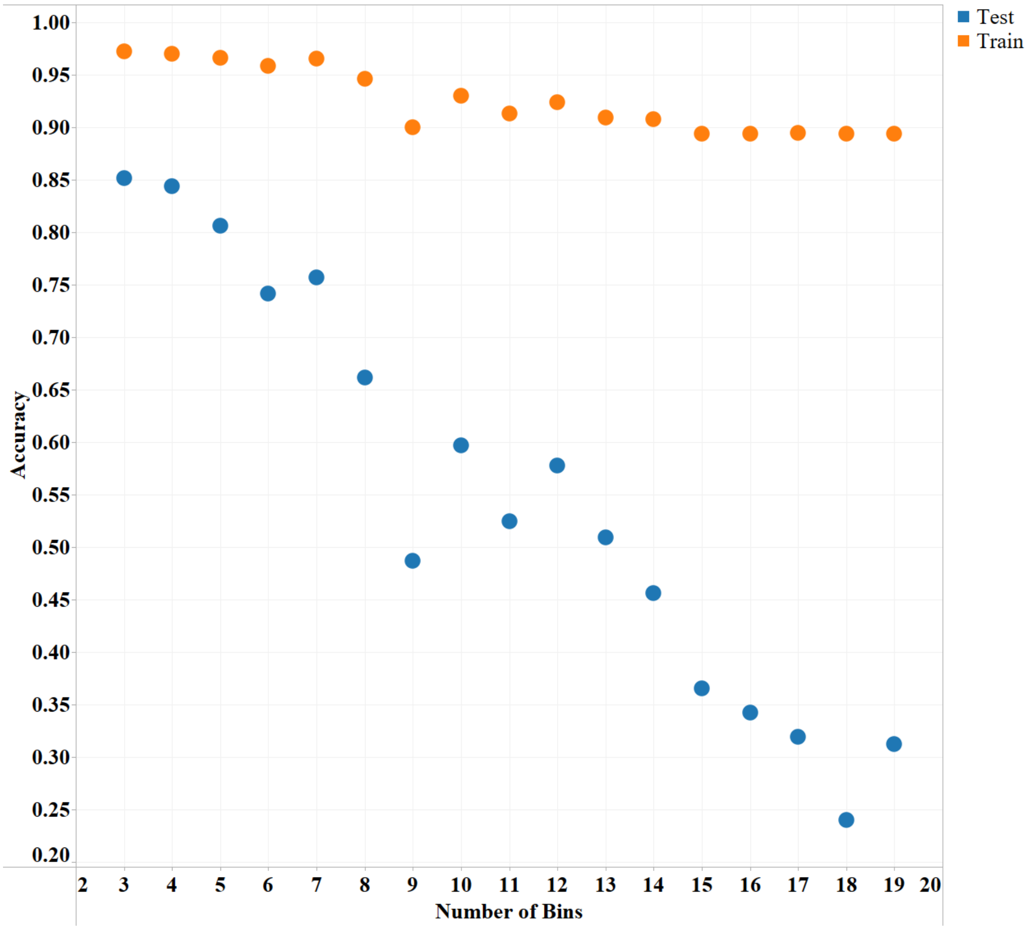

Figure 4 Variation of accuracy in the function of the number of bins for test and train sets.

Figure 4 proves that the higher the number of bins the lower the accuracy. Although it might seem that there are many more classes to predict in a regression problem, that is not the case. Classification metrics measure how accurately the model was able to predict the class; if the model correctly predicts class A and the true class is class A, it is performing its job. While in regression if the correct answer is 9 years and the model predicted 12 years, the model is penalized on the relative error (3 years in this example), it is still performing its function if the relative error is considered acceptable for the application at hand.

3.3.Effect of Dataset Size

A model was trained on train sets with different sizes, the train sets varied in size (increased from 100 datapoints to 1800 datapoint) while the test was of the same size (261 examples) and remained unchanged, the test set size was kept fixed to ensure a fair comparison among models. Figure 5 shows the trend of variation of the accuracy on the test set, while the training dataset size was varied.

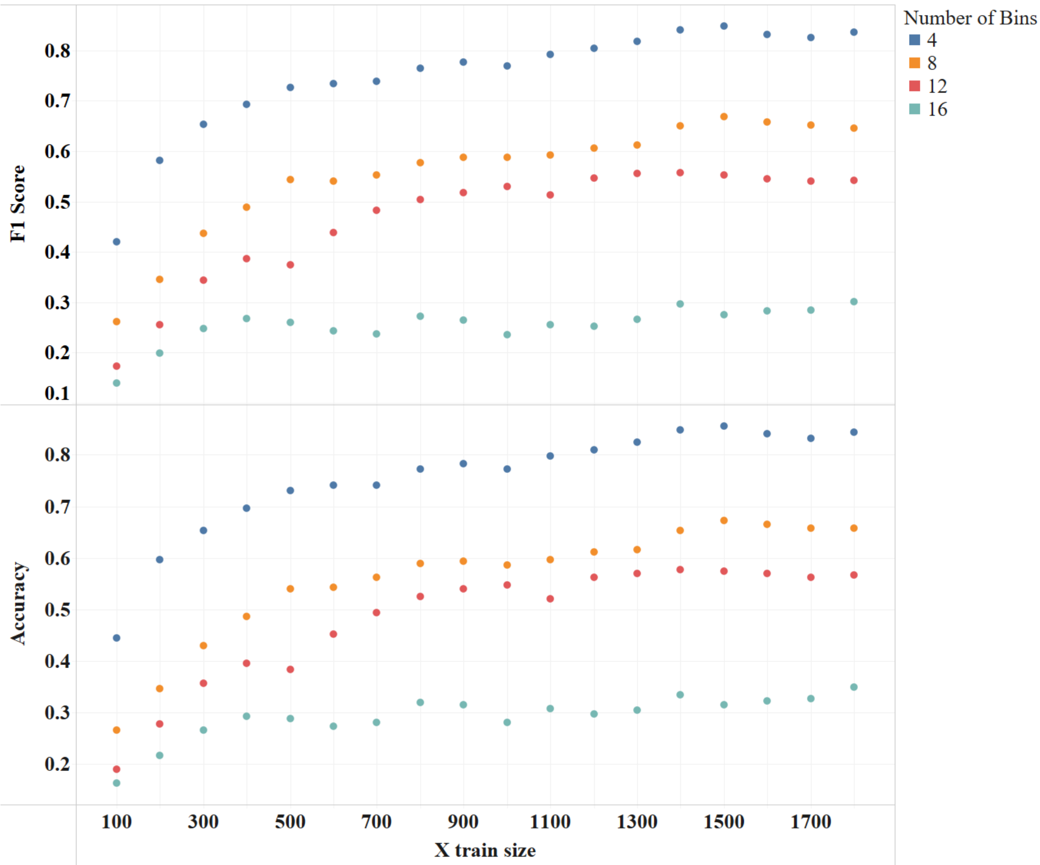

Figure 5 Variation of accuracy and F1 score with train set size for different numbers of bins.

It can be seen that the accuracy on the test set increases with the increase of the train set and does not saturate until almost the full dataset is used.

In the case of regression, as the training size increases the same increase in R-squared is noticed, but even with the full dataset used, the performance is not fully saturated and the plateau is not reached at the same speed. Figure 6 shows the variations of R-squared and MSE respectively function of the data size. The regression model needs more data points than classification to perform well. This is why this research uses the full dataset.

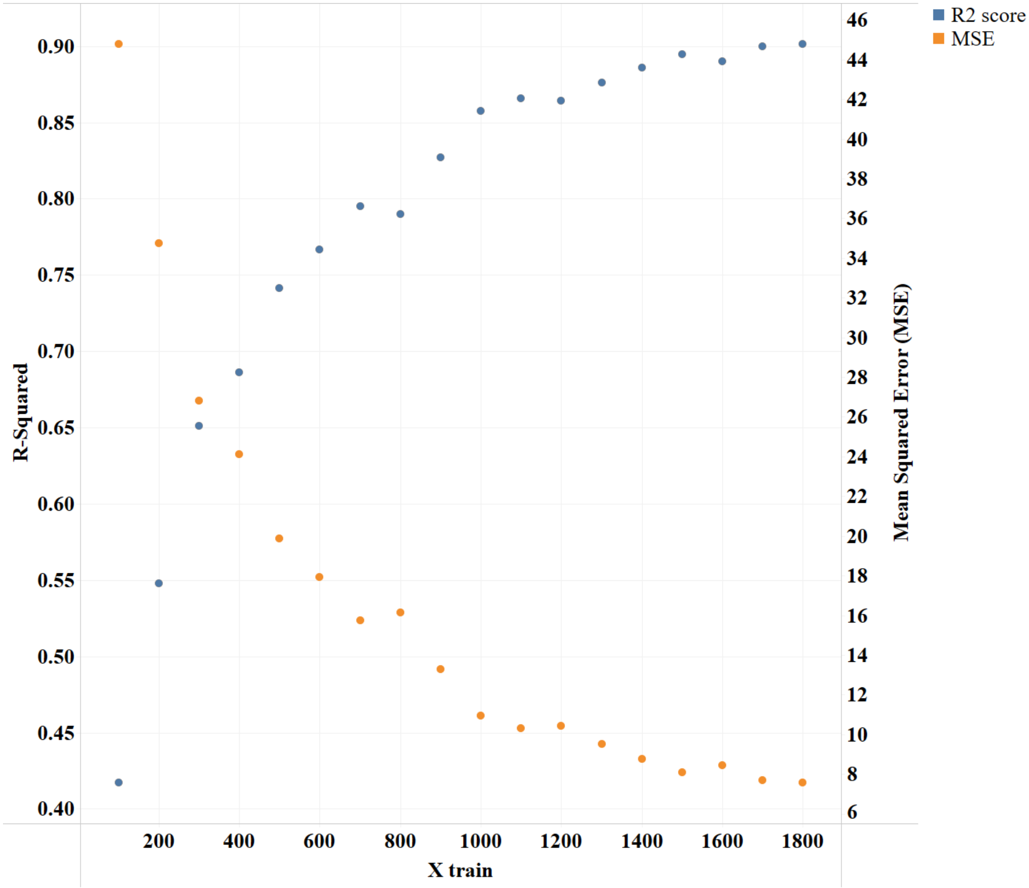

Figure 6 Variation of R-Squared and MSE with train set size.

3.4.Tuning Hyperparameters

Tuning the hyperparameters on one train/test split might lead to good values of the performance metric for classification and regression but it might not be the absolute best performing because a simple shuffling of the examples between test and train sets can ruin results. The best model is a model insensitive to the train/test split. That is achieved by using a resampling technique called cross-validation which uses several data subsets to evaluate and train a model across many iterations. A dataset of known data is typically provided to a model for training (the training dataset) and a dataset of unknown data (data never seen before) for testing (validation dataset or testing set). Cross-validation seeks to identify issues like overfitting or selection bias by testing the model's propensity to predict new data that was not utilized in its training. In addition, it gives an insight into how the model will perform on an unknown never-seen-before dataset. Partitioning a sample of data into complementary subsets, running the analysis on one subset (referred to as the training set), and validating the analysis on the other subset (referred to as the validation set or testing set) constitutes one round of cross-validation. Most approaches use numerous cross-validation rounds using various partitions to reduce variability, and the validation results are pooled (e.g., averaged) over the rounds to provide an estimate of the model's predictive performance. In addition to cross-validation, automatic tuning algorithms were used such as RandomizedSearchCV [41] and BayesSearchCV [42]. These algorithms perform hyperparameter search and return the model with the best-performing parameters. They automatically search for the designated hyperparameters, and they do cross-validation. The difference between Randomized and Bayes search is the way the hyperparameters are tried. In a randomized search of parameters only part of the parameters is tried out randomly, while in Bayes search the algorithm optimizes its parameter selection i.e., the algorithm tries relevant spaces and discards the search spaces that are not likely to provide an optimal solution. The parameters that were tuned for regression and classification are the number of estimators (number of trees in the forest), the max tree depth (maximum depth of the tree), the max number of features (number of features to consider when looking for a split) and the criterion (function to measure the quality of split). The search space for hyperparameters can be found in Table 5.

Table 5 Hyperparameter search space for tuning of regression and classification models.

Parameter	Classification	Regression
Number of Estimators	[50; 1500]	[50; 1500]
Max Depth	[4; 40]	[4; 40]
Criterion	gini, entropy, log_loss	squared error, absolute error, friedman mse, poisson
Max Features	sqrt, log2, None	sqrt, log2, None

CHAPTER 4

RESULTS AND DISCUSSION

4.1.Tuned Models

4.1.1. Regression

After tuning the XTR, it reached an R2 score of 90.36% and RMSE of 2.72 on testing. The tuned model has the following hyperparameters: criterion: Friedman mse, max depth: 33, max features: None, number of estimators:1023.

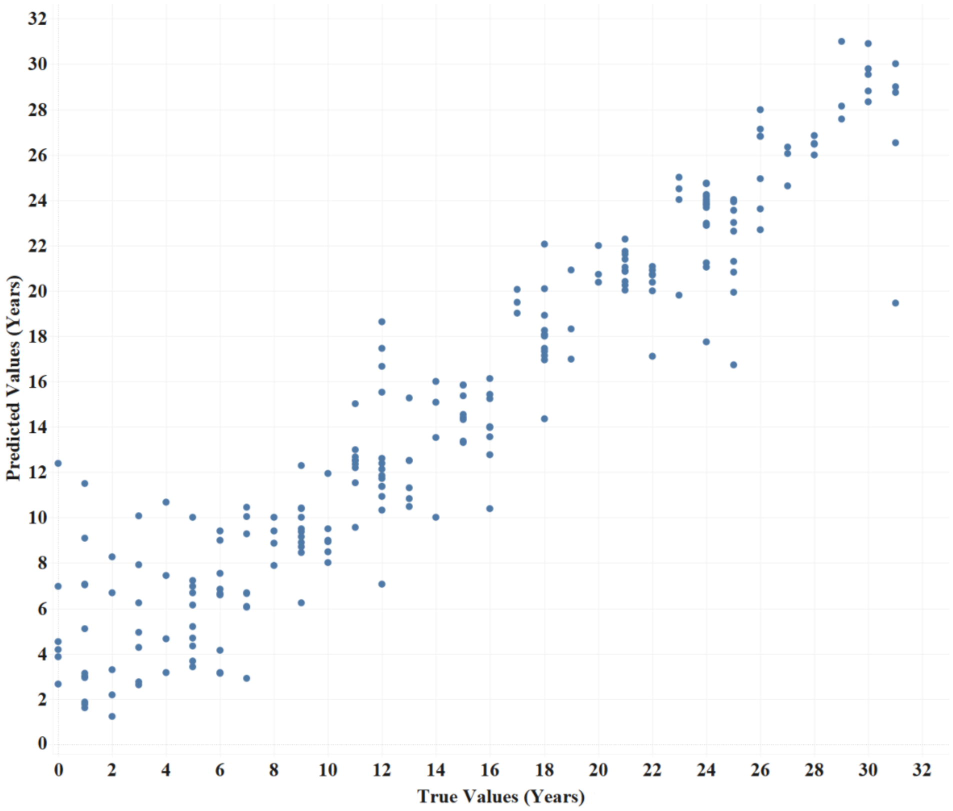

Figure 7 Predicted values function of true values for tuned XTR model.

Figure 7 shows the predicted values of the model plotted versus the true values. The closer the dots are to the diagonal at 45 degrees the more performing the model is. For values under ten years, there's a slight dispersion of the data, while after ten years the data comes closer to the diagonal. This can be due to the fact that the causes inducing pipes to fail due to corrosion at this early stage are usually uncommon and can't be always foreseen.

- ON 2/11/2017 AT APPROXIMATELY 02:54 CST, OPERATING PERSONNEL DISCOVERED A 1-GALLON SPILL UNDERNEATH A 90-DEGREE ELBOW IN ECHO STATION. IN RESPONSE, METER SKID ASSOCIATED WITH THE ELBOW WAS ISOLATED. AT APPROXIMATELY 10:15 CST IT WAS DETERMINED THAT THE REPAIR COST WOULD EXCEED $50,000 AND AN NRC NOTIFICATION WAS IMMEDIATELY MADE AT 10:25 CST. FOR REPAIR, THE ELBOW WAS CUT OUT AND THE METER SKID PLACED BACK INTO SERVICE. BASED ON THE METALLURGICAL ANALYSIS IT HAS BEEN CONCLUDED THAT THE CAUSE OF THE RELEASE WAS PREFERENTIAL WELD CORROSION. TO MINIMIZE RECURRENCE, SIMILAR LOW-FLOW AND INTERMITTENT-FLOW PIPING IN THIS FACILITY HAS BEEN IDENTIFIED AND RECEIVES FLUSHING AT SCHEDULED INTERVALS. NRC #1170795 MADE ON 2/11/2017 AT 10:25 CST

- SUBSEQUENT TO DISCOVERY THE LINE WAS EXCAVATED AND A THROUGH WALL FAILURE WAS OBSERVED AT THE 6 O'CLOCK POSITION. THE FAILED PIPE WAS CUT OUT AND REPLACED. LINE WAS RESTARTED ON SEPTEMBER 23, 2016 UNDER A PHMSA REQUIRED 20% REDUCTION IN OPERATING PRESSURE. METALLURGICAL ANALYSIS OF THE FAILED PIPE CONCLUDED THAT THE CAUSE OF THE THROUGH WALL LEAK WAS EXTERNAL PITTING/CORROSION LOCATED AT THE 6 O'CLOCK ORIENTATION, ADJACENT TO A GIRTH WELD, AT A SINGLE PIT. THE PRIMARY CONTRIBUTING FACTOR TO THE CORROSION RATE WAS DC STRAY CURRENT INTERFERENCE FROM AN ADJACENT PIPELINE.

- BASED ON FIELD ASSESSMENT, THE CAUSE OF THE INTERNAL CORROSION WAS DETERMINED TO BE MICROBIOLOGICALLY INDUCED. PART G HAS BEEN UPDATED TO REFLECT THIS INFORMATION. THIS PIPING WAS REMOVED AND REPLACED WITH ABOVEGROUND PIPING, IN ORDER TO PREVENT REOCCURRENCE.

- THE FAILURE OCCURRED DUE TO THROUGH-WALL CORROSION CAUSED BY A LOCALIZED CORROSION CELL ON THE INTERIOR SURFACE OF THE PIPE. LOW FLUID VELOCITY WITH INTERMITTENT STAGNANT CONDITIONS ON THE PIPE SEGMENT ALLOWED SOLIDS AND WATER TO ACCUMULATE ON THE BOTTOM OF THE PIPE. CHLORINE AND SULFUR WERE PRESENT AS CORROSION PROMOTERS, AND EVIDENCE OF MICROBIOLOGICALLY INDUCED CORROSION WAS FOUND.

Figure 8 Snippets of operator's narrative of incidents of pipes failing due to corrosion under four years after installation

Figure 8 shows some narrative from the operator for pipes failing due to corrosion under four years after installation [29]. This is a part of the dataset where the operator mentions his own description of the incident. At early stages the causes of the corrosion are not always foreseeable e.g., the presence of stray current and microbiological factors greatly expedite corrosion in pipes. As will be discussed later, classification is more robust to changes in factors contributing to early corrosion of pipes.

4.1.2. Classification

After tuning for 4 bins, the model reached an accuracy of 85% on testing, Figure 9 shows the confusion matrix of the test set. The tuned model has the following hyperparameters: criterion: entropy, max depth: 17, max features: sqrt, number of estimators: 1432.

The model was also tuned for 8 bins. Figure 10 shows the confusion matrix for 8 bins, an accuracy of 67% was reached after the tuning. The tuned model has the following hyperparameters: criterion: entropy, max depth: 12, max features: sqrt, number of estimators: 1484.

Table 6 and Table 7 show the precision, recall and F1 score of the models for 4 bins and for 8 bins, the support is the number of examples in each class.

Figure 9 Confusion matrix of XTC for 4 Bins.

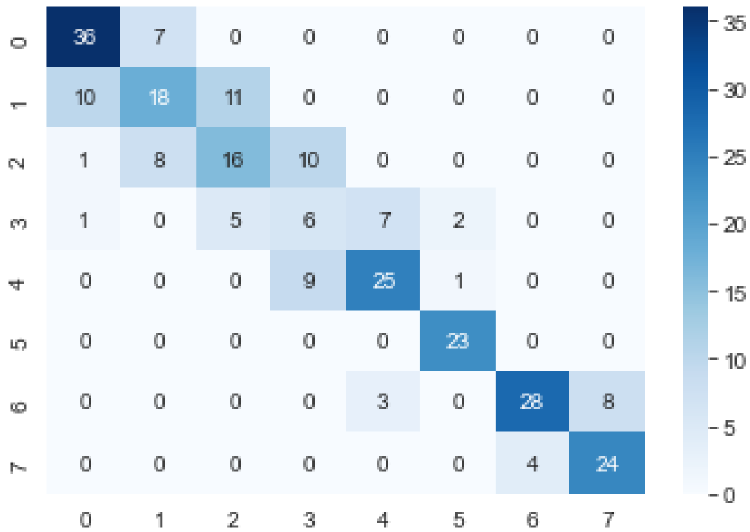

Figure 10 Confusion matrix of XTC for 8 Bins.

Table 6 Classification metrics of XTC 4 bins model.

Label	precision	recall	f1-score	support
0	0.9	0.88	0.89	82
1	0.66	0.7	0.68	56
2	0.81	0.83	0.82	58
3	1	0.97	0.98	67
Average	0.84	0.84	0.84	263

Table 7 Classification metrics of XTC for 8 bins model.

Label	precision	recall	f1-score	support
0	0.75	0.84	0.79	43
1	0.55	0.46	0.5	39
2	0.5	0.46	0.48	35
3	0.24	0.29	0.26	21
4	0.71	0.71	0.71	35
5	0.88	1	0.94	23
6	0.88	0.72	0.79	39
7	0.75	0.86	0.8	28
Average	0.66	0.67	0.66	263

To better understand how the models are making predictions, and thus explain what's happening from the input to the output, the models' interpretability will be discussed in the next section. For classification interpretability, the focus will be on the 4 bins model. Using 4 bins is more practical because the interval's width is more suitable, in the case of 8 bins the intervals become very small which makes the model redundant, especially with the presence of a regression model. In addition to that the better performance of the 4 bins model can be leveraged to better interpret the results.

4.2. Interpretability

A technique called SHAP [43] values (Shapley Additive exPlanations) is used to make ML models more transparent and understandable. It is based on cooperative game theory.

For instance, linear models can utilize their coefficients to measure the overall significance of each characteristic, but because they scale with the variable, this may cause distortions and misinterpretations. Additionally, the coefficient is unable to take into consideration the feature's local significance or how it alters with lower or higher values. The same holds true for the feature importance of tree-based models, which is why SHAP is helpful for model interpretability. It's important to note that while SHAP illustrates the value or contribution of each feature to the model's prediction, it does not assess the quality of the prediction itself. Since the models being interpreted are ensemble models a specific implementation of SHAP is used called TreeExplainer [44] which is designed to interpret the outputs of tree-based algorithms.

4.2.1. Regression

Figure 11 is the bee swarm plot of the SHAP values, the first feature is the most important while the last feature is the least important.

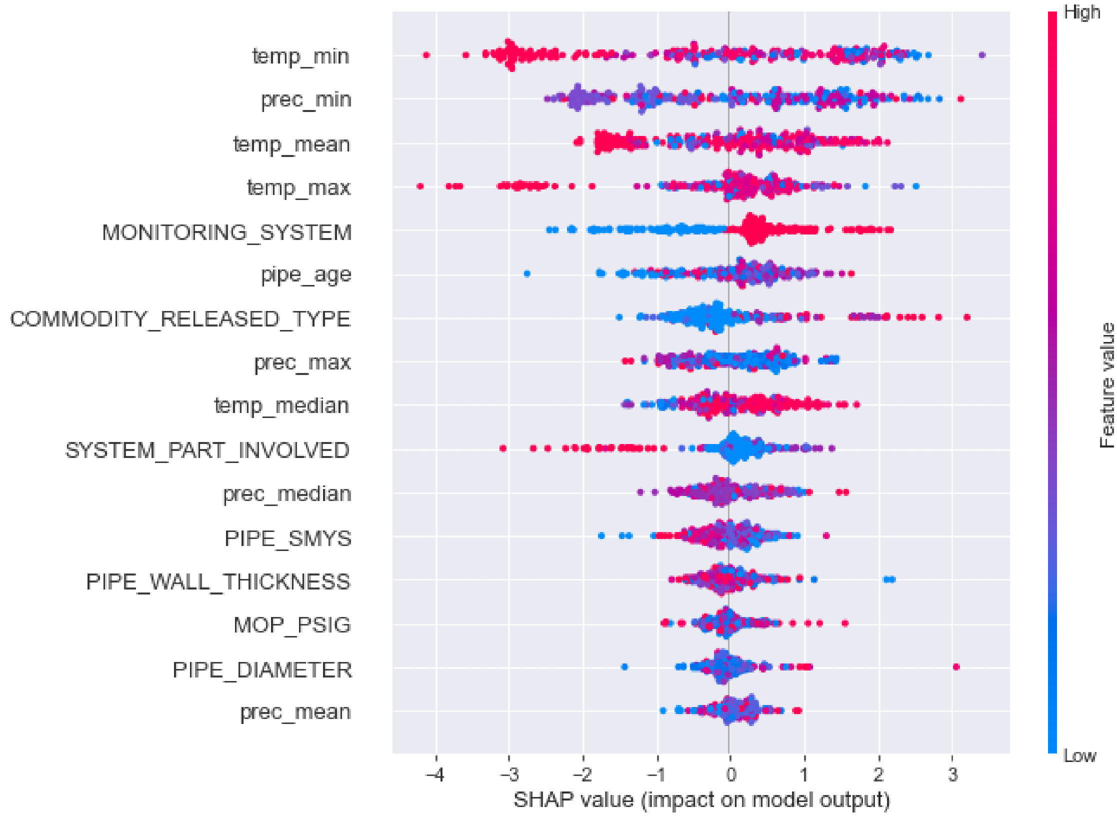

Figure 11 SHAP Values for the XTR model.

As can be seen, the temperature statistical features make a high contribution to the model's output. High values of the median temperature have a high positive contribution while lower values of the median temperature have a low negative contribution. This is consistent with the literature where it was shown that temperature can be a factor affecting the redox reaction leading to corrosion [14]. The precipitation mean has the least contribution on the model's output..

4.2.2. Classification

Figure 12 shows the contribution of the features for the classification model. The first feature is the most contributing one while the last feature is the least contributing.

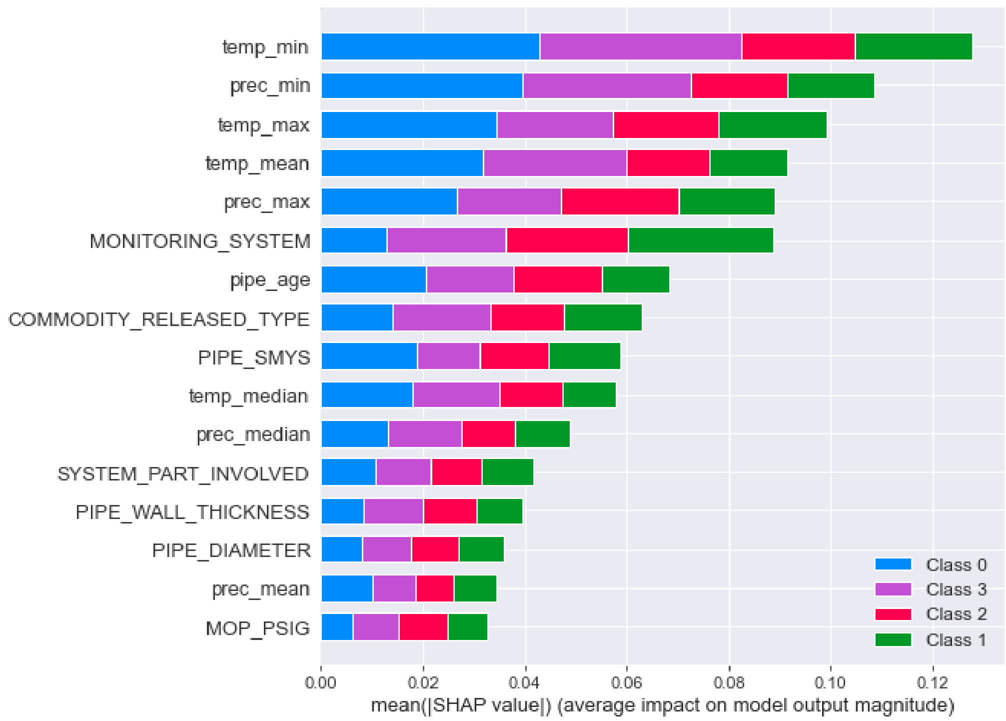

Figure 12 SHAP Values for XTC model

The minimal temperature is contributing mostly to the prediction of classes 0 and 3. In classification also the statistical properties of temperature and precipitation have a high contribution to the output of the model. This is consistent with the literature, temperature has a direct effect on the acceleration of the corrosion process [45].

Based on these explanations, the addition of the temperature and the precipitation features is beneficiary for the model performance, this validates their importance in the dataset. The presence of a monitoring system also affects positively the predictions, as well as the commodity transported by the pipe as well as the pipe age. Features like the thickness of the walls, the maximum operating pressure, specified inimal yield strength, and the pipe diameter have contribute but in lesser extents, but they still provide insights.

4.2.3. *Effect of Precipitation and Temperature*

Temperature and precipitation affect the performance of the models. To further understand the impact of these features they will be removed in the following section and the performance of the model will be monitored.

The regression model performance drops to achieve an R2 score of 86% and RMSE of 3.27 Figure 14 shows the distribution of the predicted and true values. The classification model's performance also drops to achieve an accuracy of 82.5%, Table 8 and Figure 13 show the classification report and the confusion matrix.

Given the drop in performance for regression and classification models when the temperature and precipitation features are removed, the importance of these features is further confirmed, as well as the major role they play in predicting the time left till failure for a given pipe.

Table 8 Classification metrics for XTC after removal of temperature and precipitation data.

Label	precision	recall	f1-score	support
0	0.87	0.84	0.86	82
1	0.64	0.62	0.63	56
2	0.77	0.84	0.8	58
3	0.98	0.96	0.97	67
Average	0.83	0.83	0.83	263

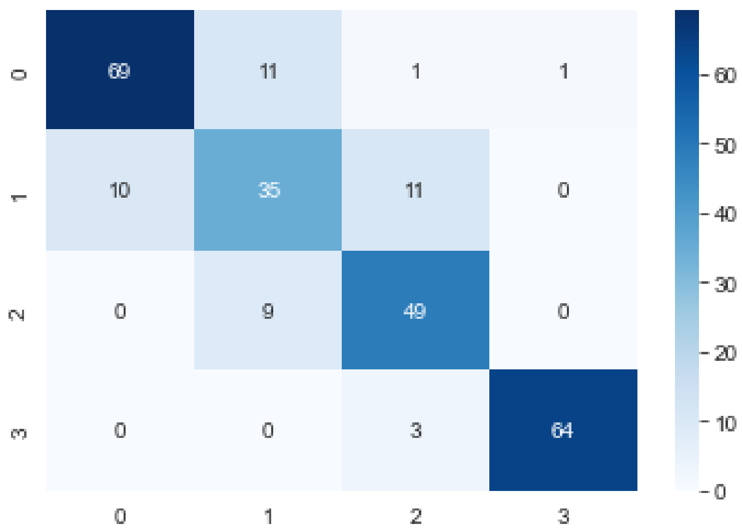

Figure 13 Confusion matrix of XTC for 4 bins without temperature and precipitation data.

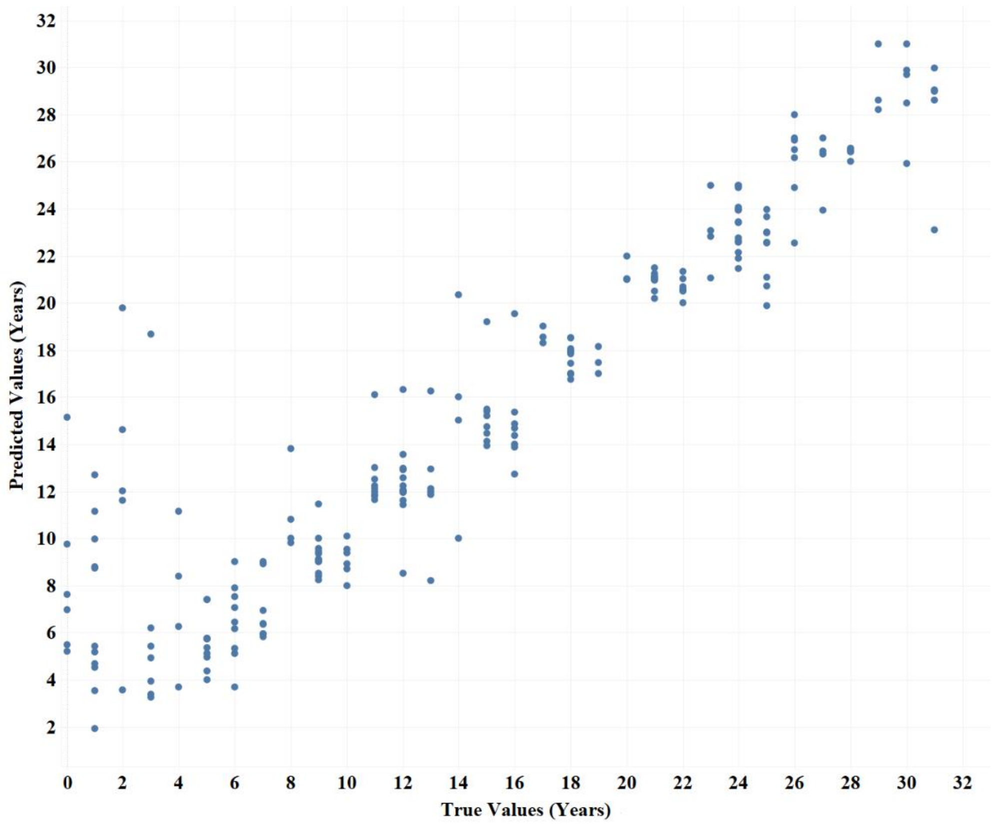

Figure 14 Predicted values function of true values of XTR without temperature and precipitation data.

CHAPTER 5

CONCLUSIONS

Emergency repairs, cleanup efforts, environmental remediation, and other adverse outcomes due to corrosion-related failures in pipelines can be mitigated with service life prediction models. In this study, the time left till the failure of a pipe due to corrosion was accurately predicted.

Two models were used: a regression model to predict the exact time left till failure and a classification model that predicts an interval within which the pipe should fail. The two models were trained on the PHMSA dataset to which was added the temperature and precipitation data from the National Centers for Environmental Information. Different ML models were compared and the XTR and XTC algorithms proved to have the best performance for regression and classification respectively.

The optimal bin sizes and the optimal data sizes were investigated, and to understand the most contributing features the outputs of the models were interpreted. The best-performing algorithm was the Extremely Randomized Trees (Extra Trees) algorithm which achieved, after tuning, an R2 score of 90.35% for regression and an f1 score of 85% for classification on the test set.

Based on the aforementioned, XTR and XTC models can be effectively used to predict the remaining service life of pipes, and that depends in great part on the temperature and precipitation conditions added to the dataset.

Future work may include predictive maintenance strategies by optimizing maintenance plans and resource allocation using the precise predictions of the time left before failure; prioritizing maintenance tasks can lower costs and boost operational effectiveness by identifying pipes that are likely to fail soon. It may also include structures other than pipes that also have a large, detailed dataset. In addition to that a prospective field might be the integration of Internet of Things (IoT) sensors to collect real-time data like humidity and temperature, and pipe wall thicknesses and add them to the existing dataset to be able to predict the time left till failure in smaller units (months or days instead of years).

REFERENCES

[1] "Global Oil and Gas Pipelines Industry Outlook to 2023 - Capacity and Capital Expenditure Outlook with Details of All Operating and Planned Pipelines," in *PR newswire*, ed: PR Newswire Association LLC, 2019.

[2] M. Hussein, "Mapping the world's oil and gas pipelines," in *ALJAZEERA*, ed, 2021.

[3] H. A. Kishawy and H. A. Gabbar, "Review of pipeline integrity management practices," *International Journal of Pressure Vessels and Piping*, vol. 87, no. 7, pp. 373-380, 2010/07/01/ 2010, doi: https://doi.org/10.1016/j.ijpvp.2010.04.003.

[4] B. Martin, "Supply Chain Disruptions: The Risks and Consequences," vol. 2023, ed, 2021.

[5] E. J. Omonbude, "The economics of transit oil and gas pipelines: a review of the fundamentals," (in English), *OPEC energy review*, vol. XXXIII, no. 2, pp. 125-139, 2009. [Online]. Available: https://go.exlibris.link/HMRmrg78.

[6] Agencies, "Corrosion outside oil pipeline caused crude spill on California coast," in *The Guardian*, ed, 2016.

[7] E. Arzaghi, B. H. Chia, M. M. Abaei, R. Abbassi, and V. Garaniya, "Pitting corrosion modelling of X80 steel utilized in offshore petroleum pipelines," *Process Safety and Environmental Protection*, vol. 141, pp. 135-139, 2020.

[8] C. P. Vetter, L. A. Kuebel, D. Natarajan, and R. A. Mentzer, "Review of failure trends in the US natural gas pipeline industry: An in-depth analysis of transmission and distribution system incidents," (in English), *Journal of loss prevention in the process industries*, vol. 60, pp. 317-333, 2019, doi: 10.1016/j.jlp.2019.04.014.

[9] M. E. A. Ben Seghier, B. Keshtegar, M. Taleb-Berrouane, R. Abbassi, and N.-T. Trung, "Advanced intelligence frameworks for predicting maximum pitting corrosion depth in oil and gas pipelines," *Process Safety and Environmental Protection*, vol. 147, pp. 818-833, 2021/03/01/ 2021, doi: https://doi.org/10.1016/j.psep.2021.01.008.

[10] "Part I: General Aspects of Corrosion, Corrosion Control, and Corrosion Prevention," in *Corrosion Atlas Case Studies*: Elsevier Inc, 2022, pp. xxv-xlii.

[11] K.-Y. Wu and A. Mosleh, "Effect of temporal variability of operating parameters in corrosion modelling for natural gas pipelines subject to uniform corrosion," *Journal of Natural Gas Science and Engineering*, vol. 69, p. 102930, 2019/09/01/ 2019, doi: https://doi.org/10.1016/j.jngse.2019.102930.

[12] corrosionpedia.com/contributors/corrview-com, "21 Types of Pipe Corrosion & Failure," in *Corrosionpedia*, ed, 2020.

[13] R. Norsworthy, "1 - Understanding corrosion in underground pipelines: basic principles," in *Underground Pipeline Corrosion*, M. E. Orazem Ed.: Woodhead Publishing, 2014, pp. 3-34.

[14] S. Harsimran, K. Santosh, and K. Rakesh, "Overview of corrosion and its control: A critical review," *Proceedings on Engineering*, vol. 3, no. 1, pp. 13-24, 2021.

[15] C. I. Ossai, "Advances in Asset Management Techniques: An Overview of Corrosion Mechanisms and Mitigation Strategies for Oil and Gas Pipelines," (in English), *ISRN Corrosion*, 2012 2012, doi: https://doi.org/10.5402/2012/570143.

[16] V. B. Sharma *et al.*, "Review of Structural Health Monitoring Techniques in Pipeline and Wind Turbine Industries," *Applied System Innovation*, vol. 4, no. 3, p. 59, 2021. [Online]. Available: https://www.mdpi.com/2571-5577/4/3/59.

[17] F. Cegla, "Ultrasonic Monitoring of Corrosion with Permanently Installed Sensors (PIMS)," in *Sensors, Algorithms and Applications for Structural Health Monitoring*. Cham: Springer International Publishing, 2017, pp. 13-20.
[18] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing,* vol. 115, pp. 213-237, 2019.
[19] M. Azimi, A. D. Eslamlou, and G. Pekcan, "Data-Driven Structural Health Monitoring and Damage Detection through Deep Learning: State-of-the-Art Review," *Sensors,* vol. 20, no. 10, p. 2778, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/10/2778.
[20] X. Fan, X. Wang, X. Zhang, and P. A. X. B. Yu, "Machine learning based water pipe failure prediction: The effects of engineering, geology, climate and socio-economic factors," *Reliability Engineering & System Safety,* vol. 219, p. 108185, 2022.
[21] P. Kumari, S. Z. Halim, J. S.-I. Kwon, and N. Quddus, "An integrated risk prediction model for corrosion-induced pipeline incidents using artificial neural network and Bayesian analysis," *Process Safety and Environmental Protection,* vol. 167, pp. 34-44, 2022.
[22] Z. Wang and S. Li, "Data-driven risk assessment on urban pipeline network based on a cluster model," *Reliability Engineering & System Safety,* vol. 196, p. 106781, 2020.
[23] Y. Yang, S. Li, and P. Zhang, "Data-driven accident consequence assessment on urban gas pipeline network based on machine learning," *Reliability Engineering & System Safety,* vol. 219, p. 108216, 2022.
[24] A. Senouci, M. Elabbasy, E. Elwakil, B. Abdrabou, and T. Zayed, "A model for predicting failure of oil pipelines," (in English), *Structure and infrastructure engineering,* vol. 10, no. 3, pp. 375-387, 2014, doi: 10.1080/15732479.2012.756918.
[25] P. D. M. Cech, W. Guijt,A. Haskamp,I. Huidobro Barrio, "Performance of European cross-country oil pipelines Statistical summary of reported spillages in 2020 and since 1971," *CONCAWE* 2022. [Online]. Available: https://www.concawe.eu/wp-content/uploads/Rpt_22-6.pdf.
[26] P. Kumari, Q. Wang, F. Khan, and J. S.-I. Kwon, "A unified causation prediction model for aboveground onshore oil and refined product pipeline incidents using artificial neural network," (in English), *Chemical engineering research & design,* vol. 187, pp. 529-540, 2022, doi: 10.1016/j.cherd.2022.09.022.
[27] G. Liu, M. Boyd, M. Yu, S. Z. Halim, and N. Quddus, "Identifying causality and contributory factors of pipeline incidents by employing natural language processing and text mining techniques," (in English), *Process safety and environmental protection,* vol. 152, pp. 37-46, 2021, doi: 10.1016/j.psep.2021.05.036.
[28] A. Ahadh, G. V. Binish, and R. Srinivasan, "Text mining of accident reports using semi-supervised keyword extraction and topic modeling," *Process Safety and Environmental Protection,* vol. 155, pp. 455-465, 2021.
[29] PHMSA. "Pipeline Incident 20 Year Trends." https://www.phmsa.dot.gov/data-and-statistics/pipeline/pipeline-incident-20-year-trends (accessed.
[30] N. C. f. E. Information, "Climate at a Glance Statewide Time Series," 2022. [Online]. Available: https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/statewide/time-series/1/tavg/all/1/1895-2022?base_prd=true&begbaseyear=1901&endbaseyear=2000.
[31] S. Z. Halim, M. Yu, H. Escobar, and N. Quddus, "Towards a causal model from pipeline incident data analysis," (in English), *Process safety and environmental protection,* vol. 143, pp. 348-360, 2020, doi: 10.1016/j.psep.2020.06.047.

[32] C. R. Harris *et al.*, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357-362, 2020.
[33] W. McKinney, "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference*, 2010, vol. 445, no. 1: Austin, TX, pp. 51-56.
[34] R. K. Singh and R. V. Sharma, "Numerical analysis for ground temperature variation," *Geothermal Energy*, vol. 5, no. 1, p. 22, 2017/11/01 2017, doi: 10.1186/s40517-017-0082-z.
[35] C. Quesada, A. Miranda, M. Hodnett, A. Santos, H. Miranda, and L. Breyer, "Seasonal and depth variation of soil moisture in a burned open savanna (campo sujo) in central Brazil," *Ecological Applications*, vol. 14, no. sp4, pp. 33-41, 2004.
[36] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, 2016. [Online]. Available: http://doi.acm.org/10.1145/2939672.2939785
http://dx.doi.org/10.1145/2939672.2939785.
[37] G. Ke *et al.*, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, vol. 30, pp. 3146-3154, 2017 2017.
[38] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," (in English), *Journal of machine learning research*, vol. 12, pp. 2825-2830, 2011, doi: 10.5555/1953048.2078195.
[39] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3-42, 2006/04/01 2006, doi: 10.1007/s10994-006-6226-1.
[40] C. Zhang and Y. Ma, *Ensemble Machine Learning: Methods and Applications*, 1. Aufl.;2012; ed. (no. Book, Whole). New York, NY: Springer-Verlag (in English U6 - ctx_ver=Z39.88-2004&ctx_enc=info%3Aofi%2Fenc%3AUTF-8&rfr_id=info%3Asid%2Fsummon.serialssolutions.com&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Abook&rft.genre=book&rft.title=Ensemble+Machine+Learning&rft.date=2012-01-19&rft.pub=Springer+New+York&rft.isbn=9781441993250&rft_id=info:doi/10.1007%2F978-1-4419-9326-7&rft.externalDocID=216493&paramdict=en-US U7 - eBook), 2012.
[41] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011 2011.
[42] *scikit-optimize/scikit-optimize*. (2021). Zenodo. [Online]. Available: https://doi.org/10.5281/zenodo.5565057
http://dx.doi.org/10.5281/zenodo.5565057
[43] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," presented at the Advances in Neural Information Processing Systems, 2017, 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.
[44] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 56-67, 2020/01/01 2020, doi: 10.1038/s42256-019-0138-9.
[45] A. V. Levy, "Erosion and Erosion-Corrosion of Metals," *Corrosion*, vol. 51, no. 11, pp. 872-883, 1995, doi: 10.5006/1.3293564.