AI SUSTEIN

An NSF EPSCoR Track-2 RII Program

Showcase #8

Predicting Natural Gas Pipeline Failures Caused by Natural Forces: An Artificial Intelligence Classification Approach

by Bright Awuku, Ying Huang, and Nita Yodo

Background

Among all the causes of pipeline failure incidents, the natural force damage on the pipeline is often inevitable and unpredictable due to the inherent properties of the natural force, such as earth movement, heavy rains, high winds, and extreme hot or cold temperatures. Although only about 3% of the overall pipeline incidents were caused by natural forces in 2022, their effect is often deliberating. It may cause catastrophic failure across the energy industry. This study highlights the potential of machine learning techniques in predicting natural gas pipeline damage caused by natural forces. The study presented in this paper aims to address the critical need for accurate forecasting of the natural force failure causes that impact natural gas pipelines. By incorporating comprehensive data on climate change into the analysis of pipeline incidents, this research represents a significant advancement in our ability to predict and mitigate the effects of natural force failures on pipeline infrastructure. The developed models can be used to assist in identifying high-risk pipelines, allowing operators to prioritize inspection and maintenance activities, leading to cost savings and improved safety.

Approach

The proposed AI algorithm is expected to contribute to preventing future incidents and minimizing catastrophic losses associated with natural gas pipeline failures caused by natural forces. The methodology employed in this study is divided into four main steps as follows.

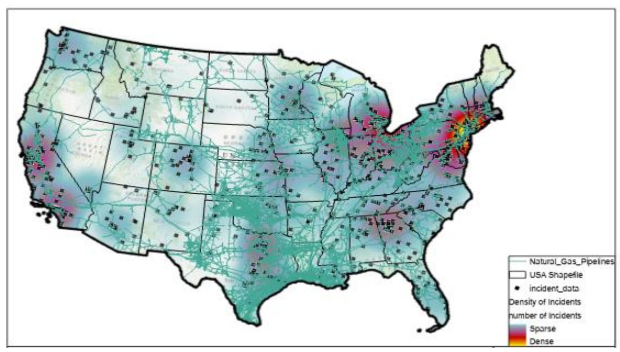

Figure 1 – Distribution of natural gas incident across U.S. based on PHMSA 20-year incident historical data.

Step 1. *Data collection*. In this study, data employed on pipeline incidents were obtained from publicly available sources provided by the Pipeline and Hazardous Material Safety Administration (PHMSA). The data highlights incidents involving natural gas transmission pipelines and spanned from 2010 to 2022, yielding 1,321 data points across the US, as shown in Figure 1. The integration of weather data into the PHMSA database was achieved by combining the location and time components as a common denominator.

Step 2. *Data pre-processing and feature selection*. After screening 96 data to remove incomplete and missing instances, the resulting dataset consisted of 81 data points and 27 features with varying magnitudes, units, and ranges; for example, the age of the pipeline is measured in years, whereas the pipe diameter is in meters. This research employed the z-score approach to scale the numerical features in the dataset and the Boruta feature selection method. Boruta is a wrapper method of feature selection, meaning it uses a model to evaluate the importance of each feature and select the most relevant ones.

Step 3. *Model selection and implementation*. The five selected classifiers are (1) k-nearest neighbors (KNN), (2) multilayer perceptron neural network (MLPNN), (3) random forest, (4) multiclass support vector machine (multiclass SVM), and (5) Extra gradient boosting classifier (XGBoost).

Step 4. *Model evaluation*. The five algorithms were evaluated using an independent testing dataset to assess their performance metrics, including accuracy, recall, precision, and F1 score.

Results and Discussion

After running the Boruta algorithm, 12 relevant features were identified, which included the pipe diameter, temperature, humidity, location, pipe age, soil type, pipe material, pipe depth, and others, as shown in Figure 2. The outcome of the feature selection process highlights the importance of incorporating the impacts of climate change stressors on natural gas pipeline damages induced by natural force.

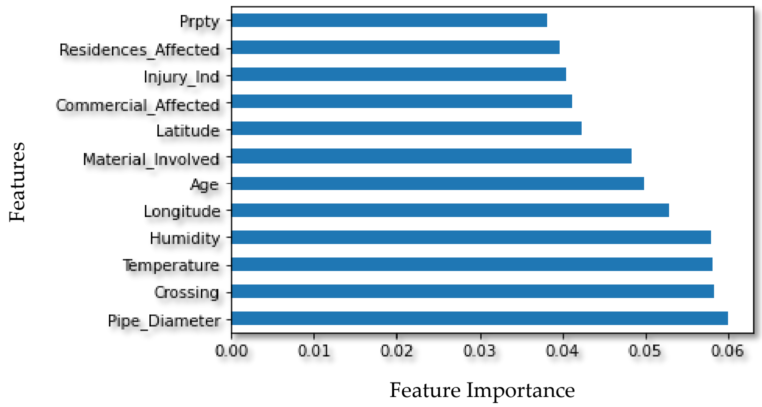

Figure 2 – Results from the feature selection process.

These 12 selected features were used in the subsequent modeling process to develop a predictive model for pipeline damage. The comparison results of in Figure 3. Results indicated that XGBoost had the highest accuracy of 92.3% on the validation dataset. Random Forest was the second-best performing algorithm with an accuracy of 92.0%, followed by SVM with 89.74% and MLPN with 87.18%. Precision and recall values were calculated for each class, and the results showed that XGBoost had the highest precision and recall values for all the classes.

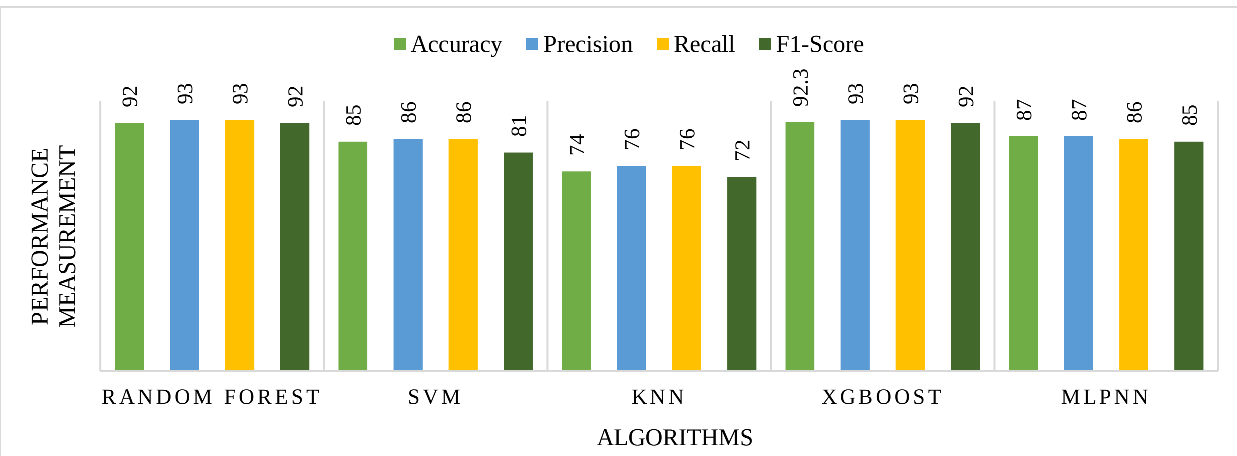

Figure 3 –Performance metrics results of algorithms used to model natural gas pipeline.

In short, XGBoost was found to be the most suitable algorithm for classifying pipeline damage caused by natural force instances with high accuracy. However, Random Forest and SVM also showed promising results and can be considered as alternative options. The findings of the study offer a framework for private and government agencies to efficiently manage their natural gas pipeline systems and preemptively avoid potential dangers. The case study results demonstrate the effectiveness of machine learning algorithms in classifying pipeline damage instances. Additionally, this research underscores the importance of leveraging the power of machine learning in predicting pipeline damage and emphasizes the need for ongoing research to enhance our understanding of the complex interactions of climate change and pipeline infrastructure monitoring and maintenance.

For more information, please refer to the full article: [Predicting Natural Gas Pipeline Failures Caused by Natural Forces: An Artificial Intelligence Classification Approach.]()