Utilizing Data-Driven Approaches To Model The Risk Of Excavation Damage To Underground Natural Gas Facilities

Dr Hamza Abusnina

Abstract: It is possible to, knowingly or unknowingly, damage underground gas services, water services, electrical services, etc. Incidents involving infrastructure damage are far more common than perceived; and these incidents result in hundreds of thousands, if not millions, of dollars in repair or replacement. Damages to underground facilities may occur by large construction contractors or by homeowners. The main objective of this research is two-fold: a) to determine the important risk factors contributing to the underground gas pipe damages; b) to identify inputs required for an effective evaluation and assessment of the risk encountered in exchange of information between different parties involved during the repair of underground gas pipelines. A predictive model will be developed based on machine learning algorithms (Logistic Regression) to be used in predicting the important risk factors affecting the underground Gas Pipe Damages. The research will systematically analyze the risk of underground gas pipeline network damage including; process the data collected from the agency, organizing/classifying the data based on certain parameters, processing the data, develop an integrated risk model and influence diagram. Next, Bayesian Network will be developed based on the derived important factors, and calculated probabilities for each attribute.

Index Terms: Risk assessment, Bayesian Network, Machine Learning

1. INTRODUCTION

Currently, in the United States, there are thousands of gas pipe miles: long grids and networks of natural gas lines across the states. Recent pipeline leaks and explosions in various regions of the US have driven the industry to re-evaluate ongoing efforts aimed at the aggressive pursuit of preventive strategies. Considering that safety and environmental risk is a major issue, particularly in cases where the underground gas line The importance of protecting underground utilities is evident, but necessary precautions are less known. Contractors and homeowners often disregard or unknowingly excavate with imminent danger in the subsurface. Out of all the underground facilities, natural gas lines pose a major threat to public health and wellbeing. Even when damages are not caused by excavation, gas pipelines may have a lasting impact if not repaired or checked upon. This study compares the failure data from various pipelines to investigate the trend for rates of failure, causes of failure, aging characteristics and relationship between the causes of damage and pipeline parameters. As the construction field continues to expand, it is important to focus on maintaining a high level of safety in order to protect occupational hazards and ensure public safety. Damages relating to excavation practices and procedures directly impact public safety due to the nature of the infrastructure system in place. It is possible to, knowingly or unknowingly, damage underground gas services, water services, electrical services, etc. Incidents involving infrastructure damage are far more common than perceived; and these incidents result in hundreds of thousands, if not millions, of dollars in repair or replacement. Damages to underground facilities may occur by large construction contractors or by homeowners. As per NTSB 2003 report, Tech Consultants was awarded similar work to be performed at1820 West 3rd Street Figure 1. The project manager surveyed the worksite and determined that sidewalk and curbing replacement was needed in front of the residences at 1816, 1818, and 1820 West 3rd Street. Then, Tech Consultants show the work to be done in front of the three addresses shown in Figure 2. On June 23, 2003, Tech Consultants issued a change order to the contractor, which included the 1820 West 3rd Street address location in a list of additional address locations. Tech Consultants did not give the 1816 and 1818 addresses or sketch to the contractor. The underground utility by mistake marked for 820 West 3rd. Therefore, the damage to the gas line happened. The primary reason for the explosion was miscommunication between the contractor and the consultants. In addition, the wrong marking was part of the problem. Also, the failure of the Tech Consultants to verify that all underground facilities were marked within the proposed dig site before beginning excavation. In responses to the gas leak, the police dispatcher received numerous reports of an explosion on West 3rd Street. The police department responded to the site by evacuating residents, conducting crowd and traffic control. Also, the fire department initially dispatched two engine companies Based on the later developed best practice, the city stated that the excavators should notify the pipeline operator immediately if their work damages a pipeline and to call 911 or another local emergency [1]. As can be seen from the gas line explosion below and missed mark out, there is a communication gap s between the involved parties in the excavation process. Due to the lack of transfer of the right information to the right party at the right time, the gas line damage happened. Even though the best practices proposed was in place after the accident, it did not solve the main problem of communication gap.

Figure 1: Locations of Mark out Locations

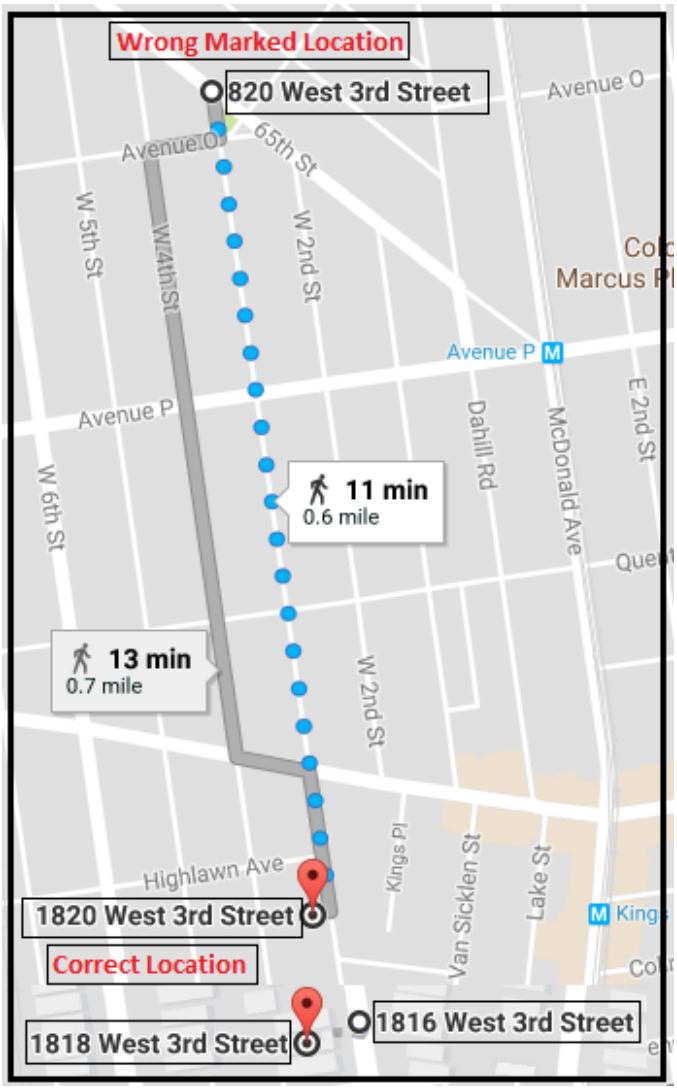

Figure 2: Locations of Mark out Locations & street #

Underground Pipelines play important role in transporting gas, water, and fuel. In addition to cooking and cleaning, the daily commute, air travel and the heating of homes and businesses are all made possible by fuels delivered through pipelines. These routine activities really add up, in terms of energy use. Natural gas and petroleum provide for 24% and 39% of our country's total energy consumption, respectively [2]. In addition, underground utilities can be hit and damaged by trucks, excavation causing the problem to underground utilities. Underground pipeline damages can be attributed to two main causes: The lack of reliable data regarding the true location of underground utilities and the lack of communicating all information. Inaccurate utility location information leads falsely instilled confidence and potentially misleads equipment operators into unintentionally utility strike, proposed ground penetration radar(GPR) to visualize and map underground utility [3]. Although this integrated system showed promise, its accuracy in locating deeply underground buried utilities is still a concern. On the other hand, many of today's underground utilities are reaching the end of their practical life and need to be replaced or repaired. Thus, precise information of underground utilities is important to utility owners, engineers, and contractors as reference for excavation [4]. Jaw & Hashim examines the accuracy of data used acquisition by scanning technique. Underground utility damage mainly occurs because of overlapping of the geospatial utility location and the movements of excavation equipment. A proposed computational detail in geometric modeling for geospatial utility data for 3D visualization and proximately monitoring to support knowledge-based excavation [5]. However, there are limitations through the different stages of the underground utility excavating cycle. It was estimated that nearly 500,000 utilities were damaged on yearly basis in the United States. The decade from 2001–2010 saw a total of 544 major excavation related accidents resulting in 37 fatalities, 152 injuries, and close to $200 million in property damage. The lack of accurate position and semantic data of buried utilities coupled with the absence of persistent visual guidance are two key problems facing excavator operators [5]. The third obstacle for safe excavation operations is the lack of real-time spatial awareness of the proximity of the digging implement to the underlying neighborhood utilities.

2. CATEGORIZE ONE CALL CENTRE PROCESS

One Call Centers serve as the clearinghouse for excavation activities that are planned close to pipelines and other underground utilities. One Call Centers help to protect underground telephone service, power lines, water and sewer pipes and energy pipelines. The processes were broken into the following four categories: One Call Center; Underground Facility Operator; Locating Company; and Excavation as shown in Figure 3 Then, the tasks broken to more details to include initiation process chart starting from excavator notifying the one-call center, the board designed one –call systems receive the request, Underground facility operator verify the location and existing utilities in the particular site, locating the pipe and mark out, and finally start excavation. In addition, this process varies between emergency cases and regular routines. Moreover, after the high level, one call center was developed, it was necessary to look for subtasks which means breaking the major steps into more details in order to map out all processes. The Operator receives and records the notice of intent to excavate provided. Then, assign a confirmation number to each notice of intent to engage in an excavation; inform the excavator or responsible contractor of the confirmation number. For each of the notices of intent, the operator maintains a register showing the name, address, and

telephone number of the excavator or responsible contractor, the site to which the notice pertains, and the assigned confirmation number. This information is promptly transmitted to the appropriate underground facility operator(s) the information received from an excavator or responsible contractor regarding intended excavation or demolition. After mark outs are made, the excavator is notified and needs to start digging within the time frame and boundaries of mark out as can be seen in Figure 4.

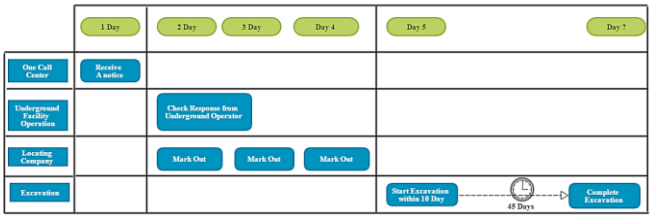

Figure 3: High-Level One Call Center Processes

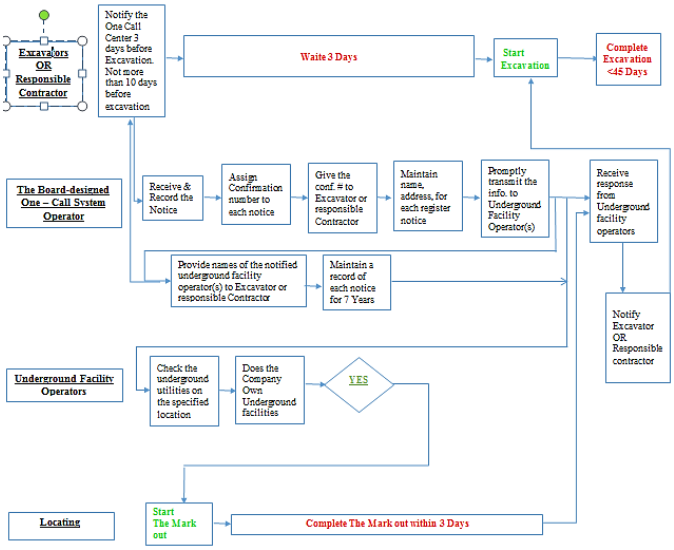

Figure 4: Shows information Flow Chart for Digging Requests

Gas Pipe Damage Request Information flow process: There are many uncertainties involved during exchanging information. This starts when the excavator requests the service from the agency, then the agency initiates the ticket request and shares information with an excavator, then the center passes the request to the Underground operator which will check which utilities are located within the parameters of the site. After determining all utilities within the area, the UG operator contact these utility companies and inform them of the request, then the utilities reply back with certain information about the Gas line location, depth, name. In the next step, the underground operator sends a person to make the mark outs, then the agency notifies the excavator that he can start excavating within a certain period depending on some condition. Finally, the excavation starts excavation. As can be seen, each and every node in these processes have the potential of miscommunication of the data which consequently will increase the probability of damage occurrence. On the other hand, the lack of coordination throughout the process and prioritizing which of the nodes are more critical makes the risk development model assess the damage very complicated with uncertain parameters involved.

3. DATA-DRIVEN APPROACHES

Data-driven analysis (DDA) is an approach to business governance that values decisions that can be backed up with verifiable data been developed through phases or stages of analysis. The success of the data-driven approach depends on the quality of the data collected and the effectiveness of its analysis and interpretation to develop a well-educated decision. In addition, data-driven analysis methods, such as independent component analysis and clustering, have been effective application in the analysis of functional magnetic resonance imaging data for identifying functionally connected risk assessment analysis. Even though independent component analysis and clustering rely on very different assumptions on the underlying distributions, both give similar results for signals with large variations. The data-driven analysis was used in many studies to explore the multivariate risk structure of the data: aiming to identify the effective components. These components may reveal structures or patterns in the data, which are difficult to identify apriority: such as unexpected activation and connection, motion-related artefacts, and drifts [6]. These data-driven analysis methods provide generalizations of connectivity analysis in situations where reference seed regions are unknown or difficult to identify reliably. One important motivation and expectation behind the use of these methods is that in many data sets, data points lie in some manifold of much lower dimensionality than that of the original data space [7]. The four most popular methods are the following: clustering; principal component analysis; independent component analysis; and probabilistic principal component analysis. Many underground gas line researchers used principal component analysis as a statistical technique to linearly transform an original set of variables into a substantially smaller set of uncorrelated variables. It is also known as the Karhunen-Loeve transform [8]. One of the main goals is to reduce the dimensionality of the original data set. In addition, a group of uncorrelated variables data is assumed to represent the underlying sources for observations and is more computationally efficient in further analysis than a larger set of correlated variables. Therefore, the principal component analysis method is often used as a pre-processing step for other data-driven analysis methods such as clustering. We are using a Gaussian latent variable model in developing a more precise probabilistic formulation of PCA. This probabilistic formulation of PCA provides a way to find a low-dimensional risk representation of higher dimensional data with a well-defined probability distribution and enables comparison to other generative models within a density estimation framework [8]. Moreover, there are some advantages of the probabilistic principal component. First, this probability model can be used to provide samples from the distribution. Second, it gives an explicit probability model of the data, in the density estimate framework: which allows us to calculate the likelihood of any observation and to compare the result of the probabilistic principal component to other exploratory data analysis methods as mentioned above. Clustering or data segmentation is another method that could be used. Clustering groups, a collection of data points into subsets

such that the points in each subset are more closely related to each other than those in other subsets, where each cluster itself is as different as possible from other clusters . In many real data cases where multiple clusters are present, a simple probability distribution is insufficient to capture the structure of the data. A linear combination of more basic distributions, known as mixture distribution, gives a better characterization by providing a framework upon which to build a more complex, richer class of density models. A comprehensive methodology that supports the entire process of determining information requirements for data warehouse users, matching information requirements with actual information supply, evaluating and homogenizing resulting information requirements, establishing priorities for unsatisfied information requirements, and formally specifying the results as a basis for subsequent phases of the data [9]. The experts' requirements for information requirements analysis in a data warehousing context call for a demand-driven approach. Since the business process-oriented approach is not applicable if the data warehouse system has to support decision processes, we focus on a 'conventional' demand-driven approach. The proposed methodology should overcome the shortcomings listed, i.e. a multi-stage approach has to be taken, users have to be supported in specifying objective (and not subjective) (Winter Author)

4. FRAMEWORK

This chapter provides an outline of the research methods, and methodology used in the study Figure 4. This chapter discusses in detail the research methodology that has been adopted in this study. This methodology involves ten steps. It starts with defining the system and collecting the data set which includes organizing the collected data in categorized attributes and compiling them in tables. In addition, this step includes exploration of data, and data set processing. Second, identifying the risk process by using the Bow-tie method. Third, Mapping out underground gas pipe damage network by using the Fault tree method, then involving machine learning algorithms such as Logistic Regression, Support vector machine, Random Forest, and KNN. Finally, processing UG gas pipe damage network with Bayesian network and defining the probability of the node. The tasks involved in the methodology are delineated in Figure 5. Defining the system and collecting data: the data used in this research was both for damaged and undamaged UG gas pipes. Most of the undamaged data contain attributes such as ticket number, date of the incident, and address of the incident. In the data structure, damaged UG gas pipe has attributes such as excavator type and type of request. The data, organized into years 2010-2014, were categorized by ticket number and address to be used later. The risk evolution process of underground gas pipe damage modelling with Bow-tie: This is a widely used graphical process for damage modelling. The Bow-tie process can present a complete accidental scenario starting from the causes and ending with the consequence. The Bow-tie method was selected for UG gas pipe risk assessment because it can identify where resources should be focused for risk reduction, i.e. prevention or mitigation. Bow-tie includes two parts, the left of bow-tie is an FT that describes the latent causes for an initial event, the right of bow-tie is an ET which describes the sequential failure of damage preventive barrier and presents the evolution process from initial event to final latent consequence. The FT and ET are linked through a pivot node that is the top event of FT and the induced event of ET. Building Underground Excavation Database: Excel and Python were used in this research to combine all the attributes of the UG gas pipe damages in columns, and rows. This step was done for many reasons. First, duplicate data can be flagged out and easily removed to get better results. Second, understating the data trend: By organizing the data in one database, the trend of the data attributes can be easily specified by performing preliminary analysis which can be used in the future steps of the research. Third, cleaning the data: the data were received as text files which were not useful. Thus by transferring the data into an excel spreadsheet, it was possible to clean the data and prepare it to be useful for the research. Exploratory and Spatial Analysis: We will use Exploratory and Spatial analysis to extract the latent risk factors. More specifically, clustering analysis will be used. We will group the data into groups containing similar attributes such as depth, pipe size, year, zip code, number of outgoing calls…etc. The goal is to observe the characteristics of each cluster and to focus on a particular set of clusters for further analysis Figure 6. The rapid miner will be used to further study different clusters patterns of the damaged data. In addition, Hot Spot analysis will be performed to identify the damaged hotspots. We will be using Arc GIS to plot the data and to generate hot spots which will enable us to focus on the most affected areas Prepare the Data of The UG Gas pipe: Several steps were followed to better prepare the data to be used in the Predictive model and risk factors identified. Since a large portion of the data set contained missing values, it was important to focus on correcting these instances. In addition, during data testing, it was evident that some gas distributors kept extra records. Often there were variables that were recorded by a single company Therefore, the damaged data was missing some attributes, the undamaged data was missing most of the attributes. The first data preprocessing step eliminated attributes that were deemed unnecessary for further computations. Eliminating attributes cleaned the data set up and assisted with the removal of a large number of missing values. Furthermore, eliminating attributes actively reduced the dimensionality of the data set and allowed for increased model performance with a reduced computation time. Preparing the data will be explained in more detail in preparing the data, and performing exploratory analysis. That will cover all the steps of data preprocessing including, incomplete data, Geo-coding the data, missing data, organizing the data, converting the text data into columns, rows in excel, data integration, cross-validation of the UG gas pipe data. The steps also include preparing the data maps, plotting the data, preparing the data attributes for damaged data, preparing the data attributes for undamaged data and deriving new attributes from the data Figure 6. Training The Data Set (80 %): The data were split into 80 % training and 20 % testing. The training data set was chosen to be large enough to yield meaningful results, and is representative of the data set as a whole including all selected data features or attributes. As part of this step, data preprocessing was conducted to make sure we have quality data and meaningful attributes to yield meaningful results. In

addition, Python Anaconda was selected to be used in the analysis of the data. Python is a popular free source tool with many add-on libraries. Jupyter Notebook was used as a tool as well. First, the operating system was imported to be used in the analysis. Python contains many Libraries, Library name called Panda is needed to perform the analysis. In order to use Pandas in Jupyter Notebook (Jupyter come with Anaconda by default). Importing a library means loading it into the memory and then it's there ready for use. In order to import Pandas, we ran import code. Then the data was saved in a CSV file, given the name " Log" in coding. We ran a code to import the data. Next, the data was displayed in Python to view all columns to make sure all the required attributes of the data were imported, and no missing columns. Selecting training Algorithm Model: Four algorithms were selected to be used in the training model, Logistic Regression Method, k-nearest neighbors Method, Support vector machine Method, and Random Forest Method. Logistic Regression analysis was selected because it indicates the strength of the impact of multiple independent variables (data attributes) on a dependent variable (Damage/ Not). In addition, Logistic Regression indicates the significant relationships between the dependent variable (Target (YES/NO)) and (Predictors (data Attributes)) independent variable. k-nearest neighbor's algorithm was selected because K nearest neighbors stores all available cases (Predictors) and classifies new cases based on a similarity measure (e.g., distance functions). In addition, KNN works as pattern recognition through specific distances for each neighborhood, then it classifies the data based on the distances between the generated neighborhoods. Random Forest Method was selected because it builds multiple decision trees for the data attributes Time, Weekdays, Months, and other attributes and merges them together to get a more accurate and stable prediction. More specifically, it creates a forest of data attributes and makes a random selection. In addition, Random Forest was used because it will be used for both regression and classification tasks. The random forest also makes it easy to view the relative importance it assigns to the input Attribute. Support Vector Machine was used because it classifies data: based on either a priori information or statistical data mined from raw data set.S. V Machine provides pattern recognition and makes it an extremely powerful tool for UG gas pipe damage data separation. Run the training Model: The process of training a predictive model involves providing a PM algorithm (that is, the learning algorithm) with training data. The term PM model refers to the model artefact that is created by the training process. The provided training data to the algorithm contains the correct answers: known as a target or target attribute. The learning algorithm finds patterns in the training data that map the input data attributes to the target (the answer that we want to predict): it outputs a PM model that captures these patterns. We will use the model to get predictions on new data for which the model does not know the target. Testing, and validation (20 %): 20 % of the data was used in testing the predictive model. The metrics used to evaluate the model were Confusion Matrix, Recall, and Precision. These three Metrics were conducted to test and validate all of the outcomes of the algorithms including Logistic Regression, Random Forest, KNN, and Bayesian. Results, Logistic Regression Model gave the best result according to Confusion Matrix, Recall, precision. The sample was used for illustration purposes. A confusion matrix is a table that describes the performance of the classification models we employed in our study (Bayesian, Logistic Regression, KNN, and random Forest) on a set of test UG gas pipe data for which the true values. The confusion matrix itself is relatively simple to understand, but the related terminology can be confusing. Therefore, these are the definitions regarding confusion matrix, true positives (TP): These are cases in which we predicted yes (they have the UG gas pipe damage), and they do have the damage. true negatives (TN): means the model predicted no, and they don't have the gas pipe damage. false positives (FP): means the model predicted yes, but they don't actually have the damage. False negatives (FN): means the model predicted no, but they actually do have the damage. Accuracy: We used accuracy measure in our study so it tells us how often is the classifier correct? Equation = (TP+TN)/total. Precession: We used accuracy measure in our study so it tells us when it predicts yes, how often is it correct? Equation =TP/predicted yes. Selecting The Model: The Logistic Regression was selected as a model to predict the future gas pipe damages and the dominant risk factors contributing to the damage of the UG gas pipe damages. The confusion matrix was produced for all algorithms, including, Logistic Regression, S.V Machine, KNN, and Random Forest. The Logistic Regression demonstrated a precision level equal to 0.89 and accuracy equal to 0.98: which is the highest as compared to the other algorithms.

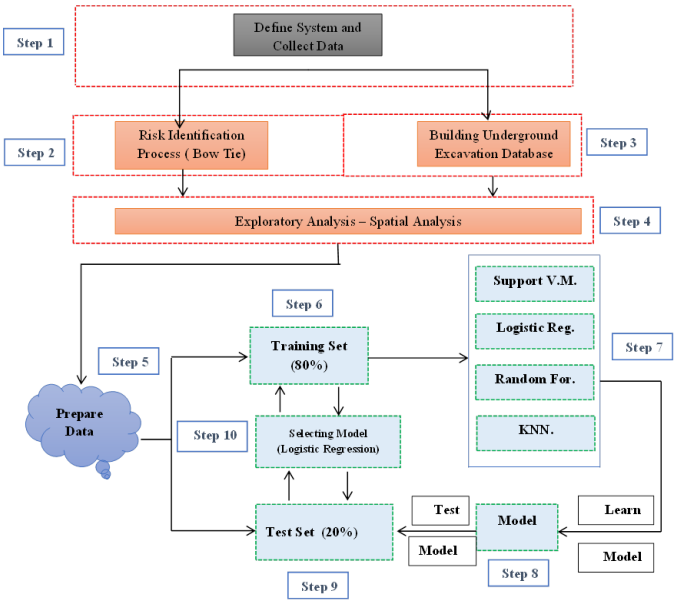

Figure 5: Overview of Components of Methodology in Predicting Damages in UG Gas pipes

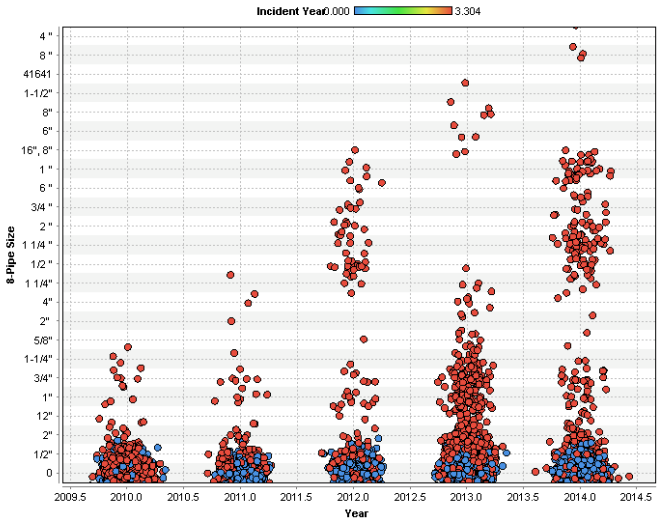

Figure 6:Cause of Damage between Pipe Size & Year (Phase 2)

5. PREDICTIVE ANALYSIS MODEL

Now as we understand the machine-learning problem want to solve for: predicting gas pipe damage, and dominant risk factors for future UG gas pipe operations. The next step is to build a model which is to employ data science methodologies like Logistic regression, Random Forest, KNN, Bayesian. Looking at the historical data we have, we want to produce a model that estimates a particular variable specific. Which is (YES/NO) damages or undamaged. The following steps explain the preliminary steps to input data into Python (Anaconda). The total number of records used in the model were 396,547 (including undamaged & damaged): see figure 7.

```
In [30]: df.shape
Out[30]: (396547, 40)
```

Figure 7: Shows the total number of data (Anaconda)

The total inputted data attributes into the model were equal to 40 attributes including (Time, am, pm, days, Mon, Tue, Wed, Thurs, Fri, Sat, Sun, Month, Jan, Feb, March, May, Jun, Jul, Aug, Sept, Oct, Nov, Dec, the Year 2010-2014, Season, Winter, Summer, Autumn, Spring, damages within 10 miles, 20 miles, and 30 miles. Then it was developed in the model into 752 attributes, cities were put in the columns and values (1, or 0) was assigned based on the ticket, damaged or not. (Zero) value replaced the undamaged tickets, and (1) value replaced the damaged tickets. This step was performed to transfer the data into numerical which make it useful for the algorithm to process it, instead of having text data. Furthermore, data cleaning was performed in Excel, and Python (Anaconda) to make sure the data was not having any missing values, repeated values, or corrupted numbers. Next, the data split was 80% Training, and 20% Testing see figure 8.

```
Xtr, Xtest, ytr, ytest = train_test_split(data, y, test_size=0.20, random_state=5)
```

Figure 8: Shows Data Split between Training, and Testing

Which means; Data in Training = 0.8 *396,547 = 317,237

Data in Testing = 0.2* 396,547= 79,309

Microanalysis was selected from (Micro, Macro, and Binary). The following chart explains the methodology of the Model sees figure 9. The process starts by preparing the data, then inputting data into Anaconda, then cleaning the data again, then splitting the data, then running the algorithms in 80 % training. Then we run the remaining 20% of testing data into testing and test the model. The following step is to select a model based on the testing metrics such as confusion matrix, precession, recall. then run the confusion matrix, precession, recall.

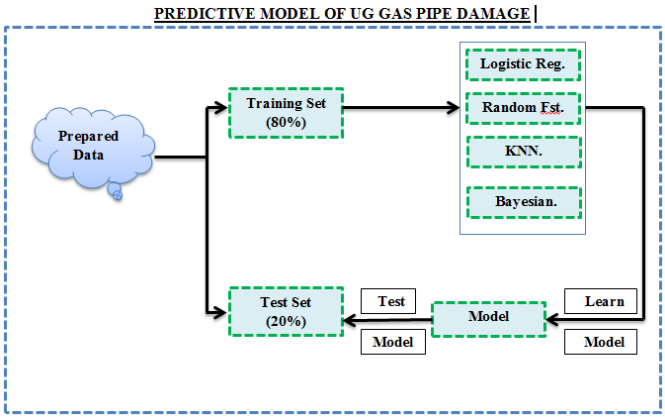

Figure 9: Predictive Model for Underground Gas Pipe Damages

6. SUMMARY AND RECOMMENDATION

The significant number of damages of underground gas pipelines and their consequences have motivated many researchers to study UG pipeline damage. A comprehensive study of these research efforts revealed the lack of a comprehensive predictive model for estimating damages. Some studies focused on corrosion or third-party failures and could not assess the effects related to information flow on damages. Most studies that consider multiple damages focus on the qualitative investigation or develop physical models that are very expensive to implement. Qualitative models rely on a survey which makes it difficult (if not impossible) to obtain due to the location of most UG pipelines. In addition to that, is the high cost of needed operators, the shortage of complete historical data on the gas pipe has been a challenge for all researchers. This research addresses the lack of effective predictive models by developing a comprehensive risk model: for the processes involved in the information flow process starting from receiving digging requests, to the completion of the excavation of gas pipelines. The model developed in this research provides an overall image throughout the digging processes of gas pipelines. A Bayesian risk model was developed to minimize pipeline damage rates by assessing and ranking the risk of various

sections of natural gas pipelines. It would explore the interaction among significant factors. Inputs are required for an effective evaluation of the risk encountered in the exchange of information between different parties involved. In addition, the past data were used to develop a risk model to study the future risk associated with the excavation requests and risk factors. This study also provides a research base by using Logistic Regression to develop a risk model to investigate the interactive effects of various factors causing underground gas line damage. In addition, regression analysis help identifies the important factors that can be used by digging agencies when they receive digging requests to minimize the possibility of gas pipe being damaged. Basically, the responsible agency can cross-reference the request attribute with the developed risk factors; then they can make an educated decision on where to be cautious and pay more attention to the digging processes in the potential areas. The risk involved in the UG gas pipe is a critical aid in the decision-making process of the underground gas pipe system. The predictive model will be useful in assisting the operators of such facilities in the maintenance and inspection planning. The model can rank the selected tools based on their probabilities of happening. This research develops a framework for the development of risk assessment models completely based on the historical damage of UG gas pipe data. The methodology is applied to the infrastructure of oil and gas pipelines. However, it can be expanded to be used in other infrastructure types. The main value of such predictive models is that they reduce the cost of damage prediction with or without an abundance of valuable data. These models assess the probability of risk of different infrastructure types and plan accordingly for the life cycle of such infrastructures.

7. ACKNOWLEDGMENT

The years at Rutgers as a Ph.D. student have been an amazing and thrilling part of my life. I have had the privilege and pleasure of being supervised by Dr. Jie Gong. I am deeply grateful for his continuous guidance, support, and encouragement throughout my entire graduate experience at Rutgers. I learned a lot from his professionalism. It is really a great honor for me to have him as my mentor and a friend.I would like to thank my friends and colleagues starting with Mr. Aravind Reddy Pasham who provided very valuable help for this dissertation. I would also like to thank Mr. Sher Khan, Mr. Kall Dalco for their support throughout my studies. I have felt so blessed to be surrounded by such great friends who made this journey enjoyable. I would like to extend my appreciation to all my colleagues who provided insightful; discussions, and scientific suggestions to help improve my dissertation. Mostly, I would like to thank my loving parents and my family members for their immense love, support, and confidence in me

8. REFERENCES

[1]. Carmody, Carol J., et al. "National Transportation Safety Board." Judiciary Committee on Primary Seatbelt Law Enforcement (2003).

[2]. Feyer, Anne-Marie, and Ann M. Williamson. "Human factors in accident modelling." Encyclopaedia of occupational health and safety. 4th edn. International Labour Organization (1998).

[3]. Li, Shuai, Hubo Cai, and Vineet R. Kamat. "Uncertainty-aware geospatial system for mapping and visualizing underground utilities." Automation in Construction 53 (2015): 105-119.

[4]. Jaw, Siow Wei, and Mazlan Hashim. "Locational accuracy of underground utility mapping using ground penetrating radar." Tunnelling and Underground Space Technology 35 (2013): 20-29.

[5]. Talmaki, Sanat, Vineet R. Kamat, and Hubo Cai. "Geometric modeling of geospatial data for visualization-assisted excavation." Advanced Engineering Informatics 27.2 (2013): 283-298.

[6]. Biswal, Bharat, et al. "Functional connectivity in the motor cortex of resting human brain using echo-planar MRI." Magnetic resonance in medicine 34.4 (1995): 537-541.

[7]. Svensén, Markus, and Christopher M. Bishop. "Pattern recognition and machine learning." (2007).

[8]. Ringnér, Markus. "What is principal component analysis?." Nature biotechnology 26.3 (2008): 303-304.

[9]. Winter, Robert, and Bernhard Strauch. "A method for demand-driven information requirements analysis in data warehousing projects." 36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the. IEEE, 2003.