DEVELOPMENT OF A MODEL TO PRIORITIZE INSPECTION AND CONDITION ASSESSMENT OF GRAVITY SANITARY SEWER SYSTEMS

By

KARTHIKEYAN LOGANATHAN

Presented to the Faculty of the Graduate School of
The University of Texas at Arlington in Partial
Fulfillment of Requirements for
the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON
August 2021

Dedication

I dedicate this thesis to my parents, Loganathan and Subbulakshmi, and my beloved wife, Sharmila, and all my friends for their wish, love, and support.

Acknowledgements

I would like to express my most sincere gratitude and appreciation to my academic advisor and mentor Dr. Mohammad Najafi, P.E., F. ASCE, Associate Professor of Civil Engineering and Director of the Center for Underground Infrastructure Research and Education (CUIRE). It was a pleasure working under the supervision of Dr. Najafi. I feel honored assisting him in several courses and research projects over the last three years at the University of Texas at Arlington. He was always there to help me when I was in need, and I am indebted for the opportunities he provided me.

I should express my heartfelt thanks to Dr. Ardeshir Anjomani, Dr. Sharareh Kermanshachi, and Dr. Melanie Sattler, for their continued support and guidance as my dissertation committee members. Their valuable comments and suggestions made a great contribution to this dissertation.

I would like to extend my sincere thanks to Ms. Sonnie LaForce, Strategic Operations Supervisor, at the City of Fort Worth, Texas, for her continued commitment in providing all the data and information required for this research study. She was always there to listen amidst her busy schedule and instantly help without any hesitation; and I am indebted to her affection and care beyond her responsibilities. I would like to thank Mr. Mark Shell, Technical Services Coordinator, City of Fort Worth Wastewater Department for his amazing presentation at CUIRE and for his valuable interview about field operations with respect to condition assessment of sanitary sewer pipelines.

I would like to extend my special thanks to Dr. Praveen Kumar Madhuri for his keen interest towards my research study and endless support throughout the model development phases. Without his training and help, this study wouldn't have been accomplished by me.

It would be ungrateful if I forget to thank my parents, mother-in-law, and especially my wife, Sharmila. She has been and will always be the most important source of support and motivation through her endless love.

August 19, 2021

Abstract

DEVELOPMENT OF A MODEL TO PRIORITIZE INSPECTION AND CONDITION ASSESSMENT OF GRAVITY SANITARY SEWER SYSTEMS

Karthikeyan Loganathan, Ph.D.

The University of Texas at Arlington, 2021

Supervising Professor: Dr. Mohammad Najafi

Underground utilities and wastewater collection systems deteriorate over time demanding utility owners to involve in continuous revisions and development of their asset management frameworks to maintain the functionality of their assets. In any asset management framework, inspection of an asset and respective condition assessment plays a vital role in successful operation and maintenance of systems. In the United States, closed-circuit television (CCTV) is the commonly used device for inspecting the inner environment of sewer pipes, which considering the large length of pipe inventory in a city, is a relatively expensive and time-consuming process. Therefore, inspection of every individual sanitary sewer pipe segment is not feasible in a short time period for any municipality owing to their large inventory of these pipes. However, sanitary sewer pipe segments in need of repair or a maintenance activity can be prioritized in advance for inspection based on their historical performance. Therefore, the primary objective of this dissertation is to develop a sanitary sewer pipe condition prediction model. Data collected from City of Fort Worth, Texas, is utilized in model development. Various supervised machine learning algorithms such as logistic regression (LR), k-nearest neighbors (k-NN) and random forests (RF) are employed. Numerous evaluation metrics such as precision, recall, F1-score and area under curve (AUC) are estimated to compare the performance of developed models. Resulted F1-score for the RF model is 0.94 while LR and k-NN models resulted 0.83 and 0.44, respectively. The results show that random forests model performed better than both LR and k-NN models. As a secondary objective of this dissertation, a decision support tool was developed for asset managers to utilize above models during inspection phase to estimate condition of their sanitary sewers for identification of critical sewers in need of immediate attention.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION AND BACKGROUND

1.1 Introduction

Underground pipeline systems in the U.S. span thousands of miles contributing a significant portion of the wastewater infrastructure assets (Najafi and Gokhale 2005). Since majority of wastewater infrastructure systems in the U.S. are more than 100 years old, any catastrophic failure to these wastewater systems could drastically disrupt the surrounding areas economically, socially and environmentally (EPA 2004). In addition to the effects on public health, emergency repair of failed sanitary pipes can cost an enormous amount to the municipality. Considering these social, economic, and environmental impacts, wastewater system must be protected from failure.

Unlike reactive maintenance practices carried out by some municipalities after failure of pipes, proactive maintenance must be accomplished that include inspection and maintenance activity in advance of failure or complete deterioration (Fenner 2000). Table 1-1 lists the common factors that could influence deterioration in sanitary sewer pipes. Generally, pipeline deterioration mechanism could be classified under any of the following:

1) Structural – cracks, fractures, breaks and so on
2) Hydraulic – flooding, encrustation, and grease
3) Corrosion – chemical and external corrosion

4) Erosion

5) Operational problems – roots, blockages, debris and so on

Table 1-1 Factors Known to Influence Sanitary Sewer Deterioration
(Davies et al. 2001)

Construction Factors	**Local External Factors**	**Other Factors**
Installation method Standard of workmanship Sewer size Sewer depth Sewer pipe material Bedding material and type Joint type and material Pipe section length Connections	Surface use Surface loading (including construction traffic) Surface type Traffic characteristics Water main bursts/leakage Ground movement Maintenance of other buried services Groundwater level Infiltration/exfiltration Soil/backfill type	Sewage characteristics Use of appropriate maintenance methods Asset age Sediment level Surcharge

Underground utilities and wastewater collection systems deteriorate over time demanding utility owners to involve in continuous revisions and development of asset management frameworks to maintain the functionality of their assets (Najafi and Kulandaivel 2005). A study by the Environmental Protection Agency (EPA) has shown that up to half of the buried assets in studied systems might be beyond midpoint of their service lives (EPA 1999). Most of the municipal sewer systems in the United States are at least 60 years old and many communities have sewers that are older than 100 years (EPA 2015). In addition, this study stated that “among public agencies in the U.S., infrastructure asset management is used most extensively in the transportation sector." It has been found that nearly several hundred operational wastewater agencies in the United States did not develop or implement an asset management program (IIMM 2006). Therefore, the importance of updating asset management policies and programs can be emphasized.

"Asset management is defined as managing infrastructure capital assets to minimize the total cost of owning and operating them, while delivering the service levels customers desire (EPA 2002)." Basic requirements for an asset management system are: (1) to maintain an inventory of assets; (2) to assess the condition of assets; and (3) to provide an estimate of required budget to maintain an asset in a serviceable condition (Daziel and Macey 2004). Inspection and condition assessment of an asset are the preliminary tasks in an asset management program (Tscheikner-Gratl et al. 2020).

Condition assessment of sanitary sewer pipes are accomplished by various methods. In the U.S., sanitary sewers are commonly assessed using the Pipeline Assessment Certification Program (PACP), first developed by National Association of Sewer Service Companies (NASSCO) in 2002. However, it should be noted that not all municipalities follow the same program for inspecting condition of their sanitary sewer pipes (NASSCO 2018). The PACP defines each possible defect with a unique code as well as a score towards the structural integrity of pipe segment on a scale of 1 – 5, while 1 and 5 refer to good and poor conditions, respectively.

Historically, design, construction, maintenance, and operation of sanitary sewer systems were addressed by municipalities (Wirahadikusumah et al. 2001). Generally, closed-circuit television (CCTV) is employed to inspect the inner environment of pipes to capture their defects and failures. A certified operator by NASSCO observes the recorded video and manually appends defect codes to a spreadsheet or a computer program. The program is predefined with scores for each type of defect and their severity. Based on operator's judgement and coded defects, a final PACP score will be generated for an inspected sewer pipe segment (NASSCO 2018). However, inspecting a sewer pipe is an expensive and time-consuming process since the PACP suggests a maximum camera speed of 30 feet or 9 meters per minute. Therefore, inspection of each sanitary sewer pipe is not feasible for any municipality or utility owner owing to their large inventory of assets (Malek Mohammadi 2019).

In addition, there are possibilities that sewer pipes in structurally good condition could be inspected by municipalities, which otherwise this considerable budget could be used for sewer pipe segments in need of repair and renewal. Efficient budget allocation could be achieved by predicting the condition of pipes based on historical performance of the same. Therefore, sanitary sewer pipes in need of repair or a maintenance activity must be prioritized in advance for inspection by predicting the future condition of sewer pipes.

1.2 Problem Statement

A wide variety of researches have been accomplished to develop prediction models for future condition of sanitary sewer pipes (Malek Mohammadi et al. 2020, Salman and Salem 2012, Kienow and Kienow 2004, Najafi and Kulandaivel 2005, and Ariaratnam et al. 2001). Developed models employ traditional statistical methods as well as advanced machine learning techniques and artificial intelligent algorithms. However, there is no single standard model that could be employed by all municipalities owing to the differences in collected historical data about their sewer pipes (McDonald and Zhao 2001). Assessing the structural condition of a sanitary sewer pipe can be achieved by identifying the critical factors influencing deterioration of sewer pipes and by developing a prediction model based on identified factors. It should be noted that the factors are not examined for their causes on failures rather for their correlations with structural condition of sewer pipes. Therefore, for a condition prediction model to be employed by a municipality or a utility owner, the model must be able to predict the sanitary sewer pipe condition based on data collected by respective municipality or utility owner.

Based on literature studies focusing on condition prediction models, following limitations and recommendations are found:

- Malek Mohammadi (2019) recommended that a prediction model must be able to predict all five condition levels individually rather than transforming to binary classes.

- Laakso et al. (2018) suggested that potentially influential variables could be further investigated.
- Vladeanu (2018) developed a model using Markov chain but inadequate data limited the validation of developed model.
- Sousa et al. (2014) concluded that machine learning and artificial intelligent models were reliable over logistic regression models and further investigation could improve the accuracy of results.
- Opila (2011) developed a multi-dimensional linear model to predict the condition of sewer pipes and recommended that the model could be improved potentially.

Studied relevant research studies and models emphasize the knowledge gap in identifying critical factors on deterioration of sewer pipes and the need for machine learning and artificial intelligence algorithms in condition prediction model development. It was found that majority of the studies were based on binary classification. In addition, it was also found that majority of the studies trained their model with a limited number of material types, which prevents the model from predicting any critical pipe of excluded material type. It would be beneficial for an agency or utility owners when a developed a model can incorporates all material types from their inventory.

Developed models in literature studies were compared based on a single evaluation metric such as accuracy or area under the curve value. This dissertation evaluates the developed models with different metrics rather than a single metric to effectively validate the performance of prediction models. Bridging the identified knowledge gap plays a key role as objectives for this study.

1.3 Objectives

The primary objectives of this research study are:

- To employ machine learning and artificial intelligent algorithms to develop a condition prediction model to predict the condition of sanitary sewers.
- To compare the results and accuracy levels of developed models; and recommend a suitable model for application.

The secondary objective of this research study is:

- To develop a decision-support tool for the asset managers and operators to effectively utilize the developed model.

1.4 Scope of Work

Table 1-2 illustrates the scope of this research study to include gravity flow sanitary sewers excluding force main sewer pipelines. To be consistent, sewer pipelines without any prior rehabilitation are considered for analysis and pipelines with a history of maintenance activity is not included. With respect to the pipe material, all 9 different types of materials, described in future sections, are included for further analysis.

Table 1-2 Scope of Work

Description	Included	Excluded
1. Pipe Functionality	Sanitary sewer pipes	-
2. Flow Type	Gravity flow pipes	Force main pipes
3. Rehabilitation	Pipes in installed condition	Pipes with any existing rehabilitation
4. Inspection Type	Pipes inspected based on PACP manual	Pipes inspected without PACP manual

1.5 Methodology

Various machine learning and artificial intelligent algorithms are employed to develop a condition prediction model, which could predict the structural condition of sanitary sewers pipelines. Numerous factors recorded during prior inspections are used to develop the models. Figure 1-1 illustrates various steps involved in the model development.

The methodology of this study starts by defining a problem statement followed by objectives and the scope for this study. A comprehensive relevant literature review is accomplished to effectively utilize the previously conducted research studies. In the next step, data required for further analysis is collected from City of Fort Worth, Texas, is processed. The processed data is then used as input for various supervised learning algorithms to train the models. As a final step, the performances of various trained models is compared based on different evaluation metrics and a reliable model for condition prediction is selected.

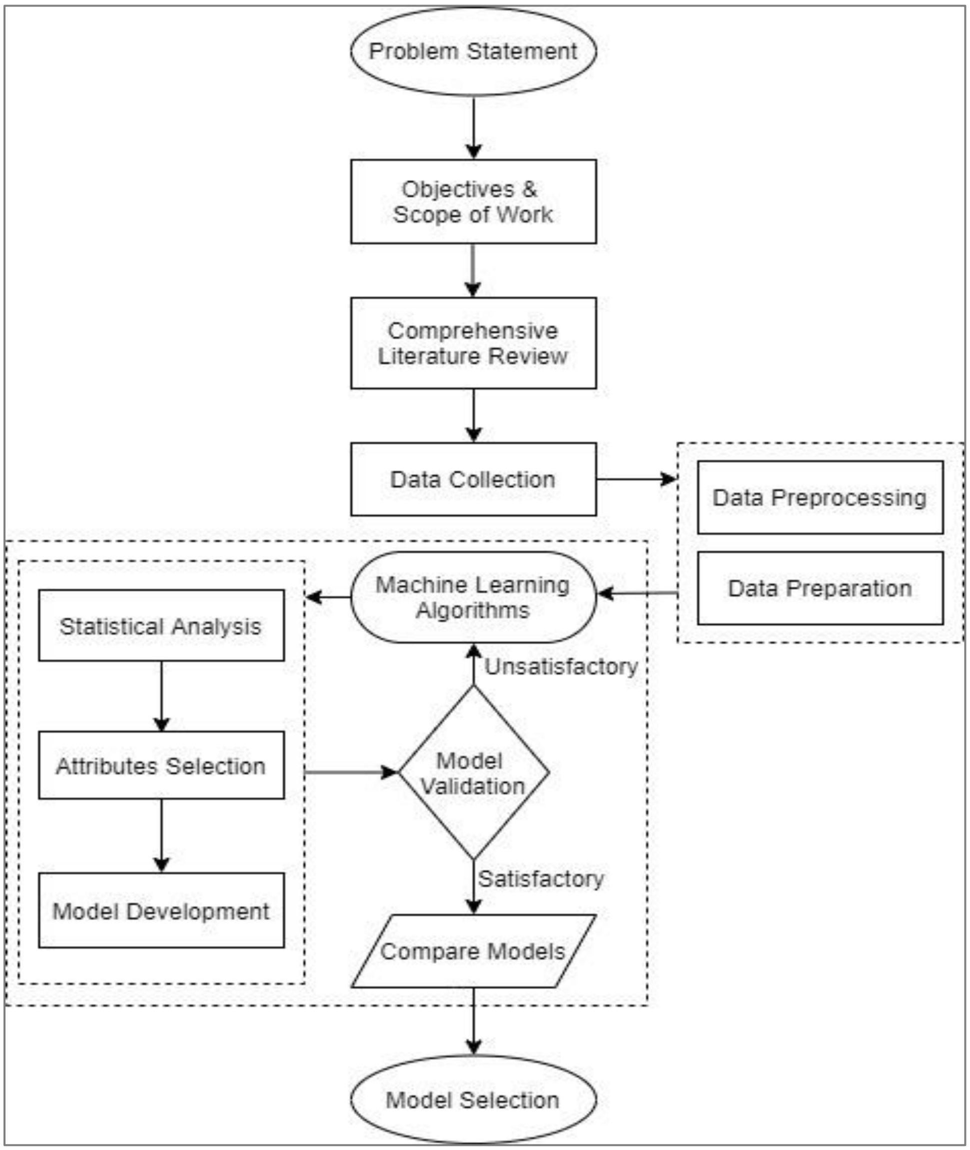

Figure 1-1 Research Methodology

1.6 Hypothesis

Factors recorded by the wastewater department such as the age, length, slope, diameter, and material of pipe, location reference and drainage basin of a pipe, could forecast the condition of sanitary sewer pipes.

1.7 Chapter Summary

Primarily, this chapter briefly discussed asset management components and current practices involved in condition assessment of sanitary sewer pipes in the United States. The knowledge gap in identifying critical factors on deterioration of sewer pipes and the need for machine learning and artificial intelligence algorithms in condition prediction is listed. Secondly, this chapter explained the research needs, objectives, scope of work followed by methodology.

CHAPTER 2

LITERATURE REVIEW

2.1 Background and Overview

The ASCE 2021 infrastructure report card ranked an overall grade of D for wastewater infrastructure utilities, which means that the system is at poor condition. In the US, public sewage pipes span over 800,000 miles and lateral sewers span around 500,000 miles, contributing a significant portion of the underground utilities and infrastructure (ASCE 2017). The report also mentioned that in the future twenty years, 56 million new users will be connected to centralized treatment systems, which would require a significant budget to satisfy current and future demands. Combined investment needs for water and wastewater systems are estimated to be \$150 billion during 2016 – 2025, and a \$105 billion investment gap is found between estimated funds and required funds (ASCE 2017). Identified investment gap emphasizes the cruciality for efficient use of budget.

The Congressional Budget Office (CBO) had compared the public spending on transportation and water infrastructure during 1956 – 2017 and found that the spending accomplished for water utilities are much lesser than that for transportation (CBO 2018). EPA had stated that most of the municipal sewer systems in the United States are at least 60 years old and many communities have sewers that are older than 100 years (EPA 2015). In addition, the report added that old and aging sewers cause at least 23,000 to 75,000 sanitary sewer overflows per year. Therefore, effective maintenance and rehabilitation strategies must be followed to maintain the functionality of sewer systems. Interestingly, it was found that several hundred operational wastewater agencies in the United States did not develop or implement an asset management program (IIMM 2006). Therefore, the municipalities and utility owners are required to revise and update their asset management methods and programs.

2.2 Asset Management

"Asset management is defined as managing infrastructure capital assets to minimize the total cost of owning and operating them, while delivering the service levels customers desire (EPA 2002)." According to Environmental Finance Center (EFC), "Asset management is an approach to manage the assets of a system that can assist the utilities with making better decisions on managing the aging assets (EFC 2006)."

New York's Department of Environmental Conservation states that "Municipal sewage system asset management (MSSAM) is the practice of managing a municipal sewage treatment plant and the associated sewage collection system's capital assets in a way that protect the public health and the environment while also minimizes the total cost of owning and operating those assets while delivering the desired levels of service (MSSAM Guide 2015)." The Orange Water and Sewer Authority's (OWASA) comprehensive asset management program was utilized to assess and prioritize infrastructure improvements needed to achieve desired customer and environmental service level (OWASA 2017).

There are various definitions given to asset management by researchers and government agencies. However, most of the asset management programs include elements such as inventory of assets, prioritization of critical assets, and financial planning to maintain their performance. The importance of asset management will increase with aging sewer systems. Compared to data-intensive disciplines such as bioinformatics and medical sciences, the value of data collected and stored for urban drainage system has not yet been fully satisfied (Tscheikner-Gratl et al. 2020). In addition, the sewer asset data handled by an agency must be stored and manipulated in such a way that it must be easily usable for operators and decision-makers as well .

2.3 Deterioration of Sewer Pipes

Utility and pipeline systems form one of the most capital-intensive infrastructure systems, especially the sanitary sewer systems, owing to their direct and indirect effect on their surroundings environmentally as well as financially (Najafi and Gokhale 2005). A study by Davies et al. (2001) claimed that some of the basic performance requirements for sewer operation are

a) Pipeline network cannot have blockages

b) Drains and sewers must be watertight to avoid leakages

c) Sewers shall not endanger existing adjacent structures

Deterioration of pipelines could be a result of different factors varying from structural loss of the pipe material to deterioration caused by the material transported by the utility system. Continued deterioration in these systems could result in failure of the pipes, which is termed as “collapse”. Deterioration in a sewer pipe could be caused by many reasons and hence, estimating the rate of deterioration is a difficult task. In addition, deterioration can be influenced by random events during the service life of a sewer pipe as well. Therefore, Water Research Center (WRc 1986) concluded that estimating the rate of deterioration is unrealistic.

2.4 Factors Influencing Deterioration

Deterioration of underground sewer pipes is a most complicated process since various pipe characteristics could play a vital role in the process (Yan and Vairavamoorthy 2003). Although estimation of the rate of deterioration for a sewer pipe is unrealistic, it is well known that any sewer pipe failure or collapse would follow any or combination of the following mechanism(s): 1) Structural, 2) Hydraulic, 3) Corrosion, 4) Erosion, and 5) Operational problems (Najafi and Gokhale 2005).

Detailed list of factors influencing the deterioration of sanitary sewer pipeline can be found in Table 2-1. These factors could be considered of primary importance during inspection processes, irrespective of the inspection technique (Malek Mohammadi et al. 2020).

Table 2-1 Factors Affecting Sewer Pipe Deterioration
(adapted from Malek Mohammadi et al. 2020)

Physical Factors	**Environmental Factors**	**Operational Factors**
End invert elevation	Backfill type Blockages	Blockages
Installation method	Bedding material	Burst history
Joint type	Ground movement	Debris
Pipe length	Groundwater level	Flow velocity
Pipe shape	pH	Hydraulic condition
Pipe slope	Road type	Infiltration/exfiltration
Pipe age	Root interference	Previous maintenance
Pipe depth	Soil corrosivity	Sediment level
Pipe material	Soil fracture potential	Sewer type
Pipe size	Soil moisture	-
Start invert elevation	Soil type	-
-	Soil sulfate level	-
-	Traffic characteristics	-
-	Vehicle flow	-

2.4.1 Pipe Age

The age of a pipe is generally estimated at the time of inspection for condition assessment from the time of installation of the same. Various studies have proved that the age of pipes could influence the condition of sewer pipes (Ariaratnam et al. 2001, Chughtai and Zayed 2008, Kienow and Kienow 2004). The serviceability of pipes decreases with time and is divided into five stages, as shown in the Figure 2-1 (Misiunas 2005). Another study by Singh and Adachi (2011) presented that pipe age is a detrimental factor in pipe failure. According to the study, pipe failure is represented in the shape of a bathtub curve, which is derived when the pipe failure rate is plotted against time. The failure rate was calculated based on historical data of the number of pipe breaks per unit time per unit length of pipe.

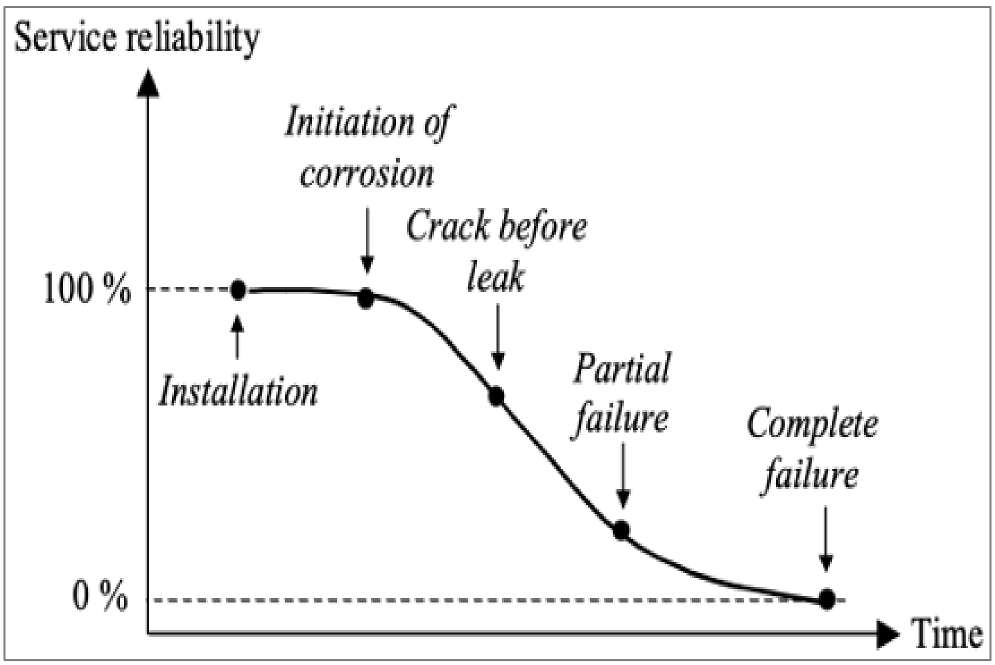

Figure 2-1 Serviceability of a Pipe
(Misiunas 2005)

However, few studies concluded that there is no relationship between deterioration and pipe age. These studies stated that the age could not be considered as a significant factor for deterioration modelling or condition prediction (Tran et al. 2007, Tafuri and Dzuray 2004, and Davies et al. 2001).

2.4.2 Pipe Material

Sewer pipes are made of wide variety of materials such as vitrified clay pipe (VCP), ductile iron (DI), cast iron (CI), polyvinyl chloride (PVC), reinforced concrete pipe (RCP), and so on. Each material has its own properties and failure mechanisms would differ as well. For instance, CI pipes and DI pipes are more susceptible to internal and external corrosion unlike plastic pipes. Plastic pipes, such as PVC or high-density polyethylene (HDPE) on the other hand, may withstand acidic and alkaline wastes, but they may deform excessively when loaded. Likewise, abrasion resistance is high in concrete pipes, and acid resistance is high in clay pipes (Singh and Adachi 2013 and Davies et al. 2001).

Material of the sewer pipes plays a major role during their service life. Pipe materials showed a direct relationship with pipe deterioration in few studies (Davies et al. 2001). Micevski et al. (2002) found that the pipe material is a significant factor of importance towards structural deterioration. The study indicated that concrete pipes are stronger and resilient compared to clay pipes. Bakry et al. (2016) stated that vitrified clay pipes are more resilient than asbestos cement and reinforced concrete pipes. The prediction model developed by Laasko et al. (2018) represented the significance of high-density polyethylene and concrete pipes. Different studies had discussed various effects of pipe material on the condition prediction models, and thus, any pipe material cannot be concluded as better than the other.

2.4.3 Pipe Size (Diameter)

Numerous studies proved that the size of pipe or the diameter of pipe is an influential factor in the deterioration process. Sewer pipes are classified as smaller sewer pipes when the diameter of pipe is between 6 and 8 inches and as larger pipes when the diameter is more than 10 inches. Based on condition prediction models developed in few studies, it was found that the rate of sewer pipe condition's degradation decreases as the pipe diameter increases, whereas few other studies found that larger diameter pipes fail more frequently. Lubini and Fuamba (2011), Salman and Salem (2012) and Bakry et al. (2016) insisted that larger diameter pipes perform well than smaller diameter pipes. Because when obstacles occurred in the larger diameter pipes, they can still run, not necessarily at the full capacity, whereas smaller pipe diameter losses the hydraulic flow. The study stated that larger pipes are buried relatively deep, which could be the reason for better structural condition of large diameter pipes. Therefore, the larger pipe diameter has lower deterioration rates as compared to smaller diameter pipes (Malek Mohammadi et al. 2020, Micevski et al. 2002, Wirahadikusumah et al. 2001, Najafi and Gokhale 2005).

In contrast, the size of pipe was found to be insignificant in a study conducted by Tran et al. in 2007. In addition, according to the study by Jeong et al. (2005), larger pipes are more likely to deteriorate because they have more surface area exposed to sewage and the surrounding soil.

2.4.4 Pipe Length

The length of a sewer pipe generally refers to the distance from entry manhole to exit manhole along the run during inspection. It is believed that shorter pipes are more likely to deteriorate faster than longer pipes. Longer pipes would have a minimum number of severe bends along the run that could result in less accumulation of debris or blockages (Davies et al. 2001, Najafi and Gokhale 2005). In contrast, longer sewer pipes were found to have a higher rate of deterioration because the likelihood of defects is higher in longer pipes (Malek Mohammadi 2019). It is also found that longer runs of sewers would eventually require a reasonable number of joints leading to the risk of infiltration (Jeong et al. 2005). Most common source of infiltration in pipelines is pipe joints, which can lead to soil and groundwater infiltration into sewer pipes.

Khan et al. (2010) discovered that the change in the pipe length has dual performance in the pipe condition. According to the study, when the pipe segment is smaller than 230 feet, the effect on the condition of sewer pipes is zero. However, when the pipe segment is longer than 230 feet, the deterioration rate was found to increase due to the end joints. The end joints were assumed to be a possible source of break, infiltration, and exfiltration. Correspondingly, Laasko et al. (2018) found that pipe segments longer than 131 feet deteriorate earlier than other pipelines in the system due to higher bending stress and potential defects in the longer pipe segments. On the other hand, Salman and Salem (2012) found that longer pipe segments perform better than smaller pipe segments.

2.4.5 Pipe Gradient (Slope)

Slope of the sewer pipe is a significant factor corresponding to the deterioration of sanitary sewer pipes (Baur and Herz 2002). Slope or gradient of a pipe can be estimated by dividing the difference between elevations from mean sea level (MSL) of pipe at start and end of the inspection by the inspected length as shown in Equation 2-1.

$$\text{Slope or Gradient (\%)} = \frac{Elevation\ at\ origin - Elevation\ at\ end}{Inspected\ Length} * 100 \qquad \text{Equation 2-1}$$

Relatively flat sewer pipes are found to deteriorate slower than pipes with greater gradient. When the slope is high, the flow rate will be high as well resulting in easier erosion (Najafi and Gokhale 2005). It is asserted that pipes with very less gradient could enable easier sediment deposition, which would lead to clogging and blockages. Sewer pipes with flat slopes tend to result in lower velocities, which would cause the wastewater to stay within the pipe for longer period resulting in natural hydrogen sulfide generation (Jeong et al. 2005).

2.4.6 Pipe Depth

Depth of a sewer pipe is generally the distance from the pipe's crown (top) to the ground surface. Numerous studies have concluded that the failure or deterioration in sewer pipes are in correlation with the depth of the sewer pipes.

Studies claimed that sewer pipes buried at shallow depths are more likely to deteriorate sooner that those buried at greater depth (Gedam et al. 2016, Harvey and McBean 2014). According to a study by Khan et al. (2010), pipe depth was found to be a significant variable, and any rise in depth has a negative impact on sewer pipe condition level. However, studies by Davies et al. (2001), Tran et al. (2006), and Ana et al. (2009), insisted that depth of sewer pipe is insignificant during model development.

2.4.7 Location of Pipe, Surface Type and Loading

It is obvious that any underground utility structure will have an impact from the surface loads above it. The amount of surface loading carried to the sewer pipe is affected by land use and type of traffic above the pipe. Though the surface loads vary in frequency, making it difficult to estimate their effect on deterioration, a relationship can be found between the surface loading type and the sewer pipe (Kley and Caradot 2013, Najafi and Gokhale 2005). According to Bakry et al. (2016), sewage pipes deteriorate more quickly when they are situated near industrial areas. Few studies affirmed that there is no significant effect of pipe location on their structural condition (Tran et al. 2007, Micevski et al. 2002).

2.4.8 Soil Type

Wirahadikusumah et al. (2001) claimed that the underlying soil has a major impact on sewer pipe degradation. In addition, the type of soil surrounding the sewer pipe is one of the most important factors that could affect frost heave, strength of soil-pipe interaction, and external corrosion, which could lead to failure mechanisms (Najafi and Gokhale 2005). It was determined that pipes installed in unstable soil experienced greater changes in condition compared to pipes installed in stable soil (Tafuri and Dzuray 2004). When there is a lack of soil support around the sewer pipe, it can be shifted. The lack of ground or soil support causes formation of voids around the pipe, making it more likely for the sewer pipe to break or deform. In contrast, Laakso et al. (2018) found that the soil type was not a significant factor in their developed model.

2.4.9 Corrosion

In general, corrosion in metallic pipes is caused by an electrochemical reaction between the exposed pipe's outer surface and the soil environment around it. It should be noted that different pipe materials have different corrosion resistance qualities. It is found that the corrosion rate is influenced by various factors such as soil acidity, resistivity, pH content, oxidation-reduction, sulfide, moisture, aeration, and so on. Longitudinal failure may occur in conjunction with pipe wall weakening due to corrosion (Najafi and Gokhale 2005).

2.4.10 Soil pH

It is important to identify the influential parameters enhancing corrosivity of the soil, which could initiate the external corrosion in pipes. Since different pH ranges induce distinct corrosion mechanisms, soil pH is an excellent indicator of external corrosion (Najafi and Gokhale 2005). In conjunction with soil pH, the material of pipe plays an important role as stated earlier. Compared to steel pipes, cast iron pipes are more likely to corrode in the same corrosive situations (Malek Mohammadi 2019).

2.4.11 Groundwater Level

One of the most common failures in sewer pipes is the infiltration of naturally available groundwater into the sewer pipes. This infiltration could cause overflows as well as soil sediments inside the sewer pipes. Malek Mohammadi et al. (2019) claimed that sewer pipes in areas where the groundwater level is considerably high are more likely to fail than sewer pipes in areas where the groundwater level is below sewer level, which was because of increase in amount of load on pipes from groundwater.

The rate of frost heave is dictated by the available free water around the pipe, which is crucial for external corrosion as well. Lack of soil support and infiltration are caused by the groundwater around the pipe (Davies et al. 2001). As a result, sewer pipes fail structurally because of lack of proper support.

Based on various literatures studied, various factors not limited to the above were identified to influence deterioration in a sewer pipe either internally or externally. However, it should be noted that it is not economically feasible for any municipality to track or collect all the listed factors to their inventory of data.

2.5 Condition Assessment of Sanitary Sewer Pipelines

In sewer system asset management, there is no single standard procedure available to evaluate the structural condition of sewer pipes. Various manuals have been developed by agencies such as Water Research Center (WRc), National Association of Sewer Service Company (NASSCO), and Water Environment Federation (WEF), to evaluate the structural condition of sewer pipes. The manuals provide a step-by-step procedure to estimate a condition rating for the sewer pipes. In the US, PACP developed by NASSCO in partnership with WRc, is the well-established manual for structural condition estimation of sanitary sewer pipes.

2.5.1 PACP Scoring System

In an aim to standardize the way of sewer pipes evaluation, the PACP was established by the NASSCO in 2002. The PACP is a flexible, customizable program to assist agencies and utility owners to record the defects and assess the condition of sewer pipes for decision-making on repair and rehabilitation. Various possible defects in the sewer pipes are uniquely coded in accordance with their severity. For a sewer pipe to be evaluated using the PACP manual, a CCTV camera is setup to record the inner environment of the pipe from one manhole to the other, as shown in Figure 2-2.

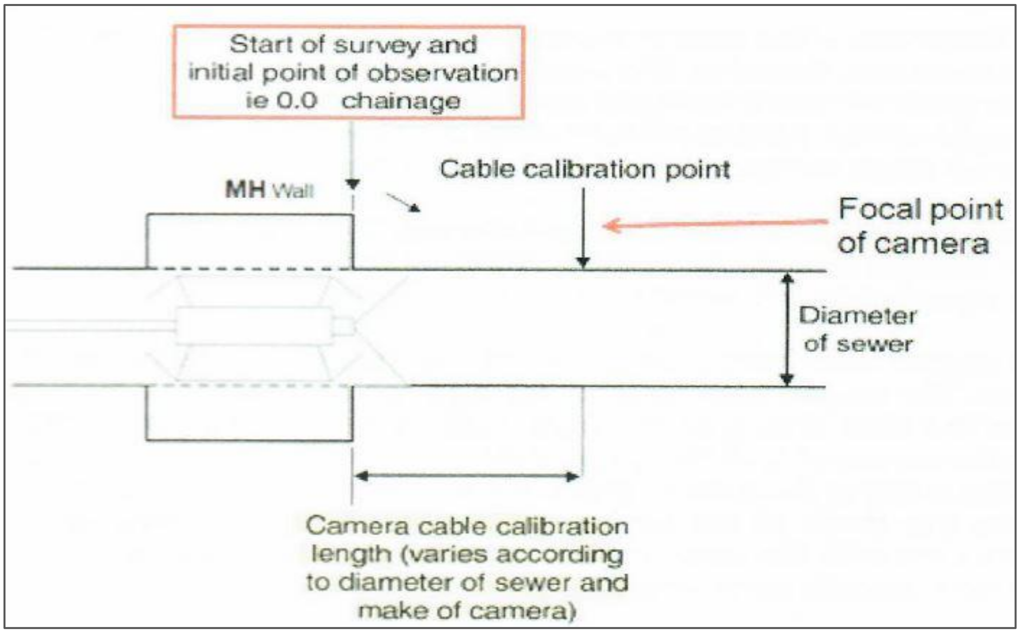

Figure 2-2 PACP Inspection Equipment Setup (NASSCO 2018)

Once the camera is setup at the origin or upstream manhole, the slack is pulled from the next manhole opening. As per the manual, the maximum speed of CCTV camera setup is 30 feet per minute. While the camera moves from origin towards the next manhole, a certified PACP operator watch the live TV stream and record the visible defects into a predefined spreadsheet program. There are nearly 230 different kinds of defects discussed in the program and a sample of the PACP inspection form is shown in Table 2-2. When the inspection is complete, computer program estimates the final condition score of each inspected pipe segment. Final score of inspected sewer pipe can be interpreted as shown in the Table 2-3. A PACP score of 1 refers to excellent condition pipe where PACP score 5 indicate that the pipe has already failed or would fail within the next 5 years.

Table 2-2 PACP Inspection Form Details Section (NASSCO 2018)

Distance (feet)	Video Ref	Code	Continuous Defect	Value			Joint	Circumferential location		Image Ref.	Remarks
		Group/ Modifier		Dimension		%					

Table 2-3 PACP Condition Rating (NASCO 2018)

PACP	Description	Estimated Time to Failure
1	Excellent	Unlikely to fail in the foreseeable future
2	Good	20 years or more
3	Fair	10 years to 20 years
4	Poor	5 years to 10 years
5	Needs Immediate Attention	Already failed or likely to fail within the next 5 years

2.5.2 Importance of Sewer Pipe Condition Prediction

It is obvious that not all the sewer pipes in an inventory would be at a structurally bad condition or near failure. In addition, inspecting every individual sewer pipe in a system would be an expensive and time-consuming process. Based on the speed of operation and involved test setup as discussed in previous section, financial requirements for every inspection operation could be understood.

Therefore, there is a need to identify critical sewer pipes for inspection among the entire inventory. Inspection of sewer pipes can be limited by scrutinizing the pipes in bad condition by predicting them in advance. This prioritization in inspection of pipes would save thousands of dollars to any municipality (Chae and Abraham 2001, Chae and Abraham 2000, Wright et al. 2006).

Such prediction of sanitary sewer pipe's condition is not a new concept. With the advancement in computer technology and advanced statistical analysis using machine learning algorithms or artificial intelligence, numerous studies have been accomplished by researchers to predict the condition of sewer pipes. However, there is no single standard model could be developed because every municipality do not record the same kind of data to their database inventory. Therefore, there is a huge demand in many municipalities for a proper asset management plan and inspection prioritization.

2.6 Machine Learning and Artificial Intelligent Models in Sewer Pipe Condition Prediction

The science (and art) of programming computers to learn from data is known as machine learning. Arthur Samuel defined machine learning in 1959 as a field of study that gives computers the ability to learn without being explicitly programmed (Géron 2017). The machine learning can be broadly classified into 2 categories such as 'supervised learning' and 'unsupervised learning'. Most of the data analysis performed in various studies related to condition prediction fall under supervised learning. In supervised learning, the computer program or algorithm is trained or directed to study the recorded historical data that includes the output or target variable. Based on the training, prediction is estimated for a new set of data or unrecorded data. Whereas with respect to unsupervised learning, the data used for training would not include the target variable.

Various training models and algorithms are being developed depending on the type of application. Most common techniques employed in machine learning models are either based on regression, classification, or combination of both. Various names are given to algorithms based on working principles of the program such as decision trees and k-Nearest Neighbors (clustering algorithm). Research studies conducted to predict the condition of sanitary sewer pipes are briefly summarized in the following section.

In a binary logistic regression, the relationship between a single nonmetric (binary) dependent variable and a set of metric or nonmetric independent variables is estimated. Model's output would be a probability of the instance being either true or false, success or failure, and zero or one. For instance, the dependent variable in sewer condition prediction models could be characterized as structurally good or bad. As a notation, sewer pipes in good condition can be classified as 1 in output and pipes in poor condition as 0 (Malek Mohammadi 2019).

Various studies have been performed to predict the structural condition of sewer pipes based on logistic regression models for decades. A study conducted by

Ariaratnam et al. in 2001 predicted the likelihood of a sewer infrastructure system to be in a structurally deficient state. The study developed a binary classification model and the developed model utilized various factors such as the age, diameter, and material of pipe, transported waste type, and average depth of installed pipe. However, the model was validated using only 86 records, which is comparatively lesser for a municipal database and therefore, the resulted model could lead to misclassification.

An expert knowledge-based support system to prioritize sewer pipeline inspection was developed in a study by Hahn at al. in 2002. The study employed the probabilistic method, Bayesian belief network, to develop the model based on interviews and case studies. A decision support tool named Sewer Cataloging, Retrieval and Prioritization System (SCRAPS) was developed based on the likelihood and consequences of failures. However, the developed SCARPS tool was based on WRc’s 1986 paradigm of pipe assessment. In addition, the study did not focus on model’s applicability in field.

Najafi and Kulandaivel in 2005 developed a condition prediction model using Artificial Neural Network (ANN) technique. Various factors such as age, length, size, material type, depth, slope, and sewer type were considered as independent variables to train the model. It was found the model performed well during training and the performance was unsatisfactory during testing. The study acknowledged that the results required thorough statistical analysis for further application. It was recommended that a model must depend on a larger and more inclusive data.

Structural and operational condition assessment models were developed in a study by Chughtai and Zayed (2008). Multiple regression models were developed based on independent variables such as age, diameter, depth, length, material, bedding factor, and street category. The study developed 3 models for 3 different pipe materials namely, concrete, asbestos cement, and PVC. Developed models were good at identifying sewer sections vulnerable to overflows and basement flooding. However, the models were not capable of quantifying the criticality of sewer pipes.

Syacharni et al. (2013) developed a decision-tree based deterioration model for sewer pipes. The study employed various techniques such as regression, decision trees, and neural networks. Models were trained with parameters such as length, slope, diameter, material, root, sludge, and debris. It was found that decision tree model outperformed both regression and neural network models. It was interesting to notice that root and sludge were found to be influential factors over pipe material.

Another study by Harvey and McBean (2014) developed a structural condition prediction model for individual sanitary sewer pipes. Machine learning technique, random forests was utilized to train the model. The model used sewer pipe age, material, diameter, depth, length, slope, sewer type, invert, and road type as attributes while training the model. The developed model found to exhibit an area under the receiver operator characteristic (ROC) curve (AUC) value of 0.81. However, a model cannot be evaluated solely based on the ROC curve and AUC values; other evaluation metrics such as precision and recall must be considered.

In 2017, Hernandez et al. developed a structural condition prediction model using various machine learning algorithms such as logistic regression, random forests, multinomial logistic regression, linear discriminant analysis, and support vector machine. The study compared the performance of various models. However, true positive rate and false positive rate were the only evaluation metrics used to compare the performance. In addition, prediction capability was found to be unsatisfactory as well.

Another study by Laakso et al. in 2018 accomplished condition prediction model development using logistic regression based on a wide variety of factors such as age, material, diameter, depth, length, slope, sewer type, location, road type, number of trees, and flow rate. The developed model resulted an accuracy of 56%. However, the model utilized around 19 predictor variables, which is not economical or practically viable for municipalities to include all variables in their inspection databases.

A study accomplished by Malek Mohammadi et al. in 2020 developed condition prediction models for sanitary sewer pipes using various machine learning models such as decision trees, random forests, and gradient boosting tree.

The developed model based on gradient boosting tree was found to have an accuracy of 87%. Though the accuracy level was acceptable, the model classified the condition rating of pipes on a binary class rather than multi-class condition ratings. The study had recommended that the future research could concentrate on multi-class condition prediction, which would be more beneficial for the municipality during inspection and condition assessment phases.

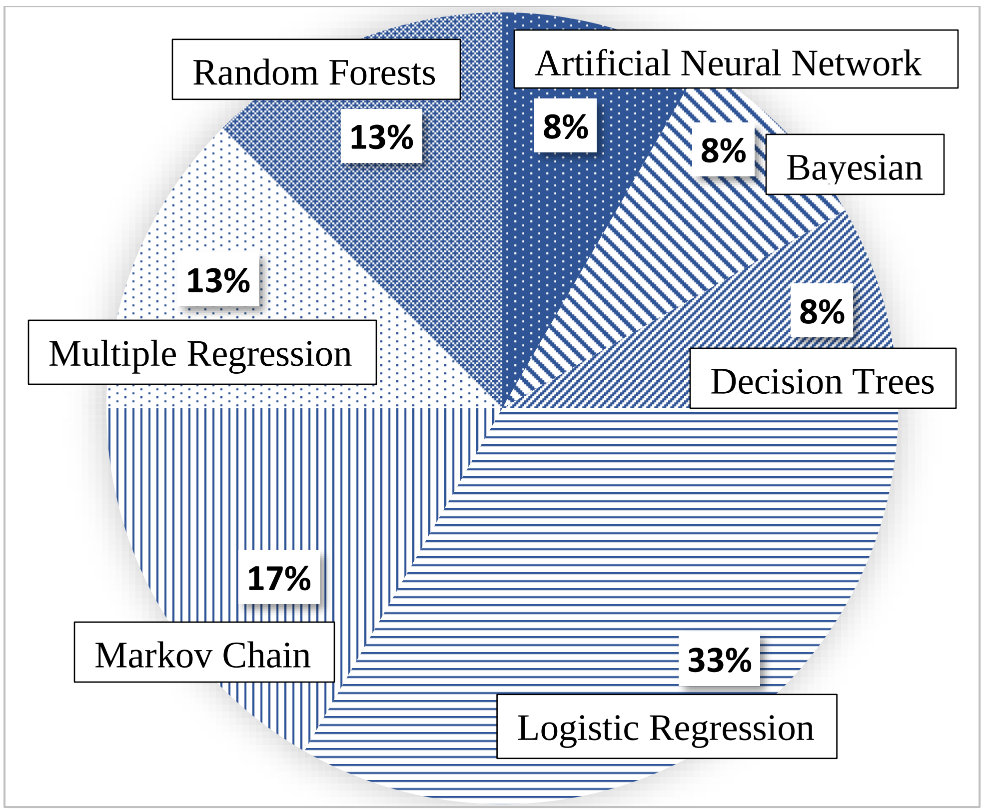

Figure 2-3 Techniques Used in Prediction Model Development

As a summary, different techniques employed in various research studies to develop condition prediction model is illustrated in the Figure 2-3. Majority of the studies considered logistic regression to develop a prediction model. A significant number of studies has given importance to machine learning algorithms like decision trees and random forests. However, studied literature studies have recommended that the application of machine learning and artificial intelligent techniques could be further expanded.

2.7 Chapter Summary

This chapter discussed various factors that could lead to structural deterioration in sanitary sewer pipes. It was found that sewer pipe failure is a complex process involving multiple factors. Inspection and condition assessment of all pipes in an inventory of wastewater agencies or municipalities on a regular basis is not economically viable. Thus, prioritizing the inspection of sanitary sewer pipes by predicting their condition could be a beneficial solution for efficient budget allocation for any municipality. Secondly, the importance for machine learning and artificial intelligence in structural condition prediction of sanitary sewer pipes was discussed based on numerous research studies. It was found that further studies are recommended by various researchers to develop multi-class prediction models rather than binary classification.

CHAPTER 3

DATA COLLECTION AND PREPARATION

3.1 Introduction

Inspection and condition assessment of sanitary sewer pipes are the critical steps involved in asset management of the system. It is well known that the CCTVs are most employed in the United States for inspection of sanitary sewer pipes (NASSCO 2018). Historical data collected by the City of Fort Worth (Texas) has been used in this study to develop a condition prediction model that could serve as a basis to prioritize future inspection of sanitary sewer pipes. In the City of Fort Worth, wastewater collection system (sewers) is separate from the storm drainage system and stormwater does not flow through the sewers. The scope of this study is limited to gravity flow sanitary sewer pipes excluding force main systems.

Like most other municipalities, CCTVs are primarily employed in the inspection and condition assessment process of sanitary sewer pipes. Based on interviews with officials, it was found that sewer pipe inspection decisions were made based on engineering judgement and operator experience. The inventory of the sewer system is stored using geographic information system (GIS) databases. The inventory of recorded database includes information but not limited to the installation details of pipes, surrounding soil type, location of pipe with respect to geographical maps, and so on. Detailed discussion of various information collected from the wastewater department is explained in following sections.

3.2 Overview of Collected Data

With more than 280-miles length of large diamter sewer interceptors whose diamters are greater than 24 inches, the wastewter system forms a complex underground infrastructure. The GIS is primarily employed to record, manage, and maintain the inventory of sewer systems. Based on the historical inspection data provided by the city, it was found that the CCTV inspection for inner environment of pipes accomplying to PACP manual started in the year 2000. During inspection, each and every pipe is given an unique name for future identification, referred as GIS ID. A sample of the data collected from the city of Fort Worth is shown in the Table 3-1. Dataset contained 32,854 number of unique pipe segment details.

Table 3-1 Sample of Data Collected for the Study

GIS_ID	INSPEC_ DATE	INSTALL_ DATE	INSPEC_ LENGTH	MAPSCO GRID	UPELEV	DOWN ELEV	SUBAREA	STYPE	DIAMETER	MATERIAL	PACP
60717	6/6/2011	7/25/1958	460	93G	551.44	551.25	VC08_01	1	39	CI	2
60718	12/12/2010	8/18/1988	844	93G	551.3	550.88	VC09_01	1	54	CONCRETE	2
60719	12/28/2017	7/17/2001	415	46L	792	790.35	MC04_04	1	8	PVC	1
60720	1/5/2018	7/1/2004	426	46H	760.1	752.45	MC04_04	1	8	PVC	1
60723	11/16/2012	12/3/1964	259	89F	673.9	668.57	CF05_03	1	6	VCP	3
60724	11/19/2012	12/9/1964	503	89F	689.06	673.9	CF05_03	1	6	CONCRETE	2
60726	8/7/2019	3/1/2005	112	119G	641.37	640.73	VC11_03	1	8	PVC	1
60728	8/9/2017	4/25/2002	444	106U	610.87	607.88	VC11_01	1	24	PVC	1
60729	4/28/2017	2/28/2002	396	106S	647.17	645.67	VC11_01	1	24	PVC	2
60732	11/13/2014	6/11/1947	203	47Y	723.7	716.86	MC03_06	1	6	VCP	2

3.3 Preliminary Data Insights

GIS_ID is an unique identification code given by the inspection operator at the time of inspection for future use. INSTALL_DATE refers to the date at which the pipe was installed for service. It was interesting to note that sewer pipes were installed as early as 1909 and the same are in service to date. Majority of the pipes installed in the early twentieth were primarily concrete or vitrified clay pipes. In addition, it was found that around 70% of the total number of pipes were installed in a span of 30 years between 1980 and 2010.

INSPEC_DATE refers to the date when the corresponding inspection was completed. Though the first inspection data dates back to the year 2000, only around 500 individual pipe segments were inspected for condition assessment until the year 2005. Difference between the installed date and inspection date would yield one of the important characteristics, age of the particular pipe segment.

INSPEC_LENGTH is the total distance in feet measurements inspected from the upstream manhole to the downstream manhole of a sanitary sewer segment. The inspected length of pipes varied from a minimum of 8 feet all the way to 5,500 feet at few instances. Inspection records show that almost 99% of the sewer pipes do not run longer than 1,000 ft.

Based on sewer design guidelines, the manual recommended a maximum of 600 ft spacing between manholes for sewer pipes with diameter greater than 27 inches (American Iron and Steel Institute 1999). However, the manual metioned that upon approval by the agency, the spacing between manholes (i.e., the inspection length) can exceed the design manual specified 600 ft. Therefore, for this study, the maximum length of sewer segment is limited to 1,500 ft.

Interestingly, none of the sewer pipes with a PACP scores of 5 and 4 had their lengths greater than 1,500 ft and 2,000 ft, respectively, as shown in the Figure 3-1.

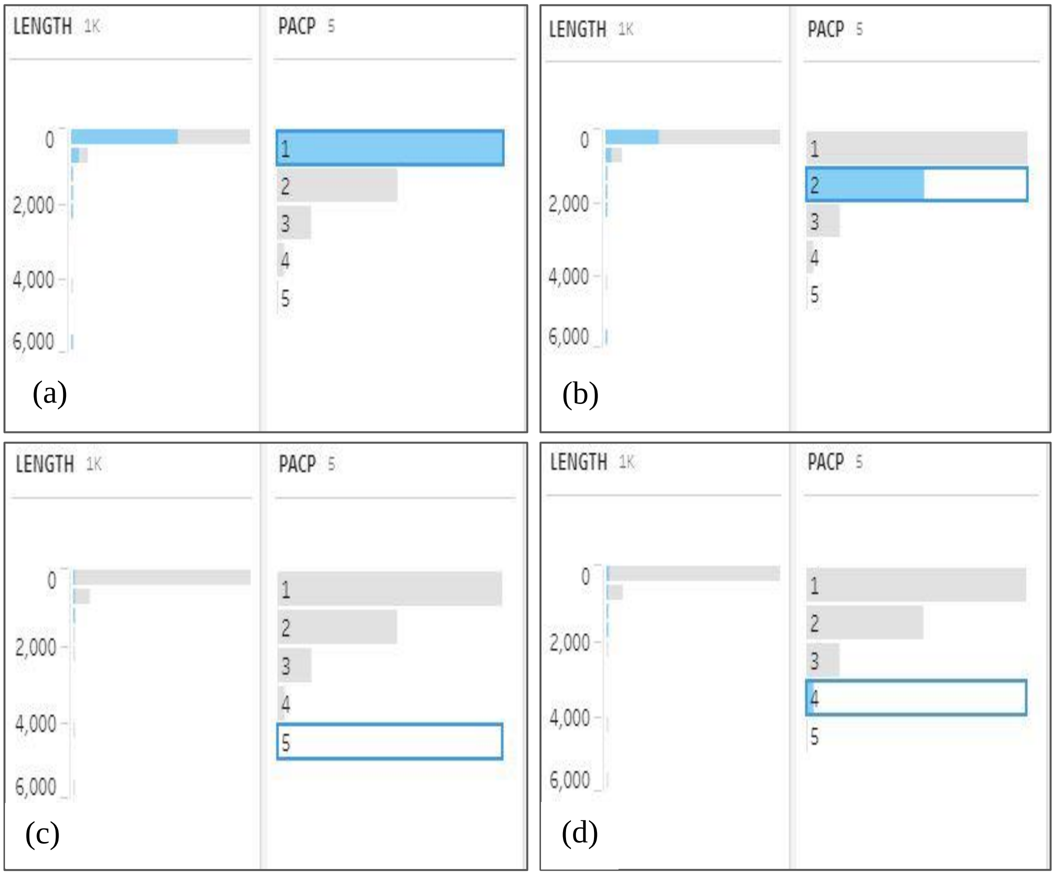

Figure 3-1 Overview of Length and PACP Scores

As illustrated in the Figure 3-1(a), highlighted number of pipes in the left are corresponding to the selected PACP score of 1 on the right. Similarly, in Figure 3-1(b, c and d), the number of pipes in different length ranges can be found correspinding to the PACP scores of 2, 4, and 5, respectively. The comparison between the PACP scores of pipes and their corresponding lengths in the illustration indicates that the shorter pipes are more prone to deterioration compared to longer pipes.

MAPSCOGRID is basically a geographical location reference for the particular pipe. Mapsco, situated in Addison, Texas, was a privately held publisher of maps and atlases, founded in 1948. Since 1952, Mapsco has become a leading maker of maps for the states of Texas, New Mexico, Oklahoma, and Colorado areas. Maps for cities were developed in the form of numbered grid systems; and were referred as Mapscogrids. An example of a Mapscogrid for a part of City of Fort Worth is shown in the Figure 3-2.

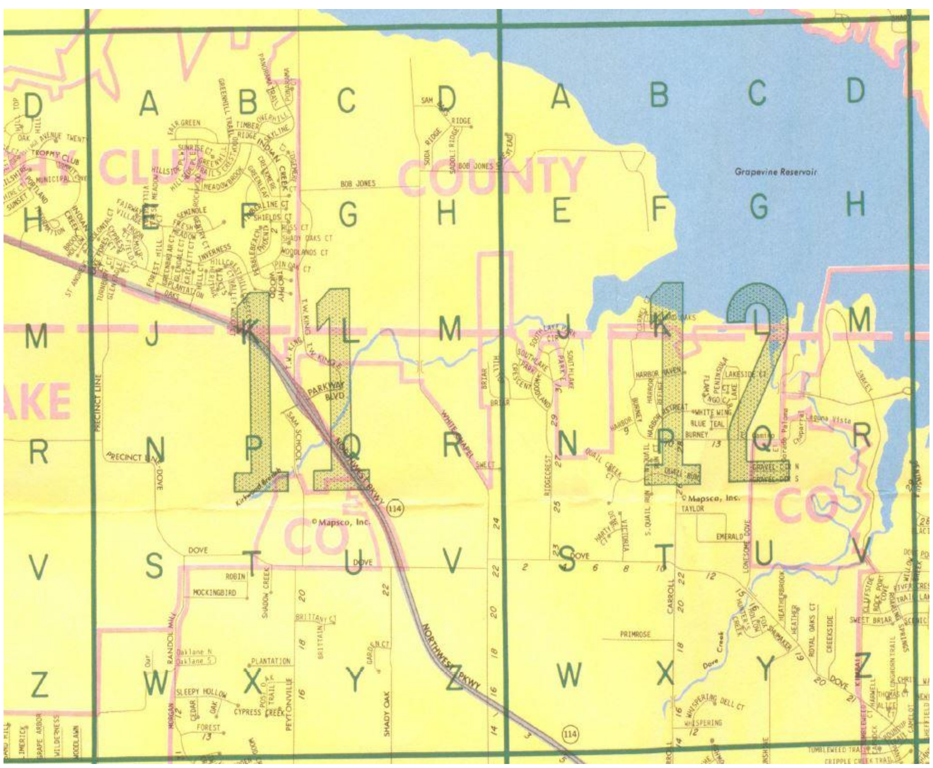

Figure 3-2 Sample of a MAPSCOGRID System in City of Fort Worth (Geography and Map Division, Library of Congress, 2020)

As shown in Table 3-1, MAPSCOGRID for each pipe segment is an unique alphanumeric code. The first numerical part refers to the rectangular box or grids in the map. Once the numerical grid is located, the pipe can be found at the following alphabet's vicinity. The sanitary sewer systems in the city of Fort Worth runs under around 988 unique values.

UPELEV and DOWNELEV are the elevations in feet above sea level at upstream manhole and downstream manhole, respectively. This information is of much importance to estimate the slope or gradient of the sewer pipe. The geography of City of Fort Worth is relatively flat and there is no huge difference between the elevations at upstream manhole and downstream manhole. Almost all of the sanitary sewer pipes have their UPELEV and DOWNELEV between 400 ft and 1,000 ft above the sea level.

SUBAREA refers to the drainage basin or the type of surrounding area. Based on the drainage basin, unique alphanumeric codes were used to refer to the subarea surrounding the sewer pipelines. The first two alphabets in SUBAREA refer to the type of the basin such as Clear Fork (CF), Village Creek (VC), Big Fossil (BF), and so on. It was noticed that majority of the sewer pipes were located at BF, CF, and VC basins.

STYPE column in the collected data refers to the types of sewer pipelines based on their flow. The types of sewer flows are either gravitational or force mains.

The gravity main sewers are termed 1 and force main sewers are termed 2. For this study, gravity main sewers are considered for further analysis.

DIAMETER column represents the size or diameter of the pipe in inches. Diameter of the sewer pipes ranged from 4 inches to 96 inches. However, almost 90% of the pipes were found to be less than 20 inches in diameter and 6% of the pipes were in the range between 20 and 40 inches. Interestingly, none of the pipes with diameter greater than 60 inches were found to have a PACP score of 5, which indicates that pipes with larger diameter were in structurally fair condition compared to smaller diameter pipes. As shown in Figure 3-3, highlighted diameter values in the left are corresponding to the selected PACP score of 5.

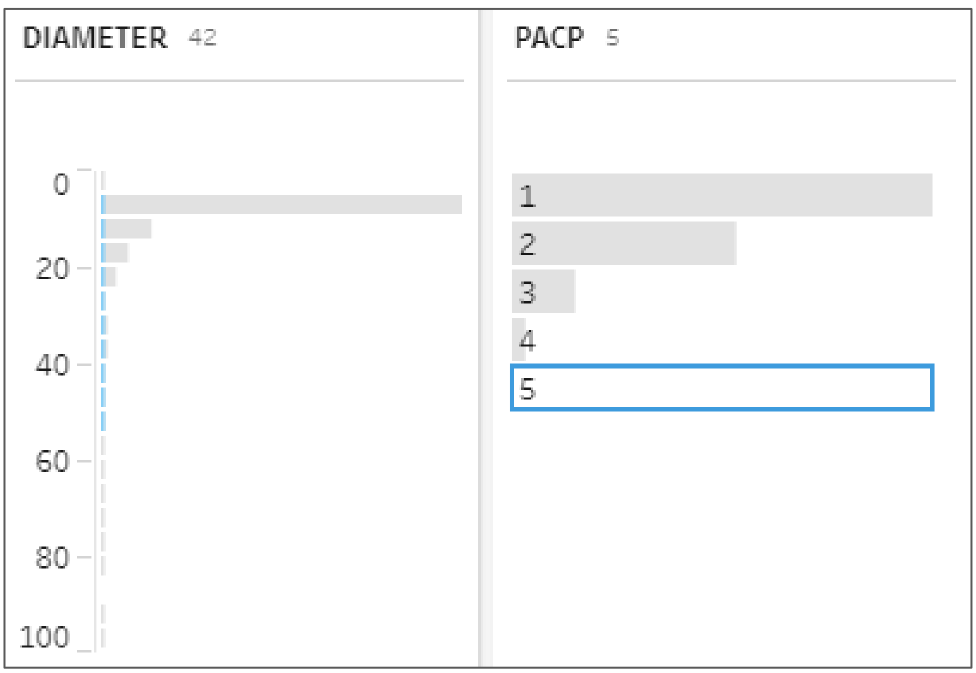

Figure 3-3 Overview of Diameter and PACP Scores

MATERIAL column in the Table 3-1 represents the type of material used to manufacture the sewer pipe. Records indicate that 9 different material types were used in sewer pipes such as armco truss (AT), cast iron (CI), concrete, ductile iron (DI), fiberglass, high density polyethylene (HDPE), polyvinyl chloride (PVC), steel, and vitrified clay pipes (VCP). It was found that a very few number of pipes made of AT, CI, fiberglass, and steel, were used in gravity main sewers while the majority of the pipes are made of PVC, VC, concrete, and DI. AT is a type of plastic material with inner and outer PVC walls filled with a lightweight material called Mearlcrete for additional pipe stiffness and compressive strength (Moore 2015).

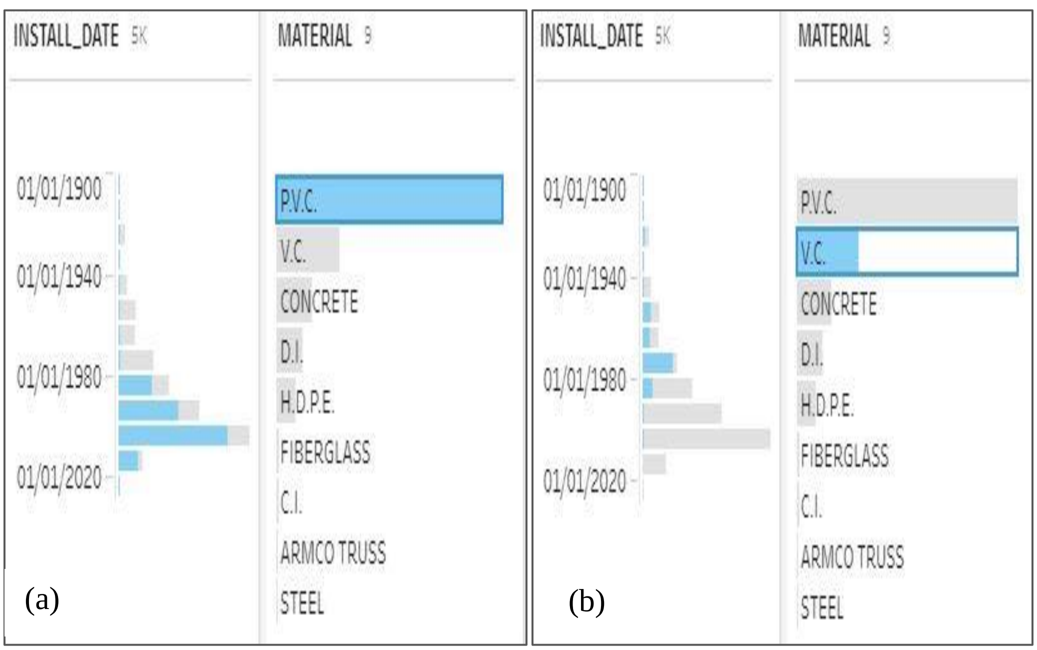

Figure 3-4 Pipe Materials and their Installation Years

It was interesting to notice that though PVC pipes constitute majority of the sewer pipes, most of them were installed after 1980 as shown in the Figure 3-4(a).

At the same time, vitrified clay pipes constitute a considerable amount of total pipes and were majorly installed before 1980. In addition, it was found that almost all types of pipes were facing deterioration as well, as shown in the Figure 3-5.

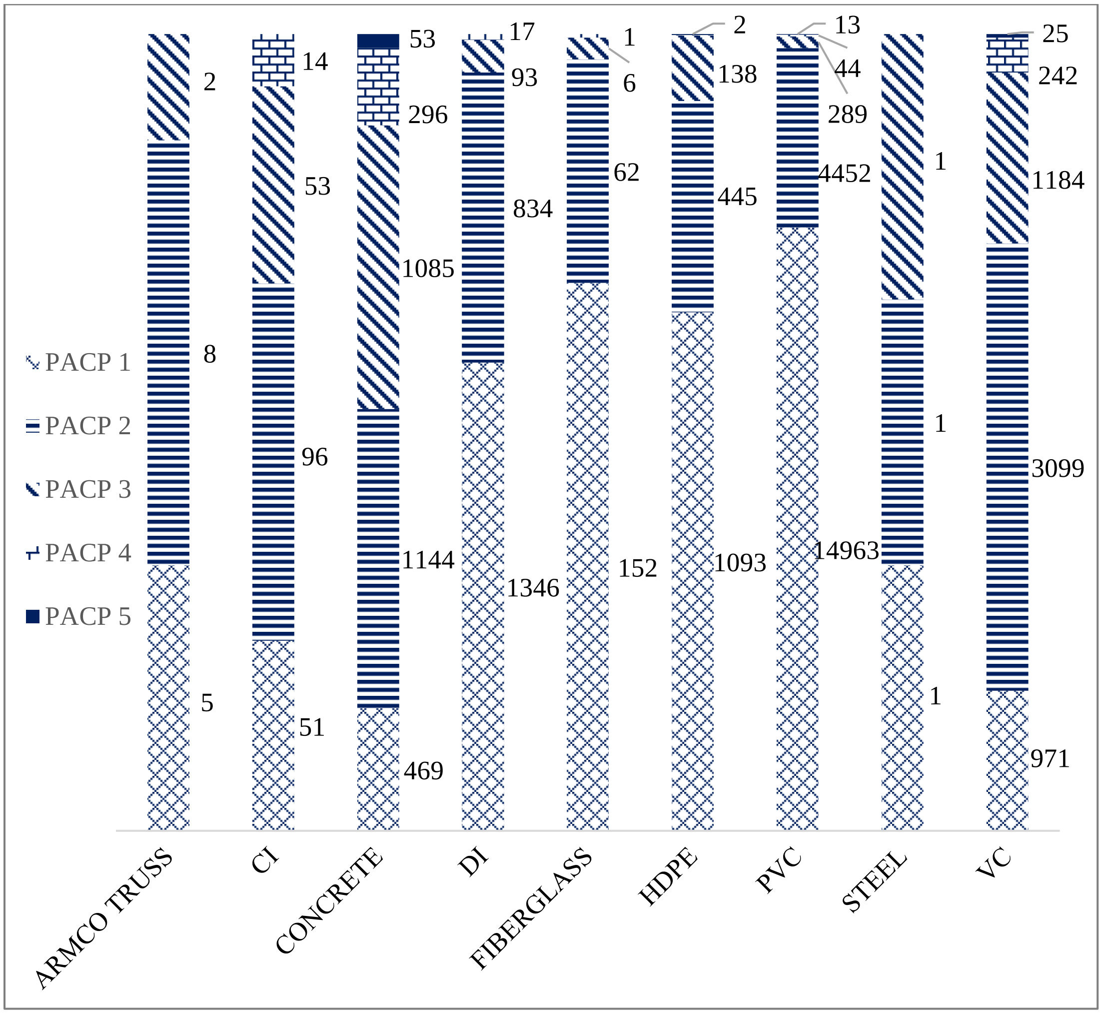

Figure 3-5 Pipe Materials and PACP Scores

The final column in the collected data lists the PACP scores for each inspected pipe segment. As already discussed, the PACP scores range on a scale from 1 to 5, where 1 refers to a structurally good condition and 5 refers to a pipe nearing failure. The Figure 3-5 illustrates the distribution of PACP scores among different pipe materials.

As shown in the Figure 3-5, structurally weak pipes are distributed in majority of the pipe materials such as concrete, PVC, VC, fiberglass, and HDPE. It was found that majority of pipes with PACP score of 5 are made of concrete and comparitively fewer pipes are made of PVC. However, it can not be concluded that a particular type of material is better than the other.

3.4 Data Preparation

The data collected from inventory of databases are in their original form and have to be processed in order to utilize them in model development or for any form of statistical analysis. Because, the GIS database collected might include erroneous or misleading data in its original form as they are manually entered to the databases. The pre-processing of collected data is one of the important step in preparing the data for further analysis.

In the pre-processing step, collected original data is refined to exclude the redundant information from the dataset for further analysis. In other words, it can be defined as the extraction of required information from the original form of data, which can be utilized as input feed to computer programs. As one of the steps in data preparation, the collected data is processed to avoid any null values. Null values found in the MAPSCOGRID field were excluded from further analysis. Therefore, final dataset with 32,751 datapoints was utilized for analysis and model development. As a next step in data preparation, individual features such as age and slope were calculated based on collected data to include in model development phase. Exploratory analysis of each feature is discussed in detail in the following section.

3.5 Exploratory Data Analysis

Processed dataset was then used to extract features for data analysis and model development. Final dataset processed for analysis included 6 independent variables and a multi-class categorical dependent variable. Extracted features include information such as age, length, diameter, slope, MAPSCOGRID, and SUBAREA. Details of individual feature is shown in Table 3-2.

Table 3-2 Details of Extracted Features

Variable Type	**Features Extracted (Variables)**	**Description (Data Types)**
Independent (Response Variables)	Age	Continuous Numerical
	Length	
	Slope	
	Diameter	
	MAPSCOGRID	Nominal
	SUBAREA	Nominal
Dependent (Target Variable)	PACP	Multi-class Categorical

3.5.1 Age

The first feature extracted from the collected data is the age of the inspected sewer pipe segment. The difference between inspection date and installation date would yield the age of the particular sewer pipe segment. The distribution of age of sewers is shown in the Figure 3-6. As illustrated in the Figure 3-6, the age of pipe was found to be from less than an year to a maximum of 107 years. Almost 73% of the pipes were found to be under 30 years in service since installation and 2% of the pipes were found to be more than 80 years of age.

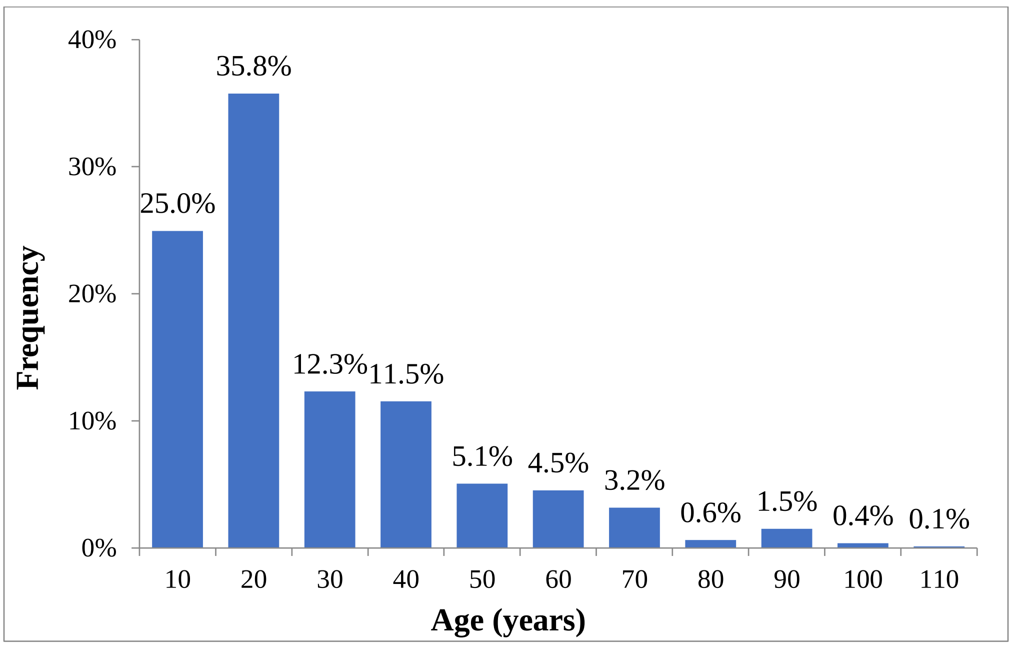

Figure 3-6 Frequency Distribution Based on Pipe Age

3.5.2 Length

Length of the sewer pipe segment is the manhole to manhole distance recorded during inspections. Distribution of frequency of pipe segment length in percentage is shown in Figure 3-7. The length of inspected sewer pipe segment varies from 8 ft to 1,500 ft. It can be seen that around 81% of total pipe segments spans less than 400 ft and a very few observations were found to run more than 1,000 ft of length.

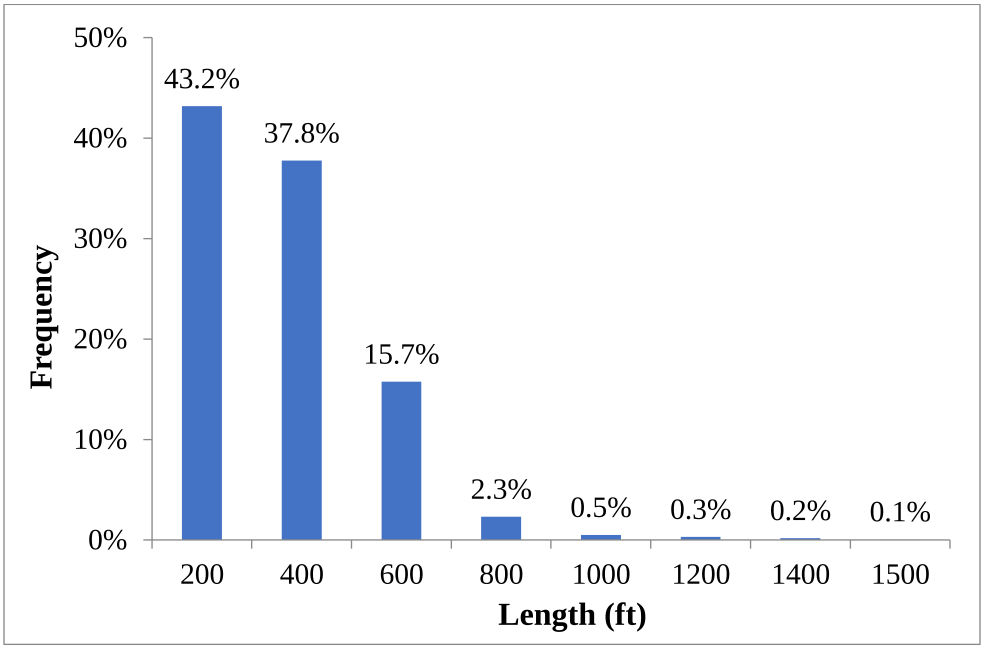

Figure 3-7 Frequency Distribution Based on Pipe Length

3.5.3 Slope or Gradient

Slope of the sewer pipe segment is calculated using the vertical and horizontal displacements. It is estimated by dividing the difference between upstream manhole elevation and downstream manhole elevation by the length of the inspected pipe segment. It was found that 99% of the pipes were relatively flat with a maximum slope of 0.2%. However, maximum slope was found to be 5%. The distribution of slope is illustrated in Figure 3-8.

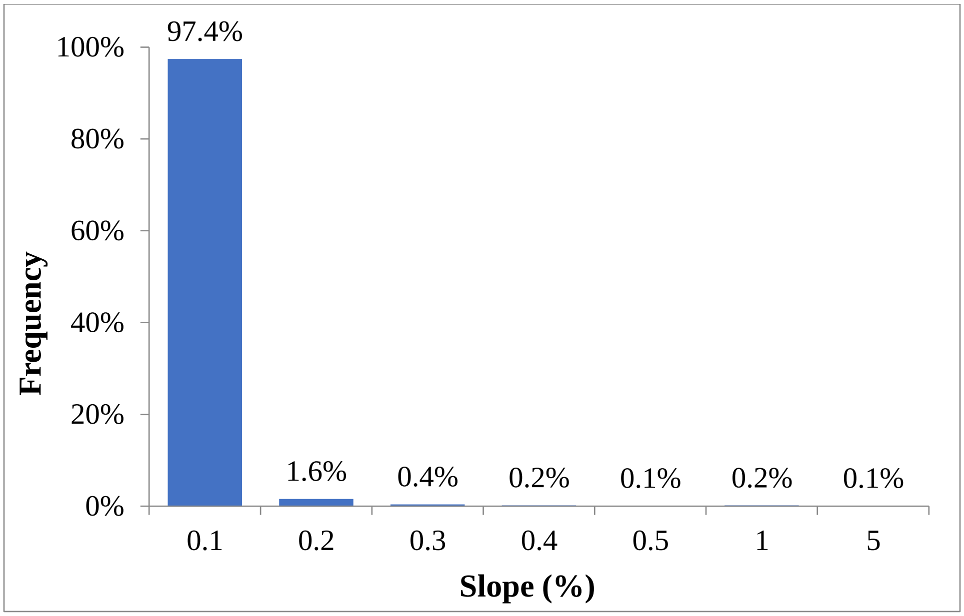

Figure 3-8 Frequency Distribution Based on Pipe Slope

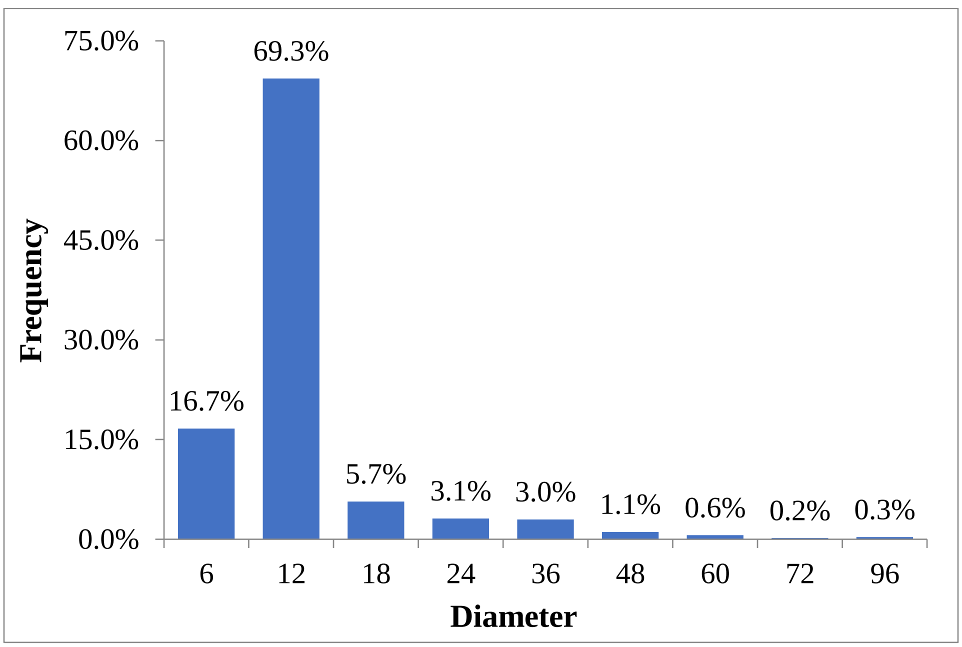

Figure 3-9 Frequency Distribution Based on Pipe Diameter

3.5.4 Diameter

Diameter of the sewer pipe segment is a basic information recorded at the time of inspection for each pipe. The diameter of the pipes ranged from as small as 4 inches to as large as 96 inches, and the distribution of the same is shown in

Figure 3-9. Based on the collected data as shown in

Figure 3-9, almost 90% of the inventory of pipes can be categorized as smaller pipes since their diameter is less than 24 inches. Roughly 1% of the pipes are found to have a diameter greater than 60 inches.

3.5.5 Material

As discussed in previous sections, sewer pipes are made of different materials and as per processed dataset in this dissertation, there are 9 different types of sewer pipe materials are identified. As shown in the Figure 3-10, 4 pipe materials such as AT, steel, CI, and fiberglass, contribute less than 2% of total dataset. It was found that PVC constitute a major portion of the sewer pipes with around 60%, followed by vitrified clay and concrete pipes with 17% and 9%, respectively. Since structurally poor conition pipes are distributed in majority types of pipe materials, all different pipe materials are included in model development.

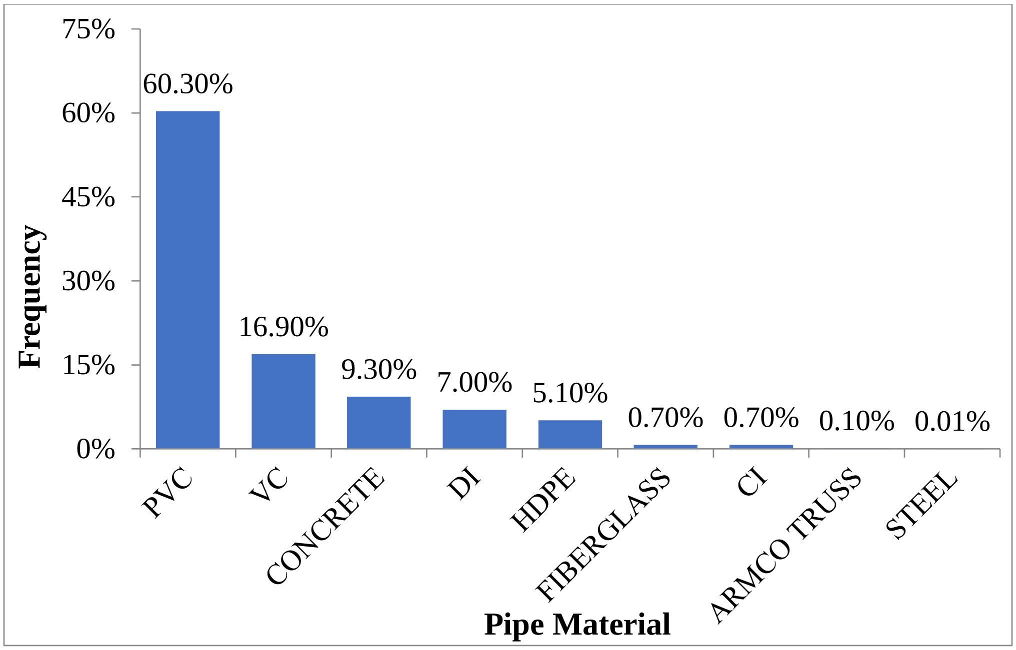

Figure 3-10 Frequency Distribution Based on Type of Pipe Material

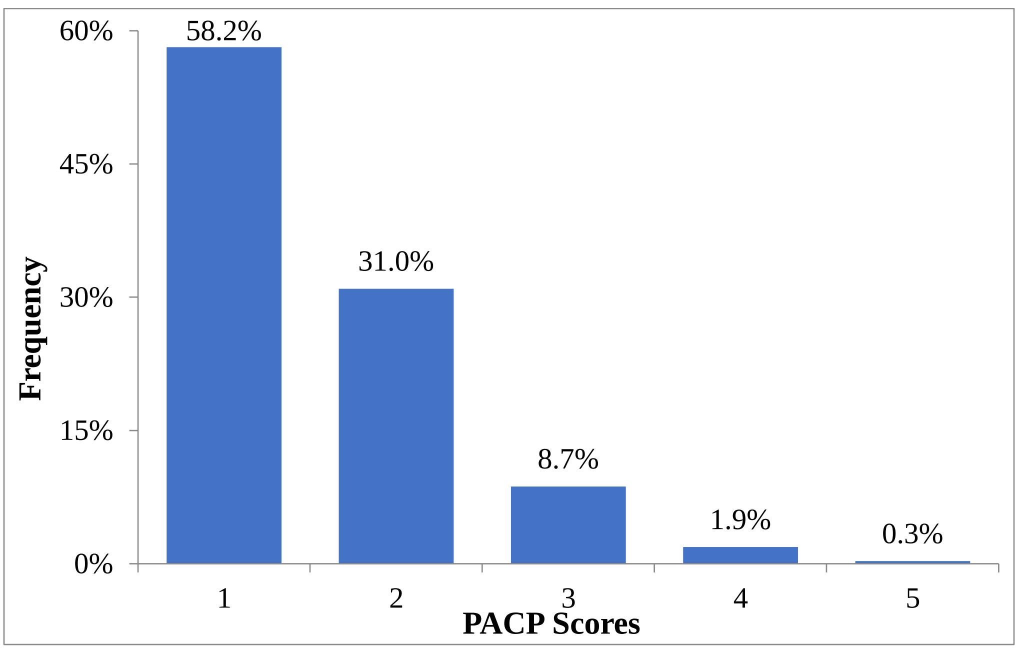

Figure 3-11 Frequency Distribution of PACP Scores

3.5.6 PACP

The last column in processed dataset is the dependent or target variable for condition prediction model development, which is the PACP score of individual pipe segment. As discussed earlier, the scores range from 1 to 5, where 1 refer to structurally sound pipes and 5 refer to pipes in the verge of failure. As shown in the Figure 3-11, almost around 90% of the pipes are in structurally sound condition with PACP scores of 1 and 2.

It was interesting to notice that only 2% of the pipes are ranked a PACP score of 4 and only 95 sewer pipe segments are in the verge of failure with a PACP score of 5. It should also be noted that more than 70% of the pipes are less than 30 years old and around 60% of the pipes are in structurally good condition. Though age can not be a single influential factor to dictate structural condition of sewer pipes, a simple linear relationship can be assumed.

Another crucial observation made from the PACP distribution is that the distribution is not even among all 5 classes. When one or more classes in a dataset is under represented compared to other classes, the scenario would be termed as class imbalance. Predictive accuracy of developed models for minority classes would be highly affected when imbalanced dataset is used in machine learning algorithms (Wallace et al. 2011). Various studies have mentioned that the imbalanced classification must be treated before training the dataset in any machine learning algroithms (Rout et al. 2018), which will be discussed in following chapter.

3.6 Chapter Summary

In this chapter, preliminary insights about the data collected for this study was discussed. It was observed that the PVC pipes constitute majority of the total sewer pipes followed by VCP and concrete material pipes. It was noticed that around 81% of total pipe segments spans less than 400 ft and almost 90% of the inventory of pipes can be categorized as smaller pipes whose diameter is less than 24 inches. In addition, a linear relationship was observed between age of the pipe segments and structural condition. Finally, it was found that the dependent variable is unevenly distributed causing a severe class-imbalance, which must be treated while developing the prediction models.

CHAPTER 4

MODEL DEVELOPMENT

4.1 Introduction

Developing a condition prediction model based on various machine learning algorithms is discussed in this chapter. Machine learning is a wide term that refers to computational algorithms that rely on prior knowledge to generate precise predictions (Mohri et al. 2018). Prior knowledge in the context refers to the recorded historical data that a computer program can learn, which is termed as training the algorithm. As discussed in previous chapters, supervised learning techniques are employed in the study and numerous machine learning techniques are in practice for developing a prediction model under supervised learning.

Classification and regression are the two major types of supervised machine learning methods. Regression method is used when a continuous dependent variable must be predicted based on various independent variables (Müller and Guido 2016). In this study, the dependent or output variable is not a continuous number rather it is categorical with 5 different classes. Hence, classification type of machine learning techniques is used in this study to develop the model. Processed data from the previous chapter is utilized as input for training the models. One of the most popular programming languages in the field of data science, Python, is used in this study to develop prediction models. The reason for using Python is that it is open-source and the availability of large number of free add-on libraries.

4.2 Imbalanced Dataset Treatments

As discussed earlier, PACP score of 5 has rare instances compared to PACP score of 1, which is termed as imbalance in the dataset. When the imbalanced or skewed data is used in traditional classification algorithms, it could often result in poor performance of trained models (Tanha et al. 2020, Teh et al. 2020, Yijing et al. 2016). Generally, minority class must be given more importance while handling imbalanced dataset because after-effects of minority class's misclassification would be exponential than the other classes. In this study, PACP score of 5 class is given more importance because misclassifying PACP score of 5 as 1 would be worse.

It was found that classification algorithms such as logistic regression, support vector machine, and decision tree are well suited for training a balanced dataset. When an imbalanced dataset is trained using these algorithms, majority class was highly represented distorting the minority class instances (López et al. 2013). Various treatment techniques are performed by researchers and data scientists to yield better results out of imbalanced dataset (Haixian et al. 2017). One of the commonly employed treatment for imbalanced dataset, data resampling technique is performed in this study. Basically, in data resampling technique, the imbalanced data is resampled to match either of the majority class or minority class by replication or removal of datapoints, respectively. Data resampling technique is divided into two categories: (1) Random under-sampling and (2) Random over-sampling.

4.2.1 Random Under-Sampling

Random under-sampling technique is nothing but the removal of observations in random from the majority class to match the minority class. As illustrated in the Figure 4-1, instances from PACP score 1 was removed in random. It can be considered a better choice when the dataset contains extensively large number of observations. However, it should be noted that valuable information might be lost during random removal of instances.

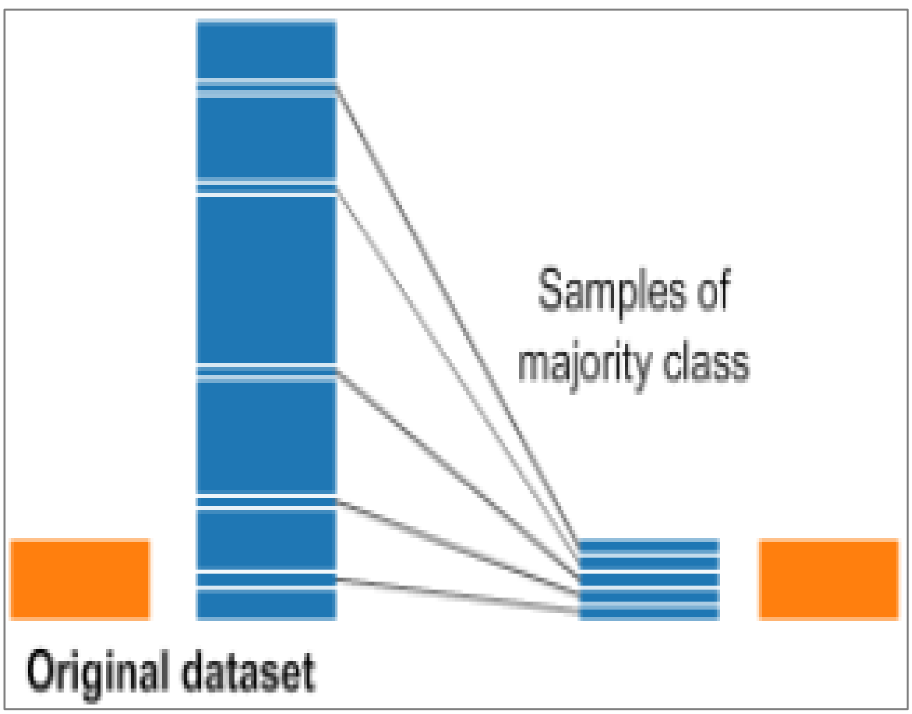

Figure 4-1 Random Under-Sampling Technique

4.2.2 Random Over-Sampling

Random over-sampling is also a resampling technique similar to random under-sampling, in which the majority class is not removed rather the minority class is replicated to match the majority class. Illustration for random over-sampling is shown in the Figure 4-2. This technique would be highly helpful when very minimal data is available. However, when a severely imbalanced dataset is replicated as a balanced dataset, it might cause the algorithms to memorize or overfit the minority class instances.

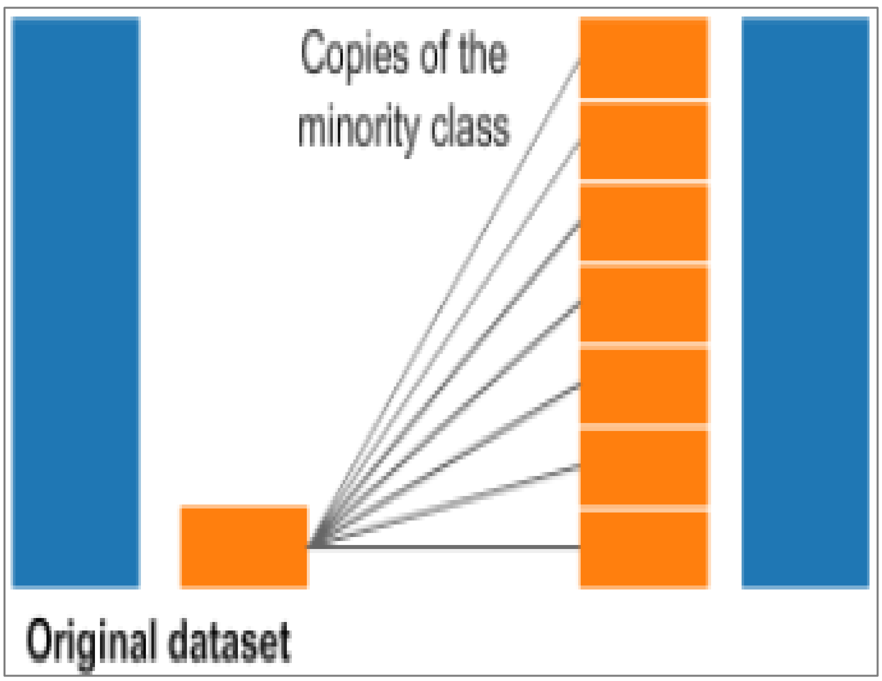

Figure 4-2 Random Over-Sampling Technique

Since both of the discussed resampling techniques are found to be effective in different studies, over-sampling technique cannot be rated as better over the under-sampling technique and vice versa (Estabrooks et al. 2004). It was found that both resampling techniques has their advantages and disadvantages. Therefore, in this study, models are trained with both resampling techniques and corresponding performances are evaluated.

4.3 Cross-Validation

Cross-validation is basically the most employed validation technique in any prediction problem. Basic concept behind cross-validation is that some portion of the input dataset is excluded while training the model and the excluded portion is used while testing the trained model. Primary reason for using cross-validation is that it would avoid overfitting and sample from all classes could be represented while training the model. The key element of cross-validation technique is that the entire dataset will be used in training and testing the model (Malek Mohammadi 2019).

For example, in a 10-fold cross-validation, the entire dataset is bagged into 10 equal parts. From 10 parts, 9 parts will be used to train the model and 1 part will be used while testing the trained model. Owing to the fact that increase in folds would decrease the number of data points in each part, 5-fold cross-validation is employed in this study, as shown in the Figure 4-3. In a 5-fold cross-validation, 4 parts or 80% of the dataset in random was used in training and the rest 20% of the data was used in testing.

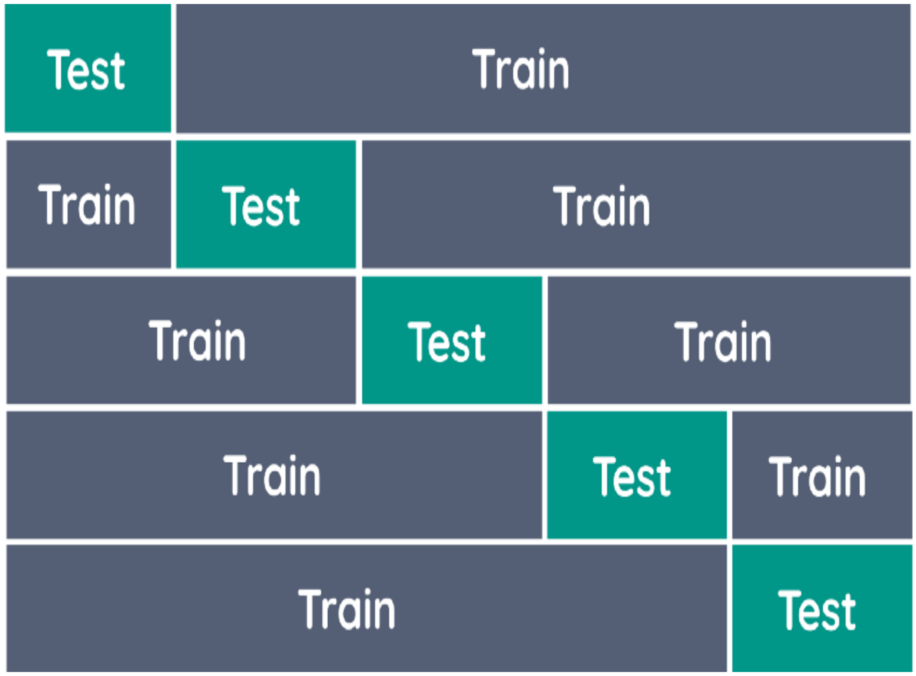

Figure 4-3 5-Fold Cross Validation

4.4 Machine Learning Methods

As discussed in earlier chapter and previous sections of this dissertation, collected data was cleaned and final dataset for further analysis is prepared. Imbalanced target variable was identified and hence, resampling techniques were employed to the dataset before it can be trained using machine learning algorithms. Therefore, the final dataset is now ready to be fed as input to various classification algorithms. As mentioned earlier, numerous open-source libraries in Python programing language are utilized to train the models. Some of the libraries used in this study are shown in the Table 4-1.

Table 4-1 Python Libraries Used in the Study

S. No	Name of the Library	Description or Functions of the Library
1	Pandas	To open spreadsheet files and manipulate numerical tables.
2	Scikit learn	This library features various classification algorithms
3	Matplotlib	It is the most common plotting library to plot graphs
4	Seaborn	It is a data visualization library
5	Streamlit	To develop an interactive decision-support tool

Using different libraries listed in the Table 4-1, various machine learning classification methods such as logistic regression, support vector machine, k-nearest neighbors, and random forests, are trained to develop the prediction models. Above mentioned methods are trained with all three sets of data, namely imbalanced dataset, under-sampled dataset, and over-sampled dataset. Each of the method employed in the study is discussed in the following sections. It should be noted that for logistic regression and k-nearest neighbors, the continuous independent variables such as age, length, slope, and diameter, must be on uniform scale and therefore, log transformed variables are used in model development.

4.4.1 Logistic Regression

Logistic regression (LR) is one of the most employed statistical methods in machine learning. Though the term used in this method is regression, the output is basically a probability of an event to happen. Generally, LR methods are used to analyze the relationship between multiple independent variables and a categorical or continuous dependent variable. In this method, the data is fit to a logistic or sigmoid curve and is used to estimate the probability of an event. When the target (dependent) variable is binary or dichotomous, binary logistic regression is utilized to create prediction models.

Dependent variables in a binary LR model have two possible values, which are mostly 0s and 1s. For example, if the pipe classification is based on either of two conditions, say good or bad, pipes in good condition could be given a label of 1 and pipes in poor condition can be labeled as 0. For a binary response variable Y and a single dependent variable X, let $\pi(X) = P(Y = 1 \mid X = x) = 1 - P(Y = 0 \mid X = x)$, the logistic regression model has linear form for the logit of this probability as shown in the Equation 4-1 (Agresti 2007). Figure 4-4 illustrates the simple logistic function used in estimating the parameter coefficients. In the illustration, the horizontal axis (x) varies from -6 to +6 and the vertical axis (f(x)) corresponds to the probability from 0 to 1.

$$\text{logit}\,[\pi(X)] = \log\left(\frac{\pi(X)}{1-\pi(X)}\right) = \alpha + \beta x \qquad \text{Equation 4-1}$$

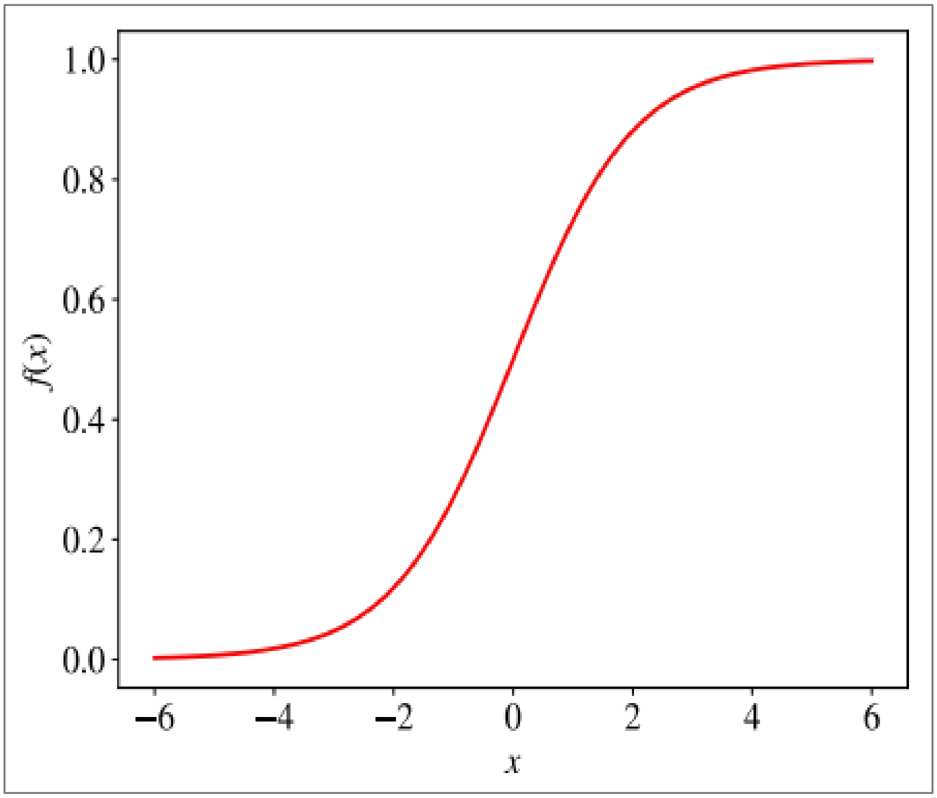

Figure 4-4 Standard Logistic Function

Logistic regression can also be employed where the dependent variable is discrete with more than two classes. This type of LR with more than 2 classes in output variable is termed as multinomial LR. While employing LR for multi-class classification, the likelihood of one class would be estimated over the rest all classes. For instance, to develop a prediction model for PACP score of 5 in this study, sewer pipes with a PACP score of 5 would be considered as one of the binary classes and all other classes such as PACP scores of 1, 2, 3, and 4, would be the other class in binary classification.

For a multinomial or multiple logistic regression, the final model would take the form as shown in the Equation 4-2.

$$\text{logit}\left[\frac{\pi}{1-\pi}\right] = \log\left(\frac{P(Y=1 \mid X_1, X_2, ..X_p}{1-P(Y=1 \mid X_1, X_2, ..X_p}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + .. + \beta_p x_p$$

Equation 4-2

Where:

$X_1, X_2, \ldots, X_p$ are independent variables

α is the intercept for i^{th} category

β is the regression coefficient

4.4.2 k-Nearest Neighbors (k-NN)

The k-NN algorithm is a supervised machine learning technique that can be utilized in both classification and regression problems. k-NN algorithm is named lazy not because of its seeming simplicity, but because it memorizes the training dataset rather than learning a discriminative function from it (Guo et al. 2003). The training dataset is all that is required to build a k-NN model. The algorithm finds the closest data points in the training dataset – its "nearest neighbors" – to classify a new data point.

In the most basic form, the k-NN algorithm only analyzes one nearest neighbor, which is the training data point that is closest to the point we wish to classify. The known output for this training point is then used to make the prediction as illustrated in the Figure 4-5. However, to increase accuracy, an arbitrary number of neighbors, k, can be considered (Müller and Guido 2016).

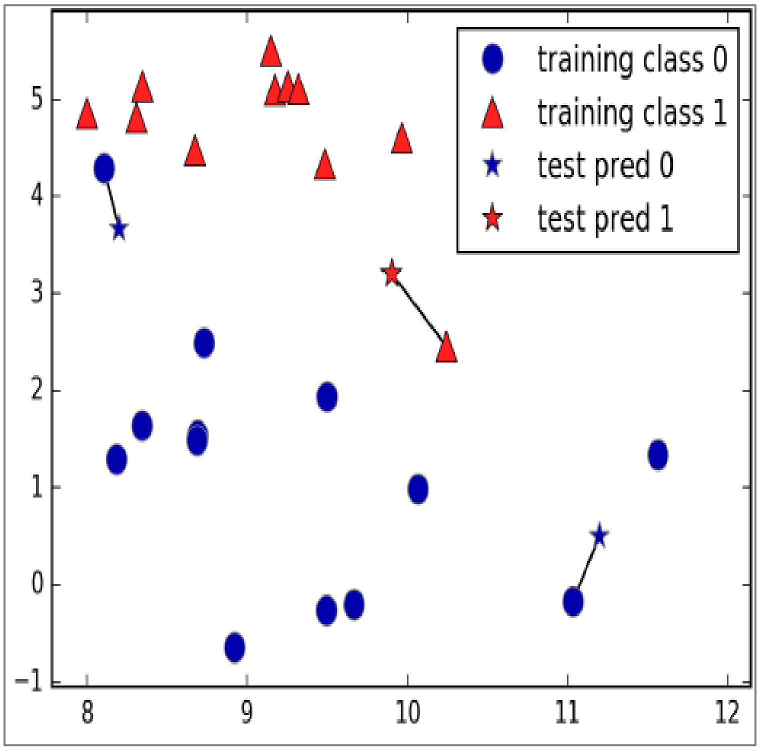

Figure 4-5 Simplest form of a k-NN Model

(Müller and Guido 2016)

For example, predictor space of the k-NN model utilized in this study is shown in the Figure 4-6. Here, the number of neighbors is selected as 3, which is k = 3. It should also be noted that though there are 7 independent variables used to train the model, only 3 variables are shown in the illustration referring to a lower-dimensional space.

When more than one neighbor is considered, a technique called voting is used to assign a label to the new data point of interest. Voting is nothing but the total count of different class labels near the data point of interest. When a majority of class labels belong to a particular class, the test data point will be assigned to that majority class. Therefore, it is always recommended to use an odd number for k, which would eventually avoid confusions during prediction based on nearest neighbors.

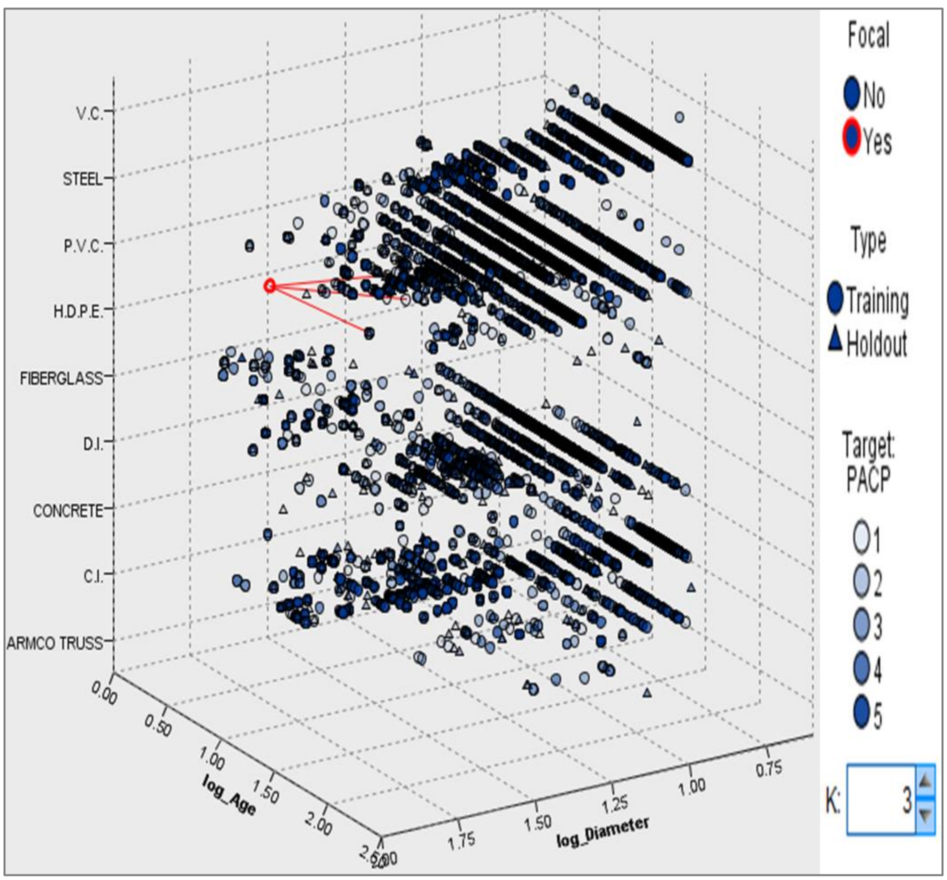

Figure 4-6 Lower-Dimensional Projection of the k-NN Predictor Space

4.4.3 Tree Based Models

Tree based models are multi-purpose machine learning algorithms that can handle classification, regression, and multi-output problems. Decision trees (DT) and random forests (RF) are most popular tree-based machine learning models. They are extremely powerful algorithms that can fit large datasets (Loh 2014). Although DT is considered as an effective supervised learning algorithm in classification problems, one of the most common limitation is that DT tend to overfit the training data (Müller and Guido 2016). Therefore, RF method is used in this study to overcome the limitation of DT.

RF is based on ensemble learning, which is a method of integrating many classifiers to solve a complicated problem and enhance the model's performance. In simple words, RF is a combination of different DT during training the data. An RF is essentially a collection of various DTs, where each tree is a little different from the others. In DT, though each tree may accomplish an acceptable job of predicting, it will almost certainly overfit on some part of the data. The amount of overfitting could then be limited by averaging the outcomes of numerous trees, which operate well and overfit in diverse ways (Estabrooks et al. 2004 and Caruana and Niculescu-Mizil 2006). The final collection of numerous DTs with retained predictive power can be collectively named as RF. A schematic illustration of RF technique is shown in the Figure 4-7.

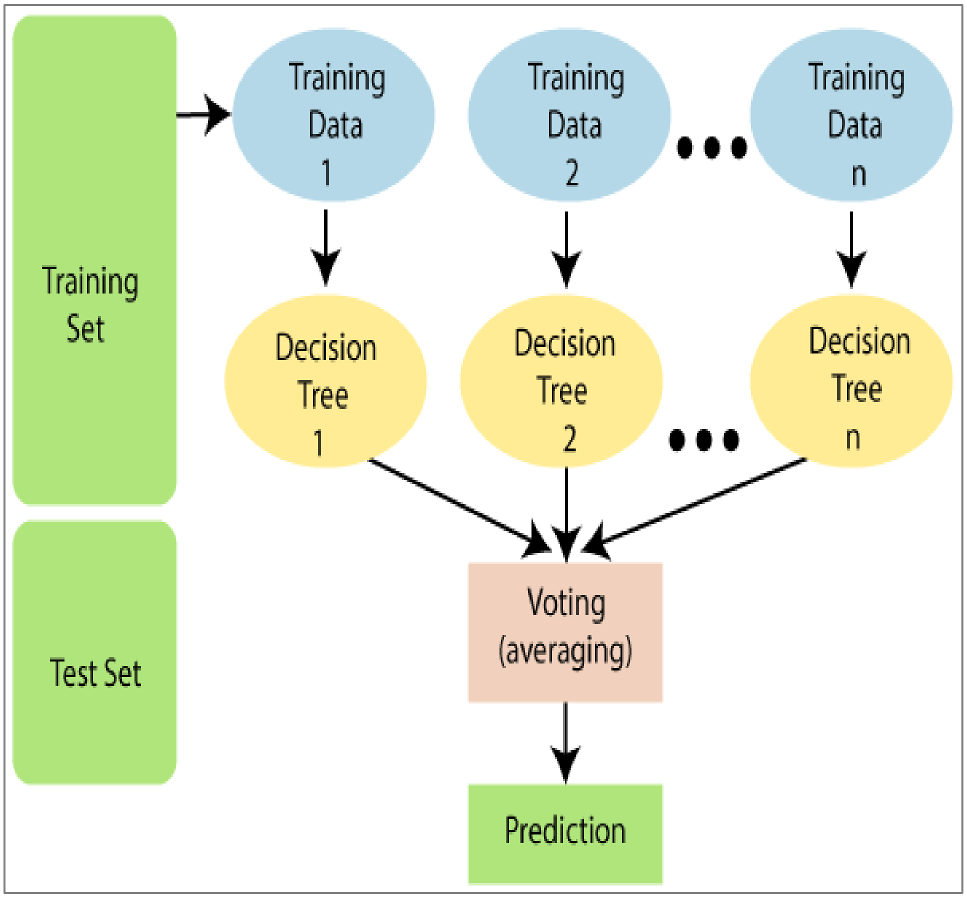

Figure 4-7 Working Structure of RF Algorithm

For tree growth, tree combination, self-testing, and post-processing, RF follows specific principles. RF is identified to be steadier in the presence of outliers and in high-dimensional parameter spaces than other machine learning methods and therefore, it is resistant to overfitting (Caruana and Niculescu-Mizil 2006). The Gini index (Gi) is a measure of the predictive capacity of variables in classification. Gi is non-parametric, which means it does not depend on data from a specific sort of distribution (Alessia et al. 2017). For a simple binary classification, the Gi of a node 'n' is calculated as:

$$Gi\ (n) = 1 - \sum_{j=1}^{2} \left(p_j\right)^2 \qquad \text{Equation 4-3}$$

Where P_j is the relative frequecy of clas j in the node n.

4.5 Chapter Summary

This chapter discussed about the imbalanced dataset treatments such as resampling techniques to prepare the data to be trained using machine learning algorithms. Under-sampled and over-sampled datasets will be utilized to train and develop the condition prediction models as discussed in following chapters. Various supervised learning algorithms such as LR, k-NN, and RF, utilized in this study are also discussed.

CHAPTER 5

MODELS PERFORMANCE COMPARISON

4.1 Introduction

In previous chapters, collected data from GIS databases was preprocessed and prepared as input for machine learning algorithms. The final dataset was utilized in training various supervised learning techniques. In this chapter, performance of the trained models in predicting the condition of sewer pipes will be discussed. To identify a better prediction model, all trained models must be validated and evaluated. There are various evaluation metrics such as confusion matrix, accuracy, precision, recall, and so on, are available to evaluate the performance of machine learning models. Various evaluation metrics used in evaluating the performance of prediction models are discussed in following sections.

4.2 Evaluation Metrics

This section of the dissertation discusses about various evaluation metrics in detail. The selection of a particular metric would be based on the type of anticipated output from the classification model.

4.2.1 Confusion Matrix

Confusion matrix is one of the most important model evaluation metrics, which is widely employed to evaluate the performance of a trained machine learning model. The number of occurrences between two raters, the true/actual classification and the predicted classification are all recorded in a cross table, which is referred as a confusion matrix. Representation of a simple binary classification confusion matrix is shown in the Figure 5-1. A confusion matrix could yield an overall understanding of the performance of a model by visual observation (Grandini et al. 2020, Hossin and Sulaiman 2015).

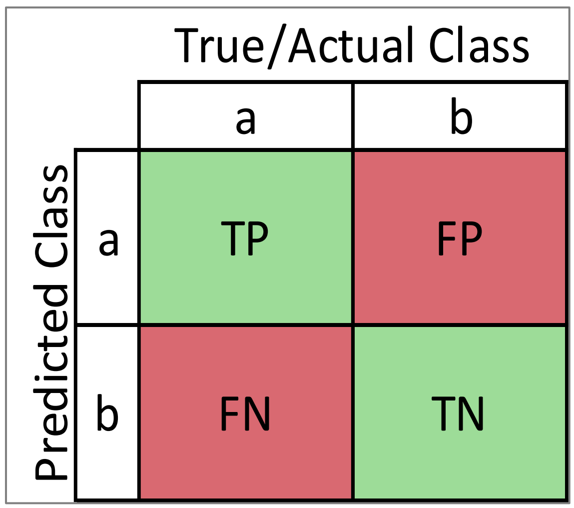

Figure 5-1 Confusion Matrix for a Binary Classification

In a confusion matrix, the correctly classified items are placed from top left to bottom right on the major diagonal, and they correlate to the number of instances the two classes agree. In the confusion matrix shown above, TP (True Positive) refers to truly predicted positive instances and TN (True Negative) refers to correctly predicted negative instances. FN (False Negative) elements are those that the model has predicted as negative but are positive and similarly, FP (False Positive) elements are those that the model has predicted as positive but are negative. It could be emphasized that the number of elements in cells other than the major diagonal cells must be minimal for a better performing model (Malek Mohammadi 2019).

The confusion matrix is considered one of the most important metrics because majority of the evaluation metrics are calculated based on the developed confusion matrix and the elements present in different cells. For a better understanding, various evaluation metrics are explained based on a binary confusion matrix and multi-class evaluation metrics will be discussed in the later sections.

4.2.2 ROC Curve and AUC

A receiver operator characteristics (ROC) graph is another commonly employed visualization, organization, and selection tool for classification-based models. ROC curve is a two-dimensional graph that displays how efficiently a classification model behaves as the discrimination cut-off value is tweaked across the predictor variable's range. In the graph, the predictive test's false positive rate is represented on the x axis and the true positive rate is represented on the y axis, as shown in the Figure 5-2 (Malek Mohammadi 2019).

True Positive Rate (TPR) – is the ratio of TP to the sum of TP and FN, which is TP / (TP + FN).

False Positive Rate (FPR) – is the ratio of FP to the sum of TN and FP, which is FP / (TN + FP).

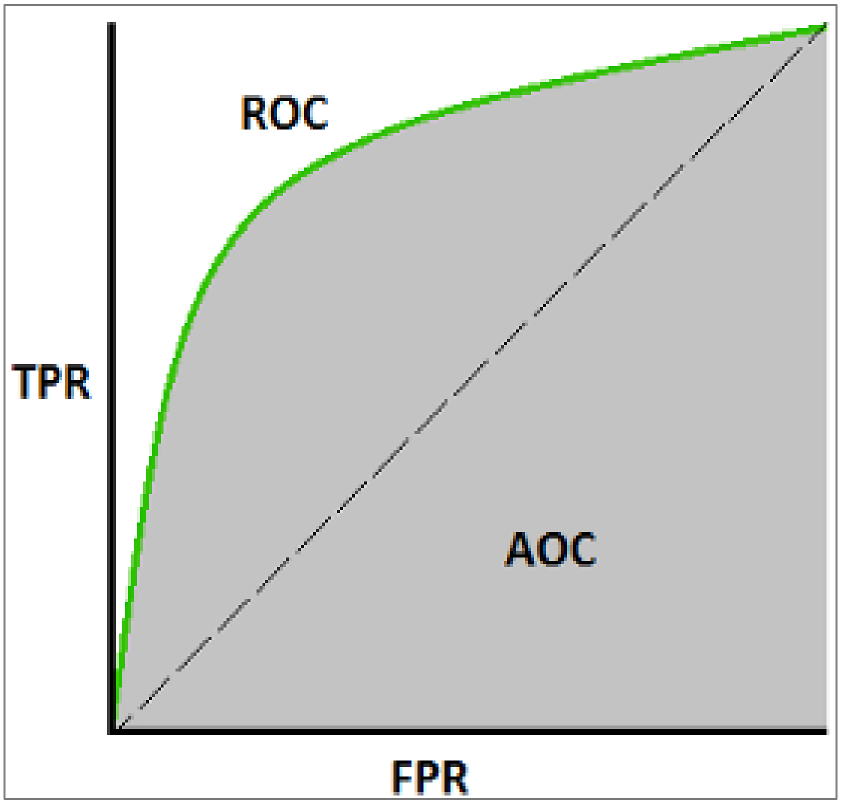

Figure 5-2 ROC Curve for a Binary Classification

The illustration shown in Figure 5-2 is based on binary classification or a classification for two classes. Area under the ROC curve (AUC) is another interesting metric extracted from the ROC curve, which is the shaded portion under the curve in the illustration. The AUC of a classifier could be defined as the likelihood that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. As the dimension of the chart is a unit square, the AUC ranges from 0 to 1. It can be concluded that higher the AUC, better would be the model performance in prediction.

4.2.3 Accuracy

Another popular metric in classification models, which is also estimated from the confusion matrix, is the accuracy. It is the ratio of sum of TP and TN to the grand total of the confusion matrix, as shown in the Equation 5-1 (Hossin and Sulaiman 2015). Since the formula for accuracy incorporates entire confusion matrix including incorrectly classified elements, it is an overall measure of the model's correct predictions. However, accuracy is found to be efficient for binary classifications compared to multi-class classification (Hossin and Sulaiman 2015).

$$\text{Accuracy} = TP + TN / (TP + TN + FP + FN) \qquad \text{Equation 5-1}$$

4.2.4 Precision

The Precision is defined as the ratio of true positive elements to the total number of positively predicted units. In other words, it is the proportion of predicted positives, which are truly positive as shown in the Equation 5-2 (Hossin and Sulaiman 2015). Based on the definition, precision of a model is significant when accuracy in the prediction is much required.

$$\text{Precision} = TP / (TP + FP) \qquad \text{Equation 5-2}$$

In simple words, the precision of a model is crucial when one class of the output variable has rare occurrences compared to the other class. Since PACP score of 5 has a comparatively lesser instances than other classes, it is more important for accurate prediction and therefore, precision would be a critical evaluation metric of interest during model selection.

4.2.5 Recall

The Recall can be expressed as the ratio of true positive elements to the total number of positively classified elements. Generally, recall yields the fraction of positive elements, which are correctly classified and is shown in the Equation 5-3 (Hossin and Sulaiman 2015).

$$Recall = TP / (TP + FN) \qquad \text{Equation 5-3}$$

It can be understood that the model's predictive accuracy for the positive class is measured by the recall. Recall of a model is significant to evaluate the model's ability to capture all positive elements in the dataset. For instance, when a model is trained, it should be capable to capture all the pipe segments with a PACP score of 5. Based on the discussion on precision and recall, it can be understood that both are relatively important and therefore, a new metric was introduced by combining both precision and recall.

4.2.6 F1-Score

Aggregating precision and recall into a single metric to assess classification model’s performance, F1-score was developed by estimating the harmonic mean of precision and recall, as shown in the Equation 5-4. The F1-score ranges on a scale from 0 to 1, where a value of 1 corresponds to a better performance of a model and vice versa.

$$\text{F1-Score} = 2 * \left(\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \right) \qquad \text{Equation 5-4}$$

Since the F1-score is calculated as a weighted average of precision and recall, both contribute equally, and hence, it can be used to identify the optimal trade-off between the two quantities. Based on the evaluation metrics, it is found that F1-score would be a significant metric to evaluate the performance of a developed model. However, it should be noted that all metrics discussed so far are based on binary classification confusion matrix.

In this study, the condition of pipe must be predicted among 5 different classes and binary classification cannot be employed. To evaluate a multi-class classification model, F1-score must account for all the classes and as a result, two different F1-scores were introduced: Micro F1-score and Macro F1-score (Grandini et al. 2020).

Predicted Class \ True / Actual Class	a	b	c	d	e
a	TN	FN	TN	TN	TN
b	FP	TP	FP	FP	FP
c	TN	FN	TN	TN	TN
d	TN	FN	TN	TN	TN
e	TN	FN	TN	TN	TN

Figure 5-3 Confusion Matrix for a Multi-Class Classification

To include all the classes in F1-score, multiple precision and recall were estimated for different classes from the multi-class confusion matrix. An example for a multi-class classification confusion matrix is shown in the Figure 5-3. The confusion matrix shown in the above figure consists of 5 output classes namely a, b, c, d, and e. Like a binary classification confusion matrix, metrics are estimated based on the confusion matrix by considering one class of interest at a time.

For instance, in the Figure 5-3, class b is considered as target class of interest. So, TP corresponds to number of correctly predicted class b elements. Similar to a binary confusion matrix, FP and FN correspond to incorrectly classified elements along row and column of class b, respectively. Finally, all other cells are referred to as TN. When a class of interest is switched from one to another, quantities are estimated again, and the confusion matrix cell labels are changed accordingly (Visani et al. 2020).

Micro F1-Score – Based on a multi-class confusion matrix with K number of classes, to estimate Micro F1-score and Macro F1-score, micro and macro average precision and recall quantities must be calculated. Micro average precision and recall are estimated using Equation 5-5 and Equation 5-6, respectively.

$$\text{Micro Average Precision} = \frac{\sum_{k=1}^{K} \mathrm{TP_k}}{\sum_{k=1}^{K} \mathrm{Total\ Column_k}} = \frac{\sum_{k=1}^{K} \mathrm{TP_k}}{\mathrm{Grand\ Total}} \qquad \text{Equation 5-5}$$

$$\text{Micro Average Recall} = \frac{\sum_{k=1}^{K} \mathrm{TP_k}}{\sum_{k=1}^{K} \mathrm{Total\ Row_k}} = \frac{\sum_{k=1}^{K} \mathrm{TP_k}}{\mathrm{Grand\ Total}} \qquad \text{Equation 5-6}$$

It is well known that F1-score is harmonic mean of precision and recall. Since micro average precision and recall are same, the harmonic mean of both quantities are also the same, which is the Micro F1-score (Visani et al. 2020).

$$\text{Micro Average F1-Score} = \frac{\sum_{k=1}^{K} TP_k}{\text{Grand Total}} \qquad \text{Equation 5-7}$$

By looking at the formula for Micro-F1 score in the Equation 5-7, it can be found that the formula is same as that of the accuracy. Since the calculation account the grand total of dataset, more importance will be given to majority classes. Hence it can be concluded that micro F1-score is not a choice of metric for this study.

Macro F1-Score – It is estimated by calculating the macro average precision and recall for each target class. Macro average precision and recall are directly estimated as the arithmetic mean of the same for individual classes, as shown in Equation 5-8 and Equation 5-9. Therefore, macro F1-score will be the harmonic mean of macro average precision and macro average recall, as shown in Equation 5-10 (Visani et al. 2020).

$$\text{Macro Average Precision} = \frac{\sum_{k=1}^{K} \frac{TP_k}{TP_k + FP_k}}{K} \qquad \text{Equation 5-8}$$

$$\text{Macro Average Recall} = \frac{\sum_{k=1}^{K} \frac{TP_k}{TP_k + FN_k}}{K} \qquad \text{Equation 5-9}$$

$$\text{Macro F1-Score} = 2 * \left(\frac{\text{Macro Precision} * \text{Macro Recall}}{\text{Macro Precision} + \text{Macro Recall}} \right) \quad \text{Equation 5-10}$$

From the formulas of macro average precision and recall, the numerators are composed of values in the range 0 – 1. This indicates that different sized classes are equally weighted and there is no impact of class size on the metric. In other words, minority class will have the same importance as that of the majority class. Therefore, it can be concluded that high Macro F1-score values depict that the trained model performs well across all classes, whereas low Macro F1-score values indicate that classes are poorly predicted by the trained model. Therefore, for this study, Macro F1-score could be considered as a significant metric for model's performance evaluation.

4.2.7 Summary of Evaluation Metrics

Choice of evaluation metric for a machine learning model would be based on the type of algorithm used and the expected outcome. In this study, classification of sewer pipes in structurally poor condition or a PACP score of 5 is of high importance. Since the number of pipes in poor condition is comparatively lesser in number over the other conditions, considered evaluation metric must be capable to capture the prediction performance of minority class. As a summary of discussed evaluation metrics, Table 5-1 displays various metrics based on importance for this study. Based on important evaluation metrics listed in the table, a better performing model over other models can be scrutinized.

Table 5-1 Summary of Evaluation Metrics

S. No	Evaluation Metric	Important	Not important
1	Confusion matrix	*	
2	ROC Curve	*	
3	AUC	*	
4	Accuracy		*
5	Precision	*	
6	Recall	*	
7	Micro F1-score		*
8	Macro F1-score	*	

4.3 Performance of Developed Models

The final dataset is trained with various supervised learning algorithms as discussed in the previous chapter. It is concluded that the metrics such as accuracy and micro F1-score are not effective for classifying the minority class. Therefore, the performances of trained models are compared based on the rest of the discussed evaluation metrics. Each algorithm is trained with 3 types of data namely:

1. Imbalanced dataset,
2. Under-sampled dataset, and
3. Over-sampled dataset

For any trained model with all the datasets, confusion matrices will be developed. In a developed confusion matrix, the rows represent actual class elements corresponding to PACP scores from 1 to 5, and the columns indicate the predicted class elements corresponding to PACP scores from 1 to 5. Evaluation metrics are estimated for each type of dataset under every algorithm based on respective confusion matrix. Prediction performance of three individual algorithms such as LR, k-NN, and RF, are discussed in the following section.

4.3.1 Logistic Regression

4.3.1.1 LR Imbalanced Dataset

One of the basic and most important evaluation metrics for classification methods, the confusion matrix for imbalanced logistic regression is shown. In the confusion matrix, it can be seen that the columns 4 and 5 corresponding to predicted PACP scores of 4 and 5, respectively, are all zeros. The zeros in two columns indicate that none of the data points predicted in the model belongs to PACP scores 4 and 5. This is because structurally poor condition pipes constitute only a 2.2% of total dataset, which was discussed in chapter 3 and illustrated in the Figure 3-11.

$$\text{Confusion matrix for Imbalanced LR} = \begin{bmatrix} 3575 & 177 & 12 & 0 & 0 \\ 1383 & 611 & 65 & 0 & 0 \\ 171 & 336 & 79 & 0 & 0 \\ 25 & 71 & 25 & 0 & 0 \\ 5 & 11 & 5 & 0 & 0 \end{bmatrix}$$

Based on the developed confusion matrix, true positive rate and false positive rate were estimated. ROC curve is plotted for individual PACP score prediction as shown in the Figure 5-4. Though classes 4 and 5 were not at all predicted by the model, AUC for the two classes were estimated as 0.89 and 0.80, respectively. This is because the formula to estimate the FPR and TPR includes entire elements in the confusion matrix.

Error in the prediction rate is illustrated in the Figure 5-5. The distribution shows the counts of each class misclassified by the model as other classes. The higher the misclassification in a model, the model would be unreliable. For instance, around 4,000 observations were predicted as PACP score 1 but more than 500 observations were belonged to PACP score of 2. Likewise, majority of the observations predicted as PACP score of 2 were incorrectly classified from PACP scores 1, 3, and 4. Since none of the observations were predicted as PACP 4 or 5, there is no counts for the two classes.

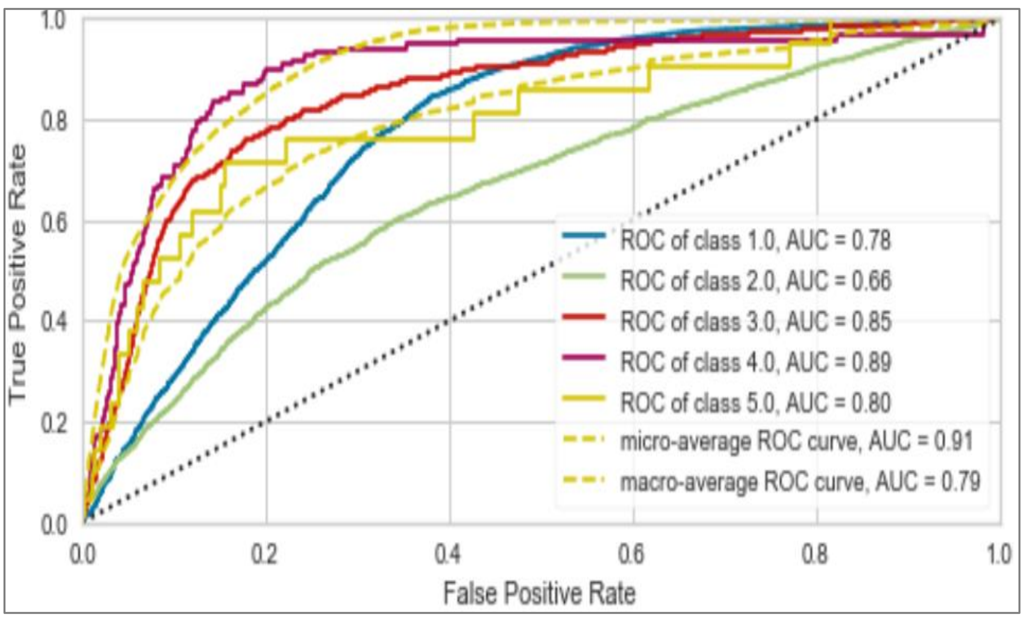

Figure 5-4 ROC Curves for Logistic Regression with Imbalanced Dataset

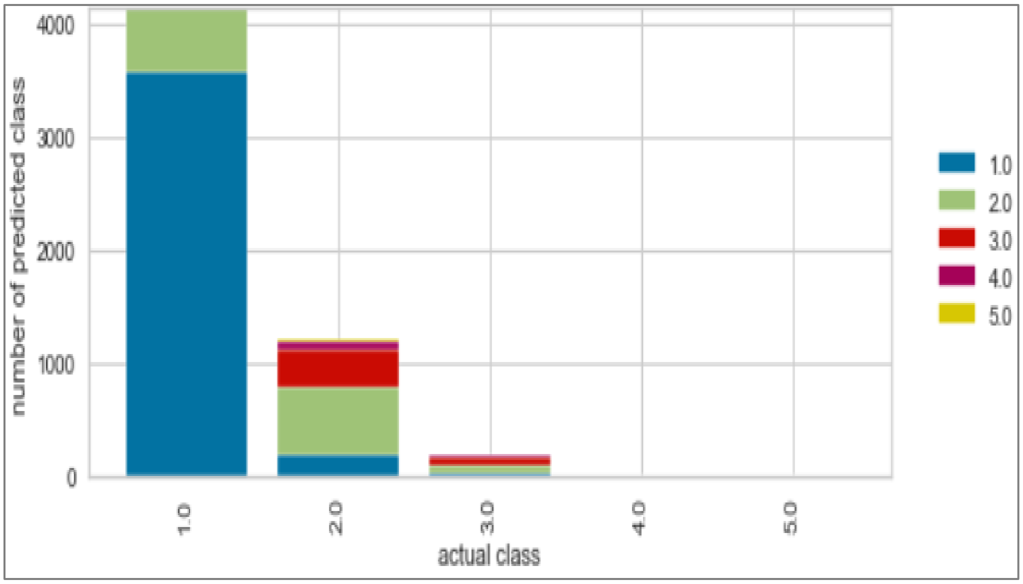

Figure 5-5 Error Prediction Rate for Imbalanced LR

Table 5-2 Precision, Recall, and F1 Metrics for Imbalanced LR

PACP Score	Macro-Precision	Macro-Recall	Macro-F1
1	0.693	0.950	0.801
2	0.507	0.297	0.374
3	0.425	0.135	0.205
4	0.000	0.000	0.000
5	0.000	0.000	0.000

Evaluation metrics such as precision, recall, and F1-score were calculated based on the confusion matrix as displayed in the Table 5-2. Since high number of observations from PACP score of 1 is correctly classified compared to observations with PACP scores 2 and 3, the precision, recall, and F1-score of PACP score 1 is greater than the other two classes. As there was no classification by the model for PACP scores 4 and 5, precision, recall, and resulting F1-score is estimated as zero.

4.3.1.2 LR Under-Sampled Dataset

Under-sampled dataset was utilized in logistic regression analysis and the obtained results are discussed in this section. In the under-sampled dataset, all 5 classes are modified to match the minority class in the dataset, which is PACP score of 5 with 95 observations. The confusion matrix for the trained LR model with under-sampled data is shown below:

$$\text{Confusion matrix} = \begin{bmatrix} 14 & 0 & 2 & 0 & 1 \\ 7 & 5 & 3 & 3 & 0 \\ 1 & 3 & 8 & 4 & 3 \\ 3 & 1 & 3 & 10 & 3 \\ 1 & 0 & 1 & 8 & 11 \end{bmatrix}$$

The pipes with PACP scores 1, 4, and 5 were correctly classified compared to PACP scores 2 and 3. This can be understood by observing the major diagonal of the confusion matrix, which is expected to have greater numbers than other cell elements. It can be seen from the confusion matrix that the model is not capable of classifying any class to a reliable extent because FP and FN for every class has integers rather than zeros.

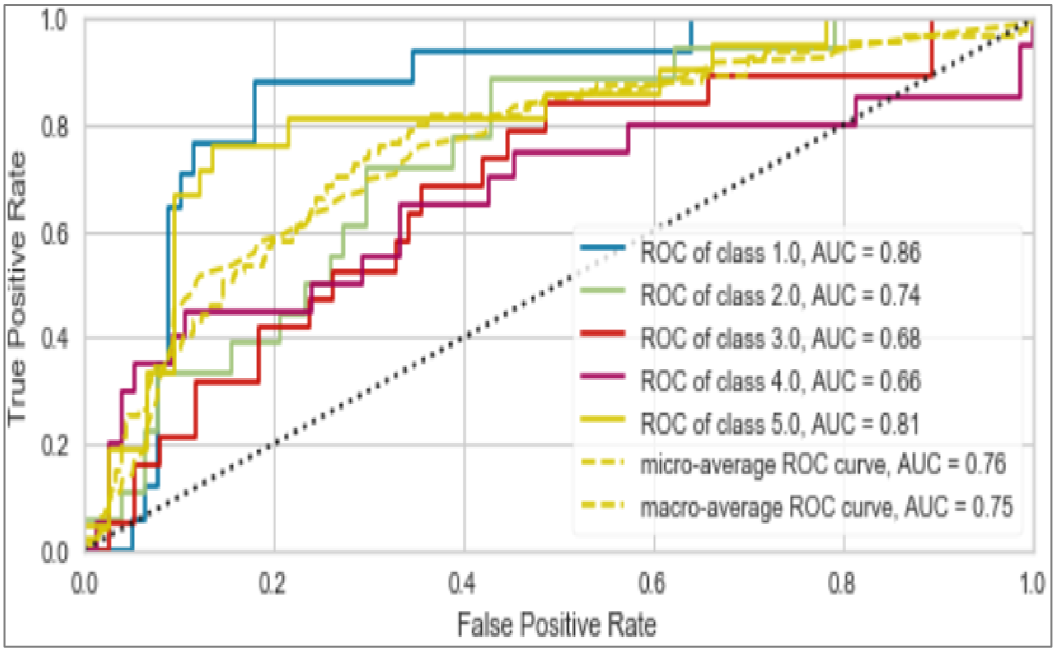

Figure 5-6 ROC Curves for LR with Under-Sampled Dataset

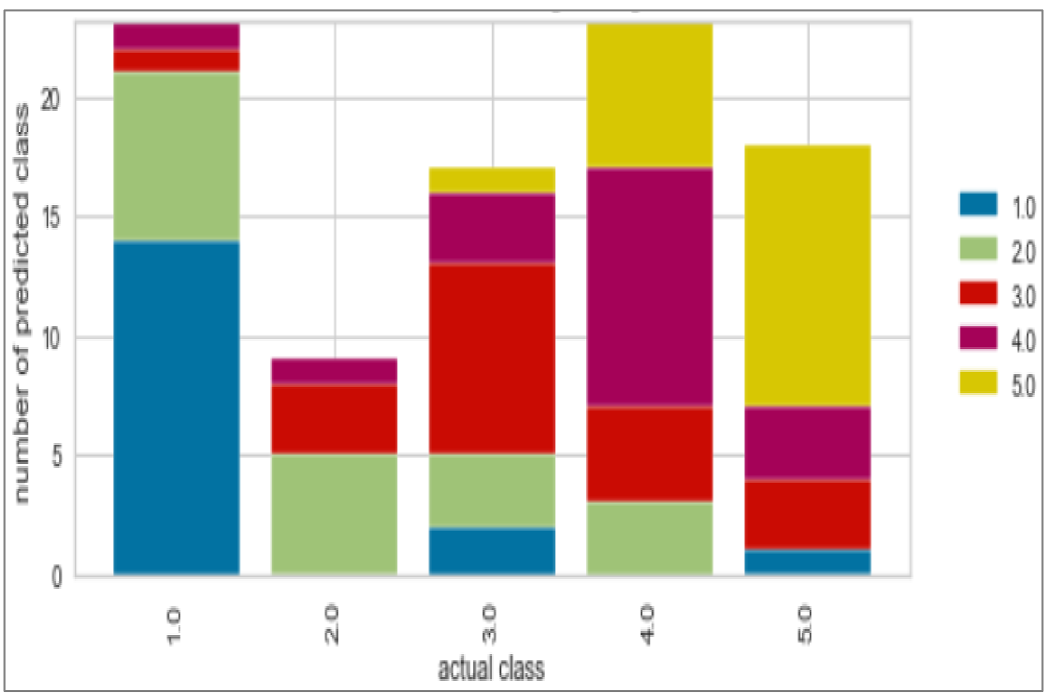

Figure 5-7 Error Prediction Rate for Under-Sampled LR

To visualize the performance of misclassification by the trained model, error prediction graph was created. Similar to LR model with imbalanced dataset, prediction of under-sampled LR model resulted in higher misclassification as well. Created error prediction bar chart is displayed in the Figure 5-7. It can be noticed that majority of the observations were classified as PACP scores 1 and 4 while most of the predictions were belonging to other classes. In addition, it can be noticed that PACP scores 1 and 5 have higher number of correctly classified instances and as a result, F1-score for these two classes is greater than other classes. Various evaluation metrics were estimated from the confusion matrix and are listed in the Table 5-3. Since none of the class has higher number of correct classifications, the precision and recall are not closer to one.

Table 5-3 Precision, Recall, and F1 Metrics for Under-Sampled LR

PACP Score	Macro-Precision	Macro-Recall	Macro-F1
1	0.538	0.824	0.651
2	0.556	0.278	0.370
3	0.471	0.421	0.444
4	0.400	0.500	0.444
5	0.611	0.524	0.564

4.3.1.3 LR Over-Sampled Dataset

Thirdly, over-sampled data was utilized to train the logistic regression and the results are discussed in this section. In over-sampled dataset, number of observations in all classes were modified to match the class with highest number of observations. All classes were made to include around 19,050 instances. Like LR models based on imbalanced and under-sampled dataset, the confusion matrix developed for the over-sampled LR model is shown below. From the confusion matrix, it can be found that false positives and false negatives are greater than the true positives of a respective class.

$$\text{Confusion matrix} = \begin{bmatrix} 2839 & 508 & 162 & 29 & 221 \\ 1455 & 1027 & 592 & 297 & 455 \\ 462 & 510 & 1041 & 1034 & 807 \\ 269 & 189 & 923 & 1356 & 1128 \\ 425 & 240 & 154 & 536 & 2391 \end{bmatrix}$$

Figure 5-8 illustrates the ROC curves for each class. ROC curves for PACP scores 1 and 5 covers larger area compared to other three classes. As a result, AUC for the 2 classes are found to be 0.87 and 0.78, respectively. However, each class comprise a high number of misclassifications resulting in higher error rate in prediction. As illustrated in Figure 5-9, majority of sewer pipes with PACP score 2 are misclassified as PACP score 1. Correspondingly, in the prediction of PACP score 5, all the other four classes constitute more than half of the total predictions as PACP score 5.

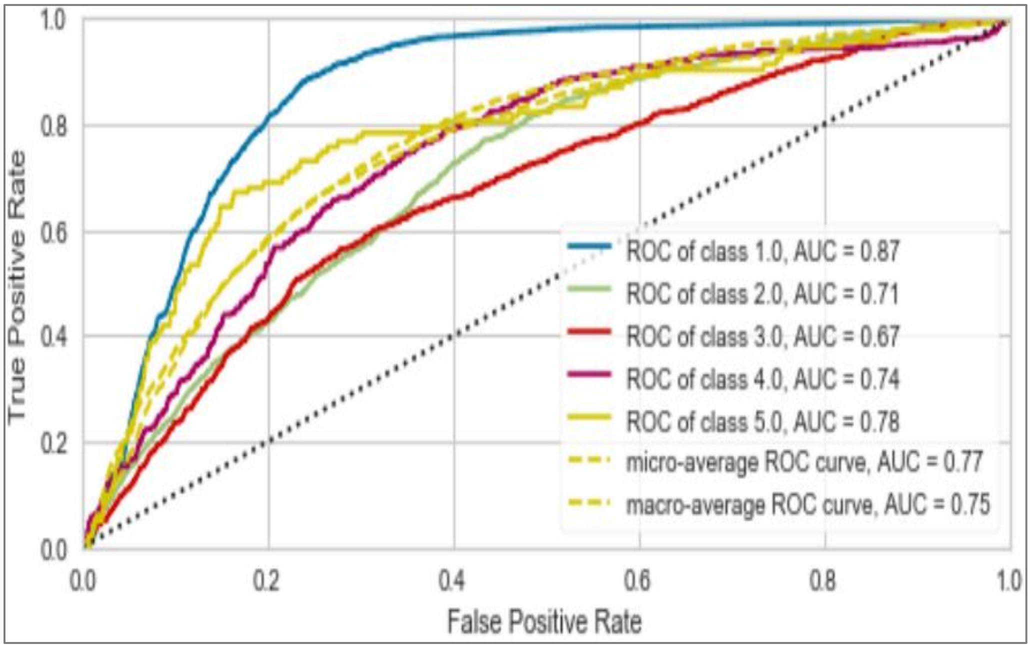

Figure 5-8 ROC Curves for LR with Over-Sampled Dataset

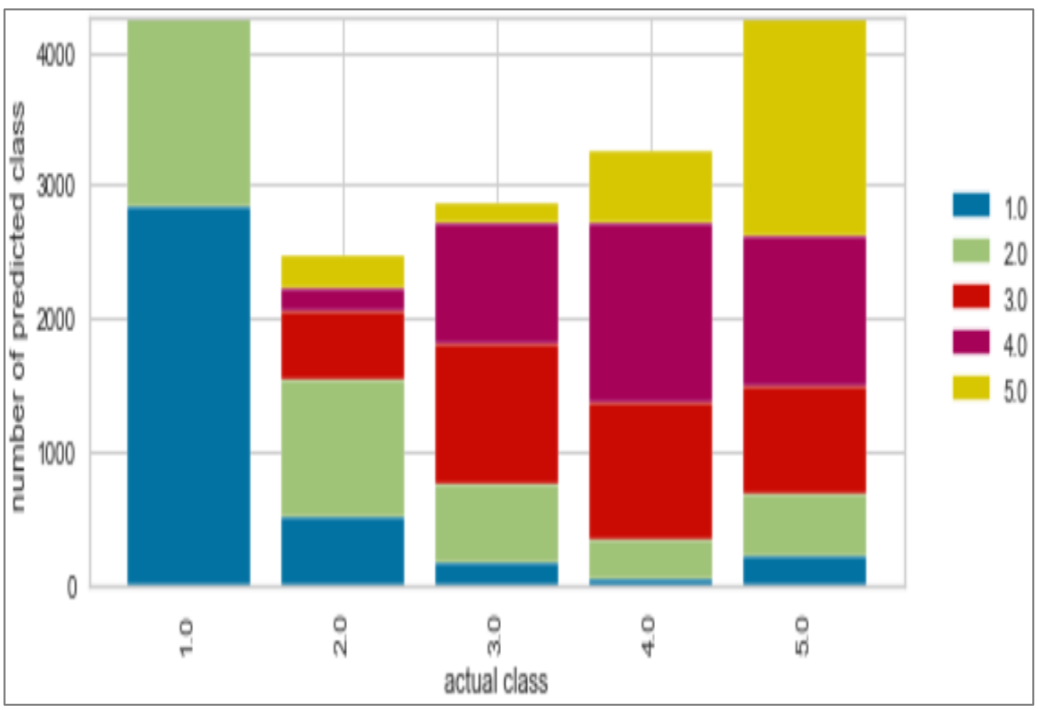

Figure 5-9 Error Prediction Rate for Over-Sampled LR

Evaluation metrics such as precision, recall, and F1-score were estimated for all 5 PACP scores. Since PACP scores 1 and 5 had relatively higher number of correctly classified observations, F-1 score is higher for the two classes than the other classes.

Table 5-4 Precision, Recall, and F1 Metrics for Over-Sampled LR

PACP Score	Macro-Precision	Macro-Recall	Macro-F1
1	0.521	0.755	0.617
2	0.415	0.268	0.326
3	0.362	0.270	0.310
4	0.417	0.351	0.381
5	0.478	0.638	0.547

4.3.1.4 Summary of LR Results

Logistic regression model based on imbalanced dataset was able to capture only PACP score 1 to a considerable extent. From the imbalanced dataset's confusion matrix, it was found that the model is inefficient to represent PACP scores 4 and 5. Under-sampled LR model performed comparatively better than the imbalanced LR model. The model resulted a maximum F1-score of 0.65 for PACP score 1 and a minimum F1-score of 0.37 for PACP score 2. Misclassification was found in all 5 classes and the model cannot be considered reliable for prediction. On the other hand, over-sampled LR model also experienced severe misclassifications in prediction.

4.3.2 k-Nearest Neighbors

Another supervised machine learning algorithm utilized in the study is k-nearest neighbors (k-NN). It is one of the simplest methods in machine learning. Similar to LR, k-NN models are also trained for 3 different datasets and the performance of models are discussed in the following sections.

4.3.2.1 k-NN Imbalanced Dataset

Imbalanced dataset is used as input features for the k-NN algorithm and confusion matrix is generated as shown. It can be seen that all predicted elements for PACP score 5 are zeros. Unlike the LR model with imbalanced dataset, PACP score 4 is represented by the k-NN model. Since the number of instances in false positives and false negatives outnumber the true positives, it can be concluded that there is a higher chance for incorrect classification than correct classification.

$$\text{Confusion matrix} = \begin{bmatrix} 3248 & 509 & 37 & 5 & 0 \\ 1094 & 832 & 106 & 4 & 0 \\ 149 & 249 & 145 & 9 & 0 \\ 30 & 64 & 40 & 8 & 0 \\ 8 & 10 & 3 & 1 & 0 \end{bmatrix}$$

Metrics such as FPR and TPR were estimated from the confusion matrix and resulting ROC curves were plotted for each PACP score, as shown in the Figure 5-10. As expected from the observation of confusion matrix, AUC of ROC curve for PACP score 5 has the least value of 0.54 while the curves for PACP scores 1 and 3 resulted an AUC value of 0.76. However, visualizing the errors in prediction would be of much importance and therefore, error rate in prediction is illustrated in the Figure 5-11. It can be observed that around 80% of the predictions as PACP score 1 is PACP 1 and the rest 20% was misclassified from PACP 2. Whereas predictions classified as PACP 2 and 3 had majority of misclassified observations from the rest of the classes. On the other hand, model did not predict any observation as PACP score 5 and henceforth, there is no error prediction rate shown for the class 5.

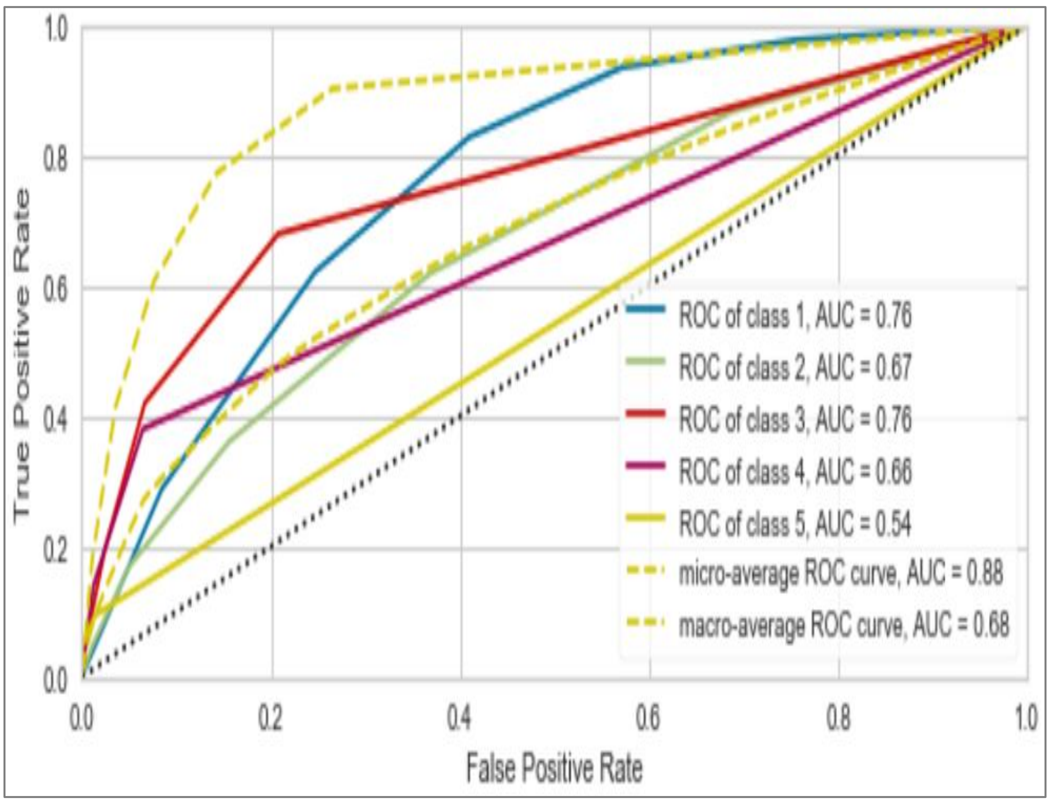

Figure 5-10 ROC Curves for k-NN with Imbalanced Dataset

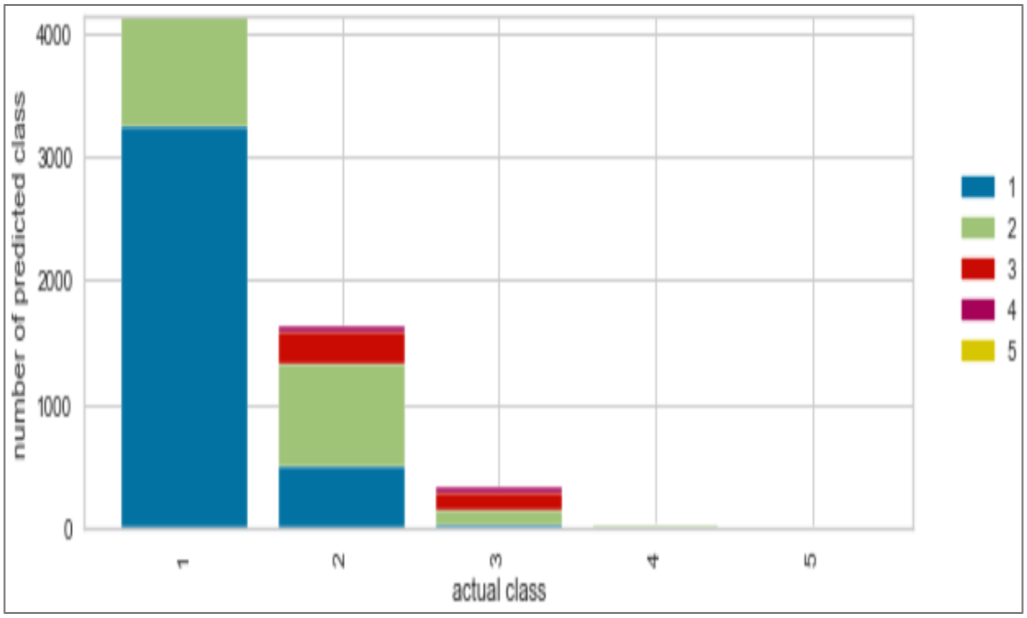

Figure 5-11 Error Prediction Rate for Imbalanced k-NN

To quantify the performance of trained model for comparison, evaluation metrics were estimated and listed as in Table 5-5. As observed in the confusion matrix and Figure 5-11, majority of class 1 or pipes under PACP score 1 category was correctly classified and hence, the F1-score for class 1 was found to be the maximum among all 5 classes with a value of 0.777.

Table 5-5 Precision, Recall, and F1 Metrics for Imbalanced k-NN

PACP Score	Macro-Precision	Macro-Recall	Macro-F1
1	0.709	0.859	0.777
2	0.503	0.399	0.445
3	0.436	0.242	0.344
4	0.250	0.058	0.094
5	0.500	0.048	0.087

It should be noted that most of predicted classification under PACP score 4 was not actually from PACP score 4, resulting in a minimal recall score and therefore, the F1-score is near zero as well. Similarly, there were no true positives predicted by the model and thus, the F1-score for PACP score 5 is the least among all classes.

4.3.2.2 k-NN Under-Sampled Dataset

Secondly, the k-NN algorithm was trained with under-sampled dataset. The results obtained from the model is discussed in this section. Primary evaluation metric for a trained model is the confusion matrix and is shown below. Almost all elements in the matrix other than the major diagonal are non-zeros resulting in high misclassification. PACP score 1 has the highest true positives followed by PACP score 2. However, PACP score 1 constitutes a significant number of false positives as well.

$$\text{Confusion matrix} = \begin{bmatrix} 8 & 3 & 6 & 0 & 0 \\ 7 & 4 & 5 & 1 & 1 \\ 3 & 6 & 3 & 2 & 5 \\ 3 & 5 & 6 & 3 & 3 \\ 3 & 5 & 6 & 5 & 2 \end{bmatrix}$$

ROC curves were plotted for all 5 classes as displayed in Figure 5-12. It is interesting to notice that almost all the curves are near the diagonal line representing an AUC value of 0.5. The AUC values of different classes range from a minimum of 0.54 to a maximum of 0.61, which indicates that there is no significant difference in model performance between classes.

For a better understanding, it is important to verify the error rates in classification and is illustrated in the Figure 5-13. It can be noticed that classes 2, 3 and 4 are misclassified in all other classes. For instance, out of 18 observations classified as PACP 1, 4 belonged to PACP 2, 3 were from PACP 3, 4 observations from PACP 4. More than half of the classified predictions were misclassified. Similarly, errors in classified predictions can be noticed, which indicates that the trained model is not reliable for prediction. In addition, the AUC values depicted that the overall performance of this model is not reliable as well.

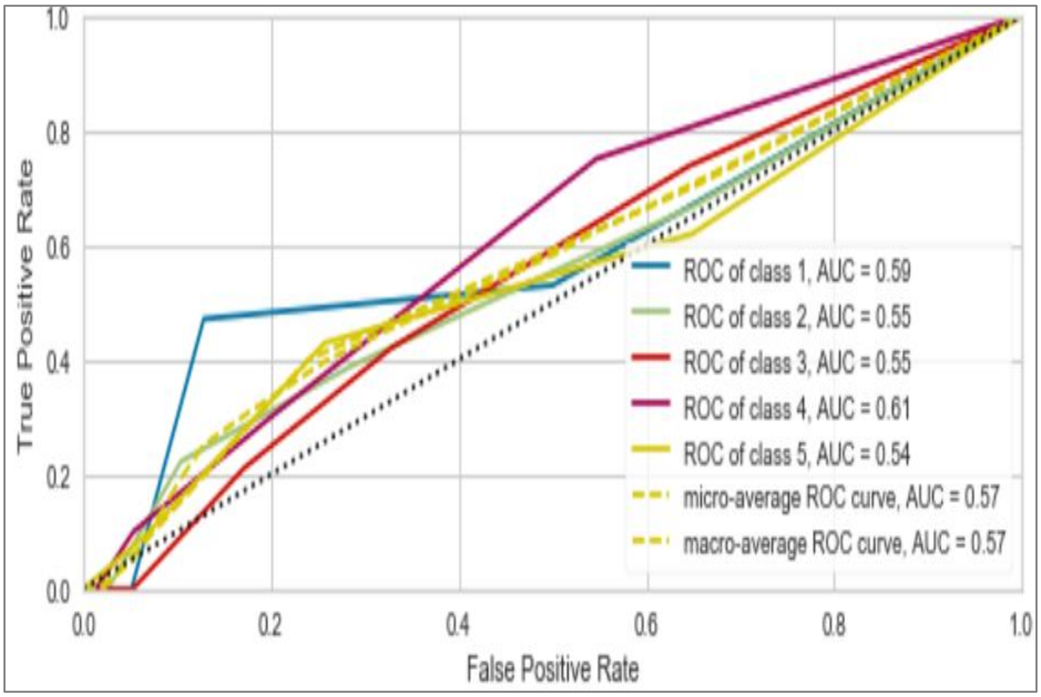

Figure 5-12 ROC Curves for k-NN with Under-Sampled Dataset

As expected, based on confusion matrix and prediction error rate, estimated metrics such as precision, recall, and F1-score were minimum. The F1-score ranged from a minimum of 0.22 to a maximum of 0.4. The maximum value of precision was 0.389 for pipes with PACP score 1, which indicates that around 39% of the data was correctly classified. In addition, the precision was less than 0.3 for all other classes leading to a conclusion that the model cannot be relied for prediction.

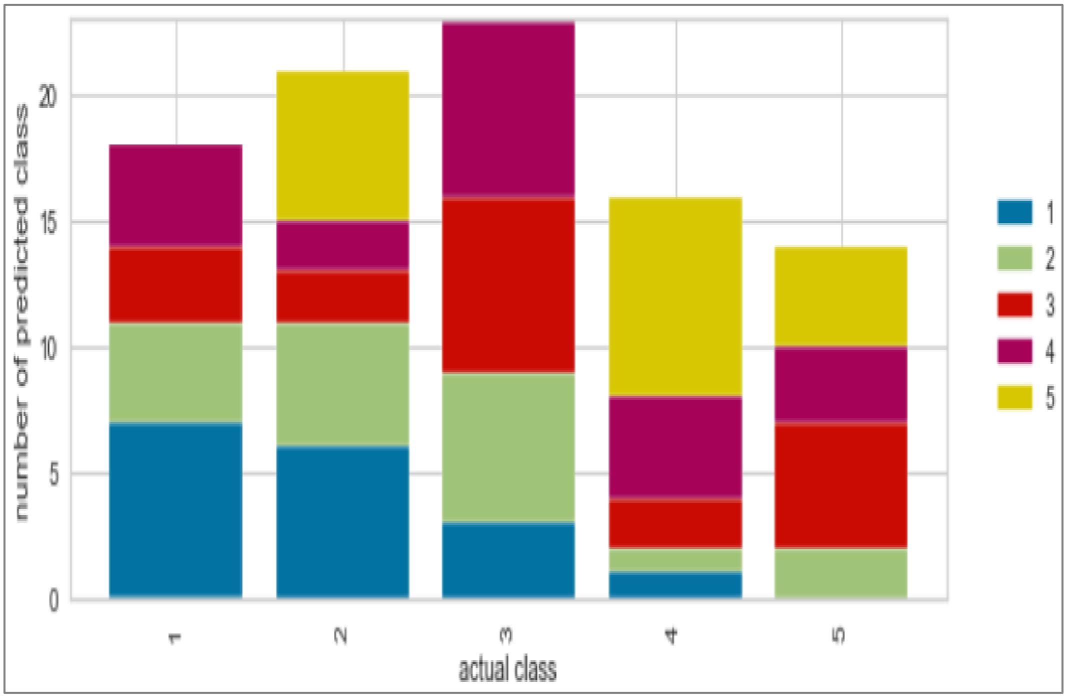

Figure 5-13 Error Prediction Rate for Under-Sampled k-NN

Table 5-6 Precision, Recall, and F1 Metrics for Under-Sampled k-NN

PACP Score	Macro-Precision	Macro-Recall	Macro-F1
1	0.389	0.412	0.400
2	0.238	0.278	0.256
3	0.269	0.368	0.311
4	0.250	0.200	0.222
5	0.286	0.190	0.229

4.3.2.3 k-NN Over-Sampled Dataset

In this section, performance of k-NN algorithm trained with over-sampled dataset is discussed. In the over-sampled dataset, number of instances in different classes are matched to the majority class. Hence, all classes in an over-sampled dataset contain around 19,000 observations. Confusion matrix was generated based on the trained k-NN model as shown below.

$$\text{Confusion matrix} = \begin{bmatrix} 2477 & 963 & 290 & 45 & 10 \\ 889 & 2267 & 531 & 95 & 15 \\ 46 & 105 & 3650 & 44 & 8 \\ 0 & 0 & 0 & 3866 & 0 \\ 0 & 0 & 0 & 0 & 3749 \end{bmatrix}$$

It is interesting to notice that there is a significant change in performance of the model with over-sampled data compared to imbalanced and under-sampled dataset. In the confusion matrix, almost all cells under the major diagonal are zeros, which is a good indication that the model is performing better. However, it should be noted that the cells above the major diagonal are all non-zeros, which are misclassified observations.

To better understand the performance of the k-NN model, ROC curves were plotted as shown in the Figure 5-14. AUC values for curves corresponding to PACP scores 3, 4, and 5, are almost unity, which indicates that the model is highly accurate at predicting these classes. AUC for classes 1 and 2 were found to be 0.90 and 0.91, respectively. The reason for the difference can be understood by examining the classification error prediction rates illustrated in the Figure 5-15.

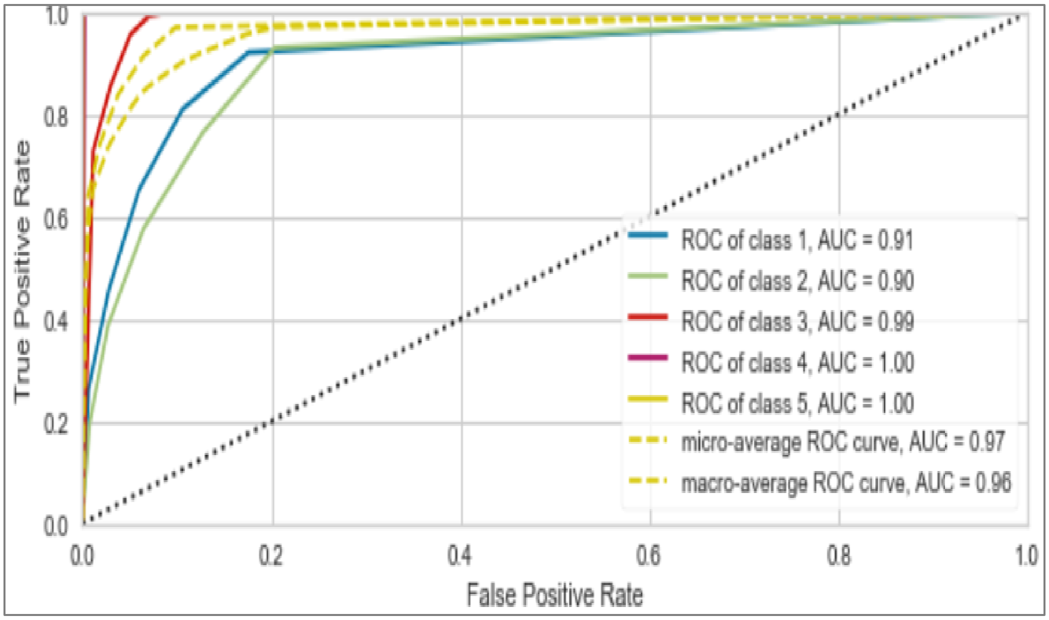

Figure 5-14 ROC Curves for k-NN with Over-Sampled Dataset

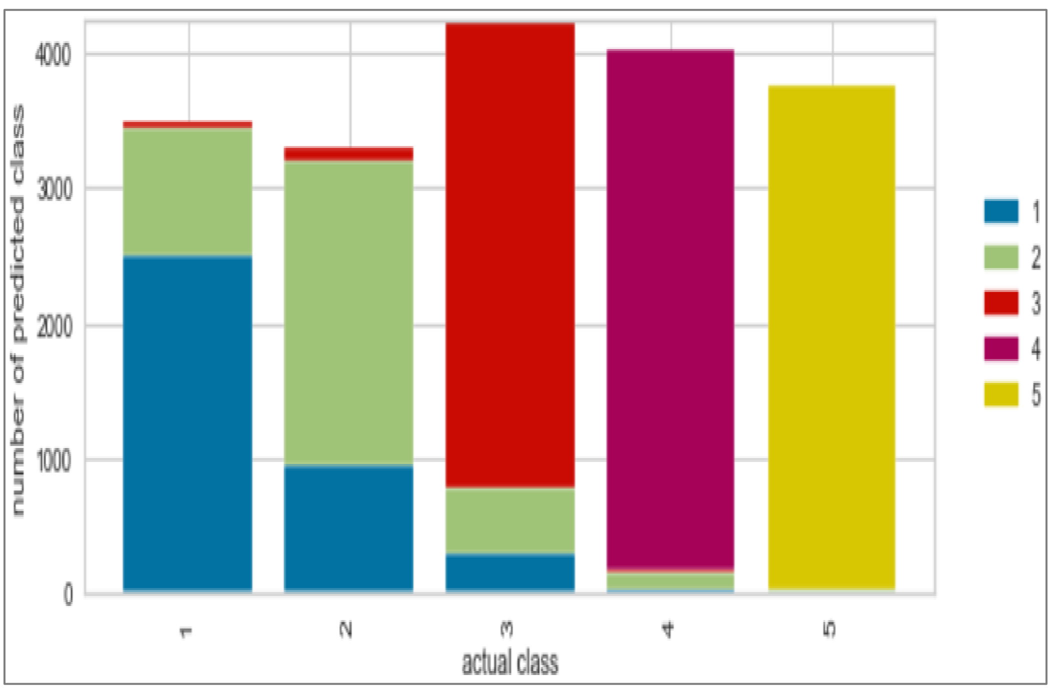

Figure 5-15 Error Prediction Rate for Over-Sampled k-NN

As observed in the Figure 5-15, PACP scores 1 and 2 are majorly misclassified among each other than any other class in the graph. This could be the reason for lower AUC compared to AUC values of classes 3, 4, and 5. However, various metrics are estimated from the confusion matrix to evaluate the performance of the trained model and are listed in the Table 5-7. It can be noticed that class 5 do not have any misclassified observation in it and as a result, the precision and recall are near unity. It is interesting to notice the F1-score of 0.68 for PACP score 1 even though the AUC was 0.91 for the same. The importance of different metrics to evaluate the performance of a model can be understood.

Table 5-7 Precision, Recall, and F1 Metrics for Over-Sampled k-NN

PACP Score	Macro-Precision	Macro-Recall	Macro-F1
1	0.715	0.664	0.688
2	0.684	0.589	0.633
3	0.827	0.957	0.887
4	0.957	1.000	0.978
5	0.995	1.000	0.997

4.3.2.4 Summary of k-NN Results

It was interesting to notice that the k-NN model with over-sampled dataset had a better overall performance compared to the imbalanced and under-sampled k-NN models. PACP scores 1 and 2 were found to be misclassified within each other. However, F1-scores for classes 4 and 5 were found to be near unity, which indicates that the model is much reliable in predicting the sewer pipes in structurally poor condition.

4.3.3 Random Forests (RF)

Finally, another most commonly employed supervised machine learning technique, random forests (RF) algorithm is used in this study. RF is considered a powerful algorithm that can fit complex datasets (Géron 2017). Visualizing options of RF makes it an effective and easily interpretable machine learning algorithm. As discussed earlier, the model is trained based on sequential if/else questions. Similar to other two methods, RF is also trained with three datasets and the obtained results are discussed in this section.

4.3.3.1 RF Imbalanced Dataset

One of the primary evaluation metrics for the trained RF model with imbalanced dataset, the confusion matrix is shown below. Like the results from other two models, almost all classes have considerable false positives and false negatives. Interestingly, there are some observations predicted as PACP score 5 by the RF imbalanced dataset model while the other two models did not classify any observation as PACP score 5.

$$\text{Confusion matrix} = \begin{bmatrix} 3388 & 368 & 41 & 2 & 0 \\ 737 & 1176 & 117 & 6 & 0 \\ 75 & 236 & 222 & 18 & 1 \\ 20 & 40 & 53 & 27 & 2 \\ 5 & 8 & 5 & 3 & 1 \end{bmatrix}$$

ROC curves were plotted for individual class to compare the performance of model for each class prediction as shown in the Figure 5-16. Unlike the other two models with imbalanced dataset, AUC values for all classes are found to be higher than 0.75, which indicate that the model has a better performance than other two models. However, each class is constituted by a considerable total of misclassified observations, which must be examined.

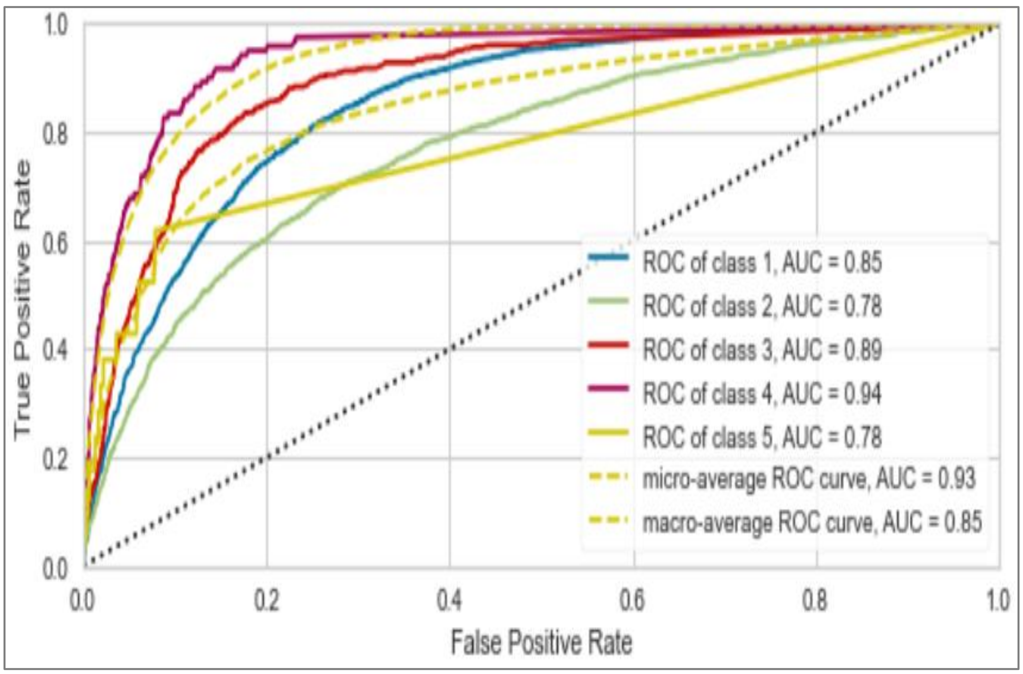

Figure 5-16 ROC Curves for RF with Imbalanced Dataset

To examine individual class prediction performance, errors in prediction rates were estimated and illustrated in the Figure 5-17. It can be observed that misclassification of PACP 2 as 1 is relatively lesser than other two models. PACP scores 2 and 3 are significantly misclassified within each other. However, it should be noted that PACP score 5 predictions are accurate without any misclassification. In addition, prediction of PACP score 4 is found to have a minimal misclassification. However, other evaluation metrics must be considered before concluding the performance of a model.

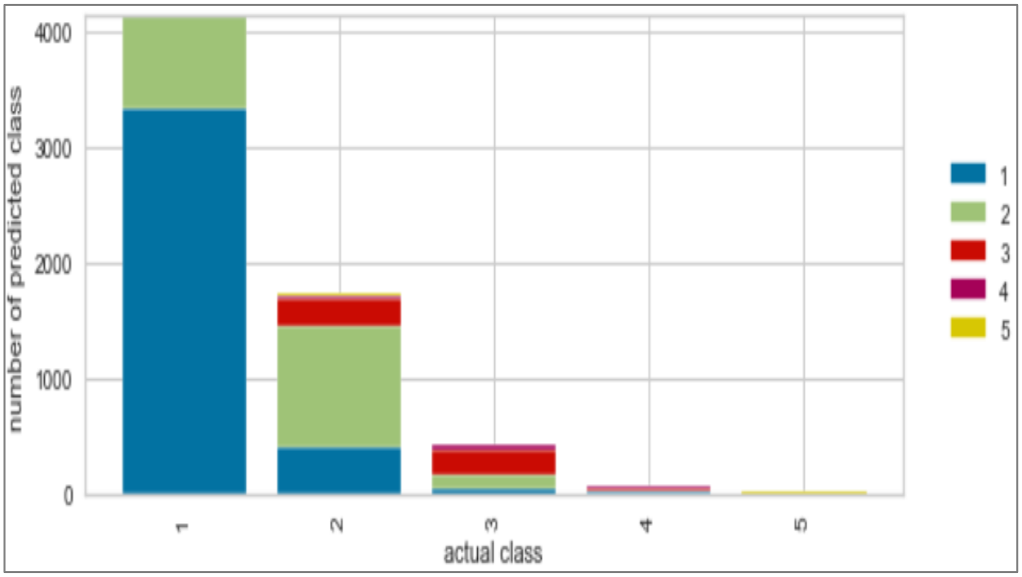

Figure 5-17 Error Prediction Rate for Imbalanced RF

Table 5-8 Precision, Recall, and F1 Metrics for Imbalanced RF

PACP Score	Macro-Precision	Macro-Recall	Macro-F1
1	0.772	0.886	0.825
2	0.617	0.518	0.563
3	0.517	0.382	0.440
4	0.382	0.215	0.275
5	0.250	0.048	0.080

Estimated metrics such as precision, recall, and F1-scores for imbalanced RF model is displayed in the Table 5-8. It can be seen that class 1 has a higher F1-score of 0.825. Even though PACP score 5 had no misclassification, the F1-score was estimated as 0.08, indicating that the model is unreliable. This phenomenon is because of the minimal recall value of 0.048, which is because of significantly higher total of false negatives compared to true positives.

4.3.3.2 RF Under-Sampled Dataset

The RF algorithm is trained with the under-sampled dataset and the obtained results are discussed in this section. As a primary evaluation metric, generated confusion matrix is shown below.

$$\text{Confusion matrix} = \begin{bmatrix} 13 & 3 & 0 & 0 & 1 \\ 6 & 5 & 4 & 0 & 3 \\ 2 & 2 & 6 & 6 & 3 \\ 1 & 1 & 10 & 6 & 2 \\ 2 & 3 & 2 & 7 & 7 \end{bmatrix}$$

It was interesting to notice that all classes have considerable observations. However, there are a major number of observations in false positives and false negatives. ROC curves for all classes were plotted based on FPR and TPR estimated from the confusion matrix as shown in the Figure 5-18. ROC curve of class 1 covers larger area whereas ROC curves of PACP scores 3 and 4 covers a relatively smaller area. Hence, the AUC for PACP 1 is higher than that of the other classes. However, the false positives and false negatives in the confusion matrix must be accounted for the performance evaluation of the model. Therefore, the errors in prediction rate for different classes are examined as shown in the Figure 5-19.

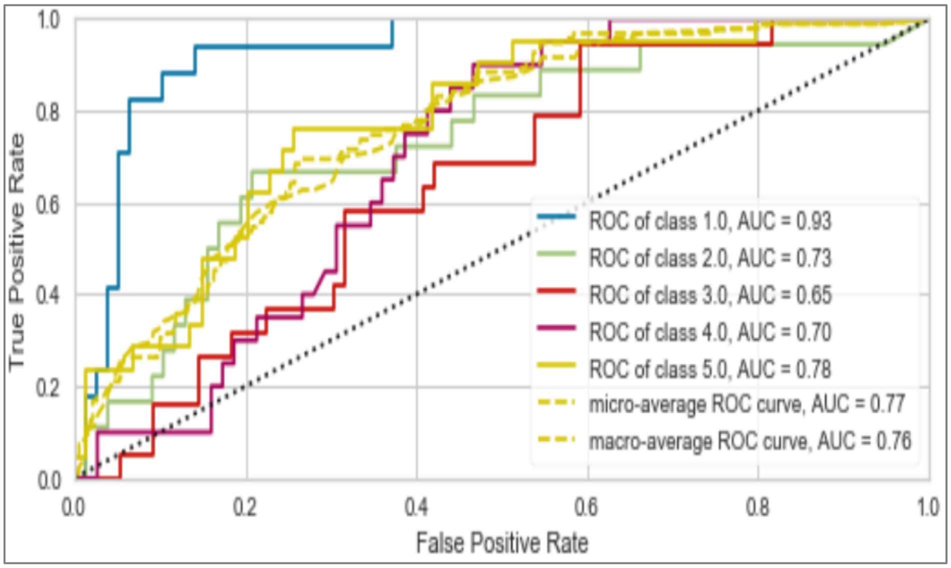

Figure 5-18 ROC Curves for RF with Under-Sampled Dataset

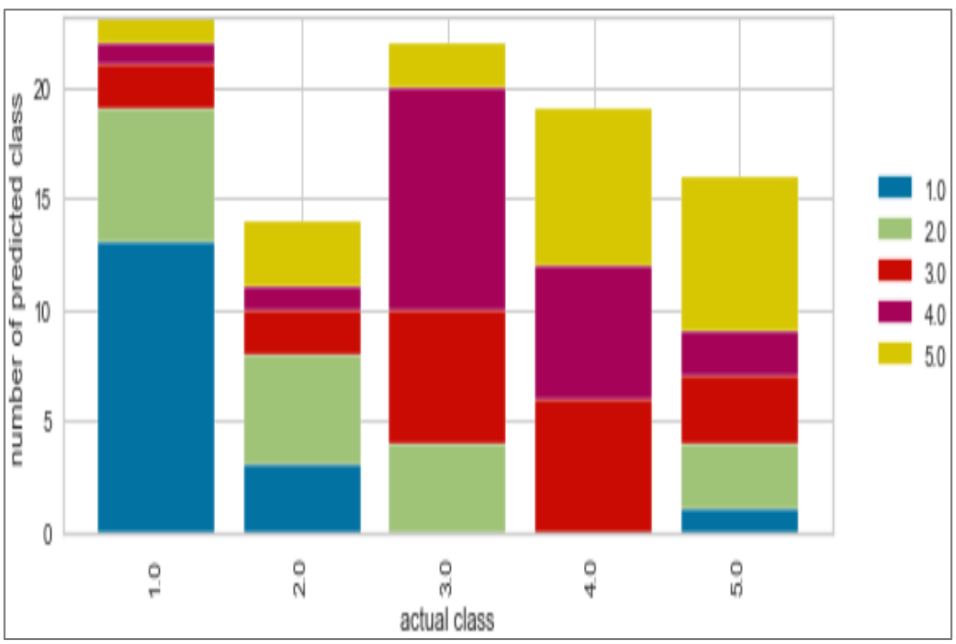

Figure 5-19 Error Prediction Rate for Under-Sampled RF

It can be seen in the error prediction chart that all the predicted classes are contributed by major misclassifications. None of the predicted classes have a majority of correct classification. Considering PACP score 3 for instance, 6 observations from PACP 3 were correctly classified as 3 but, 10 observations and 2 observations from PACP scores 4 and 5, respectively, were predicted as PACP 3 as well. Similarly, 7 instances from PACP score 5 was predicted correct while 9 instances from other 4 classes were misclassified as PACP 5. Therefore, better insight on other metrics is required to quantify the performance evaluation.

Metrics such as precision, recall, and corresponding F1-scores were estimated for all classes and are shown in the Table 5-9. As observed in the error prediction chart for PACP score 1, majority of the pipes with PACP score 1 was correctly classified as 1 and hence, the F1-score is higher than the other 4 classes. Majority of the predicted PACP score 3 classifications were misclassified from other classes and therefore, the F1-score of class 3 is the minimal.

Table 5-9 Precision, Recall, and F1 Metrics for Under-Sampled RF

PACP Score	Macro-Precision	Macro-Recall	Macro-F1
1	0.542	0.765	0.634
2	0.357	0.278	0.313
3	0.273	0.316	0.293
4	0.316	0.300	0.308
5	0.438	0.333	0.378

4.3.3.3 RF Over-Sampled Dataset

Finally, the RF algorithm was trained with the over-sampled data in which number of observations in all classes are matched to the class with higher number of observations, which is PACP score 1 with 19,050 observations. Results of the RF model trained with over-sampled dataset is discussed in this section. As a basic evaluation metric, confusion matrix was generated and is shown below.

$$\text{Confusion matrix} = \begin{bmatrix} 3204 & 500 & 73 & 7 & 1 \\ 348 & 3336 & 112 & 1 & 0 \\ 7 & 5 & 3841 & 0 & 0 \\ 0 & 0 & 0 & 3866 & 0 \\ 0 & 0 & 0 & 0 & 3749 \end{bmatrix}$$

It was interesting to notice in the confusion matrix that majority of elements other than the major diagonal are zeros, which is a good indication for a better performing model. Especially with the PACP score 5 predictions, only one observation was found to be misclassified from other classes. Similarly, considering PACP score 4 predictions, only a few observations are misclassified as well. The total false positives and false negatives for the classes 4 and 5 are almost negligible, which indicates that the model is capable to correctly predict all the pipes in structurally poor condition. However, PACP scores 1, 2, and 3 were found to have minor misclassifications and hence, their performance can de compared using ROC curves.

ROC curves for all classes were plotted as shown in the Figure 5-20. It can be seen that the AUC for ROC curves of PACP scores 3, 4, and 5, are unity. ROC curves of PACP scores 1 and 2 are visible because of minor misclassification observed in the confusion matrix. However, as discussed earlier, error in prediction rate and other metrics such as precision, recall, and F1-scores must be estimated to evaluate the performance of a prediction model in detail.

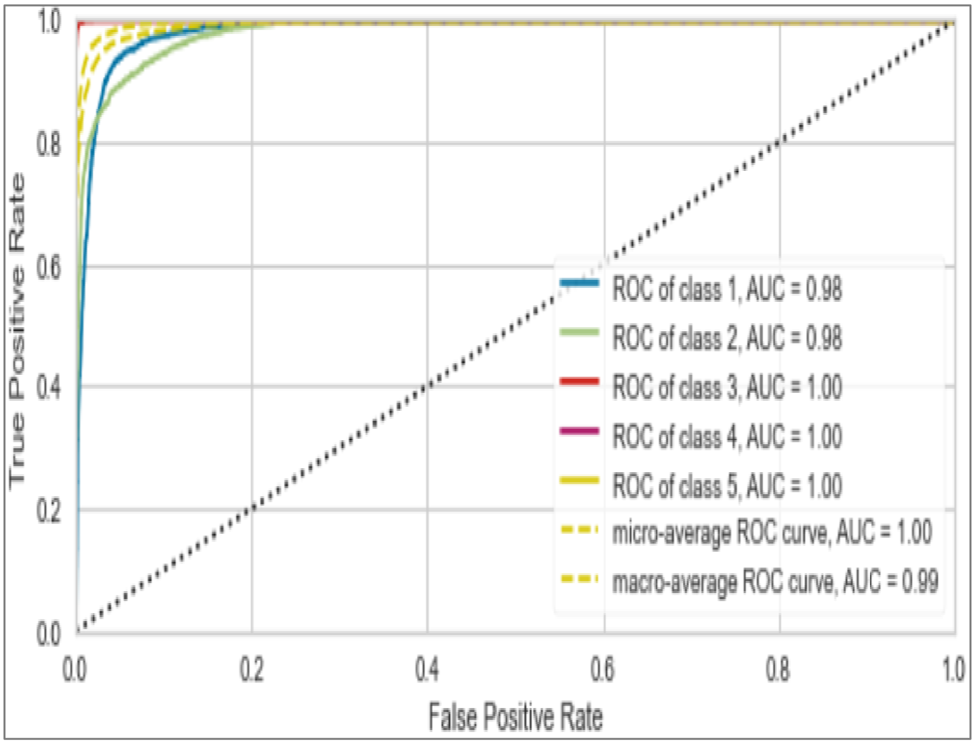

Figure 5-20 ROC Curves for RF with Over-Sampled Dataset

Errors in classified predictions are illustrated in the Figure 5-21. Classes 1 and 2 were found to have minor misclassifications within each other, and PACP score 3 had a minor misclassification from classes 1 and 2 as well. However, it should be noticed that PACP scores 4 and 5 had zero misclassification from other classes in their predictions.

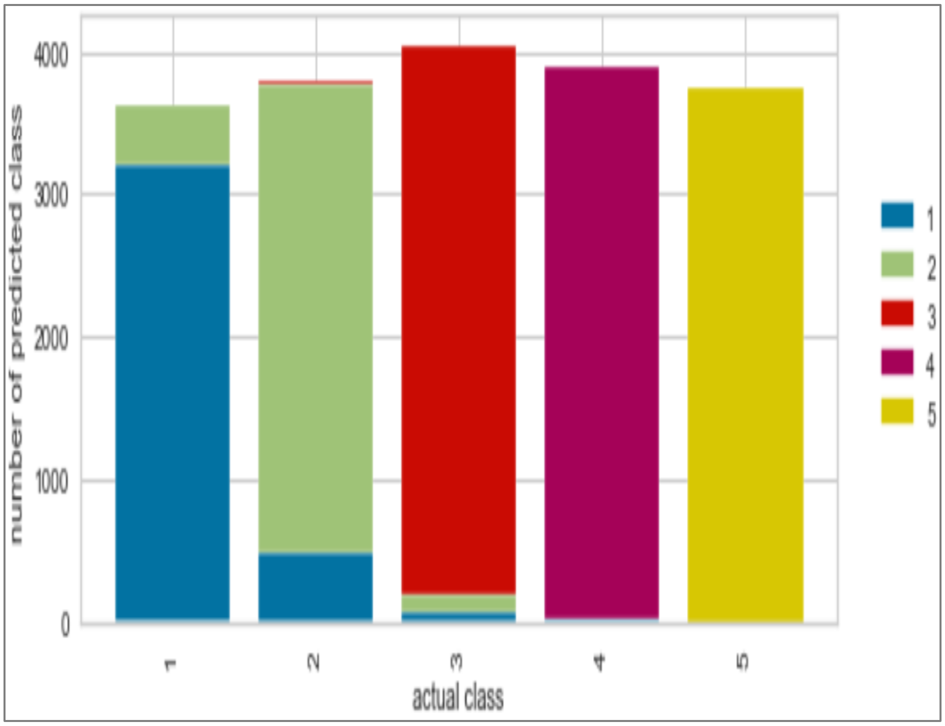

Figure 5-21 Error Prediction Rate for Over-Sampled RF

As discussed in earlier chapters, performance of developed models can be compared based on evaluation metrics. Therefore, various metrics were calculated and are listed in the Table 5-10. It is obvious from the confusion matrix that the evaluation metrics would yield better scores compared to the imbalanced and under-sampled dataset models. As expected, based on confusion matrix and error in prediction chart, precision, recall, and F1-scores for PACP 4 and 5 were found to be almost unity, which indicates that the model could perform better than the models trained with imbalanced and under-sampled datasets. The F1-scores for PACP scores 1 and 2 are around 0.87, which indicate that the model is reliable for future prediction.

Table 5-10 Precision, Recall, and F1 Metrics for Over-Sampled RF

PACP Score	Macro-Precision	Macro-Recall	Macro-F1
1	0.888	0.853	0.870
2	0.871	0.860	0.866
3	0.955	0.998	0.976
4	0.995	1.000	0.998
5	1.000	1.000	1.000

4.3.3.4 Summary of RF Results

The RF model is trained with all 3 kinds of datasets to develop a condition prediction model and respective results were discussed in previous sections. RF model with imbalanced dataset couldn't successfully classify the pipes in poor condition and the F1-scores were found to be 0.08 and 0.275 for PACP scores 5 and 4, respectively. Though the F1-scores of PACP scores 4 and 5 were improved to 0.308 and 0.378, respectively, in under-sampled dataset trained RF model, it is not reliable because of high misclassification rate. It was found that the trained RF model with over-sampled dataset performed better than the other two models. Based on the errors in prediction and F1-scores, it can be concluded that the RF model trained with over-sampled dataset would be a reliable model to accurately predict the sewer pipes in structurally poor condition.

4.4 Chapter Summary

Primarily, various evaluation metrics to compare the performance of different machine learning models are discussed in detail. Relevant metrics for reliable prediction of pipes in poor condition are scrutinized. Secondly, machine learning algorithms such as logistic regression, k-nearest neighbors, and random forests are trained with imbalanced, under-sampled, and over-sampled datasets, individually. Each method yielded three different results and evaluation metrics for all 9 results were calculated.

Based on the confusion matrix developed for all 3 methods, it was found that the models trained with imbalanced dataset failed to classify structurally poor condition pipes. Though the methods trained with under-sampled dataset was able to classify the pipes with PACP score 5, there was a considerable amount of misclassification that resulted the models to be unreliable. It was found that all 3 methods relatively performed better than imbalanced when trained with over-sampled dataset.

CHAPTER 6

RESULTS AND CONCLUSIONS

6.1 Introduction

In the previous chapters, collected sanitary sewer dataset was processed and prepared for training with various supervised machine learning algorithms. As discussed in chapter 4, different algorithms were selected for the required classification output in this study. As a result of training various algorithms, their performances were discussed in detail in the previous chapter. In this chapter, a machine learning model will be selected as a suitable model for prediction application.

It should be noted that the evaluation metrics were discussed based on individual class for each trained algorithm. It is possible that the performance of one model to predict class 5 would be better than the other model while the same model would perform poor to predict another class and vice versa. Therefore, there is a necessity to average the metrics of all 5 classes to compare the performance of different models. Hence, evaluation metrics from different algorithm results such as precision, recall, and F1-scores, were averaged for all 5 classes.

The following sections will discuss the performances of all three algorithms based on 3 different datasets namely imbalanced dataset, under-sampled dataset, and over-sampled dataset.

6.2 Results from Logistic Regression

Results from trained logistic regression models are illustrated in the Figure 6-1. It can be seen that the F1-score was found to be a maximum of 0.49 for under-sampled dataset and a minimum of 0.28 for the imbalanced dataset. Likewise, precision and recall were also found to be maximum in under-sampled dataset compared to imbalanced and over-sampled datasets. It can be concluded from the figure that under-sampled dataset trained model performs better than the other two. However, the precision score of 0.52 indicates that the model is capable to correctly predict the given data to an extent of 52%.

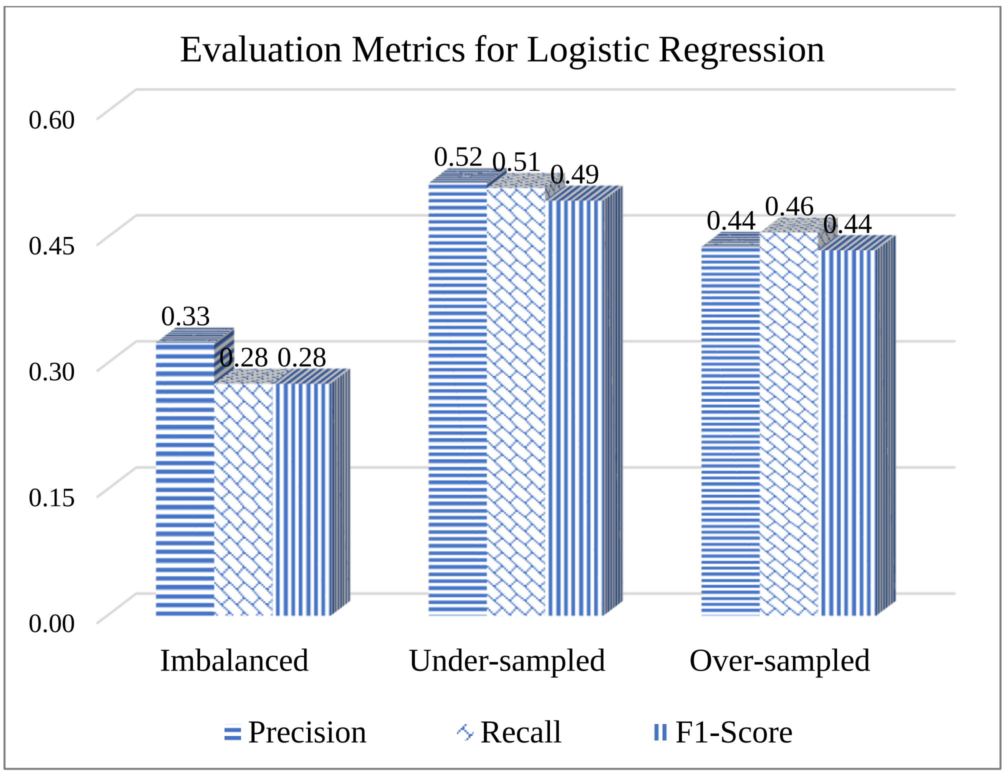

Figure 6-1 Summary of Results from Logistic Regression

6.3 Results from k-Nearest Neighbors

Obtained results from k-NN trained models are illustrated in the Figure 6-2. Unlike the logistic regression results, the over-sampled dataset trained model was found to perform better. It was found that the trained model with under-sampled dataset resulted a poor performance compared to imbalanced and over-sampled dataset trained models. The maximum F1-sccore was found to be 0.83 from over-sampled dataset. Recall score of 0.84 indicates that 84% of the given test data would be correctly captured by the model. Though the F1-score of 0.83 is considered reliable, it could be further improved.

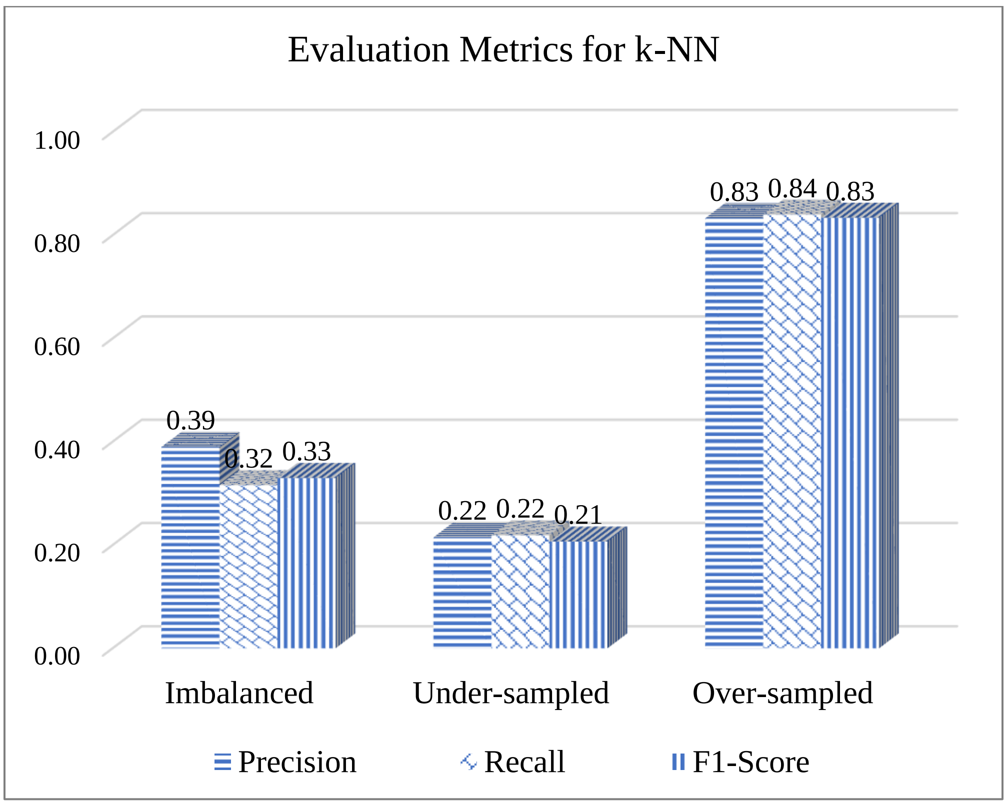

Figure 6-2 Summary of Results from k-NN

6.4 Results from Random Forests

As illustrated in the Figure 6-3, it was found that the results of random forests models outperformed both logistic regression and k-NN models. Precision and recall scores of 0.94 indicates that the model could correctly predict the given data and almost all correct observations would be predicted by the model. Interestingly, the F1-score was found to be a maximum of 0.94 for over-sampled dataset, which indicates that the model is greatly reliable. It should also be noted that the results from imbalanced dataset and under-sampled dataset did not differ in great extent as seen in other two algorithms.

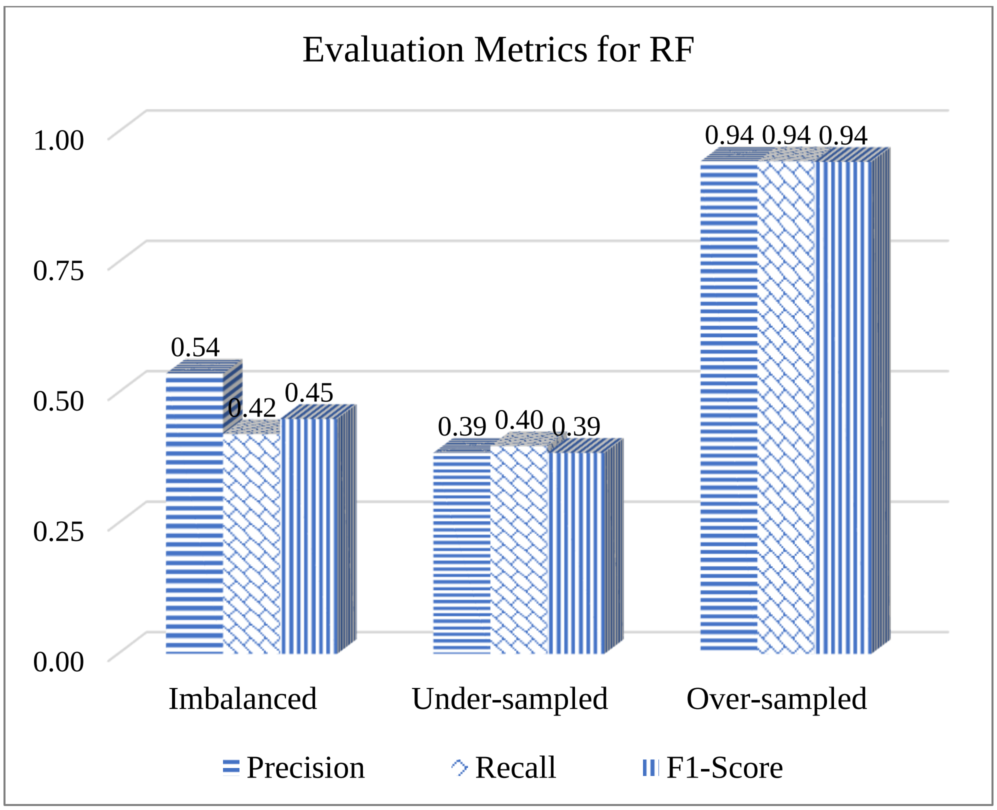

Figure 6-3 Summary of Results from RF

6.5 Area Under the Curve (AUC) and F1-scores Comparison

Other critical parameters in evaluating a prediction model such as AUC and F1-scores are compared between various algorithms in this section. As discussed in earlier sections, the AUC would provide an overall performance of a trained model. Resulted AUCs for various algorithms are compared as shown in the Figure 6-4 and RF models were found to perform better than that of LR and k-NN. However, it should be noted that the effect of misclassification is not accounted by the AUC and hence, F1-score is considered for model evaluation.

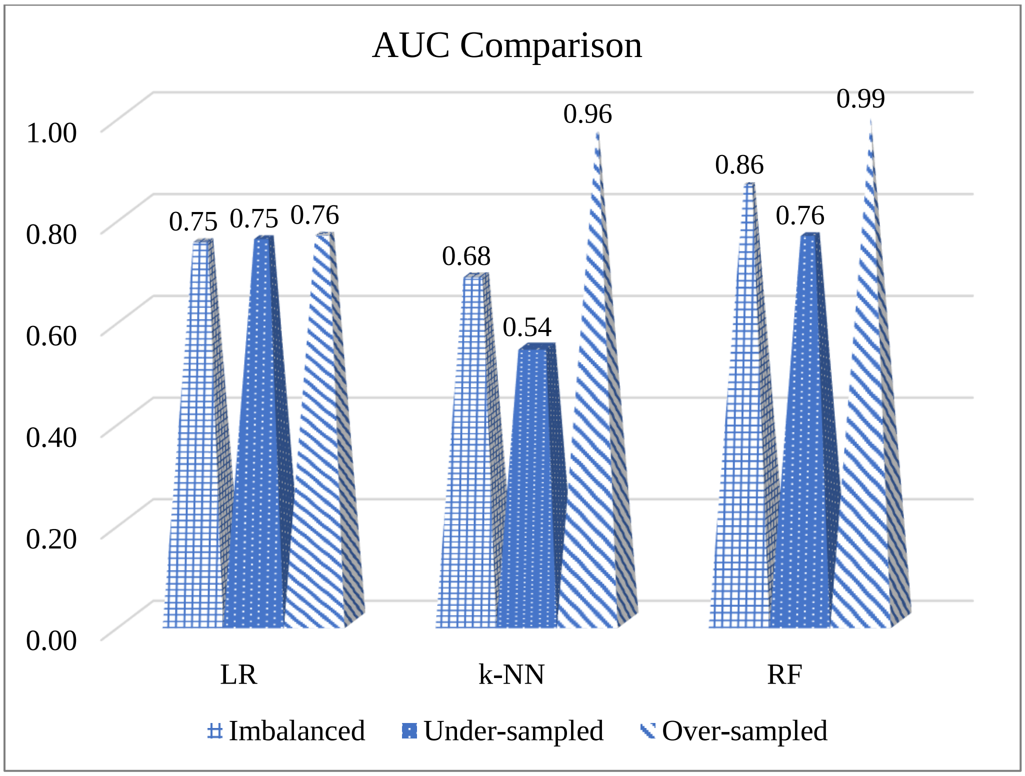

Figure 6-4 AUC Comparison between LR, k-NN, and RF

As a part of evaluating the trained models, F1-score was compared as shown in the Figure 6-5. F1-score was observed to drastically improve from 0.44 with LR model to 0.94 with RF model for the over-sampled dataset. The F1-score could be considered as an important metric to evaluate the performance of a model since it includes both precision and recall as discussed earlier.

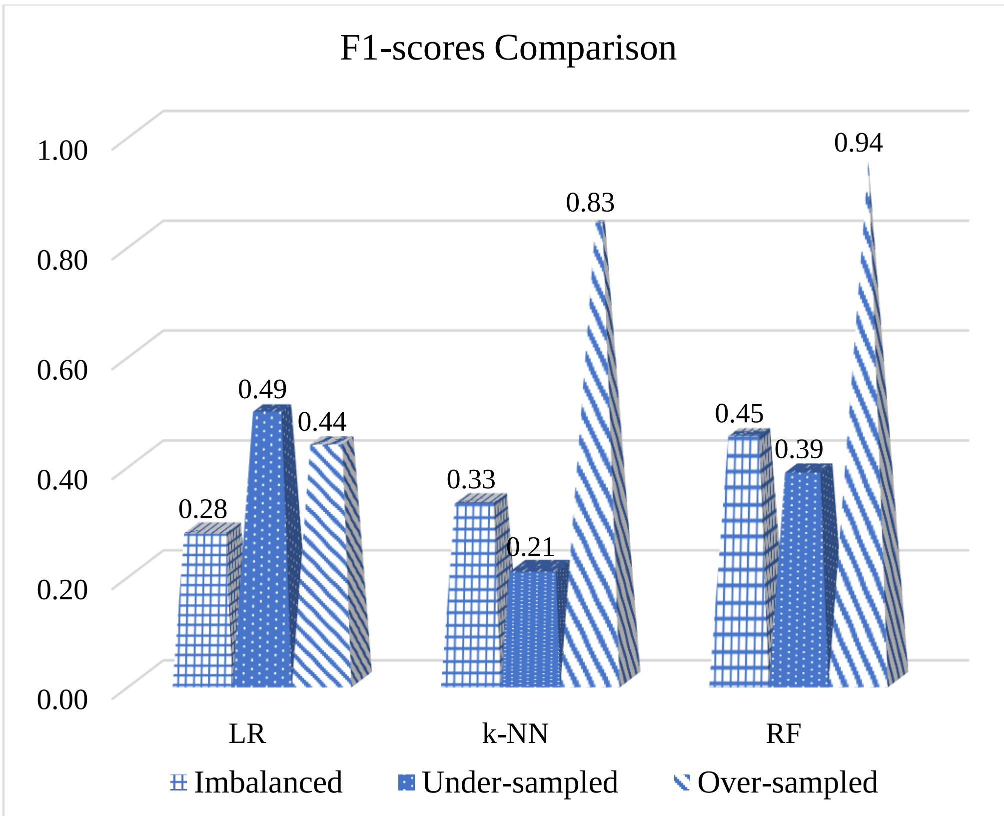

Figure 6-5 F1-score Comparison between LR, k-NN, and RF

Based on the evaluation metrics discussed from results of various trained models, it can be concluded that the over-sampled RF model perform better than the other two models. The trained RF model is capable to correctly predict every observation from each class with high reliability.

CHAPTER 7

PRACTICAL APPLICATIONS

To effectively utilize the developed condition prediction model, a decision-support tool was developed. Developed web-based application as shown in the Figure 7-1 utilizes an open-source library from Python, called "Streamlit". Application file can be shared with the utility owners to estimate the condition of uninspected pipes in their inventory.

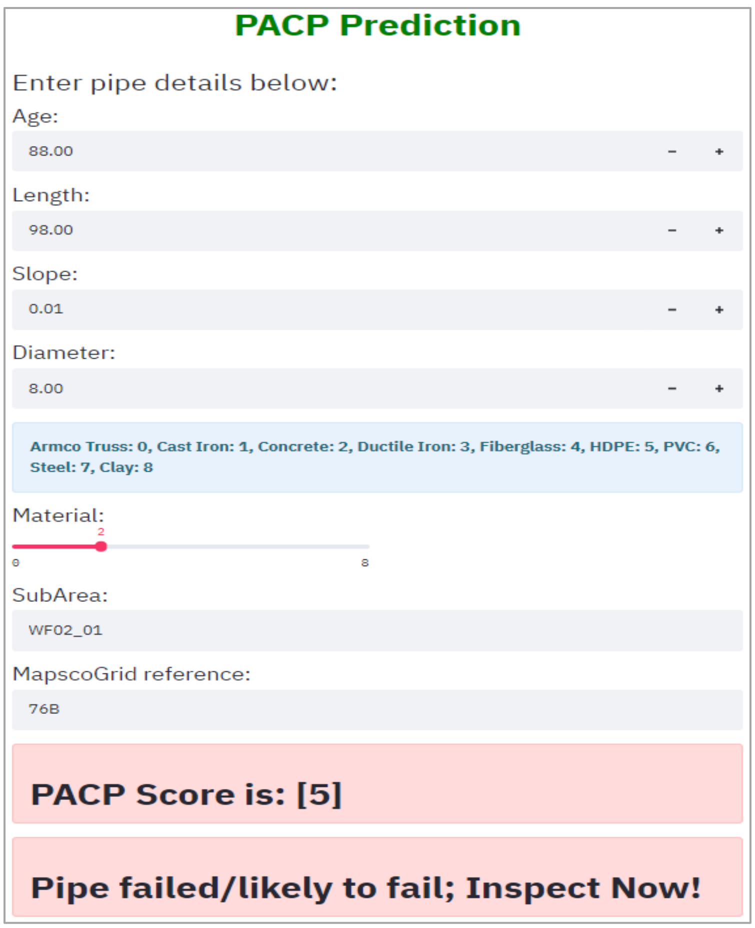

Figure 7-1 Decision-Support Tool for PACP Prediction

The web-application can be used by the asset managers to forecast the condition of their sewer pipes and hence, critical sewer pipes can be prioritized for inspection. Based on the data used for this dissertation, the PACP score of an uninspected sanitary sewer pipe segment could be assessed by the utility managers without an extensive CCTV operation in field. The following details are needed as inputs for the decision support tool:

1) Age
2) Length
3) Diameter
4) Slope
5) Pipe Material
6) location of the pipe such as MAPSCOGRID reference
7) SUBAREA of the pipe

Based on inputs given to the program, condition of the pipe in PACP score will be provided as output. For instance, sewer pipes predicted as PACP score 5 would require immediate attention whereas sewer pipe with PACP score 3 would need inspection after 15 years.

CHAPTER 8

RECOMMENDATIONS FOR FUTURE STUDIES

Additional research studies could be accomplished to further improve the research work discussed in this dissertation. Potential future development could include but not limited to the following:

- Number of joints in inspected sanitary sewer pipe segment could be analyzed and included in the model development.
- Surface loads such as the traffic and population intensity could be investigated as influential factors.
- Cost-benefit analysis could be accomplished to examine the cost savings for the municipality or utility owners.
- Other machine learning algorithms and artificial intelligent techniques could be utilized to further investigate effects of influential factors.
- Studies could intend to include sanitary sewer pipe segments with history of maintenance activities such as CIPP.
- Developed model in this dissertation could be validated on inspection data from a different municipality.
- It should be noted that the model developed in this dissertation is based on data collected from the City of Fort Worth. The program must be modified corresponding to any other agency's data inventory prior to practical application.

- Integration of Global Positioning System (GPS) during inspection of pipe segment would help in mapping the critical pipelines during inspection and condition assessment phases.

CHAPTER 9

REFERENCES

Alessia, S., Antonio, C., and Aldo, Q. (2017). Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review. *Frontiers in Aging Neuroscience, 9*, 329.

AMERICAN IRON AND STEEL INSTITUTE. (1999). *Modern Sewer Design.* Washington, DC.

Ana, E. V., and Bauwens, W. (2010). Modeling the structural deterioration of urban drainage pipes: the state-ofthe-art in statistical methods. *Urban Water Journal*, 47-59.

Ariaratnam, S. T., El-Assaly, A., and Yang, Y. (2001). Assessment of Infrastrucure Inspection Needs Using Logistic Models. *Journal of Infrastructure Systems*, 160 - 165.

Bakry, I., Alzraiee, H., Masry, M. E., Kaddoura, K., and Zayed, T. (2016). Condition Prediction for Cured-in-Place Pipe Rehabilitation of Sewer Mains. *Journal of Performance of Constructed Facilities, 30*(5), 04016016, 1-12.

Baur, R., and Herz, R. (2002). Selective inspection planning with ageing forecast for sewer types. *Water Science and Technology*, 389-396.

Caruana, R., and Niculescu-Mizil, A. (2006). An Empirical Comparison of Supervised Learning Algorithms. *Proceedings of the 23rd International Conference on Machine Learning*, (pp. 161-168). Pittsburgh.

Chae, M. J., and Abraham, D. M. (2000). Automated Condition Assessment of Sanitary Sewer Pipelines. *Computing in Civil and Building Engineering* , 1196-1203.

Chae, M. J., and Abraham, D. M. (2001). NEURO-FUZZY APPROACHES FOR SANITARY SEWER PIPELINE CONDITION ASSESSMENT. *JOURNAL OF COMPUTING IN CIVIL ENGINEERING*, 4-14.

Chughtai, F., and Zayed, T. (2008). Infrastructure Condition Prediction Models for Sustainable Sewer Pipelines. *Journal of Performance of Constructed Facilities,* 333-341.

Davies, J., Clarke, B., Whiter, J., and Cunningham, R. (2001). Factors influencing the structural deteriorationand collapse of rigid sewer pipes. *Urban Water*, 73-89.

Daziel, A., and Macey, C. (2004, February). The benefits of using quality sewer condition data in the asset management process. *Trenchless Technology*, pp. 36-39.

DEC. (2015). *Municipal Sewage System Asset Management Guide (MSSAM).* Albany, NY: Department of Environmental Conservation.

Dzuray, E. J., and Tafuri, A. N. (2000). Sewer Pipeline Performance Indicators: Learning from the European Experience. *Joint Conference on Water Resource Engineering and Water Resources Planning and Management 2000.* Minneapolis: ASCE.

EFC. (2006). *A Guide for Water and Wastewater Systems.* New Mexico: Environmental Finance Center.

EPA. (2002). *Asset Management for Sewer Collection Systems.* Washington, DC: Office of Wastewater Management.

EPA. (2004, August). *Report to Congress on Impacts and control of CSOs and SSOs.* Washington, D.C.: Office of Water. Retrieved from EPA: https://www.epa.gov/npdes/2004-npdes-cso-report-congress

Estabrooks, A., Jo, T., and Japkowicz, N. (2004). A Multiple Resampling Method for Learning from Imbalanced Data Sets. *Computational Intelligence*, 18-36.

Fenner, R. A. (2000). Approaches to sewer maintenance: a review. *Urban Water*, 343-356.

Gedam, A., Mangulkar, S., and Gandhi, B. (2016). Prediction of sewer pipe main condition using the linear regression approach. *Journal of Geoscience and Environment Protection*, 100-105.

Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow.* Sebastopol: O'Reilly.

Grandini, M., Bagli, E., and Visani, G. (2020). *METRICS FOR MULTI-CLASS CLASSIFICATION: AN OVERVIEW.* Bologna (BO), Italy: Università degli Studi di Bologna.

Guo, G., Wang, H., Bell, D., Bi, Y., and Greer, K. (2003). k-NN Model-Based Approach in Classification. *On The Move to Meaningful Internet Systems*, 986-996.

Hahn, M., Palmer, R., Merrill, S., and Lukas, A. (2002). Expert System for Prioritizing the Inspection of Sewers: Knowledge Base Formulation and Evaluation. *Journal of Water Resources Planning and Management*, 121-129.

Haixiang, G., Yijing, L., Shang, J., Mingyun, G., and Yuanyue, H. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert Systems With Applications*, 220-239.

Harvey, R. R., and McBean, E. A. (2014). Comparing the utility of decision trees and support vector machines when planning inspections of linear sewer infrastructure. *Journal of Hydroinformatics*, 1265-1279.

Hernandez, N., Caradot, N., Sonnenberg, H., Rouault, P., and Torres, A. (2017). Support Tools to Predict the Critical Structural Condition of Uninspected Sewer Pipes in Bogota D.C. *The Leading Edge Sustainable Asset Management of Water and Wastewater Infrastructure Conference.* Trondheim, Norway.

Hossin, M., and Sulaiman, M. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining and Knowledge Management Process*.

(2006). *International Infrastructure Management Manual.*

Jeong, H. S., Baik, H.-S., and Abraham, D. M. (2005). An ordered probit model approach for developing Markov chain-based deterioration model for wastewater infrastructure systems. *ASCE Pipelines*, 649-661.

Khan, Z., and Zayed, T. M. (2010). Structural Condition Assessment of Sewer Pipelines. *Journal of Performance of Constructed Facilities, 24*(2), 170-179.

Kienow, K. E., and Kienow, K. K. (2004). Risk Management... Predicting your Next Concrete Pipe Sewer Failure Before it Happens. *ASCE Pipeline Engineering and Construction*.

Kley, G., and Caradot, N. (2013). *D 1.2 Review of Sewer Deterioration Models.* Berlin: Kompetenzzentrum Wasser Berlin.

Laakso, T., Kokkonen, T., Mellin, I., and Vahala, R. (2018). Sewer Condition Prediction and Analysis of Explanatory Factors. *Water*, 1239.

Loh, W.-Y. (2014). Fifty Years of Classification and Regression Trees. *International Statistical Review, 82*(3), 329-348.

López, V., Fernández, A., García, S., Palade, V., and Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences*, 113-141.

Malek Mohammadi, M., Najafi, M., Kaushal, V., Serajiantehrani, R., Salehabadi, N., and Ashoori, T. (2019). Sewer Pipes Condition Prediction Models: A State-of-the-Art Review. *Infrastructures*.

Malek Mohammadi, M., Najafi, M., Kermanshachi, S., Kaushal, V., and Serajiantehrani, R. (2020). Factors Influencing the Condition of Sewer Pipes: State-of-the-Art Review. *Journal of Pipeline System and Engineering*, 11(4).

Malek Mohammadi, M., Najafi, M., Salehabadi, N., Serajiantehrani, R., and Kaushal, V. (2020). Predicting Condition of Sanitary Sewer Pipes with Gradient Boosting Tree. *ASCE Pipelines*, 80-89.

Malek Mohammadi, M., Najafi, m., Tabesh, A., Riley, J., and Gruber, J. (2019). Condition Prediction of Sanitary Sewer Pipes. *ASCE Pipelines*, 117-126.

McDonald, S. E., and Zhao, J. Q. (2001). Condition assessment and rehabilitation of large sewers. *International Conference on Underground Infrastructure Research* (pp. 361-369). Waterloo: National Research Council Canada.

Micevski, T., Kuczera, G., and Coombes, P. (2002). Markov Model for Storm Water Pipe Deterioration. *Journal of Infrastructure Systems*, 49-56.

Misiunas, D. (2005). *Failure monitoring and asset condition assessment in water supply systems.* Sweden: Lund University.

Mohri, M., Rostamizadeh, A., and Talwalkar, A. (2018). *Foundations of Machine Learning (second edition).* Cambridge, MA: The MIT Press.

Moore, Bob. (2015, December 01). *TRUSS PIPE for Sanitary Sewers – 52 Years and Counting*. Retrieved from Contech: https://www.conteches.com/pipe-article/article/26/truss-pipe-for-sanitary-sewers-52-years-and-counting

Moteleb, M. (2010). *Risk Based Decision Making Tools for Sewer Infrastructure Management.* Ohio: the University of Cincinnati.

Müller, A. C., and Guido, S. (2016). *Introduction to Machine Learning with Python: A Guide for Data Scientists.* Boston: O'Reilly.

Najafi, M., and Gokhale, S. (2005). *Trenchless Technology.* New York: McGraw-Hill.

Najafi, M., and Kulandaivel, G. (2005). Pipeline Condition Assessment Prediction Using Neural Network Models. *ASCE Pipelines*, 767-781.

NASSCO. (2018, January). Pipeline Assessment Certificate Program .

Opila, M. C. (2011). *Structural Condition Scoring of Buried Sewer Pipes for Risk-Based Decision-Making.* Delaware: University of Delaware.

OWASA. (2017). *Asset Management Program.* Carrboro, NC: Orange Water and Sewer Authority.

Rout, N., Kuhoo, Mishra, D., and Mallick, M. K. (2018). Analysing the Multi-class Imbalanced Datasets using Boosting Methods and Relevant Information. *International Journal of Pure and Applied Mathematics, 120*(6), 6171-6191.

Singh, A., and Adachi, S. (2011). Bathtub curves and pipe prioritization based on failure rate. *Built Environment: Project Asset Management*, 105-122.

Sousa, V., Matos, J. P., and Matias, N. (2014). Evaluation of artificial intelligence tool performance and uncertainty for predicting sewer structural condition. *Automation in Construction*, 84-91.

Syachrani, S., Jeong, H. S., and Chung, C. S. (2013). Decision Tree–Based Deterioration Model for Buried Wastewater Pipelines. *ASCE Journal of Performance of Constructed Facilities*, 633-645.

Tafuri, A. N., and Dzuray, E. J. (2004). Sewer Pipeline Performance Indicators: Learning from the European Experience. *Building Partnerships, Water Resources 2000*, 1-10.

Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N., and Asadpour, M. (2020). Boosting methods for multi-class imbalanced data classification: an experimental review. *Journal of Big Data*, 1-47.

Teh, K., Armitage, P., Tesfaye, S., Selvarajah, D., and Wilkinson, I. D. (2020). Imbalanced learning: Improving classification of diabetic neuropathy from magnetic resonance imaging. *Oversampling improvements in diabetic neuropathy MRI*.

Tran, D., Ng, A. W., Perera, B. J., Burn, S., and Davis, P. (2007). Application of probabilistic neural networks in modelling structural deterioration of stormwater pipes. *Urban Water Journal*, 175-184.

Tscheikner-Gratl, F., Caradot, N., Cherqui, F., Leitão, J. P., Ahmadi, M., Langeveld, J. G., . . . Clemens, F. (2020). Sewer asset management – state of the art and research needs. *Urban Water Journal*, 662-675.

Visani, G., Bagli, E., and Grandini, M. (2020). *Metrics for Multi-class Classification: An Overview.* Italy: Università degli Studi di Bologna.

Vladeanu, G. J. (2018). *Wastewater Pipe Condition and Deterioration Modeling for Risk-Based Decision Making.* Louisiana: Louisiana Tech University.

Wallace, B. C., Small, K., Brodley, C. E., and Trikalinos, T. A. (2011). Class Imbalance, Redux. *11th IEEE International Conference on Data Mining*, 754-763.

Water Reseach Center. (1986). *Sewerage Rehabilitation Book 2nd edition.* United Kingdom.

Wirahadikusumah, R., Abraham, D., and Iseley, T. (June 2001). Challenging Issues in Modeling Deterioration of Combined Sewers. *ASCE Journal of Infrastructure Systems*, 77-84.

WRc. (1986). *Sewerage Rehabilitation Manual.* England.

Wright, L. T., Heaney, J. P., and Dent, S. (2006). Prioritizing Sanitary Sewers for Rehabilitation Using Least-Cost Classifiers. *ASCE JOURNAL OF INFRASTRUCTURE SYSTEMS* , 174-183.

Yan, J., and Vairavamoorthy, K. (2003). Fuzzy approach for pipe condition assessment. *New Pipeline Technologies*, 466-476.

Yijing, L., Haixiang, G., Xiao, L., Yanan, L., and Jinling, L. (2016). Adapted ensemble classification algorithm based on multiple classifier system and feature selection for classifying multi-class imbalanced data. *Knowledge-Based Systems*, *94*, 88-104.

Appendix A

Abbreviations

ANN – Artificial Neural Network

ASCE – American Society of Civil Engineering

AT – Armco Truss

AUC – Area under the Curve

BF – Big Fossil

CBO – Congressional Budget Office

CCTV – Closed-Circuit Television

CF – Clear Fork

CI – Cast Iron

DI – Ductile Iron

DT – Decision Trees

EFC – Environmental Finance Center

EPA – Environmental Protection Agency

FN – False Negative

FP – False Positive

FPR – False Positive Rate

Gi – Gini Index

GIS – Geographic Information System

GPS – Global Positioning System

HDPE – High-Density Polyethylene

IIMM – International Infrastructure Management Manual

k-NN – k-Nearest Neighbors

LR – Logistic Regression

MSSAM – Municipal Sewage System Asset Management

NASSCO – National Association of Sewer Service Company

OWASA – Orange Water and Sewer Authority

PACP – Pipeline Assessment Certification Program

PVC – Polyvinyl Chloride

RF – Random Forests

ROC – Receiver Operator Characteristic

SCRAPS – Sewer Cataloging, Retrieval and Prioritization System

TN – True Negative

TP – True Positive

TPR – True Positive Rate

VC – Village Creek

VCP – Vitrified Clay Pipe

WEF – Water Environment Federation

WRc – Water Research Center

Appendix B

Data Sample (1,000 pipe segments)

Sl. No	Age	Length	MAPSCO GRID	Slope	SUBAREA	Size	Material	PACP
1	22.3	844	93G	0.0005	VC09_01	54	Concrete	2
2	16.4	415	46L	0.004	MC04_04	8	PVC	1
3	13.5	426	46H	0.018	MC04_04	8	PVC	1
4	48	259	89F	0.0206	CF05_03	6	VCP	3
5	47.9	503	89F	0.0301	CF05_03	6	Concrete	2
6	14.4	112	119G	0.0057	VC11_03	8	PVC	1
7	15.3	444	106U	0.0067	VC11_01	24	PVC	1
8	15.2	396	106S	0.0038	VC11_01	24	PVC	2
9	67.4	203	47Y	0.0337	MC03_06	6	VCP	2
10	33.3	95	103H	0.0168	SC09_05	6	VCP	3
11	35.4	382	103D	0.0183	SC09_02	6	PVC	1
12	20.2	431	78K	0.0257	SC10_03	8	HDPE	1
13	35.7	180	103H	0.0563	SC09_02	6	PVC	1
14	17.6	71	103D	0.0331	SC09_01	12	PVC	1
15	15.1	417	103D	0.0142	SC09_02	21	PVC	1
16	7.5	345	71L	0.0038	PCF15_01	36	PVC	1
17	78.3	17	48Y	0.0118	MC03_01	6	Concrete	2
18	9.5	271	62E	0.0316	MC06_04	8	PVC	1
19	3	65	62E	0.0365	MC06_04	8	PVC	1
20	25.3	195	75Y	0.0018	CF05_01	30	Concrete	1
21	30.8	379	63Y	0.0014	BF04_01	15	VCP	1
22	14.3	521	90A	0.0059	CF04_03	8	PVC	2
23	82.5	252	90B	0.0041	CF04_03	8	Concrete	4
24	82.6	142	90A	0.0222	CF04_03	8	Concrete	4
25	11.4	301	61C	0.0219	MC03_06	8	PVC	1
26	28.8	531	61H	0.088	WF01_02	6	VCP	3
27	7.8	81	89D	0.004	CF04_05	8	HDPE	1
28	7.8	502	89D	0.0042	CF04_05	8	HDPE	3
29	7.8	141	89D	0.0065	CF04_05	8	HDPE	3
30	62.2	30	90E	0.1133	CF04_05	8	Concrete	4
31	16.3	295	74Y	0.0039	CF12_02	8	CI	1
32	12.9	460	92G	0.0061	VC09_04	8	HDPE	1
33	12.4	439	92L	0.011	VC09_03	8	PVC	1
34	12.5	363	92F	0.004	VC09_04	8	PVC	2
35	7.5	242	72N	0.0019	CF14_02	36	PVC	1

36	1.1	327	78F	0.0638	SC10_04	8	PVC	3
37	13.4	460	19Z	0.0055	BF05_08	8	PVC	1
38	10.7	552	19Z	0.0089	BF05_08	8	PVC	1
39	61	412	74N	0.0194	CF12_03	8	VCP	4
40	27.5	230	103J	0.0104	CF09_04	10	PVC	1
41	29.9	376	88T	0.004	CF08_02	8	PVC	1
42	20	163	87Z	0.0141	CF08_05	8	PVC	1
43	10.6	437	103X	0.006	CF09_03	8	PVC	1
44	88.8	565	73X	0.0121	CF13_01	15	Concrete	3
45	5.9	316	21X	0.0099	BF09_03	8	PVC	1
46	10.2	288	21X	0.0197	BF09_03	8	PVC	1
47	17.9	488	61D	0.0438	MC03_06	8	PVC	2
48	8.1	80	89D	0.0041	CF04_05	8	HDPE	1
49	8.1	304	89D	0.0055	CF04_05	8	HDPE	1
50	7.7	73	89H	0.0069	CF04_05	8	HDPE	1
51	7.7	290	89H	0.0052	CF04_05	8	HDPE	1
52	26.3	634	74Y	0.0023	CF12_02	10	PVC	1
53	64.9	121	88A	0.006	CF12_01	12	Concrete	3
54	4.8	625	23J	0.004	DC03_01	30	PVC	1
55	14.8	48	31H	0.0015	MC05_07	36	PVC	1
56	12.9	408	72D	0.0146	WF05_02	8	PVC	1
57	4.1	100	35S	0.0219	BF05_03	8	PVC	1
58	33.2	663	47J	0.0023	MC04_04	30	VCP	3
59	88.3	98	76B	0.0061	WF02_01	8	Concrete	5
60	13.8	69	76D	0.0033	CF01_04	27	DI	2
61	14.3	240	76D	0.0072	CF01_04	27	DI	3
62	15.6	187	76D	0.0235	CF01_05	24	DI	1
63	20.5	45	63W	0.0071	CF01_07	10	HDPE	1
64	8	463	76D	0.0069	CF01_04	10	HDPE	1
65	15.8	295	103N	0.0051	CF09_02	15	PVC	1
66	16.4	238	102R	0.0408	CF09_02	8	PVC	1
67	10.8	221	78A	0.0013	SC10_01	42	PVC	3
68	44.1	94	50Y	0.0143	BF02_01	30	Concrete	3
69	49	99	62L	0.0017	MC03_01	45	Concrete	2
70	28.3	900	65B	0.0007	BF01_04	54	Concrete	2
71	22.3	14	93D	0.0007	VC08_01	54	Concrete	1
72	34.1	10	93G	0.55	VC08_01	8	VCP	1

73	13.5	84	46H	0.0032	MC04_04	10	PVC	1
74	26.9	422	106U	0.0033	VC11_01	24	Concrete	3
75	39.6	374	72C	0.0175	SC06_01	12	VCP	2
76	17.5	412	74Y	0.0039	CF12_02	8	PVC	1
77	12.5	281	92H	0.0009	VC09_02	18	PVC	1
78	43.6	632	64H	0.0015	BF02_01	35	Concrete	3
79	44.2	1109	64C	0.0035	BF02_01	30	Concrete	5
80	25.3	535	72B	0.0474	WF05_05	6	PVC	1
81	17	363	32P	0.0183	MC05_02	8	PVC	1
82	13.5	507	93U	0.0003	VC10_01	48	Concrete	1
83	12.5	38	92H	0.0053	VC09_02	18	PVC	1
84	7.1	73	63W	0.0137	CF01_03	12	PVC	1
85	27	311	102R	0.0174	CF09_04	8	PVC	1
86	62.5	157	74N	0.0241	CF12_03	8	VCP	2
87	27.5	308	103J	0.0388	CF09_04	8	PVC	1
88	15.7	306	103N	0.008	CF09_02	8	PVC	1
89	15.7	47	103P	0.0268	CF09_02	8	PVC	1
90	46.2	474	93Y	0.0004	VC10_01	36	Concrete	1
91	6	342	78F	0.0258	SC10_04	8	PVC	1
92	6	103	78G	0.0407	SC10_04	8	PVC	1
93	13.5	204	46H	0.0032	MC04_04	10	PVC	1
94	13.5	102	46H	0.1121	MC04_04	10	PVC	1
95	22.3	600	93L	0.0004	VC09_01	54	Concrete	2
96	15	291	106S	0.0023	VC11_01	24	PVC	2
97	14.7	320	79F	0.0148	VC06_03	15	PVC	1
98	10.8	376	78M	0.0133	VC07_03	8	PVC	1
99	19.2	158	66Y	0.0246	VC03_02	6	PVC	1
100	43.3	196	120B	0.0012	VC11_02	36	VCP	3
101	13.5	411	119Z	0.0039	VC11_03	8	PVC	1
102	35	562	80F	0.006	VC04_02	6	VCP	2
103	12.8	168	81E	0.0482	VC01_01	8	PVC	1
104	38.3	600	80H	0.0439	VC03_03	6	VCP	2
105	12.5	187	92H	0.0016	VC09_02	18	PVC	1
106	9.8	139	47Y	0.035	MC03_06	8	PVC	1
107	17.8	253	78L	0.0049	SC10_03	8	HDPE	1
108	57.9	76	74P	0.0442	CF12_05	6	Concrete	2
109	35.7	308	103H	0.036	SC09_02	6	PVC	1

110	63.2	130	74P	0.0546	CF12_05	6	Concrete	3
111	7.1	480	71F	0.002	PCF15_01	36	PVC	2
112	7.5	400	71L	0.0038	PCF15_01	36	PVC	1
113	7.5	356	72N	0.0021	CF14_02	36	PVC	3
114	2.2	155	77T	0.0173	SC02_05	8	PVC	1
115	1.6	228	80N	0.0033	VC01_05	24	DI	1
116	51.6	470	89C	0.0029	CF05_02	24	Concrete	3
117	46.2	370	93U	0.0002	VC10_01	36	Concrete	1
118	16.3	212	74Y	0.0029	CF12_02	10	CI	1
119	7.7	24	93C	0.0613	VC08_01	24	Concrete	1
120	16.6	165	88C	0.0031	CF12_01	18	PVC	1
121	52.2	233	89L	0.03	CF05_03	8	Concrete	3
122	50.5	43	89L	0.0861	CF05_03	6	VCP	2
123	52.7	180	90N	0.0256	SC08_02	6	Concrete	3
124	15.7	119	88Q	0.0347	CF07_03	8	PVC	2
125	12.1	236	76Q	0.0416	CF02_02	8	PVC	1
126	64.9	220	87D	0.0081	CF12_01	12	Concrete	1
127	9	266	48F	0.0029	MC02_04	10	PVC	1
128	9.2	291	74K	0.0003	WF04_02	8	PVC	1
129	35.7	170	103G	0.0832	SC09_02	6	PVC	1
130	35.7	302	103G	0.0424	SC09_02	6	PVC	1
131	14.4	559	74L	0.0051	CF11_04	8	HDPE	3
132	16.1	138	21X	0.008	BF09_03	15	DI	1
133	0.2	52	78Q	0	SC11_04	8	HDPE	4
134	35.7	42	103H	0.0245	SC09_02	6	PVC	1
135	3	141	62E	0.0745	MC06_04	8	PVC	1
136	53	494	63T	0.0004	MC01_01	68	Concrete	3
137	60	243	80A	0.0288	VC04_04	8	Concrete	3
138	24.5	369	78E	0.0192	SC10_02	8	PVC	1
139	83.1	370	61D	0.0458	MC03_05	6	Concrete	3
140	13.4	202	19Z	0.0185	BF05_08	8	PVC	1
141	8.9	150	49W	0.0015	MC02_03	8	PVC	1
142	17.9	101	35X	0.0025	BF05_02	8	PVC	2
143	25	86	48U	0.019	MC02_04	6	DI	1
144	7.5	487	71R	0.0035	PCF15_01	36	PVC	1
145	6	84	74Y	0.0012	CF12_02	10	PVC	1
146	6	98	74Y	0.001	CF12_02	10	PVC	1

147	10.9	128	76H	0.004	CF01_06	8	PVC	1
148	25.3	261	72B	0.0307	WF05_05	6	PVC	1
149	30.7	494	65G	0.0008	BF01_04	54	Concrete	2
150	15.1	314	103D	0.0057	SC09_02	21	PVC	1
151	7.1	250	71F	0.0036	PCF15_01	36	PVC	1
152	7.5	367	71L	0.0038	PCF15_01	36	PVC	1
153	7.8	549	77T	0.013	SC02_05	8	PVC	1
154	26.6	132	102D	0.0369	CF07_04	8	PVC	1
155	6	35	74Y	0.01	CF12_02	10	PVC	1
156	61.3	219	48Y	0.0161	MC03_01	6	Concrete	4
157	3	180	62E	0.0421	MC06_04	8	PVC	1
158	27	147	78B	0.052	SC10_01	8	VCP	2
159	48.5	289	80C	0.0228	VC04_01	6	Concrete	2
160	46.1	51	93X	0.0024	VC10_01	36	Concrete	2
161	83.8	111	90M	0.0865	SC04_02	6	Concrete	3
162	26.3	129	74Y	0.0038	CF12_02	10	PVC	1
163	26.3	366	88C	0.0029	CF12_02	10	PVC	3
164	41	405	79T	0.003	VC07_02	18	Concrete	3
165	8.5	68	79S	0.0028	VC07_02	10	DI	4
166	69.4	247	88B	0.003	CF12_01	12	Concrete	3
167	12.1	128	76Q	0.0577	CF02_02	8	PVC	1
168	16.8	8	74X	0.2825	CF12_02	16	DI	3
169	35.6	226	74X	0.1231	CF12_02	6	VCP	1
170	2.9	113	89M	0.0099	CF06_07	8	PVC	1
171	2.9	276	90N	0.0065	CF06_07	8	PVC	1
172	2.9	127	90N	0.0161	CF06_07	8	PVC	1
173	22.8	297	36L	0.0202	BF06_05	8	PVC	1
174	13.7	385	90N	0.0636	CF06_07	8	PVC	1
175	0.3	88	75G	0.0265	CF10_04	8	PVC	1
176	9.3	82	76K	0.0111	CF02_02	8	PVC	1
177	23.4	40	76P	0.0425	CF03_01	10	HDPE	2
178	24.5	72	76P	0.0117	CF03_02	21	PVC	1
179	11.5	264	76K	0.0061	CF02_02	8	PVC	1
180	13.8	481	76Q	0.0062	CF02_02	8	PVC	1
181	19.5	385	76P	0.0112	CF03_02	8	VCP	3
182	24.5	35	76T	0.0034	CF03_02	21	PVC	1
183	13.8	27	76T	0.0219	CF03_02	10	DI	1

184	33.4	280	80D	0.0464	VC03_03	6	VCP	2
185	65.1	186	74P	0.0097	CF12_05	8	Concrete	3
186	35.7	400	103G	0.0238	SC09_02	8	VCP	2
187	59.1	559	74J	0.0201	WF04_03	8	PVC	2
188	7.1	415	71E	0.0021	PCF15_02	36	PVC	1
189	7.1	250	71E	0.0021	PCF15_01	36	PVC	1
190	7.5	150	72N	0.0023	CF14_02	36	Fiberglass	1
191	30.1	191	73K	0.0201	CF13_02	8	VCP	3
192	12.5	101	88U	0.0462	CF08_02	8	PVC	1
193	82.9	170	90B	0.0082	CF03_04	10	Concrete	2
194	12.1	303	103E	0.0198	SC09_04	8	PVC	1
195	15.5	231	104A	0.0089	SC09_01	24	DI	1
196	38.7	380	103H	0.0292	SC09_02	6	PVC	1
197	38.7	192	103C	0.0059	SC09_02	6	VCP	2
198	38.7	295	103C	0.0059	SC09_02	6	VCP	2
199	59.9	127	74K	0.0638	CF12_05	6	Concrete	2
200	7.1	400	71F	0.0021	PCF15_01	36	PVC	1
201	7.5	400	71L	0.0038	PCF15_01	36	PVC	1
202	52.1	108	73Z	0.0735	CF12_09	6	Concrete	3
203	61.5	27	64Z	0.0148	SC10_01	6	Concrete	2
204	37.5	311	80H	0.0172	VC03_03	8	VCP	2
205	40	400	103G	0.0344	SC09_03	6	VCP	2
206	30.1	480	103F	0.0099	SC09_04	8	PVC	1
207	39.4	247	76T	0.0326	CF03_01	8	HDPE	1
208	43.7	347	76K	0.0021	CF04_01	24	CI	1
209	14	168	76Y	0.0139	CF03_05	8	DI	1
210	82.9	28	90B	0.0071	CF03_04	10	Concrete	2
211	27.6	30	76V	0.1617	CF02_04	8	PVC	2
212	12.1	383	76L	0.0218	CF02_02	8	DI	1
213	16.6	77	76K	0.006	CF03_01	8	PVC	2
214	12.7	85	106W	0.0053	VC11_06	10	PVC	1
215	0.1	22	62P	0.0032	MC06_02	8	DI	1
216	14.8	21	62K	0.0052	MC06_02	15	PVC	1
217	3.5	248	22B	0.004	DC02_03	8	PVC	1
218	46.2	915	93Q	0.0008	VC09_05	34	Concrete	2
219	46.1	740	93X	0.0002	VC10_01	36	Concrete	2
220	87	55	76Y	0	CF03_05	6	Concrete	3

221	39.7	161	72D	0.0186	WF05_02	18	VCP	2
222	3.4	9	103P	0.0322	CF09_05	8	PVC	1
223	0.4	483	79M	0.0242	VC01_05	8	PVC	1
224	12.6	32	88T	0.1681	CF08_02	8	PVC	1
225	12.1	237	103E	0.0136	SC09_04	8	PVC	1
226	14.8	222	103N	0.0402	CF09_05	8	PVC	1
227	14.8	544	103K	0.01	CF09_05	8	PVC	1
228	11.7	78	92H	0.0019	VC09_02	8	DI	2
229	12.9	78	35U	0.036	BF09_02	8	PVC	1
230	12.9	208	35U	0.0289	BF09_02	8	PVC	1
231	3.6	25	21Y	0.0088	BF08_04	8	PVC	1
232	20	97	49Q	0.0112	BF02_03	8	DI	2
233	3.4	254	21T	0.0158	BF09_03	10	PVC	1
234	43.3	778	120B	0.0012	VC11_02	36	VCP	4
235	69.1	308	63E	0.0091	MC02_01	6	Concrete	3
236	9.3	268	76K	0.0209	CF02_02	8	PVC	1
237	17.4	258	76K	0.0248	CF02_02	8	PVC	1
238	31.1	409	76K	0.0111	CF03_01	8	PVC	1
239	12.9	288	72D	0.0163	WF05_02	8	PVC	1
240	12.9	504	72C	0.0211	WF05_02	8	PVC	1
241	18.4	405	66W	0.0041	VC04_04	8	PVC	1
242	60.6	142	79D	0.087	VC04_04	6	Concrete	3
243	54.9	403	81A	0.0062	VC03_03	18	VCP	2
244	12.7	568	106W	0.0024	VC11_06	10	PVC	1
245	12.7	539	106W	0.0028	VC11_06	10	PVC	1
246	13.8	222	76Q	0.0023	CF02_02	8	PVC	3
247	12.2	213	76Q	0.0057	CF02_02	8	HDPE	3
248	20.7	82	66N	0.0063	VC02_03	8	PVC	3
249	85.7	300	63Y	0.0003	BF04_01	15	Concrete	4
250	7.5	472	72N	0.002	PCF15_01	36	Fiberglass	1
251	6	92	74Y	0.0051	CF12_02	8	PVC	1
252	61.9	43	89H	0.0091	CF04_05	6	Concrete	4
253	12.8	360	92G	0.0209	VC09_03	8	HDPE	1
254	51.5	36	89C	0.0044	CF05_02	24	Concrete	2
255	46.1	38	93X	0	VC10_01	36	Concrete	2
256	60.8	457	79X	0.0045	VC07_02	10	CI	4
257	7.6	155	119Z	0.003	VC11_03	10	PVC	1

258	13.2	437	92G	0.011	VC09_03	8	HDPE	1
259	14	50	92L	0.0082	VC09_04	8	DI	1
260	12.2	408	92L	0.0116	VC09_03	8	HDPE	1
261	13.2	125	92L	0.0039	VC09_03	8	PVC	1
262	11.7	345	92M	0.0063	VC09_02	8	HDPE	2
263	7.6	507	92M	0.0099	VC09_02	8	PVC	1
264	13.9	434	73U	0.0129	CF13_01	8	PVC	1
265	37.5	427	73V	0.0141	CF12_09	8	VCP	2
266	50.1	60	62L	0.0757	MC06_01	21	Concrete	3
267	10.3	377	21X	0.005	BF09_03	8	PVC	1
268	25.3	128	72B	0.0363	WF05_05	8	PVC	1
269	28.6	12	65G	0.0008	BF01_04	54	Concrete	2
270	13.5	203	78X	0.0232	SC11_03	8	PVC	1
271	28.2	375	35S	0.0022	BF05_04	36	Concrete	1
272	46.4	294	89N	0.0228	CF07_03	18	Concrete	2
273	38.8	606	47J	0.004	MC04_04	21	VCP	2
274	10.6	381	62F	0.0084	MC06_04	12	PVC	1
275	30.7	204	73P	0.055	CF13_02	6	VCP	2
276	16.4	218	103N	0.0136	CF09_02	8	PVC	1
277	14.8	98	103P	0.0244	CF09_05	8	PVC	1
278	9.9	37	103W	0.0784	CF09_03	8	PVC	1
279	27.4	113	66S	0.0904	VC02_03	8	PVC	1
280	86.1	280	78A	0.2229	SC10_01	6	VCP	3
281	15.2	11	90A	0.0082	CF04_03	15	PVC	1
282	3.5	264	21P	0.004	BF09_03	8	PVC	1
283	58.4	3	80B	4.9033	VC04_03	6	Concrete	3
284	18.9	121	61D	0.004	MC03_06	8	PVC	1
285	8.1	297	89D	0.0057	CF04_05	8	HDPE	1
286	7.8	56	89D	0.0045	CF04_05	8	HDPE	1
287	6.9	109	74X	0.0655	CF12_02	8	PVC	1
288	16.8	9	74X	0.09	CF12_02	16	DI	1
289	22.3	712	93G	0.0005	VC08_01	54	Concrete	2
290	58.2	145	92D	0.0052	VC08_05	10	VCP	2
291	53	389	62M	0.0004	MC02_01	69	Concrete	3
292	3.4	217	21T	0.006	BF09_03	10	PVC	1
293	46.3	291	93U	0.0007	VC10_01	36	Concrete	2
294	46.1	119	93X	0.0004	VC10_01	36	Concrete	2

295	60.8	708	79X	0.004	VC07_02	10	VCP	4
296	5.3	248	90S	0.0351	SC08_02	8	HDPE	3
297	50.4	240	89G	0.006	CF05_03	8	Concrete	4
298	61.9	84	89H	0.007	CF04_05	6	Concrete	3
299	12.9	118	92G	0.0081	VC09_04	8	HDPE	2
300	12.2	341	92L	0.0072	VC09_03	8	HDPE	1
301	12.5	528	92F	0.006	VC09_04	8	PVC	1
302	10.6	226	93J	0.0121	VC09_01	8	PVC	1
303	28.7	456	74Z	0.0012	CF12_04	21	Concrete	2
304	22.3	473	92N	0.0048	SC05_02	8	PVC	1
305	15.9	180	87D	0.0208	CF12_09	8	PVC	1
306	15.9	75	73Z	0.0184	CF12_09	8	PVC	1
307	11.4	182	75G	0.0141	CF10_04	8	PVC	1
308	48.7	250	62F	0.0018	MC03_02	39	Concrete	3
309	85.1	602	62Y	0.0012	WF02_01	12	Concrete	3
310	58.4	600	79D	0.0445	VC06_01	6	VCP	4
311	53	1431	63U	0.0004	MC01_01	68	Concrete	3
312	48.5	289	80C	0.032	VC04_01	6	Concrete	1
313	22.8	237	76Q	0.0125	CF02_02	8	PVC	1
314	13.2	126	92L	0.0041	VC09_03	8	PVC	1
315	10.6	374	93J	0.0397	VC09_01	8	PVC	2
316	4.4	5	63W	0.008	MC01_01	8	PVC	1
317	7.1	472	63W	0.0199	CF01_03	12	DI	1
318	32.2	172	74Y	0.182	CF12_02	6	VCP	1
319	22.3	334	92N	0.005	SC05_02	8	PVC	1
320	11.9	195	92H	0.004	VC09_02	8	PVC	2
321	1.6	152	80N	0.0032	VC01_05	24	DI	1
322	1.6	191	80N	0.0033	VC01_05	24	DI	1
323	34.3	77	89F	0.032	CF05_03	6	VCP	2
324	52.4	314	76N	0.006	CF04_01	6	Concrete	3
325	10.9	320	119Z	0.0084	VC11_03	8	PVC	1
326	35.4	133	66P	0.0309	VC02_03	6	VCP	1
327	59.2	309	74T	0.067	CF12_02	6	VCP	3
328	1.6	325	76B	0.005	WF02_01	8	PVC	1
329	56.4	196	74X	0.0792	CF12_02	6	VCP	3
330	29.1	302	88W	0.0093	CF08_05	10	PVC	1
331	27.5	247	102M	0.0463	CF09_04	8	PVC	1

332	11.2	89	87Z	0.004	CF08_04	8	DI	1
333	35.4	217	103H	0.0254	SC09_02	6	PVC	1
334	35.7	395	103G	0.0094	SC09_02	8	VCP	2
335	34	263	103G	0.0152	SC09_02	6	VCP	1
336	35.7	285	103G	0.0161	SC09_02	8	VCP	2
337	38.7	503	103C	0.0441	SC09_02	6	VCP	3
338	38.7	289	103C	0.0087	SC09_02	6	VCP	2
339	19.6	200	74L	0.011	CF11_04	8	HDPE	2
340	7.5	300	71L	0.0039	PCF15_01	36	PVC	1
341	14.9	219	67Y	0.0049	VC03_02	8	PVC	1
342	52.8	587	93G	0.0005	VC09_01	39	Concrete	2
343	5.5	270	46G	0.0075	MC04_02	8	PVC	1
344	44.3	993	106V	0.0031	VC11_01	33	VCP	3
345	14.4	320	119G	0.0378	VC11_03	8	PVC	1
346	33.3	284	72C	0.0154	WF05_05	8	VCP	1
347	33.3	251	80D	0.0039	VC03_03	8	VCP	3
348	63	116	74T	0.2303	CF12_02	10	VCP	5
349	16.7	321	74T	0.0027	CF12_02	16	DI	1
350	23.8	750	89Q	0.0026	CF06_02	20	DI	1
351	44.1	217	89S	0.0068	CF07_04	18	Concrete	2
352	3.2	94	75W	0.0038	CF11_06	8	PVC	1
353	1.6	143	75H	0.075	CF10_03	8	PVC	1
354	2.1	287	22W	0.0109	BF08_02	8	PVC	1
355	3.1	528	19W	0.0039	PBF10_01	8	PVC	1
356	4.2	10	19Z	0.004	BF05_08	8	PVC	1
357	4.2	9	19Z	0.0044	BF05_08	8	PVC	1
358	14.3	241	76D	0.0078	CF01_04	30	DI	1
359	20.5	95	63W	0.0221	CF01_07	10	HDPE	1
360	13.8	230	76D	0.0022	CF01_04	30	DI	1
361	49.6	101	91C	0.003	SC03_01	45	Concrete	5
362	43.6	1053	64C	0.0025	BF02_01	35	Concrete	3
363	30.6	360	66D	0.0064	BF01_03	8	PVC	1
364	18	187	78H	0.007	SC10_05	8	HDPE	1
365	4	182	63Z	0.0032	BF04_01	8	PVC	1
366	13.8	359	67Q	0.071	VC02_01	8	PVC	1
367	6	95	74Y	0.0011	CF12_02	10	PVC	1
368	8.3	104	65U	0.0032	VC05_01	16	DI	2

369	31.6	356	102D	0.0095	CF07_06	6	VCP	2
370	2.9	150	90N	0.0087	CF06_07	8	PVC	1
371	2.9	450	90N	0.0378	CF06_07	8	PVC	1
372	2.9	293	90N	0.0149	CF06_07	8	PVC	1
373	2.9	21	90N	0.0276	CF06_07	8	PVC	1
374	7.6	88	21L	0.0225	DC02_04	8	PVC	2
375	2.9	30	90N	0.02	CF06_07	8	PVC	1
376	2.9	29	90N	0.0103	CF06_07	8	PVC	1
377	15.7	340	103S	0.021	CF09_02	8	PVC	1
378	14.7	437	103S	0.0045	CF09_02	8	PVC	1
379	14.8	444	103N	0.0179	CF09_05	8	PVC	1
380	10.2	291	103X	0.004	CF09_03	8	PVC	1
381	13	265	103S	0.011	CF09_02	8	PVC	1
382	2.9	155	90N	0.0054	CF06_07	8	PVC	1
383	2.9	108	90N	0.0546	CF06_07	8	PVC	1
384	54.1	428	90W	0.0072	SC08_06	12	Concrete	3
385	15.7	347	90V	0.0071	SC08_01	35	Concrete	3
386	2.9	56	89R	0.07	CF06_07	6	HDPE	1
387	3.5	237	75M	0.004	CF10_05	8	PVC	1
388	2.9	52	90N	0.005	CF06_07	8	PVC	1
389	0.8	147	77T	0.0195	SC03_05	8	PVC	1
390	73.2	11	62G	0.1909	MC03_02	10	PVC	5
391	5	138	63L	0.0583	BF04_02	8	HDPE	1
392	12.5	28	63L	0.1436	BF04_02	8	PVC	1
393	32.7	77	92S	0.0807	VC10_01	10	PVC	3
394	6	42	74Y	0.0024	CF12_02	10	PVC	1
395	46.5	331	76N	0.0204	CF04_01	6	Concrete	3
396	14.8	245	32E	0.004	MC05_06	24	DI	2
397	12.7	60	89G	0.2803	CF05_03	8	PVC	1
398	33.1	501	66N	0.0322	VC02_03	6	VCP	2
399	35.6	269	74X	0.0137	CF12_02	6	VCP	2
400	66.3	224	74T	0.0037	CF12_02	10	VCP	3
401	31	232	89N	0.0039	CF07_03	8	PVC	2
402	46.2	8	93X	0.0013	VC10_01	36	Concrete	1
403	14.8	286	103P	0.005	CF09_05	8	PVC	1
404	51.6	54	79M	0.0074	VC01_05	10	Concrete	4
405	51.5	302	79M	0.0034	VC01_05	10	Concrete	3

406	15.9	153	76K	0.0079	CF02_02	8	PVC	1
407	83.5	420	79K	0.0157	VC06_03	6	VCP	4
408	67.3	565	79F	0.0046	VC06_03	6	Concrete	3
409	16.2	170	90X	0.0027	SC08_03	33	PVC	2
410	18.2	377	76Q	0.0193	CF02_02	10	HDPE	3
411	86.1	31	78A	0.321	SC10_01	6	VCP	3
412	17.2	208	36R	0.0158	BF06_02	10	PVC	1
413	12.3	350	62D	0.0153	MC02_02	24	PVC	2
414	0.9	306	90N	0.0061	CF06_07	8	HDPE	1
415	2.9	121	89M	0.035	CF06_07	8	PVC	1
416	2.9	244	89R	0.0051	CF06_07	8	PVC	1
417	15.7	25	90V	0.0088	SC08_01	35	Concrete	1
418	29.1	89	76K	0.0367	CF03_01	8	PVC	1
419	23.2	228	76P	0.0233	CF03_01	8	PVC	1
420	5.8	147	76P	0.0081	CF03_02	8	HDPE	1
421	76.8	107	76P	0.0561	CF03_02	6	VCP	3
422	78.6	581	76S	0.0455	CF03_01	8	HDPE	2
423	26.8	914	76K	0.0012	CF03_01	30	Concrete	1
424	28.8	61	76J	0.0013	CF04_01	30	DI	1
425	0.3	306	75G	0.0229	CF10_04	8	PVC	1
426	0.6	287	62W	0.004	WF02_03	8	PVC	1
427	21.6	176	72A	0.0434	WF05_05	8	PVC	1
428	48.6	245	89K	0.0486	CF05_03	6	Concrete	3
429	13.4	23	80K	0.2444	VC01_04	12	DI	2
430	13.4	98	80J	0.0041	VC01_04	8	DI	2
431	11.3	112	75K	0.0433	CF11_07	12	PVC	1
432	34.7	381	103F	0.0082	SC09_04	10	VCP	2
433	33.5	403	103F	0.0043	SC09_04	8	VCP	2
434	6.9	437	63Q	0.0097	BF04_03	8	HDPE	1
435	52.9	547	93C	0.0004	VC08_01	39	Concrete	2
436	3.2	75	75S	0.0144	CF11_06	10	PVC	1
437	7.5	550	71R	0.0022	PCF15_01	36	PVC	1
438	6.2	466	78F	0.0147	SC10_04	8	PVC	1
439	8.3	142	76J	0.026	CF10_05	8	PVC	1
440	10.7	40	90J	0.0538	SC05_05	8	HDPE	1
441	63.7	61	79D	0.0643	VC06_01	6	VCP	2
442	18.6	65	76D	0.0155	CF01_05	8	DI	1

443	35.7	269	91G	0.002	SC04_01	27	VCP	2
444	0.6	17	62W	0.0535	WF02_03	15	PVC	1
445	7.8	45	62X	0.0091	WF02_03	24	DI	1
446	42.1	3	62W	0.7567	WF02_03	8	Concrete	2
447	9.6	221	67U	0.0101	VC03_01	8	PVC	1
448	13.8	381	67V	0.0074	VC02_01	8	PVC	1
449	3.9	291	21N	0.0051	BF09_03	8	PVC	1
450	2.9	118	89M	0.0208	CF06_07	8	PVC	1
451	2.9	139	90N	0.0069	CF06_07	8	PVC	1
452	2.9	73	90N	0.0715	CF06_07	8	PVC	1
453	2.9	27	90N	0.0111	CF06_07	8	PVC	1
454	15.7	377	90V	0.005	SC08_01	35	Concrete	2
455	16.2	286	90X	0.0041	SC08_03	27	PVC	2
456	2.9	26	90N	0.0223	CF06_07	8	PVC	1
457	13.5	56	76L	0.0004	CF02_02	6	PVC	1
458	19.5	152	76T	0.023	CF03_02	8	VCP	3
459	24.5	161	76P	0.0037	CF03_02	21	PVC	1
460	13.8	51	76T	0.0165	CF03_02	12	PVC	1
461	12	378	76Q	0.0116	CF02_02	8	HDPE	1
462	20.7	79	66N	0.0062	VC02_03	8	PVC	1
463	85.7	234	63Y	0.0034	BF04_01	15	Concrete	4
464	90.6	297	63Y	0.0019	BF04_01	15	Concrete	3
465	16.5	236	46H	0.0139	MC04_04	8	PVC	1
466	22.3	795	79Z	0.0005	VC07_01	54	Concrete	2
467	3.5	133	93L	0.001	VC09_05	54	Fiberglass	2
468	15.2	400	106S	0.0032	VC11_01	24	PVC	2
469	54.9	329	81A	0.007	VC03_03	18	VCP	3
470	1.8	496	73Z	0.0106	CF12_08	8	PVC	1
471	7.8	17	62X	0.0035	WF02_03	24	DI	1
472	6.2	127	62W	0.0039	WF02_03	8	DI	2
473	7.5	352	76B	0.0047	WF02_01	12	PVC	1
474	0.8	260	79H	0.0058	VC01_05	8	PVC	1
475	49.8	370	91G	0.0009	SC03_01	45	Concrete	2
476	28.6	1200	65G	0.0008	BF01_04	54	Concrete	3
477	46.1	591	93X	0.0003	VC10_01	36	Concrete	1
478	55.4	1023	63Z	0	BF04_01	72	Concrete	4
479	17.1	191	32T	0.1382	MC05_02	8	PVC	1

480	17	80	32N	0.1088	MC05_02	8	PVC	1
481	86.6	492	76N	0.0012	CF04_01	24	Concrete	5
482	33.4	154	80D	0.0468	VC03_03	6	VCP	2
483	13.2	114	76V	0.0487	CF02_04	8	PVC	1
484	48.5	187	89K	0.03	CF05_03	6	Concrete	2
485	13.7	112	80K	0.0222	VC01_03	10	PVC	2
486	48.7	95	89G	0.0121	CF05_03	6	VCP	2
487	82.5	695	62B	0.021	MC03_02	6	Concrete	4
488	15.9	200	49U	0.0053	BF02_02	8	PVC	1
489	17.5	121	49U	0.0052	BF02_02	8	PVC	1
490	90.9	479	62D	0.0129	MC02_02	10	VCP	2
491	16.2	450	90X	0.0042	SC08_03	27	PVC	2
492	13.7	251	90N	0.0477	CF06_07	8	PVC	1
493	1.1	441	75M	0.0126	CF10_02	8	PVC	1
494	3.6	504	75L	0.0396	CF10_05	8	PVC	1
495	0.1	181	62X	0.0033	WF02_03	10	DI	4
496	66.2	7	62W	0.0071	WF02_03	6	Concrete	1
497	43.2	77	89F	0.0812	CF05_03	6	Concrete	2
498	19	223	103D	0.0022	SC09_06	15	PVC	1
499	35.7	274	103H	0.0289	SC09_02	6	PVC	1
500	63.2	78	74P	0.0513	CF12_05	6	Concrete	3
501	7.5	395	71L	0.0039	PCF15_01	36	PVC	1
502	7.5	400	71Q	0.0038	PCF15_01	36	PVC	1
503	45.7	97	89W	0.0078	CF07_05	12	Concrete	2
504	2.1	199	22W	0.01	BF08_02	8	PVC	1
505	3.5	178	22B	0.0039	DC02_03	8	PVC	1
506	34.4	257	31H	0.0023	MC05_04	12	VCP	1
507	26.6	130	62N	0.01	MC06_02	6	VCP	1
508	13.6	242	93U	0.0003	VC10_01	48	Concrete	1
509	6	233	76D	0.0028	CF01_04	10	PVC	2
510	13.8	128	76D	0.0345	CF01_05	24	DI	2
511	18.5	354	76D	0.0635	CF01_04	10	HDPE	1
512	44.1	406	64C	0.0027	BF02_01	33	Concrete	2
513	28.7	746	65B	0.0008	BF01_04	54	Concrete	2
514	26.3	98	74Y	0.0028	CF12_02	10	DI	1
515	41	504	79X	0.003	VC07_02	18	Concrete	3
516	41	385	79T	0.003	VC07_02	18	Concrete	3

517	15.9	74	88C	0.0015	CF12_01	12	PVC	1
518	10.9	84	119Z	0.0066	VC11_03	8	PVC	1
519	46.2	18	90A	0.0567	CF04_03	15	Concrete	3
520	96.5	127	76D	0.01	CF01_07	6	VCP	2
521	0.3	184	75G	0.0051	CF10_04	8	PVC	1
522	0.8	279	79H	0.0157	VC01_05	8	PVC	1
523	13.8	161	76Q	0.0086	CF02_02	8	PVC	3
524	20.7	292	66N	0.0199	VC02_03	8	PVC	1
525	19.1	211	62B	0.015	MC03_02	8	PVC	2
526	86.5	118	62D	0.0403	MC02_02	10	VCP	4
527	0.9	168	90N	0.0136	CF06_07	8	HDPE	1
528	12.9	22	35U	0.1277	BF09_02	8	PVC	1
529	12.9	63	35U	0.0656	BF09_02	8	PVC	1
530	3.6	16	21Y	0.5338	BF09_03	8	PVC	1
531	28.4	70	76C	0.0007	CF10_01	54	Concrete	4
532	65.8	147	77A	0.0495	CF01_07	8	VCP	2
533	14.6	150	32B	0.0073	MC05_08	24	DI	2
534	3.2	57	75W	0.004	CF11_06	8	PVC	1
535	3.2	349	75W	0.0029	CF11_06	10	PVC	1
536	34.7	162	36W	0.004	BF05_01	8	VCP	1
537	2.1	502	22W	0.005	BF08_02	8	PVC	1
538	2.1	146	22W	0.0199	BF08_02	8	PVC	1
539	28.7	501	65G	0.0008	BF01_04	54	Concrete	3
540	87.9	176	62F	0.0035	MC06_04	18	Concrete	2
541	22.2	239	93L	0.0008	VC09_05	54	Concrete	1
542	43.6	525	64D	0.0018	BF02_01	35	Concrete	2
543	44.1	679	64C	0.0027	BF02_01	33	Concrete	3
544	4	459	92C	0.0465	VC08_04	8	PVC	4
545	45	383	63Y	0.004	BF04_01	8	VCP	2
546	13.5	200	78X	0.0191	SC11_03	8	PVC	1
547	13.4	274	80K	0.0025	VC01_04	16	DI	1
548	48.5	9	66Y	0.0233	VC04_01	6	Concrete	2
549	27.6	82	65V	0.039	VC02_04	6	VCP	3
550	19.8	71	78J	0.0051	SC11_01	27	PVC	1
551	22.3	1384	93C	0.0005	VC08_01	54	Concrete	3
552	0.2	418	75H	0.0058	CF10_03	8	DI	2
553	2.1	414	22W	0.01	BF08_02	8	PVC	1

554	0.1	71	62P	0.0652	MC06_02	8	PVC	1
555	0.5	57	62N	0.079	MC06_02	8	PVC	1
556	13.8	65	76D	0.1142	CF01_04	30	DI	1
557	14.2	533	63W	0.0049	CF01_07	24	PVC	1
558	4.6	35	76D	0.0046	CF01_05	8	DI	1
559	18.3	338	77A	0.0002	CF01_07	8	Concrete	2
560	15.2	101	76D	0.0045	CF01_05	8	DI	1
561	10.5	329	75R	0.0123	CF04_01	32	DI	1
562	0.9	891	62B	0.0196	MC03_02	8	PVC	2
563	3.5	105	21P	0.0592	BF09_03	8	PVC	1
564	3.4	272	21T	0.0091	BF09_03	10	PVC	1
565	3.4	278	21T	0.0087	BF09_03	10	PVC	1
566	7.5	35	72N	0.0046	CF14_02	36	PVC	2
567	6	80	74Y	0.0013	CF12_02	10	PVC	1
568	6	52	74Y	0.0019	CF12_02	10	PVC	1
569	33.2	165	65U	0.0061	VC02_05	6	VCP	3
570	34.6	271	103F	0.004	SC09_04	8	VCP	1
571	34.6	245	103F	0.004	SC09_04	8	VCP	1
572	38.2	413	103F	0.0042	SC09_04	8	VCP	2
573	39.9	322	103C	0.0334	SC09_06	6	VCP	2
574	30.7	14	65G	0.0021	BF01_04	54	Concrete	4
575	6.3	60	78M	0.0165	SC10_06	8	HDPE	1
576	13.8	95	67U	0.036	VC02_01	8	PVC	1
577	35.5	108	67U	0.0235	VC03_01	6	VCP	1
578	9.6	275	21S	0.0185	BF09_03	8	PVC	1
579	61.7	133	77A	0.0137	CF01_07	8	VCP	3
580	31.9	1303	66G	0.0008	BF01_03	78	Concrete	3
581	45.3	1301	66G	0.0007	BF01_03	96	Concrete	3
582	22.3	232	93G	0.0005	VC08_01	54	Concrete	2
583	100.4	792	62Y	0.0005	CF01_04	48	Concrete	3
584	47.7	291	80D	0.008	VC04_01	6	Concrete	2
585	57.3	373	80A	0.0327	VC04_04	6	Concrete	3
586	73.2	280	62G	0.0018	MC03_02	10	PVC	2
587	13	35	76L	0.1686	CF02_02	8	PVC	1
588	8.8	95	35Y	0.0098	BF05_02	8	PVC	2
589	66.8	182	78M	0.0125	SC10_05	6	VCP	3
590	11.4	39	75R	0.0018	CF04_01	48	PVC	2

591	5.4	62	62F	0.0182	MC03_02	8	DI	1
592	15.9	199	49U	0.005	BF02_02	8	PVC	1
593	17.2	507	36R	0.0125	BF06_02	10	PVC	1
594	67.3	137	62X	0.0058	WF02_03	6	VCP	2
595	14.7	175	62J	0.059	MC06_02	8	PVC	1
596	13.4	787	93Y	0.0003	VC10_01	42	Concrete	3
597	3	174	62E	0.0101	MC06_04	8	DI	1
598	108.7	312	62L	0.0058	MC06_01	6	VCP	2
599	7.2	377	63N	0.0039	MC01_01	12	PVC	1
600	25.3	39	75Y	0.0015	CF05_01	30	Concrete	2
601	4.1	205	35S	0.0512	BF05_03	8	PVC	1
602	87.1	91	90A	0.0033	CF04_03	10	Concrete	2
603	16.8	152	74X	0.0077	CF12_02	16	DI	1
604	29.1	66	89S	0.0117	CF07_04	18	Concrete	2
605	31	462	89E	0.0089	CF05_03	6	PVC	1
606	9.7	144	78H	0.0106	SC10_05	8	HDPE	1
607	66.8	109	78M	0.0275	SC10_05	6	VCP	3
608	8.8	74	35Y	0.0105	BF05_02	8	PVC	2
609	7.9	137	76H	0.0073	CF01_05	8	PVC	1
610	10.9	274	76L	0.0107	CF02_02	8	PVC	1
611	31.1	33	76K	0.0482	CF03_01	8	PVC	1
612	23.2	168	76P	0.022	CF03_01	8	PVC	2
613	13.8	203	76Q	0.0217	CF02_02	8	PVC	3
614	2	107	76P	0.0067	CF03_02	8	PVC	1
615	23.2	173	76P	0.0291	CF03_01	8	PVC	1
616	7.7	130	76T	0.0153	CF03_02	8	PVC	1
617	76.5	138	76N	0.0609	CF03_01	6	VCP	3
618	3.1	353	92D	0.0059	VC08_05	8	DI	1
619	13.5	41	78X	0.0181	SC11_03	8	DI	2
620	102.8	48	62Y	0.0042	CF01_04	21	Concrete	2
621	59.1	304	90N	0.0201	CF06_07	8	PVC	1
622	43.3	950	120C	0.0013	VC11_02	36	VCP	4
623	3	298	79R	0.0064	VC01_06	15	PVC	1
624	11	37	22C	0.0049	DC02_03	8	PVC	1
625	34.4	423	31H	0.0473	MC05_04	6	VCP	2
626	34.4	166	31H	0.0023	MC05_04	12	VCP	1
627	22.2	657	93Q	0.0009	VC09_05	54	Concrete	2

628	13.4	746	93U	0.0004	VC10_01	42	Concrete	1
629	35.4	374	103H	0.0059	SC09_02	6	PVC	1
630	38.7	126	103C	0.0319	SC09_02	6	VCP	2
631	14.4	443	74K	0.0089	CF11_04	8	HDPE	2
632	7.5	400	71L	0.0038	PCF15_01	36	PVC	1
633	6	107	74Y	0.0037	CF12_02	10	PVC	1
634	60.3	205	63D	0.0132	BF03_04	6	Concrete	3
635	41.9	1025	61Q	0.0013	WF01_05	36	Concrete	2
636	3.2	425	75S	0.0148	CF11_06	10	DI	1
637	2.1	103	22W	0.0145	BF08_02	8	PVC	1
638	4.4	431	19Y	0.0035	BF05_08	18	PVC	1
639	0.1	66	62P	0.0379	MC06_02	8	PVC	1
640	3.5	160	22B	0.0039	DC02_03	8	PVC	1
641	0	275	62F	0.032	MC06_01	8	PVC	1
642	22.3	319	93D	0.0005	VC08_01	54	Concrete	1
643	52.6	710	93D	0.0004	VC08_01	39	Concrete	2
644	32.2	17	93C	0.0153	VC08_01	8	DI	2
645	41.9	1146	61Q	0.0013	WF01_05	36	Concrete	2
646	3.2	182	75S	0.0181	CF11_06	8	PVC	1
647	2.1	206	22W	0.005	BF08_02	8	PVC	1
648	4.6	177	62P	0.004	MC06_02	8	PVC	1
649	5.2	54	62N	0.0326	MC06_02	8	PVC	1
650	13.4	37	19Z	0.0995	BF05_08	8	PVC	1
651	14.4	478	17Y	0.0092	MC05_07	8	PVC	1
652	17.9	261	35X	0.0038	BF05_02	10	PVC	2
653	8.9	77	49W	0.0033	MC02_03	8	DI	1
654	12	290	49D	0.0045	BF05_02	8	PVC	1
655	13.8	193	67Q	0.0099	VC02_01	8	PVC	1
656	35.6	308	67U	0.0101	VC03_01	6	VCP	3
657	14	11	49C	0.0191	BF02_02	8	PVC	2
658	19.5	207	80N	0.0291	VC01_05	8	PVC	1
659	26.3	282	74Y	0.0021	CF12_02	10	PVC	1
660	62.8	201	77J	0.0164	CF01_02	6	VCP	3
661	64.9	409	88B	0.0022	CF12_01	12	Concrete	3
662	46.7	27	88H	0.0263	CF07_02	24	Concrete	3
663	4	22	88H	0.0086	CF07_02	24	DI	1
664	31.2	189	88H	0.0064	CF07_02	8	VCP	2

665	39.6	179	67V	0.0182	VC02_01	6	VCP	2
666	9.6	274	21T	0.0066	BF09_03	10	PVC	1
667	4.6	295	76D	0.0105	CF01_05	8	DI	1
668	7.9	106	76H	0.0381	CF01_05	8	PVC	1
669	21.6	365	76K	0.0203	CF02_02	10	PVC	1
670	15.9	192	76K	0.0072	CF02_02	8	PVC	3
671	15.9	169	76K	0.0071	CF02_02	8	PVC	1
672	12	57	76Q	0.0074	CF02_02	8	HDPE	1
673	24.5	25	76P	0.004	CF03_02	21	PVC	1
674	25.3	331	76T	0.009	CF03_01	6	VCP	3
675	87.8	72	62G	0.0043	MC06_04	18	VCP	2
676	24.5	150	72B	0.0099	WF05_05	10	PVC	1
677	6.7	135	90Q	0.0104	SC05_03	8	PVC	1
678	13.8	144	67U	0.0398	VC02_01	8	PVC	1
679	13.8	156	67U	0.0299	VC02_01	8	PVC	1
680	9.6	175	21T	0.0049	BF09_03	10	PVC	1
681	9.6	271	21N	0.0039	BF09_03	8	PVC	1
682	52.9	591	93D	0.0004	VC08_01	39	Concrete	1
683	3.2	351	75S	0.0029	CF11_06	10	PVC	1
684	1.6	3	75H	3.7233	CF10_03	8	PVC	1
685	35.9	133	67V	0.0088	VC03_01	6	VCP	1
686	9.6	277	21T	0.0158	BF09_03	8	PVC	1
687	3.9	344	21N	0.005	BF09_03	8	PVC	1
688	3.9	265	21N	0.005	BF09_03	8	PVC	1
689	21.6	183	76K	0.0028	CF02_02	10	PVC	1
690	23.2	471	76P	0.001	CF03_01	30	Concrete	1
691	19.1	314	76P	0.0055	CF03_01	8	PVC	2
692	24.5	518	76P	0.0039	CF03_02	21	PVC	1
693	13.8	196	76T	0.0039	CF03_02	8	PVC	1
694	13.8	144	67U	0.0313	VC02_01	8	PVC	1
695	35.5	506	67U	0.0317	VC03_01	10	VCP	1
696	35.6	538	67U	0.0393	VC03_01	6	VCP	3
697	9.6	273	21S	0.0277	BF09_03	8	PVC	1
698	10.5	331	21S	0.0204	BF09_03	8	PVC	1
699	10.5	475	21S	0.0316	BF09_03	8	PVC	1
700	13.4	291	20W	0.0093	BF05_08	8	PVC	1
701	11.4	36	35X	0.0042	BF05_03	8	PVC	1

702	4.2	1	35Z	0.42	BF05_02	8	PVC	1
703	15.7	216	91W	0.0173	SC06_01	8	PVC	1
704	26.8	21	76K	0.0005	CF04_01	24	DI	1
705	10.6	425	76X	0.0151	CF03_04	12	PVC	2
706	51.6	511	89C	0.004	CF05_02	24	Concrete	2
707	8.8	40	81J	0.0015	VC01_02	54	DI	1
708	28.8	526	81J	0.0014	VC01_02	54	Concrete	1
709	76.5	154	76W	0.0188	CF03_03	8	PVC	2
710	62.8	115	77J	0.0244	SC01_03	6	VCP	3
711	7.7	44	93C	0.0034	VC08_01	24	Concrete	2
712	16.6	187	88C	0.0026	CF12_01	21	PVC	1
713	6.9	351	75P	0.0356	CF11_02	8	PVC	1
714	6	45	74Y	0.0098	CF12_02	10	PVC	1
715	6	103	74Y	0.0101	CF12_02	10	PVC	1
716	8.8	136	89X	0.0147	CF05_05	8	PVC	1
717	51.2	580	63B	0.0043	MC02_05	8	VCP	3
718	62.5	60	74N	0.05	CF12_03	6	VCP	2
719	25.5	364	93N	0.0231	VC10_01	6	PVC	1
720	4.4	346	63P	0.0015	MC01_01	18	PVC	1
721	15.7	66	77A	0.0174	CF01_03	16	DI	1
722	8.8	239	35Y	0.0091	BF05_02	8	PVC	2
723	17.4	199	78M	0.0309	SC10_05	8	HDPE	3
724	27.8	172	63Z	0.0034	BF04_01	8	VCP	2
725	14.5	214	88H	0.0074	CF07_02	24	DI	2
726	38.9	526	31H	0.003	MC05_04	12	VCP	2
727	22.2	832	93Q	0.0006	VC09_05	54	Concrete	2
728	0.1	405	78R	0.0654	SC11_04	8	HDPE	2
729	35.7	520	103H	0.0173	SC09_02	6	PVC	1
730	35.8	329	103C	0.0273	SC09_02	6	VCP	1
731	7.1	492	71F	0.002	PCF15_01	36	PVC	2
732	7.1	477	71F	0.0021	PCF15_01	36	PVC	1
733	7.5	198	72N	0.0163	CF14_02	36	Fiberglass	1
734	50.4	13	89G	0.1308	CF05_03	8	Concrete	3
735	16.3	482	90X	0.0205	SC08_03	8	PVC	1
736	16.8	258	74X	0.0014	CF12_02	16	DI	1
737	59.2	13	74T	0.1023	CF12_02	6	VCP	3
738	18.9	131	76Q	0.0295	CF02_02	8	PVC	1

739	23.8	298	89Q	0.002	CF06_02	20	DI	1
740	29.1	11	89S	0.0427	CF07_04	18	Concrete	2
741	26	79	89S	0.0029	CF07_04	15	PVC	1
742	13.4	450	74P	0.0834	CF12_05	8	PVC	1
743	29.7	285	88T	0.0218	CF08_02	8	PVC	1
744	27.5	305	102M	0.0279	CF09_04	8	PVC	1
745	16.2	900	90X	0.0025	SC08_03	33	Concrete	1
746	12.7	192	119D	0.0189	VC11_06	10	PVC	1
747	13.5	4	20W	0.005	BF05_08	8	PVC	1
748	3.9	123	21N	0.005	BF09_03	8	PVC	1
749	5	68	63L	0.0382	BF04_02	8	HDPE	1
750	12.5	322	92K	0.008	VC09_04	8	HDPE	1
751	14.6	358	76J	0.0132	CF10_05	8	PVC	1
752	17.3	173	90D	0.0189	SC04_04	8	PVC	1
753	16.9	21	79K	0.0095	VC06_03	8	PVC	1
754	1.2	54	64T	0.0054	BF03_01	8	PVC	1
755	55.4	239	90T	0.005	SC08_02	10	Concrete	2
756	1.5	74	90T	0.0151	SC08_02	8	PVC	1
757	40	354	103G	0.0136	SC09_03	6	VCP	2
758	35.4	93	103H	0.1139	SC09_02	6	PVC	2
759	35.7	197	103G	0.0112	SC09_02	8	VCP	3
760	7.5	350	71L	0.0038	PCF15_01	36	PVC	1
761	11.4	198	77S	0.0482	SC02_05	8	PVC	1
762	6	70	74Y	0.0014	CF12_02	10	PVC	1
763	6.2	411	76T	0.0056	CF03_01	8	HDPE	1
764	12.5	431	76S	0.0488	CF04_01	8	HDPE	1
765	14.2	302	76X	0.0048	CF03_03	18	PVC	2
766	15.1	34	76W	0.0024	CF04_03	8	PVC	1
767	6.9	168	90B	0.0004	CF03_04	8	PVC	1
768	61.2	101	74N	0.0268	CF12_03	8	VCP	2
769	27.5	190	103J	0.0295	CF09_04	8	PVC	1
770	24.7	608	91T	0.0107	SC05_03	8	DI	2
771	25	128	93P	0.0791	VC09_05	6	PVC	1
772	52.9	679	93C	0.0004	VC08_01	39	Concrete	1
773	55.5	111	61Q	0.0045	WF01_05	12	Concrete	3
774	0.6	88	75D	0.0256	CF10_02	8	HDPE	1
775	38	65	76K	0.1663	CF03_01	6	VCP	2

776	33.6	140	76K	0.01	CF02_01	6	CI	1
777	2.7	123	64N	0.005	BF03_02	8	PVC	1
778	5.6	59	75H	0.0324	CF10_03	8	PVC	4
779	3.1	523	19W	0.004	PBF10_01	8	PVC	1
780	14.7	165	62K	0.0052	MC06_02	15	PVC	1
781	14.8	229	62J	0.0228	MC06_02	15	PVC	1
782	11.1	132	22B	0.0049	DC02_03	8	PVC	1
783	0.6	135	62P	0.0028	MC06_02	8	PVC	1
784	0.5	198	62P	0.0034	MC06_02	8	PVC	1
785	15.8	220	62J	0.005	MC06_03	15	PVC	1
786	13.1	450	93Q	0.0003	VC09_05	48	Concrete	1
787	46.2	734	93U	0.0004	VC10_01	36	Concrete	1
788	46.2	795	93U	0.0004	VC10_01	36	Concrete	2
789	16	275	19Z	0.0054	BF05_08	27	PVC	1
790	32.5	1398	67R	0.0005	BF01_01	96	Concrete	2
791	82.3	95	76B	0.033	WF02_01	6	Concrete	4
792	12.9	67	92B	0.0179	SC03_04	8	PVC	2
793	6.7	96	90Q	0.0442	SC05_03	6	PVC	4
794	52.1	15	73Z	0.0727	CF12_09	6	Concrete	1
795	59.5	86	74V	0.1138	CF11_06	6	Concrete	3
796	46.2	392	107B	0.0003	VC10_02	36	Concrete	2
797	26.9	592	106R	0.0082	VC11_01	21	PVC	2
798	22.7	943	80N	0.0005	VC01_04	54	Concrete	5
799	10.2	175	21X	0.034	BF09_03	8	PVC	1
800	11.2	220	76H	0.0085	CF01_05	8	PVC	1
801	34.4	31	103F	0.0394	SC09_04	8	VCP	2
802	15.9	78	76H	0.0076	CF01_06	8	PVC	1
803	3	90	62E	0.0073	MC06_04	8	PVC	2
804	15.2	336	103S	0.0206	CF09_02	15	PVC	1
805	15.7	91	88Q	0.004	CF07_03	8	PVC	2
806	12.5	396	88U	0.0444	CF08_02	8	PVC	1
807	12.1	337	103E	0.0178	SC09_04	8	PVC	1
808	16.4	317	103N	0.0602	CF09_02	8	PVC	1
809	14.8	324	103N	0.0151	CF09_05	8	PVC	1
810	14.8	408	103N	0.0253	CF09_05	8	PVC	1
811	16.8	265	103K	0.0101	CF09_05	8	PVC	2
812	12.3	191	103J	0.0318	SC09_04	8	PVC	1

813	10.6	25	72D	0.0252	WF05_02	18	PVC	2
814	36.4	492	67U	0.091	VC03_01	6	VCP	1
815	13.8	287	67U	0.0264	VC02_01	8	PVC	1
816	14	175	49D	0.0224	BF05_02	8	PVC	2
817	11.4	573	75F	0.0327	CF10_04	8	PVC	1
818	78.7	219	76B	0.0082	WF02_01	6	Concrete	3
819	89.9	99	76G	0.0009	CF02_01	30	Concrete	2
820	52.8	21	79D	0.4586	VC06_01	6	VCP	3
821	26.5	269	63W	0.0126	CF01_07	12	DI	1
822	3.8	78	91W	0.0041	SC06_01	8	PVC	1
823	55.1	141	91Y	0.0186	SC05_07	6	VCP	2
824	10.3	509	21X	0.005	BF09_03	8	PVC	2
825	15.7	214	103N	0.032	CF09_02	8	PVC	1
826	9.8	238	73J	0.0127	CF13_03	12	PVC	1
827	9.9	144	73J	0.0097	CF13_03	12	PVC	1
828	14.8	123	103J	0.0252	CF09_05	8	PVC	1
829	13	90	103N	0.0103	CF09_02	8	PVC	1
830	12.3	205	103E	0.0173	SC09_04	8	PVC	1
831	13	243	103N	0.0538	CF09_02	8	PVC	1
832	31	98	73N	0.0304	CF13_03	6	VCP	1
833	17.6	154	87D	0.0248	CF12_08	8	PVC	1
834	16.3	417	103R	0.004	SC07_02	8	PVC	1
835	0.4	166	79H	0.018	VC01_05	8	DI	1
836	10	312	76K	0.0121	CF02_02	8	HDPE	1
837	2.1	338	22W	0.005	BF08_02	8	PVC	1
838	5.9	292	76E	0.0143	CF10_02	20	DI	1
839	16.5	476	90T	0.0025	SC08_02	8	PVC	1
840	55.4	255	90T	0.004	SC08_02	10	Concrete	1
841	6.3	249	75D	0.0102	CF10_02	8	PVC	1
842	3.2	298	75S	0.004	CF11_06	8	PVC	1
843	2.1	422	22W	0.005	BF08_02	8	PVC	1
844	4.2	277	19Z	0.0122	BF05_08	18	PVC	2
845	0.6	212	62P	0.0492	MC06_02	8	PVC	1
846	3.5	172	22B	0.004	DC02_03	8	PVC	1
847	16.1	17	19Z	0.0018	BF05_08	18	PVC	1
848	10.1	25	67U	0.0104	VC03_01	8	PVC	1
849	24.4	149	104G	0.078	SC06_05	8	PVC	1

850	9.6	265	21S	0.0286	BF09_03	8	PVC	1
851	10.5	188	21N	0.0141	BF09_03	8	PVC	1
852	18	406	47Y	0.0059	MC03_06	8	PVC	2
853	7.8	374	89D	0.0088	CF04_05	8	HDPE	3
854	7.8	123	89D	0.0051	CF04_05	8	HDPE	1
855	7.7	123	89H	0.007	CF04_05	8	HDPE	1
856	61.8	44	90E	0.0068	CF04_05	6	Concrete	3
857	12.9	92	92G	0.0159	VC09_04	8	HDPE	1
858	42.4	299	67U	0.022	VC03_01	6	VCP	2
859	14	220	49C	0.0202	BF05_02	8	PVC	2
860	18	259	66D	0.0017	BF01_03	18	PVC	1
861	9.6	206	21S	0.0134	BF09_03	8	PVC	1
862	3.9	275	21N	0.0041	BF09_03	8	PVC	1
863	61	195	74N	0.0039	CF12_03	8	VCP	1
864	53.4	17	74T	0.0477	CF12_05	6	VCP	3
865	61.2	412	74N	0.0214	CF12_03	8	VCP	2
866	27.5	141	103J	0.0043	CF09_04	8	PVC	1
867	18.6	45	90S	0.0067	SC08_02	8	HDPE	2
868	58.2	141	76J	0.0064	CF04_01	8	Concrete	2
869	6	173	76P	0.0254	CF03_02	8	HDPE	1
870	6	168	76P	0.021	CF03_02	8	HDPE	1
871	14	477	76T	0.0016	CF03_03	18	PVC	1
872	23.2	98	76P	0.0022	CF03_01	15	PVC	1
873	33	111	72C	0.0103	WF05_05	6	VCP	1
874	17.6	45	74Y	0.0093	CF12_02	18	PVC	1
875	12.5	172	92H	0.0026	VC09_02	18	PVC	1
876	43.2	118	89K	0.1057	CF05_03	6	Concrete	2
877	1.1	344	92D	0.0082	VC08_05	8	DI	4
878	33.2	874	47J	0.0023	MC04_04	30	VCP	2
879	56.7	31	73Z	0.0065	CF12_09	6	Concrete	2
880	48.4	222	80C	0.056	VC04_01	6	Concrete	2
881	25.4	358	72B	0.003	WF05_05	10	PVC	1
882	28.3	187	76C	0.0064	CF10_01	36	DI	1
883	11.4	25	76C	0.3732	CF02_01	6	DI	2
884	62.7	534	79D	0.0305	VC06_01	6	VCP	3
885	14	40	62Z	0.009	CF01_07	8	DI	1
886	13.8	268	90D	0.0034	SC04_04	10	PVC	1

887	1.2	70	64T	0.004	BF03_01	8	PVC	1
888	55.4	214	90T	0.0053	SC08_02	10	Concrete	1
889	6.5	57	22T	0.0797	BF08_02	8	PVC	1
890	15.3	396	63Z	0.0041	BF04_01	8	PVC	1
891	22.7	284	77J	0.017	SC01_03	8	PVC	1
892	46.7	250	88H	0.0091	CF07_02	24	Concrete	2
893	9.7	307	90C	0.0235	CF03_04	8	HDPE	1
894	35.9	282	67V	0.0147	VC03_01	6	VCP	1
895	32.7	158	67U	0.0863	VC03_01	6	VCP	3
896	14.3	65	62Y	0.0045	CF01_04	30	DI	1
897	6.2	50	63W	0.032	CF01_07	4	PVC	1
898	4.6	81	76D	0.0496	CF01_05	8	DI	1
899	16.3	53	77A	0.0085	CF01_07	8	DI	1
900	14.8	278	103P	0.015	CF09_05	8	PVC	2
901	14.8	149	103P	0.0101	CF09_05	8	PVC	1
902	0.4	494	79M	0.0216	VC01_05	8	PVC	1
903	19.4	40	91B	0.0155	SC04_01	10	PVC	1
904	17.2	512	90D	0.006	SC04_04	8	PVC	1
905	12.1	470	76K	0.008	CF02_02	8	HDPE	3
906	0.6	108	76E	0.0357	CF10_02	8	PVC	1
907	8.5	289	20V	0.004	BF05_05	8	PVC	1
908	61.8	643	76G	0.0232	CF02_01	6	Concrete	3
909	35.9	429	67V	0.0062	VC03_01	6	VCP	2
910	68.1	234	74P	0.0681	CF12_05	6	Concrete	2
911	27.5	153	103J	0.0046	CF09_04	8	PVC	1
912	15.7	324	103N	0.0146	CF09_02	8	PVC	1
913	16.4	170	102R	0.0555	CF09_02	8	PVC	1
914	63	217	76C	0.0369	CF01_04	6	Concrete	1
915	46.9	1152	67R	0.0005	BF01_01	96	Concrete	3
916	17	24	72D	0.0021	WF05_02	15	PVC	1
917	71.5	163	75M	0.0061	CF10_05	6	VCP	3
918	2.9	180	90N	0.0237	CF06_07	8	PVC	1
919	22.9	72	36L	0.0051	BF06_05	10	PVC	1
920	7.5	226	47L	0.0109	MC07_01	8	PVC	1
921	15.3	381	90U	0.0071	SC08_01	33	PVC	1
922	16.2	302	90X	0.0093	SC08_03	27	PVC	2
923	13.7	181	89R	0.0504	CF06_07	8	PVC	1

924	5.5	183	75G	0.0197	CF10_04	8	PVC	1
925	0.3	187	75G	0.0166	CF10_04	8	PVC	1
926	1	278	80G	0.0342	VC04_02	8	PVC	1
927	0.8	314	79H	0.0162	VC01_05	8	PVC	1
928	5.1	271	79H	0.0156	VC06_01	8	PVC	2
929	12.5	366	92F	0.0061	VC09_04	8	HDPE	1
930	51.6	229	79M	0.0128	VC01_05	10	Concrete	3
931	17.2	495	90D	0.006	SC04_04	8	PVC	1
932	1.2	208	64T	0.004	BF03_01	8	PVC	1
933	34.3	3	91X	0.42	SC06_01	10	CI	3
934	28.2	364	93N	0.0576	VC10_01	6	PVC	1
935	5.5	83	49M	0.0052	BF02_02	10	PVC	1
936	14	64	92K	0.0066	VC09_04	16	DI	1
937	39.7	430	72D	0.0096	WF05_02	18	VCP	2
938	32.7	287	91V	0.0051	VC10_01	10	PVC	3
939	0.4	261	78U	0.0417	SC11_05	8	PVC	1
940	6	329	20V	0.004	BF05_05	8	PVC	2
941	1.5	322	22T	0.0449	BF08_02	8	PVC	1
942	10.6	95	92L	0.0688	VC09_03	8	DI	1
943	52.9	355	92H	0.006	VC09_02	18	Concrete	3
944	17	238	73Z	0.0114	CF12_09	12	DI	1
945	26.1	112	74Y	0.0021	CF12_02	12	PVC	1
946	12.8	45	75M	0.11	CF10_02	12	PVC	1
947	15.9	94	76H	0.0072	CF01_06	15	PVC	1
948	3	181	48Y	0.0113	MC03_01	12	PVC	1
949	82.6	225	90A	0.0079	CF04_03	8	Concrete	4
950	23.2	75	76P	0.0029	CF03_01	15	PVC	1
951	3.4	83	76T	0.0039	CF03_02	8	PVC	1
952	12.9	128	76W	0.0032	CF04_03	8	PVC	1
953	11.3	33	76A	0.0112	WF02_02	6	PVC	1
954	62.7	460	79D	0.0529	VC06_01	6	VCP	1
955	14.4	67	76D	0.0091	CF01_05	24	DI	2
956	13	153	103N	0.0249	CF09_02	8	PVC	1
957	11.2	412	103T	0.0175	CF09_03	8	PVC	1
958	11.2	403	103T	0.0051	CF09_03	8	PVC	1
959	13.7	446	103N	0.011	CF09_05	8	PVC	1
960	27.5	160	103J	0.0263	CF09_04	8	PVC	1

961	16.6	126	87Z	0.0049	CF08_05	8	PVC	1
962	27	185	103N	0.0213	CF09_04	8	PVC	1
963	15.7	276	103P	0.0059	CF09_02	8	PVC	1
964	16.4	463	102R	0.0324	CF09_02	8	PVC	1
965	9.8	164	73J	0.0093	CF13_03	12	PVC	1
966	14.8	475	103P	0.009	CF09_05	8	PVC	1
967	13	359	103N	0.0598	CF09_02	8	PVC	1
968	34.1	131	73J	0.0061	CF13_03	6	VCP	2
969	15.8	190	76D	0.0077	CF01_05	24	DI	1
970	17.9	103	61D	0.0097	MC03_06	8	PVC	1
971	8.1	125	89D	0.0049	CF04_05	8	HDPE	1
972	16.3	296	74Y	0.0029	CF12_02	10	CI	1
973	13.2	120	92G	0.025	VC09_03	8	HDPE	1
974	12.4	530	92L	0.0063	VC09_03	8	PVC	1
975	13.2	127	92L	0.004	VC09_03	8	PVC	1
976	10.4	265	92H	0.0229	VC09_01	8	HDPE	1
977	46	303	73U	0.0492	CF12_09	6	VCP	2
978	33	28	67U	0.0064	VC03_01	6	VCP	1
979	13.8	143	67R	0.049	VC02_01	8	PVC	1
980	14	305	49C	0.0114	BF05_02	8	PVC	2
981	3.9	344	21N	0.0322	BF09_03	8	PVC	1
982	9.6	264	21N	0.0041	BF09_03	8	PVC	1
983	27.6	164	103N	0.0123	CF09_04	8	PVC	1
984	27.5	96	103J	0.0309	CF09_04	8	PVC	2
985	45.2	179	62L	0.0112	MC06_01	21	Concrete	1
986	42.2	27	65U	0.0074	VC05_01	6	VCP	2
987	25.3	92	75Y	0.0019	CF05_01	30	Concrete	1
988	4.1	244	35S	0.0059	BF05_03	8	PVC	1
989	12	48	78D	0.2383	SC10_01	8	HDPE	1
990	9.9	27	73N	0.0044	CF13_03	8	PVC	2
991	16.5	312	103T	0.0382	SC07_06	8	PVC	1
992	12.3	308	103J	0.0098	SC09_04	8	PVC	1
993	10.6	102	103T	0.0051	CF09_03	8	PVC	1
994	11.2	337	103T	0.005	CF09_03	8	PVC	1
995	9.9	298	103W	0.0049	CF09_03	8	PVC	1
996	16.3	248	103R	0.004	SC07_02	8	PVC	1
997	11.8	301	103T	0.0162	CF09_03	8	PVC	1

998	7.7	101	93C	0.0033	VC08_01	24	Concrete	1
999	31.3	473	80S	0.0024	VC07_01	6	VCP	4
1000	46.2	327	93X	0.0004	VC10_01	36	Concrete	2