Arab American University
Faculty of Graduate Studies

Fair Water Distribution Network Using An Intelligent Spatial Support System

By

Doha Abu Arra

Supervisor

Prof. Amjad Rattrout

Co- Supervisor

Dr. Jacqueleen Joubran

This thesis was submitted in partial fulfillment of the requirements for the master`s degree in Computer Science
July / 2024

Thesis Approval

Fair Water Distribution Network Usingan Intelligent Spatial Support System

By
Doha Abu Arra

This thesis was defended successfully on 6 / 7 / 2024 and approved by:

Committee members Signature

1. Dr. Amjad Rattrout : Supervisor
2. Dr. Jacqueleen Joubran: Co- Supervisor
3. Prof . Adwan Yasin: Internal Examiner
4. Dr. Rashid Jayousi: External Examiner

Declaration

I certify that the work provided in this thesis, unless otherwise referenced, is the researchers work and has not been submitted for a higher degree to any other university or institution.

Student: Doha Abu Arra

Student ID: 202120167

Signature: Doha

Date: 7 / 10 / 2024

Dedication

First of all thanks God for being alive to use my abilities to help my country and solve its problems concerning water.

I dedicate this thesis to my parents, whose unwavering support and encouragement have been my guiding light throughout this journey [Mohammad Rushdi Abu Arra & Eman Ahmad Abu Arra], for their unending love, sacrifices, and belief in me. To my siblings, [Tuqa, Rushdi, Fatima, Wafa’a, Wala’a and Abdulrahman], and the lovely baby [Elias] for their constant encouragement and understanding.

To all my Doctors and colleagues who have supported me along the way, thank you for your companionship and inspiration. Even though it was hard to forget, for my friends, I am so very thankful. Who joined me on this trip, and for all the anonymous soldiers who successfully contributed this behind-the-scenes work, I'm grateful to you all; if you had not been there, I could not have done this.

Acknowledgments

My heartfelt appreciation goes out to everyone who helped and advised me while I finished this thesis.

First and foremost, I would like to thank my supervisors, [Prof Amjad Rattrout], [Dr Jacqueleen Joubran] for their invaluable guidance, insightful feedback, and constant encouragement. Their expertise and support have been crucial to this work's success.

I am also grateful to [Esra’a] for her time, constructive criticism, and insightful recommendations that enhanced the caliber of this study.

Special thanks to my colleagues and friends at [Arab American University] for their camaraderie and assistance during challenging times. Their collaboration and support have been a source of motivation and inspiration.

I would like to acknowledge the municipality of Nablus, especially the engineer [Areej] and the municipality of Aqqaba, especially the engineer [Salam], which made this research possible.

Finally, I want to express my sincere gratitude to my family for their constant support, tolerance, and love during this journey. My biggest strength has been their support.

Thank you all.

Abstract

The existence of water is vital because it is necessary for all living things. Due to their fixed and limited allocations, Palestinians currently face a number of obstacles in their quest for sufficient and safe drinking water. Many municipalities have water data, but it is not arranged in a way that makes decision-making and problem-solving easier. Therefore, all water data needs to be collected, edited, and managed to be analyzed and useful information can be drawn from it.

Water resource management is the management, development, distribution, planning and of water resources to guarantee their optimal utilization. A water distribution network's awareness is increasing through the use of GIS. Using water network analysis and GIS, we can locate and identify pipe breaks as soon as they happen. In addition to handling all of the data related to the sporadic water supply networks in Palestine and other nations. My goal is to locate a system that can evaluate Nablus' water distribution system.

Data from the study area (industrial zone) were gathered for my thesis. All of the data I have gathered were not GIS data; rather, they were converted from AutoCAD format to GIS data, which meant that they included many issues that needed to be fixed. Data must be organized in order for the developer to perform analysis. I have made numerous edits to the data, including defining simple, complex, and connected edges and junctions. Water data consist of Customer meters (water destination), Source data, Valves data, and Pipe data (main pipes, House Connection pipes). ArcMap software and (GIS) were combined to initialize a water network that supports the intermittent supply system. The purpose of the geometric network is to identify loops and other issues like finding disconnected or connected network elements, locating disconnected network parts and figuring out how to reconnect them, as well as to perform network analysis and verify the connectivity between network components.

To get a fair water distribution system, it was necessary to first ensure the safety and correctness of the connection of the network components. Secondly, to establish a predicting system based on historical data for maintenance operations, this system used C# programming language to apply data mining algorithms, and it used ArcMap software, to apply spatial data mining analyses that take into account the exact coordinates of each component and the spatial relations of the components around it, in particular the Ordinary least squares analysis, Lastly, the network is examined using the breadth-First algorithm and isolation algorithms to identify each network component's controlled valve, which the intelligent system will then propose to close in the event that the related component malfunctions.

The results showed that the use of data mining techniques and spatial data mining techniques in the establishment of a prediction system based on historical data for maintenance operations to determine which components may need maintenance. The results also showed that the use of visualization algorithms in network analysis makes the system able to identify the controlling valve in each element to propose its closure in case the element breaks down; these results help save time and money and reduce water loss.

Table of Contents

Thesis Approval..I
Declaration..II
Dedication..III
Acknowledgments..IV
Abstract..V
List of Tables..X
List of Figures..XI
List of Appendices..XV
List of Abbreviation..XVI
Chapter One..1
Introduction..1
1.1 Historical Perspective..1
1.2 Motivation..2
1.3 Research problem..4
1.4 Contribution..6
1.5 Study Area..10
1.6 Usage software..11
1.7 Methodology..12
Chapter Two..15
State of Art..15
2.1 Water supply system..15
2.2 Related Work..17
2.3 Geographic Information system..31

2.4 water network as a distributed system and complex adaptive system (CAS) 34
Conclusion 37
Chapter Three 38
Our Approach for building water network 38
3.1 Water System Geo database 38
3.2 Water Geometric Network 39
Conclusion 46
Chapter Four 47
Data Preprocessing 47
4.1 Spatial Data Preprocessing 47
4.2 Topology Rules 57
4.3 Meters Maintenance Data: 61
4.4 House Connection Pipes Maintenance Data 63
4.5 Calculation of Approximate Cost of Maintenance: 65
Conclusion: 68
Chapter Five 69
Data Mining and Spatial Data Mining 69
5.1 Data Mining 69
Conclusion: 74
5.2: Spatial Data Mining: 75
Conclusion: 95
Conclusion: 105
5.3 Control Valve: 105
Conclusion: 112
Chapter Six 113

Our Desktop Application: 113
6.1 The Application 113
6.2: Testing and Validation: 138
Conclusion 141
Chapter Seven 142
3D Simulation 142
Chapter Eight 146
Conclusion & Recommendations 146
Conclusion 146
Recommendation 147
References 149
Appendices 170
الملخص 201

List of Tables

Table 1: Main Pipe Prices (HP: high-performance polypropylene). 66

Table 2: House Connection Prices (HDPE: High-density polyethylene). 67

Table 3: Meters Prices. 67

Table 4: Valves Prices. 67

Table 5: Testing and Validation Results. 140

List of Figures

Figure 1: Nablus City [108] 2

Figure 2: : (b) Nablus City and (a) its boundaries [108] 11

Figure 3: Spatial Decision Support System. 34

Figure 4: The Geometric Network Component How it is Manipulate And connect to provide the final Geometric network with all connected component [44]. 42

Figure 5: The Geometric network build errors. 45

Figure 6: An example of error 11 in GN & Figure 7: An example of error 16 in GN 45

Figure 8: Detecting Loops. 46

Figure 9: Data Before Modified. 48

Figure 10: (a) Meters without connections to main pipe. (b) The map with the aerial image. 49

Figure 11: The Map Complexity. 50

Figure 12: Random distributed Pipes 51

Figure 13: Disconnected Meters. 52

Figure 14: a: Loops in Pipes 53

Figure 15: b: Loops in Pipes. 53

Figure 16: a: Bent Pipes. 54

Figure 17: b: Bent Pipes. 55

Figure 18: The Geometric Network Errors Number. 56

Figure 19: Building Errors Table 56

Figure 20: Building Errors Number in the first processing phase. 57

Figure 21: Track the correct water flow in the network. 60

Figure 22: The Geometric Network Connected and Disconnected using the Utility Network Analyst toolbar. 61

Figure 23: Meter Maintenance Data Before Modified. 62

Figure 24: Meter Maintenance Data After Modified. 63

Figure 25: House Connection Pipes Historical Maintenance Data before Modification. 64

Figure 26: House Connection Pipes Historical Maintenance Data after Modification. 65

Figure 27: How the Decision Tree Algorithm Work [95]. 72

Figure 28: The Workflow Structure to Obtain Results by OLS. 79

Figure 29: Meters Original Data 81

Figure 30: The contour line layer. 82

Figure 31: Convert Contour Layer into Raster. 83

Figure 32: Extract Values to Point. 84

Figure 33: Spatial Join between the meters and the height points. 84

Figure 34: Meter Numerical parameters. 85

Figure 35: Near tool to calculate distance between meters and height point. 86

Figure 36: Reading_12023 statistics. 87

Figure 37: The OLS analysis for the Meters Pipes. 88

Figure 38: OLS meters analyst results. 91

Figure 39: Histogram of Standardized Residuals for meters analyst. 93

Figure 40: Residual vs. Predicted Plot for meters analyst. 94

Figure 41: Spatial Distribution for Meters Analyst Result. 95

Figure 42: House Connection Original Data. 96

Figure 43: Nablus Street Data. 97

Figure 44: Using Near Tool to Find the Nearest Street to the House Connection Pipe. 98

Figure 45: The Parameter which should be deleted. 99

Figure 46: The House Connection Parameters Statistics. 100
Figure 47: The OLS analysis for the House Connection Pipes. 101
Figure 48: The OLS analysis Results for house connection pips. 102
Figure 49: Histogram of Standardized Residuals for house connection analyst. 103
Figure 50: Residual vs. Predicted Plot for house connection analyst. 104
Figure 51: Spatial distribution for house connection analyst result. 105
Figure 52: The control Valve Not the Nearest Valve. 108
Figure 53: The graph for the distribution. 111
Figure 54: The App diagram. 113
Figure 55: View the network in the app. 115
Figure 56: Saved Data in the app dataset. 116
Figure 57: Application interface. 116
Figure 58: street background. 117
Figure 59: Aerial Image background. 118
Figure 60: Topographic Background. 118
Figure 61: Maintenance Pop-up in many cases. 121
Figure 62: Fix History. 121
Figure 63: Generate Valves Dependency. 123
Figure 64: Generate Valves Dependency pop-up. 124
Figure 65: Generate Valves Dependency Tree. 124
Figure 66: The Massage will be send to the consumers. 125
Figure 67: The network tracing. 127
Figure 68: The Meters Factors Relations. 129
Figure 69: The House Connection Factors Relations. 130
Figure 70: Predict Setup Tool. 131

Figure 71: Predict Setup Window. 132
Figure 72: Predict to Maintenance Tool 133
Figure 73: Predict to Maintenance Tool Results. 134
Figure 74: Meters Data tree with max depth = 5 135
Figure 75: Meters Data tree with max depth = 6 136
Figure 76: Fixes History 137
Figure 77: Searching in Fixes History. 137
Figure 78: Converting the Industrial Zone Contour Layer into Raster. 143
Figure 79: Activate Base Heights Property for the Raster 144
Figure 80: Network 3D Simulation. 145

List of Appendices

Appendix	Page
Appendix 1 Ordinary least square tests	170
Appendix 2 Apps code snippets	173
Appendix 3 Diagrams	179

List of Abbreviation

No	Acronym	Full Name
1	PCBS	Palestinian Central Bureau of Statistics
2	GIS	Geographic information system
3	3D	3 Dimension
4	ASAP	as soon as possible
5	masl	meters above sea level
6	SDK	software development kit
7	LCA	life-cycle assessment
8	WDNs	water distribution networks
9	DT	Digital Twins
10	AMR	Automated Meter Reading
11	CMMS	Computerized Maintenance Management System
12	WUs	Water Utilities
13	AI	artificial intelligent
14	LoRa	Long Range
15	RFID	Radio Frequency Identification
16	IoT	internet of thing
17	IFIS	Integrated Financial Information System
18	DSS	decision support system
19	ANFIS	Adaptive Neuro-fuzzy Inference System
20	SF	smart farming
21	ILP	Identification, Localization, Pinpoint
22	IFC	Industry Foundation chapters
23	BIM	Building Information Modeling
24	CityGML	City Geography Coding Language
25	ADE	Application Domain Extensions
26	DBMS	Database Management System
27	CAD	Computer-Aided Design
28	UKE	University Medical Center Hamburg-Eppendorf
29	dEBC	demand edge between central
30	SDSS	spatial decision support system
31	DEMs	Digital Elevation Models
32	CAS	complex adaptive system
33	GPS	global position system
34	DMA	District Metered Area
35	SQL	Structured query language
36	SR	spatial relation
37	HP	high-performance polypropylene
38	HDPE	High-density polyethylene
39	DT	Decision tree
40	RF	Random Frost
41	SVM	support vector machines
42	NN	neural network

43	FP-Growth	Frequent Pattern Growth
44	LOF	local outlier factor
45	t-SNE	t-Distributed Stochastic Neighbor Embedding
46	PCA	Principal component analysis
47	DBSCAN	Density-based spatial clustering of applications with noise
48	ID3	Iterative Dichotomies
49	Weka	Waikato Environment for Knowledge Analysis
50	KNIME	Konstanz Information Miner
51	OLS	Ordinary Least Squares
52	AICc	Akaike's Information Criterion
53	BFS	Breadth-First Search
54	MST	Minimal Spanning Tree
55	ESRI	Environmental Systems Research Institute

Chapter One

Introduction

1.1 Historical Perspective

Water is essential for everything, and because of development and population growth, we must conserve and deal with water resources wisely. Water loss in Palestine is serious, the Palestinian Central Bureau of Statistics (PCBS) cited that in 2021, the amount of water available in the case study (Nablus) is 14.4 mm3, and the losses are about 5.4 mm3, the percentage of loss is about 39% which is a significant proportion. GIS (Geographic information system) is used as a query instrument for vast amounts of geographical data, which has become increasingly relevant in various sectors such as the water department, telecommunications networks, gas networks and transportation. All water agencies need a system to handle all their data , detect problems, facilitate the service for the consumers and the suppliers and save effort money and time, Because water systems contain spatial data, GIS is considered as the major tool for water management.

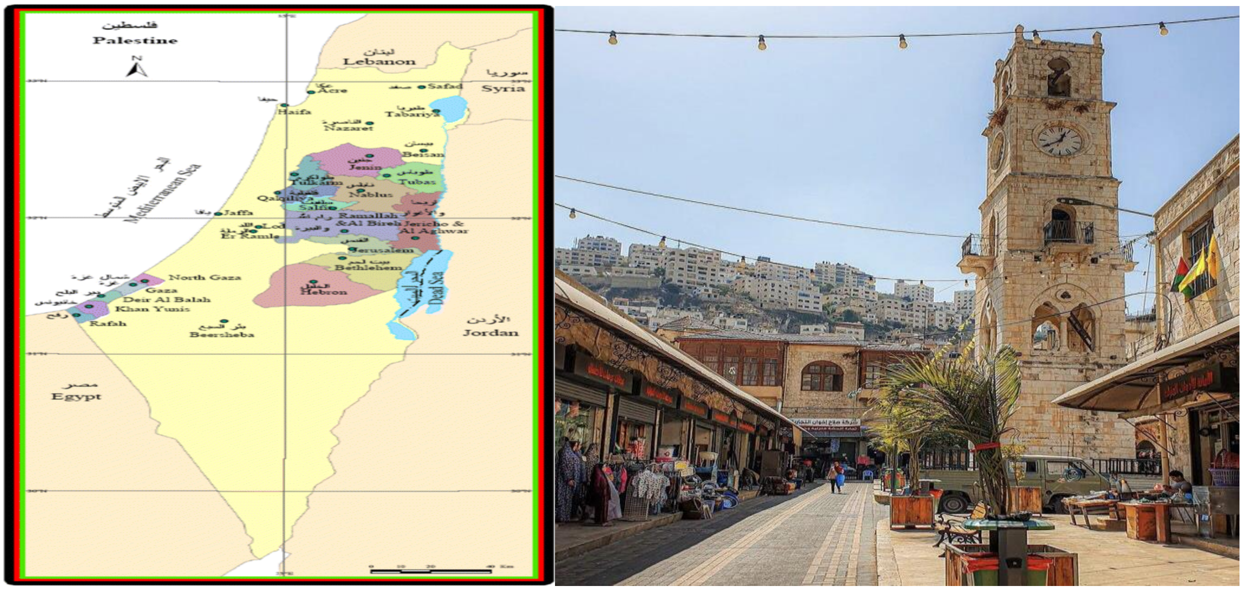


Figure 1: Nablus City [108]

1.2 Motivation

In fact, water data has been kept as mere deaf data, which cannot be used for analysis or decision-making, which are critical concerns for improving existing water systems and solving current problems such as water distribution. There are many severe issues affecting the Palestinian water system, including poor infrastructure, Israeli occupation, unethical use of water or waste, the consequences of climate change, and poor municipal water management.

Network theory is the study of charts, which reflects relationships between separate objects. Several factors prompt me to use GIS technology alongside a 3D (3 Dimension) analyst to build a coordinated water network. When the network regulates and controls the water system's data, we will be able to perform many analyses and make many judgments.

Water distribution systems in large cities, in particular, need ongoing follow-up and monitoring to ensure equitable access to all users, and to ensure the network's safety and components. Therefore, in this research, we propose a fair distribution system based on advanced smart spatial decision support systems. We'll use Nablus city as a case study.

Our approach is to propose new mechanisms for prudently managing the water engineering system and solving distribution and maintenance problems using spatial and historical data. The proposed research methodology is the use of a smart resolution spatial support system that builds a special architecture for the engineering network to connect the components of the system considering the components' characteristics, water direction and engineering and physical rules. GIS is used to model the network, taking into account all components of the network such as locations, quantities of water, physical components, etc. We will consider conflicts of interest and priorities according to the types of service providers, etc.

Each feature is displayed in the water distribution system will be worked on and collected. All this relevant data will be preserved and managed in a geodatabase where each type of data will be preserved as a layer as a form file in the basic GIS format. The data consists of geometric forms and every geometric feature connected to its record in the database as a theme table. Layers for modeling water distribution systems are: (pipes, junctions, valves, and meters) monthly database of meters (these will all be connected to geographical location).

Our approach proposes GIS techniques to preserve, digitize, analyze, manage and visualize the layers handled from spatial data, 3D models will be built for the study case area, and the proposed model will detect problems and optimal solution with minimum cost and minimum effort. The optimal solution that will reveal the source of the problem and the minimum area that should be affected and that will be effectively and wisely addressed, photographed and presented to give a better view of the system of equitable distribution of water and the surrounding spatial effects.

Management of the water distribution network through a spatial engineering network will use all analysis methods to obtain important solutions to solve problems of interruption or water loss or pipe failure.

C# will be used as a programming language for network analysis as a tree structure for calculating and managing water and pressure trends. But research will focus on water distribution and advanced calculations to find a formula for equitable and intelligent distribution. Spatial data mining methods will also be applied to analyze historical data of problems and maintenance problems to be able to design, predict and propose smart future plans, and the decision support system will rely on machine learning techniques to obtain the final decision.

Beside to the ability to monitor, analyze, and manage restructure the network in a way that increases the effectiveness of water distribution and smart plans, helping to reduce economic and technical loss and maintenance work.

1.3 Research problem

Water management is critical since the demand for water is growing in a variety of human and economic activities. This study attempts to leverage current data sources to derive meaningful information about the network and user actions. Furthermore, the project will leverage existing data on the pipe network to create a model of the network, or at least a subset of it, this is to follow the water flow and propose solutions to the problems.

Data in governments or other organizations such as electricity, telecommunications and transportation companies require an intelligent decision-making system. We proposed a concept to create an engineering network for the water department with an intermittent supply system. The engineering network helps the water network developer locate all parts connected or not connected to the network and determine how to manage them. In case of intermittent supply, the municipality strives to distribute water to all regions with the quantity (demand) needed for all consumers in order to achieve fair water distribution. The pipes in the municipality vary in age, length and diameter; all the above factors affect pipe performance. In ArcMap, we utilized a computer model to simulate the water distribution network. The municipality receives recommendations from the research findings.

Water governance should depend on substantive aspects such as spatial location, determination of user behavior (customer meters), pipe network modeling, etc. The Nablus Municipal Water Department keeps records of customers' use (individually and collectively). In fact, its staff organizes regular visits meters to check the water readings on the meters.

Maintenance problems in municipalities are one of the most important problems to pay special attention to because they require a great deal of effort to determine the part to be maintained without the financial cost and time needed for maintenance and are lost until the completion of the operation, which means that water services are interrupted for many consumers. This is done using various methods, including data mining techniques and spatial data mining techniques that will be relied upon in this research, especially the linear regression algorithm and ordinary least square analysis.

We can formulate my research question as follows:

Can we rely on data mining algorithms and spatial data mining algorithms to create a prediction system that can identify components that may need maintenance based on historical maintenance data?

Can the system analyze the water distribution network and propose a controlling valve in each component to close this valve if the component fails?

1.4 Contribution

Water distribution systems in large cities, in particular, need ongoing follow-up and monitoring to ensure that water reaches all users fairly and to confirm the system safety through periodic maintenance and repair of any failure of a component as soon as possible (ASAP). Therefore, in this research, we propose a fair distribution relied on advanced intelligent spatial decision support systems. We'll use Nablus as a case study.

We assume in this research several hypotheses:

1 Using data mining algorithms to create a prediction system capable of identifying components at risk of failure and requiring maintenance based on historical maintenance data.

2 We can use spatial data mining analyses to identify and quantify the impact of various parameters on analysis results which determine if the component needs to maintenance or not according to the maintenance historical data.

3 The network can be analyzed using the isolation algorithms and the theory graph algorithms to determine the controlling valve for each component of the network that have to be closed if the component breaks down, which helps save time and effort and contributes to the conservation of water from leakage.

Our solution is unique as it proposes new mechanisms to wisely manage the water engineering network while resolving distribution and maintenance concerns using geographical and historical data. The proposed study technique is to use an intelligent resolution spatial support system that creates a specific engineering network structure for connecting the components of the system taking into account component features, water direction and engineering and physical principles. GIS is used to model the network, considering all elements of the network such as locations, water sizes, physical components, etc. Conflicting interests and priorities will be considered based on service provider type, etc.

The data of each feature of the data is stand for in the water distribution system will be gathered and worked on. All relevant data will be recorded and stored in a geodatabase, with each type of data preserved as a layer in the form of a file format in a basic GIS format. The data consist of geometric forms, with each geometric feature attached to its own record in the database via the attribute table. Pipes, junctions, valves and meters are the layers used to represent water distribution systems. Monthly database of meters (all of this will be connected to the geographic location).

GIS approaches are provided for the preservation, digitization, analysis, management and visualization of the spatial data layers addressed. 3D models will be created for the study case area, and the proposed model will detect problems and propose ideal solution with minimal expense and work. The optimal solution will discover the cause of the problem and the smallest area to be affected, this solution will be presented efficiently and intelligently to provide a clearer perspective on the system of equitable water distribution and the surrounding spatial impact.

Water Distribution Network Management will apply all analysis tools to get useful answers to problems for example if we have a pipe failure a number of users will automatically be deprived from water. The system will anticipate the failure process as accurately as possible to try to handle it in advance, and in case of unexpected damage, the repair will be done as quickly as possible and alert affected users.

The network will be analyzed as a tree structure using C# programming language in order to calculate, compress and regulate water trends. To find a formula for fair and prudent distribution and ensure the integrity of the network and its components, research will focus on water distribution and complex calculations. Furthermore, historical data on challenges and maintenance issues will be examined using geographical data exploration techniques to model, predict and advise on prudent future plans. To make a decision, the decision support system will use machine learning courses. It can be used to monitor, analyze, reorganize and rebuild the network in a way that promotes intelligent planning and effective water distribution. It also helps reduce technical and financial loss as well as maintenance work.

The Nablus Water Service's engineering network integrates all aspects in a rational manner. Feature categories form the engineering network. Feature categories are a set of geographical features with the same geometry and species. All features in the same feature category have the same attributes, so in ArcGIS we can inquire and choose a particular field based on a particular property or the result of a particular hydraulic model. We modify current formats and build topological links between water distribution network feature categories relying on the test of basic aspects of current data. The hydraulic model used to simulate the function of the actual water distribution network of various components in the intermittent water delivery system.

Spatial analysis tools will be relied upon by ArcMap to analyze data and establish a prediction system to contribute to the solution of maintenance problems. ArcMap program contains a number of spatial statistical analyses, including what depends on time such as space time pattern mining tools, and what derives from modeling like spatial static tools, measurement and many other tools and analyses that facilitate the organization of control, modeling and analysis of the components of the network and the rules of association of its components.

As for the C# programming language, it will be relied upon to analyze the network, conduct data mining for one of the mining algorithms, utilize it to create an expectation system, show results on the network, and be used to create desktop applications, user interfaces, and others will be mentioned later.

Using spatial data characterized by the distinctive structure of the study area (industrial area in Nablus City) after its processing and the establishment of an integrated geometric network, consideration is taken of spatial attributes such as location, component size, size of elements compared to each other, height from ground level, distance between them and road, etc.

This proper and coherent network is an integrated system so that each element is aware of its function and qualities and the functioning of the system is the integration of the functions of all components. Topological rules are applied to ensure the proper integration of the system so that we get a network documented spatially with precise coordinates of each component.

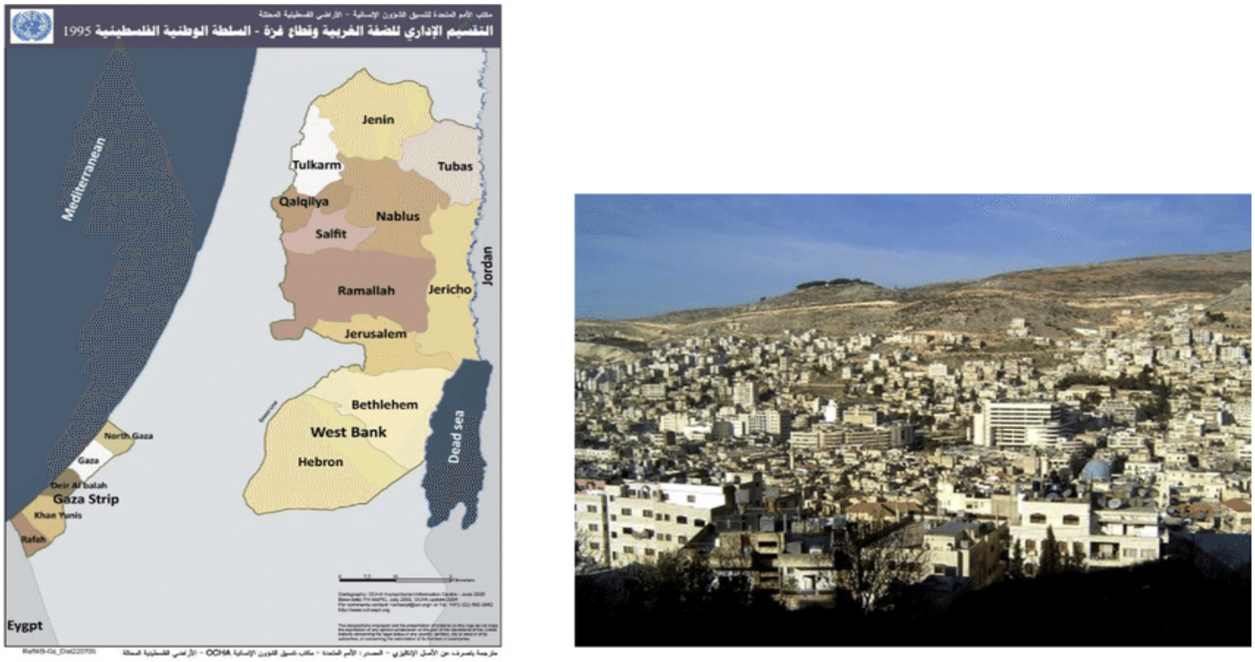
We built a tree structure of data within this network by identifying the main root and sons of each father and by doing so; we can identify the controller of each component and determine the direction of the water flow in the network so that the process of spatial data analysis becomes easier, more accurate, and more effective.

The analysis and decision-making process was based on a number of AI algorithms with the aim of harnessing computer science in general and AI in particular to analyze spatial data and a decision-making industry related to the network system for genuine integration between GIS science and computer science. A number of network services like viewing the map, handle the components, include a prediction system for components that may need maintenance using data mining algorithms and a network analysis system using visualization algorithms to construct the tree structure of the valves and identify the controlling valve for each component to be proposed by the intelligent system to be closed if the element breaks down, and so we have proven the hypotheses valid and achieved the objective of the study.

1.5 Study Area

Nablus is located between Gerizim Mount 881 masl (meters above sea level) in the south and Ibal Mount (940 masl) in the north It is also referred to as the Queen of Palestine and Little Damascus. Geographically speaking, Nablus is the most industrial and commercial city in Palestine and one of the most populated as well. It has a population of about 384,900. Nablus is situated in the West Bank's northern region. Its limits are shown in the figure 2 below. The average temperature is between 6.2 and 29.4 degrees Celsius. Nablus has a hot dry summer and mild wet winter. According to the Nablus meteorological station, the annual average rainfall in the city of Nablus ranged from 1975 to 2023 between 350 and 1200 mm. Well Ras al-Ayn, Ayn Dafna, Al-Quriyat, Dir Sharf, Ayn Beit 'a, Ayn al-Asal, Beir al-Ba'dan, Beir Sebastian, Beir al-Fara 'a, and Beir Udala are the city's primary water sources. I did my research in a certain zone. The zone is a part of the city of Nablus which is known as the industrial zone.

The city suffers from scarcity of water and resources. The basic idea behind the water problem in the city of Nablus is that as the population grows, demand increases, yet new resources are not generated to meet the city's demands. Nablus is the economic center of Palestine and one of the Palestinian cities built around 3600 BC. It has an area of 85 km2.



a b

Figure 2: : (b) Nablus City and (a) its boundaries [108]

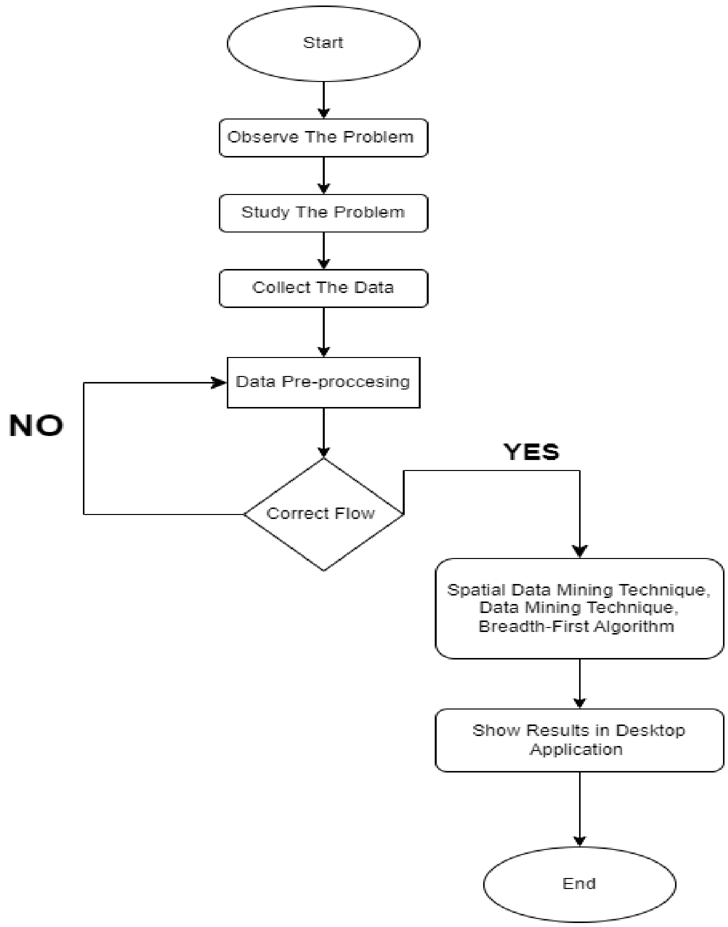
1.6 Usage software

I used a computer with an Intel® Core (TM) i7-6500U CPU, 8GB of installed RAM, and a 64-bit operating system. ArcGIS 10.7 was used for data handling and analysis and the establishment of the engineering network that logically connects the components of the water distribution network. ArcGIS 10.7 via the ArcObject SDK (software development kit) to build and customized GIS applications ,I used ArcScene as hydraulic modeling software, ArcMap used to visualize data and build the network properly and control it and perform statistical analysis .

The desktop application was constructed in its entirety using the C# programming language via Visual Studio 2022, and data mining using the linear regression algorithm was performed.

Excel has been used to save process and clean data. Excel 2010 has been used from Microsoft 2010 package.

1.7 Methodology



Personal and informal interviews were used to identify diverse needs and perspectives on the subject of my study in order to achieve research objectives. I identified my study problem using the results of these interviews, as well as reviewing literature in an objective area. The information obtained from these interviews was examined to help me achieve my goal. The main benefit of data analysis is that it helps to organize, understand and simplify the data obtained. On the other hand, I got results proving some data inaccurate and some important data missing. Data analysis helps me structure the acquired data in a way that meets the requirements of my thesis. Brainstorming, current scientific knowledge (methodological literature, past work and related function), and the industrial environment are all part of the first step of my research. It gained new information, identified a topic, research strategy, and a goal to study based on preliminary data collected from books, the Internet and related materials, as well as other data sources. The second component included topic analysis, learning the GIS concepts we used to achieve my study's objective, and then choosing Nablus water management as a case study. Preliminary statistics of Nablus water management have been collected. All data obtained contain many problems to be solved, updated and controlled. It was difficult to determine the study area and its components associated with the water system, as well as to correct current data inaccuracies.

We need to produce a large number of shapefiles, including pipes, meters, water source and valves. The water source, valve, water meters and all the point features are under the point feature data type category. Then, with all shapes, which include main pipes, house connections, valves, sources and customer meters, we must build a relationship between them, such as topology and connection. Therefore, we use the engineering network, but it has a number of problems, such as incorrect water distribution, water stopping in some places for long periods of time, and breakdown caused by damage to a physical component of the network. Therefore, using GIS and Excel software, we must prepare each customer's spatial data and track source-to-consumer water flow to ensure access and safety of pipelines and valves.

Network topology is the physical form of the network. The method of passing information between nodes is known as logical topology. Topology in GIS describes how categories such as point, line and ribbed engage in engineering. For example, it takes into account the connection between the water pipe and its associated control valve. The industrial zone was selected as a residential area based on the credibility of its statistics. As a result, the focus of the study will be on it. The third step involves the use of GIS capabilities to manage, modify and generate additional required formats and the engineering network. The fourth and last step is the creation of an intelligent system based on spatial data After it is represented and ensured that it is properly connected, the data is processed and joined by a spatial join tools and cleaned so that it is ready for analysis, The system becomes capable of analyzing data based on data mining algorithms and techniques and spatial data mining and proposing appropriate solutions taking into account all spatial data. And then create the desktop app that shows the network and the results of the analyses.

Chapter Two

State of Art

2.1 Water supply system

Water is essential for life and industry, including agriculture, industry and local consumption. Aquatic knowledge is necessary to understand how water and humans interact. The water supply system consists of hydrological components designed to provide water. The percentage of freshwater is 3%, the rest is saltwater. Many issues affect the supply of this resource in Palestine, starting with loss, pollution, infrastructure and Israel's occupation of current water supplies.

Water problems affect each country's environmental and economic situation. In Palestine, we suffer from poor water distribution; in 2021, 14.4 mm3 is Nablus' total supply resources. Aquatic knowledge is necessary to understand how water and humans interact. Many reasons, like population expansion and global climate change, contribute to the global water issue, requiring prudent management of water resources. Water resources engineering deals with the design of systems to manage water distribution systems regarding planning, time, quality and various water volumes that must be provided to different regions in order to give citizens the sufficient amount of water with the least amount of waste. All of these items should be checked in all areas where we may have inadequate or excessive water, and may be placed in the wrong place. Water resources may be dirty or expensive; Information about water resources must therefore be collected and stored for use in monitoring and managing water systems.

Several studies have been conducted to manage, monitor, and control all elements affecting water supply, treatment, and distribution. Limos and others use a life-cycle assessment (LCA) approach to examine urban water systems. Takes into account all stages of treatment, such as the abstraction, treatment, distribution, and management of water. The quality of the data used in this investigation is classified into three levels, Data of the highest caliber: electricity, water and wastewater flows consumption through sewage collection, water distribution, and during the stages of administration, fuel consumption in the stage of administration, and pipe consumption. The following data have a medium quality level: outputs of N and P in the wastewater that has been treated; outputs of sludge (overall weight of wet without composition); outputs to the landfill of inert waste and fats; and electricity utilized for treating and abstracting water, and the stages of the WWTP; inputs of lime, CO2, chlorine, and sodium hypochlorite. Poor-quality data: inputs of polymers and ferric chloride, N and P contents in the sludge, and emissions of CH4 and N2O.They investigate the effects of each level of water treatment. [5]

GIS is a spatial database containing all locations and distance measurements, as well as a lot of other information for each feature. It is used to create maps, which is an effective means of communication. Geographic Information System (GIS) provides a database of features used for reporting and detection of relationships between features. It is used to make judgments and choose the best options. GIS has different uses in water resources engineering, starting with watershed identification and presentation of location and geographical data, as well as watershed characteristics such as percentage rainfall and land-use change. GIS stores a range of geographically referenced map layers. Each set of maps is presented in a dataset and devoted to common projections. Class is the name given to each map. Class is a chip depicting a particular component in a given area, such as a political boundary layer or a road layer. [6] Water sources such as tanks, main distribution devices, fittings, faucets, systems valves, pump tanks, treatment plants, and side devices such as home and service delivery, are all part of the water system. All running pipes are fed by main distribution. They are used to deliver water to end consumers.

Fittings are used to join different pipe sections of the water system. In the water system, valves are used to manage and control water flow. As a result, each type of data in the water system is represented as a layer in GIS software.

2.2 Related Work

A person often faces many difficulties, concerns and challenges that require support and aid in addressing and responding to them. Decisions must be made. For example, the urban planner may be asked to choose the ideal location for a new urban assemble. The farmer must decide on the most important features. Soil in private agricultural sites, traders faces the dilemma to open a new store in order to maximize profits [2] GIS improves spatial thinking skills by providing tools for the integration, presentation and evaluation of data [3].

The advantages of using dual digital technology in the water distribution system were studied in [4]. This study looks at the digital transformation that will be implemented in water distribution networks (WDNs). The qualities of smart water management were evaluated through the use of a case study that was realistic, i.e., the results of the use of Digital Twins (DT. A district-level measured area in Lisbon's water consumption system was selected as a case study. Due to a combination of SCADA, GIS, and smart sensors technologies, real-time monitoring of digital twins leads to rapid loss or recognition and repair, allowing for significant water savings. This has huge benefits, including reducing water loss and operational costs, as well as fewer social and environmental implications. The DT application was simulated by running an optimization technology to find the best valve regulations to reduce water loss within the examined network. When leakage is found, the need for improvement develops, and it is necessary to have a way of restricting the amount of leaking water. Once the leak is detected, it can take several hours to search for the best settings section, along with intervention. According to these results, the total leaked volume declines from 1498 cubic meters every day to 1272 cubic meters every day, resulting in 15% savings. According to the results of this study, the potential benefits of digital twins in the field of water for digital transformation and enhancing system effectiveness are enormous, benefiting the water economy immediately, increasing customer experience, providing improved maintenance and environmental safeguards. DT technology also allows to predicting demand during peak demand, improving pressure and flow control, and greatly enhancing the flexibility of operation and system reliability. Furthermore, it should be noted that the substance DT has made a permanent home in the WDS area. Many water companies seek to build and maintain the water distribution system as a secondary target, but it is not a simple process. However, the data and technologies available today make this possible. This paper describes Global Omnium, GoAigua, and UPV's efforts to achieve this goal- being one of the primary facilities to be linked to DT to be Key data sources for companies. The many advantages ensure that they become common in some of the most advanced urban areas. GO2HydNet software may generate a complete hydraulic model of data saved on the company's extensive data platform, which collects data from GIS, AMR (Automated Meter Reading), CMMS (Computerized Maintenance Management System), and field sensors. Algorithms have been developed with the aim of maximizing efficiency. As a result, a comprehensive model can be GO2HydNet with 325,000 knots in less than a minute. Hydraulic simulation loading techniques are designed and implemented from data collected in the field, as well as successfully prepared and filtered. DT enables you to replicate any previous scenario as well as simulate immediate conditions and receive expected future progress. However, there is still domestic work to be done. At present, the comprehensive model and the strategically intensive model are not connected. More work is needed to develop an automatic mechanism to link both models.

Furthermore, water hygiene model has not been updated, or even it hasn't been created yet. As WUs (Water Utilities) are increasingly concerned about managing and promoting water-related issues, it is important to take advantage of the potential benefits of the DT quality model. [13]

In this study they examined the hydraulic efficiency of the Yika Abadu water supply area in Addis Ababa. Analysis of the system's capacity in 2021 and 2044 for a continuous supply scenario shows that the system met capacity requirements in 2021 but will be limited in meeting future need in the study area as demand for flow increases significantly between 2021 and 2044. According to the results of the typical study, the flow requirements in the network increase significantly from 285 to 1024 liters per second. Negative results for 2044 indicate that demand will exceed the system's capacity. The sheer volume of flow going through each pipe is the root of this issue. Since the flow is directly related to the loss of the pipe head, the increased flow will raise the loss of the pipe head. Extending the system may be one of the solutions to the problem. the device would lessen each tube's individual flow, reduce head loss, and stabilize pressure [7].

They investigated the usefulness of the family profile and individual comments to improve decision-making on water use in this study report [8]. Their research recommended the establishment of an individual recommendation system relied on data collected by multiple sensors planted in different features of homes, with a particular focus on water outlets such as the basin (cold water), sink (hot water) and toilet. The deployment of an (artificial intelligent) AI-driven recommendation system employing two strategies: collaborative filtering and with rule-based proposals has proved very useful in providing feedback relied on consumption patterns seen in similar households. One important outcome of their study was the identification of families using water more than similar families. This determination allows the rules-based system to make personal recommendations, like replacing leaky taps in basins in bathrooms, to reduce excessive water consumption. On the other hand, External incentives can be used to encourage low-use families to continue their water-saving practices, which will form the focus of the future study. In addition, they are aware of the possibility of expanding the system to assess the effectiveness of the advice provided. As a result, the dataset may be extended to train different machine learning techniques and offer suggestions to additional families. Continuous analysis of data, improved rules based on experts' factual knowledge, and interactive user links are needed to improve accuracy of outcomes and respond to changing household consumption trends. [8]

This research report introduces the Android app and a LoRa-based (Long Range) platform. When installed in fields at different agricultural sites, the prototype displayed is adept at monitoring crop factors. For an effective intelligent farming method designed to monitor and control environmental ingredients in agriculture, we used the LoRa module link. The low-cost system that is constantly proposed tracks accurate statistics and sends them to farms via the Android app. The results that show the proposed structure also support it, as rapid use and supervision of all farm sites are made possible by the availability of real-time information. The proposed system also provides for an automatic adjustment of irrigation and environmental parameters by introducing wireless data transfer capabilities into the intended application. [9]

The implementation of an intelligent global system in agriculture may help farmers achieve sustainable development goals by improving their financial position and increasing agricultural yields, which will result from the widespread implementation of the proposed system. Future developments may allow the expected automated system to be integrated with Sharma's strategy and others with the help of drones [10]. Moreover, a different thermal imaging camera can be used to monitor the farm, as can drone technology. With the help of a drone a liquid spraying tank can be used to spray insecticides in the affected area of the crop farm if the pest feeds on crop leaves.

This study examined [11] sources written by magazines and indexed in Scopus databases from 2000 and 2021 that included relevant publications on RFID (Radio Frequency Identification) technology and decision support systems in the context of Industry 4.0. More research should focus on how to improve other associated technologies, like (internet of things) IoT and artificial intelligence, for supply chain performance. The review covered a variety of topics, such as organizational strategies for the adoption of new technologies, such as RFID, equitable growth, contemporary methods, the IT approach for the design and management of the supply chain, and other advantages of RFID and Industry 4.0. The results reveal potential savings of up to 84% of total transit time and late time, as well as a nearly 60% decrease in the time required to supply materials, which generally improves the efficiency.by of about 80% of the supply chain. RFID has proven itself as a global technology that provides industrial benefits and has seen widespread use across industries, especially health care supply chains. This study made recommendations for improving the efficiency of supply chain operation while reducing expenses associated with the deployment of RFID technology. The paper's main contribution is its analysis and assessment of the various RFID implementation techniques in supply chains in order to save time and cost efficiency. The focus of this study is limited to investigating the impact of RFID on supply chains. Future research should look at the interaction between RFID-IoT and the physical internet, as well as whether real-time data processing may continue to improve supply chain decision-making.

Coagulation and flow are among the most well-known actions in water surface treatment to eliminate colloidal debris and turbidity. The clotted dose should be adjusted based on logical criteria by operators. Manufacturers are interested in improving chemical components and energy consumption to remove high-efficiency turbidity. As a result, they need reliable and sophisticated patterns and techniques to examine glamour and clotting behaviors. A correlative analysis of both the inputs and the results of the water treatment process were conducted in this paper. The amount of iron chloride introduced into the water was one of the inputs. The results included the outlet's turbulence and the amount of energy spent on the procedure. To systematize this practice, the Integrated Financial Information System (IFIS) model was used. The large determination factor (over 80 %) indicates a coherent relationship between inputs and outputs. In the last phase of this project, Petri Net modeling is used to apply (decision support system) DSS at wastewater treatment facility. According to the technology given, each smart order is associated with turbidity changes via Petri Net modeling. Based on the detected DSS, the additional coagulation component (iron salt) should be set at, 60–85, and 40–60 kg/day as higher as and lower than inlet ruffle values. In general, the main benefits of their ANFIS (Adaptive Neuro-fuzzy Inference System) model are its excellent accuracy and strength, while the main downside is its long calculation time. It goes beyond indicating that there are multiple proposals to improve the study's impact on future studies. First, more elements related to water treatment, such as employment opportunities and social justice for workers, can be examined in our model. Another good idea is to apply the concepts of improvement and uncertainty to our water systems. Finally, in order to enhance the effectiveness of their ANFIS model, alternative programming approaches such as genetic algorithms and adaptable research methods can be recommended [12].

This research [14] showed the usefulness of smart agriculture in boosting agricultural productivity and increasing it to help close the food need gap. IoT is the core of smart technology for agriculture because it connects every aspect of smart systems, not just those in agriculture. However, there are additional requests. With regard to the IoT in agriculture, it can be used in a variety of disciplines, like farm monitoring, irrigation, pest management, harvesting, etc. IoT connects different sensors to processing units, analyses data, and makes appropriate decisions in real time. This paper examined the integration of IoT with drones and robot systems operated using AI approaches, as well as the limitations of their implementation in underdeveloped areas. In recent years, achieving SF (smart farming) performance has been linked to fast data transmission. As a result of its very high speed compared to first-generation networks, 5G enabled smart farming and provided flexible and effective solutions. The use of smart agricultural equipment benefits developing countries, including some Egyptian roads that reflect the beginning of this proliferation. Technology can help those countries expand their agricultural sectors and achieve agricultural sustainability. Lastly, governments in third-world countries must promote smart technology at the small farm level while striving to boost production while promoting effective of land and water resources management.

The need to provide expensive resources and prevent the effects of spills has led to an increase in the field and industry of study devoted to locating pipeline leaks [15]. Immediate identification of leaks can prevent large gas leaks, water leaks to the ground under highways and sewers, and infrastructure damage. Avoid harming the environment around you, as well as saving money. The ILP system (Identification, Localization, and Pinpoint) divides leak detection into three steps, but this method is thought to be unclear, so a new approach, the ILP system, is proposed. The ILP method defines the boundaries between research areas. The first step is identification, which determines whether leakage occurs or not and distinguishes leakage data collected from additional data sources. They can also locate and localize the place. But dynamic leak detection systems are used to confirm leaks by deploying devices and workers to suspected locations. This method gives a quick clarification of the leak and accuracy determination, but allows for increased leaks if the reaction is delayed. The best practice right now is the combination of the two systems, where the fixed system provides initial detection and the dynamic system, on the other hand, provides localization and accuracy. Fixed leak detection techniques will be able to do the entire ILP within seconds, according to studies in the fixed detection sector. With regard to technological applications, the topic of identifying leaks in pipelines is expanding as new technologies are always developed to explore more rapid and effective aspects. The profession is likely to expand further as demand grows in the water and gas systems industries.

In the management of underground facilities, the storage of data in two-dimensional form can be difficult, there is no single database, and data records are inconsistent. The majority of facility cases are unknown to managers. As a result, the decision on maintenance and inspection of facilities is ineffective. All obstacles can be traced back to the lack of a unified platform for storing, analyzing, managing, and updating information. Based on the combination of the building information model and GIS, this study [16] established a common structure to facilitate facility-level facility management and spatial network components. The integrated BIM-GIS (Building Information Modeling) framework is implemented by mapping IFC (Industry Foundation chapters) and CityGML (City Geography Coding Language). As the current IFC and CityGML lack entities, the standard utility data model has been created to serve as a reference for the expansion of the IFC and CityGML Utility Network ADE (Application Domain Extensions). Data model entities are used to show data from five major categories: First network data, Second component information, Third engineering data, Fourth status data, and Finally miscellaneous data. The facilities' data model is being developed. Relied on the data involved and needed at various stages of the facilities project, including the semantics and engineering data required for the design of facilities and the inspection data and status required for the maintenance of facilities,. The applicable utility model also includes references for the creation of the important Database Management System (DBMS), which stores and updates all relevant information. By integrating BIM and CAD (Computer-Aided Design), rich and useful information can be distributed. GIS, the integrated platform, may assist in the management of facilities in several ways, including assessing the situation, assisting in decision-making, and influencing the perception of facilities. A case study has been conducted to show how the suggested framework can be applied. The functionality of the existing framework was verified by introducing the system to possible users and gathering feedback from them through scanning and in-depth conversation, the results of which supported the completeness and usefulness of their proposed framework. The utility degradation model was not fully built and tested because of the lack of long-term historical utility data. More will be added in the future. The system designed to collect historical inspection data will be used. The basis for the development of the underground facility degradation model is that the potential for collapse can be assessed in order to achieve the maintenance of predictive facilities.

As people become aware of the necessity of water and the need to maintain its near-depletion sources, a number of solutions and systems have emerged, such as [17]. It proposes a framework for the design of a multi-standard system to support spatial decision-making aimed at establishing a link between the subject area and the evaluation area. Its methodology is based on creating a network model and selecting the highest-priority characteristics for building links.

Monitoring and decision support systems were most needed in particular, such as monitoring systems and the scheduling of irrigation [18]. This idea emerged in the 1970s to make the irrigation process more effective. Irrigation water management is required on farms due to the use of a computer to link water quality conditions, soil, and crops. It can be used to analyze and determine the amount of water needed and the time to be given thereafter.

Understanding the concept of water loss is the first and most important step in its management and control journey [19], which is why it provides strategies that support good understanding.

To reduce water leakage and loss, this research [20] provided a methodology for analyzing water distribution and simulating its screening process using the genetics algorithm and MATLAB software. New tools and methods were used in the form of advanced hydraulic simulation models and improved, resulting in a reduction of 73% of actual leakage calculations.

The water distribution area was not the only one designed using GIS. It was stated in [21] that the use of GIS infrastructure and applications in electricity distribution systems. They developed an application on how to perform automatic reporting in the GIS environment with county integration and used the C# programming language.

In [22], the authors review the models and decision support systems that are currently in place at the basin level for the distribution of public water supplies. They also offer real-world examples and conduct a thorough literature review on the ideas that arise from the definition of integrated water resources management. They have considered two environmental concepts in this study: sustainability and managing water quality. The use of the decision support system has expanded to include energy distribution. One of the main challenges in energy informatics is that electricity supply must be balanced with demand at all times. This, in addition to balancing manufacturing between demand and purchase, has become a powerful challenge between decision support systems and human decision-making capabilities.

In [23], the authors propose a new model-based decision system to strengthen operators' manufacturing or procurement decisions under different levels of data quality in terms of availability, detail, and timeliness. The research indicates a saving of up to 40% in cost adjustment to support the ambitious municipal development efforts we study in our realistic assessment.

In agriculture, there is fierce competition in the agricultural digital data market, and although farms are the key element in data collection, their presence remains weak. In [24], they conducted the study on a data platform that can be developed, deleted, and added to, even for small landowners. They have selected the fruit and vegetable industry in Spain as their case study, and their goal is to develop a multifaceted, multi-layered platform under the guidance of the theory of superior governance. In order to obtain efficiency, transparency, and good results, decision support systems and tools are used to help farmers make better decisions.

A qualitative shift is occurring in the management of water resources. This shift requires sufficient effective techniques to ensure good results. Therefore, studies are needed not only for intelligence department but also for assessing the available tools and integrating them with existing management models. This thesis [25] provides a decision support system for supply networks that helps managers make decisions on the basis of objective information rather than intuition or experience. In general, open source software and hardware are used, and the development of this thesis should be emphasized. They are permitted to use the approaches that water companies have suggested thanks to this particular feature. Regardless of size or financial resources, the entire system is only part of it will be adapted for the company's operation.

In [26], authors proposed a system of spatial decision support for geographical disaster management. To take advantage of the benefits of geospatial techniques in disaster management, despite the significant challenge that remains in aligning sources such as data algorithms due to gaps such as knowledge gaps and network gaps, this approach aimed at supporting disaster management through the integration of geospatial resources and mission chains. This approach has been implemented as a geospatial platform in China that provides effective and timely responses to natural disasters.

Monte Carlo's approach to this work [27] has been used to address the problem of improvement. It was implemented in the UKE laboratory network (University Medical Center Hamburg-Eppendorf.). In addition, three mathematical models (EPANET, the AZRED model based on Romero-Gomez and Choi formulations, and EPANET-DD model) were considered to see how they could influence the best position of water quality sensors in terms of detection probability, detection time, and redundancy. The following examples highlight the research paper's accomplishments:

1- Sensors were placed in places with high Reynolds numbers, where flow systems are usually instable and transitional, whenever adjective technology is used to address the improvement problem.

2- Using distracted AZRED technology; sensors were placed in a linear manner and covered the majority of the network.

3- The EPANET-DD model performed a 95% probability of detecting contaminated events, a 70% frequency, and an average detection time of about 9 minutes.

4- Whatever model is used to solve the improvement problem, different configurations of the sensor location are achieved, as a result of the different detection efficiency of the target functions. The parameters determined by the adjective model, for example, are much lower than those obtained through the dynamic dispersion model (EPANET-DD).

The purpose of this study [28] was to consider the spatial distribution and gates of nitrate and fluoride quantities in Behbehan's drinking water distribution network, as well as associated health risks. From 2018 to 2019, their focus was calculated through sensitivity analysis and Monte Carlo simulation. The current investigation determined that the maximum and minimum levels of nitrate and fluoride in the study area were 26.5 and 1.39 mg/L, 7.2 and 0.36 mg/L, respectively. Furthermore, the average concentration of nitrates was higher during the rainy season, than during the extraordinary winter. Nitrate concentrations across the North, South, and Southwest decreased, according to the zoning map. In Behbehan, the largest and lowest volumes of fluoride in water-based beverage sources were 1.39 and 0.36 mg/L, respectively. During the low rainy season, the average fluoride content in the study area was higher more than during the rainy season. Furthermore, three samples (6.7%) and two samples (4.4%) showed fluoride contents below the minimum during the low rainy season. Fluoride concentrations decreased from north to south during the heavy rainy season. The headquarters produced by ion fluoride were larger in children than in others. During the rainy season, nitrate headquarters levels in the groups investigated were 0.5, 0.4, 0.34, and 0.07 for children, adolescents, adults, and infants, respectively. Nitrate ions 95 P and nitrate-induced fluoride remained greater than one in all age groups, showing that nitrates and fluoride do not pose any cancer risk to 95% of the population at the research site.

Evolutionary algorithms for improvement seem to give superior results, albeit at a much higher computing cost. However, determining the amount of time required to reach an optimal solution is difficult. The dual characteristics of 100 ideal WDNs from Pareto, especially the path of improvement from random solutions to optimal Pareto solutions, were explored in this paper [29] for three separate real-case studies (a total of 14,300 design solutions). The improvement was implemented using a multi-purpose design technique, with each case study reaching 500,000 generations and an average community size of 100 individuals. Two topics are being studied: (1) what constitutes the characteristics of the global development networks' optimal dual chart, and (2) when and how the important phase of improvement is achieved and in what way this can be assessed. A dual graph technique has been created that uses diameter as a generalization model to distinguish between the optimal and non-optimal approaches to design global defense networks. Dual features are thoroughly explored using a 14,300-point dual graph to find aspects of ideal WDNs and assess whether dual properties can be used as a measure of how close existing optimal solutions are. The total number of double nodes in double charts decreased as one approached ideal solutions. The distribution of global development network requirements identifies the smallest possible number of twin contracts. Flow categories resulting from the demand edge between central (dEBC) have been found to be a good indicator of the lowest possible double node score in the improvement process. As a result, the dual version of the primitive chart can be used to determine whether an evolutionary improvement method can produce superior results. As a result, the number of double nodes with a lower value attainable based on demand distribution characteristics (dEBC) can be seen as a measure of the progress of the improvement approach and can be applied to the evolutionary improvement of global development networks to improve this technology. The proposed method does not add any significant computing expenses and can be applied to any assembly engine used to plan the water distribution network. However, when other goals are added to the less network structure-based improvement strategy and order distribution, the (dEBC) scale may become less effective, requiring additional research. The dual mapping method can also be useful in determining the effectiveness of search engines during the improvement process. For some generations in this study, the range of possibilities was somewhat limited at the beginning of the improvement procedure. A future study in this area could focus on a systematic analysis of the effectiveness of research operators using the proposed dual method.

2.3 Geographic Information system

2.3.1 Definition of Geographic Information system

Freshwater is the second most vital resource in human life after food, and access to drinking water is a human right. The difficulties of accessing drinking water affect people's standard of living throughout the world, and in Palestine, we suffer from water loss. Statistics from 2021 reveal that 5.4 mm3 of the total supply resources were wasted in Nablus. GIS is a type of information system that collects, stores, verifies, manipulates, analyzes, and displays data that specifically refers to Earth. [30] [31] Dr. Tomllinson published the first recorded use of the word GIS in 1968. [32] Tomllinson is often seen as a "GIS parent." In the early 1960s, he established the first computer-based geographic information system for use by Canada's land inventory. His efforts in the GIS sector earned him the Canadian government's highest civilian award. "He pioneered its use around the world to collect, manage, and manipulate geographical data and change the face of geography as a system." [33] GIS facilitates spatial analysis by conceptualizing sites and items from different fields of study, such as modeling the complete cycle of the water supply network from its originator to its dwelling. GIS provides an interface for spatial data analysis and calculations, as well as users' ability to ask questions and receive responses. [34] Because of the diversity of uses and discrepancies in defining the system's objectives, there is no single definition of GIS. It is a computerized map that is associated with databases that enables us to add, delete, and change data, as well as carry out search operations and analysis to make the best judgments.

GIS is a robust web and information interconnection system. It is a powerful data visualization tool. In [35] the data is classified into two categories: metadata and spatial data. Metadata provides descriptions of features related to feature category characteristics, it's separated into tables. Spatial data is a description of the location, shape, and coordinates of geographical features. It connects relationships with each other. Spatial data can be represented in GIS in two ways: vector and raster. I'll select raster data and vectors using GIS. The vector data model uses points, lines, and polygons to stand for the world. It is used to retain data with distinct limits, such as power lines, water pipes, or streets. Deng et al. A method is used to integrate vector data with landscape visualization, where vector data is used in many applications to analyze and manage nature. [36] The world is represented using the Raster data model as a layer divided into pixels (or cells) organized into rows and columns (or networks). Each pixel is connected to a value, such as temperature. Raster images come from satellites, digital images, or scanned maps. It is used to store data about these changes over time, such as soil type, land use, chemical concentration surface, precipitation, or elevation surface. On the altitude surface, for example, pixel values may reflect elevation above sea level, which appears as a network of cells (often square). During the first step of my research, I gathered information from previous scientific studies, such as methodological literature, previous works, and industrial attitudes.

Identifying the problem and mastering GIS principles is part of the second phase. Water difficulties vary depending on the research area's infrastructure, access to water in the study area, the method used to transport water in that area, and the components and characteristics of the water system. GIS helps describe all elements of the water network and create relationships between all components in order to identify problems, try multiple solutions, and choose the best one of all options.

2.3.2Geographic Information System as Decision Support System

GIS can be used as a spatial decision support system (SDSS) tool. SDSS is complex because it includes a number of goals and limits. It's a way to choose the best option from a range of substitutes as well as an intelligent and integrated tool to organize information and knowledge in the desired field. SDSS requires spatial data, organized or unregulated knowledge, and human governance. To determine the solution, the decision-making process can use traditional (organizational) solution processes, or there is no specific (unregulated) action, or a combination of both (standard procedures and human provisions). Information from SDSS is derived from a variety of sources, including maps and remote sensing. This information is often incomplete or complete and changes over time and space. As a result, we must use human understanding to manipulate it in a successful and effective manner. Like if then rule- which is a rule-based system that can be used to represent human knowledge. We must prevent rings in the set of rules and separation in proportions when designing this rules-based system.

Spatial decision-making included information and knowledge aspects [37] as shown in the figure below. SDSS requires the integration of geographical data with algorithms, technologies or other tools in order to optimize the choice between different alternatives. One of these tools is (GIS). [38] GIS is a science which connects digital maps to databases. Uslu a et al. Processing, assessment and analysis of data from the Ondkuz Mays water distribution system, an area of Samson City, Turkey, using the Urban Information System as well as geographic information system capabilities. They used GIS technology to visualize the water distribution system by introducing pipe data, but they did not create a network or provide topology between layers in order to make the system smarter. They use GIS technology to automatically calculate node heights from DEMs (Digital Elevation Models). [39]

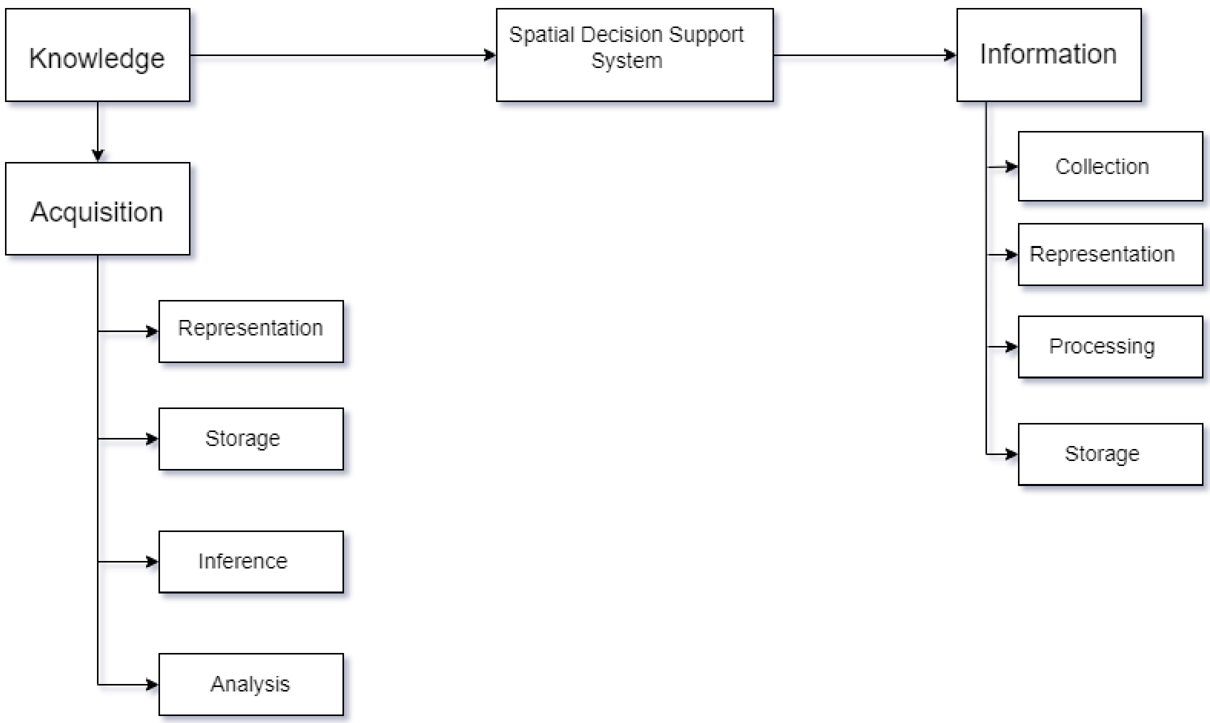


Figure 3: Spatial Decision Support System.

Any software may be open-source or owned. There are many examples of open-source GIS software, including SAGA GIS, gvSIG, and QGIS. [40] ArcGIS, chosen for this thesis, is an example of closed software. ESRI's ArcMap 10.7.1 GIS software was used. ArcMap 10.7.1 is GIS software that allows users to question raster-based graphical and non-graphical data. ArcMap offers high-quality mapping presentations to users through its functions of presenting, updating, inquiring, analyzing, and displaying current graphic and oral data.

2.4 water network as a distributed system and complex adaptive system (CAS)

The water system is a network of distributed components, each with a specialized duty. Water systems consist of many nodes, including customer meters, bulk meters, pumps, pipelines, and valves, which are often placed in huge geographical areas for remote operations. In my model I have collected data for the industrial area in Nablus city and build a geometric network to do optimizations in order to meet the problems faced by the employees of the institutions to make the interest easier, which means providing better services more fairly like maintenance problems, The water system is a flexible system because it can accommodate a large number of contracts (customers) without affecting the overall system, its performance, or its effectiveness.

Transparency in water systems is the main aim of the distributed system is to hide the fact that the system's components and devices are widely spread in distinct logical areas and make them look to consumers as one coherent system. In general, the idea of transparency can be applied to several features of a distributed system. As an instance of this in water distribution system, I will address the most important systems and what can be achieved in a built-in distributed system. The term "node localization" refers to the process of establishing location awareness across every water system node in deployment.

One of the most critical difficulties in the water system is location information, which allows the system to recognize events and route packets between pipes, valves, and sink nodes in order to share information and readily discover faults. Location transparency in the water system, which is a distributed and complex system, for example, across a big building or a large area, one of the major goals of this network is to provide water from water sources to the customers meters, so the position of this meter and the other network components must be identified in the network, and we have two methods to find the positions of these components. The first way is by GPS (global position system). But this is an expensive option. Another method is to employ predetermined location node information with GIS technology. In contrast, with a distributed embedded system, the scalability of a water network means that it can still scale and maintain its efficiency even if we add a large number of nodes to the network. The scalability of a water network allows consumers or nodes inside a building to be included in such a way that the system continues to perform its tasks without affecting system performance. The water network requires scalability, which allows the network to add a large number of nodes without reducing the network's efficiency or performance.

In this section, I will examine the water system from a CAS (complex adaptive system) perspective, as well as how to construct an intelligent system utilizing GIS technology. A. Rattrout et al. [41] describe CAS as a collection of integrating nodes or agents that come together to form a behavioral pattern. CAS may be used in a variety of domains. There are certain CAS characteristics that are prevalent; I will examine some of these qualities and how they apply to the water system. The complex adaptive system components combine to generate the overall system behavior. The water system has several agents or components, such as users, pipelines, wells, customer meters, and pumps etc. The CAS component must adjust to changes in the water environment. Due to their temporal-spatial arrangement, CAS level water adjusts and modifies their internal model and behavior. For example, as the water cycle begins, the network responds to any change smartly in case of any failure in any element. Water networks are dynamic systems that change over time as a result of internal changes or external factors. Water pressure and flow, for example, alter in an unforeseen way over time.

Aggregation, flows, internal models, tagging, diversity, non-linearity, and building blocks are some of the aspects of CAS as defined by John Holland. Each component in a water system that uses GIS technology is recognized by a particular tag that indicates the tagging mechanism. The water system is a non-linear system since the change in output is not caused by a change in input, and data changes from zone to zone and from water cycle to water cycle. As an example, when we pumped water into the zone, not all of it reached the customers, and part of it was lost as a result of real or commercial loss. Water system components form the building blocks for distributed system services. [41]. determining how to construct a water network that can adapt to changes in structure while maintaining low communication costs will be the problem (localization). Node localization is the process of making all deployed nodes aware of their location. I've grouped the water network into many zones, each of which represents a DMA (District Metered Area).

Conclusion

GIS is one of the most effective solutions for managing data since it integrates all geographic data with data from many sources, it can now assist in spatial decisions. As a result, GIS has become one of the most effective tools for managing water systems.

Chapter Three

Our Approach for building water network

3.1 Water System Geo database

Databases are commonly used to store, manage, and modify data. A geo database is a database that stores spatial or geometric data in addition to conventional or multimedia data. Geo databases contain datasets such as feature classes, raster datasets, and network datasets such as geometric networks or network datasets. All of these datasets may be modified and operated using a Database Management System (DBMS) like SQL (Structured query language) or ArcGIS. When a geo database is saved as a spatial type, SQL may access, modify, and update it.

A storage model exists for the geodatabase. We have many feature classes with various data types, like point, line, and polygon. The water meters and the tank feature class are recorded in the geodatabase as a point feature class. It is saved in the form of a table containing attributes and records. Every attribute is a characteristic, and every record is a feature. The point geometry was saved in the shape. There are several benefits for storing data in a geodatabase. Subtypes can be defined in a geodatabase. Subtypes are branches of the primary type, such as land use, commercial, industrial, religious, and agricultural.

Water resource information is intrinsically geographical. The global water distribution network is constantly increasing. Pipelines are used to transport water. However, this does not imply they are risk-free. The water distribution leaks are among the most serious issues that must be evaluated and recognized.

Before repairing the water network, it is required to verify the network's difficulties, such as the connection relationships between the components of the network, the life span of the components, etc.

3.2 Water Geometric Network

Geometric Network is a novel participatory decision-making process that is based on the development of a geographic information system database. The essential components of a water system are the water source, main pipes, laterals, service connection, fittings, and water sink. There are two types of water distribution systems that vary by nation in terms of water availability and water infrastructure: continuous supply and intermittent supply.

Geometric networks enable the representation of typical networks in the real-life world, like water distribution, water flow, and electrical cables. All of these networks may be modeled and analyzed using a geometric network. I constructed geometric networks but not network datasets in my water network because geometric networks enable agents to go in only one direction at a time on edges. As a result, the agent is unable to pick which path to take. External factors, including electromagnetism and water pressure, determine the destination. As a result, the programmer has the power to manage the flow of the agent by manipulating how external influences affect it. Network datasets, on the other hand, such as street, allow the agent to traverse on edges in both directions.

A geometric network is made up of a number of feature classes. All feature classes included in the geometric network have to be recorded in the same feature dataset, which is constructed under the file Geodatabase. ArcCatalog allows me to build new feature classes and subtypes to make editing easier. A feature class is just a table in a database that contains records, including geometric data. A type is assigned to each record. Geometry is a specific discipline that includes concepts such as points, lines, and polygons.

When I add them to ArcMap, they constitute a layer, and the layer cannot exist without data, a table, a feature class, or a source. When the geometric network is constructed, the geometric network icon will appear inside the feature dataset in the ArcCatalog Tree. Since all edges finish with a junction, geometric networks such as water, electricity, sewage, and gas networks are made up of edges (line feature classes). These include both distance and orientation. Point feature classes (Junctions) Junctions can exist in two states: one as the termination point of an edge and the other as an intersection of two edges. If the point in the point feature class is not accessible to act as a junction, an orphan junction in the junction feature class is produced. Simple or complicated edges and junctions can exist. Simple edges have two junctions at each end and correspond to a single edge component in the logical network. Because a complex has not less than two intersections throughout its length, it has one or more edge elements in the geometric network.

Simple junctions are junctions that correspond to a point in the point feature classes, which are divided into two types: Source-location junction where the flow begins and Sink-location junction where the flow stops. The second type of junction is the complicated junction (orphan junction), which is generated when the initial edge is introduced to the geometric network rather than based on user source data.

In the real world, there are several networks that consist of water distribution, electrical connections, and transportation networks. GIS is utilized as a tool for modeling all of these networks so that they may be manipulated, managed, queried, and analyzed. Many studies have been conducted using GIS technology in all forms of these networks. (GIS) are repeatedly utilized in urban planning. GIS is used to design the location of many different network elements, like energy supply systems, gas supply networks, and water systems.[1] Jiang et al. employed GIS technology to depict and forecast traffic flow in a transportation network. Jiang et al. employed topological representations based on streets. Other modeling, such as pedestrian modeling and crime analysis, was applied, allowing for more analysis and improved results [43]. Many more studies make use of GIS technology to track, manage, simulate, and evaluate water delivery networks. Ruggeri assesses the Byzantine water supply system using satellite data, GPS data, and a GIS environment [44].

3.2.1 Geometric Network Component

A geometric network has edges and junctions, which are lines and points that are given spatial names. When we developed the geometric network, we defined the edges as simple or complicated. The feature dataset generates a geometric network. Feature classes in feature datasets will have new behavior because it may alter. It is more than just a line and a point. Edges and junctions are so named because edges know what junctions are linked to, and junctions clarify what edges are connected to. So when network feature classes are in a geometric network, they exhibit some extra behavior, which is why we utilize these spatial terms. A connectivity link exists between the features classes specified in my network in a geometric network. Now, a key point about this is that network connectivity is built on geometric coincidence, which is why we call it a geometric network. So, when you add or remove features in a geometric network, the way those features are connected is determined by their X and Y coordinates.

The connectivity index is a set of tables that reside behind the scenes in the geo database. It stores all connection information between network features. If you have primary information stored in a connectivity index regarding a specific feature and decide to delete this feature data, it will be deleted automatically without user intervention because the connectivity data in these tables is not displayed in ArcGIS. We do not interface directly with ArcCatalog or ArcMap with these tables that we utilize behind the scenes to maintain connection and keep it up-to-date.

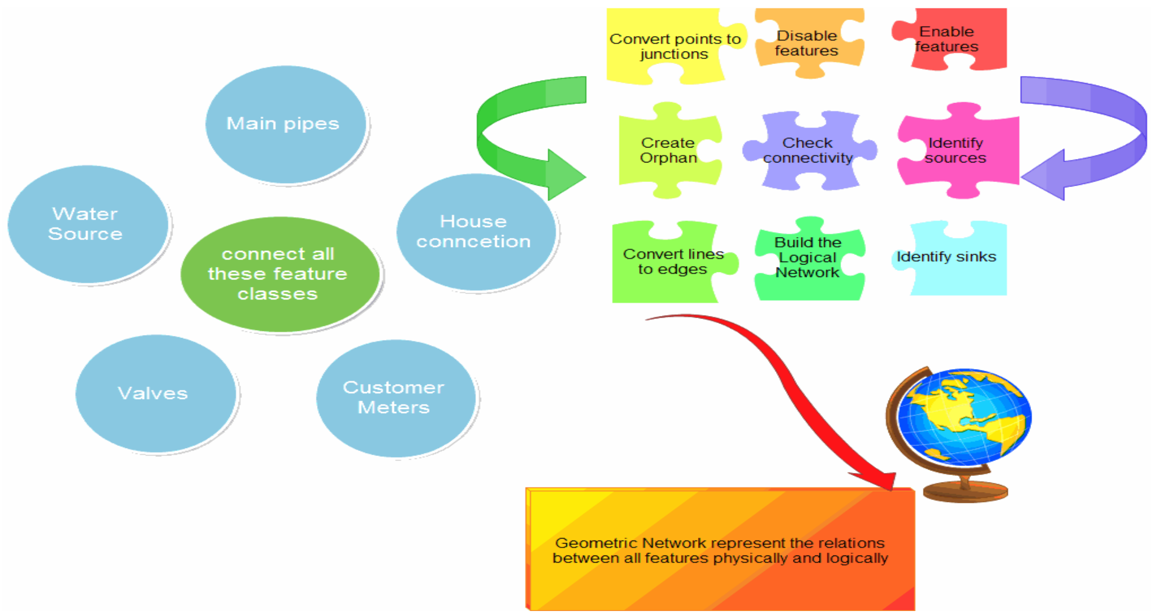


Figure 4: The Geometric Network Component How it is Manipulate And connect to provide the final Geometric network with all connected component [44].

3.2.2Geometric Network Analysis

The data obtained for the industrial zone consists of five layers of Tank water sources. Each source has an identification number and data showing its type and statues. The (2,133) water meters have identification number, meter number, coordinates and meter reading for specific months; Construction number; a pumping area; a pressure area, etc., valves that control water access and cutting; 18 valves have identification number and diameter; and its condition and type have counting types as follows: Air-valve, Ball, Butterfly, Bypass, check, cone, gate, hydraulic, Isolation-valve, plug and Wash-out.

The main pipes drenched at a low depth of 202 have an identification number and the length of each tube and the material made of it are its external and internal diameters, as well as its spatial coordinates pumping area and pressure zone. The household connections pipes were 99% non-existent. There were a small number of connections. The remainder was manually added with an identification number, length and material, of which 2,244 were house connections.

The geometric network faces two types of problems: the first is the lack of connection between the customer meter and the house connection. Thus, there were many meters that were not connected to the network pipes, and this is not acceptable in the geometric network. On the other hand, if we draw the service connection, the model will involve a large number of pipes, which means a complex process. The second problem is that many pipes are not connected to any edges. I have solved these problems and prepared shapefiles that support the network using the following scenario:

First of all, I used an aerial image of the study area to solve problems in a guaranteed way because converting data from a CAD format into shapefiles resulted in some unacceptable errors in the geometric network.

Secondly, I processed data errors to access a perfect geometric network.

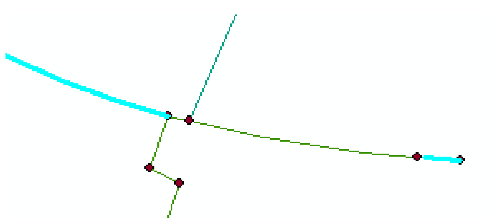
The geographical database contains a number of form files, first, The tank (water source) supplying the area (ain difna), Secondly, the main pipes that are distributed over the area of the study area are Third, home connections are subcontracts that connect the main pipes and consumers , Fourthly, consumers' water counts, which perform the task of measuring their water consumption, and finally, water valves that are distributed over the course of the main pipes that are used to control the water flow. Each of these files was displayed as a separate layer in the ArcMap.

Using the geometric network editing tool, the geometric network was constructed, maintained, and the errors in the water pipes were corrected. The feature dataset's geometric network can identify seven different types of errors, which are listed in a table in figure (5). They will be detailed in the next chapter

Two types of issues with the Nablus City data have come up for me: error 11, which indicates that "The feature's geometry has multiple parts." Since the parts of the two highlighted geometric shapes are discontinuous or disconnected, figure (6) below provides an example. Error 16—"The junction is not connected to any other edge feature"—is the second type, as seen in figure (7).The geometric network editing toolbar and the NET_BUILDER table, which are used to identify all network issues, are shown in Figure. The errors are kept in this table.

Error Code	Error description
10	The feature has an empty geometry.
11	The feature's geometry has multiple parts.
12	The feature's begin and end vertex are the same.
13	The feature has a geometry with zero length.
14	The edge feature is associated with the same from–to junction feature.
15	The junction is coincident with an edge–feature vertex having a different z-value.
16	The junction is not connected to any other edge feature.
17	Edge features prevented from collapsing on themselves because their length is near the snapping tolerance.

Figure 5: The Geometric network build errors.



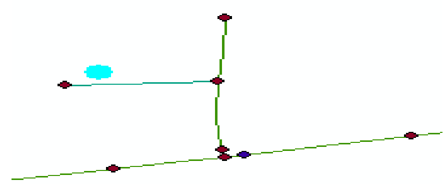


Figure 6: An example of error 11 in GN

Figure 7: : An example of error 16 in GN

I have located loops and connected and disconnected components in the network using the utility network analyst. The figure (8) below provides an example of detecting loops.

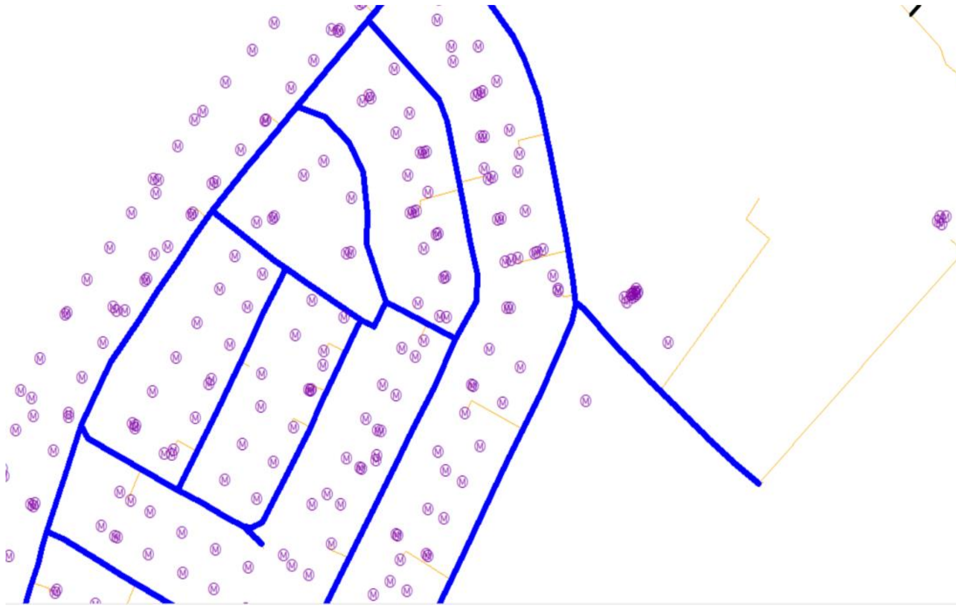


Figure 8: Detecting Loops.

Conclusion

A geodatabase allows for the ordering of water systems. It connects every element of water network using the geodatabase for the water system. I have examined the water network using geometric networks to identify issues and provide solutions.

Chapter Four

Data Preprocessing

4.1 Spatial Data Preprocessing

We have had a number of problems during the course of this research, the most important of which is the quality of the data. These data are archived and preserved in the municipality in CAD format, and when they are converted to shapefiles, this leads to an impairment of its validity. Also, the process of cumulative data preservation and not updating the old version of the data to collect two copies of the data for the same point is inconsistent with the requirement that this type of data is sensitive and must remain with high accuracy and higher quality.

For the data in our study area, several problems and issues have emerged that should have been resolved once and for all before the network was built (figure 9), the data before modified

Figure 9: Data Before Modified.

The network is confronted with some sorts of issues: The house connections layer was 99% non-existent or disconnected, so the missing connections between the main pipe and the consumer meter had to be added. That's why we used an aerial image of the area to make corrections closer to reality. (Figures 10a, b)

(a)

Figure 10: (a) Meters without connections to main pipe. (b) The map with the aerial image.

The following figure (11) shows the shape of the map after the addition of house connections pipes and the complexity of the network.

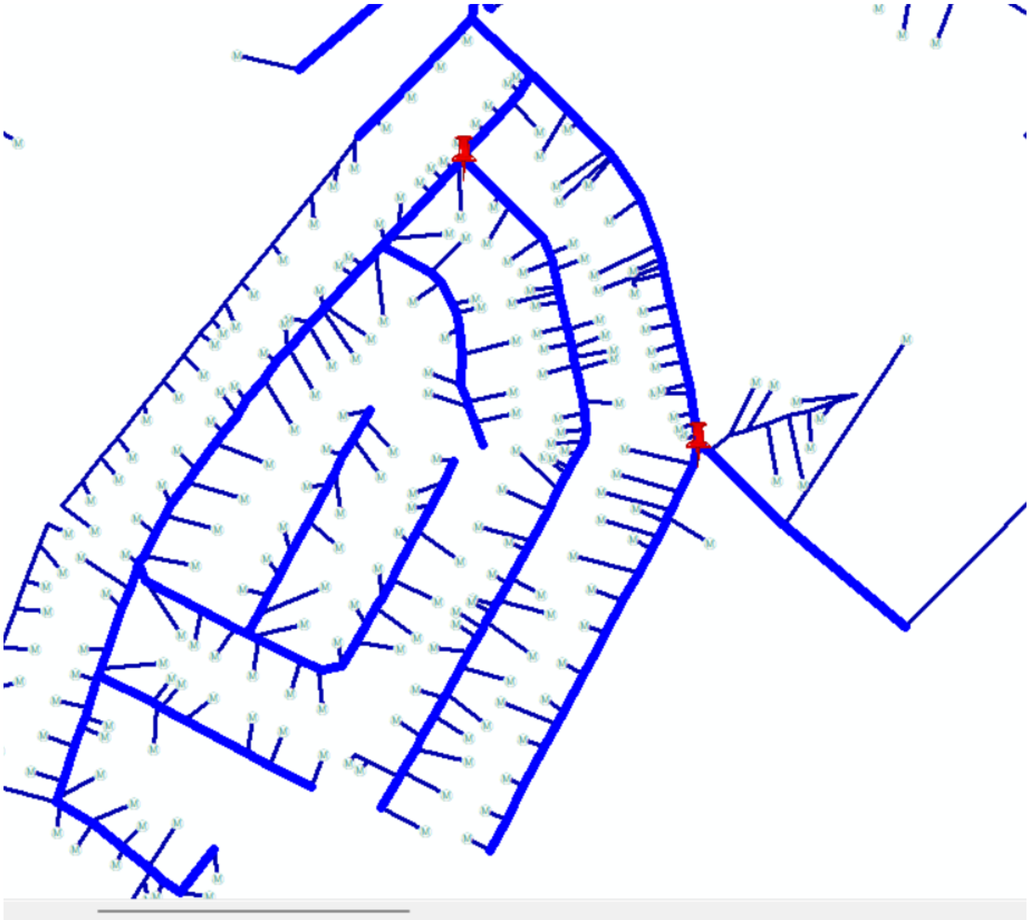


Figure 11: The Map Complexity.

Converting the data from CAD to shapefiles, some pipes appeared to be cut off from both sides without contact with any lines which led us to the deletion of some of them, as shown in figure (12), so we deleted the excess of it and connected it logically and closer to reality.

Figure 12: Random distributed Pipes

Some water meters had their geographical coordinates away from the buildings; like houses for example, because of the lack of a house connection layer, we manually returned the meters to their right place and connected them to the main pipes using house connection pipes.as shown figure (13).

Figure 13: Disconnected Meters.

The network contained a number of loops and this is a fatal error in the geometric network because it hinders tracking the flow of water and estimating users' consumption. As shown in figures (14 a), (14 b).

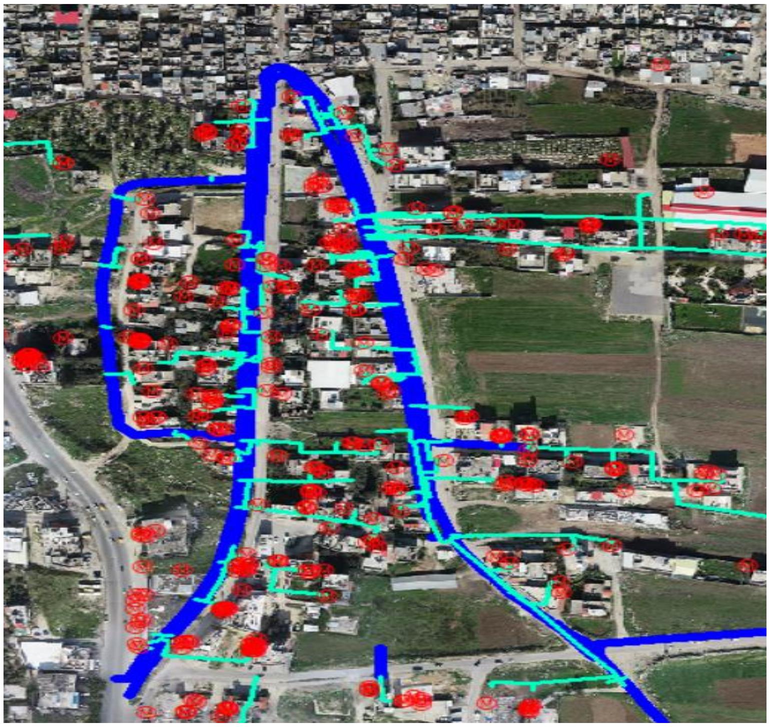


Figure 14: a: Loops in Pipes

Figure 15: b: Loops in Pipes.

During the processing of main pipes, we noticed that pipes are of two layers, such as a long, bent pipe attached to the next pipe of the bending joint. This is also a fatal error that impedes the tracking of water flow. As shown in the figures (15 a), (15 b), colored pipes were under the main pipes.

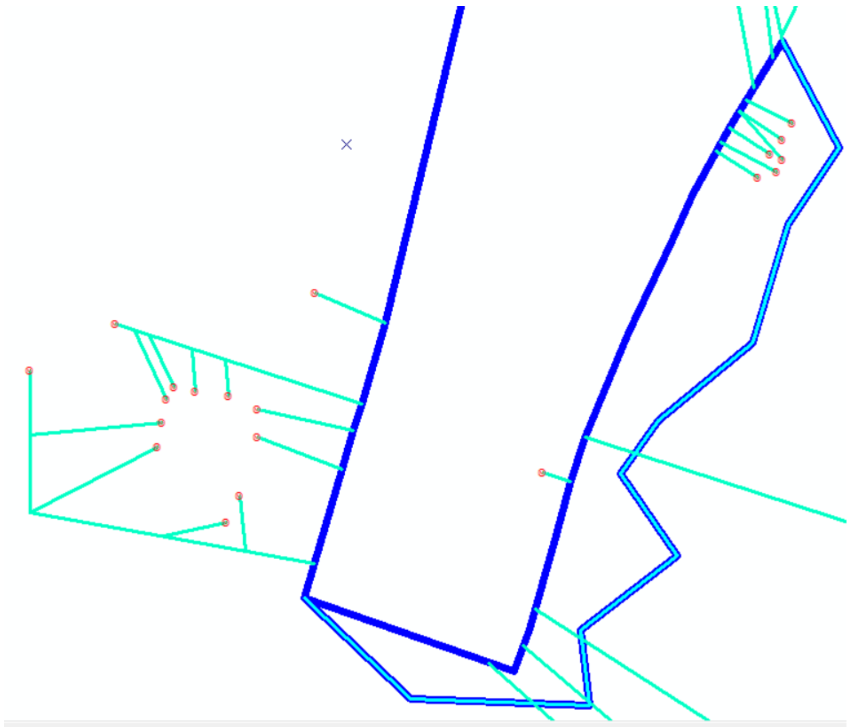


Figure 16: a: Bent Pipes.

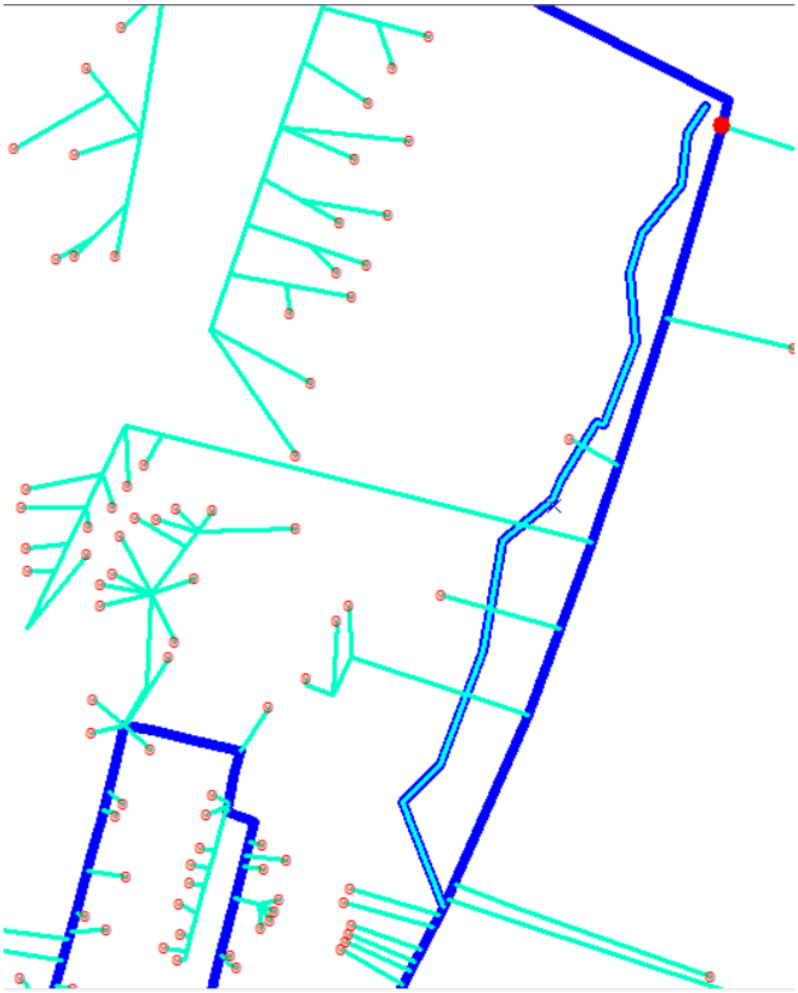


Figure 17: b: Bent Pipes.

At the beginning of the work, we set up a geometric network using data before any modification, producing a network with more than (2500) errors, the figure (16) building errors number. This network contained more than (2500) errors monitored in building error table, Figure (17) shows building errors table.

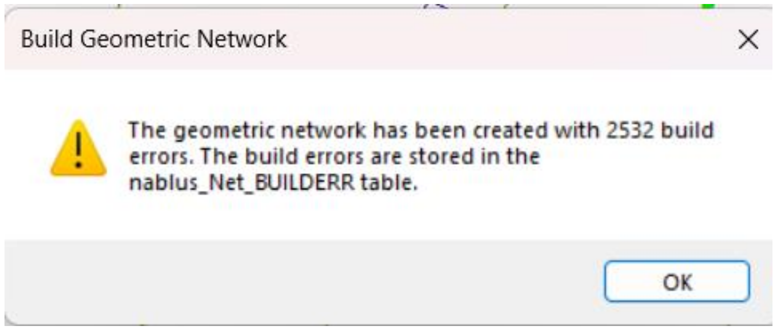
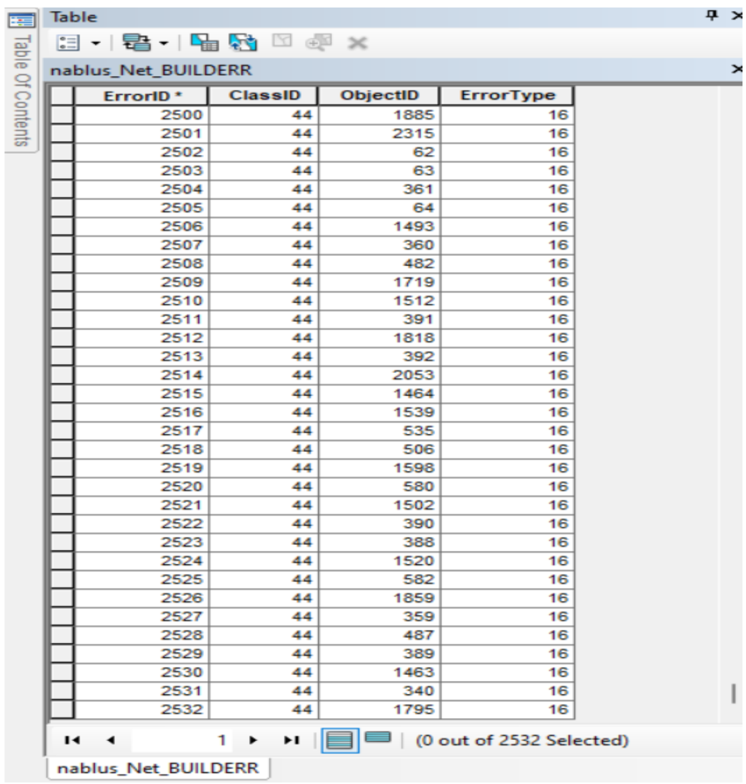


Figure 18: The Geometric Network Errors Number.



ErrorID *	ClassID	ObjectID	ErrorType
2500	44	1885	16
2501	44	2315	16
2502	44	62	16
2503	44	63	16
2504	44	361	16
2505	44	64	16
2506	44	1493	16
2507	44	360	16
2508	44	482	16
2509	44	1719	16
2510	44	1512	16
2511	44	391	16
2512	44	1818	16
2513	44	392	16
2514	44	2053	16
2515	44	1464	16
2516	44	1539	16
2517	44	535	16
2518	44	506	16
2519	44	1598	16
2520	44	580	16
2521	44	1502	16
2522	44	390	16
2523	44	388	16
2524	44	1520	16
2525	44	582	16
2526	44	1859	16
2527	44	359	16
2528	44	487	16
2529	44	389	16
2530	44	1463	16
2531	44	340	16
2532	44	1795	16

Figure 19: Building Errors Table

The data processing process was done in two stages, in the first stage the number of errors of the geo-network building process decreased from more than (2,500) errors to (67) errors as shown in figure (18) showing the number of errors after the first phase.

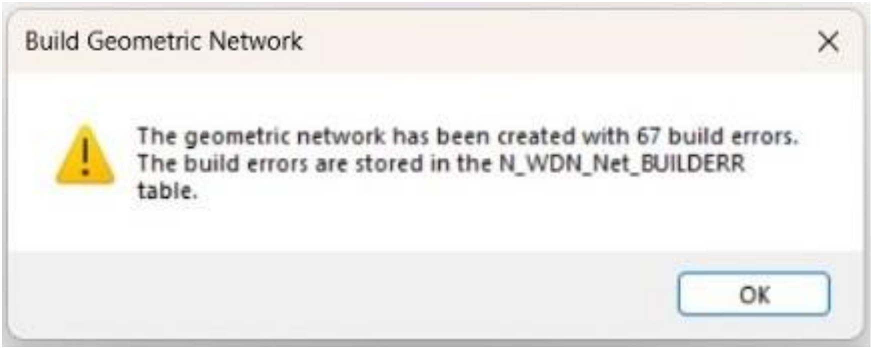


Figure 20: Building Errors Number in the first processing phase.

In the second stage, the number of errors decreased to zero and we got the Geometric network without any errors, errors in the network were processed according to the table of errors from ESRI (Environmental Systems Research Institute).

4.2 Topology Rules

For spatial data in a GIS to remain accurate, consistent, and intact, topology rules are necessary. They offer a framework that facilitates trustworthy spatial analysis and decision-making, guarantee high-quality data, and validate spatial relationships.

In GIS, topology rules are used to specify and impose constraints and spatial relationships between various features in a geospatial dataset. These guidelines support the preservation of data consistency, accuracy, and integrity—all of which are essential for decision-making and spatial analysis. Here are some typical applications for topology rules:

1- Data Integrity:

Data integration means integrating consistency and harmonizing data from several sources into a single, comprehensible format that can be utilized for a range of operational, analytical, and decision-making objectives [73].

- Harmonizing Datasets: Topology rules assist in ensuring that features, such as roads from various sources connecting smoothly, align correctly when integrating multiple datasets.
- Consistency across Layers: Make sure that various feature classes are spatially consistent with one another, for example, by making sure that river and buildings do not cross.

2- Spatial Analysis and Modeling:

Spatial analysis means investigation entities by inspection, assessment, evaluation, and modeling of spatial data features—such as locations, attributes, and their relationships—that disclose the geometric or geographic properties of the data [74].

- Accurate Analysis means accurate results from spatial analyses, like overlay operations and proximity analysis are guaranteed by stable spatial relationships.
- Model Building means to create and analyze spatial models, such as hydrological models that need accurate stream connectivity, maintain accurate topology.

3- Editing and Quality Control: Quality assurance and editing are essential parts of managing GIS data. They guarantee the accuracy, consistency, and dependability of spatial data, which are necessary for efficient modeling, analyzing and decision-making. High standards of data integrity and quality can be upheld by GIS specialists by employing powerful editing tools and strict quality control procedures [75] [76]. Such as Automated Error Detection — Errors during data editing can be automatically found and flagged by topology rules; Quality Assurance—Make sure edits follow specified spatial rules so that high standards for data quality are maintained.

4- Data Validation and Error Detection:

Validating entry data decreases the possibility of errors entering the system and increases data input accuracy by guaranteeing that data satisfies quality standards prior to storage, it offers users real-time guidance and lessens the need for subsequent data correction and maintenance [77].

- Ensuring Data Integrity: Errors in spatial data, like gaps, overlaps, or misalignments, can be found and fixed with the aid of topology rules.
- Consistency Checks: it makes sure certain requirements are met by the spatial relationships between features, like no overlapping polygons in a land use dataset.

5- Cartography and Visualization:

Mapping techniques are used to transmit data that include spatial information about on the surface of the earth like points, paths [78]. This application provides:

- Clean Map Representation: To provide a clear and accurate representation of spatial data, make sure that maps are free from visual errors such as gaps or overlapping features.
- Semiology and Labeling: In order to properly arrange labels and symbols and prevent conflicts and overlaps, topology rules can be helpful.

6- Maintaining Spatial Relationships:

spatial relations (SR) is the ability to determine an object's position in relation to a previous position by taking into account its size, distance, or any other distinguishing sign. [79].this application helps to obtain:

- Adjacent Features: Making sure that there are no spaces or overlaps in the common boundary between neighboring features.

- Connectivity: connecting all network features—such pipelines correctly for a reliable network analysis,

There are a number of topology rules, but we don't need all of them. We have only applied what our data needs. Here are the two needed rules:

- Must Not Self-Intersect this rule is to make sure that no two lines features cross over one another.
- Must Be Covered by Endpoint Of this rule is to make sure that point features, like junctions, cover the endpoints of line features, like pipes.

These two rules we had to make sure they were properly applied in the geometric network to ensure that all components and network work correctly. We checked using utility tools to track the flow and the network appeared as shown in the figure (19).

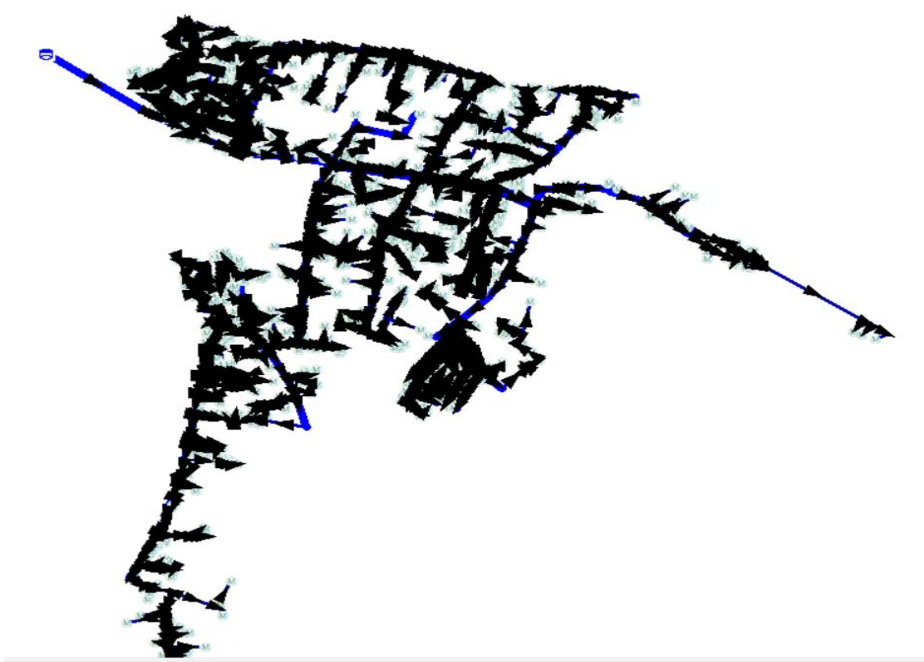


Figure 21: Track the correct water flow in the network.

Based on Geometric Network, many analyses can be done on the network such as identifying connected areas and disconnected areas in case of breaking. (Figure 20)

Figure 22: The Geometric Network Connected and Disconnected using the Utility Network Analyst toolbar.

4.3 Meters Maintenance Data:

Water meters in the water system are used to measure consumers' consumption and have several types that are installed under municipal supervision, including Actars, Baylan and others. The shelf life of these meters is five years, but during the period of operation, they may cease to operate due to emergency breakdowns, which require either repairing the meter or replacing it with a new one, depending on the type of failure.

After considering that the study area (industrial zone) is a relatively small area and the maintenance cases in which it is not enough to analyze and anticipate had to find a solution, the solution was to rely fully on the maintenance data of the city of Nablus to train the model and adopt the data of the industrial area for testing.

For the municipality of Nablus, maintenance data for meters is kept in Excel sheet so handling it needed special attention, although there is a common field between maintenance data and our network data, which is the meter's number, maintenance data needed to be thoroughly processed before it was ready to be analyzed and utilized. As shown in figure (21), Arabic was used in the documentation process, And in the address field, the same place was almost encrypted in more ways than one (Many ways to write the address), this is what I need to process and convert data into English and address all maintenance cases correctly.

A	B	C	D	E	F	G	H	I	J	K	L
التاريخ	رقم الخدمة	العنوان	العمل المطلوب	قطر ا	نوع العداد	المصدر	السبب	نتيجة الفحص	منطقة الضخ	ملاحظات	نوع العداد الجديد
02/01/2023	15584	نابلس!/ش!عمربنالخ	تغيير عداد		Kent	القارئ	العداد متوقف عن التسج	تم تغيير العداد	W2B		Kent
02/01/2023	2075	نابلس!/ش!قدري	تغيير عداد		اكتارس	القارئ	العداد متوقف عن التسج	تم تغيير العداد	W2B		Kent
02/01/2023	38479	ش!التعاون/داررابع	تغيير عداد		اراد اسرائيلي حجمي	القارئ	العداد متوقف عن التسج	تم تغيير العداد	W4_1		بايلان
02/01/2023	37478	التعاون/طابقرابع	تغيير عداد		بايلان	القارئ	العداد متوقف عن التسج	تم تغيير العداد	W4_1		بايلان
02/01/2023	36994	رفيديا/دارتسويه	تغيير عداد		بايلان	القارئ	العداد متوقف عن التسج	تم تغيير العداد	W4_1		بايلان
02/01/2023	54032		تغيير عداد		بايلان	القارئ	العداد متوقف عن التسج	تم تغيير العداد	W3		بايلان
02/01/2023	36367	المخفيهطابقخامس!	تغيير عداد		Kent	القارئ	العداد متوقف عن التسج	تم تغيير العداد	W3		Kent
02/01/2023	18943	نابلس!/رفيديا	تغيير عداد		اراد اسرائيلي حجمي	القارئ	العداد متوقف عن التسج	تم تغيير العداد	W3		بايلان
02/01/2023	34135	المخفيه/دارثاني	تغيير عداد		بايلان	القارئ	العداد متوقف عن التسج	تم تغيير العداد	W3		بايلان
04/01/2023	58101		تغير من مسبق دفع الى مسبق	0	YLAN AK-311 مسبق دفع	البلدية	العداد تالف	تم تغيير العداد	C1W		مسبق دفع BAYLAN AK-311
03/01/2023	43953	زواتا	تغيير عداد		بايلان	القارئ	العداد متوقف عن التسج	تم تغيير العداد	Zawata_V		بايلان
03/01/2023	19669	نابلس!/المعاجين	تغيير عداد		Kent	القارئ	العداد متوقف عن التسج	تم تغيير العداد	W-1		بايلان
03/01/2023	29509	مفرقزواتا/دارثانيقبليه	تغيير عداد		فرنسي TD8	القارئ	العداد متوقف عن التسج	تم تغيير العداد	W-1		بايلان
03/01/2023	29510	مفرق زواتا دار ارضي !	تغيير عداد		بايلان	القارئ	العداد متوقف عن التسج	تم تغيير العداد	W-1		بايلان
03/01/2023	21640	زواتا	تغيير عداد		Kent	القارئ	العداد متوقف عن التسج	تم تغيير العداد	Zawata_V		بايلان
03/01/2023	30427	زواتا/دارارضي	تغيير عداد		بايلان	القارئ	العداد متوقف عن التسج	تم تغيير العداد	Zawata_V		بايلان
03/01/2023	15290	نابلس!/مفرقزواتا	تغيير عداد		بايلان	القارئ	العداد متوقف عن التسج	تم تغيير العداد	NW0		بايلان
03/01/2023	29954	زواتا/دارثاني	تغيير عداد		بايلان	القارئ	العداد متوقف عن التسج	تم تغيير العداد	NW0		بايلان
03/01/2023	23748	نابلس!0ش!طولكرم	تغيير عداد		اراد اسرائيلي حجمي	القارئ	العداد متوقف عن التسج	تم تغيير العداد	W-1		بايلان
03/01/2023	18528	نابلس!/زواتا	تغيير عداد		Kent	القارئ	العداد متوقف عن التسج	تم تغيير العداد	Zawata_V		Itron
03/01/2023	20935	نابلس!/اول زواتا	تغيير عداد		Kent	القارئ	العداد متوقف عن التسج	تم تغيير العداد	W-1		بايلان
03/01/2023	4080	نابلس!/المركزالتجاري	تغيير عداد		بايلان	القارئ	العداد متوقف عن التسج	تم تغيير العداد	C1W		بايلان
03/01/2023	12154	نابلس!/ش! حيفا	تغيير عداد		بايلان	القارئ	العداد متوقف عن التسج	تم تغيير العداد	W0	معظم وقت المشترك عند والده والبيت	بايلان
03/01/2023	929	نابلس!/ش! فيصل	تغيير عداد	0	Itron	القارئ	العداد متوقف عن التسج	تم تغيير العداد	C1W		Itron
04/01/2023	25760	ش!التعاون/داراولجنو	تغير من مسبق دفع الى عداد		YLAN AK-311 مسبق دفع	السيد رئيس	بناء على تعليمات رئيس ا	تم تغيير العداد	S4	يوجد رصيد بالكرت 80 شيكل لاحتسابه	بايلان
04/01/2023	57693		تغير من مسبق دفع الى مسبق		YLAN AK-311 مسبق دفع	المشترك	العداد تالف	تم تغيير العداد	W4_1	المشترك له رصيد بالعداد القديم 24 شي	مسبق دفع LXHZ-15
04/01/2023	11093	نابلس!/ مبلاطه	تغيير عداد		Itron	القارئ	العداد متوقف عن التسج	تم تغيير العداد	E0.3		بايلان
05/01/2023	18617	نابلس!/الباذان	تغيير عداد		بايلان	القارئ	العداد متوقف عن التسج	تم تغيير العداد	BADANRESER		بايلان
05/01/2023	30857	الباذان	تغيير عداد		Kent	القارئ	العداد متوقف عن التسج	تم تغيير العداد	BADANRESER		بايلان
05/01/2023	37652	الباذان	تغيير عداد		Itron	القارئ	العداد متوقف عن التسج	تم تغيير العداد	BADANRESER		بايلان

Figure 23: Meter Maintenance Data Before Modified.

The data have been transformed into data that is valid for analysis and utilization of all the parameters these data provide in the analysis and data mining processes to create a predictive system for meter malfunction based on historical data of previous maintenance in Nablus after the necessary adjustments and processing. The data include the date of maintenance, meter number, address, procedure required, meter diameter, type, source, cause of failure, inspection result, pump area and new meter type. Figure (22) shows the data after processing processes.

Date	service num	address	work	meter type	resours	reason	result	erea	type new meter
2/1/2023	15584	omar ibn alkhattab	changing the meter	Kent	reader	The meter is discontinued	changing the meter	W2B	Kent
2/1/2023	2075	qadri street	changing the meter	aktars	reader	The meter is discontinued	changing the meter	W2B	Kent
2/1/2023	38479	alta'won street	changing the meter	arad	reader	The meter is discontinued	changing the meter	W4_1	baylan
2/1/2023	37478	alta'won street	changing the meter	baylan	reader	The meter is discontinued	changing the meter	W4_1	baylan
2/1/2023	36994	rafidia	changing the meter	baylan	reader	The meter is discontinued	changing the meter	W4_1	baylan
2/1/2023	54032	nablus	changing the meter	baylan	reader	The meter is discontinued	changing the meter	W3	baylan
2/1/2023	36367	almakhfia	changing the meter	Kent	reader	The meter is discontinued	changing the meter	W3	Kent
2/1/2023	18943	rafidia	changing the meter	arad	reader	The meter is discontinued	changing the meter	W3	baylan
2/1/2023	34135	almakhfia	changing the meter	baylan	reader	The meter is discontinued	changing the meter	W3	baylan
4/1/2023	58101	nablus	change from prepaid t	baylan	municipality	The meter is damaged	changing the meter	C1W	baylan
3/1/2023	43953	zawata	changing the meter	baylan	reader	The meter is discontinued	changing the meter	Zawata_V	baylan
3/1/2023	19669	alma'jeen	changing the meter	Kent	reader	The meter is discontinued	changing the meter	W-1	baylan
3/1/2023	29509	zawata	changing the meter	French TD8	reader	The meter is discontinued	changing the meter	W-1	baylan
3/1/2023	29510	zawata	changing the meter	baylan	reader	The meter is discontinued	changing the meter	W-1	baylan
3/1/2023	21640	zawata	changing the meter	Kent	reader	The meter is discontinued	changing the meter	Zawata_V	baylan
3/1/2023	30427	zawata	changing the meter	baylan	reader	The meter is discontinued	changing the meter	Zawata_V	baylan
3/1/2023	15290	zawata	changing the meter	baylan	reader	The meter is discontinued	changing the meter	NW0	baylan
3/1/2023	29954	zawata	changing the meter	baylan	reader	The meter is discontinued	changing the meter	NW0	baylan
3/1/2023	23748	tulkarm street	changing the meter	arad	reader	The meter is discontinued	changing the meter	W-1	baylan
3/1/2023	18528	zawata	changing the meter	Kent	reader	The meter is discontinued	changing the meter	Zawata_V	Itron
3/1/2023	20935	zawata	changing the meter	Kent	reader	The meter is discontinued	changing the meter	W-1	baylan
3/1/2023	4080	nablus	changing the meter	baylan	reader	The meter is discontinued	changing the meter	C1W	baylan
3/1/2023	12154	haifa street	changing the meter	baylan	reader	The meter is discontinued	changing the meter	W0	baylan
3/1/2023	929	faisl street	changing the meter	Itron	reader	The meter is discontinued	changing the meter	C1W	Itron
4/1/2023	25760	alta'won street	change from prepaid t	baylan	municipality	Upon the instructions of the re	changing the meter	S4	baylan
4/1/2023	57693	nablus	change from prepaid t	baylan	shared	The meter is damaged	changing the meter	W4_1	LXHZ-15
4/1/2023	11093	balata	changing the meter	Itron	reader	The meter is discontinued	changing the meter	E0.3	baylan
5/1/2023	18617	bathan	changing the meter	baylan	reader	The meter is discontinued	changing the meter	BADANRESERVOIR	baylan
5/1/2023	30857	bathan	changing the meter	Kent	reader	The meter is discontinued	changing the meter	BADANRESERVOIR	baylan
5/1/2023	37652	bathan	changing the meter	Itron	reader	The meter is discontinued	changing the meter	BADANRESERVOIR	baylan

Figure 24: Meter Maintenance Data After Modified.

4.4 House Connection Pipes Maintenance Data

House connection pipes in water distribution systems are the most vulnerable physical component of breakdowns and fractures due to the small volumes and the little depth on which they are dumped and the materials from which they are made. This is why, when reviewing the historical data on pipe maintenance, most of them are 95% due to the piping of house connections, so they had to receive special attention in this study.

Because the case study was very small we needed to use Historical data for the maintenance of house connections in the city of Nablus which was obtained from the municipality in the form of an Excel sheet, but it had several problems that needed to be addressed before it began to be used in data mining.(figure 23).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	رقم المهمة	تاريخ الصيانة	الشارع	العنوان	القطر	الفرقة	الفني	نوع الصيانة	العمق	مياه مهدور	السبب	النوع	القطعة المستخدمه	طقة الضخ
2	84645	01/01/2023	شارع داخلي	البلده القديمه شارع النصر محل حبيشه للدواجن	1/2	ذيب احمد علي المسيمي	ذيب احمد علي المسيمي	صيانة شبكة	0	0	هريان	برابيش	نبل 1/2 ، مجوز(ء	C1
3	84646	01/01/2023	شارع داخلي	نزلة الماتورات مسجد الحاج عدلي / تم فتح المحبس ووص	1/2	ذيب احمد علي المسيمي	ذيب احمد علي المسيمي	ازالة سدد	0	0	ازالة سدد	فولاذ		C1
4	84647	01/01/2023	شارع داخلي	الضاحيه الوسطى عند بيت مرمش / شبك بربيش 20 ملم	20mm	محمد ديب محمود منسي	محمد ديب محمود منسي	شبك خط	0	0	شبك بربيش	برابيش	كوع مجلفن بسن ا	SE2
5	84648	01/01/2023	شارع داخلي	نابلس الجديده قبل فيلا بانا	1 inch	محمد ديب محمود منسي	محمد ديب محمود منسي	صيانة شبكة	0	0	سيلان خط	فولاذ	شد وصل 1(عدد)	S5
6	84649	01/01/2023	شارع داخلي	شارع المياه عند بيت القدع	1/2	محمد ديب محمود منسي	محمد ديب محمود منسي	صيانة شبكة	0	0	سيلان خط	فولاذ	محبس جل 1/2 (	مساكن الشعبية
7	84650	01/01/2023	شارع داخلي	المساكن الشعبيه العليا مقابل بيت المحامي علي عيساوي /	1 inch	محمد ديب محمود منسي	محمد ديب محمود منسي	فصل	0	0	فصل خط مؤقت	فولاذ	مسد مجلفن سن .	مساكن الشعبية
8	84651	01/01/2023	شارع داخلي	المساكن الشعبيه امام عمارة العرندي / ازالة سدد	1/2	خالد جمال ياسين الاغبر	خالد جمال ياسين الاغبر	ازالة سدد	0	0	ازالة سدد	فولاذ		مساكن الشعبية
9	84662	02/01/2023	شارع داخلي	المساكن الشعبية بجانب فلل المدفع	1/2	علي عصام محمد دويكات	علي عصام محمد دويكات	صيانة شبكة	0	0	سيلان خط	فولاذ	شد وصل 1/2 (ع	مساكن الشعبية
10	84665	02/01/2023	شارع داخلي	مفرق زواتا الاول بعد مركز فحص كورونا	16mm	محمد اشرف محمد حشاش	محمد اشرف محمد حشاش	صيانة شبكة	0	0	سيلان خط	بلاستيك بولي اثيلين	محبس جل 1(عد	Zawata_V
11	84668	02/01/2023	شارع داخلي	نابلس الجديده قرب مسجد البرقاوي	25mm	رائد غالب صدقي الخطيب	رائد غالب صدقي الخطيب	صيانة شبكة	0	0	سيلان بربيش	برابيش	شد وصل 25 ملم	S5
12	84669	02/01/2023	شارع داخلي	المساكن الشعبيه عمارة العرندي	1 inch	ذيب احمد علي المسيمي	ذيب احمد علي المسيمي	صيانة شبكة	0	0	سيلان خط	فولاذ	شد وصل 1(عدد)	مساكن الشعبية
13	84670	02/01/2023	شارع داخلي	طلعة الفاطميه ديوان ال النابلسي	1/2	حكمت عمر زكي هارون	حكمت عمر زكي هارون	صيانة شبكة	0	0	سيلان خط	فولاذ	نبل حديد 1/2" 0	C1
14	84703	02/12/2023	شارع داخلي	كطمال جنبلاط / تركيب كوبلييج بحاجة الى تغيره الى درسر	2 inch	ودت بن ناصر البشتاوي	ناصر" محمدفاروق "جودت بن	صيانة شبكة	50	0	كسر خط من قبل متعهد	فولاذ		W2a
15	84704	02/01/2023	شارع داخلي	شارع جمال عبد الناصر مقابل مخبز نصار	1/2	علي عصام محمد دويكات	علي عصام محمد دويكات	فصل	0	0	فصل خط ملغي	فولاذ	مف مجلفن 1/2 (	C1
16	84706	02/12/2023	شارع داخلي	رفيديا علاء الدين عمارة الاخرس / تبين انم الخط بعد العد	1/2	خالد جمال ياسين الاغبر	خالد جمال ياسين الاغبر	صيانة شبكة	0	0	سيلان خط	فولاذ		W2a
17	84707	03/01/2023	شارع داخلي	الضاحيه التحتا عند بيبت جاموس / هريان خط 3/4" شب	32mm	رائد غالب صدقي الخطيب	رائد غالب صدقي الخطيب	تغيير خط مهري	30	0	هريان	بلاستيك بولي بروبلين	ترانزشن 32 ملم 1	SE1
18	84708	03/01/2023	شارع داخلي	البلده القديمه سوق البصل محل عزام طبيله للعطور / ص	1/2	ودت بن ناصر البشتاوي	ناصر" محمدفاروق "جودت بن	صيانة شبكة	0	0	سيلان خط	فولاذ	محبس جل 1/2 (	C1
19	84719	03/01/2023	شارع داخلي	البلده القديمه سوق البصل محل عزام طبيله للعطور / ص	1/2	ودت بن ناصر البشتاوي	ناصر" محمدفاروق "جودت بن	صيانة شبكة	0	0	سيلان خط	فولاذ	محبس جل 1/2 (	C1
20	84720	03/01/2023	شارع داخلي	مفرق زواتا الثاني قرب فيتامين مول	1/2	ودت بن ناصر البشتاوي	ناصر" محمدفاروق "جودت بن	صيانة شبكة	0	0	سيلان خط	فولاذ	محبس جل 1/2 (	W-1
21	84721	03/01/2023	شارع داخلي	رفيديا كمال جنبلاط قرب الخزان / شبك خط 63 ملم اخط	63mm	سامر نعيم خليل سايح	سامر نعيم خليل سايح	تغيير خط مهري	50	0	هريان	بلاستيك بولي بروبلين	ترانزشن 63 ملم 2	W2a
22	84722	03/01/2023	شارع داخلي	حارة الحبله عند فرن ابو الامين كلبونه / هريان خط حديد	32mm	علي عصام محمد دويكات	علي عصام محمد دويكات	تغيير خط مهري	30	0	هريان	بلاستيك بولي بروبلين	مف لحام كهرباء 2	C1
23	84733	04/01/2023	شارع داخلي	مخيم عسكر الجديد المطاحن العربيه مدرسة الوكاله / تبي	1/2	ذيب احمد علي المسيمي	ذيب احمد علي المسيمي	صيانة شبكة	0	0	سيلان خط	فولاذ		E0.2
24	84734	04/01/2023	شارع داخلي	دوار عصيره مقابل مدرسة الحاجه رشده على درج البلديه	3/4	رائد غالب صدقي الخطيب	رائد غالب صدقي الخطيب	صيانة شبكة	0	0	سيلان خط	فولاذ	محبس جل 3/4 (	W0
25	84735	04/01/2023	شارع داخلي	طلوع الفاطميه مدرسة الفاطميه عند ديوان ال النابلسي	3/4	غسان بسام ناصر ابو شلبك	غسان بسام ناصر ابو شلبك	فصل	0	0	فصل خط ملغي	فولاذ	مسد مجلفن سن .	C1
26	84736	04/01/2023	شارع داخلي	شارع ابو عبيده درج هواش بيت علام التيي / تغير شبكات	40mm	حكمت عمر زكي هارون	حكمت عمر زكي هارون	تغيير خط مهري	40	0	هريان	بلاستيك بولي بروبلين	تحويل 1 على 3/4	C1
27	84744	04/01/2023	شارع داخلي	رفيديا مقابل مضخة كمال جنبلاط عمارة المعتصم سوبرما	2 inch	ودت بن ناصر البشتاوي	ناصر" محمدفاروق "جودت بن	صيانة شبكة	0	0	تغير محبس تالف	فولاذ	محبس جل 2(عد	W2a
28	84745	04/01/2023	شارع داخلي	شارع تل قرب مسجد النور قرب عمارة وضاح عبده / تم ت	1/2	محمد اشرف محمد حشاش	محمد اشرف محمد حشاش	صيانة عداد	0	0	سيلان عداد	فولاذ		W3
29	84746	04/01/2023	شارع داخلي	المخفيه عمارو اشبيليه / تصليح سيلان عداد 2" وتم عمل	2 inch	خالد جمال ياسين الاغبر	خالد جمال ياسين الاغبر	صيانة عداد	0	0	سيلان عداد	فولاذ		W3
30	84747	05/01/2023	شارع داخلي	حاره الحبله بجانب فرن ابو الامين كلبونه / شبك خط 32	32mm	غسان بسام ناصر ابو شلبك	غسان بسام ناصر ابو شلبك	تغيير خط مهري	0	0	هريان	بلاستيك بولي بروبلين	مف لحام كهرباء 2	C1
31	84748	05/01/2023	شارع داخلي	شارع عصيره مقابل مدرسة الحاجه رشده	3/4	رائد غالب صدقي الخطيب	رائد غالب صدقي الخطيب	تغيير خط مهري	0	0	هريان	فولاذ	انابيب مياه سن مع	W0

Figure 25: House Connection Pipes Historical Maintenance Data before Modification.

As can be seen in the figure (23), maintenance data have been archived in Arabic. This is inconsistent with the possibility of making use of it. Therefore, initially it had to be converted into English, and then the different similar data were consolidated in the way it is expressed. As shown in figure (24)

	A	B	C	D	E	F	G	H	I	J	K
1	num_id	task No	maintenance data	Street	diameter	maintenance type	depth	reson	pipe type	Pumping Area	maint
2	1	84645	1/1/2023	Inner Street	2-Jan	network maintenance	0	filly Line	Plastic Polyethylene	C1	1
3	2	84646	1/1/2023	Inner Street	2-Jan	remove repayment	0	remove repayment	steel	C1	1
4	3	84647	1/1/2023	Inner Street	20mm	connect pipe	0	connect pipe	Plastic Polyethylene	SE2	1
5	4	84648	1/1/2023	Inner Street	1 inch	network maintenance	0	leak line	steel	S5	1
6	5	84649	1/1/2023	Inner Street	2-Jan	network maintenance	0	leak line	steel	almsakin	1
7	6	84650	1/1/2023	Inner Street	1 inch	cancelled	0	cancelled	steel	almsakin	1
8	7	84651	1/1/2023	Inner Street	2-Jan	remove repayment	0	remove repayment	steel	almsakin	1
9	8	84662	2/1/2023	Inner Street	2-Jan	network maintenance	0	leak line	steel	almsakin	1
10	9	84665	2/1/2023	Inner Street	16mm	network maintenance	0	leak line	Plastic Polyethylene	Zawata_V	1
11	10	84668	2/1/2023	Inner Street	25mm	network maintenance	0	leak line	Plastic Polyethylene	S5	1
12	11	84669	2/1/2023	Inner Street	1 inch	network maintenance	0	leak line	steel	almsakin	1
13	12	84670	2/1/2023	Inner Street	2-Jan	network maintenance	0	leak line	steel	C1	1
14	13	84703	2/12/2023	Inner Street	2 inch	network maintenance	50	break line	steel	W2a	1
15	14	84704	2/1/2023	Inner Street	2-Jan	cancelled	0	cancelled	steel	C1	1
16	15	84706	2/12/2023	Inner Street	2-Jan	network maintenance	0	leak line	steel	W2a	1
17	16	84707	3/1/2023	Inner Street	32mm	change filly pipe	30	filly Line	Plastic Polypropylene	SE1	1
18	17	84708	3/1/2023	Inner Street	2-Jan	network maintenance	0	leak line	steel	C1	1
19	18	84719	3/1/2023	Inner Street	2-Jan	network maintenance	0	leak line	steel	C1	1
20	19	84720	3/1/2023	Inner Street	2-Jan	network maintenance	0	leak line	steel	W-1	1
21	20	84721	3/1/2023	Inner Street	63mm	change filly pipe	50	filly Line	Plastic Polypropylene	W2a	1
22	21	84722	3/1/2023	Inner Street	32mm	change filly pipe	30	filly Line	Plastic Polypropylene	C1	1
23	22	84733	4/1/2023	Inner Street	2-Jan	network maintenance	0	leak line	steel	E0.2	1
24	23	84734	4/1/2023	Inner Street	4-Mar	network maintenance	0	leak line	steel	W0	1
25	24	84735	4/1/2023	Inner Street	4-Mar	cancelled	0	cancelled	steel	C1	1
26	25	84736	4/1/2023	Inner Street	40mm	change filly pipe	40	filly Line	Plastic Polypropylene	C1	1
27	26	84744	4/1/2023	Inner Street	2 inch	network maintenance	0	change broken valve	steel	W2a	1
28	27	84745	4/1/2023	Inner Street	2-Jan	meter maintenance	0	leak meter	steel	W3	1
29	28	84746	4/1/2023	Inner Street	2 inch	meter maintenance	0	leak meter	steel	W3	1
30	29	84747	5/1/2023	Inner Street	32mm	change filly pipe	0	filly Line	Plastic Polypropylene	C1	1
31	30	84748	5/1/2023	Inner Street	4-Mar	change filly pipe	0	filly Line	steel	W0	1

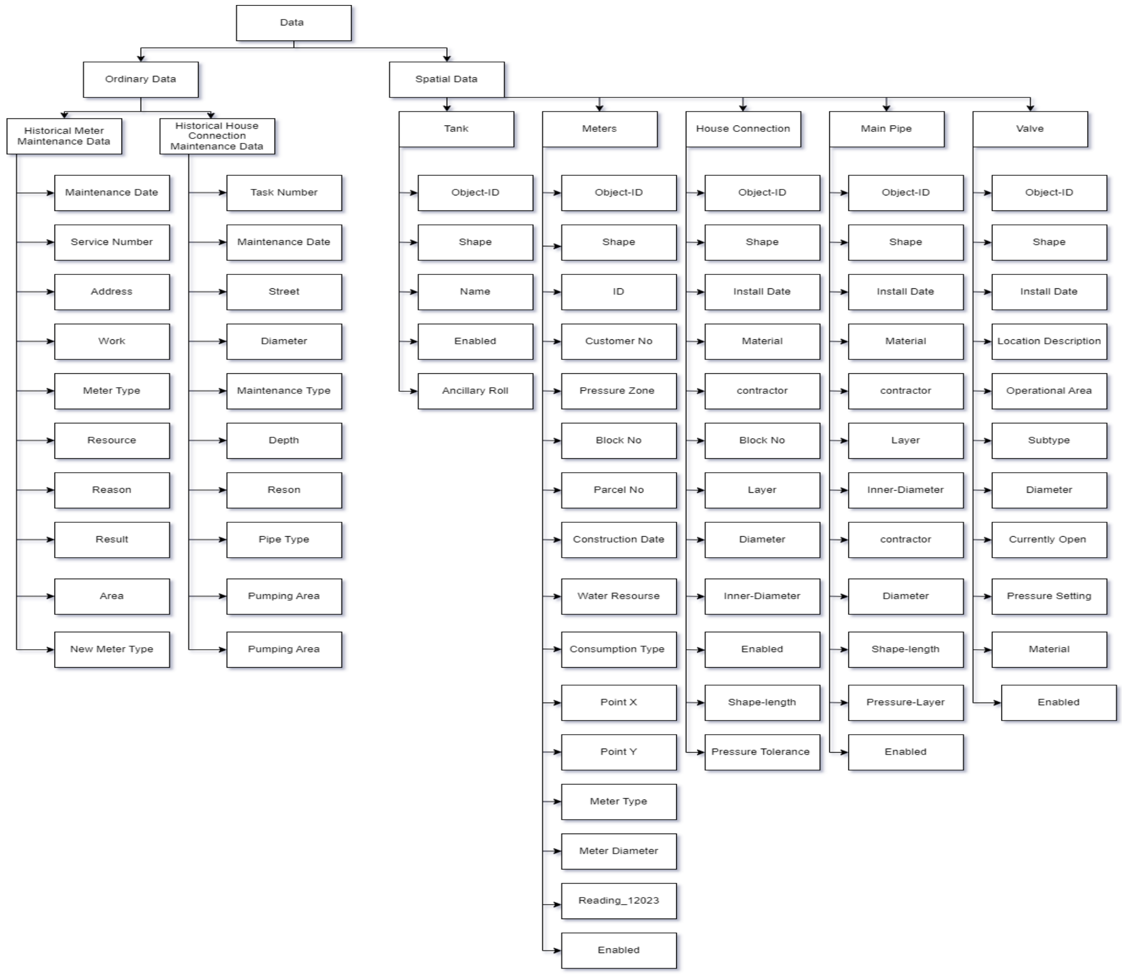


Figure 26: House Connection Pipes Historical Maintenance Data after Modification.

4.5 Calculation of Approximate Cost of Maintenance:

Water Distribution Networks (WDNs) are vital infrastructure for the transportation of water for home and industrial purposes. WDNs are vulnerable to failure due to their subsurface location and several internal and external variables [53, 54]. A single component failure in a water distribution network (WDN) might affect the system's ability to supply appropriate water quality and quantity [55]. When a single component of a WDN (e.g., pipeline, pump, valve) fails or is damaged, a segment (small collection of nodes and pipes) is separated by closing isolation valves to perform maintenance. Segment isolation may cause accidental cut-off of additional segments downstream of the isolated segment [56], depriving them of supply from the primary source(s).

This makes segment isolation in WDNs highly important and must be handled properly to lessen its effect.

WDN assets, such as tankss, valves, and reservoirs, are not always operational because of the dynamic nature of WDNs. The way that flow is distributed throughout the network varies greatly based on how these assets are operating. Furthermore, nodes' pressure and demand also change dramatically over time. This suggests that the WDN's cumulative demand shortfall brought about by a segment being isolated for maintenance becomes time-dependent. If segment I is isolated at time, for example, its effect may be different from that of segment I isolated at time. Therefore, the length of time a segment is isolated for rehabilitation establishes how much of an impact it has on the WDN [57].

Depending on the part that needs to be maintained, the financial cost of each maintenance process will change. The length of the depth of the material made of excavation works and other materials induced to ensure its safety and installation properly [58].Palestinian municipalities deal with private companies to obtain the necessary materials and pieces based on special agreements between them and according to these companies, the prices were as shown in the table(1)(2)(3)(4)

Table 1: Main Pipe Prices (HP: high-performance polypropylene).

Pipe Type	Material	Diameter	Price\unit
Main Pipe	Steel	8 inch	250 ₪
		6 inch	140 ₪
		4 inch	95 ₪
		3 inch	75 ₪
		2 inch	47 ₪
	HP	110 mm	40 ₪
		90 mm	32 ₪
		75 mm	20 ₪
		63 mm	17 ₪
		50 mm	14 ₪

Table 2: House Connection Prices (HDPE: High-density polyethylene).

Pipe Type	Material	Diameter	Price\unit
House Connection	HDPE	63 mm	17 ₪
		32 mm	6 ₪
	steel	63 mm	32 ₪
		32 mm	15 ₪

Table 3: Meters Prices.

Meter Type	Price\unit
Arad	140 ₪
Baylan	130 ₪
Actars	150 ₪
Kent	120 ₪
Sattco	170 ₪
Itron	150 ₪
French TD8	160 ₪

Table 4: Valves Prices.

Valve type	Diameter	Price\unit
Air-valve	6 inch	800 ₪
	10 inch	1000 ₪
Gate	63 mm	20 ₪
	1 inch	80 ₪
	2 inch	150 ₪
	3 inch	300 ₪
	4 inch	400 ₪
	8 inch	700 ₪
	10 inch	1000 ₪

The cost of maintenance for each material component is calculated based on a mathematical equation as follows:

- Meters total maintenance cost =Meter Price + installation costs. ……………..(1)
- Valve total maintenance cost= Valve Price + installation costs. ……………..(2)
- House Connection pipes = pipe price (unit price * length) + ……………..(3)

Excavation quantity price (unite price * quantity) +

Sand price (unit price* quantity) +

Base course (unit price* quantity) +

Bitumen (unit price * quantity)

- Main pipes = pipe price (unit price * length) + ……………..(4)

Excavation quantity price (unite price * quantity) +

Sand price (unit price* quantity) +

Base course (unit price* quantity) +

Bitumen (unit price* quantity)

These prices are obtained from the companies concerned and can be adjusted depending on the situation.

Conclusion:

The aim of all the data collected to study this research, is to look for a solution for the problem and to prove the validity of hypotheses, whether it is spatial data or ordinary data, needs to be processed, as we indicated earlier that spatial data is saved as deaf data and converted from one formula to another, which has led to an imbalance in their validity. And when we represent it, it must be true without any problems. Frequently occurring data that has problems as a result of improper preservation and archiving practices.

Chapter Five

Data Mining and Spatial Data Mining

5.1 Data Mining

Finding patterns, correlations, and insights in massive data sets through a variety of statistical, machine learning, and database management techniques is known as data mining. It is a crucial part of database knowledge discovery (KDD) and data analysis [80].

Data mining's main goals are as follows: Finding relationships, trends, and patterns in data is known as pattern recognition, Prediction: Forecasting future occurrences by utilizing past data, Classification is the process of dividing up a dataset's items into predetermined groups or classes, Clustering: Assembling comparable items into groups without established classifications, and Association Rule Learning: Finding fascinating correlations between various variables in sizable databases [81].

Data mining employs a number of approaches, each of which is appropriate for a particular type of data and analysis objective:

- For Classification there are several algorithms such as Decision tree (DT), random forests (RF), support vector machines (SVM), and neural networks (NN) [82], which are used for image recognition, Spam detection, and credit scoring [83].
- For Regression, data mining provides many of algorithms like linear regression, polynomial regression, and ridge regression [68], that is beneficial for risk assessment, Stock price prediction, and resource allocation [84].

- Algorithms for Association Rule Learning are Apriori, Eclat, and FP-Growth (Frequent Pattern Growth) [85]; it's used for bioinformatics, Market basket analysis, and recommendation systems [86].
- Anomaly Detection to obtain results in this domain there's diverse algorithms are allowed like Isolation forest, one-class SVM, and local outlier factor (LOF) [87] which are useful for Fraud detection, quality control and network security [88].
- Dimensionality Reduction feature extraction, Data visualization, and noise reduction, are available services we can obtain it by applying one of these algorithms Principal component analysis (PCA), autoencoders, and t-SNE (t-Distributed Stochastic Neighbor Embedding) [89].
- Clustering: about this topic, data mining provide the following algorithms hierarchical clustering, K-means, and DBSCAN (Density-based spatial clustering of applications with noise) [90], which we can use for social network analysis and Customer segmentation [91].

Numerous industries have employed data mining for a variety of purposes like Finance, Healthcare, Telecommunications, Manufacturing and Retail, on the other hand, there are many advantages for data mining, but there are also some challenges such as Privacy and Security of the data, Interpretability and make the model understandable to non-experts, Data Quality which means that it is accurate, complete, and reliable; Ethical Issues and fair use and Scalability with any volume of data [89], [87], [80].

If we want to apply the data mining process there are steps we have to follow:

1- Data Collection: This step was already implemented and data was collected from the municipality and the institutions concerned at the beginning of the research for other objectives, such as data representation in ArcMap at chapter IV previously.

2- Data Preprocessing: This step was explained in detail in the previous chapter IV, and all the problems experienced by the data were explained, whether network data or historical data for maintenance and how the data was processed.

3- Modeling: As a water system engineer responsible for municipal water systems, you want to count the physical components that are prone to failure. This is how you can save your budget, time, and effort. The process of classifying ingredients into a range of safe and fault-prone ingredients is defined as a classification problem.

The process of classification has two stages: learning and prediction. During the learning stage, the model is constructed using the supplied training data. The model is used in the prediction step to forecast the response to the supplied data. One of the simplest and most widely used classification algorithms for comprehending and interpreting data is the decision tree. It can be applied to problems involving regression as well as classification [92].

A decision tree is a type of tree structure that resembles a flowchart, with each leaf node representing the result, the branch representing a decision rule, and the internal node representing a feature or attribute [92].

One can refer to the root node as the highest node in a decision tree. It gains the ability to divide data according to attribute values. Recursive partitioning is the method by which it divides the tree recursively. This structure, which resembles a flowchart, aids in decision-making. It is a visual representation which resembles human thought processes, much like a flowchart diagram. For this reason, decision trees are easy to understand and evaluate. [93].

Machine Learning algorithms of the white box variety include decision trees. It shares internal reasoning for making decisions, which is not present in black box algorithms like neural networks. When compared to the neural network algorithm, its training time is faster [94].

The quantity of records and attributes in the provided data determines the decision tree's time complexity. The decision tree approach is non-parametric and distribution-free; it does not depend on assumptions about possibility distributions. High-dimensional data can be handled by decision trees accurately [92].

Any decision tree algorithm's has this fundamental concept [93]:

1- To split the records, use feature selection method to determine which attribute is best.

2- Classify the dataset into more manageable subsets by using that attribute as a decision node.

3- For each child, repeat this process recursively until one of the conditions matches to begin building the tree:

- Every tuple is associated with a single attribute value.
- None of the attributes are left.
- No further occurrences exist.

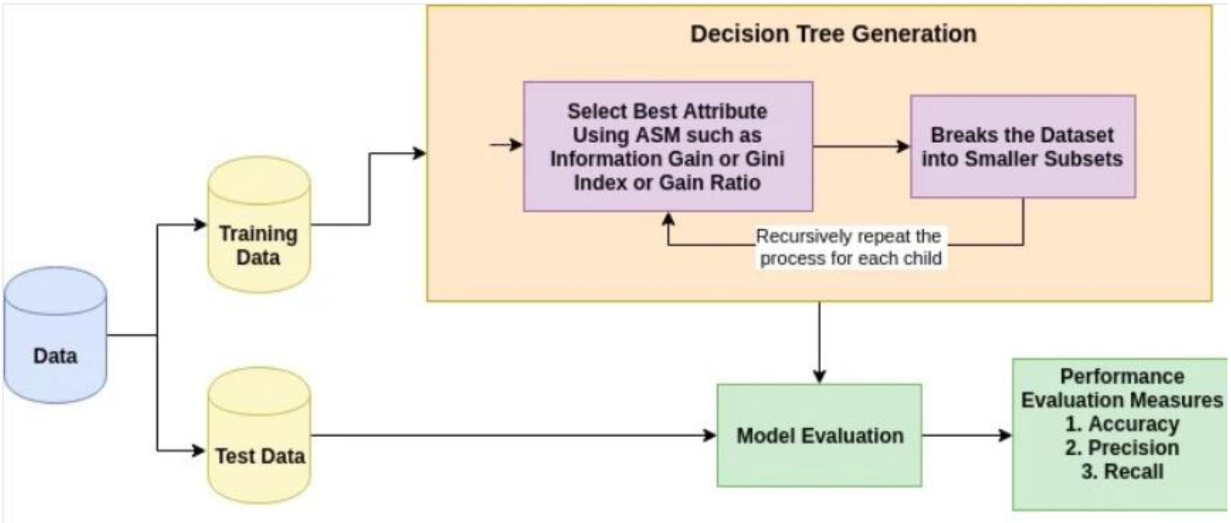


Figure 27: How the Decision Tree Algorithm Work [95].

The feature selection strategy is a heuristic used to select the splitting criterion that best divides data. Because it helps us in identifying breakpoints for tuples on a particular node, it is also known as splitting rules. It uses an explanation of the provided dataset to assign a rank to each feature (or attribute). As a splitting attribute, the highest scoring attribute will be chosen. Whenever a feature has a continuous value, split points for its branches must also be defined. The Gini Index, Gain Ratio, and Information Gain are the most widely used selection metrics [96], [95].

Entropy is a concept developed by Claude Shannon that quantifies the degree of impurity in an input set. It is the randomness or impurity in a system and is used in mathematics and physics. It alludes to the impurity in a collection of examples in information theory. The reduction in entropy is known as information gain. Information gain calculates the difference between the average entropy following the split of the dataset relied on specified attribute values and the entropy prior to the split. Information gain is used in the ID3 (Iterative Dichotomiser) decision tree algorithm [97].

The feature with multiple outcomes is skewed in the information gained. It shows that it favors the feature with a high number of different values. Such as, take into consideration an attribute, (Meter_ID) that has zero info (D) due to pure partition and has a unique identifier. This produces pointless partitioning while optimizing the information gain [97].

The following formula represents the mathematical formula of the ratio:

$$SplitInfo_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} * \log_2\left(\frac{|D_j|}{|D|}\right) \qquad \ldots\ldots\ldots\ldots\ldots..(5)$$

Where:

- $\frac{|D_j|}{|D|}$ Acts as the partition's weight in j_i .
- v Is the number of discrete values in attribute A.

The splitting attribute is determined by selecting the attribute with the highest gain ratio [95].

4- Evaluation: evaluating the models with respect to metrics such as F1 score, recall, accuracy, and precision.

5- Deployment: Implementing models in real-world applications as we will see in next chapter.

6- Monitoring and Maintenance: monitoring the model's performance and updating it as necessary.

Conclusion:

Large datasets can be mined for insightful information using a potent technique called data mining. It includes a range of methods and tools that can greatly improve how decisions are made in a variety of sectors. Notwithstanding the difficulties, continuous technological developments keep expanding the possibilities for data mining.

In order to create a prediction system that can estimate the failure ratio of any component relied on historical data for maintenance and display the results of this system on our network through the desktop application that will be covered in detail in the next chapter, the decision tree algorithm has been chosen to be applied to historical data for maintenance of both counters and house connections.

5.2: Spatial Data Mining:

Spatial data mining is finding intriguing, previously undiscovered, but possibly helpful patterns in spatial datasets; spatial data mining considers the attributes, geometry, and spatial relationships of data objects, in contrast to traditional data mining [59].

Fundamental Ideas in Spatial Data Mining [60]:

1- Types of Spatial Data: polygons, lines, points, and raster data are among them. Mining spatial patterns requires an understanding of the various forms of spatial data.

2- Topological relationships (adjacency, containment, etc.) as well as distance and directional relationships are examples of spatial relationships. These relationships are frequently used by spatial data mining techniques to identify patterns.

In real life, spatial data mining is used in a variety of different fields as follow [61]:

- Transportation: using geographic data on vehicle movements and road networks to optimize routes, control traffic flows, and enhance public transportation systems.
- Retail and marketing: To improve sales tactics and advertising campaigns, analysis of consumer demographics, purchase trends, and store locations is conducted.
- Public Health: The field of public health involves the identification of disease patterns, the location of healthcare facilities, and the comprehension of the correlation between environmental factors and health outcomes.
- Environmental Monitoring: Tracking changes in land cover, managing natural resources, and anticipating environmental hazards like floods and forest fires are all made possible with the help of spatial data mining.

- Urban planning: Planners can decide on zoning, public services, and transportation networks by using spatial data on population density, land use, and infrastructure.

Spatial data mining faces many challenges, including the complexity of such data. The complexity of multiple relationships and dimensions that characterize spatial data makes analysis difficult, and data quality is a difficult task. Because spatial data can be noisy, lacking, or inconsistent, the quality of the data is essential to achieving trustworthy mining results. They are also scalable and sophisticated handling large quantities of spatial data need strong algorithms and high computational resources, and finally, its interpretability for practical applications, it is essential that the results of spatial data extraction have to be intelligible to non-experts [62].

Spatial data mining can be accomplished with a variety of tools and software packages; GIS software, for example, offers powerful platforms for spatial data analysis and visualization through software like QGIS, an open-source program, and ArcGIS; data mining tools, like Weka (Waikato Environment for Knowledge Analysis), RapidMiner, and KNIME (Konstanz Information Miner), can be extended with spatial data mining capabilities. With libraries like Shapely, GeoPandas, and Scikit-learn, Python provides robust spatial data mining capabilities [63] [64] [65].

The Ordinary Least Squares (OLS) method is a fundamental statistical technique used for estimating the parameters in a linear regression model. It was one of several spatial mining analyses offered by the ArcMap program that we selected [66]. By minimizing the sum of the squared differences between the observed values and the values predicted by the linear model, OLS seeks to identify the best-fitting line through the data points [67].

A linear regression model's fundamental form is [68]:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \quad \ldots\ldots\ldots\ldots\ldots(6)$$

Where:

- Y is the variable that is dependent.
- $\beta_0, \beta_1, \ldots, \beta_k$ are to be estimated parameters, or coefficients.
- $x_1, x_2, \ldots, x_k$ the variables that are independent.
- ϵ represents the discrepancy between the observed and predicted values and is the error term.

The difference between the actual and predicted values, or the sum of the squared residuals, is what OLS aims to minimize. This can be expressed mathematically as [69]:

$$\min_{\beta 0, \beta 1, \ldots, \beta k} \sum_{i=1}^{n} (y_i - y_i^{\wedge})2 \quad \ldots\ldots\ldots\ldots\ldots(7)$$

Where:

$y_i^{\wedge} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$ Represent the predicted values.

As we know each Statistical analysis has an assumption, OLS's underlying assumptions is as follows [70]:

- Linearity: The independent and dependent variables have a linear relationship.
- Independence: The observations standalone from one another.
- Homoscedasticity: At each level of the independent variables, the error terms' variance remains constant.
- Normality: means the distribution of the error terms is normal, which is significant for inference.
- Lack of multicollinearity: There is little to no correlation between the independent variables.

A mainstay of regression analysis, the Ordinary Least Squares method is popular because it is easy to use and efficient at estimating linear relationships. Its foundational principles and application make it a powerful tool in statistical modeling and data analysis, even though there are some assumptions that need to be verified [70].

OLS Regression Steps:

1- Create the Model [70]: Indicate the structure of the dependent and independent variables in a linear regression model

2- Determine Parameters[71]:

Matrix algebra can be used to obtain the OLS estimates of the parameters. Given a vector of observed values (Y), a matrix of independent variables (X), and a vector of coefficients (β), the OLS estimator can be expressed as follows:

$$\boldsymbol{\beta}^{\wedge} = (\mathbf{X}^{\boldsymbol{T}}\mathbf{X})^{-1}\mathbf{X}^{\boldsymbol{T}}\boldsymbol{Y} \qquad \dots\dots\dots\dots\dots..(8)$$

3- Examine the Model[72]:

Using the following metrics, the model can be examined:

- Goodness of Fit: Evaluate the model's ability to fit the data by calculating R-squared and Adjusted R-squared.
- Statistical Significance: To ascertain the significance of the predictors, run hypothesis tests (e.g., t-tests for individual coefficients, F-test for overall model significance).
- Residual Analysis: Examine the residuals to make sure the OLS assumptions—such as constant variance and the absence of patterns in the residuals—are met.

4- Interpretation of the Model[71]:

To comprehend the way that independent and dependent variables are related, interpret the estimated coefficients. For instance, the change in the dependent variable is represented by the coefficient β_j.

Holding all other variables constant, y represents a one-unit change in the independent variable x_j.

This statistical analysis was selected for application to Nablus city data after adding historical maintenance data for both meters and house connection pipes. The workflow was followed for results, as shown in the figure (26).

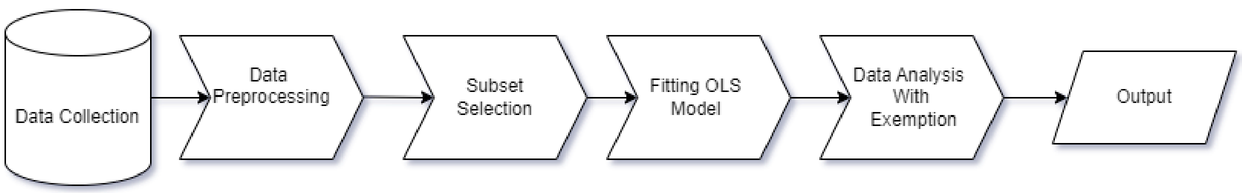


Figure 28: The Workflow Structure to Obtain Results by OLS.

1- Data Collection:

The data collection process began with the first steps of this study and obtained spatial data of the industrial area and spatial data of the city of Nablus completely individually. These data consist of several layers so that each layer expresses a physical component of the network, and the network contains a so-called attribute table that shows the characteristics of each part of the layer, such as the date of its composition, size, diameter, etc., which will be relied upon as factors in the analysis. These data are represented as points or lines according to the type of each component and layer.

As mentioned in the previous chapter, the data needed spatially processed processes, and the processing processes were detailed to access the correct geometric network without errors using the data of the industrial area. The data used in the analysis is that of the city of Nablus, but it also needed processing processes. Since OLS analysis only relies on numerical values, other factors such as indicators and texts have been excluded. Let's explain in more detail. We will talk about each analysis separately.

A-Meter OLS Analysis:

2A- Data Pre-processing:

Nablus City Network data contains more than 59 thousand water meters preserved and archived, each meter saved each (shape, meter number, pressure zone, bi number, block number, parcel number, installation date, Consumption type, x coordinates, y coordinates, water source, Meter diameter, meter type, meter location, meter reading for January 2023, February in 2023 and march 2023), some of these factors were unfortunately empty and had to be excluded, as shown in figures (27).

The meter maintenance data are described in detail in Chapter IV (previous), and the methods of processing them are explained after they are ready to be joined to the meter layer to become a single layer that can be analyzed using the instrument of join between the meter layer and the maintenance spreadsheet.

meter

OBJECTID_1	Shape *	Customer_No	Pressure_Zone	Bl_Num	BLOCK_NM	PARCEL_NM	Construction_Date	نوع الأستهلاك
1	Point	8059	E0.2	2210	Balata_12	14	5/21/1970	منزلي داخل الحدود
2	Point	34856	E0.2	2178	Balata_12	14	10/31/2005	منزلي
3	Point	29350	E0.2	1667	Balata_12	28	11/14/2000	موظف صحه داخل الحدود
4	Point	21020	E0.2	<Null>	Balata_12	59	4/14/1994	منزلي داخل الحدود
5	Point	26072	E0.2	2097	Balata_12	61	11/3/1998	منزلي داخل الحدود
6	Point	27742	E0.2	2599	Balata_12	1/A	10/24/1999	تجارى
7	Point	17758	E0.2	2534	Balata_12	69	8/13/1988	تجارى
8	Point	24758	E0.2	3683		13	10/4/1997	تجارى
9	Point	24438	E0.2	3701	Balata_11	15	7/8/1997	تجارى
10	Point	15282	E0.2	3312	Balata_11	65	11/7/1983	صناعي
11	Point	4675	E0.2	3462	Balata_11	57	4/1/1962	منزلي داخل الحدود
12	Point	32304	E0.2	3442	Balata_11	57	11/9/2003	موظف مياه
13	Point	32305	E0.2	3427	Balata_11	57	11/9/2003	منزلي داخل الحدود
14	Point	16720	E0.2	<Null>	Balata_11	63	10/11/1986	صناعي غير غذائي
15	Point	25954	E0.2	3503	Balata_11	73	9/7/1998	تجارى
16	Point	36346	E0.2	3490	Balata_11	73	12/2/2006	تجارى
17	Point	14757	E0.2	3619	Balata_11	A/99+100+101+102+103+104	9/13/1982	تجارى
18	Point	6306	E0.2	3712	Balata_9	23	9/27/1965	صناعي غير غذائي
19	Point	8387	E0.2	3931	Balata_10	31	8/10/1971	صناعي غير غذائي
20	Point	15163	E0.2	3847	Balata_11	97	8/8/1983	صناعي غير غذائي
21	Point	19397	E0.2	3830	Balata_11	97	9/2/1991	منزلي داخل الحدود
22	Point	16478	E0.2	<Null>	Balata_11	98	6/28/1986	تجارى
23	Point	6733	E0.2	3927	Balata_11	72	8/17/1966	تجارى
24	Point	23196	E0.2	3991	Balata_11	44/4	6/11/1996	تجارى
25	Point	20402	E0.2	3998	Balata_11	a/26/45+27/45	4/28/1993	تجارى
26	Point	20885	E0.2	3976	Balata_11	a/26/45+27/45	11/20/1993	منزلي داخل الحدود
27	Point	33501	E0.2	3969	<Null>	<Null>	11/27/2004	منزلي داخل الحدود
28	Point	33502	E0.2	3962	<Null>	<Null>	11/27/2004	تجارى
29	Point	35819	E0.2	<Null>	Balata_11	126+127+128	7/10/2006	تجارى
30	Point	15271	E0.2	<Null>	Balata_11	125	10/27/1983	تجارى
31	Point	16503	E0.2	4130	Balata_11	125	7/7/1986	تجارى

meter

POINT_X	POINT_Y	Water_Re	نوع العداد	قطر العداد	موقع العداد	Reading_12023	Reading_22023	R_32023
178370.646118	179127.673096	Ein Dafna	ITRON	<Null>	<Null>	8	<Null>	<Null>
178369.055908	179123.556885	Ein Dafna	مسبق الدفع_ساتكو	0.5_ inch	<Null>	<Null>	<Null>	<Null>
178280.115295	179023.013916	Ein Dafna	TD8 فرنسي	0.5_ inch	<Null>	18	<Null>	<Null>
178242.164307	179079.602722	Ein Dafna	Kent	0.5_ inch	<Null>	15	<Null>	<Null>
178260.29071	179110.801331	Ein Dafna	Kent	0.5_ inch	<Null>	3	<Null>	<Null>
178260.658508	179197.931702	Ein Dafna	Kent	0.5_ inch	<Null>	8	<Null>	<Null>
178314.30188	179183.656128	Ein Dafna	مسبق الدفع_ساتكو	0.5_ inch	<Null>	<Null>	<Null>	<Null>
178471.761108	179394.529907	Ein Dafna	Kent	0.5_ inch	<Null>	3	<Null>	<Null>
178497.32312	179397.605103	Ein Dafna	اكتارس	0.5_ inch	<Null>	2	<Null>	<Null>
178291.600098	179324.724304	Ein Dafna	مسبق الدفع_ساتكو	0.5_ inch	<Null>	<Null>	<Null>	<Null>
178243.01532	179349.127502	Ein Dafna	<Null>	<Null>	<Null>	15	<Null>	<Null>
178242.759705	179346.059875	Ein Dafna	Kent	0.5_ inch	<Null>	2	<Null>	<Null>
178242.504089	179342.99231	Ein Dafna	TD8 فرنسي	0.5_ inch	<Null>	21	<Null>	<Null>
178259.962891	179382.739319	Ein Dafna	اراد اسرائيلي	1 inch	<Null>	19	<Null>	<Null>
178160.944885	179355.658325	Ein Dafna	Kent	0.5_ inch	<Null>	8	<Null>	<Null>
178160.433472	179352.846313	Ein Dafna	بايلان تركي	0.5_ inch	<Null>	0	<Null>	<Null>
178121.331299	179383.295715	Ein Dafna	Kent	0.5_ inch	1	8	<Null>	<Null>
178020.869507	179399.743103	Ein Dafna	<Null>	<Null>	<Null>	0	<Null>	<Null>
178060.432495	179445.646301	Ein Dafna	بايلان تركي	1/2	1	37	<Null>	<Null>
178095.770081	179427.862915	Ein Dafna	Kent	0.5_ inch	<Null>	18	<Null>	<Null>
178095.258911	179426.329102	Ein Dafna	Kent	0.5_ inch	<Null>	22	<Null>	<Null>
178112.104126	179410.1203	Ein Dafna	Kent	0.5_ inch	<Null>	4	<Null>	<Null>
178153.799072	179444.990295	Ein Dafna	Kent	1/2	<Null>	9	<Null>	<Null>
178380.862305	179465.008301	Ein Dafna	Kent	0.5_ inch	<Null>	26	<Null>	<Null>
178403.660095	179467.728271	Ein Dafna	Kent	0.5_ inch	<Null>	17	<Null>	<Null>
178425.396118	179459.266907	Ein Dafna	Kent	0.5_ inch	<Null>	0	<Null>	<Null>
178424.485718	179456.535706	Ein Dafna	TD8 فرنسي	0.5_ inch	<Null>	0	<Null>	<Null>
178423.802917	179453.804321	Ein Dafna	TD8 فرنسي	0.5_ inch	<Null>	0	<Null>	<Null>
178479.24292	179503.268921	Ein Dafna	Kent	0.5_ inch	<Null>	0	<Null>	<Null>
178457.004272	179512.628906	Ein Dafna	Kent	0.5_ inch	<Null>	2	<Null>	<Null>
178440.50708	179516.840088	Ein Dafna	Kent	0.5_ inch	<Null>	2	<Null>	<Null>

Figure 29: Meters Original Data

One factor to consider is the rise of the ground level on which the meter stands because it affects the value of the water pressure on the meter, so we added contour data that shows the height of each area from the sea level in the form of lines, as shown in the figure (28).

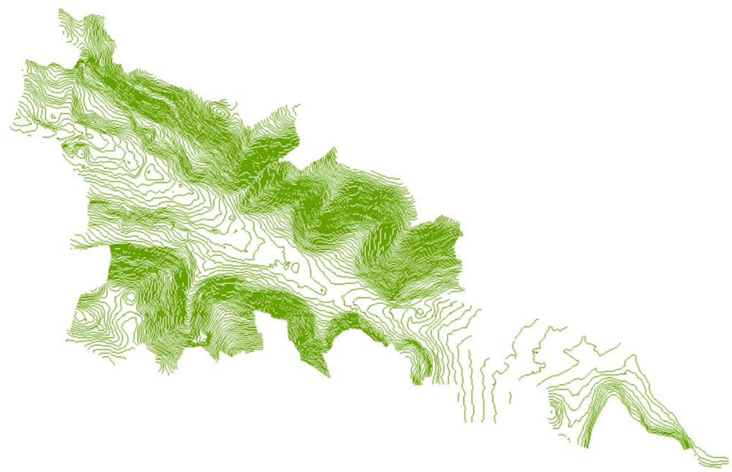


Figure 30: The contour line layer.

The meter layer reflecting the contour line layer had to be converted into points for use in the analysis. The first step was to convert it to Raster using ArcMap tools, by using (to conversation ->feature to raster) tool, as shown in figure (29).

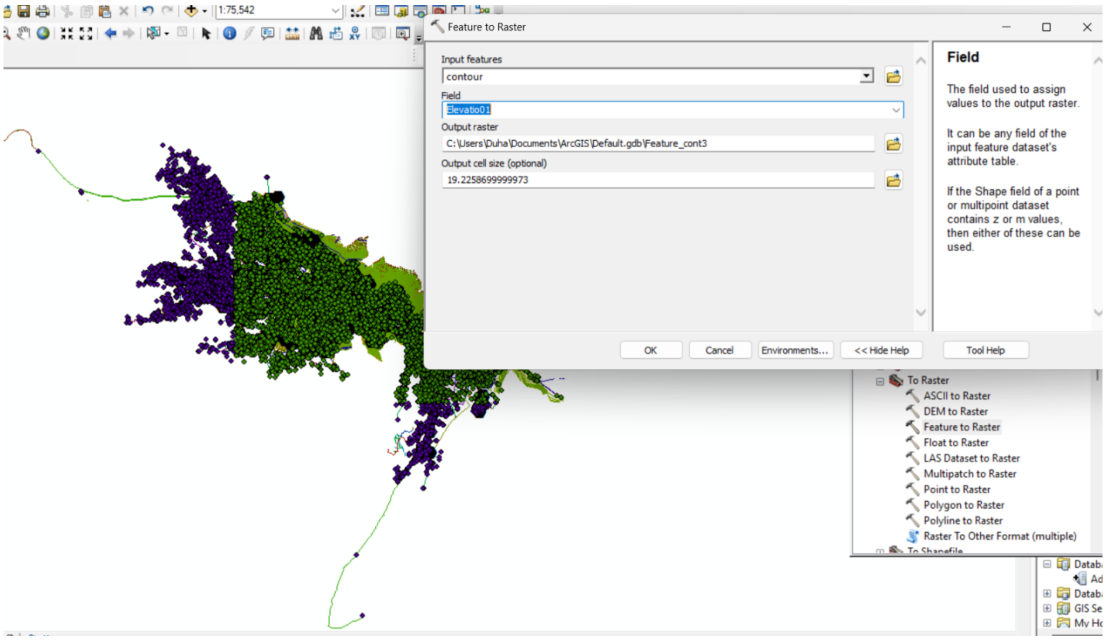


Figure 31: Convert Contour Layer into Raster.

The second step required converting the result of the first step (RASTER) to points associated with the layer of meters maintained by one of the ArcMap tools, it done by using (spatial analysis tools -> extraction -> extract values to point) tool . As shown in figure (30). Thus, the data available in the contouring layer is ready for use in the analysis process.

After converting to points, it is now possible to calculate the height of the ground for each meter. This was done using (Analysis tool -> overly -> spatial join) tool, as shown in figure (31) to result in a new column in the meter layer that expresses the height of the area on which the meter is located. Thus, the meter layer has full meter data in the city of Nablus and meter data on the maintenance in the city and the height of the ground on which each meter is located.

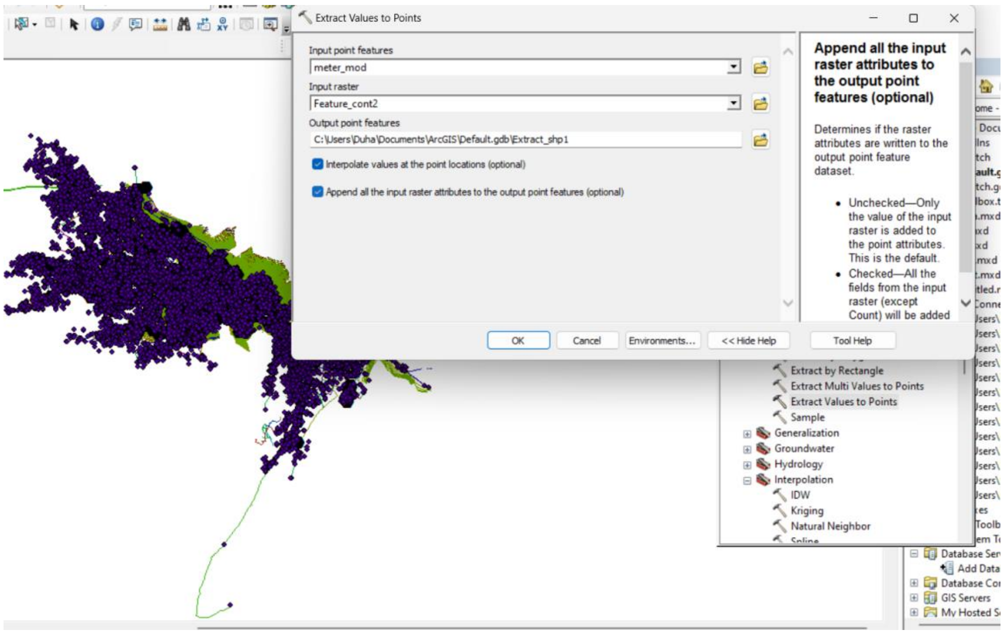


Figure 32: Extract Values to Point.

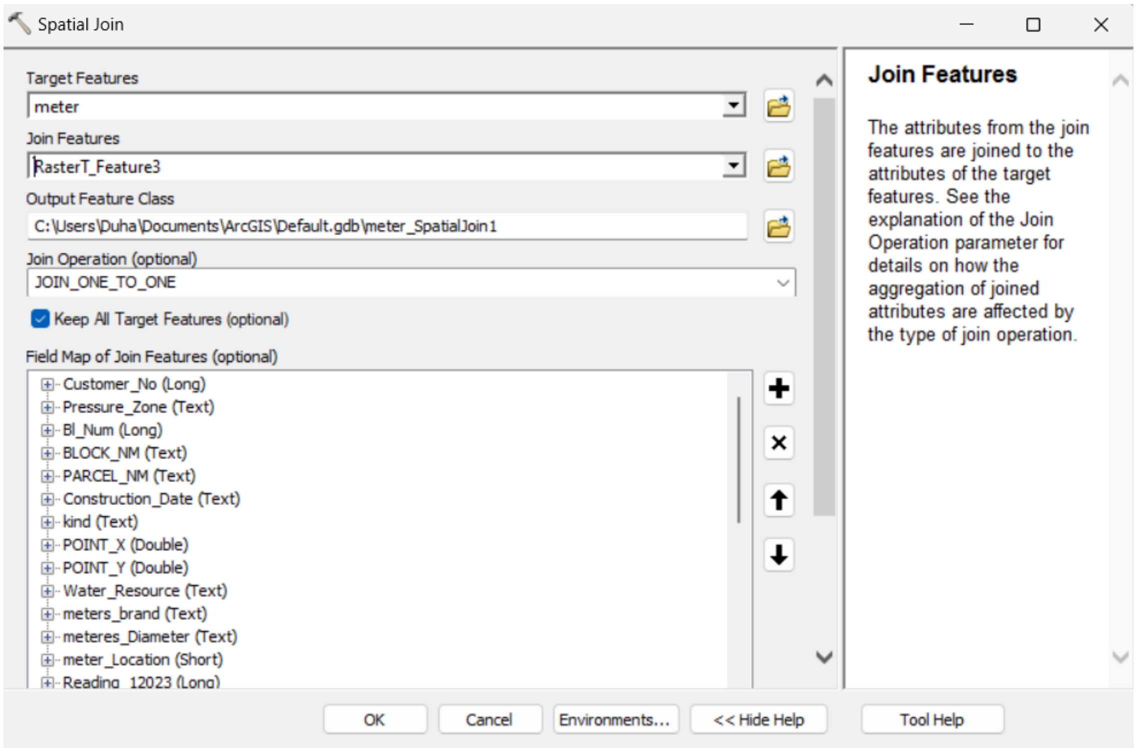


Figure 33: Spatial Join between the meters and the height points.

I also mentioned the analysis of the statistical OLS dealing only with numerical values, so it was necessary to select the appropriate transactions, process them and clean them. Initially, the table must be extracted and converted into an excel file to modify the values and clean the data.

As a first step: Selection of appropriate parameters, all non-numerical values such as pointer data (Object_ID* 1), dates (install date 01\01\1990, 12:00:00 AM) and blank columns (Reading 22023 <Null>), (Reading 32023 <Null>) should be deleted, have to remain just the numerical values. As shown in the figure (32).

H1

	A	B	C	D	E	F	G
1	OBJECTID	Customer_	POINT_X	POINT_Y	Reading_1	grid_code	maint
2	1	34856	178359.9	178359.9	35	470	1
3	2	17758	178307	178307	97	470	1
4	3	15282	178291.6	178291.6	36	465	1
5	4	36346	178160.4	178160.4	86	465	1
6	5	6306	178020.9	178020.9	94	470	1
7	6	20885	178425.4	178425.4	52	465	1
8	7	33501	178424.5	178424.5	99	465	1
9	8	33502	178423.8	178423.8	50	465	1
10	9	35819	178470.7	178470.7	85	465	1
11	10	20767	178357.5	178357.5	69	465	1
12	11	35031	178428.5	178428.5	45	465	1
13	12	32929	178373.3	178373.3	98	470	1
14	13	13913	178465	178465	34	465	1
15	14	32737	178755.6	178755.6	60	465	1
16	15	13188	178747.8	178747.8	16	465	1
17	16	13023	178771.9	178771.9	24	465	1
18	17	32443	178769.7	178769.7	69	465	1
19	18	19245	178880	178880	91	465	1
20	19	27573	178814.2	178814.2	49	465	1
21	20	20922	178811.2	178811.2	72	465	1
22	21	30394	178832.2	178832.2	23	465	1
23	22	24961	178773.2	178773.2	45	465	1
24	23	6606	177281.9	177281.9	33	485	1
25	24	5422	177263.3	177263.3	80	485	1
26	25	23108	177292.4	177292.4	34	485	1
27	26	14304	177391.4	177391.4	73	480	1
28	27	14443	178102.1	178102.1	64	475	1
29	28	6136	177596.7	177596.7	98	475	1
30	29	7504	177767.6	177767.6	92	475	1
31	30	29411	177624.5	177624.5	44	495	1

Figure 34: Meter Numerical parameters.

Second step: is the cleaning stage, the cleaning process involves the elimination of null values. The missing records of (Point_X) and (Point_Y) were calculated using geometric calculation tool to obtain the exact coordinates. after this step we can generate a new parameter, the (Near_Dis) column which is a parameter displays the nearest horizontal distance between each meter and the Earth's elevation point from sea level, it has been calculated by using (Analysis tool -> proximity -> Near) tool, as shown in the figure (33). Thus, all parameters become complete, ready for the next stage.

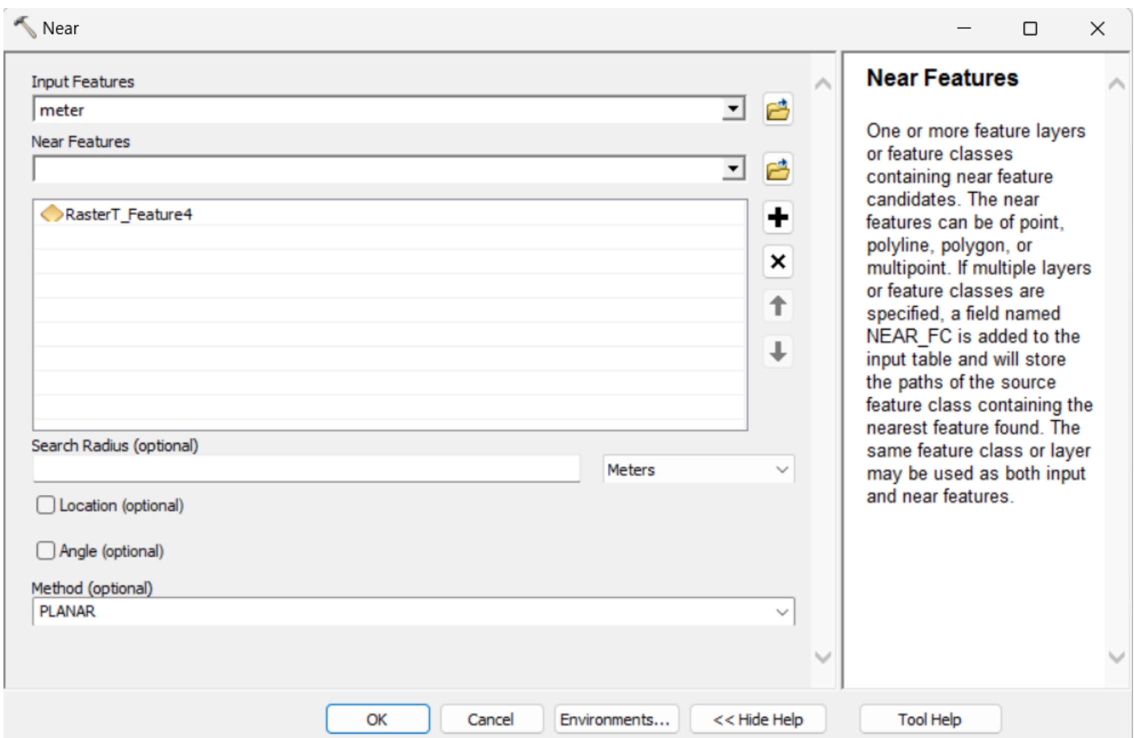


Figure 35: Near tool to calculate distance between meters and height point.

Third Step: Removing Outlier Data, Sometimes we find in the data an extraterritorial value that is considered out of range, either for a special case or incorrectly in the input of values, This can be detected by checking the statistics of the data and ranking values from the largest to the smallest and vice versa, and during the test of (Reading_12023) the statistics was as shown in the figure (34).

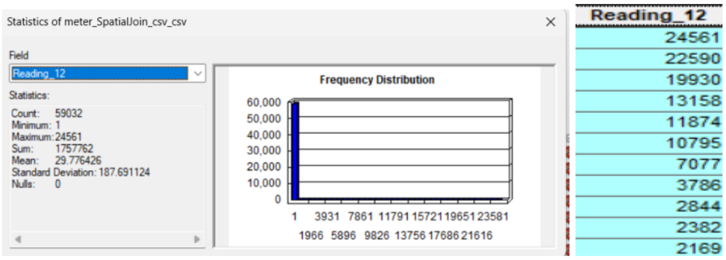


Figure 36: Reading_12023 statistics.

The figure shows that there are a number of values outside the range (24561, 22590, 19930, 13158, 11874, and 10795) these values have been replaced with logical values that fall within the coefficient's range because it appears that they do not make sense.

Reclassify Step: To determine the impact of each factor on the outcome of the analysis, so we created a decimal order to give us the ratio of each value per record of ten.by the following formula:

$$X = \left(\frac{(X_i - \text{MIN})}{(\text{MAX} - \text{MIN})} \right) * 10 \quad \ldots\ldots\ldots\ldots\ldots..(9)$$

Where:

X , is the value ratio.

X_i , Is the original value.

MAX ,the maximum value in the parameter column.

MIN ,the minimum value in the parameter column.

This value has been calculated for all the numerical factors that will be used in the analysis, each value has been represented as (F_Dist).

3A- Subset selection:

As shown in the figure (35), we use the (meter_SpatialJoin) layer to analyze the data, the (Object_ID) parameter as a unique ID field because it represents a unique number for each record, (maint) parameter as a dependent variable since it represents the target, it shows that the meter has been maintained or not, and each of (F-y, F-R, F-DIST,F-GRID,F, CUSTOMER_N, POINT_Y, READING_12, NEAR_DIST and GRID-CODE) as an explanatory variables.

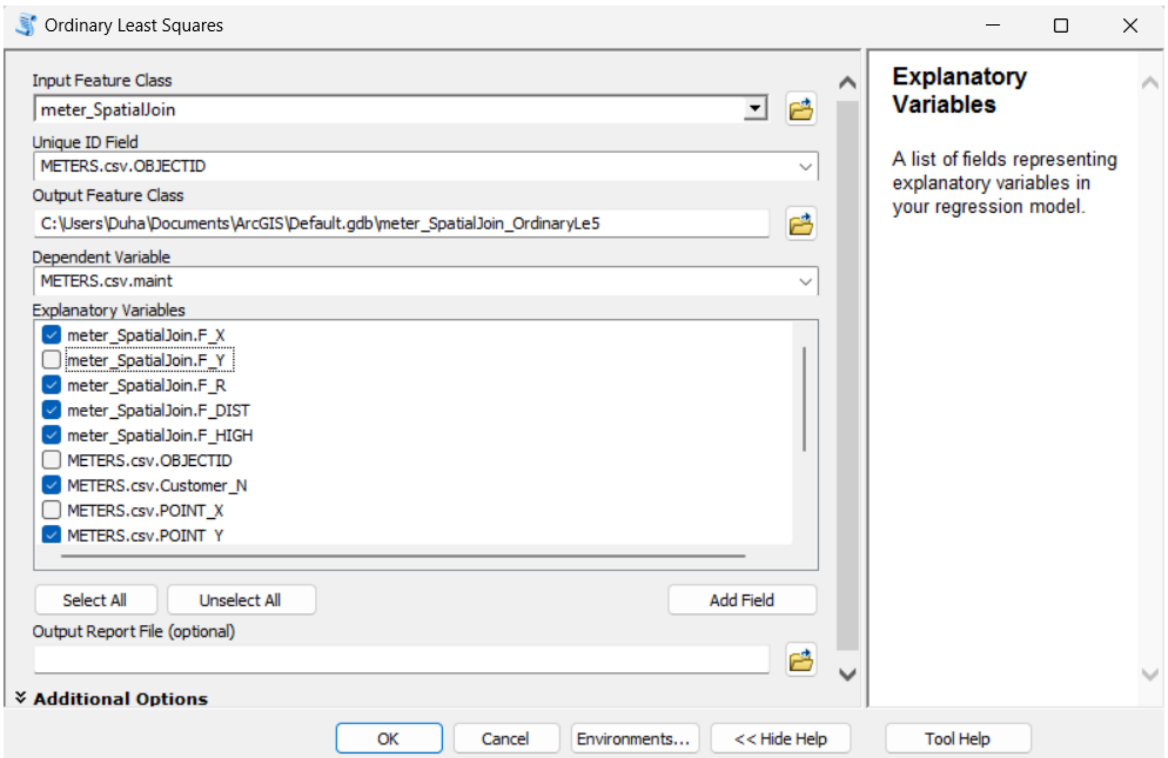


Figure 37: The OLS analysis for the Meters Pipes.

β0: This is the baseline value of the dependent variable (maint) when all explanatory variables are zero.

Coefficients (β1, β2,..., βk): These values show how the dependent variable changes when each explanatory variable changes by one unit while keeping the other variables unchanged.

3A – Fitting OLS model:

The global linear regression of Ordinary Least Squares (OLS) is used to model the relationships of the variable to a range of interpretive variables and to produce predictions. We used this analysis to study and predict the occurrence of warranted maintenance and the impact of each factor on the outcome of the analysis.

- Model specification: to specify the form of the model

$$Y_i^{\wedge} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} \quad \ldots\ldots\ldots\ldots\ldots..(10)$$

Where:

$Y_i^{\wedge}$ is (maint) parameter (The dependent parameter)

$\beta_0, \beta_1, \ldots, \beta_k$ are to be estimated parameters, or coefficients.

$x_1, x_2, \ldots, x_k$ the variables that are explanatory parameter.

ϵ is the error term (residual).

- Estimation of Parameters:

Estimating the values of the parameters ($\beta 0, \beta 1, \ldots, \beta n$), by using this equation, to make the predicted values ($Y^{\wedge}$) are as close as feasible to the dependent variable's observed values(Y).

- Ordinary Least Squares (OLS) Analysis:

The fitting procedure for OLS regression entails reducing the total squared discrepancies between the observed and predicted values. In terms of mathematics, this entails identifying the coefficients that reduce the subsequent sum of squares:

$$\text{Sum of Squared Residuals (SSR)} = \min_{\beta 0, \beta 1, \ldots, \beta k} \sum_{i=1}^{n} (y_i - y_i^{\wedge})2 \quad \ldots\ldots\ldots..(11)$$

Where

y_i Is the observed value and $y_i^{\wedge}$ is the predicted value.

- Calculating the Coefficients:

The coefficients (β) are calculated using the following formula in matrix notation:

$$\beta^{\wedge} = (X^T X)^{-1} X^T Y \quad \ldots\ldots\ldots (12)$$

Where:

X is the matrix of independent variables (explanatory variables).
Y is the vector of the dependent variable.
X^T is the transpose of X.
$(X^TX)^{-1}$ is the inverse of X^TX.

- Evaluating the Fit:
  Assessing the model's ability to accurately describe the data is crucial after it has been fitted. This entails examining:

  1- R-squared (R^2): shows the percentage of the dependent variable variance that can be predicted based on the independent variables.

  2- Adjusted R-squared: Adjusts the (R^2) value for the number of predictors in the model.

  3- F-statistic: evaluates the model's overall significance.

  4- P-values for coefficients: evaluates each independent variable's significance.

  5- Residuals analysis: Analyzes the variations between observed and predicted values to look for trends that could indicate an inadequate model.

4A –OLS model Results:

The results report was attached in the appendix, and let’s looks at the results as follows:

- Number of Observations: 59,032

- The Key Metrics and its Interpretation:

1- Akaike's Information Criterion (AICc): -21,839.10(A lower value indicates a better fit of the model).

2- Multiple R-Squared: 0.0373.

3- Adjusted R-Squared: 0.0372, (2, 3) (These values indicate that approximately 3.73% of the variability in the dependent variable (MAINT) is explained by the model. This is quite low, suggesting that other factors not included in the model may be important.)

4- Joint F-Statistic: 254.14 (p-value: 0.0000)

5- Joint Wald Statistic: 376.08 (p-value: 0.0000) (4, 5) (Both are highly significant, indicating that the overall model is statistically significant.)

6- Koenker (BP) Statistic: 2,644.46 (p-value: 0.0000) (Significant, indicating issues with non-stationarity or heteroskedasticity. Thus, the robust probabilities should be considered for inference.)

7- Jarque-Bera Statistic: 859,729.99 (p-value: 0.0000) (Significant, suggesting that the residuals are not normally distributed, indicating possible model bias.)

1- Coefficient Estimates and Significance

Each coefficient represents the relationship between an explanatory variable and the dependent variable as shown in the figure (36).

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	-0.544304	0.254543	-2.138359	0.032477*	0.248116	-2.193747	0.028241*	——
METER_SPATIA	-0.007448	0.002719	-2.739580	0.006154*	0.002599	-2.865801	0.004167*	15.179269
METER_SPATIA	-3.001002	0.066098	-45.401997	0.000000*	0.185434	-16.183623	0.000000*	35.650676
METER_SPATIA	0.012984	0.004499	2.885870	0.003912*	0.004585	2.831874	0.004634*	13.469244
METER_SPATIA	0.012482	0.003147	3.966010	0.000082*	0.003070	4.066217	0.000055*	50.299699
METERS.CSV.C	-0.000000	0.000000	-6.627627	0.000000*	0.000000	-1.100604	0.271067	1.006237
METERS.CSV.P	0.000004	0.000002	2.786621	0.005330*	0.000001	2.857272	0.004280*	15.358289
METERS.CSV.R	0.001212	0.000026	46.034332	0.000000*	0.000074	16.473815	0.000000*	35.651411
METERS.CSV.N	-0.000023	0.000005	-4.401098	0.000014*	0.000005	-4.504351	0.000009*	13.506896
METERS.CSV.G	-0.000322	0.000066	-4.873679	0.000002*	0.000064	-5.019392	0.000001*	49.914479

Figure 38: OLS meters analyst results.

And that mean:

1- F_Y: Negative effect, significant

2- F_R: Large negative effect, highly significant

3- F_DIST: Positive effect, significant

4- F_HIGH: Positive effect, significant

5- CUSTOMER_N: Negative effect, significant

6- POINT_Y: Positive effect, significant

7- READING_12023: Positive effect, highly significant

8- NEAR_DIST: Negative effect, significant

9- GRID_CODE: Negative effect, significant

2- Diagnostic Plots:

1- Residuals Histogram: The result of the Jarque-Bera statistic is confirmed by the residuals' imperfect conformity to a normal distribution, as shown in the figure (37)

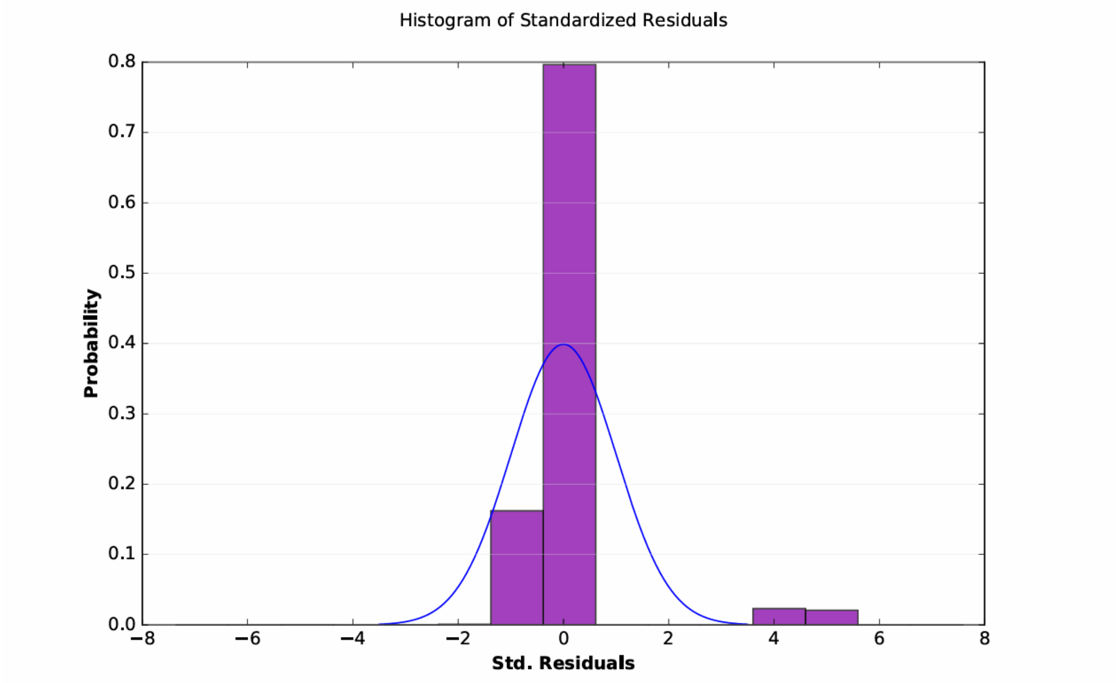


Figure 39: Histogram of Standardized Residuals for meters analyst.

2- Residual vs. Predicted Plot: Ideally, this plot should show no pattern, indicating that the model errors are randomly distributed. Organized patterns imply incorrect model specifications as shown in the figure (38).

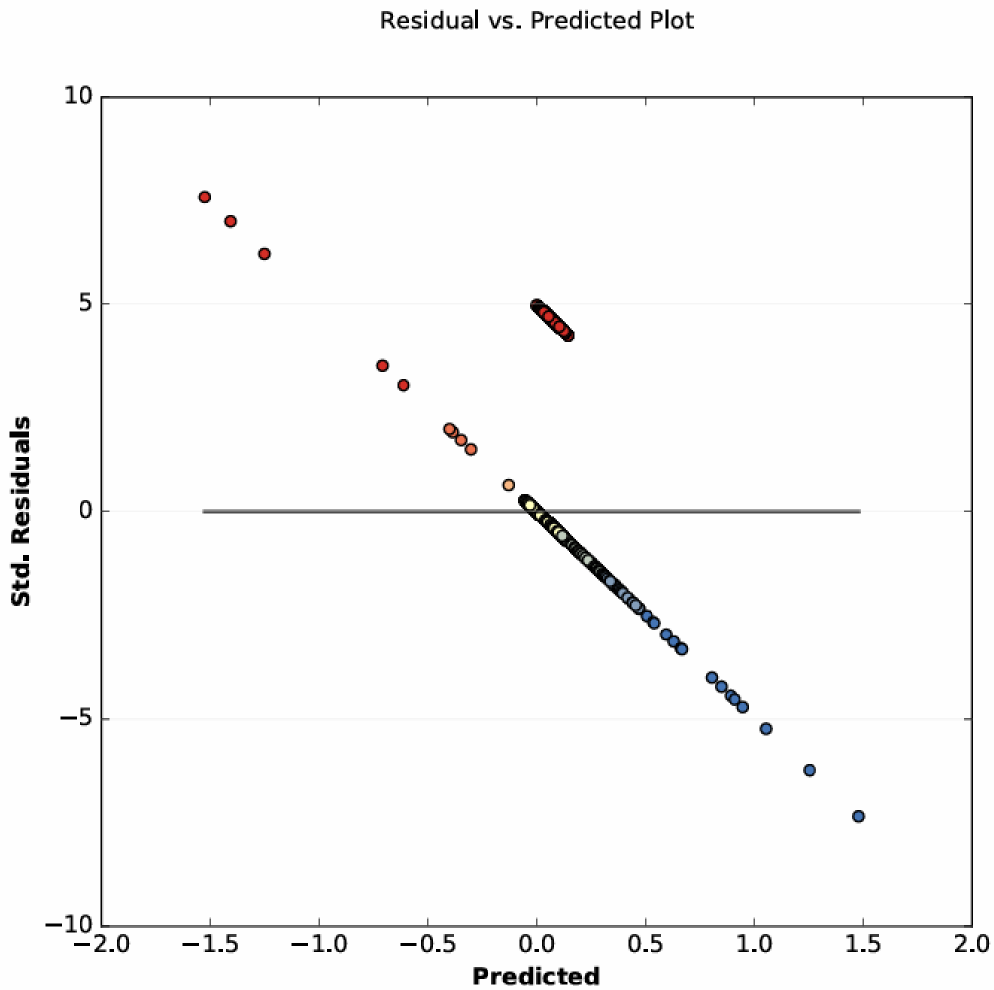


Figure 40: Residual vs. Predicted Plot for meters analyst.

Spatial Distribution:

The map uses standard deviation to visualize the distribution of data points. Points that are significantly different from the mean (either much higher or much lower) are shown in distinct colors.as shown in figure (39), this type of visualization helps identify outliers and understand the distribution of the dataset.

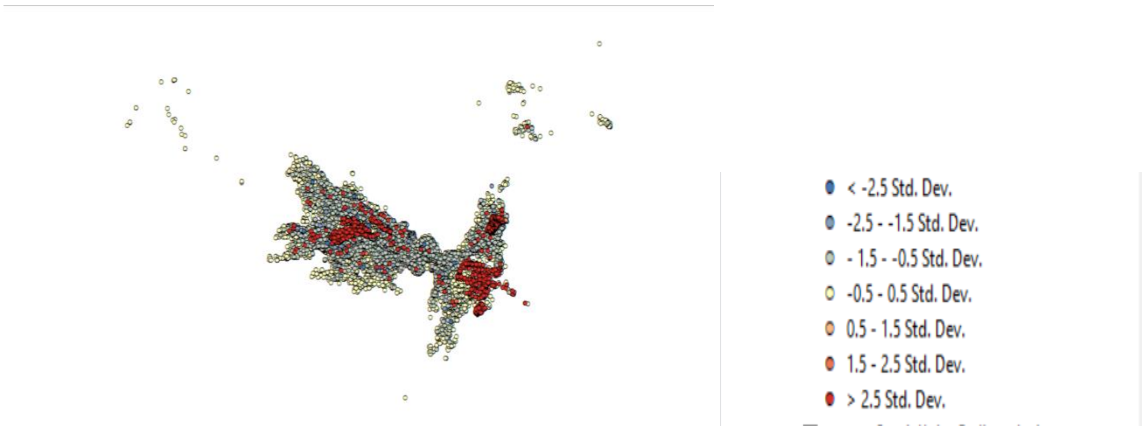


Figure 41: Spatial Distribution for Meters Analyst Result.

Conclusion:

In summary, the OLS model provides a statistically significant but weak explanation of the variance in (MAINT) there are issues with non-normal residuals and multicollinearity, which point to possible areas for model improvement.

B-House Connection OLS Analysis:

2B- Data Pre-processing:

Nablus City Network data contain more than 3200 house connection pipes preserved and archived, each house connection saved each (Object_ID, shape, length, pressure zone, material, diameter, inner diameter, subtype, project_name, fund, and x,y coordinate) some of these factors were unfortunately text and had to be excluded, as shown in figures (39).

InstallDa	Subtype	Material	Contracter	Project_	layer	Link	Type	Inner_Dia
1/2/2010	25 mm	HDPE	Brothers	17st	25mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	20.4_mm
1/2/2010	63 mm	HDPE	Brothers	17st	63mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	51.4_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
1/2/2010	32 mm	HDPE	Brothers	17st	32mm_PE	C:\c-projects\C1\c1_final2011\16ST.dwg	Contract_1	26_mm
9/20/2010	32 mm	HDPE	Brothers	agamei	32mm_PE	C:\c-projects\C1\c1_final2011\agamei.dwg	Contract_1	26_mm
9/20/2010	32 mm	HDPE	Brothers	agamei	32mm_PE	C:\c-projects\C1\c1_final2011\agamei.dwg	Contract_1	26_mm

Figure 42: House Connection Original Data.

The House Connection maintenance data are described in detail in Chapter IV previously, and the methods of processing them are explained after they are ready to be joined to the house connection layer to become a single layer that can be analyzed using the instrument of join between the house connection layer and the maintenance spreadsheet.

One factor to consider is the distance between the house connection pipe and the nearest street, so we added street data that shows the street of Nablus in the form of lines, as shown in the figure (40).

Figure 43: Nablus Street Data.

Using this tool, X coordinates (NEAR_X) and Y coordinates (NEAR_Y) were obtained for the nearest domestic connection street, each in a separate column, and also in another column, the closest value distance between them (NEAR_DIST), it has been calculated by using (Analysis tool -> proximity -> near) tool, as shown in the figure (41).

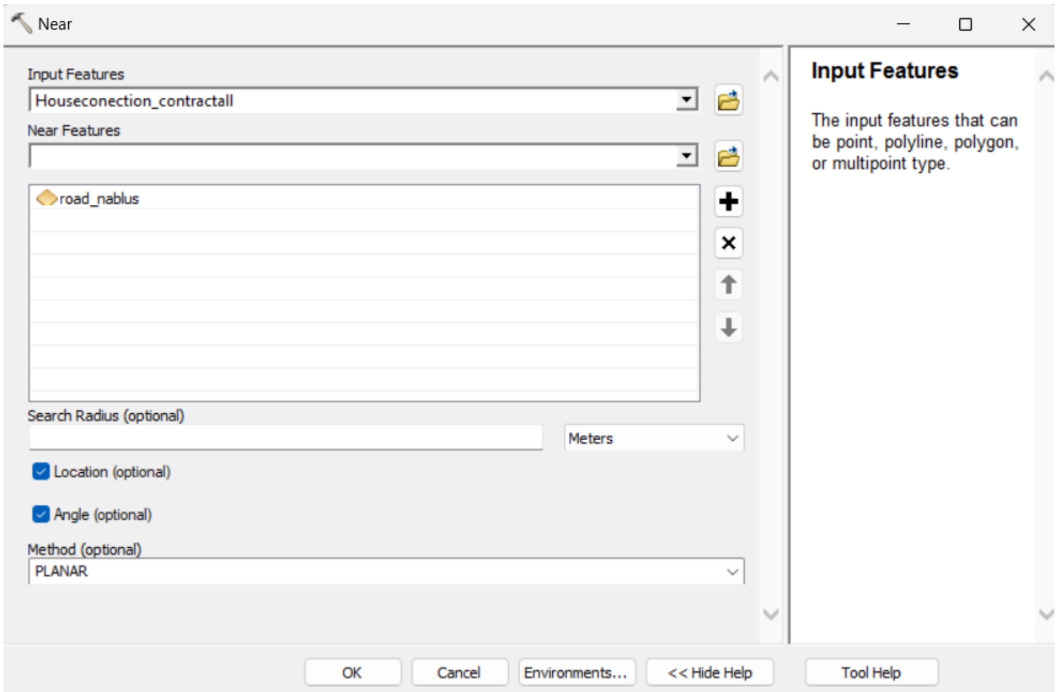


Figure 44: Using Near Tool to Find the Nearest Street to the House Connection Pipe.

I also discussed the statistical OLS analysis that only dealt with numerical values, necessitating the selection, processing, and cleaning of the relevant transactions. To change the values and clean the data, the table must first be extracted and converted into an Excel file.

As a first step: Selection of appropriate parameters, all non-numerical values such as pointer data (Object_ID* 1), dates (install date 01\01\1990, 12:00:00 AM) and blank columns (Link <Null>), (Type<Null>) should be deleted, have to remain just the numerical values. As shown in the figure (42).

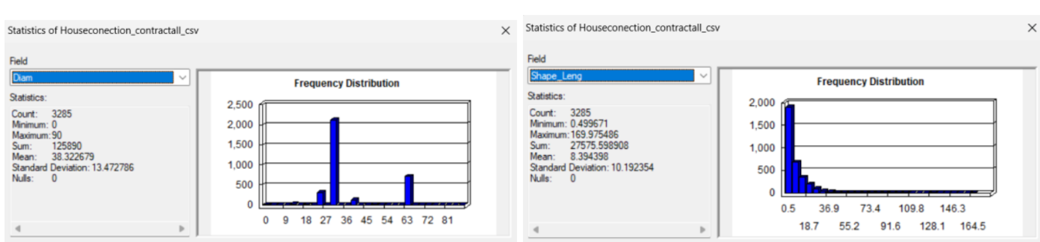
Houseconection_contractall

Enabled	InstallDate	Subtype	Material	Contracter	Project_name	layer	Link	Type	Inner_Dia_r
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	se2	<Null>	<Null>	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	se2	<Null>	<Null>	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	se2	<Null>	<Null>	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	se2	<Null>	<Null>	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	se2	<Null>	<Null>	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	<Null>	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	<Null>	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	se2	<Null>	<Null>	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	se2	<Null>	<Null>	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	se2	<Null>	<Null>	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	se2	<Null>	<Null>	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>
True	1/6/2014	32 mm	HDPE	Dar_Albenaa	WLRP-Phase2b	32mm	<Null>	C6	<Null>

Figure 45: The Parameter which should be deleted.

Second step: is the cleaning stage, the cleaning process involves the elimination of null values. The missing records of (Point_X) and (Point_Y) were calculated using geometric calculation tool to obtain the exact coordinates. After this step we can regenerate a new parameter, the (Near_Dist) column for the records which didn't have correct coordinates at the previous step, thus, all parameters become complete, ready for the next stage.

Third step include checking the statistics for each parameter to examine if any parameter have an out-of-range values by displaying the maximum, minimum and values histogram as shown in the below figures (43).



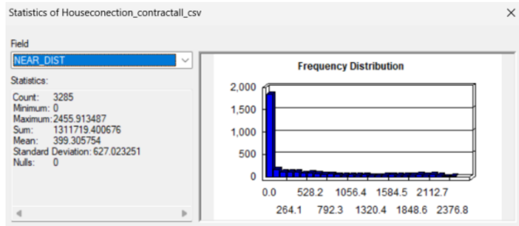


Figure 46: The House Connection Parameters Statistics.

Statistics, as seen in figures, indicate that parameters do not contain values out-of-range.

Reclassify Step: To determine the impact of each factor on the outcome of the analysis, so we created a decimal order to give us the ratio of each value per record of ten.by the following formula:

$$X= \left(\frac{(X_i - \text{MIN})}{(\text{MAX} - \text{MIN})} \right)*10 \qquad \dots\dots\dots\dots\dots(13)$$

Where:

X , is the value ratio.

X_i , Is the original value.

MAX, the maximum value in the parameter column.

MIN , the minimum value in the parameter column.

This value has been calculated for all the numerical factors that will be used in the analysis, each value has been represented as

(F_Dist),(F_DIAM),(F_LENGTH),(F_X),(F_Y).

3B- Subset selection:

As shown in the figure (44), we use the (houseconnection-contractall) layer to analyze the data, the (Object_ID) parameter as a unique ID field because it represents a unique number for each record, (maint) parameter as a dependent variable since it represents the target, it shows that the pipe has been maintained or not, and each of (shape-length, diam, F-Nearst, Near-x, f-depth, and f-y) as an explanatory variables.

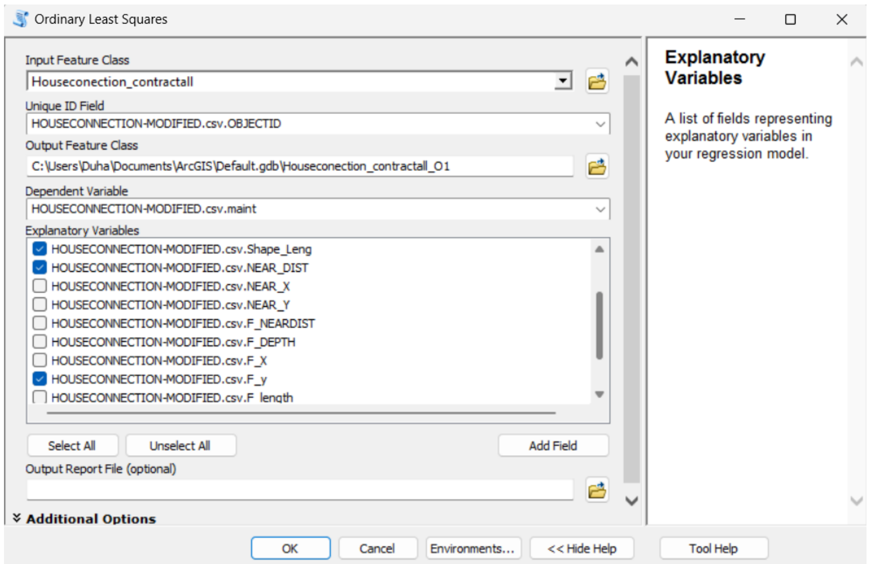


Figure 47: The OLS analysis for the House Connection Pipes.

β0: This is the baseline value of the dependent variable (maint) when all explanatory variables are zero.

Coefficients (β1, β2,..., βk): These values show how the dependent variable changes when each explanatory variable changes by one unit while keeping the other variables unchanged.

3B – Fitting OLS model:

The fitting model process for this analysis is the same as the previous analysis of the count; the analysis of house connections has been fitted in the same way as the previous modeling and calculation of value for transactions.

4B– OLS model Results:

The results report was attached in the appendix, and let’s looks at the results as follows:

3- Number of Observations: 3,280

4- Key Metrics and its Interpretation:

1. Akaike's Information Criterion (AICc): 2524.48, (A measure of model fit; lower values indicate a better fit)
2. Multiple R-Squared: 0.4151 (41.51% of the variability explained by the model)
3. Adjusted R-Squared: 0.4140, (2, 3) (These values indicate that approximately 41.51% of the variability in the dependent variable (MAINT) is explained by the model.)
4. Joint F-Statistic: 387.15 (p-value: 0.0000, significant)
5. Joint Wald Statistic: 2955.89 (p-value: 0.0000, significant), (4, 5) (Highly significant, showing that the overall model is statistically significant.)
6. Koenker (BP) Statistic: 258.05 (p-value: 0.0000, significant), (Significant, suggesting issues with non-stationarity or heteroskedasticity. Hence, use robust probabilities for inference.)
7. Jarque-Bera Statistic: 57.09 (p-value: 0.0000, significant), (Significant, indicating that the residuals are not normally distributed, suggesting model bias.)

- Coefficient Estimates, Significance and Interpretation:

Each coefficient represents the relationship between an explanatory variable and the dependent variable, as shown in the figure (45).

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	31.076112	0.676183	45.958152	0.000000*	0.679391	45.741109	0.000000*	———
HOUSECONECTI	-0.085939	0.018161	-4.732047	0.000004*	0.021068	-4.079148	0.000053*	> 1000.0
HOUSECONECTI	-0.002918	0.000621	-4.701133	0.000004*	0.000612	-4.764995	0.000003*	1.010749
HOUSECONNECT	-0.000171	0.000004	-45.790384	0.000000*	0.000004	-46.054190	0.000000*	1.182137
HOUSECONNECT	0.014939	0.002561	5.832234	0.000000*	0.002059	7.256139	0.000000*	1.142937
HOUSECONNECT	-0.009461	0.005749	-1.645559	0.099963	0.005838	-1.620575	0.105218	1.001607
HOUSECONNECT	0.654708	0.139957	4.677922	0.000004*	0.162002	4.041364	0.000062*	> 1000.0

Figure 48: The OLS analysis Results for house connection pips.

1. Intercept: The baseline level of the dependent variable when all predictors are zero.
2. DIAM: Negative effect, significant, indicating that as this variable increases, the dependent variable decreases.
3. SHAPE_LENGTH: Negative effect, significant.
4. NEAR_X: Negative effect, highly significant.
5. F_NEARDIST: Positive effect, highly significant.
6. F_DEPTH: Negative effect, not significant.
7. F_SUBTYPE: Positive effect, significant.

- Model Diagnostics:

1- Residuals Histogram: it ideally should follow a normal distribution; deviations suggest model bias, as shown in the figure (46).

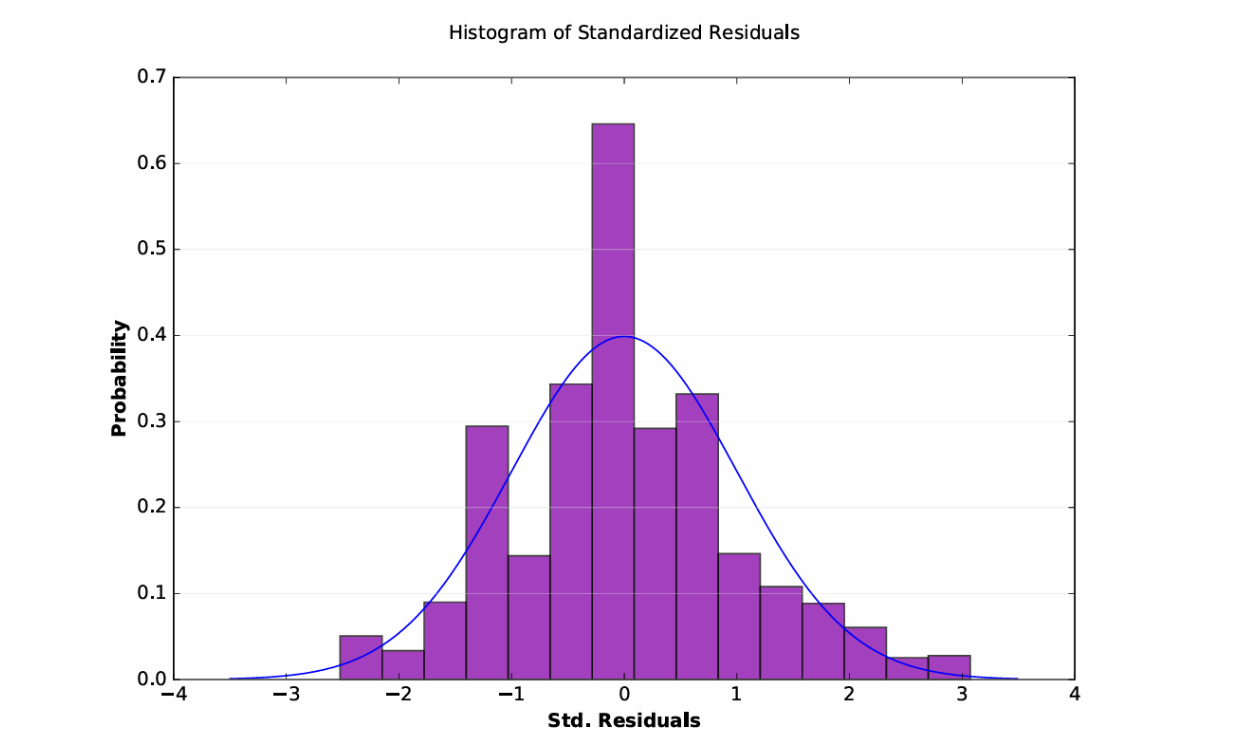


Figure 49: Histogram of Standardized Residuals for house connection analyst.

- Residual vs. Predicted Plot: Random distribution of residuals indicates a well-specified model. Patterns suggest model issues, as shown in the figure (47).

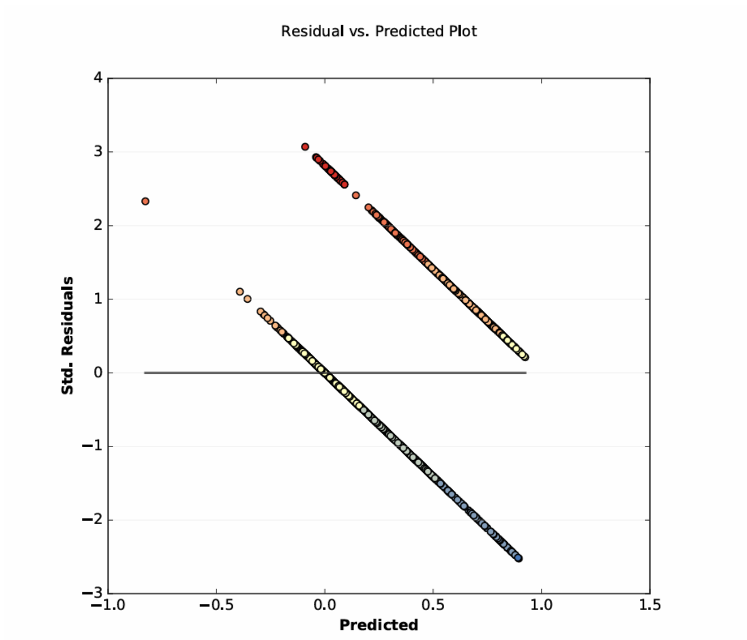


Figure 50: Residual vs. Predicted Plot for house connection analyst.

Spatial Distribution:

Different clusters and patterns can be seen in the point distribution across the map, signifying regions with varying degrees of mean deviation.

High deviation areas (highlighted in red or dark blue) may represent anomalies or notable differences in the data.as shown in figure

Figure 51: Spatial distribution for house connection analyst result.

Conclusion:

The OLS model accounts for a moderate amount of the variability in (MAINT) and is statistically significant. On the other hand, problems with non-normal residuals and possible multicollinearity point to locations that need more validation and model improvement.

5.3 Control Valve:

When there is a fracture in one of the main pipes or house connections and the detection of a leak of water, the flow of water must be stopped to save it, prevent its waste, and complete the maintenance process and replace the broken pipe. The process of locating the broken pipe remains important and needs special attention because precise identification accelerates the maintenance process and helps in the decision-making process to choose the best valve to close and cut off water from the area concerned.

The process of selecting the control valve for closure may be done using several methods, but not all of them with the required accuracy and realism, including the closest valve (the closest valve of the broken pipe) using the shortest path algorithm.

The shortest path algorithm is a fundamental method for determining the shortest route connecting two nodes in a network with non-negative edge weights ;it is frequently utilized in several applications, including network routing, transportation planning, and logistics.

To clear the problem definition of this algorithm we can say if we had given a weighted graph $G= (N, E)$, where N stands for the set of nodes and E represents the set of edges (connections between nodes), along with non-negative weights assigned to each edge, the aim is to determine the shortest path from a source node s to a destination node d [45].

The process starts with initializing a data structure to keep track of the shortest distances from the source node (the broken pipe) to all other nodes (the valves) in the graph; and assigning a zero distance to the source node and an infinite distance to all other nodes. Storing nodes sorted by their estimated distances from the source node, and finally initializing a priority queue or other comparable data structure [46].

I need to repeat the following steps to reach the intended node or all the accessible nodes are processed [47]:

- Extract the node at the lowest temporary distance from the priority queue.
- Explore the adjacent nodes (neighboring nodes) and update their temporary distances if a shorter path is found through the current node.
- In order to account for changes in temporary distances, update the priority queue.

- Proceed with the procedure until either the priority queue is cleared or the destination node is reached.

The algorithm ends when either the destination node is reached or all attainable nodes have been processed. By using the recorded shortest distances and predecessors to return to the destination node to the source node, the shortest route from the source node to the destination node can be reconstructed [48].

The most popular uses for shortest path algorithms are in network routing, navigation systems, logistics, transportation planning, and other domains where optimization problems need to be solved. They assist in determining the best routes between points, reducing travel time or distance, and allocating resources optimally across a range of applications [49].

The shortest path algorithm can be made more effective by utilizing a variety of optimization strategies. For example, a priority queue or min-heap can be used to quickly extract the node with the smallest potential distance [50].

A variety of shortest path algorithm variations, like Floyd-Warshall algorithm, Bellman-Ford algorithm, and Dijkstra's algorithm, may be employed, depending on the particular application and requirements [51][52].

Despite all that said and all the benefits of this algorithm, it was not a viable solution in the process of choosing the best valve to close in the event of a broken single pipe or water leak instead of shutting down the main source and denying all consumers access to water, It is not always the closest valve in terms of distance that is responsible for supplying the broken pipe with water as shown in figure (48) , although the valve number (1) is the closest to the broken pipe, the valve number (2) is its provider. In the event of a break or leak, the responsible valve must be closed valve number (2) and not the nearest (1).

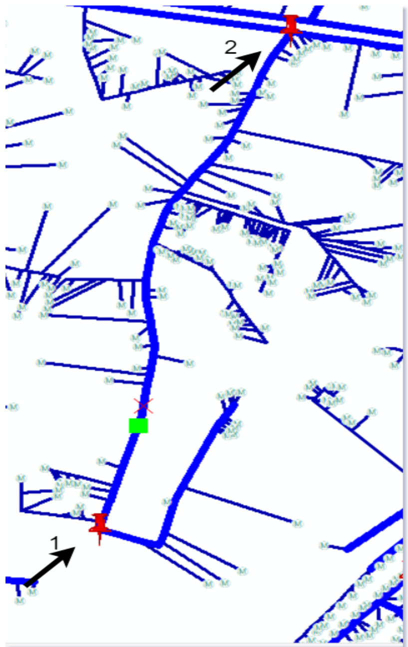


Figure 52: The control Valve Not the Nearest Valve.

To choose the controlling valve, the network had to be analyzed using specialized algorithms because selecting the valve manually is useless in case the data was changed to another network, so we searched for an algorithm that analyzed the network and selected the controlling valve, and among several options, we selected Network Isolation Algorithms with graph theory algorithm .

Based on predetermined criteria, network isolation algorithms are used to locate and isolate particular network components. The objectives may be to stop breakdowns from spreading, divide the network for security reasons, or improve resource management [98].

In order to prevent cascading breakdowns, this algorithm is used for fault tolerance, which entails isolating failing nodes or segments. Security is the process of separating hacked nodes from the rest of the network to keep it safe. And resource management, which divides the network into segments to control traffic and resources effectively. In a water distribution network, for instance, separating a section where a leak has been found will stop water loss and pressure problems in the remaining network segments [99].

However, we must first graph the network. Graphs are mathematical structures that stand for pairwise relations between objects. Graph theory algorithms are a collection of procedures used to solve problems relating to graphs. Edges show the relationships between nodes, which stand in for entities [100].

Various techniques are permitted to graph the networks, and each technique has its own algorithm. The shortest route between two nodes can be found by applying algorithms like Dijkstra's and A* [101]. Minimal Spanning Tree (MST) that is to identify a subset of edges that connects every node with the least amount of whole edge weight by using algorithms such as Kruskal's and Prim's [102]. Connectivity: Component identification and network connectivity algorithms for graphs [103]. Flow Algorithms: To determine the maximum flow in a network, algorithms such as Ford-Fulkerson are utilized [104]. These algorithms are useful in a many fields, including transportation networks (which employ them to determine shortest path between two locations) and utility networks (which use them to make sure all network components are effectively connected and functioning) [101].

The graph traversal algorithm is known as Breadth-First Search (BFS) which begins at a chosen node (root) and investigates every one of its neighbors at the current depth level before moving on to nodes at the next depth level [105].

BFS has two characteristics: the capacity to determine the shortest route from the root node to any other node in an unweight graph, and the ability to determine the Shortest route in unweight graphs in general. And Level Order Traversal, which can look into nodes at a level-by-level and it can be used for Network Analysis and Shortest Path Finding. It can also be used in network isolation to help isolate sub-networks by identifying all nodes that can be reached from a starting node. [106].

In order to apply the Breadth-First Search algorithm, we must do the following steps [103], [105]:

1. Begin with the root node.
2. Put the root node in queue.
3. As long as the line is not empty:

-A node should be dequeued from the front.

-Go to the node and designate it as explored.

- Enqueue every nearby, untouched node.

BFS and other graph theory algorithms are essential when examining a network for isolation purposes. And the way they are related is as follows [107]:

- Detection: Locate a node or edge (such as a leaking pipe or breakdown meter) that requires isolation.
- Traversal: Investigate every node and edge that can be reached from the point of interest using BFS. This aids in determining how much of the affected area is affected.

In mathematical terms we can explain the fundamental of the BFS as follow:

1- Graph representation:

The following figure (49) shows the graph of the valves distribution in our data each valve is marked through the (Object_ID) field.

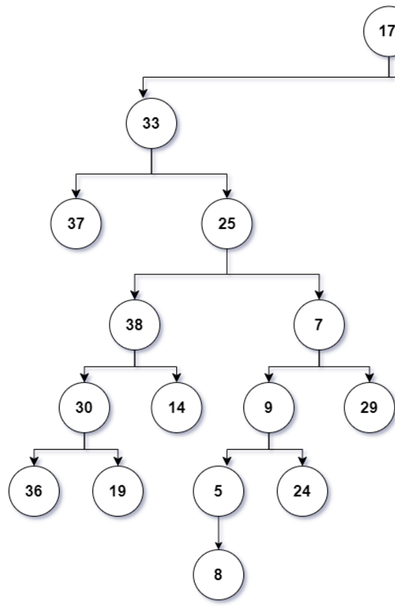


Figure 53: The graph for the distribution.

Let:

$$G = (V, P) \qquad \ldots\ldots\ldots\ldots\ldots\ldots.(14)$$

G is a graph with P representing the set of edges (pipes) and V representing nodes (valve).

For each V , N(neighbors) which represent the number of neighbors.

And:

$|X|$ represents how many vertices there are.

$|Y|$ represents how many edges there are.

2- Initialization:

We have to create a queue (Q) to manage the valves to be explored, and create (D) as a set which used to keep track of visited nodes.

Traversal process can be done by these steps:

- Start from a node S (sources).
- In Q store the D set.

Q=[S]

D={S}

- While Q Is not empty:

1- Dequeue a nodeV_i.(I = number of valves).

2- For each neighbor V_i∈ N(V_i):

IF V_i is not ∈ D :

Enqueue V_i and add it to D.

3- Complexity:

- BFS time complexity of is $O(|V| + |P|)$, This is due to the fact that every valve(node) is examined and dequeued exactly once, as well as every pipe (edge).
- Creating the queue and visited set takes constant time, $O(1)$.
- Node Processing: Every node's enqueuing and dequeuing takes time, $O(1)$.
- Verifying a node's neighbors requires examining its neighboring edges. This is to proportionate the total edges number in the graph, $|P|$.

Thus, the overall complexity is

$$\text{Complexity Time} = \sum_1^i O(|V|) + \sum_1^j O(|P|) \quad \ldots\ldots\ldots\ldots\ldots\ldots(15)$$

Where i is the valves number , j is the pipes number.

Conclusion:

Although BFS lacks a conventional mathematical formula, an understanding of its algorithmic phases and time complexity offers a clear picture of how it functions. When it comes to tasks like control valve placement in water distribution networks, the BFS algorithm's simplicity of implementation and efficiency make it an indispensable tool for achieving the best possible network isolation and management.

Chapter Six

Our Desktop Application:

6.1 The Application

We found it possible to find a practical solution that contributes to the decision-making process with less time and effort and higher quality. This is because of the many challenges facing water systems and the scarcity of technology and applications that help municipalities and water service institutions save time and effort, and that there are real problems in these institutions.

We have created a desktop application aimed at facilitating the handling of water distribution systems through the algorithms and analyses described in detail in the previous chapters.

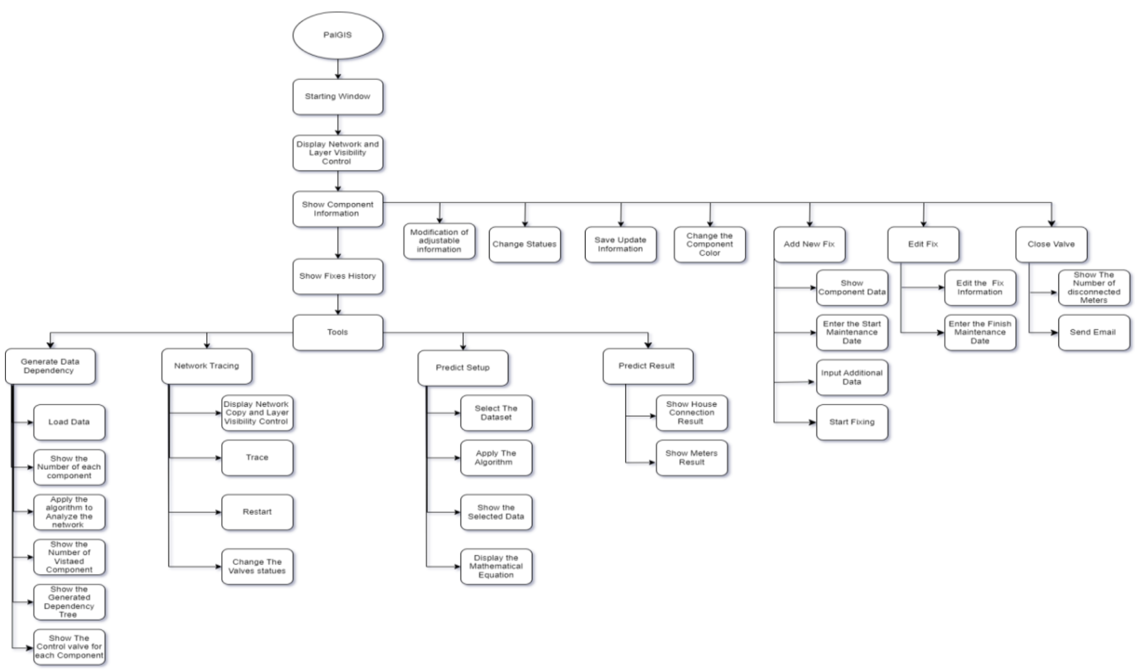


Figure 54: The App diagram.

The previous diagram shows the capabilities of the desktop application and the services it provides. The application is an integrated system that receives the processed spatial data, displays it in the form of a network in the correct coordinates on the map and allows the user to recognize the attributes and information of each element and to control it. The application also includes a network analysis system using theory algorithms to build the tree structure of data and identify the controlled valve in each element. A forecast system based on data mining algorithms is used to determine the component expected to require maintenance based on historical maintenance data.

This app was created using the programming language C# and the Video Studio Code 2022 software, mainly relied on GIS libraries provided by ISRI , especially ArcGIS SDK.NET Library, which is free, light, convenient to build a desktop application, suitable for a mobile version whether Android or iPhone, and effective in providing facilities for building code structures.. This program generally aims to display data as a network to facilitate its management handling and modification. As shown in the figure

Figure 55: View the network in the app.

These data have been converted from shapefiles to Geofiles so that C# can read and represent them based on ArcGIS SDK.NET Library, after we have read and saved the data in dataset as tables to make it easier to read and view. As shown in the figure

	OBJECTID_1	Id	Line_ID	OBJECTID	Customer_No	Pressure_Zone	Bl_Num	BLOCK_NM	PARCEL_NM	Construction_Dat	kind	POINT_X	POINT_Y	BUILDING_
	Filter	Filter	Filter	61	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	61	NULL	NULL	61	7504	E0.2	3996	Balata_9	A/3/18+1/19	1/1/1969	صناعي غذائي	177767.59869...	179466.23492...	NULL
2	161	0		161	35360	E0.2	496	Roujeeb_3	52/237	3/23/2006	منزلي داخل ال...	178079.92633...	178318.37408...	496E0.2
3	261	0		261	18111	E0.2	730	Roujeeb_3	41/237	7/26/1989	منزلي داخل ال...	178183.51110...	178488.23907...	727E0.2
4	361	943305946	NULL	361	40595	E0.2	1138	Balata_12	35/33+34+36	3/4/2010	ورشه بناء	178238.87988...	178796.92932...	1129E0.2
5	461	852771245	NULL	461	43406	E0.2	4066	Balata_6	79	NULL	NULL	177200.25512...	179487.33911...	4065E0.2
6	561	0		561	6582	E0.2	6050	Balata_5	35	5/16/1966	موظف كهرباء	177073.12072...	179893.97772...	6050E0.2
7	610	0		610	27585	E0.2	NULL	Balata_5	13	NULL	NULL	177156.21771...	179950.89288...	NULL
8	611	0		611	22078	E0.2	NULL	Balata_5	13	9/16/1995	منزلي داخل ال...	177155.98608...	179950.80267...	NULL
9	612	0		612	23192	E0.2	6102	Balata_5	1/51	6/6/1996	تجاري	177226.50250...	179900.47509...	6093E0.2
10	613	0		613	9694	E0.2	5883	Balata_5	85+86	11/24/1974	تجاري	177334.73010...	179869.16809...	5883E0.2
11	614	0		614	16820	E0.2	5569	Balata_5	19	11/29/1986	تجاري	177316.31872...	179828.72949...	5555E0.2
12	615	0		615	13682	E0.2	5886	Balata_5	85+86	5/24/1980	منزلي داخل ال...	177334.15069...	179869.409729	5883E0.2
13	616	0		616	22232	E0.2	6405	Balata_5	82+83	7/15/1995	تجاري	177356.66107...	179932.06750...	6405E0.2
14	617	0		617	12624	E0.2	6408	Balata_5	82+83	5/22/1978	تجاري	177355.94128...	179932.34429...	6405E0.2
15	618	0		618	6175	E0.2	6411	Balata_5	82+83	7/11/1965	تجاري	177356.05187...	179933.17492...	6405E0.2
16	619	0		619	38350	E0.2	4263	Balata_6	93	6/7/2008	منزلي داخل ال...	177228.98187...	179557.31188...	4263E0.2
17	661	0	بلاطه البلد	661	10436	E0.2	5757	Balata_4	NULL	9/14/1976	منزلي داخل ال...	176885.97167...	179854.51092...	5757E0.2
18	761	960370146	NULL	761	42248	E0.2	794	Roujeeb_1	174	5/8/2011	ورشه بناء	177485.50427...	178566.91528...	746E0.2
19	861	0		861	9816	E0.2	1031	Kfor_Qaleel 11	66	4/19/1975	منزلي داخل ال...	177107.22967...	178737.07489...	1024E0.2
20	961	0		961	14475	E0.2	5806	Balata_5	72	9/10/1981	منزلي داخل ال...	177249.70648...	179859.36370...	5797E0.2
21	1061	NULL	NULL	1061	35481	E0.2	1151	Kfor_Qaleel 11	A/48+50	5/2/2006	منزلي داخل ال...	177206.97851...	178804.00750...	1151E0.2
22	1161	158	NULL	1161	33558	E0.2	6742	Asker_16	NULL	12/11/2004	تجاري	177808.55987...	179966.31512...	6742E0.2
23	1261	32483	NULL	1261	40337	E0.2	870	Roujeeb_1	248	1/16/2010	منزلي داخل ال...	177327.53833...	178638.51232...	870E0.2
24	1361	957251549	NULL	1361	45077	E0.2	5648	بلاطه 5	73	3/30/2013	منزلي داخل ال...	NULL	NULL	5638E0.2
25	1461	950348797	NULL	1461	48039	E0.2	NULL	Roujeeb_1	60	19/3/2015	منزلي داخل ال...	NULL	NULL	NULL
26	1561	907747372		1561	50402	E0.2	0	Balata_10	1/4	14/8/2016	منزلي داخل ال...	0.0	0.0	
27	1610	852551704	NULL	1610	51841	E0.2	NULL	Balata_5	2/78	NULL	NULL	NULL	NULL	NULL
28	1611	955547757	NULL	1611	51917	E0.2	NULL	Balata_12	117	NULL	منزلي داخل ال...	NULL	NULL	NULL

Figure 56: Saved Data in the app dataset.

These data are displayed as layers each layer is displayed on a panel on the left of the network, and the appearance of the layer is controlled whether to disappear or appear through a check box. On the right, another panel shows information on any component selected, such as its name, ID number and specifications by component type. It also includes fixes history which the historical maintenance of the specific component will be displayed. As shown in the figure (52).

Figure 57: Application interface.

It is possible to change the icon for any item easily through the code. Some of the attributes which is shown on the information tab are adjustable, and the rest are not. If any of the editable features are modified, the new data can be saved by clicking the Save button, each item changes color depending on its condition; for example, the main tubes are blue when they are valid, yellow when they are under maintenance and red when they break down.

It is possible also to change the background of the apparent network of the user through the list of map that appears at the top so that you can add streets, aerial image of the area or topography as shown in the images

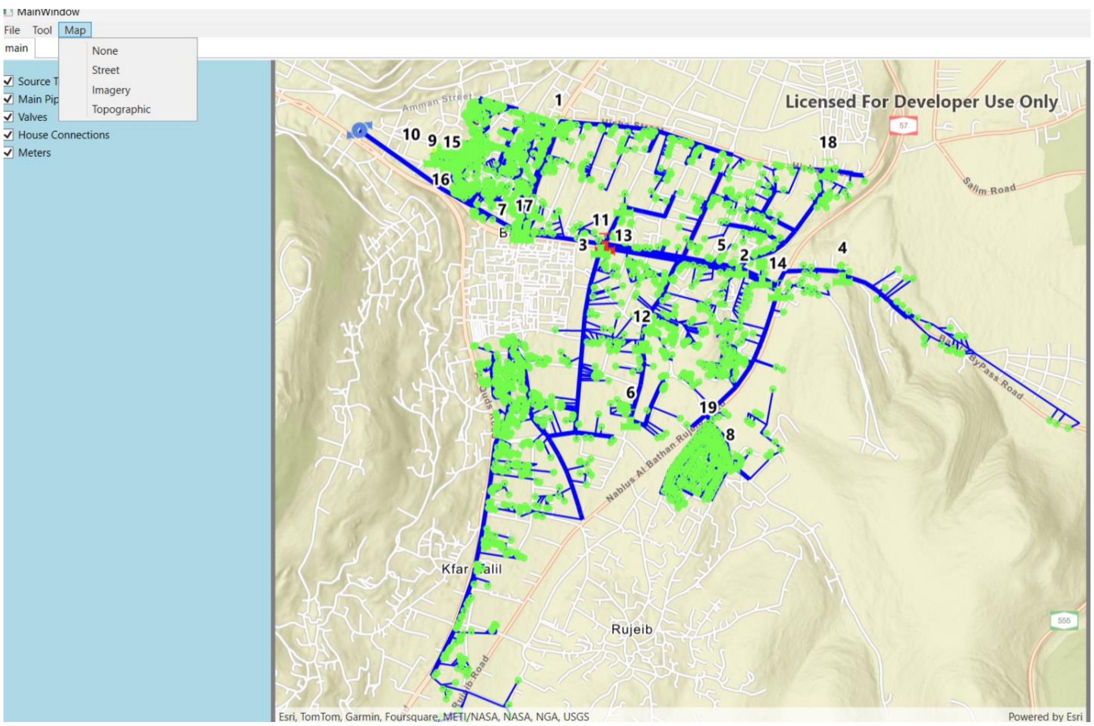


Figure 58: street background.

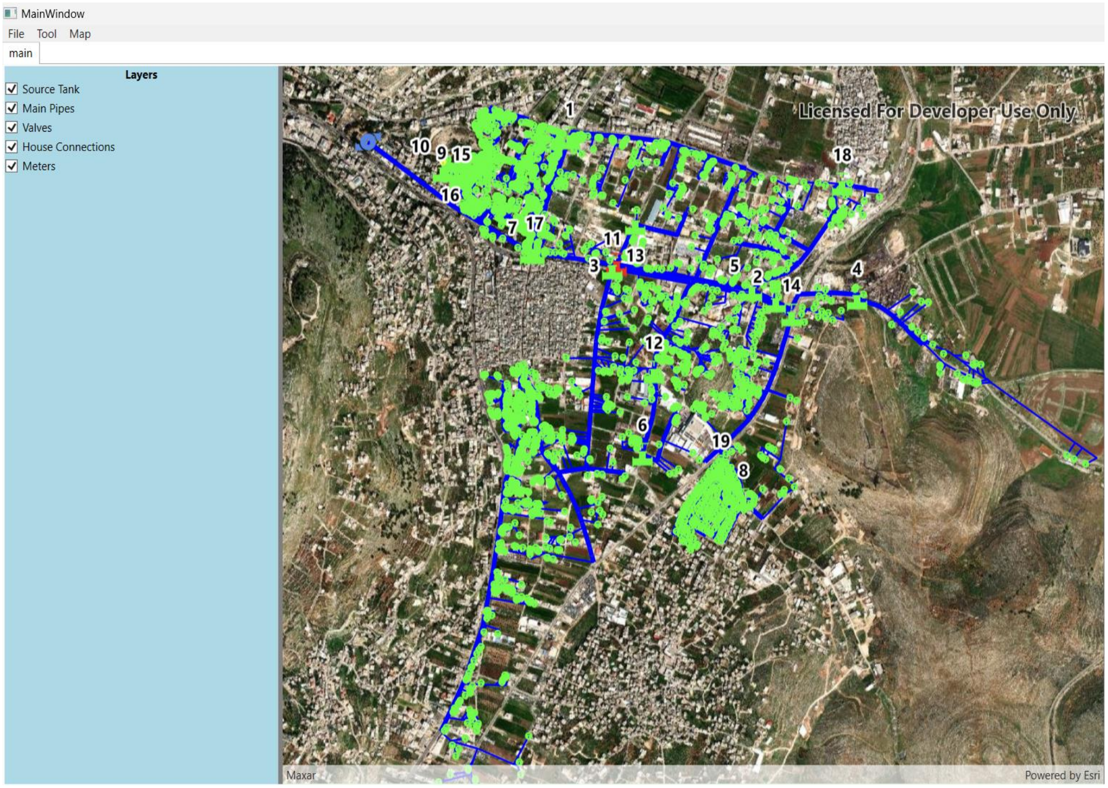


Figure 59: Aerial Image background.

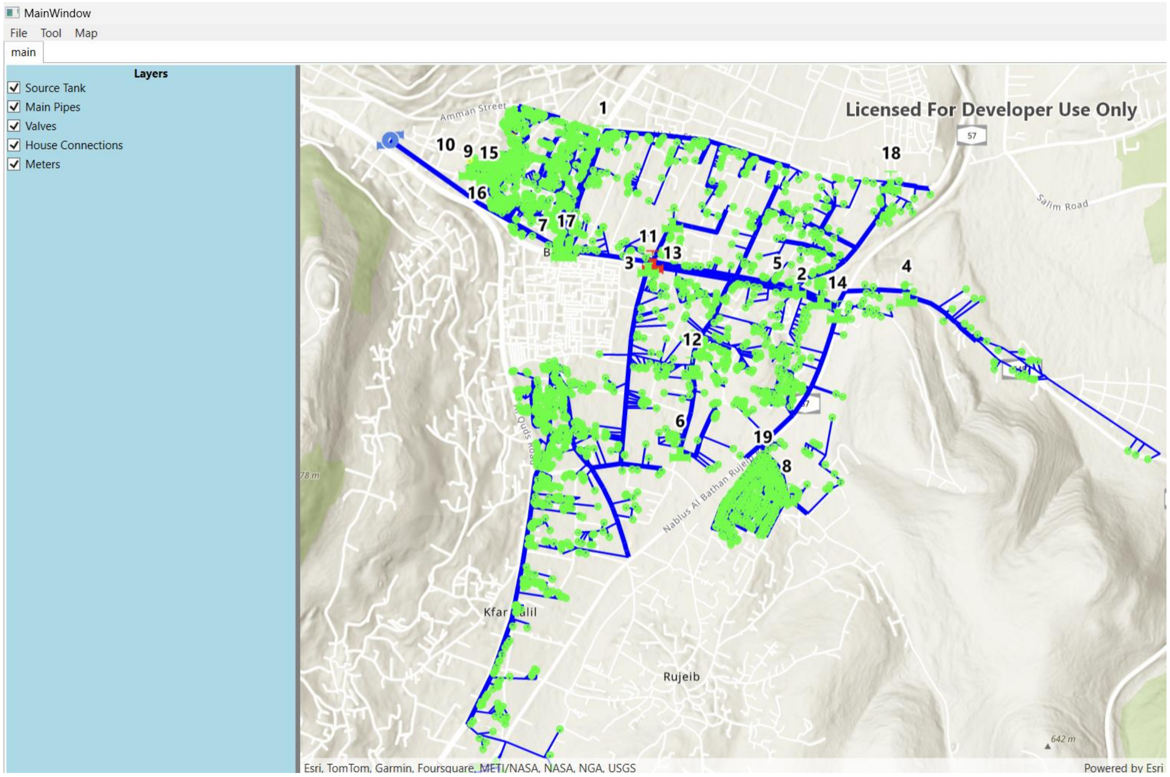
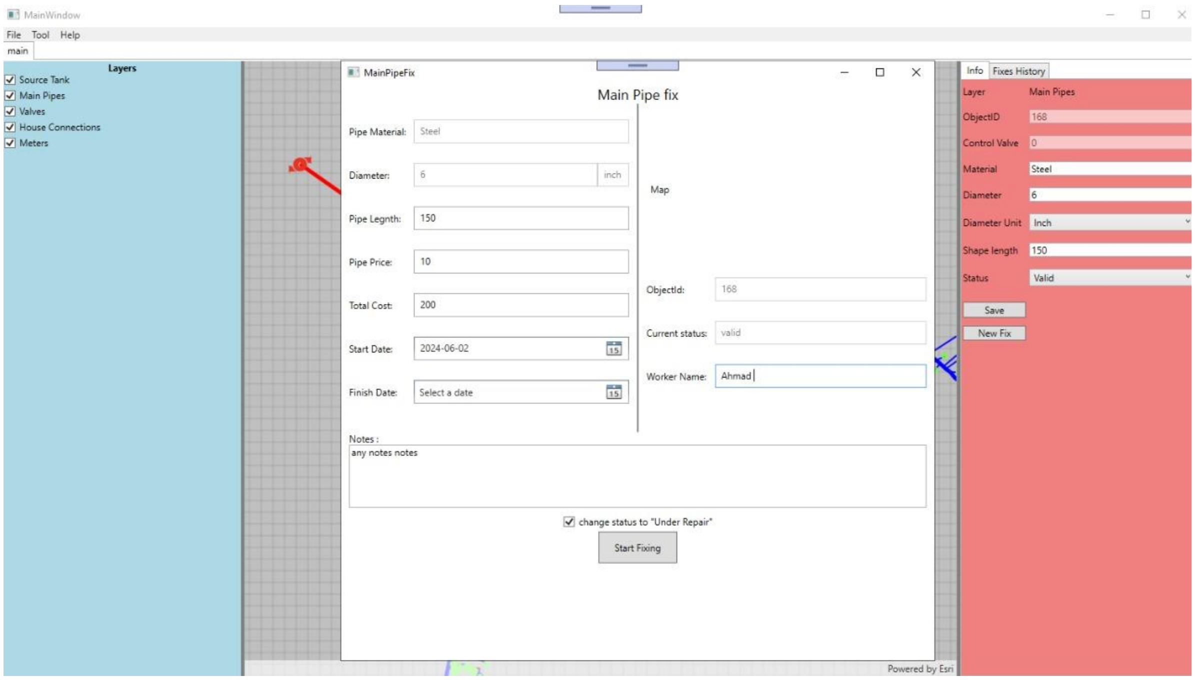


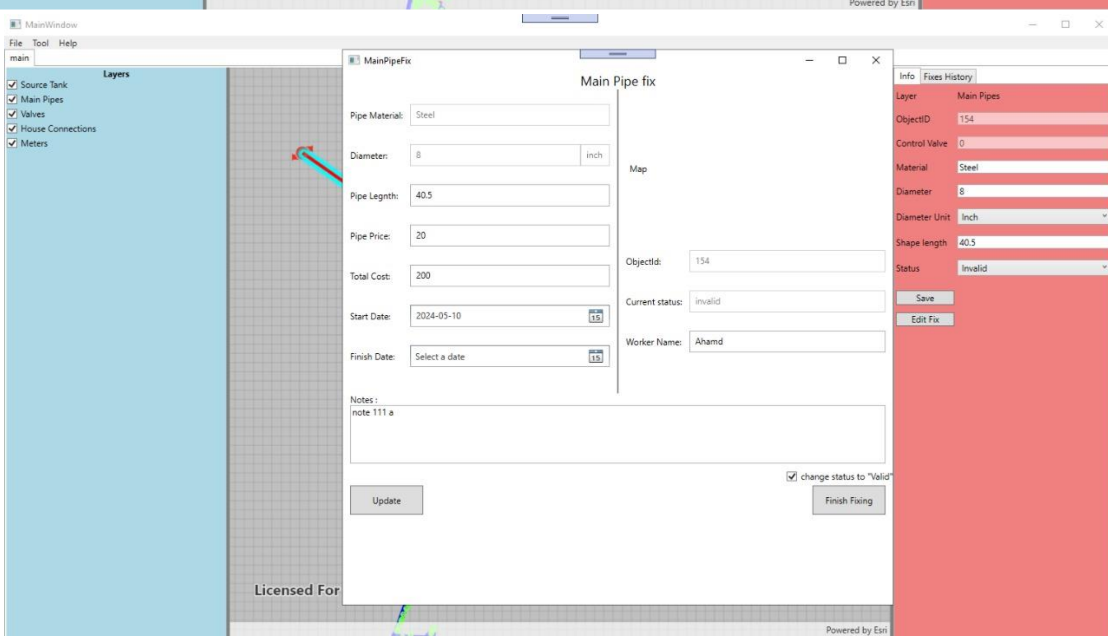
Figure 60: Topographic Background.

In case the component needs maintenance you have to press (Fix) button, this will lead to pop-up to display the details of the maintenance process, The pop-up contains details about the selected component, the price of this component, the approximate price of the total maintenance process which was detailed in chapter Four and the start and end date of the maintenance process so as to allow the user to modify the maintenance process data as long as the end date of the operation is not entered, during this period, the button (Fix) button turns into a (Edit Fix). As shown in the figure (53)

The maintenance process start date is the date of the establishment of the maintenance process by default, but it is adjustable if the process will start in another day. It also appears in the pop-up name of the worker so that it is entered by the user, automatically converting the component status from (valid) to (under repair) is activated, and it is adjustable by the check box at the top of the (start fixing) button.

If the maintenance process is approved, the statues of the component will be converted into (under repair). The data in the maintenance operation can be modified until the date of completion and termination of the operation is entered. It will become non-adjustable and will be archived in the database as a GIS. The user will be able to view the previous maintenance history through the tab (Fix History).as shown in the figure


MainWindow
File Tool Help
main
Layers
Source Tank
Main Pipes
Valves
House Connections
Meters
MainPipeFix
Main Pipe fix
Pipe Material: Steel
Diameter: 6 inch
Pipe Legnth: 150
Pipe Price: 10
Total Cost: 200
Start Date: 2024-06-02
Finish Date: Select a date
Map
ObjectId: 168
Current status: valid
Worker Name: Ahmad
Notes :
any notes notes
change status to "Under Repair"
Start Fixing
Info Fixes History
Layer Main Pipes
ObjectID 168
Control Valve 0
Material Steel
Diameter 6
Diameter Unit Inch
Shape length 150
Status Valid
Save
New Fix
Powered by Esri


MainWindow
File Tool Help
main
Layers
Source Tank
Main Pipes
Valves
House Connections
Meters
MainPipeFix
Main Pipe fix
Pipe Material: Steel
Diameter: 8 inch
Pipe Legnth: 40.5
Pipe Price: 20
Total Cost: 200
Start Date: 2024-05-10
Finish Date: Select a date
Map
ObjectId: 154
Current status: invalid
Worker Name: Ahamd
Notes :
note 111 a
change status to "Valid"
Update
Finish Fixing
Licensed For
Info Fixes History
Layer Main Pipes
ObjectID 154
Control Valve 0
Material Steel
Diameter 8
Diameter Unit Inch
Shape length 40.5
Status Invalid
Save
Edit Fix
Powered by Esri

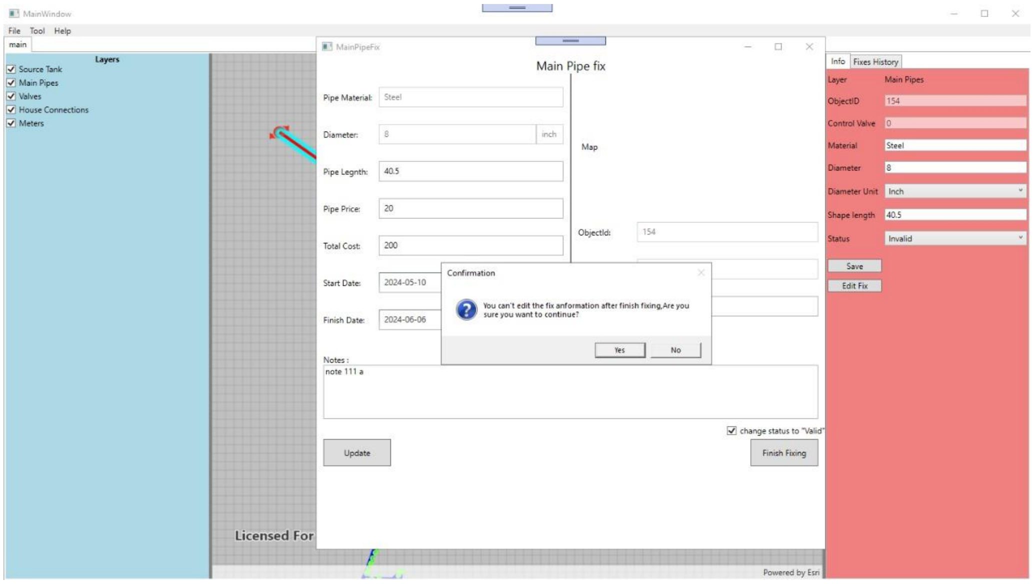


Figure 61: Maintenance Pop-up in many cases.

The note box is included to add any additional note other than the data that can be entered by the user, such as the increase in the total cost value due to the need for an additional piece or displaying this component of rust, but it is still working, i.e., the notes of the exact component that cannot be circulated.

Figure 62: Fix History

The process of determining the controlled valve helps save time and effort and reduces water loss. As we indicated in Chapter V previously, when there is a breakdown of one of the components in the network, the controlling valve must be specified, which must be closed to stop water pumping in that area first to conserve water and secondly to complete the maintenance process.

The process of selecting the controlled valve is done in two ways: first manually and practically on our data, but it is not a useful solution in case the data are changed to another area and the location of one of the valves is changed, removed, or temporarily stopped. Second, using one of the algorithms that analyze the network and build a hierarchical model, considering that the water source is the main root and the valves are the sons, the isolation algorithm and the Breadth-First Search algorithm that are explained in detail in Chapter V have been applied in the analysis and identification of the controlling valve in each component of the network, the system is proposed to be closed if the selected component fails.

From (Tools -> Generate Valves Dependency) we can apply the isolation algorithm and the Breadth-First Search algorithm on the data. As shown in the figure

Figure 63: Generate Valves Dependency.

This will open a new pop-up .as shown in the figure (56) this pop-up contains the number of each feature in the network, When the button (load) is pressed the app load the data (because it may change in whole or in part), When the button (start) is pressed the algorithm starts to analyze and determine the number of visited feature to be sure every component is visited, and build the tree to determine the control valve for each component.as shown in the figure (57) which represents the tree and the components with their control valve . After this step the value of control valve will be save in the dataset so when we select any feature, its control valve number will appear on (info) panel.

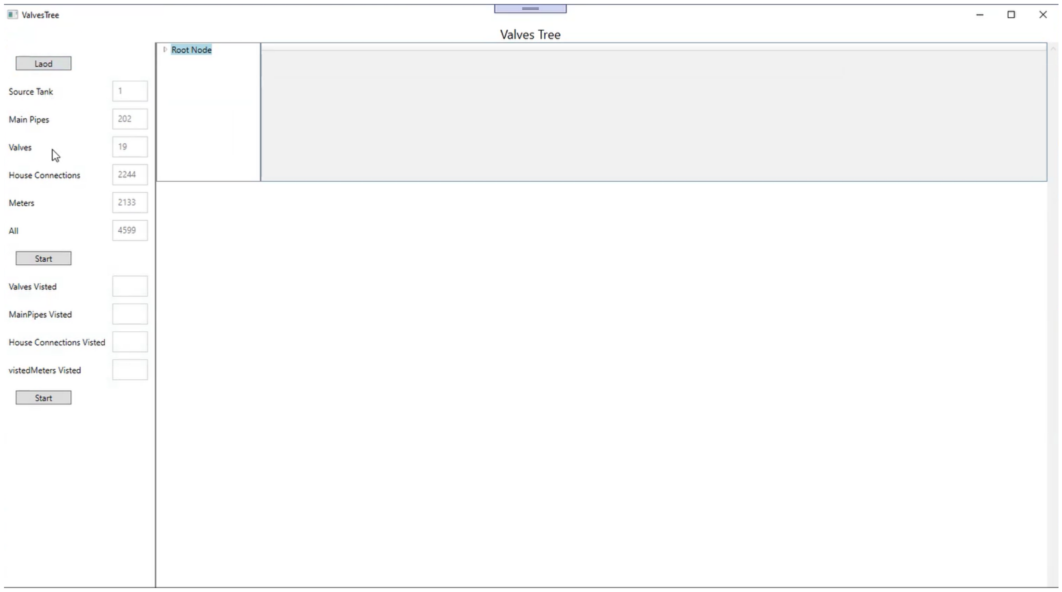


Figure 64: Generate Valves Dependency pop-up.

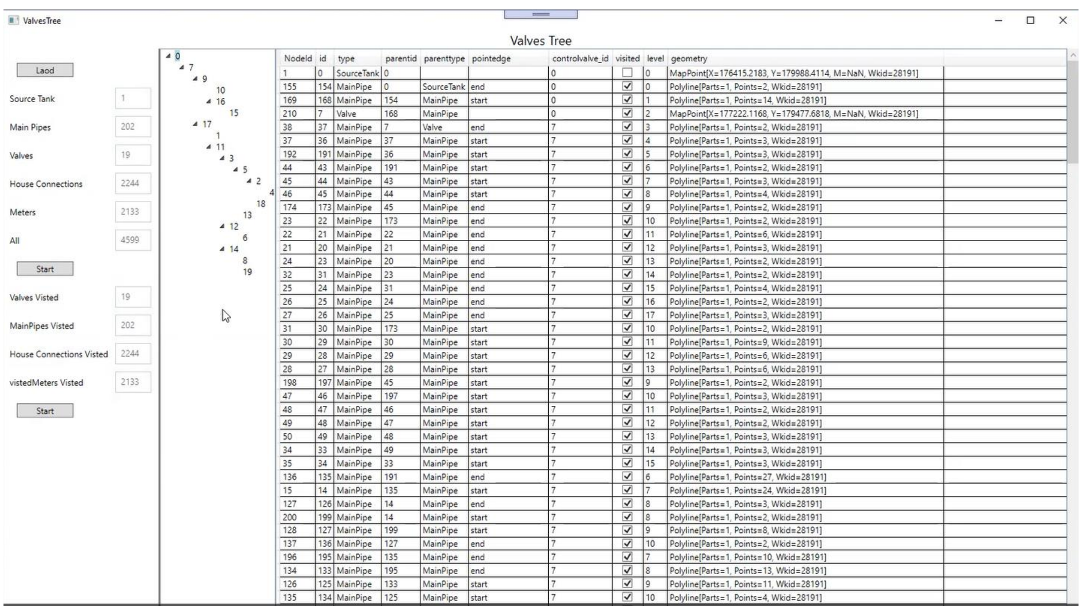


Figure 65: Generate Valves Dependency Tree.

the controlled valve number for each meter has now become known , and the set of meters that will be disconnected Upon closing the controlled valve, so when the valve status is turned into (invalid) or (under repair), a pop-up will appear to ask the user if he wants to send a message to the consumers of these meters that mean is (We apologize for the disruption of your water access service for an emergency that will be dealt with as soon as possible, Thank you for your patience and understanding.) as shown in the figurer

Figure 66: The Massage will be send to the consumers.

Sometimes the user may need to make several scenarios to trace the water flow of the network and know the components that water connects in each scenario, so we created a network tracing tool from (Tool -> Network Tracing). This tool will display a new pop-up version of the displayed network to ensure that no erroneous modifications are made to the main interface (to stay in the safe side). This pop-up contains the same details as the main interface to control the appearance of the components, their invisibility, and the possibility of changing their attributes values, as well as the (trace) button. When this button is pressed, all the components that the water will reach will be highlighted according to the scenario data, such as if one of the valves is invalid, one of the pipes is invalid, or one of the meters is discontinued. As shown in the figure

Figure 67: The network tracing.

The use of data mining algorithms offers the ability to create a prediction system based on historical data for maintenance operations to predict components at risk of being broken and that may need maintenance. This topic was discussed in detail in Chapter V previously.

The decision tree algorithm was selected to build a prediction system for both meters and house connections based on historical data for both maintenance operations as mentioned in the previous chapters, each explained in detail as follows:

1- Meter Prediction System:

The prediction system is based on the following four factors that the algorithm will analyze and rely on in the final decision: First, the date of installation that results the age of the meter is considered by the algorithm to be the root because it has the information gain highest value. Second, pressure tolerance if the meter can withstand pressure, this mean it has less chance for needs to maintain. Third, the flow rate for each meter, and finally the remainder of its life span, as it is known that each meter has a life span after which it loses its efficiency to operate. The shorter the meter's life span, the greater the chance it needs maintenance. The following figure shows the factors relations.

The flow rate per meter has been calculated according to the monthly reading available in the data for the month of January 2023 (Reading-12023). A new field has been added to the data, including the calculated value based on the following formula:

$$Flow\ Rate = \frac{Reading_{12023}}{(22 * 8)} \qquad \dots\dots\dots\dots\dots\dots.(16)$$

We multiplied the number of working hours in a day by the number of working days in a month, and then divided the total reading for the month by that number, to arrive at the flow rate.

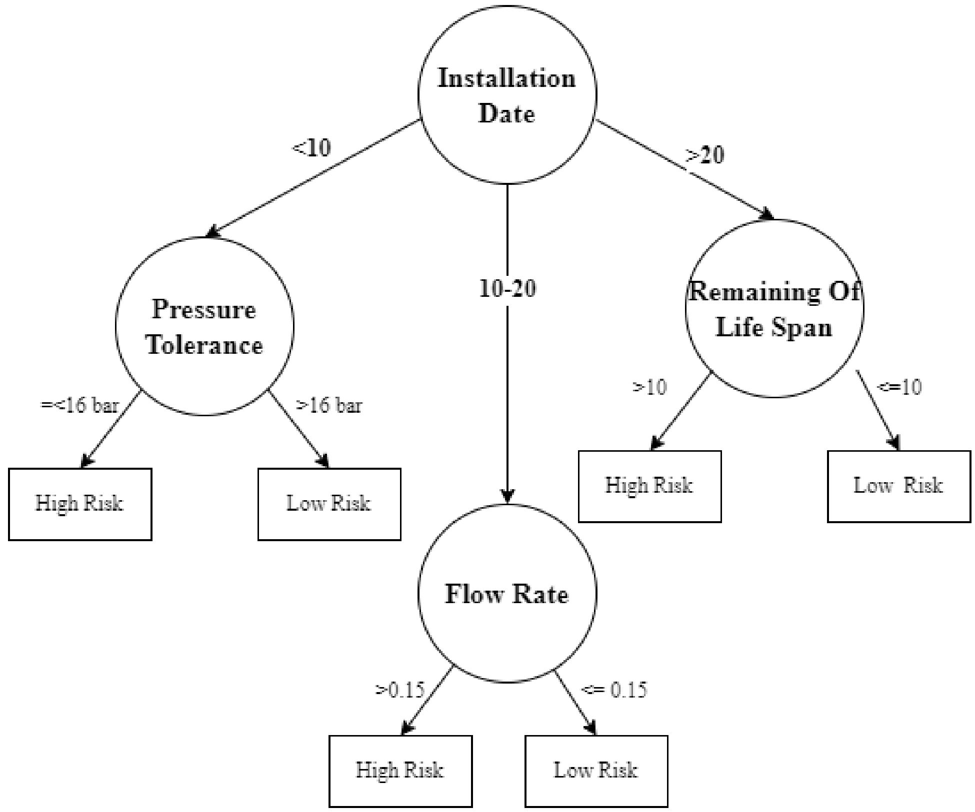


Figure 68: The Meters Factors Relations.

2- House Connection Prediction System:

The house connection prediction system was created based on four factors, analyzed by the algorithm to obtain the final decision: First, the pipe date of installation which presents the pipe age and has been considered root because it has the information gain highest value. Second, the pipe is able to withstand pressure; the less likely it is to need maintenance. Thirdly, the pipe's material—steel has a lower likelihood of breaking. Finally, the longer pipe is the greater chances of being exposed to malfunction. As shown in the figure

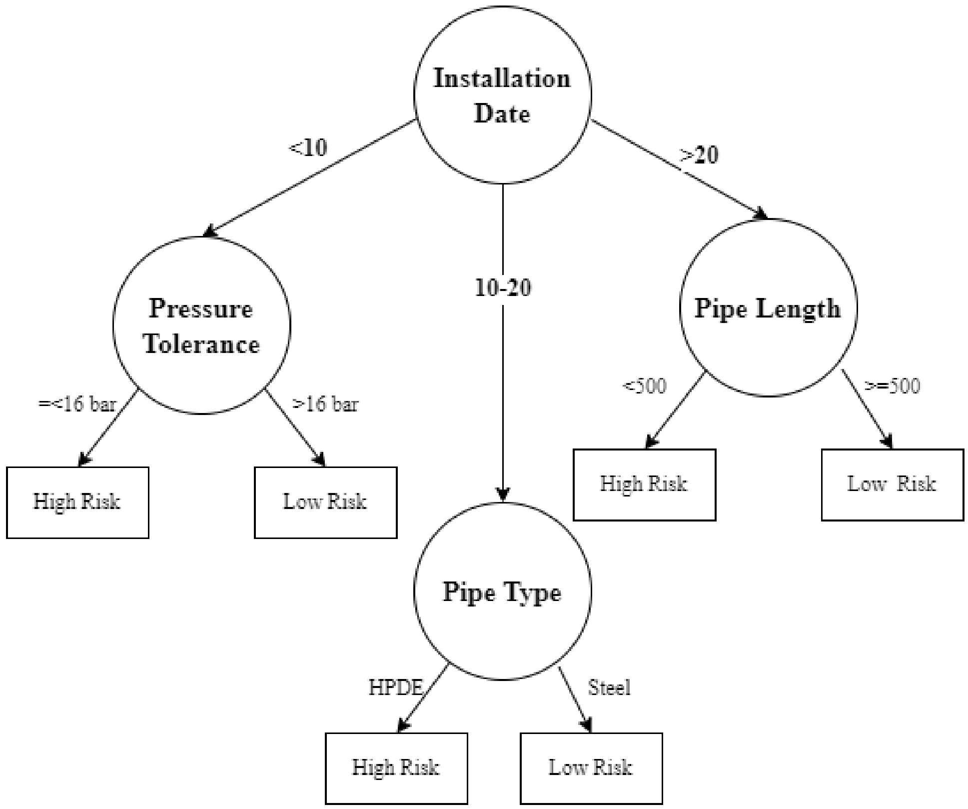


Figure 69: The House Connection Factors Relations.

Since the projection process factors may vary from month to month depending on the data saved - bearing in mind that the accumulation of data generates new factors such as the time difference between each maintenance and the next or the number of times of maintenance operations for the same component - the user can determine the file that will be relied upon in the projection process through (tools-> predict).

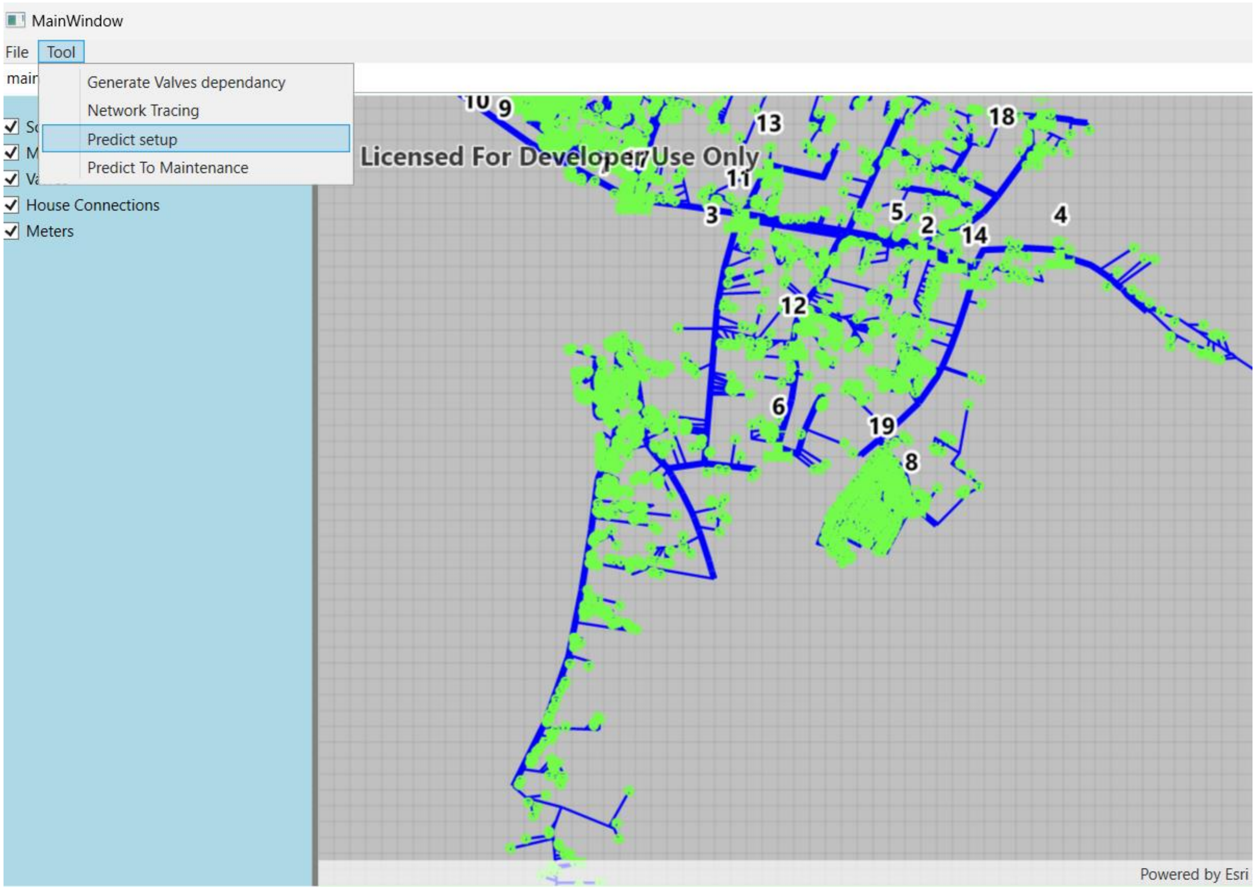


Figure 70: Predict Setup Tool.

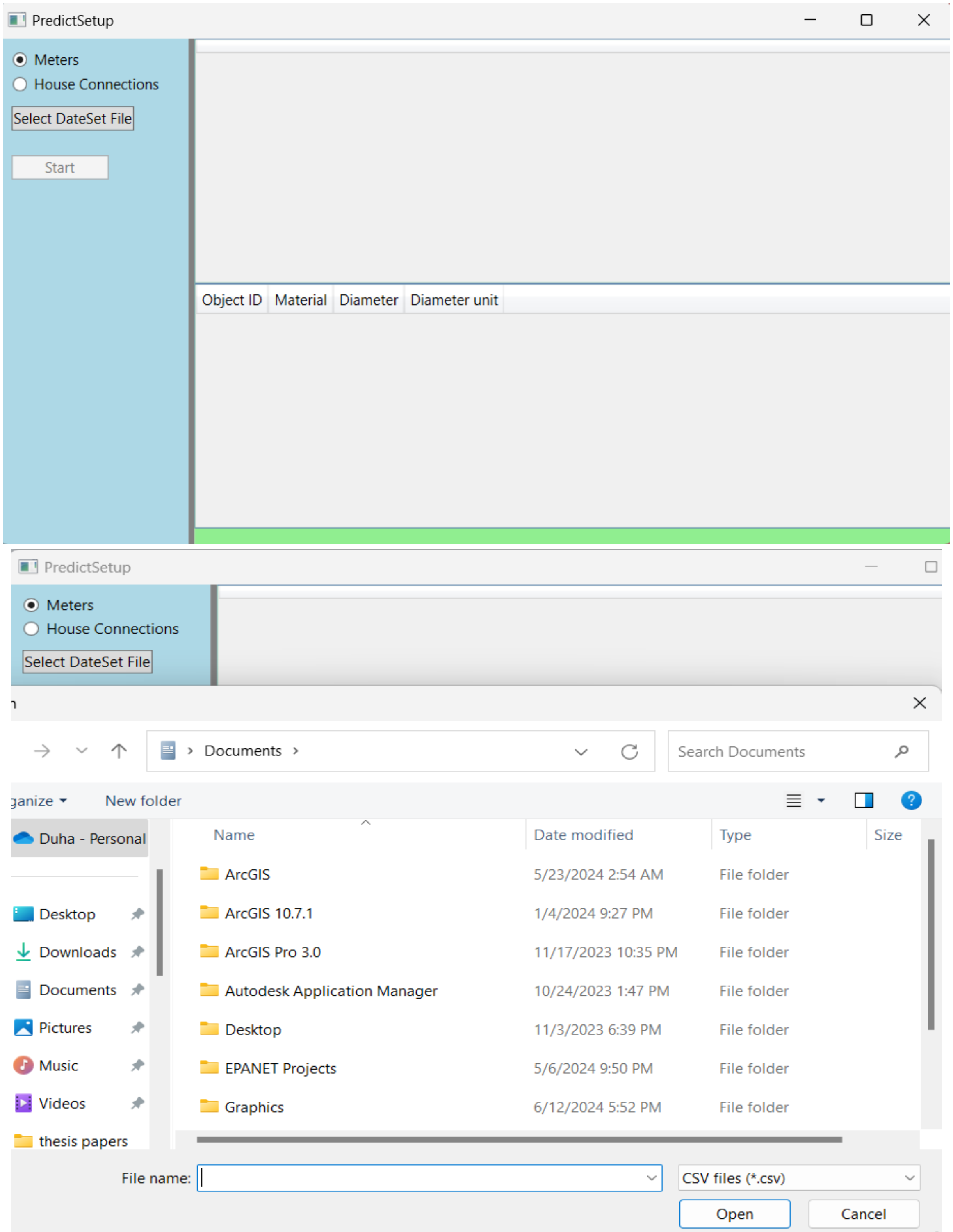


Figure 71: Predict Setup Window

Two prediction systems have been created in our app. We have created a new prediction tool (tool->predict to maintenance) as shown in the figure (64) that, once selected, starts the algorithm by analyzing meter data, house connection data and showing in a new pop-up meters and house connections data which is expected to need maintenance. As shown in the figure

Figure 72: Predict to Maintenance Tool

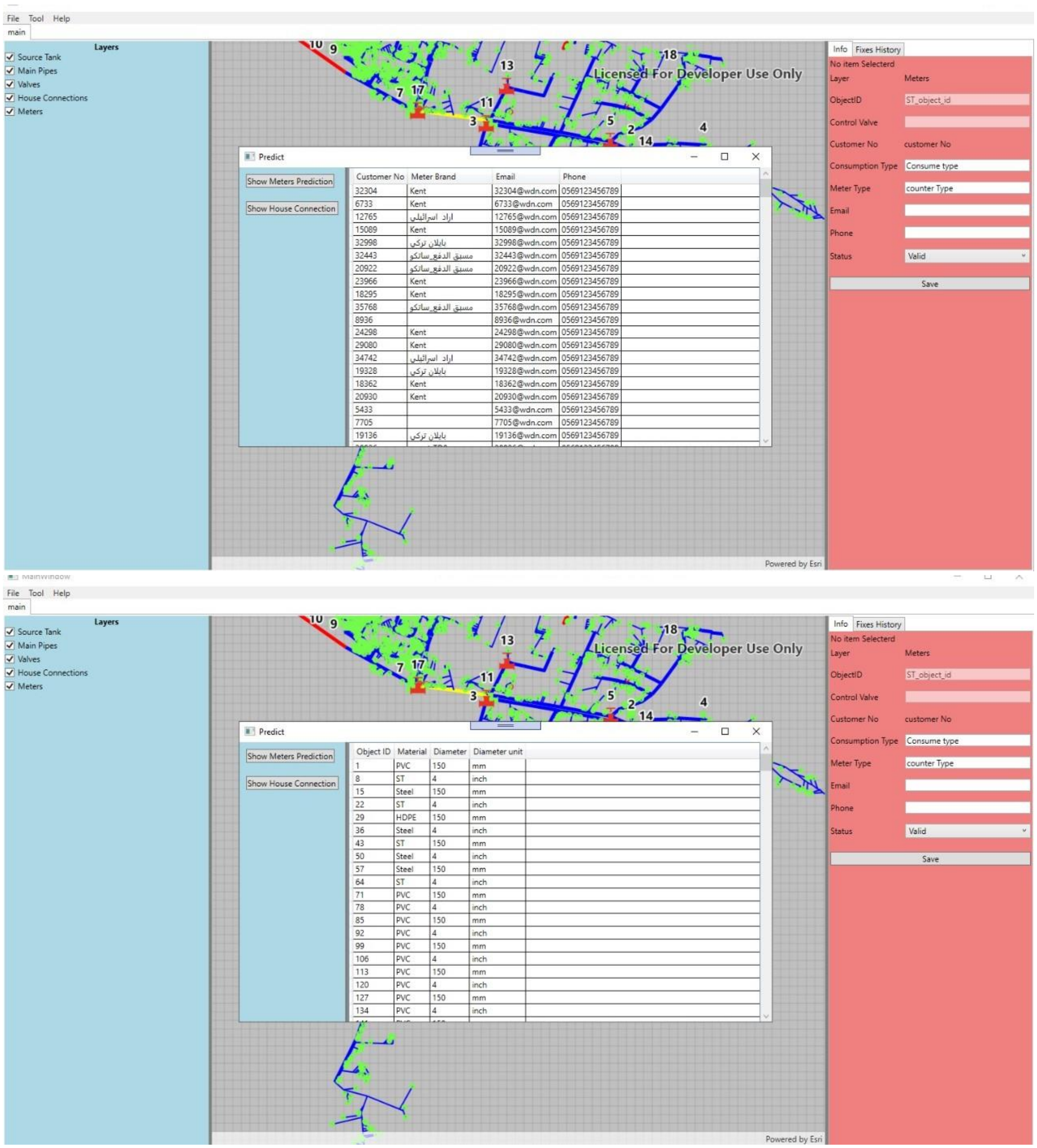


Figure 73: Predict to Maintenance Tool Results.

One factor that is supposed to be illustrated in the decision tree algorithm is the max depth value of the tree, The depth of the tree is determined by the experiment to reach the desired accuracy in the result of the algorithm and it is clear from the figuers that all that increased the depth value increased the resolution accuracy, but in return, the time needed to reach the decision is complicated by this in our application, making the process of determining the depth value depending on the user.

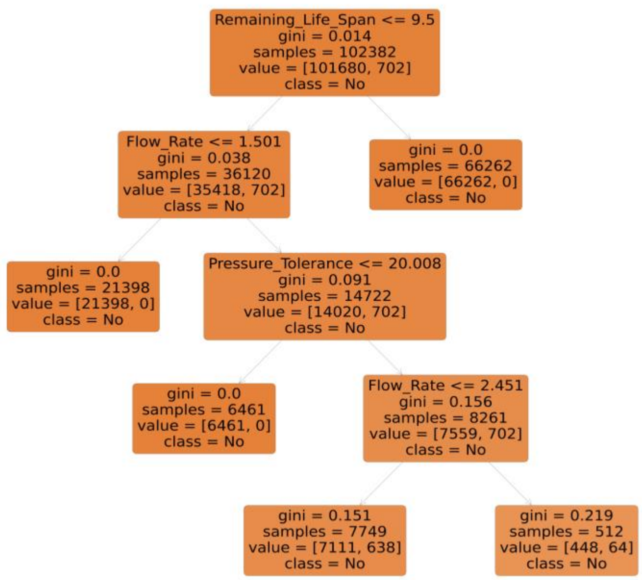


Figure 74: Meters Data tree with max depth = 5

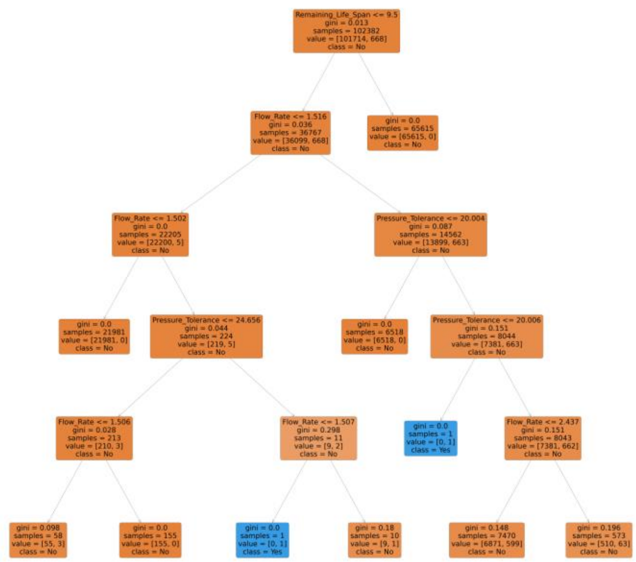


Figure 75: Meters Data tree with max depth = 6

Note that the depth difference led to different results, so when it was equal to 5, there was no yes decision, but when the depth became equal to 6, some decisions appeared yes.

The application also provides the advantage of searching the database of items maintained within a specified period of time so that the start date and end date of the period are determined and then pressing the search button.

Figure 76: Fixes History

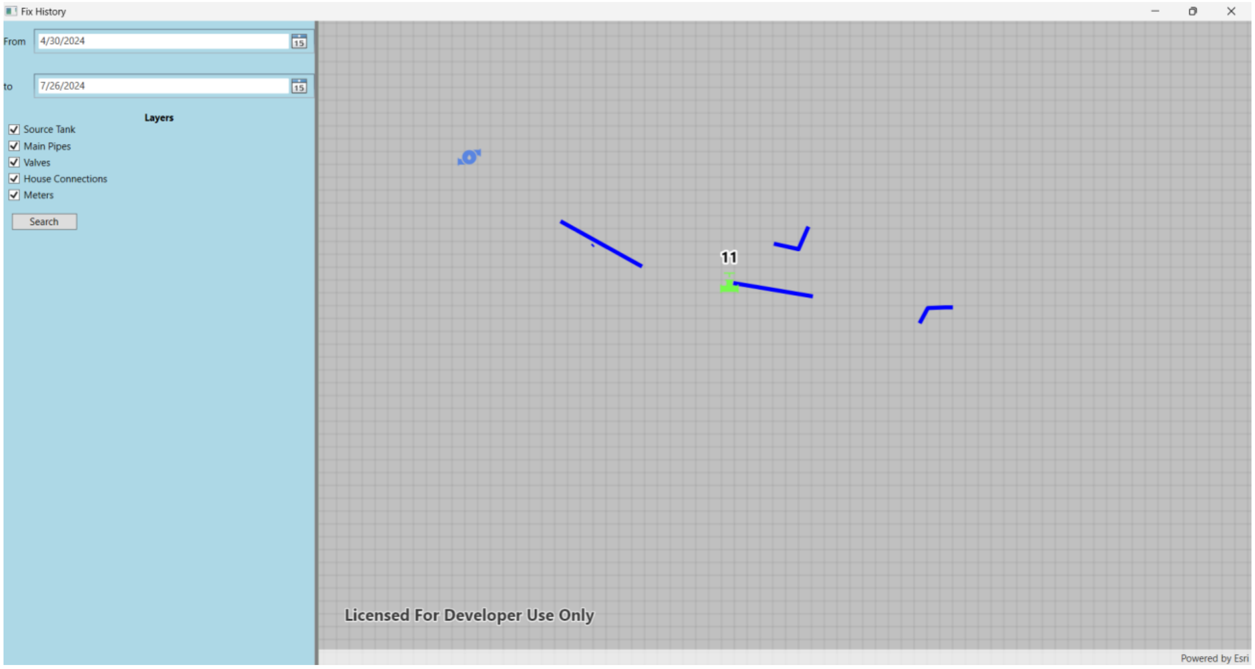


Figure 77: Searching in Fixes History.

As for the code snippet you will them in the appendix.

6.2: Testing and Validation:

In order to inspect the capabilities of the system, we have applied the functional test, as will be shown in detail

Functional testing is a crucial phase in the software testing lifecycle to ensure that a system operates according to its specified requirements.

The application-based prediction system using C# language and based on the decision tree algorithm aims to identify components of the network that may need maintenance in the near next period so that the system expects elements that need maintenance after analyzing historical data of past maintenance operations and devising as a mathematical equation that is applied to elements of the study area network and through the result of the equation that is limited to (0 or1) which expresses that the element needs maintenance or not. Through the testing phase we will test the application capabilities through several cases to estimate the accuracy of the algorithm in prediction.

The test phase includes a number of scenarios that represent changing the ratio of training data to testing data, apply the algorithm by using python language, the libraries were used: numpy, pandas, scikit-learn (DecisionTreeClassifier, train_test_split, accuracy_score) and then comparing accuracy and error rate for each scenario where :

- **Accuracy**: The ratio of correctly predicted instances to the total instances.
- **Error Rate**: The ratio of incorrectly predicted instances to the total instances.

Cross-validation involves multiple rounds of splitting the data into k subsets (folds) and systematically using each subset as a test set while the remaining subsets are used for training. The model is evaluated k times, and the results are averaged to provide a more stable estimate of performance.

The purpose of using cross-validation in machine learning is to evaluate the performance of a model in a more reliable and robust manner compared to a simple train-test split. Here are the key reasons for using cross-validation

1-More Accurate Performance Estimate:

Cross-validation gives a more accurate approximation of a model's performance by averaging the results across numerous training and testing cycles. This decreases variation and provides a more accurate overall evaluation.

2-Improved data utilization:

It enables the model to be trained and evaluated on several subsets of data. This guarantees that every data point is utilized for both training and testing, which is especially beneficial when the dataset is limited.

3-Reduced over fitting:

Cross-validation aids in the detection of overfitting by evaluating the model across different data splits. A model that performs well across all cross-validation folds is less likely to be overfitted for a single subset of the data.

4-Model Selection:

Cross-validation allows you to compare different models or hyperparameters. By comparing each model or parameter configuration across numerous folds, you can determine which one has the best average performance.

5-Reliable Performance Metrics:

It displays a distribution of performance parameters (such as accuracy, precision, and recall) across the folds. This aids in understanding the stability and variability of the model's performance. As shown in the table

Table 5: Testing and Validation Results.

Test Case	Testing Data Ratio	k-fold	Accuracy	Error Rate
1	0.3	10	0.9946	0.0054
2	0.3	5	0.9941	0.0059
3	0.15	10	0.9217	0.0783
4	0.15	5	0.9182	0.0818
5	0.09	10	0.9196	0.0804
6	0.09	5	0.9164	0.0836
7	0.05	10	0.9159	0.0841
8	0.05	5	0.9158	0.0842
9	0.01	10	0.9158	0.0842
10	0.01	5	0.9117	0.0883

From the results in the table, it is evident that the system achieves high accuracy in most test runs, with an accuracy range between 0.9117 and 0.9946. The error rate is relatively low, ranging from 0.0054 to 0.0883. The variation in results suggests that while the system performs well overall, there might be instances where prediction accuracy could be further improved.

Conclusion

PalGIS application demonstrates strong predictive capabilities with high accuracy and low error rates across multiple test runs. Future enhancements could focus on further reducing the variability in accuracy and improving the system's robustness.

Chapter Seven

3D Simulation

Water distribution network 3D simulation is an effective tool that improves planning, customer service, maintenance, operational effectiveness and training. Utilities and stakeholders can make well-informed decisions, optimize resource allocation and guarantee a sustainable and dependable water supply by utilizing the comprehensive and interactive network representation that it offers.

To improve maintenance and emergency response, as well as visualization and analysis, we use the ArcScene software to create a 3D model of our network by following these steps:

1- Import Data:

We import industrial zone data which we represent in ArcMap previously.

2- Semiology all Features:

Each feature was given a clear symbol, indicating its meaning and function.

3- Classify Main Pipe Layer and House Connection Layer:

Pipe layers (main and house connections) are classified into several colors according to the size and diameter of each pipe so that each pipe is represented by a specific color according to its size and diameter

4- Convert Contour Layer into Raster:

As we did earlier in Chapter IV, but this time in ArcScene, We scraped a part of the industrial zone.as shown in the figure.

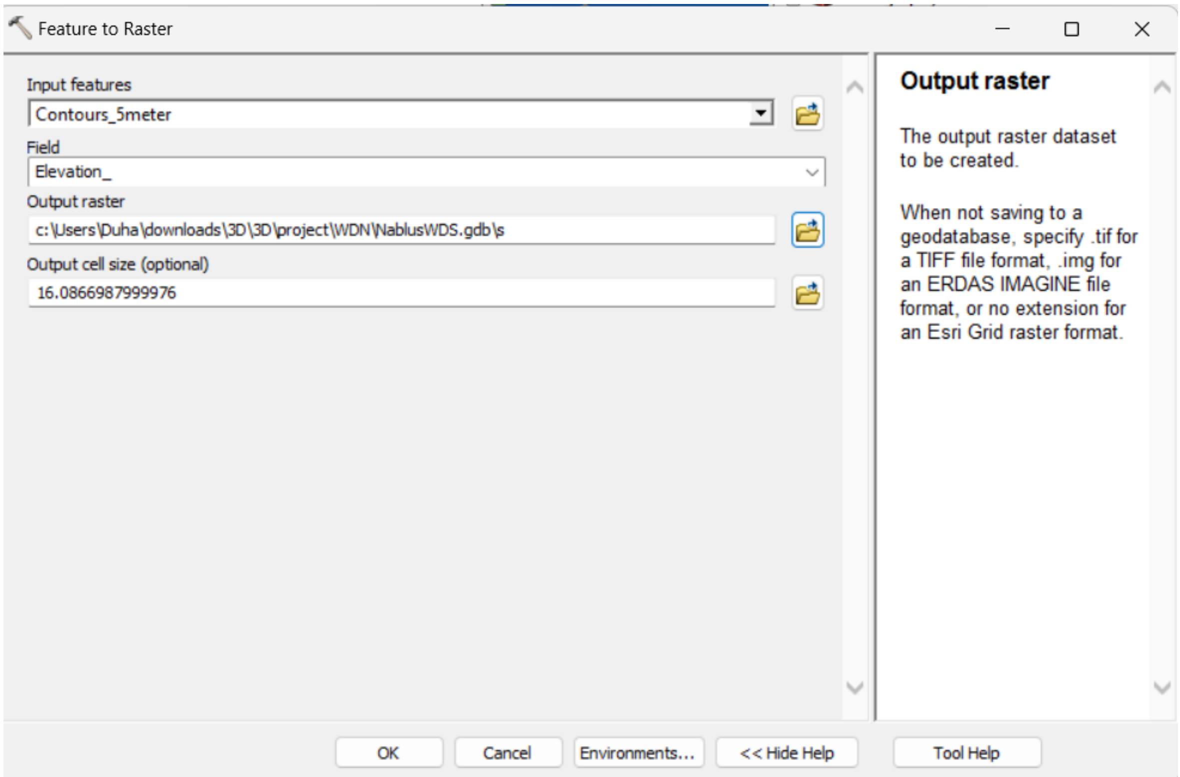


Figure 78: Converting the Industrial Zone Contour Layer into Raster.

5- The layer turned into a raster, thus we can activate the Base heights property for the layer to give the area altitude milestones, as shown in the figure

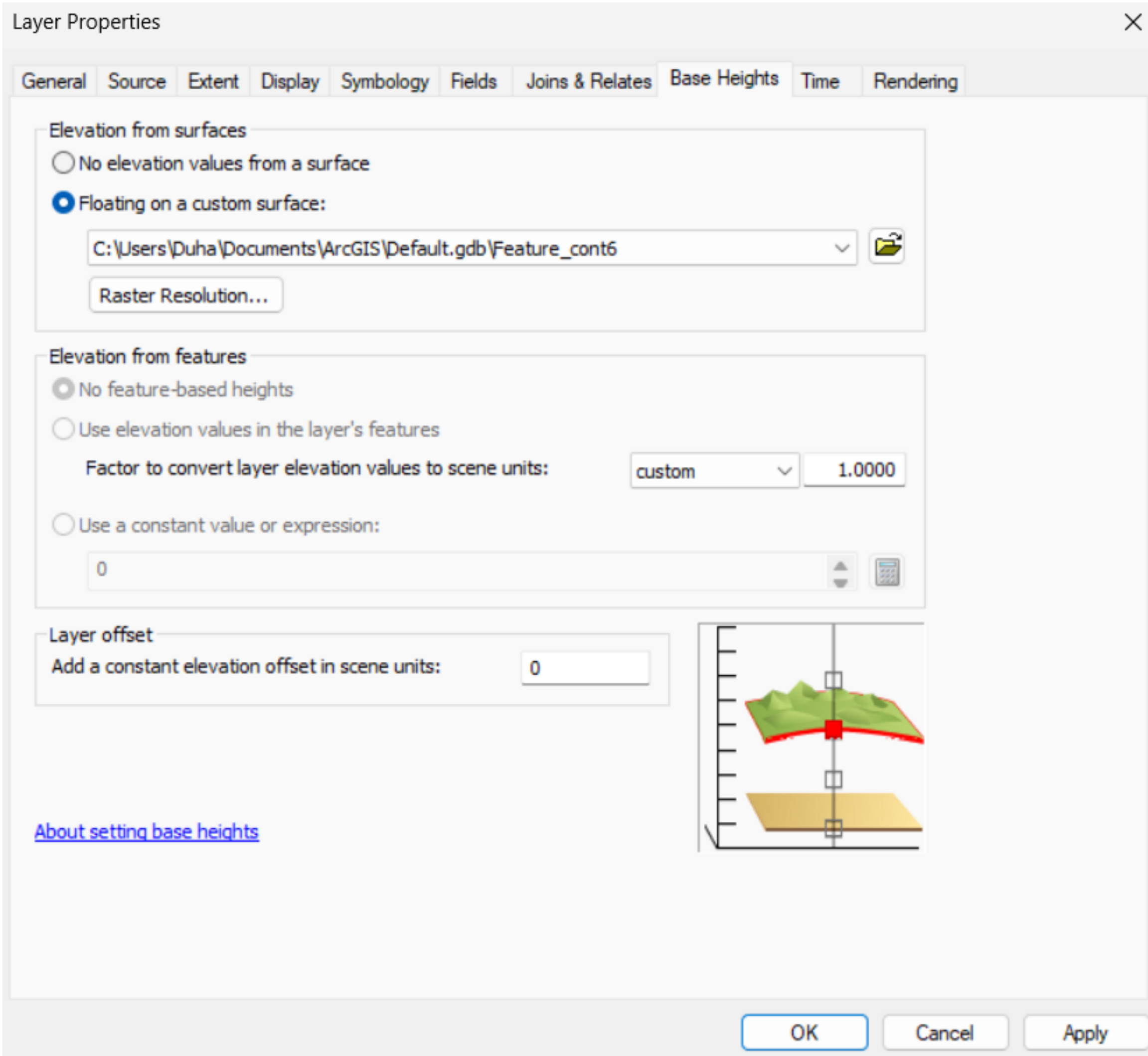


Figure 79: Activate Base Heights Property for the Raster

6- To Obtain Altitude Milestones for all Components, we have to activate Base Heights Property to all Layers, in order to integrate them with the raster.as shown in the following figures

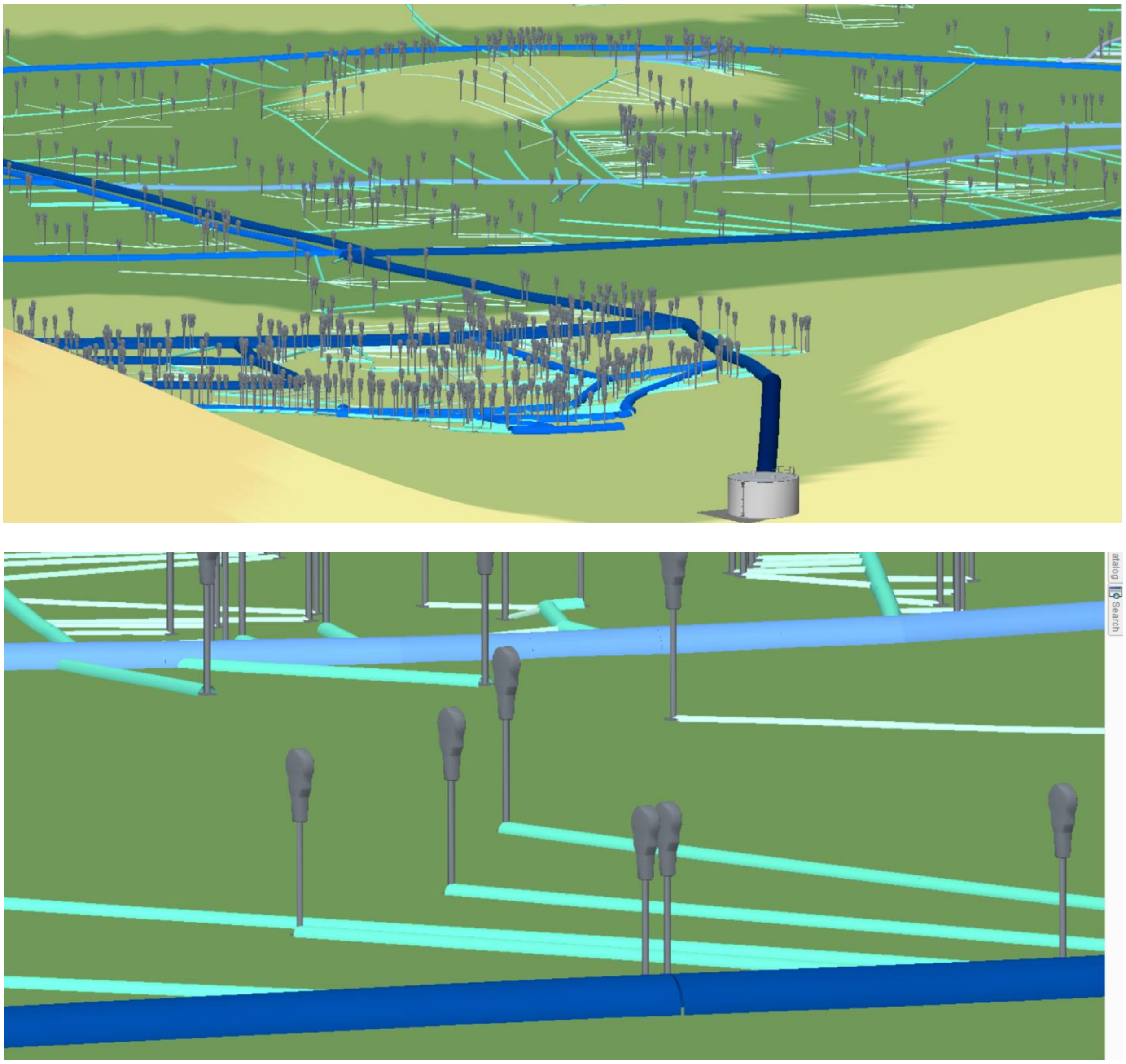


Figure 80: Network 3D Simulation.

Chapter Eight

Conclusion & Recommendations

Conclusion

In this thesis, we have developed a number of hypotheses that need to be validated in a number of consecutive phases, including processing several types of data in a number of different ways until we have reached valid data for use in analyses and algorithms to eventually reach a coherent network whose components have been correctly linked. Algorithms have been applied to regular data and spatial data, and these data have been represented in two and three dimensions to several objectives, such as facilitating the planning process, understanding complex data and helping to assess data validity.

We were finally able to present a practical and reliable solution that could validate the theories outlined at the beginning of this study. This solution contributes to the decision-making process and improves the quality of decisions in several areas, including saving time, effort and reduce water loss through its ability to analyze the network and propose the controlling valve for any physical component in the network if it breakdown. Its ability to predict house connection pipes that need maintenance through a predicting system relied on data mining algorithms and historical data for the maintenance of such pipes, and finally anticipates meters that need maintenance also relied on appropriate data mining algorithms and historical data for meter maintenance.

All of these results are displayed in a desktop application capable of displaying the network on the map, allowing the ability to control, monitor and modify the status of its components, as well as to archive maintenance data in the GIS format and help determine the approximate cost for each maintenance process. The application also shows on the network meters and house connections expected to need maintenance according to the decision of the prediction system that was established based on data mining algorithms, The application also allows a live simulation of the network analysis process to select the controlled valve in the event of a break, identify the areas to which water access will be interrupted and inform consumers present in these areas of an emergency occurrence leading to the interruption of water services.

This app is capable of receiving and handling the data of any region, whatever its size, is easy to use and clear even for non-GIS specialists, and no need to be connect with the Internet.

Recommendation

We determined the following paths for the future:

1- Explore our solution in other WDNs, in Palestine, since the app is able to handle the data of any region.

2- The relevant institutions, which stand to gain from GIS facilities, give careful consideration to the storage and conservation of spatial data.

3- Using other algorithms in the analysis process and perhaps in a timely manner, there is properly archived data to take advantage of new factors such as the time difference between maintenance operations.

4- Documenting maintenance in this format makes it easier to know what features have been maintained before so as to take it into account and make data mining operations more accurate.

5- More broadly familiarize organizations with geographical information systems, their capabilities and advantages, and the recruitment of specialists in the field.

6- Enhancing the ability to take into account more spatial variables; data, layers and factors.

7- Take into account important variables affecting physical network components, such as life span and the quality of the manufactured material.

References

A. Yazdani, P. Jeffrey, Water distribution system vulnerability analysis using weighted and directed network models, Water Resour. Res. 48 (6) (2012), https://doi.org/10.1029/2012WR011897

Acharya, D., & Mukhopadhyay, A. (2024). A comprehensive review of machine learning techniques for multi-omics data integration: challenges and applications in precision oncology. Briefings in Functional Genomics, elae013.

Akhiat, Y., Touchanti, K., Zinedine, A., & Chahhou, M. (2024). IDS-EFS: Ensemble feature selection-based method for intrusion detection system. Multimedia Tools and Applications, 83(5), 12917-12937.

Ali, T. H., & Saleh, D. M. (2022). Proposed hybrid method for wavelet shrinkage with robust multiple linear regression model: With simulation study. Qalaai Zanist Journal, 7(1), 920-937.

Alokhina, A., & van den Brand, J. (2024). Fully dynamic shortest path reporting against an adaptive adversary. In Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA) (pp. 3027-3039). Society for Industrial and Applied Mathematics.

Arsene, D., Predescu, A., Truica, C. O., Apostol, E. S., & Mocanu, M. (2023). Decision Support Strategies for Household Water Consumption Behaviors Based on Advanced Recommender Systems. Water 2023, 15, 2550.

Arshad, J., Aziz, M., Al-Huqail, A. A., Husnain, M., Rehman, A. U., & Shafiq, M. (2022). Implementation of a LoRaWAN based smart agriculture decision support system for optimum crop yield. Sustainability, 14(2), 827.

Aruna, R., Kushwah, V. S., Praveen, S. P., Pradhan, R., Chinchawade, A. J., Asaad, R. R., & Kumar, R. L. (2024). Coalescing novel QoS routing with fault tolerance for improving QoS parameters in wireless Ad-Hoc network using craft protocol. Wireless Networks, 30(2), 711-735.

B. Jiang, C. Liu, "Street-based topological representations and analyses for predicting traffic flow in GIS," International Journal of Geographical Information Science, vol. Volume 23, no. Issue 9, pp. Pages 1119-1137 , Augest 2007.

Badeenezhad, A., Darabi, K., Heydari, M., Amrane, A., Ghelichi-Ghojogh, M., Parseh, I., ... & Javanmardi, P. (2023). Temporal distribution and zoning of nitrate and fluoride concentrations in Behbahan drinking water distribution network and health risk assessment by using sensitivity analysis and Monte Carlo simulation. International Journal of Environmental Analytical Chemistry, 103(13), 3163-3180.

Bentert, M., Nichterlein, A., Renken, M., & Zschoche, P. (2023).Using a Geometric Lens to Find-Disjoint Shortest Paths. SIAM Journal on Discrete Mathematics, 37(3), 1674-1703.

Bilaşco, Ş., & Man, T. C. (2024). GIS-Based Spatial Analysis Model for Assessing Impact and Cumulative Risk in Road Traffic Accidents via Analytic Hierarchy Process (AHP)—Case Study: Romania. Applied Sciences, 14(6), 2643.

Birdi, P. K., Kale, K., & Ajith, V. (2021). Spatial data mining of agricultural land area using multi-spectral remote-sensed images. In Data intelligence and cognitive informatics: Proceedings of ICDICI 2020 (pp. 457-469). Springer Singapore.

Borrero, J. D., & Mariscal, J. (2022). A case study of a digital data platform for the agricultural sector: A valuable decision support system for small farmers. Agriculture, 12(6), 767.

Burton, A. L. (2021). OLS (Linear) regression. The encyclopedia of research methods in criminology and Criminal Justice, 2, 509-514.

Candido, L. A., Coêlho, G. A. G., de Moraes, M. M. G. A., & Florêncio, L. (2022). Review of decision support systems and allocation models for integrated water resources management focusing on joint water quantity-quality. Journal of Water Resources Planning and Management, 148(2), 03121001.

Chechik, S., & Zhang, T. (2023).Faster deterministic worst-case fully dynamic all-pairs shortest paths via decremental hop-restricted shortest paths.In Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA) (pp. 87-99).Society for Industrial and Applied Mathematics.

Chen, H., Wu, C., Wu, M., Xu, B., & Fan, P. (2024). Survey of Connectivity Restoration in 3D Wireless Ad Hoc/Sensor Networks. International Journal of Distributed Sensor Networks, 2024.

Chou, J. S., & Chen, K. E. (2024). Optimizing investment portfolios with a sequential ensemble of decision tree-based models and the FBI algorithm for efficient financial analysis. Applied Soft Computing, 158, 111550.

Conejos Fuertes, P., Martínez Alzamora, F., Hervás Carot, M., & Alonso Campos, J. C. (2020). Building and exploiting a Digital Twin for the management of drinking water distribution networks. Urban Water Journal, 17(8), 704-713.

Dandrifosse, S., Jago, A., Huart, J. P., Michaud, V., Planchon, V., & Rosillon, D. (2024). Automatic quality control of weather data for timely decisions in agriculture. Smart Agricultural Technology, 8, 100445.

Deng, Baosong, et al, "Visualization of vector data on global scale terrain," in Proceedings International Conference on Computer Science and Electronics Engineering, 2013.

Deng, Baosong, et al, "Visualization of vector data on global scale terrain," in Proceedings International Conference on Computer Science and Electronics Engineering, 2013.

Desai, S., & Rajapara, G. (2021). Leakage Optimization of Water Distribution Network Using Artificial Intelligence. In Advanced Modelling and Innovations in Water Resources Engineering: Select Proceedings of AMIWRE 2021 (pp. 265-284). Singapore: Springer Singapore.

Dhouib, S. (2024). Innovative method to solve the minimum spanning tree problem: The Dhouib-Matrix-MSTP (DM-MSTP). Results in Control and Optimization, 14, 100359.

Diao, K., Sweetapple, C., Farmani, R., Fu, G., Ward, S., & Butler, D. (2016). Global resilience analysis of water distribution systems. Water research, 106, 383-393.

Dorotea, T., Riuzzi, G., Franzago, E., Posen, P., Tavornpanich, S., Di Lorenzo, A.,... & Ferrè, N. (2023). A Scoping Review on GIS Technologies Applied to Farmed Fish Health Management. Animals, 13(22), 3525.

Duarte, L., Teodoro, A. C., & Gonçalves, H. (2022). Evaluation of spatial thinking ability based on exposure to Geographical Information Systems (GIS) concepts in the context of higher education. ISPRS International Journal of Geo-Information, 11(8), 417.

El-Zahab, S., & Zayed, T. (2019). Leak detection in water distribution networks: an introductory overview. Smart Water, 4(1), 1-23.

Embury, J., Nara, A., Rey, S., Tsou, M. H., & Ghanipoor Machiani, S. (2024). Detecting synthetic population bias using a spatially-oriented framework and independent validation data. International Journal of Geographical Information Science, 1-27.

F. Ruggeri, M. Crapper, J. R. Snyder and J. Crow, "A GIS-based assessment of the Byzantine water supply system of Constantinople," Water Science and Technology: Water Supply IWA, vol. Volume 17, no. Issue 3, April 2017.

Fang, Z., Wang, Y., Wang, Z., Zhang, J., Ji, X., & Zhang, Y. (2024). Mammil: Multiple instance learning for whole slide images with state space models. arXiv preprint arXiv:2403.05160.

Fang, Z., Yue, P., Zhang, M., Xie, J., Wu, D., & Jiang, L. (2023). A service-oriented collaborative approach to disaster decision support by integrating geospatial resources and task chain. International Journal of Applied Earth Observation and Geoinformation, 117, 103217

Faramondi, L., Oliva, G., & Setola, R. (2020). Multi-criteria node criticality assessment framework for critical infrastructure networks. International Journal of Critical Infrastructure Protection, 28, 100338.

Gao, J., Li, P., Laghari, A. A., Srivastava, G., Gadekallu, T. R., Abbas, S., & Zhang, J. (2024). Incomplete multiview clustering via semidiscrete optimal transport for multimedia data mining in IoT. ACM Transactions on Multimedia Computing, Communications and Applications, 20(6), 1-20.

Ghavami, S. M. (2019). Multi-criteria spatial decision support system for identifying strategic roads in disaster situations. International Journal of Critical Infrastructure Protection, 24, 23-36.

Gholami, H., Mohamadifar, A., & Collins, A. L. (2020). Spatial mapping of the provenance of storm dust: Application of data mining and ensemble modelling. Atmospheric Research, 233, 104716.

Giustolisi, O., & Savic, D. (2010). Identification of segments and optimal isolation valve system design in water distribution networks. Urban Water Journal, 7(1), 1-15.

Hara, T., & Sasabe, M. (2024). Practicality of in-kernel/user-space packet processing empowered by lightweight neural network and decision tree. Computer Networks, 240, 110188.

Hidayat, I. (2023). Decision Support System at PDAM Tirta Sari City of Binjai Using the Decision Tree Method in Determining Maintenance Actions for Clean Water Distribution Networks. Journal of Artificial Intelligence and Engineering Applications (JAIEA), 3(1), 454-462.

Hu, J., Yang, P., Li, Q., Wang, M., Feng, J., Gao, Z., & Liu, J. (2024). Evaluation of Groundwater Vulnerability of Yishu River Basin Based on DRASTIC-GIS Model. Water, 16(3), 429.

Hwang, Sungsoon. (2013). Placing GIS in sustainability education. Journal of Geography in Higher Education. 37. 276-291. 10.1080/03098265.2013.769090.

James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). Linear regression. In An introduction to statistical learning: With applications in python (pp. 69-134). Cham: Springer International Publishing.

Jerie, S., Shabani, T., Mudyazhezha, O. C., & Shabani, T. (2024).A review towards developing a hierarchical model for sustainable hospital solid waste management in rural areas of Zimbabwe.Environmental Monitoring and Assessment, 196(3), 308.

Jiang, Q. (2024). Communication network security situation analysis based on time series data mining technology. Open Computer Science, 14(1), 20230104.

Karadirek, I. E., & Aydin, M. E. (2022). Water Losses Management in Urban Water Distribution Systems. In Water and Wastewater Management: Global Problems and Measures (pp. 53-65). Cham: Springer International Publishing.

Khadri, S.F.R ,Chaitanya Pande, "Urban Water Supply Systems - A Case Study On Water Network Distribution in Chalisgaon City in Dhule District Maharashtra Using Remote Sensing & GIS Techniques," , NAGPUR, 2014.

Kumar, K. P., Kumar, B. H., & Bhuvanesh, A. (2024). Spectral clustering algorithm based web mining and quadratic support vector machine for learning style prediction in E-learning platform. Measurement: Sensors, 31, 100962.

LeJeune, D., Javadi, H., & Baraniuk, R. (2020, June). The implicit regularization of ordinary least squares ensembles. In International Conference on Artificial Intelligence and Statistics (pp. 3525-3535). PMLR.

Lemos, Diogo, et al, "Environmental assessment of an urban water system," Elsevier, vol. Volume 54, pp. Pages 157–165, September 2013

Lumia, G., Modica, G., & Cushman, S. (2024). Using simulation modeling to demonstrate the performance of graph theory metrics and connectivity algorithms. Journal of Environmental Management, 352, 120073.

Lynn E Johnson, Geographic information system in water resource engineering. London , 209.

M., M. Sun, and J.C.S. Lui, DroidRay Zheng, "a security evaluation system for customized android firmwares," in Proceedings of the 9th ACM symposium on Information, computer and communications security, Japan, 2014, pp. Pages 471-482.

Maguire, David J, "An overview and definition of GIS," in Geographical information systems: Principles and applications, 1991.

Mahanty, C., Kumar, R., & Mishra, B. K. (2021). Analyses the effects of COVID-19 outbreak on human sexual behaviour using ordinary least-squares based multivariate logistic regression. Quality & quantity, 55(4), 1239-1259.

Malczewski, Jacek, "GIS-based land-use suitability analysis: a critical overview," Elsevier, vol. Volume 62, no. Issue 1, pp. Pages 3–65, July 2004.

Maurya, S. K., Singh, D., & Maurya, A. K. (2024). Deceptive opinion spam detection using feature reduction techniques. International Journal of System Assurance Engineering and Management, 15(3), 1210-1230.

Merlin, M. L., & Chen, Y. (2021). Analysis of the factors affecting electricity consumption in DR Congo using fully modified ordinary least square (FMOLS), dynamic ordinary least square (DOLS) and canonical cointegrating regression (CCR) estimation approach. Energy, 232, 121025.

Mohamed, E. S., Belal, A. A., Abd-Elmabod, S. K., El-Shirbeny, M. A., Gad, A., & Zahran, M. B. (2021). Smart farming for improving agricultural management. The Egyptian Journal of Remote Sensing and Space Science, 24(3), 971-981.

Mohammadpour, N., Jahanishakib, F., & Asadolahi, Z. (2023). Modeling the Supply of Habitat Service and Spatial Data mining of Hotspots in Arid Ecosystems. Journal of Environmental Studies, 49(1), 33-49.

Mokashi, M. S., Perdita Okeke, and Uma Mohan, "Study on the Use of Geographic Information Systems (GIS) for Effective Transport Planning for Transport for London (TfL)." Proceedings of the International Conference on Data Engineering and Communication Tec," Springer, vol. volume 469, pp. pp 719-728, 2017.

Morabit, M., Desaulniers, G., & Lodi, A. (2023).Machine-learning–based arc selection for constrained shortest path problems in column generation.INFORMS Journal on Optimization, 5(2), 191-210.

Nair, B. B., Mohandas, V. P., & Sakthivel, N. R. (2010). A genetic algorithm optimized decision tree-SVM based stock market trend prediction system. International journal on computer science and engineering, 2(9), 2981-2988.

Nemati, Z., Mohammadi, A., Bayat, A., & Mirzaei, A. (2024). Metaheuristic and Data Mining Algorithms-based Feature Selection Approach for Anomaly Detection. IETE Journal of Research, 1-15.

Olgu, K., Kenter, T., Nunez-Yanez, J., & Mcintosh-Smith, S. (2024, April). Optimisation and Evaluation of Breadth First Search with oneAPI/SYCL on Intel FPGAs: from Describing Algorithms to Describing Architectures. In Proceedings of the 12th International Workshop on OpenCL and SYCL (pp. 1-11).

Patel, P., Sivaiah, B., Patel, R., & Choudhary, R. (2024). Association Rule Mining for Healthcare Data Analysis. In Computational Intelligence in Healthcare Informatics (pp. 127-139). Singapore: Springer Nature Singapore.

Pérez-Padillo, J. (2023). Intelligent management of water distribution systems.

Perkins, J. (2022). Object Recognition and Virtual Retail Algorithms, Metaverse and Immersive Technologies, and Simulation Modeling and Spatial Data Acquisition Tools in Extended Reality Environments. Analysis and Metaphysics, (21), 142-158.

Piazza, S., Sambito, M., & Freni, G. (2023). Analysis of optimal sensor placement in looped water distribution networks using different water quality models. Water, 15(3), 559.

Pouresmaeil, H., Faramarz, M. G., ZamaniKherad, M., Gheibi, M., Fathollahi-Fard, A. M., Behzadian, K., & Tian, G. (2022). A decision support system for coagulation and flocculation processes using the adaptive neuro-fuzzy inference system. International Journal of Environmental Science and Technology, 19(10), 10363-10374.

Prihozhy, A. A., & Karasik, O. N. (2023).Advanced heterogeneous block-parallel all-pairs shortest path algorithm.

Priyadi, A., Pujiantara, M., & Purnomo, M. H. (2024). Parallel Hybrid Particle Swarm-Grey Wolf Algorithms for Optimal Load-Shedding in An Isolated Network. Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI, 13(1).

Prof.Dr.yee leung, InItelligent Spatial Decision Support Systems. hong kong: Springer Science & Business Media, 2012.

Quan, X. X., Yang, J. F., & Luo, Z. (2024). RETRACTED ARTICLE: Models in digital business and economic forecasting based on big data IoT data visualization technology. Personal and Ubiquitous Computing, 28(Suppl 1), 11-11.

Rajan, R., Rajest, S., & Singh, B. (2021). Spatial data mining methods databases and statistics point of views. Innovations in Information and Communication Technology Series, 103-109.

Ramos, H. M., Morani, M. C., Carravetta, A., Fecarrotta, O., Adeyeye, K., López-Jiménez, P. A., & Pérez-Sánchez, M. (2022). New challenges towards smart systems' efficiency by digital twin in water distribution networks. Water, 14(8), 1304.

Reger, Cian Michael, "Bringing GIS to a Small Community Water System," 2017.

Remembering Dr. Roger Tomlinson, the inventor of computer-based GIS.: UCGIS, 2014.

Rezk, A. A., & Hendawy, M. (2024). Informative cartographic communication: A framework to evaluate the effects of map types on users' interpretation of COVID-19 geovisualizations. Cartography and Geographic Information Science, 51(2), 257-274.

Rupert, Maya, Amjad Rattrout, Salima Hassas, "The web from a complex adaptive systems perspective," Journal of Computer and System Sciences , vol. 74, no. 2, pp. 133-145, march 2008.

Saggi, M. K., & Jain, S. (2022). A Survey Towards Decision Support System on Smart Irrigation Scheduling Using Machine Learning approaches. Archives of Computational Methods in Engineering, 29(6), 4455-4478.

Santos, S., Monteiro, F. A., Coutinho, B. C., & Omar, Y. (2023). Shortest path finding in quantum networks with quasi-linear complexity. IEEE Access, 11, 7180-7194.

Selvarajan, S., Manoharan, H., Goel, S., Akili, C. P., Murugesan, S., & Joshi, V. (2024). SCMC: Smart city measurement and control process for data security with data mining algorithms. Measurement: Sensors, 31, 100980.

Selvida, D., Pulungan, A. F., & Elveny, M. (2024, April). Data mining decision tree algorithm C4. 5 classification of student personality characteristics. In AIP Conference Proceedings (Vol. 2987, No. 1). AIP Publishing.

Sharma, V.; You, I.; Pau, G.; Collotta, M.; Lim, J.D.; Kim, J.N. LoRaWAN-based energy-efficient surveillance by drones for intelligent transportation systems. Energies 2018, 11, 573.

Shirowzhan, S., Lim, S., Trinder, J., Li, H., & Sepasgozar, S. M. (2020). Data mining for recognition of spatial distribution patterns of building heights using airborne lidar data. Advanced Engineering Informatics, 43, 101033.

Singh, R., Kumar, A., Lather, V., Sharma, R., & Pandita, D. (2024). Identification of novel signal of Raynaud's phenomenon with Calcitonin Gene-Related Peptide (CGRP) antagonists using data mining algorithms and network pharmacological approaches. Expert opinion on drug safety, 23(2), 231-238.

Sitzenfrei, R., Hajibabaei, M., Hesarkazzazi, S., & Diao, K. (2023). Dual graph characteristics of water distribution networks—how optimal are design solutions?. Complex & Intelligent Systems, 9(1), 147-160.

Szymon SzewraNski, Jan Kazak, Marta Sylla , Małgorzata Swiąder, "Spatial Data Analysis with the Use of ArcGIS and Tableau Systems," Springer, pp. pp 337-349, 2017.

Tao, H., Abba, S. I., Al-Areeq, A. M., Tangang, F., Samantaray, S., Sahoo, A., ... & Yaseen, Z. M. (2024). Hybridized artificial intelligence models with nature-inspired algorithms for river flow modeling: A comprehensive review, assessment, and possible future research directions. Engineering Applications of Artificial Intelligence, 129, 107559.

Tornyeviadzi, H. M., Mohammed, H., & Seidu, R. (2022). Dynamic segment criticality analysis: A precursor to scheduling of maintenance routines in water distribution networks. Alexandria Engineering Journal, 61(12), 9261-9272.

Unhelkar, B., Joshi, S., Sharma, M., Prakash, S., Mani, A. K., & Prasad, M. (2022). Enhancing supply chain performance using RFID technology and decision support systems in the industry 4.0–A systematic literature review. International Journal of Information Management Data Insights, 2(2), 100084.

Ünverdi, A. B. (2021). Elektrik dağıtım sisteminde cbs uygulamaları ve gerçekleşen kesintilerin cbs entegrasyonu ile raporlanması (Master's thesis, Ondokuz Mayıs Üniversitesi, Lisansüstü Eğitim Enstitüsü).

V.Maniezzo, I.Mendes, M. Paruccini, "Decision Support for Siting problems," Elsevier, vol. Volume 23, no. Issue 3, pp. Pages 273–284, July 1998.

Wahyuni, W. I., Yusfrizal, Y., & Fatmaira, Z. (2024). Data Mining Correlation Of Student Discipline With Level Of Achievement Using Apriori Method. Journal of Engineering, Technology and Computing (JETCom), 3(1), 43-50.

Walliser, F. (2024). Exploring Edge Orientation Algorithms in the Static and Dynamic Case.

Wang, L., Hou, H., & Weng, J. (2020). Ordinary least squares modelling of urban heat island intensity based on landscape composition and configuration: A comparative study among three megacities along the Yangtze River. Sustainable Cities and Society, 62, 102381.

Wang, M., Deng, Y., Won, J., & Cheng, J. C. (2019). An integrated underground utility management and decision support based on BIM and GIS. Automation in Construction, 107, 102931.

Wederhake, L., Schlephorst, S., & Zyprian, F. (2022). Make or buy: IT-based decision support for grid imbalance settlement in smarter electricity networks. Energy Informatics, 5(1), 1-27.

Workneh, Z. G., Gebremedhin, E. K., Beyene, N. A., & Nigatu, N. M. (2023). Hydraulic performance evaluation of Yeka Abado supply zone of Addis Ababa's water distribution network. AQUA—Water Infrastructure, Ecosystems and Society, 72(9), 1697-1706.

Wu, W., Hayashi, T., Haruyasu, K., & Tang, L. (2023). Exact algorithms based on a constrained shortest path model for robust serial-batch and parallel-batch scheduling problems. European Journal of Operational Research, 307(1), 82-102.

Xia, L. (2022, December). Traffic Spatial Data Management System Based on Data Mining. In International Conference on Big Data Analytics for Cyber-Physical System in Smart City (pp. 17-25). Singapore: Springer Nature Singapore.

Xiaojing, W., Yanling, Y., Wenyu, L., & Ping, T. (2024). Research value of risk assessment of multidrug resistant organisms based on big data mining. Electronic Journal of Emerging Infectious Diseases, 9(1), 31.

Yuliana, Y., Noviyanti, N., & Qulub, M. (2024). Implementasi Algoritma Depth-First Search Dan Breadth-First Search Pada Dokumen Akreditasi. Journal Of Science And Social Research, 7(1), 197-204.

Zeitz, T. (2023).NP-hardness of shortest path problems in networks with non-FIFO time-dependent travel times.Information Processing Letters, 179, 106287.

Zha, W., Ji, Y., & Liang, C. (2024). Short-term load forecasting method based on secondary decomposition and improved hierarchical clustering. Results in Engineering, 22, 101993.

Zhang, Z., Dong, S., Li, D., Liu, P., & Wang, Z. (2024). Prediction and Diagnosis of Electric Vehicle Battery Fault Based on Abnormal Voltage: Using Decision Tree Algorithm Theories and Isolated Forest. Processes, 12(1), 136.

Appendices

Appendix 1

Ordinary Least Square Results:

Meters Results:

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	0.158876	0.027923	5.689849	0.000000*	0.027051	5.873312	0.000000*	--------
METER_SPATIA	-3.002761	0.066076	-45.443808	0.000000*	0.185355	-16.200067	0.000000*	35.623323
METER_SPATIA	0.013483	0.004495	2.999934	0.002713*	0.004568	2.951535	0.003173*	13.440229
METER_SPATIA	0.012484	0.003103	4.022935	0.000065*	0.003049	4.094389	0.000049*	48.899936
METERS.CSV.C	-0.000000	0.000000	-6.686622	0.000000*	0.000000	-1.104700	0.269288	1.003721
METERS.CSV.R	0.001213	0.000026	46.076228	0.000000*	0.000074	16.488400	0.000000*	35.624830
METERS.CSV.N	-0.000023	0.000005	-4.525382	0.000008*	0.000005	-4.643744	0.000005*	13.476088
METERS.CSV.G	-0.000318	0.000065	-4.865348	0.000002*	0.000064	-4.993445	0.000001*	48.852874

OLS Diagnostics

Input Features:	meter_SpatialJoin	Dependent Variable:	METERS.CSV.MAINT
Number of Observations:	59032	Akaike's Information Criterion (AICc) [d]:	-21835.300433
Multiple R-Squared [d]:	0.037179	Adjusted R-Squared [d]:	0.037065
Joint F-Statistic [e]:	325.599390	Prob(>F), (7,59024) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	357.173162	Prob(>chi-squared), (7) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	2637.197666	Prob(>chi-squared), (7) degrees of freedom:	0.000000*
Jarque-Bera Statistic [g]:	860232.034094	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.01$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

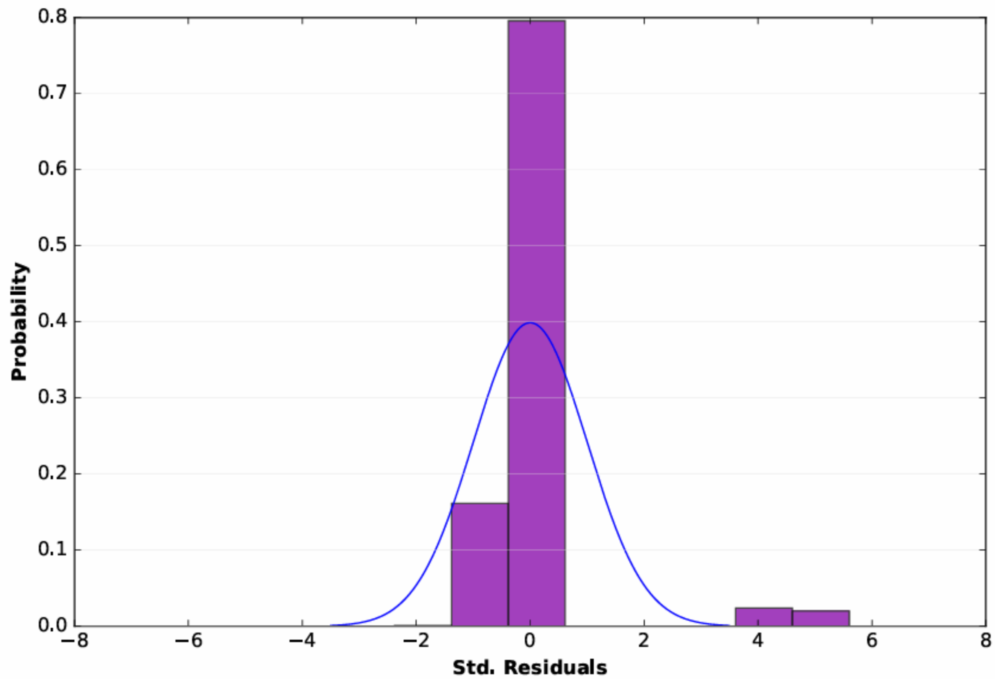
[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.01$) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships (Cont.)

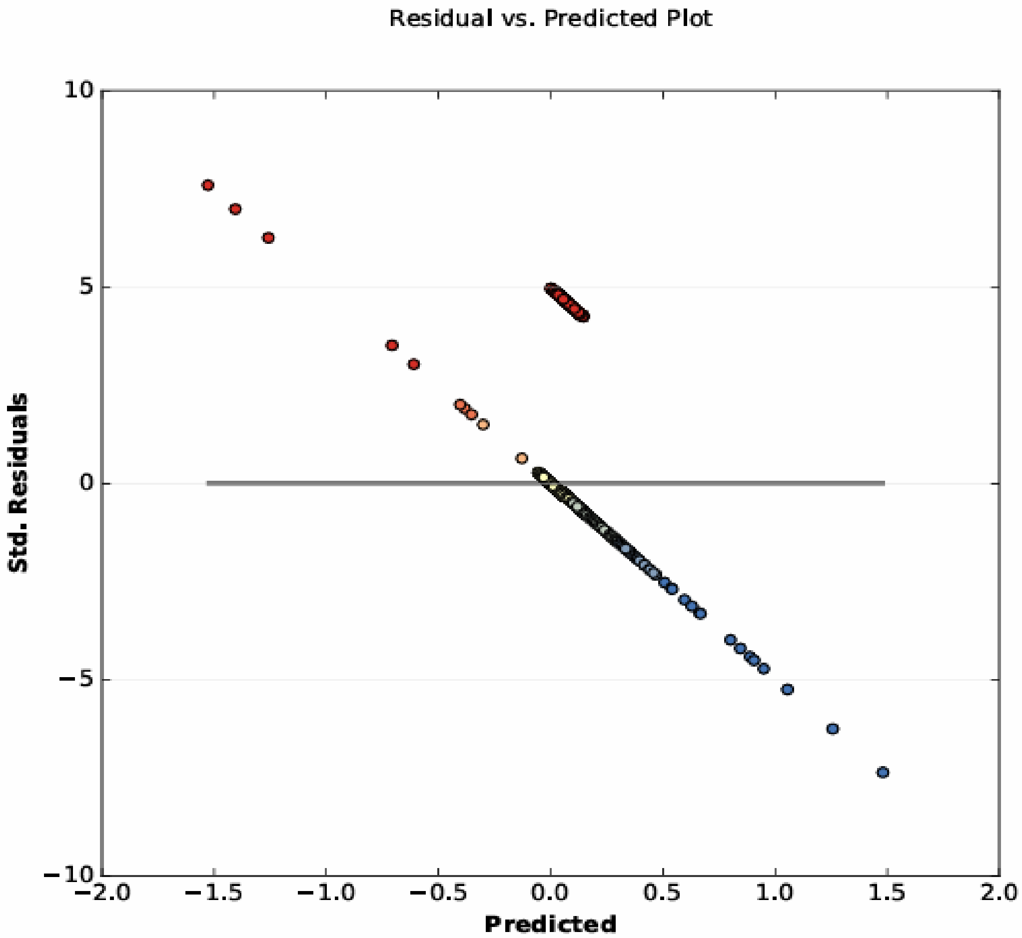
The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.

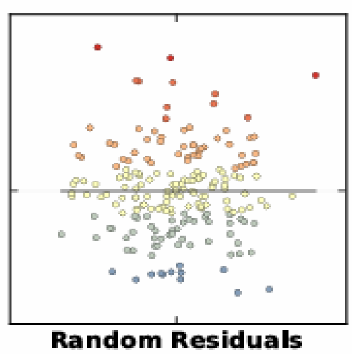
Histogram of Standardized Residuals



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Appendix 2

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	31.076112	0.676183	45.958152	0.000000*	0.679391	45.741109	0.000000*	———
HOUSECONECTI	-0.085939	0.018161	-4.732047	0.000004*	0.021068	-4.079148	0.000053*	> 1000.0
HOUSECONECTI	-0.002918	0.000621	-4.701133	0.000004*	0.000612	-4.764995	0.000003*	1.010749
HOUSECONNECT	-0.000171	0.000004	-45.790384	0.000000*	0.000004	-46.054190	0.000000*	1.182137
HOUSECONNECT	0.014939	0.002561	5.832234	0.000000*	0.002059	7.256139	0.000000*	1.142937
HOUSECONNECT	-0.009461	0.005749	-1.645559	0.099963	0.005838	-1.620575	0.105218	1.001607
HOUSECONNECT	0.654708	0.139957	4.677922	0.000004*	0.162002	4.041364	0.000062*	> 1000.0

OLS Diagnostics

Input Features:	Houseconection_contract	Dependent Variable:	HOUSECONNECTION-MODIFIE
Number of Observations:	3280	Akaike's Information Criterion (AICc) [d]:	2524.476641
Multiple R-Squared [d]:	0.415108	Adjusted R-Squared [d]:	0.414036
Joint F-Statistic [e]:	387.151474	Prob(>F), (6,3273) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	2955.891469	Prob(>chi-squared), (6) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	258.054709	Prob(>chi-squared), (6) degrees of freedom:	0.000000*
Jarque-Bera Statistic [g]:	57.092969	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value (p < 0.01).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant (p < 0.01); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

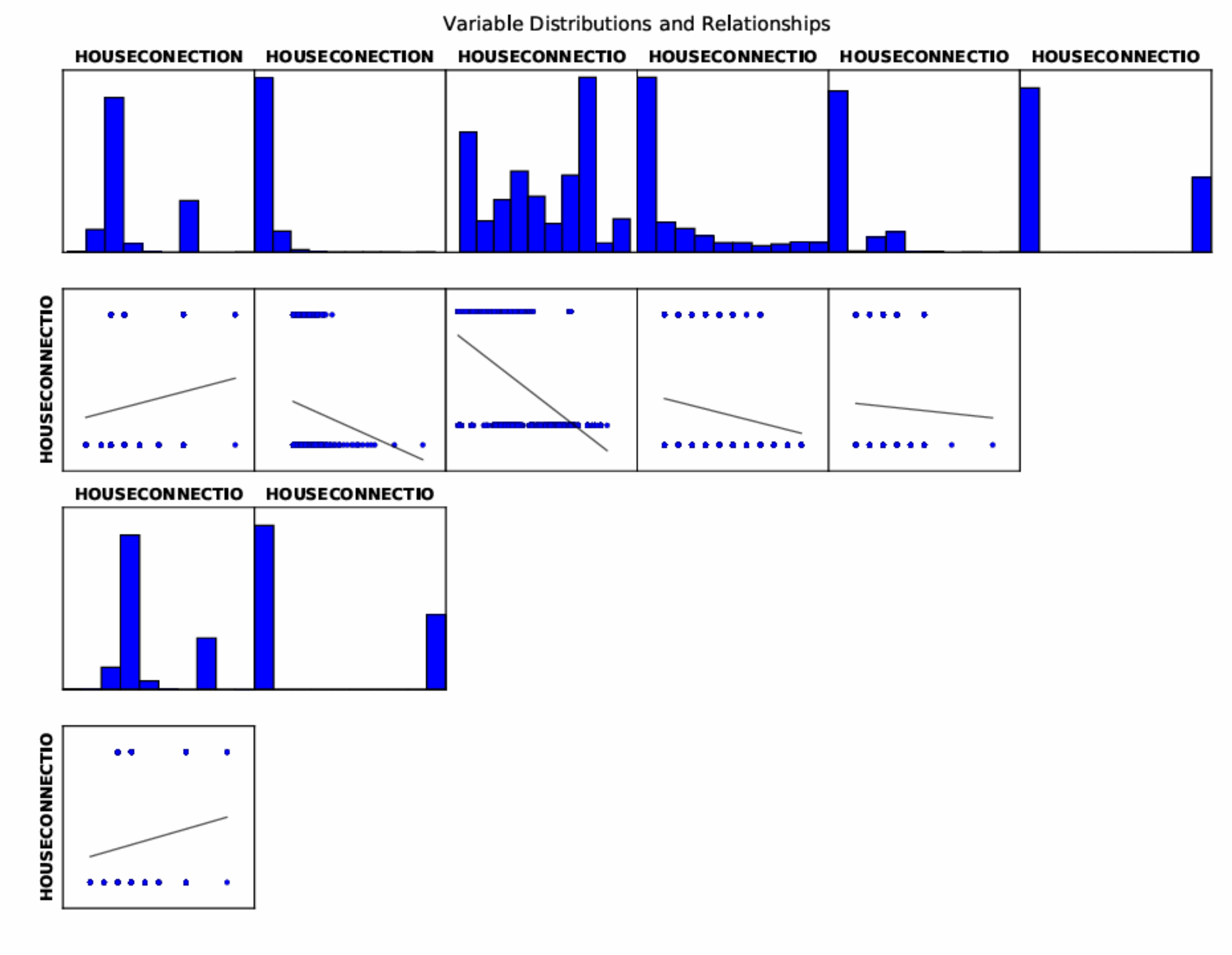
[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance (p < 0.01); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant (p < 0.01), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

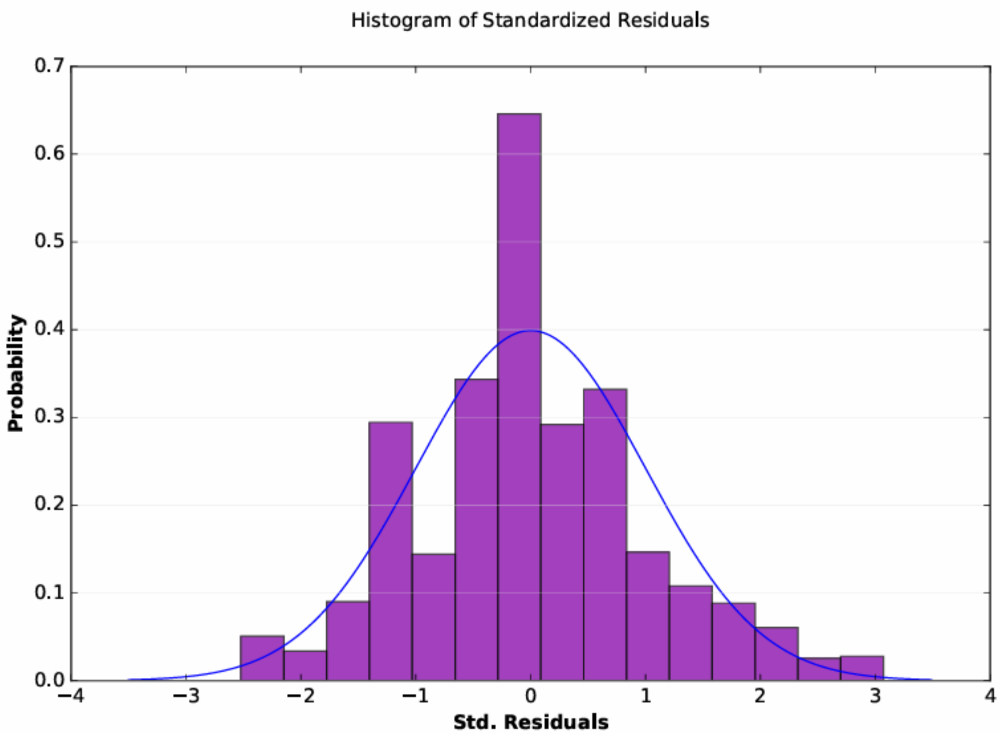
[g] Jarque-Bera Statistic: When this test is statistically significant (p < 0.01) model predictions are biased (the residuals are not normally distributed).


Variable Distributions and Relationships
HOUSECONECTION
HOUSECONECTION
HOUSECONNECTIO
HOUSECONNECTIO
HOUSECONNECTIO
HOUSECONNECTIO
HOUSECONNECTIO
HOUSECONNECTIO
HOUSECONNECTIO
HOUSECONNECTIO

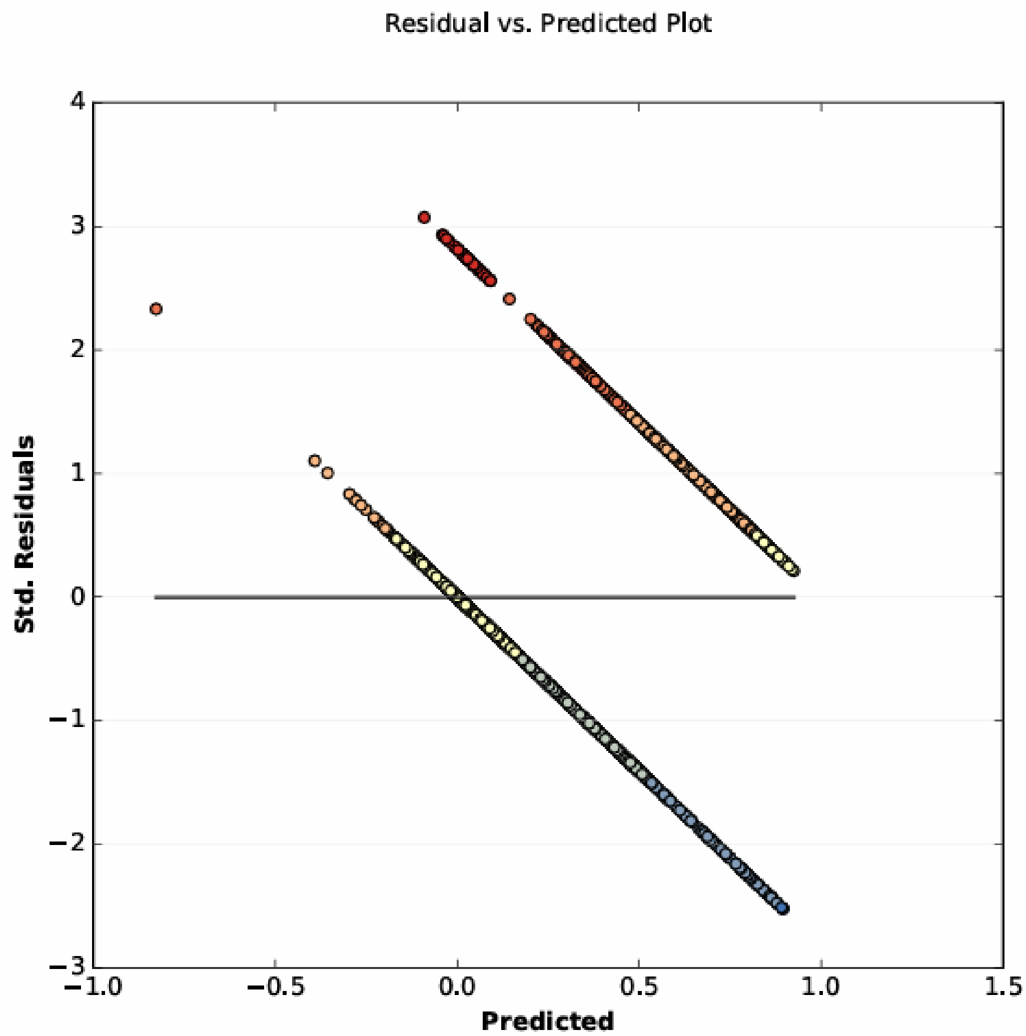
Variable Distributions and Relationships (Cont.)

The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

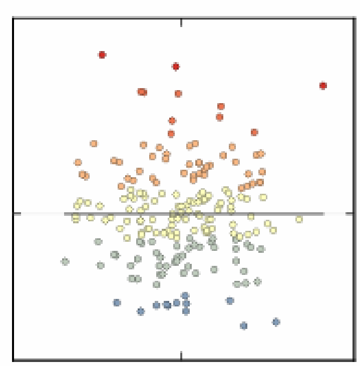
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



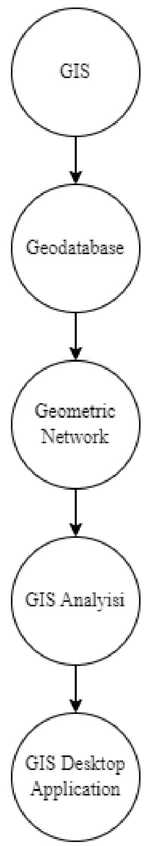
Random Residuals

Ordinary Least Squares Parameters

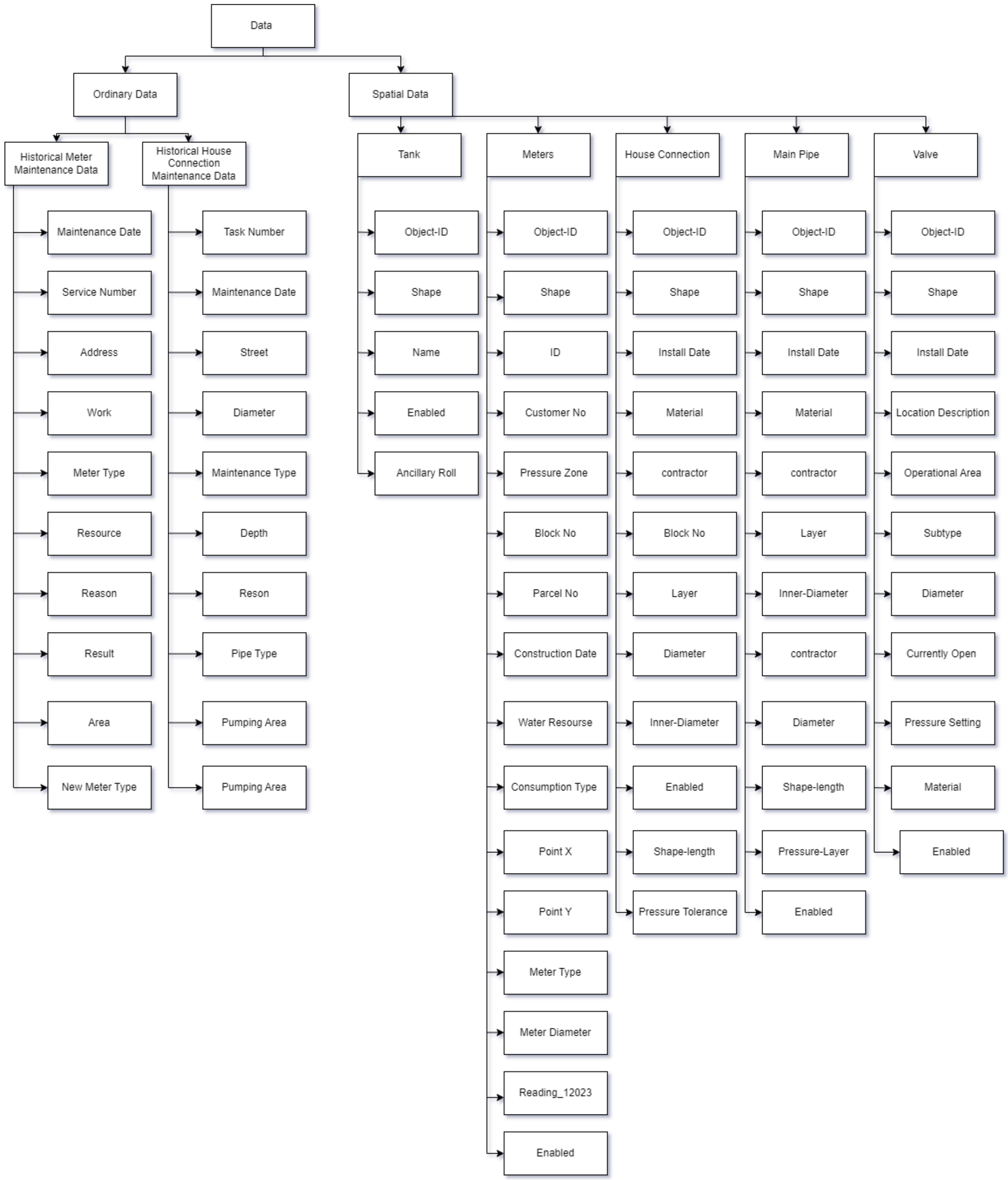
Parameter Name	Input Value
Input Features	Houseconection_contractall
Unique ID Field	HOUSECONNECTION-MODIFIED.csv.OBJECTID
Output Feature Class	None
Dependent Variable	HOUSECONNECTION-MODIFIED.CSV.MAINT
Explanatory Variables	HOUSECONECTION_CONTRACTALL.DIAM
	HOUSECONECTION_CONTRACTALL.SHAPE_LENGTH
	HOUSECONNECTION-MODIFIED.CSV.NEAR_X
	HOUSECONNECTION-MODIFIED.CSV.F_NEARDIST
	HOUSECONNECTION-MODIFIED.CSV.F_DEPTH
	HOUSECONNECTION-MODIFIED.CSV.F_SUBTYPE
Selection Set	False

Appendix 3

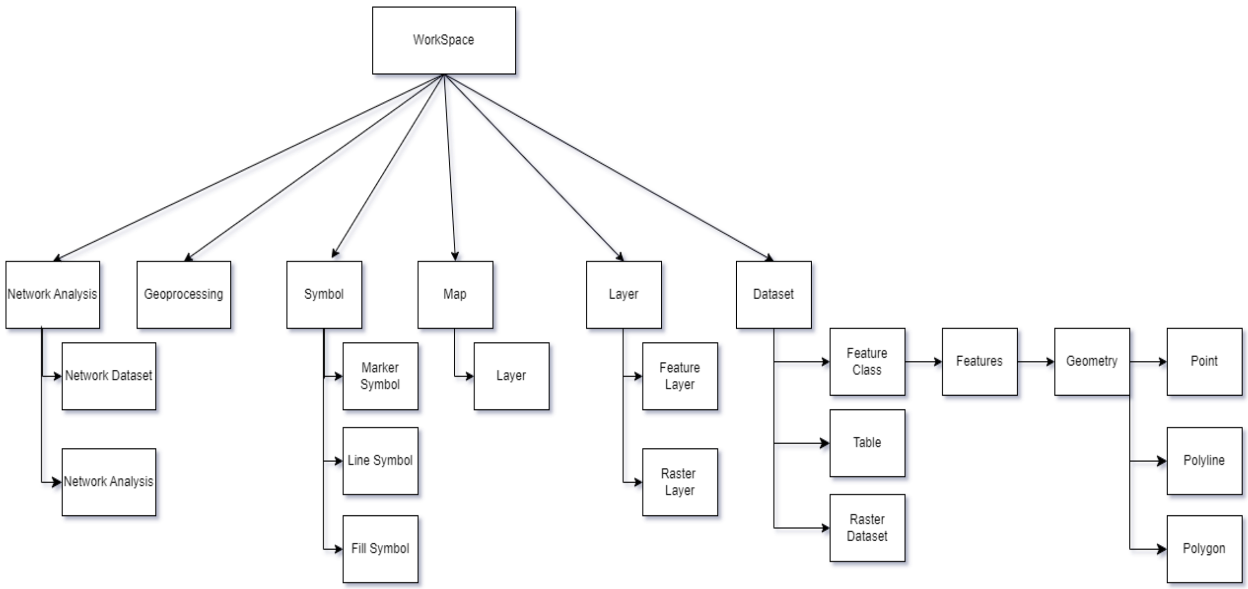
1- Diagrams:



Methodology Diagram


Data
Ordinary Data
Spatial Data
Historical Meter Maintenance Data
Historical House Connection Maintenance Data
Maintenance Date
Service Number
Address
Work
Meter Type
Resource
Reason
Result
Area
New Meter Type
Task Number
Maintenance Date
Street
Diameter
Maintenance Type
Depth
Reson
Pipe Type
Pumping Area
Pumping Area
Tank
Object-ID
Shape
Name
Enabled
Ancillary Roll
Meters
Object-ID
Shape
ID
Customer No
Pressure Zone
Block No
Parcel No
Construction Date
Water Resourse
Consumption Type
Point X
Point Y
Meter Type
Meter Diameter
Reading_12023
Enabled
House Connection
Object-ID
Shape
Install Date
Material
contractor
Block No
Layer
Diameter
Inner-Diameter
Enabled
Shape-length
Pressure Tolerance
Main Pipe
Object-ID
Shape
Install Date
Material
contractor
Layer
Inner-Diameter
contractor
Diameter
Shape-length
Pressure-Layer
Enabled
Valve
Object-ID
Shape
Install Date
Location Description
Operational Area
Subtype
Diameter
Currently Open
Pressure Setting
Material
Enabled

App Diagram



ArcObject Diagram

2-Code Snippets:

```csharp
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Threading.Tasks;

namespace WpfApp1.Models
{
    public static class Database
    {
        public static string ConnectionString = "";
    }
}
```

```csharp
using System;
using System.Collections.Generic;
using System.Linq;
using System.Reflection;
using System.Text;
using System.Threading.Tasks;
using Esri.ArcGISRuntime.Data;
using Microsoft.Data.Sqlite;

namespace WpfApp1.Models
{
    public class HouseConnection
    {
        public int OBJECTID { get; set; }
        public string Material { get; set; }
        public int diameter { get; set; }
        public string diameter_unit { get; set; }
        public short Diam { get; set; }
        public string layer { get; set; }
        public int control_valve_id { get; set; }
        public string shape_length { get; set; }
        //public int Diam { get; set; }
        //public string Material { get; set; }
        public string status { get; set; }
        //public int diameter { get; set; }
        //public string diameter_unit { get; set; }
        public bool HasUnFinishedFix { get; set; }
        public List<HousePipeFix> fixes;
```

```
public HouseConnection() {

    HasUnFinishedFix = false;
}

2 references
public void setDatAFromFeature(ArcGISFeature feature, string[] attributes)
{
    for (int i = 0; i < attributes.Length; i++)
    {
        string attribute = attributes[i];
        if (attribute.Length == 0)
        {
            continue;
        }
        object outObject;
        PropertyInfo propertyInfo = typeof(HouseConnection).GetProperty(attribute);
        if (propertyInfo != null && propertyInfo.CanWrite)
        {
            if (propertyInfo.PropertyType.Name == "String")
            {
                propertyInfo.SetValue(this, "");
            } else
            {
                propertyInfo.SetValue(this, null);
            }


            if (feature.Attributes.TryGetValue(attribute, out outObject))
            {
                if (!(outObject is null))
                {
                    if (outObject is Int64)
                    {
                        propertyInfo.SetValue(this, Convert.ToInt32(outObject));
                    }
                    else
                    {
```

WpfApp1 | WpfApp1.Models.HouseConnection

```
                    }
                }
            }
            HasUnFinishedFix = false;
            if (!(fixes is null))
            {
                fixes.Clear();
            }

        }

        1 reference
        internal void getFixes()
        {
            fixes = HousePipeFix.GetAllByPipeID(OBJECTID);
            for (int i = 0; i < fixes.Count; i++)
            {
                if (fixes[i].Finished == 0)
                {
                    HasUnFinishedFix = true;
                    break;
                }
            }
        }

        1 reference
        public HousePipeFix getNotFinishedFixe()
        {
            for (int i = 0; i < fixes.Count; i++)
            {
                if (fixes[i].Finished == 0)
                {
                    return fixes[i];
                }
            }
            return null;
        }

```

This code snippet is for modeling, representing and handling house connection pipes.

```
WpfApp1    WpfApp1.Models.HousePipeFix    Insert()

        1 reference
        public HousePipeFix(SqliteDataReader reader)
        {
            ConnectionString = Database.ConnectionString;
            // (pipe_id, pipe_price, total_cost, start_date, finish_date, finished, pipe_length, notes, worker_name, material)
            Id = reader.GetInt32(reader.GetOrdinal("id"));
            PipeId = reader.GetInt32(reader.GetOrdinal("pipe_id"));
            PipeLength = reader.GetDouble(reader.GetOrdinal("pipe_length"));
            PipePrice = reader.GetDouble(reader.GetOrdinal("pipe_price"));
            TotalCost = reader.GetDouble(reader.GetOrdinal("total_cost"));
            StartDate = reader["start_date"].ToString();
            FinishDate = reader["finish_date"].ToString();
            Finished = reader.GetInt16(reader.GetOrdinal("finished"));
            Notes = reader["notes"].ToString();
            WorkerName = reader["worker_name"].ToString();
            Material = reader["material"].ToString();
        }

        // Create (Insert)
        1 reference
        public bool Insert()
        {
            try
            {
                using (SqliteConnection connection = new SqliteConnection(Database.ConnectionString))
                {
                    connection.Open();
                    string query = "INSERT INTO house_pipe_fixes (pipe_id, pipe_price, total_cost, start_date, finish_date, finished, pipe_length, notes, worker_name
                    using (SqliteCommand command = new SqliteCommand(query, connection))
                    {
                        command.Parameters.AddWithValue("@PipeId", PipeId);
                        command.Parameters.AddWithValue("@PipePrice", PipePrice);
                        command.Parameters.AddWithValue("@TotalCost", TotalCost);
                        command.Parameters.AddWithValue("@StartDate", StartDate);
                        command.Parameters.AddWithValue("@FinishDate", FinishDate);
                        command.Parameters.AddWithValue("@Finished", Finished);
                        command.Parameters.AddWithValue("@PipeLength", PipeLength);
                        command.Parameters.AddWithValue("@Notes", Notes);
                        command.Parameters.AddWithValue("@WorkerName", WorkerName);
                        command.Parameters.AddWithValue("@Material", Material);
                        command.ExecuteNonQuery();
                        return true;
```

```
WpfApp1    WpfApp1.Models.HousePipeFix    Insert()

        // Read (Select) - Retrieve all records
        1 reference
        public static List<HousePipeFix> GetAllByPipeID(int PipeId)
        {
            List<HousePipeFix> fixes = new List<HousePipeFix>();
            using (SqliteConnection connection = new SqliteConnection(Database.ConnectionString))
            {
                connection.Open();
                string query = "SELECT * FROM house_pipe_fixes WHERE pipe_id = @id";
                using (SqliteCommand command = new SqliteCommand(query, connection))
                {
                    command.Parameters.AddWithValue("@id", PipeId);
                    using (SqliteDataReader reader = command.ExecuteReader())
                    {
                        while (reader.Read())
                        {
                            fixes.Add(new HousePipeFix(reader));
                        }
                    }
                }

            }
            return fixes;
        }

        // Read (Select) - Retrieve by Id
        0 references
        public static HousePipeFix GetById(int id)
        {
            using (SqliteConnection connection = new SqliteConnection(Database.ConnectionString))
            {
                connection.Open();
                string query = "SELECT * FROM house_pipe_fixes WHERE Id = @Id";
                using (SqliteCommand command = new SqliteCommand(query, connection))
                {
                    command.Parameters.AddWithValue("@Id", id);
                    using (SqliteDataReader reader = command.ExecuteReader())
                    {
                        if (reader.Read())
                        {
                            //return new HousePipeFix(reader);
                        }
                    }
```

WpfApp1 | WpfApp1.Models.HousePipeFix | Insert()

```
public bool Update()
{
    try
    {
        using (SqliteConnection connection = new SqliteConnection(Database.ConnectionString))
        {
            connection.Open();
            // (pipe_id, pipe_price, total_cost, start_date, finish_date, finished, pipe_length, notes, worker_name, material)
            //(@PipeId, @PipePrice, @TotalCost, @StartDate, @FinishDate, @Finished, @PipeLength, @Notes, @WorkerName, @Material)

            string query = "UPDATE house_pipe_fixes SET pipe_price = @PipePrice, total_cost = @TotalCost, start_date = @StartDate, finish_date = @FinishDate,
            using (SqliteCommand command = new SqliteCommand(query, connection))
            {
                command.Parameters.AddWithValue("@Id", Id);
                command.Parameters.AddWithValue("@PipePrice", PipePrice);
                command.Parameters.AddWithValue("@TotalCost", TotalCost);
                command.Parameters.AddWithValue("@StartDate", StartDate);
                command.Parameters.AddWithValue("@FinishDate", FinishDate);
                command.Parameters.AddWithValue("@PipeLength", PipeLength);
                command.Parameters.AddWithValue("@Finished", Finished);
                command.Parameters.AddWithValue("@Notes", Notes);
                command.Parameters.AddWithValue("@WorkerName", WorkerName);
                command.Parameters.AddWithValue("@Material", Material);
                command.ExecuteNonQuery();
                return true;
            }
        }
    }
    catch (SqliteException ex)
    {
        Console.WriteLine("SQLite Error: " + ex.Message);
    }
    catch (Exception ex)
    {
        Console.WriteLine("Error: " + ex.Message);
    }

    return false;
}

// Delete
0 references
```

WpfApp1 | WpfApp1.Models.MainPipe | OBJECTID

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Reflection;
using System.Text;
using System.Threading.Tasks;
using Esri.ArcGISRuntime.Data;
using Microsoft.Data.Sqlite;

namespace WpfApp1.Models
{
    5 references
    public class MainPipe
    {
        4 references
        public int OBJECTID { get; set; }
        1 reference
        public int control_valve_id { get; set; }
        4 references
        public double Shape_Leng { get; set; }
        0 references
        public int Diam { get; set; }
        4 references
        public string Material { get; set; }
        15 references
        public string status { get; set; }
        4 references
        public int diameter { get; set; }
        4 references
        public string diameter_unit { get; set; }
        4 references
        public bool HasUnFinishedFix { get; set; }
        public List<MainPipeFix> fixes;
        1 reference
        public MainPipe() {

            HasUnFinishedFix = false;
        }

        1 reference
        public void setDatAFromFeature(ArcGISFeature feature, string[] attributes)
        {
            for (int i = 0; i < attributes.Length; i++)
            {
                string attribute = attributes[i];
```

WpfApp1 | WpfApp1.Models.MainPipe | OBJECTID

```
    {
        string attribute = attributes[i];
        if (attribute.Length == 0)
        {
            continue;
        }
        object outObject;
        PropertyInfo propertyInfo = typeof(MainPipe).GetProperty(attribute);
        if (propertyInfo != null && propertyInfo.CanWrite)
        {
            if (propertyInfo.PropertyType.Name == "String")
            {
                propertyInfo.SetValue(this, "");
            } else
            {
                propertyInfo.SetValue(this, null);
            }


            if (feature.Attributes.TryGetValue(attribute, out outObject))
            {
                if (!(outObject is null))
                {
                    propertyInfo.SetValue(this, outObject);
                }
            }
        }
    }
    HasUnFinishedFix = false;
    if (!(fixes is null))
    {
        fixes.Clear();
    }

}

1 reference
internal void getFixes()
{
    fixes = MainPipeFix.GetAllByPipeID(OBJECTID);
    for (int i = 0; i < fixes.Count; i++)
    {
        if (fixes[i].Finished == 0)
        {
            HasUnFinishedFix = true;
```

83 % | No issues found

Output

WpfApp1 | WpfApp1.Models.MainPipe

```
        for (int i = 0; i < fixes.Count; i++)
        {
            if (fixes[i].Finished == 0)
            {
                HasUnFinishedFix = true;
                break;
            }
        }
    }

    1 reference
    public MainPipeFix getNotFinishedFixe()
    {
        for (int i = 0; i < fixes.Count; i++)
        {
            if (fixes[i].Finished == 0)
            {
                return fixes[i];
            }
        }
        return null;
    }

    0 references
    public string getStatus()
    {
        if (status == "valid")
        {
            return "Valid";
        }
        if (status == "invalid")
        {
            return "Invalid";
        }
        if (status == "under_repair")
        {
            return "Under Repair";
        }

        return status;
    }
```

WpfApp1 WpfApp1.Models.MainPipeFix

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using Microsoft.Data.Sqlite;

namespace WpfApp1.Models
{
    20 references
    public class MainPipeFix
    {
        4 references
        public int Id { get; set; }
        3 references
        public int PipeId { get; set; }
        5 references
        public double PipePrice { get; set; }
        5 references
        public double TotalCost { get; set; }
        5 references
        public string StartDate { get; set; }
        6 references
        public string FinishDate { get; set; }
        11 references
        public short Finished { get; set; }
        5 references
        public double PipeLength { get; set; }
        5 references
        public string Notes { get; set; }
        5 references
        public string WorkerName { get; set; }
        5 references
        public string Material { get; set; }
        10 references
        public MainPipe MainPipe { get; set; }

        // Database connection string (adjust as needed)
        private string ConnectionString = "";

        1 reference
        public MainPipeFix() {
            ConnectionString = Database.ConnectionString;
        }

        1 reference
```

```
        1 reference
        public MainPipeFix(SqliteDataReader reader)
        {
            ConnectionString = Database.ConnectionString;
            // (pipe_id, pipe_price, total_cost, start_date, finish_date, finished, pipe_length, notes, worker_name, material)
            Id = reader.GetInt32(reader.GetOrdinal("id"));
            PipeId = reader.GetInt32(reader.GetOrdinal("pipe_id"));
            PipeLength = reader.GetDouble(reader.GetOrdinal("pipe_length"));
            PipePrice = reader.GetDouble(reader.GetOrdinal("pipe_price"));
            TotalCost = reader.GetDouble(reader.GetOrdinal("total_cost"));
            StartDate = reader["start_date"].ToString();
            FinishDate = reader["finish_date"].ToString();
            Finished = reader.GetInt16(reader.GetOrdinal("finished"));
            Notes = reader["notes"].ToString();
            WorkerName = reader["worker_name"].ToString();
            Material = reader["material"].ToString();
        }

        // Create (Insert)
        1 reference
        public bool Insert()
        {
            try
            {
                using (SqliteConnection connection = new SqliteConnection(Database.ConnectionString))
                {
                    connection.Open();
                    string query = "INSERT INTO main_pipe_fixes (pipe_id, pipe_price, total_cost, start_date, finish_date, finished, pipe_length, notes, worker_name
                    using (SqliteCommand command = new SqliteCommand(query, connection))
                    {
                        command.Parameters.AddWithValue("@PipeId", PipeId);
                        command.Parameters.AddWithValue("@PipePrice", PipePrice);
                        command.Parameters.AddWithValue("@TotalCost", TotalCost);
                        command.Parameters.AddWithValue("@StartDate", StartDate);
                        command.Parameters.AddWithValue("@FinishDate", FinishDate);
                        command.Parameters.AddWithValue("@Finished", Finished);
                        command.Parameters.AddWithValue("@PipeLength", PipeLength);
                        command.Parameters.AddWithValue("@Notes", Notes);
                        command.Parameters.AddWithValue("@WorkerName", WorkerName);
                        command.Parameters.AddWithValue("@Material", Material);
                        command.ExecuteNonQuery();
                        return true;
```

WpfApp1 | WpfApp1.Models.MainPipeFix | Id

```
                        return true;
                    }
                }
            }
            catch (SqliteException ex)
            {
                Console.WriteLine("SQLite Error: " + ex.Message);
            }
            catch (Exception ex)
            {
                Console.WriteLine("Error: " + ex.Message);
            }

            return false;
        }

        // Read (Select) - Retrieve all records
        2 references
        public static List<MainPipeFix> GetAllByPipeID(int PipeId)
        {
            List<MainPipeFix> fixes = new List<MainPipeFix>();
            using (SqliteConnection connection = new SqliteConnection(Database.ConnectionString))
            {
                connection.Open();
                string query = "SELECT * FROM main_pipe_fixes WHERE pipe_id = @id";
                using (SqliteCommand command = new SqliteCommand(query, connection))
                {
                    command.Parameters.AddWithValue("@id", PipeId);
                    using (SqliteDataReader reader = command.ExecuteReader())
                    {
                        while (reader.Read())
                        {
                            fixes.Add(new MainPipeFix(reader));
                        }
                    }
                }

            }
            return fixes;
        }

        // Read (Select) - Retrieve by Id
        0 references
        public static MainPipeFix GetById(int id)
```

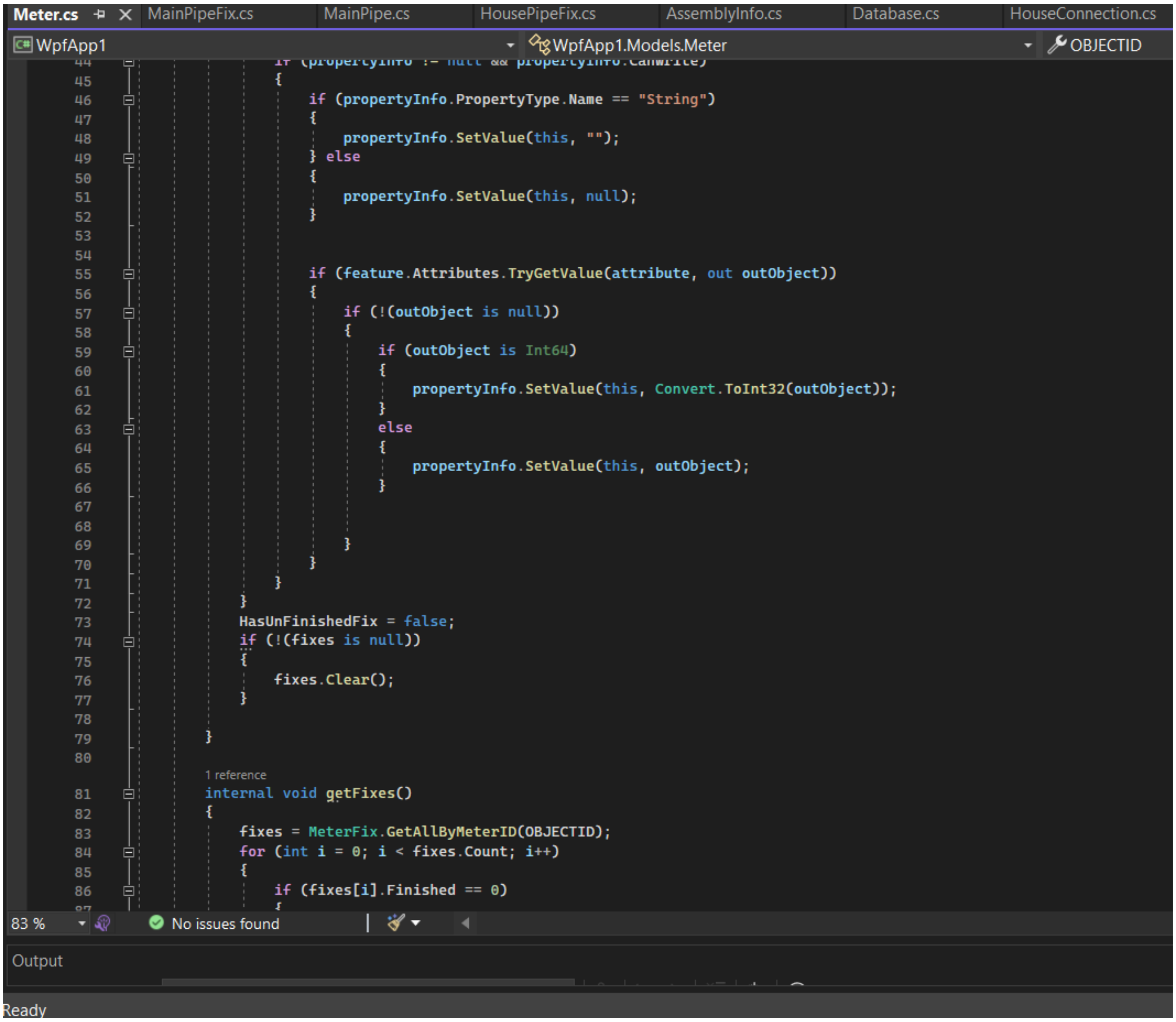
Meter.cs | MainPipeFix.cs | MainPipe.cs | HousePipeFix.cs | AssemblyInfo.cs

WpfApp1 | WpfApp1.Models.Meter

```
namespace WpfApp1.Models
{
    public class Meter
    {
        public int OBJECTID { get; set; }
        public int Customer_No { get; set; }
        public string meters_brand { get; set; }
        public string email { get; set; }
        public string phone { get; set; }
        public int control_valve_id { get; set; }
        public string kind { get; set; }
        //public int Diam { get; set; }
        //public string Material { get; set; }
        public string status { get; set; }
        //public int diameter { get; set; }
        //public string diameter_unit { get; set; }
        public bool HasUnFinishedFix { get; set; }
        public List<MeterFix> fixes;
        public Meter() {

            HasUnFinishedFix = false;
        }

        public void setDatAFromFeature(ArcGISFeature feature, string[] attributes)
        {
            for (int i = 0; i < attributes.Length; i++)
            {
                string attribute = attributes[i];
                if (attribute.Length == 0)
                {
                    continue;
                }
                object outObject;
                PropertyInfo propertyInfo = typeof(Meter).GetProperty(attribute);
```


Meter.cs
MainPipeFix.cs
MainPipe.cs
HousePipeFix.cs
AssemblyInfo.cs
Database.cs
HouseConnection.cs
WpfApp1
WpfApp1.Models.Meter
OBJECTID
if (propertyInfo != null && propertyInfo.CanWrite)
{
if (propertyInfo.PropertyType.Name == "String")
{
propertyInfo.SetValue(this, "");
} else
{
propertyInfo.SetValue(this, null);
}
if (feature.Attributes.TryGetValue(attribute, out outObject))
{
if (!(outObject is null))
{
if (outObject is Int64)
{
propertyInfo.SetValue(this, Convert.ToInt32(outObject));
}
else
{
propertyInfo.SetValue(this, outObject);
}
}
}
}
}
HasUnFinishedFix = false;
if (!(fixes is null))
{
fixes.Clear();
}
}
1 reference
internal void getFixes()
{
fixes = MeterFix.GetAllByMeterID(OBJECTID);
for (int i = 0; i < fixes.Count; i++)
{
if (fixes[i].Finished == 0)
83 %
No issues found
Output
Ready

MeterFix.cs | Meter.cs | MainPipeFix.cs | MainPipe.cs | HousePipeFix.cs | AssemblyInfo.cs | Database.cs | HouseConnection.cs

WpfApp1 | WpfApp1.Models.MeterFix | Id

```csharp
        1 reference
        public MeterFix(SqliteDataReader reader)
        {
            ConnectionString = Database.ConnectionString;
            // (pipe_id, pipe_price, total_cost, start_date, finish_date, finished, pipe_length, notes, worker_name, material)
            Id = reader.GetInt32(reader.GetOrdinal("id"));
            MeterId = reader.GetInt32(reader.GetOrdinal("meter_id"));
            MeterPrice = reader.GetDouble(reader.GetOrdinal("meter_price"));
            TotalCost = reader.GetDouble(reader.GetOrdinal("total_cost"));
            StartDate = reader["start_date"].ToString();
            FinishDate = reader["finish_date"].ToString();
            Finished = reader.GetInt16(reader.GetOrdinal("finished"));
            Notes = reader["notes"].ToString();
            WorkerName = reader["worker_name"].ToString();
        }

        // Create (Insert)
        1 reference
        public bool Insert()
        {
            try
            {
                using (SqliteConnection connection = new SqliteConnection(Database.ConnectionString))
                {
                    connection.Open();
                    string query = "INSERT INTO meter_fixes (meter_id, meter_price, total_cost, start_date, finish_date, finished, notes, worker_name) VALUES (@Me
                    using (SqliteCommand command = new SqliteCommand(query, connection))
                    {
                        command.Parameters.AddWithValue("@MeterId", MeterId);
                        command.Parameters.AddWithValue("@MeterPrice", MeterPrice);
                        command.Parameters.AddWithValue("@TotalCost", TotalCost);
                        command.Parameters.AddWithValue("@StartDate", StartDate);
                        command.Parameters.AddWithValue("@FinishDate", FinishDate);
                        command.Parameters.AddWithValue("@Finished", Finished);
                        command.Parameters.AddWithValue("@Notes", Notes);
                        command.Parameters.AddWithValue("@WorkerName", WorkerName);
                        command.ExecuteNonQuery();
                        return true;
                    }
                }
            }
            catch (SqliteException ex)
            {
                Console.WriteLine("SQLite Error: " + ex.Message);
```

```
                }
            }
            catch (SqliteException ex)
            {
                Console.WriteLine("SQLite Error: " + ex.Message);
            }
            catch (Exception ex)
            {
                Console.WriteLine("Error: " + ex.Message);
            }

            return false;
        }

        // Read (Select) - Retrieve all records
        1 reference
        public static List<MeterFix> GetAllByMeterID(int PipeId)
        {
            List<MeterFix> fixes = new List<MeterFix>();
            using (SqliteConnection connection = new SqliteConnection(Database.ConnectionString))
            {
                connection.Open();
                string query = "SELECT * FROM meter_fixes WHERE meter_id = @id";
                using (SqliteCommand command = new SqliteCommand(query, connection))
                {
                    command.Parameters.AddWithValue("@id", PipeId);
                    using (SqliteDataReader reader = command.ExecuteReader())
                    {
                        while (reader.Read())
                        {
                            fixes.Add(new MeterFix(reader));
                        }
                    }
                }

            }
            return fixes;
        }

        // Read (Select) - Retrieve by Id
        0 references
        public static MeterFix GetById(int id)
        {
```

WpfApp1 WpfApp1.Models.SourceTank

```
public class SourceTank
{
    1 reference
    public int OBJECTID { get; set; }
    0 references
    public int control_valve_id { get; set; }
    //public double Shape_Leng { get; set; }
    //public int Diam { get; set; }
    //public string Material { get; set; }
    1 reference
    public string status { get; set; }
    //public int diameter { get; set; }
    //public string diameter_unit { get; set; }
    3 references
    public bool HasUnFinishedFix { get; set; }
    public List<MainPipeFix> fixes;
    1 reference
    public SourceTank() {

        HasUnFinishedFix = false;
    }

    1 reference
    public void setDatAFromFeature(ArcGISFeature feature, string[] attributes)
    {
        for (int i = 0; i < attributes.Length; i++)
        {
            string attribute = attributes[i];
            if (attribute.Length == 0)
            {
                continue;
            }
            object outObject;
            PropertyInfo propertyInfo = typeof(SourceTank).GetProperty(attribute);
            if (propertyInfo != null && propertyInfo.CanWrite)
            {
                if (propertyInfo.PropertyType.Name == "String")
                {
                    propertyInfo.SetValue(this, "");
                } else
                {
                    propertyInfo.SetValue(this, null);
                }

```

WpfApp1 | WpfApp1.Models.SourceTank

```
                {
                    if (!(outObject is null))
                    {
                        propertyInfo.SetValue(this, outObject);
                    }
                }
            }
        }
        HasUnFinishedFix = false;
        if (!(fixes is null))
        {
            fixes.Clear();
        }

    }

    0 references
    internal void getFixes()
    {
        fixes = MainPipeFix.GetAllByPipeID(OBJECTID);
        for (int i = 0; i < fixes.Count; i++)
        {
            if (fixes[i].Finished == 0)
            {
                HasUnFinishedFix = true;
                break;
            }
        }
    }

    0 references
    public MainPipeFix getNotFinishedFixe()
    {
        for (int i = 0; i < fixes.Count; i++)
        {
            if (fixes[i].Finished == 0)
            {
                return fixes[i];
            }
        }
        return null;
    }

    /*
```

```csharp
using Esri.ArcGISRuntime.Geometry;
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Threading.Tasks;

namespace WpfApp1.Models
{
    public class TreeNode
    {
        public int NodeId { get; set; } = -1;
        public int id { get; set; } = -1;
        public string type { get; set; }
        public int parent_id { get; set; }
        public string parent_type { get; set; }
        public string point_edge { get; set; }
        public int control_valve_id { get; set; } = -1;
        public bool visited { get; set; } = false;
        public int level { get; set; } = 0;
        public Geometry geometry { get; set; }

        public override string ToString()
        {

            return $"NodeId: {NodeId} \n, id: {id} \n, type: {type} \n,";
        }
    }
}
```

WpfApp1 | WpfApp1.Models.Valve | OBJECTID

```
public int OBJECTID { get; set; }
2 references
public int control_valve_id { get; set; }
//public double Shape_Leng { get; set; }
4 references
public double Diameter { get; set; }
7 references
public int Subtype { get; set; }
14 references
public string status { get; set; }
3 references
public string Material { get; set; }
4 references
public string diameter_unit { get; set; }
4 references
public bool HasUnFinishedFix { get; set; }
public List<ValveFix> fixes;
2 references
public Valve() {

    HasUnFinishedFix = false;
}

2 references
public void setDatAFromFeature(ArcGISFeature feature, string[] attributes)
{
    for (int i = 0; i < attributes.Length; i++)
    {
        string attribute = attributes[i];
        if (attribute.Length == 0)
        {
            continue;
        }
        object outObject;
        PropertyInfo propertyInfo = typeof(Valve).GetProperty(attribute);
        if (propertyInfo != null && propertyInfo.CanWrite)
        {
            if (propertyInfo.PropertyType.Name == "String")
            {
                propertyInfo.SetValue(this, "");
            } else
            {
                propertyInfo.SetValue(this, null);
            }

```

WpfApp1 — WpfApp1.Models.Valve

```
                            {
                                propertyInfo.SetValue(this, outObject);
                            }


                        }
                    }
                }
            }
            HasUnFinishedFix = false;
            if (!(fixes is null))
            {
                fixes.Clear();
            }

        }

        1 reference
        internal void getFixes()
        {
            fixes = ValveFix.GetAllByValveID(OBJECTID);
            for (int i = 0; i < fixes.Count; i++)
            {
                if (fixes[i].Finished == 0)
                {
                    HasUnFinishedFix = true;
                    break;
                }
            }
        }

        1 reference
        public ValveFix getNotFinishedFixe()
        {
            for (int i = 0; i < fixes.Count; i++)
            {
                if (fixes[i].Finished == 0)
                {
                    return fixes[i];
                }
            }
            return null;
        }

        1 reference
```

```
WpfApp1    WpfApp1.Models.ValveFix    Id

namespace WpfApp1.Models
{
    17 references
    public class ValveFix
    {
        4 references
        public int Id { get; set; }
        3 references
        public int ValveId { get; set; }
        5 references
        public double ValvePrice { get; set; }
        5 references
        public double TotalCost { get; set; }
        5 references
        public string StartDate { get; set; }
        6 references
        public string FinishDate { get; set; }
        9 references
        public short Finished { get; set; }
        5 references
        public string Notes { get; set; }
        5 references
        public string WorkerName { get; set; }
        8 references
        public Valve Valve { get; set; }

        // Database connection string (adjust as needed)
        private string ConnectionString = "";

        1 reference
        public ValveFix() {
            ConnectionString = Database.ConnectionString;
        }

        1 reference
        public ValveFix(SqliteDataReader reader)
        {
            ConnectionString = Database.ConnectionString;
            // (pipe_id, pipe_price, total_cost, start_date, finish_date, finished, pipe_length, notes, worker_name, material)
            Id = reader.GetInt32(reader.GetOrdinal("id"));
            ValveId = reader.GetInt32(reader.GetOrdinal("valve_id"));
            ValvePrice = reader.GetDouble(reader.GetOrdinal("valve_price"));
            TotalCost = reader.GetDouble(reader.GetOrdinal("total_cost"));
            StartDate = reader["start_date"].ToString();
            FinishDate = reader["finish_date"].ToString();
```

```
WpfApp1    WpfApp1.Models.ValveFix    Id

        1 reference
        public ValveFix(SqliteDataReader reader)
        {
            ConnectionString = Database.ConnectionString;
            // (pipe_id, pipe_price, total_cost, start_date, finish_date, finished, pipe_length, notes, worker_name, material)
            Id = reader.GetInt32(reader.GetOrdinal("id"));
            ValveId = reader.GetInt32(reader.GetOrdinal("valve_id"));
            ValvePrice = reader.GetDouble(reader.GetOrdinal("valve_price"));
            TotalCost = reader.GetDouble(reader.GetOrdinal("total_cost"));
            StartDate = reader["start_date"].ToString();
            FinishDate = reader["finish_date"].ToString();
            Finished = reader.GetInt16(reader.GetOrdinal("finished"));
            Notes = reader["notes"].ToString();
            WorkerName = reader["worker_name"].ToString();
        }

        // Create (Insert)
        1 reference
        public bool Insert()
        {
            try
            {
                using (SqliteConnection connection = new SqliteConnection(Database.ConnectionString))
                {
                    connection.Open();
                    string query = "INSERT INTO valve_fixes (valve_id, valve_price, total_cost, start_date, finish_date, finished, notes, worker_name) VALUES (@Val
                    using (SqliteCommand command = new SqliteCommand(query, connection))
                    {
                        command.Parameters.AddWithValue("@ValveId", ValveId);
                        command.Parameters.AddWithValue("@ValvePrice", ValvePrice);
                        command.Parameters.AddWithValue("@TotalCost", TotalCost);
                        command.Parameters.AddWithValue("@StartDate", StartDate);
                        command.Parameters.AddWithValue("@FinishDate", FinishDate);
                        command.Parameters.AddWithValue("@Finished", Finished);
                        command.Parameters.AddWithValue("@Notes", Notes);
                        command.Parameters.AddWithValue("@WorkerName", WorkerName);
                        command.ExecuteNonQuery();
                        return true;
                    }
                }
            }
            catch (SqliteException ex)
            {
                Console.WriteLine("SQLite Error: " + ex.Message);
```

الملخص

وجود الماء أمر حيوي لأنه ضروري لجميع الكائنات الحية. ويواجه الفلسطينيون حاليا، بسبب مخصصاتهم الثابتة والمحدودة، عددا من العقبات في سعيهم للحصول على مياه الشرب الكافية والمأمونة. لدى العديد من البلديات بيانات عن المياه، لكنها غير مرتبة بطريقة تجعل صنع القرار وحل المشكلات أمرا سهلا . لذلك، يجب جمع جميع بيانات المياه وتحريرها وإدارتها من أجل تحليلها، ويمكن استخلاص معلومات مفيدة منها.

إدارة الموارد المائية هي تخطيط الموارد المائية وتطويرها وتوزيعها وإدارتها لضمان استخدامها على أفضل وجه. ويزداد وعي شبكة توزيع المياه من خلال استخدام نظم المعلومات الجغرافية. باستخدام تحليل شبكة المياه ونظام المعلومات الجغرافية، يمكننا تحديد مواقع فواصل الأنابيب وتحديد هويتها بمجرد حدوثها. بالإضافة إلى التعامل مع جميع البيانات المتعلقة بشبكات إمدادات المياه المتفرقة في فلسطين ودول أخرى،. هدفي هو تحديد موقع نظام يمكنه تقييم نظام توزيع المياه في نابلس.

تم جمع البيانات من منطقة الدراسة (المنطقة الصناعية) لأطروحتي. جميع البيانات التي جمعتها لم تكن بيانات نظام المعلومات الجغرافية ؛ بدلاً من ذلك، تم تحويلها من تنسيق AutoCAD إلى بيانات GIS، مما يعني أنها تضمنت العديد من المشكلات التي تحتاج إلى إصلاح. يجب تنظيم البيانات حتى يقوم المطور بإجراء التحليل. لقد أجريت العديد من التعديلات على البيانات، بما في ذلك تحديد الحواف والتقاطعات البسيطة والمعقدة والمتصلة. تتكون بيانات المياه من عدادات العملاء (وجهة المياه)، وبيانات المصدر، وبيانات الصمامات، وبيانات الأنابيب (الأنابيب الرئيسية، وأنابيب الوصلات المنزلية). وتم الجمع بين برمجيات ArcMap ونظام المعلومات الجغرافية لإنشاء شبكة مياه تدعم نظام الإمداد المتقطع. الغرض من الشبكة الهندسية هو تحديد الحلقات والمسائل الأخرى مثل العثور على عناصر الشبكة المنفصلة أو المتصلة، وتحديد مواقع أجزاء الشبكة المنفصلة ومعرفة كيفية إعادة توصيلها، وكذلك إجراء تحليل الشبكة والتحقق من الاتصال بين مكونات الشبكة.

للحصول على نظام توزيع المياه العادل، كان من الضروري أولاً ضمان سلامة وصحة توصيل مكونات الشبكة. ثانيا، إنشاء نظام للتنبؤ يستند إلى البيانات التاريخية لعمليات الصيانة، استخدم هذا النظام لغة برمجة #c لتطبيق خوارزميات استخراج البيانات، واستخدمت برنامج ArcMap، لتطبيق تحليلات تعدين البيانات المكانية التي تأخذ في الاعتبار الإحداثيات الدقيقة لكل مكون والعلاقات المكانية للمكونات المحيطة

به، أخيرًا، يتم فحص الشبكة باستخدام خوارزمية (Breadth First) وخوارزميات العزل (isolation) لتحديد الصمام المتحكم في كل مكون من مكونات الشبكة، التي يقترح النظام الذكي بعد ذلك إغلاقها في حالة تعطل العنصر ذي الصلة.