EasyChair Preprint

№ 9230

Leakage Detection of Pipelines in Water Distribution Network Using Machine Learning Algorithm

M Saravanabalaji, S Shakthi Raagavi, E Praveen and S Samritha

November 3, 2022

LEAKAGE DETECTION OF PIPELINES IN WATER DISTRIBUTION NETWORK USING MACHINE LEARNING ALGORITHM

Saravanabalaji.M, Shakthi Raagavi.S, Praveen.E, Samritha.S
saravanabalaji.m.eie@kct.ac.in,shakthiraagavi.20ei@kct.ac.in,
praveen.20ei@kct.ac.in, samritha.20ei@kct.ac.in
Department of Electronics and Instrumentation Engineering
Kumaraguru College of Technology, Coimbatore

INTRODUCTION

Technology plays a vital role in our day today life. We are searching for the best solution to solve our regular problems with less effort[1]. One of the emerging techniques that changes the way we share and store information is cloud technology simply called as cloud. One of the largest advantage of using cloud is that it allows users to develop IoT application. The efficiency, accuracy and speed of IoT applications will be achieved by using cloud computing. There is a saying in the internet called "If you have not been living beneath the land, then you must have heard about IoT (i.e. Internet of Things)". This statement tells us how important IoT is! IoT along with Wireless Sensor Networks (WSN) and small nodes that communicate by wireless links can be used to collect information of various operations, conditions and performance of any system / operation / task / environment that is far away from the controlling unit i.e. remotely controlled[2]. One such remotely controlled scenario is water distribution networks (WDN). Water distribution network is the source of water supply in many cities in the globe. Water is a pivotal natural resource which is broadly mishandled by human beings. A report says around 40% of clean water is wasted due to leakages in the WDN[3]. According to McKenzie in South Africa (so called a water – stress country) due to its limited water resources, steadily increasing population and growing water demand loss / leakage of water plays an important role as its municipal water service is threatened with an appraised value which is more than R7 billion per annum [4]. Thus it's important to conserve water loses in the form of leaks in WDN. We will further discuss leakage detection techniques using Machine learning algorithms.

METHODOLOGY

EPANET is open source software used for modeling distribution networks and monitoring water quality issues. EPANET is used to generate a dataset of different pressure values in each node of WDN. By increasing the value of emitter coefficient we can design leak in the WDN and be able to obtain pressure drops in each node. From this data we can develop models using different machine learning algorithms.

PROPOSED SYSTEM

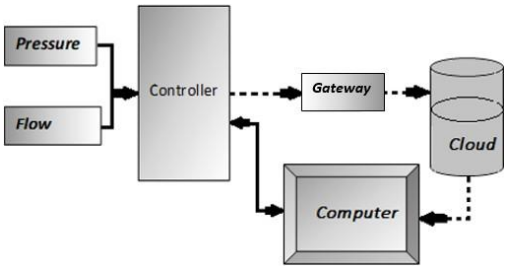

Fig.1.Block diagram of proposed model

COMPONENTS

A leakage detection system for monitoring water is designed for water distribution network. Our water leakage detection model to detect leakage comprises of the following components:

Flow/pressure sensor

A wireless sensor is placed in pipelines at a fixed distance representing nodes of our model and pressure / flow at that point is monitored and transmitted at regular intervals.

Gateway

G ateway is a net work node that connec ts different networks with different transmission output together. Mainly used for the purpose of data lagging and data transfer.

WSN

Wireless Sensor Network (WSN) is a wireless network that is deployed in large number of wireless sensors to monitor the system or environment. Sensor nodes in WSN are provided with on-board processor that helps to manage and monitor environmental conditions. These sensor nodes are connected to the base station to share data.

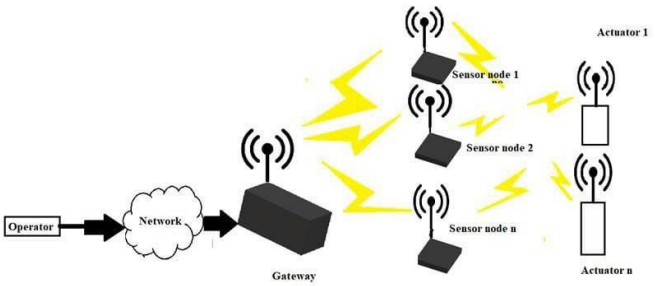

Fig.2. Wireless sensor network

CLOUD

Cloud is nothing but delivering different types of services over the internet. Starting from software, analytics to secure and safe data storage and networking resources everything can be delivered by using cloud. Thus it stores every data transmitted by wireless sensors and sends it to computer to perform ML algorithms.

Implementation

Implementation of all these things explained above will be pictorially represented as:

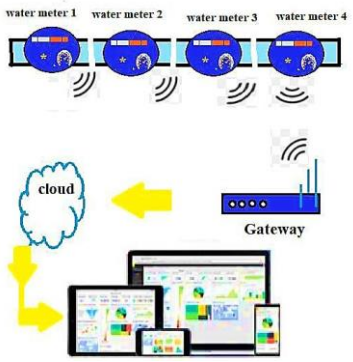

Fig.3. Pictorially representation for cloud platform

WATER DISTRIBUTION NETWORK:

A Water Distribution Network is a hydraulic infrastructure that contains a complex assembly of hydraulic control elements together to distribute quantities of water from sources to consumers that are implemented to face peak demands. The main focus of a WDN is to supply drinking water to a group of people with suitable pressure, quantity and quality. The four principle techniques to design a WDN are

- Dead end
- Gridiron
- Circular
- Radial systems

A water distribution network consist of components such as reservoirs, tanks, pumps, pipes and valves.

Reservoirs: Reservoirs are the source of water from which water is collected and distributed. For Example Rivers, lakes, bore wells etc.

Tank

Depending on demand and need the choice of choosing number of storage tanks may vary. Tanks are temporary storage units where the water transmitted from reservoirs are stored for distributing in times of peak demand.

Pumps

Pumps are electrical components that are used to pressure up the water in WDN during distribution.

Valves

There are various types of valves including stop valves, gate valves, check valves and so on that are used in WDN for different purposes like controlling direction of water and soon.

Pipes

Pipes of different diameters are used to pump water to different areas in WDN.

Most commonly used method in WDN was gravity and pump where the water from the reservoir is stored in tank and is pumped to the user in time of demand or need [5].

LEAKAGES

Some quantity of water or gas that get away from the pipe or a closed container due to crack, hole or any faults is called as Leakage.

Causes

Leakage in pipeline may occur due to following reasons:

- High water pressure – when pressure of the water flowing through the pipe increases or direction of flow of water changes suddenly pipes can't be able to bear, hence leads to damage in wear and tear, bursting or may result in leaks.
- Tree roots – tree roots burrow their way to pipe lines buried in ground where water is leaked in form of vapour due to cracks.
- Corrosion – still copper or galvanised steel piping is used in some older homes, where there are possibilities of rusting may lead to leak.
- Differential temperature – crack and leak may be caused due to temperature changes also, especially at winter due to extremely cold weather.
- Shifting soil under the surface.
- Ground freezing and thawing.

"A drop a day wastes the water away"

According to the above saying we don't like to hear the drip, drip sound. Even size of leak is small it grabs our attention. We never know the worth of water till the well is dry. But every problem has its own solution, and here is an idea to solve this problem too.

EPANET

The prior thing required to build a model that detec ts leak age in WDN system is dataset (containing hydraulic details such as direction of flow of water in the pipe, pressure of water at each node, emitter property, velocity, elevation, and so on)that is used to train and test the model. But unfortunately the dataset is not leaked anywhere for security purposes. To overcome this and to generate dataset we move on to a software called EPANET. It is a free software application that helps students and research people to model water distribution system and monitor water properties throughout the EPANET network allows users to perform simulation hydraulic and water quality behaviour of different water distribution networks that consist of reservoirs, tanks, valves, pumps, nodes, and pipes.

Fig.4.Water detection network using EPANET is our WDN designed using the software. This software allows user to design parameters of their own and observe the simulation and view results in various formats like time series graphs, data tables, charts, colour – coded network maps, energy usage reaction, calibration, and profile and counter plots. Although the software doesn't have a direct tool to design Leak in the network, we can use a property called emitter coefficient actually designed to model fire hydrants and sprinkler can be used to design leak at required nodes. If emitter coefficient was zero, there was no leak at the nodes (junction),if emitter coefficient is greater than zero then there occurs a leak. Thus by changing the emitter coefficient leak is designed and dataset is generated [6].

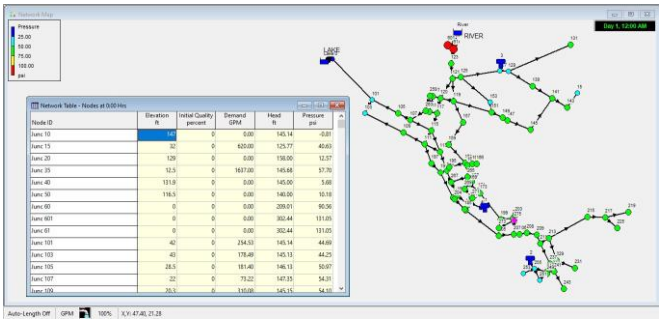

Fig.4.Water detection network using EPANET

MACHINE LEARNING

Machine lear ning is a subset of Ar tificial intelligence, which rests on the concept that system can learn from input data, identify patterns, attachment or relationship between them and makes decision with least human participation [7]. In other words machine learning is one the most exciting technologies that gives computer potential to learn without being explicitly programmed. As it is evident from the name: The ability to learn, make computer similar to humans [8]. It was started from pattern recognition and the theory which makes the computer to learn without being programmed to do a specific task. But now extended its arm over wide range of applications as they make computer learn from previous computations to provide accurate, reliable, repeatable decisions

and results. In greater fields like fraud detection, spam filtering, business process automation, predictive maintenance, financial ser vices (like banks and various financial industries to identify important insights in data and prevent fraud etc.), health care (in health care industries to help experts analyse and identify things in order to improve diagnosis and treatments), retail (in websites like Netflix and Amazon to provide recommendation by analysing history, for customer insights and so on.), transportation (in delivery companies and public transportation to analyse data to identify patterns and make routes shortest and efficient and soon), oil and gas (to identify resources, streamlining oil distribution etc.), government (in government agencies, public safety for insights, detect fraud, minimize identity threat etc.) and still expanding its application over large fields [7].

Types of machine learning

Machine learning is classified based on how the algorithm learns to predict the output more accurate and reliable. The four basic types are:

- Supervised learning
- Unsupervised learning
- Reinforcement learning

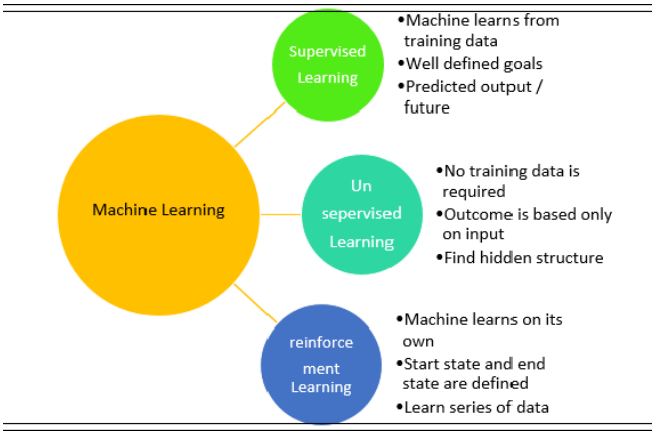

Fig.5.Type of machine learning

Supervised learning

Supervised learning algorithms are used when labelled dataset is available for training the model (Labelled dataset contains input along with its desired output). The algorithm learns the relationship between input parameters and the output, from the dataset and then compares the actual output with correct output to find errors and finally adjusts the model accordingly to get best results [7]. At the end, the algorithm learns how the data works and relationship between input and output. Then the algorithm uses this learning to predict the output of additional unlabelled data. Thus super vised leaning algorithm is used for discovering new patterns and relationships [6]. Supervised learning uses methods like:

- Classification
- Prediction
- Regression
- Gradient boosting

For learning patterns of labelled dataset and predicting the output of unlabelled data. Hence mostly used in application where historical data used for predicting similar future events [7].

Unsupervised learning

Like supervised leaning, there is no labelled dataset in unsupervised learning to predict the output but in contrary it creates hidden layers and adjusts it dynamically to predict the output. Unsupervised learning algorithm explores the dataset (usually unlabelled dataset) and checks for possible connections or structures within input parameters in order to predict the output. The techniques used to predict output are:

- K-means clustering
- Nearest –neighbour
- Singular value decomposition
- Self-organizing maps

Few application of unsupervised learning include item recommendation, customer segmenting, clustering variables and so on.

Reinforcement learning

Reinforcement learning almost behaves like humans (i.e. it makes the model to find best-suited path or possible behaviour it should take in a specific situation). It is used to train a machine or model to per form a multi-step process that have a set of clearly defined rules. This algorithm trains a model to make a sequence of decisions to take suitable action in order to maximize rewards in a particular situation. In reinforcement learning the algorithm uses trial and error method to finds solution that owes the highest rewards. Thus reinforcement learning works by putting the algorithm in an environment

that has an interpreter and a reward system. In every iteration, the output of the algorithm is given to the interpreter which decides whether the output is favourable or not. If the output is favourable the algorithm is encouraged with a reward else it is discouraged with punishment. Reinforcement learning differs from supervised learning in a way that supervised learning has dataset with answer key to train the model to predict output of unlabelled data whereas in reinforcement learning there is no labelled dataset provided by the programmer but reinforcement agent decides what action to perform next in the given task or environment[9].

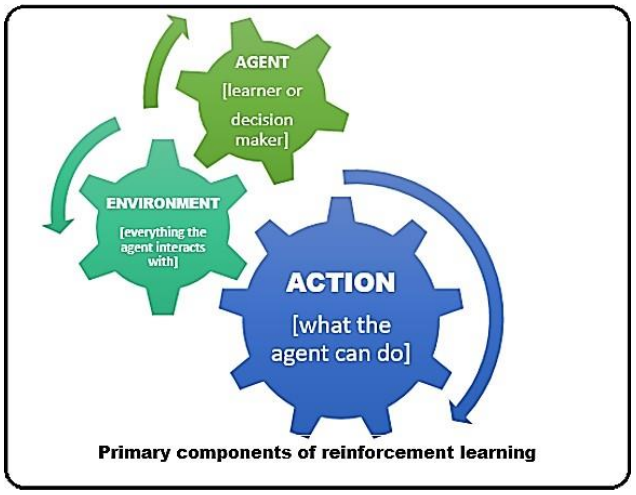

Fig.6.primary components of reinforcement learning

Thus this type of algorithm is used for robotics, gaming, and navigation and so on.

Advantages

- Almost every process can be automated by using machine learning, thus reducing time and work.
- Today ML can be applied everywhere from medical, business, bank ing, cust omer interaction to science and technology.
- Helps improve both software and hardware.
- Have capability to deal any type of data.
- Can be implemented in eduction and shopping too.

Disadvantages

- Although ML can automate ever ything selection of ML algorithm has to be done manually.
- In case of advanced and large data the system will take more time to process.
- Possibility of high error.
- Data acquisition.

Steps for choosing an appropriate machine learning model

- Select and align the input that are to be considered for the solution of the problem.
- Collect, format and label the data inputs if it's necessary.
- Choose a appropriate ML algorithm and test for its accuracy.
- Tune the output to get the required level of accuracy.

Algorithm chosen

Algorithms chosen for our problem considered (leakage detection in WDN) are:

- Logistic regression
- Support vector machine
- Decision tree

Logistic regression

L og istic r eg r ession is a t ype of binar y classification used for modelling the probabilities of problems with two possible outcomes. We can say that it is an extension of linear regression model for classification problems. Logistic regression uses a more complex cost function called " Sigmoid Function" or also known as " Logistic Function". Logistic function is defined as:

One of the advantages of logistic regression model is that in addition to binary classificatio it also gives you the probabilities. O ther classification output will tell you whether the output is true or false (0 or 1) but it will also gives you probabilities (i.e. for knowing an instance has

$$\text{Logistic}(x) = \frac{1}{1+e^{(-x)}}$$

51% of probability for a class compared to 99% is not the same it makes a big difference). But when it comes to interpretation it struggles with restrictive expressiveness.

Support vector machine

SVM is a type of supervised ML method used for problems like classification, regression and outliers detection. SVM model is nothing but the representation of different classed in hyperplane

in multidimensional space. The hyperplane will be generated in an iterative manner in order to minimize the error. Usually this model uses classification algorithms for two group classification problems.

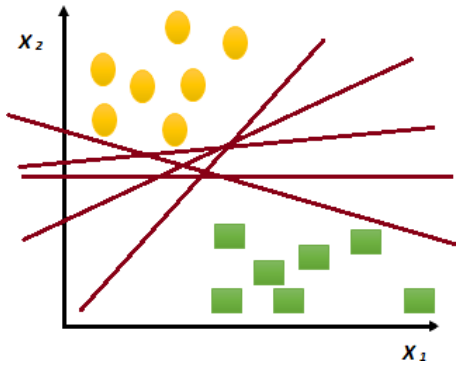

Fig.7.SVM graph model

To separate the two class of data points, we may choose many possible hyperplanes. But our objective is to find a hyperplane with minimum margin. I f i t has minimum marigin then the distance between data points of both classes will be maximum. Hence maximizing the margin distance will provide some reinforcement to classify the data points with higher accuracy in future. Two different types of SVM used are:

- Simple SVM and
- KernelSVM

Decision tree

Decision tree is a type of supervised ML technique that can be used to solve both regression and classification problems. But mostly used to solve classification problems. This algorithm is names so as it is similar to a tree i.e. starting with root it expands multiple branches and hold leaves and constructs a tree like structure. Thus Decision tree algorithm uses tree representation to solve the problem where the data is split continuously by a certain parameter.it contains two nodes called “Decision Node” and “Leaf Node”. Decisions are made in Decision Nodes and it contains multiple branches where as Leaf nodes represents final outcome of the decision made at decision nodes and it does not have further branches. Structure formed during splitting a node into branches is called sub-tree. Representation of structure of Decision tree algorithm is shown below:

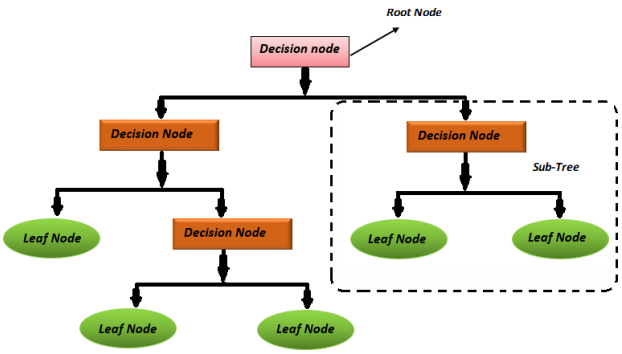

Fig.8.Structure of decision tree algorithm

12. RESULT

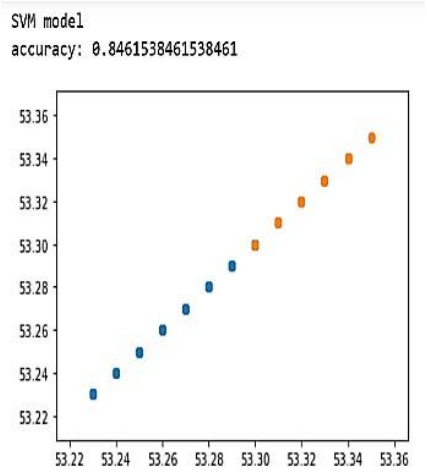

Fig.9. SVM Model result

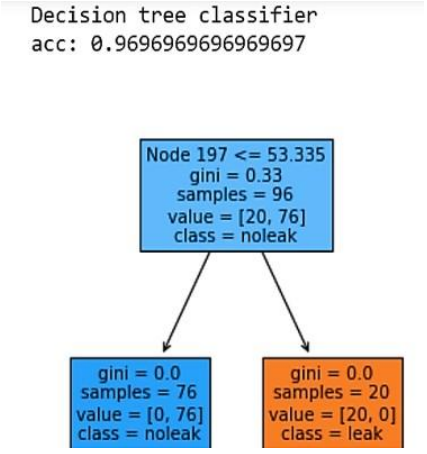

Fig.10.Decision tree result

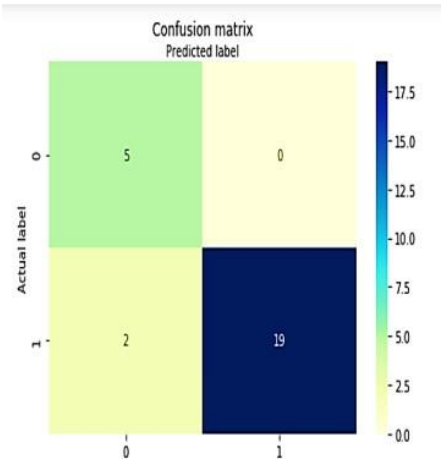

Fig.11.confusion matrix

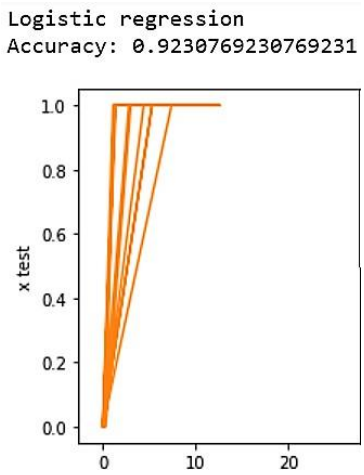

Fig.12.logistic regression result

CONCLUSION

From graph theory, it says that concept of nodes based on algorithms. Algorithms find the number of sensors and location in pipeline networks. For leakage detection, they have a set of patterns and regularities in supervised machine learning algorithms. The algorithms provide a better resolution for water leakage detection. Here we build models on algorithms like SVM, Logistic Regression and Decision Tree. We get better accuracy with Decision Tree. By using these algorithms accuracy can be achieved above 84% which is enough to detect the leak. The obtained results are found to be good in the proposed technique.

FUTURE SCOPE

This is an small approach to a problem and it can be extended further to solve many real life problems like leakage detection in oil pipelines that are transported over longer distance where there is a chances that leaks are unnoticed in remote areas, to find blocks in sewages in developing countries where it is still done manually by daily wage workers, water quality monitoring in WDN in order to prevent drinking water contamination in many rural areas and any liquid transportation in underground pipelines and so on.

REFERENCES

[1]. Al-fuqaha, A.; Member, S.; Guizani, M.; Mohammadi, M.; Member, S. Internet of Things: A Survey on Enabling . IEEE Commun. Surv. Tutor. 2015, 17, 2347–2376

[2]. IoT. IoT Application Areas. Available online: https://www.fractal.com/en/blog/the-9-most-importantapplications- of-the-internet-of-things (accessed on 5 November 2020).

[3]. Joao Alves Coelho,Andre Gloria,Pedro Sebastiao,"PRECISEWATER LEAK DETECTION USING MACHINE LEARNING AND REAL-TIME SENSOR DATA" published in 3 december 2020 on MDPI,474-493

[4].R. Mckenzie, Z. N. Siqalaba and W. A. Wegelin, " The state of non-revenue water in South Africa", 2012.Show Context Google Scholar

[5].https://www.fs.fed.us/eng/pubs/pdf/waterguide/hi_res/distribution.pdf

[6]. Reference - GAGAN GUPTA, "Monitoring Water Distribution Network using Machine Learning", Published on 2017:163 ISSN 1653-5146

[7]. https://www.sas.com/en_in/insights/analytics/machine-learning.html#:~:text=Machine%20learning%20is%20a%20method,decisions%20with%20minimal%20human%20intervention.

[8] Website https://www.geeksforgeeks.org/machine-learning/

[9]. https://www.geeksforgeeks.org/what-is-reinforcement-learning/

[10].Eduardo Oliveira, Mário da Fonseca, Daniele Kappes, "Leak Detection System using Machine Learning Techniques" published on June 2018 in an international conference on Automation in mining.325825850

[11].Li Cheng, Du Kuna, Tu Jia-pengb, "Optimal placement of pressure sensors in water distribution systems based on clustering analysis of pressure sensitive matrix", published on XVIII International Conference on Water Distribution Systems Analysis, WDSA2017.

[12].Thilagaraj.M, Stephy Akkara, Aakil Habibullah. M, "Water Leakage Detection and Management System Using IOT", published on International Journal of Advanced Science and Technology
Vol. 29, No. 9s, (2020), pp. 683-689

[13].Sewerin. Stethophon 04—Compact Sound Detector for Water Leak Detection. Available online: https://www. sewerin.com/fileadmin/redakteure/Prospekte/pro_stethophon_04_en.pdf (accessed on 5 November 2020).

[14].Electrics, G. AquaTransTM AT868Panametrics Liquid FlowUltrasonic Transmitter. Available online: http: //www.panametria.cz/produkty_pdf/AT868br.pdf (accessed on 5 November 2020).

[15].Ross, M.; Graves, C.A.; Campbell, J.W.; Kim, J.H. Using support vector machines to classify student attentiveness for the development of personalized learning systems. In Proceedings of the 2013 12th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 4–7 December 2013. [CrossRef]

[16].Saravanan, R.; Sujatha, P. A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification. In Proceedings of the 2nd International Conference on Intelligent Computing and Control Systems (ICICCS), Secunderabad, India, 27–28 June 2019. [CrossRef]

[17]. Mamdouh, M.; Elrukhsi, M.A.; Khattab, A. Securing the Internet of Things and Wireless Sensor Networks via Machine Learning: A Survey. In Proceedings of the 2018 International Conference on Computer and Applications (ICCA), Beirut, Lebanon, 27–28 June 2018. [CrossRef]

[18].Li, L.; Chou, W.; Zhou, W.; Luo, M. Design Patterns and Extensibility of REST API for Networking Applications. IEEE Trans. Netw. Serv. Manag. 2016. [CrossRef]

[19].Kazeem B. Adedeji; Yskandar Hamam; Bolanle Tolulope Abe; Adnan M. Abu-Mahfouz “Towards Achieving a Reliable Leakage Detection and Localization Algorithm for Application in Water Piping Networks” published on 27 September 2017

[20].J. Thornton, R. Sturm and G. Kunkel, Water Loss Control, New York, NY, USA:McGraw-Hill, 2008.

[21].R. Mckenzie, Z. N. Siqalaba and W. A. Wegelin, "The state of non-revenue water in South Africa", 2012.

[22].XunWang Mohamed S.Ghidaoui “Identification of multiple leaks in pipeline: Linearized model, maximum likelihood, and super-resolution localization” published on 17 February 2018.