# TABLE OF CONTENTS

# ABSTRACT

Leakage detection and localization in pipelines has become an important aspect of water management systems. Since monitoring leakage in large-scale water distribution networks (WDNs) is a challenging task, the need to develop a reliable and robust leak detection and localization technique is essential for loss reduction in potable WDNs. Some of the existing techniques for water leakage detection are discussed and open research areas and challenges are highlighted. It is concluded that despite the numerous research efforts and advancement in leakage detection technologies, a large scope is still open for further research in this domain. One such area is the effective detection of background type leakages that have not been covered fully in the literature. The utilization of wireless sensor networks for leakage detection purposes, its technical challenges as well as some future research areas are also presented. In a general remark, practical application of these techniques for large-scale water distribution networks is still a major concern. In this paper, an overview of this important problem is addressed.

# ACKNOWLEDGMENT

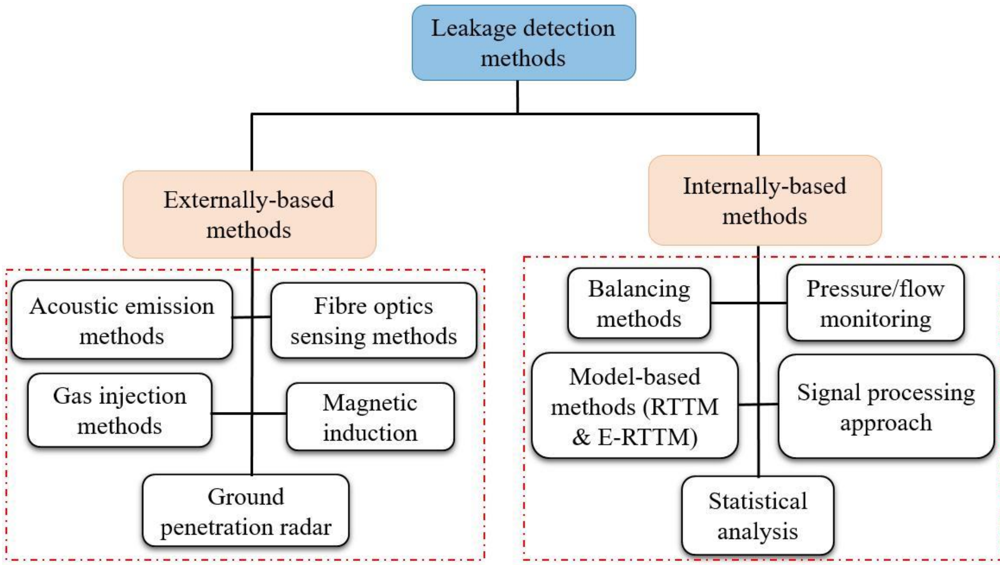
We would like to express deep gratitude to our guide **Asst. Prof. DR. U. SRINIVASULU REDDY** for his constant guidance, encouragement and gracious support throughout the course of our work. For their motivation that encouraged us to work in this area and for their faith in us at every stage of this research.
We would like to thank all the students and staff of the Spastic Society, also I would like to thank our fellow group mates for their co-ordination which lead to the successful completion of the project. Lastly, we would like to extend our sincere thanks to our Academic Dean and Project Coordinator **Mr. Balaji Chandrasekaran** for their valuable insights and support throughout the project.

# CHAPTER 1: INTRODUCTION

## 1.1 MOTIVATION

Pipelines transporting are part of a country's important assets, contributing to the nation's economy. These pipelines which are usually installed to serve cities across the country, experience failures along their length. The failure of pipelines is usually attributed to the aging infrastructure, severe environmental conditions and third party damage. As a result, a portion or parts of a pipeline wall is perforated over time, thus causing leakage. When this happens, a loss (for instance water) through leaks is observed. Water loss through leakages is recognized as a costly problem worldwide, due to the waste of precious natural resources, as well as from the economic point of view. Survey work conducted revealed that almost 20% of the US water supply is lost through leaking pipes. In South Africa, a water-scarce country with limited water resources and steadily growing water demand, high water losses threaten its municipal water service with an estimated value of more than 7 billion annually. Furthermore, the global demand for water is increasing due to improved standard of living, and resources are diminishing. In the last few years, this has rendered the pipeline leakage detection problem increasingly prominent. Leakages through piping network can be classified as reported, unreported and back-ground type leakages. The former often surfaces on the ground and reported by the public or utility personnel. This type of leakage such as pipe burst is detectable by applying an appropriate leakage detection technique discussed in the next sections. The unreported leakage often does not surface on the ground but in a similar manner to the reported one, can be detected applying a suitable leakage detection method. When both type of leakage (reported and unreported) occur in piping networks, pressure reduction is usually noticed at the downstream of the pipes. The background type leakage is not characterized by pressure drop and are difficult to detect. These types of leakages are hidden and run continuously in distribution networks. Since they are difficult to detect and go unreported for a very long period of time, such kind of leakages posed the largest threat to water distribution networks. Nonetheless, modelling a distribution piping network to analyze rows through the pipelines including background leakage rows should give a remarkable breakthrough in detecting such kind of leakages.

Numerous studies, reports and standards that deal with this problem have been published. Of particular interest are the various standards and regulations developed. In order to obtain a good leak detection system, API 1555 sets performance requirements. These include sensitivity, accuracy, reliability and robustness. Further research efforts led to the classification of leakage detection systems based on their technical nature as internally-based and externally-based system as illustrated in Fig, each having relative advantages and certain limitations.




The traditional leakage detection technique requires huge human involvement due to visual inspection by personnel, and thereby, is time consuming, labor intensive, and has low reliability in detecting leaks. Moreover, externally-based methods use sensors installed outside the pipe to generate a leak alarm. System costs and complexity of installation are usually high for this type of leakage detection methods and therefore, their application is limited to high-risk areas. Internally-based systems on the other hand, use sensors to monitor internal pipeline parameters such as pressure and temperature. These signals are used for inferring a leak.The review of some leakage management involving pressure reduction methodologies as well as wireless sensor networks for water networks are presented. It is important to emphasize that the focus of this study is only on pressurized piping systems. Previous and recent research progress as well as some open research area are discussed. Technical research challenges and design specifications using WNS for leakage detection and localization in piping networks are also discussed.

## 1.2 OBJECTIVES

The key objectives for the project were identified as:

- To develop an optimal technique to find a leak in the flow.
- Perform prediction analysis to predict a leakage on a given dataset.
- To minimize the loss or damage due to the leakage of water, oil, gas, etc.
- To reduce the damage when a leak arises.
- To create an alert system.

# CHAPTER 2: LITERATURE REVIEW

## 2.1 Machine Learning

Machine Learning is a class of methods in data analysis that learns patterns and hidden insights in the data without being explicitly programmed for it. Thanks to the better and more powerful computing devices, there is an increasing trend to apply ML in various cases recently, such as equipment failure detection , pattern and image recognition, email spam filtering, and fraud detection.ML algorithms performs predictive analysis. When the output variables takes continuous values, it is termed as Regression whereas when it takes class labels it is called Classification. Regression refers to estimating a response whereas Classification refers to identifying group membership.

Machine learning techniques have shown to be powerful tools for different purposes. Roughly speaking, their key objective is to automatically extract information from data, using statistical methods. Inference processes might be inductive-deductive or transductive; the former builds a function model, while the latter not. In any case, such tools require a set of training data in order to make their predictions. According to the desired outcome, the machine learning algorithm could be based on:

**Supervised learning**: a function is generated to map inputs to desired outputs according to training data.

**Unsupervised learning**: patterns are sought based on self-organization of data.

**Reinforcement learning**: the learning process involve an interaction with the environment in order to adapt its behavior.

When used for classification, supervised learning is perhaps the most popular choice. For this purpose, algorithms are trained in order to classify data as 1 (belong to a given class) or 0 (it doesn’t). This classification is called binary classification. When a multiclass classification is required (several classes) then different approaches are used, mainly based on several binary classifications.

## 2.2 Problem Formulation

The main goal of this thesis work is to come up with an appropriate solution for automatic detection and localization of pipeline leakages, and estimation of the size of leakages in water distribution networks. Thesis objective can be further sub-divided into two distinct problems. First, it is required to come with a judicious rule to place the sensor node in the water distribution network. Secondly, it is required to find the best way to use various ML techniques on the obtained data to find out the leakage information in the network. In the following subsection both of these problems are discussed in detail.

### 2.2.1 Additional Information

Leakage in a distribution systems can be caused by several ways. It could be due to faulty connection and corrosion of the pipe, high system pressure, or it could be due to damage caused by ground movement in winter, excavation, or even due to poor quality of workmanship. Many practical deployed water leakage detection systems measure various hydraulic and acoustic parameters such as flow, pressure and water quality parameters.

Out of these, pressure change due to leakage is the most noticeable and hence it is of primary interest for localizing leaks in the network. Leakage of water from the pipe is directly proportional to system pressure on the pipe. Since the pipes in a WDN are pressurized, events such as pipe bursts, valve opening/closure results in a sudden change in the flow through the pipe causing a pressure transient that propagates along the the pipelines.

The pattern of the pressure changes can be analyzed computationally to detect the location and the size of leakage in real-time. In this section, a short literature review of the available leakage detection methods are provided.

Hardware based methods using acoustic or non-acoustic equipment are existent. however, such methods are expensive, time consuming, and labor intensive. Then there are classified transient- based methods such as detecting negative pressure waves caused by pipe bursts. still, the method fails to locate leakage correctly when pipe burst near boundary of pipe-joint and also suffers with false alarms caused by normal pipeline operation. Alternatively, there also exists a methodology called 'Inverse Transient Analysis (ITA). In this technique, system state is already known but system parameters are not known. The inverse problem is solved for parameters such as leak size and its position. In simple words, leakage detection are also performed through the analysis of subsequent behavior of burst-induced transient signals. However, the burst-induced pressure signals may be masked by background noise or other events in the network.

Model based leak detection is also one of the ways in which researchers have approached this problem. Using this way, the problem can be formulated into a classical least square estimation problem. Nonetheless, due to relatively small number of pressure measurement readings available, it is a challenging task to solve for complex non-linear models of any city's WDN. The other techniques that have gained traction in recent times is the use of pattern recognition in the data to find the leakage in the distribution system.

Several data driven approaches to solve this problem have been published in the area of leakage detection in WDN.

This article have classified the data-driven approach into three categories a) Classification method b) Prediction-Classification method and c) Statistical Method. Methods that use classification techniques on the physical parameters are placed in the first category. In Prediction-Classification methods, classification is carried out after a stage of prediction. In statistical method, the burst detection completely relies on statistical theory such as 'Statistical Process Control'. In this thesis work, Classification methods are used to detect and localize the leakage in the system and the literature review corresponding to it is mentioned in the following paragraph.

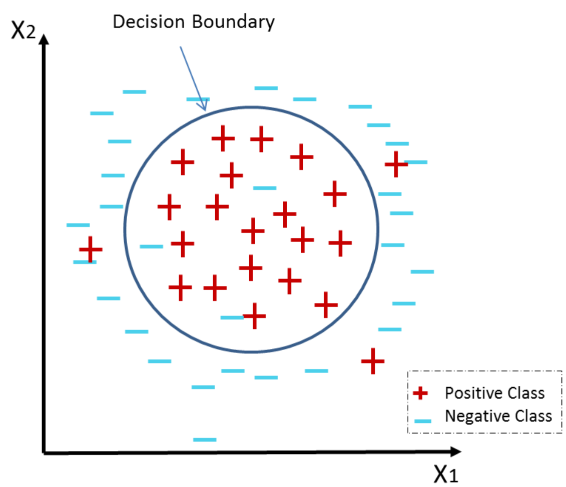



**Figure 1**: Binary Classification: *A circular (Non-linear) decision boundary depicts the region belonging to the two class. Any future data points lying within the marked circle would be classified as positive and vice-versa.*

A leak detection using analysis of monitored pressure values is performed by support vector machines. However, it places the monitoring nodes in the network randomly and does not provide in- formation on placement of sensor nodes. A Novelty detection is performed based on WDN time series data of pressure and flow values. This novelty could include variety of events such as pipe bursts, hydrants flushing and sensor failure. This detects the leakage but does not provide information on the localization of leakage, a data-driven novelty detection system was used to solve the problem of leakage detection using Multiregional Principal Component Analysis (PCA). However, a major limitation comes from the linear nature of the method that applies PCA. As a result, the system led to limited sensitivity or a high false alarms rate. From IoT perspective, it is desirable to transmit limited data to the central hub. In this regard, a study is done to extract the feature from the pressure data in the water distribution system. However, how to use the information further in WDN framework has not been looked upon. In the next section the problem formulation for the thesis work is provided.

With the help of water monitoring agencies, water supply data to various localities in a particular area can be gathered and further this project can be extended by the usage of SVM (Support Vector Machine) and RVM (Relevance Vector Machine) techniques. Which, helps in the field of Prediction Analysis.

## 2.3 Experimental setup

An experimental water pipeline model with a length of 2 ft.-long plastic pipes with a radius of 1 cm has been implemented with three wireless water flow sensors are equipped at three different locations. The system has been designed to allow changing the deploying locations of the sensors by adding extra connection points, and these have been used during test processes. In order to supplement pressurized water into the system, a pump motor of 1.5 HP power and a hydrophore unit to adjust water pressure to five bars are connected. Sensed water pressure data from the wireless nodes have been transferred to a computer through base station. And the data have been recorded and analyzed at this computer by a multi-label classification approach. If any leak and leak location is detected through the epochs notification.

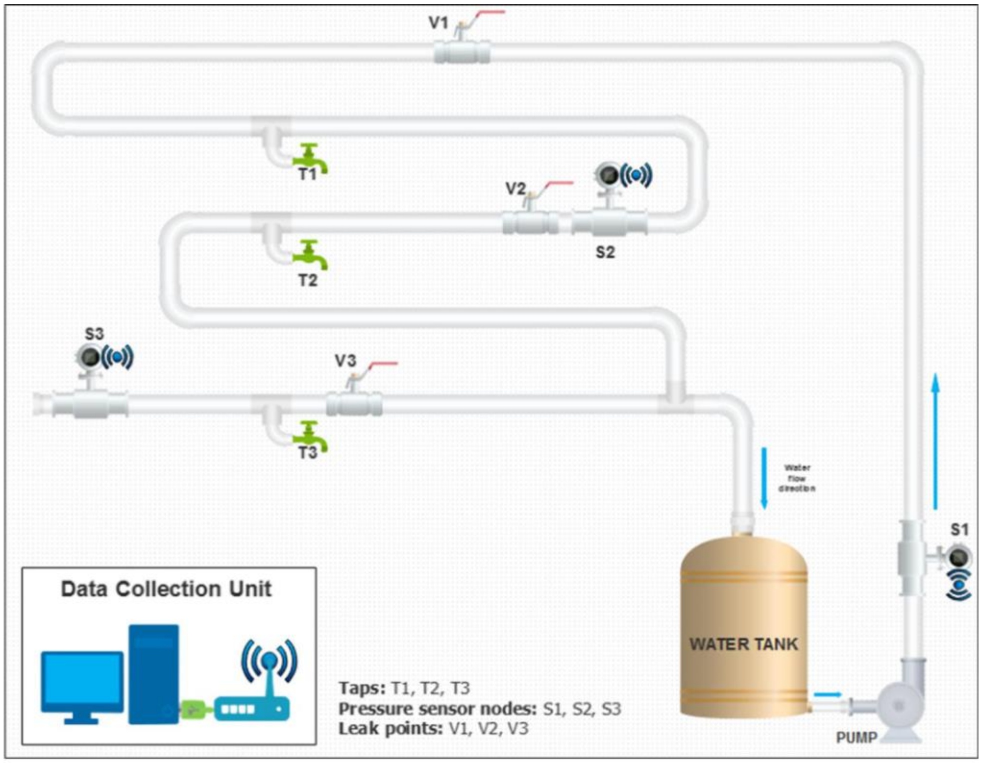



**Figure 2**: Plan of experimental pipeline system

In Figure. 2, taps are illustrated as T1, T2, and T3 to simulate water uses on pipeline system. Besides, wireless pressure sensors are illustrated as S1, S2, and S3. And V1, V2, and V3 illustrate valves to be used as leak points.

**Figure 3**: Water Flow Rate Sensor

## 2.4 Data collection

Three wireless pressure sensor nodes are connected to the system to gather volume data of the pressurized water in the pipes with a period of six seconds. The data are sent through a base station to the computer and stored in a database to be used for analyzing. The locations of the sensor nodes, water pressure data collecting time period, manually created leakage, and normal water usage locations through taps are changed many times to test and validate the results of the system. The system also has the potential to monitor other types of data which can be wanted on water pipeline systems if suitable sensors are connected like chlorine, pH, and turbidity.

| Signal no. | Sensor Values (mL/min) | | | Leakage symbols (leak=1, no leak=0) | | |
|---|---|---|---|---|---|---|
| | *S*1 | *S*2 | *S*3 | *V*1 | *V*2 | *V*3 |
| 1 | 5064.04 | 5319.218 | | 5542.8960 | 0 | 0 |
| 2 | 2977.427 | 2556.998 | | 2904.6441 | 0 | 0 |
| 3 | 3057.01 | 2808.066 | | 3062.9711 | 0 | 0 |
| 4 | 3181.466 | 2883.4 | | 3112.1691 | 0 | 0 |
| 5 | 3265.07 | 2935.469 | | 3166.2281 | 0 | 0 |
| 6 | 3301.746 | 2951.87 | | 3193.6621 | 0 | 0 |
| 8 | 5239.3 | 5494.564 | 2923.33 | 0 | 0 | 0 |
| 9 | 3368.805 | 1853.396 | | 2179.9360 | 1 | 0 |
| 10 | 3973.977 | 2192.166 | | 2459.9730 | 1 | 0 |
| 11 | 4412.585 | 2394.455 | | 2645.3290 | 1 | 0 |
| 12 | 4761.109 | 2604.703 | | 2793.7140 | 1 | 0 |
| 25 | 4488.539 | 4742.926 | | 4967.0360 | 0 | 0 |
| 26 | 4486.069 | 4741.033 | | 4954.7250 | 0 | 0 |
| 27 | 4392.295 | 1377.193 | | 1389.2520 | 1 | 1 |
| 28 | 2959.901 | 736.71 | 588.71 | 0 | 1 | 1 |
| 29 | 3080.695 | 743.296 | | 601.8430 | 1 | 1 |
| 196 | 4487.11 | 4739.272 | | 4995.9490 | 0 | 0 |
| 197 | 4485.043 | 2349.795 | | 1264.6661 | 1 | 1 |
| 198 | 2846.582 | 354.948 | | 390.5131 | 1 | 1 |

In the above Table, the valves V1, V2, and V3 refer to locations near the sensors S1, S2, and S3. The indication of leak is represented by 1, and no leak is represented by 0.

## 2.5 Leakage detection and localization

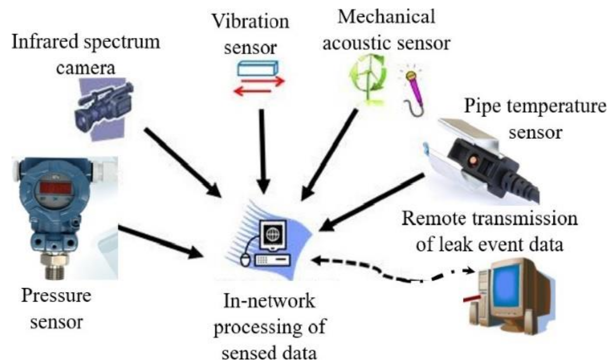



**FIGURE 4**. Sensor array for pipeline health monitoring.

Pressure data gathered from wireless pressure sensors have been used for leakage detection and location system. Because of this, a model using pressure data as inputs and leak occurrence information as outputs has been designed. Also, an output to locate the leak position that the leakage is near any of the sensors should be created. As decreases on pressure data on multi sensors can be seen at the same time, it must be considered for the designed model.

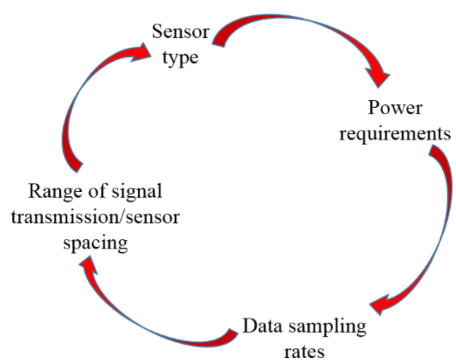



**FIGURE 5**. WSN specification for pipeline health monitoring.

Designed model should be capable of producing outputs referring to multi-leak locations at the same time. When the leak detection and location problem is analyzed carefully, as the inputs should be mapped to more than one class or label at the same time, it can be considered as a multi-label learning problem. It is important to note that this problem cannot be solved with multiclass classification which assumes that each sample is assigned to one and only one label or class at the same time.

The water Flow data gathered from the sensors have been used as inputs of the classification method. The results of classification method give two information about the leakage. The first information is about whether there exist any leakages on the pipeline system. The other one is about the location of the occurring leakage.

## 2.6 Data Analysis for leakage information

Once the number of sensor nodes, its location, and corresponding leakage zone are identified, it is possible to use the network for data generation and train various machine learning models. However, this requires various consideration at different levels which are discussed below:

The next natural progression after sensor placement in the thesis work is to generate training data required for training ML models. How the data is generated would determine the time taken in detecting the occurrence of leakage in the system. Furthermore, from an IoT perspective, it is desired to minimize the energy consumption of the sensors. This calls for data reduction or compression technique that would minimize the data to be sent by each sensor over the cloud to the central hub.

To address this, in the thesis various pre-processing approaches are tested and a suitable way of extracting relevant parameters is used which reduces the data to be transmitted by the sensor. By using the preprocessed data various ML algorithms are trained and tested. In this thesis, Logistic Regression, Support Vector Machines, Artificial Neural Network, and Random Forest are used. Out of these, Logistic Regression being the simplest algorithm, it is used to find leakage in the system as well as to understand the performance of pre-processed data as compared to the raw data. Support Vector Machines and Artificial Neural Network are the most used algorithm in the literature for leakage detection, thus they are included in the discussion. In the end, ensemble of the learned models is made to form a better performing model for leakage detection and localization. This ensemble model is further compared to ensemble learning model called Random Forest algorithm for performance. For leak-size estimation, only Support Vector Regression is used in the thesis as it provided reasonable accurate results for practical purposes.

This thesis is laid out as follows. Following this current chapter, description of solution approach used for data collection and analysis in the thesis. Chapter 4 provides the numerical results obtained using Logistic regression Algorithms.

# CHAPTER 3: METHODOLOGY

The water leakage detection system can be deployed in the already existing plumbing with flow rate sensors attached in the path of the water flow. The sensor does not obstruct the water flow but just collects the data of flow rate. Actuators like solenoid valve is needed to control the water flow in the event of a leak.

The proposed system uses a microcontroller which constantly reads the data from multiple flow rate sensors thereby constantly monitoring the water flow. It compares the flow rate by calculating the difference in data from subsequent sensors and takes the necessary action. If the difference is greater than the set threshold, microcontroller signals the solenoid valve to stop the water flow and sends alert information to the user. This minimizes the water wastage. On the other hand, if the difference is less than the threshold, it sends the sensor data to the cloud for data logging and the process continues as shown in Fig. 1. Online data logging allows the user to keep track of the water usage and take necessary decisions to conserve the water.

## 3.1 Hardware

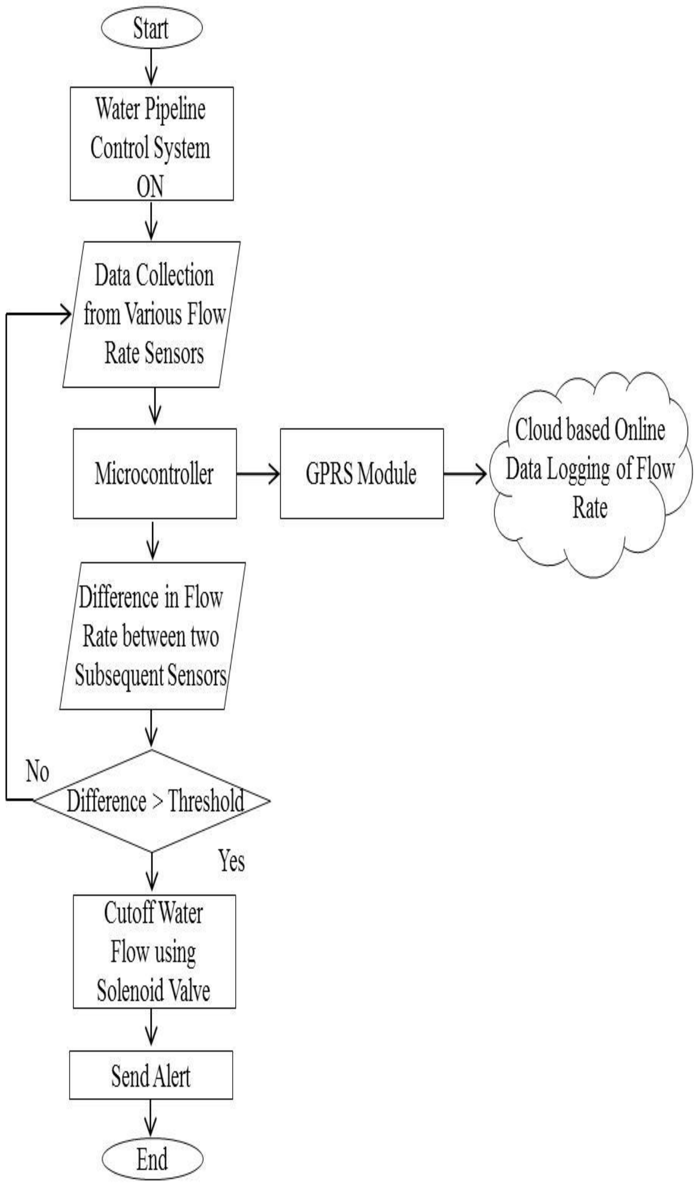
This section deals with the explanation about the hardware components used in the implementation of smart water leakage detection system.

**Flow sensor**

The main input to the microcontroller is the flow sensor (as shown in Fig. 2) which is used to find the flow rate of a liquid. There are various kinds of flow sensors, and the one which is deployed in the proposed system is based on the principle of Hall Effect. Electric potential is generated perpendicular to both electric current flowing along a conducting material and an external magnetic field applied at right angles to the current upon application of the external disturbance. This principle is directly used to measure the flow rate [6]. When the liquid moves through the pipe it rotates a small turbine attached to the hall sensor which produces pulses whenever the blade is rotated and the analog signal produced is sent to the microcontroller for calculations and the same is transmitted to the cloud GPRS module. The flow sensor is capable of detecting up to 30L/Min and a pressure of 2.0MPa

**Solenoid Valve**

The control action is taken by means of solenoid valve during the event of water leak by stopping the water flow.



Water Flow Rate Sensor


Arduino uno

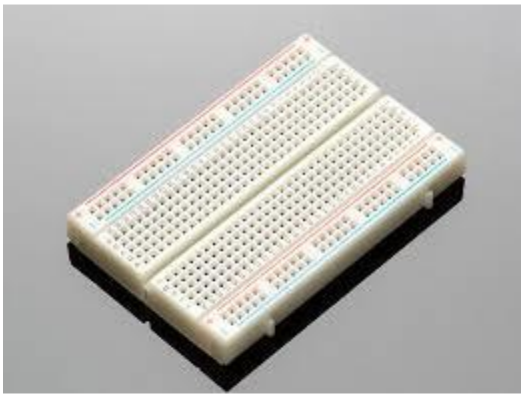
## Arduino Uno

The Arduino Uno is a microcontroller board based on the ATmega328. It has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz crystal oscillator, a USB connection, a power jack, an ICSP header, and a reset button. It contains everything needed to support the microcontroller; simply connect it to a computer with a USB cable or power it with a AC-to-DC adapter or battery to get started. The Uno differs from all preceding boards in that it does not use the FTDI USB-to-serial driver chip. Instead, it features the Atmega8U2 programmed as a USB-to-serial converter.

**JUMPER WIRES**

A jump wire (also known as jumper wire, or jumper) is an electrical wire, or group of them in a cable, with a connector or pin at each end (or sometimes without them – simply "tinned"), which is normally used to interconnect the components of a bread board or other prototype or test circuit, internally or with other equipment or components, without soldering.

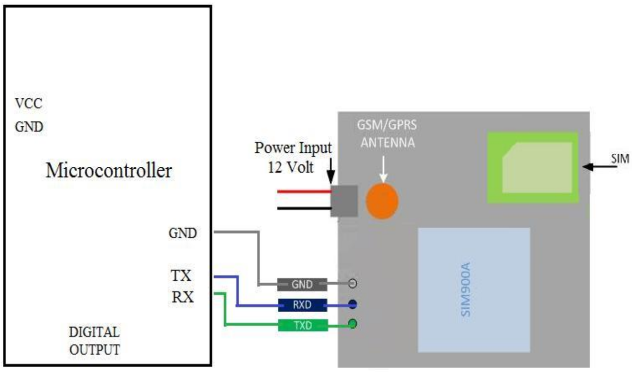
## Bread Board



A breadboard is a solder less device for temporary prototype with electronics and test circuit designs. Most electronic components in electronic circuits can be interconnected by inserting their leads or terminals into the holes and then making connections through wires where appropriate.

## Integrated Water Cut-off System

The automatic water cut-off system is very useful to stop the leakage of water at various points if a leakage is detected. The Monitoring system detects the leakage of water and sends an alert signal. In addition to this the water supply is stopped with the help of solenoid value connected to the water pipelines.

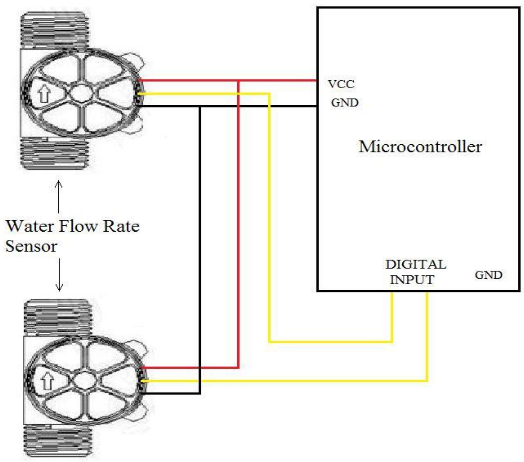



GSM/GPRS module setup

The solenoid valve is connected in series with a transistor such that the transistor works as a switch. Above figure shows the circuitry involved in the solenoid valve set up. The Solenoid valve is normally closed type. Whenever the base signal is applied to the transistor, makes the solenoid value normally open allowing the flow of water. But in case of water leakage, the solenoid value should have normally closed configuration for blockage of water supply. Hence the base signal to the transistor is removed which stops the water flow thus saving wastage of water at initial stage itself.

## Flow rate monitoring system

All the flow rate measurement sensors pertaining to a particular area are connected to microcontroller for as shown below. Couple of microcontrollers is connected to the Network such that we can monitor and control water supply for the whole region. Each Flow rate sensor sends the amount of water passing through it to microcontroller. The microcontroller collects the data from flow sensors and sends the values to the Cloud using GPRS connected to the internet. The flow rate measurements are logged into a sensor cloud which can be utilized for later use. This method is commonly known as data logging.




Microcontrollers connected to sensor

# 3.2 SOFTWARE

### 3.2.1 Pseudo code for leak Detection:

```
Keep the no.of wheel rotation sensors();
 Calculate the water_flow_rate();
   if(water flow rate == leak water flow rate)
          if((water flow rate == leak water flow rate) &(sending of leak
                detection data!=1))
              calculate normal velocity();
              calculate leak velocity();
              calculate water acceleration();
              calculate location of leak();
          else
               leak water flow rate = water flow rate;
               increase time;
   else if(water flow rate > normal flow water rate)
           normal water flow rate = water flow rate;
           time=0;
           sending of leak detection data=0;
   else
           time=0;
           sending of leak detection data=0;
Send data to serer();
```

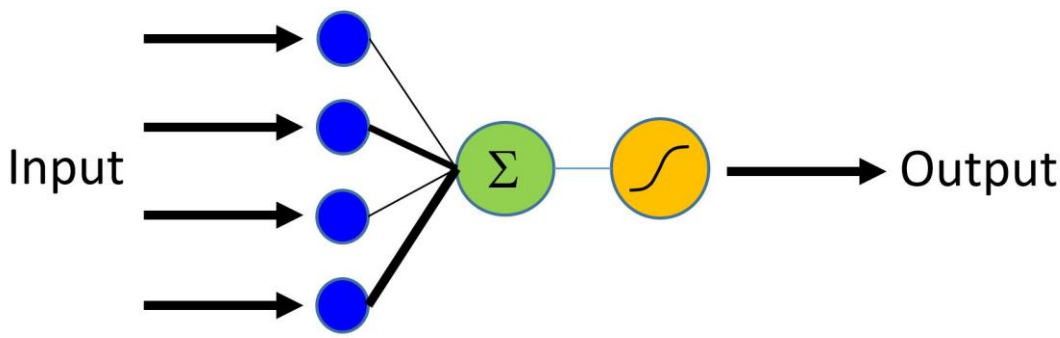
Here,we are using the following Machine Learning Algorithms to predict/detect the leaks:
1.Logistic Regression
2.Support Vector Machine

### 3.2.2 Logistic Regression:

**Logistic regression** is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). Like all regression analyses, the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.
Sometimes logistic regressions are difficult to interpret; the Intellectus Statistics tool easily allows you to conduct the analysis, then in plain English interprets the output.

**Binary logistic regression major assumptions:**

1. The dependent variable should be dichotomous in nature (e.g., presence vs. absent).
2. There should be no outliers in the data, which can be assessed by converting the continuous predictors to standardized scores, and removing values below -3.29 or greater than 3.29.




3. There should be no high correlations (multicollinearity) among the predictors. This can be assessed by a correlation matrix among the predictors. Tabachnick and Fidell (2013) suggest that as long correlation coefficients among independent variables are less than 0.90 the assumption is met.

At the center of the logistic regression analysis is the task estimating the log odds of an event. Mathematically, logistic regression estimates a multiple linear regression function defined as:

logit(p)
for i = 1…n .

**Overfitting.** When selecting the model for the logistic regression analysis, another important consideration is the model fit. Adding independent variables to a logistic regression model will always increase the amount of variance explained in the log odds (typically expressed as $R^2$). However, adding more and more variables to the model can result in overfitting, which reduces the generalizability of the model beyond the data on which the model is fit.

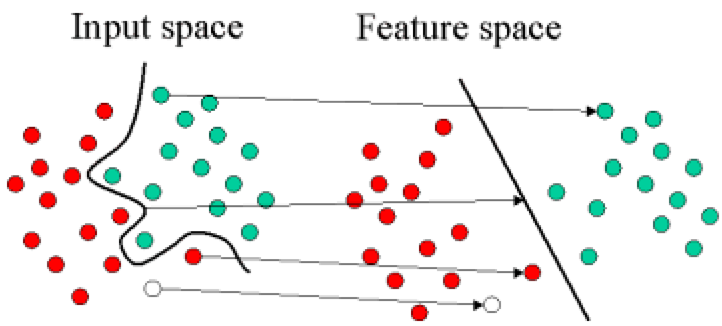
**Reporting the R2**. Numerous pseudo-R2 values have been developed for binary logistic regression. These should be interpreted with extreme caution as they have many computational issues which cause them to be artificially high or low. A better approach is to present any of the goodness of fit tests available; Hosmer-Lemeshow is a commonly used measure of goodness of fit based on the Chi-square test.

**3.2.3 Algorithm :**Logistic Regression Steps

**Input:**Training data Matrix$\boldsymbol{X}$, number of sensor node$n_s$.
**Output:**Theta parameter matrix$\boldsymbol{P}$, mean vector$\boldsymbol{\mu}$and variance vector$\boldsymbol{\sigma}$.
1:Normalize the columns of the training matrix$\boldsymbol{X}$excluding the last column which corresponds to the label in the data. Store the mean value in vector$\boldsymbol{\mu}$and variance in$\boldsymbol{\sigma}$
2:Randomly reshuffle the row of matrix$\boldsymbol{X}$.
3:Form new matrix for training$\boldsymbol{T}$, validation$\boldsymbol{V}$and test data$\boldsymbol{T}$.
4:**for**$i$=1 to$i$=$n_s$***do***
5: Change the labelaccordingly.
6: Reduce cost function using gradient descentoptimization.
7: Tunethe regularization parameter accordingly to obtain best result on the test data$\boldsymbol{T}$.
8: Update the parameter matrix$\boldsymbol{P}$**.**
9: $i$=$i$+1
10:**end for**
11:**return**$\boldsymbol{P}$,$\boldsymbol{\mu}$,$\boldsymbol{\sigma}$

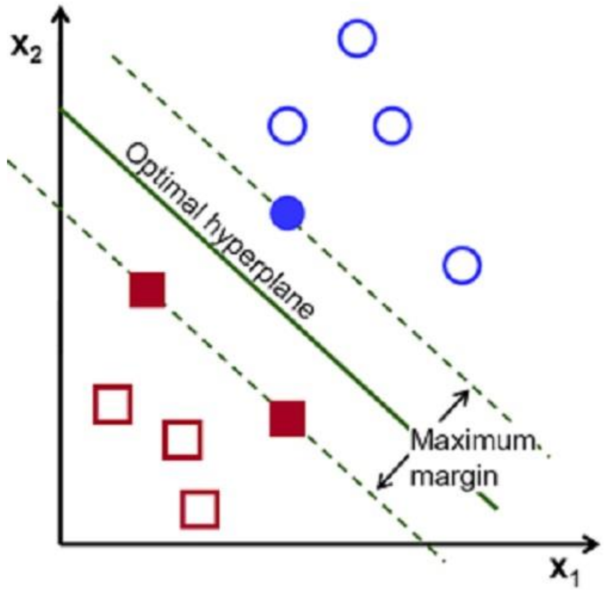
### 3.2.4 Support Vector Machines:

Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. A schematic example is shown in the illustration below. In this example, the objects belong either to class GREEN or RED. The separating line defines a boundary on the right side of which all objects are GREEN and to the left of which all objects are RED. Any new object (white circle) falling to the right is labeled, i.e., classified, as GREEN (or classified as RED should it fall to the left of the separating line).



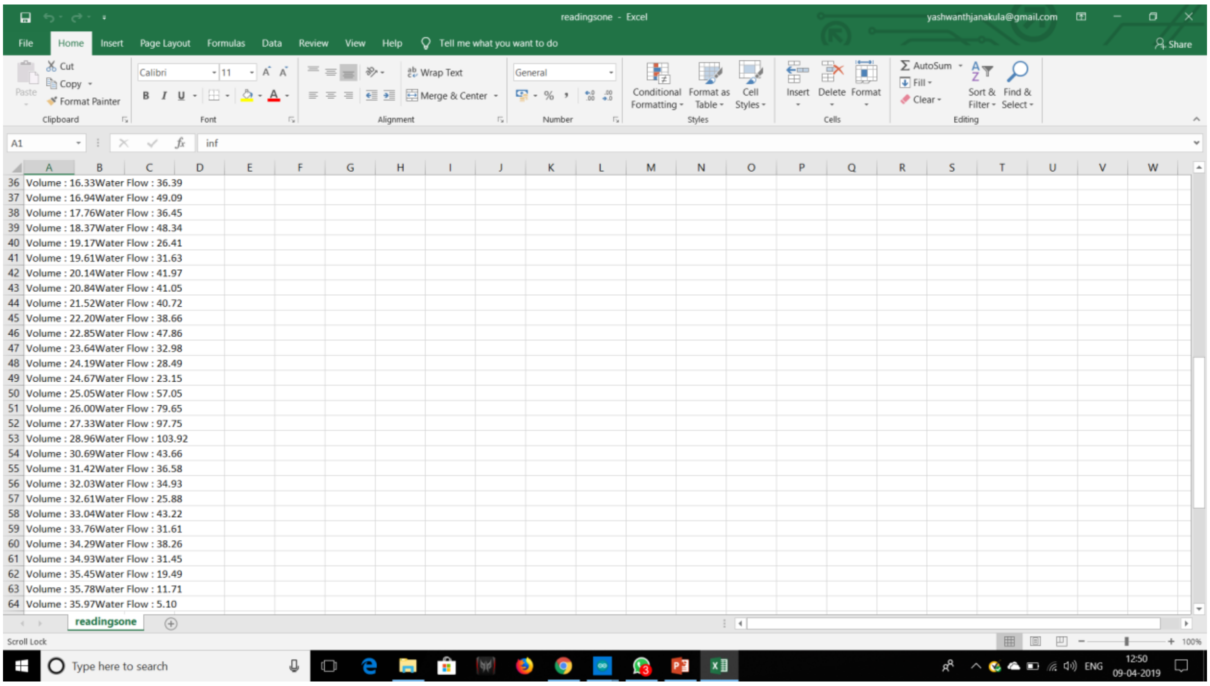

The above is a classic example of a linear classifier, i.e., a classifier that separates a set of objects into their respective groups (GREEN and RED in this case) with a line. Most classification tasks, however, are not that simple, and often more complex structures are needed in order to make an optimal separation, i.e., correctly classify new objects (test cases) on the basis of the examples that are available (train cases). This situation is depicted in the illustration below. Compared to the previous schematic, it is clear that a full separation of the GREEN and RED objects would require a curve (which is more complex than a line). Classification tasks based on drawing separating lines to distinguish between objects of different class memberships are known as hyperplane classifiers. Support Vector Machines are particularly suited to handle such tasks.

The illustration below shows the basic idea behind Support Vector Machines. Here we see the original objects (left side of the schematic) mapped, i.e., rearranged, using a set of mathematical functions, known as kernels. The process of rearranging the objects is known as mapping (transformation). Note that in this new setting, the mapped objects (right side of the schematic) is linearly separable and, thus, instead of constructing the complex curve (left schematic), all we have to do is to find an optimal line that can separate the GREEN and the RED objects.



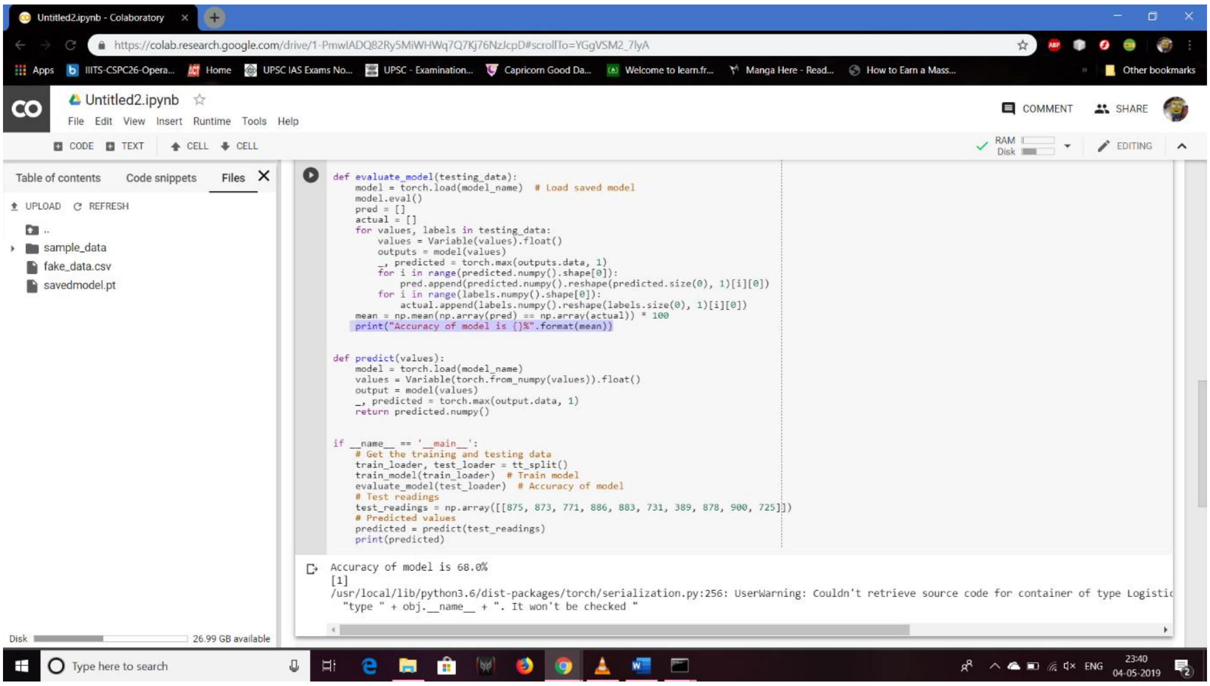

**3.2.5 Algorithm :**SVM Steps
**Input:**Training data Matrix***X***, Number of sensor node$n_s$.
**Output:**Model as struct***M***, mean vector$\boldsymbol{\mu}$and variance vector$\boldsymbol{\sigma}$.
1:Normalize the columns of the training matrix***X***excluding the last column which corresponds to the label in the data. Store the mean value in vector$\boldsymbol{\mu}$and variance in$\boldsymbol{\sigma}$
2:Randomly reshuffle the row of matrix***X***.
3:Form new matrix for training***T***, validation***V***and test data***T***.
4:**for**$i$=1 to$i$=$n_s$***do***
5: Change the label accordingly.Fora positive class the label is marked as 1 and for negative class the label is marked as-1.
6: Choose Gaussian Kernel function as shown in Equation (2.18b).7: Tunethe hyper-parameter C and gamma, using LIBSVMlibrary.8: Update the model***M.***
9: $i$=$i$+1
10:**end for**
11:**return**Model***M***

# 4. RESULTS AND DISCUSSIONS

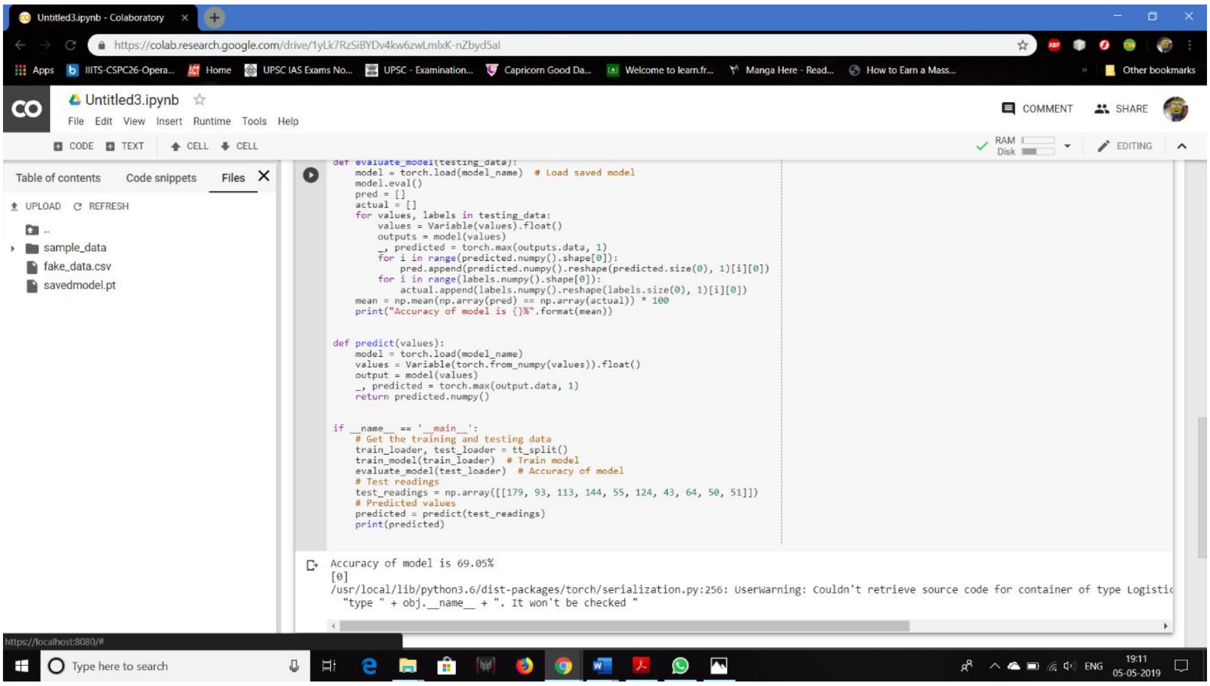



The above figure shows the sensor data which we use for further processing and also for training and testing the model

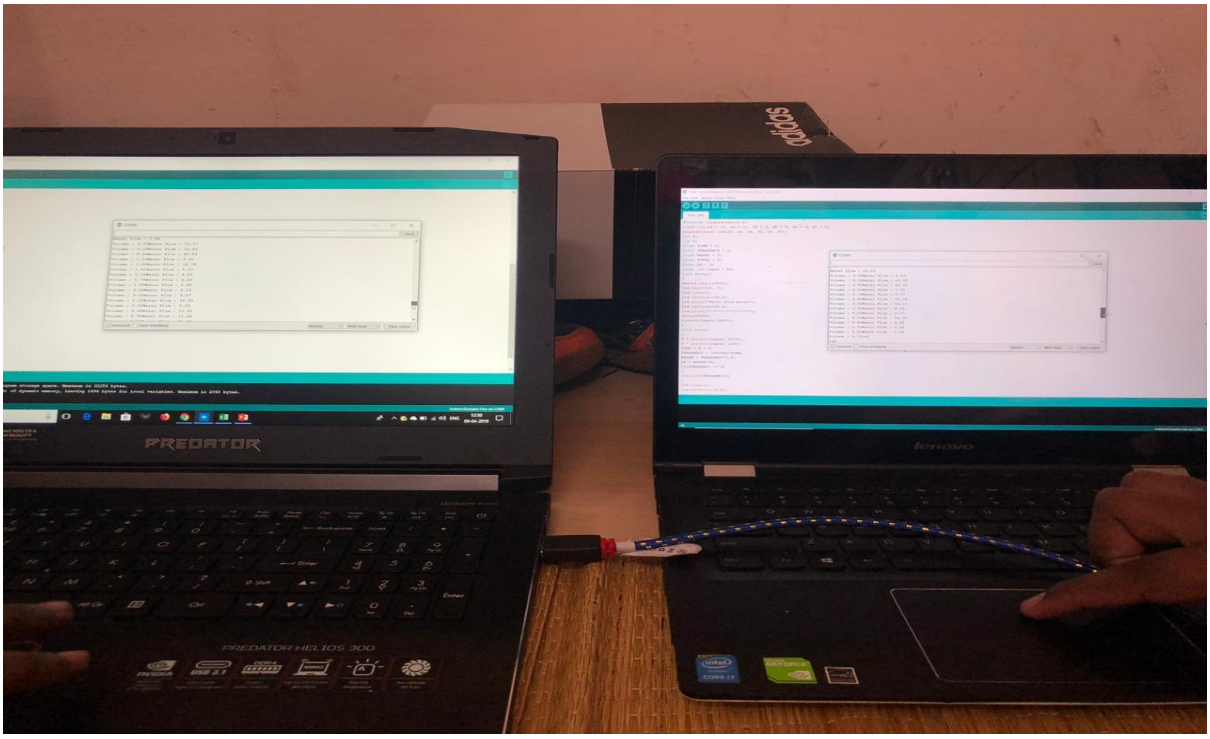
We did a 80-20 division for training and testing the data and then we input unpredicted values to find if there is a leakage in the system. The below are two examples of doing that.




This is the python code execution that gives the final result either 0 or 1 and also the accuracy of our model in percentage. Here it is showing an accuracy of 68% and the result is 1 that means there is a leakage in the network.




The above code shows an accuracy of 69% and the result is 0 that means there is no water leakage in the network according to the values entered.



The above figure shows readings from two different sensors that we use for gathering the information then training and testing the model.

# 5. SUMMARY AND CONCLUSION

Water for domestic purposes is always very essential and it is mandatory to prevent it from getting wasted due to any pipeline leaks. Hence the designed prototype is an effective solution for monitoring the flow of water as well as detecting for leaks in the pipelines. The smart water leakage detection system can help in water distribution process by remote activation of solenoid valves. Usage of cloud logging technique enables the data acquisition and analysis in any point of the pipeline. This makes the system cost efficient and simple.

The system is capable of detecting leaks between any sensor nodes rather than the exact location of the leak. The sensors require lengthy wiring for power supply and data transmission. This reduces the area under observation. The sensors and actuators can be powered by batteries or solar panel. Wireless transceivers can be fitted to acquire the data from the sensor and send command signals to actuator. This sensor network based system may increase the system cost, but it adds the advantage of monitoring a huge area with minimal human power.

# 6. REFERENCES


- Ramleela Khare, Dr. Filliperodrigues & E. mela "Automation of water distribution plant" International journal of Research in engineering and advanced Technology, vol.2 pp 1-6,2014.
- Ayamal Alinhussein & Mohammed Adedalati; "A Supervisory Control and Data Acquisition (ACADA) for water pumping station of GAZA". The Islamic University Journal, Vol-19, pp:303-321,2011.
- Gowri Shankar; "Control of Boiler Operation Using PLC & SCADA", International Journal of Multi Conference of Engineering and computer Scientists, Vol.2 pp:1-6-, 2008.
- Water sector in India: Overview and focus areas for the future, PanIIT Conclave 2010, Available from: https://www.kpmg.de/docs/Water_sector_in_India.pdf
- Guidelines For Improving Water Use Efficiency in Irrigation, Domestic & Industrial Sectors, Central water commission, Ministry of water resources, Government of
- India, November 2014, Available from: http://wrmin.nic.in/writereaddata/Guidelines_for_improvi ng_water_use_efficiency.pdf
- Adam Openshaw, Kalvin Vu, Irrigation Leak Detection-Using Flow Rate Sensors to Detect Breaks in an Irrigation System, Available from http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?ar ticle=1010&context=cpesp
- Teddy Ariyatham, water leak detection, california state university, northridge, Available from: http://scholarworks.calstate.edu/bitstream/handle/10211.3 /132789/Ariyatham-Teddy-thesis-2015.pdf?sequence=1