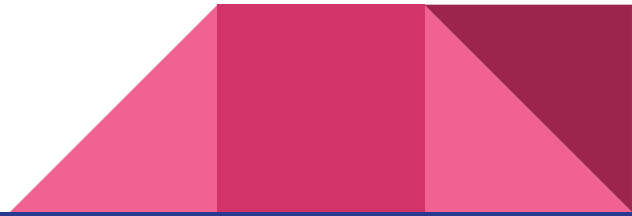
# HackFest 2022

# Team Name- Digital Destroyers

Team Members:

1. Vikky Anand
2. Arunim Basak
3. Anubhav Rajak
4. Aneesha Das
5. Saswati Subhalaxmi

Theme: Water Leakage



# Goal: Real time detection of water loss in transmission channels using sensory devices

**Motivation**: Water is supplied to our home by water distribution network (WDN), which is owned and maintained by water utility companies. Leaks in resource transmission pipelines is a growing concern for the water transmission industry. This creates a need to prevent the threat of leaks and minimize their damages through extensive research in leak detection technology.

Hence, using leak detection techniques, it is possible to monitor leakages using fewer number of sensors in order to help in reduce water loss as well as damage to pipes.

# Dataset

**DESCRIPTION-**

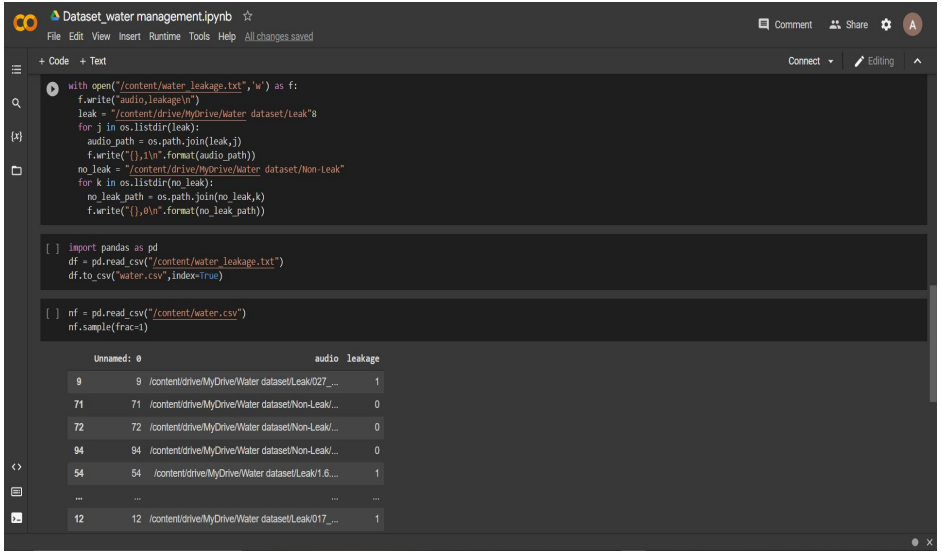
This data acoustic signals recorded from real and buried water distribution networks (WDNs) in **Hong Kong**. The signals were collected from about **ninety leak sites** throughout a period of a year using three different non-destructive technologies, namely **noise loggers, hydrophones, and micro-electro-mechanical-system (MEMS) accelerometers.** The leak signals were collected when leakage is reported, while the no-leak signals were recorded when a previously visited leak site is repaired. Both **leak** and **no-leak** signals were recorded from metal and non-metal pipes.

In PipeNet, a system based on wireless sensor networks, was proposed. It aims to monitor water flow and detect leaks by attaching acoustic and vibration sensors to large bulk-water pipelines and pressure sensors to normal pipelines. This theory is the major basis for origin of the dataset that we are using from Google dataset.

**Link to Dataset:** Acoustic Based Data Acquisition for Leak Detection of Water Distribution Networks

# Solution/Approach

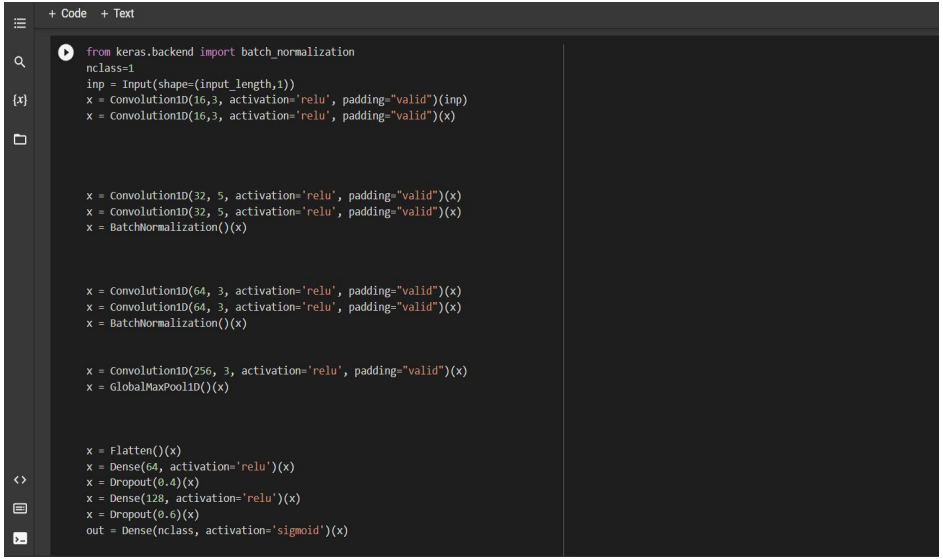
- After downloading the dataset , it was organised such that all the leak and no-leak audio files are separated in two folders. This dataset is then uploaded in the google drive and merged with google colab. A water_leakage.txt file is created and then using os package ,the paths of all the audio files are added in the first column of that text file and then in the 2nd column , those audio files are labeled as 0 and 1 representing non_leak and leak respectively. Then using pandas , the txt file is converted into csv using the dataframe.to_csv() function.



# Solution/Approach

- The dataset is then used for feature extraction, and then divided into training and test set. Then a convolutional neural network model of 11 layers is built using tensorflow library and training set is then trained on it , also showing loss, training_accuracy, validation_loss, validation_accuracy at each epoch. Other models based on Artificial Neural Network (ANN) for classification have been tried but discarded due to a low accuracy.

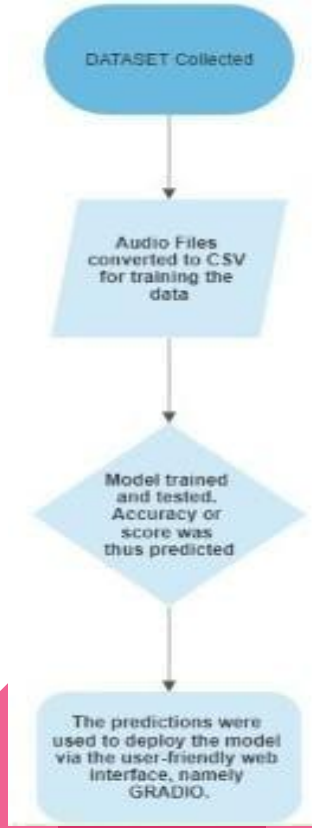
- The model is then deployed by building an easy-to-use online GUI using Gradio. The Interface class has 3 parameters-the leak_detector function, the input components being Time Series data and the output component showing whether there is a leak or not.



# Methodology

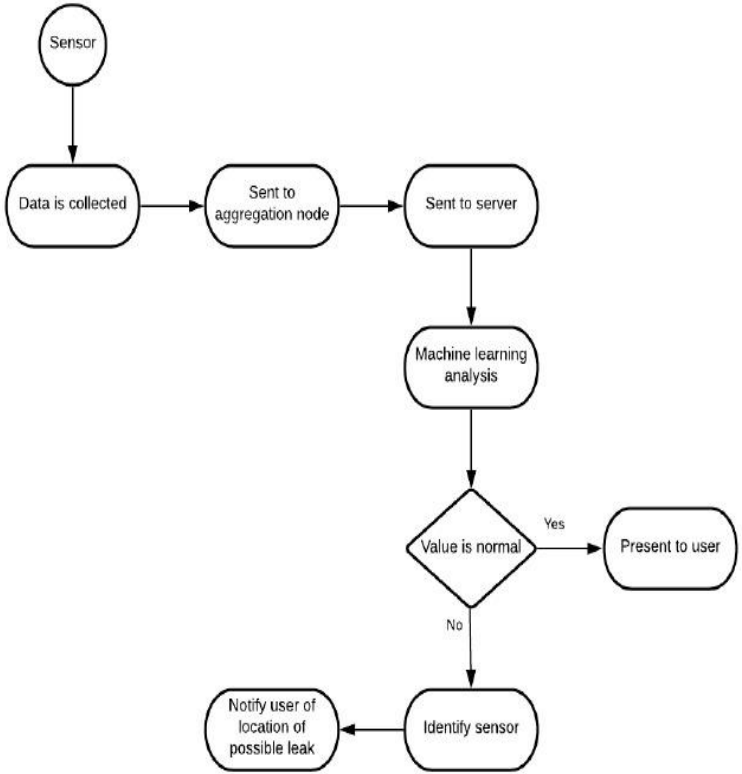
## TechStack

1. Librosa -> used for audio importing and its analysis.
2. Pandas -> used to modify the dataset according to requirements.
3. Numpy -> used for high performance multidimensional array calculations.
4. Matplotlib -> used for plotting graph for analysis.
5. skLearn -> used for importing models and splitting datas.
6. Tensorflow -> used to build convolutional neural network layers.
7. Keras -> Keras is used for distributed training of deep learning models.
8. Gradio -> an open source Python library used for creating an user interface for deployment of the built model

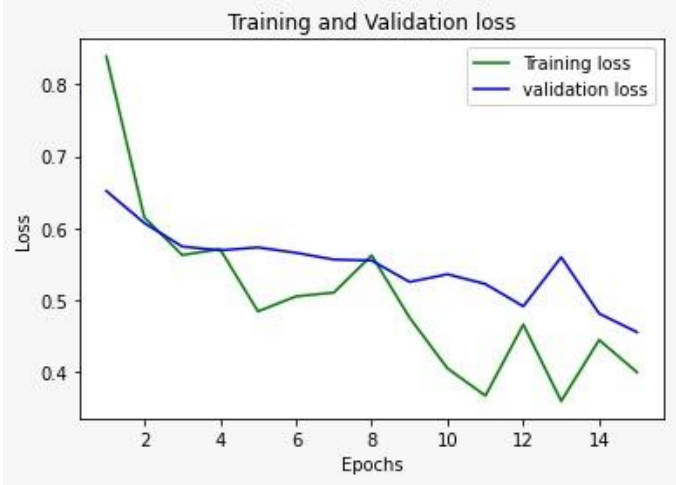


# Methodology

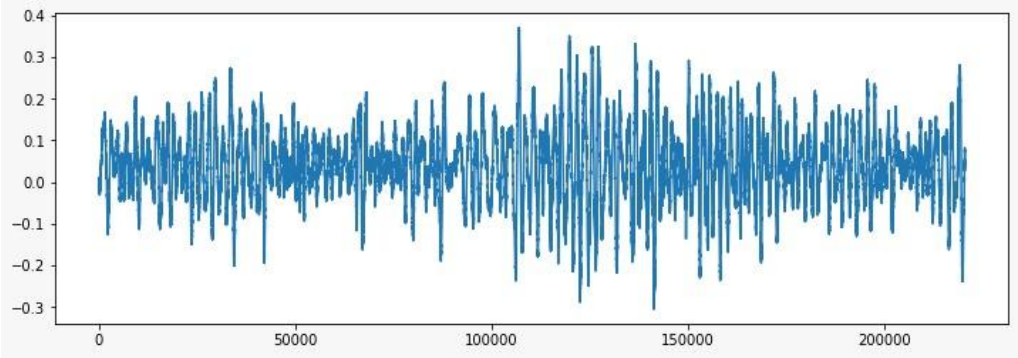
A walk-through to our solution Approach. The data is collected, sent to aggregation mode. Machine Learning analysis is done by building, training and testing the model. Deployment of the model is done at the end.



# Methodology




Training and Validation Loss at each epoch




A plot of the Librosa audio data

[Google Colab]

[GitHub]

# Societal Impact/Novelty

Water transmission pipelines periodically lose an average of 20% to 30% of the water transmitted through them. There are multiple causes for loss of water in transmission pipelines which include leakage, metering errors, public usage such as firefighting, and theft . The most critical route for losses is a leak, as they are considered to contribute an estimated of 70% of water loss in water transmission systems.

The built ML Model using CNN Model training aims to identify the leaks in the water transmission system, which shall save the industries millions of rupees. Water, an inevitable natural resource, is managed and thus saved for other uses.

The model is also deployed using Gradio with a friendly web interface so that it can be easily accessed by anyone, without having to go through the code!

# Future Scope

- The **usage of the data of the MEMS Accelerometers** have been restricted in this project to determine the leakage in the water transmission systems. The only problem that might arise with our chosen data of Acoustics technology is in case of extreme noise or drastic flow conditions, just the sounds and vibrations won't give us accurate results. This is the point where the use of MEMS Accelerators comes, which are used wherever there is a need to measure linear motion, either movement, shock or vibration but without a fixed reference. The difference in acceleration values will thus help to analyse the leak or no-leak condition in such cases where the sound vibrations are vague.
- Use of Convoluted Neural Networks is done to reach at the desired accuracy of the model. **More hyperparameter tuning** can also be exploited in the future to improve the score further.

# Thank You!