JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Early Earthquake Warning Using Machine Learning

Rikesh Bhakta,
17bmiit054@gmail.com
Masters in Information Technology,
Parul Institute of Engineering and Technology-MCA
Parul University

Asst.Prof. Dharmendrasinh Rathod
dharmendrasinh.rathod42072@paruluniversity.ac.in
Faculty of IT and Computer Science
Parul University

***Abstract*—The earthquake early warning system uses a high-speed computer network to transmit information about earthquakes to the population center prior to the arrival of destructive seismic waves. Traditional EEW seismometric methods do not accurately identify large earthquakes due to their sensitivity to the speed of ground movement. Precision GPS stations, on the other hand, are ineffective in identifying average earthquakes due to their tendency to generate noisy data.**

An early warning system is primarily required to set off an alarm so that critical facilities can be evacuated or closed, rather than determining the exact parameters of the earthquake. Therefore, the early warning system must be carried out independently and the government and other authorities must immediately publish accurate information on earthquakes.

The properties required for early warning systems can be summarized as follows:

- **Fully automatic: As the time frame is limited, the facility must be controlled directly without human judgment.**
- **Fast and Reliable: Since there is limited time to respond to the movement of an earthquake, this type of system must be fast and reliable.**
- **Small and Inexpensive - For easy installation, the system must be small and inexpensive.**
- **Independence - In order to issue fail-safe alarms, the system must be independent of other systems.**
- **Easy to connect network: In order to provide the earthquake information, the system must be easy to connect to the network.**
- **Accuracy is better: The accuracy of the information is not such a serious problem for the alarm.**

In this document, I use machine learning methods to address the most pressing challenges facing EEW systems. Several data sources are integrated in real time to cover the entire spectrum of potentially damaging earthquakes (medium and large). Our solution is based on two types of complementary sensors (GPS stations and seismometers).

Introduction

Early Earthquake Warning (EEW) systems are expected to automatically detect and characterize earthquakes as they occur and issue warnings before the ground movement actually reaches sensitive areas so that protective measures can be taken. Evacuating or closing key facilities rather than determining the exact parameters of the earthquake seismometers, which have long been the stronghold of seismology for detecting earthquakes, are struggling to cope with large earthquakes due to a known saturation problem caused by their sensitivity to earthquakes to recognize and characterize the speed of the ground movement. As a result, earthquakes greater than 7.5 tend to be underestimated. The earthquake detection solution must be developed in collaboration with experts in distributed computing and cyber infrastructure to enable real-time alerts. The earthquake early warning system (EEW) was built with the following functions in mind:

1) Rapid detection of earthquakes. Installing seismometers far from the target (e.g., in an urban area) is the easiest way to create enough time to escape. The time is caused by the difference in speed between telecommunications (300,000 km / s) and the seismic wave (8 km / s). Even if the system can detect P waves and determine the parameters of the earthquake or estimate the risk of movement of the earthquake, the time span is longer.

2) Automatic Management. All early warning and alert procedures should be carried out automatically as human assessment can take time and cause errors in assessment.

3) Education and training. It is necessary to educate the public about the importance of the information or alarm of the early warning system. It is also important to train staff on how to behave in the event of an early warning, and manuals for the country to promote to Take countermeasures.

4) In order for the possibility of false positives and information errors to be recognized, the organizations using the alarm system must understand the risk as there is always the possibility of triggering a false alarm. Obviously, they should try to reduce the possibility of false positives.

- **Application Areas**
  Earthquake
  Engineering

Earthquake engineering is the scientific field concerned with protecting society, the natural environment, and the man-made environment from earthquakes by limiting the seismic risk to socio economically acceptable levels.

- **Pattern Recognition and Machine Learning**

A learning procedure then generates a model that attempts to meet two sometimes conflicting objectives: Perform as well as possible on the training data, and generalize as well as possible to new data.

- **Data analysis and seismogram interpretation.**

The knowledge of the velocity structure of the earth and of the various types of seismic sources is the result of interpreting seismograms. Seismograms are a complicated mixture of source radiation effects, such as the spectral content and relative amplitude of the primary- and secondary-wave energy.

- **Seismology (Earth structure)**

Seismologists map the Earth's interior structure by looking at changes in the way seismic waves, produced by earthquakes or explosions, travel through the various layers in the Earth.

- **Engineering Geology**

The realm of the engineering geologist is essentially in the area of earth-structure interactions, or investigation of how the earth or earth processes impact human made structures and human

activities.

- **Geophysics**

The term geophysics sometimes refers to solid earth applications, Earth's shape, its gravitational and magnetic fields, its internal structure and composition, its dynamics and their surface expressions in plate tectonics, the generation of magmas, volcanism and rock formation.

I. Methodologies

The Propagation model (Seismometers and GPS station) to detect early earthquake warning

An earthquake occurs due to the shaking of the surface of the Earth caused by seismic waves. Among these seismic waves, two types stand out: Primary waves (P-waves) and Secondary waves (S-waves).

However, P-waves travel 1.7 times faster than S-waves which propagate through Earth's interior. In addition, only S-waves are responsible for the severe damages. P-waves cause soft shaking due to their longitudinal shape (they move sideways), whereas S-waves are slanting waves (they move up and down). Therefore, an Earthquake Early Warning (EEW) system, which aims to provide an alarm before the damaging effects reach sensitive areas, relies on the detection of the P- wave before the S-wave arrives.

Usually, inertial seismometers are used to detect primary waves. The inertial mass is designed to remain stationary following sudden movements while the frame and drum move with the ground to record waves. However, during large earthquakes, ground motion velocity causes the inertial mass to be displaced above the allowed span. This effect is called saturation. As a result, earthquakes over magnitude 7.5 (Richter scale) tend to be underestimated.

To overcome this problem, GPS satellites can be used because GPS satellites are not affected by earthquakes, so a GPS receiver station on Earth can be used to assess Large earthquakes (above 7.5 on Richter Scale). However, one downside of using GPS is it is unable to characterize medium earthquakes as GPS is sensitive to a variety of noise sources, mostly of atmospheric region. So, we need to combine both these sensors to estimate the P-wave arrival time on each sensor (seismometers and GPS stations) according to its distance to the epicentre with the propagation model.

Multivariate Time Series Classification

A time series is multivariate if a sequence of multivariate measurements is available**.** Multivariate time series (MTS) obtained from GPS stations (3 dimensions: east-west**,** north-south and above**)** and seismometers (3 dimensions: east-west**,** north-south and above**)** are divided into 3 classes according to the potential damage from ground movements**:** normal activity, medium earthquakes, large earthquakes. Therefore**,** earthquake detection can be formulated as an MTS classification problem. MTS classifiers are composed of 3 categories:

A. Similarity based

II. Feature-based

III. Deep learning methods

Similarity-based methods make use of similarity measures (e.g., Euclidean distance) to compare two MTS. Dynamic Time Warping (DTW) has been shown to be the best similarity measure to use along k-Nearest Neighbours (kNN). There are two versions of kNN-DTW for MTS: dependent (DTWD) and independent (DTWI). Neither dominates over the other. DTWI measures the cumulative distances of all dimensions independently measured under DTW. DTWD uses a similar calculation with a single-dimensional time series; it considers the squared Euclidean cumulated distance over the multiple dimensions.

Feature-based methods include shapelets and bag-of- words (BoW) models. Shapelets models use shapelets to transform the original time series into a lower-dimensional that is easier to classify. They relax the major limiting factor of the time to find discriminative shapelets in mul- tiple dimensions (shapelet discovery) by randomly selecting shapelets. On the other hand, WEASEL+MUSE (Schafer and Leser 2017)) convert time series into a bag of discrete words, and use a histogram of word representation to perform the classification.

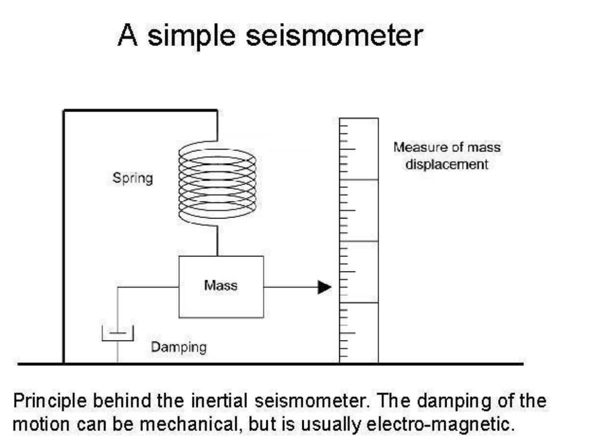
Deep learning methods use Long-Short Term Memory (LSTM) and/or Convolutional Neural Networks (CNN) to extract latent features.

- **Targeting a Distributed Cyberinfrastructure**

A cyber infrastructure is the set of logical and physical computer systems in which a scientific application is deployed. In the context of EEW, a cyber infrastructure must support the processing of large amounts of data generated by geographically distributed seismic sensors such as GPS stations and seismometers. Cyberinfrastructure has two main levels: the sensor level and the central level. The sensor level consists of sensor devices (ie GPS stations and seismometers) with limited computing capacity. The central level consists of well-equipped computer systems that meet high computing requirements. (e.g. cloud data centers).

Machine Learning Solutions

There are a couple of studies (Yoon et al. 2015; Li et al. 2018; Perol, Gharbi, and Denolle 2018) using machine learning methods for earthquake characterization based on P-wave detection (EEW). However, none of them used a combination of GPS and seismometers data so the whole spectrum of earthquakes with damaging potential is not appropriately covered.



Principle behind the inertial seismometer. The damping of the motion can be mechanical, but is usually electro-magnetic.

Fig. 1

IV. Literature Review

Research paper-1: UrEDAS, the Earthquake Warning System:

Author: Yutaka Nakamura, Jun Saita

Urgent Earthquake Detection and Alarm System, is the first real-time P-wave alarm system in practical use in the world. It is able to process digitized waveforms step by step without storing the waveform data. As the amount of processing does not differ whether or not an earthquake occurs, system failure due to overload will not occur. This system detects the earthquake in real-time using a P-wave, it detects the following earthquake in the world.

The 1994 Northridge earthquake

The 1995 Kobe earthquake

The 2003 Miyagiken-Oki earthquake

The 2004 Niigata Ken Chuetsu earthquake

Main UrEDAS functions are estimation of magnitude and location, vulnerability assessment and warning within a fewof initial P wave motion at a s

in issuing the alarm due to absence of Machine Learning, vulnerable to issue a false alarm, and accuracy is also the concern.the existing automatic seismic observation systems, UrEDAS does not have to transmit the observed waveform in real time to a remote processing or centralized system and thus the system can be considerably simplified.

Strength: It can estimate the earthquake parameters and issues an alarm in 3 seconds. This research provides a base to detect earthquakes, the limitation of this is that it slows

Research paper-2: Learn to Detect: Improving the Accu- racy of Earthquake Detection

Publish Year: 2019

Publication: IEEE

Accuracy is one of the most important issues for EEW systems since false alarms may generate unnecessary panic and cause significant economic loss. Unfortunately, sensor readings are usually corrupted by noise. Simple detection schemes could easily misread certain vibrations as earth- quakes. However, false alarms could still occur from time to time since the selection of thresholds highly depends on human experiences.

In this paper, they talk about various models of earth- quakes using different combinations to measure the accuracy of earthquakes. Machine learning is the technique that can make a variety of decisions about the observations based on the extracted knowledge from the historical data. In this paper, the learning-based schemes are exploited to identify the presence of earthquakes. The features of the seismic waves collected from historical events are used to train the classifier for earthquake detection. Three learning-based schemes are built in this paper to perform the verification of earthquake events, namely, the KNN, classification tree, and SVM.

Strength: From the experiments, the detection performance of the learning-based schemes outperforms the traditional criterion-based method. In particular, one can envision that the reliability of earthquake detection can be dramatically increased if the learning-based schemes are adopted. Further studies about the fusion of local predictions and epicentre localization are worth exploring.

Research paper-3: Smartphones Used to Detect an Earth- quake Using a Machine Learning Approach to Identify an Earthquake Event

Publish year: 2014

Publication: IEEE

The possibility of using smartphone accelerometers to detect earthquakes is investigated in this research. Accelerometer has become a common part of a smartphone. Experiments are designed to learn the pattern of an earthquake signal recorded from a smart phone's accelerometer. Many earthquake alarm systems have been proposed. The recent ones are using a smartphone to detect an earthquake event.

The use of machine learning to process smartphone's ac- celerometer output is not new. In this paper they used different algorithms to process and classify the accelerometer output. The approach they proposed in using machine learning to the signal pattern from a smartphone accelerometer is described in this paper. This research shows that methods used in this can distinguish between movement caused by an earthquake and

•

movement caused by other reasons such as walking, etc.

Strength: They have concluded that during the detection of earthquakes they faced three challenges: different types of sensors, insufficient data, and bandwidth limitations. The third challenge, the bandwidth limitation, means that the detection cannot be performed successfully if the data is not completely received by the server. They addressed this problem as the main concern in their research because in disaster areas network failure is very likely to happen.

Research Paper- 4 : A Distributed Multi-Sensor Machine Learning Approach to Earthquake Early Warning

Publish Year: 2018

Publication: AAAI

This research paper is based on concurrent detection of both medium and large earthquakes using a cyberinfrastructure with the help of machine learning algorithms. Their research aims to improve the accuracy of Earthquake Early Warning (EEW) systems by means of machine learning. Conventional EEW methods based on seismometers fail to accurately iden- tify large earthquakes due to their sensitivity to the ground motion velocity. The recently introduced high-precision GPS stations, on the other hand, are ineffective to identify medium earthquakes due to its propensity to produce noisy data. In addition, GPS stations and seismometers may be deployed in large numbers across different locations and may produce a significant volume of data consequently, affecting the response time and the robustness of EEW systems.

In this paper, they introduce the Distributed Multi-Sensor Earthquake Early Warning (DMSEEW) system, a novel ma- chine learning-based approach that combines data from both types of sensors (GPS stations and seismometers) to detect medium and large earthquakes. They show that DMSEEW approach is more accurate than both the conventional approach and used to detects all large earthquakes with a precision of 97%.

Strength: Instead of relying on fully centralized processing of sensor data, they assume that their approach of using the distributed data processing based on geographically distributed cyberinfrastructure, significantly reduces the large amount of data transmitted in the network which also meets the real-time requirement while increasing reliability of EEW Systems.

Algorithms

I. Distributed Machine Learning Approach to Earthquake Early Warning

A Distributed Machine Learning Approach to Earthquake Early Warning (DMSEEW) takes sensor-level class predictions (normal activity, medium earthquake or large earthquake) based on the data gathered by each individual sensor (GPS stations and seismometers). It then aggregates those sensor- level class predictions using a bag-of-words representation in order to calculate a final prediction for the earthquake category.

Step 1 – Predicting the MTS Category at the Sensor-Level:

There are 2 sorts of devices - GPS stations and seismome- ters, and that we train one MTS classifier per sensor type. The classifiers are trained employing a dataset composed of a statistic of three dimensions (east-west, north-south and up-down) and glued time length (60 seconds). we have a tendency to illustrate this opening within the higher part of our approach in Figure 1. so as to predict the earthquake class at the individual sensor level, we use the WEASEL+MUSE (Schafer ¨and Leser 2017) MTS classifier. WEASEL+MUSE fits our approach as a result of

(i) its symbolic representation filters out noise (related to GPS and seismometers sensors) from the dataset.

(ii) it is phase invariant, i.e., features generated do not have to appear at the same time across different MTS.

(iii) It keeps the interplay of dimensions since features generated by WEASEL+MUSE contain the identifier of the dimension, which allows the characterization of co-occurrence of events on different dimensions

Step 2 – Detecting Earthquakes by Combining Sensor- level Predictions:

we have a tendency to collect the category predictions from the various devices (GPS stations and seismometers) and perform a bag-of-words representation.

every sensor foretold class is taken into account to be a word and also the frequency vector of the words from each earthquake is employed to classify its category. This frequency vector is normalized by the number of instances (number of MTS per earthquake, i.e. number of sensors) to get the relative frequency vector.

The last step consists of mixing the bag-of-words of GPS stations and seismometers to characterize the entire spectrum of earthquakes with damaging potential. we have a tendency to illustrate this second step of our approach within the lower a part of Figure 1.

Distributed execution.

The first step of the algorithm is performed on the sensor-level part of the infrastructure. There, an MTS classifier is running on each individual sensor (GPS stations and seis-mometers) in order to generate sensor-level class predictions based on data produced by each sensor.

`

Then, the output of the MTS classifier from each sensor is transferred over the network to the central level part of the cyberinfrastructure. There, the second part of the algorithm is run, i.e., a machine learning method combines all the class predictions from GPS stations and seismometers to form a final class prediction.

•This approach drastically reduces the amount of data over the network since most of the data produced by a sensor is not related to an earthquake event and thus can be filtered out. Moreover, a sensor-level prediction is, in fact, an aggregation of data, hence, it also helps reduce the amount of data sent to central to level data centres.

Tools & Technologies

- **Hardware Necessities**
- **PC with Higher Performance**

In order to detect earthquakes as early as possible we require a stronger base at the hardware side that will give instant alarms to people to make immediate actions.

So, a PC with 250 GB of hard disk capacity, minimum 8 GB of RAM, and with processor of intel i7 and above is required for better performance to give instant alarms whenever it detects casualties.

- **Seismometers**

Seismometers are the most important sensors which will help us to detect the P-wave, which in the end is the base to detect earthquakes.

- **GPS Station**

The limitation of seismometers is that they cannot detect large earthquakes because of saturation. So, In order to detect a large earthquake the GPS is required.

- **Battery enclosure**

Battery Enclosure is used to keep the seismometers charged up to date.

- **Alarms/Sirens**

Whenever the system detects an earthquake, the system should warn the people by turning on the alarms or sirens so they can take the protective measures.

- **Solar Panel**

We can't rely on one source of energy when it comes to detection of earthquakes so we need constant energy so use of solar energy is always a good idea.

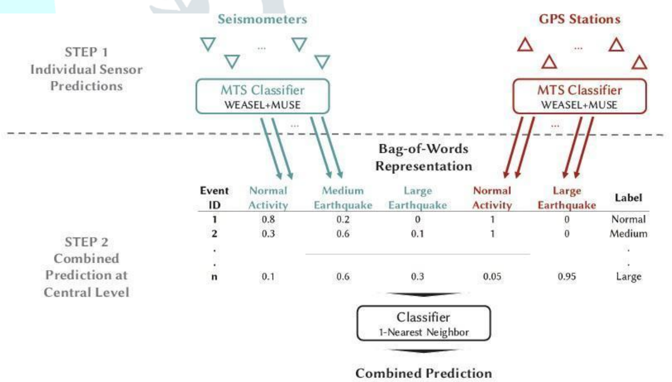


Figure 1: Distributed Multi-Sensor Earthquake Early Warning Algorithm (DMSEEW).

- **Hardware Necessities**

TABLE I

Number	Description
1	PC with 250 GB or more Hard disk.
2	PC with 8 GB or more RAM.
3	PC with intel i7 or Above.
4	Number of seismometers.
5	Alarms / Sirens.
6	GPS Station.
7	Battery enclosure.
8	Solar panel

- **Technology Used**
- **Machine Learning Algorithms**

We use a distributed machine learning approach for earth- quake early warning (DMSEEW) to detect moderate to severe earthquakes. The reason to use a distributed approach is because seismometers cannot detect large earthquakes (above 7.5 Richter scale). So, whenever, Seismometer's mass exceeds the scale, the algorithm automatically triggers the GPS station. A DMSEEW takes sensor-level class predictions (normal activity, medium earthquake or large earthquake) based on the data gathered by each individual sensor (GPS stations and seismometers. Then add these sensor-level predictions using multi-word representation to compute the final prediction for the earthquake type.

By using large datasets of data, the algorithm itself becomes accurate and can detect both medium and large earthquakes at almost 97% accuracy.

I. Conclusion

The use of machine learning methods in seismology is still in its infancy. One area of development where it demonstrated promising results is earthquake early warning (EEW), i.e. the characterization of an earthquake before it reaches sensitive areas. Current state-of-the-art methods based on seismometers data only demonstrated an applicability limited to medium earthquakes. In contrast, GPS based methods are only suitable for large earthquake detection.

We propose DMSEEW, a novel stacking ensemble approach for characterizing the whole spectrum of earthquakes with damaging potential by combining both GPS and seismometer data. Our evaluation on a real-world dataset collected with domain experts demonstrates that the proposed distributed stacking ensemble approach improves the detection of both Medium and large earthquakes compared to traditional seis- mometer only approach and the combined sensors (GPS and seismometers) baseline approach that uses the rule of relative strength (F1 value: + 7% and + 6% for medium-sized earth- quakes, + 45% and + 27% for large earthquakes). In addition, DMSEEW detects all major earthquakes with 100% accuracy. While existing solutions rely on fully centralized processing of the sensor data, this approach assumes distributed data processing based on a geographically distributed cyberinfras- tructure. This design significantly reduces the volume of data transmitted in the network, meets the real-time requirements while increasing reliability of the EEW system.

Current / Latest R&D works in the field

Research on earthquake prediction has never achieved re- sults that are so convincing as to assert itself over other methodologies for an approach to the question of defence against earthquakes. In the coming years, the field of earth- quake system of structures is most likely to witness the following significant developments:

- Performance-based design processes will take centre stage, making conventional descriptive codes obsolete.
- The acceptable risk criterion for design purposes will be prescribed in terms of performance objectives and hazard levels.
- The development of new structural systems and devices will continue for base-isolation, passive energy dissipation and active control systems, along with the proliferation of non- traditional civil engineering materials and techniques.
- Analytical tools for reliable prediction of structural re- sponse (essential tools in performance-based design processes) will continue to improve and be updated frequently to include new devices and materials.
- **Innovating with AI**

The team named their AI system ConvNetQuake, and it's the first neural network designed to detect and locate earth- quakes. The specialized algorithm can look at ground motion measurements known as seismograms and determine whether or not the seismic activity is just "noise" or an earthquake. However, while it is superior to other earthquake detection methods, ConvNetQuake can only detect earthquakes — it can't predict them.

Also, the innovation in deep learning on data can be a breakthrough in the near future in prediction of an earthquake more efficiently and accurately. The development of Machine Learning algorithms can be enhanced with a deeper study related to seismic data and datasets. The following can be a breakthrough in the future in terms of technology related to earthquakes.

- Data-driven seismic prediction systems.

Deep learning-enhanced seismology in the internet of things platform.

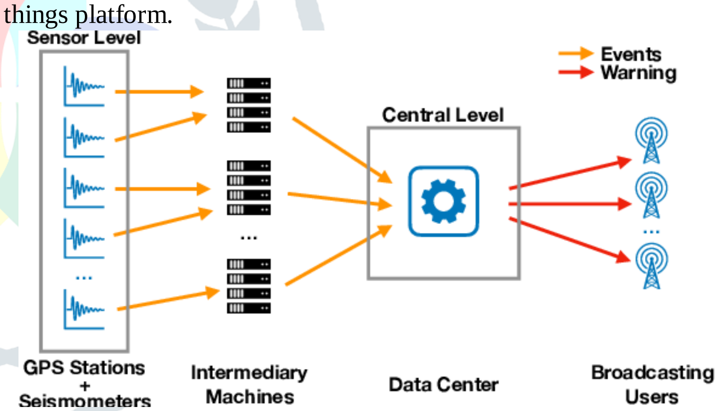


Fig. 1. **2: High-level architecture and data workflow for the Earthquake Early Warning System**

II. Bibliography

[1.] General description about Seismic data detection (March 2021) - https://grillo.io/data/

[2.] Machine Learning Model (April 2021) - https://openeew.com/docs/machine-learning

[3.] Learnt about implementation of earthquake network (April 2021) - https://sismo.app/

[4.] Working Model (June 2021) - https://grillo.io/impact/#openeew

[5.] Case Study on ShakeAlert - An Earthquake Early Warning System for the West Coast of the United States (June 2021) – https://www.shakealert.org/implementation/shakealert-phase-1

III. References

[1.] Allen, R. M., and Melgar, D. 2019. Earthquake Early Warning: Advances, Scientific Challenges, and Societal Needs. Annual Review of Earth and Planetary Sciences 47:361– 388.

[2.] Application to Earthquake Early Warning. Geophysical Research Letters 45:4773–4779.

[3.] Baydogan, M. G., and Runger, G. (2014.) 'Learning a Symbolic Representation for Multivariate Time Series Classification'. Data Mining and Knowledge Discovery 29:400–422.

[4.] Burkett, Erin R.; Given, Douglas D.; Jones, Lucile M. (February 2017) *ShakeAlert: an earthquake early warning system for the United States West Coast*. Fact Sheet 2014–3083. U.S. Geological Survey. p. 4. doi:10.3133/fs20143083. ISSN 2327-6932.

[5.] Karlsson, I.; Papapetrou, P and Bostrom, H. (2016.) 'Generalized Random Shapelet Forests.' Data Mining and Knowledge Discovery 30:1053–1085

[6.] Kevin Fauvel.; Diego Melgar, Manish Parashar. (2018). "A Distributed Multi-Sensor Machine Learning Approach to Earthquake Early Warning"

[7.] Schafer, P., and Leser, U. (2017.) 'Multivariate Time Series" Classification with WEASEL+MUSE.

[8.] Yoon, C.; O'Reilly, O.; Bergen, K. J.; and Beroza, G. C. (2015.) 'Earthquake Detection Through Computationally Efficient Similarity.' Science Advances.

Rikesh Bhakta Studying Masters Of Science in Information & Technology, Parul Institute of Engineering & technology, Vadodara