A Project Report on

“ F- ALERT: Fire Disaster Management Using Machine Learning ”

Submitted in partial fulfillment of the requirement for
Degree in Bachelor of Engineering (Information Technology)

By

Group Members:
Purva Gharat(501519)
Elvis D'Costa(501508)
Jincy Jacob(501524)
Kunal Das(501511)

Guided by:
Prof. Archana Shirke

Department of Information Technology
Fr. Conceicao Rodrigues Institute of Technology
Sector 9A, Vashi, Navi Mumbai – 400703
University of Mumbai

2018-2019

CERTIFICATE

This is to certify that the project entitled

F-Alert : Fire Disaster Management using Machine Learning

Submitted By
Purva Gharat
Elvis D'Costa
Jincy Jacob
Kunal Das

In partial fulfilment of degree of **B.E.** in **Information Technology** for term work of the project is approved.

External Examiner **Internal Examiner**

Internal Guide

Head of the Department **Principal**

Date: - **College Seal**

DECLARATION

I declare that this written submission represents our ideas in our own words and where others ideas or words have been included; I have adequately cited and referenced the original sources. I also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Purva Gharat

Elvis D'Costa

Jincy Jacob

Kunal Das

ABSTRACT

It is estimated that more than USD 7 trillion worth of economic damage and 8 million deaths via natural disasters have occurred since the start of the 20th century and have caused a great havoc in the society, killing thousands of people and destroying lives and properties. Farming, irrigation, or domestic uses of water are normally highly impacted during droughts whereas natural disasters like floods cause loss of human life, damage to property, destruction of crops, loss of livestock, and deterioration of health conditions owing to waterborne diseases. Due to high rains in late July 2018 and heavy monsoon rainfall, the Indian state of Kerala was affected with severe flooding, resulting in 445 deaths. Also the year 2018 continues to prove catastrophic for Mumbai when it comes to combating fire incidents.

Even as development in the state continues to overlook disaster vulnerability, the lack of a proper system to inspect constructions and to ensure their compliance with disaster mitigation guidelines, building codes and other development control regimes appear glaring. Introducing scientific and technological advances into disaster management is potentially of great value. Improved early warning systems to move people out of harm's way, sound land use to avoid hazards, and construction practices to better withstand loads imposed by hazards are examples of more effective mitigation strategies based on new knowledge. Despite the advances in technology, no one can tell with complete accuracy when a place will catch fire, or how powerful a hurricane will be on landfall. However, observation and data are powerful tools, so prediction is getting better and faster. This data analysis is achieved by using various technologies such as Data Mining and Big Data, using which we can check for certain vulnerabilities from the data collected and analyze the possible disaster that can occur due to these vulnerabilities and also list out the different steps to mitigate the disaster.

The first benefit of being prepared for calamities and disasters is the fact that, in a society it can save one’s life or health even after the disaster has stricken and thus can save millions of lives who otherwise would be in grave danger. Thus with preparedness, response and recovery techniques, disaster can be managed to a certain extent.

INDEX

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1. INTRODUCTION

Disaster management essentially deals with management of resources and information towards a disastrous event and is measured by how efficiently, effectively and seamlessly one coordinates these resources. The ability to effectively deal with disasters has become a challenge to modern technology.

Figure 1.1: Steps involved in any disaster management system [15]

1.1 Background

Disaster management can be defined as "The organization and management of resources and responsibilities for dealing with all humanitarian aspects of emergencies, in particular preparedness, response and recovery in order to lessen the impact of disasters"[1].

There have been countless disasters both natural and man-made, where millions of lives have been lost before and since Indian Independence. In 2001 as a whole, India suffered 73.1% of all disaster-related deaths in Asia (IFRC 2002). According to the statistics, 68% of India's land is prone to drought, 60% to earthquakes, 12% to floods and 8% to cyclones, making India one of the most disaster prone countries in the world, affecting overall 85% of Indian land and more than 50 million people as shown in Fig 1.2. Since 2005, though there have been achievements, there have also been major challenges [2].

As shown in Fig 1.1, disaster management includes three basic phases: preparedness, response, and recovery. The efficient data collection, archiving and analytics is essential for effective disaster management. Big data can play a major role in all phases of disaster management. Big data can be defined as the collection of data sets which are large and complex and beyond the capacity of the conventional processing system[3] . The preparedness phase is challenging for multi-hazard because it is hard to model all consequence of a disaster due to the dependencies on many variables and uncertainties. This kind of dependency modeling requires big data with high accuracy (Pu and Kitsuregawa, 2013) for better modeling. It is possible to project the probable scenario of a catastrophe in big data era. Many system of systems platform are present for forecasting natural hazards including Dartmouth Flood Observatory (Revilla-Romero et al., 2015), Near Real-time Thermal Monitoring Global Hot-spots (Harris et al., 2000) for forest fire and volcanic monitoring, National Hurricane Centre Active Tropical Cyclone, Global Agricultural Drought Monitoring. These monitoring and forecasting systems use tons of remote sensing data and data from other sources. These systems need a high-performance computational system to process huge data daily. The NASA center for climate simulation, for instance, is providing big data management and analytical service to support NASA modeling groups, funded investigator and extended research community (Schnase et al., 2011) for climate modeling and simulation[3].

Figure 1.2: Major disasters occurred in India between 1980 and 2009

1.2 Motivation

In the past couple of years, many people have been killed by earthquakes, fire hazards, tsunamis, volcanic eruptions, landslides, floods, tropical storms, drought, locust invasions, and other natural and man-made calamities. These disasters have also inflicted injury, disease, homelessness, and misery on many people, and caused colossal material damages. There is nothing natural about the disasters, what can be considered natural is the ‘event’ not the ‘losses’. There were many techniques implemented to curb the occurrence of disasters, which could be done only to a certain extent. Crowdsourcing data has also proved its usefulness in emergency response to catastrophes such as the earthquake in Haiti and Nepal [3].

The year 2018 continues to prove catastrophic for Mumbai when it comes to combating fire incidents. In 6 months, Mumbai has seen 12 major fires, 22 deaths, which are shown in the Table I.

Table I: Major fires happened in Mumbai in the year 2018

Date	Place of Disaster
June 13, 2018	A major fire broke out at Beaumonde Towers in Mumbai's Worli

May 30, 2018	Five dead in Goregaon fire
April 6, 2018	Fire in a commercial-cum-residential complex in Mumbai's Pydhonie area, eight people rescued
March 17, 2018	Army building in Colaba catches fire
March 3, 2018	Chemical factory gutted in Asalfa village in Ghatkopar, a central Mumbai suburb, no casualties
February 8, 2018	Two injured in fire at Malad
February 7, 2018	Fire in cloth mill in Goregaon, no casualties
January 22, 2018	Two injured as fire erupts in Industrial Area
January 8, 2018	Fire at Sessions Court
January 6, 2018	One dead in fire at Cine Vista film studio, Kanjurmarg
January 4, 2018	Four dead, 5 injured in fire Maimoon Building, Marol
December 18, 2017	12 killed in a fire in SakiNaka

Big data can play a major role in all phases of disaster management. The unstructured and semi structured data were previously been considered as “dead” data, however, with the advancement of big data technologies, these data can be analyzed to bring valuable information to effective emergency management. As big data is characterized by 5Vs (Volume, Velocity, Variety, Veracity, and Values), which make big data more challenging to deal with. It is important to understand how these characters are related to disaster management.

Big data can contribute to early detection of many aspects of flood such as probable inundation, duration, river discharge of the flood through the integration of various data format from different sources. Both the remote sensing data and data from social media can play a major role to flood prediction and impact analysis. Earthquake is an onset natural disaster, and until now, there is no proper early warning system for an earthquake. Big Data analysis, however, can contribute to earthquake forecasting using satellite and atmospheric data combined with statistical analysis. The meteorological disasters such as tropical storm and tsunami are not as quick as an earthquake, and are possible to make alert with lead-time. Big data has proven its usefulness for forecasting and warning system for this meteorological and climate hazard [3].

All these possibilities and the advantages gained from them act as a motivation to create many other effective disaster management techniques especially for fire related hazards, thus saving the lives of mankind and livestock.

1.3 Overview

Disaster management, at the individual and organizational level, deals with issues of planning, coordinating, communication and risk management. Planning for the unforeseen can be the key to disaster preparedness since accurate and timely information can reduce the loss from the catastrophe [3].

This Disaster management project will involve collecting household level data ie. Big data of Vashi (Navi Mumbai, India). After collecting data from multiple sources and integrating them together, the possible vulnerabilities of the area for any occurrence of a disaster are detected. Then mitigation plans are provided for the same. Here the huge amount of data is analyzed for their different characteristics, which may give an idea about the possibility of a disaster, primarily fire disasters.

Big Data Analysis comprises of taking existent high volume data sets and predicting results for the future. Data analysis and management of this variety of unstructured and semi-structured data add more challenge in disaster management.

1.4 Problem definition

It is needless to say that natural disasters like floods, famine, and drought, etc. result in loss of property, livelihood and most importantly loss of lives. Hence a disaster management system that can help reduce the damage done by such hazards is the necessity of the hour. Being prepared for such disasters and having the right solutions once these disasters occur are key in disaster management systems. This project aims to predict any such occurrence of a hazard beforehand by analyzing the collected data of a particular area and suggest the most effective mitigation plan for the same.

1.5 Aim and Objectives

1.5.1 Aim

The aim of this project is to develop disaster based system that collects, integrate, visualizes the relevant data and analyze vulnerabilities and suggest best mitigation plans.

1.5.2 Objectives

1. To collect relevant data from various sources and integrate all these data.
2. To clean the collected data and analyze which data is necessary to predict a particular disaster.
3. To decide which disaster is to be taken into consideration based on the cleaned data.
4. To identify the vulnerabilities for a particular disaster.
5. To suggest the best mitigation plan for a disaster in any specific area.

1.6 Scope

- The purpose of this disaster management project is to build an effective technique for detecting the possible vulnerabilities of disasters that can take place in a particular area. The disaster that we will be mainly focusing on is fire- related disaster.
- Machine learning will provide solutions that address a wider spectrum of problems, such as situational awareness and real-time threat assessment using diverse streams of data.
- Reliability :

  The reliability of the project will depend upon the data that will be gathered, more accurate the data; more efficient the vulnerability assessment and mitigation plans will be achieved.
- Scalability :

  This project deals with fire-related hazards, but with further advancements, other disasters can also be detected. Initially, the project is used for a particular area, which can be scaled to larger areas as well.
- Predictability:

  To provide larger-scale behavior and movement data, run through predictive machine learning models that will help officials to take more appropriate action. Predictive models will offer a promising new approach to disaster relief.

- Google BigTable, Simple DB, NoSQL are some of the tools and techniques that will be used for data management of this disaster management system. Since disaster management involves Big Data, which differs from the traditional data and lacks the structure of data, it cannot be stored in a single machine. For Big Data, tools that will be used are Hadoop or Big Table as they effectively process large amounts of data efficiently in a timely manner. Hadoop emphasizes discovery from the perspective of scalability and analysis to realize nearly-impossible achievements.

1.7 Issues/ Limitations

In the case of floods, rainfall varies with latitude, elevation, topography, seasons, distance from the sea, and coastal sea-surface temperature. These constraints are not considered and analysis was not drawn for the complete monsoon season.

Food crises are slow-onset disasters. They emerge over a period of months and are routinely tracked and anticipated by famine early warning systems – specialist units that monitor and forecast risk factors such as food prices, health indicators, rainfall and crop production. These systems provide governments and humanitarian actors with the chance to take early action and prevent the situation from escalating into an emergency. Cost-benefit analyses indicate that, compared with emergency response, early action offers significant cost savings in the long run. Yet all too often the link between early warning and early action fails and the opportunity to mitigate a gathering crisis is lost. Famine early warning systems have a good track record of predicting food crises but a poor track record of triggering early action.

Policy makers, disaster response organizations, are not very familiar with computer generated data. Thus, these data have been ignored in the decision-making process.

Accurate forecasting and early warning system is an important aspect of disaster preparedness if these forecasting systems fail to do a good job, there is a chance to lose the public trust, for instance false information may lead to resource allocation in wrong place or may result in the improper rescue operation.

However, a large volume of data can help to increase the accuracy and reliability of these model simulation but high volume data not necessarily mean the accurate data. Therefore, the noise removal from these big data is crucial to get results right.

Data privacy could be another challenge to deal with. Many social media data are considered as private. Apart from this privacy concern, access to some parts of social media data related to an emergency might be helpful.

Another challenge could be the resilience of big data system itself. The dependency on these big data will increase with the frequent use of big data in disaster management. It is important to make system architecture resilient to disaster, so that, data can be recovered and retrieved without experiencing any problems. Pu and Kitsuregawa, (2013) mentioned some ways to achieve resilience such as having a redundant backup system, cloud based distributed storage system and distributed computing systems. Data storage with multiple formats might also help to reduce the loss of data. [3]

Chapter 2

Literature Survey

2. LITERATURE SURVEY

Disaster management in urban areas is particularly complex owing to urbanization and its associated impacts which often increase the exposure of people and economic assets to hazards and create new patterns of risks. Fire disasters in particular are a common occurrence in cities across the world.

2.1 Related Work

2.1.1 Forest Fire Prediction and Alert System Using Big Data Technology [10]

Aim:

Forest fire is considered as one of the major natural disaster. Our method is to collect and analyze the data from wireless sensor using Hadoop tool to predict the forest fire before it occurs. Here we use machine learning tool called as Mahout which is used for clustering and filtering the datasets and it can be able to predict the valid output. By using GSM we can give alert message to people so that they can relocate to a safety place immediately, when fire occurs. Signal and Infrared image processing is used to monitor the signals and images of the entire forest for every 30 min and those data will be stored in datasets, by using those data we can be able to predict the forest fire in advance.

Method:

In forest fire prediction, the data will be collected and stored in Hadoop as unstructured form. Hadoop ecosystem components are pig, hive, MapReduce, HDFS. In forest fire prediction Map Reduce it is used to reduce the large datasets into simpler datasets. Hive is the data warehouse infrastructure built on top of Hadoop for providing data, query, and analysis. Apache hive supports analysis of large datasets stored in Hadoop's HDFS. I provides an SQL like language called HiveQL and it is use to convert data into map reduce. To accelerate queries in Hadoop it provides indexes and bitmap indexes. We use Hadoop tool to store the data of forest fire prediction for analyzing. Hive is used to analyze the datasets of forest fire that are stored in hadoop HDFS. We can use machine learning tool to collect the temperature, rare trees, weather, gas, etc. it is used to store the scalability of large datasets. Mahout tool is one of the machine learning tools used in hadoop ecosystem. Its OS is independent. Mahout tool is used to filter and classify the datasets based on keyword. Cameras are used for monitoring the entire forest. Signal processing and infrared image processing is also used to monitor the signal and image of the entire forest. By using the signal processing and multi sensor we can get an alert message from the server and can be able to predict the fire in advance.

Algorithm:

1. All the nodes should be initialized and synchronized to same clock
2. A cluster of nodes will be connected to a base station, and all the base station are connected to the control center.
3. When humidity of air is high, LM35 senses the temperature and transmit it to the base station every 30 minutes
4. When the humidity of the air reduces there is more possibility for the fire hence the rate of measurement will be increased to every 15 minutes
5. If the temperature is less than the threshold value then the node enters the sleep state else the sensor continuously senses the temperature and transmits the result to the base station
6. When a node senses fire it sends a danger packet to its neighboring nodes and the timer is started and it will run till it gets a fire alert. This is to calculate the rate of spread of fire and the direction of spread
7. The base station collects all the values and calculates the rate and direction of spread of fire
8. Through the GSM, alert messages is sent to nearby villages to relocate the people to a safe locality This is a simple method where we have a less overhead in the data packets, and this topology is easy to expand. The energy consumption is also less as the node senses the parameters only on certain intervals which are controlled by the base station.

Advantage:

To reduce the damage and destruction that is caused by the forest fire to the life and property of humans and also wild animals. Apart from early detection of forest fire they have also attempted to predict the fire in advance with the help of the data obtained from the sensors that are deployed in the forest.

2.1.2 A Mobile Disaster Management System using the Android Technology [11]

Aim:

The disaster management system Android application known as MyDisasterDroid determines the optimum route along different geographical locations that the volunteers and rescuers need to take in order to serve the most number of people and provide maximum coverage of the area in the shortest possible time.

Method:

The process of disaster management involved four phases: mitigation, preparedness, response, and recovery. The mitigation phase is the attempt to reduce disaster risks by focusing on long-term measures of eliminating disasters. The preparedness phase is the development of an action plan for an upcoming disaster. The response phase is the mobilization of services and relief when disaster strikes and the recovery phase is the restoration of the affected area to its previous state.

SAHANA is a free and open source disaster management system. It is a web-based collaboration tool that addresses common coordination problems during a disaster. It is a set of pluggable, web-based disaster management solution that provides solutions to problems caused by the disaster and it is designed to help during the relief phase of a disaster. Wireless technologies as well as the android environment for mobile development were used.

Algorithm:

Genetic algorithms are computational models inspired by evolution that provides a potential solution to a specific problem. It has a wide range of applications from optimization, test pattern generation, voice recognition, and image processing.

Genetic algorithm starts with a set of solutions (represented by chromosomes) called population. Solutions from one population are taken and used to form a new population. Its motivation is hope that the new population will be better than the old population. Solutions which are selected to form new are selected according to their fitness – the more suitable solution has a higher chance to reproduce. However, this depends on the selector operator used. The next phase is the crossover phase wherein the selected individuals are mated pair by pair to form a new offspring. Finally, some of the offspring are mutated. This is repeated until some condition (e.g. no. of populations or improvement of the best solution) is satisfied.

Advantages:

A disaster management system that facilitates the logistics for the rescue and relief operations during a disaster known as MyDisasterDroid was implemented in an Android- based mobile phone. Geographic locations of the people in need were sent via SMS or inputted directly to MyDisasterDroid. Determining the optimum route along the different geographic locations is similar to solving the travelling salesman problem wherein the geographic locations correspond to the cities and the rescuers or volunteers correspond to the travelling salesman. Using genetic algorithm, an optimum route along the given geographic locations was determined. Different genetic algorithm parameters were varied and based on the results, the best combination of genetic algorithm operators to use are the tournament selector operator with a tournament size of 20 and a swapping mutation operator with a mutation rate of 20 for 40 evolutions.

2.1.3 A Secure Big Data Stream Analytics Framework for Disaster Management on the Cloud [12]

Aim:

The paper presents a framework that supports emergency event detection and alert generation by analyzing the data stream, which includes efficient data collection, data aggregation and alert dissemination. One of the goals for such a framework is to support an end-to-end security architecture to protect the data stream from unauthorized manipulation as well as leakage of sensitive information. The proposed system provides support for both data security punctuation and query security punctuation. This paper presents the proposed architecture with a specific focus on data stream security. It also briefly describes the implementation of security aspects of the architecture.

Method:

There are two common approaches in evaluating or processing the data: batch processing and stream processing. Batch processing works in a store-and-process fashion, but for real-time event detection it is important to process the data on data streams. Therefore, batch processing is not capable to detect events in real time. Data stream processing is an emerging computing paradigm where a huge amount of data (Big Data) must be processed in real time (with minimal delay). Alert message generation in the cloud depends on the data set available for analysis. If the data is being modified in transit or the data is from malicious sources, there is a possibility of false alarms being generated or a failure to detect a real emergency event. In such situations, maintaining originality of the data is very important. It is of fundamental importance to develop a system which is sensitive and able to effectively recognize hazardous conditions. At the same time, a system should be intelligent enough not to overreact and trigger false alarms. So it is mandatory to deploy security verification mechanisms to collect only original data at a cloud data center for processing. There are several solutions proposed for communication between the source and the centralized station (cloud) and data analysis on cloud for disaster management. Figure 2.1 shows the complete architecture of big data stream query processing with possible security attacks, and the DSM (Data Stream Manager) structure for a security framework. Data transfer to stream, clustering and Bayesian network are standard data processing, but we did not address all these in our architecture.

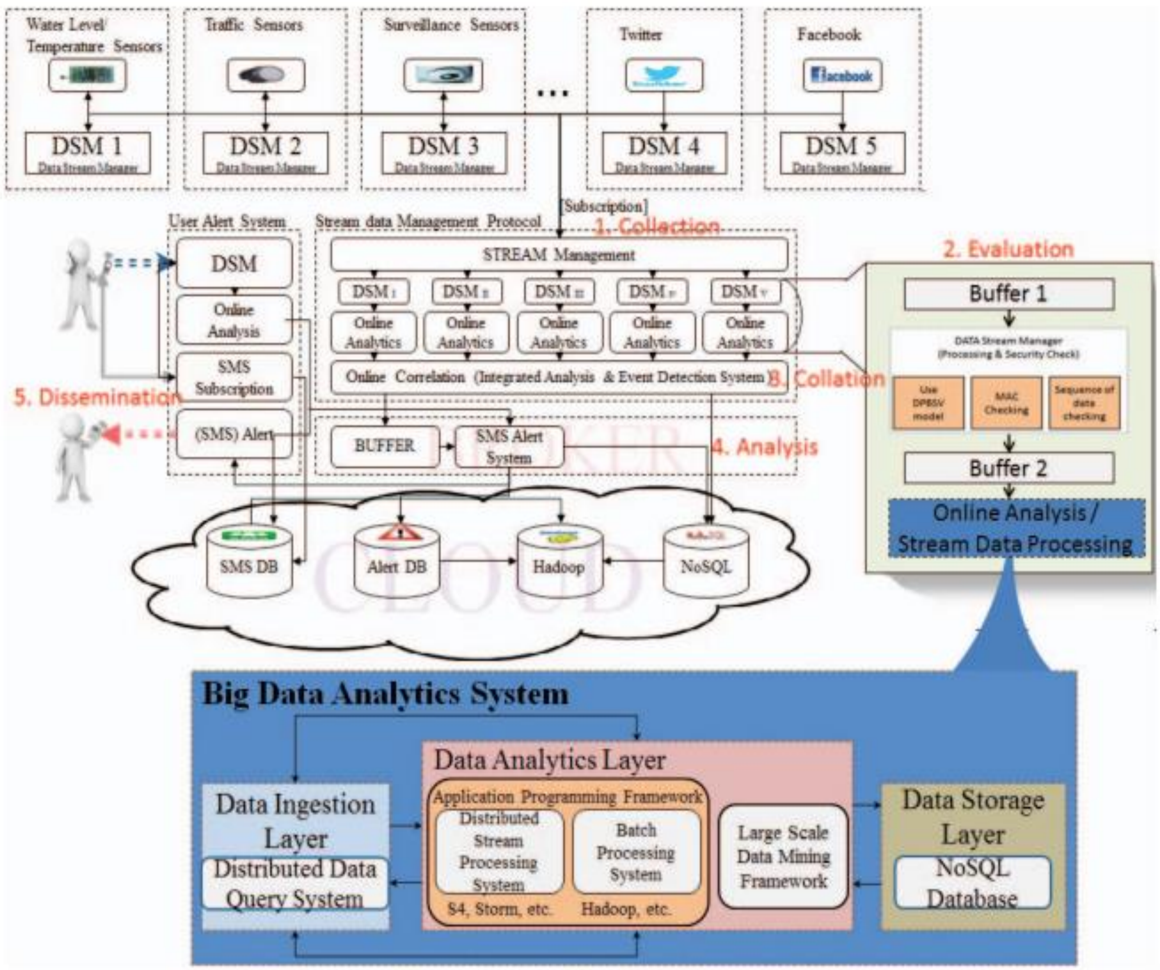

Figure 2.1: Architectural diagram for classification of streaming sources to alert generation with security verifications, access control and stream data analysis

The above figure 2.1 shows the complete architecture of the system from source device to cloud for data analytics, security framework and alert message generation and distribution. These five steps are defined as follows:

1. Collection: Data collected from different sources such as sensors, mobile devices, and social media for data analysis and event detection.

2. Evaluation: Stream data verified for security evaluation to maintain the originality of the data and go for online stream query processing.
3. Collation: Evaluate data from different DSMs and aggregate together for event detection and alert generation. Data also move to the cloud for batch processing.
4. Analysis: Analyze the data to detect event and generate alert messages. It includes two types of analysis: stream processing and batch processing.
5. Dissemination: This step is the output of the data analysis and distributes emergency alert messages if necessary.

Advantage:

Diversity of existing big data analytics frameworks in cloud makes the process of emergency decision making for software engineers, solution architects, or infrastructure administrators challenging. To address such an issue, this study explores the holistic system framework from data collection to security framework, data analytics and alert message distribution.

Limitation:

1. Many of the data sources emit data as a stream, which is an ordered sequence of data instances that can be read in real time using limited computing and storage capabilities.
2. Collecting data from different sources, verifying the originality of data, and processing data to extract actionable information in near real time is a challenging problem that demands a secure big data analytics framework.

2.1.4 The role of big data in disaster management

Aim:

Big data from these sources is getting attention from disaster managers and researchers. Management and analytics of big data can play a vital role in all phases of disaster management. Though the challenges are also remaining to use these big data in an emergency, but big data already proves its usefulness. As big data application in disaster management is comparatively new, many improvements in capacity level, management level, and research level will help to get the maximum benefits from this modern opportunity. [3]

Method:

The disaster management includes three basic phases: preparedness, response, and recovery. Big data can be defined as the collection of data sets which are large and complex and beyond the capacity of the conventional processing system (Deng and Di, 2013).Big data is available from dedicated sensor networks such as seismograph networks, remote sensing sensors (Pu and Kitsuregawa, 2013). With the development of internet and web service, data from the internet of things (IOT), social media and smartphones in the heterogeneous format are available for disaster management. Various methods are used for processing big data as mentioned in Table III. Crowdsourcing data is also proved its usefulness in emergency response. Big data is characterized by 5Vs (Volume, Velocity, Variety, Veracity, and Values). Planning for the unforeseen can be the key to disaster preparedness since accurate and timely information can reduce the loss from the catastrophe. The preparedness phase is challenging for multi-hazard because it is hard to model all consequence of a disaster due to the dependencies on many variables and uncertainties. Earthquake is an onset natural disaster, and there is no proper early warning system for an earthquake. Big Data analysis, however, can contribute to earthquake forecasting using satellite and atmospheric data combined with statistical analysis. For instance, Terra Seismic explores the potentialities to predict major earthquakes using satellite data anywhere in the world (Marr, 2015). This system measuring atmospheric anomalies causes by energy release or gas release by processing a large volume of satellite as well as ground based instrument.

Three aspects where big data can enhance disaster response are the critical area, real time situation analysis, and identification of most efficient response from past experiences. With the increasing number of smartphones, social media platforms, mobile application, and a large volume with the various formats of data can be gathered during a disaster. The powerful data stream from these sources brings the opportunity to improve disaster management through visualization and GIS mapping. Big data can help to identify the area where people need help and where to allocate resources. With the aid of advanced GPS system, it is possible to provide a guideline to the public for evacuation from the hazardous area.

The disaster recovery phase is mainly post-disaster activities including but not limited to damage assessment, information archiving from a lesson learned, and research activity to enhance resilience. Hong et al., (2015) implemented parallel processing of Unmanned Aerial Vehicle (UAV) imagery for three dimensional building damage detection to get more precise damage information instead of traditional two-dimensional damage assessment. As UAV image is big data and the rapid information is crucial for emergency management, the prime objective of their research is to reduce the processing time of big volume data. [3]

Table II: Big Data Paradigm [14]

Area	Subarea	Description
Methods	1. Data mining 2. Machine learning 3. Statistical analysis 4. Analytical tools	1. Process to explore large amount of data to find meaningful patterns and rules. 2. An area of Artificial Intelligence (AI), used to discover knowledge and make intelligent decisions. 3. Deals with the collection, analysis, interpretation, presentation and organization of data 4. Examines raw data to uncover hidden patterns, correlations and insights
Processing	1. Hadoop engines. 2. Real-time analytics 3. Cloud sourcing	1. Open source Java Framework technology to store and process big data at less cost with high degree of fault tolerance and high scalability 2. It refers to a level of responsiveness that a system process and make useful insights out of it as immediate or nearly immediate upon receiving. 3. Outsourcing of IT resources to reduce costs and get access to technological expertise with fewer resources.
Representation	Visualizations	It provides at-a-glance views of key performance indicators for a particular objective. E.g. Tableau software is used to make dashboards based on real-time data, Google charts, etc.

Algorithm:

Machine learning algorithms actively adapt with and learn the problem without relying on statistical assumptions about data distribution. Machine learning algorithms have an overall improved accuracy compared to traditional classification and change detection methods. They work with nonlinear datasets, allow learning with limited training data, and successfully solve classification problems that are difficult to distinguish. [13]

Advantages:

The disaster managers and policy makers will get more confidence on the usefulness of big data in disaster management. More data we gather, more new analytical tools and techniques will be developed to get appropriate information from these data to enhance disaster management. The organizations involved with disaster management will strengthen their capacity to utilize this growing opportunity. [3]

Limitations:

1. Policy makers, disaster response organizations, are not very familiar with computer generated data. Thus, these data have been ignored in the decision-making process.
2. Accurate forecasting and early warning system is an important aspect of disaster preparedness if these forecasting systems fail to do a good job, there is a chance to lose the public trust, and for instance false information may lead to resource allocation in wrong place or may result in the improper rescue operation.
3. However, a large volume of data can help to increase the accuracy and reliability of these model simulation but high volume data not necessarily mean the accurate data. Therefore, the noise removal from these big data is crucial to get results right.
4. The language barrier is often hindering the disaster management process. Crowdsourcing and data mining approaches also suffer from language and cultural difference.
5. Many disaster management specifications are contextual basis; thus, this information and contents might need to be translated. Crowdsourcing approach has sometimes been used to translate necessary information for common understanding, translation of text message during Haiti earthquake can be a great example.
6. Data privacy could be another challenge to deal with. Many social media data are considered as private. Apart from this privacy concern, access to some parts of social media data related to an emergency might be helpful.

7. Another challenge could be the resilience of big data system itself. The dependency on these big data will increase with the frequent use of big data in disaster management. However, what will happen if these systems are affected by the disaster? It is important to make system architecture resilient to disaster, so that, data can be recovered and retrieve without experiencing any problems. Pu and Kitsuregawa, (2013) mentioned some ways to achieve resilience such as having a redundant backup system, cloud based distributed storage system and distributed computing systems. Data storage with multiple formats might also help to reduce the loss of data.[3]

2.2 Existing System

2.2.1 A mobile disaster management system using the Android technology [11]

Aim:

The disaster management system Android application known as MyDisasterDroid determines the optimum route along different geographical locations that the volunteers and rescuers need to take in order to serve the most number of people and provide maximum coverage of the area in the shortest possible time.

Method:

The process of disaster management involved four phases: mitigation, preparedness, response, and recovery. The mitigation phase is the attempt to reduce disaster risks by focusing on long-term measures of eliminating disasters. The preparedness phase is the development of an action plan for an upcoming disaster. The response phase is the mobilization of services and relief when disaster strikes and the recovery phase is the restoration of the affected area to its previous state.

SAHANA is a free and open source disaster management system. It is a web-based collaboration tool that addresses common coordination problems during a disaster. It is a set of pluggable, web-based disaster management solution that provides solutions to problems caused by the disaster and it is designed to help during the relief phase of a disaster. Wireless technologies as well as the android environment for mobile development were used.

Algorithm:

Genetic algorithms are computational models inspired by evolution that provides a potential solution to a specific problem. It has a wide range of applications from optimization, test pattern generation, voice recognition, and image processing.

Genetic algorithm starts with a set of solutions (represented by chromosomes) called population. Solutions from one population are taken and used to form a new population. Its motivation is hope that the new population will be better than the old population. Solutions which are selected to form new are selected according to their fitness – the more suitable solution has a higher chance to reproduce. However, this depends on the selector operator used. The next phase is the crossover phase wherein the selected individuals are mated pair by pair to form a new offspring. Finally, some of the offspring are mutated. This is repeated until some condition (e.g. no. of populations or improvement of the best solution) is satisfied.

Advantages:

A disaster management system that facilitates the logistics for the rescue and relief operations during a disaster known as MyDisasterDroid was implemented in an Android- based mobile phone. Geographic locations of the people in need were sent via SMS or inputted directly to MyDisasterDroid. Determining the optimum route along the different geographic locations is similar to solving the travelling salesman problem wherein the geographic locations correspond to the cities and the rescuers or volunteers correspond to the travelling salesman. Using genetic algorithm, an optimum route along the given geographic locations was determined. Different genetic algorithm parameters were varied and based on the results, the best combination of genetic algorithm operators to use are the tournament selector operator with a tournament size of 20 and a swapping mutation operator with a mutation rate of 20 for 40 evolutions.

2.2.2 A mobile disaster management system on mobile phones using Google maps [16]

Aim:

It is a disaster management system and evacuation system for people using Google Map (GM). The system is implemented on android mobile phone because of the burgeoning growth of smartphones in world. Android device with our application installed on it and user .User can register the multiple receiver or family member or friends to send SMS at a time to send notification for help.

Method:

Our proposed location based disaster preparedness system consists of a GPS supported android mobile phones with our proposed application installed on it and users having national id. The user of our proposed android application can also register the number of family members, relative, friends to send message for help. Through network provider or GPS provider, mobile phone gets the current location of its user and sends it to register member by the user in its application.

Algorithm:

1. Our application gets the current position through GPS or network provider from the user mobile phone send the latitude and longitude of user's current position.

2.When the user in a disaster zone, it will send a message notification along with path to the member who can help.in this message user send text message and path to reach for help.The path contain the distance from helper to the user. The distance can be given with the help of Google Map to show the distance by walk, by railway or by bus.

3. When the user in a disaster zone, it will send a text message.

4. Weather forecast detection also in this application

5. The application uses GPS to detect the phone's current position and sends this data through the text messages to the member which user already insert in to the application for help. Our application will run on android 2.3.3 supported mobile phones and upgraded versions. The mobile phone should have GPS supported location identification facility. The application provides disaster warning in Manual Notification. The application also has a button which send messages to the member from which user can get help.

Advantages:

GoogleMap (GM) is a rapidly growing open source map of the world because of the availability of map information across the world and the advent of low-cost convenient GPS devices [3]. So, the popularity to using Google map increases. The demand of location based services is also increasing day by day with the burgeoning growth of smartphones. The proposed system is developed for the normal people. The usability of google Map (GM) is ensured for all users as it is free. Users of our application will send text message along the direction on the map to one or more family member which number user register in that application.

2.2.3 Location Based Early Disaster Warning and Evacuation System on Mobile Phones Using OpenStreetMap [17]

Aim:

It is a location based early disaster warning and evacuation system for both normal and blind people using OpenStreetMap (OSM). The application recognizes the user in probable disaster zone then application will disseminate visual and audio disaster warning and evacuation guideline including shortest path of shelter or safe zone on the map of the application.

Method: Disaster Management Server (DMS), a third party server, stores disaster prone areas and the details about the users in its database. Regional weather offices can access the database to update the disaster prone areas. The user of the proposed android application can also register Android Cloud to Device Messaging (C2DM) server to get notification on application only when there is any update in the database of DMS. Through network provider or GPS provider, mobile phone gets the current location of its user and sends it to server. Using this current position, our system determines whether the user is in probable disaster affected area or not.

Algorithm:

1. Register/Unregister to C2DM Server: Android Cloud to Device Messaging (C2DM) server is a Google server involved in taking messages from the third-party application server and sending them to the device. Cloud to Device Messaging (C2DM) is also a service to send data from server to the applications on Android devices. The service provides a simple, lightweight mechanism that servers can use to tell mobile application to contact the server directly, to fetch updated data.After registering on this server, application receives a registration id which is stored into the Disaster Management Server (DMS).
2. Communication with Disaster Management Server (DMS): Our application gets the current position through GPS or network provider from the user mobile phone and application communicates with DMS to send the latitude and longitude of user’s current position.
3. Probable Disaster Affected Area Determination: The DMS server runs Ray casting algorithm to determine the probable disaster affected areas. The algorithm counts the number of intersections for a ray passing from the exterior of the polygon to any point. If the count value is odd, it shows that the point is inside of the polygon. And if the count value is even, it shows that the point is outside of the polygon.

4. Dissemination of Disaster Warning and Evacuation Guideline: When the application finds its user in a disaster zone, it will give him a notification along with an audio message and vibration.
5. Tracking Evacuation Progress: If the user is in any disaster zone, the application will automatically start another service on the mobile phone to track the user so that the authority can rescue him from probable disaster affected area.

Advantages:

The proposed system is developed for both the normal and blind people. The usability of OpenStreetMap (OSM) is ensured for all users as it is free. Users of our application will get both text and audio warning message and evacuation direction on the map. OpenStreetMap (OSM) is a rapidly growing open source map of the world because of the availability of map information across the world and the advent of inexpensive portable GPS devices instead of using restricted Google map.

Chapter 3

System Design

3.1 Architectural Diagram

The proposed system is a cloud based fire prediction system that collects, organizes, manages several fire hazard databases and helps in the detection of fire hazards. The system will be available to the users in the form of a web application.

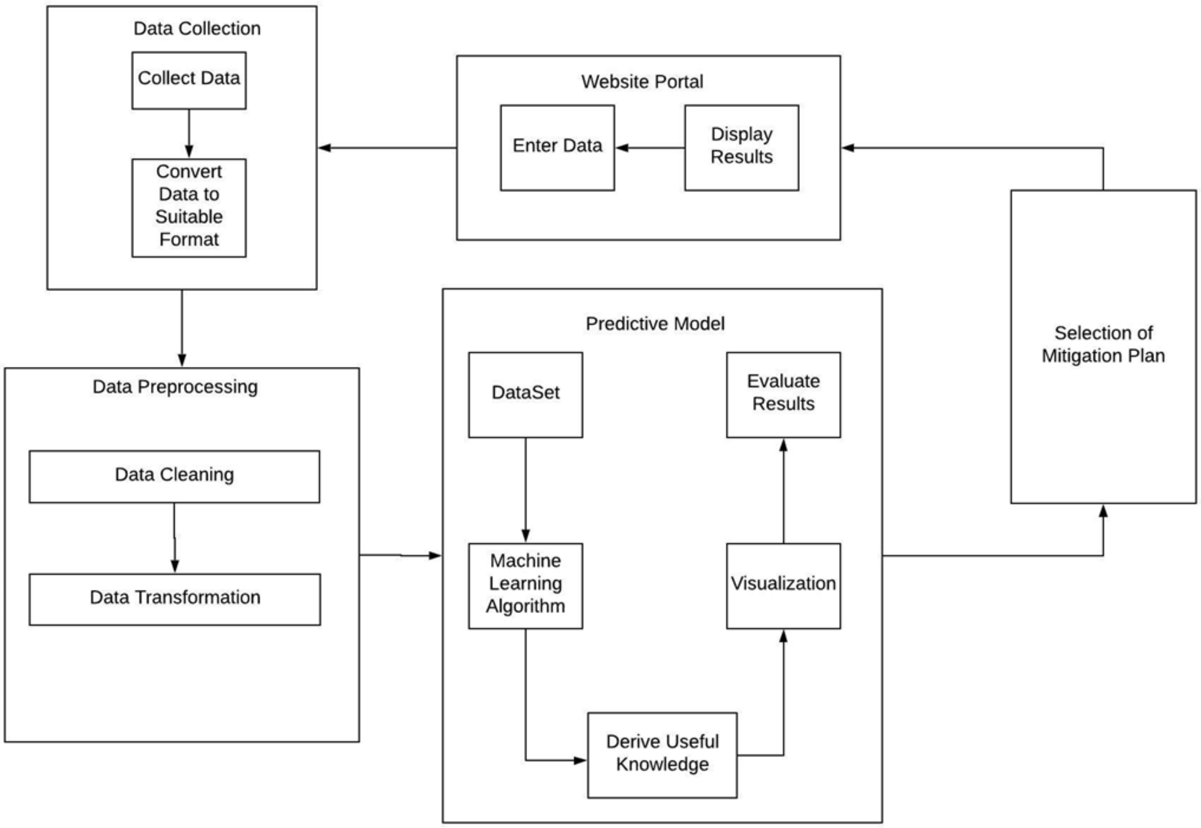

Figure 3.1: System Architecture Block Diagram

3.2 Working of the System

The figure 3.1 explains the block level representation of the project. It shows how different modules will be interacting with each other and also the entire flow of the system.

a. Data Collection Module: In this phase we collect relevant data from different sources. Data are collected from various sources for analysis and event detection. Data sources considered for obtaining fire related data are the fire stations. Collected data are in different formats such as handwritten, Marathi Script in paragraph format. The corresponding data is aggregated to a csv file.

b. Data Preprocessing Module: The data acquired through the data collection process is processed to convert it into a format which is favorable for the Machine Learning Process. In this module data cleaning and data transformation is performed.

c. Predictive Model Module: This module uses a machine learning algorithm to implement a Predictive model. The typical workflow for machine learning module consists of many phases:

- Identifying a fire related problem to solve and a metric for measuring results
- Identifying the best attributes for the collected data for prediction of fire
- Using models to generate predictions, and provide mitigation steps.
- Deriving useful knowledge that was previously unknown.
- Visualizing models using graphs, plots, etc.

d. Selection of Mitigation Plan Module: Based on the results obtained from the Predictive model, this module will decide as to what mitigation plan will be the most suitable one. **e. Website Portal Module**: The final results of our project will be displayed on a website Portal. This application consists of a login module for the fire officer who wants to use it to enter new fire incidents. Mitigation plan for the particular disaster will also be displayed on website for the users.

3.3 DFD

3.3.1 Level 0

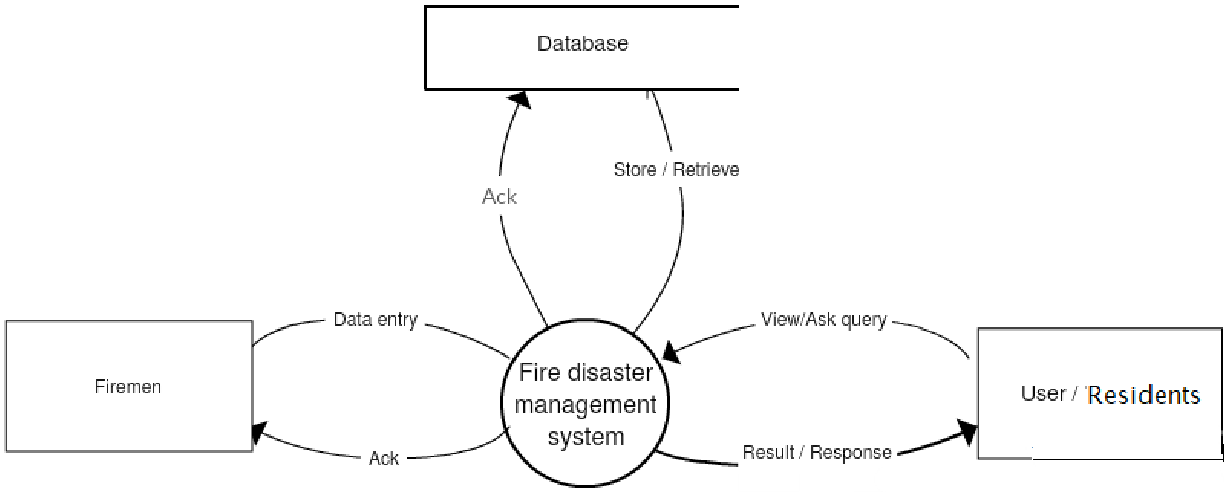

Figure 3.2: Level 0 DFD diagram

The figure 3.2 shows the implementation of the F-Alert System from an external persons Point of view. It shows how data is exchanged between system's major entities- residents and firemen.

3.3.2 Level 1.0

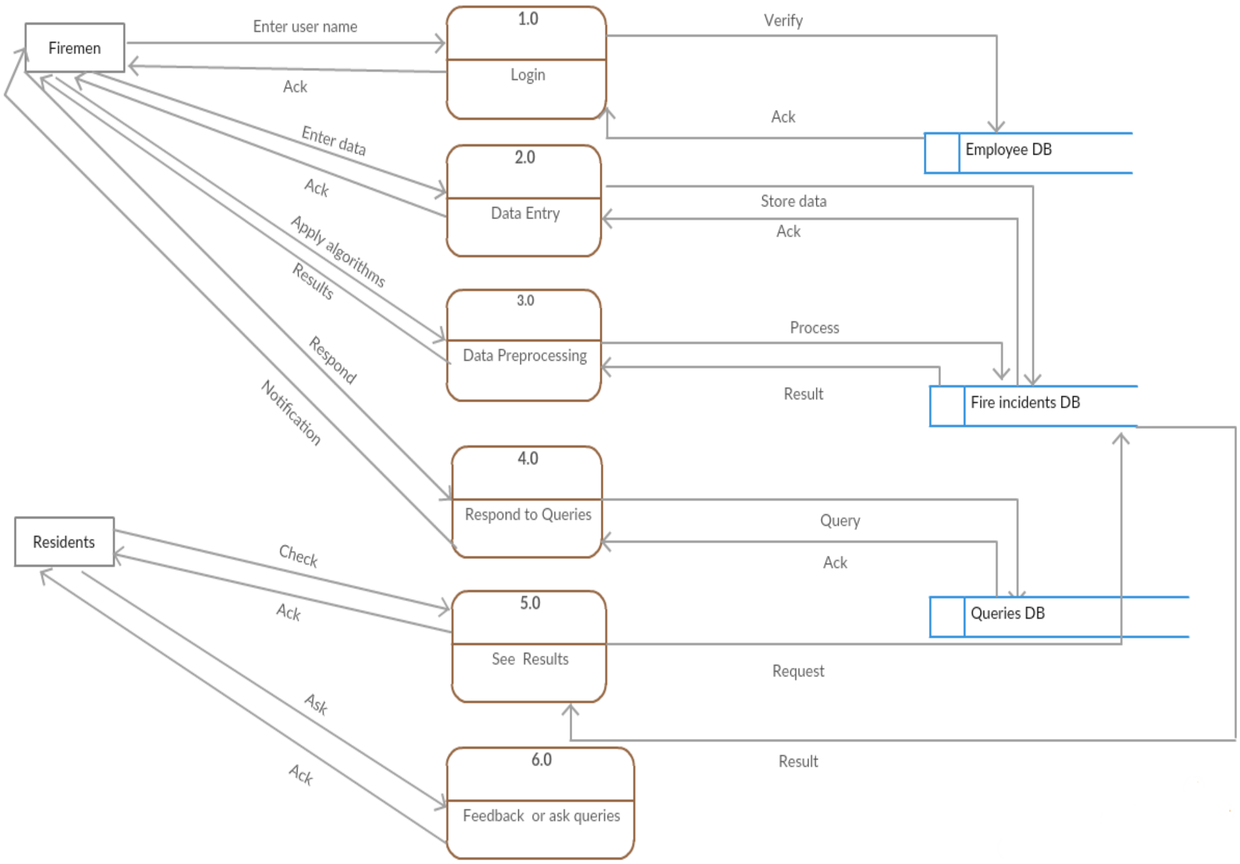

Figure 3.3: Level 1 DFD diagram

The figure 3.3 shows Level 1.0, which shows the different stages within the F-ALERT system and the data which is being passed from one stage to another.It also shows the data flow and processes involved within the Data Preprocessing Module and Machine Learning/ Predictive Model Module.

3.3.3 Level 2

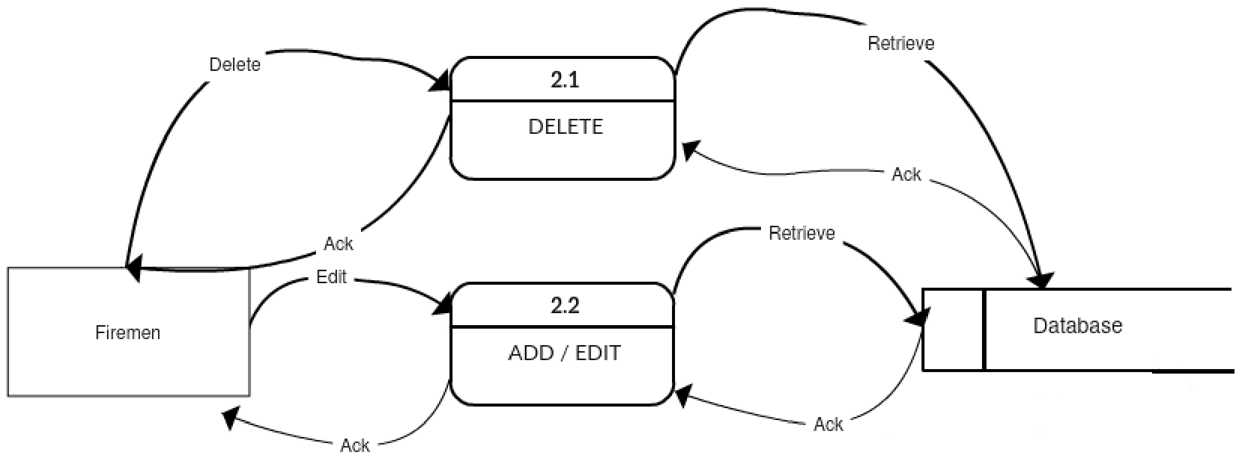

Figure 3.4: Level 2 DFD diagram

The figure 3.4 shows the operations that can be performed by the admin firemen in the data entry stage involved in level 1.

3.4 Flow chart

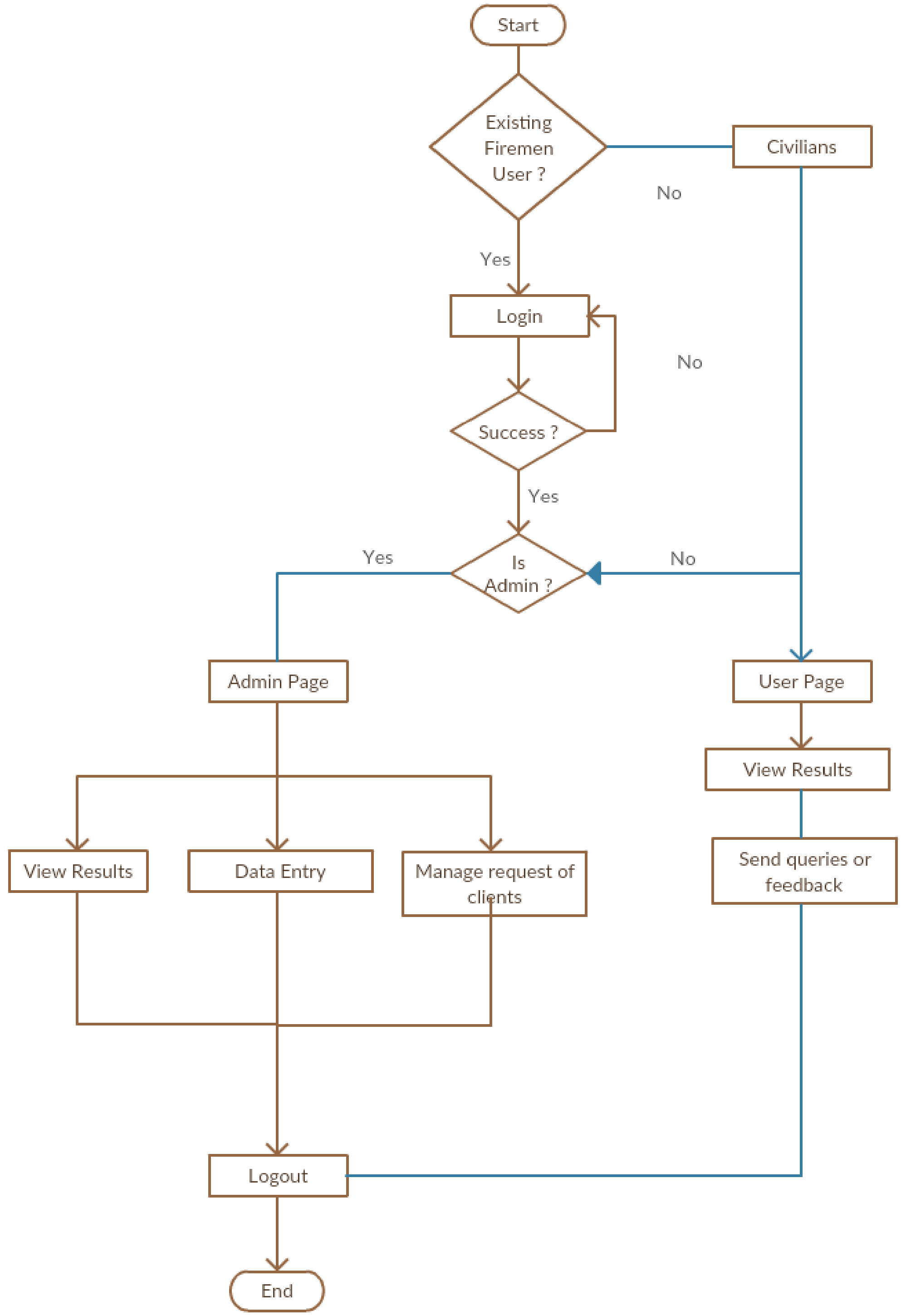

Figure 3.5: Flow chart

3.5 Use Case diagram

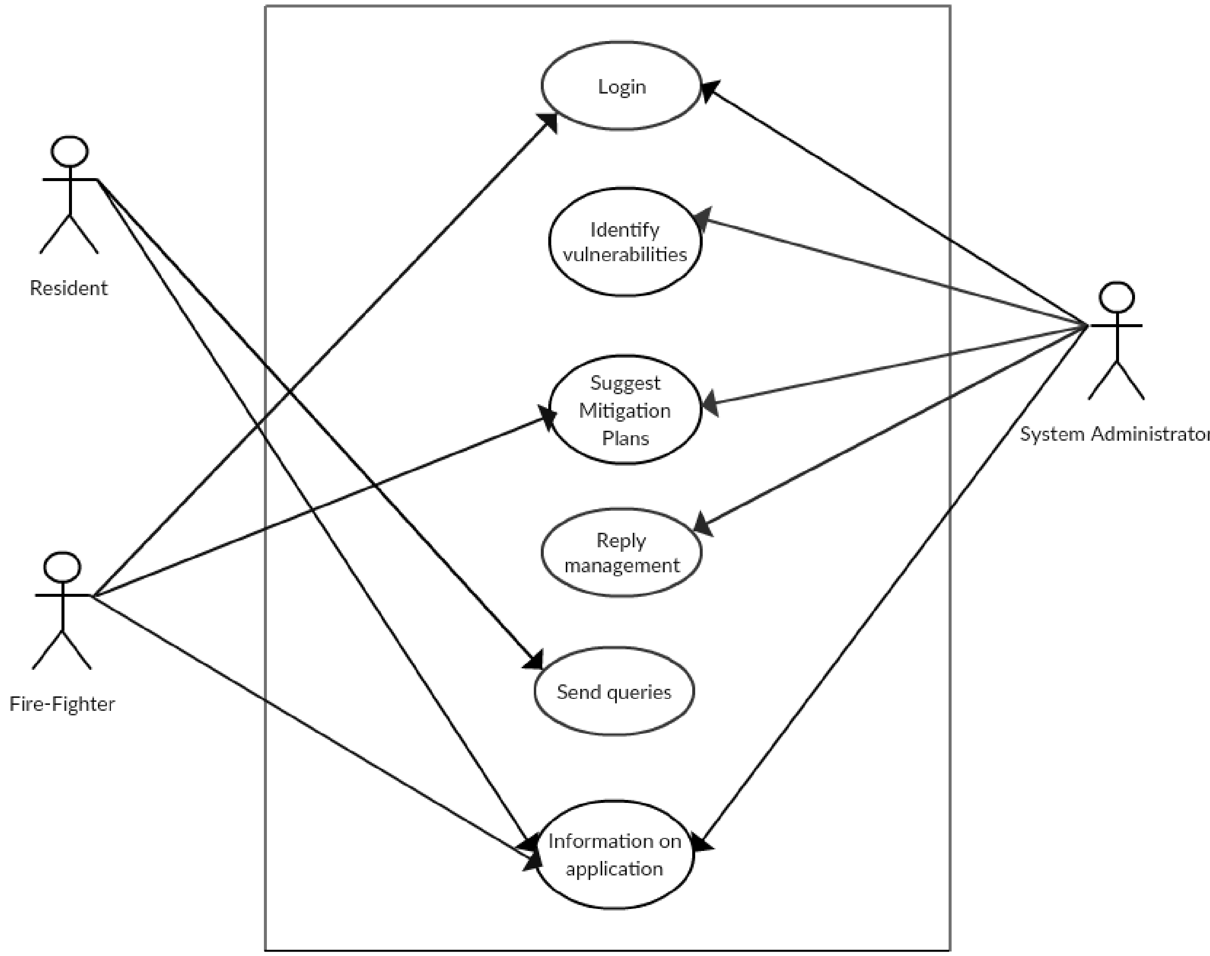

Figure 3.6: Use Case diagram

Chapter 4

Implementation Details

4.1 System Requirements

4.1.1 Software

- Python development environment eg. Enthought Canopy
- Jupyter Notebooks
- Heatmaps
- Browser with support for HTML5 and CSS3
- Microsoft Excel (for data cleaning and data transformation)

4.1.2 Hardware

- System : Core i3
- Monitor: 14’ Colour Monitor.
- Mouse: Optical Mouse.
- Ram: 2 GB.
- Processor: i3 2.0 Ghz +
- Primary Memory : 8GB RAM
- Hard Disk: 40 GB.

4.2 Algorithms used

The different types of algorithms that we are going to use in our system are as follows:

4.2.1 Multivariate regression [7]

Multivariate regression will be utilized in our system as it belongs to the class of supervised learning. There are various problems addressed by this regression like if the data is collected and some of the variables are defined, the relationship between them can be determined.

- Selecting the features: Finding the features on which a response variable depends (or not) is one of the most important steps in Multivariate Regression. To make our analysis we have selected the features to be the call time which denotes the time when there is a fire-breakout and the extinguish time when the fire was extinguished.
- Normalizing the features: The call time is then scaled in the range of (0,24) to represent in the 24 hour format in order to make better analysis.

So this algorithm will be used in analyzing the data, getting best possible outcome for the input and for forecasting and finding out cause and effect relationship between variables and will also help to detect outliers.

4.2.2 Clustering

Clustering is performed to group a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups.

a) K-Means

K-means algorithm will be used for clustering in our system. It is used for clustering of call time (reporting time of any fire that has occurred). This is used to find out which time of the day is most prone to fire accidents as derived from the past incidents. This algorithm is also used to cluster Severity Scores and Extinguishing Time (Time required to extinguish the fire). This is a crucial step in order to determine more prone areas from lesser prone areas.

b) DB Scan

DBScan is another clustering algorithm that is used for visual purposes. It is used to plot regions or zones in the map based on the severity of fire hazards that have taken place in that area. Severity is shown on the map with the help of the color of the zones that are plotted using the clustering algorithm.

Chapter 5

Test Cases

TEST CASES

The system was thoroughly tested for following test cases:

Test Case ID	Test Case	Test Case Description	Tester	Test Result
TC1	**User verification**	**The user with the valid credentials is allowed to login or not**	**Purva**	**PASS**
TC2	**Login**	**While login, password is hidden or not and prompt message is shown on screen**	**Jincy**	**PASS**
TC3	**Software Setup**	**Use the valid credentials and then try to check whether whole website works as per requirements**	**Elvis**	**PASS**
TC4	**Slow Connection**	**Testing the network connection**	**Kunal**	**PASS**

Chapter 6

Experimental Results

6.1 GUI

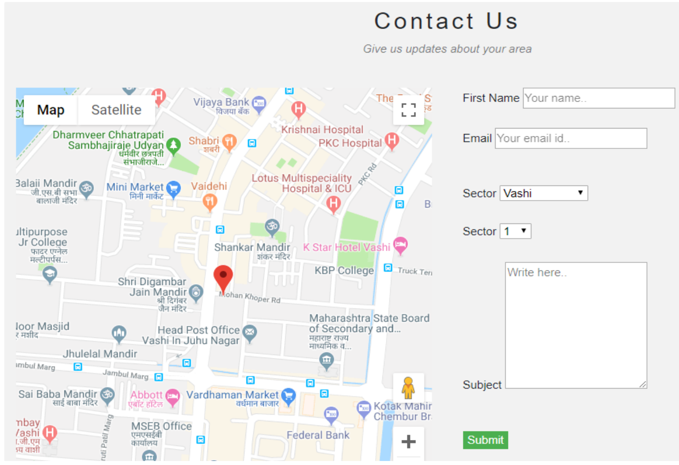

Figure 6.1: GUI of the system

The above figure 6.1 includes the contact us page of the website where the residents can send the queries regarding the fire safety of the areas.

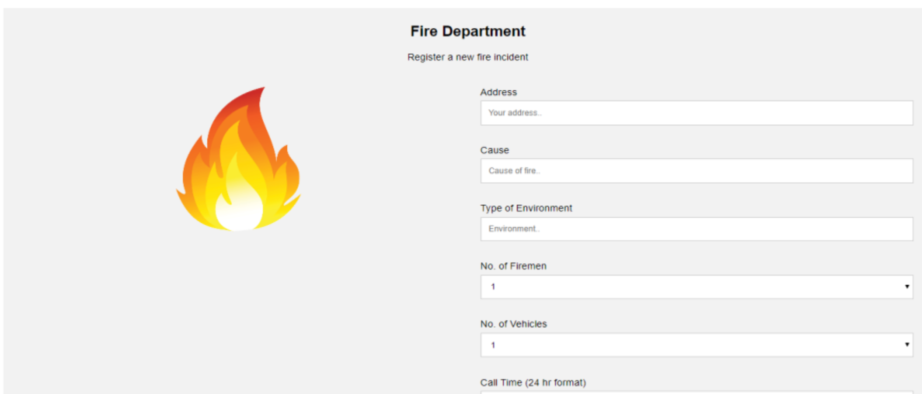

Figure 6.2: Data Entry System

The figure 6.2 includes the dataset which includes the following attributes: "Location" of the fire incident ,the "Cause of fire", the "Type of environment" where the fire has occurred (road , residential ,office) , "Number of firemen" sent to extinguish the fire "Number of vehicles" required , "Call time" represents at what time the fire occurred, the "Return time" helps to identify the time required to extinguish the fire, "Second call" depicts the fire was severe and more vehicles were needed to extinguish it ,any "Casualties occurred", and the "Action taken" by civilians or firemen.

The data is classified based on the attributes of the dataset. This dataset will be updated by the firemen by entering details through a web-form, while encountering with new fire incidents.

The website consists of an admin webpage that takes the input from firemen on the incidents noted by them on a daily basis and this will be directed to the csv file. This data can be used for further analysis to get more accurate results.

6.2 Result analysis

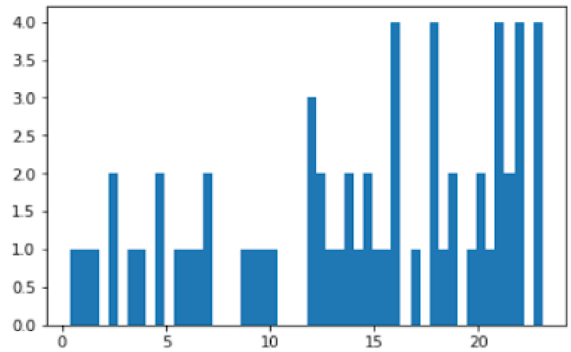

Figure 6.3: Histogram

The figure 6.3 represents a histogram which tells the frequency of fire occurrences at different times of the day. Here the graph shows the relationship between distribution of time of the day (x-axis) and frequency of occurrence (y-axis). From this histogram we can infer that a there is a sizeable increase in number of cases of fire from noon to midnight as compared to number of cases that occur from midnight to noon.

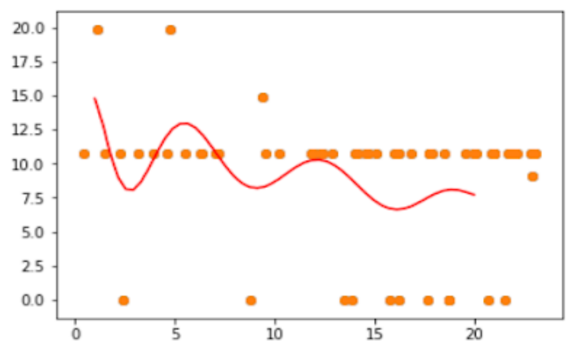

Figure 6.4: Multivariate regression graph

The figure 6.4 represents a graph for the multivariate regression which shows the relationship between distributions of severity score (x-axis) and the time of occurrence in a day (y-axis). The graph helps to predict that throughout the day at what time the fire incident occurred was more severe. From this plot we can see that the cases that occurred from midnight to noon were more severe as compared to the cases that occurred from noon to midnight.

Hence we come across an interesting pattern, from the figure 6.3 and figure 6.4; we can see that even though there are more number cases that occur during the day time from noon to midnight, these cases are usually less severe as compared to those that occur at night i.e. from midnight to noon, which are more severe.

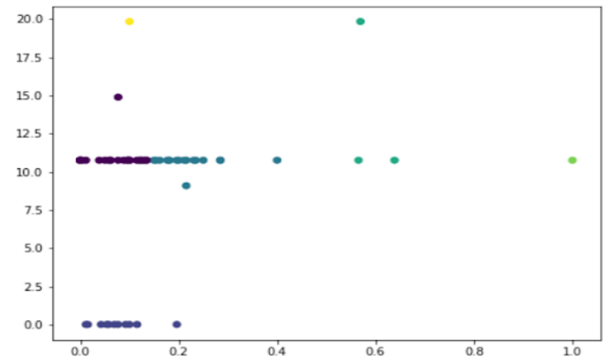

Figure 6.5: K-means Clustering

The above figure 6.5 represents clustering is done by using parameters as extinguish time of fire with respective the severity score of it, it tells the time at which most of the incidents occurs. This is an important step in determining zones that are more severe as compared to zones that are lesser severe to fire hazards.

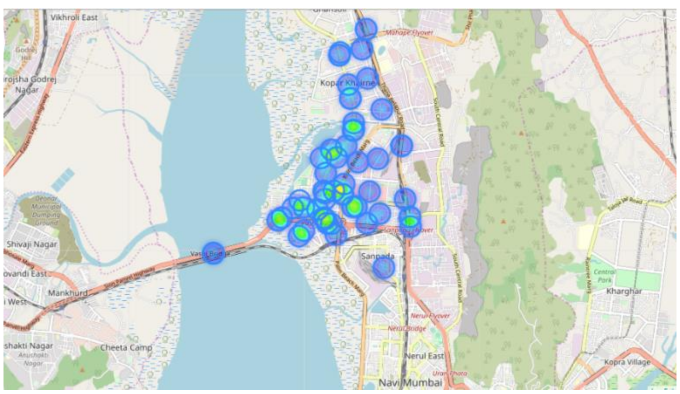

Figure 6.6: Map showing areas prone to occurrence of fire

The above figure 6.5 represents a map of Vashi it shows all the fire incidents that were happened in 2018 and the blue area shows the low possible occurrence regions and the green areas depicts the high possible occurrence regions in that area.

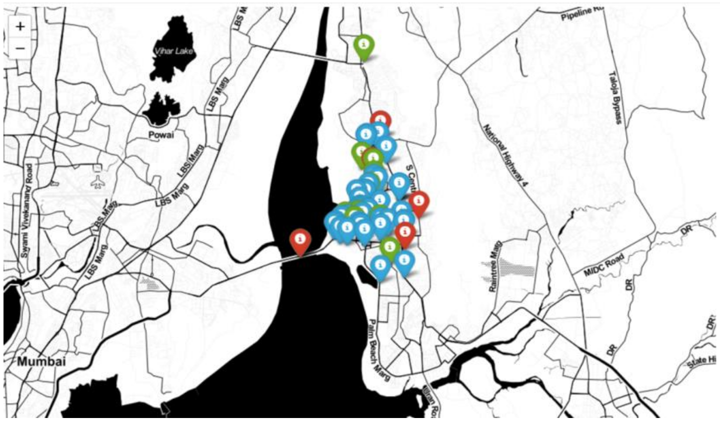

Figure 6.7: Map showing areas prone to different levels of severity of fire

The above figure 6.6 represents a map of Vashi, it shows all the fire incidents that were happened in 2018 and the red markers indicate highly severe fire zones, blue markers indicate moderately severe fire zones and green markers indicate low severe fire zones.

Chapter 7

Conclusion

7. Conclusion and Future scope

7.1 Conclusion

In conclusion, we can state that our system is capable of creating useful results which can be used for precautionary efforts. In addition to the data analytics performed on the collected data, our project also helps in maintaining a digital record of all the incidents that occur. This is advantageous in the sense that digital data is more useful as compared to handwritten journals. This is an important feature of our project as there is no digital record of data for fire incidents in India. All the data is present in the form of handwritten journals which cannot be used for data analytics purpose. In conclusion, this project achieved its objective and goals in producing useful results that prove to be a key in avoiding fire related disasters.

7.2 Future Scope

This system is based on collecting data, determining parameters that lead to a problem and draw important results through data analytics. Hence this system can be modified to predict other disasters such as road accidents, cases in which an ambulance is required, types of crimes in an area. The applications of systems like these are vast as data analytics through data mining and machine learning provide new improved ways of predicted future outcomes or predicting the possibilities of certain desirable or undesirable events.

References

[1] Author: Satish Modh, Edition First, VES Institute of Management, Studies and Research, Mumbai, Publisher: Macmillan Publishers India, dated January 2010

[2] Md. Shahinoor Rahman, Liping Di and Md. Esraz-Ul-Zannat ,WORLD FOCUS special issue “Disaster Management In India” dated July 2014

[3] Md. Shahinoor Rahman, Liping Di , Md. Esraz-Ul-Zannat : The Role of Big Data in Disaster Management ,September 23 - 24, 2017

[4]Marion,Unsplash,Online:https://www.eartheclipse.com/natural-disaster/causes-and-effects-of-drought.html,Accessed on 27 September 2018

[5] The International Federation of Red Cross and Red Crescent Societies (IFRC) Online:http://www.ifrc.org/en/what-we-do/disaster-management/about-disaster-management/ Accesed on 18 September 2018

[6]Karlsruhe Institute of Technology Online:https://www.sciencedaily.com/releases/2016/04/160418092043.htm, Published :18 April, 2016

[7] Neeraj Kumar; Alka Agrawal; R. A. Khan : Emergency Information System Architecture for Disaster Management: Metro City Perspective, May-June 2017

[8]Frank Press, Robert M. Hamilton, Online : http://science.sciencemag.org/content/284/5422/1927, Published :28 September 2018

[9]NorthropGrummanCorporation,Online:http://now.northropgrumman.com/tsunamis-hurricanes-eruptions-predicting-a-natural-disaster , Accessed on 30 September 2018

[10] Mr.T.Rajasekaran, J.Sruthi, S.Revathi , N.Raveena :FOREST FIRE PREDICTION AND ALERT SYSTEM USING BIG DATA TECHNOLOGY, Accepted: 31 July 2015.

[11] Jovilyn Therese B. Fajardo, Carlos M. Oppus, A mobile disaster management system using the Android technology, Department of Electronics, Computer, and Communications Engineering Ateneo de Manila University, 15 Jan , 2014

[12] Deepak Puthal, Surya Nepal , Rajiv Ranjan and Jinjun Chen :A Secure Big Data Stream Analytics Framework for Disaster Management on the Cloud, on 31 May 2016.

[13] Manzhu Yu , Chaowei Yang ID and Yun Li :Big Data in Natural Disaster Management: A Review, 4400 University Drive, Fairfax, VA 22030, USA, Published: 5 May 2018

[14] Muhammad Arslan, Ana Roxin, Christophe Cruz, Dominique Ginhac. A Review on Applications of Big Data for Disaster Management. The 13th International Conference on Signal Image Technology Internet, Dec 2017, Jaipur, India.

[15] CFS fire academy ,Online :www.campussafetyexpo.com ,Accessed on 30 September 2018

[16] Varsha S. Sonwane,A mobile disaster management system on mobile phones using Google maps,International Journal of Computer Science and Information Technologies, Vol. 5 (5) , 2014

[17] Khandaker Mustakimur Rahman; Tauhidul Alam; Mahfuzulhoq Chowdhury, Location Based Early Disaster Warning and Evacuation System on Mobile Phones Using OpenStreetMap, 21-24 Oct. 2012

[18]IndiaToday.in,Online-https://www.indiatoday.in/india/story/in-6-months-mumbai-has-seen 12-major-fires-22-deaths-1259486-2018-06-13, Updated on 13 June, 2018

APPENDIX A: CODE SAMPLE

Code Sample for displaying maps:

```
import folium
from folium import plugins
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import pandas as pd
df = pd.read_csv("titan.csv")
df.head()
m = folium.Map([19.077515, 72.998388], zoom_start=13)
m
# mark each station as a point
for index, row in df.iterrows():
    folium.CircleMarker([row['latitude'], row['longitude']],
                        radius=15,
                        popup=row['formatted_address'],
                        fill_color="red", # divvy color
                       ).add_to(m)
# convert to (n, 2) nd-array format for heatmap
stationArr = df[['latitude', 'longitude']].as_matrix()
# plot heatmap
m.add_child(plugins.HeatMap(stationArr.tolist(), radius=15))
m
```

Code for k-means clustering:

```
import pandas as pd
df = pd.read_csv("F-Alert1.csv")
df.head(30)

%matplotlib inline

from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
from sklearn.preprocessing import scale
from numpy import random, float
```

```
data = df2.values[:,:]

model = KMeans(n_clusters=6)

# Note I'm scaling the data to normalize it! Important for good results.
model = model.fit(scale(data))

# We can look at the clusters each data point was assigned to
print(model.labels_)

# And we'll visualize it:
plt.figure(figsize=(8, 6))
plt.scatter(data[:,1], data[:,0], c=model.labels_.astype(float))
plt.show()
```

APPENDIX B: PAPER PUBLISHED AND CERIFICATES

F- ALERT: Fire Disaster Management Using Machine Learning

Purva Gharat+
purvagharat28@gmail.com

Kunal Das+
KunalGTN7100@gmail.com

Jincy Jacob+
jincysarajacob2597@gmail.cm

Prof. Archana Shirke+
archanashirke25@gmail.com

Elvis D'Costa+
elvisdcosta44@gmail.com

Information Technology+
Fr. C. Rodrigues Institute of Technology
Vashi, Navi Mumbai, India

***Abstract* — It is estimated that more than USD 7 trillion worth of economic damage and 8 million deaths via natural disasters have occurred since the start of the 20th century and have caused bad impact on the society, destroying lives and properties. Also the year 2018 continues to prove catastrophic for Mumbai when it comes to combating fire incidents. Introduction of technological advances into disaster management has proven to be of great value. Despite these advances, no one can tell with complete accuracy when a place will catch fire, or how powerful a hurricane will be on landfall. However, observation and data are powerful tools, so prediction is getting better and faster. This data analysis is achieved by using various technologies such as Data Mining and Big Data, using which we can check for certain vulnerabilities from the data collected and analyze the possible disaster that can occur due to these vulnerabilities and also list out the different steps to mitigate the disaster. The benefit of being prepared for calamities and disasters is the fact that, in a society it can save one's life or health even after the disaster has stricken. Thus with preparedness, response and recovery techniques, disaster can be managed to a certain extent.**

***Keywords* — Fire management,Prediction, Big data ,mitigation plan**

I. INTRODUCTION

Disaster management essentially deals with the management of resources and information towards a disastrous event and is measured by how efficiently, effectively and seamlessly one coordinates the resources.

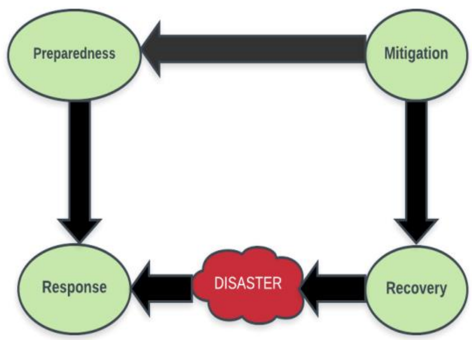

Figure 1: Steps involved in any disaster management system

As shown in Figure 1, disaster management includes three basic phases: preparedness, response, and recovery. The efficient data collection, archiving and analytics is essential for effective disaster management, for which big data plays a very important part. Big data can be defined as the collection of large and complex data sets that are beyond the capacity of the conventional processing system. The preparedness phase is challenging for multi-hazard because it is hard to model all the consequences of a disaster due to the dependencies on many variables and uncertainties. This kind of dependency modeling requires big data with high accuracy for better modeling. It is possible to project the probable scenario of a catastrophe in big data era [2].

The previous year 2018 proved to be disastrous for Mumbai when it came to combating fire incidents. Ever since the tragic fire that broke out at Kamala Mills which killed 14 people in December 2017, the city has seen 12 major fire tragedies claiming about 22 lives, besides injuring many and destroying properties. A fire break out at a snack shop in Saki Naka area of the city in December 2017 caused the life of twelve people. On January 4, four people died after a fire broke out in Maimoon Building that housed residential complexes in the suburban part of Marol in Mumbai. Cine Vista, a film studio in KanjurMarg (West) was gutted in a massive fire on January 6. A fire was reported in the Mumbai’s Sessions Court building on the morning of January 8. Two firemen were injured in a massive fire that broke out on January 22 in an industrial area in Mumbai's Andheri. A fire broke out at a four-storey Army building in south Mumbai's Colaba area on March 17 and many more fire incidents have taken place in the year 2018 in Mumbai itself. Hence a fire disaster management system seems to be the need of the hour.

House fire is one of the major concerns for designers, builders, and residents of property.

Some of the causes of household fire break-outs are:

- Cooking Equipment: The damages or scratches on the gas cylinders can lead to leakage of gas which can catch fire.
- Heater: Keep portable heaters at least one metre away from anything that could easily catch fire such as furniture, curtains, laundry, clothes
- Electrical Equipment: A switch that is overloaded with double adapter plugs can cause a fire from an overuse of electricity. An extension cord for switch can also be a fire hazard if not used appropriately.
- Faulty Wiring: Homes and buildings with inadequate wiring can cause fires from electrical hazards.
- Flammable liquids: should be careful while handling flammable liquids like petrol, kerosene kept in the houses or garages.

II. RELATED WORK

Various disaster management information systems are using different technologies to be a meaningful in typical disaster management. This section presents a light on different technologies discussed below:

Jovilyn Therese B. Fajardo have described a disaster management system Android application known as MyDisasterDroid that determines the optimum route along different geographical locations that the volunteers and rescuers need to take in order to serve the most number of people and provide maximum coverage of the area in the shortest possible time.[9]

Mr.T.Rajasekaran has described a Forest fire prediction and alert system using big data,this method involves collecting and analyzing the data collected from the wireless sensors embedded in the forest corresponding data is aggregated then methods are applied to predict the forest fire.They have used a machine learning tool called Mahout for clustering and filtering the datasets and predict the output.GSM technology is used for providing the alert message to the people so that proper relocation occurs to a safe place. Signal and Infrared image processing is used to monitor the signals and images of the entire forest for every 30 min and those data will be stored in datasets, by using those data we can be able to predict the forest fire in advance.[11]

Mohammad Sultan Mahmud, Md. Shohidul Islam and Md. Ashiqur Rahman suggested a solution for individual sensors that were used for detecting fire but could not detect the level of fire and notify the emergency response units. They have proposed an intelligent early fire detection system that would not only detect the fire by using integrated sensors but also notify the appropriate authorities including fire department, ambulance services, and local police station. Signals from the embedded detectors e.g., heat, smoke, and flame go through the machine learning algorithms to check the potentiality of the fire as well as broadcast the predicted result to various parties using a GSM modem. The output of this development minimized false alarms, thus making this system more reliable.[10]

Md. Shahinoor Rahman described the role of big data in disaster management; big data from these various sources is getting attention from disaster managers and researchers. Management and analytics of big data can be effective in all phases of disaster management. Even though there are challenges using these big data in an emergency, but big data already proves its effectiveness. As the big data application in disaster management is going to affect many improvements in the capacity level, management level, and research level which will help to get the maximum benefits from this modern opportunity [8].

III. PROPOSED SYSTEM

The proposed system is a cloud based fire detection that collects, organizes, manages several fire hazard databases and helps in the detection of fire hazards.

A. Modules

The system will be available to the users in the form of a web application.

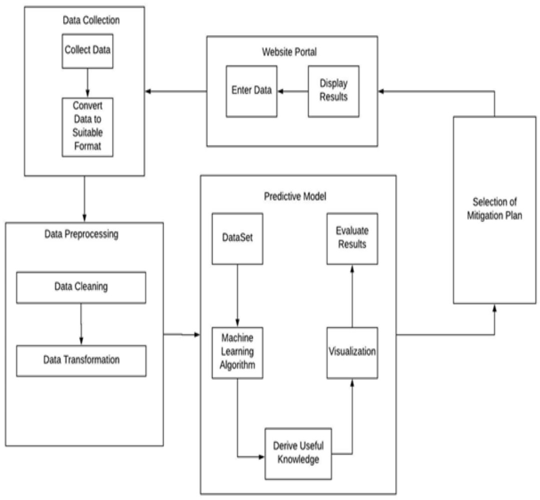

Figure 2: System Architecture Block Diagram

The above figure 2 explains the block level representation of the project. It shows how different modules will be interacting with each other and also the entire flow of the system.

The system consists of the 5 Modules as shown in the Figure 2

1. **Data Collection Module**: In this phase we collect relevant data from different sources. Data are collected from various sources for analysis and event detection. Data sources considered for obtaining fire related data are the fire stations. Collected data are in different formats such as handwritten, sensor data, websites, etc. The corresponding data is aggregated to a csv file.

2. **Data Preprocessing Module**: The data acquired through the data collection process is processed to convert it into a format which is favorable for the

Machine Learning Process. In this module data cleaning and data transformation is performed.

3. **Machine Learning/ Predictive Model Module**: This module uses a machine learning algorithm to implement a Predictive model. The typical workflow for machine learning module consists of many phases:

- Identifying a fire related problem to solve and a metric for measuring results.
- Finding, cleaning, and preparing appropriate data.
- Identifying the best attributes for the collected data for prediction of fire
- Building, evaluating, and tuning models.
- Using models to generate predictions, and provide mitigation steps.

4. **Selection of Mitigation Plan Module**: Based on the results obtained from the Predictive model, this module will decide as to what mitigation plan will be the most suitable one for the given situation.

5. **Website Portal Module**: The final results of our project will be displayed on a website Portal. This application consists of a login module for the fire officer who wants to use it to enter new fire incidents. Mitigation plan for the particular disaster will also be displayed on website for the users.

B. Algorithm

The different types of algorithms that we are going to use in our system are as follows:

1. Multivariate regression [7]

Multivariate regression will be utilized in our system as it belongs to the class of supervised learning. There are various problems addressed by this regression like if the data is collected and some of the variables are defined the relationship between them can be determined. :

a) Selecting the features: Finding the features on which a response variable depends (or not) is one of the most important steps in Multivariate Regression. To make our analysis we have selected the features to be the call time which denotes the time when there is a fire-breakout and the extinguish time when the fire was extinguished.

b) Normalizing the features: The call time is then scaled in the range of (0,24) to represent in the 24 hour format in order to make better analysis.

c) Identifying Hypothesis and Cost function and Minimizing the Cost function

d) Testing the hypothesis: So this algorithm will be used in analyzing the big data, getting best possible outcome for the input and for forecasting and finding out cause and effect relationship between variables and will help to detect outliers.

2. Clustering

Clustering is performed to group a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups. K-means algorithm will be used for clustering in our system. It is used for clustering of call time (Time of the reporting of any fire that has occurred). This is used to find out which time of the day is most prone to fire accidents as derived from the past incidents.

DBScan is another clustering algorithm that is used for visual purposes. It is used to plot regions or zones in the map based on the severity of fire hazards that have taken place in that area. Severity is shown on the map with the help of the color of the zones that are plotted using the clustering algorithm.

C. Working

- Identifying vulnerabilities from collected data

After collecting fire related data, regression and clustering are performed on the data and the severity score of the fire is calculated based on the four parameters- casualty, second call to the fire station, extinguish time and who extinguished fire (firemen or civilians).

- Display on website

The system will be made accessible to the firemen and civilians through the website. The firemen of a particular area will have their login account in the website, through which they can enter the new incidents occurring in that area. The users can know the vulnerabilities of their area through the various patterns displayed in the site.

- Prediction of fire

It is necessary for us to detect the fire as early as possible and it would be better if it is predicted in advance. The system will help them to identify the probability of fire at a particular time in the day and also its severity. The site will also display a map indicating the highly and moderately fire-prone areas through blue and green zones respectively.

D. Applications

It predicts the occurrence of a fire break-out in a region based on the vulnerabilities found after the analysis of the collected data. It helps the people to identify a region that is likely to catch fire. It also provides the people with proper mitigation plan to avoid the fire.

IV. RESULTS

For the dataset the real-time data from fire-station is collected. The data is classified based on the attributes of the dataset.

The dataset includes the following attributes: "Location" of the fire incident ,the "Cause of fire", the "Type of environment" where the fire has occurred (road , residential ,office) , "Number of firemen" sent to extinguish the fire "Number of vehicles" required , "Call time" represents at what time the fire occurred, the "Return time" helps to identify the time required to extinguish the fire, "Second call" depicts the fire was severe and more vehicles were needed to extinguish it ,any "Casualties occurred", and the "Action taken" by civilians or firemen.

This dataset will be updated by the firemen by entering details through a webform, while encountering with new fire incidents.

The website consists of an admin webpage that takes the input from firemen on the incidents noted by them on a daily basis and this will be directed to the csv file. This data can be used for further analysis to get more accurate results. The website also has a contact us page where residents can input there address to check specifically if that region or area is safe from fire or not.

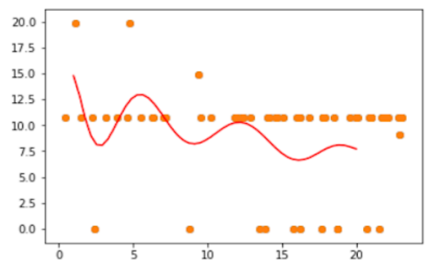

Figure 3: Multivariate regression graph

The figure 3 represents a graph for the multivariate regression which shows the relationship between distributions of severity score (x-axis) and the time of occurrence in a day (y-axis). The graph helps to predict that throughout the day at what time the fire incident occurred was more severe.

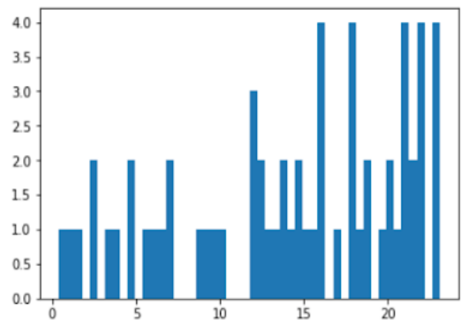

Figure 4: Histogram

The figure 4 represents a histogram which tells the frequency of fire occurrences at different times of the day.

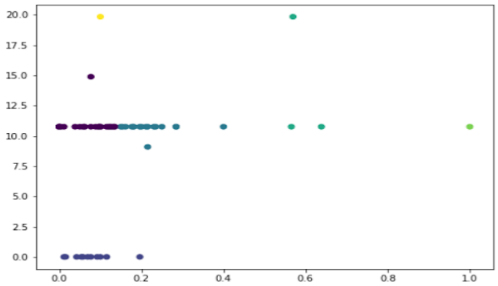

Figure 5: K-means Clustering

The above figure 5 represents clustering is done by using parameters as extinguish time of fire with respective the severity score of it, it tells the time at which most of the incidents occurs.

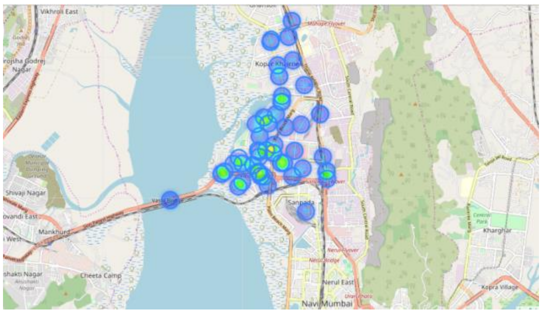

Figure 6: Map showing fire prone areas.

The figure 6 represents a map of Vashi it shows all the fire incidents that were happened in 2018 and the blue spots shows the less severe regions and the green spots depicts the more severe regions in that area.

Figure 7: Map showing areas prone to different levels of severity of fire.

The figure 7 represents a map of Vashi, it shows all the fire incidents that were happened in 2018 and the red markers indicate highly severe fire zones, blue markers indicate moderately severe fire zones and green markers indicate low severe fire zones.

V. CONCLUSION

The objective of this paper is to aware the civilians about the various household fires that can take place in their areas and also provides mitigation plan. Big data from the fire stations are taken and these data help the policy makers and first responders to come up with quick and concrete decision on the number of people affected, type and nature of the damage and where to allocate the resource. Crowdsourcing and cloud computing approach can be used to get required information for emergency management by analyzing big data. Machine learning approach and parallel processing approach saves the valuable processing time during an emergency. With the help of big data the disaster management can be done more efficiently at organization level. Hence, any person of a particular area can easily use the application to get mitigation plans and know the vulnerabilities for a particular disaster, primarily fire hazards.

Future studies will focus on expanding the applicability of this system to not only predicting fire disasters, but also various other disasters through Web or mobile application services, and perform preventive actions for optimal disaster recovery.

VI. REFERENCES

[1] Author: Satish Modh, Edition First, VES Institute of Management, Studies and Research, Mumbai, Publisher: Macmillan Publishers India, dated January 2010

[2] Md. Shahinoor Rahman, Liping Di , Md. Esraz-Ul-Zannat : The Role of Big Data in Disaster Management ,September 23 - 24, 2017

[3]Marion,Unsplash,Online:https://www.eartheclipse.com/natural-disaster/causes-and-effects-of-drought.html,Accessed on 27 September 2018

[4] The International Federation of Red Cross and Red Crescent Societies (IFRC) Online:http://www.ifrc.org/en/what-we-do/disaster-management/about-disaster-management/ Accessed on 18 September 2018

[5]RohtithGandhi,Online:https://towardsdatascience.com/introduction-to-machine-learning-algorithms-linear-regression-14c4e325882a, Published :18 April, 2016

[6] Neeraj Kumar; Alka Agrawal; R. A. Khan : Emergency Information System Architecture for Disaster Management: Metro City Perspective, May-June 2017

[7]Geeks for geeks, Online:https://www.geeksforgeeks.org/multivariate-regression/Published :28 September 2018

[8] Muhammad Arslan, Ana Roxin, Christophe Cruz, Dominique Ginhac. A Review on Applications of Big Data for Disaster Management. The 13th International Conference on Signal Image Technology Internet, Dec 2017, Jaipur, India.

[9] Varsha S. Sonwane,A mobile disaster management system on mobile phones using Google maps,International Journal of Computer Science and Information Technologies, Vol. 5 (5) , 2014

[10] Mohammad Sultan Mahmud, Md. Shohidul Islam and Md. Ashiqur Rahman :Smart Fire Detection System with Early Notifications Using Machine Learning, Vol.16 (2) 2017.

[11] Mr.T.Rajasekaran, J.Sruthi, S.Revathi , N.Raveena :Forest Fire Prediction and Alert System using big data technology, 31 July 2015

Certificates:

DATTA MEGHE COLLEGE OF ENGINEERING

SECTOR-3, AIROLI, NAVI MUMBAI-400708 MAHARASHTRA

www.dmce.ac.in, www.eesadmce.com

सा विद्या या विमुक्तये

CERTIFICATE

This is to certify that Dr. /Mr. /Mrs. /Ms. Purva Gharat has presented a paper titled F-Alert : Fire Disaster Management using Machine learning in 3rd International Conference on Advanced Trends in Engineering (ICATE-2019) organised by Dept. of Electronics Engg, Datta Meghe College of Engineering, Airoli on 4th and 5th April 2019.

Prof. S.R. Jajoo
Organising Secretary

Dr. D. J. Pete
Organising Chair

Dr. S. D. Sawarkar
Principal

DATTA MEGHE COLLEGE OF ENGINEERING

SECTOR-3, AIROLI, NAVI MUMBAI-400708 MAHARASHTRA

www.dmce.ac.in, www.eesadmce.com

सा विद्या या विमुक्तये

CERTIFICATE

This is to certify that Dr. /Mr. /Mrs. /Ms. Elvis D'costa has presented a paper titled F-Alert : Fire Disaster Management Using Machine learning in 3rd International Conference on Advanced Trends in Engineering (ICATE-2019) organised by Dept. of Electronics Engg, Datta Meghe College of Engineering, Airoli on 4th and 5th April 2019.

Prof. S.R. Jajoo
Organising Secretary

Dr. D. J. Pete
Organising Chair

Dr. S. D. Sawarkar
Principal

DATTA MEGHE COLLEGE OF ENGINEERING

सा विद्या या विमुक्तये

SECTOR-3, AIROLI, NAVI MUMBAI-400708 MAHARASHTRA

www.dmce.ac.in, www.eesadmce.com

CERTIFICATE

This is to certify that Dr. /Mr. /Mrs. /Ms. Jincy Jacob has presented a paper titled F-Alert : Fire Disaster Management using Machine learning in 3rd International Conference on Advanced Trends in Engineering (ICATE-2019) organised by Dept. of Electronics Engg, Datta Meghe College of Engineering, Airoli on 4th and 5th April 2019.

Prof. S.R. Jajoo Organising Secretary	Dr. D. J. Pete Organising Chair	Dr. S. D. Sawarkar Principal

DATTA MEGHE COLLEGE OF ENGINEERING

सा विद्या या विमुक्तये

SECTOR-3, AIROLI, NAVI MUMBAI-400708 MAHARASHTRA

www.dmce.ac.in, www.eesadmce.com

CERTIFICATE

This is to certify that Dr. /Mr. /Mrs. /Ms. Kunal Das has presented a paper titled F-Alert : Fire Disaster Management using Machine Learning in 3rd International Conference on Advanced Trends in Engineering (ICATE-2019) organised by Dept. of Electronics Engg, Datta Meghe College of Engineering, Airoli on 4th and 5th April 2019.

Prof. S.R. Jajoo Organising Secretary	Dr. D. J. Pete Organising Chair	Dr. S. D. Sawarkar Principal

ACKNOWLEDGEMENTS

The making of the project Disaster Management involves contribution of many people.

We owe a great thanks to many people who have helped and supported us during the course of our project. We express our gratitude to Dr. S. M. Khot, Principal of Fr. C. Rodrigues Institute of Technology, H.O.D. of IT department Prof. Hari Chavan, the Coordinators of IT department Mrs. Dhanashree Hadsul and Mrs. Kalpana Wani for extending their inevitable and valuable support to us.

We would take this opportunity to thank our guide Mrs. Archana Shirke for guiding and correcting various documents of ours with attention and care. This synopsis could not have been completed without their knowledge and expertise.

We would also thank our Institution and our faculty members without whom this project would have been a distant reality. We also extend our heartfelt thanks to our families and well-wishers.

Yours sincerely,

Purva Gharat

Elvis D'Costa

Jincy Jacob

Kunal Das