B.E. Project Report

On

Connecting People to Avail the Resources During Crisis Through Twitter Using Machine Learning

by

Muskan Singh	**Exam No. B150368586**
Shruti Nikam	**Exam No. B150368591**
Areeba Shaikh	**Exam No. B150368613**
Pooja Shingewad	**Exam No. B150368620**

Under the guidance of

Prof. S.A. Nagtilak

Department of Information Technology

Smt. Kashibai Navale College of Engineering, Pune-41

SAVITRIBAI PHULE PUNE UNIVERSITY

2021-2022

Sinhgad Institutes

CERTIFICATE

This is to certify that the project report entitled

Connecting People to Avail the Resources During Crisis Through Twitter Using Machine Learning

Submitted by

Muskan Singh	**Exam No. B150368586**
Shruti Nikam	**Exam No. B150368591**
Areeba Shaikh	**Exam No. B150368613**
Pooja Shingewad	**Exam No. B150368620**

are bonafide students of this institute and the work has been carried out by them under the supervision of **Prof. (Mrs) S. A. Nagtilak** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of **Bachelor of Engineering** (Information Technology).

(Prof. (Mrs) S. A. Nagtilak)
Project Guide
Department of Information Technology

(Prof. M. L. Bangare)
Head of the Department
Department of Information Technology

(Dr. A. V. Deshpande)
Principal
Smt. Kashibai Navale College of Engineering Pune - 41

Place: Pune
Date: 24/05/2022

ABSTRACT

Social media has become the most used and active way of communication; however, studies that look at social media use in crisis management are in their growing stage. Thus, this study analyzes this growing body of knowledge that researched into social media and crisis management. A review was undertaken between October 2017 to January 2018 which entails sourcing and retrieving materials from an electronic database. The findings of this review confirmed that the growth of social media has transformed the crisis communication landscape because it allows more interactivity. Still, a crisis could also be catalyzed by social media because of its nature. This signifies that the crisis can be created in social media, as well as distributed by social media, respectively. Yet, the potential of social media as a crisis resolution tool is indisputable. It has the potential to prove a statement, to dispel untrue rumors or just to showthe fact. As most of the crisis problems were reported via twitter.

However, most of the problem reported and corresponding responses via twitter were not successfully exchanged between victim's and resource organizations. As a result, most of the tweets were not getting help. Thus, we designed a platform where people can avail the resources of crisis through tweets matching concept using machine learning

Keywords:- Crisis, Machine Learning, Twitter Dataset, SVM, KNN, RF Techniques

ACKNOWLEDGMENTS

We are very thankful to all those who have provided us valuable guidance towards the completion of this Seminar Report on "Connecting People to Avail the Resources During Crisis Through Twitter Using Machine Learning" as part of the syllabus of bachelor's course. We express our sincere gratitude towards the cooperative department who has provided us with valuable assistance and requirements for the system development. We are very grateful and want to express our thanks to **Prof. (Mrs) S.A. Nagtilak** for guiding us in the right manner, correcting our doubts by giving us their time whenever we required, and providing their knowledge and experience in making this project.

We are also thankful to the HOD of our Information Technology department **Prof. Dr. M. L. Bangare** for his moral support and motivation which has encouraged us in making this project. The acknowledgement will be incomplete if we do not thank our Principal **Prof. Dr. A.V. Deshpande**,Vice Principal **Prof. Dr. K. R. Borole,** who gave us his constant support and motivation which has been highly instrumental in making our project.

We hereby wish to take this opportunity to express our gratitude to our teachers and friends and all who have helped toward the completion of our project.

Muskan Singh	B150368586
Shruti Nikam	B150368592
Areeba Shaikh	B150368613
Pooja Shingewad	B150368620

(Information Technology)

ACRONYMS

HTML	HyperText Markup Language
KNN	K-nearest Neighbors
PHP	Hypertext Preprocessor
RAM	Random Access Memory
RF	Random Forest
SM	Social Media
SVM	Support Vector Machine
UML	Unified Modeling Language
XSA	Extensible Markup Language Software Auto update
ML	Machine Learning
XSA	Explainable Sentiment Analysis
CART	Classification and regression trees
GUI	Graphical user interface
DFD	Data flow diagram

LIST OF FIGURES

4.1	Block Diagram	13
5.1	DFD Level 0	16
5.2	DFD Level 1	16
5.3	Use-Case Diagram	17
5.4	Activity Diagram	18
5.5	Sequence Diagram	20
5.6	Class Diagram	21
5.7	Component Diagram	21
5.8	Deployment Diagram	22
7.1	Importing Library	30
7.2	Reading CSV File	30
7.3	Platted Graph	31
7.4	Platted Graph	31
7.5	Getting first 5 records	32
7.6	Cleaning Dataset	32
7.7	Applying SVM Algorithm	33
7.8	Random Forest Implementation	33
7.9	KNN Classifier	34
7.10	Database Connectivity	34
7.11	Home Page	35
7.12	Sign Up Page	35
7.13	Login Page	36
7.14	Dashboard after Login	36
7.15	Change Password Option	37
7.16	Profile Page	37
7.17	Prediction Page	38
9.1	Timeline Chart	47

1.1 LIST OF TABLES

2.1 LiteratureSurvey Table 8

6.2 Training Accuracy Table 28

6.3 Testing Accuracy Table 28

8.5 Test Case Table 30

9.1 Effort Estimation Table 43

9.2 Project Scheduling Table 43

9.3 Module Table 44

INDEX

Abstract **I**

Acknowledgement **II**

ACRONYMS **III**

List of Figures **IV**

1 INTRODUCTION **1**

1.1 Background 1

1.1.1 Aim / Motivation 2

1.1.2 Justification of Problem 2

1.1.3 Need for System 3

1.2 Organization of project report 3

2 LITERATURE SURVEY **5**

2.1 Literature Survey 5

3 PROJECT REQUIREMENT SPECIFICATIONS **9**

3.1 Software Requirements (Platform Choice) 9

3.2 Hardware Requirements 9

3.3 Tools and Technology Used 9

4 SYSTEM ARCHITECTURE **12**

4.1 Existing System Architecture 12

5 HIGH LEVEL SYSTEM DESIGN 16

5.1 Data Flow Diagram 16

5.1.1 Level 0 Data Flow Diagram 16

5.2 UML Diagram 17

5.2.1 Use-cases 17

5.2.2 Activity Diagram 18

5.2.3 Sequence Diagram: 19

5.2.4 Class Diagram 20

5.2.5 Component Diagram 21

5.2.6 Deployment Diagram 22

6 SYSTEM IMPLEMENTATION 23

6.1 Assumptions and Dependencies 23

6.2 Algorithm 23

6.2.1 SVM 23

6.2.2 RANDOM FOREST 27

6.2.3 KNN 27

6.3 Methodology 28

6.4 Overview of Project Modules 29

7 GUI / WORKING MODULES / EXPERIMENTAL RESULTS 30

7.1 Outcomes 30

8 TESTING 39

8.1 Test Plans 39

8.2 Testing Procedure 39

8.3 Test Strategy 40

8.3.1 Unit Testing 40

8.3.2 Integration Testing 40

8.3.3 Performance Testing 41

8.3.4 Regression Testing 41

8.4 Test Cases 41

8.5 Test Result 42

9 PROJECT PLAN **43**

9.1 Project Estimates 43

9.1.1 Effort Estimate Table: 43

9.1.2 Project Description: 43

9.1.3 Estimation of KLOC: 44

9.2 Risk Management 44

9.2.1 Overview of Risk Mitigation, Monitoring, Management 44

9.3 Project Schedule 47

9.3.1 Project task 47

9.3.2 Timeline Chart 47

10 CONCLUSION **48**

10.1 Conclusions 48

10.2 Future Scope 48

10.3 Limitations of Project 48

REFERENCES **49**

ANNEXTURE **50**

A. List of Publications / Paper Published and Certificates 50

B. Plagiarism report 50

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

The primary goal of crisis management is to determine concrete steps that must be taken before (prevention, preparedness), during (response), and after a crisis has occurred (recovery and mitigation). It is beneficial to use data from multiple sources, including the public as witnesses to emergency situations, in order to complete these tasks effectively. Such information will enable emergency operations centers to plan and execute rescue and response operations. A number of research studies have looked at the use of social media as a source of knowledge for effective crisis management in recent years. People use social media (SM) to explain and address various circumstances in which they find themselves, such as crisis. As a result, it is worthwhile to use SM content to aid crisis management, especially by disclosing useful and previously unknown details about crises in real time. Here proposed Social Media Analysis to support Crisis Management Using online Machine Learning which classifies the relevant and irrelevant data Due to the ever increasing reach and influence of the mass media in today's society, they have become an inseparable part of any social situation. Mass media facilitates access to information for policymakers, managers and citizens alike and increases the speed with which new information is gathered, accumulated and distributed, further increasing its role in managing situations such as natural disasters. Mass media can have a negative role in a crisis, increasing its scope or play a positive role and help solve the crisis. Nevertheless, the role of mass media in any crisis cannot be denied. Mass media and its tools play an important role in what we think about, how we think about it, and what, how and why effects our emotions. In the eye of normal people, regardless of their education level, the belief that mass media has unimaginable power is widespread. Most people think that mass media can change philosophical and political opinions, give thoughts a new form and guide all our actions. These show the undeniable role of the mass media in human societies especially in crisis situations in which they can help the victims or instead increase their problems and suffering.

1.2 Aim / Motivation

Aim: The aim of our project is to help people who are suffering due to crisis related complications or due to lockdown by connecting them to people who can provide resources or time in doing good.

Motivation: In covid-19 times, social media platforms are playing a major role and in particular Twitter where people are posting requests, updates, and helpful resources to find the latest information and access the resources faster. As a global platform Twitter has seen a tremendous surge of users posting updates about help supplies and more such situations in the past few weeks. The motivation behind this project has been to gather information in a structured way related to the required resources. It will help people to find correct information on Twitter related to any specific tweet and provide resources not only during covid but also in different crises.

1.3 Justification of Problem

The proposed system uses applied machine learning methods in the field of artificial intelligence to analyze data collected from Twitter. the proposed system examined crisis related tweets.

Social network and sentiment analysis were also conducted to determine the social network of dominant topics and depending on the tweets to found of any one needs help in surrounding cities.

1.4 Need for System

As discussed in above session the problems, if the system is designed the people can easily find availability of Covid-19 bed or vaccination centers or any other helprequired in the surroundings

1.5 ORGANIZATION OF PROJECT REPORT

- Chapter 2: Deals with the Project Related Work i.e., Literature Survey. It synthesizes related research done on this topic. Literature survey includes different research done methods used their algorithms and which area the research lacks along with that it also includes existing system architecture and its detailed explanation and working.
- Chapter 3: This chapter gives an overall view of the model which we are going to use along with the algorithm of the system. The algorithm which weare

  going to be used are SVM, RF and KNN.
- Chapter 4: This chapter gives an overall view of tools and techniques used inthe systemand the information related tothem.
- Chapter 5: Deals with System Design. We need system design for defining the architecture, modules, interfaces, and data for a system to satisfy specified requirements. There are different types of system diagrams in this chapter such as architecture diagrams, class diagrams, activity diagrams, sequencediagram.
- Chapter 6: This part includes the implementation. Putting strategies into place, utilizing resources and executing parts of the system.
- Chapter 7: This chapter contains the detailed information about the system implementation of the project assumptions which are taken, the algorithm which we have used in the project such as SVM, KNN and RF. Also explained the flowchart and methodologies in the chapter
- Chapter 8: This chapter deals with the testing of the project. Testing part includes different test cases, test plans, and the result of thetest
- Chapter 9: It contains the detailed project planning of the project in the form of a timeline chart.
- Conclusion and at last references.

CHAPTER 2

LITERATURE SURVEY

2.1 LITERATURE SURVEY

(1) Douglas Cirqueria and Gultekin Cakir, “Explainable Sentiment Analysis Application for Social Media Crisis Management in Retail”,2020

The proposed study developed an Explainable Sentiment Analysis (XSA) application for Twitter data, and proposes research propositions focused on evaluating such application in a hypothetical crisis management scenario. Particularly, they evaluate, through discussions and a simulated user experiment, fcrfthe XSA support for understanding customer’s needs, as well as if marketing analysts would trust such an application for their decision-making processes

(2) Azzam Mourad , Ali Srour and Mohamad Arafeh, “Critical Impact of Social Networks Infodemic on Defeating Coronavirus COVID-19 Pandemic: Twitter-Based Study and Research Directions”, Transaction on Network and Service Management, IEEE 2020

Authors investigated the COVID-19 infodemic's negative impact on the major efforts to defeat the pandemic through a novel large-scale Twitter-based study, which provided quantitative assessment using real-life experiments re- flecting the actual environments. The empirical analysis of 1 million COVID- 19-related tweets belonging to 288K unique users illustrated the severe impact of misleading people and spreading unreliable information

(3) Umar Ali Bukar and Fatimah Sidi, “Crisis Informatics in the Context of Social Media Crisis Communication: Theoretical Models, Taxonomy, and Open Issues”, IEEE Access 2020

The proposed system aims to review and analyze the relationship of social media-based crisis communication in the context of crisis informatics and its taxonomy and the related crisis communication theoretical models to derive the challenges and limitations. The result of the finding shows that stakeholder interaction is an understudied field, while information reliability and processing for decision-making purposes, the wider application of social media sites.

(4) Jayashree Domala and Vinit Masrani, "Automated Identification of Disaster News for Crisis Management using Machine Learning and Natural Language Processing", International Conference on Electronics and Sustainable Communication System, IEEE 2020

In the proposed system the objective was to automatically scrape news from English news websites and identify disaster relevant news using natural language processing techniques and machine learning concepts, which can further be dynamically displayed on the crisis management websites. The complete model was automated and requires no manual labor at all. The architecture was based on Machine Learning principles that classifies news scraped from top news websites using a spider-scraper into two categories, one being disaster relevant news and other being disaster irrelevant news and eventually displaying the relevant disaster news on the crisis management website.

(5) Tejas Shah, Zhenyu Wen and Divya Pullarkatt, "Use of Social Media Data in Disaster Management: A Survey", AI and IoT technologies in smart cities, MDPI 2020

This survey includes the methodologies for social media data classification and event detection as well as spatial and temporal information extraction. Furthermore, a taxonomy of the research dimensions of social media data management and analysis for disaster management was also proposed, which was then applied to a survey of existing literature and to discuss the core advantages and disadvantages of the various methodologies.

(6) Vedant Dhurve, Krutika Hedaoo, Himanshu Itankar, Jayesh Lanjewar, “Survey on Content Based Disaster Management Using Social Media”, International Journal of Scientific Research Engineering Trends, 2021

Proposed a natural disaster analysis interface that solely makes use of tweets generated by the twitter user during the event of disasters. They observe that their analysis of data from social media provides a viable, economical, uncen- sored and real-time alternative to traditional methods for disaster analysis and the perception of affected populations towards a natural disaster.

(7) Anita Saroj and Sukomal Pal, “Use of social media in crisis management: A survey”, Elsevier 2020

In the proposed system they made an attempt to see and analyze the relationship between emergencies and online social media, especially Twitter, Facebook, and Youtube. Specifically, they look at three important issues. First, try to see the effect of occurrence of emergencies on social media. Second, when there was a sudden surge of posts in social media due to the occurrence, how that deluge of data can be effectively extracted and processed to create situational awareness and minimize the damage due to the disaster. Third, how different social media posts can help different governments and other agencies to get prepared and to take necessary steps to manage emergencies in order to minimize the loss.

(8) Christian Reuter and Amanda Lee Hunges, “Social Media in Crisis Management: An Evaluation and Analysis of Crisis Informatics Research”, Research Article 2018 They evaluate and analyze crisis informatics research by looking at case studies of social media use in emergencies, outlining the types of research found in crisis informatics, and expounding upon the forms of interaction that have been researched. Finally, they summarize the achievements from a human–computer interaction perspective and outline trends and challenges for future search.

Table 2.1 Literature Survey

Sr.no	Paper Title /Publication Year	Techniques used	Problem Found	Algorithms
1	Explainable Sentiment Analysis Application for Social Media Crisis Management in Retail. IEEE 2020	Proposed research focuses on evaluating such applications in crisis management scenarios.	Does not use world knowledge. Totally dependent on exact word matches in the tweet	RF, XG-Boost and MLP
2	Critical Impact of Social Networks Infodemic on Defeating Coronavirus COVID-19 Pandemic: Twitter-Based Study and Research Directions. IEEE 2020	The research provides quantitative assessment using real-life experiment reflecting the actual environment	Does not take into account relation between labeling function	NLP (Natural Language Processing)
3	Crisis Informatics in the Context of Social Media Crisis Communication: Theoretical Models, Taxonomy, and Open Issues 2019	System aims to review and analyze the relationship of social media base crisis inthe context of crisis information using matching tweets.	Does not match problem tweets to aids tweets.	SVM (Support Vector Machine)
4	Automated Identification of Disaster News for Crisis Management using Ma- chine Learning and Natural Language Processing IEEE 2019	Objective was to automatically scrap the news from English news website and identify disaster relevant news	Identification not done properly as the news management was not done properly	Naive Bayes, Linear Classifier and Support Vector Machine.
5	Use of Social Media Data In Disaster Management: A Survey IEEE 2018	The survey includes the methodologies for social media data classification and event detection	The survey not includes every concept of social media and classification not done properly	Linear Classifier
6	On Identifying Disaster-Related Tweets: Matching-based or Learning-based? (2017)	For representing each tweet as an embedding vector, TF-IDF for penalizing high-frequency words, latent semantic indexing for dimension reduction, and logistic regression for classifying tweets into .	Consider only geo-tagged tweets. Only 1-2 % of all tweets are geotagged	Collaborative filtering, naïve bays

Table 2.1 Literature Survey

Sr.no	Paper Title /Publication Year	Techniques used	Problem Found	Algorithms
7	Named Entity Recognition for Code-Mixed Indian Corpus using Meta Embedding IEEE 2017	Uses meta-embedding in the Transformer to encode the sentence and CRF takes the encoded information from the Transformer to predict the tags of Named Entity.	Assumption that corpus is available for related languages. Script needs to be transliterated	Hierarchical clustering with wards algorithm
8	Worldwide COVID-19 outbreak data analysis and prediction. [preprint] Bull World Health Organ. E-pub. 2015	To trail the out the economic, political and health related impact on the people as envisaged through Corona Tracker website	The tracker was not able to take the covid related information and the most of the data was anonymized	SVM (Support Vector Machine)

2.2 Problem Definition

Due to the ever increasing reach and influence of the mass media in today's society, they have become an inseparable part of any social situation. Mass media facilitates access to information for policymakers, managers and citizens alike and increases the speed with which new informationis gathered, accumulated and distributed, further increasing its role in managing situations such as natural disasters. A tweet that informs about the possibility or emergence of a problem during a crisis that requires a treatment or countermeasure. Thus, we need to develop a system which makes use of tweets to help people in obtaining the resources during a crisis.2

CHAPTER 3

SYSTEM SPECIFICATION

3.1 SOFTWARE REQUIREMENTS (PLATFORM CHOICE)

- Tools - Python IDE
- Programming Language - Python, PHP
- Software Version - Python 3.5

3.2 HARDWARE REQUIREMENTS

- Processor - Pentium IV/Intel I3 core
- Speed - 1.1 GHz
- RAM - 512 MB (min)
- Hard Disk - 20GB
- Keyboard - Standard Keyboard
- Mouse - Two or Three Button Mouse
- Monitor - LEDMonitor

3.3 TOOLS AND TECHNOLOGY USED

Python

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace.

Python is a programming lan- guage that includes features of C and Java. It provides the style of writing an elegant code like C, and for object-oriented programming, it offers classes and objects like Java.

- Python was developed in the late eighties, i.e., late 1980's by Guido van Rossum at the National Research Institute for Mathematics and Computer Sci- ence in the Netherlands as a successor of ABC language capable of exception handling and interfacing.
- Python is derived from programming languages such as ABC, Modula 3, small talk, Algol-68. Van Rossum picked the name Python for the new language from a TV show, Monty Python's Flying Circus.
- Python page is a file with a .py extension that contains could be the combina- tion of HTML Tags and Python scripts.
- In December 1989 the creator developed the 1st python interpreter as a hobby and then on 16 October 2000, Python 2.0 was released with many new features. On 3rd December 2008, Python 3.0 was released with more testing and includesnew features.
- Python is an open source scripting language., which means that anyone can download it freely from www.python.org and use it to develop programs.

Features of Python

(1) Easy to Learn and Use.

(2) Expressive Language.

(3) Interpreted Language.

(4) Cross-platform Language.

(5) Free and Open Source.

(6) Object-Oriented Language.

(7) Large Standard Library

CHAPTER 4

SYSTEM ARCHITECTURE

4.1 EXISTING SYSTEM ARCHITECTURE

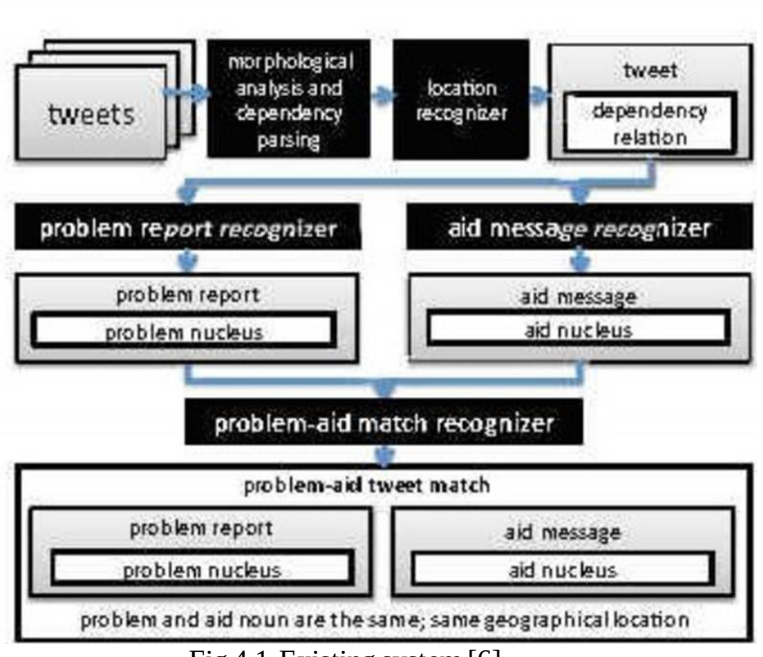


Fig 4.1 Existing system [6]

Machine learning based systems to recognize problem reports, aid messages and problem-aid tweet matches. Illustrates the whole system. First, location names in tweets are identified by matching tweets against our location dictionary. Then, each tweet is paired with each dependency relation in the tweet, which is a candidate of problem/aid nuclei and given to the problem report and aid message recognizers. A tweet-nucleus-candidate pair judged as problem report is combined with another tweet-nucleus-candidate pair recognized as an aid message if the two nuclei share the same noun and the tweets share the same location name, and given to the problem-aid match recognizer [6].

System Overview

- **Problem report:** A tweet that informs about the possibility or emergence of a problem that requires a treatment or countermeasure.

- **Aid message:** A tweet that informs about situations or actions that can be a remedy or solution for a problem, or informs that the problem is solved or is about to be solved.

- **Problem-aid tweet match:** A tweet pair is a problem-aid tweet match if the aid message informs how to overcome the problem, if the aid message informs about the settlement of the problem, or if the aid message provides information which contributesto the settlement of the problem

4.2 PROPOSED SYSTEM ARCHITECTURE

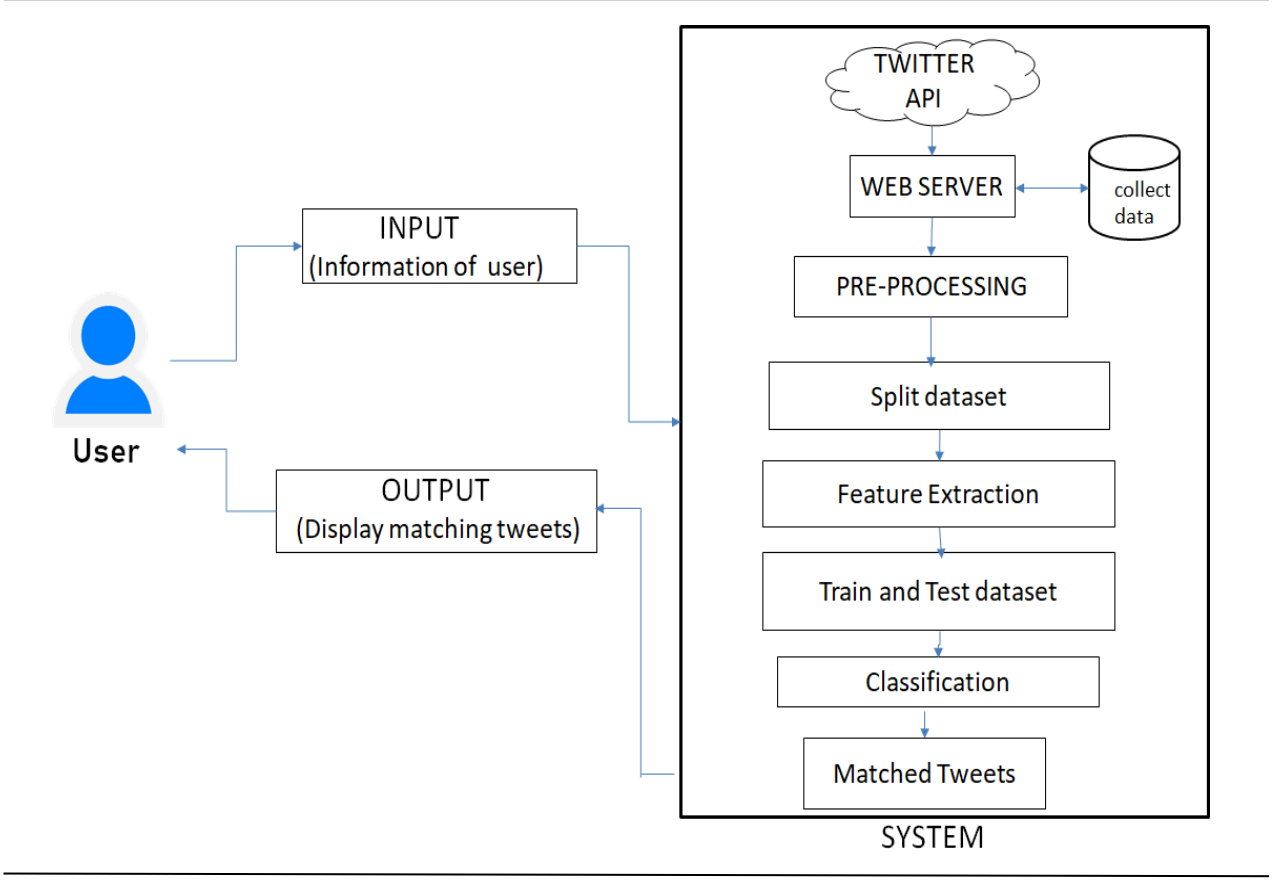


Figure 4.2 Block Diagram

The proposed system is web-based application build using php and css as front end and python for backend. The connectivity is done using MySQL database. This proposed system presents a comparison of machine learning (ML) algorithms: Support machine vector(SVM), Random Forest(RF), K Nearest Neighbour (KNN) search. For the implementation of the ML algorithms, the dataset was partitioned into the training set and testing set.The dataset is passed in pre-processing state where unwanted data or null values are removed. Later on, in next step the features are been extracted using various machine learning techniques such as support vector machine, KNN and Random Forest techniques. A comparison between all the 3 algorithms will be made. The algorithm that gives the best results will be supplied as a model to the website. The website will be made from a python framework, called flask. And it will host the database on Xampp, Python and its libraries.

The process of the proposed system is as follows,

1. First the user registered in the website.
2. After registration, user login using email-id and password .
3. Then the home page of the website opens, in which there is Profile tab and the Resources tab.
4. In the profile tab user information is provided in which name ,id, email , address is given.
5. In the Resources tab users can find the resources they needed during the crisis. User have to give an input example :# Bed needed , #Oxygen etc.
6. After giving an input the system will find the resources an if resources are available , then name contact number and the address of the person who is providing the resources is shown.
7. Along withthat email containing resources detail is send to the user through the system.

4.3 Gap Analysis

During this covid-19 pandemic, there were development of so many technologies and techniques to minimize the spread of covid-19. Some of this technology mention in this were the City Digital Twin, Policies Tracker, Fuzzy Logical Approach, as this technology tried to minimize the spread of covid-19 but it was not fully supported toward the minimization of covid-19 Effect.The digital Twin help us to minimize the spread of covid-19 but it not supports fully due to minimum storage of data; Fuzzy Logical Approach is developed to system to help the safety of health-related issues of the patient's condition according to the changes of environment, but it only supported for a particular symptom such as Breathing could and Policies Tracker it helps us to understand the policies taken by different government but it cannot fully filter the policies taken by the government .

CHAPTER 5

HIGH LEVEL SYSTEM DESIGN

5.2 DATA FLOW DIAGRAM

5.2.2 Level 0 Data Flow Diagram

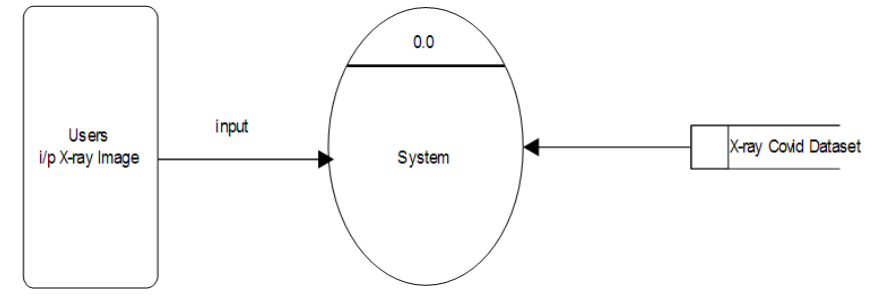


Figure 5.1: DFD Level 0

5.2.2.1 Level 0 Data Flow Diagram

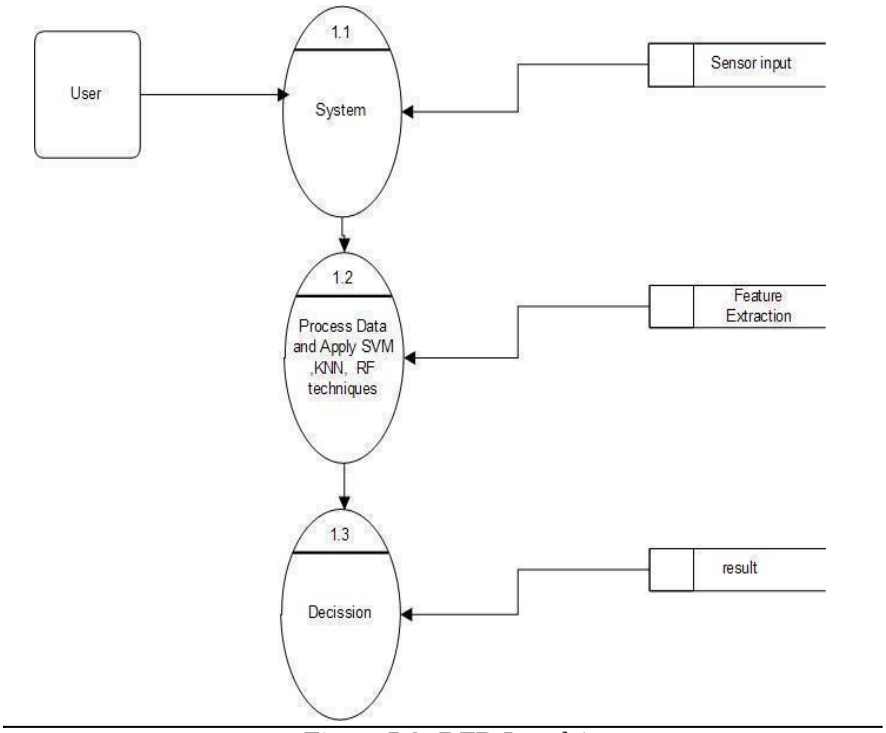


Figure 5.2: DFD Level 1

5.3 UML DIAGRAM

5.3.2 Use-cases

A use case diagram is the primary form of system/software requirements for a new software program underdeveloped. Use cases specify the expected behavior, and not the exact method of making it happen. Use cases once specified can be denoted both textual and visual representation i.e., use case diagram. A key concept of use case modeling is that it helps us design a system from the end user's perspective. It is an effective technique for communicating system behavior in the user's terms by specifying all externally visible system behavior. It only summarizes some of the relationships between use cases, actors, and systems. In the above use case diagram users first have to register through his/or her information. Then login using email- id and password during this process authentication takes place. After login user has to set the profile, the profile includes all the information such as contact number, mail-id, address and the resources they need or the resources that they can provide during the crisis. After that, through the check result tweet matches and the result is displayed.

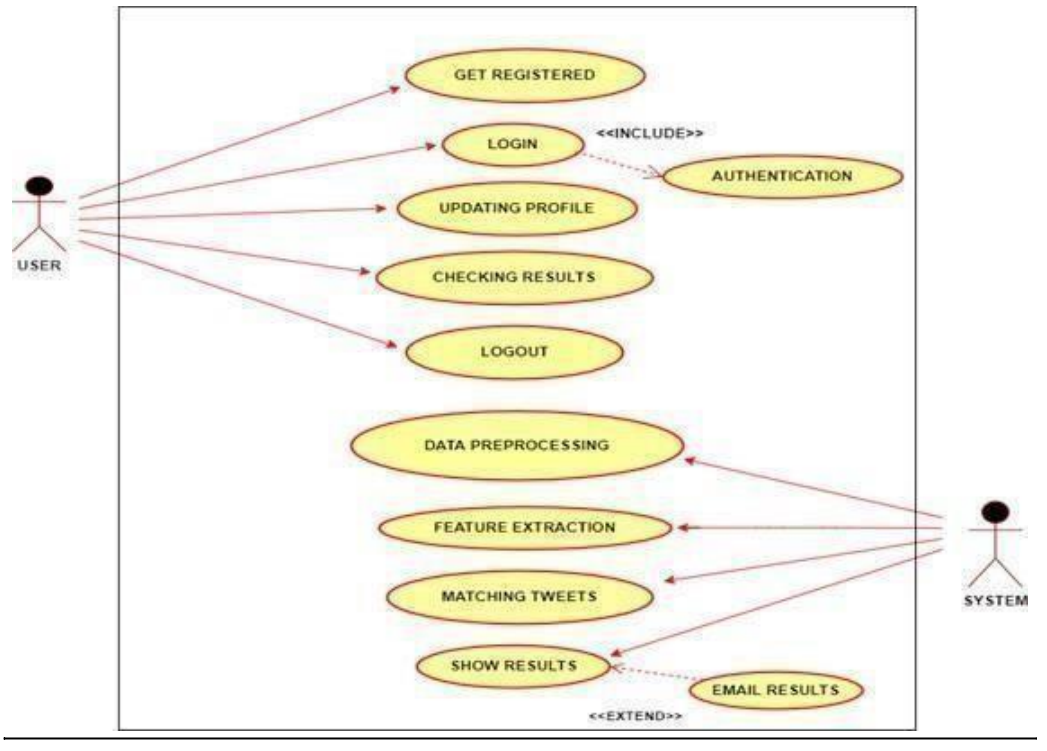


Figure 5.3: Use Case Diagram

5.3.3 Activity Diagram:

An activity diagram is a behavioral diagram i.e., it depicts the behavior of a system. An activity diagram portrays the control flow from a start point to a finish point showing the various decision paths that exist while the activity is being executed. We can depict both sequential processing and concurrent processing of activities using an activity diagram.In the activity diagram the circle represents the start. Through twitter APIs geolocated tweets are taken then the geolocated tweets go under pre-processing and the pre-process tweets are then future taken for feature extraction. And the tweets which are not processed again go through the pre-processing part. After that the classification of tweets takes place, learning algorithms are applied and the tweets are divided into two parts that can help tweets and need help tweets and if the tweets matchthe result is displayed as output. And if the tweets are not matched again it goes under theprocess

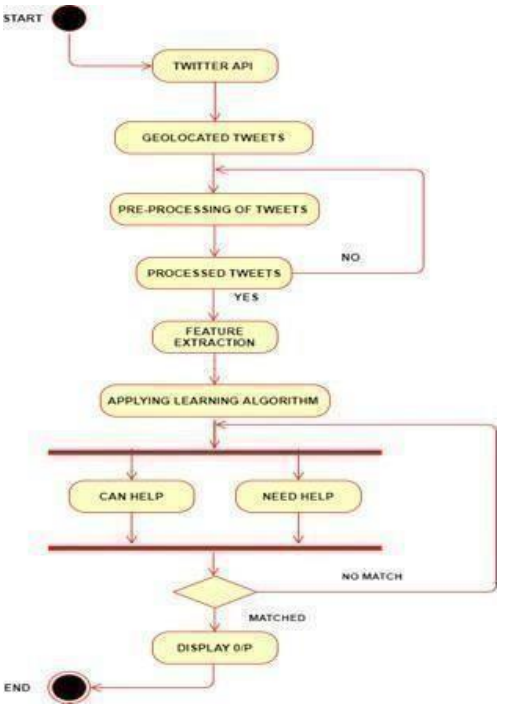


Figure 5.4: Activity Diagram

5.3.4 Sequence Diagram:

A sequence diagram or system sequence diagram shows object interactions arranged in time sequence in the field of software engineering. It depicts the objects involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. Sequence diagrams are typically associated with use case realizations in the logical view of the system under development. Sequence diagrams are sometimes called event diagrams or event scenarios. A sequence diagram shows, as parallel vertical lines (lifelines), different processes or objects that live simultaneously, and, as horizontal arrows, the messages exchanged between them, in the order in which they occur. Messages, written with horizontal arrows with the message name written above them, display interaction. Solid arrowheads represent synchronous calls, open arrowheads represent asynchronous messages, and dashed lines represent reply messages. If a caller sends a synchronous message, it must wait until the message is done, such as invoking a sub

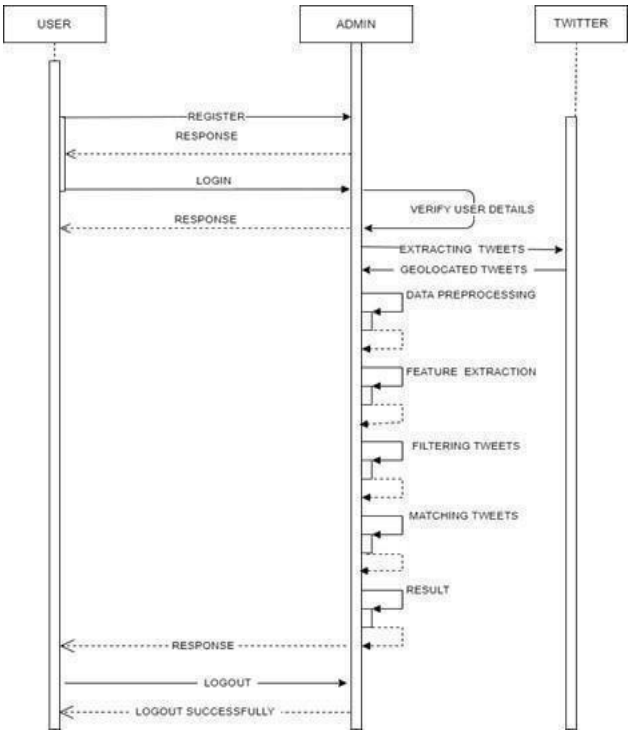


Figure 5.5: Sequence Diagram

5.3.5 Class Diagram:

Class diagram describes the attributes and operations of a class and also the constraints imposed on the system. The class diagrams are widely used in the modeling of object-oriented systems because they are the only UML diagrams, which can be mapped directly with object-oriented languages. Class diagram shows a collection of classes, interfaces, associations, collaborations, and constraints. It is also known as a structural diagram. The purpose of the class diagram is to model the static view of an application. Class diagrams are the only diagrams which can be directly mapped with object-oriented languages and thus widely used at the time of construction. UML diagrams like activity diagram, sequence diagrams can only give the sequence flow of the application, however the class diagram is a bit different. It is the most popular UML diagram in the coder community.

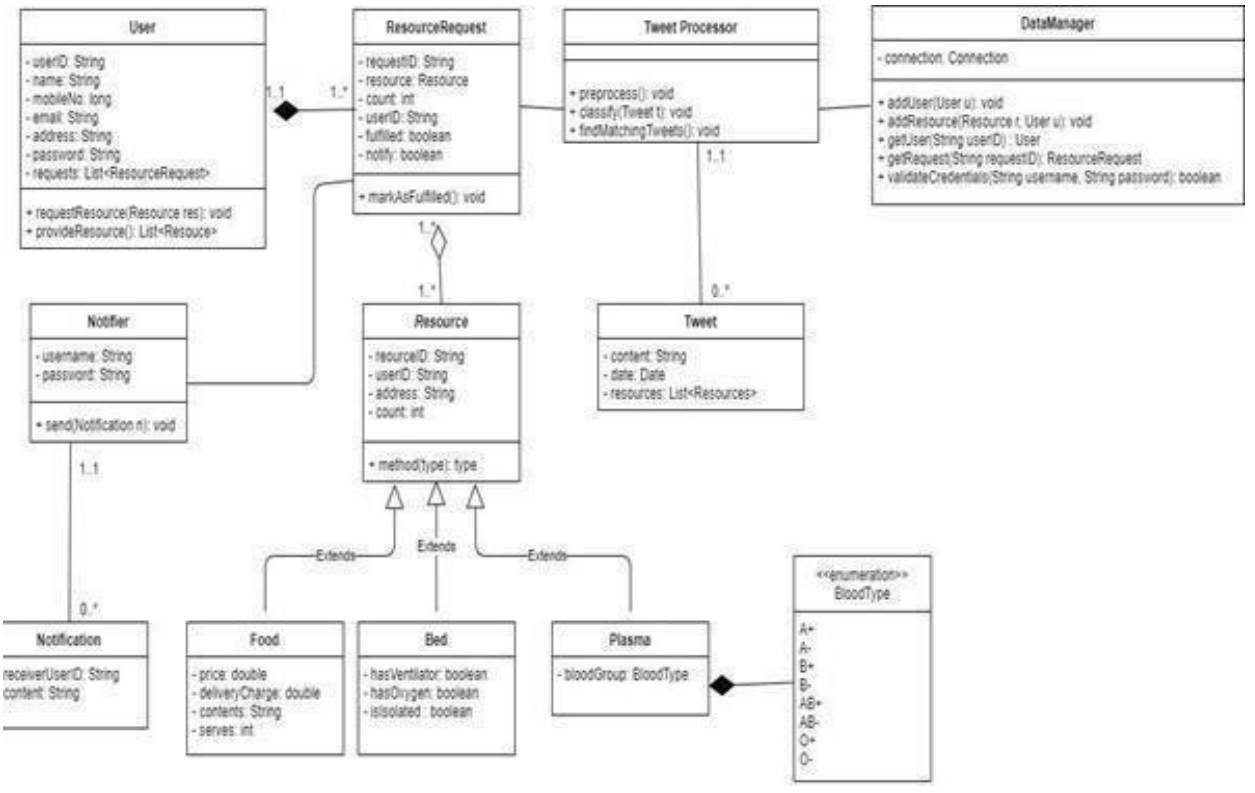


Figure 5.6: Class Diagram

\

5.3.6 Component Diagram:

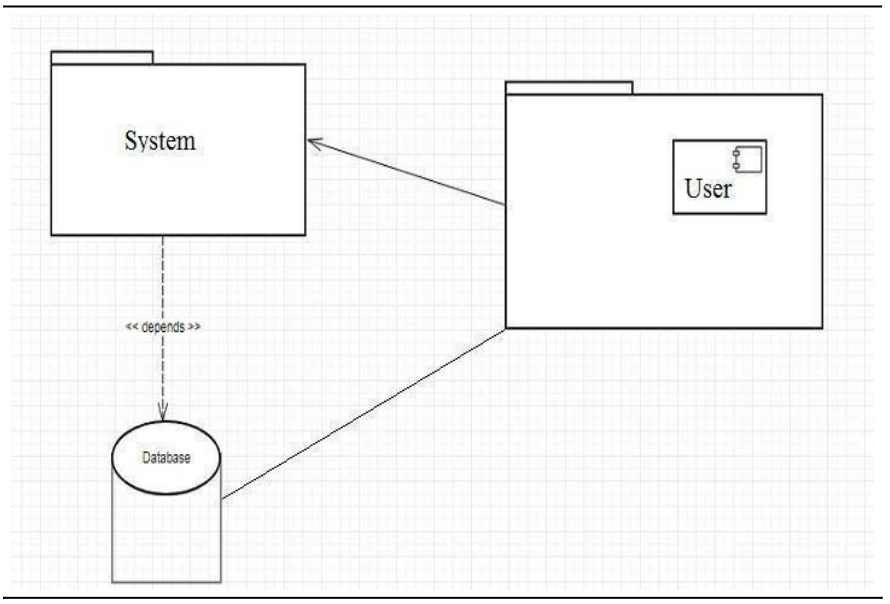


Figure 5.7: Component Diagram

5.3.7 Deployment Diagram:

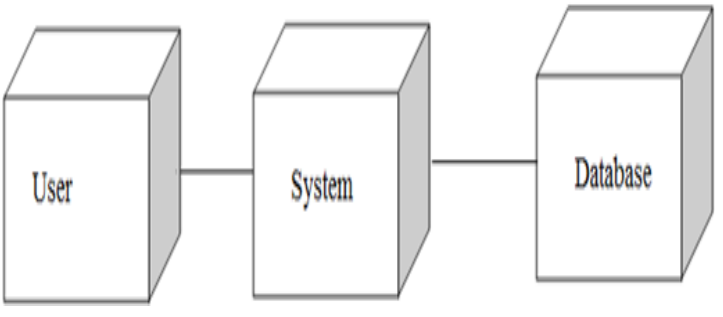


Figure 5.8: Deployment Diagram

CHAPTER 6

SYSTEM IMPLEMENTATION

6.1 ASSUMPTIONS AND DEPENDENCIES

- Assumptions:

  (1) The product must have an interface which is simple enough to understand.

  (2) All the software such as python, mysql, php are installed and running on the computers

- Dependencies:

  (1) All necessary software is available for implementing and use of the system.

  (2) The proposed system would be designed, developed and implemented based on the software requirements specifications document.

  (3) End users should have basic knowledge of computer and we also assure that the users will be given software training documentation and reference material.

  (4) Well Trained dataset

6.2 ALGORITHM

6.2.1 SVM

Support Vector Machine also known as svm is a supervised machine learning algorithm. Svm is the most popular classification technique. Svm creates a hyperplane that separates two classes. It can create a hyperplane or set of hyperplanes in high dimensional space. This hyperplane can be used for classification or regression also. Svm differentiates instances in specific classes and can also classify the entities which are not supported by data. Separation is done by a hyperplane that performs the separation to the closest training point of any class.

Algorithm-

- Select the hyper plane which divides the class better.
- To find the better hyper plane you have to calculate the distance between the planes and the data which is called Margin.
- If the distance between the classes is low then the chance of miss conception is high and vice versa.
- Select the class which has the highest margin. Margin = distance to positive point + Distance to negative point
- Examples of SVM boundaries Selecting best hyperplane for our classification. We will show data from 2 classes. The classes are represented by triangles and circles .

Case 1:

Consider the case in Fig 1, with data from 2 different classes. Now, we wish to find the best hyperplane which can separate the two classes. Please check Fig 1. On the right to find which hyperplane best suits this use case. In SVM, we try to maximize the distance between hyperplanes nearest data points. This is known as margin. Since the 1st decision boundary is maximizing the distance between classes on left and right. So, ourmaximum margin hyperplane will be "1st ".

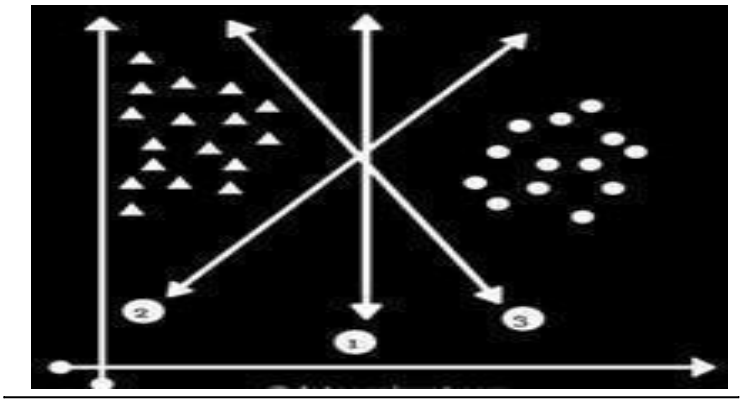


Figure 6.1 Case1

Case 2:

Consider the figure 4, we will learn about outliers in SVM. We wish to find the best hyperplane which can separate the two classes. Data is not evenly distributed on left and right. Some of them are on the right too. In the real world, you may find few valuesthat correspond to extreme cases i.e., exceptions.

These are outliers. While selecting hyperplanes, SVM will automatically ignore these and select best-performing hyperplane.1st 2nd decision boundaries are separating classes but 1st decision boundary shows maximum margin in between boundary and support vectors.

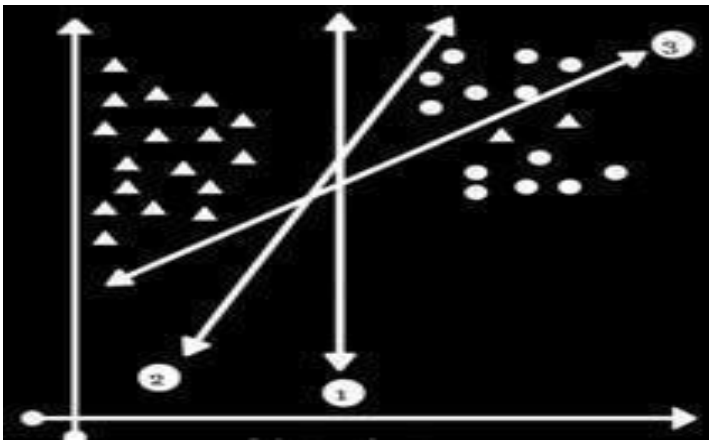


Figure 6.2. Case 2

Case 3:

We will learn about non-linear classifiers. Please check the figure 5 on right. It' showing that data can't be separated by any straight line, i.e, data is not linearly separable. SVM possess the option of using Non-Linear classifier. We can use differenttypes of kernels like Radial Basis Function Kernel, Polynomial kernel etc. We have shown a decision boundary separating both the classes. This decision boundary resembles a parabola

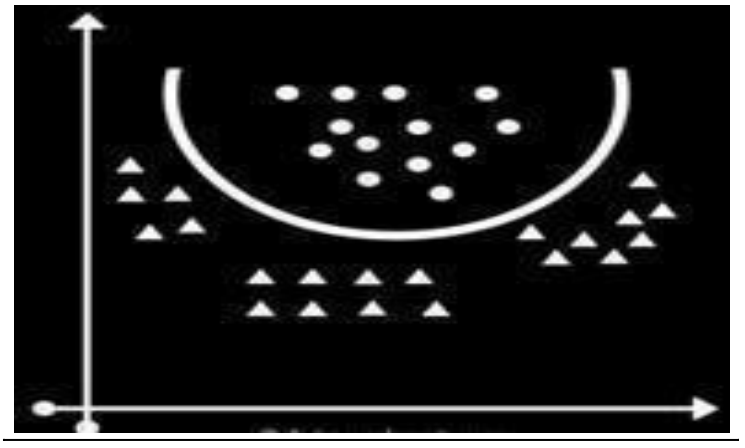


Figure 6.3. Case 3

6.2.2 RANDOM FOREST

It is a type of ensemble learning method and also used for classification and regression tasks. This method can easily handle large datasets. Random Forest was developed by Leo Bremen. Random Forest Improves Performance of Decision Trees by Reducing variance. It operates by constructing a multitude of decision trees at training time and outputs the class that is the mode of the classes or classification or mean prediction (regression) of the individual trees.In the RF method, trees are created by CART (classification and regression trees) algorithms and boot bagging combination method. The data set is divided into training and test data. From the training data set, samples are selected as bootstrap (resampled and sampled) tech- nique, which will form trees (in a bag) and data that will not build trees (out of the bag). 1/3 of the training set is divided into data that will not form trees, and 2/3 of them will be data that will build trees.

Algorithm-

- The first step is to select the R features fromthetotal features m where R<<M.
- Among the R features, the node uses the best split point.
- Split the node into sub nodes using the best split.
- Repeat a toc steps until the number of nodes has beenreached.
- Built forest by repeating steps a to d for a number of times to create n-number oftrees.

6.2.3 KNN

KNN is also a supervised machine learning algorithm. KNN helps to solve both the classification and regression problems. KNN is a lazy prediction technique.KNN assumes that similar things are near to each other. Many times data points which are similar are very near to each other.KNN helps to group new work based on similarity measure.KNN algorithm records all the records and classifies them according to their similarity measure. For finding the distance between the points uses a tree-like structure. To make a prediction for a new data point, the algorithm finds the closest data points in the training data set to its nearest neighbors. Here K= Number of nearby neighbors, it's always a positive integer. The Neighbor's value is chosen from the set of classes. Closeness is mainly defined in terms of Euclidean distance. The Euclidean distance between two points P and Q i.e. P (p1,p2, . Pn) and Q (q1, q2,..qn) is defined by the following equation:-

Algorithm-

- Take a sample dataset of columns and rows.
- Take a test dataset of attributes and rows.
- Find the Euclidean distance by the help of formula.
- Then, Decide a random value of K. is the no. of nearest neighbors.
- Then with the help of these minimum distance and Euclidean distance find out the nth column of each.
- Find out the same output values.

Table 6.2 Testing Accuracy

ALGORITHMS	PREDICTION(0/1)	PRECISION	RECALL	F1-SCORE	ACCURACY
SVM	0	0.93	0.74	0.83	0.79
	1	0.63	0.89	0.74	
KNN	0	0.08	0.64	0.71	0.65
	1	0.48	0.68	0.56	
RF	0	0.95	0.71	0.81	0.76
	1	0.54	0.91	0.68	

Table 6.3 Training Accuracy

ALGORITHMS	PREDICTION (0/1)	PRECISION	RECALL	F1-SCORE	ACCURACY
SVM	0	0.99	0.92	0.95	0.94
	1	0.88	0.98	0.93	
KNN	0	0.87	0.74	0.8	0.75
	1	0.58	0.77	0.66	
RF	0	1	0.82	0.9	0.87
	1	0.7	0.99	0.82	

6.3 METHODOLOGY

This is the most important phase which includes model building for matching of tweets.

Procedure of Proposed Methodology-

Step1: Import required libraries, Import twitter dataset.

Step2: Pre-process data to remove missing data.

Step3: Performa percentage split of 80% to divide the dataset as Training set and 20% to Test set.

Step4: Select the machine learning algorithm i.e. KNN, SVM and Random Forest algorithm.

Step5: Build the classifier model for the mentioned machine learning algorithm based on training set.

Step6: Test the Classifier model for the mentioned machine learning algorithm based on the test set.

Step7: Perform Comparison Evaluation of the experimental performance results obtained for each classifier.

Step8: After analyzing based on various measures conclude the best performing algorithm.

6.4 OVERVIEW OF PROJECT MODULES

The proposed system undergoes some modules such as :-

[1] Preprocessing

[2] Feature Extraction

[3] Classification

- Data preprocessing: - It is a technique used in data mining that involves transforming rawdata into an understandable format. The data is cleansed through processes such as filling in missing values, smoothing the noisy data, or re-solving the inconsistencies in the data. As it contains some missing value, the dataset is cleaned, and decimal values are converted into proper float values

- Data splitting: - The new dataset is split into two, training set and testing set. The splitting is done in an 80-20 ratio. 80% of the dataset is taken as the Training Set which is used to train the model. The remaining 20% becomes the Test Set which is used to test the model, to analyze its accuracy. The testing set is never used for training, which could otherwise lead to overfitting the mode

- Feature Selection: - The data features that used to train machine learning models have a huge influence on the performance of the model. Irrelevant or partially relevant features can negatively impact model performance.

- Classification: - The model is trained by fitting the training set to the classifier model. The classifier model upon testing, classifies the air quality into good or bad. The classifications are fairly close to the testingset.

CHAPTER 7

GUI / WORKING MODULES / EXPERIMENTAL RESULTS

7.1 OUTCOMES

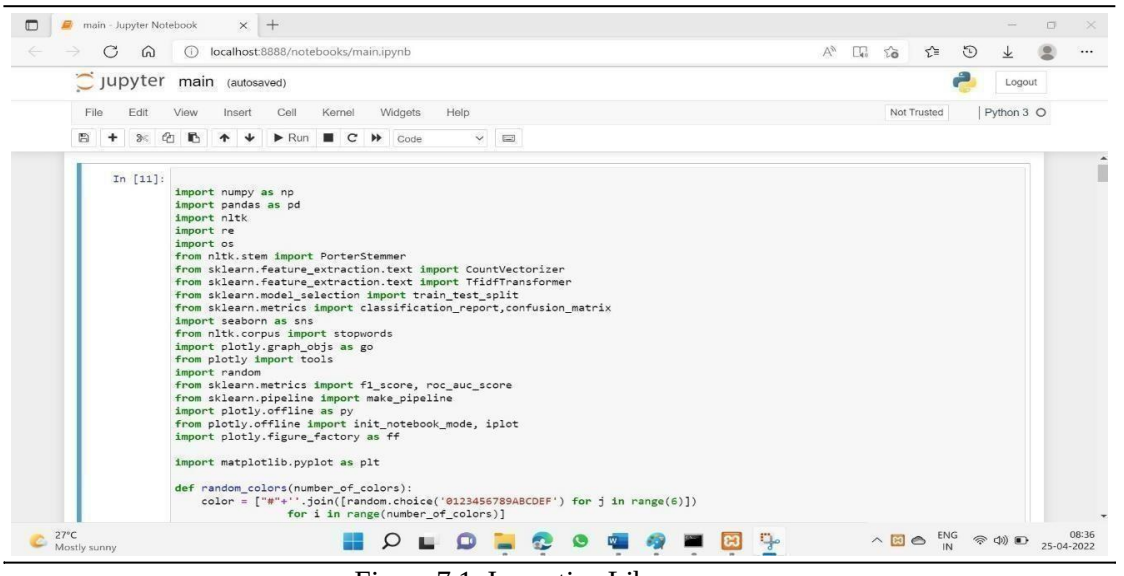


Figure 7.1. Importing Library.

Importing all the required libraries for the project (numpy, pandas,nltk, sklearn, seaborn, plotly.graph_objs, random, plotly.offline, plotly.figure_factory, matplotlib.pyplot,re,os)

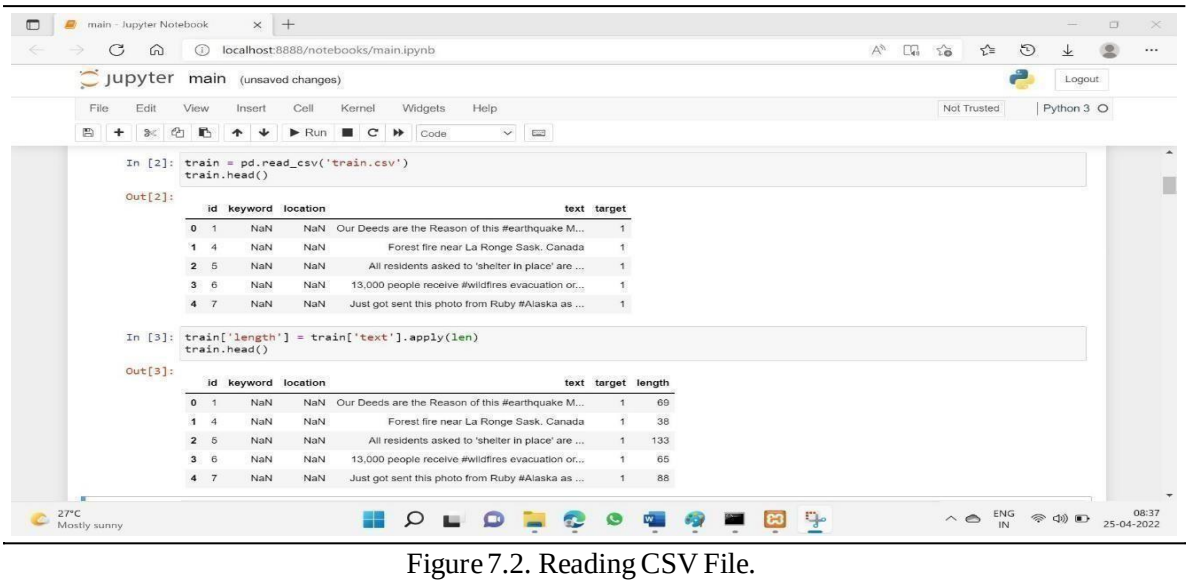


Figure 7.2. Reading CSV File.

Reading the CSV by using the pd.read_csv function and printing the id, keyword, location, text, target of the tweet. Calculate the length of each text and print the id, keyword, location,text, target of the tweet.

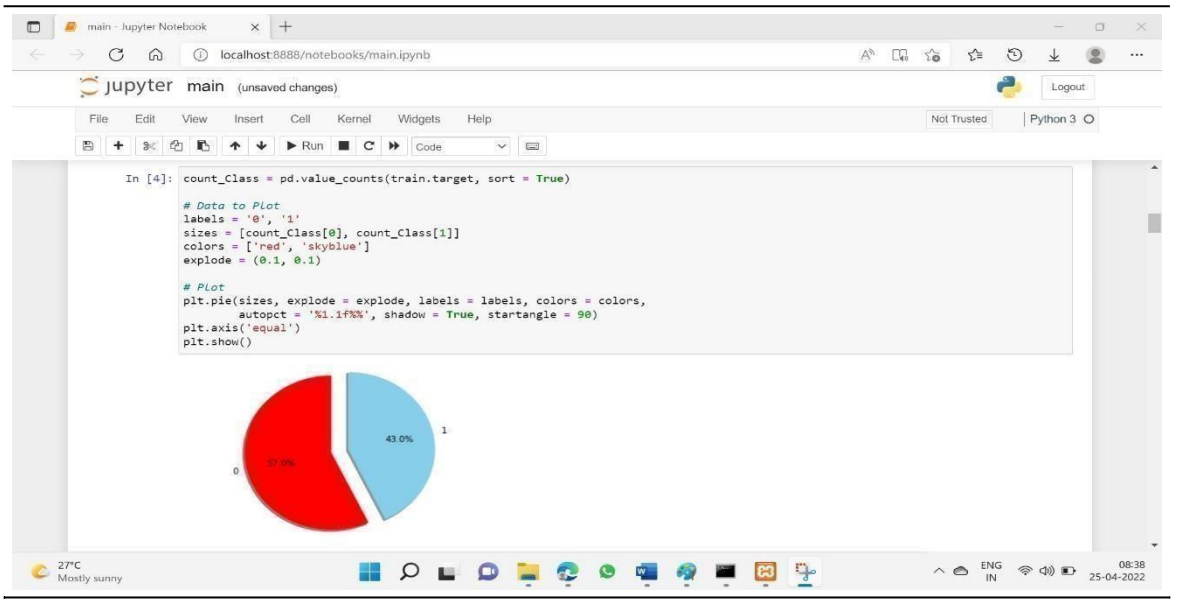


Figure 7.3. Plotted Graph.

Data visualization: Depicting the target 0/1 in the form of a pie chart (43.0 % target 1, 57.0 % target 0)

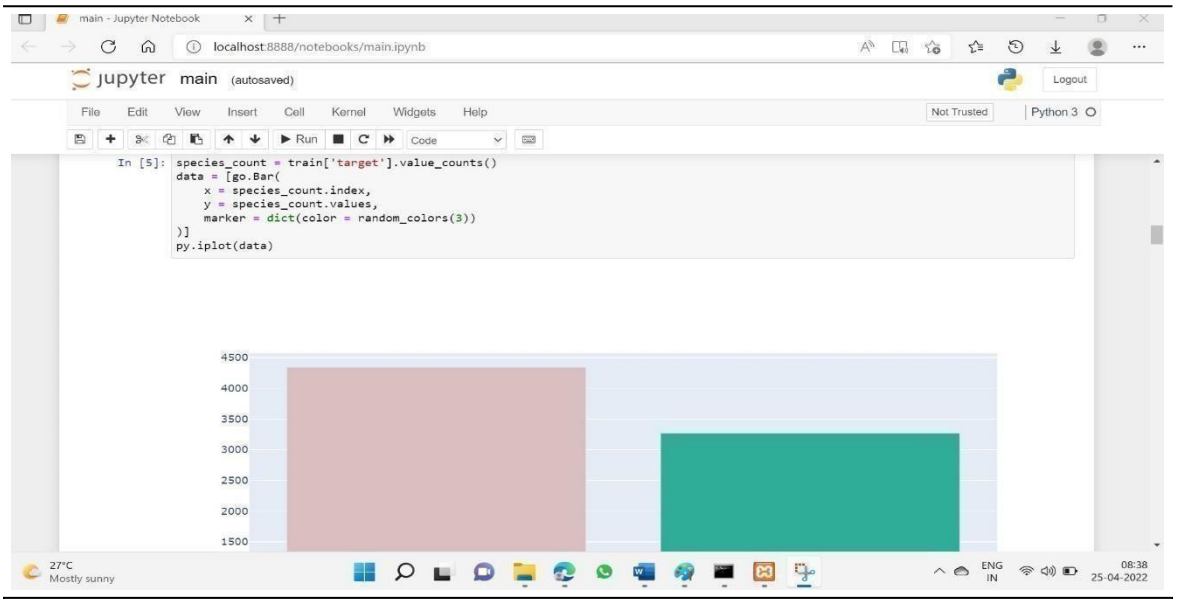


Figure 7.4. Plotted Graph.

Data visualization: Depicting the target 0/1 in the form of bar graph

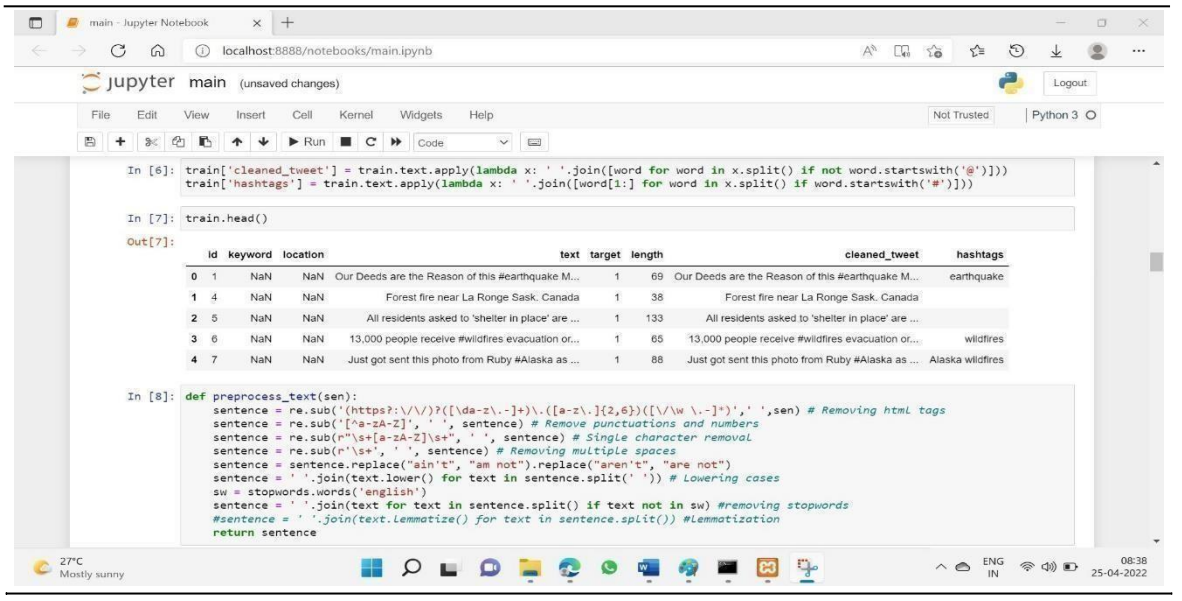


Figure 7.5. Getting first 5 records.

Removing symbolic data from the tweets and displaying the word that has been hashtagged and also the cleaned tweet. Then by preprocessing the entire csv remove http symbolic data fromthe tweet.

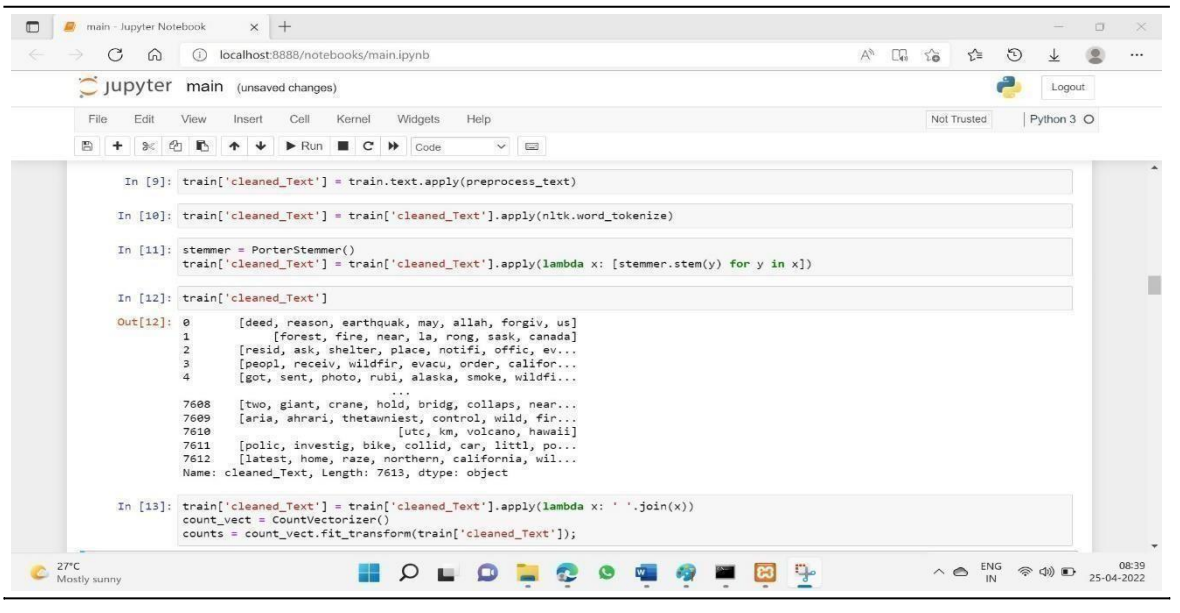


Figure 7.6. Cleaning Dataset.

Store the preprocessed text into "cleaned_Text". Convert each word into token and apply PorterStemmer and print the output

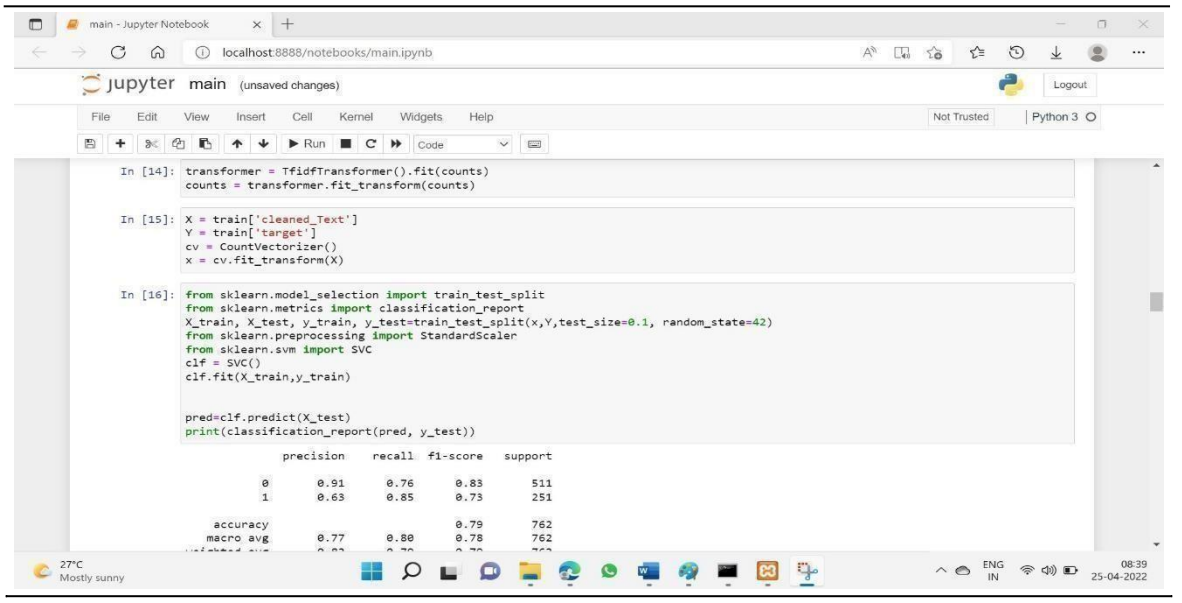


Figure 7.7. Applying SVM Algorithm.

Divide the dataset into 2 parts for training, testing by using the function train_test_split. Fit the training dataset and the trained dataset is tested and lastly the classification repriotis printed(precision, recall, f1-score, support)

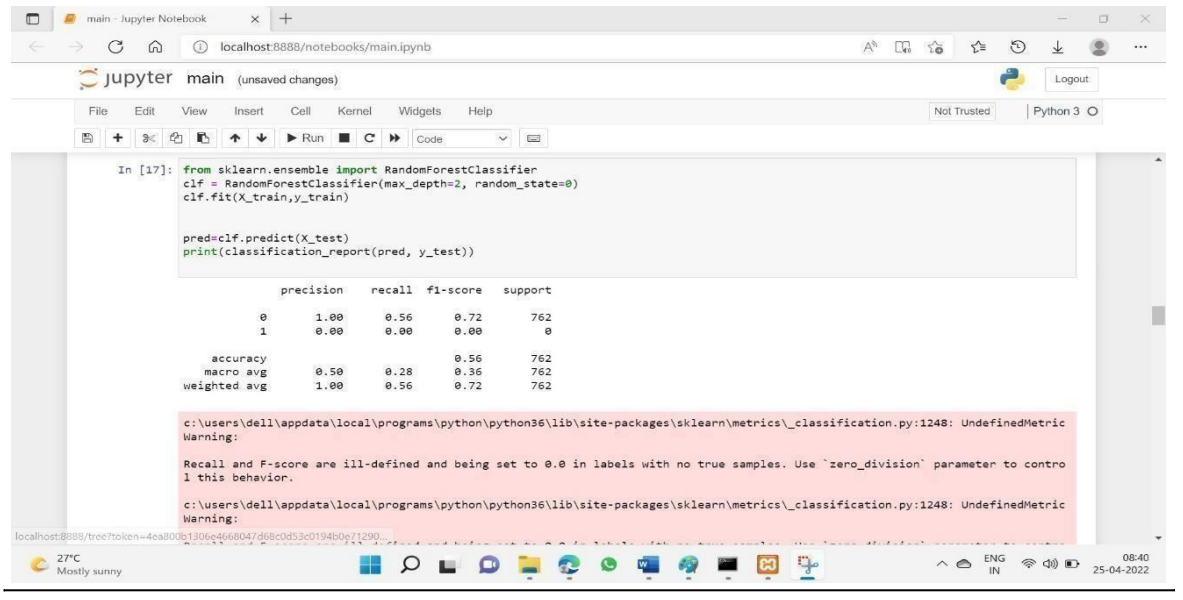


Figure 7.8. Random Forest Implementation.

Import the Random Forest Classifier for Random Forest and fit the training data, test the testing data and calculate and print the parameters (precision, recall, f1-score, support)

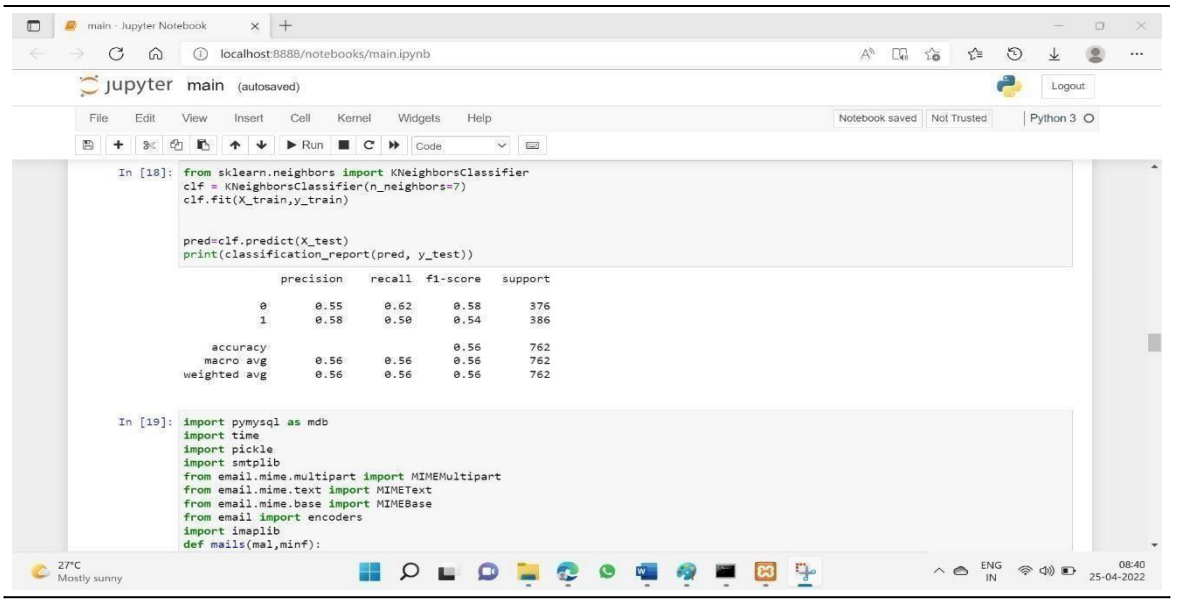


Figure 7.9. KNN Classifier.

Import the K-Neighbors Classifier for KNN classification and fit the training data,test the testing data and calculate and print the parameters (precision, recall, f1- score, support)

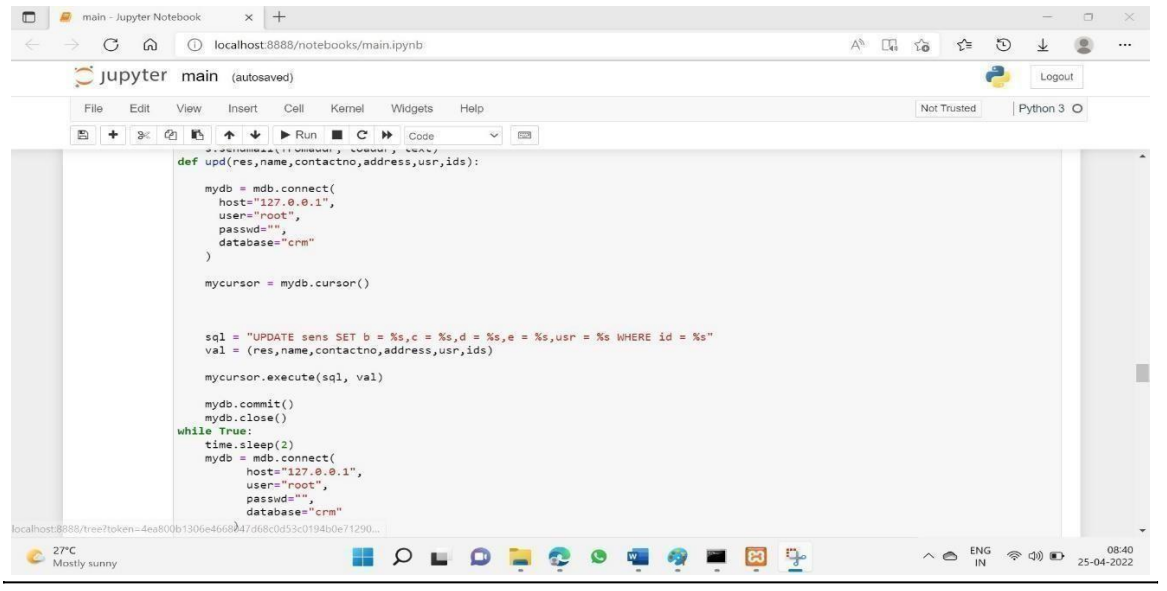


Figure 7.10. Database Connectivity.

Connecting the frontend with the MySQL database

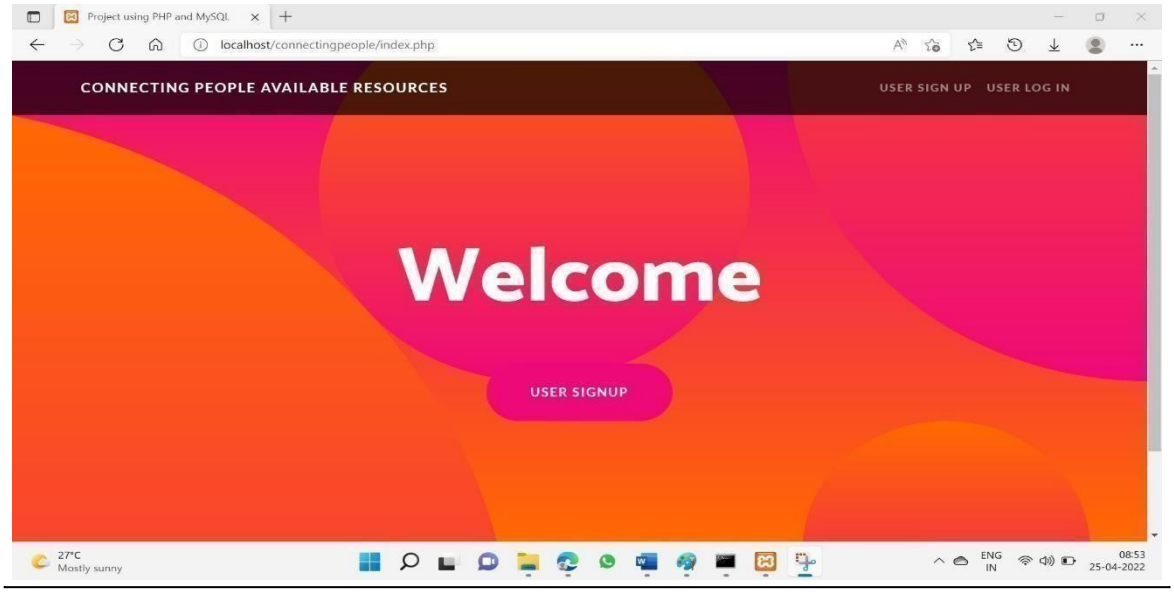


Figure 7.11. Home Page.

This is the Homepage of the app where a user can click on "User Signup" button for creating an account

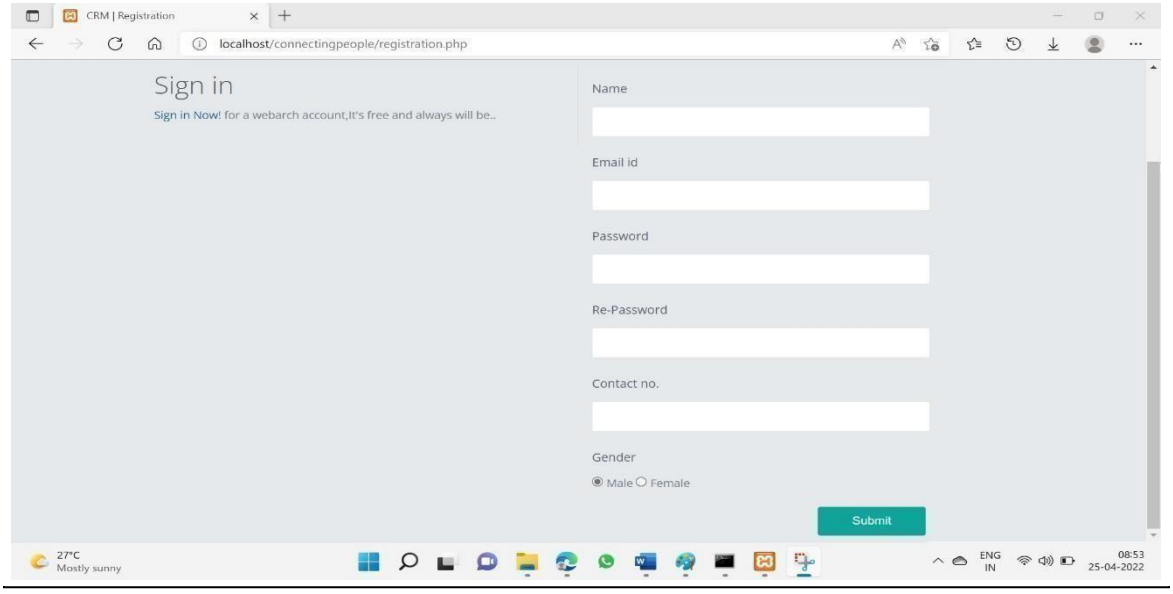


Figure 7.12. Sign Up Page.

This is the registration page of the app, where a user needs to fill in the information like Name, Email id, Password, Contact number, Gender for creating an account.

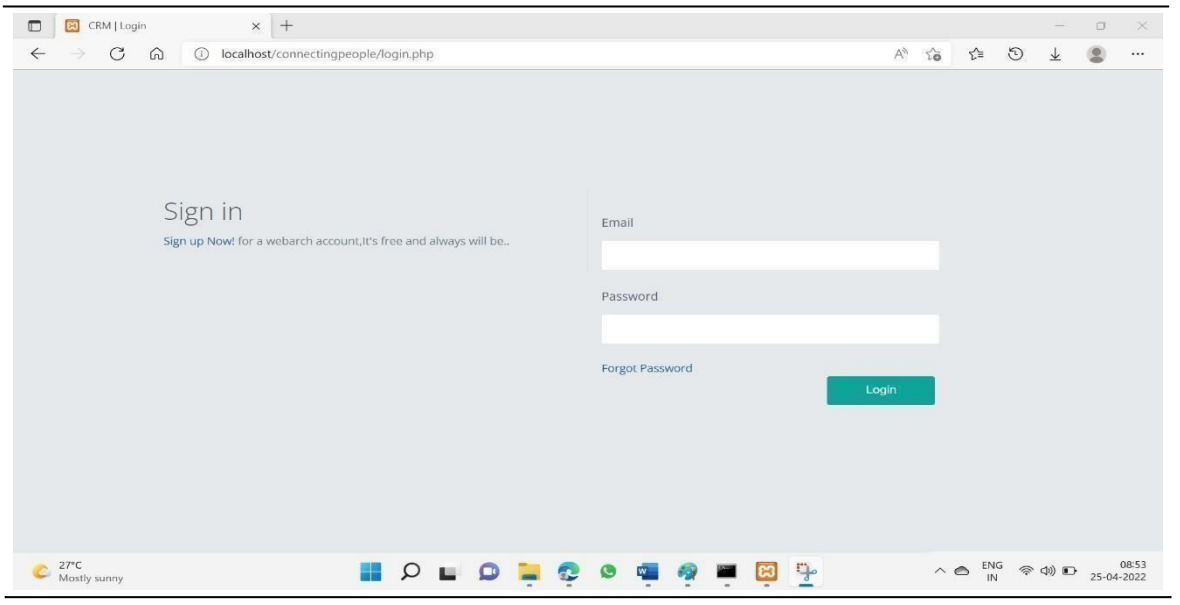


Figure 7.13. Login Page.

This is the Sign in page of the app where after the user has created an account he can put details like email, password to login into the app.

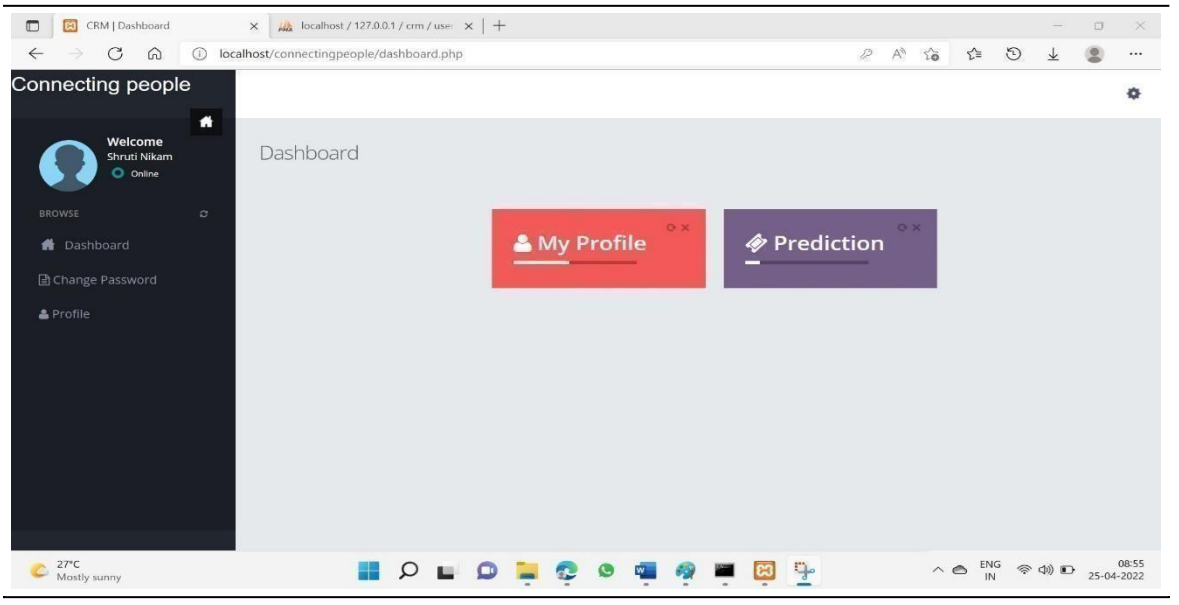


Figure 7.14. Dashboard after Login.

This is the dashboard visible to the user after he has logged in in the app. Here he can navigate to the “My Profile” section to update his profile or he can go to the prediction tab for getting the details of the person whom he can contact for help.

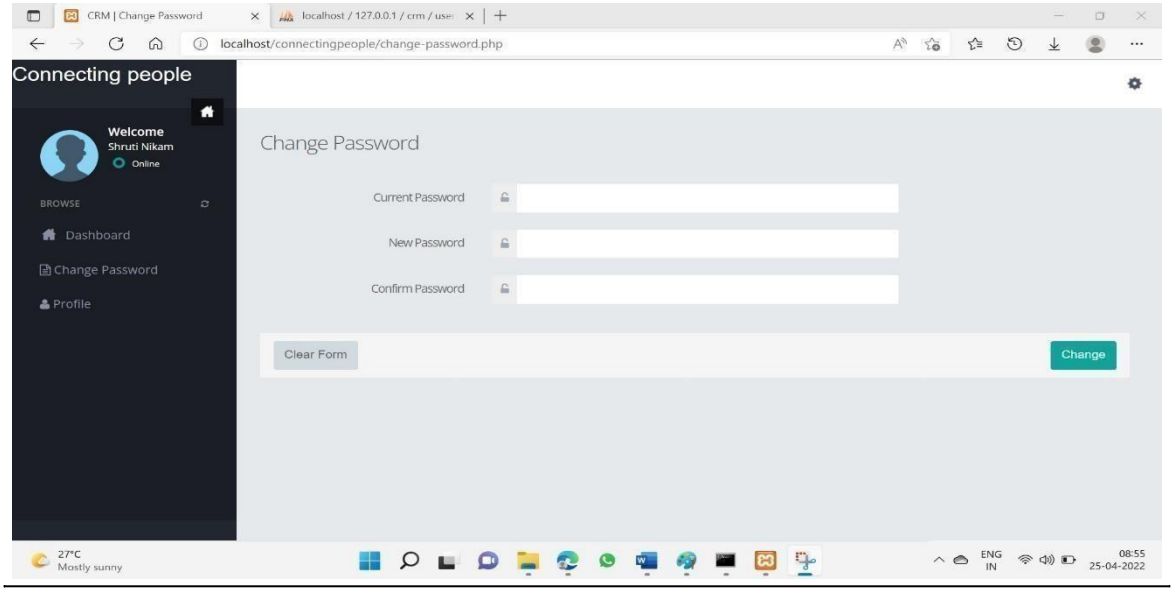


Figure 7.15. Change Password Option.

Here user can update their existing password with the new password easily

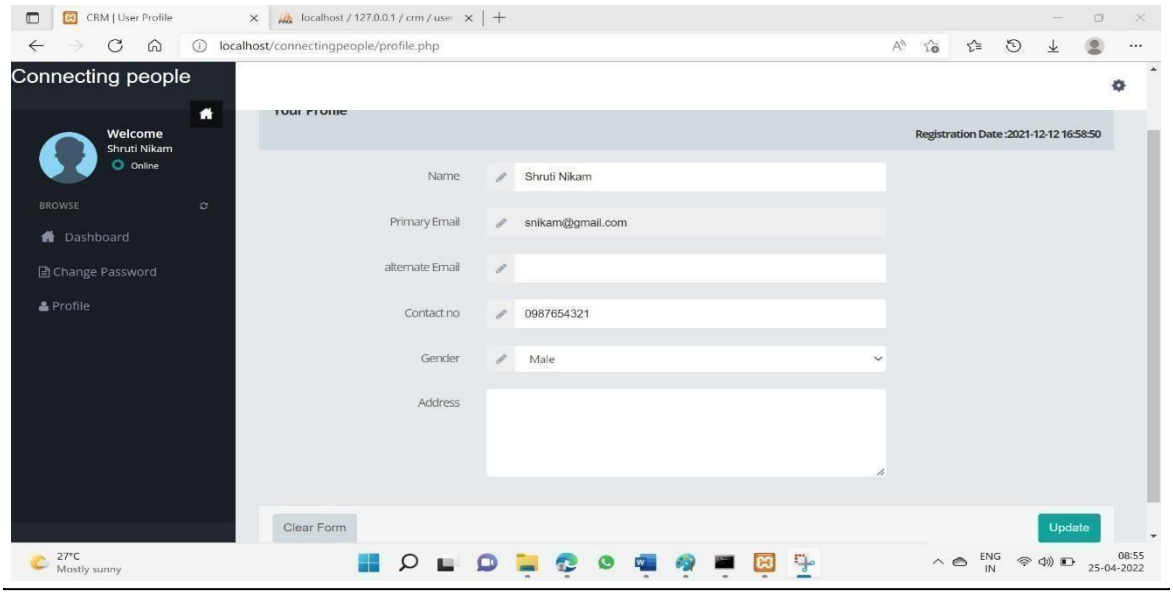


Figure 7.16. Profile Page.

This is the profile page of the user signed in the app. Here the user can update their name, alternate email, contact number, gender and address

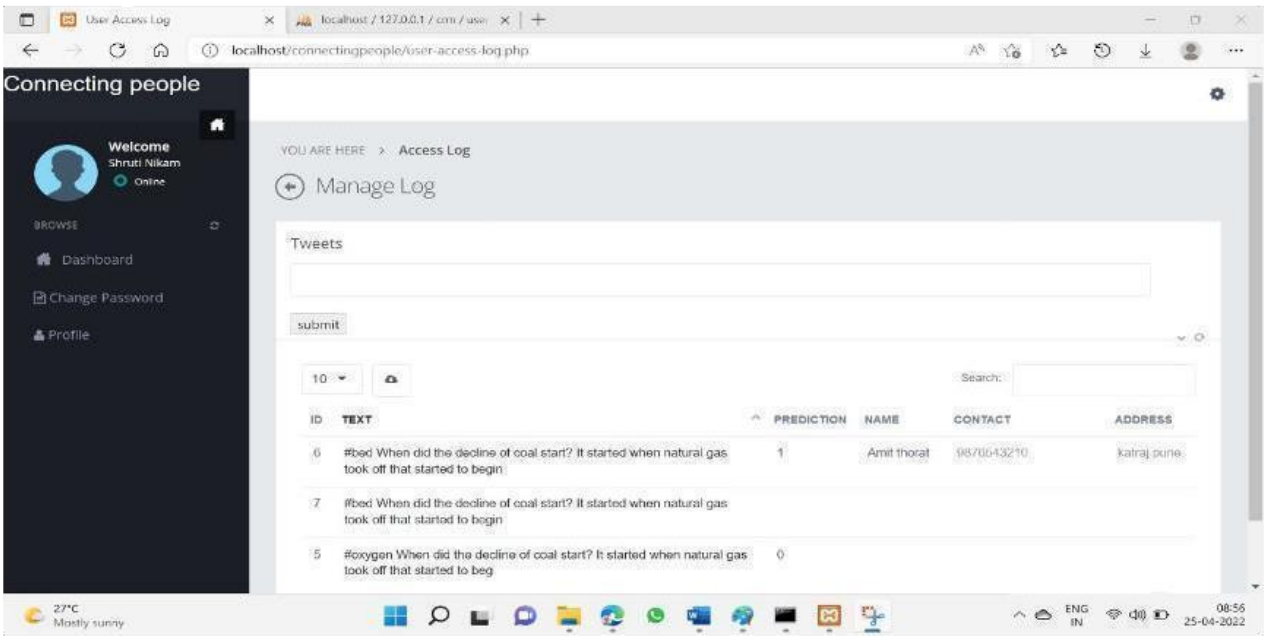


Figure 7.17. Prediction Page.

This is the page where the user can search for the tweet to avail the resources the person needs. For this the user needs to fill the tweet box with Ex. #Beds needed and click on the search button. By this action tweets that can provide beds will be displayed with the person name, contact number and address.

CHAPTER 8

TESTING

8.1 TEST PLANS

A system should always be tested thoroughly before implementing it, as regards its individual programs. This is because implementing a new system is a major job which requires a lot of man hours and a lot of other resources, so an error not detected before implantation may cost a lot. Effective testing early in the process translates directly into long term cost saving from reduced number of errors. This is also necessary because in some cases, a small error is not detected and corrected before installation, which may explode into a much larger problem. Programming and testing is followed by the stage of installing the new computer based system. Actual implementation of the system can begin at this point using either a parallel or a direct changeover plan, or some blend of two. Testing and implementation of fire fighting robot controlled using android application is carried out as below. Software testing is a critical element of Software Quality Assurance (SQA) and represents the ultimate review of specification, design and coding. The purpose of product testing is to verify and validate the various work products viz. units, integrate unit, final product to ensure that they meet their respective requirements.

8.2 TESTING PROCEDURE

Software Testing is the critical element of the Software Quality Assurance and represents the ultimate review of specification, design and coding. Testing is the process of checking whether software works according to the specification. Testing will be performed by running the program using the test data. Testing is vital to the success of the system. It will also test whether the system identifies the problemcorrectly.

System is tested by following steps:

- Unit Testing: Each program is tested individually using dummy records to see whether that programproduces satisfactory reports.
- Sequential Testing: The program, whose output will affect the processing done by another program, will be tested using dummy records. Testing: The system is corrected in such a way that it does not affect the forced system failure. This testing is done with low volumes of data

8.3 TEST STRATEGY

The test strategy consists of a series of different tests that will fully exercise the system. The primary purpose of the test is to uncover the system limitations. Following are the several tests that will be conducted:

8.3.1 Unit Testing:

Testing conducted to verify the implementation of the design for one software element (e.g., unit, module) is called unit testing. The purpose of unit testing is to ensure that the program logic is complete and correct and ensuring that the component works as designed. In this module, each unit will go through Unit testing after the completion of the module. The bugs in module testing will be reported in the Test Log document and will be reported to the developers. After fixing the bug successfully, one more iteration of module testing (Regression Testing) is done. This process is repeated till all critical test cases pass.

8.3.2 Integration Testing:

Testing conducted in which software elements, hardware elements, or both are combined and tested until the entire system has been integrated. The purpose of integration testing is to ensure that design objectives are met and ensures that the software, as a complete entity, complies with operational requirements. This type of testing will be done after all module test cases are passed through module testing, security testing, performance testing, user interface testing and regression testing

8.3.3 Performance Testing:

In developing the system, we are going to use Java which will reduce the response time. In Performance Testing, We are going to test Response time for each Screen. It is a type of non-functional testing. Performance testing is testing that is performed; to determine how fast some aspect of a system performs under a particular workload. It can serve different purposes like it can demonstrate that the system meets performance criteria. It can compare two systems to find which performs better. Or

it can measure what part of the system or workload causes the system to perform badly. This process can involve quantitative tests done in a lab, such as measuring the response time or the number of MIPS (millions of instructions per second) at which a system functions.

8.3.4 Regression Testing:

Testing done to ensure that the changes to the application have not adversely affected previously tested functionality. Here testing will take care that the test cases passed during the first module testing will not be affected in the subsequent rounds of module testing.

8.4 TEST CASES

The listed tests were conducted in the software at the various development stages. Unit testing was conducted. The errors were debugged and regression testing was performed. The integration testing will be performed once the system is integrated with other related systems like Inventory, Budget etc. Once the design stage was over the Black Box and White Box Testing was performed on the entire application. The results were analyzed and the appropriate alterations were made. The test results proved to be positive and henceforth the application is feasible and test approved.

8.5 TEST RESULTS

The listed tests were conducted in the software at the various development stages. Unit testing was conducted. The errors were debugged and regression testing was performed. The integration testing will be performed once the system is integrated with other related systems like Inventory, Budget etc. Once the design stage was over the Black Box and White Box Testing was performed on the entire application. The results were analyzed and the appropriate alterations were made. The test results proved to be positive and henceforth the application is feasible and test approved.

Table 8.1: Test Result

Sr.No	Description	Test Case I/P	Actual Result	Expected	Test Criteria (P/F)
1	Install Python	Python Exe	Should get install properly	Proper Installed	P
2	Installing Libraries	Library command for install	Should Get installed	Library Installed Successfully	P
3	Training Dataset	Dataset Training	Error in Training Model	Trained Model	F
4	Training Dataset	Dataset Training	Trained Model	Trained Model	P
5	Login Credentials	User Name and Password	Login Unsuccessful	Unsuccessful Login	F
6	Login Credentials	User Name and Password	Login Successful	Successful Login	P
7	Password	Current and NewPassword	Password Updated	Update Password	P
8	Select Limited Dataset	Number of Rows	Should select and train the selected data	Trained Model	P
9	Prediction	URL as in-Put	Should Predict the result	Result Predicted	P

CHAPTER 9

PROJECT PLAN

9.1 PROJECT ESTIMATES

9.1.1 Effort Estimate Table:

Table 9.1: Effort Estimate Table

Task	Effort weeks	Deliverables	Milestones
Analysis of existing systems & compare with proposed one	4 weeks		
Literature survey	1 weeks		
Designing & planning	2 weeks		
System flow	1 weeks		
Designing modules & its´ deliverables	2 weeks	Modules: design document	
Implementation	7 weeks	Primary system	
Testing	4 weeks	Test Reports	Formal
Documentation	2 weeks	Complete project report	Formal

9.1.2 Project Description:

Table 9.2: Project Scheduling

Phase	Task	Description
Phase 1	Analysis	Analyze the information given in the IEEE paper.
Phase 2	Literature survey	Collect raw data and elaborate on literature surveys.
Phase 3	Design	Assign the module and design the process flow control.
Phase 4	Implementation	Implement the code for all the modules and integrate all the modules.
Phase 5	Testing	Test the code and overall process weather the the process works properly.
Phase 6	Documentation	Prepare the document for this project with conclusion and future enhancement.

9.1.3 Estimation of KLOC:

The number of lines required for implementation of various modules can be estimated as follows:

Sr. No.	Modules	KLOC
1	Graphical User Interface	0.20
2	Back-end Algorithm Implementation	1.2
3	Front-Side Coding	1.2
4	Back-end Connectivity	0.6

Thus the total number of lines required is approximately 2.60 KLOC. D

= (Total KLOC / SLOC in a day) / 30

= (3.6/0.025)/30

= 4.8

9.2 RISK MANAGEMENT

9.2.1 Overview of Risk Mitigation, Monitoring, Management

Risk management organizational role

Each member of the organization will undertake risk management. The development team will consistently be monitoring their progress and project status to identify present and future risks as quickly and accurately as possible. With this said, the members who are not directly involved with the implementation of the product will also need to keep their eyes open for any possible risks that the development team did not spot. The responsibility of risk management falls on each member of the organization, while William Lord maintains this document.

Business Impact Risk

- Amount and quality of documentation that must be produced and delivered to the customer will be supplied with a complete online help file and users manual for Game Forge. Coincidentally, the customer will have access

to all development documents for Game Forge, as the customer will also be grading the project.

- Governmental constraints in the construction of the product are known.
- Costs associated with late delivery Late delivery will prevent the customer from issuing a letter of acceptance for the product, which will result in an incomplete grade for the course for all members of the organization.
- Costs associated with a defective product Unknown at this time.

Customer Related Risks

- Have you worked with customers in the past? Yes, All team members have completed at least one project for the customer, though none of them have been to the magnitude of the current project.
- Does the customer have a solid idea of what is required? Yes, the customer has access to both the System Requirements Specification, and the Software Requirements Specification.
- Will the customer agree to spend time in formal requirements gathering meetings to identify project scope? Unknown. While the customer will likely participate if asked, the inquiry has not yet been made.

Process Risks

- Does senior management support a written policy statement that emphasizes the importance of a standard process for software development? N/A. PA Software does not have a senior management. It should be noted that the structured method has been adopted. At the completion of the project, it will be determined if the software method is acceptable as a standard process, or if changes need to be implemented.
- Has your organization developed a written description of the software process to be used on this project? Yes.

- Are staff members willing to use the software process? Yes. The software process was agreed upon before development work began.
- Is the software process used for other products? N/A. PA Software has no other projects currently.

Technical Issues

- Are facilitated application specification techniques used to aid in communication between the customer and the developer? The development team will hold frequent meetings directly with the customer. No formal meetings are held (all informal). During these meetings the software is discussed and notes are taken for future review.
- Are specific methods used for software analysis? Special methods will be used to analyze the software progress and quality. These are a series of tests and reviews to ensure the software is up to speed. For more information, see the Software Quality Assurance and Software Configuration Management documents.
- Do you use a specific method for data and architectural design? Data and architectural design will be mostly object oriented. This allows for a higher degree of data encapsulation and modularity of code.

Technology Risk

- Is the technology to be built new to your organization? No
- Does the software interface with new or unproven hardware? No
- Is a specialized user interface demanded by the product requirements? Yes.

Development Environment Risks

Is a software project management tool available? No. No software tools are to be used. Due to the existing deadline, the development team felt it would be more productive to begin implementing the project than trying to learn new software tools. After the completion of the project software tools may be implemented for future projects.

9.1 PROJECT SCHEDULE

9.1.1 Project task

- Getting Basic Knowledge of python
- Going through the previous existing system
- Looking for required libraries in python
- Collecting Dataset
- Training Model using ML algorithm
- Building GUI for better outlook
- Dividing the assigned task among group members.

9.1.2 Timeline Chart

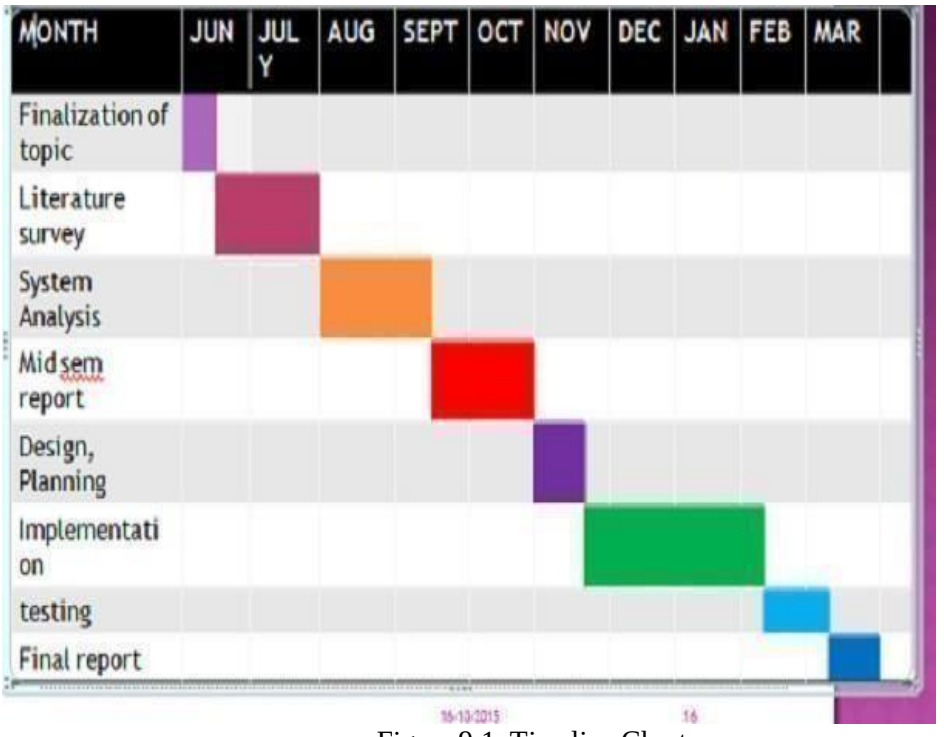


Figure 9.1: Timeline Chart

CHAPTER 10

CONCLUSION

10.1 Conclusion

The main aim of this project was to design and implement an application to avail resources during crisis using machine learning methods and performance analysis of the methods and it has been achieved successfully. The proposed system presents the matches between problems and aid problems through tweets in large scale disasters or crisis. The proposed approach used various classifications and ensemble learning methods in which Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Random Forest classifiers are used and This project will be advantageous to reduce fatality providing basic Necessities such as food, shelter and medicines, it will also provide a particular resource needed according to the crises and reduce death.

10.2 Future Scope

Future work concerns deeper analysis of particular mechanisms, new proposals to try different methods or simply curiosity.

10.3 Limitations of Project

- Unable to access Twitter data due to pending access to Twitter API.
- The spelling of the locations must be correct in order to matchthem.

REFERENCES

[1] Douglas Cirqueria and Gultekin Cakir , "*Explainable Sentiment Analysis Application for Social Media Crisis Management in Retail*",2020.

[2] Azzam Mourad, Ali Srour and Mohamad Arafeh, "*Critical Impact of Social Networks Infodemic on Defeating Coronavirus COVID-19 Pandemic: Twitter- Based Study and Research Directions*",Transaction on Network and Service Management, IEEE 2020.

[3] Umar Ali Bukar and Fatimah Sidi, "*Crisis Informatics in the Context of Social Media Crisis Communication: Theoretical Models, Taxonomy, and Open Issues*", IEEE Access 2020.

[4] Jayashree Domala and Vinit Masrani, "*Automated Identification of Disaster News for Crisis Management using Machine Learning and Natural Language Processing*", International Conference on Electronics and Sustainable Communication System, IEEE 2020.

[5] Tejas Shah, Zhenyu Wen and Divya Pullarkatt, "*Use of Social Media Data in Disaster Management: A Survey*", AI and IoT technologies in smart cities,
MDPI 2020.

[6] Vedant Dhurve, Krutika Hedaoo, Himanshu Itankar, Jayesh Lanjewar , "*Survey on Content Based Disaster Management Using Social Media*", International Journal of Scientific Research Engineering Trends , 2021.

[7] Anita Saroj and Sukomal Pal, "*Use of social media in crisis management: A survey*", Elsevier 2020.

[8] Christian Reuter and Amanda Lee Hughes, "*Social Media in Crisis Management: An Evaluation and Analysis of Crisis Informatics Research*", IEEE 2018.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology(IJARSCT)
Volume 2, Issue 3, April 2022 DOI: 10.48175/IJARSCT-3937
Impact Factor: 6.252

Connecting People to Avail the Resources During Crisis Through Twitter Using Machine Learning

Shaikh Areeba[1], Muskan Singh[2], Pooja Nandukumar Shingewad[3], Shruti Nikam[4], Prof. (Mrs) S. A. Nagtilak[5]
Students, Department of Information Technology[1,2,3,4]
Faculty, Department of Information Technology[5]
Smt Kashibai Navale College of Engineering (SKNCOE), Pune. Maharashtra, India

Abstract: *Although social media has become the most widely utilized and active form of communication, research on its usage in crisis management is still in its early stages. As a result, this research examines the rising body of knowledge on social media and crisis management. [1] Between October 2017 and January 2018, a review was conducted, which included locating and retrieving records from an electronic database. The outcomes of this study indicated that the rise of social media has altered the landscape of crisis communication by allowing for greater engagement. However, due to its nature, social media might also be used to spark a crisis. This means that the crisis can be both produced and disseminated through social media. Nonetheless, social media's promise as a crisis-resolution tool is undeniable. It has the capability of proving a claim, dispelling false rumors, or just demonstrating a fact. As a result, practitioners should understand how social media works and how to best use it to interact with their stakeholders. This study also includes other findings, limits, and useful suggestions for scholars and practitioners interested in learning more about the role of social media on crisis communication and management. As most of the crisis problem were reported via twitter. However, most of the problem reported and corresponding responses via twitter were not successfully exchanged between victim's and resource organization. As a result, most of the tweets were not getting help. Thus, we designed a platform where people can avail the resources of crisis through tweets matching concept using machine learning.*

Keywords: Crisis, Machine Learning, Twitter Dataset, SVM , RF, KNN

I. INTRODUCTION

The main purpose of crisis management is to figure out what tangible activities need to be taken before (prevention, preparedness), during (response), and after a crisis occurs (recovery and mitigation). In order to execute these jobs efficiently,it is beneficial to use data from numerous sources, including public witnesses to emergency situations. Emergency operations centers will be able to organize and execute rescue and response activities using this information. In recent years,a number of research projects have looked into the use of social media as a source of information for effective crisis management. People utilize social media (SM) to explain and respond to various situations they find themselves in, such as emergencies. As a result, using SM material to enhance crisis management is worthwhile, particularly in terms of revealing relevant and previously unknown details regarding crises in real time. As a result, using SM material to enhance crisis management is worthwhile, particularly in terms of revealing relevant and previously unknown details regarding crises in real time. Using Twitter and Machine Learning, we suggested a Social Media Analysis for Connecting People to Resources During a Crisis.

II. LITERATURE SURVEY

In 2) proposed concentrate on fostered an Explainable Sentiment Analysis (XSA) application for Twitter information, and proposes research suggestions zeroed in on assessing such application in a speculative emergency the board situation. Especially, they assess, through conversations and a reproduced client analyze, the XSA support for getting client's necessities, as well as if marketing experts would trust such an application for their dynamic cycles.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology(IJARSCT)
Volume 2, Issue 3, April 2022 DOI: 10.48175/IJARSCT-3937
Impact Factor: 6.252

In 3) explored the COVID-19 data demic adverse consequence on the significant endeavors to overcome the pandemic through a clever huge scope Twitter-based study, which gave quantitative appraisal utilizing genuine tests mirroring the real conditions. The experimental investigation of 1 million COVID-19-related tweets having a place with 288K exceptional clients delineated the extreme effect of deluding individuals and spreading temperamental data.

4) to audit and investigations the relationship of web-based media-based emergency correspondence with regards to emergency informatics and its scientific categorization and the connected emergency correspondence hypothetical models to infer the difficulties and constraints. The consequence of the finding shows that partner association is an understudied field, while data dependability and handling for dynamic purposes, the more extensive use of online media destinations.

5) framework the goal was to consequently scratch news from English news sites and recognize debacle pertinent news utilizing normal language handling strategies and AI ideas, which can additionally be powerfully shown on the emergency the board sites. The total model was robotized and requires no physical work by any means.

6) To review and analyses the relationship of social media-based crisis communication in the context of crisis informatics and its taxonomy and the related crisis communication theoretical models to derive the challenges and limitations. The result of the finding shows that stakeholder interaction is an understudied field, while information reliability and processing for decision- making purposes, the wider application of social media sites.

7) This paper proposed a method to discover matches between problem reports and aid messages from tweets in large - scale disasters. Through a series of experiments, we demonstrated that the performance of the problem-aid matching can be improved with the usage of semantic orientation of excitation polarities,

8) study incorporates the techniques for online media information grouping and occasion location as well as spatial and worldly data extraction. Besides, a scientific categorization of the exploration aspects of online media information the executives and investigation for calamity the board was likewise proposed, which was then applied to a review of existing writing and to talk about the center benefits and detriments of the different philosophies.

III. Problem Definition

Because of the mass media's ever-expanding reach and influence in today's society, they have become an inescapable part of any social situation. Mass media enables information access for policymakers, managers, and citizens alike, as well as boosting the pace with which new information is received, accumulated, and disseminated, enhancing its role in disaster management. As a result, we need to create a system that uses tweets to assist individuals in acquiring supplies during a crisis.

IV. Implementation Details Of Module

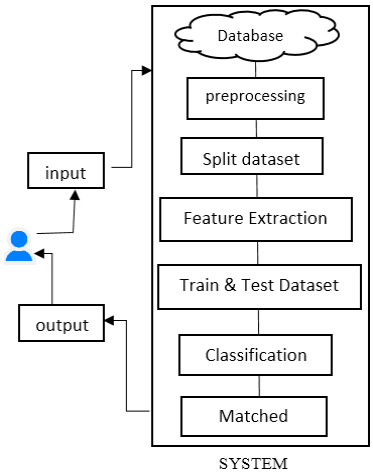


Figure: System Architecture

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology(IJARSCT)
Volume 2, Issue 3, April 2022 DOI: 10.48175/IJARSCT-3937
Impact Factor: 6.252

The proposed system is web-based application build using php and css as front end and python for backend. The connectivity is done using MySQL database. This proposed system presents a comparison of machine learning algorithm: SVM, RF, KNN. For the implementation of the ML algorithms, the dataset was partitioned into the training set and testingset. The dataset is passed in pre-processing state where unwanted data or null values are removed. Later on, in next step the features are been extracted using various machine learning techniques such as support vector machine, KNN and Random Forest techniques. A comparison between all the 3 algorithms will be made. The algorithm that gives the best results willbe supplied as a model to the website. The website will be made from a python framework, called flask. And it will host the database on Xampp, Python and its libraries. The process of the proposed system is as follows,

1. First the user registered in the website.
2. After registration, user login using email-id and password.
3. Then the home page of the website opens, in which there is Profile tab and the Resources tab.
4. In the profile tab user information is provided in which name ,id, email , address is given.
5. In the Resources tab users can find the resources they needed during the crisis. User have to give an input example :# Bed needed , #Oxygen etc.
6. After giving an input the system will find the resources an if resources are available , then name contact number and the address of the person who is providing the resources is shown.
7. Along with that email containing resources detail is send to the user through the system.

V. CONCLUSION

The main aim of this project was to design and implement an application to avail resources during crises using machine learning methods and performance analysis of the methods and it has been achieved successfully. The proposed system presents he matches between problems and aid problems through tweets in large scale disasters or crises. The proposed approach used various classifications and ensemble learning methods in which Support Vector Machine(SVM), K-Nearest Neighbor (KNN) and Random Forest classifiers accuracy has been achieved. This project will be advantageous to reduce fatalities by providing basic Necessities such as food, shelter and medicines, it will also provide a particular resource needed according to the crises and reduce death.

REFERENCES

[1]. Douglas Cirqueria and Gultekin Cakir , "Explainable Sentiment Analysis Application for Social Media Crisis Management in Retail",2020

[2]. Azzam Mourad , Ali Srour and Mohamad Arafeh, "Critical Impact of Social Networks Infodemic on Defeating Coronavirus COVID-19 Pandemic: Twitter-Based Study and Research Directions",Transaction on Network and Service Management, IEEE 2020

[3]. Umar Ali Bukar and Fatimah Sidi, "Crisis Informatics in the Context of Social Media Crisis Communication: Theoretical Models, Taxonomy, and Open Issues", IEEE Access 2020

[4]. Jayashree Domala and Vinit Masrani, "Automated Identification of Disaster News for Crisis Management using Machine Learning and Natural Language Processing", International Conference on Electronics and SustainableCommunication System ,IEEE 2020

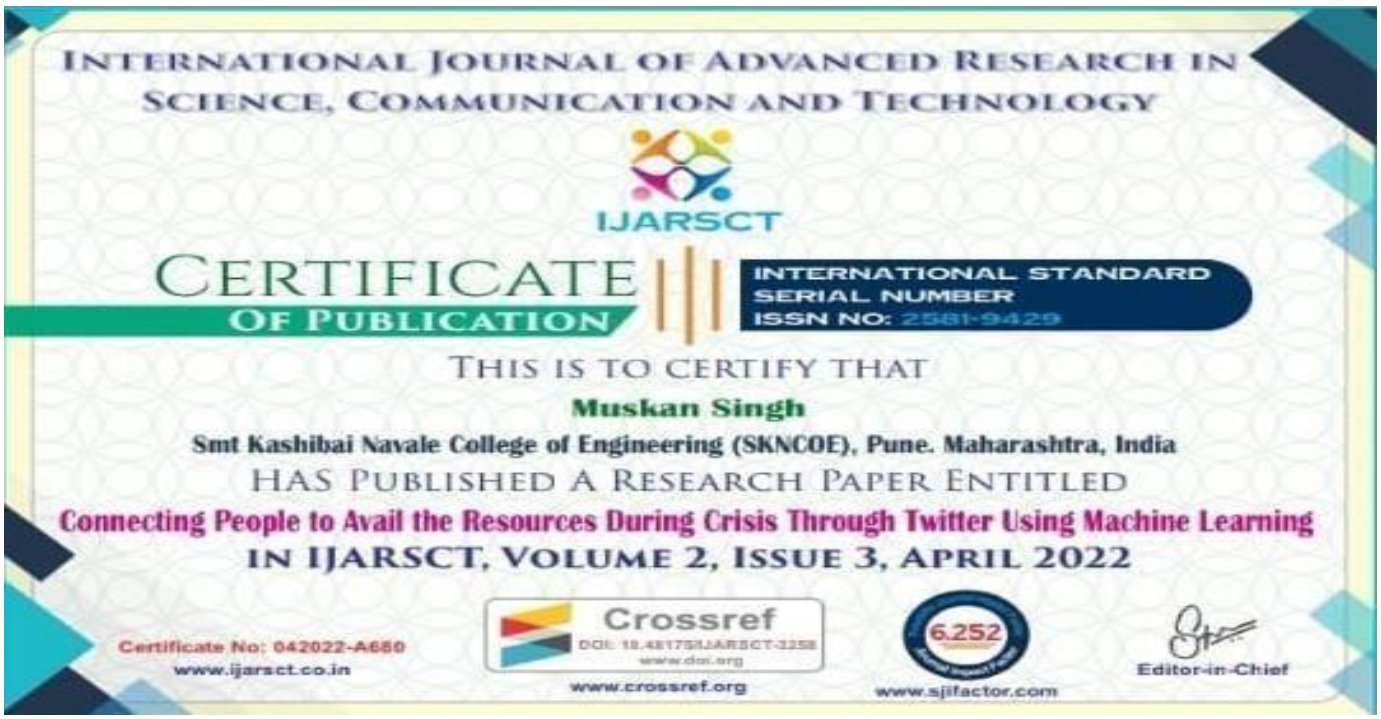
[5]. Tejas Shah, Zhenyu Wen and Divya Pullarkatt, "Use of Social Media Data in Disaster Management: A Survey", AI and IoT technologies in smart cities, MDPI 2020

[6]. Vedant Dhurve, Krutika Hedaoo, Himanshu Itankar, Jayesh Lanjewar , "Survey on Content Based Disaster Management Using Social Media", International Journal of Scientific Research & Engineering Trends , 2021

[7]. Anita Saroj and Sukomal Pal, "Use of social media in crisis management: A survey", Elsevier 2020

[8]. Christian Reuter and Amanda Lee Hunges, "Social Media in Crisis Management: An Evaluation and Analysis of Crisis Informatics Research", Research Article 2018


INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN SCIENCE, COMMUNICATION AND TECHNOLOGY
IJARSCT
CERTIFICATE OF PUBLICATION
INTERNATIONAL STANDARD SERIAL NUMBER ISSN NO: 2581-9429
THIS IS TO CERTIFY THAT
Shaikh Areeba Abdul Hakeem
Smt Kashibai Navale College of Engineering (SKNCOE), Pune. Maharashtra, India
HAS PUBLISHED A RESEARCH PAPER ENTITLED
Connecting People to Avail the Resources During Crisis Through Twitter Using Machine Learning
IN IJARSCT, VOLUME 2, ISSUE 3, APRIL 2022
Certificate No: 042022-A679
www.ijarsct.co.in
Crossref
www.crossref.org
6.252
www.sjifactor.com
Editor-in-Chief


INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN SCIENCE, COMMUNICATION AND TECHNOLOGY
IJARSCT
CERTIFICATE OF PUBLICATION
INTERNATIONAL STANDARD SERIAL NUMBER ISSN NO: 2581-9429
THIS IS TO CERTIFY THAT
Muskan Singh
Smt Kashibai Navale College of Engineering (SKNCOE), Pune. Maharashtra, India
HAS PUBLISHED A RESEARCH PAPER ENTITLED
Connecting People to Avail the Resources During Crisis Through Twitter Using Machine Learning
IN IJARSCT, VOLUME 2, ISSUE 3, APRIL 2022
Certificate No: 042022-A680
www.ijarsct.co.in
Crossref
www.crossref.org
6.252
www.sjifactor.com
Editor-in-Chief


INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN SCIENCE, COMMUNICATION AND TECHNOLOGY
IJARSCT
CERTIFICATE OF PUBLICATION
INTERNATIONAL STANDARD SERIAL NUMBER ISSN NO: 2581-9429
THIS IS TO CERTIFY THAT
Pooja Nandukumar Shingewad
Smt Kashibai Navale College of Engineering (SKNCOE), Pune. Maharashtra, India
HAS PUBLISHED A RESEARCH PAPER ENTITLED
Connecting People to Avail the Resources During Crisis Through Twitter Using Machine Learning
IN IJARSCT, VOLUME 2, ISSUE 3, APRIL 2022
Certificate No: 042022-A681
www.ijarsct.co.in
Crossref
www.crossref.org
6.252
www.sjifactor.com
Editor-in-Chief


INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN SCIENCE, COMMUNICATION AND TECHNOLOGY
IJARSCT
CERTIFICATE OF PUBLICATION
INTERNATIONAL STANDARD SERIAL NUMBER ISSN NO: 2581-9429
THIS IS TO CERTIFY THAT
Shruti Shrikant Nikam
Smt Kashibai Navale College of Engineering (SKNCOE), Pune. Maharashtra, India
HAS PUBLISHED A RESEARCH PAPER ENTITLED
Connecting People to Avail the Resources During Crisis Through Twitter Using Machine Learning
IN IJARSCT, VOLUME 2, ISSUE 3, APRIL 2022
Certificate No: 042022-A682
www.ijarsct.co.in
Crossref
www.crossref.org
6.252
www.sjifactor.com
Editor-in-Chief

Sentence

Social media has become the most used and active way of communication; however, studies that look at social media use in crisis management are in their growing stage. Thus, this study analyzes this growing body of knowledge that researched into social media and crisis management. A review was undertaken between October 2017 to January 2018 which entails sourcing and retrieving materials from an electronic database. The findings of this review confirmed that the growth of social media has transformed the crisis communication landscape because it allows more interactivity. Still, a crisis could also be catalyzed by social media because of its nature. This signifies that the crisis can be created in social media, as well as distributed by social media, respectively. Yet, the potential of social media as a crisis resolution tool is indisputable. It has the potential to prove a statement, to dispel untrue rumors or just to show the fact. As most of the crisis problems were reported via twitter.However, most of the problem reported and corresponding responses via twitter were not successfully exchanged between victim's and resource organizations. As a result, most of the tweets were not getting help. Thus, we designed a platform where people can avail the resources of crisis through tweets matching concept using machine learningWe are very thankful to all those who have provided us valuable guidance towards the completion of this Seminar Report on "Connecting People to Avail the Resources During Crisis Through Twitter Using Machine Learning" as part of the syllabus of bachelor's course. We express our sincere gratitude towards the cooperative department who has provided us with valuable assistance and requirements for the system development. We are very grateful and want to express our thanks to Prof. (Mrs) S.A. Nagtilak for guiding us in the right manner, correcting our doubts by giving us their time whenever we required, and providing their knowledge and experience in making this project.We are also thankful to the HOD of our Information Technology department Prof. Dr. M. L. Bangare for his moral support and motivation which has encouraged us in making this project. The acknowledgement will be incomplete if we do not thank our Principal Prof. Dr. A.V. Deshpande,Vice Principal Prof. Dr. K. R. Borole, who gave us his constant support and motivation which has been highly instrumental in making our project. WeThe primary goal of crisis management is to determine concrete steps that must be taken before (prevention, preparedness), during (response), and after a crisis has occurred (recovery and mitigation). It is beneficial to use data from multiple sources, including the public as witnesses to emergency situations, in order to complete these tasks effectively. Such information will enable emergency operations centers to plan and execute rescue and response operations. A number of research studies have looked at the use of social media as a source of knowledge for effective crisis management in recent years. People use social media (SM) to explain and address various circumstances in which they find themselves, such as crises. As a result, it is worthwhile to use SM content to aid crisis management, especially by disclosing useful and previously unknown details about crises in real time. As a result, it is worthwhile to use SM content to aid crisis management, especially by disclosing useful and previously unknown details about crises in real time. Here proposed Social Media Analysis to support Crisis Management Using online Machine Learning which classifies the relevant and irrelevant data Due to the ever increasing reach and influence of the mass media in today's society, they have become an inseparable part of any social situation. Mass media facilitates access to information for policymakers, managers and citizens alike and increases the speed with which new information is gathered, accumulated and distributed, further increasing its role in managing situations such as natural disasters. Mass media can have a negative role in a crisis, increasing its scope or play a positive role and help solve the crisis. Nevertheless, the role of mass media in any crisis cannot be denied. Mass media and its tools play an important role in what we think about, how we think about it, and what, how and why effects our emotions. In the eye of normal people, regardless of their education level, the belief that mass media has unimaginable power is widespread.Most people think that mass media can change philosophical and political opinions, give thoughts a new form and guide all our actions. These show the undeniable role of the mass media in human societies especially in crisis situations in which they can help the victims or instead increase their problems and suffering.Aim: The aim of our project is to help people who are suffering due to crisis related complications or due to lockdown by connecting them to people who can provide resources or time in doing good.Motivation: In covid-19 times, social media platforms are playing a major role and in particular Twitter where people are posting requests, updates, and helpful resources to find the latest information and access the resources faster. As a global platform Twitter has seen a tremendous surge of users posting updates about help supplies and more such situations in the past few weeks. The motivation behind this project has been to gather information in a structured way related to the required resources. It will help people to find correct information on Twitter related to any specific tweet and provide resources not only during covid but also in different crises.Chapter 2: Deals with the Project Related Work i.e., Literature Survey. It synthesizes related research done on this topic. Literature survey includes different research done methods used their algorithms and which area the re- search lacks along with that it also includes existing system architecture and its

detailed explanation and working.Chapter 3: This chapter gives an overall view of the model which we are going to use along with the algorithm of the system. The algorithm which we aregoing to be used are SVM, RF and KNN. Chapter 4: This chapter gives an overall view of tools and techniques used in the system and the information related to them.Chapter 5: Deals with System Design. We need system design for defining the architecture, modules, interfaces, and data for a system to satisfy specified requirements. There are different types of system diagrams in this chapter such as architecture diagrams, class diagrams, activity diagrams, sequence diagram.Chapter 6: This part includes the implementation. Putting strategies into place, utilizing resources and executing parts of the system. Chapter 7: This chapter contains the detailed information about the system implementation of the project assumptions which are taken, the algorithm which we have used in the project such as SVM, KNN and RF. Also explained the flowchart and methodologies in the chapterSMT.Due to the ever increasing reach and influence of the mass media in today's society, they have become an inseparable part of any social situation. Mass media facilitates access to information for policymakers, managers and citizens alike and increases the speed with which new information is gathered, accumulated and distributed, further increasing its role in managing situations such as natural disasters. A tweet that informs about the possibility or emergence of a problem during a crisis that requires a treatment or countermeasure. Thus, we need to develop a system which makes use of tweets to help people in obtaining the resources during a crisis.2The process of the proposed system is as follows, 1. First the user registered in the website. 2. After registration, user login using email-id and password . 3. Then the home page of the website opens, in which there is Profile tab and the Resources tab. 4. In the profile tab user information is provided in which name ,id, email , address is given. 5. In the Resources tab users can find the resources they needed during the crisis. User have to give an input example :# Bed needed , #Oxygen etc. 6. After giving an input the system will find the resources an if resources are available , then name contact number and the address of the person who is providing the resources is shown. 7. Along with that email containing resources detail is send to the user through the system.During this covid-19 pandemic, there were development of so many technologies and techniques to minimize the spread of covid-19. Some of this technology mention in this were the City Digital Twin,Policies Tracker, Fuzzy Logical Approach, as this technology tried to minimizethe spread of covid-19 butit was not fully supported toward the minimization of covid-19 Effect.The digital Twin help us to minimizethe spread of covid-19 but it notsupports fully due to minimum storage of data; Fuzzy Logical Approach isdeveloped to system to help the safety of health-related issues of the patient's condition according to thechanges of environment, but it only supported for a particular symptom such as Breathing could andPolicies Tracker it helps us to understand the policies taken by different government but it cannot fully filterthe policies taken by the government . If the distance between the classes is low then the chance of miss conception is high and vice versa. So we need to Select the class which has the highest margin. Margin = distance to positive point + Distance to negative point Examples of SVM boundaries Selecting best hyperplane for our classification. We will show data from 2 classes. The classes are represented by triangles and circlesThe main aim of this project was to design and implement an application to availresources during crises using machine learning methods and performance analysisof the methods and it has been achieved successfully. Importing all the required libraries for the project (numpy, pandas,nltk, sklearn, seaborn, plotly.graph_objs, random, plotly.offline, plotly.figure_factory, matplotlib.pyplot,re,os) The proposed system presents the matches between problems and aid problems through tweets in large scale disasters or crises.This project will be advantageous to reduce fatalitiesby providing basic Necessities such as food, shelter and medicines, it will also provide a particular resource needed according to the crises and reduce death.Future work concerns deeper analysis of particular mechanisms, new proposals to try different methods or simply curiosity.

Report Title:	Plagiarism report
Report Link: (Use this link to send report to anyone)	https://www.check-plagiarism.com/plag-report/77313b1869f40c6a305e36a788e679b152a1d1653209783
Report Generated Date:	22 May, 2022
Total Words:	1657
Total Characters:	10461
Keywords/Total Words Ratio:	0%
Excluded URL:	No
Unique:	86%
Matched:	14%