A Project report

Analysis of women safety in Indian cities

using Machine Learning on Tweets

Submitted in partial fulfillment of the

Requirements for the award of the Degree of

BACHELOR OF COMPUTER SCIENCE (B.Sc)

By

N.Kalyan(194312)

Ch.Syam prasad(194316)

V.Swaroop(194306)

Under the esteemed guidance of

Mr. Ravi Shankar
Lecturer, Department of Computer Science

DEPARTMENT OF COMPUTER SCIENCE

P. B. SIDDHARTHA COLLEGE OF ARTS AND SCIENCE
VIJAYAWADA – 10
An Autonomous college under the jurisdiction of Krishna University, Machilipatnam.

(A college with potential for excellence ; Re – accredited at the level 'A+' by NAAC, Banglore)2021-2022

P. B. SIDDHARTHA COLLEGE OF ARTS AND SCIENCE

VIJAYAWADA – 10

An Autonomous college under the jurisdiction of Krishna University, Machilipatnam.

(A college with potential for excellence ; Re – accredited at the level 'A' by NAAC, Banglore)

DEPARTMENT OF COMPUTER SCIENCE

CERTIFICATE

This is to certify that the project entitled **"ANALYSIS OF WOMEN SAFETY ON INDIAN CITIES USING MACHINE LEARNING ON TWEETS ",** is bonafied work of N.KALYAN(194312), CH.SYAM PRASAD(194316), V.SWAROOP(194306) submitted in partial fulfillment of the requirements for the award of the degree of **BACHELOR OF COMPUTER SCIENCE(B.SC)** from
P. B. Siddhartha College of Arts and Science, Vijayawada.

K. SudhirM.Sc (IT), M.Tech

Internal Guide

K. SridharM.Sc, M.Tech

Head of the Department

Department of Computer Science

ACKNOWLEDGEMENT

With great pleasure I want to take this opportunity to express my heartfelt gratitude to all the people who helped in making this project work a grand success.

First of all I would like to express my profound sense of gratitude to the Principal

Dr. M. RAMESH of or giving me the permission to carry out this project.

I express my deep gratitude and regards to **Mr. K. Sridhar,** Head of the Department of Computer Science**,** and project guide, **Mr.Ravi shankar,** Lecturer, Department in Computer Science for their encouragement and valuable guidance in bringing shape to this dissertation.

I am thankful to all the **Technical Staff and Non-teaching staff** in the department for their teachings and academic support.

N.Kalyan(194312)

Ch.Syam prasad(194316)

V.Swaroop(194306)

Analysis of Women Safety in Indian Cities Using MachineLearning on Tweets

ABSTRACT:

Women and girls have been experiencing a lot of violence and harassment in public places in various cities starting from stalking and leading to abuse harassment or abuse assault. This research paper basically focuses on the role of social media in promoting the safety of women in Indian cities with special reference to the role of social media websites and applications including Twitter platform Facebook and Instagram. This paper also focuses on how a sense of responsibility on part of Indian society can be developed the common Indian people so that we should focus on the safety of women surrounding them. Tweets on Twitter which usually contains images and text and also written messages and quotes which focus on the safety of women in Indian cities can be used to read a message amongst the Indian Youth Culture and educate people to take strict action and punish those who harass the women. Twitter and other Twitter handles which include hash tag messages that are widely spread across the whole globe sir as a platform for women to express their views about how they feel while we go out for work or travel in a public transport and what is the state of their mind when they are surrounded by unknown men and whether these women feel safe or not?

INDEX

LIST OF FIGURES

INTRODUCTION:

There are certain types of harassment and Violence that are very aggressive including staring and passing comments and these unacceptable practices are usually seen as a normal part of the urban life. There have been several studies that have been conducted in cities across India and women report similar type of sexual harassment and passing off comments by other unknown people. The study that was conducted across most popular Metropolitan cities of India including Delhi, Mumbai and Pune, it was shown that 60 % of the women feel unsafe while going out to work or while travelling in public transport.

Women have the right to the city which means that they can go freely whenever they want whether it be too an Educational Institute, or any other place women want to go. But women feel that they are unsafe in places like malls, shopping malls on their way to their job location because of the several unknown Eyes body shaming and harassing these women point Safety or lack of concrete consequences in the life of women is the main reason of harassment of girls. There are instances when the harassment of girls was done by their neighbours while they were on the way to school or there was a lack of safety that created a sense of fear in the minds of small girls who throughout their lifetime suffer due to that one instance that happened in their lives where they were forced to do something unacceptable or was sexually harassed by one of their own neighbor or any other unknown person.

Safest cities approach women safety from a perspective of women rights to the affect the city without fear of violence or sexual harassment. Rather than imposing restrictions on women that society usually imposes it is the duty of society to imprecise the need of protection of women and also recognizes that women and girls also have a right same as men have to be safe in the City.

Analysis of twitter texts collection also includes the name of people and name of women who stand up against sexual harassment and unethical behaviour of men in Indian cities which make them uncomfortable to walk freely. The data set that was obtained through Twitter about the status of women safety in Indian society was for the processed through machine learning algorithms for the purpose of smoothening the data by removing zero values and using Laplace and porter’s theory is to developer method of analyzation of data and remove retweet and redundant data from the data set that is obtained so that a clear and original view of safety status of women in Indian society is obtained.

Literature survey

1. Sentiment analysis of top colleges in India using Twitter data:

In today's world, opinions and reviews accessible to us are one of the most critical factors in formulating our views and influencing the success of a brand, product or service. With the advent and growth of social media in the world, stakeholders often take to expressing their opinions on popular social media, namely Twitter. While Twitter data is extremely informative, it presents a challenge for analysis because of its humongous and disorganized nature. This paper is a thorough effort to dive into the novel domain of performing sentiment analysis of people's opinions regarding top colleges in India. Besides taking additional preprocessing measures like the expansion of net lingo and removal of duplicate tweets, a probabilistic model based on Bayes' theorem was used for spelling correction, which is overlooked in other research studies. This paper also highlights a comparison between the results obtained by exploiting the following machine learning algorithms: Naïve Bayes and Support Vector Machine and an Artificial Neural Network model: Multilayer Perceptron. Furthermore, a contrast has been presented between four different kernels of SVM: RBF, linear, polynomial and sigmoid.

2. Study of Twitter sentiment analysis using machine learning algorithms on Python

Twitter is a platform widely used by people to express their opinions and display sentiments on different occasions. Sentiment analysis is an approach to analyze data and retrieve sentiment that it embodies. Twitter sentiment analysis is an application of sentiment analysis on data from Twitter (tweets), in order to extract sentiments conveyed by the user. In the past decades, the research in this field has consistently grown. The reason behind this is the challenging format of the tweets which makes the processing difficult. The tweet format is very small which generates a whole new dimension of problems like use of slang, abbreviations etc. In this paper, we aim to review some papers regarding research in sentiment analysis on Twitter, describing the methodologies adopted and models applied, along with describing a generalized Python based approach.

3. Determining the sentiment of opinions

Identifying sentiments (the affective parts of opinions) is a challenging problem. We present a system that, given a topic, automatically finds the people who hold opinions about that topic and the sentiment of each opinion. The system contains a module for determining word sentiment and another for combining sentiments within a sentence. We experiment with various models of classifying and combining sentiment at word and sentence levels, with promising results.

4. sentiment detection on twitter from biased and noisy data

In this paper, we propose an approach to automatically detect sentiments on Twitter messages (tweets) that explores some characteristics of how tweets are written and meta-information of the words that compose these messages. Moreover, we leverage sources of noisy labels as our training data. These noisy labels were provided by a few sentiment detection websites over twitter data. In our experiments, we show that since our features are able to capture a more abstract representation of tweets, our solution is more effective than previous ones and also more robust regarding biased and noisy data, which is the kind of data provided by these sources.

5. Classifying sentiment in microblogs: is brevity an advantage.

Microblogs as a new textual domain offer a unique proposition for sentiment analysis. Their short document length suggests any sentiment they contain is compact and explicit. However, this short length coupled with their noisy nature can pose difficulties for standard machine learning document representations. In this work we examine the hypothesis that it is easier to classify the sentiment in these short form documents than in longer form documents. Surprisingly, we find classifying sentiment in microblogs easier than in blogs and make a number of observations pertaining to the challenge of supervised learning for sentiment analysis in microblogs.

EXISTING SYSTEM

People often express their views freely on social media about what they feel about the Indian society and the politicians that claim that Indian cities are safe for women. On social media websites people can freely Express their view point and women can share their experiences where they have faced abuse harassment or where we would have fight back against the abuse harassment that was imposed on them . The tweets about safety of women and stories of standing up against abuse harassment further motivates other women data on the same social media website or application like Twitter. Other women share these messages and tweets which further motivates other 5 men or 10 women to stand up and raise a voice against people who have made Indian cities and unsafe place for the women. In the recent years a large number of people have been attracted towards social media platforms like Facebook, . It is a common practice to extract the information from the data that is available on social networking through procedures of data extraction, data analysis and data interpretation methods. The accuracy of the Twitter analysis and prediction can be obtained by the use of behavioral analysis on the basis of social networks.

DISADVANTAGES

1. Twitter and Instagram point and most of the people are using it to express their emotions and also their opinions about what they think about the Indian cities and Indian society.
2. There are several method of sentiment that can be categorized like machine learning hybrid and lexicon-based learning.
3. Also there are another categorization Janta presented with categories of statistical, knowledge-based and age wise differentiation approaches

PROPOSED SYSTEM

Women have the right to the city which means that they can gofreely whenever they want whether it be too an Educational Institute, or any other place women want to go. But women feel that they are unsafe in places like malls, shopping malls on their way to their job location because of the several unknown Eyes body shaming and harassing these women point Safety or lack of concrete consequences in the life of women isthe main reason of harassment of girls. There are instances when the harassment of girls was done by their neighbours while they were on the way to school or there was a lack of safety that created a sense of fear in the minds of small girls who throughout their lifetime suffer due to that one instance that happened in their lives where they were forced to do something unacceptable or was abusely harassed by one of their own neighbor or any other unknown person. Safest cities approach women safety from a perspective ofwomen rights to the affect the city without fear of violence or abuse harassment. Rather than imposing restrictions on women that society usually imposes it is the duty of society to imprecise the need of protection of women and also recognizes that women and girls also have a right same as men have to be safe in the City.

ADVANTAGES

1. Analysis of twitter texts collection also includes the name of people and name of women who stand up against abuse harassment and unethical behaviour of men in Indian cities which make them uncomfortable to walk freely.
2. The data set that was obtained through Twitter about the status of women safety in Indian society

FEASIBILITY STUDY

Preliminary investigation examine project feasibility, the likelihood the system will be useful to the organization. The main objective of the feasibility study is to test the Technical, Operational and Economical feasibility for adding new modules and debugging old running system. All system is feasible if they are unlimited resources and infinite time. There are aspects in the feasibility study portion of the preliminary investigation:

- Technical Feasibility
- Operational Feasibility
- Economical Feasibility

ECONOMIC FEASIBILITY

A system can be developed technically and that will be used if installed must still be a good investment for the organization. In the economical feasibility, the development cost in creating the system is evaluated against the ultimate benefit derived from the new systems. Financial benefits must equal or exceed the costs.

The system is economically feasible. It does not require any addition hardware or software. Since the interface for this system is developed using the existing resources and technologies available at NIC, There is nominal expenditure and economical feasibility for certain.

OPERATIONAL FEASIBILITY

Proposed projects are beneficial only if they can be turned out into information system. That will meet the organization's operating requirements. Operational feasibility aspects of the project are to be taken as an important part of the project implementation. Some of the important issues raised are to test the operational feasibility of a project includes the following: -

- Is there sufficient support for the management from the users?
- Will the system be used and work properly if it is being developed and implemented?
- Will there be any resistance from the user that will undermine the possible application benefits?

This system is targeted to be in accordance with the above-mentioned issues. Beforehand, the management issues and user requirements have been taken into consideration. So there is no question of resistance from the users that can undermine the possible application benefits.

The well-planned design would ensure the optimal utilization of the computer resources and would help in the improvement of performance status.

TECHNICAL FEASIBILITY

The technical issue usually raised during the feasibility stage of the investigation includes the following:

- Does the necessary technology exist to do what is suggested?
- Do the proposed equipments have the technical capacity to hold the data required to use the new system?
- Will the proposed system provide adequate response to inquiries, regardless of the number or location of users?
- Can the system be upgraded if developed?
- Are there technical guarantees of accuracy, reliability, ease of access and data security?

Earlier no system existed to cater to the needs of ‘Secure Infrastructure Implementation System’. The current system developed is technically feasible. It is a web based user interface for audit workflow at NIC-CSD. Thus it provides an easy access to the users. The database’s purpose is to create, establish and maintain a workflow among various entities in order to facilitate all concerned users in their various capacities or roles. Permission to the users would be granted based on the roles specified. Therefore, it provides the technical guarantee of accuracy, reliability and security. The software and hard requirements for the development of this project are not many and are already available in-house at NIC or are available as free as open source. The work for the project is done with the current equipment and existing software technology. Necessary bandwidth exists for providing a fast feedback to the users irrespective of the number of users using the system.

HARDWARE REQUIREMENTS:

System	:	Pentium i3/i5.
Hard Disk	:	500 GB.
Monitor	:	15'' LED
Input Devices	:	Keyboard, Mouse
Ram	:	4 GB

SOFTWARE REQUIREMENTS:

Operating system	:	Windows 8/10.
Coding Language	:	Python

What is SDLC?

SDLC stands for Software Development Life Cycle. A Software Development Life Cycle is essentially a series of steps, or phases, that provide a model for the development and lifecycle management of an application or piece of software.

SDLC is the process consisting of a series of planned activities to develop or alter the software products.

Benefits of the SDLC Process

The intent of a SDLC process it to help produce a product that is cost-efficient, effective, and of high quality. Once an application is created, the SDLC maps the proper deployment and decommissioning of the software once it becomes a legacy. The SDLC methodology usually contains the following stages: Analysis (requirements and design), construction, testing, release and maintenance (response). Veracode makes it possible to integrate automated security testing into the SDLC process through use of its cloud based platform.

1. Requirements Gathering:

   In this phase we gather all the requirements from the client, i.e. what are the client expected input, output……

2. Analysis:

   In this phase based upon the client requirements we prepare one documentation is called "High Level Design Document". It contains Abstract, Functional

   Requirements, Non Functional Requirements, Existing System, Proposed System, SRS,………

3. Design:

   It is difficult to understand the High Level Design Document for all the members, so to understand easily we use "Low Level Design Document". To design this document we use UML (Unified Modeling Language). In this we have Use case, Sequence, Collaboration……..

4. Coding:

   In this phase we develop the coding module by module. After developing all the modules we integrate them.

5. Testing:

   After developing we have to check weather client requirements are satisfied or not. If not we are again going to develop.

6. Implementation:

   In testing phase if client requirements are satisfied, we go for implementation. i.e. we need to deploy the application in some server.

7. Maintenance:

   After deployment, if at all any problems come from the client side; we are providing maintenance for that application.

Introduction to SOFTWARE VALIDATION?

The concept of validation has been in vogue for centuries. Irrespective of the industry or products, validation ensures various critical aspects of a product and guarantees its success in the market as well as among users. Likewise, software validation plays an immensely significant role during the software development life cycle (SDLC) and helps the testing and development teams to create a quality product. Therefore, in this article, we shall discuss the various aspects of software validation.

What is Software Validation?

Software Validation is a process of evaluating software product, so as to ensure that the software meets the pre-defined and specified business requirements as well as the end users/customers' demands and expectations.

It is basically, performed with the intent to check that whether the developed software is built as per pre-decided software requirement specifications (SRS) and if it caters to fulfil the customers' actual needs in the real environment.

Both, the verification and validation is a software testing activity, and verification is followed by the validation. Validation is usually carried out at the end of the software development.

FUNCTIONAL REQUIREMENTS

In software engineering and systems engineering, a **functional requirement** defines a function of a system or its component, where a function is described as a specification of behavior between outputs and inputs.[1]

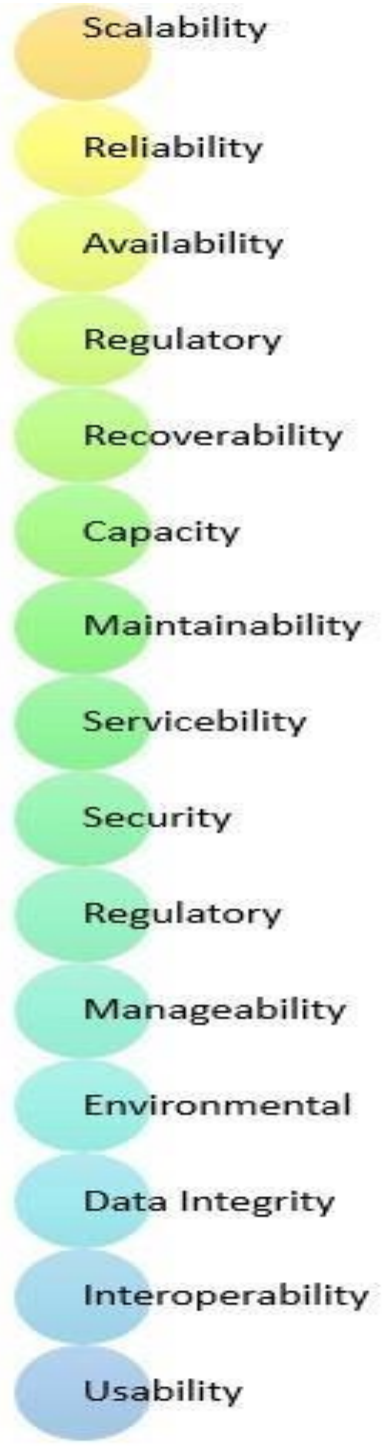
Functional requirements may involve calculations, technical details, data manipulation and processing, and other specific functionality that define what a system is supposed to accomplish.[2] Behavioral requirements describe all the cases where the system uses the functional requirements, these are captured in use cases. Functional requirements are supported by non-functional requirements (also known as "quality requirements"), which impose constraints on the design or implementation (such as performance requirements, security, or reliability). Generally, functional requirements are expressed in the form "system must do <requirement>," while non-functional requirements take the form "system shall be <requirement>." The plan for implementing functional requirements is detailed in the system design, whereas *non-functional* requirements are detailed in the system architecture.[4][5]

As defined in requirements engineering, functional requirements specify particular results of a system. This should be contrasted with non-functional requirements, which specify overall characteristics such as cost and reliability. Functional requirements drive the application architecture of a system, while non-functional requirements drive the technical architecture of a system.[4]

In some cases a requirements analyst generates use cases after gathering and validating a set of functional requirements. The hierarchy of functional requirements collection and change, broadly speaking, is: user/stakeholder request → analyze → use case → incorporate. Stakeholders make a request; systems engineers attempt to discuss, observe, and understand the aspects of the requirement; use cases, entity relationship diagrams, and other models are built to validate the requirement; and, if documented and approved, the requirement is implemented/incorporated.[6] Each use case illustrates behavioral scenarios through one or more functional requirements. Often, though, an analyst will begin by eliciting a set of use cases, from which the analyst can derive the functional requirements that must be implemented to allow a user to perform each use case.

NON-FUNCTIONAL REQUIREMENT (NFR) specifies the quality attribute of a software system. They judge the software system based on Responsiveness, Usability, Security, Portability and other non-functional standards that are critical to the success of the software system. Example of nonfunctional requirement, *"how fast does the website load?"* Failing to meet non-functional requirements can result in systems that fail to satisfy user needs.

Non-functional Requirements allows you to impose constraints or restrictions on the design of the system across the various agile backlogs. Example, the site should load in 3 seconds when the number of simultaneous users are> 10000. Description of non-functional requirements is just as critical as a functional requirement.



- Usability requirement
- Serviceability requirement
- Manageability requirement
- Recoverability requirement Security requirement

- Data Integrity requirement
- Capacity requirement
- Availability requirement
- Scalability requirement
- Interoperability requirement
- Reliability requirement
- Maintainability requirement
- Regulatory requirement
- Environmental requirement

Advantages of Non-Functional Requirement Benefits/pros of

Non-functional testing are:

- The nonfunctional requirements ensure the software system follow legal and compliance rules.
- They ensure the reliability, availability, and performance of the software system They ensure good user experience and ease of operating the software.
- They help in formulating security policy of the software system.

Disadvantages of Non-functional requirement

Cons/drawbacks of Non-function requirement are:

- None functional requirement may affect the various high-level software subsystem
- They require special consideration during the software architecture/high-level design phase which increases costs.
- Their implementation does not usually map to the specific software sub-system, It is tough to modify non-functional once you pass the architecture phase.

KEY LEARNING

- A non-functional requirement defines the performance attribute of a software system.
- Types of Non-functional requirement are Scalability Capacity, Availability, Reliability, Recoverability, Data Integrity, etc.
- Example of Non Functional Requirement is Employees never allowed to update their salary information. Such attempt should be reported to the security administrator.
- Functional Requirement is a verb while Non-Functional Requirement is an attribute
- The advantage of Non-functional requirement is that it helps you to ensure good user experience and ease of operating the software
- The biggest disadvantage of Non-functional requirement is that it may affect the various high-level software subsystems.

SYSTEM REQUIREMENTS:

HARDWARE REQUIREMENTS:

- System : Pentium i3.
- Hard Disk : 500 GB.
- Ram : 4GB.

SOFTWARE REQUIREMENTS:

- Operating system : Windows 8/10.
- Coding Language : PYTHON

ARCHITECTURE

UML DIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Metamodel and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

GOALS:

The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.

2. Provide extendibility and specialization mechanisms to extend the core concepts.

3. independent of particular programming languages and development process.

4. Provide a formal basis for understanding the modeling language.

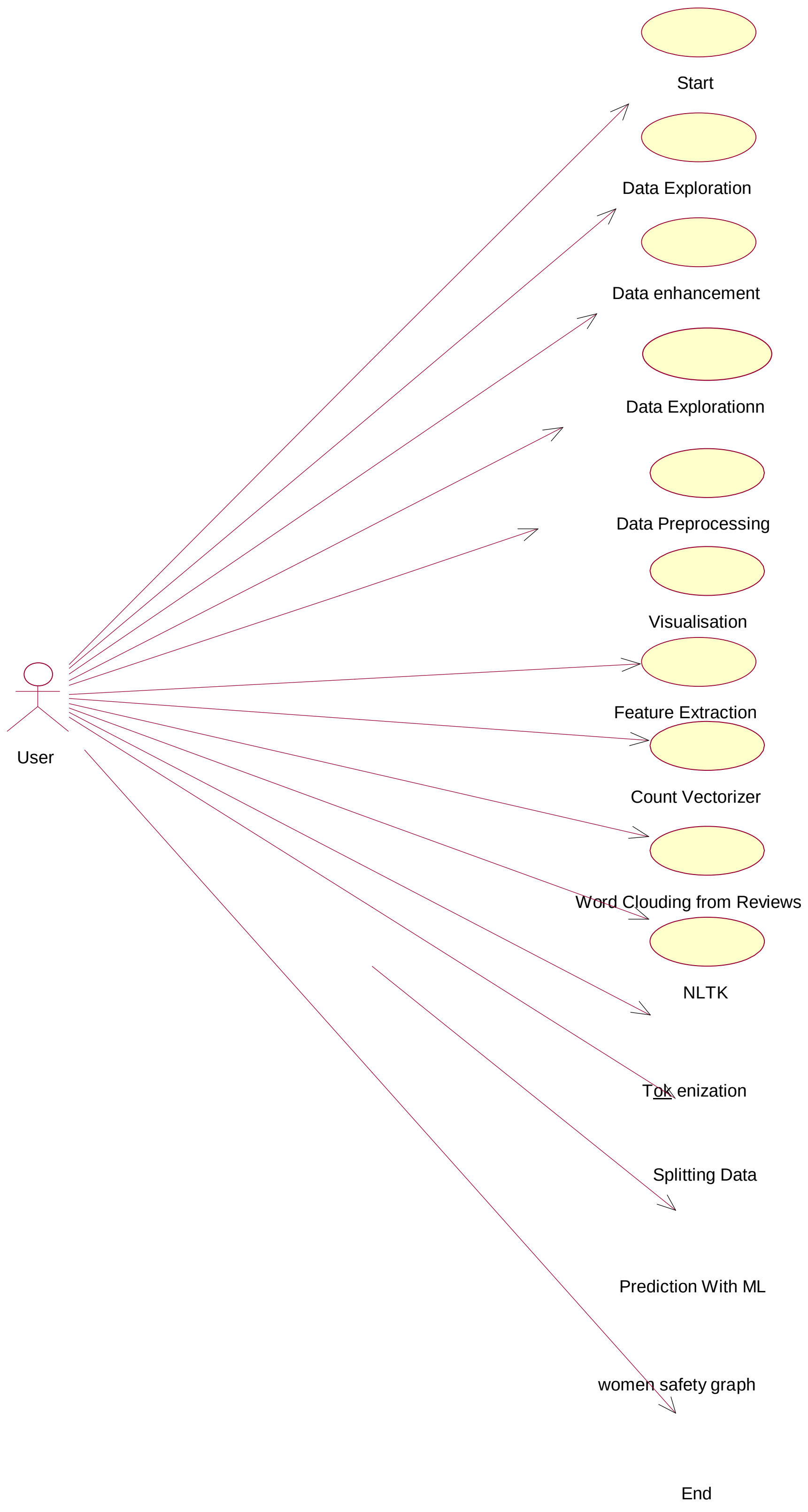
5. Encourage the growth of OO tools market.

6. Support higher level development concepts such as collaborations, frameworks, patterns and components.

7. Integrate best practices.

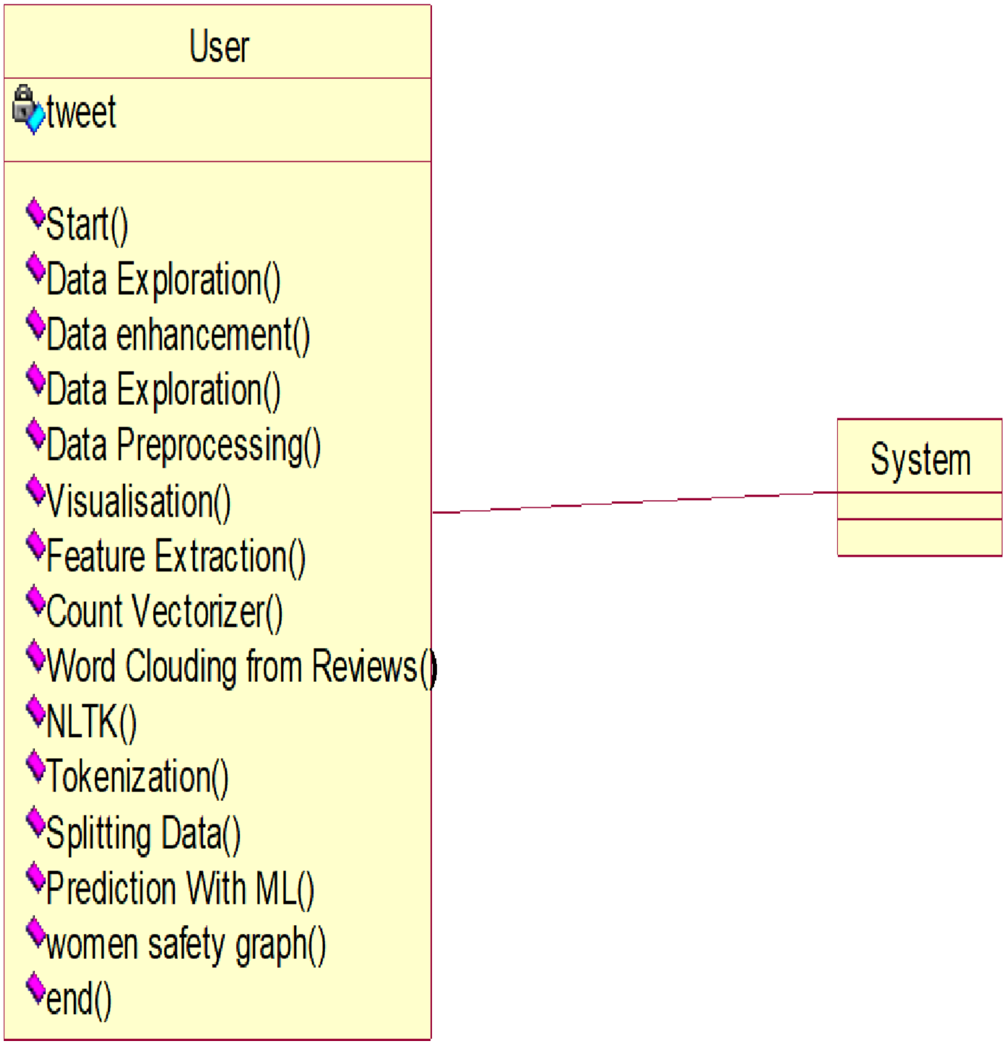
USE CASE DIAGRAM

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted


Start
Data Exploration
Data enhancement
Data Explorationn
Data Preprocessing
Visualisation
Feature Extraction
Count Vectorizer
Word Clouding from Reviews
NLTK
Tok enization
Splitting Data
Prediction With ML
women safety graph
End
User

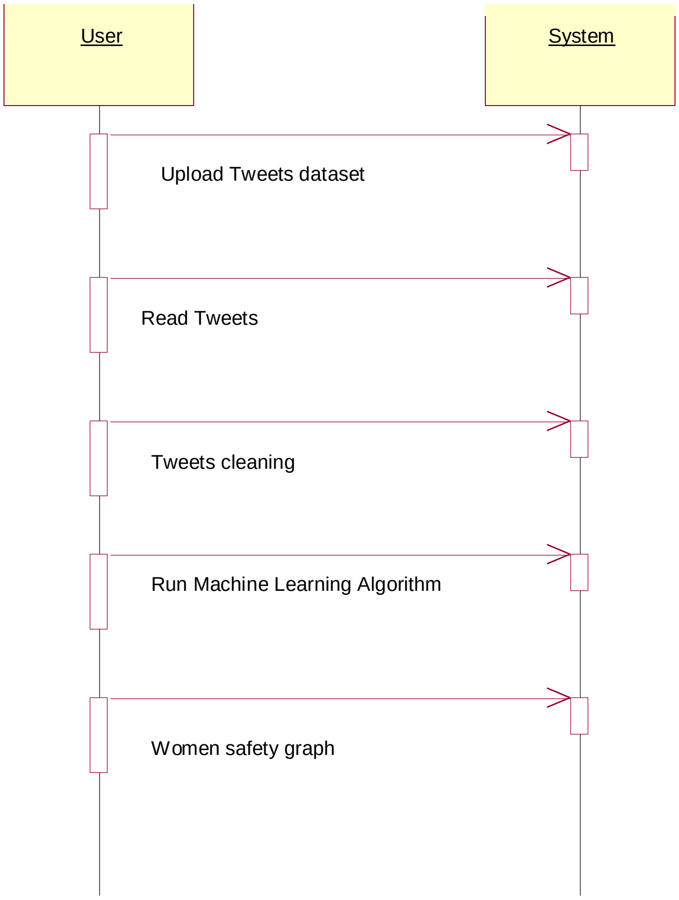
CLASS DIAGRAM:

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.



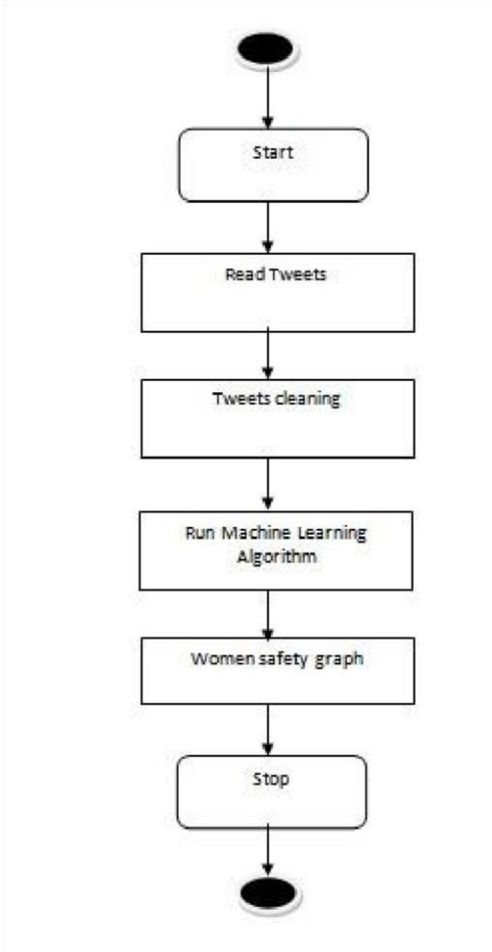
SEQUENCE DIAGRAM:

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.



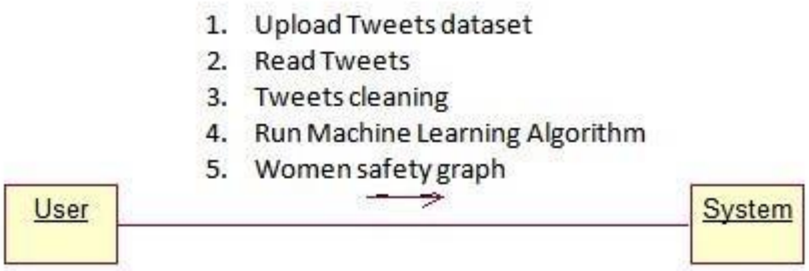
ACTIVITY DIAGRAM:

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.



COLLABORATION DIAGRAM

The collaboration diagram is used to show the relationship between the objects in a system. Both the sequence and the collaboration diagrams represent the same information but differently. Instead of showing the flow of messages, it depicts the architecture of the object residing in the system as it is based on object-oriented programming. An object consists of several features. Multiple objects present in the system are connected to each other. The collaboration diagram, which is also known as a communication diagram, is used to portray the object's architecture in the system.



Following modules to analyse women safety.

TWEEPY package from python to download tweets from twitter but every time INTERNET will not available to download tweets online so we downloaded MEETOO tweets on women safety and safe inside dataset folder. Application will read this tweets to detect women's sentiments.

NLTK (natural language tool kit) to remove special symbols and stop words from tweets and to make them clean.

TEXTBLOB corpora package and dictionary to count positive, negative and neutral polarity and the tweets which has polarity value less than 0 will consider as negative as and greater than 0 and less than 0.5 will consider as neutral and polarity greater than 0.5 will consider as positive.

Modules:

DATA COLLECTION and Upload : kaggle.com

Preprocesinng :

Applyalgorithm:

Visualization:

SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the

Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTS Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

Unit Testing:

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed

- All links should take the user to the correct page.

Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

SYSTEM TESTING

TESTING METHODOLOGIES

The following are the Testing Methodologies:

○ **Unit Testing.** ○

Integration Testing. ○

User Acceptance

Testing. ○ **Output Testing.**

○ **Validation Testing.**

Unit Testing

Unit testing focuses verification effort on the smallest unit of Software design that is the module. Unit testing exercises specific paths in a module's control structure to ensure complete coverage and maximum error detection. This test focuses on each module individually, ensuring that it functions properly as a unit. Hence, the naming is Unit Testing.

During this testing, each module is tested individually and the module interfaces are verified for the consistency with design specification. All important processing path are tested for the expected results. All error handling paths are also tested.

Integration Testing

Integration testing addresses the issues associated with the dual problems of verification and program construction. After the software has been integrated a set of high order tests are conducted. The main objective in this testing process is to take unit tested modules and builds a program structure that has been dictated by design.

The following are the types of Integration Testing:

1)Top Down Integration

This method is an incremental approach to the construction of program structure. Modules are integrated by moving downward through the control hierarchy, beginning with the main program module. The module subordinates to the main program module are incorporated into the structure in either a depth first or breadth first manner.

In this method, the software is tested from main module and individual stubs are replaced when the test proceeds downwards.

2. Bottom-up Integration

This method begins the construction and testing with the modules at the lowest level in the program structure. Since the modules are integrated from the bottom up, processing required for modules subordinate to a given level is always available and the need for stubs is eliminated.
The bottom up integration strategy may be implemented with the following steps:

- The low-level modules are combined into clusters into clusters that perform a specific Software sub-function.
- A driver (i.e.) the control program for testing is written to coordinate test case input and output.
- The cluster is tested.
- Drivers are removed and clusters are combined moving upward in the program Structure

The bottom up approaches tests each module individually and then each module is module is integrated with a main module and tested for functionality.

User Acceptance Testing

User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.

Output Testing

After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format. Asking the users about the format required by them tests the outputs generated or displayed by the system under consideration. Hence the output format is considered in 2 ways – one is on screen and another in printed format.

Validation Checking

Validation checks are performed on the following fields.

Text Field:

The text field can contain only the number of characters lesser than or equal to its size. The text fields are alphanumeric in some tables and alphabetic in other tables. Incorrect entry always flashes and error message.

Numeric Field:

The numeric field can contain only numbers from 0 to 9. An entry of any character flashes an error messages. The individual modules are checked for accuracy and what it has to perform. Each module is subjected to test run along with sample data. The individually tested modules are integrated into a single system. Testing involves executing the real data information is used in the program the existence of any program defect is inferred from the output. The testing should be planned so that all the requirements are individually tested.

A successful test is one that gives out the defects for the inappropriate data and produces and output revealing the errors in the system.

Preparation of Test Data

Taking various kinds of test data does the above testing. Preparation of test data plays a vital role in the system testing. After preparing the test data the system under study is tested using that test data. While testing the system by using test data errors are again uncovered and corrected by using above testing steps and corrections are also noted for future use.

Using Live Test Data:

Live test data are those that are actually extracted from organization files. After a system is partially constructed, programmers or analysts often ask users to key in a set of data from their normal activities. Then, the systems person uses this data as a way to partially test the system. In other instances, programmers or analysts extract a set of live data from the files and have them entered themselves.

It is difficult to obtain live data in sufficient amounts to conduct extensive testing. And, although it is realistic data that will show how the system will perform for the typical processing requirement, assuming that the live data entered are in fact typical, such data generally will not test all combinations or formats that can enter the system. This bias toward typical values then does not provide a true systems test and in fact ignores the cases most likely to cause system failure.

Using Artificial Test Data:

Artificial test data are created solely for test purposes, since they can be generated to test all combinations of formats and values. In other words, the artificial data, which can quickly be prepared by a data generating utility program in the information systems department, make possible the testing of all login and control paths through the program.

The most effective test programs use artificial test data generated by persons other than those who wrote the programs. Often, an independent team of testers formulates a testing plan, using the systems specifications.

The package "Virtual Private Network" has satisfied all the requirements specified as per software requirement specification and was accepted.

USER TRAINING

Whenever a new system is developed, user training is required to educate them about the working of the system so that it can be put to efficient use by those for whom the system has been primarily designed. For this purpose the normal working of the project was demonstrated to the prospective users. Its working is easily understandable and since the expected users are people who have good knowledge of computers, the use of this system is very easy.

MAINTAINENCE

This covers a wide range of activities including correcting code and design errors. To reduce the need for maintenance in the long run, we have more accurately defined the user's requirements during the process of system development. Depending on the requirements, this system has been developed to satisfy the needs to the largest possible extent. With development in technology, it may be possible to add many more features based on the requirements in future. The coding and designing is simple and easy to understand which will make maintenance easier.

TESTING STRATEGY :

A strategy for system testing integrates system test cases and design techniques into a well planned series of steps that results in the successful construction of software. The testing strategy must cooperate test planning, test case design, test execution, and the resultant data collection and evaluation .A strategy for software testing must accommodate low-level tests that are necessary to verify that a small source code segment has been correctly implemented as well as high level tests that validate major system functions against user requirements.

Software testing is a critical element of software quality assurance and represents the ultimate review of specification design and coding. Testing represents an interesting anomaly for the software. Thus, a series of testing are performed for the proposed system before the system is ready for user acceptance testing.

SYSTEM TESTING:

Software once validated must be combined with other system elements (e.g. Hardware, people, database). System testing verifies that all the elements are proper and that overall system function performance is achieved. It also tests to find discrepancies between the system and its original objective, current specifications and system documentation.

UNIT TESTING:

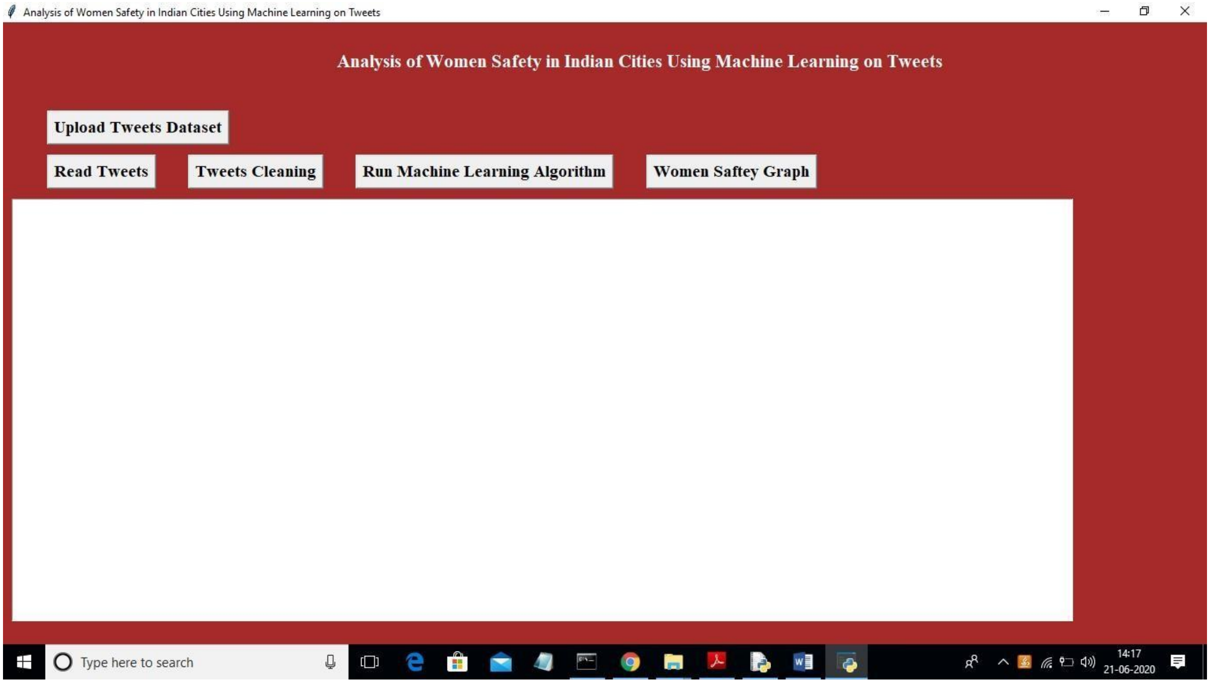
In unit testing different are modules are tested against the specifications produced during the design for the modules. Unit testing is essential for verification of the code produced during the coding phase, and hence the goals to test the internal logic of the modules. Using the detailed design description as a guide, important Conrail paths are tested to uncover errors within the boundary of the modules. This testing is carried out during the programming stage itself. In this type of testing step, each module was found to be working satisfactorily as regards to the expected output from the module.

In Due Course, latest technology advancements will be taken into consideration. As part of technical build-up many components of the networking system will be generic in nature so that future projects can either use or interact with this. The future holds a lot to offer to the development and refinement of this project.

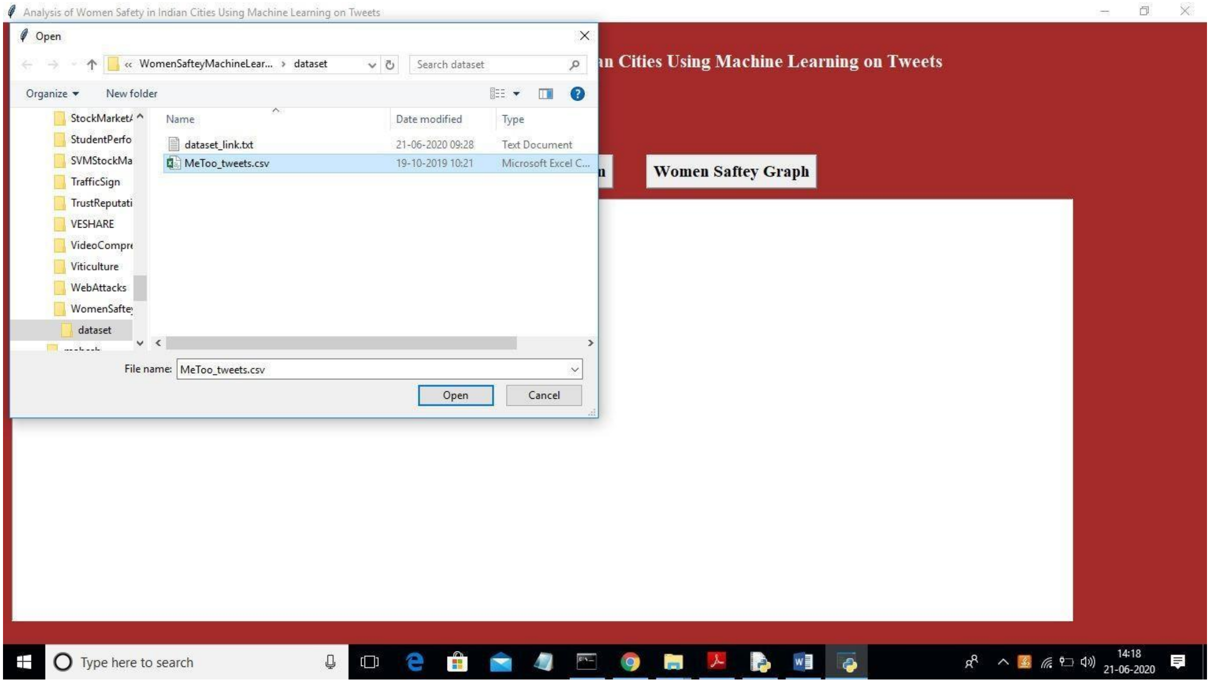
MachineLearning.py

```
import tkinter
from textblob import TextBlob
from tkinter import *
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from string import punctuation
from nltk.corpus import stopwords

main = tkinter.Tk()
main.title("Analysis of Women Safety in Indian Cities Using Machine
Learning on Tweets") #designing main screen
main.geometry("1300x1200")

global filename
tweets_list = []
clean_list = []
global pos, neu, neg

def tweetCleaning(doc):
    tokens = doc.split()
    table = str.maketrans('', '', punctuation)
    tokens = [w.translate(table) for w in tokens]
    tokens = [word for word in tokens if word.isalpha()]
    stop_words = set(stopwords.words('english'))
    tokens = [w for w in tokens if not w in stop_words]
    tokens = [word for word in tokens if len(word) > 1]
    tokens = ' '.join(tokens) #here upto for word based
    return tokens

def upload(): #function to upload tweeter profile
    global filename
    filename = filedialog.askopenfilename(initialdir="dataset")
    pathlabel.config(text=filename)
    text.delete('1.0', END)
    text.insert(END,filename+" loaded\n");

def read():
    text.delete('1.0', END)
    tweets_list.clear()
    train = pd.read_csv(filename,encoding='iso-8859-1')
    for i in range(len(train)):
        tweet = train.get_value(i, 'Text')
        tweets_list.append(tweet)
        text.insert(END,tweet+"\n")
    text.insert(END,"\n\nTotal tweets found in dataset is :
"+str(len(tweets_list))+"\n\n\n")

def clean():
    text.delete('1.0', END)
    clean_list.clear()
    for i in range(len(tweets_list)):
        tweet = tweets_list[i]
        tweet = tweet.strip("\n")
        tweet = tweet.strip()
        tweet = tweetCleaning(tweet.lower())
        clean_list.append(tweet)
        text.insert(END,tweet+"\n")
    text.insert(END,"\n\nTotal tweets found in dataset is : 
"+str(len(clean_list))+"\n\n\n")


def machineLearning():
    text.delete('1.0', END)
    global pos, neu, neg
    pos = 0
    neu = 0
    neg = 0
    for i in range(len(clean_list)):
        tweet = clean_list[i]
        blob = TextBlob(tweet)
        if blob.polarity <= 0.2:
            neg = neg + 1
            text.insert(END,tweet+"\n")
            text.insert(END,"Predicted Sentiment : NEGATIVE\n")
            text.insert(END,"Polarity Score      : 
"+str(blob.polarity)+"\n")

text.insert(END,'=========================================================
============================\n')
        if blob.polarity > 0.2 and blob.polarity <= 0.5:
            neu = neu + 1
            text.insert(END,tweet+"\n")
            text.insert(END,"Predicted Sentiment : NEUTRAL\n")
            text.insert(END,"Polarity Score      : 
"+str(blob.polarity)+"\n")

text.insert(END,'=========================================================
============================\n')
        if blob.polarity > 0.5:
            pos = pos + 1
            text.insert(END,tweet+"\n")
            text.insert(END,"Predicted Sentiment : POSITIVE\n")
            text.insert(END,"Polarity Score      : 
"+str(blob.polarity)+"\n")

text.insert(END,'=========================================================
============================\n')

def graph():
    label_X = []
    category_X = []
    text.delete('1.0', END)
    text.insert(END,"Saftey Factor\n\n")
    text.insert(END,'Positive : '+str(pos)+"\n")
    text.insert(END,'Negative : '+str(neg)+"\n")
    text.insert(END,'Neutral  : '+str(neu)+"\n\n")
    text.insert(END,'Length of tweets  : '+str(len(clean_list))+"\n")
    text.insert(END,'Positive : '+str(pos)+' / '+ str(len(clean_list))+'
= '+str(pos/len(clean_list))+'%\n')
    text.insert(END,'Negative : '+str(neg)+' / '+ str(len(clean_list))+'
= '+str(neg/len(clean_list))+'%\n')
    text.insert(END,'Neutral  : '+str(neu)+' / '+ str(len(clean_list))+'
= '+str(neu/len(clean_list))+'%\n')
    label_X.append('Positive')
    label_X.append('Negative')
    label_X.append('Neutral')
    category_X.append(pos)
    category_X.append(neg)
    category_X.append(neu)

    plt.pie(category_X,labels=label_X,autopct='%1.1f%%')
    plt.title('Women Saftey & Sentiment Graph')
    plt.axis('equal')
    plt.show()

font = ('times', 16, 'bold')
title = Label(main, text='Analysis of Women Safety in Indian Cities Using
Machine Learning on Tweets')
title.config(bg='brown', fg='white')
title.config(font=font)
title.config(height=3, width=120)
title.place(x=0,y=5)

font1 = ('times', 14, 'bold')
uploadButton = Button(main, text="Upload Tweets Dataset", command=upload)
uploadButton.place(x=50,y=100)
uploadButton.config(font=font1)

pathlabel = Label(main)
pathlabel.config(bg='brown', fg='white')
pathlabel.config(font=font1)
pathlabel.place(x=370,y=100)

readButton = Button(main, text="Read Tweets", command=read)
readButton.place(x=50,y=150)
readButton.config(font=font1)

cleanButton = Button(main, text="Tweets Cleaning", command=clean)
cleanButton.place(x=210,y=150)
cleanButton.config(font=font1)

mlButton = Button(main, text="Run Machine Learning Algorithm",
command=machineLearning)
mlButton.place(x=400,y=150)
mlButton.config(font=font1)

graphButton = Button(main, text="Women Saftey Graph", command=graph)
graphButton.place(x=730,y=150)
graphButton.config(font=font1)
```

```
font1 = ('times', 12, 'bold')
text=Text(main,height=25,width=150)
scroll=Scrollbar(text)
text.configure(yscrollcommand=scroll.set)
text.place(x=10,y=200)
text.config(font=font1)


main.config(bg='brown')
main.mainloop()
```


Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets
Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets
Upload Tweets Dataset
Read Tweets
Tweets Cleaning
Run Machine Learning Algorithm
Women Saftey Graph
Type here to search
14:17
21-06-2020

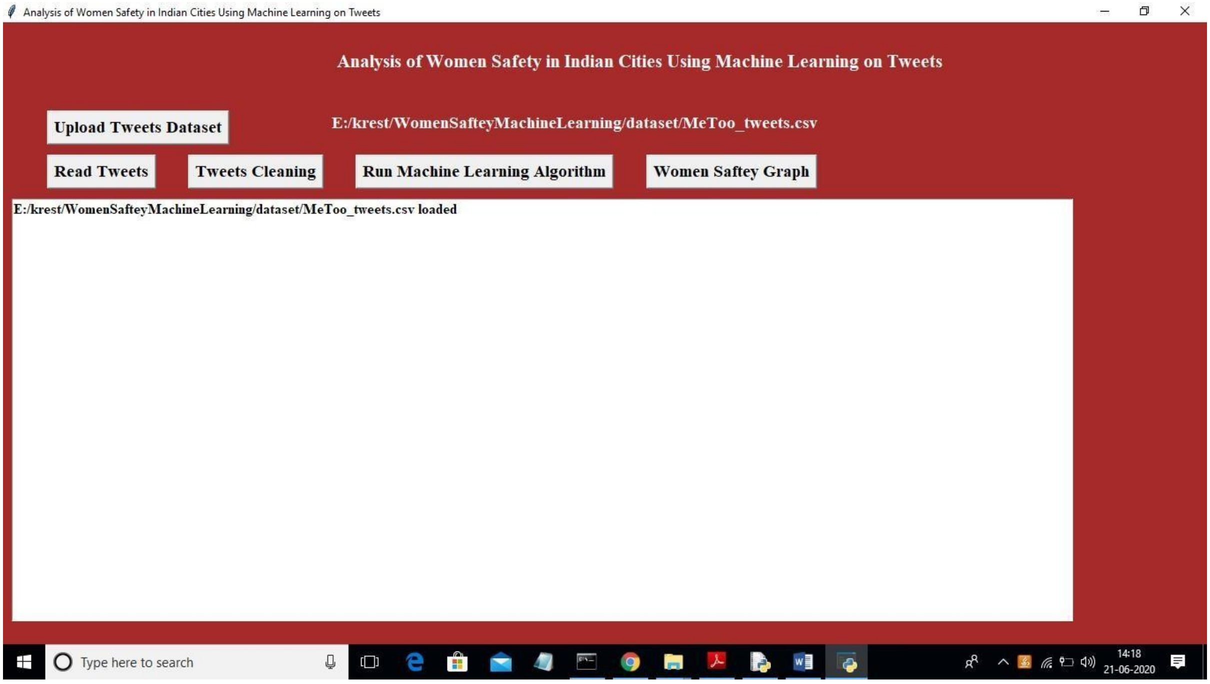
In above screen click on ‘Upload Tweets Dataset’ button and upload tweets



Name	Date modified	Type
dataset_link.txt	21-06-2020 09:28	Text Document
MeToo_tweets.csv	19-10-2019 10:21	Microsoft Excel C...

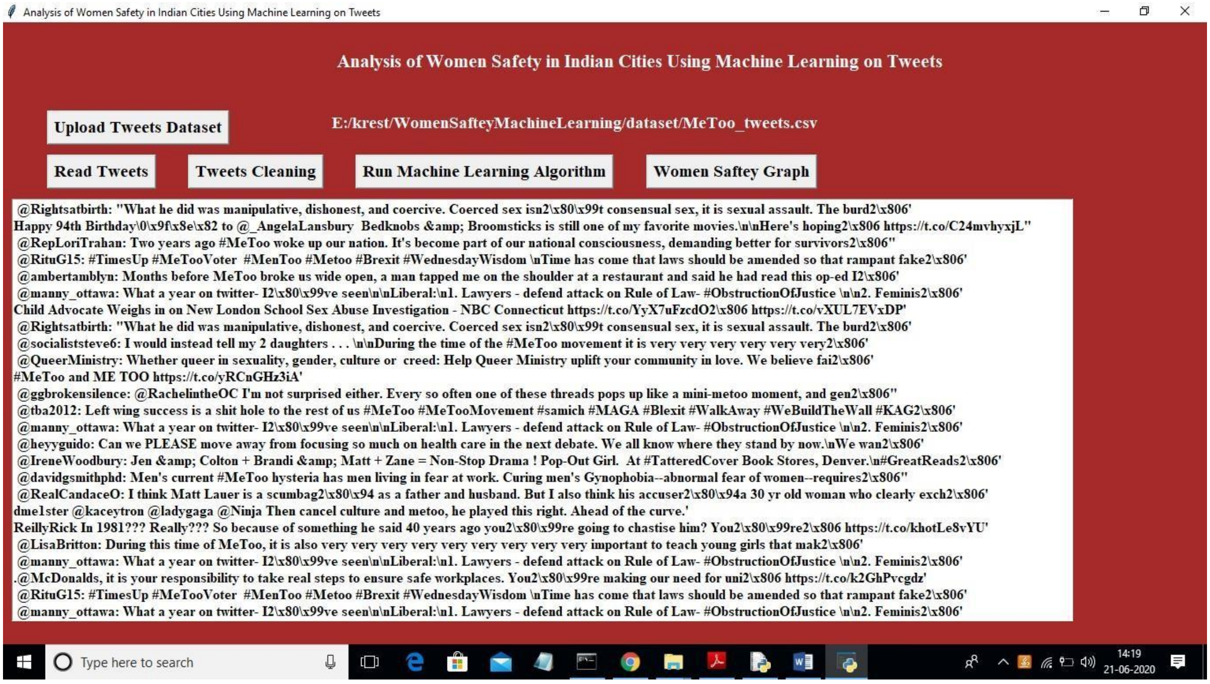
File name: MeToo_tweets.csv

In above screen uploading MeeToo_tweets.csv file and then click on ‘Open’ button to load dataset and to get below screen


Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets
Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets
Upload Tweets Dataset
E:/krest/WomenSafteyMachineLearning/dataset/MeToo_tweets.csv
Read Tweets
Tweets Cleaning
Run Machine Learning Algorithm
Women Saftey Graph
Type here to search
14:18
21-06-2020

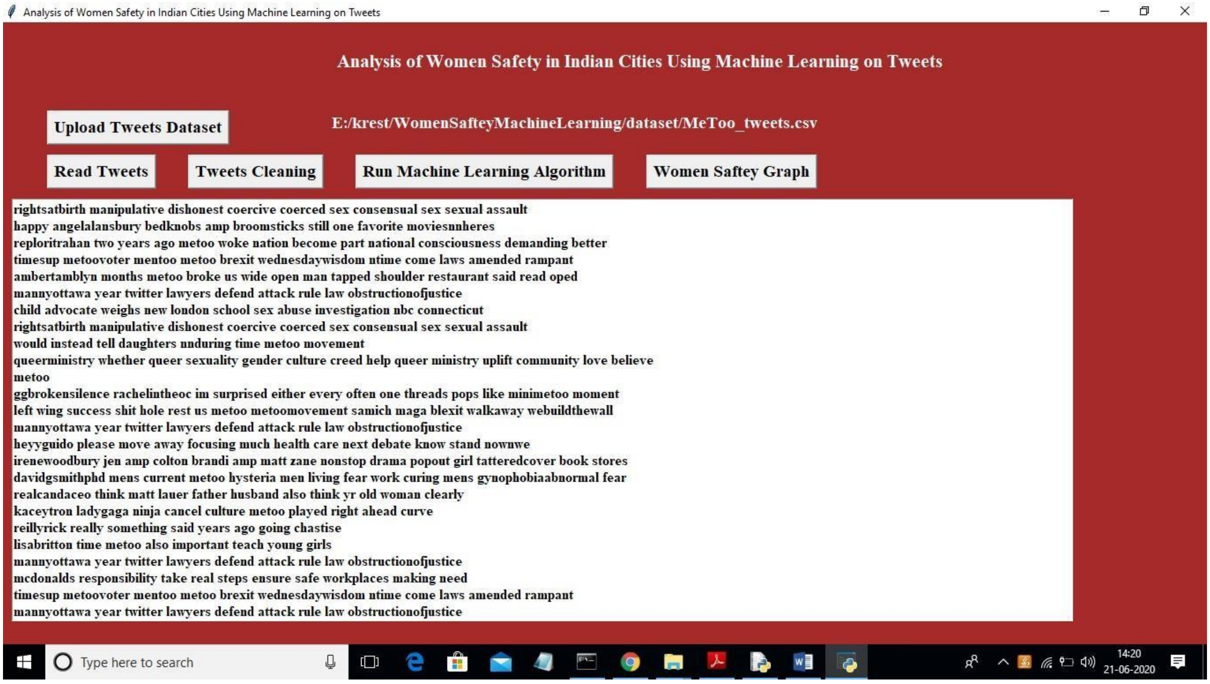
```
E:/krest/WomenSafteyMachineLearning/dataset/MeToo_tweets.csv loaded
```

In above screen tweets dataset loaded and now click on 'Read Tweets' button to read tweets from dataset


Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets
Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets
Upload Tweets Dataset
E:/krest/WomenSafteyMachineLearning/dataset/MeToo_tweets.csv
Read Tweets
Tweets Cleaning
Run Machine Learning Algorithm
Women Saftey Graph
Type here to search
14:19
21-06-2020

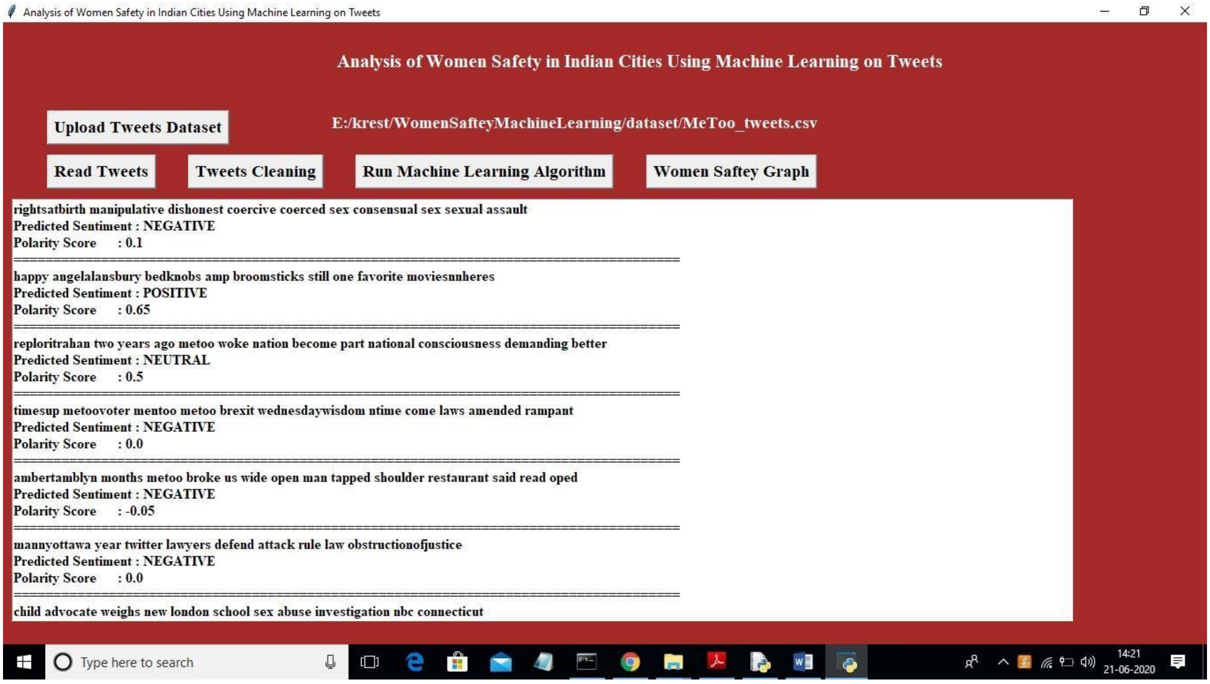
```
 @Rightsatbirth: "What he did was manipulative, dishonest, and coercive. Coerced sex isn2\x80\x99t consensual sex, it is sexual assault. The burd2\x806'
Happy 94th Birthday\0\x9f\x8e\x82 to @_AngelaLansbury  Bedknobs &amp; Broomsticks is still one of my favorite movies.\n\nHere's hoping2\x806 https://t.co/C24mvhyxjL"
 @RepLoriTrahan: Two years ago #MeToo woke up our nation. It's become part of our national consciousness, demanding better for survivors2\x806"
 @RituG15: #TimesUp #MeTooVoter  #MenToo #Metoo #Brexit #WednesdayWisdom \nTime has come that laws should be amended so that rampant fake2\x806'
 @ambertamblyn: Months before MeToo broke us wide open, a man tapped me on the shoulder at a restaurant and said he had read this op-ed I2\x806'
 @manny_ottawa: What a year on twitter- I2\x80\x99ve seen\n\nLiberal:\n1. Lawyers - defend attack on Rule of Law- #ObstructionOfJustice \n\n2. Feminis2\x806'
Child Advocate Weighs in on New London School Sex Abuse Investigation - NBC Connecticut https://t.co/YyX7uFzcdO2\x806 https://t.co/vXUL7EVxDP'
 @Rightsatbirth: "What he did was manipulative, dishonest, and coercive. Coerced sex isn2\x80\x99t consensual sex, it is sexual assault. The burd2\x806'
 @socialiststeve6: I would instead tell my 2 daughters . . . \n\nDuring the time of the #MeToo movement it is very very very very very very2\x806'
 @QueerMinistry: Whether queer in sexuality, gender, culture or  creed: Help Queer Ministry uplift your community in love. We believe fai2\x806'
#MeToo and ME TOO https://t.co/yRCnGHz3iA'
 @ggbrokensilence: @RachelintheOC I'm not surprised either. Every so often one of these threads pops up like a mini-metoo moment, and gen2\x806"
 @tba2012: Left wing success is a shit hole to the rest of us #MeToo #MeTooMovement #samich #MAGA #Blexit #WalkAway #WeBuildTheWall #KAG2\x806'
 @manny_ottawa: What a year on twitter- I2\x80\x99ve seen\n\nLiberal:\n1. Lawyers - defend attack on Rule of Law- #ObstructionOfJustice \n\n2. Feminis2\x806'
 @heyyguido: Can we PLEASE move away from focusing so much on health care in the next debate. We all know where they stand by now.\nWe wan2\x806'
 @IreneWoodbury: Jen &amp; Colton + Brandi &amp; Matt + Zane = Non-Stop Drama ! Pop-Out Girl.  At #TatteredCover Book Stores, Denver.\n#GreatReads2\x806'
 @davidgsmithphd: Men's current #MeToo hysteria has men living in fear at work. Curing men's Gynophobia--abnormal fear of women--requires2\x806"
 @RealCandaceO: I think Matt Lauer is a scumbag2\x80\x94 as a father and husband. But I also think his accuser2\x80\x94a 30 yr old woman who clearly exch2\x806'
dme1ster @kaceytron @ladygaga @Ninja Then cancel culture and metoo, he played this right. Ahead of the curve.'
ReillyRick In 1981??? Really??? So because of something he said 40 years ago you2\x80\x99re going to chastise him? You2\x80\x99re2\x806 https://t.co/khotLe8vYU'
 @LisaBritton: During this time of MeToo, it is also very very very very very very very very very important to teach young girls that mak2\x806'
 @manny_ottawa: What a year on twitter- I2\x80\x99ve seen\n\nLiberal:\n1. Lawyers - defend attack on Rule of Law- #ObstructionOfJustice \n\n2. Feminis2\x806'
.@McDonalds, it is your responsibility to take real steps to ensure safe workplaces. You2\x80\x99re making our need for uni2\x806 https://t.co/k2GhPvcgdz'
 @RituG15: #TimesUp #MeTooVoter  #MenToo #Metoo #Brexit #WednesdayWisdom \nTime has come that laws should be amended so that rampant fake2\x806'
 @manny_ottawa: What a year on twitter- I2\x80\x99ve seen\n\nLiberal:\n1. Lawyers - defend attack on Rule of Law- #ObstructionOfJustice \n\n2. Feminis2\x806'
```

In above screen each line represents one tweet and you can scroll down above screen text area to view all tweets. In above screen we can see all tweets contains special symbols and stop words and to clean those tweets click on 'Tweets Cleaning' button


Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets
Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets
Upload Tweets Dataset
E:/krest/WomenSafteyMachineLearning/dataset/MeToo_tweets.csv
Read Tweets
Tweets Cleaning
Run Machine Learning Algorithm
Women Saftey Graph
Type here to search
14:20
21-06-2020

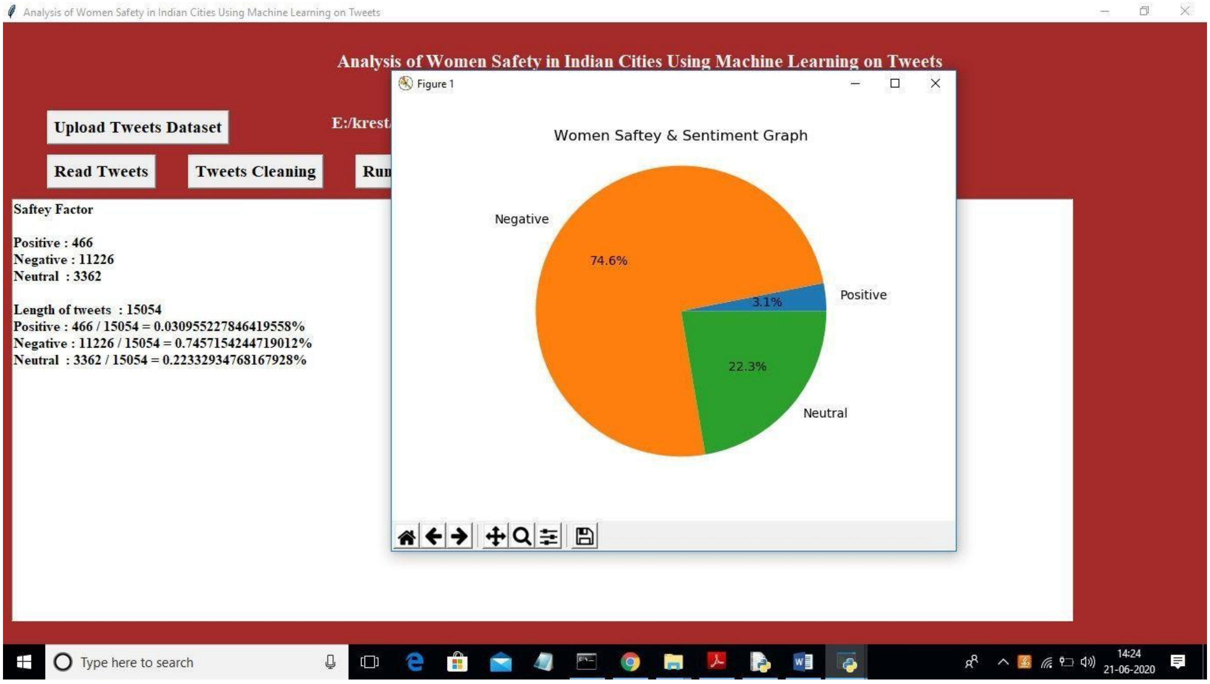
```
rightsatbirth manipulative dishonest coercive coerced sex consensual sex sexual assault
happy angelalansbury bedknobs amp broomsticks still one favorite moviesnnheres
reploritrahan two years ago metoo woke nation become part national consciousness demanding better
timesup metoovoter mentoo metoo brexit wednesdaywisdom ntime come laws amended rampant
ambertamblyn months metoo broke us wide open man tapped shoulder restaurant said read oped
mannyottawa year twitter lawyers defend attack rule law obstructionofjustice
child advocate weighs new london school sex abuse investigation nbc connecticut
rightsatbirth manipulative dishonest coercive coerced sex consensual sex sexual assault
would instead tell daughters nnduring time metoo movement
queerministry whether queer sexuality gender culture creed help queer ministry uplift community love believe
metoo
ggbrokensilence rachelintheoc im surprised either every often one threads pops like minimetoo moment
left wing success shit hole rest us metoo metoomovement samich maga blexit walkaway webuildthewall
mannyottawa year twitter lawyers defend attack rule law obstructionofjustice
heyyguido please move away focusing much health care next debate know stand nownwe
irenewoodbury jen amp colton brandi amp matt zane nonstop drama popout girl tatteredcover book stores
davidgsmithphd mens current metoo hysteria men living fear work curing mens gynophobiaabnormal fear
realcandaceo think matt lauer father husband also think yr old woman clearly
kaceytron ladygaga ninja cancel culture metoo played right ahead curve
reillyrick really something said years ago going chastise
lisabritton time metoo also important teach young girls
mannyottawa year twitter lawyers defend attack rule law obstructionofjustice
mcdonalds responsibility take real steps ensure safe workplaces making need
timesup metoovoter mentoo metoo brexit wednesdaywisdom ntime come laws amended rampant
mannyottawa year twitter lawyers defend attack rule law obstructionofjustice
```

In above screen we can see all special symbols and stop words remove from tweets and only clean words are there and now click on 'Run Machine Learning Algorithm' button to predict sentiments from tweets


Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets
Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets
Upload Tweets Dataset
E:/krest/WomenSafteyMachineLearning/dataset/MeToo_tweets.csv
Read Tweets
Tweets Cleaning
Run Machine Learning Algorithm
Women Saftey Graph
Type here to search
14:21
21-06-2020

```
rightsatbirth manipulative dishonest coercive coerced sex consensual sex sexual assault
Predicted Sentiment : NEGATIVE
Polarity Score     : 0.1
=====================================================================================
happy angelalansbury bedknobs amp broomsticks still one favorite moviesnnheres
Predicted Sentiment : POSITIVE
Polarity Score     : 0.65
=====================================================================================
reploritrahan two years ago metoo woke nation become part national consciousness demanding better
Predicted Sentiment : NEUTRAL
Polarity Score     : 0.5
=====================================================================================
timesup metoovoter mentoo metoo brexit wednesdaywisdom ntime come laws amended rampant
Predicted Sentiment : NEGATIVE
Polarity Score     : 0.0
=====================================================================================
ambertamblyn months metoo broke us wide open man tapped shoulder restaurant said read oped
Predicted Sentiment : NEGATIVE
Polarity Score     : -0.05
=====================================================================================
mannyottawa year twitter lawyers defend attack rule law obstructionofjustice
Predicted Sentiment : NEGATIVE
Polarity Score     : 0.0
=====================================================================================
child advocate weighs new london school sex abuse investigation nbc connecticut
```

In above screen each tweet having tweet text and then displaying tweets sentiments with polarity score. Scroll down above text area to see all tweets. Now click on 'Women Saftey Graph' button to get below results and by seeing that result user can easily understand whether area is safe or not. If area is safe then more peoples will express either positive or neutral tweets and if not safe then more peoples will discuss negative tweets.


Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets
Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets
Upload Tweets Dataset
E:/krest
Read Tweets
Tweets Cleaning
Run
Figure 1
Women Saftey & Sentiment Graph
Negative
74.6%
Positive
3.1%
22.3%
Neutral
Type here to search
14:24
21-06-2020

```
Saftey Factor

Positive : 466
Negative : 11226
Neutral  : 3362

Length of tweets  : 15054
Positive : 466 / 15054 = 0.030955227846419558%
Negative : 11226 / 15054 = 0.7457154244719012%
Neutral  : 3362 / 15054 = 0.22332934768167928%
```

In above screen 0.74 multiply by 100 will give 74% which means 74% peoples are talking negative and area is not safe and only 22 and 3% peoples are talking positive and neutral.

CONCLUSIONS

Throughout the research paper we have discussed about various machine learning algorithms that can help us to organize and analyze the huge amount of Twitter data obtained including millions of tweets and text messages shared every day. These machine learning algorithms are very effective and useful when it comes to analyzing of large amount of data including the SPC algorithm and linear algebraic Factor Model approaches which help to further categorize the data into meaningful groups. Support vector machines is yet another form of machine learning algorithm that is very popular in extracting Useful information from the Twitter and get an idea about the status of women safety in Indian cities.

REFERENCES

[1] Agarwal, Apoorv, Fadi Biadsy, and Kathleen R. Mckeown. "Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams." Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2009.

[2] Barbosa, Luciano, and Junlan Feng. "Robust sentiment detection on twitter from biased and noisy data." Proceedings of the 23rd international conference on computational linguistics: posters. Association for Computational Linguistics, 2010.

[3] Bermingham, Adam, and Alan F. Smeaton. "Classifying sentiment in microblogs: is brevity an advantage?." Proceedings of the 19th ACM international conference on Information and knowledge management. ACM, 2010.

[4] Gamon, Michael. "Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis." Proceedings of the 20th international conference on Computational Linguistics. Association for Computational Linguistics, 2004.

[5] Kim, Soo-Min, and Eduard Hovy. "Determining the sentiment of opinions." Proceedings of the 20th international conference on Computational Linguistics. Association for Computational Linguistics, 2004.

[6] Klein, Dan, and Christopher D. Manning. "Accurate unlexicalized parsing." Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1. Association for Computational Linguistics, 2003..

[7] Charniak, Eugene, and Mark Johnson. "Coarse-to-fine n-best parsing and MaxEnt discriminative reranking." Proceedings of the 43rd annual meeting on association for computational linguistics. Association for Computational Linguistics, 2005.

[8] Gupta, B., Negi, M., Vishwakarma, K., Rawat, G., & Badhani, P. (2017). Study of Twitter sentiment analysis using machine learning algorithms on Python. International Journal of Computer Applications, 165(9), 0975-8887.

[9] Sahayak, V., Shete, V., & Pathan, A. (2015). Sentiment analysis on twitter data. International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2(1),

178- 183.

[10] Mamgain, N., Mehta, E., Mittal, A., & Bhatt, G. (2016, March). Sentiment analysis of top colleges in India using Twitter data. In Computational Techniques in Information and Communication Technologies (ICCTICT), 2016 International Conference on (pp. 525-530). IEEE.