ANALYSIS OF WOMEN SAFETY IN INDIAN CITIES USING MACHINE LEARNING ON TWEETS

K. CHITRAANJAN
PG SCHOLAR, Dept of CSE
SWARNANDHRA COLLEGE OF ENGINEERING AND TECHNOLOGY
Dr BOMMA RAMAKRISHNA
Professor, Dept of CSE
SWARNANDHRA COLLEGE OF ENGINEERING AND TECHNOLOGY

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.

INTRODUCTION

There are certain types of harassment and Violence that arevery 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

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 [1]. On social media websites people can freely Express their view point and women can share their experiences where they have faced sexual harassment or where we would have fight back against the sexual harassment that was imposed on them[2] . The tweets about safety of women and stories of standing up against sexual 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 likeFacebook, 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. There are several method of sentiment that can be categorized like machine learning hybrid and lexicon-based learning. [5] Also there are another categorization Janta presented with categories of statistical, knowledge-based and age wise differentiation approaches. 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 ofbehavioural analysis on the basis of social networks.

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 ofpeople and name of women who stand up against abuseharassment and unethical behaviour of men in Indian citieswhich make them uncomfortable to walk freely.

2.The data setthat was obtained through Twitter about the status of womensafety in Indian society
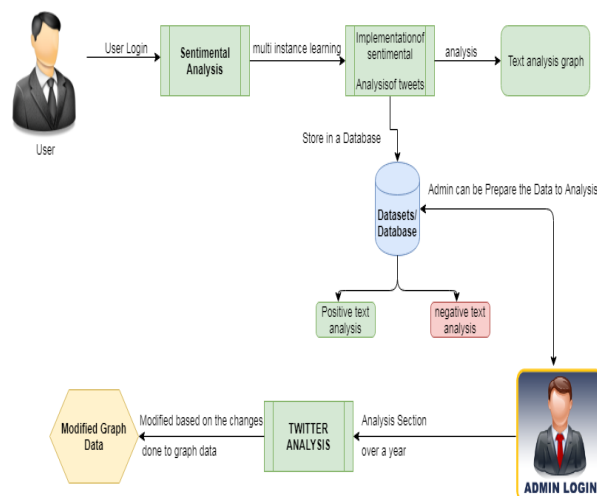
SYSTEM ARCHITECTURE

Fig 1: System Architecture

TWITTER ANALYSIS

People communicate and share their opinion actively onsocial medias including Facebook and Twitter, Social network can be considered as a perfect platform to learn about people's opinion and sentiments regarding different events. There exists several opinion-oriented information gathering and analytics systems that aim to extract people's opinion regarding different topics.

IMPLEMENTATION OF SENTIMENTAL ANALYSIS OF TWEETS

Report the tweets picked upfrom Twitter API provided by Twitter itself. Due to the presence of Twitter API, there are many techniques available for sentimental analysis of data on Social media. In this project a set of available libraries has been used.

GRAPH

A Depressed interaction graph $G\_$ is generated via some social graph model,minimizing the distance between the real and Depressed interaction graphs.An *interaction graph G* is extracted from the input (real) social media data. An interactiongraph represents how social network actors interact witheach other [25], [26]. Entities and their interactions in socialmedia are identified, and an interaction graph is built with avertex set V, including entities, an edge set E representinginteractions, and an attribute set A, which includes bothvertex (entity) attributes and edge (interaction) attributes

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:aMeta-model 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 projec

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 actorsinthe system can be depicted.
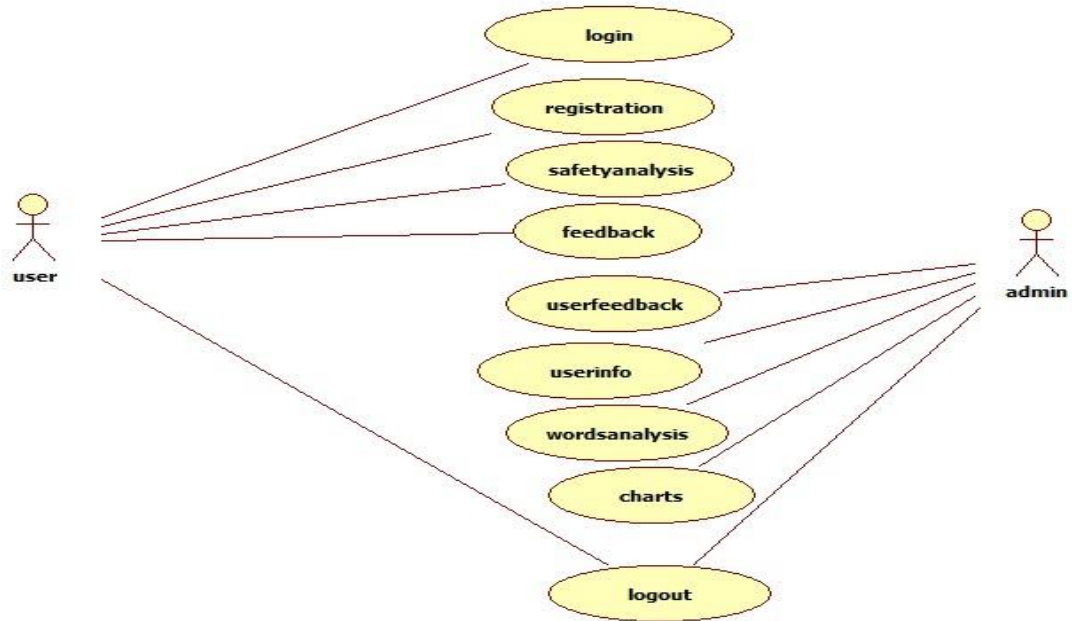
Fig 2: Use case Diagram

Class diagram:

In software engineering, a class diagram in the Unified Modeling Language (UML) is atype 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.
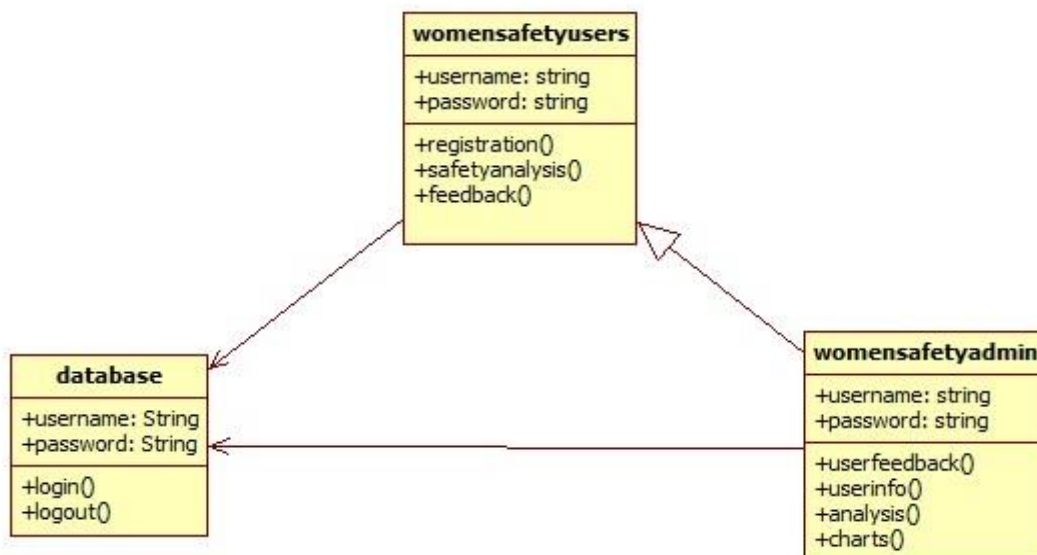

Fig 3:Class Diagram

Sequence diagram:

A sequence diagram in Unified Modelling 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
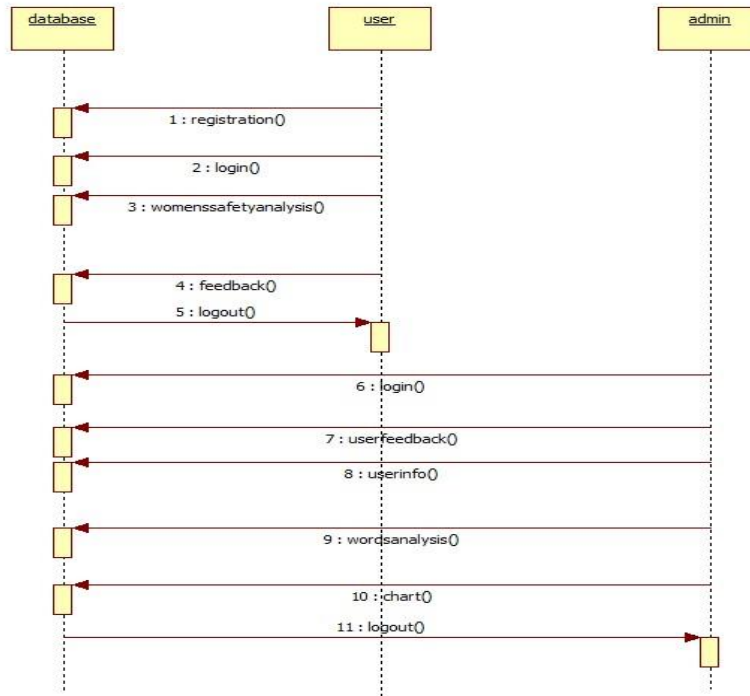

Fig 4: Sequence diagram

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
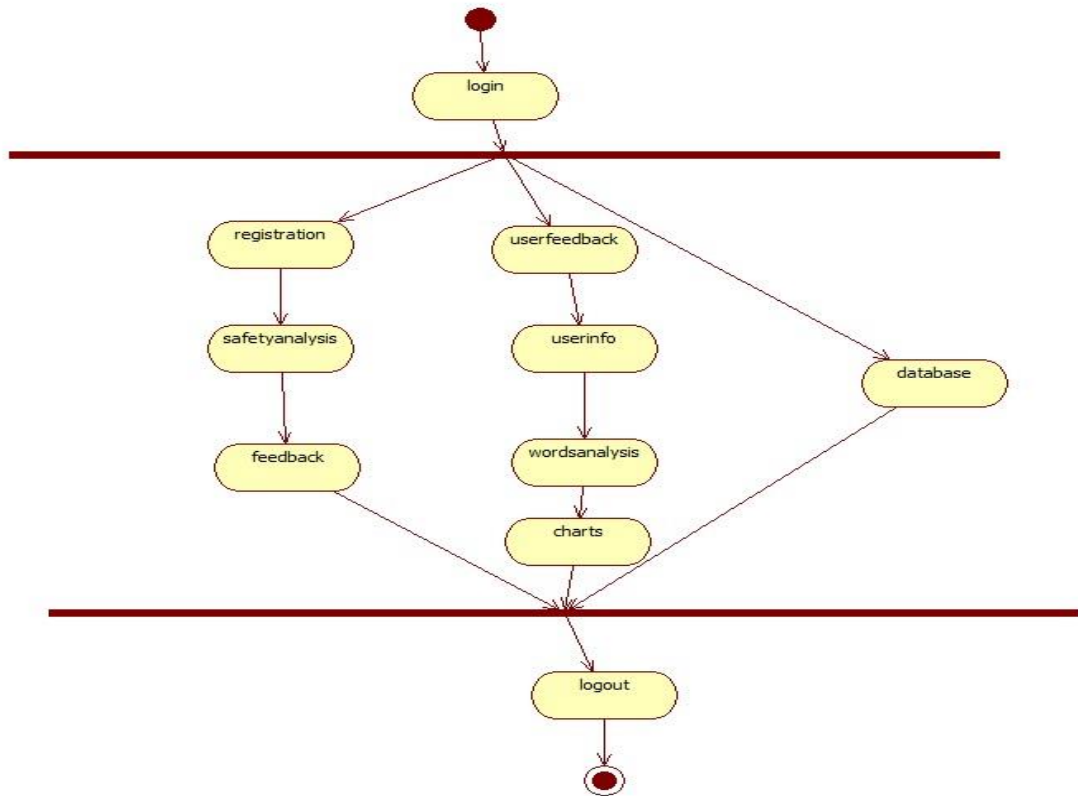
Fig 5: Activity Diagram

CONCLUSION

Throughout the research paper we have discussed aboutvarious 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 dataincluding 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 affectscoring and syntactic n-grams." Proceedings of the 12thConference of the European Chapter of the Association forComputational Linguistics. Association for ComputationalLinguistics, 2009.

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

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

[4]Gamon, Michael. "Sentiment classification on customerfeedback data: noisy data, large feature vectors, and the role oflinguistic analysis." Proceedings of the 20th internationalconference on Computational Linguistics. Association forComputational Linguistics, 2004.

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

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

[7] Charniak, Eugene, and Mark Johnson. "Coarse-to-fine n-bestparsing and MaxEntdiscriminative reranking." Proceedings ofthe 43rd annual meeting on association forcomputationallinguistics. Association for Computational Linguistics, 2005.

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

[9] Sahayak, V., Shete, V., & Pathan, A. (2015). Sentimentanalysis on twitter data. International Journal of InnovativeResearch 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 usingTwitter data. In Computational Techniques in Information and Communication Technologies (ICCTICT), 2016 InternationalConference on (pp. 525-530). IEEE.