

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

<sup>1</sup> A.Sudhawali, M.Tech, Assistant Professor, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.

<sup>2</sup> Sri Harshitha Buddam, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.

<sup>3</sup> Nageswara Sarma Veerubotla, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.

<sup>4</sup> Pavan Sai Perni, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.

<sup>5</sup> Saran Ganesh Kasagani, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.

**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 focuses on the role of social media in promoting women's safety in Indian cities with special reference to the role of social media websites and applications including Twitter platforms Face book and Instagram. This paper also focuses on how a sense of responsibility on the part of Indian society can be developed by the common Indian people so that we should focus on the safety of women surrounding them. Tweets on Twitter which usually contain images and text and also written messages and quotes that 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 they go out for work or travel on public transport and what the state of their mind when they are surrounded by unknown men and whether these women feel safe or not?

## 1. INTRODUCTION

Twitter in this modern era has emerged as the ultimate microblogging social network consisting of over a hundred million users and generating over five hundred million messages known as 'Tweets' every day. Twitter with such a massive audience has magnetized users to emit their perspectives and judgemental about every existing issue and topic of the internet, therefore Twitter is an informative source for all the zones like institutions, companies, and organizations. On Twitter, users will share their opinions and perspectives in the tweets section. This tweet can only contain 140 characters, thus making the users compact their messages with the help of abbreviations, slang, short forms, emoticons, etc. In addition to this, many people express their opinions by using polysemy and sarcasm. Hence twitter language can be termed as unstructured. From the tweet, the sentiment behind the message is extracted. This extraction is done by using the sentimental analysis procedure. Results of the sentimental analysis can be used in many areas like sentiments regarding a particular brand or release of a product, analyzing public opinions on government policies, people's thoughts on women, etc. To perform the classification of tweets and analyze the outcome, a lot of studies have been done on the data obtained by Twitter. We also review some studies on machine learning in this paper and research how to perform sentimental analysis using that domain on Twitter data. The paper's scope is restricted to machine learning algorithms and models. Staring at women and passing comments can be certain types of violence and harassment and these

unacceptable practices are usually normal, especially on the part of urban life. Many researches that have been conducted in India show that women have reported sexual harassment and other practices as stated above. Such studies have also shown that in popular metropolitan cities like Delhi, Pune, Chennai, and Mumbai, most women feel they are unsafe when surrounded by unknown people. On social media, people can freely express what they feel about Indian politics, society, and many other thoughts. Similarly, women can also share their experiences if they have faced any violence or sexual harassment and this brings innocent people together to stand up against such incidents. The analysis of tweets text collection obtained by Twitter includes names of people who have harassed women and also names of women or innocent people who have stood against such violent acts or unethical behavior of men thus making them uncomfortable to walk freely in public.

## 2. LITERATURE SURVEY

2.1 Agarwal, Aproov, 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 were proposed a classifier to predict the contextual polarity of subjective phrases in a sentence. Our approach features lexical scoring derived from the Dictionary of Affect in Language (DAL) and extended through WordNet, allowing us to automatically score the vast majority of

words in our input avoiding the need for manual labeling. We augment lexical scoring with n-gram analysis to capture the effect of context. We combine DAL scores with syntactic constituents and then extract ngrams of constituents from all sentences. We also use the polarity of all syntactic constituents within the sentence as features. Our results show significant improvement over a majority class baseline as well as a more difficult baseline consisting of lexical n-grams.

2.2 D. Kumar and S. Aggarwal, "Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets", *2019 Amity International Conference on Artificial Intelligence (AICAI)* were proposed the safety analysis and monitoring of women using various social media platforms in Indian cities. The posts on Facebook and Instagram, as well as tweets on Twitter, that abuse women are considered and show the percentage of threats that women face from social media, which aids in understanding by the youth of India who misuse the women's safety and harass them in social medias via tweets, posts, and text should face strict action. Algorithms such as Nave Bayes (NB) and XGBoost are used in the analysis of women's safety on various social media sites. The goal is to use classification techniques to categorize or forecast the Type based on dataset properties.

2.3 D. Madhubala, M. Rajendiran and D. Elangovan, "A Study on Effective analysis of Machine Learning algorithm towards the Women's safety in Social Media", *2020 4th International Conference on Electronics Communication and Aerospace Technology (ICECA)* were proposed Social media is the most challenging one for analyzing the text and the social media application is a tool for expressing the opinion without facing any fear. At the same time, there will be a lot of crime and violence occurs on the internet. Especially the women who are in a popular position like celebrities, politicians, and also the commonpeople are facing the threats of violence are in social media. So they feel insecure in society. Still, there will be a problem for analyzing the text. It is very too difficult to classifying the text on social media. Through the machine learning technique, it would be done effectively. It filtered the unwanted text and then it reports to the respective office.

2.4 P. Kohli and K. Singh, "Analysis of Woman Safety Parameters in Smart and Non Smart Cities", *2021 9th International Conference on Reliability Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)* were proposed the era of Information and Communication Technology (ICT). Due to this the demands of people and industry is growing day by day. So quality of service in cities and urban area is greatly affected. Both Government and private organizations took

initiatives to migrate traditional cities into Smart cities. Development of Smart cities is focused by parameters: Technology, Safety, Privacy, Education, Unemployment, Crime, Health, Social, Legal, Economic, Traffic and Sustainability. Smart cities are considered safer cities against non-smart cities. These cities utilizes 'enterprise' LTE (eLTE) technology, a private adaptation of the 4G/5G (LTE) networks, IoT technology, secure wireless broadband connectivity etc for their advancement. These Smart cities are providing best resources to the people but Women Safety, Crime and 18 Violence is still a big challenge. Weather a city is smart or non-smart, the crime against women is never ended. However, in reality cases like rape, harassment, teasing, sexual assault, molestation, domestic violence etc. increases very rapidly.

2.5 K. Ashok, A. B. Gurulakshmi, M. B. Prakash, R. Poornima, N. S. Sneha and V. Gowtham, "A Survey on Design and Application Approaches in Women-Safety Systems", *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)* were proposed 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.

2.6 B. Patel and M. C. Zala, "Crime Against Women Analysis & Prediction in India Using Supervised Regression", *2022 First International Conference on Electrical Electronics Information and Communication Technologies (ICEEICT)* were proposed Women's safety and protection is must important but crime happens against women is common and serious issue. Problem of rising crimes that happens with women like rape & gang-rape cases, sexual harassment with women, cruelty by her husband or family, dowry death cases, acid attack on women and many more are increasing day by day. That is critical issue in not only in India but also various countries. The huge amount of data collection is generated based on crime reporting. This data can be very useful for appraise and predict crime and can help us to some degree to stop the crime. Data analysis is a process of examining, cleansing, transforming, and modelling data to establish useful information, reporting conclusions, and sustaining decision-making. This research shows a contextual

investigation of the women's crime examination in India using supervised learning. Information was pre-prepared to eliminate the anomalies, fix invalid locations, and ascertain the longitudes and scopes. The expressive investigation is created for dissect the women's crime per crime type and district and create heat maps for the crime dispersion.

2.7 D. A. J. Bonny, M. Jahan, Z. F. Tuna, A. Al Marouf and S. M. T. Siddiquee, "Sentiment Analysis of User-Generated Reviews of Women Safety Mobile Applications", 2022 First International Conference on Electrical Electronics Information and Communication Technologies (ICEEICT) were proposed Data analysis is a process of examining, cleansing, transforming, and modeling data to establish useful information, reporting conclusions, and sustaining decision-making. This research shows a contextual investigation of the women's crime examination in India using supervised learning. Information was pre-prepared to eliminate the anomalies, fix invalid locations, and ascertain the longitudes and scopes. The expressive investigation is created for dissect the women's crime per crime type and district and create heat maps for the crime dispersion. The outcomes help decision-makers with important experiences into women's crime forecast and avoidance. Application of this For Find the spatial criminal hotspot using a set of analysis and identify the crimes.

2.8 Muqem Ahmed, Mohd Dilshad Ansari, Ninni Singh, Vinit Kumar Gunjan, Santhosh Krishna B.V and & Mudassir Khan, "Rating-Based Recommender System Based on Textual reviews Using IoT Smart Devices", *Mobile Information Systems*, vol. 2022 were proposed Recommender system (RS) is a unique type of information clarification system that anticipates the user's evaluation of items from a large pool based on the expectations of a single stakeholder. The proposed system is highly useful for getting expected meaning suggestions and guidance for choosing the proper product using artificial intelligence and IoT (Internet of Things) such as chatbot. The current proposed technique makes it easier for stakeholders to make context-based decisions that are optimal rather than reactive, such as which product to buy, news classification based on high filtering views, highly recommended wanted music to choose, and desired product to choose. Recommendation systems are a critical tool for obtaining verified information and making accurate decisions. As a result, operational efficiency would skyrocket, and the risk to the company that uses a recommender system would plummet.

2.9 J. Refonaa, A. Porselvi, S.L.J. Shabu, BV Santhosh Krishna and Kazy Noor-E-Alam Siddiquee, "Characterisation of Intelligent Autonomous Agents Inspired by Biological Theory in CognitiveEnvironment", *Mobile Information Systems*, vol. 2022, pp. 1-7, ISSN 1875-905X.A. Mitra, "Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset)", *Journal of Ubiquitous Computing and Communication Technologies* were proposed a naturally motivated meaning of wise shrewd self-ruling specialists of humans. Knowledge is identified with whether the conduct of a framework adds to its self-upkeep. Conduct turns out to be clearer (or adapts to more biological issue factors) when it is able to make and utilize portrayals. The thought of portrayal ought not to be confined to formal articulations with neuro hypothetical semantics. The element at different degrees of canny frameworks assumes a fundamental part in shaping portrayals. The paper investigates an organically roused meaning of shrewd self-ruling specialists of humans. Insight is identified with whether the conduct of a framework adds to its self-upkeep. Conduct turns out to be keener when it is proficient to make and utilize portrayals. The idea of portrayal also focused on formal articulations with sentimental hypothetical semantics.

### 3. EXISTING SYSTEM

People often express their views freely on social media about what they feel about Indian society and the politicians who claim that Indian cities are safe for women. On social media websites, people can freely express their viewpoints and women can share their experiences where they have faced abuse and harassment or where they would have fought back against the abuse and harassment that was imposed on them. The tweets about the safety of women and stories of standing up against abuse and harassment further motivate other women to 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 an unsafe place for the women. In recent years a large number of people have been attracted to social media platforms like Facebook, It is a common practice to extract 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 based on social networks.

### 4. PROPOSED SYSTEM

Women have the right to the city which means that they can go freely whenever they want whether it be to an Educational Institute or any other place women want to go. But women feel that they are unsafe in places like malls, and shopping malls on their way to their job

location because of the several unknown Eyes body shaming,g, and harassment these women point to Safety or lack of concrete consequences in the life of women the main reason of harassment of girls. There are instances when the harassment of girls was done by their 21 neighbors 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 were abused or harassed by one of their neighbor or any other unknown person. Safest cities approach women's safety from the perspective of women's rights to affect the city without fear of violence abuse harassment. Rather than imposing restrictions on women that society usually imposes it is the duty of society to imprecise the need for the protection of women and also recognize that women and girls also have a right same as men have to be safe in the City.

**5. UML DIAGRAMS**

**5.1. CLASS DIAGRAM**

Class diagram is a static diagram. It represents the static view of an application. Class diagram is not only used for visualizing, describing, and documenting different aspects of a system but also for constructing executable code of the software application. 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 modelling of object oriented systems because they are the only UML diagrams, which can be mapped directly with object-oriented languages. It is also known as a structural diagram. Class diagram contains • Classes • Interfaces • Dependency, generalization and association.

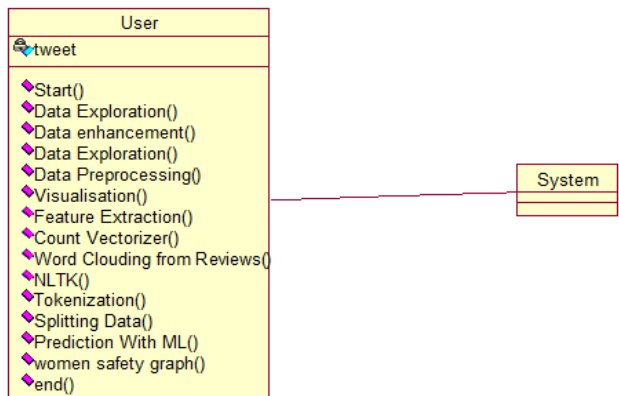


Fig 5.1 shows the class diagram of the project

**5.2. USECASE DIAGRAM:**

Use Case Diagrams are used to depict the functionality of a system or a part of a system. They are widely used to illustrate the functional requirements of the system and its interaction with external agents (actors). In

brief, the purposes of use case diagrams can be said to be as follows

- Used to gather the requirements of a system.
- Used to get an outside view of a system.
- Identify the external and internal factors influencing the system.

Use case diagrams commonly contains

- Use cases
- Actors
- Dependency, generalization and association relationships.

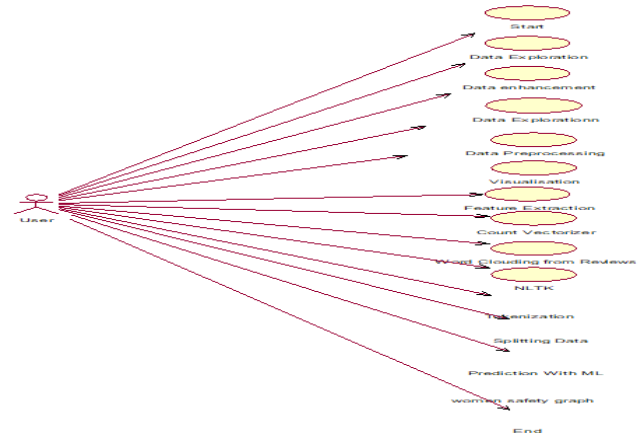


Fig 5.2 Shows the Use case Diagram

**5.3. SEQUENCE DIAGRAM:**

A sequence diagram simply depicts interaction between objects in a sequential order i.e. the order in which these interactions take place. We can also use the terms event diagrams or event scenarios to refer to a sequence diagram. Sequence diagrams describe how and in what order the objects in a system function. Sequence diagrams are used to formalize the behaviour of the system and to visualize the communication among objects. These are useful for identifying additional objects that participate in the use cases. These diagrams are widely used by businessmen and software developers to document and understand requirements for new and existing systems.



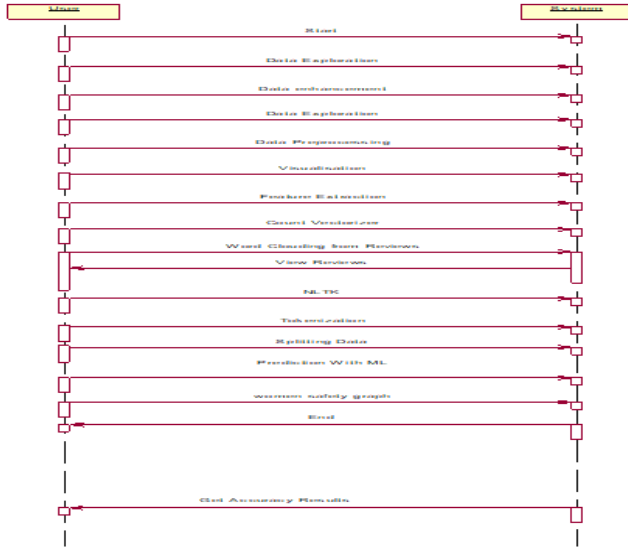


Fig 5.3 Shows the Sequence Diagram

6. RESULTS

6.1 Output Screens

In below screen click on ‘Upload Tweets Dataset’ button and upload dataset

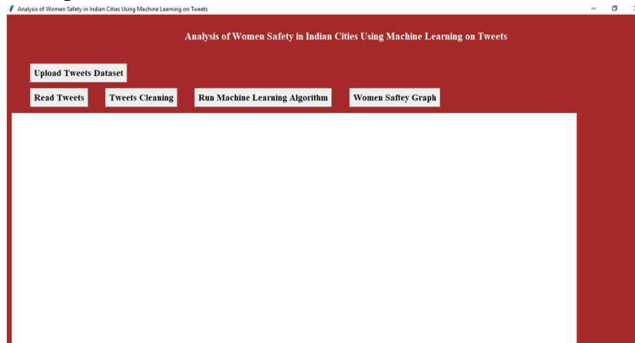


Fig 6.1 Upload the Tweets Dataset

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

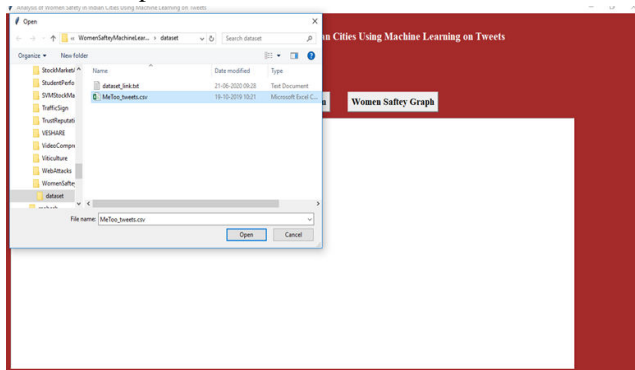


Fig 6.2 Load dataset

In the above screen upload the MeeToo\_tweets.csv file and then click on the ‘Open’ button to load the dataset.

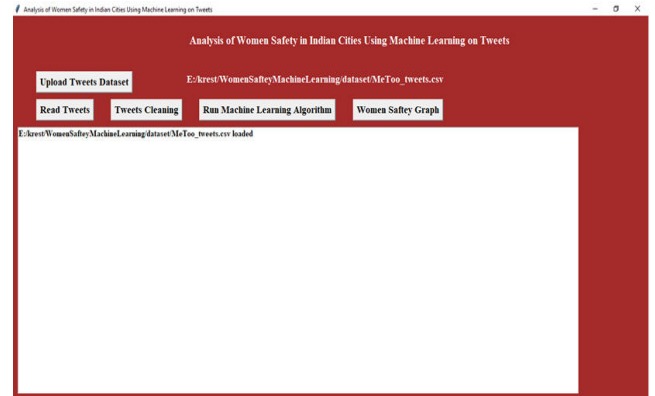


Fig 6.3 Reads Dataset

In the above screen tweets dataset is loaded and now click on the ‘Read Tweets’ button to read tweets from the dataset.



Fig 6.4 Tweets Cleaning

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



Fig 6.5 Run Machine Learning Algorithm

In the above screen we can see all special symbols and stop words removed from tweets and only clean words are there now click on the ‘Run Machine Learning Algorithm’

button to predict sentiments from tweets.

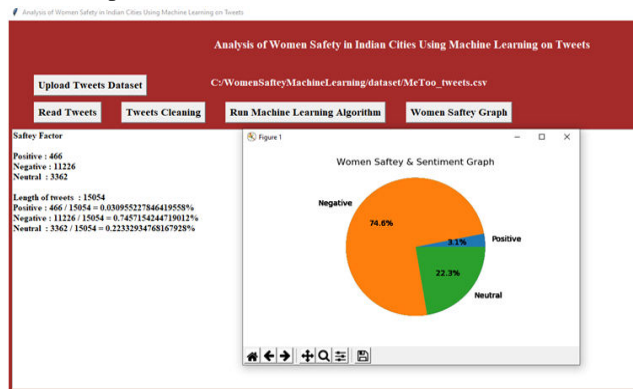


Fig 6.6 Women Safety Graph

Now click on the 'Women Safety Graph' button to get the below results and by seeing that result user can easily understand whether the area is safe or not. If the area is safe then more people will express either positive or neutral tweets and if not safe then more people will discuss negative tweets.

## 7. CONCLUSION

Throughout the paper, various algorithms have been discussed about deep learning and machine learning which can help in analyzing the huge amount of data accumulated via Twitter to help determine the safety of women in society. The machine learning algorithms used are very effective and work efficiently on various platforms when it comes to handling large amounts of data from social media platforms. These algorithms can help make a dent in women's safety extract information and create various datasets to work with. We look forward to working more and tweaking it to work even more efficiently shortly.

## FUTURE SCOPE

In the future, the analysis of women's safety in Indian cities using machine learning on tweets could expand by enhancing data collection methodologies to capture a more diverse range of perspectives and experiences. This would involve refining data preprocessing techniques to handle noise and language variations effectively and exploring advanced feature engineering approaches to extract nuanced information from tweets. Additionally, integrating geospatial analysis and real-time monitoring capabilities could provide valuable insights into spatial patterns of safety concerns and enable timely interventions. Engaging with local communities and stakeholders to validate findings and translate insights into actionable policy recommendations and intervention strategies would be crucial. Moreover, ensuring ethical considerations, mitigating biases, and conducting longitudinal studies to evaluate the long-term

impact of interventions are essential steps toward creating safer environments for women in Indian cities.

## 8. REFERENCES

- [1] Agarwal, Aproov, Fadi Biadisy, 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] D. Kumar and S. Aggarwal, "Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets", 2019 Amity International Conference on Artificial Intelligence (AICAI).
- [3] D. Madhubala, M. Rajendiran and D. Elangovan, "A Study on Effective analysis of Machine Learning algorithm towards the Women's safety in Social Media", 2020 4th International Conference on Electronics Communication and Aerospace Technology (ICECA).
- [4] P. Kohli and K. Singh, "Analysis of Woman Safety Parameters in Smart and Non-Smart Cities", 2021 9th International Conference on Reliability Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO).
- [5] K. Ashok, A. B. Gurulakshmi, M. B. Prakash, R. Poornima, N. S. Sneha and V. Gowtham, "A Survey on Design and Application Approaches in Women-Safety Systems", 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS).
- [6] B. Patel and M. C. Zala, "Crime Against Women Analysis & Prediction in India Using Supervised Regression", 2022 First International Conference on Electrical Electronics Information and Communication Technologies (ICEEICT).
- [7] D. A. J. Bonny, M. Jahan, Z. F. Tuna, A. Al Marouf and S. M. T. Siddiquee, "Sentiment Analysis of User-Generated Reviews of Women Safety Mobile Applications", 2022 First International Conference on Electrical Electronics Information and Communication Technologies (ICEEICT).
- [8] Muqeem Ahmed, Mohd Dilshad Ansari, Ninni Singh, Vinit Kumar Gunjan, Santhosh Krishna B.V and & Mudassir Khan, "Rating-Based Recommender System Based on Textual Reviews Using IoT Smart Devices", *Mobile Information Systems*, vol. 2022.
- [9] J. Refonaa, A. Porselvi, S.L.J. Shabu, BV Santhosh Krishna and Kazy Noor-E-Alam Siddiquee, "Characterisation of Intelligent Autonomous Agents Inspired by Biological

Theory in CognitiveEnvironment", *Mobile Information Systems*, vol. 2022, pp. 1-7, ISSN 1875-905X.A. Mitra, "Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset)", *Journal of Ubiquitous Computing and Communication Technologies*.