

Text Detection and Analyzing Sentiment Using Text Classification Approach

¹Ms. Mrunali D. Chaurse, ²Prof. Jayant P. Mehare

¹Student of Computer Science and Engineering, ²Assistant Professor
G. H. Rasoni University, Amravati, Maharashtra, India

Abstract: Text Classification, Text Mining and Sentiment Analysis have received huge attention, especially because of the availability of large volume of textual data on social media, e-commerce websites, blogs and other similar sources. This data is usually unstructured that becomes complex and expensive to extract knowledge from them to make some intelligent decisions. There is a growing need for developing different methodologies and models for efficiently processing the texts and extracting useful information from it. One way to extract information is text classification and its analysis that helps in decision making in various field of our lives. One of such a growing field is Social Media. This project performs the study and analysis of Social Media post by using different technique of text analysis and opinion mining along with sentiment analysis. This results in finding out the positive, negative users post and further analyzes negative category as violence, vulgar, offensive and hate on social media platform by performing text classification and analysis techniques.

Index Terms: Decision Making, Natural Language Processing (NLP). Social Networking, Social Media, Text classification, Text Analysis, Unstructured Data.

I. INTRODUCTION

As all knows, Social Media becomes the integral part of people's life, they used to obtain and share different types of information updates proactively on 24/7 basis. Social media includes social networking sites, some business applications and blogs where people easily connect with each other that helps in growing business ideas and decision making. With the social media platform people can share their ideas and interact with other people, from old friends to strangers. As the application areas of social media platform are much more and always growing on increasing, it plays a vital role in transforming people's life style. With this such sites have become a day to day routine for people of all age group.

But at the same time, some of us have to experience the darker side of social media uses. News reports of cyber bullying, violence, criminal activity, and suicide fueled by social media are very shocking and troubling. As per the study Social media is used by billions of people worldwide and becomes an inherent part of Internet. This interaction reveals a lot of information, may include personal information visible or can be shared publically. Now a days children started more and more use of such social media platform as more surrounded by mobile devices. So it becomes necessary to avoid this drawback and improve the detection of violence posting on the social media. This project helps in detecting such violence posting and restricts such users in doing so or uses of social media platform for further.

This project mainly studies the text analysis techniques that can be helpful for the detection and prevention of positive, negative and violence type of post on social media sites. The proposed system developed Social Networking Application Evaluator for detecting Malicious Post through Social Networking Application, a suite of efficient classification techniques for identifying whether the users post is malicious or not. This is the comprehensive study for focusing on malicious Social networks apps that focuses on quantifying, profiling, and understanding malicious apps, and synthesizes this information into an effective detection approach. The application develop with the text analysis and classification approach will becomes the prototype to avoid such type of violence posting by social media users.

1.1 Problem Definition

Many of the social media sites are most popular on the web. Some social media sites have transformed the way where people communicate and socialize on the web. Social networking sites render the opportunity for people to reconnect with their old friends, colleagues and mates. It also helps people to make new friends, share content, pictures, audios, videos amongst them. Social media also changes the life style of a society. One of the negative effects of social media is that it makes people addicted. People spend lots of time in social networking sites which can divert the concentration and focus from the particular task. It is also abusing the society by invading on people's privacy. Social lies like family ones also weaken as people spend more time connecting to new people. Social media can easily affect the kids; the reason is sometimes people shares photos, videos on media that contain violence and negative things which can affect the behavior of kids. This needs to be overcome with possible ways of growing technologies.

1.2 Aim

The best entertainment for the younger generation is given in the form of Social Networking sites. The Online Social Networks mainly helps an individual to connect with their friends, family and the society online in order to gather and share new experiences with others. Now days, the Online Social Networks are facing the problem of the people posting the indecent messages on any individual's wall which annoys other people on seeing them. The proposed system aims to perform blocking the malicious post form the social networking application.

II. METHODOLOGY

Here, Brute Force Algorithm is used for searching positive, negative and neutral words from user’s post. With this algorithms all character found to match or mismatch is detected. First the user is login and set the post. The posts mainly contains only textual data. Splitter is used to spread out word to word data from the user post and then malicious word detected. With this methodology system is able to detect positive, negative and neutral post. And for negative post, there are four types of categories of the post such as violence, vulgar, offensive and hate. Determining the category of the post and find out the average percentage value of malicious post on the social media which make more secured from spreading such type of violence thought on the social networking. And according to the percentage negative posting users are blocked and restricted to this in future.

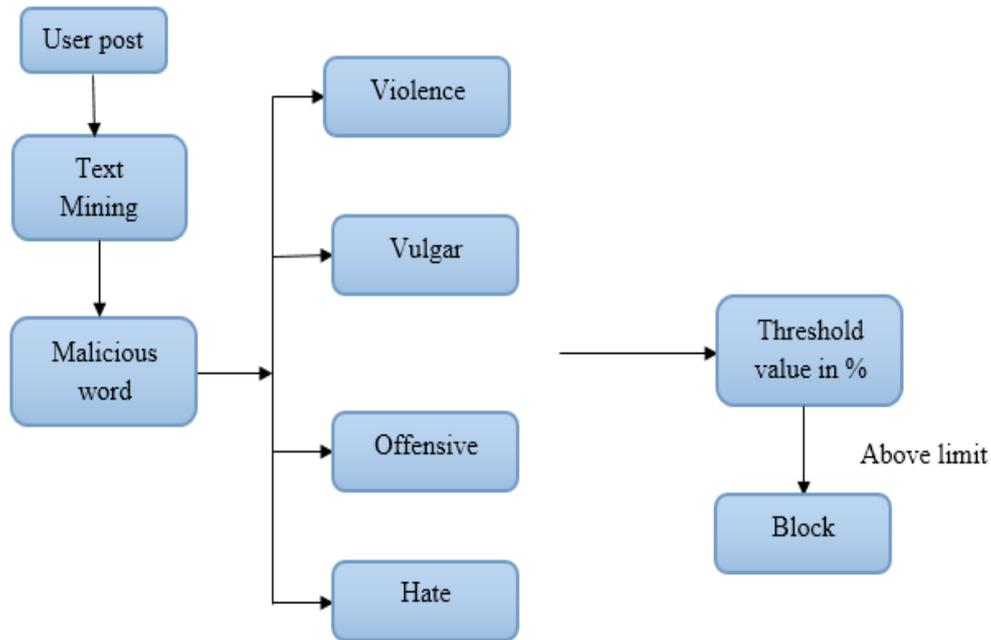


Fig. 1: Diagram of working system

In Figure 1 proposed system which will check for the user post links post by them and their content in post as well as link and determining the abusing of that which help to make social media to make more secured from spreading such type of volatile thoughts on web. Social networking sites render the opportunity for people to reconnect with their old friends, colleagues and mates. It also helps people to make new friends, share content, pictures, audios, videos amongst them. Social media also changes the life style of a society. One of the negative effects of social media is that it makes people addicted. People spend lots of time in social networking sites which can divert the concentration and focus from the particular task. Social media can easily affect the kids. So to avoid this this project filter the negative post in these 4 categories and further decisions are made based on the average percentage calculation.

III. ALGORITHM USED

Brute Force Algorithm

Brute force algorithms are all character found to match or mismatch is detected. It is also checking at all the position of text between 0 and nm. Whether an occurrence of the pattern start there or not. While a brute force search is simple to implement, and will always find a solution if it exists, its cost is proportional to the number of candidate solution which in many practical problem tends to grow very quickly as the size of the problem increases. Therefore, brute force search is typically used when the problem size is limited, or when there are problem specific heuristics that can be used to reduce the set of candidate solution to a manageable size. The method is also used when the simplicity is more important than speed.

```

Brute Force String Match (T[0..n-1],P[0..m-1])
//Implements brute-force string matching
//Input: An array T[0..n-1] of n characters representing a text and an array
P[0..m-1] //Output: The index of the first character in the next that starts a
matching substring // or -1 if the search is unsuccessful

For i←0 to n-m do
    j←0
    While j<m and P[j]=T[i+j] do
        j←j+1
    if j=m return i
return -1

```

In order to apply brute-force search to a specific class of problems, one must implement four procedures, first, next, valid, and output. These procedures should take as a parameter the data P for the particular instance of the problem that is to be solved, and should do the following:

1. First (P): generate a first candidate solution for P.
2. Next (P, c): generate the next candidate for P after the current one c.
3. Valid (P, c): check whether candidate c is a solution for P.
4. Output (P, c): use the solution c of P as appropriate to the application.

The next procedure must also tell when there are no more candidates for the instance P, after the current one c. A convenient way to do that is to return a "null candidate", some conventional data value Λ that is distinct from any real candidate. Likewise the first procedure should return Λ if there are no candidates at all for the instance P. The brute-force method is then expressed by the algorithm.

IV. IMPLEMENTATION & RESULT ANALYSIS

The proposed system has done analysis on well-known Social media violence posting using the effective proposed approach and finally displays the results. System methodology implements and determine the negative category of the post and find out the average percent value of malicious post. This helps social media to become more secure by blocking user that performs spreading of such type of violence thoughts on the social networking sites. In implementation phase, to avoid mistake if the innocent user gets block by his mistake. The functionality is available to send the request to the administrator for unblocking. The request is receive as the message by the administrator, with this request administrator checks the behavior of the user from history of his uploaded posts. If the administrator wants to give chance to that user, then he unblock him and all his previous malicious post gets deleted. And if administrator don't wants to give chance to requested user then he simply reject the request and users is restricted to use the portal for future.

- Parameters for Performance

The system is analyzing the post which is posted by the user and calculating how much category malicious percentage. Whether it is malicious and non-malicious.

By using the following formula:

$$\text{Category Malicious percentage} = \frac{\text{Itemsets count} * 100}{\text{Itemsets length}}$$

Where, Itemsets count – Number of words,

Itemsets length – Length of words

The system is analyzing the post which is posted by the user and calculating how much malicious percentage.

The system runs proposed approach and base approach on the social networking. At the end compute various parameters to conclude the system results. The results obtained are compared using a graph and table that presents the actual research work with different approaches. In result analysis performed in the proposed system calculates different types of ratios of violence post present

on the social media. Based on the different ratios and values of different categories of words, users are successfully evaluated based on their malicious contents. That helps to block the malicious user from posting contents on the social networking portal.

Comparative Analysis:

Following table shows the comparative analysis between the Existing and the Proposed System:

Table 1: Comparative Analysis

Existing System	Proposed System
Analysis of post is done After Post is uploaded	Post is uploaded only after analysis is done and inform to User
Only Post is block if some maliciousness is found	Malicious User can directly blocked if maliciousness reaches threshold limit.

VI. CONCLUSION

As the technology is growing the social media has become the routine for each and every person, people are seen addicted with these technology every day. With different fields its impact is different on people. Social media has increased the quality and rate of collaboration for students. Business uses social media to enhance an organization's performance in various ways such as to accomplish business objectives, increasing annual sales of the organization. As we have seen the use of Social media platforms by all age people is growing on increasing, people are seen addicted with these technology every day. So it becomes necessary to develop some techniques and methods for analyzing customer's post as they are present in thousands. There are different terms and methods like opinion mining, sentiment analysis, word alignment model, etc. associated with this concept. Here, the technique develop and implement helps to stop violence data posting on the social media by find out the average percent value of malicious post and blocks such negative users.

REFERENCES

- [1] Said A. Salloum, Mostafa Al-Emran, Azza Abdel Monem, Khaled Shaalan, "A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives", *Advances in Science, Technology and Engineering Systems Journal*, Vol. 2, No. 1, 127-133 (2017).
- [2] RIZWANAFIFA, CRISTINE K. KING, "A Survey on Text Mining in Social Networks", *The Knowledge Engineering Review*, Vol. 00:0, 1-14 © 2004, Cambridge University Press
DOI: 10.1017/S0000000000000000.
- [3] Sorensen, L. 2009. "User managed trust in social networking comparing facebook, myspace and linkdin", *In Proceedings of 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic System Technology*, (Wireless VITAE 09), Denmark, 427-431.
- [4] Puja Munjal, Aditi Gupta, Mahima Abrol, Hema Banati and Sandeep Kumar, "Social Media Based Opinion Mining Using Lexical Sentiment Analysis", *International Conference on Paradigm Shift in World Economies: Opportunities and Challenges - ISBN :978-1-63535-729-5* (2017).
- [5] João Filipe Figueiredo Pereira, "Social Media Text Processing and Semantic Analysis for Smart Cities", FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO, arXiv:1709.03406v1 [cs.SI] 11 Sep 2017.
- [6] Ahmed Imran KABIR, Ridoan KARIM, Shah NEWAZ, Muhammad Istiaque HOSSAIN, "The Power of Social Media Analytics: Text Analytics Based on Sentiment Analysis and Word Clouds on R", *Research Gate, Informatica Economică* vol. 22, no. 1/2018. DOI: 10.12948/issn14531305/22.1.2018.03.
- [7] Vibhuti Patel, Mital Panchal, "A survey on Opinion Mining Methods from Online Reviews", *International Journal of Scientific Research in Science, Engineering and technology*, In December, 2015.
- [8] Aggarwal, C. 2011. Text mining in social networks. In *Social Network Data Analytics*. 2nd edn. Springer, 353-374. Baumer.
- [9] T. Anwar and M. Abulaish, "Ranking Radically Influential Web Forum Users," in *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 6, pp. 1289-1298, June 2015.
doi: 10.1109/TIFS.2015.2407313.
URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7050292&isnumber=7084215>
- [10] Bo Pang, Lillian Lee, "Opinion Mining and Sentiment Analysis", *Foundations and Trends in Information Retrieval* Vol. 2, Nos. 1-2 (2008).
- [11] E. Mouhssine and C. Khalid, "Social Big Data Mining Framework for Extremist Content Detection in Social Networks," *2018 International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*, Rabat, Morocco, 2018, pp. 1-5. doi: 10.1109/ISAECT.2018.8618726.
- [12] Fangtao Li, Sinno Jialin Pan, Ou Jin, Qiang Yang and Xiaoyan Zhu, "Cross-Domain Co-Extraction of Sentiment and Topic Lexicons", *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 410-419, Jeju, Republic of Korea, 8-14 July 2012.

- [13] Terrorism detection from social media, available [online]:https://www.leadingindia.ai/downloads/projects/SMA/sma_6.pdf
- [14] Srijan Kumar, Francesca Spezzano and V.S. Subrahmanian, "Identifying Malicious Actors on Social Media", available [online]: <https://cs.stanford.edu/~srijan/badactorstutorial/>
- [15] Guadalupe Obdulia Gutiérrez-Esparza, Maite Vallejo-Allende and José Hernández-Torruco, "Classification of Cyber-Aggression Cases Applying Machine Learning", *Applied Science*, May 2019, 9(9), 1828; <https://doi.org/10.3390/app9091828>.
- [16] From Research page: What is Text Analytics? [Online] available: <https://www.predictiveanalyticstoday.com/text-analytics/>
- [17] K. Gordon. Number of social mediausers worldwide from 2010 to 2021 (in billions) [Online]. Available: <https://www.statista.com/statistics/278414/number-of-worldwide-socialnetwork-users/>
- [18] Unknown. Daily time spent on social networking by internet users worldwide from 2012 to 2017 (in minutes) [Online]. Available: <https://www.statista.com/statistics/433871/daily-social-media-usageworldwide/>
- [19] Unknown. Percentage of global population using Facebook as of June 2017, by region [Online]. Available: <https://www.statista.com/statistics/241552/share-of-global-populationusing-facebook-by-region/>
- [20] Unknown. Number of monthly active Facebook users worldwide as of 3rd quarter 2017 (in millions) [Online]. Available: <https://www.statista.com/statistics/264810/number-of-monthly-activefacebook-users-worldwide/>
- [21] K. Vance, W. Howe, and R. P. Dellavalle, "Social internet sites as a source of public health information," *Dermatologic clinics*, vol. 27, pp. 133-136, 2009.
- [22] J. S. Brown and P. Duguid, *The Social Life of Information: Updated, with a New Preface*: Harvard Business Review Press, 2017.

