International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN (online): 2347-5552, Volume-11, Issue-3, May 2023
https://doi.org/10.55524/ijircst.2023.11.3.3
Article ID IJIRD-1244, Pages 12-19
www.ijircst.org

Machine Learning Prospects: Insights for Social Media Data Mining and Analytics

Anu Sharma[1], and Vivek Kumar[2]

[1,2]Assistant Professor, Department of Computer Science & Engineering, Teerthanker Mahaveer University, Moradabad, India

Correspondence should be addressed to Anu Sharma; er.anusharma18@gmail.com

ABSTRACT: Social network has increased surprising consideration in the most recent decade .Social network deals with enormous volume of composite as well as unstructured data and they are very hard to handle. Due to expanding dimensions and demand, one of the encouraging and interesting research field becomes social network. Data Mining affirms to get knowledge by discovery patterns among data.

We have discussed social media mining and Social Media analytics. We have insights on the social media effect of our lives, some facts and reports from various sources. We have Integrated this growing research field of social networks with Machine Learning with one simple example of sentiment analysis of Twitter data using Machine Learning. We have also proposed the algorithms to improve the social media analytics results using Machine Learning. In this paper, we will exhibit how machine learning will utilizing for social networking systems like Twitter. In this procedure, a framework is proposed that will collect the tweets messages from the and we will inspect the item's input to show the positive, negative, or nonpartisan tweets, for this this purpose we have proposed new machine learning algorithms Naive Bayes, maximum entropy to find these outputs. Our proposed Model will help new researchers, companies, Industries, business community, practitioners, new integrated application designers, and the global community to solve the new research problem and may reducing design failure rate of 80% by large through social media mining and networks.

KEYWORDS: Data mining, Social media data, deep learning, Machine learning, Naïve Bayes, Maximum entropy

I. INTRODUCTION

Now a day's social media is the great source of information for all the business community and individuals with multimedia options. Using these approach users may connect with any targeted groups to share their data and information in today's competitive environment.[2] [5]

Social media is an excellent source of information and a fantastic communication tool. Instead of only sharing photographs and videos on the site, businesses and individuals can make the most of it. The platform allows users to effortlessly and wonderfully connect with their target audience. Learners, professionals, scientists, and project managers can use social media mining to better comprehend the basics and potentials of social media mining by using social media platforms, social network analysis, and data mining. [3][6]

Social networks are characterized as virtual areas where set of all ages can connect, distribute information as well as ideas, and establish bonds [7][9]. People are represented and connected through a social network community. It also allows users to stay in touch with friends, create personal profiles, browse other people's profiles and connections, chat, and exchange personal data [10]. Twitter, Facebook, MySpace are utmost frequently access social network sites.[41][42]

The development of the blogosphere has given the average customer the capacity to impact public image and brand profitability. As a result, marketing companies must be aware of what is being said on influential blogs, how the expressed thoughts may affect their business, and how to extract business intelligence and value from these blogs. [4][8]

We will use sentiment analysis in this study, which is critical for data mining. The computational handling of suppositions is known as assessment analysis. The great majority of the general public uses Twitter to share information. [1][11].

II. RELATED TECHNOLOGIES

A. Data Mining

In figure 1, Initial step is to prepare the data. Data is chosen and processed with the help of a domain specialist. Further, the prepared input is processed using a data mining technique. The third process is to determine whether the data mining algorithms produced important information. [14][41]. Figure 2 describe the Predictive and Descriptive data mining Algorithm and Fig 3 describe the data mining Technique.

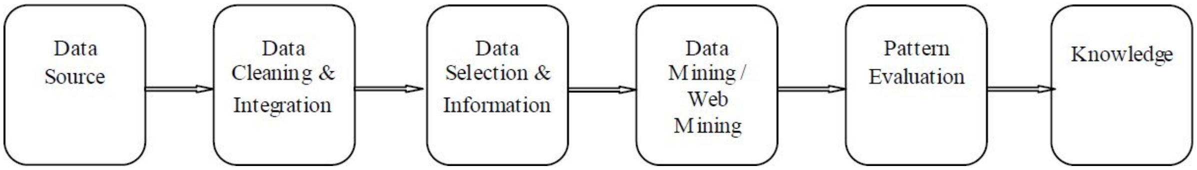


Figure 1: Steps in Data Mining Process

B. Data Mining Algorithms

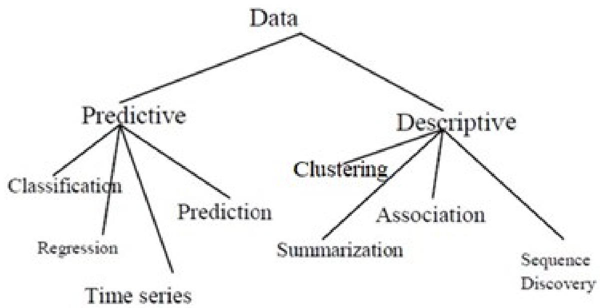


Figure 2: Modeling of Data Mining algorithms

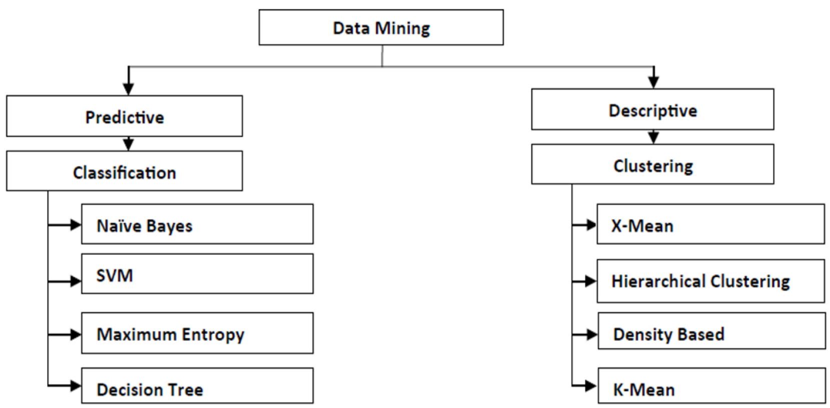


Figure 3: Data Mining Techniques

III. DEEP LEARNING PROJECTION FOR SOCIAL MEDIA ANALYTICS

Due to the rapid growth and extensive scope of social media (SM), assessing these data with classical techniques and technology has become difficult. [4]The solution to this problem is discovered to be DL. We explore the practiced DL architectures in depth by offering a taxonomy-oriented summary in this study (SMA). Nonetheless, rather than focusing on technical details, this work focuses on describing SMA-oriented challenges and their DL-based solutions. Figure 4 describe the input, hidden and output layer.

Linking DL with SMA can expose reminiscent insights. The following are the key questions that this study aims to answer in terms of contribution:

- Contribute deep learning approaches that can be used to create a roadmap for extracting useful insights for SMA.
- Prepare a classification that determines essential elements for studying the semantics of the problem, which could be useful in building better future vision in a variety of SMA application.
- Examines the advantages and disadvantages of present approaches.
- Illustrates the most common application scope being using DL.
- Identifies significant research issues including directions for future study.

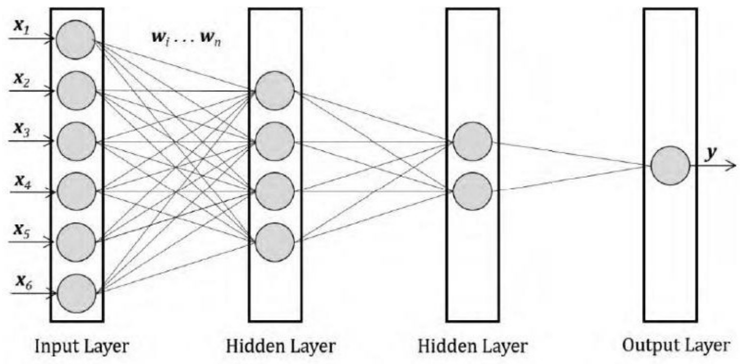


Figure 4: Deep learning[47]

A social net is a linked creation of social factor whichever nodes are social objects like people, organisations, web pages, and so on, and links are formed as relationships between them [24][28].Using statistical, graphical, and visual methodologies, researchers have developed many methods for network knowledge exploration and representation. They utilised degree to assess an individual's direct relevance or social relationship; community to measure a group's importance; and place to analyze people or groups capable of information flow across a network [32][34][41]. This is because the importance of employing centrality measurements to identify key persons, clustering techniques to identify subgroups, and network visualisations to characterise online conversions and marketplaces has been recognised. Traditional network analysis approaches have grown less successful as a result of the extraordinary accumulation of social-related data, fueled by the expansion of social media websites and inherent discrepancy as well as intricacy[15]. As a result, it is clear that machine learning will become an indispensable tool for social network mining in the next years. Psychologists work alongside machine learning (ML) to develop computer models for activities such as recognition, prediction, planning, and analysis, alike unpredictable conditions [21][23]. As a result, it's critical to investigate the synergy of machine learning approaches in social network analysis, with an emphasis on pragmatic operations and novel investigation. Fig 5 describe the Deep Learning in Social Media.

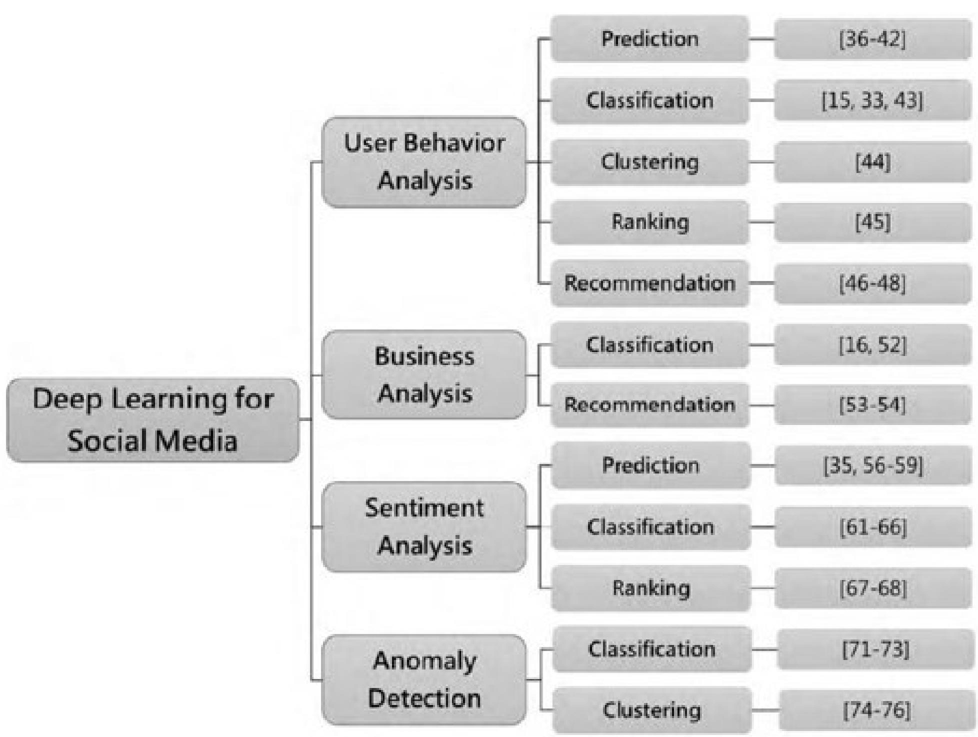


Figure 5: Deep Learning In Social Media[47]

IV. MACHINE LEARNING FOR SOCIAL MEDIA ANALYTICS

Machine learning algorithms like KNN, K-means, SVM can be used to accomplish this. In some circumstances, statistical methods are also considered non-machine learning methods for discovering patterns. "Statistical techniques are driven by the data and are used to find trends and develop predictive models,". [16][17].

A. Representation of network data [62]

Fig 6 and Fig 7 describe Graphs and matrices are used to represent networks. One of the most extensively used skeletal net illustration and observable analytical investigative is the graph based view, which employs graph theory. Distinct colors, hues, symbols, and sizes are utilised to indicate the attributes or categories of different actors [18][20]. This aids in the comprehension of overlay actor relationships. To avoid overlapping, ties are sometimes depicted with curves rather than straight lines.

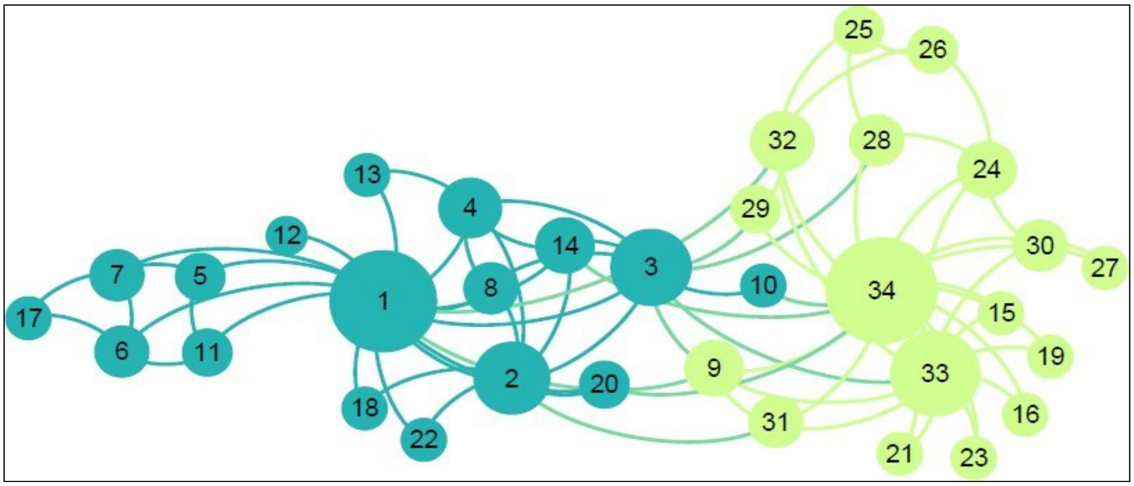
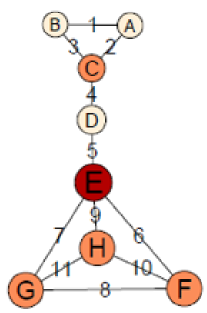


Figure 6: Zachary's karate club is depicted as a sociogram.



(a) Sociogram

	A	B	C	D	E	F	G	H
A	0	1	1	0	0	0	0	0
B	1	0	1	0	0	0	0	0
C	1	1	0	1	0	0	0	0
D	0	0	1	0	1	0	0	0
E	0	0	0	1	0	1	1	1
F	0	0	0	0	1	0	1	1
G	0	0	0	0	1	1	0	0
H	0	0	0	0	1	1	0	0

(b) Adjacency matrix

	1	2	3	4	5	6	7	8	9	10	11
A	1	1	0	0	0	0	0	0	0	0	0
B	1	0	1	0	0	0	0	0	0	0	0
C	0	1	1	1	0	0	0	0	0	0	0
D	0	0	0	1	1	0	0	0	0	0	0
E	0	0	0	0	1	1	1	0	1	0	0
F	0	0	0	0	0	1	0	1	0	1	0
G	0	0	0	0	0	0	1	1	0	0	1
H	0	0	0	0	0	0	0	0	1	1	1

(c) Incident matrix

	Adjacent nodes
A	B, C
B	A, C
C	A, B, D
D	C, E
E	D, F, G, H
F	E, H, G
G	E, F, H
H	E, F, G

(d) Adjacency list

	Incident edges
A	1, 2
B	1, 3
C	2, 3, 4
D	4, 5
E	5, 6, 7, 9
F	6, 8, 10
G	7, 8, 11
H	9, 10, 11

(e) Incident list

Vertices	A, B, C, D, E, F, G, H		
Edge	**Source**	**Target**	**Props(Weight,...)**
1	A	B	-
2	A	C	-
3	B	C	-
4	C	D	-
5	D	E	-
6	E	F	-
7	E	G	-
8	F	G	-
9	E	H	-
10	F	H	-
11	G	H	-

(f) Vertex-Edge list

Figure 7: Social network graphical and mathematical representations

Crisp characteristic function

$$V'_{i,j} = \begin{cases} 1, & v_i \text{ is related to } v_j \\ 0, & v_i \text{ is not related to } v_j \end{cases} \tag{1}$$

Fuzzy characteristic function

$$V'_{i,j} = \begin{cases} 1, & v_i \text{ has strong positive relationship with } v_j \\ \gamma = f(V_{i,j}) \in]0,1[, & v_i \text{ is related to a certain extent with } v_j \\ 0, & v_i \text{ is not related to } v_j \end{cases} \tag{2}$$

V. SKELETON

This model collects comments (tweets) against long-distance informal communication sites as well as provides a business context. The sentiment analysis has two layers. They are the sentiment analysis layer and the data processing layer. The main layer is in charge of data collection and mining, as opposed to the subsequent layer employs an operation to establish the info mining's aftereffects.

A. Collecting Data and Preprocessing

We will physically compile a list of tweets or comments on varied items, and then visit informal community gathering spots to collect tweets. All of the gathered tweets can be saved in a database and processed later. Words and their implications will be reviewed during the assessment process. Tweets concerning tattles or disconnected information will be discarded based on social semantic analysis, and true material will be vigilantly erased. Fig 8 shows the architecture for sentiment analysis system.

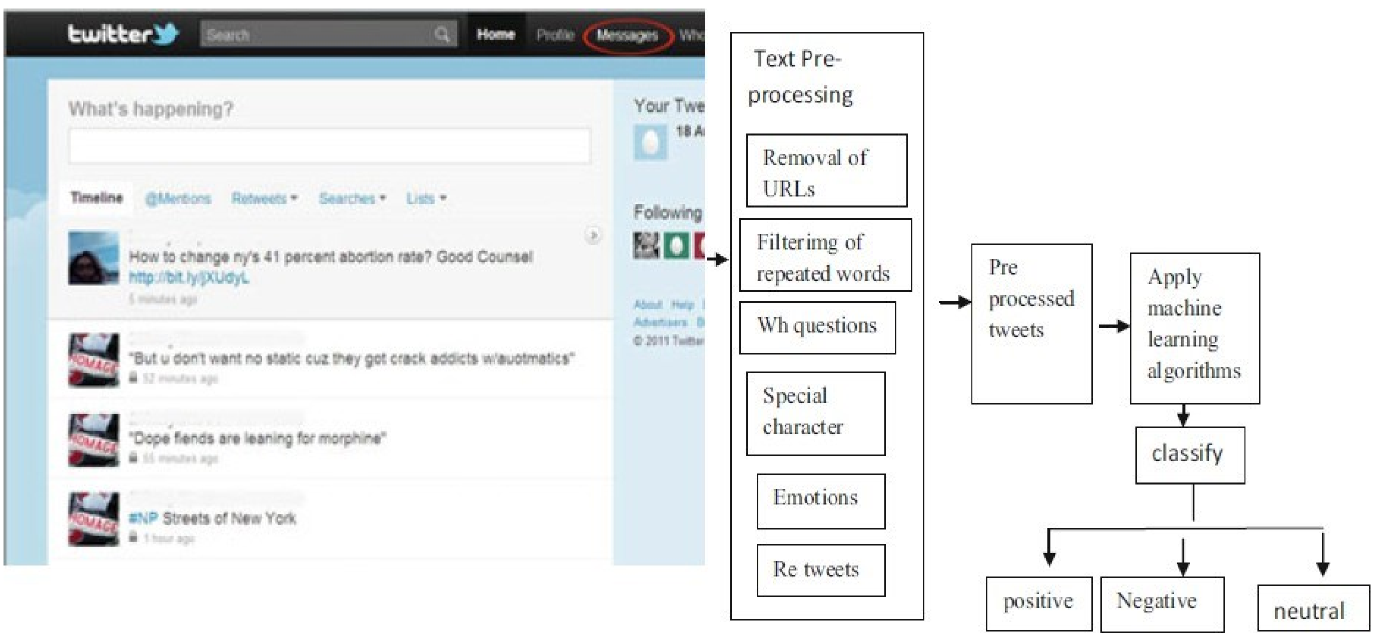


Figure 8: Architecture for sentiment analysis system

B. Proposed Naive Bayes Classifier

Knowledge: messages *Msg* _ {*ms*1, *ms*2, *ms*3 . . . *msn*},
Database: Baye's Table ByT
Yield: Great1 messages G _ (*g* 1, *g* 2 . . .),
Non-Great1 messages *NG*_ {*ng*1, *ng*2, *ng*3 . . .},
nonpartisan1 messages *nu* _ {*nu*1, *nu*2, *nu*3 . . .} *Msg* _ {*ms*1, *ms*2, *ms*3 . . .}

Step 1:Dividing the tweet message into words t*mi* _ {*wo*1, *wo*2,*wo*3 . . .}, *i* = 1, 2, . . . n

Step 2: If *woi* ByT
+ive extremity as well as -ive extremity should be returned.

Step 3:We can use the following equation to determine the general extremity of a word
Log (+ive extremity) - log(−ive extremity).

Step 4: Reiterate step 2 for each of the words until they are all finished.

Step 5: To calculate the message's overall extremity, we must take into account the clash of all of the message's expressions.

Step 6: The message could be positive or negative, depending on the extremity.

Step 7: We must repeat step 1 for each of the messages as far as M1 is NULL..

C. Proposed Maximum Entropy Classifier

Info: Mess *Msg* _ {*ms*1, *ms*2,*ms*3 . . . ,*msn*},
Database: Entropy Table EnT
Output:
Great1_ mess G*F* _ { g*f* 1,*g f* 2 . . .},
NonGreat1 mess N*M*_ {n*m*1,*nm*2, *nm*3....}
neutral mess nu _ {nu1, nu2, nu3 .... }
M _ {*ms*1,*ms*2,*ms*3...}

Step 1: Separating a tweet message into lexis *mli* _ {*wl*1, *wl*2, *wl*3 . . .}, *i* = 1, 2,... *n*

Step 2: If wli NyT.
+ive extremity and -ive extremity should be returned

Step 3: Associate equation {(+ive polarity) * log(1/+ive extremity) −(−ive polarity) * log(1/−ive extremity)}.

Step 4: Rerun the step 2 prior for each one of the lexis till end of the lexis

Step 5: We have to rehash the step 1 as far as M1 is NULL, for each and every one of the messages

VI. CONCLUSION

Proposed approach in fig 10 is to preprocess the tweets to find the client estimation. The data will be saved in a database after the preparation stage. Machine learning algorithms will make use of the data that has been saved. We have proposed new machine learning algorithms i.e Proposed Naive Bayes, maximum entropy to find these outputs.

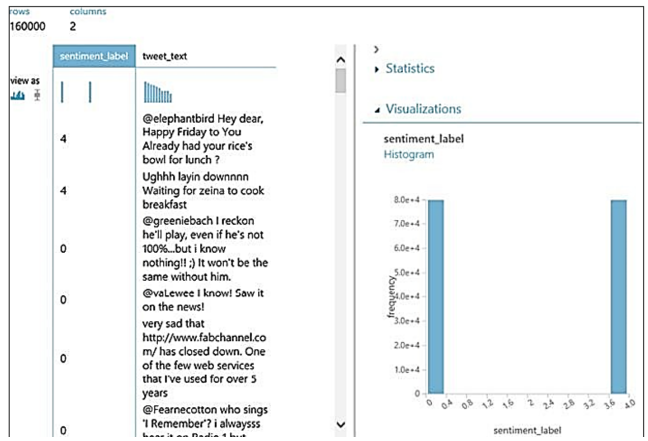


Figure 9: Preprocess the Twitter reviews

Our proposed Model will help new researchers, companies, Industries, business community, practitioners, new integrated application designers, and the global community to solve the new research problem and may reducing design failure rate of 80% by large through social media mining and networks.

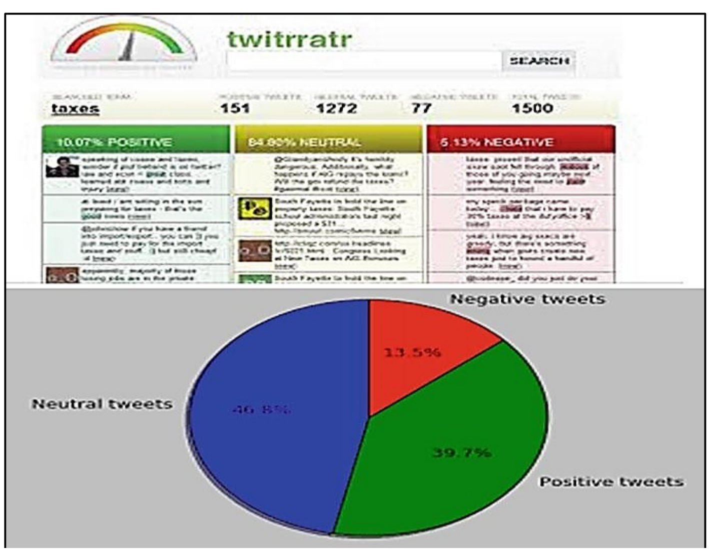


Figure 10: Messages classified as positive, negative, and neutral

REFERENCES

[1] Relationships and Technology Lab, College of Liberal Arts & Sciences: https://randtlab.ku.edu/jeffrey-hall. [accessed on: January 18, 2021]

[2] Twenge JM, Spitzberg BH, Campbell WK. Less in-person social interaction with peers among U.S. adolescents in the 21st century and links to loneliness. Journal of Social and Personal Relationships. 2019;36(6):1892-1913. doi:10.1177/0265407519836170.

[3] Social media's growing impact on our lives https://www.apa.org/members/content/social-media-research- series. [accessed on: January 18, 2021]

[4] Y. Zhao et al., "IEEE Access Special Section Editorial: Advanced Data Mining Methods for Social Computing," in IEEE Access, vol. 8, pp. 228598-228604, 2020, doi: 10.1109/ACCESS.2020.3043060.

[5] Tajeuna Etienne Gael, Bouguessa Mohamed, and Wang Shengrui. Modeling and predicting community structure changes in time evolving social networks. IEEE Transactions on Knowledge Data Engineering, pages 1–1.

[6] Zhaoyi Li , Fei Xiong , Ximeng Wang, Hongshu Chen and Xi Xiong, Topological Influence-Aware Recommendation on Social Networks, Complexity, 2019, Article no. 6325654.

[7] Barabasi, A.L. (2002). Linked: The New Science of Networks. Cambridge, MA: Perseus

[8] Maddon, M. & Fox, S. (2006). Riding the waves of "Web2.0": More than a buzzword, but still not easy to define. Pew Internet Project. Retrieved from http://www.pewinternet.org/pdfs/PIP_Web_2.0.pdf

[9] E.K Clemons., " The Future of Advertising and the Value of Social Network Websites: Some Preliminary Examinations". Minneapolis, Minneosta, USA, 2007.

[10] D. Boyd., "Social Network Sites: Definition, History, and Scholarship",Journal of Computer-Mediated Communication 13 (1), 2007.

[11] "Social Network Marketing: The Basics" Available: http://www.labroots.com/Social_Networking_th

[12] Nandi,G., Das, A.:Asurvey on using datamining techniques for online social network analysis. Int. J. Comput. Sci. Issues (IJCSI) 10(6), 162 (2013)

[13] [13]. Batrinca, B., Treleaven, P.C.: Social media analytics: a survey of techniques, tools and platforms. Ai Soc.30(1), 89–116 (2015).

[14] J.W. Seifert. "Data Mining: An Overview", CRS Report for Congress, 2004 [Online] Available: http://www.fas.org/irp/crs/RL31798.pdf [Sep 05 2013]

[15] H. Kob, G, Tan, "Data Mining Applications in Healthcare" ,Nov 10 2010.

[16] M Hart. "Progress of organisational data mining in South Africa", Department of Information Systems, University of Cape Town, South Africa, 2006. [Online] Available: http://intranet.inria.fr/international/arima/00

[17] Neelamadhab Padhy1, Dr. Pragnyaban Mishra 2, and Rasmita Panigrahi3, "The Survey of Data Mining Applications And feature scope", in International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, Issue.3, 2012.

[18] Smita1, Priti Sharma, " Use of Data Mining in Various Field: A Survey Paper", in IOSR Journal of Computer Engineering (IOSR-JCE), Volume 16, Issue 3, PP 18-21, 2014.

[19] Mrs. Bharati M. Ramageri," Data Mining Techniques And Applications", in Indian Journal of Computer Science and Engineering, Vol. 1 Issue. 4, PP: 301-305.

[20] Annan Naidu Paidi " Data Mining: Future Trends and Applications" in International Journal of Modern Engineering Research (IJMER), Vol.2, Issue.6, PP:4657-4663, 2012.

[21] Umamaheswari. K, S. Niraimathi "A Study on Student Data Analysis Using Data Mining Techniques", in International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 8, PP:117-120, 2013.

[22] E.Veerman, T.Lachev, D.Sarka, Microsoft SQL Server 2008- Bussiness Intelligence Development and Manitenance, Microsof, PHI Learning Private Limited, pp 372 – 380, 2009.

[23] Nikita Jain, Vishal Srivastava "*DATA MINING TECHNIQUES: A SURVEY PAPER*" in IJRET: International Journal of Research in Engineering and Technology, Volume: 02, Issue: 11, PP:116-119, 2013. [24]. Ranshul Chaudhary1, Prabhdeep Singh2, Rajiv Mahajan3, "*A SURVEY ON DATA MINING TECHNIQUES" i*n International Journal of Advanced Research in Computer and Communication

[24] Engineering, Vol. 3, Issue 1, PP:5002-5003, 2014.

[25] Social Network Usage & Growth Statistics: How Many People Use Social Media in 2020? [Online] Available: https://backlinko.com/social-media-users [Jan 23, 2021]

[26] GLOBAL SOCIAL MEDIA OVERVIEW [OCT 2020] [Online] Available: https://datareportal.com/social- media-users [Jan 23, 2021]

[27] Global social media research summary August 2020 [Online] Available: https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/ [Jan 23, 2021]

[28] Adedoyin-Olowe, M., Gaber, M. M., & Stahl, F, "A Survey of Data Mining Techniques for Social Media Analysis" in Journal of Data Mining & Digital Humanities, PP:1-27, 2014.

[29] S.G.S Fernando et.al "Empirical Analysis of Data Mining Techniques for Social Network Websites" in COMPUSOFT, An international journal of advanced computer technology, Volume-III, Issue-II PP:582- 592, 2014.

[30] Thabit Zatari, "Data Mining in Social Media" in International Journal of Scientific & Engineering Research, ISSN 2229-5518 Volume 6, Issue 7, PP:152-154, 2015.

[31] Dr.B.Umadevi, P.Surya, "A Review on Various Data Mining Techniques in Social Media", in International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 4, PP: 8082-8086, 2017.

[32] Mohammad Noor Injadat, Fadi Salo, Ali Bou Nassif "Data Mining Techniques in Social Media: A Survey", NEUCOM17295, Volume 214, PP:654-670, 2016.

[33] Rahman, M. M, "Mining Social Data to Extract Intellectual Knowledge", in International Journal of Intelligent Systems and Applications(IJISA), vol.4, no.10, PP:15-24, 2012.

[34] XimingWang · Panos M. Pardalos, “A Survey of Support Vector Machines with Uncertainties”, © Springer-Verlag Berlin Heidelberg 2015, Ann. Data. Sci. (2014) 1(3–4) PP:293–309, 2014

[35] D. M. Boyd and N. B. Ellison, “Social network sites: Definition, history, and scholarship” Journal of computer-mediated communication, vol. 13, no. 1, pp. 210-230, 2007.

[36] J. F. Sánchez-Rada and C. A. Iglesias, “Social context in sentiment analysis: Formal definition, an overview of current trends and framework for comparison” Information Fusion, vol. 52, pp. 344-356, 2019.

[37] C. Haythornthwaite, “Social network analysis: An approach and technique for the study of informationexchange” Library & information science research, vol. 18, no. 4, pp. 323-342, 1996.

[38] Z. Yang, R. Zheng and Y. Ma, “Parallel Heuristics for Balanced Graph Partitioning Based on Richness of Implicit Knowledge” in IEEE Access, vol. 7, pp. 96444-96454, 2019, doi: 10.1109/ACCESS.2019.2926753. [39]. Z. Zhuang, C. Wei, B. Li, P. Xu, Y. Guo and J. Ren, “Performance Prediction Model Based on Multi-Task Learning and Co-Evolutionary Strategy for Ground Source Heat Pump System” in IEEE Access, vol. 7, pp. 117925-117933, 2019, doi: 10.1109/ACCESS.2019.2936508.

[39] X. Gong and S. Zhu, “Person Re-Identification Based on Two-Stream Network With Attention and Pose Features” in IEEE Access, vol. 7, pp. 131374-131382, 2019, doi: 10.1109/ACCESS.2019.2935116.

[40] Anu Sharma et.al, “Literature Review and Challenges of Data Mining Techniques for Social Network Analysis”, Advances in Computational Sciences and Technology ISSN 0973-6107 Volume 10, Number 5 (2017) pp. 1337-1354

[41] Anu Sharma et.al , “Hybrid Neuro-Fuzzy Classification Algorithm for Social Network”, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-6, August 2019 https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8673951&tag=1