Available Online at www.ijcsmc.com/journal

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 11, Issue. 9, September 2022, pg.34 – 40

SENTIMENT ANALYSIS OF REDDIT COMMENTS

Abiona A.A.[1]; Ajinaja M.O.[2]; Olaniyan O.L.[3]

[1, 2, 3]Department of Computer Science, Federal Polytechnic Ile-Oluji, Ondo State Nigeria

Corresponding Email: [1] akeabiona@fedpolel.edu.ng

[2] micajinaja@fedpolel.edu.ng

[3] olaolaniyan@fedpolel.edu.ng

DOI: https://doi.org/10.47760/ijcsmc.2022.v11i09.004

Abstract: A significant amount of data is generated as well as being made available to internet users thanks to the development and growth of online technologies. The internet has developed into a forum for online education, idea sharing, and opinion exchange. Social networking services like Twitter, Facebook, and Google+ are quickly gaining popularity as a result of the ability for users to share and express their opinions on many subjects, engage in conversation with various communities, and broadcast messages globally. Reddit data sentiment analysis has been the subject of a great deal of research. This study focuses on Reddit comments from the AMC, Clover, and GameStop subreddits using sentiment analysis. We provide study on Reddit comments where a radar chart was produced displaying comparison between AMC, Clover, and GameStop using machine learning methods Network analysis.

Keywords: sentiment analysis, network analysis, reddit

© 2022, IJCSMC All Rights Reserved

1. Introduction

Sentiment analysis, often known as opinion mining, is a natural language processing (NLP) method for identifying the positivity, negativity, or neutrality of data [1]. Businesses frequently do sentiment analysis on textual data to track the perception of their brands and products in customer reviews and to better understand their target market. Businesses frequently employ it to analyze social media data for sentiment, evaluate brand repute, and comprehend clientele. It is also the process of identifying positive or negative sentiment in text. It is primarily concerned with a text's polarity (positive, negative, or neutral), but it also extends beyond polarity to identify certain moods and emotions (angry, joyful, sad, etc.), urgency (urgent, not urgent), and even intents (interested v. not interested).

The world of social media is altering as a result of web 2.0's introduction. Not only are people using social media to connect, exchange information, and express their personal opinions with others, but businesses can also utilize social media to connect, learn about, and improve their products and services [2]. Every day, more people use social media, and it is predicted that by 2019, there will be 2.77 billion users worldwide [3]. Social media is full with raw, unprocessed data, and technological advancements, particularly in machine learning and artificial intelligence, have made it possible to process and turn that data into meaningful information that may help the majority of commercial organizations [4].

Automatic sentiment analysis (SA) is an issue that is becoming a more popular research area. Despite the importance of SA and the variety of applications it has now, there are a number of difficulties with natural language processing that must be overcome (NLP). Recent sentiment analysis research is nevertheless plagued by theoretical and procedural problems that reduce their overall accuracy in polarity detection [5,6]. Hussein et al. [6] investigated how these problems related to the sentiment structure and how they affected the reliability of the findings. This study demonstrates that accuracy is a major concern in the most recent works on sentiment analysis and that it is hampered by issues like handling negation or domain dependence.

Therefore, in this paper, we focused on using SA to determine users' opinion about reddit comments datasets. The remaining part of this paper is organized as follows: Section 2 presents the review of related works, section 3 provides the methodology, data source, text pre-processing used in the study. Results and discussion are provided in Section 4 and Section 5 presents the conclusion.

2. Related Works

There has been lot of work in the field of sentiment analysis on different dataset. [8] worked on Sentiment Analysis of Twitter Data while giving a detailed survey technique. This study primarily focuses on sentiment analysis of twitter data, which proved useful for analyzing information in tweets where opinions are very unstructured, varied, and occasionally neutral. Additionally, the study compares and contrasts evaluation measures with existing methods for opinion mining, including lexicon-based and machine learning approaches. The study examined general issues and uses of sentiment analysis on Twitter and presented research on twitter data streams using several machine learning algorithms, including Naive Bayes, Max Entropy, and Support Vector Machine. In [2], a systematic literature review was done to see how sentiment analysis in social media and various application. Using the following reliable databases, including ACM, Emerald Insight, IEEE Xplore, Science Direct, and Scopus, a systematic review of studies published between 2014 and 2019 was conducted. Out of the 77 articles that were reviewed, 24 were chosen after the initial and thorough screening of the papers. The findings revealed that the majority of publications used the opinion-lexicon method to analyze text sentiment in social media, extracting data from microblogging sites, particularly Twitter, and using sentiment analysis in business, politics, healthcare, and other fields. [8] worked extensively on sentiment analysis has been a fascinating new area of study in artificial intelligence that combines psychology, machine learning, and natural language processing.

The experimental findings of [9]'s work on Chinese Text Sentiment Analysis Based on Extended Sentiment Dictionary demonstrated that the suggested sentiment analysis method based on extended sentiment dictionary has some degree of viability and accuracy, and that the research is significant for the sentiment recognition of the comment texts. A study on sentiment analysis for reviews of Chinese e-commerce products based on a sentiment lexicon and deep learning can be found in [10]. The study suggested a brand-new sentiment analysis model called SLCABG, which combines Convolutional Neural Networks (CNN) and attention-based Bidirectional Gated

© 2022, IJCSMC All Rights Reserved

Recurrent Units and is based on the sentiment lexicon (BiGRU). The experimental findings demonstrated that both models can significantly boost text sentiment analysis performance.

[11] developed a method for analyzing the sentiment of Danmaku videos using naive Bayes and a sentiment dictionary. By the end of the research, the technique had proven to be quite effective in determining a danmaku video's general emotional orientation and making predictions about its level of popularity. Study [12] was based Sentiment Analysis Using Refined Global Word Embeddings Based on Sentiment Concept. Text sentiment analysis using frequency-based vigorous features was undertaken by [13] where the research presented a three-step Sentiment Analysis and Prediction (SAP) solution of Text Trend through K-Nearest Neighbor (KNN). Experiments results illustrated satisfactory improvement as compared to existing solutions. [14] spent a lot of time working on sentiment analysis using the data crawler's naive bayes algorithm (Twitter). To categorize social classes or the intensity of societal attitudes, the Nave Bayes approach was employed. According to the study's findings, the relationship between Jokowi and Ma'ruf Amin had a positive sentiment polarity score of 45.45% and a negative sentiment score of 54.55%, while the relationship between Prabowo and Sandiaga had a positive sentiment score of 44.32% and a negative sentiment score of 55.68%. The accuracy was 80.90% 80.1% when the combined data was examined using the training data utilized for each presidential contender. In this study, naive bayes, svm, and K-Nearest Neighbor (K-NN) approaches were compared. These methods were tested using RapidMiner, which produced accuracy values for naive bayes of 75.58%, svm of 63.99%, and K-NN of 73.34%. [15] worked on improving the performance of sentiment analysis of tweets containing fuzzy sentiment using the feature ensemble model.

3. Methodology

The dataset contains have 18 different columns and 272 observations which has various columns like author, user, comments score and so on. Data was pulled by using the Clov package available on R studio. Figure 1 shows screenshot of the dataset.

↕	post_score ↕	author ↕	user ↕	comment_score ↕	controversiality ↕	comment ↕	title ↕	post_text ↕	link ↕
0.75	2664	OPINION_IS_UNPOPULAR	Just_a_Userrr	3	0	CLNE==<a8>>u	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	praisomnisf	1	0	Does anyone have a good graph of the short interest?	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	Organization‧road759	10	0	Buy CLNE	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	rephxsun	15	0	CLOV about to cross the line =<aa> ><8d>	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	tibbiz	13	0	CLOV ARMY IS HODLING TOGETHER!!! SO PROUD TO ...	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	rephxsun	12	0	CLOV lets go!! Edit: ><8d> forgot about TLRY	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	LimpingWhale	19	0	Ive done 0 research. CLOV TO THE MOON =<80>=<...	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	thethirdeye_	19	0	We're getting short attacked, CLOV ARMY assemble, j...	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	twitch760	34	0	So why the suddenly loss of interest in CLOV? I've sol...	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	fkypete	8	0	people see red and freak out man..... sads :(	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	AlathargicMoose	11	0	CLOV puts really printed today	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	monsterdealdough	13	0	Alright you monkeys I got into clne calls 6/18 13c.	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	ProtectYOURshelves	21	0	Officially a CLOV bag holder. >&<<fd> 8	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	Newtonhog	6	0	About to buy a couple 12$ contracts CLNE call for De...	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	CoVLifter	1	0	Like this dude said, IV is high. This weeks gonna be c...	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	lithium182	2	0	At what cost? IV is high.	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	BullishBearishBiz	12	0	Im already in for 100 shares and 5K worth of calls, an...	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	lemenick	2	1	Be glad that you didnt. News broke out that that a lar...	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red
0.75	2664	OPINION_IS_UNPOPULAR	BullishBearishBiz	12	0	Its a ridiculous fud story, that largest holder is an inst...	Daily Popular Tickers Thread for June 09, 2021 - CLO...		https://www.red

Fig. 1 – Reddit comments dataset

© 2022, IJCSMC All Rights Reserved

4. Results and Discussion

Generating the network diagram in figure 2, it was observed at the center which person posted the main comment and generated lots of interest.

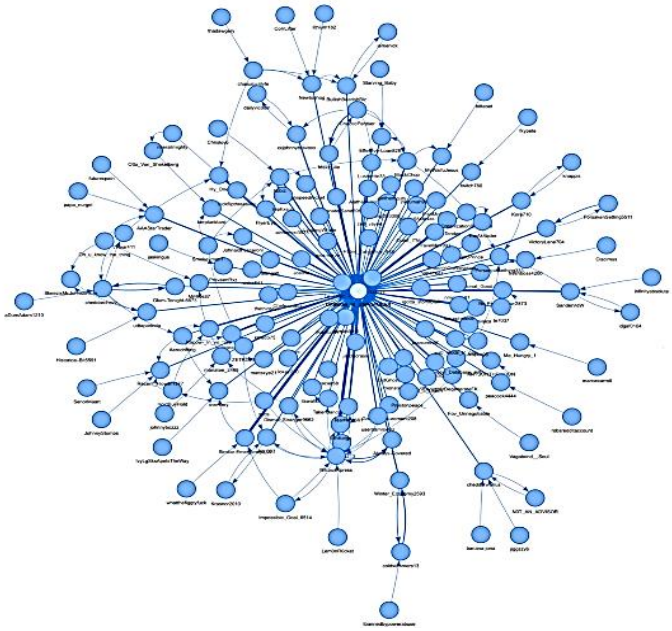

Fig 2 – Network diagram

Although, the main comment user can be seen in the center and generated lot of interests from many other participants, the network diagram looks congested and difficult to locate the branches. In other to reduce the rowdiness, the include author is set to false and the network diagram is plotted again. It is observed that the network was very well connected with many other branches and basically users seemed to be engaged very actively as seen in figure 3.

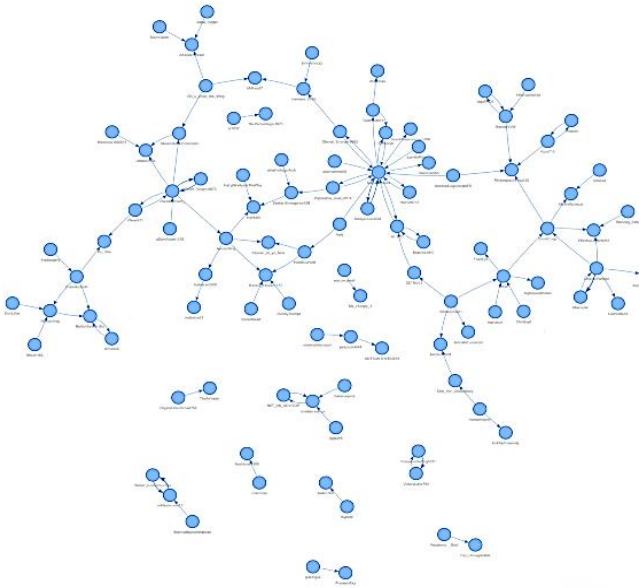

Fig 3 – network diagram with less rowdiness

© 2022, IJCSMC All Rights Reserved

Setting the network diagram to True, figure 4 showed the main user or the person who wrote the first comment and then how it generated discussions involving many other users.

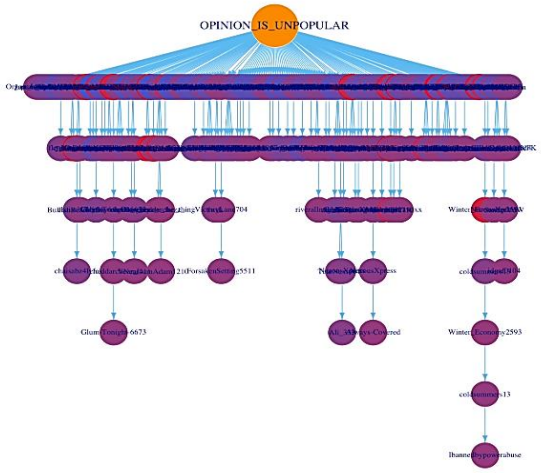

Fig 4 – Network diagram set to True

Coming to sentiment analysis, the comments are converted using icov library on R studio. The sentiments course was ran using NRC lexicon and the function that was used gate NRC sentiment. Basically, sentiment is classified into 8 categories and then there were two broad categories - positive or negative. The study considered 3 products which was AMC, Clover and GameStop. In figure 5 for Clover, it was observed in the first comment, for example, all zeros meaning there was nothing to say whether the comment is positive or negative. And second one indicated anticipation, joy, surprise, trust and had score of 2 for positive. This was a very approximate scoring method because when there is lot of data, it can give some direction about what's going on in terms of sentiment. The column sum was done and divided by the overall to convert it into a percentage by multiplying this by hundred and these was repeated for Clover, AMC and GameStop.

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
1	0	0	0	0	0	0	0	0	0	0
2	0	1	0	0	1	0	1	1	0	2
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0
9	2	2	0	1	1	1	2	2	2	2
10	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0
12	0	1	0	0	0	0	0	1	1	0
13	0	0	0	0	0	0	0	0	0	0
14	0	3	0	0	1	0	0	0	0	3
15	1	0	0	1	0	1	0	0	1	0
16	0	0	0	0	0	0	0	0	0	0
17	0	2	0	1	1	0	0	2	1	2
18	0	1	0	1	1	1	0	0	1	1
19	2	0	1	1	1	0	0	1	3	2
20	0	0	0	0	0	0	0	1	0	1
21	0	2	0	0	0	0	0	1	1	2
22	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0

Fig 5 – Clover Sentiment

© 2022, IJCSMC All Rights Reserved

The row names are finally extracted for the 3 products. The row names are basically the sentiments in figure 6 with positive the highest for Clover.

	Clover	AMC	GameStop
anger	6.260032	8.301405	7.453754
anticipation	14.446228	13.026820	12.840044
disgust	2.086677	3.703704	2.230686
fear	4.654896	6.385696	4.678999
joy	9.309791	7.662835	9.303591
sadness	4.654896	5.747126	5.821545
surprise	7.223114	5.108557	6.909684
trust	15.890851	13.282248	16.376496
negative	12.841091	17.241379	13.112078
positive	22.632424	19.540230	21.273123

Fig 6 – Clover, AMC and GameStop Sentiment Analysis

The radar chart can be generated and it is shown in figure 7. It should be noted that this is an approximate sentiment analysis so a very generic lexicon has been used but overall, it gave some direction if clover was compared with AMC. Red one is clover while green one is AMC. It was observed one thing seemed to stand out was that negative comments for AMC were beginning to be slightly on the higher side than for both Clover and GameStop.

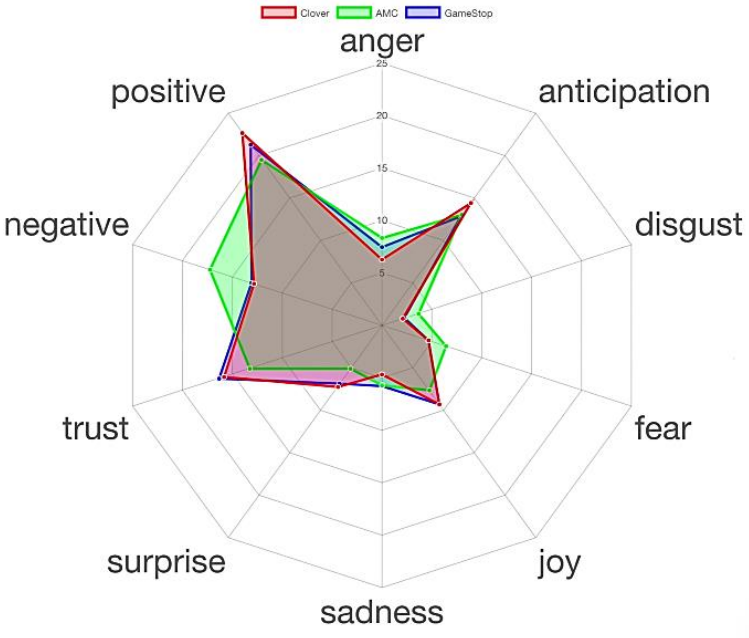

Fig 7 – Radar Chart for AMC, Clover and GameStop

They were more or less at a smaller value compared to AMC. Similarly, percentage of positive comments for Clover seemed to be the highest followed by GameStop and AMC was lagging behind a little bit. Also, GameStop and

© 2022, IJCSMC All Rights Reserved

clover had higher trust values, higher joy in comments and anticipation were all nearby. Fear was slightly high for AMC which gave a very broad overview of what people were discussing in their comments and specially those comments, which are engaging more and more users.

5. Conclusion

In this work we have shown the analysis of reddit comments on 3 different products. Our empirical results were obtained on Reddit messages gathered from AMC, Clover and GameStop subreddits. We extracted reddit comments and then classified them into 8 different sentiment categories amd also created a network diagram. The radar chart generated observed that negative comments for AMC were beginning to be slightly on the higher side than for both Clover and GameStop. The majority of the positive remarks appeared to be for Clover, followed by GameStop and a slight lag for AMC. Additionally, Clover and GameStop had greater trust ratings, greater happiness in comments, and anticipation were all close by. Fear was slightly elevated for AMC, which provided a very thorough picture of what was being discussed in comments, particularly in those comments that were drawing in an increasing number of users.

References

[1] Retrieved from https://monkeylearn.com/sentiment-analysis/

[2] D. Zulfadzli and K. Haliyana. Sentiment Analysis in social media and Its Application: Systematic Literature Review. The Fifth Information Systems International Conference 2019, Procedia Computer Science 161 (2019) 707–714

[3] Statista. (2019) Number of social media users worldwide 2010-2021. Available from: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users

[4] Sivarajah, Uthayasankar, Muhammad Mustafa Kamal, Zahir Irani, and Vishanth Weerakkody. (2017) "Critical Analysis of Big Data Challenges and Analytical Methods." Journal of Business Research 70: 263-286.

[5] Cambria, E.; Das, D.; Bandyopadhyay, S.; Feraco, A. A Practical Guide to Sentiment Analysis; Springer: Berlin, Germany, 2017.

[6] Hussein, D.M.E.-D.M. A survey on sentiment analysis challenges. J. King Saud Univ. Eng. Sci. 2018, 30, 330–338.

[7] Vishal A. Kharde & S.S. Sonawane. Sentiment Analysis of Twitter Data: A Survey of Techniques: International Journal of Computer Applications (0975 – 8887). Vol. 139(11), 2016

[8] P. Bhattacharyya, "Sentiment Analysis," 2013 1st International Conference on Emerging Trends and Applications in Computer Science, 2013, pp. i-i, doi: 10.1109/ICETACS.2013.6691379.

[9] G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng and X. Wu, "Chinese Text Sentiment Analysis Based on Extended Sentiment Dictionary," in IEEE Access, vol. 7, pp. 43749-43762, 2019, doi: 10.1109/ACCESS.2019.2907772.

[10] L. Yang, Y. Li, J. Wang and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," in IEEE Access, vol. 8, pp. 23522-23530, 2020, doi: 10.1109/ACCESS.2020.2969854.

[11] Z. Li, R. Li and G. Jin, "Sentiment Analysis of Danmaku Videos Based on Naïve Bayes and Sentiment Dictionary," in IEEE Access, vol. 8, pp. 75073-75084, 2020, doi: 10.1109/ACCESS.2020.2986582.

[12] Y. Wang, G. Huang, J. Li, H. Li, Y. Zhou and H. Jiang, "Refined Global Word Embeddings Based on Sentiment Concept for Sentiment Analysis," in IEEE Access, vol. 9, pp. 37075-37085, 2021, doi: 10.1109/ACCESS.2021.3062654.

[13] A. Razzaq et al., "Text sentiment analysis using frequency-based vigorous features," in China Communications, vol. 16, no. 12, pp. 145-153, Dec. 2019, doi: 10.23919/JCC.2019.12.011.

[14] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," 2019 Fourth International Conference on Informatics and Computing (ICIC), 2019, pp. 1-5, doi: 10.1109/ICIC47613.2019.8985884.

[15] H. T. Phan, V. C. Tran, N. T. Nguyen and D. Hwang, "Improving the Performance of Sentiment Analysis of Tweets Containing Fuzzy Sentiment Using the Feature Ensemble Model," in IEEE Access, vol. 8, pp. 14630-14641, 2020, doi: 10.1109/ACCESS.2019.2963702.

© 2022, IJCSMC All Rights Reserved