Sentiment Analysis: A Case Study among the selected Government Agencies in the Philippines

Julieta M. Umali[1], John Paul P. Miranda[2], Anicia L. Ferrer[3]
[1]Don Honorio Ventura State University, Philippines, jemumali17.new@gmail.com
[2]Don Honorio Ventura State University, Philippines, johnpaul.p.miranda@gmail.com
[3]Don Honorio Ventura State University, Philippines, annethferrer@yahoo.com

ABSTRACT

This paper describes a sentiment analysis method to analyze comments polarity and to facilitate selected government agencies such as DepEd, DPWH, DILG, DND, DSWD, and DOH to describe quantitatively the opinions and comments of active users and citizens on social networks such as YouTube. The method used was from the latest comments from the first 100 were selected from each video to form the final datasets used in the study. The comments were extracted from 1450 videos and contain a total number of 14,506 posts. The objectives of the study were (1) to analyze the citizen's attitude among the six government agencies in the category of positive and negative polarity. (2) to utilized the results in developing a decision support system.

Key words : Sentiment analysis, social media, text mining, lexicon-based analysis.

1. INTRODUCTION

Sentiment analysis is mainly used for data mining and analytics, particularly in judging human behavior [1-2]. It is also used to extract meaningful insights among people's opinions and feelings [3], [4]. Sentiment analysis also provides shared results of thoughts of the masses based on the enormous text data that can be found and are readily available over the internet [5], [6], [7], [35]. Multiple studies have used sentiment analysis to discover important context within a collection of social discourse [5], [8], [9], [10], as a tool to retrieve and extract contents and features relating to specific topics [4] [11], [12]. Different social networking platforms is used and has a huge potential as a source for retrieving vast amounts of information [13-14] are mostly unorganized [15], [34], [35] and in particular, it can be used to determine the opinion of the public on a variety of subjects [9]. Also, the widespread use of social networks makes its users to actively involve with social issues and trends, and through comments, feedbacks and posts, it encourages analytical analysis and discussion among its user.

It is increasingly important to learn about citizens' emotions through online text comments. Such analysis of citizen's emotional feeling has been considered as a problem of polarity classification. Literatures cited that almost 71% of the Philippines are active users of social media and the most used platform is the Facebook. Filipino are known to have spent the most number of hours in social media with an average of four (4) hours per individual a day [16]. Social networking sites have also become a tool to allow the public access to information as well as a means to communicate with the head of government agencies. Posting of comments online is an effective way to express individual feelings.

YouTube is one of the leading video contents sharing site used by 87% of Filipino internet users watch online video content [17]. The English language is the preferred and best executed for sentiment analysis [18]. However, since the scope of the study is within the Philippines, Filipino in social media uses code switch between English and Tagalog or more commonly known as "Taglish" because the majority of them are raised using these two languages [19] and this was used for sentiment analysis. With this, the study used a lexicon-based classification to identify the sentiment of the comments. The study combined two (2) lexicons, an English lexicon [20] and the Tagalog lexicon [21], latter was created using graph propagation based on knowledge graph where the assumption is that words that are closely linked to each other on a knowledge graph have the same polarity [21].

The study adds to the current research works and discussions that might help policy makers in understanding the views and attitudes of the public towards a particular government department. About this, the study gathered users' comments from YouTube videos from the top six (6) government agencies in the Philippines based on the 2019 national budget [22]. The main objective of the study is to analyze YouTube users' comments through sentiment analysis. Specifically, the study sought to answer the following: 1. How can the users' comments on YouTube videos about the six (6) government departments be described in terms of volume, most occurring words, lexicon-based classification, and its overall sentiment? 2. How the results be utilized in developing a decision support system?

2. LITERATURE REVIEW

Websites are becoming huge source of customer voices as online platforms expand. Recently, the gathering and study of public opinions from social media platforms has experienced a major burst of research activities [23], [24]. One of the emerging fields is the study of sentiments [25]. Hundreds of articles were subsequently released on the topic. Among these papers, we concentrate on the work discussed in this paper regarding sentiment analysis as follows: several researches adopted a somewhat different approach by splitting into two tasks of sentence classification namely: (1) subjectivity and objectivity, and (2) polarity. [25], [26], [27]. In addition, a few studies followed an approach involving determining the polarity of sentences using multiclass classifiers in a single step; that is, a 3-class classifier was used. [28], [29], [30].

Sentiment analysis has a variety of applications, including but not limited to understanding the opinions from YouTube learners towards programming [36], political campaigns [37], product and services [38], public emergency [39], and government public policy [40]. Analysis of opinions can also be useful for government agencies responsible for home land protection issues. Buladaco et al [24] has used sentiment analysis for analyzing land transportation infrastructure. In similar case, monitoring spikes in negative feelings against a particular government institution or authority or even country, agencies may gather valuable information on emerging threats. In 2010, sentiment analysis was suggested by Gartner through Open Government Maturity Model as a way to achieved collaboration for governments [31]. It echoes the pattern, Forrester Research [32] observed that the United States Federal Government tracked citizens' feelings on Twitter and advised the use of social networking sites to achieve collective budgeting and trend discovery where citizen sentiment analysis could play a significant role in these platforms. Studies from these sites from the Malaysian government are significant to achieve their social media operations, providing the most effective public-to-user contact systems. Through doing this, policymakers should keep an eye on potential public opinion and use those that show how their people respond to services [33], [34].

3. METHODS

The study's objective is to perform sentiment analysis to a collection of YouTube user's comments relating to different Philippine government agencies and compare the results to one another.

3.1 Data Collection Procedure, Sample Size, and Sampling Design

YouTube comments were extracted using the python program and analyzed through sentiment analysis. Videos are divided

and selected based on the top 6 recipients of the General Appropriations Act of 2019 in the Philippines. Different keywords were used to search for the videos such as "DepEd Philippines" for the Department of Education (DepEd), "DPWH Philippines" for the Department of Public Works and Highways (DPWH), Department of Interior and Local Government (DILG), "DND Philippines" for the Department of National Defense (DND), "DSWD Philippines" for the Department of Social Welfare and Development (DSWD) and "DOH Philippines" for the Department of Health (DOH). Videos that can viewed only in the Philippines from January 1, 2018 were searched and sorted by relevance and only the videos whose title or description contains the keywords mentioned above (e.g., "DepEd", "DPWH", and "DILG") were used in the study. The latest comments from the first 100 were selected from each video to form the final datasets used in the study. Furthermore, the extracted comments were then combined then stored into a different spreadsheet. It should also be mentioned that the researchers manually checked and removed the videos that were not related to the keyword and were not part of the Philippines, and then the videos were replaced by new video after the last selected video with the same criteria mentioned above. The comments were extracted from 1450 videos and contain a total number of 14,506 comments shown in Table 1. Additionally, the researchers acknowledged the possible biases in the results of the study.

3.2 Data Cleaning and Data Pre-processing

As first step in sentiment analysis, the data needs to be cleaned and preprocesses so that it will be easier to extract insights from the dataset. One of the objectives for data cleaning is to reduce the noise, dimensionality, and inconsistency within the datasets. Since the data contains mixed English and Tagalog texts, a Tagalog stop words published in Node Package Manager (NPM) were used in this study. The English stop words were then combined. The combined stop words, punctuation, unrecognizable characters, special characters, and numbers within the dataset were removed, as they do not carry much weight in the context of the study. The Tagalog stop words came from the bower library. Words that are equal or less than three characters were also removed. Lastly, the data were then stemmed, transformed to lowercase and tokenized

3.3 Data Analysis

For data analysis, the researcher combined two (2) positive and negative lexicons. The first lexicon came from [17] which consist of approximately 6800 English words and the second lexicon came from [16] [18] is exactly 1858 Tagalog and English words. The two lexica were then combined the duplicates were then removed to form the final lexicon for the

study, which consists of 8206 words. The final lexicon was used to classify the sentiments in each dataset. The top five (5) most frequently used words per department and its overall sentiment was reported. Likewise, the sentiments of each department were extracted and were compared to one another. Word clouds were used to illustrate the sentiment of each department.

4. RESULTS AND DISCUSSION

Excluding the keywords used in searching for the videos, "good", were the most frequently used word by DepEd and DPWH, while "kapa", "fake", "pera", "thank", are DILG, DND, DSWD, and DOH respectively. As shown in table 1, DILG tops the departments with the highest number in terms of comment volume while DND has the greatest number of videos. By combining the lexicons of [16] and [17], the study was able to improve the results of the study and solved the problem with bilingual comments. The word "good" were present across all departments and it is within their top 5 in terms of frequency (tables 2 -8). The polarity of words between positive and negative in all departments are seemingly balanced and the overall sentiment is slightly positive (figures 1 – 6).

Table 1: Departments sorted by the number of extracted comments

Rank	Department	No. of Videos	Comments Counts
1	DILG	344	4341
2	DPWH	58	2810
3	DND	418	2703
4	DSWD	260	2205
5	DepEd	297	1940
6	DOH	73	507
	TOTAL	1450	14,506

Table 2: Five Most Occurring Words in DepEd Videos

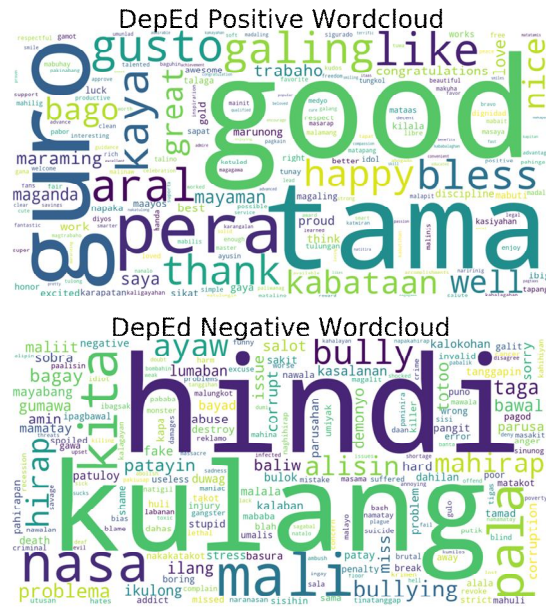| Rank	DepEd	Frequency	Classification
1	good	70	Positive
2	tama	52	Positive
3	hindi	51	Negative
4	kulang	42	Negative
5	guro	39	Positive

Figure 1: Word clouds of Comments on YouTube Videos about DepEd sorted by Sentiment

Table 3: Five Most Occurring Words in DPWH Videos

Rank	DPWH	Frequency	Classification
1	good	119	Positive
2	mabuhay	96	Positive
3	basura	85	Negative
4	thank	72	Positive
5	galing	69	Positive

Figure 2: Word clouds of Comments on YouTube Videos about DPWH sorted by Sentiment

Table 4: Five Most Occurring Words in DILG Videos

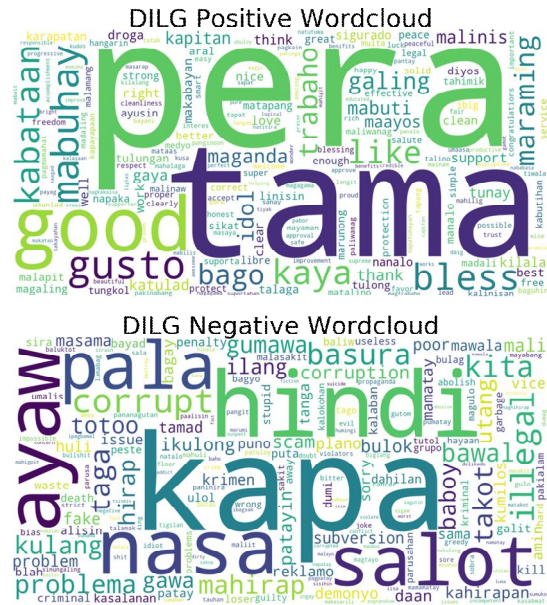| Rank	DILG	Frequency	Classification
1	kapa	389	Negative
2	pera	161	Positive
3	tama	147	Positive
4	good	146	Positive
5	hindi	144	Negative

Figure 3: Word clouds of Comments on YouTube Videos about DILG sorted by Sentiment

Table 5: Five Most Occurring Words in DND Videos

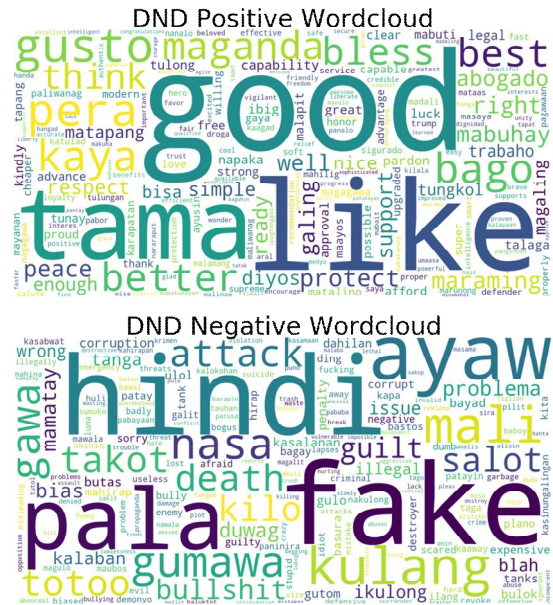| Rank	DND	Frequency	Classification
1	fake	85	Negative
2	hindi	81	Negative
3	good	67	Positive
4	like	54	Positive
5	tama	53	Positive

Figure 4: Word clouds of Comments on YouTube Videos about DND sorted by Sentiment

Table 6: Five Most Occurring Words in DSWD Videos

Rank	DSWD	Frequency	Classification
1	pera	165	Positive
2	tama	117	Positive
3	pala	100	Negative
4	mabuhay	92	Positive
5	good	87	Positive

Figure 5: Word clouds of Comments on YouTube Videos about DSWD sorted by Sentiment

Table 7: Five Most Occurring Words in DOH Videos

Rank	DOH	Frequency	Classification
1	thank	28	Positive
2	good	26	Positive
3	gusto	21	Positive
4	support	19	Positive
5	gamot	19	Positive

Figure 6: Word clouds of Comments on YouTube Videos about DOH sorted by Sentiment

Table 8: Polarity of Words

Department	Positive	Negative	Overall	Lexicon-based Sentiment
DepEd	1345	1119	226	Positive
DPWH	2152	1781	371	Positive
DND	1614	1887	-273	Negative
DILG	3657	4023	-366	Negative
DSWD	2646	2174	472	Positive
DOH	528	428	100	Positive
TOTAL	11942	11412	530	Positive

5. CONCLUSION

In this study, researchers presented that sentiment analysis is one of the best ways of interpreting social media comments of users that are based on their own personal opinion and experience. Since the study aims to understand the sentiment of the YouTube users from the Philippines, the study collected 14,506 comments from 1450 videos viewable from the Philippines from January 01, 2018 wherein DILG rank as top 1 with the highest number of comments and it contains the highest number of both positive and negative sentiments. Most of the departments presented in the study reported positive sentiments in which DSWD reported highest overall sentiment while DILG reported the lowest among the other departments. The study was able to present the most frequently used words of YouTube users through Word clouds per department using lexicon-based approach. Using lexicon-based sentiment analysis, the study presented a way to obtain the perception of the YouTube users towards government organizations with bilingual comments.

This study contributes towards better understanding the perceptions of YouTube users in the Philippines towards the videos related to the top six (6) government agencies through their comments. The study concludes that lexicon-based approach for sentiment analysis can solve the problem with bilingual comments like in the case of the Philippines, where most YouTube users often used "Taglish" or a mixture of English and Tagalog. Additionally, the study proposes that future researchers should add more Tagalog words and a neutral "Tagalog" lexicon should be developed and an additional keyword other than mentioned in the methodology should be used to improve and acquire more reliable data from YouTube videos. Finally, the results of this study may be utilize by decision makers to understand the views of its citizens and it is suggested that government organization should consider in developing a decision support system that supports bilingual content that will help them understand the sentiments of its citizens based on the methodology mentioned in this study.

REFERENCES

1. K. Chakraborty, S. Bhattacharyya, R. Bag, and A. A. Hassanien. **Sentiment Analysis on a Set of Movie Reviews Using Deep Learning Techniques**, *Social Network Analytics*, pp. 127-147, 2019. https://doi.org/10.1016/B978-0-12-815458-8.00007-4
2. R. Kumar and S. Garg. **Aspect-Based Sentiment Analysis Using Deep Learning Convolutional Neural Network**, in *Tuba M., Akashe S., Joshi A. (eds)*

Information and Communication Technology for Sustainable Development. Advances in Intelligent Systems and Computing, Singapore, 2019, vol. 933, pp. 43-52.

3. B. Liu. **Sentiment Analysis and Opinion Mining**, *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp.1-167, May 2012.

4. F. Bisio, C. Meda, P. Gastaldo, R. Zunino, and E. Cambria. **Sentiment-Oriented Information Retrieval: Affective Analysis of Documents Based on the SenticNet Framework**, in *Pedrycz W., Chen SM. (eds) Sentiment Analysis and Ontology Engineering. Studies in Computational Intelligence*, vol. 639, pp. 175-197, Cham, March 2016.
   https://doi.org/10.1007/978-3-319-30319-2_8

5. M. Godsay. **The Process of Sentiment Analysis: A Study**, *International Journal of Computer Applications*, vol. 126, no. 7, pp. 26-30, Sept. 2015.
   https://doi.org/10.5120/ijca2015906091

6. E. Cambria, B. Schuller, Y. Xia, and C. Havasi. **New Avenues in Opinion Mining and Sentiment Analysis**, *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15-21, April 2013.
   https://doi.org/10.1109/MIS.2013.30

7. P. Baid, A. Gupta, and N. Chaplot. **Sentiment Analysis of Movie Reviews using Machine Learning Techniques**, *International Journal of Computer Applications*, vol. 179, no. 7, pp. 45-49, Dec. 2017.

8. Z.Kokkinogenis, J. Filguieras, S. Carvalho, L. Sarmento, and R. J. F. Rossetti. **Mobility Network Evaluation in the User Perspective: Real-Time Sensing of Traffic Information in Twitter Messages**, *Advances in Artificial Transportation Systems and Simulations*, ch. 12, pp. 219-234, 2015.

9. B. Dahal, S. A. P. Kumar, and Z. Li. **Topic modeling and sentiment analysis of global climate change tweets**, *Social Network Analysis and Mining*, vol. 9, pp. 9-24, June 2019.

10. M. Tsytsarau and T. Palpanas. **Survey on mining subjective data on the web**, *Data Mining and Knowledge Discovery*, vol. 24, pp. 478-514, May 2012.

11. M. Ravichandran, G. Kulanthaivel, and T. Chellatamilan. **Intelligent Topical Sentiment Analysis for the Classification of E-Learners and Their Topics of Interest**, *The Scientific World Journal*, pp. 1-8, Feb. 2015.
    https://doi.org/10.1155/2015/617358

12. P. Ficamos and Y. Liu. **A Topic based Approach for Sentiment Analysis on Twitter Data**, *International Journal of Advanced Computer Science and Applications,* vol. 7, no. 12, pp. 201-205, 2016.

13. M. Injadat, F. Salo, and A. B. Nassif. **Data mining techniques in social media: A survey**, *Neurocomputing*, vol. 214, pp. 654-670, Nov. 2016.

14. H. Chen, R. Chiang, V. Storey. **Business Intelligence and Analytics: From Big Data to Big Impact**, *MIS Quarterly*, vol. 36, no. 4, pp. 1165-1188, Dec. 2012.

15. A. Kavanaugh, E. Fox, S. Sheetz, S. Yang, L. T. Li, D. Shoemaker, A. N. Xie. **Social media use by government: From the routine to the critical**, *Government Information Quarterly*, vol. 29, no. 4, pp.480-491, Oct. 2012.
    https://doi.org/10.1016/j.giq.2012.06.002

16. G. Gonzales. **Filipinos spend most time online, on social media worldwide – report**, *Rappler*, Feb. 2019

17. T. Smith. **Trends 17: The Trends to Watch in 2017**, 2017.

18. J. Contreras, M. Ballera, A. Lagman, and J. Raviz. **Lexicon-based Sentiment Analysis with Pattern Matching Application using Regular Expression in Automata**, in *ICIT 2018: Proceedings of the 6th International Conference on Information Technology: IoT and Smart City,* Hongkong, pp. 31-36, Dec. 2018.
    https://doi.org/10.1145/3301551.3301596

19. R. Thompson. **Filipino English and Taglish. Language Switching from Multiple Perspectives,** *John Benjamins,* pp. 301-328, 2003.

20. M. Hu and B. Liu. **Mining Opinion Features in Customer Reviews,** in *AAAI'04: Proceedings of the 19th national conference on Artificial intelligence,* pp.755-760, July 2004.

21. Y. Chen and S. Skiena. **Building Sentiment Lexicons for All Major Languages**, in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics,* pp. 383-389, June 2014.
    https://doi.org/10.3115/v1/P14-2063

22. Department of Budget and Management. **Reforming the Philippine Budgeting System**, pp. 5. 2019

23. S. Shukri, R. Yaghi, I. Aljarah, and H. Alsawalqah. **Twitter sentiment analysis: A case study in the automotive industry**, in *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT),* Nov. 2015.

24. M. V. Buladaco, J. Buladaco, and L. Cantero. **Sentiments Analysis On Public Land Transport Infrastructure in Davao Region using Machine Learning Algorithms**, *International Journal of Advanced Trends in Computer Science and Engineering,* vol. 9, no. 1, pp. 658-690, Feb. 2020
    https://doi.org/10.30534/ijatcse/2020/97912020

25. B. Pang and L. Lee. **Opinion mining and sentiment analysis**, *Foundations and Trends in Information Retrieval*, vol. 2, no.1-2, pp.1-135, 2008.

26. E. Marrese-Taylor, J. Velásquez, F. Bravo-Marquez, and Y. Matsuo. **Identifying Customer Preferences about Tourism Products Using an Aspect-based Opinion Mining Approach**, *Procedia Computer Science,* vol. 22, pp. 182-191, Oct. 2003.
    https://doi.org/10.1016/j.procs.2013.09.094

27. E. Riloff and J. Wiebe. **Learning Extraction Patterns for Subjective Expressions**, in *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP-03),* United States, pp. 105-112, 2003.

28. G. Ganu, N. Elhadad, and A. Marian. Beyond the Stars: Improving Rating **Predictions using Review Text Content**, in *12th International Workshop on the Web and Databases,* Rhode Island, pp. 1-6, June 2009.

29. D. Gräbner, M. Zanker, G. Fliedl, and M. Fuchs. **Classification of Customer Reviews based on Sentiment Analysis**, in *Fuchs M., Ricci F., Cantoni L. (eds) Information and Communication Technologies in Tourism 2012*, Vienna, pp. 460-470, May 2012.

30. J. Brob. **Aspect-Oriented Sentiment Analysis of Customer Reviews Using Distant Supervision Techniques**, Ph.D. dissertation, Fachbereich Für Mathematik Und Informatik, Freie Universität Berlin, Berlin, 2013.

31. A. Di Maio. **Gartner Open Government Maturity Model**, June 2010.

32. C. Gliedman, A. Cullen, and K. Naton. **Industry Innovation: US Federal Government**, May 2011.

33. A. Mainka, S. Hartmann, W. Stock, and I. Peters. **Government and Social Media: A Case Study of 31 Informational World Cities,** in *Proceedings of the 2014 47th Hawaii International Conference on System Sciences,* Jan. 2014. https://doi.org/10.1109/HICSS.2014.219

34. J. Pinto, V. Murari, and S. Kelur. **Twitter Sentiment Analysis: A Political View**, *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 1, pp. 723-729, Feb. 2020. https://doi.org/10.30534/ijatcse/2020/103912020

35. M. Syamala and N. J. Nalini. **A Deep Analysis on Aspect based Sentiment Text Classification Approaches**, *International Journal of Advanced Trends in Computer Science and Engineering,* vol. 8, no. 5, pp. 1795-1801, Oct. 2019. https://doi.org/10.30534/ijatcse/2019/01852019

36. R. Bringula, J. N. Victorino, M. De Leon, and M. R. Estuar. **Cluster and Sentiment Analyses of YouTube Textual Feedback of Programming Language Learners to Enhance Learning in Programming**, In *Arai K., Bhatia R., Kapoor S. (eds) Proceedings of the Future Technologies Conference (FTC) 2019. FTC 2019, Advances in Intelligent Systems and Computing, vol 1070. Springer, Cham* vol. 1070, pp. 913-924, Oct. 2019.

37. R. Sandoval-Almazan and D. Valle-Cruz. **Facebook impact and sentiment analysis on political campaigns**, In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, pp. 1-7, May 2018. https://doi.org/10.1145/3209281.3209328

38. I. Juanatas, R. Fajardo, E. Manansala, A. Pasilan, J. Tabor, and H. D. Balmeo. **Sentiment Analysis Platform of Customer Product Reviews**, In *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, Dubai, pp. 230-234, Dec. 2019.

39. Z. Wu and Y. Lu. **A study on micro-blog sentiment analysis of public emergencies under the environment of big data**, In *2017 29th Chinese Control and Decision Conference (CCDC), Chongqing, 2017*, pp. 4435-4438, May 2017.

40. Y. W. Syaifudin and D. Puspitasari. **Twitter data mining for sentiment analysis on people's feedback against government public policy,** *MATTER: International Journal of Science and Technology*, vol. 3, no. 1, pp. 110-122, Jan. 2017. https://doi.org/10.20319/Mijst.2017.31.110122