The impact of sentiment analysis from user on Facebook to enhanced the service quality

Daniel D. Albesta, Michael L. Jonathan, Muhammad Jawad, Oktovianus Hardiawan, Derwin Suhartono
Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Article Info	ABSTRACT
	Facebook's influence on the modern social media platform is undoubtedly enormous. While it has gotten a backlash for its inability to control its influence over important affairs, there are still many questions regarding people's perception of Facebook and their sentiment over Facebook. This paper's role in this ongoing debate is to give a glimpse of people's sentiment and perception of Facebook in recent times. By collecting samples data from Facebook's Top Page, this paper hopes to represent a significant amount of people's aspirations towards this company. By processing the data with a processing tool to construct and model out the data and a sentiment analyzer tool helps determine the sentiment, this paper can deduce a 600-comment worth of processed data. The results from the 600 sampled comments concluded that the sentiments towards Facebook are 41.50% negative comments, 22.83% neutral comments, and 35.67% positive comments.
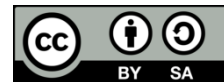
Corresponding Author:

Derwin Suhartono
Computer Science Department, School of Computer Science
Bina Nusantara University
K. H. Syahdan 9 street, Palmerah, 11480, West Jakarta, Indonesia
Email: dsuhartono@binus.edu

1. INTRODUCTION

Sentiment analysis in [1] is a text-based data quarrying method. Applied in an array of chars implied ASCII code to procure core aim: To tag, extract, and bind in-depth subjectively information of a set of the source of hid hint on info, specifically in a snippet of digital information. Millions of information currently stream out from the online social communication application such as Facebook as a part of big data complicity. Based on the research fact, up to December 2019, Facebook has been hitting up to billions of monthly active accounts, and it would keep rising in 2020 and beyond. Based on [2], over half a million comments and almost one-third of million of posts exist newly inside it regularly. Mostly, all the conveyed data inserted by the user modeled as a compilation of an interpretation of information called opinions. An opinion needed to be set-based processed, in view, it would be enhanced the analyzing progress for purposes, but the authors' field pointed to the business concern of Facebook. Enormous opinions produced each minute on the Facebook scoop may have been influencing the development of Facebook company.

An Opinion can be derived as explicitly or implicitly a sense form of satire. Both senses have a consequent on the development of the company, can be disastrously or triumphantly determined by the effectiveness of the opinion itself, so the point of the problem occurred is what the consequences of a user opinion on the Facebook platform to the Facebook service quality itself, are they would make the service be worsening or bettering? An efficient method can act as a delimiter on a view, e.g., sentiment and intent analysis. Sentiment analysis is dealing with explicit information and denotes the highlight object to be

determined if it is classified as the positive, neutral, or maybe negative sentiment. The concept of intent analysis veiled inside [3] takes a whole new level in dealing with analysis data by doing analytics procedures applied in diverse user's messages to determine whether it has any intention chemistry associates with positive or negative argumentation. As mentioned before the application of both analysis type have the potential to perform huge consequential for the company (Facebook) itself, e.g., As the source taken from [4]. As it implicitly tells us that Facebook is at the position of the lowest bound of the visualization of the graphic chart provided of user satisfaction of Facebook services, which means that the service as the primer factor does not fulfill the qualified percentage needed for the user satisfaction rate. It gives rise to the user's opinion to be bad enough for the company; at this instant, the entanglement of sentiment analysis on millions of data evolves into a primary essential task to handle corrupted opinion for the sake of Facebook's future development. A simple but comprehensive method is needed as a delimiter desire of opinion by utilizing the sentiment analysis method. This research elaborates on the analytical of sentiment objects implied within Facebook's essential elements such as posts and commentary about Facebook services as a source of Sentiment data as it is represented in the [5].

2. RELATED WORKS

Liu [6] defines the sentiment analysis as the computational study of opinions, sentiments, and emotions that are communicated through the text medium. While Balahur and Turchi [7] defines it as the task of detecting, extracting, and classifying opinions and sentiments concerning different topics, as expressed in textual input. The desires and personal attractiveness of customers stem from an emotion that arises either positively or negatively in which it provides an essential factor in determining a decision as in [8] if the content of the user comment gets detail, which means more abstract less, would be produced more emotions on it, based on the evidence occurred. As stated in [9], for some cases, there is a time when the candidates-people communicate channel utilized Facebook as their other communication media in Mexico. Even though social media is majorly used as part of the political campaign, in this era, the implementation of the sentiment analysis has no chance to predict the result of an election.

Although the analytical using the sentimental of a user has a significant impact on the latest developments, this would not happen if it does not have a method that is efficient in the process. The model is as mentioned in [10]; the authors used an action-object approach for post-classification to understand social media marketing, implied within such as the sentimentally of a user on a Facebook brand page by following some of the coding development strategies which consist of two steps, tagging, and integrating. The tagging objective is to identify and then split up different keywords on a Facebook post or comments. While integration is the part where they classify the tagging results that the authors concluded were that to run a successful Facebook page, the marketing team must be more concerned about what things commonly shared. Although the existing data processing can works with other tools as in [11], the authors are trying to achieve how to use open source technology with Hadoop based utilization to capture inputs about a brand's information and feedback through Facebook posts and comments. They outlined six steps to acquire the data, and it was already tested with six different brands. The accuracy of sentiment predictions of their testing ranges from 53.33% up to 76.76%. As the analytical using the sentimental of a user, based on a process by expressing the user opinions in a file, following that, it would continue to another process, which is the identification and categorization, these aims to predict of the writer's attitude, whether it is a lively, neutral, or maybe negative sentiment. By doing the analytical process, it would be easy to recognize whether the post contains any sentimental polarity value, also to know of what most discussed topic, it is likely to apply the K-means clustering method within it. Also, the analytical process by using a user sentiment is often used to evaluate the content within the social media platform and to forecast any emotion that occurs within the consumer for incidents that are observed on a full scale in real-time [12]. Although sentiment analysis is also used to predict personality based on users behaviour [13].

Sentiment analysis does have a significant impact today. They can comprehend other personally from their behavior [14]. From previous research, they implied an approach within the analysis with a lexicon algorithm. The fundamental theory within this approach rests on the idea, which is to comprehend the concept of lexical itself and generate some lexical sentences as pieces as an essential part of learning a language. The sentiment analysis and Lexicon based dictionary method have shared an effective rate of effort in such kinds of works, which produces a lot of functional benefits [15]. Moreover, one method to find the personality traits of the user is to read a text and to understand them. Also, the same for [16], which using Facebook because it explains which is exceedingly known well all over the world, these platforms hold a vital role in personality analysis based on users' activities. Because the concept of personality is a little bit complicated, it would presently be an issue known to be challenging to be solved. As the issue which cannot be easy to be solved, documentation comprehension can be puzzled out by using the supervised nor

unsupervised method of the machine learning algorithm [17]. Natural language processing is one of them which implemented in [18]. It is commonly used for a simple way to implement sentiment analysis when there is a small chance that learning syntax would be necessary. Based on the previous papers that discussed the various applications of sentiment analysis on multiple different subjects, all of them have somewhat the same step by step that researchers generally used when researching this subject. The implementation of sentiment analysis varies from its usage to predict personality modelling gathered from Facebook [19], to exploring its usage on Arabic slang [20], to its application in e-learning [21], and more. The other approach is to use SentiCircle and Lexicon-Based algorithms as in [22] implementing sentiment analysis in semantic contextual in the Twitter platform in 2016, which is supported by the Semicircle workflow model. To make the algorithm much better processing.

Besides, there is a way to improvise against sentiment analysis as in [23] with the sentence type classification with the implementation of the BiLSTM-CRE and convolutional neural network (CNN) algorithm in 2017. It has a variety of algorithms to help break down problems such as co-ranking, back-propagation, and other algorithms that share the way it works. There are still many other ways of using sentiment analysis as [24] there is a unique method of utilizing the hybrid cuckoo search method in the year 2017. Not only by implementing algorithms such as K-mean, cuckoo search, and other algorithms that resemblance to it, the implementation of architecture models such as the hybrid cuckoo search method flowchart model. There is also another way in the application of the sentiment analysis that is with [25] the implementation of general knowledge (commonsense knowledge) into an attentive long short-term memory (LSTM) in 2018, which has an algorithm of deep neural network and has its architectural model for semantic network. Admittedly, there are supporting methods to implement the applied algorithm that is the implementation of the method: Target-level attention, Sentic LSTM, and other similar methods.

Based on the preceding researched, mixed methods and appliances approached have been done and consummated; all of them have appropriateness and applicability to be applied to various errands. From previous research, sentiment analysis commonly exists on the Feedback activity in the scope of service apps (e.g., Facebook). Most of Facebook's users have communicational interacted through the news posts that have been posted by typing their thoughts of the comment section instead rather than doing it with personal chat. From the previous studies, one of the social media platforms used to perform text processing was Facebook. Furthermore, it explains that any automatically generated content application has a feature that is capable of being as guidance for deducing the character behavior of the user.

3. RESEARCH METHOD

The method of working can be composed of three primer objectives, which divide into variants of prominent diverge segmentations-the dataset collection segment, the data pre-processing segment, and the analyzed data visualization segment. To operate the algorithms of the methodology, as utilized the Python integrated development environment (Python IDE ver.3.7.7.) as the base of the programming language to conduct the experiments with the Python package index (PIP) as the repository of the software. Besides that, the method of working helped by with the assistance of web-based application such as open-document format based on Jupyter-notebook for Windows platform (JSON), text-analysis and machine learning web-based service (MonkeyLearn on MonkeyLearn.com), Facebook data comment's extractors via web-based services, and Microsoft Excel 2016 as the data container. For the information, it applied the method of working based on the existing methods from the internet, the method [26] applied based on a GitHub Post.

The aimed method of working at the three selectively Facebook's posts comments:
- 1st Post, Facebook Official Page-https://www.Facebook.com/Facebook/videos/190403452280238/
- 2nd Post, Mark Zuckerberg Official Page-https://www.Facebook.com/zuck/posts/10111707373631191/
- 3rd Post, Mark Zuckerberg Official Page-https://www.Facebook.com/zuck/posts/10111462102202101/

Which all of them acted as clarity of data resources. As it analyzed, it derived a much potential array of the raw data from it as the primary sources, which are loaded by approximately 600 comments combined as the sample raw data to be analyzed furthermore.

3.1. Data collection segment

As it used an online algorithm of Python that can generate the sentiment graph from a comment scrapper website (like DataExtractor.io), which scrapped the user comments through multiple Facebook Pages, as mentioned above. The utilized web scraping literally in the form of automatic tools as it is much easier and accessible rather than the manual process, which is much more complexities in its procedural occurred. In the automatic tools used, the tool would load the inputted URLs and render the entire aimed page. The tool would generate the desired result by inserting the three URLs of the dataset in earlier

mentioned in an efficient and quick feasible format without any codes involved within it. As the comments extracted, moreover, converted into a comma separated value (CSV) file formatted. Figure 1 in this case, by the website's features, the three variants of posts comments are implemented as a single source of the data. For the element of authenticity itself, the utilized datasets based on the three URLs source above, there is less, or no data changed during the data scrape process, as it would keep the data integrity over time.
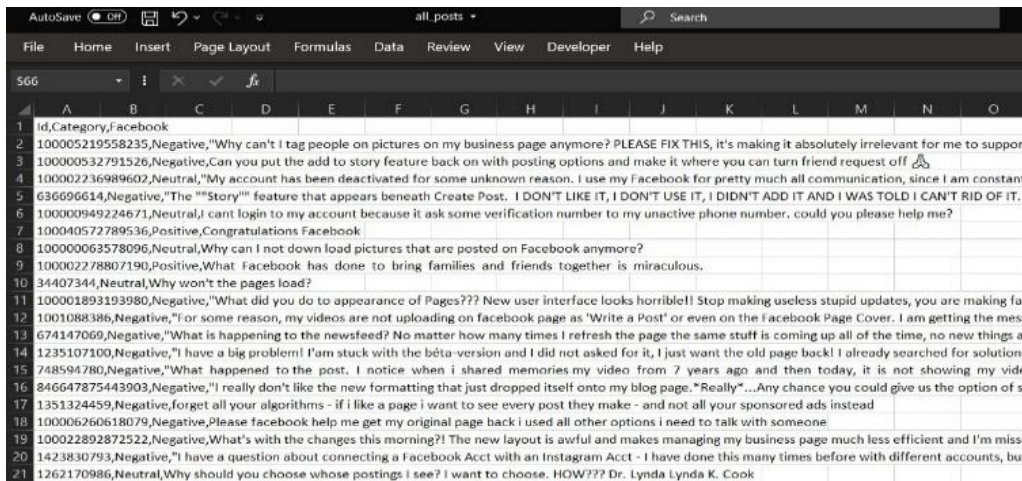
Figure 1. Sample of combined of three Facebook post's comments data

Each of the sample data output generates by the program, which for further processing to load the data form the given file, or just read from the cache already pre-processed file. Also, there is a distinction between testing processing and training data. They are being eliminated as the (CSV) file formatted was full of empty entries with the support of the practice of functioned properties such as data model, wordlist, other similar properties functioned to import the outputted raw data of variant of datasets that represented in the percentage of negative comments, neutral comments, or positive comments, represented in Figure 2.

From the first post, it purposely uses the posts filled with the most negative comments, positive comments, and mixed comments that have been showing sequential visualize. The graphical in Figure 3 is exported by the implementation of Plotly's Python graphing library, which is called by using the renderers framework) to generate the graphic chart of the raw data. In order to collect the visualization graphical, it implied by using Python codes implemented in Jupyter-notebook IDE.

Figure 2. A combined sample of three Facebook post's comments data tuples visualizations

3.2. Data pre-processing segment

The sample Facebook Post's Comment raw data firstly pre-processed by utilizing the PIP libraries include Pandas, NumPy, Plotly, and Scikit-learn library along with the implementation of other additional

libraries. Pandas correctly implemented for open-source data analysis and manipulation tools such as for opened a *.csv* extension file as in the case and for data framing and labeling of the sample raw data. NumPy is suitable as cultivating for mathematical functions, which later used. Plotly best deal in mainly for graphical visualization based on the existing data. Scikit-learn marvelous with learning or validation process occurred in the analyzing process. These libraries are mandatory for the experiment as it loads many critical functions to generate the data as aimed. Following the raw data are done imported; if the raw data want to be present to the public, it requires a complete procedure to display the raw data entirely clear.

3.2.1. Dataset cleansing

The objective of dataset cleansing segment is to extract a value of the raw data strings contained from the Facebook posts which use iteration method to identify, search, and separated each word occurred in the selective sentence on the data of tuple-n-also being cleansed from any variants of unique characters, escaped characters, and the rest itself. As with the help of utility of Regex library functions, e.g., escape checked and compiled functions, following the iteration, e.g., using lambda function() which take an argument of r: regex to executed its function with the user-defined based expression (",", ":", "\", "=", and the rest itself) to make a list of a word as desired, following the data saved and generated in an interprets of a separated array of words. If it a success, then it would throw an output consist of a cleaned, separated array of words data for approximately 600 arrays of words. As the algorithms are executed, it generates the following algorithm's product, as they represent in Figure 4.

Figure 3. Sample of the graphic chart of combined of three Facebook post's comments data tuples visualizations

Figure 4. Sample of the Facebook post's comments data following the cleansing process

3.2.2. Data tokenizing and stemming

The objective of data tokenizing or known as lexing (lexical analysis) and stemming segment, is to convert a sequence of characters that occurred into a form of a sequence of tokens (strings with an assigned and thus identified meaning as wikipedia.org source said). These processes make a partitioned of text data as a sequence of words. Which consists of an identifier (x, y, z), a keyword (case, break, return), a separator

([, {, ;), an operator (+, -, =), a literal (1.00e64, "data", true) or a comment (line, block); each word in the reduced text data classify into one or more stems. The word that identifies as a "word" is checked in this process. The text processing would occur in these steps and involve one of the most common libraries that exist in Python, which is the natural language tool kit (NLTK). It has a function named NLTK.word_tokenize function. The function work as the word_tokenize module is imported from the NLTK library; following with the argument of "text" initialized with each row of data text inside it, within the row() function itself; the row() function is passed in word_tokenize module and represents it outputs. Each row of data reduces into a minimalized the redundancy of word, then are joined back as a list of one or more stems because the text of data is in the form of English, the stemming method done using the Potters Stemmer algorithm in the process. The result represents in Figure 5, with the four-column variables to distinguished the processed data.

Out[9]:

	id	emotion	text	tokenized_text
0	100005219558235.0	Negative	[whi, cant, i, tag, peopl, on, pictur, on, my,...	[Why, cant, I, tag, people, on, pictures, on, ...
1	100000532791526.0	Negative	[can, you, put, the, add, to, stori, featur, b...	[Can, you, put, the, add, to, story, feature, ...
2	100002236989602.0	Neutral	[my, account, ha, been, deactiv, for, some, un...	[My, account, has, been, deactivated, for, som...
3	636696614.0	Negative	[the, stori, featur, that, appear, beneath, cr...	[The, Story, feature, that, appears, beneath, ...
4	100000949224671.0	Neutral	[i, cant, login, to, my, account, becaus, it, ...	[I, cant, login, to, my, account, because, it,...
...	...	...	...	...
595	100013996875293.0	Negative	[whi, peopl, charg, for, their, facebook, grou...	[Why, people, charging, for, their, Facebook, ...
596	100000965959078.0	Positive	[felicidad, mark, ✓]	[Felicidades, Mark, ✓]
597	100000109264793.0	Positive	[nice, one]	[Nice, one]
598	100007719192250.0	Positive	[♥♥]	[♥♥]
599	1750766168.0	Positive	[thank, you, for, your, updat]	[Thank, you, for, your, updates]

600 rows × 4 columns

Figure 5. Sample of the Facebook post's comments data following the tokenizing and stemming process

3.2.3. Construct the wordlist

Following the candidate data after the tokenized and stemmed procedure, as they represented in Figure 6 occurred, the candidate data supposed to count as it would produce the information about the most common occurrence of the words concatenated in the testing sample sentence. The candidate data supposed to count as it would produce the information with more restrict qualification with typical English stopwords (additional requirement added). However as the previously objective is to determine whether the dataset is as interpret of the part of sentiment analysis (positive, neutral or negative) cluster, a word such as "not" and "n't" comes big deal as it can influence the resulting process of the experiment of sentiment analysis significantly. Hence it big deals with these, then the words ("not" and "n't") would be whitelisted in the whitelist variable.

```
Out[10]: [('to', 556),
         ('the', 509),
         ('i', 472),
         ('and', 371),
         ('you', 351),
         ('a', 308),
         ('facebook', 266),
         ('it', 252),
         ('of', 236),
         ('is', 226),
         ('my', 222),
         ('for', 214),
         ('that', 188),
         ('thi', 167)]
```

Figure 6. Sample of the Facebook post's comments data following the calculation of word occurrences process

Following the candidate data after being listed of all the occurrence of words and stored all the candidate data, as they represent in Figure 7, visualized all the distinct occurrence of all words implied on a

Facebook post comments data. The graph is generated by the utilization of the Plotly library of the Python, which has a great deal with graphic analytical as the interpretation of the selective candidate data. The word roughly ordered even though already shown the occurrence of each of the words. It needs to be simplified to know that there is a connection between word occurrence and the sentiment. Each list of the individual word in the previous graph is stored, also connected within the base [emotion], which results in several occurrences words existed, with the [thanks] word variable is the most occurrence word (up to more than 70 repetitions).
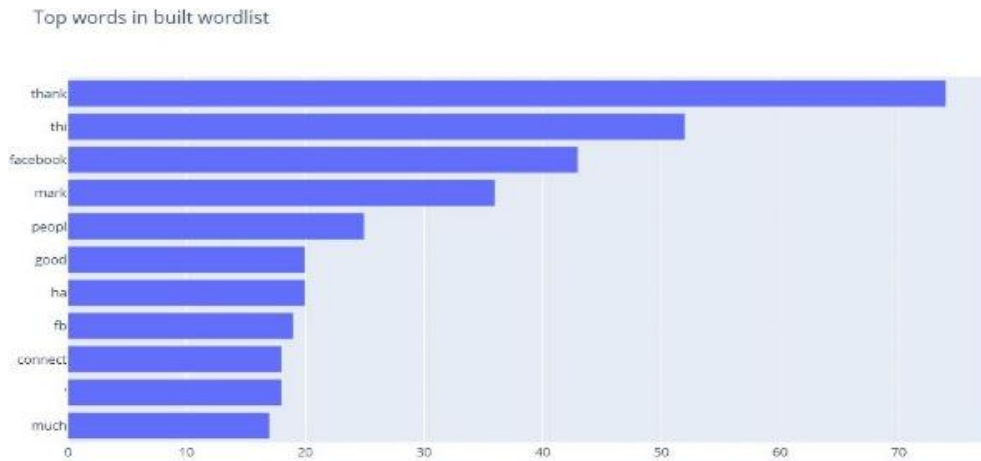

Figure 7. Sample of the graphic chart of the occurrence of the Facebook post's comments data visualization

4. RESULTS AND DISCUSSION

This study method involved Facebook's posts, and comments are evaluated with a held-out test set, of which 600 posts and comments samples data. This posts and comments uploaded in a range time of early 2020; these data evaluate near the date these posts posted to make it easier to see Facebook services improvement that is reflected from these posts and comments. The scenarios are divide into three phases, and each of the phases represents the post posted at a different time. The Figure 8 represents the post on January 30, 2020; Figure 9 represents the post on February 26, 2020; Figure 10 represents the post on March 25, 2020. It showed in Figure 10 that this post acquires more words as a negative word; these statistics tell that the quality of the services on January 30, 2020, more said to have decreased. Also, in Figure 9, each word that has many types of sentiment representation are summarized, and the quality of the services on February 26, 2020, yet still reaping negative replies-in Figure 10, already shown that on March 25, 2020, the most collected word is the word that represents the positive sentiment. It signed that the data distribution has a random pattern. So, from that three-phase can be summarized, the quality of the services on the date mentioned is increasing and reaping positive replies. The implications of this research will undoubtedly be useful for the development of the Facebook app in terms of the quality of service provided to Facebook users.
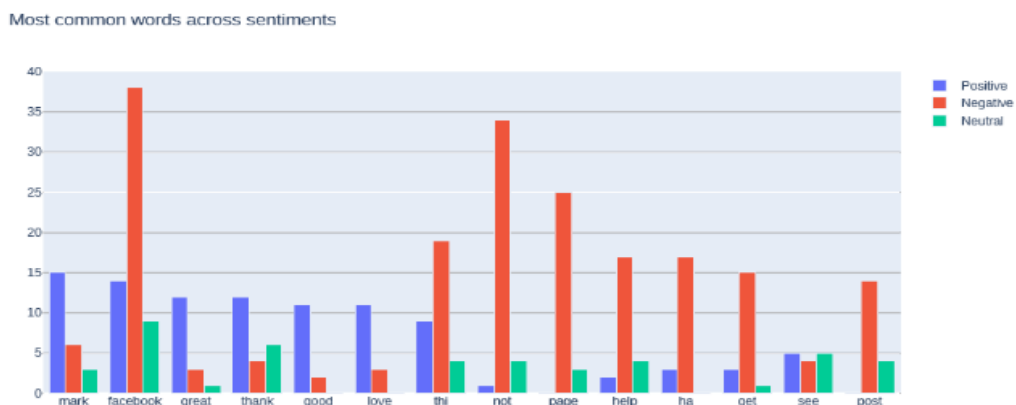

Figure 8. Sample of sentiment analysis of the Facebook post's comments data
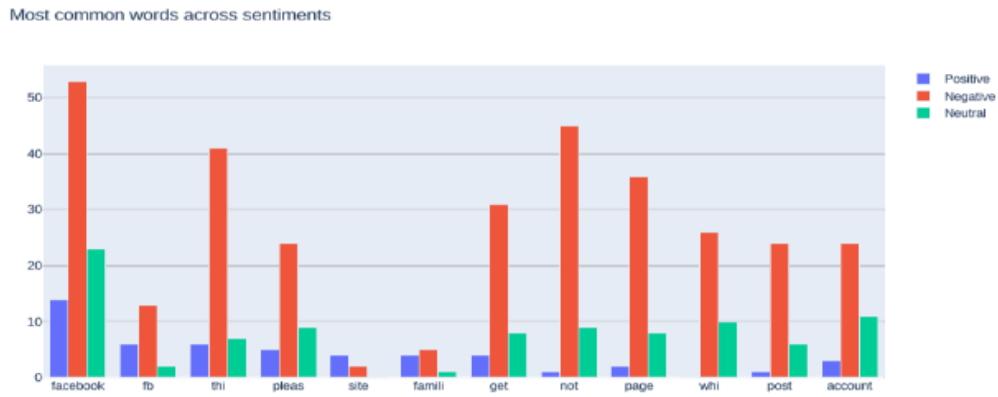
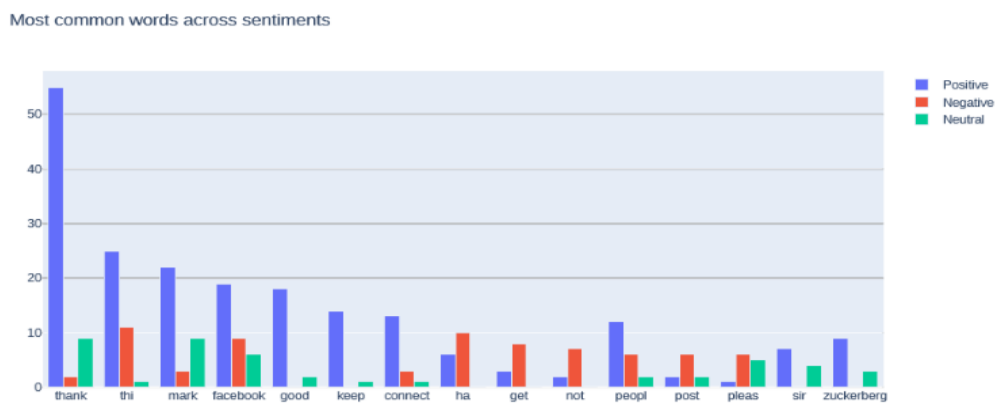Figure 9. Sample of sentiment analysis of the Facebook post's comments data

Figure 10. Sample of sentiment analysis of the Facebook post's comments data

Indirectly, the assessment of services that utilize the concept of analysis based on sentimental users has quite a complexity in the analysis that is not very complicated, with the support of python programming base that implements natural language processing algorithm and functions in machine learning of NLTK, and Scikit-learn library. There are two shortcomings of the implemented algorithm, resulting from the data distribution process in this study, which is the skewed data such as the variables [post] and [account] in Figure 9. It also happens similarly to positive and neutral comments; this shortcoming would lead to the distinguish problematic between the positive, neutral, nor negative comments from one another. Followed by the incomplete analytical process, which has not included in the data processing method, is the determination of data precision. Without the precision method involved, it also would lead to the data distrust and inconsistent problematic for the user who analyzed. Nevertheless, the strong point of these is less complexity time, more comfortable to be coded and implemented, also generates a simple but straightforward graphics rather than other algorithms such as Naïve Bayes for the data classification method or GloVe-trained model with word2vec format.

5. CONCLUSION

As this research paper is completed, it reveals the implication that occurs as it is presented in the authors' current research. The result of this paper shows an improvement advisor for Facebook's services quality future from a glimpse of people's sentiment and perception of Facebook in recent times. This current study would predict the improvement advice of Facebook's services development in the future. The future developments of this study may utilize a large testing dataset, which would allow the test to predict more accurately as improvement advice for Facebook's services development. Furthermore, the research will be conducted in terms of precision of data processing accuracy, which includes three main factors in determining the level of sentimental accuracy of the user, namely accuracy, Recall, and F1 Score factor. By focusing on these factors, it will result in a fair, correct, and accurate analytic process.

REFERENCES

[1] Gupta, Shashank, "Sentiment Analysis: Concept, Analysis, and Applications," *Towardsdatascience.com*, [Online]. Available: https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17.

[2] Noyes, Dan, "The Top 20 Valuable Facebook Statistics-Updated January 2020," *zephoria.com*, [Online]. Available: https://zephoria.com/top-15-valuable-Facebook-statistics.

[3] Pathak, Lalit, Mulchandani, Sanjana, Kate, Mayuri, Khatwani, Kajal, Khedkar, Sujata, "Sentiment and Intent Analysis for Business Intelligence," *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI),* vol. 6, no. 1, 2018, pp. 348-354, doi: 10.15680/IJIRCCE.2018.0601059.

[4] Richter, Felix, "Toxic Content Runs Rampant on Facebook," *Statista.com*, [Online] Available: https://www.statista.com/chart/13875/Facebooks-spring-cleaning/, [accessed March 2, 2020].

[5] Richter, Felix. "Facebook Keeps on Growing," *Statista.com*, [Online] Available: https://www.statista.com/chart/10047/Facebooks-monthly-active-users

[6] Liu, B., "Sentiment analysis and subjectivity," *Handbook of natural language processing,* vol. 2, no. 2010, pp. 627-666, 2010.

[7] Balahur, A., and Turchi, M., "Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis," *Computer Speech and Language,* vol. 28, no. 1, pp. 56-75, 2014.

[8] Franzoni, V., Li, Y., and Mengoni, P., "A path-based model for emotion abstraction on Facebook using sentiment analysis and taxonomy knowledge," *Proceedings of the International Conference on Web Intelligence-WI,* 2017, pp. 947-952, doi:10.1145/3106426.3109420.

[9] Sandoval-Almazan, R., and Valle-Cruz, D., "Facebook impact and sentiment analysis on political campaigns," *Proceedings of the 19th Annual International Conference on Digital Government Research Governance in the Data Age – Dgo,* 2018, pp. 1-7, doi:10.1145/3209281.3209328.

[10] Pletikosa, Irena and Michahelles, Florian, "Understanding social media marketing: A case study on topics, categories and sentiment on a Facebook brand page," 2011, doi:10.1145/2181037.2181066.

[11] S. S. Dasgupta, S. Natarajan, K. K. Kaipa, S. K. Bhattacherjee and A. Viswanathan, "Sentiment analysis of Facebook data using Hadoop based open source technologies," *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA),* 2015, pp. 1-3.

[12] Rodríguez, A., Argueta, C., and Chen, Y. L., "Automatic detection of hate speech on facebook using sentiment and emotion analysis," *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC),* 2019, pp. 169-174.

[13] Yoo, S., Song, J., and Jeong, O., "Social media contents based sentiment analysis and prediction system," *Expert Systems with Applications,* vol. 105, pp. 102-111, 2018.

[14] Tadesse, M. M., Lin, H., Xu, B., and Yang, L., "Personality predictions based on user behavior on the facebook social media platform," *IEEE Access,* vol. 6, pp. 61959-61969, 2018.

[15] Akter, S., and Aziz, M. T., "Sentiment analysis on facebook group using Lexicon based approach," In *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT),* 2016, pp. 1-4.

[16] Imran, A., Faiyaz, M., and Akhtar, F., "An enhanced approach for quantitative prediction of personality in facebook posts," *International Journal of Education and Management Engineering (IJEME),* vol. 8, no. 2, pp. 8-19, 2018.

[17] Bagheri, H., and Islam, M. J., "Sentiment analysis of twitter data," *arXiv preprint arXiv:1711.10377,* 2017.

[18] Skowron, M., Tkalčič, M., Ferwerda, B., and Schedl, M., "Fusing social media cues: personality prediction from twitter and instagram," In *Proceedings of the 25th international conference companion on world wide web,* 2016, pp. 107-108.

[19] Markovikj, D., Gievska, S., Kosinski, M., and Stillwell, D. J., "Mining facebook data for predictive personality modelling," In *Seventh International AAAI Conference on Weblogs and Social Media,* 2013.

[20] Soliman, T. H., Elmasry, M. A., Hedar, A., and Doss, M. M., "Sentiment analysis of Arabic slang comments on facebook," *International Journal of Computers and Technology,* vol. 12, no. 5, pp. 3470-3478, 2014.

[21] Dasgupta, S. S., Natarajan, S., Kaipa, K. K., Bhattacherjee, S. K., and Viswanathan, A., "Sentiment analysis of Facebook data using Hadoop based open source technologies," In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA),* 2015, pp. 1-3.

[22] Saif, H., He, Y., Fernandez, M., and Alani, H., "Contextual semantics for sentiment analysis of Twitter," *Information Processing and Management,* vol. 52, no. 1, pp. 5-19, 2016.

[23] Chen, T., Xu, R., He, Y., and Wang, X., "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," *Expert Systems with Applications,* vol. 72, pp. 221-230, 20017.

[24] Pandey, A. C., Rajpoot, D. S., and Saraswat, M., "Twitter sentiment analysis using hybrid cuckoo search method," *Information Processing and Management,* vol. 53, no. 4, pp. 764-779, 2017.

[25] Ma, Y., Peng, H., and Cambria, E., "Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM," In *Proceedings of the AAAI Conference on Artificial Intelligence,* 2018.

[26] Zabłocki, Marcin, "Jupyter Notebook + Python Code of Twitter Sentiment Analysis," *Github.com*, [Online] Available: https://github.com/marrrcin/ml-twitter-sentiment-analysis.

BIOGRAPHIES OF AUTHORS

Daniel Demetrius Albesta is a motivated person to do study in the database field, especially in the multiple subjects related to the Database Administration and System Analysis. Take educational courses on the Bina Nusantara University, and he got guidance on his practice, which includes the Big Data comprehensive within it. Not only that, but Daniel also has a skill in the UI/ UX Design of an application he made on the Bina Nusantara Festival (Bifest) annual event. Currently, he works on the new research paper, which specifically about Big Data and sentiment analysis flow within the IMDb contents. He is presently finishing his Bachelor of Science in Computer Science, and from this point forward, the ambition he has is to be a data engineer in an information technology (IT) department.

Michael Liong Jonathan, study at Bina Nusantara University for almost two years. In these past years, he dedicated himself to learn database and how it works within the computer science world. Until now, he is also working as a full-time junior programmer at Bina Nusantara University always to keep up and learning more about database when on the other side trying to find more experiences of having a career. While pursuing his career, currently, he is completing and aiming for his Bachelor of Computer Science.

Muhammad Jawad is a Bina Nusantara University learner who is currently undergoing his bachelor's degree in computer science. The interest he has is in the field related to the Internet of Things (IoT) and other fields such as social media and digital marketing.

Oktovianus Hardiawan is a Bina Nusantara University learner-based at Jakarta, Indonesia, and at the age of eleven, he was genuinely keen on the coding things. Introduced to the system of the computer has, by his uncle, make him got interested in that. He will still recall the feeling that he had to learn how machines applied within the computer, why it is needed, and what the capability it has. That makes him interested in Computer Science and Mathematics. After graduated from high school, he took a computer science major. At Bina Nusantara University, he learned to Analysis of Algorithms, Data Structures, Programming Language, and Mathematics. He has interested in Database systems related to System Analysis, and Programming focused on mobile programming.

Derwin Suhartono is a faculty member of Bina Nusantara University, Indonesia. He got his Ph.D. degree in computer science from Universitas Indonesia in 2018. His research fields are natural language processing. Recently, he is continually researching argumentation mining and personality recognition. He actively involves in the Indonesia Association of Computational Linguistics (INACL), a national scientific association in Indonesia. He has his professional memberships in ACM, INSTICC, and IACT. He also takes the role of reviewer in several international conferences and journals.