Sentiment Analysis of Twitter User Opinions Related to Metaverse Technology Using Lexicon Based Method

Wahyu Azriel Akbari[1]*, Tukino[2], Baenil Huda[3], Muhamad Muslih[4]
[1,2,3] Universitas Buana Perjuangan Karawang, [4]Universitas Nusa Putra, Indonesia
[1]si19.wahyuakbari@mhs.ubpkarawang.ac.id, [2]tukino@ubpkarawang.ac.id, [3]baenil88@ubpkarawang.ac.id,
[4],muhamad.muslih@nusaputra.ac.id

Abstract: The technology of the metaverse has been a topic of discussion in the tech sector recently, and the implementation of the metaverse in the future is one of the most anticipated things for internet users around the world, allowing users to become the avatar they want in the virtual universe. To see how someone approaches metaverse technology, we can analyze it by looking at various opinions that emerge from the sentiment of internet users, one of which is from the social media Twitter. This study aims to analyze the sentiment of Twitter users' opinions, which is one of the most popular internet social media platforms. Then, the analysis of the sentiment of Twitter user opinions is carried out using the Lexicon Based method, which is a method that uses a dictionary of words to categorize the sentiment of a text. The results of this study show that the Lexicon Based method can classify the sentiment of Twitter user opinions related to metaverse technology with an accuracy value of 84%. This shows that the results of sentiment analysis using the Lexicon Based method have a good level of accuracy in performance in analyzing the sentiment of Twitter user opinions related to metaverse technology.

Keywords: Lexicon Based; Metaverse Technology; Sentiment Analysis; Twitter; Python.

INTRODUCTION

The digital era has brought the internet and the digital revolution to a new level. With the availability of the internet and easy access to technology, such as smartphones and other smart devices, people spend a lot of time online, including on social media. The internet has provided various benefits to users worldwide by allowing easy interaction. This interaction has led to the development of technologies such as augmented reality (AR), which lets users experience the physical world with virtual information, and virtual reality (VR), which provides users with a fully immersive experience in a virtual world. The combination of AR and VR technologies has resulted in mixed reality (MR), which allows users to experience the real and virtual worlds simultaneously(Lee et al., 2021). In this context, the metaverse is the most anticipated technology for users' internet experiences. The metaverse allows users to enter a truly immersive VR world where they can become the avatar of their choice in a virtual universe. Users can then experience the sensation of entering their virtual world or enjoying the experience in the virtual universe with their friends through the metaverse(Buhalis & Karatay, 2022).

In 1991, Tim Berners-Lee created the first website, marking the start of the internet's development(Tukino, 2019). In the 1990s, the internet experienced rapid growth and became a crucial technology for the metaverse, which allows users to connect without boundaries. The term metaverse was first introduced in Neal Stephenson's science fiction novel Snow Crash, published in 1992. The novel describes a 3D virtual world inhabited by avatars controlled by users. Recently, the metaverse has become a keyword in the technology sector(Comer, 2018). One of the factors of metaverse implementation that makes this technology development so fast is the condition of the covid-19 pandemic, which has significantly changed community interaction and also had an impact on direct learning that was abolished and replaced by online learning, and various work also changed were working in the office was increasingly abandoned changed to digitalization. Until now, some continue with increasingly diverse online work supported by renewable technology such as the metaverse(Mann et al., 2018). During this period, metaverse technology has provided a concept that can gain the interest of internet users because it allows them to share new experiences through virtual world meetings. In 2020, American rapper Travis Scott held the first metaverse concert on the online video game platform Fortnite. The following year,

*name of corresponding author

Microsoft introduced Mesh, a platform for virtual meetings with mixed reality. In 2021, Facebook changed its logo and product name to Meta, signaling its intention to build metaverse technology(Kraus et al., 2022). The metaverse then became a hot topic on social media and briefly trended on Twitter, indicating the anticipation of internet users for the potential influence of metaverse implementation in everyday life. By analyzing the responses to metaverse technology on social media, such as Twitter, we can see the sentiments of internet users towards this technology.

Sentiment analysis is an analysis technique that can be used to determine and measure the attitudes or emotions of each user toward a topic. In this context, sentiment analysis of Twitter users regarding metaverse technology may involve collecting and analyzing tweets about the metaverse to determine whether Twitter users generally show a positive, negative, or neutral attitude toward the topic(Ahmad & Gata, 2022). Previous research on Lexicon-based sentiment analysis usually focuses on how this method can be used for sentiment analysis. The Lexicon Based method is a commonly used sentiment analysis method. It can classify text as positive, negative, or neutral using a dictionary containing words or phrases associated with positive or negative feelings. Previous research has shown that the Lexicon Based algorithm method can provide relatively accurate results in classifying text, primarily when used in conjunction with other methods, such as machine learning(Sumitro et al., 2021).

METHOD
The Flow Model of Sentiment Analysis Stages on Twitter
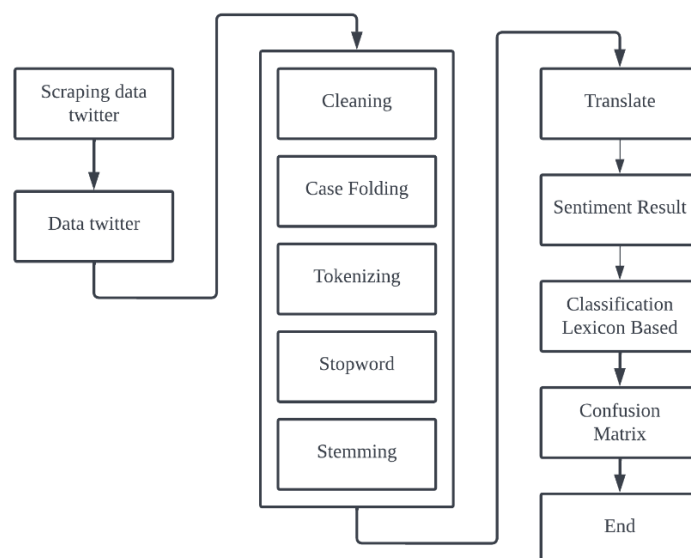Figure 1 below shows the flow model of the stages of sentiment analysis on Twitter that was conducted in this research.

Fig. 1 Flow Model of Sentiment Analysis Stages on Twitter

Twitter Data Collection
In this section, the process of automatically collecting data from the Twitter platform using data scraping on snscrape, a library from the Python programming language. Snscrape provides various features for collecting data from Twitter in bulk, such as collecting tweets from a specific Twitter account, collecting account information, or collecting interactions between accounts(Wakamiya et al., 2019). The process of collecting Twitter data using snscrape requires the keyword "Metaverse Technology" using a certain period of data collection from 01 December 2021 to 01 December 2022, after which scraping data through snscrape can produce 1500 tweet data from Twitter users consisting of data frame results containing Datetime, TweetId, Tweet, and UserId.

Preprocessing Data
After scraping data in the previous stage, the preprocessing data stage is carried out to process the data before further analysis. Preprocessing data usually includes several steps, such as cleaning the data from unwanted characters (cleaning text), breaking the text into smaller parts called tokens (tokenize), removing meaningless words (stopwords), and changing the word to its basic form (stemming or lemmatization). The purpose of

*name of corresponding author

preprocessing data is to produce clean data ready for analysis so that the analysis results are more accurate and valid(Magriza, 2022).

Translate Data

In this section, translating the preprocessed tweet data into English is carried out through the Python programming language using the googletrans library. This library implements the Google Translate API that provides features for translating text into various languages. Translating tweets into English is to provide maximum results in applying the Lexicon Base sentiment analysis method, using the Vader Sentiment library, which can only be used in English to provide polarity values of each sentiment tweet from Twitter users.

Classification of Sentiment Analysis

In this stage, classification of processed sentiment tweet data from the previous stage is performed(Lia Hananto et al., 2021). The classified tweet data is classified using the Lexicon Based algorithm, which will be used to classify the tweet data as a positive, negative, or neutral sentiment value using a library containing words or phrases associated with positive, negative, or neutral emotions. The results of the classification are then displayed using a confusion matrix as the result of the sentiment analysis and displaying the accuracy, precision, and recall of the Lexicon Based algorithm(Matulatuwa et al.,2017). To detect classification or sentiment, this study utilizes the Python library with a polarity score > 0 is a positive sentiment, then polarity score = 0 is a neutral sentiment, and polarity score < 0 is a negative sentiment. The classification of sentiment can be done with the following equation (1).

$$Sentence\ sentiment \begin{cases} positif\ if\ Spositif > Snegatif \\ netral\ if\ Spositif = Snegatif \\ negatif\ if\ Spositif < Snegatif \end{cases} \quad (1)$$

RESULT

This sentiment analysis research on metaverse technology through social media Twitter aims to understand various opinions that emerge from internet users, one of which is from Twitter social media regarding the metaverse technology that has recently been discussed. In data collection to understand Twitter user opinions, the researcher used the scraping data method on the snscrape library using the Python programming language by entering the keyword "metaverse technology" and successfully obtained 1500 data tweets from Twitter users from 01 December 2021 to 01 December 2022. The obtained data is converted into a data frame table that contains the Datetime, TweetId, Tweet, and UserId attributes. Then after being converted into a data frame table, data is further extracted only from the tweet content to make it easier to process in the next data preprocessing stage. The following is a figure of the result of scraping tweet data using the snscrape library:

	Datetime	TweetId	Tweet	UserId
0	2022-05-31 23:50:50+00:00	1531785304650096640	Jujur desain futuristic tuh bukan tipe guee bg...	oreinjis
1	2022-05-31 21:10:57+00:00	1531745070374068224	Sibuk² metaverse, dunia krisis bahan pangan da...	alfy_desty
2	2022-05-31 17:55:30+00:00	1531695883972030464	hari hari ada aja yak kelakuan seleb seleb dun...	maltness
3	2022-05-31 16:45:18+00:00	1531678215970095104	Kalau Ant-Man membuka wacana ke publik tentang...	Robin_Blac
4	2022-05-31 15:50:47+00:00	1531664498037497856	Tarian yang akan membawamu ke dunia metaverse ...	Muk_Gozali
...	...	...	...	...
1496	2022-02-24 01:56:33+00:00	1496665319980486661	Emak-emak Indosiar menuju Metaverse dan dunia ...	AarwanRamsey
1497	2022-02-24 01:47:22+00:00	1496663007501623300	Y.B. Hariantono memastikan BNI pun telah memil...	BUMN_Biak_Ayo
1498	2022-02-24 01:46:43+00:00	1496662844682960898	Gandeng WIR Group, BNI Siap Ekspansi Bisnis Di...	BUMN_Biak_Ayo
1499	2022-02-24 01:37:54+00:00	1496660625900335104	/sms/ ini apalagi yg direncanain SM yang udah ...	Karrina
1500	2022-02-24 01:14:43+00:00	1496654793674473472	/ae/ mys, SM tuh bener2 gak main2 buat metaver...	aespafess_

Fig. 2 Results of Scraping Tweet Data Using the Snscrape Library

After scraping data in the previous stage, the data preprocessing stage is carried out to process the data before further analysis. Preprocessing includes several steps such as cleaning the data from unwanted characters (cleaning text), breaking the text into smaller parts called tokens (tokenize), removing meaningless words (stopwords), and converting words to basic form (stemming or lemmatization), the researcher uses the NLTK and Sastrawi libraries in data preprocessing, after the preprocessing stage and duplicate tweet cleaning produces

*name of corresponding author

1429 results of data tweets from Twitter user opinions(Agastya, 2018). The following is a figure of the result of preprocessing the cleaned tweet data:

Fig. 3 Results of Preprocessing Tweet Data

After passing the data preprocessing stage to provide the best results in implementing the sentiment analysis method using the Lexicon Based algorithm to provide the polarity value of each sentiment tweet from Twitter users, the translation is performed into English using the googletrans library. This library is an implementation of the Google Translate API that provides the feature of translating the text into various languages. The following is a picture of the result of translating tweet data into English:

Fig. 4 Results of Translate Data Tweet into English

After preprocessing the data and translating the clean tweets into English, the classification stage is performed. In this research, sentiment classification is performed using the Lexicon Based algorithm method using a library or Lexicon as the primary language in determining the sentiment of each tweet using the Vader Sentiment library, which produces a compound score or sentiment polarity value. Sentiments with a polarity score > 0 are positive sentiment values, polarity scores < 0 are negative sentiment values, and polarity scores = 0 are neutral sentiment values. The following is a figure of the result of the sentiment using the Lexicon Based method on the Vader Sentiment library:

*name of corresponding author

198

	Tweet	Compound Score	Sentiments
0	to be honest the futuristic design is my type	0.5106	Positif
1	busy metaverse world food energy crisis ironic	-0.5423	Negatif
2	just like the behavior of cyber celebrities th...	0.5994	Positif
3	antman opens public discourse on technology na...	-0.2023	Negatif
4	dance to bring the metaverse world	0.0000	Netral
...	...	...	...
1425	metapets coins make metaverse big not an onlin...	0.5994	Positif
1426	workforce indosior matters from world	0.0258	Positif
1427	partnering with wir group bni digital business...	0.0000	Netral
1428	the sms that was planned sm is beyond reason e...	0.2023	Positif
1429	ae mys sm really doesn play metaverse the conc...	0.2449	Positif

Fig. 5 Sentiment Results Using the Lexicon Based Method

In sentiment analysis using the Lexicon Based method, a comparison is made between manual classification and the results of classification using the Lexicon Based method, resulting in a percentage of positive sentiment polarity values of 50.1%. In comparison, neutral sentiment produces a polarity value of 35.62%, and negative sentiment produces a polarity value of 14.28%. The classification results show that the opinion of Twitter users on metaverse technology gives a positive sentiment value, indicating that metaverse technology has become something that most internet users worldwide, including Twitter users(Muhamad, 2019). The following is a figure of the histogram display and percentage of each sentiment of the Lexicon Based method:
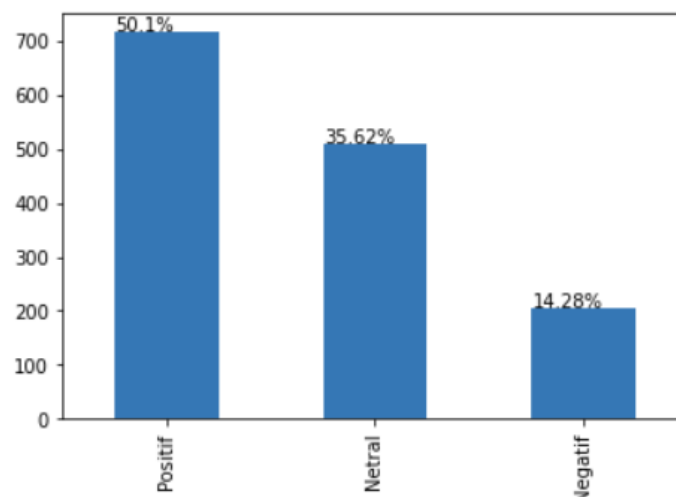
Fig. 6 Histogram Display and Percentage of Each Lexicon Based Sentiment Method

Based on the analysis of sentiment based on previous polarity values, the visualization process is then continued into the form of a word cloud image, which is based on data from sentiment tweets of Twitter users through how many words are output, each word that comes out is then visualized in the form of an image and the words that are many, in the visualized image point towards support for this metaverse technology, which indicates that Twitter users give most of the sentiment tweet values as positive. The following is a word cloud image from the results of using the Lexicon Based method:

*name of corresponding author

Fig. 7 Word Cloud Image from the Results of Using the Lexicon Based Method

In this section, testing is performed using the confusion matrix method with the Lexicon Based method. The accuracy testing is performed by comparing the actual classification with the classification results generated by the model. Testing is performed by displaying the classification report, which calculates the accuracy, precision, recall, f1-score, macro avg, and weighted avg. Here is the figure of the result of the confusion matrix using the Lexicon Based method:
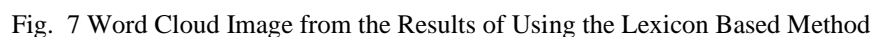
	precision	recall	f1-score	support
Negatif	0.86	0.81	0.84	216
Netral	0.74	0.88	0.81	429
Positif	0.91	0.83	0.87	784
accuracy			0.84	1429
macro avg	0.84	0.84	0.84	1429
weighted avg	0.85	0.84	0.84	1429

Fig. 8 Result of the Confusion Matrix Using the Lexicon Based Method

DISCUSSIONS

Based on the classification results on the confusion matrix using the lexicon-based method, the precision, recall, and f1-score values for positive sentiment can be obtained in sequence as 91%, 83%, and 87% with support 784. In comparison, the neutral sentiment is 74%, 88%, and 81% with support 429, and for negative sentiment is 86%, 81%, and 84% with support 216. From the above data, the accuracy for the f1-score is obtained as 84% and support 1429, then for the macro average for precision 84%, recall 84%, f1-score 84% and support 1429, while the weighted average for precision 85%, recall 84%, f1-score 84% and support 1429. Therefore, this shows that sentiment analysis using lexicon-based has a high level of performance accuracy because it has an accuracy value above 80%(Normawati & Prayogi, 2021).

CONCLUSION

The arrival of the metaverse technology a while ago caused a sensation and became something anticipated among internet users. In this study, sentiment analysis was carried out using the Lexicon Based method to find out various opinions from internet users, one of which is from the social media Twitter, which produces positive, negative, or neutral sentiments from a total of 1500 tweets obtained the process of scraping data using the library from snscrape in the python programming language, data was obtained as many as 1429 tweets after undergoing the data preprocessing stage, then continued to the process of translating into English for further sentiment analysis using the Lexicon Based method by producing a sentiment analysis value at an accuracy of 84% showing a positive response from Twitter user opinions. Therefore, this shows that the sentiment analysis using the Lexicon Based method has a good performance accuracy value.

REFERENCES

Agastya, I. M. A. (2018). Pengaruh Stemmer Bahasa Indonesia Terhadap Peforma Analisis Sentimen Terjemahan Ulasan Film. *Jurnal Tekno Kompak*, *12*(1), 18. doi:10.33365/jtk.v12i1.70

Ahmad, A., & Gata, W. (2022). Sentimen Analisis Masyarakat Indonesia di Twitter Terkait Metaverse dengan Algoritma Support Vector Machine. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, *6*(4),

*name of corresponding author

548–555. doi:10.35870/jtik.v6i4.569

Buhalis, D., & Karatay, N. (2022). Mixed Reality (MR) for Generation Z in Cultural Heritage Tourism Towards Metaverse. *Information and Communication Technologies in Tourism 2022*, 16–27. doi:10.1007/978-3-030-94751-4_2

Comer, D. E. (2018). *The Internet Book Everything You Need to Know about Computer Networking and How the Internet Works*. *The Internet Book*. doi:10.1201/9780429447358

Kraus, S., Kanbach, D. K., Krysta, P. M., Steinhoff, M. M., & Tomini, N. (2022). Facebook and the creation of the metaverse: radical business model innovation or incremental transformation? *International Journal of Entrepreneurial Behaviour and Research*, *28*(9), 52–77. doi:10.1108/IJEBR-12-2021-0984

Lee, L.-H., Braud, T., Zhou, P., Wang, L., Xu, D., Lin, Z., … Hui, P. (2021). All One Needs to Know about Metaverse: A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda, *14*(8), 1–66. Retrieved from http://arxiv.org/abs/2110.05352

Lia Hananto, A., Priyatna, B., Fauzi, A., Yuniar Rahman, A., Pangestika, Y., & Tukino. (2021). Analysis of the Best Employee Selection Decision Support System Using Analytical Hierarchy Process (AHP). *Journal of Physics: Conference Series*, *1908*(1). doi:10.1088/1742-6596/1908/1/012023

Magriza, I. A. (2022). Analisis Penerapan Algoritma Pada Preprocessing Data, *2*(5), 1–9.

Mann, S., Furness, T., Yuan, Y., Iorio, J., & Wang, Z. (2018). Virtual reality in Metaverse for future mental health-helping profession: an alternative solution to the mental health challenges of the COVID-19 pandemic, 1–2. Retrieved from http://arxiv.org/abs/1804.08386

Matulatuwa, F. M., Sediyono, E., & Iriani, A. (2017). Text mining dengan metode lexicon based untuk sentiment analysis pelayanan PT. Pos Indonesia melalui media sosial Twitter. *Jurnal Masyarakat Informatika Indonesia*, *2*(3), 52–65.

Muhamad, A. N. (2019). Analisis Sentimen Positif Dan Negatif Komentar Video Youtube Menggunakan Metode Naïve Bayes - Support Vector Machine (Nbsvm) Classifier. *Skripsi*, 17.

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, *5*(2), 697–711.

Sumitro, P. A., Rasiban, Mulyana, D. I., & Saputro, W. (2021). Analisis Sentimen Terhadap Vaksin Covid-19 di Indonesia pada Twitter Menggunakan Metode Lexicon Based. *J-ICOM - Jurnal Informatika Dan Teknologi Komputer*, *2*(2), 50–56. doi:10.33059/j-icom.v2i2.4009

Tukino, R. &. (2019). Penggunaan website sebagai media promosi. *Jurnal Pengabdian Masyarakat Multidisiplin*, *2*(2), 113–118. Retrieved from http://jurnal.univrab.ac.id/index.php/jpm/article/view/702/493

Wakamiya, S., Morita, M., Kano, Y., Ohkuma, T., & Aramaki, E. (2019). Tweet classification toward twitter-based disease surveillance: New data, methods, and evaluations. *Journal of Medical Internet Research*, *21*(2). doi:10.2196/12783

*name of corresponding author