

SENTIMENT ANALYSIS CHATGPT USING THE MULTINOMINAL NAÏVE BAYES CLASSIFIER (NBC) ALGORITHM

Dwi Sri Rahayu
Information Systems
Faculty of Science and Technology
Sultan Syarif Kasim Riau State Islamic University
Pekanbaru, Indonesia
12050320393@students.uin-suska.ac.id

Rice Novita
Information Systems
Faculty of Science and Technology
Sultan Syarif Kasim Riau State Islamic University
Pekanbaru, Indonesia
rice.novita@uin-suska.ac.id

Tengku Khairil Ahsyar
Information Systems
Faculty of Science and Technology
Sultan Syarif Kasim Riau State Islamic University
Pekanbaru, Indonesia
Tengkukhairil@uin-suska.ac.id

Zarnelly
Information Systems
Faculty of Science and Technology
Sultan Syarif Kasim Riau State Islamic University
Pekanbaru, Indonesia
zarnelly@uin-suska.ac.id

Abstract— Chatbots have become one of the most popular solutions to improve customer service. One well-known chatbot is ChatGPT, a language model developed by OpenAI. As time goes by and more and more people use ChatGPT, sentiment analysis is needed regarding user opinions regarding the ChatGPT service. Therefore, it is necessary to carry out sentiment analysis of the ChatGPT service on Twitter to find out how users respond to this chatbot service. One algorithm of the sentiment analysis method that can be used is the Multinomial Naïve Bayes Classifier (NBC) algorithm. The advantage of this algorithm is that it is relatively easy to implement and runs fast in processing. In this research, the results showed a positive sentiment of 57%, a negative sentiment of 29%, and a neutral sentiment of 14%. Obtained topics for each sentiment and sentiment prediction results from 40% of test data, with results of 96% positive, 3.5% negative, and 0.5% neutral with a test accuracy of 63%.

Keywords— ChatGPT, Google colab, Multinomial Naïve Bayes Classifier, Sentiment Analysis

I. INTRODUCTION

In an increasingly advanced digital era, chatbots have become one of the most popular solutions for improving customer service. One well-known chatbot is ChatGPT, a language model developed by OpenAI [1]. ChatGPT model GPT-3.5 was first introduced by OpenAI on November 30, 2022 [2], and was immediately available for use by users worldwide after its release, including Indonesia [3]. However, the detailed timing of when this model will begin to be widely used in Indonesia may vary depending on acceptance by users and companies in Indonesia. Several other chatbots, like ChatGPT, are BERT (Bidirectional Encoder Representations from Transformers), developed by Google; XLNet, which is also the result of research from Google Brain; and TransformerXL, developed by Salesforce Research [4][5]. Although there are differences in architecture and approaches, they all have the same goal of understanding and generating text contextually. However, of the many chatbots, ChatGPT is currently the most popular and widely used [6].

ChatGPT is used by a wide range of audiences, including researchers, software developers, businesses, writers, and general users who have access to the internet, including students, teachers, professionals, social media users, translators, users of text-based services, and many more [7]. GPT is used for research, application development, customer service, content creation, and communicating with machines using natural language [8]. Research on the use of ChatGPT chatbots in serving users on the Twitter

platform is very important to find out the extent of user response to this service. With sentiment analysis, companies can find out whether the service provided by the ChatGPT chatbot is satisfactory or not [9]. Therefore, it is necessary to carry out sentiment analysis of the ChatGPT service on Twitter to find out how users respond to this chatbot service [10].

To evaluate the ChatGPT chatbot service, one method that can be used is sentiment analysis. In the context of chatbot services, sentiment analysis can provide valuable insights into the user experience and help businesses better understand and meet the needs of their users [11]. In this research, sentiment analysis will be carried out on tweets containing user interactions with ChatGPT. In this case, one of the algorithms that can be used is the Multinomial Naïve Bayes Classifier (NBC) algorithm. The advantage of this algorithm is that it is relatively easy to implement and runs quickly [12] [13]. This makes it suitable for sentiment analysis in situations where execution time is important, such as social media monitoring [14]. This algorithm generally produces good results when used for text data, such as product reviews, tweets, or blog posts.

In the research of Dedy Atmajaya and friends (2023), this research discusses ChatGPT sentiment analysis from content created by users on platforms such as Twitter using the Support Vector Machine (SVM) and Naive Bayes algorithms. The results of this research show clear differences in model performance. SVM, combined with Vader, achieves an accuracy rate of 59%. Using Vader, Naïve Bayes achieved an accuracy rate of 47%. Meanwhile, using RoBERTa with Naive Bayes shows a decrease in accuracy to 43% [15]. Apart from that, in another similar study conducted by Muhammad Nanda Fahriza and Noviana Riza (2023), the following results were obtained: The best K value. In this study, after carrying out an evaluation, it was found that the best K value was 9, namely 96.6% [6].

The aim of this research is to implement the Multinomial NBC Algorithm in ChatGPT Chatbot sentiment analysis. This research will use a tweet dataset provided by Twitter and NLP techniques to carry out sentiment analysis on tweets that contain interactions with ChatGPT. This research will help to find out how users respond to the ChatGPT service and can provide input for the development of chatbot services in the future.

II. METHOD

This study use the sentiment analysis method. Sentiment analysis is the practice of categorizing extremes of text inside a document or sentence into positive, negative, or neutral sentiment. Data mining is used in sentiment analysis to evaluate, process, and extract text data from a given service, product, individual, phenomena, or topic [16] [17]. The algorithm used in this research is the Multinomial Naïve Bayes Classifier (NBC) algorithm. This algorithm works on the concept of term recurrence which implies the number of times a word shows up in a document. Multinomial naive Bayes could be a strategy that works by calculating the recurrence of each term within the data [18] [19]. In multinomial Naive Bayes, the sequence of events in which words appear in the document is not taken into account, so the document is considered like a "*bag of words*," so that each word is processed using a multinomial distribution [20]. The methodology of this research is as follows:

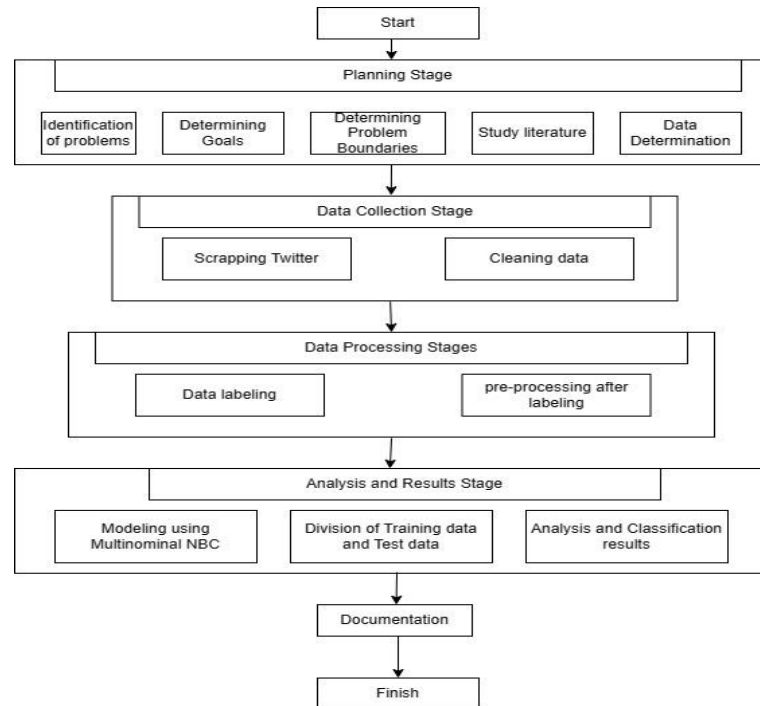


Figure 1. Research Metodology

The data used is data scraped from Twitter about GPT chat, which was collected from January 1, 2023, to August 30, 2023. Then, the data that was obtained was processed and cleaned using Google Colab. For data labeling, use the lexicon dictionary from GitHub, validated by Indonesian language experts. After labeling, the next process continues until the modeling process uses the multinomial Naïve Bayes algorithm, so that the results and accuracy of testing in this research are obtained.

Multinomial Naïve Bayes can be expressed in the following way:

Formula 1:

$$P(p|n) \propto P(p) \prod_{1 \leq k \leq n} P(t_k|p)$$

Where $P(t_k|p)$ is the probability of the occurrence of a text document (t_k), where n is the number of documents and p is the polarity. Then, the polarity or papers with similarities are formulated as follows:

Formula 2.

$$P(t_k|p) = \frac{\text{count}(tk|p)+1}{\text{count}(tp)+|V|}$$

Where $(t_k | p)$ represents the number of t_k in a text document with polarity p , and (tp) represents the number of tokens in news stories with polarity p [21].

A. Data Collection Process

This research takes data from Twitter, namely comment data from users of the ChatGPT chatbot service, using a scrapping technique with the keyword chatgpt. Scrapping was carried out using Google Colab tools, and 2090 pieces of data were obtained. Scrapping results data can be seen in the following table:

Table 1.
Preliminary data

NO.	Created_at	Username	Text
-----	------------	----------	------

1.	Sun Aug 13 05:41:07 +0000 2023	ken_aryadh arma	@piotrj @cakasana Sempet nanya ke ChatGPT gmn pandangan etika soal pembuat aplikasi pengukur polusi udara yg ngerangkap jd makelar air purifier? Jawabannya cukup menarik. Point utamanya: harus ada bukti kuat klo memang air purifier terbukti efektif mengur
2.	Fri Aug 11 11:21:38 +0000 2023	ihsanbianto ro	Sepakat. Belajar game design, kurikulum nya dibuatin sama chatGPT, dicariin sumber2 buku atau teori yg mesti dipelajarin. Efektif bgt. Pngen minta mentorin temen tapi saya bukan prioritas ðŸ˜...
....			
2090.	Tue Sep 26 16:03:46 +0000 2023	vencheio	@ilyasastra Tipe pacar yang lucu, kalo bucin suka ngelawak, tapi bisa juga brubah ala ala chatgpt can yu plis give me love bucin in english. Bentar gw bingung apakah lu cemburuan atau lebih ke tipe yang santuy? Tapi lu orangnya bulol. Apkh WOA?

B. Data processing process

- Cleaning. The data is processed, and the text will be purged first of the same symbols, hashtags, usernames, URLs, and comments.
- Data Labeling, Labeling the sentiment of comment text data for each criterion is divided into three, namely positive, negative, and neutral. with label information 0 = neutral, 1 = positive, and 2 = negative.
- Then do tokenize, remove stop words, stemming, etc. until the data is completely clean and ready for modeling. Data is processed by breaking the text into smaller pieces, such as words or symbols, to facilitate text analysis and natural language processing (tokenize), then removing common words that usually do not provide important information in text analysis (remove stopword). Next, word endings are removed so that words that are different but have the same root can be treated as the same form (Stemming). The following is the pre-processing data, as seen in the table below:

Table 2.

Data resulting from pre-processing after labeling

NO	Stemming	Label
1.	['sempet', 'nanya', 'chatgpt', 'gm', 'pandang', 'etika', 'buat', 'aplikasi', 'ukur', 'polusi', 'udara', 'yg', 'ngerangkap', 'jd', 'makelar', 'air', 'purifier', 'jawab', 'tarik', 'point', 'utama', 'bukti', 'kuat', 'klo', 'air', 'purifier', 'bukti', 'efektif']	1

2.	['sepakat', 'ajar', 'game', 'design', 'kurikulum', 'nya', 'dibuat', 'chatgpt', 'dicari', 'sumber', 'buku', 'teori', 'yg', 'mesti', 'dipelajari', 'efektif', 'bgt', 'ken', 'mentorin', 'temen', 'prioritas']	1
....		
2071	['suka', 'upchar', 'chatgpt', 'diving', 'mah', 'aja', 'udah', 'alhamdulillah']	1

- d) TF –IDF Weighting At this stage, TF-IDF gives weight to each word. Each word in a document is different using Term Frequency and Inverse Document Frequency.
- e) Division of Training Data and Test Data The next stage is to divide the training data and test data using the 60:40 *Hold Out* technique (Comparison of 60% for training data and 40% for test data).

III. RESULT AND DISCUSSION

From the research that the author has done, the following sentiment results are obtained:

- Neutral sentiments: 281 data (14%)
- Positive sentiments: 1179 data (57%)
- Negative sentiments: 611 data (29%)

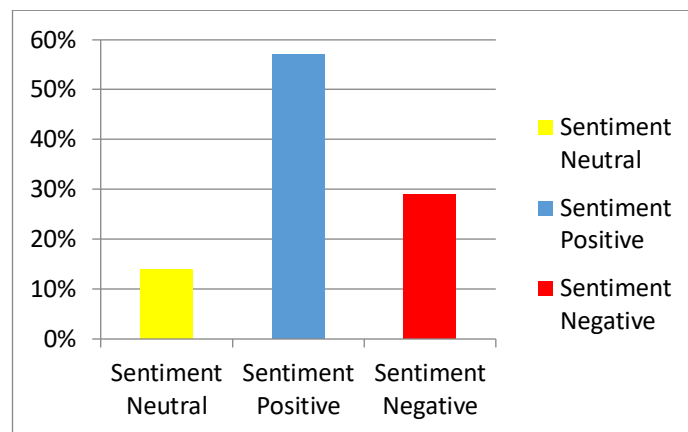


Figure 2. Sentiment Yield Chart

On each label or sentiment there is some similarity in the topic of the topic discussed in the comments. The topic for each sentiment can be seen as shown in the table below:

Positive Topic	Negative Topic	Neutral Topic
a) Membantu pekerjaan	a) Ragu-ragu	a) Jawaban
b) Efisien	(Jawaban yang diberikan berbeda)	tergantung inputan pengguna
c) Responsif, komunikatif dan informatif	b) Tidak bisa input gambar, tabel dan suara.	
	c) Tidak ada sumber referensi	
	d) Ketergantungan	

Furthermore, sentiment prediction results were obtained from test data using the multinominal naive Bayes classifier algorithm. The sentiment prediction results can be seen in the following table:

Table 4.
Sentiment prediction results

NO	Sentiment	NO	Sentiment	NO	Sentiment
1.	1	6.	2	11.	1
2.	1	7.	1	12.	1
3.	1	8.	1	13.	1
4.	1	9.	1		
5.	1	10.	1	828.	1

From the table above, the sentiment prediction results are as follows:

- Neutral: 4 (0,5%)
- Positive: 798 (96%)
- Negative: 26 (3,5%)

Then modeling was carried out using the multinominal naïve Bayes classifier algorithm, and a fairly good test accuracy was obtained, namely 63%. The recapitulation of test accuracy results using the multinominal Naïve Bayes classifier algorithm is as follows:

Table 5.
Testing accuracy

Classifiastion report					
	Precision	Recall	f1-score	support	
0	1.00	0.04	0.07	106	
1	0.61	1.00	0.76	489	
2	1.00	0.11	0.20	234	
accuracy			0.63	829	
macro avg	0.87	0.38	0.34	829	
weighted avg	0.77	0.63	0.51	829	

Multinomial Naïve Bayes classifier In this research, the K-fold Cross Validation test on Multinomial Naïve Bayes uses a dataset that has been preprocessed and labeled from comment data. The data will be divided into training data and test data. Repeat k times to randomly divide a set into k independent subsets, leaving one set for test data for each repetition and the rest for training data. The following accuracy results for each topic with K 10 Fold can be seen in the following table:

Table 6.
percentage of k-fold cross validations

	percentage
K Fold 1	57 %
K Fold 2	62 %
K Fold 3	62 %
K Fold 4	60 %
K Fold 5	63 %

K Fold 6	60 %
K Fold 7	61 %
K Fold 8	57 %
K Fold 9	62 %
K Fold 10	61 %
Average	60 %

Then, the visualization results for the words that appear most frequently can be seen in the following Wordcloud:

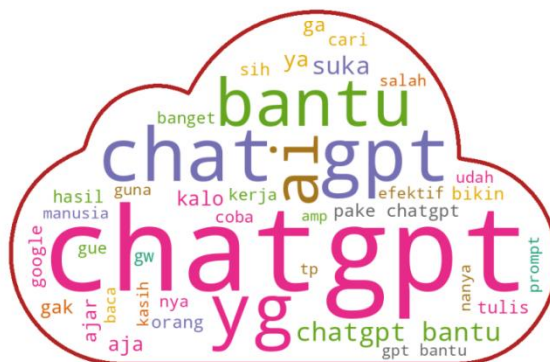


Figure 3. Word Clouds

Overall, based on the visualization of the top 40 words, it can be seen that the words that often appear in user comments on Twitter are "chatgpt" with 2014 words, "help" with 1003, like with 338 words, and "effective" with 134 words.

IV. CONCLUSION

The results of this research show that the sentiment of ChatGPT users on Twitter is predominantly positive, namely 57%. However, it turns out that there are several topics of negative and neutral sentiment that we need to pay attention to. For data processing and modeling, the algorithm used is also quite good, with a test accuracy of 63%. Thus, this research can be used as recommendations or literature for ChatGPT users and for further research.

REFERENCES

- [1] B. Huang, Y. Ou, and K. M. Carley, “Aspect level sentiment classification with attention-over-attention neural networks,” in *Social, Cultural, and Behavioral Modeling: 11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, July 10-13, 2018, Proceedings 11*, 2018, pp. 197–206.
- [2] M. Abdullah, A. Madain, and Y. Jararweh, “ChatGPT: Fundamentals, applications and social impacts,” in *2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2022, pp. 1–8.
- [3] T. Hadian, M. Pkim, and E. Rahmi, *Berteman dengan ChatGPT: Sebuah Transformasi dalam Pendidikan*. Edu Publisher, 2023.
- [4] U. Chakraborty, S. Roy, and S. Kumar, *Rise of Generative AI and ChatGPT: Understand how Generative AI and ChatGPT are transforming and reshaping the business world (English Edition)*. BPB Publications, 2023.
- [5] Y. Cao *et al.*, “A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt,” *arXiv Prepr. arXiv2303.04226*, 2023.
- [6] M. N. Fahriza and N. Riza, “ANALISIS SENTIMEN PADA ULASAN

- APLIKASI CHAT GENERATIVE PRE-TRAINED TRANSFORMER GPT MENGGUNAKAN METODE KLASIFIKASI K-NEAREST NEIGHBOR (KNN),” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1351–1358, 2023.
- [7] R. S. Y. Zebua et al., *FENOMENA ARTIFICIAL INTELLIGENCE (AI)*. PT. Sonpedia Publishing Indonesia, 2023.
 - [8] D. Kalla and N. Smith, “Study and analysis of chat GPT and its impact on different fields of study,” *Int. J. Innov. Sci. Res. Technol.*, vol. 8, no. 3, 2023.
 - [9] E. H. Muktafin, K. Kusriani, and E. T. Luthfi, “Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing,” *J. Eksplora Inform.*, vol. 10, no. 1, pp. 32–42, 2020.
 - [10] V. Taecharunroj, ““What can ChatGPT do?” Analyzing early reactions to the innovative AI chatbot on Twitter,” *Big Data Cogn. Comput.*, vol. 7, no. 1, p. 35, 2023.
 - [11] D. Tang, B. Qin, and T. Liu, “Document modeling with gated recurrent neural network for sentiment classification,” in *Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015, pp. 1422–1432.
 - [12] P. S. Zalukhu, T. Handhayani, and M. Sitorus, “Analisis Sentimen terhadap Kenaikan BBM di Indonesia pada Media Sosial Twitter Menggunakan Metode Naïve Bayes,” *Simtek J. Sist. Inf. dan Tek. Komput.*, vol. 8, no. 1, pp. 65–69, 2023.
 - [13] P. U. Maharani, N. Amalita, A. A. Putra, and F. Fitri, “Sentiment Analysis og Goride Services on Twitter Social Media Using Naive Bayes Algorithm,” *UNP J. Stat. Data Sci.*, vol. 1, no. 3, pp. 134–139, 2023.
 - [14] P. I. Anakku, E. Erizal, and F. N. Hasan, “Visualisasi Dashboard Business Intelligence Untuk Analisa Ketersediaan Tenaga Kesehatan Pada Saat Covid-19 Di Jakarta Menggunakan Tableau,” *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 4, no. 4, pp. 1006–1019, 2023.
 - [15] D. Atmajaya, A. Febrianti, and H. Darwis, “Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter,” *Indones. J. Comput. Sci.*, vol. 12, no. 4, 2023.
 - [16] A. M. Zuhdi, E. Utami, and S. Raharjo, “Analisis sentiment twitter terhadap capres Indonesia 2019 dengan metode K-NN,” *J. Inf. J. Penelit. dan Pengabd. Masy.*, vol. 5, no. 2, pp. 1–7, 2019.
 - [17] E. R. Subhiyakto, Y. P. Astuti, N. Alexander, and E. Kartikadarma, “Analisis Sentimen Menggunakan Metode Naïve Bayes Untuk Mengetahui Respon Masyarakat Terhadap Vaksinasi,” *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 02, pp. 179–188, 2022.
 - [18] R. L. Atimi and E. E. Pratama, “Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia,” *J. Sains Dan Inform.*, vol. 8, no. 1, pp. 88–96, 2022.
 - [19] E. Mulyani, F. P. B. Muhamad, and K. A. Cahyanto, “Pengaruh N-Gram terhadap Klasifikasi Buku menggunakan Ekstraksi dan Seleksi Fitur pada Multinomial Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 5, no. 1, pp. 264–272, 2021.
 - [20] S. Fanissa, M. A. Fauzi, and S. Adinugroho, “Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking,” *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, vol. 2, no. 8, pp. 2766–2770, 2018.
 - [21] N. Hidayah and S. Sahibu, “Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 820–826, 2021.