Data & Metadata. 2023; 2:45
doi: 10.56294/dm202345

ORIGINAL

Implementation of Naive Bayes classification algorithm for Twitter user sentiment analysis on ChatGPT using Python programming language

Implementación del algoritmo de clasificación Naive Bayes para el análisis del sentimiento de los usuarios de Twitter en ChatGPT utilizando el lenguaje de programación Python

Adhitia Erfina[1], M Rifki Nurul Ramdani Alamsyah[1]

[1]Universitas Nusa putra. Sukabumi, Indonesia.

Cite as: Erfina A, Rifki Nurul M. Implementation of Naive Bayes classification algorithm for Twitter user sentiment analysis on ChatGPT using Python programming language. Data & Metadata. 2023;2:45. https://doi.org/10.56294/dm202345

Submitted: 17-04-2023 **Revised:** 30-04-2023 **Accepted:** 06-06-2023 **Published:** 07-06-2023

Editor: Prof. Dr. Javier González Argote

ABSTRACT

ChatGPT (Generative Pre-Trained Transformer) is a chatbot that is being widely used by the public. This technology is based on Artificial Intelligence and is capable of having conversational interactions with its users just like humans, but in the form of automated text. Because of this capability, online forums such as Brainly and the like can be overtaken by these smart chatbots. Therefore, this study was conducted to determine the positive and negative sentiments towards ChatGPT using Naive Bayes Classification algorithm on 5000 Twitter users. Data was collected by scraping technique and Python programming language was used in data analysis. The results showed that the majority of Twitter users had a positive sentiment of 57,6 % towards ChatGPT, while the negative sentiment reached 42,4 %. The resulting classification model had an accuracy of 80 %, indicating a good classification model in determining sentiment probabilities. These findings provide a basis for the development of better AI chatbot technology that can meet user needs. The results of this study provide insights into user sentiment towards ChatGPT and can be used as a reference for future AI chatbot development.

Keywords: Chatgpt; Sentiment Analysis; Naive Bayes Classifier; Programming Language; Algorithm.

RESUMEN

ChatGPT (Generative Pre-Trained Transformer) es un chatbot muy utilizado por el público. Esta tecnología se basa en la Inteligencia Artificial y es capaz de mantener interacciones conversacionales con sus usuarios igual que los humanos, pero en forma de texto automatizado. Debido a esta capacidad, foros en línea como Brainly y similares pueden verse superados por estos chatbots inteligentes. Por lo tanto, este estudio se llevó a cabo para determinar los sentimientos positivos y negativos hacia ChatGPT utilizando el algoritmo de clasificación Naive Bayes en 5000 usuarios de Twitter. Los datos se recogieron mediante la técnica de scraping y se utilizó el lenguaje de programación Python para su análisis. Los resultados mostraron que la mayoría de los usuarios de Twitter tenían un sentimiento positivo del 57,6 % hacia ChatGPT, mientras que el sentimiento negativo alcanzó el 42,4 %. El modelo de clasificación resultante tuvo una precisión del 80 %, lo que indica un buen modelo de clasificación para determinar las probabilidades de sentimiento. Estos resultados proporcionan una base para el desarrollo de una mejor tecnología de chatbot de IA que pueda satisfacer las necesidades de los usuarios. Los resultados de este estudio proporcionan información sobre el sentimiento de los usuarios hacia ChatGPT y pueden utilizarse como referencia para el desarrollo de futuros chatbot de IA.

Keywords: Chatgpt; Análisis De Sentimiento; Clasificador Naive Bayes; Lenguaje De Programación; Algoritmo.

INTRODUCTION

Currently, global internet users have reached 5,15 billion people as of January 2023. This number accounts for 64,4 % of the global population, which totalled 8,01 billion people. The number of global internet users in January 2023 increased by 1,9 % compared to the same period last year (year-on-year/yoy), which was still 5,01 billion people.[1]

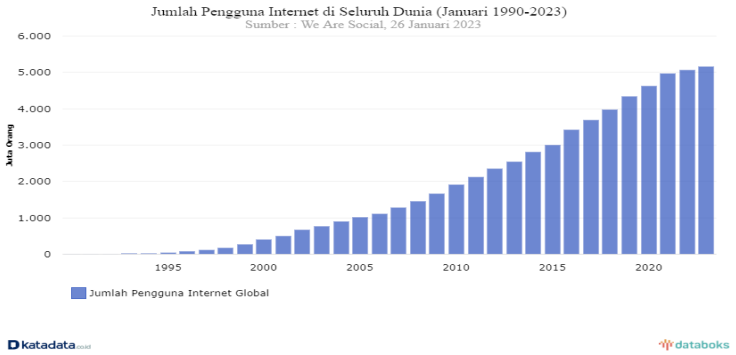

Figure 1. Graph of internet users around the world

One of the technologies that is currently being used by the public is AI chatbot technology called ChatGPT. ChatGPT stands for Generative Pre-Trained Transformer, a chatbot based on Artificial Intelligence technology that can carry out conversational interactions with its users in a sophisticated manner. This chatbot is able to answer user questions with the same steps as humans but in the form of automated text.[2] But on the other hand, there are allegations of the impact of the presence of ChatGPT, namely that some learning forums will be less attractive because they will start to switch through the help of ChatGPT.

Not only learning forums, human jobs may be replaced by Artificial Intelligence technology, one of which is ChatGPT, and can damage the homework system because students may cheat on answers directly from ChatGPT without trying to do the assignment independently. Because of these allegations, it is necessary to analyse the sentiment of Twitter users towards ChatGPT. Thus, it can be known the positive and negative views of ChatGPT which is very useful for the community in using ChatGPT, as well as helping ChatGPT developers in developing the application.

The Naïve Bayes classification algorithm was used with the research subject in the form of Twitter users who made uploads (tweets) containing chatGPT words. The Naïve Bayes algorithm was chosen because its reliability has been proven by previous researchers, regarding "Sentiment analysis of economic recovery in Indonesia after the covid-19 pandemic on twitter using the naive bayes classifier algorithm". The results showed that the Naive Bayes Classifier algorithm was able to classify tweet data with an accuracy value of 78 %, class precision positive predictions 96 % while negative predictions 31% and recall obtained from true positive of 78 % while true negative of 75 %. The results obtained for sentiment classification using Naive Bayes Classifier on public tweets have achieved maximum expectations.[3]

METHODS

Method the implementation of the Naive Bayes classification algorithm of twitter users against chatGPT involves several stages shown in the diagram below.

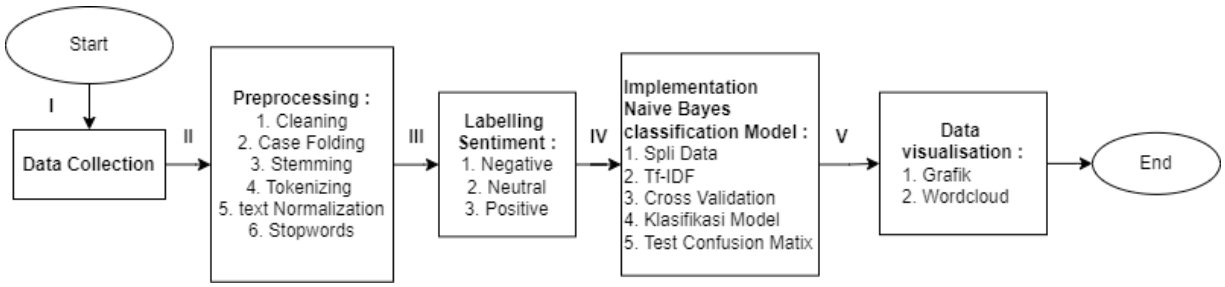

Figure 2. Research method

1. **Data collection stage:** at this stage, third-party software called Snscrape is used to collect data from Twitter through web scraping techniques. Although Snscrape is not part of the official Twitter API, it is reliable for retrieving items such as user profiles, hashtags, or relevant search results. “Snscrape is a scraper for social

network services (SNS) that has the ability to retrieve posts that are relevant and related to the research objectives".[4] The stages of data collection are in the diagram below.

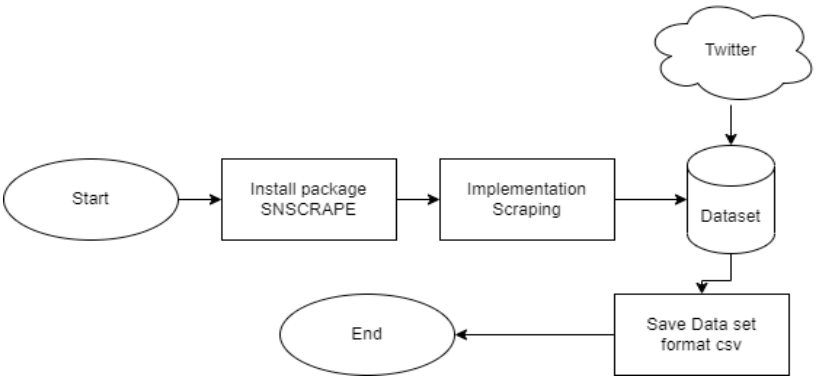

Figure 3. Data collection stages

2. Preprocessing stage: is a technique used to transform raw data in a useful and efficient format. This stage is necessary because raw data is often incomplete and has an inconsistent format. Data quality itself has a direct correlation with the success of any project involving data analysis.[5] In preprocessing, there are also several stages including case folding, cleaning, tokenizing, text normalisation, stopword removal, stemming, removing duplicate data, and removing NaN (not at number) data. The preprocessing stages are in the diagram image below.

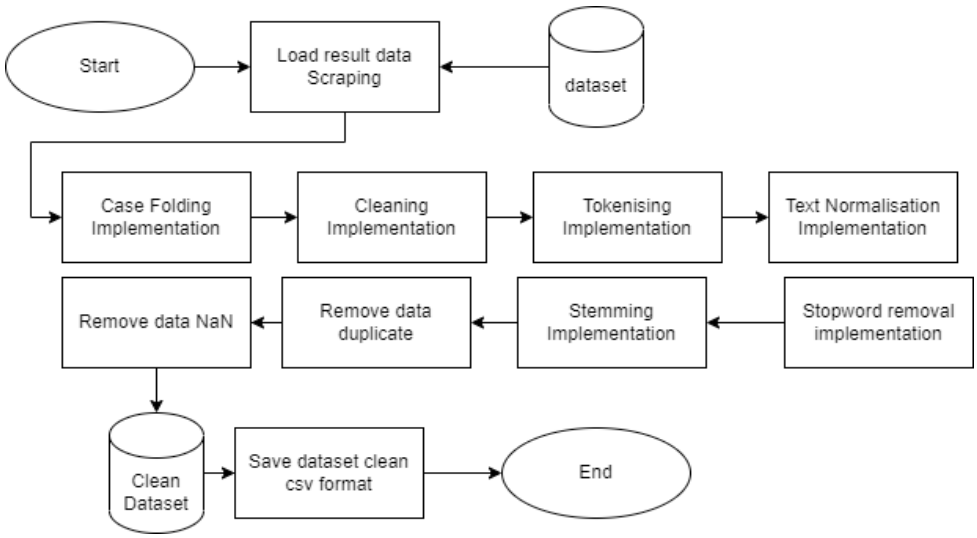

Figure 4. Preprocessing stages

3. Sentiment labelling stage: after preprocessing is complete. Sentiment labelling uses the Opinion Lexicon approach. Opinion lexicon is a list of words or phrases categorised by sentiment polarity, such as positive, negative, or neutral. Opinion lexicon is usually used in sentiment analysis to help machines or computer programs understand and extract opinion polarity or sentiment from text. The opinion lexicon used by hu and liu consists of approximately 6800 words.[6]

4. Implementation of Naive Bayes Classification Model stage: after the tweet data has been labelled, the next step is to proceed to the implementation stage of the Naive Bayes classification model. The purpose of this stage is to evaluate the extent to which the model that has been built can be accurate in determining positive and negative sentiments in the sentiment analysis of Twitter users towards ChatGPT. Naïve Bayes has the ability to fast in modelling, has predictive ability and also provides a new method of exploring and understanding data.[7] Due to the large amount of tweet data, it is necessary to add more optimisation in the classification model, namely cross validation, Cross-validation (CV) is a statistical method that can be used to evaluate the performance of a model or algorithm where the data is separated into two subsets, namely learning process data and validation / evaluation data. The model or algorithm is trained by the learning subset and validated by the validation subset. Furthermore, the selection of CV type can be based on the size of the

https://doi.org/10.56294/dm202345

dataset. Usually, K-fold CV is used because it can reduce computation time while maintaining the accuracy of the estimation.[8]

The following below is a diagram of the stages in implementing the Naive Bayes classification model.

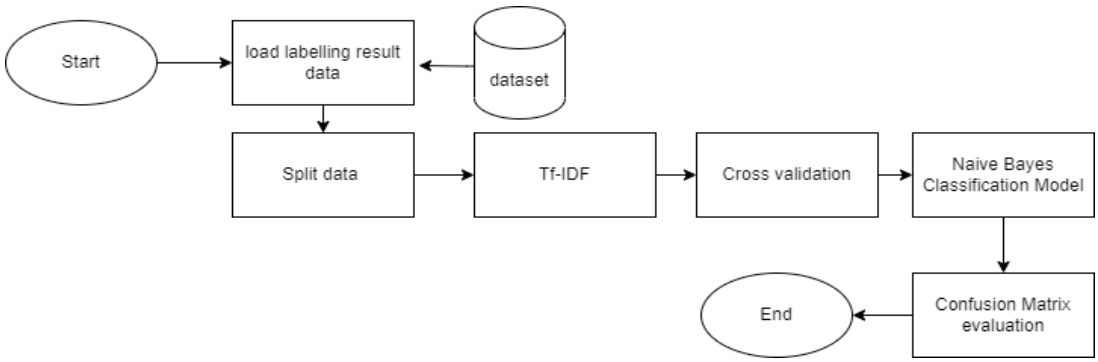

Figure 5. Implemention the Naive Bayes classification model stages

5. **Data visualisation stage**: a bar chart graph will be used to see the top 10 most frequent words in the tweet data, so that the most frequent words in the data can be identified. In addition, wordcloud will also be used for visualisation. A word cloud is one of the visualisation techniques that consists of a collection of words that appear the most when the dataset is analysed. The size of the letters is determined by the intensity with which the word is used. The more often it is used, the larger the letter size of the word.[9] the information obtained from this visualisation will be useful for visualising the data from the research, making it easier for researchers to understand the data that has been collected.

RESULTS AND DISCUSSION

1. **Data collection results:** the data taken amounted to 5000 tweet data with the initial time set of 1 January 2023 to 09 March 2023 the data taken in the form of tweet dates, usernames, and the tweets themselves with the keyword #ChatGPT, the contents of the tweets are in English After the data is successfully scraped and then save it into csv format. 5000 tweets about chatGPT were collected and ready to enter the preprocessing stage. Following table 1 is the result of data collection.

Table 1. Data collection result

Date	Username	Tweet
09/03/2023	OGVernon	"new technology won't replace us! people need a human touch in these complex situations!" ~telephone operators, 1930 ~elevator operators, 1960 ~radiologists, 2023 #ai #radiologists #radiology #chatGPT
09/03/2023	Ben_Ccarter	The most copied human expression is the smile! #ChatGPT
09/03/2023	TessRachel	Hey @Jeep - I think I may have come up with the perfect name for your all electric vehicle: JeePT #ChatGPT #EV #AI #TheFutureIsNow #jeep #yourewelcome #dontmesswithtess
09/03/2023	EdAttacked	#ChatGPT called me lame in an educational way lol
09/03/2023	mrosalesconvol1	I asked #ChatGPT if he could give me an idea to solve the twin prime conjecture Could not He told me that it was an unresolved problem and he threw me a "historical talk" So what is its intelligence? I didn't ask him to solve it, just to give an idea... nothing new under the sun.
		Up to 5000 records

2. **Preprocessing results:** in addition to the previously mentioned noise, scraped datasets may also contain unwanted characters or irregular sentences due to abbreviated writing. These characters can interfere with the quality and accuracy of the dataset, while irregular sentences can make data processing more difficult.

The following are the results of preprocessing including casedolding, cleaning, tokenizing, text normalisation, stopword removal, stemming, remove duplicate data and remove NaN (not in number) data.

Table 2 shows the results of case folding, Case folding is a process in text preprocessing that is done to uniform the characters in the data. Case folding is the process of converting all letters into lowercase letters. In this process the characters 'A'-'Z' contained in the data are converted into 'a'-'z' characters.[10]

Table 2. Sample data Casefolding result

Tweet	Casefolding
The most copied human expression is the smile! #ChatGPT	the most copied human expression is the smile! #chatgpt
Hey @Jeep - I think I may have come up with the perfect name for your all electric vehicle: JeePT #ChatGPT #EV #AI #TheFutureIsNow #jeep #yourewelcome #dontmesswithtess	hey @jeep - i think i may have come up with the perfect name for your all electric vehicle: jeept #chatgpt #ev #ai #thefutureisnow #jeep #yourewelcome #dontmesswithtess
#ChatGPT called me lame in an educational way lol	#chatgpt called me lame in an educational way lol

Table 3 shows the results of cleaning, which removes noise characters such as links, @, hashtags, etc. Data cleaning in sentiment analysis is the process of removing redundant and incorrect values in the data intended for analysis. It is an important step in the sentiment analysis process.[11]

Table 3. Sample data Cleaning result

Casefolding	Cleaning
the most copied human expression is the smile! #chatgpt	most copied human expression smile chatgpt
hey @jeep - i think i may have come up with the perfect name for your all electric vehicle: jeept #chatgpt #ev #ai #thefutureisnow #jeep #yourewelcome #dontmesswithtess	think have come with perfect name your electric vehicle jeept chatgpt thefutureisnow jeep yourewelcome dontmesswithtess
#chatgpt called me lame in an educational way lol	chatgpt called lame educational

Table 4 shows the results of tokenising, which breaks down tweets into words to facilitate the text normalisation process. Tokenising is the operation of separating text into pieces in the form of tokens, which can be pieces of letters, words, or sentences, before further analysis. Entities that can be referred to as tokens include words, numbers, symbols, punctuation marks, etc.[12]

Table 4. Sample data Tokenizing result

Cleaning	Tokenizing
Most copied human expression smile chatgpt	['most', 'copied', 'human', 'expression', 'smile', 'chatgpt']
Think have come with perfect name your electric vehicle jeept chatgpt thefutureisnow jeep yourewelcome dontmesswithtess	['think', 'have', 'come', 'with', 'perfect', 'name', 'your', 'electric', 'vehicle', 'jeept', 'chatgpt', 'thefutureisnow', 'jeep', 'yourewelcome', 'dont-messwithtess']

Table 5 shows the results of text normalisation, which is converting text into a standard or common form that can be processed and interpreted more easily. converting mistyped words into the correct form. With normalisation, text can become more consistent, easier to read, and easier to process.

Table 6 shows the results of Stopword removal, which is a process that removes words that do not function but often appear in the tweets that have been obtained.

Table 5. text notmalisation result

Tokenizing	Normalization
['technology', 'wont', 'replace', 'people', 'need', 'human', 'touch', 'these', 'complex', 'situations', 'telephone', 'operators', 'elevator', 'operators', 'radiologists', 'radiologists', 'radiology', 'chatgpt']	['technology', 'will', 'not', 'replace', 'people', 'need', 'human', 'touch', 'these', 'complex', 'situation', 'telephone', 'operator', 'elevator', 'operator', 'radiologist', 'radiologist', 'radiology', 'chatgpt']

Table 6. Sample data Stopword removal result

Normalization	Stopword
['think', 'have', 'come', 'with', 'perfect', 'name', 'your', 'electric', 'vehicle', 'jeept', 'chatgpt', 'thefutureisnow', 'jeep', 'yourewelcome', 'dont-messwithtess']	['think', 'come', 'perfect', 'name', 'electric', 'vehicle', 'jeept', 'chatgpt', 'thefutureisnow', 'jeep', 'yourewelcome', 'dontmesswithtess']

Table 7 shows the results of stemming is Stemming is a process to change the sentence into basic words, namely by removing words that contain affixes at the beginning of the sentence or at the end of the sentence.

Table 7. Sample data stemming result

No	Stemming
1	Good place 'great afford miss chatgp
2	Build product chatgpt chatgpt damn product welcome
3	Chatgpt good project
4	Maybe chatgpt source

After the stemming process is complete, the next step is to remove duplicate data, the amount of data before deleting duplicates is 5000 and the amount of data after deletion is duplicate: 4950, meaning there are 50 duplicate data then remove NaN data (Not at Number) the results obtained are empty data totalling 0, meaning that the data is very clean and ready to be used for sentiment labelling.

3. Sentiment labelling results: in the sentiment labelling process is using supervised learning method. Supervised learning is a machine learning approach that uses labelled data or datasets that are already known by the designer. These pre-designed data are expected to train "supervise" algorithms for classification or prediction of a case accurately.[13] Obtained tweet data with positive sentiment as much as 3681 and negative sentiment as much as 918 so the total amount of data from the neutral data elimination is 4599. Table 8 show sample of the results of the sentiment labelling process.

Table 8. Sample data Labelling result

Tweet_Clean	Sentiment	Value Positif	Value Negatif
Copy human expression smile chatgpt	positive	8	6
Openai chatgpt information get information late mother little know accurate detail unsure also incorrect information confused	negative	9	11

4. Implementation of Naive Bayes Classification Model result: after sentiment labelling, it means that it has entered the final stage of data analysis of twitter user sentiment towards chatGPT, in the implementation of the naive bayes classification model there are several sub-stages including data split, Tf-IDF word weighting, cross validation optimization, Naive bayes model classification, and confusion matrix evaluation.

Split data or Data splitting is a method of dividing data into two or more parts that form a subset of data. Generally, data splitting separates two parts, one part is used to evaluate or test the data and the other is used to train the model.[14] The dataset is divided into training and testing subsets with a ratio of 80:20, where 20 % of the data is used for testing. The data division was done randomly using random_state with a random number seed of 42. From the results of the division, it was found that the train data amounted to 3679 and the test data amounted to 920.

Tf-IDF word weighting, Naïve Bayes classification model using the results of TF-IDF gives better average

accuracy than without using TF-IDF. accuracy is better than without using TF-IDF.[15] there are 920 documents in the test dataset each represented by a vector with 8486 dimensions (columns), which represents the unique words in the whole dataset. Sample output of X_test_tfidf: such as (0,8230) 0,17023511025684257 shows that in the first row and 8230th column of the matrix, the TF-IDF weight value of the word represented by that column is 0,17023511025684257. for the results of sample data from Tf-IDF word weighting can be seen in figure 6 below.

```
Shape of X_test_tfidf: (920, 8486)
Sample of X_test_tfidf:
  (0, 8230)     0.17023511025684257
  (0, 7428)     0.2653983924969395
  (0, 7418)     0.15448853558764725
  (0, 7194)     0.22289034936547725
  (0, 7134)     0.2786638077407959
  (0, 5876)     0.18410084582156813
  (0, 5265)     0.24868594577599956
  (0, 5258)     0.12429085282133165
  (0, 4656)     0.25598643059695764
  (0, 4519)     0.21742491779830997
  (0, 4478)     0.21742491779830997
  (0, 3955)     0.20173494615282841
  (0, 3119)     0.2653983924969395
  (0, 3070)     0.1478565163726533
  (0, 2845)     0.1900657762444665
  (0, 2804)     0.17753559507780065
```

Figure 6. Results of sample data from Tf-IDF word weighting

The results of the Cross Validation Optimisation resulted in Best parameters: {'alpha': 0,1, 'fit_prior': True} refers to the best parameters selected when tuning the hyperparameters in the Naive Bayes model. Hyperparameter tuning is done to find the optimal parameters that can improve model performance. In this case, the tuned parameters are alpha and fit prior.

The classification results of the Naive Bayes model show an accuracy of 80 %, which indicates that this model is able to classify sentiment on the dataset well. for more detail, it will be shown from the confusion matrix test results, figure 7, shows the results of the confusion matrix can be seen in the picture below.

	precision	recall	f1-score	support
negatif	0.63	0.14	0.23	194
positif	0.81	0.98	0.89	726
accuracy			0.80	920
macro avg	0.72	0.56	0.56	920
weighted avg	0.77	0.80	0.75	920

Figure 7. Results of Confusion matrix evalution

The figure shows the model performance evaluation results in the Precision column where the precision value for negative sentiment is 63 % and for positive sentiment is 81 %. Precision measures how accurate the model is in identifying the correct sentiment, and the recall value for negative sentiment is 14 % and for positive sentiment is 98 %. Recall measures how much text with a particular sentiment can be identified by the model. And the F1-Score value for negative sentiment is 23 % and for positive sentiment is 89 %.

The model shows that the precision and recall values are higher for positive sentiment than negative sentiment, so the model can identify positive sentiment better. The F1-score value for positive sentiment is also higher than that for negative sentiment, indicating better performance in classifying positive sentiment and the low recall for negative sentiment indicates that Twitter users tend to give more positive responses to ChatGPT.

5. Data visualisation Result: the following figure 8 is a visualisation of the top 10 words contained in the dataset that have been labelled with sentiment.

https://doi.org/10.56294/dm202345

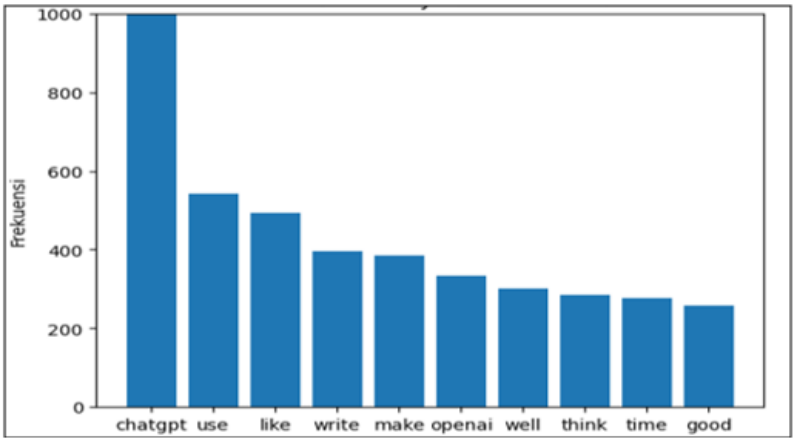

Figure 8. Frequency of occurrence of the top 10 words

From the picture above, the word "chatgpt" is the most mentioned word with a total of 5458 words, the word "use" is 543 words, the word "like" with a total of 493 words, the word "write" is 395 words, "make" is 385 words, "openai" is 333 words, "well" is 302 words, "think" is 284 words, "time" is 277 words, and "good" is 258 words. In wordcloud visualisation, it can be seen what words are often used by twitter users against ChatGPT by skip it text "chatgpt" then it will look like figure 9 below.

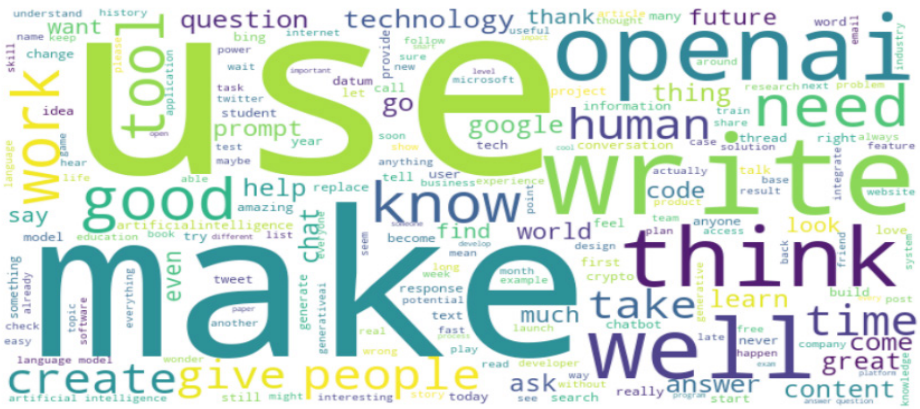

Figure 9. Wordcloud Dataset

From the visualisation, it means that twitter users use ChatGPT technology from Open AI for tools in work needs efficiently because this technology can think for itself so that it saves time.

The following in figure 10 is a visualisation of the percentage between positive and negative sentiment in a pie chart.

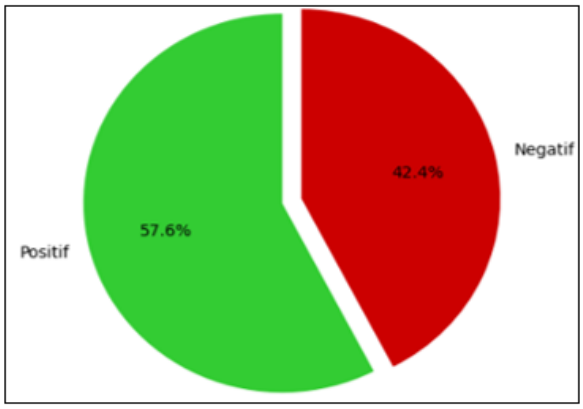

Figure 10. Pie chart of positive and negative sentiment in the dataset

For future researchers who wish to conduct research related to ChatGPT, it is recommended to include an assessment of the weaknesses of ChatGPT and summarise them in a questionnaire that can be filled out by the

public. This will help to see if people still have a positive sentiment towards ChatGPT despite its weaknesses being exposed. please try to carry out such research.

CONCLUSIONS

The implementation of the Naive Bayes algorithm to classify positive and negative sentiments on tweet data against ChatGPT resulted in an accuracy value of 80%, which means that the model built has been able to represent the classification well. Information obtained from sentiment analysis and classification in this study shows that positive sentiment is 57.6% and negative sentiment is 42.4%. From this percentage, it can be concluded that with ChatGPT, Twitter users tend to have more positive sentiments. In addition, the graph visualisation of the top 10 most spoken words and word cloud shows the word "ChatGPT" is most used, followed by words such as "use", "like", "write", "make", "OpenAI", "well", "think", "time", and "good". To prove the usefulness of ChatGPT, the author asked ChatGPT to combine these 10 words into a complete sentence. ChatGPT's answer shows that ChatGPT is an advanced language model created by OpenAI that is used for various tasks such as writing and generating content, thus making it a good tool with good performance and the ability to think at any time.

REFERENCES

1. Annur Cindy Mutia. Jumlah Pengguna Internet Global Tembus 5,16 Miliar Orang pada Januari 2023 [Internet]. databooks. 2023. https://databoks.katadata.co.id/datapublish/2023/02/03/jumlah-pengguna-internet-global-tembus-516-miliar-orang-pada-januari-2023

2. Suud Hefty. Apa Itu Chat GPT? Ini Arti Kata, Keunggulan dan Cara Menggunankannya, Chatbot Viral di Media Sosial. TribunJatim.com. 2023. https://jatim.tribunnews.com/2023/02/04/apa-itu-chat-gpt-ini-arti-kata-keunggulan-dan-cara-menggunankannya-chatbot-viral-di-media-sosial

3. A. Erfina et al., "Indonesia's Economic Recovery Post Covid-19 Pandemic Sentiment Analysis," 2022 IEEE 8th International Conference on Computing, Engineering and Design (ICCED), Sukabumi, Indonesia. 2022; pp. 1-4. https://doi.org/10.1109/ICCED56140.2022.10010695.

4. JustAnotherArchivist. A social networking service scraper in Python [Internet]. Github. 2018. https://github.com/JustAnotherArchivist/snscrape

5. Jain Deepak. Data Preprocessing in Data Mining [Internet]. https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/

6. Sutami C. PERBANDINGAN METODE KLASIFIKASI NAIVE BAYES CLASSIFIER DAN LEXICON BASED DALAM ANALISIS SENTIMEN (Studi Kasus: Twitter). 2015;

7. Komputer JS, Buatan K, Ridwan A. Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus. 2020.

8. Cross Validation: Teknik Evaluasi Machine Learning, 6 Metode [Internet]. Available from: https://digitalpolar.com/cross-validation/

9. Agrani A, Rikumahu B. PERBANDINGAN ANALISIS SENTIMEN TERHADAP DIGITAL PAYMENT "GO-PAY" DAN "OVO" DI MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN WORD CLOUD COMPARISON OF SENTIMENT ANALYSIS AGAINST DIGITAL PAYMENT "GO-PAY" AND "OVO" IN SOCIAL MEDIA TWITTER USING NAÏVE BAYES ALGORITHM AND WORD CLOUD. Agustus. 2020;7(2):2534.

10. Resti J, Selva Jumeilah F. Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian. 2017;1.

11. Role of Data Cleaning in Sentiment Analysis [Internet]. https://www.repustate.com/blog/data-cleaning-in-sentiment-analysis/

12. Rizki Aditiya. #BelajarPython 9: Operasi "Tokenizing" pada Teks Berbahasa Indonesia - Aditya Rizki's Note [Internet]. Blog. 2020. https://adityarizki.net/belajarpython-9-operasi-tokenizing-pada-teks-berbahasa-indonesia/

13. Baskoro Hayo. Supervised vs Unsupervised Learning: Apa Bedanya? [Internet]. 2022. https://pacmann.io/blog/supervised-dan-unsupervised-learning

14. Trivusi. Data Splitting: Pengertian, Metode, dan Kegunaannya - Trivusi [Internet]. 2022. https://www.trivusi.web.id/2022/08/data-splitting.html

15. Rahmayanti Setyaning Nastiti V, Basuki S. Klasifikasi Sinopsis Novel Menggunakan Metode Naïve Bayes Classifier. 2019;1(2):125–30.

FINANCING

The authors did not receive financing for the development of this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Adhitia Erfina, M Rifki Nurul R A.
Data curation: Adhitia Erfina, M Rifki Nurul R A.
Formal analysis: Adhitia Erfina, M Rifki Nurul R A.
Acquisition of funds: Adhitia Erfina, M Rifki Nurul R A.
Research: Adhitia Erfina, M Rifki Nurul R A.
Methodology: Adhitia Erfina, M Rifki Nurul R A.
Project management: Adhitia Erfina, M Rifki Nurul R A.
Resources: Adhitia Erfina, M Rifki Nurul R A.
Software: Adhitia Erfina, M Rifki Nurul R A.
Drafting - original draft: Adhitia Erfina, M Rifki Nurul R A.
Writing - proofreading and editing: Adhitia Erfina, M Rifki Nurul R A.