The Indonesian Journal of Computer Science
www.ijcs.net
Volume 13, Issue 5, October 2024
https://doi.org/10.33022/ijcs.v13i5.4278

Public Sentiment Analysis About Neuralink from Twitter Using Naïve Bayes: Multinomial, Gaussian and Complement

Azwan Triyadi[1], Purnawansyah[2], Herdianti Darwis[3]
azwantriyadi.labfik@umi.ac.id[1], purnawansyah@umi.ac.id[2], herdianti.darwis@umi.ac.id[3]
[1,2,3] Universitas Muslim Indonesia

Article Information

Received : 8 Jul 2024
Revised : 20 Sep 2024
Accepted : 30 Oct 2024

Keywords

Neuralink, sentiment, SMOTE, TF-IDF, Naïve Bayes.

Abstract

Elon Musk owns the business Neuralink, which attempts to build brain-machine interfaces. This study categorizes public opinion towards the use of Neuralink goods, including whether people agree (positive), disagree (negative), or feel neither way. Without accessing the Twitter API, the Twint Python Libraries were utilised to retrieve a dataset of 3000 using the keyword "neuralink". What datasets are included in positive, neutral, or negative categories are designated using RoBERTa. Term Frequency Inverse Document Frequency (TF-IDF) is utilized for feature extraction, while Synthetic Minority Over-sampling Technique (SMOTE) is employed to handle class imbalance. Complement Naive Bayes, achieved accuracy of 81%, followed by Multinomial Naive Bayes, which achieved accuracy of 80%, and Gaussian Naive Bayes, which achieved accuracy of 75%. The model Complement Naïve Bayes was used in this study to attain the maximum accuracy, and accuracy increases when employing SMOTE compared to other Naïve bayes variants.

A. Introduction

As the 20th century gives way to the 21st, technology is advancing in a number of areas, one of which is the medical field. As Elon Musk stated in July 2019, the business planned to create agar brain implant technology so that a technical object, like a computer, could be controlled by nothing more complicated than electrical neurons [1]. Additionally, to comprehend brain disorder and treat it, as well as to design the future in harmony with technological intelligence [2]. 96 probes, each with 32 electrodes, are implanted as part of the Neuralink device, placing a total of 3,072 electrodes throughout the brain [3]. Due to their flexible shape, robotic devices must be used in order to increase probe rigidity and assure correct placement throughout the brain [4]. Robotic testing results in an average success rate of 87.1% across 19 mouse model tests [2]. Benefits in the medical field include the ability to communicate with patients who have medical disabilities like locked-in syndrome, treat lost nerve connections like Alzheimer's disease, monitor and train human psychophysiological state and cognitive abilities, and assist in the treatment of drug-resistant epilepsy as well as the prevention of seizures [1], [5], [6]. This corporation declared on May 26, 2023, via a tweet, that the US Food and Drug Administration (FDA) has given its permission to conduct human testing.

The Multinomial Naïve Bayes (MNV), Gaussian Naïve Bayes (GNV), and Complement Naïve Bayes (CNV) approaches have been employed in numerous prior investigations. One of the researchers used word counting based on sentiment categories generated, and researchers calculated the average polarity and subjectivity, and researchers calculated the accuracy using GNV and Logistic Regression. The study was carried out in Indonesia. It is obvious from the results of the experiments that the suggested GNV approach is better than existing methods [7]. For each type of vaccine dataset, the accuracy findings from the GNV approach revealed a high average of 97.48% [7]. In another study, it was discovered that researchers used TF-IDF feature extraction to analyze a general analytic dataset in Indonesian, and they concluded that MNV performed the best. The General Analysis Dataset Indonesia is subjected to several naive bayes, which yields the maximum accuracy rating on MNV [8]. And determines the min-df and max-df. Other academics in this field look at sentiment analysis of reviews from the Internet Movie Database (IMDB), which is divided into positive and negative evaluations. The goal of the study is to locate the most widely used and accurate model. Linear Model and Naive Bayes were the models utilized, with TF-IDF being the feature extraction method. According to the test results, CNV had the two reports' highest accuracy (75.13%) [9]. In the next investigation, the team used the string-based MNV algorithm to identify feelings towards the online diabetic community. Using the Emoticons Lexicon (EMOLEX), researchers establish that there are eight different emotions: anger, sadness, fear, joy, surprise, belief, anticipation, and disgust. Researchers also employed the MNV algorithm. With an average accuracy of 74%, the results obtained demonstrate that the proposed strategy beats all models that were examined, including Emolex, Naive Bayes, and MNV [10]. Additional study use the Bernoulli, Multinomial, Gaussian, and Complements variation of the entire Naive Bayes. Between 2004 and 2005, researchers used news data from the BBC website. Term Frequency Inverse

Document Frequency (TF-IDF) feature extraction is also used by researchers. Using CNB the best accuracy score of 97.604 percent was demonstrated [11].

With the use of Neuralink technology, persons with medical issues like paralysis can communicate (talk or type) via an implanted brain. With the help of this implant, people could become as intelligent as computers. The FDA's major impact on human implants, in accord, generated popular sentiment that spread on Twitter. As a result, the researcher brought up the research issue and presented a text-mining strategy that was appropriate while also learning from earlier researchers. Researchers who employ the Naïve Bayes approach are motivated by the researcher. Numerous studies conducted in the past only employed one Naïve Bayes method or compared variables using different Naïve Bayes variations. Researchers will contrast them in this study with Naïve Bayes variations such as Multinomial Naïve Bayes, Gaussian Naïve Bayes, and Complement Naïve Bayes. Since Bernoulli Naïve Bayes can only classify data into two of three possible categories—positive, neutral, and negative—it will not be used in this study. Additionally, RoBERTa and SMOTE applications are available for this study.

The public's perception on the use of technology from This Elon Mask, particularly the FDA's approval of implanting "The Link" on the human brain, will be covered in this study to determine whether they agree, disagree, or are undecided. Using the MNV method, GNV, and CNV, the accuracy of the calculation procedure will be categorized as positive, neutral, or negative. The precision attained by the models put into practise is anticipated to be used as learning material, a reference point, or a source of reference for upcoming studies.

B. Research Method

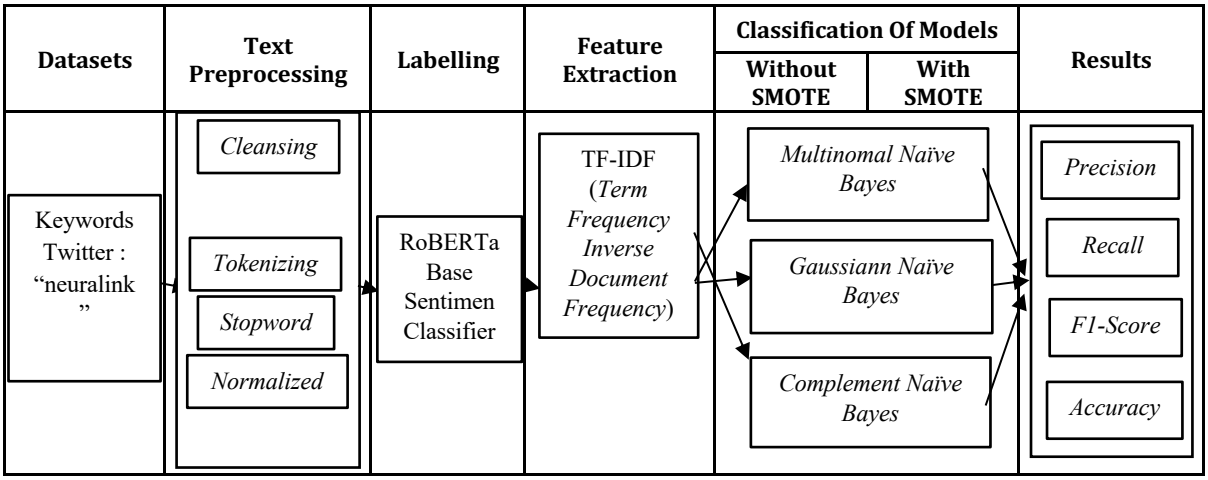


Figure 1. Research Flow

Figure 1 uses Tweets from Twitter that are in the English language. An approach to obtaining data from a website and storing the extracted data in files and databases is called scraping [12]. Twitter Intelligence Tool (Twint) is an open-source library of python that is used to collect data from Twitter [13]. To obtain the necessary precise data, the library uses the Twitter operator [13]. Without utilizing the Twitter API, we may get data from Twitter using twints. The dataset

used in this study includes tweets with the English language hashtag “neuralink” that were sent between January 1, 2023, and June 27, 2023.

The following steps are included in the preprocessing stage for English text: case folding, which is useful for changing all words to lowercase, tokenization, which is used to insert words and expand contractions (for example, we change "we're" to "we are" or "don't" to "do not"), deletion of characters, which involves changing characters such as é and ï to e and i, deletion stopword, which involves removing stopwords based on word frequency or stopword lists, remove mark punctuation and numbers: punctuation and numbers are removed entirely or replaced with special symbols, and finally stemming: which aims to reduce every word to its basic form [14]. The next step is to utilize RoBERTa to categorize sentiment as positive, neutral, or negative. The most recent model, RoBERTa, which BERT created using training data, produced significantly superior outcomes [15]. One of the tasks supported by the RoBERTa model is text classification [15]. Huggingface libraries aid in the tagging procedure. Pretraining and fine-tuning are the two stages of the RoBERTa labeling process. The RoBERTa model was trained in the "masked language modeling" challenge during the pretraining stage. This task requires the model to predict which word in a sentence is hidden based on context because some words in a sentence are hidden. A trained RoBERTa model pretraining was retrieved and customized to the intended particular natural language processing task during the step of fine-tuning. The success rate of each learning machine is determined by process feature extraction [16]. A well-known technique for determining the significance of words in a document is the TF-IDF [17]. In order to give words that frequently appear in a document but infrequently elsewhere in the corpus more weight, TF-IDF combines the values of Term Frequency (TF) and Inverse Document Frequency (IDF). Equation (1) is used to determine TF, with the i TF being the frequency of the i phrase appearing in the j text. IDF, which is defined technically in equation (2), is the logarithm of the ratio to the total number of documents in the corpus that contain the number of documents in each term. Mark was calculated by multiplying the two, as expressed in equation (3).

$$\boxed{tf_i} = \boxed{\frac{freq_i(dj)}{\sum_{i=1}^{k} freq_i(dj)}} \quad (1)$$

$$\boxed{tf_i} = \boxed{log\ log\ \frac{|D|}{|\{d:t_i \in d\}|}} \quad (2)$$

$$\boxed{(tf - idf)_{ij}} = \boxed{tf_i(d_j)} * \boxed{idf_i} \quad (3)$$

The Bayes' theorem, which asserts that each pair to be classified is independent of the other, forms the basis of the Naive Bayes classification method [18]. Text categorization works well with Multinomial Naive Bayes (MNB), a naive Bayes classifier [18].

$$\boxed{t_w} = \boxed{log\ log\ \left(\frac{\sum_{n=1}^{N} doc_n}{doc_w}\right)} \quad (4)$$

$$\boxed{Pr(c)} = \boxed{log\ log\ \pi_c\ +\ \prod_{w=1}^{|V|} f_w\ log\ log\ (t_w Pr(c))} \quad (5)$$

To account for stop words, added here Inverse Document Frequency (IDF), where t the frequency of the term in the document doc, word t_w is the number of times term t appears in a document doc in equation (4), then for further processing entered into equation (5) [19]. The distribution usually is used to calculate the Gaussian Naïve Bayes [7]. Gaussian Naive Bayes allows for the classification of categorical data and numerical data with Gaussian distribution [20]. The calculation of the Gaussian Naive Bayes in (6)

$$\boxed{P(Z)} = \frac{p(c)\,X\,p(C)}{p(z)} \tag{6}$$

According to Equation (6), C is the class label, Z is the applied attribute, temporary PZ is the probability of the preceding class, P(C) is the likelihood that it appears in the class label, and P(Z) is the likelihood that the attribute is applied [7]. With the exception of class C, all data for the Complement Naive Bayes algorithm are trained from classes [21]. Formulas Complement Naïve Bayes show in (7):

$$\boxed{CNV(d)} = \boxed{argmax_c \left[log\ log\ p(\vec{0}_c) - \sum_i \quad f_i\ log\ log\ \frac{Nc'i+ai}{Nc'+a} \right]} \tag{7}$$

In equation (7), ´ is the number of times the word is occurs in a document in a class other than c, Nc´ means the total number of word occurrences in classes other than c [21].

C. Result and Discussion

1. Datasets

When the keyword "neuralink" is used to crawl data, the results can generate up to 3000 tweets. The crawled data is kept in an Excel file with a CSV version that has three columns that are vital to the subsequent procedure: date, tweets, and username. The outcomes of using the Twitter dataset are displayed in Table 1.

Table 1. Data Crawling

Date	Tweet	Username
2023-06-22 23:59:59+00:00	@elonmusk @lexfridman My prediction: Elon uses Neuralink to summon a robot army. Zuckerberg tries to use VR to no effect. Things just got real.	MoffittWarren
2023-06-22 23:59:06+00:00	@jswatz @SamLMontano One of the thousand young of Shub-Niggurath had become a trailblazing neurosurgeon and leading researcher at NeuraLink, and CthulhuCare(tm) came through. It's a feel-good story all the way around! Er, except for the sudden dismemberment.	stevebloom55
2023-06-22 23:55:58+00:00	@TeslaAndDoge Imagine the epic battle between Zuckerberg's BJJ and Musk's Kung Fu, powered by Neuralink! I'd definitely pay to watch that!" See my profile plz	MWinter58235836
2023-06-22 23:54:55+00:00	@girldrawsghosts That's just the expression you'll make with one of them Neuralink chips in your	deadened

	brain.	
2023-06-22 23:53:52+00:00	@shivon I've never seen elon hit anything lmao 🤣🤣🤣	NeuralinkShahel

2. Text Preprocessing

Table 2 shows the data before and after preprocessing. First, the dataset will be cleaned of mentions and hashtags. Next, punctuation marks and emoticons in the form of images or code symbols will be removed. Case folding, tokenizing, stopword removal, normalization words, and stemming will then be applied to the dataset.

Table 2. Before and after Text-preprocessing

Before	**After**
@elonmusk @lexfridman My prediction: Elon uses Neuralink to summon a robot army. Zuckerberg tries to use VR to no effect. Things just got real.	my prediction elon uses neuralink to summon robot army zuckerberg tries to use vr to no effect things just got real
@jswatz @SamLMontano One of the thousand young of Shub-Niggurath had become a trailblazing neurosurgeon and leading researcher at NeuraLink, and CthulhuCare(tm) came through. It's a feel-good story all the way around! Er, except for the sudden dismemberment.	one of the thousand young of shubniggurath had become trailblazing neurosurgeon and leading researcher at neuralink and cthulhucaretm came through its feelgood story all the way around er except for the sudden dismemberment
Before	**After**
@TeslaAndDoge Imagine the epic battle between Zuckerberg's BJJ and Musk's Kung Fu, powered by Neuralink! I'd definitely pay to watch that!" See my profile plz	imagine the epic battle between zuckerbergs bjj and musks kung fu powered by neuralink id definitely pay to watch that see my profile please
@girldrawsghosts That's just the expression you'll make with one of them Neuralink chips in your brain.	thats just the expression youll make with one of them neuralink chips in your brain
@shivon I've never seen elon hit anything lmao 🤣🤣🤣	ive never seen elon hit anything lmao

3. Labelling

RoBERTa is used to categorize the data, classifying the opinions expressed in tweets as positive, neutral, or negative. Roberta was used to label the data in Table 3.

Table 3. Results of dataset labelling

tweet_remove_punct	**label**
my prediction elon uses neuralink to summon robot army zuckerberg tries to use vr to no effect things just got real	neutral

one of the thousand young of shubniggurath had become trailblazing neurosurgeon and leading researcher at neuralink and cthulhucaretm came through its feelgood story all the way around er except for the sudden dismemberment	positive
imagine the epic battle between zuckerbergs bjj and musks kung fu powered by neuralink id definitely pay to watch that see my profile please	positive
hell waddle in and activate some sort or neuralink device and all thatll be left of zuckerborg is shredded android parts	negative
ive never seen elon hit anything lmao	neutral

4. Data visualization

According to the statistics in Figure 1, 59% of the reviews or up to 1759 reviews were classified as neutral, 25% of the reviews or up to 767 reviews were classified as negative, and the remaining 16% or 474 reviews were classed as positive.

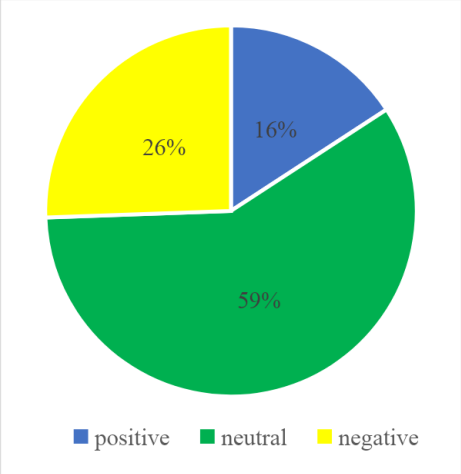


Figure 1. Percentage of public sentiment measured by RoBERTa

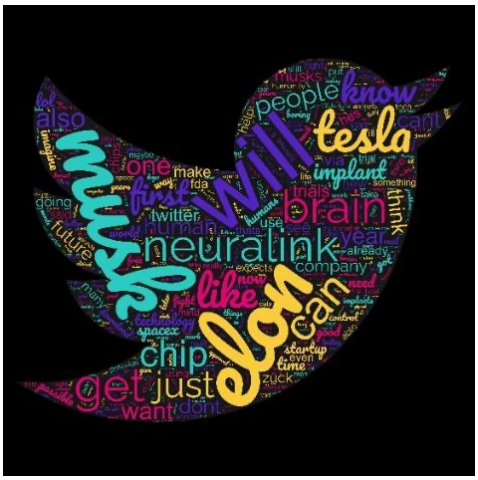


Figure 2. Employing WordCloud

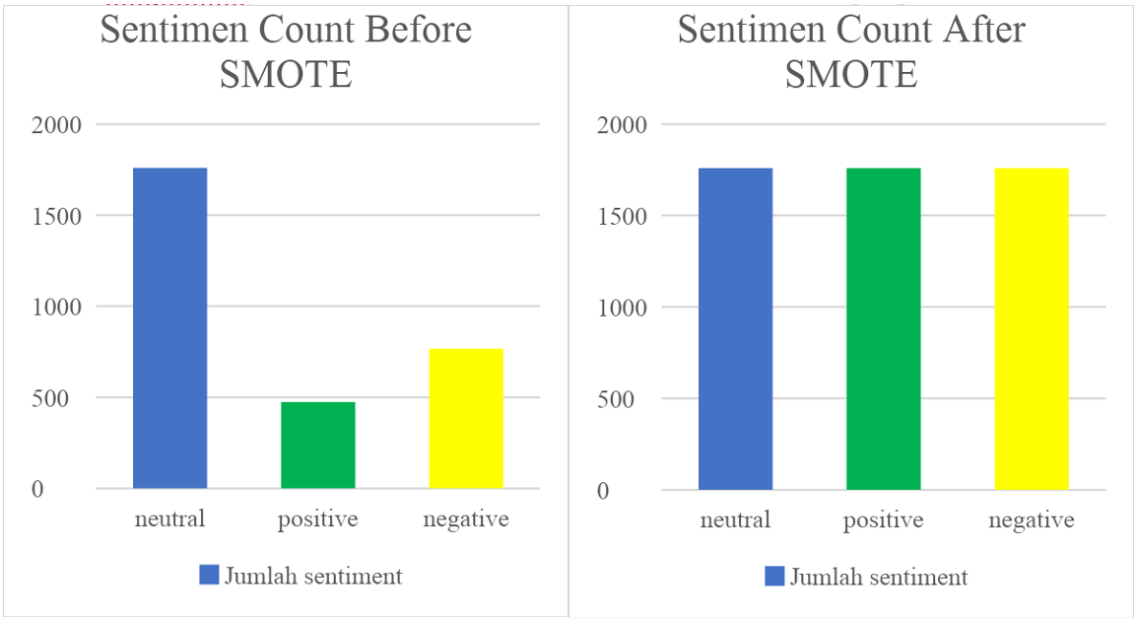
Figure 2 provides a summary of the opinions expressed by reviewers who used the Twitter keyword "neuralink" in their evaluations. The most frequently occurring words are "Elon", "Musk", "neuralink", "brain", etc.

5. Classification of sentiments

In this study, a comparison ratio of 80% : 20% was employed to divide the training and testing data.

a. Data Imbalance Handling

There is a disparity in class in Figure 3. Utilizing the oversampling technique is one way to solve the class imbalance issue [22]. One of the most popular methods used nowadays to address class imbalance with oversampling is SMOTE [23]. In order to produce synthetic data on the minority class of its closest neighbors, SMOTE makes use of distance Euclidean [24].



(a) (b)

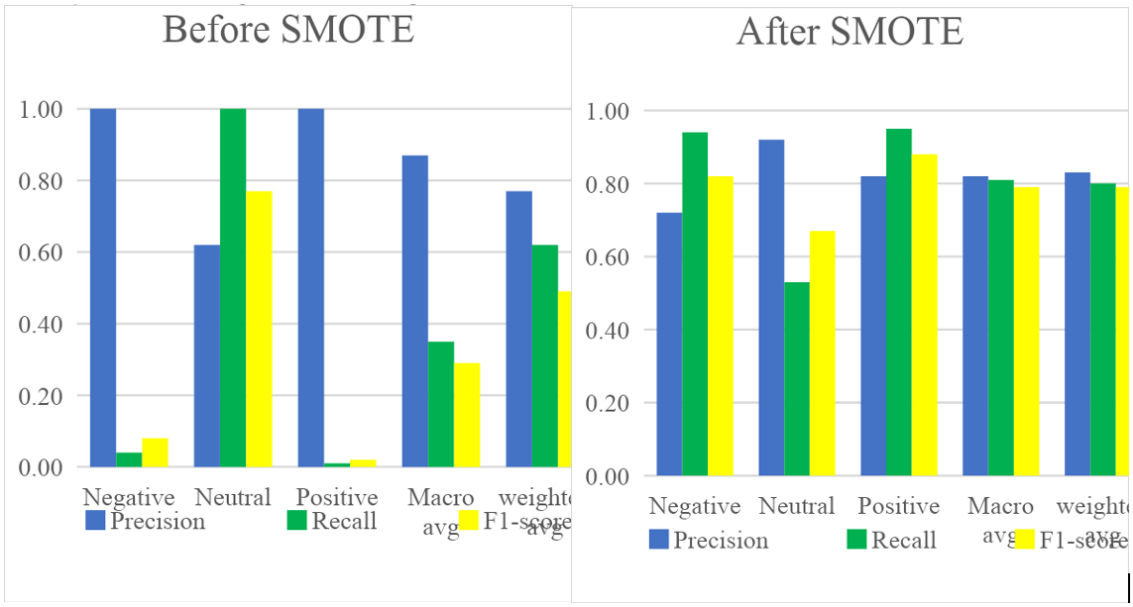
Figure 3. (a) Before SMOTE Oversampling, (b) After SMOTE Oversampling

Figure 3 depicts the total amount of feelings before the SMOTE procedure is applied. There are 1759 neutral sentiments, 474 positive sentiments, and 767 negative sentiments. Oversampling using SMOTE will provide fake data for the minority class on its nearest neighbor after the operation is complete. As a result, after the SMOTE process is completed, the number of sentiments in the neutral class remains at 1759, while the number of sentiments in the positive and negative classes likewise rises to 1759.

b. Multinomial Naïve Bayes

Figure 4 displays the outcomes of categorization performance levels using the MNV model. Figure 4 depicts precision, recall, and the f1-score before and after the SMOTE process, respectively. The effects of class attendance are 100%, 4%, and 8% prior to SMOTE. 62%, 100%, and 77% of respondents chose neutral. Positive in the class up to 100%, 1%, and

2%. On average, 87%, 35%, and 29% of the class passed. Class weighted average scores last with 77%, 62%, and 49%. Additionally, 72%, 94%, and 82% of the class are negative when the SMOTE technique was used. Received 92%, 53%, and 67% for neutral. Positive for the class up to 82%, 95%, and 88%. macro class averages of 82%, 81%, and 79%. Finally, class weighted average scores are 83%, 80%, and 79%.

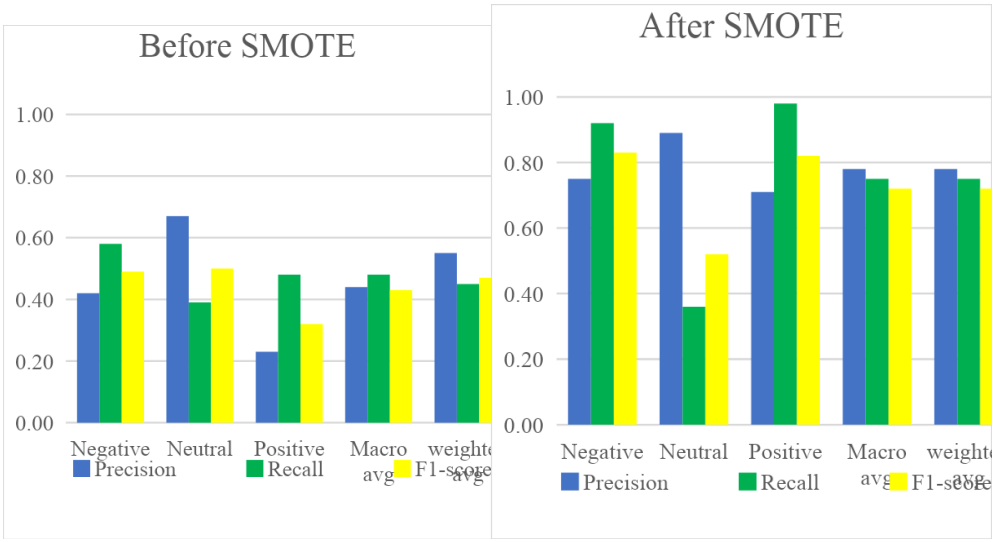


(a) (b)

Figure 4. MNV Performance (a) Before SMOTE, (b) After SMOTE.

c. Gaussian Naïve Baye

Figure 5 displays the outcomes of categorization performance levels using the GNV model. Figure 5 shows precision, recall, and f1-score before and after the SMOTE process for each. Class attendance rates were 42%, 58%, and 49% lower before SMOTE. Received 67%, 39%, and 50% for neutral. Class positive rates range from 23% to 48% to 32%. On average for the class, 44%, 48%, and 43%. Finally, class weighted average scores are 55%, 45%, and 47%. Additionally, 75%, 92%, and 83% of the samples are negative after the SMOTE process. Gains for neutral were 89%, 36%, and 52%. Up to 71%, 98%, and 82% of the class were positive. macro average for the class of 78%, 75%, and 72%. Final results: 78%, 75%, and 72% on a class weighted average.

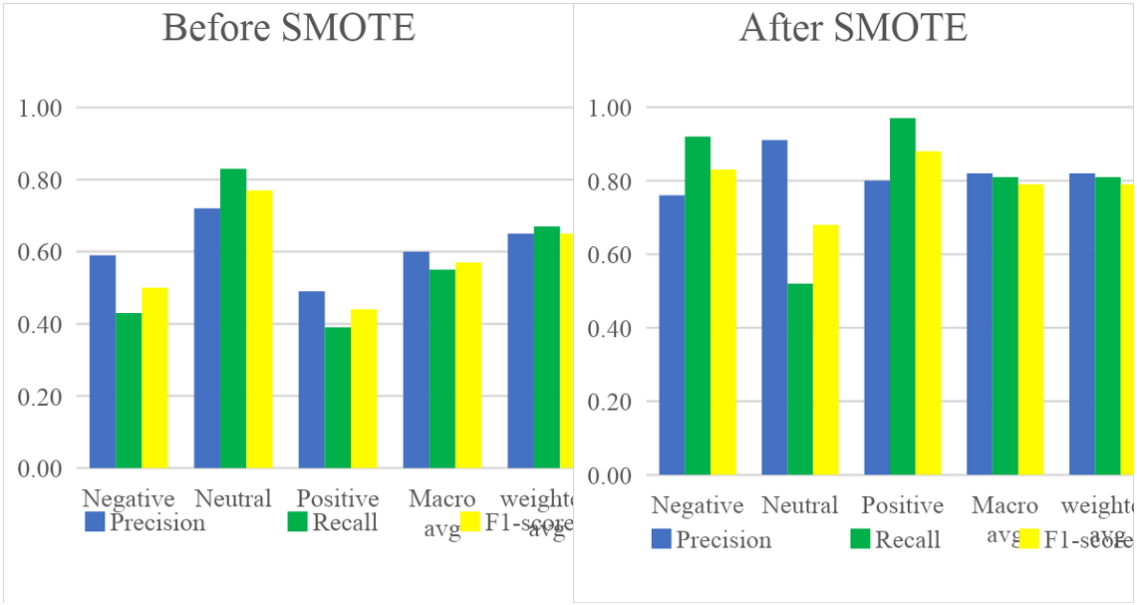


(a) (b)

Figure 5. GNV Performance (a) Before SMOTE, (b) After SMOTE.

d. Complement Naïve Bayes

Figure 6 displays the outcomes of categorization performance levels using the CNV model. Figure 6 shows the precision, recall, and f1-score measurements prior to and during the SMOTE process. Class attendance is negative by 59%, 43%, and 50% prior to SMOTE. 72%, 83%, and 77% of respondents chose neutral. Positive class ratings of up to 49%, 39%, and 44%. 60%, 55%, and 57% on the class macro average. Avg scores from the previous class were 65%, 67%, and 65%. Additionally, the class negative is 76%, 92%, and 83% following the SMOTE process. The results for neutral were 91%, 54%, and 68%. Positive in the class up to 80%, 97%, and 88%. On average for the class, by 82%, 81%, and 79%. Weighted average from last class: 82%, 81%, and 79%.



(a) (b)

Figure 6. CNV Performance (a) Before SMOTE, (b) After SMOTE.

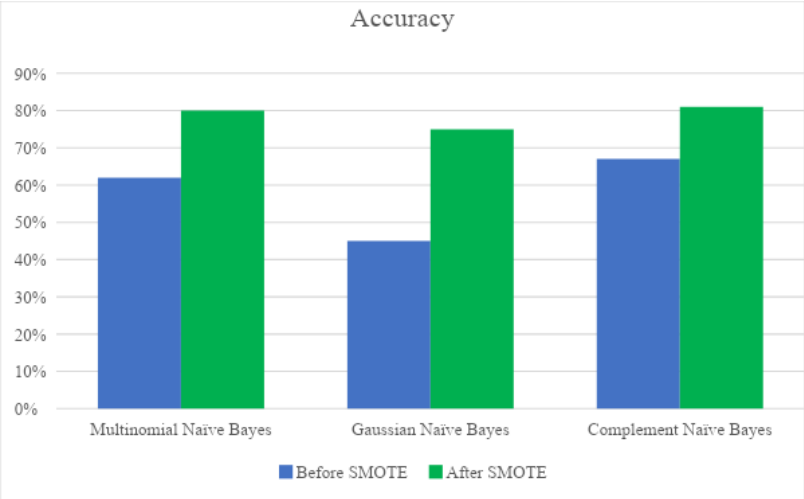


Figure 7. Multinomial, Gaussian, and Complement Naïve Bayes Model Accuracy

The accuracy outcomes from using all three Naïve Bayes models are shown in Figure 7. Complement Naïve Bayes models yield the best accuracy, followed by Multinomial Naïve Bayes models, while Gaussian Naïve Bayes models yield the worst accuracy.

D. Conclusion

The Complement Naive Bayes model used in this investigation generates the highest amount of accuracy, with accuracy prior to SMOTE being 67% and accuracy post-SMOTE being 81%. After that, use the Multinomial Naive Bayes model, which yielded 80% accuracy after SMOTE and 62% accuracy before SMOTE. Finally, employing the Gaussian Naive Bayes model before SMOTE yields 45% accuracy and 75% accuracy after SMOTE. The accuracy attained is still less accurate than the other two models, however the Gaussian Naïve Bayes model saw the most increases once SMOTE was used, increasing by 30%. As can be observed, applying techniques like oversampling and SMOTE increases accuracy. Future research can employ a labeling procedure that can get a balance of data to provide higher accuracy. This study still relies data imbalance managing strategies such as oversampling.

E. Acknowledgment

The researcher would like to thank the facilities given by the Research Laboratory Faculty of Computer Science at the Muslim University of Indonesia for their assistance in carrying out this research.

F. References

1. É. Fourneret, “The Hybridization of the Human with Brain Implants: The Neuralink Project,” Cambridge Quarterly of Healthcare Ethics, vol. 29, no. 4, pp. 668–672, Oct. 2020, doi: 10.1017/S0963180120000419.
2. A. Kulshreshth, A. Anand, and A. Lakanpal, “Neuralink- An Elon Musk Start-up Achieve symbiosis with Artificial Intelligence,” Proceedings - 2019

International Conference on Computing, Communication, and Intelligent Systems, ICCCIS 2019, vol. 2019-January, pp. 105–109, Oct. 2019, doi: 10.1109/ICCCIS48478.2019.8974470.

3. E. Musk and Neuralink, "An Integrated Brain-Machine Interface Platform With Thousands of Channels," J Med Internet Res 2019;21(10):e16194 https://www.jmir.org/2019/10/e16194, vol. 21, no. 10, p. e16194, Oct. 2019, doi: 10.2196/16194.
4. B. Fiani, T. Reardon, B. Ayres, D. Cline, and S. R. Sitto, "An Examination of Prospective Uses and Future Directions of Neuralink: The Brain-Machine Interface," Cureus, Mar. 2021, doi: 10.7759/cureus.14192.
5. T. Dadia and D. Greenbaum, "Neuralink: The Ethical 'Rithmatic of Reading and Writing to the Brain," https://doi.org/10.1080/21507740.2019.1665129, vol. 10, no. 4, pp. 187–189, Oct. 2019, doi: 10.1080/21507740.2019.1665129.
6. A. N. Pisarchik, V. A. Maksimenko, and A. E. Hramov, "From Novel Technology to Novel Applications: Comment on 'An Integrated Brain-Machine Interface Platform With Thousands of Channels' by Elon Musk and Neuralink," J Med Internet Res 2019;21(10):e16356 https://www.jmir.org/2019/10/e16356, vol. 21, no. 10, p. e16356, Oct. 2019, doi: 10.2196/16356.
7. N. G. Ramadhan and F. D. Adhinata, "Sentiment analysis on vaccine COVID-19 using word count and Gaussian Naïve Bayes," Indonesian Journal of Electrical Engineering and Computer Science, vol. 26, no. 3, pp. 1765–1772, Jun. 2022, doi: 10.11591/ijeecs.v26.i3.pp1765-1772.
8. N. Umar and M. Adnan Nur, "Application of Naïve Bayes Algorithm Variations On Indonesian General Analysis Dataset for Sentiment Analysis," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 6, no. 4, pp. 585–590, Aug. 2022, doi: 10.29207/resti.v6i4.4179.
9. C. Dewi and R. C. Chen, "Complement Naive Bayes Classifier for Sentiment Analysis of Internet Movie Database," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 13757 LNAI, pp. 81–93, 2022, doi: 10.1007/978-3-031-21743-2_7/COVER.
10. V. Balakrishnan and W. Kaur, "String-based Multinomial Naïve Bayes for Emotion Detection among Facebook Diabetes Community," Procedia Comput Sci, vol. 159, pp. 30–37, Jan. 2019, doi: 10.1016/J.PROCS.2019.09.157.
11. H. Abbas Salman and T. Hameed Obaida, "Bbc News Data Classification Using Naïve Bayes Based On Bag Of Word," 2021.
12. G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6. pp. 734–749, Jun. 2005. doi: 10.1109/TKDE.2005.99.
13. V. Santoni and S. Rufat, "How fast is fast enough? Twitter usability during emergencies," Geoforum, vol. 124, pp. 20–35, Aug. 2021, doi: 10.1016/J.GEOFORUM.2021.05.007.
14. Z. Rahimi and M. M. Homayounpour, "The impact of preprocessing on word embedding quality: a comparative study," Lang Resour Eval, vol. 57, no. 1, pp. 257–291, Mar. 2023, doi: 10.1007/S10579-022-09620-5/FIGURES/9.

15. G. Reddy Narayanaswamy, “Exploiting BERT and RoBERTa to Improve Performance for Aspect Based Sentiment Analysis,” 2021, doi: 10.21427/3w9n-we77.
16. E. Hossain, O. Sharif, and M. Moshiul Hoque, “Sentiment Polarity Detection on Bengali Book Reviews Using Multinomial Naïve Bayes,” Advances in Intelligent Systems and Computing, vol. 1299 AISC, pp. 281–292, 2021, doi: 10.1007/978-981-33-4299-6_23/COVER.
17. R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, “The Impact of Features Extraction on the Sentiment Analysis,” Procedia Comput Sci, vol. 152, pp. 341–348, Jan. 2019, doi: 10.1016/J.PROCS.2019.05.008.
18. A. Sachdeva and I. Kashyap, “Empirical Analysis of Support Vector Machine and Multinomial Naive Bayes,” 2022. [Online]. Available: www.ijraset.com
19. M. Abbas, K. Ali, A. Jamali, K. Ali Memon, and A. Aleem Jamali, “Multinomial Naive Bayes Classification Model for Sentiment Analysis Overview of China View project Compact Equivalent Circuit Model for RF CMOS Transistors View project Multinomial Naive Bayes Classification Model for Sentiment Analysis,” IJCSNS International Journal of Computer Science and Network Security, vol. 19, no. 3, p. 62, 2019, doi: 10.13140/RG.2.2.30021.40169.
20. J. H. Lau and T. Baldwin, “An empirical evaluation of doc2vec with practical insights into document embedding generation,” Proceedings of the Annual Meeting of the Association for Computational Linguistics, pp. 78–86, 2016, doi: 10.18653/v1/w16-1609.
21. B. Seref and E. Bostanci, “Performance of Naïve and Complement Naïve Bayes Algorithms Based on Accuracy, Precision and Recall Performance Evaluation Criterions,” 2019. [Online]. Available: http://www.meacse.org/ijcar
22. S. Feng, J. Keung, X. Yu, Y. Xiao, and M. Zhang, “Investigation on the stability of SMOTE-based oversampling techniques in software defect prediction,” Inf Softw Technol, vol. 139, p. 106662, Nov. 2021, doi: 10.1016/J.INFSOF.2021.106662.
23. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/JAIR.953.
24. M. Umer et al., “Scientific papers citation analysis using textual features and SMOTE resampling techniques,” Pattern Recognit Lett, vol. 150, pp. 250–257, Oct. 2021, doi: 10.1016/J.PATREC.2021.07.009.