Twitter Sentiment Analysis for Exploring Public Opinion on the Merdeka Belajar-Kampus Merdeka (MBKM) 2023 with the Naïve Bayes Classifier Algorithm

by - -

Submission date: 01-Jul-2023 09:00AM (UTC+0300)
Submission ID: 2273017079
File name: Public_Opinion_on_the_Merdeka_Belajar-Kampus_Merdeka_MBKM.docx (271.14K)
Word count: 4891
Character count: 27532

ICEBIT
The 4th International Conference On Economics , Business and Information Technology
June 10, 2023 pp. 852-860

Twitter Sentiment Analysis for Exploring Public Opinion on the Merdeka Belajar-Kampus Merdeka (MBKM) 2023 with the Naïve Bayes Classifier Algorithm

Mas'ud Hermansyah
Institut Teknologi dan Sains Mandala
Jl. Sumatra No.118-120, Jember
+62 813-3646-6463
masudhermansyah@itsm.ac.id

M. Faiz Firdausi
Institut Teknologi dan Sains Mandala
Jl. Sumatra No.118-120, Jember
+62 852-5832-6262
faizfirdausi@itsm.ac.id

Abdul Wahid
Institut Teknologi dan Sains Mandala
Jl. Sumatra No.118-120, Jember
+62 852-5707-6330
abdulwahid@itsm.ac.id

Nur Andita Prasetyo
Institut Teknologi dan Sains Mandala
Jl. Sumatra No.118-120, Jember
+62 823-2920-1769
nurandita.prasetyo69@itsm.ac.id

ABSTRACT

The Merdeka Belajar-Kampus Merdeka (MBKM) as one of the public policies by the Ministry of Education, Culture, Research and Technology, cannot be separated from public opinion. Opinion directions are divided into three categories, positive opinion, negative opinion, and neutral opinion. Utilizing tweet data from Twitter with the keyword Merdeka Belajar-Kampus Merdeka (MBKM) in May 2023, a sentiment analysis was carried out to identify public opinion on the Merdeka Belajar-Kampus Merdeka (MBKM) program. Through this research, it is hoped that opinions, opinion factors, and problems that may arise from the implementation of the Merdeka Belajar-Kampus Merdeka (MBKM) policy can be identified as early as possible. The Naïve Bayes Classifier is used to classify the direction of a person's sentiment towards the Merdeka Belajar-Kampus Merdeka (MBKM), both positive sentiment, neutral sentiment, and negative sentiment. Dataset collection and preparation begins with feature selection, eliminating duplication and tweet selection, then pre-processing is carried out, namely case folding, tokenizing, character cleaning, normalization to stemming for use in labeling. Of the 1,212 data used, the model managed to classify 453 sentiments as neutral, 477 negative sentiments, and 282 positive sentiments. With these results, information is obtained that the community still has many pros and cons regarding the implementation of the Merdeka Belajar-Kampus Merdeka (MBKM) program. In this study from the results of Twitter sentiment analysis on the Merdeka Belajar-Kampus Merdeka (MBKM) program, the Naïve Bayes Classifier algorithm produces an accuracy value of 74.25%.

Keywords : Sentiment Analysis; Twitter; Merdeka Belajar-Kampus Merdeka (MBKM); Naïve Bayes Classifier

INTRODUCTION

Higher education plays an important role in the development of human resources and economic growth of a country. In the Indonesian context, the Merdeka Belajar-Kampus Merdeka (MBKM) program is the government's effort to renew higher education by giving freedom to students in determining the curriculum, choosing cross-disciplinary courses, and developing creativity and entrepreneurship. This program has the goal of producing graduates who have skills that are relevant to the needs of the world of work and are able to compete globally.

The main policy programs of the MBKM are facilitating the opening of new research programs, improving higher education accreditation schemes, facilitating official state universities to become integrated tertiary institutions, and freedom for students to study outside the study program for three semesters. Students are given the freedom to attend lectures outside the curriculum, three semesters in the form of 1 (one) semester opportunity to attend lectures outside the curriculum, and two semesters to carry out study activities outside the tertiary institution. Various types of learning activities outside the university, including internships (practice work in business or other workplaces), community service projects in villages, teaching in school units, engaging in student exchanges, conducting research, conducting entrepreneurial activities, conducting academic studies/projects, and participate in humanitarian service. Based on the lecturer's instructions, all of these assignments must be carried out. The hope is that the MBKM policy can provide opportunities for students to contextually gain experience off campus which can strengthen students' abilities as a whole, ready to work or build a new career. (Rochana et al., 2021), (Pratiwi et al., 2016).

However, to ensure the successful implementation of the MBKM program, it is important to understand the views, perceptions and sentiments that arise from various stakeholders, especially students who are program participants. The MBKM program has pros and cons from the community, especially students who have experienced the impact and have perceptions of the MBKM policy. As one of the new policies, it is necessary to analyze and evaluate the programs being implemented to improve their performance

through feedback from the public (Pipin & Kurniawan, 2022). Sentiment analysis is an effective approach to gain insight into how the program is received and responded to by program participants, as well as their views regarding the benefits and challenges that arose during its implementation. (Indrayuni, 2016). Sentiment analysis is needed to obtain information in the form of text that can be used as a reputation management and evaluation tool for companies, agencies or brands, this information can be found in online media such as Twitter (Oktasari et al., 2016).

Twitter is one of the social media used for expression, sharing statements and opinions in the form of short text (tweets). Based on a we are social survey in February 2022, there were 18.45 million Indonesian users using Twitter (We Are Social). In using social media, especially Twitter, every individual has needs that they seek to obtain satisfaction through social media. In this case, users consciously consume media based on the individual's desire to fulfill their needs so that satisfaction will be achieved within them (Pangestika & Kusna, 2023). It is predicted that the number of Twitter users will continue to grow because Indonesia has a penetration of 93.4 million internet users. This allows Twitter to have a very fast dissemination and distribution of information (Prasetyo et al., 2021). Through Twitter, users can express their opinions freely. Twitter also provides a trending topic feature to find out the hottest news that is currently being discussed. The very rapid development of Twitter is very interesting to analyze various kinds of opinions and phenomena (Irawansyah et al., 2022).

Research (Rochana et al., 2021) and research (Pipin & Kurniawan, 2022) expressed various criticisms and public responses regarding the MBKM program coming from various disciplines and users. This program has been widely discussed by Indonesian Twitter users since the program was published in early 2020, related to issues of curriculum design and the mechanism for awarding credits, unclear MBKM learning outcomes, availability of cooperation with partners and partners' needs, issues of MBKM program funding, allowance apprenticeships, to the issue of exploitation of cheap labor in industry through apprenticeship programs.

The large number of Twitter users who convey comments or opinions can be utilized by classifying tweet reviews so that information is obtained. One method that can be used to analyze opinions in tweets is sentiment analysis. The data obtained from the tweets is processed by text mining, followed by classifying the tweet data into three classes, positive, negative and neutral. In this study the classification uses the Naïve Bayes Classifier (NBC) algorithm (Fransiska & Yolanda, 2019) with the hope of being able to recognize and classify public opinion on twitter with a high degree of accuracy.

2. RESEARCH METHOD

In analyzing sentiment and knowing its accuracy, there are several steps to get the best results. The steps taken are data collection and labeling, pre-processing, sentiment analysis and evaluation to obtain accuracy and sentiment data visualization. The entire sentiment analysis methodology that was carried out can be seen in Figure 1 below.

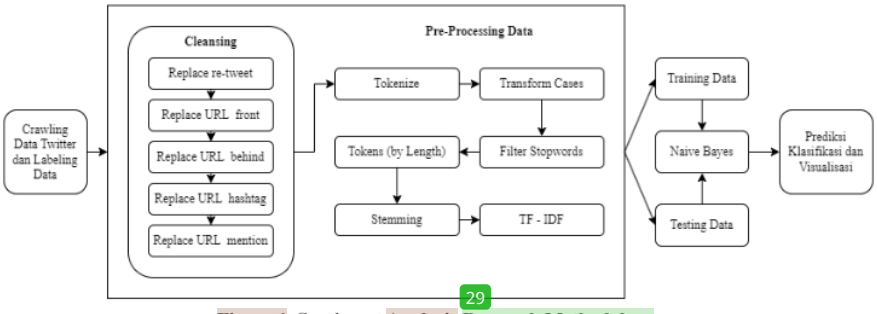


Figure 1. Sentiment Analysis Research Methodology

2.1 Data Collection

In the data collection process there are 3 parts, namely:

1. Twitter data collection (crawling) the tweet data obtained is data that is directly taken from the Twitter site with keywod "mbkm, kampus merdeka, kampus mengajar, msib, pertukaran mahasiswa merdeka (pmm), praktisi mengajar, wirausaha merdeka" using the Rapidminer application.
2. Save data to the Microsoft Excel application. The crawled data is imported into the Rapidminer application with the help of the Naïve Bayes Classifier method for processing.

3. Provide data labels/sentiments to several Twitter comments that have been successfully crawled. There are 3 (three) sentiments, namely: positive, negative, and neutral.

2.2 Pre-Processing Data

There are several stages in preprocessing, including:

1. Repetitive tweet deletion.
2. Data cleaning (cleaning) to delete URLs, both the front URL and the back URL.
3. Data deletion (cleaning) to remove punctuation marks, hashtags, mentions
4. Tokenize, processes to divide text which can be in the form of sentences, paragraphs or documents, into certain tokens/parts.
5. Case folding, the case folding process is the process of changing all letters to lowercase.
6. Elimination of stop words (slang words), is a step taken to remove non-descriptive words.
7. Stemming, is the process of mapping and decomposing the form of a word into its basic word form.
8. Filter token by length, is a part to display the number of letters in 1 word with certain limitations.
9. Term frequency (a word that often appears) and its visualization.

2.3 Classification with the Naive Bayes Classifier Method

Data classification and labeling use functions available in the Rapidminer application package with Bayes algorithm parameters. One of the tasks of Data Mining is data classification, namely mapping (classifying) data into one class or several classes that have been defined previously. One of the methods in data classification is Naïve Bayes, which is a machine learning method that utilizes probability and statistical calculations put forward by the British scientist Thomas Bayes, the way Naive Bayes works is to predict future probabilities based on previous experience. (Darwis et al., 2021). The basis of the Naïve Bayes Classifier used in programming is the Bayes formula as in the equation (1) (Ling et al., 2014):

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Information:
X = Data sample with unknown class (label).
H = Hypothesis that X is data with class (label) C
P(H|X) = Probability that the hypothesis is true (valid) for observed X sample data
P(X|H) = Probability of sample data X, if it is assumed that the hypothesis is true (valid)
P(H) = Probability of hypothesis H
P(X) = Probability of observed sample data

The purpose of this test is to be able to automatically classify tweets as positive, negative or neutral tweet sentiment. In the Naive Bayes Classifier method, the process of classifying text is carried out based on training data that has previously been stored. In its implementation there are three stages, namely making a list of training data, making a list of testing data, and making a classifier (Ratnawati, 2018).

After the classification process, an evaluation is needed to determine the quality of the process carried out. At this evaluation stage, the performance of the classification process that has been carried out will be tested with accuracy, precision, and recall parameters. The average values of accuracy, precision, recall can be obtained from the confusion matrix table as shown in Table 1 below.

Table 1. Confussion Matrix

	True Yes	True No
Prediction Yes	TP	FP
Prediction No	FN	TN

Source: (Widowati & Sadikin, 2021)

Accuracy is the number of documents that are classified correctly, when the prediction results from the model with a predetermined sentiment, with 3 types of sentiment, namely: positive, negative, and neutral. The accuracy value can be calculated by equation 2.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2)$$

Evaluation of this system analysis is carried out by calculating the level of accuracy of a method in analyzing opinions.

2.4 Sentiment Visualization

Data visualization is presenting data in a simpler and more attractive form. This is done to make it easier to understand directly the information from the data that has been processed. Data visualization is divided into two, namely data presentation in the form of a wordcloud and presentation in the form of a histogram (Nurulbaiti & Subekti, 2018).

Word cloud is a text mining method that displays a graphic of word frequency which further emphasizes the words that appear more often in the source text. The bigger the word size in the visual, the more common the word is in the

document (Josen Limbong et al., 2019). Sentiment visualization modeling uses the package provided in the Rapidminer application.

3. RESULT AND DISSCUSION

This section describes the results of the research conducted. This study aims to determine the sentiment analysis of opinion tweets on the implementation of MBKM. Then measure the accuracy of the classification results with the Naive Bayes Classifier method. This study uses a dataset of tweets which totals 2.235 tweets. After passing through the cleaning stage, there are 1.212 data tweets that can be used for sentiment analysis. By using a comparison of training data and testing data of 8:2. Generates processing data details of 970 for training data and 242 for testing data.

3.1 Crawling Process

The first stage in the sentiment analysis process is data collection. Data was taken from Twitter with a crawling process that was carried out using the Twitter API with Rapidminer Studio tools version 9.10.008. The tweets taken were tweets with the keywords mbkm, independent campus, teaching campus, msib, independent student exchange (pmm), teaching practitioners, and independent entrepreneurs. Then the crawled data is labeled manually, labels for positive, negative and neutral sentiments. The model used for crawling Twitter data is shown in Figure 2.

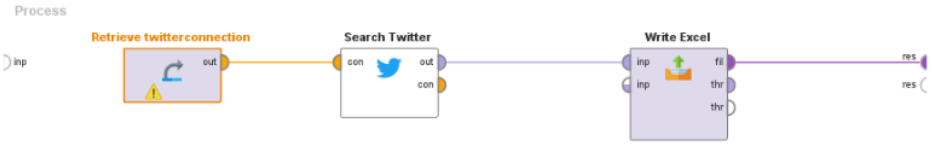


Figure 2. Process Crawling Data

Figure 2 shows data retrieval from Twitter using the "Retrive Twitter Connection" operator which functions to connect Rapidminer with Twitter's social media. Next is the "Search Twitter" package to search for Twitter data according to the specified parameters, then save the data in an excel file using the "Write Excel" operator. Twitter data crawling related to MBKM was carried out for 30 days in May 2023. In that month several MBKM programs have been running and are entering the final phase. There are many sentiments given by the community regarding the implementation of MBKM in 2023.

3.2 Labeling Data

The next stage is labeling, labeling is done to divide the data into 3 (three) sentiment classes to be used, positive, negative and neutral. The criteria used in labeling tweets are if the tweets contain positive words, sentences that are expressed positively and intend to approve, support, and appreciate the MBKM program are categorized as positive. If the tweet contains negative words and intends to disapprove or protest about the implementation of MBKM, then the tweet will be categorized as negative. While tweets that only provide information or contain neutral words, the tweets will be categorized as neutral tweets. It looks like Table 2 below is an example of a dataset that has been labeled:

Table 2. Contoh Dataset dengan Label

Text	Sentiment
Program MSIB yang diusung pemerintah ini sangat bermanfaat bagi mahasiswa https://t.co/hxd5WQR1Xq	Positif
BBH MSIB ni emg parah bgt... telat terus pdhl tuntutan pekerjaan di kantor jg udh selevel karyawan...	Negatif
@fleurlovinxxx @collegemenfess Halo kak, kakak ikut kampus mengajar pas semester berapa?	Netral
RT @unimudasorong: Selamat kepada 30 Dosen Pengampu yang lolos dalm Program Praktisi Mengajar Angkatan 2 di UNIMUDA Sorong https://t.co/Luj…	Positif
RT @collegemenfess: [cm] kalian udah pada baca ini belum? kasian msib batch ini ada lagi kendalanya. bantu up guys, kasian anak magang…	Negatif
Sosialisasi MSIB, Bangkit dan Kampus Mengajar Semester Ganjil 2023/2024. https://t.co/boQQWcUiYA	Netral

Source: Crawling Data Twitter

As many as 300 data were labeled manually, the remaining data that did not have labels was done automatically using the Rapidminer application. The labeling model in an automatic way can be seen in Figure 3.

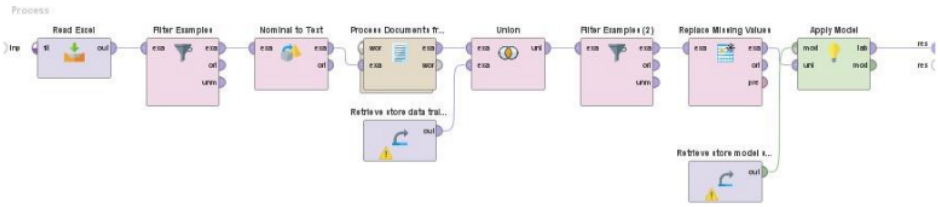


Figure 3. Pelabelan Otomatis dengan Rapidminer

Figure 3 describes several packages that are used in automatic labeling, including::

1. Read Excel: used to read all data that will be used for labeling. In this data consists of data that already has a label and does not have a label.
2. Filter Examples: used to filter data that already has a label and data that does not yet have a label.
3. Nominal to Text: used to change nominal data to text.
4. Process Document from Data: contains several processes for cleaning data.
5. Union: used to combine training data with testing data.
6. Retrive store training data: used to read training data.
7. Filter Example 2: used to delete data that has multiple attributes.
8. Retrive store model: used to retrieve the classification model for labeling automatically.
9. Apply model: used to apply the labeling classification model automatically.

3.3 Pre-Processing Data

After labeling the data, the next step is pre-processing. Both training data and testing data must go through a pre-processing process first. Pre-processing is required before the classification process so that the dimensions of the vector space model become smaller. By reducing the dimensions of the vector space model, the classification process will be faster. The purpose of this pre-processing is to homogenize and reduce the volume of words (Adi, 2018). This stage is carried out to make the data easily and feasible for analysis. The stages in this preprocessing are divided into two stages, in the subprocess before the data is divided into training data and data testing, and also pre-processing in the document process. The cleansing steps carried out in the subprocess are deleting the RT writing, deleting the URL on the front and back, deleting the hashtag, deleting the mention (@), and deleting the symbols in the tweet. The operator for the first stage of pre-processing is shown in Figure 4 below:

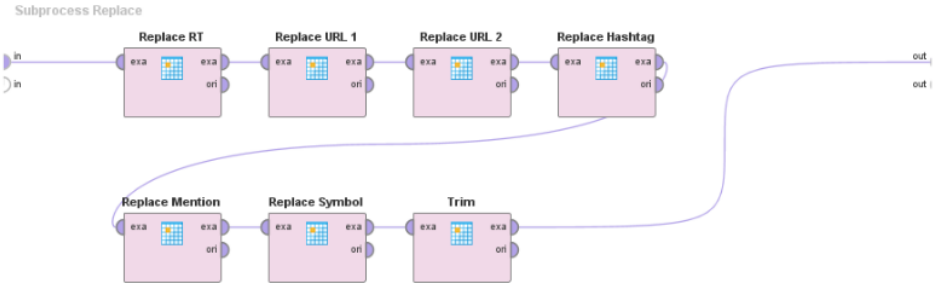


Figure 4. Process Cleansing Data

While the second stage, preprocessing in "Document Process" uses the operators "Tokenize", "Transform Case", "Filter Token by Length", "Filter Stopword (Dictionary)" and "Stem (Dictionary)" which use a dictionary in Indonesian. As seen in Figure 5 below:

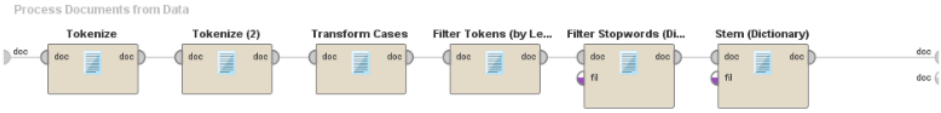


Figure 5. Proses Pre-processing dalam Document Process

The following is an explanation of the preprocessing stages above:

1. Tokenize
   String truncation based on each word that makes up the document. After separating each word from the document, case folding is performed to remove numbers, punctuation, glyphs, and other characters not used by word processors. Case folding results must go through the cleaning stage, namely the process of

removing components such as HTML tags, links, urls, hashtags (#), username (@), and website addresses. (Nirwandani et al., 2021).

2. Transform Cases
   The case conversion feature makes it possible to automatically convert all letters in a text to lowercase or to uppercase as a whole. In this study, most of the text is in the form of opinions and most of the letters are lowercase, so all letters are changed to all lowercase. (Jaka, 2015).
3. Filter Token (by Length)
   Token filtering (by length) is the process of extracting significant words from the resulting words. Words of a certain length are deleted. In this study, researchers used words with a minimum of 4 letters and a maximum of 25 letters.
4. Filter Stop Words (Dictionary)
   Filter stop words (dictionary) is the process of removing words that appear frequently but do not affect the extraction of sentiment from reviews. Words that contain verbs and interrogations (Lilyani Asri Utami, 2017).
5. Stemming
   Is the stage for carrying out the process of changing words that are rich in infixes and suffixes into a basic word that will contain more meaning to obtain information so that comments will become more specific in categorizing (Salam et al., 2018).

All of the above cleansing processes are part of the TF-IDF algorithm model, which is a weighting algorithm composed of two values derived from two algorithms with different weights, namely Term Frequency (TF) and Inverse Document Frequency (IDF). (Februariyanti et al., 2020). The results of the TF-IDF document process can be seen in Table 3.

Table 3. Preprocessing Results to be Entered for Term Frequency

Preprocessing	Result
Early tweet	Program MSIB yang diusung pemerintah ini sangat bermanfaat bagi mahasiswa https://t.co/hxd5WQR1Xq #mbkm #kampusmerdeka
Tokenize	Program MSIB yang diusung pemerintah ini sangat bermanfaat bagi mahasiswa
Transform cases	program msib yang diusung pemerintah ini sangat bermanfaat bagi mahasiswa
Filter token (by length)	program msib yang diusung pemerintah sangat bermanfaat bagi mahasiswa
Filter stopwords (dictionary)	program msib diusung pemerintah sangat bermanfaat mahasiswa
Stemming	program msib usung pemerintah sangat manfaat mahasiswa

Source: Data Preprocessing

Basically TF-IDF works by calculating the relative frequency of a word that appears in a document compared to the inverse proportion of the word that appears in the entire document set. Intuitively, this calculation can be used to find out how relevant the word is in a particular document.

3.4 Process Classification with the Naive Bayes Method

After pre-processing the data, the next step is the classification of sentiment analysis. This stage is the stage for providing training and implementing various machine learning algorithms. Figure 6 shows the contents of the "Cross Validation" operator in the Rapidminer application. In this process, the Naïve Bayes Classifier classification algorithm is used. After the experiment and the proposed model have been made, then an experiment is carried out by testing the existing model with the dataset that has been grouped into training data and testing data. "Performance" operator is an operator to display accuracy results.

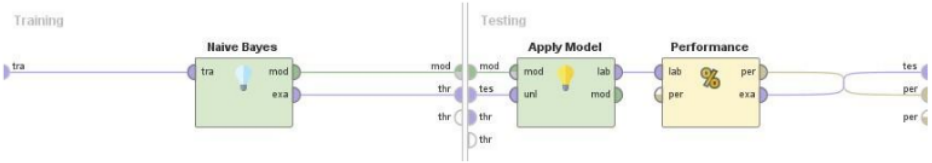


Figure 6. Sentiment Analysis Classification

Open in Turbo Prep Auto Model Filter (1,212 / 1,212 examples): all

Row No.	Label	prediction(L...	confidence(...	confidence(...	confidence(...	aaaaa	aaaaaaaaaa...	aamiin	aam
1	negatif	netral	0	1	0	0	0	0	0
9	negatif	negatif	0	0	1	0	0	0	0
30	negatif	negatif	0	0	1	0	0	0	0
36	negatif	negatif	0	0	1	0	0	0	0
39	negatif	netral	0	1	0	0	0	0	0
41	negatif	negatif	0	0	1	0	0	0	0
45	negatif	netral	0	1	0	0	0	0	0

Figure 7. Sentiment Prediction Results

Figure 7 shows the results of sentiment prediction in the RapidMiner application Comparing the predetermined sentiment label with the predicted result label from the process. Classification using the Naive Bayes Classifier method results in an accuracy value of 74.25%. Accuracy describes how much accurate the model that has been made can classify the data correctly. Accuracy is obtained from calculating the ratio of correct predictions to all data. By knowing the magnitude of the value of accuracy on the overall performance of the system, it can be stated that the level of the system's ability to find the accuracy between the information that the user wants and the answers given by the system. The results of calculating accuracy, precision and recall can be seen in Table 4.

Tablr 4. The Calculation Results Confusion Matrix

n = 242	Actual Neutral	Actual Negative	Actual Positive
Neutral Predictions	88	8	14
Negative Predictions	18	38	10
Positive Predictions	8	4	54

Source: Data Processing

$$accuracy = \frac{88 + 38 + 64}{242} = 74{,}25\%$$

Performance evaluation is carried out on the experimental results of the sentiment analysis system and on the results of the respondent's sentiment analysis. The result of accuracy produces a value from 0 to 1 and a better result is a value that is closer to 1 (Annisa Rahmaniar Dwi Pratiwi & Erwin Budi Setiawan, 2020).

3.5 Data Visualization

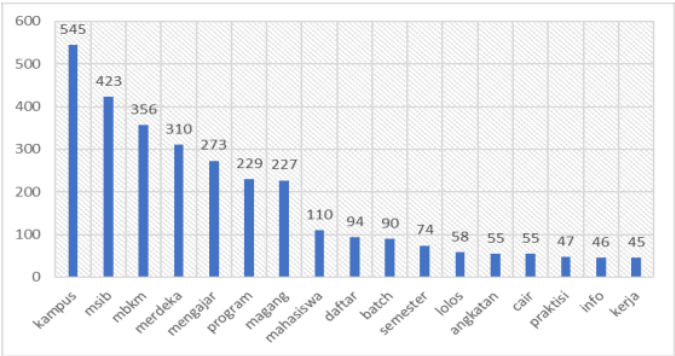
Based on the results of the tokenization carried out, the writer also wants to know the frequency of words that are widely discussed by Twitter users, for this reason the writer visualizes them in the form of wordcloud in Figure 7..

Figure 7. Popular Word Visualization with Wordcloud

Wordcloud will display the word with the most frequency, the more frequency the word will stand out more than the other words. The frequency with which the word appears in these tweets shows information about the topics most frequently discussed by the public on the social network Twitter related to MBKM. These topics include campus, mbkm, merdeka, msib, megajar, programs, and internships. The 7 (seven) words are words that are often discussed by Twitter users related to the implementation of MBKM. Words appear on wordcloud vary in size from the largest to the smallest color with the intention of showing the order of the frequency of occurrence of words in the data from the highest frequency.

The number of word frequencies related to the impmentation sentiment of MBKM can be seen in Figure 8 by displaying the numerical values for words that often appear in the sentiments given by the community. The histogram

in Figure 8 shows the most frequently mentioned words in MBKM's tweets. In addition, there are also 17 other words whose frequency of occurrence follows the most words.



Figur 8. Popular Word Visualization with Graph

In addition to data visualization from wordcloud, there is also a visualization derived from the number of sentiment divisions resulting from predictions and labeling using the Naïve Bayes Classifier algorithm. As seen in Figure 9.

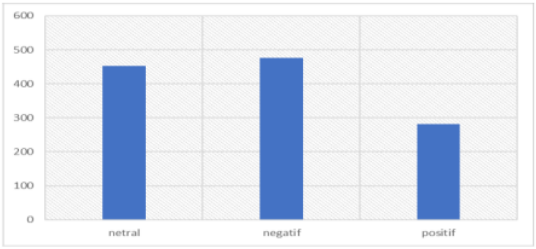


Figure 9. Sentiment Data Visualization Results

Figure 8 above shows the model analysis of 1,212 new data resulting from cleansing and modeling the Naïve Bayes Classifier algorithm. Of the 1,212 data used, the model managed to classify 453 sentiments as neutral, 477 negative sentiments, and 282 positive sentiments. With these results, information was obtained that the Twitter community in May 2023 gave more negative sentiments, namely 39% of all sentiments which were greater for the MBKM program.

It can also be seen that the value of neutral sentiment is also quite large, namely 453 sentiments which shows that as much as 37% of sentiments do not lead to agree or disagree with the MBKM program, or sentiments that do not lead to the topic of the MBKM program, so it is categorized as neutral. From the positive sentiment data there are 282 data or 23% sentiment which shows the number of sentiments that support the MBKM program is less than the other two classes.

CONCLUTION

Based on the results of the discussion, it can be concluded that the sentiment analysis of the MBKM program from user tweets in May 2023 illustrates that sentiment analysis information using the Naïve Bayes Classifier can be applied to predict the direction of one's sentiment towards MBKM, whether positive sentiment, neutral sentiment, or negative sentiment. Based on the results of an analysis of the patterns contained in the research data, of the 1,212 data used, the model succeeded in classifying 453 sentiments as neutral, 477 sentiments negative, and 282 sentiments positive. With these results, information is obtained that the community still has many pros and cons regarding the implementation of the MBKM program. In this study, from the results of Twitter sentiment analysis on the MBKM program, the Naïve Bayes Classifier algorithm produces an accuracy value of 74.25%.

5. REFERENCES

Adi, S. (2018). Perancangan Klasifikasi Tweet Berdasarkan Sentimen Dan Fitur Calon Gubernur DKI Jakarta 2017. *Journal Of Informatic Pelita Nusantara*, *3*(1), 10–16.

Annisa Rahmaniar Dwi Pratiwi, & Erwin Budi Setiawan. (2020). Implementation of Rumor Detection on Twitter Using the SVM Classification Method. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *4*(5), 782–789. https://doi.org/10.29207/resti.v4i5.2031

Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review

Data Twitter BMKG Nasional. *Jurnal Tekno Kompak*, *15*(1), 131–145. https://doi.org/10.33365/jtk.v15i1.744
Februariyanti, H., Firmansyah, M., Wibowo, J. S., & Utomo, M. S. (2020). Analisis Sentimen Tanggapan Terhadap Aplikasi Layanan Informasi Penginapan Menggunakan Metode Klasifikasi Naive Bayes. *SemanTIK*, *6*(2), 115–124.
Fransiska, S., & Yolanda. (2019). Analisis Sentimen Twitter Untuk Review Film Menggunakan Algoritma Naive Bayes Classifier (NBC) Pada Sentimen R Programming. *Jurnal Siliwangi*, *5*(2), 68–71.
Indrayuni, E. (2016). Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Evolusi*, *4*(2), 20–27.
Irawansyah, R. S., Irfan, L. A. S., & Wiriasto, G. W. (2022). Analisis Sentimen Terhadap Program Merdeka Belajar-Kampus Merdeka (MBKM) pada Twitter Menggunakan Algoritma Naive Bayes Classifier (NBC). *Jurnal Repository Unram*.
Jaka, A. T. (2015). Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining. *Jurnal Informatika UPGRIS*, *1*, 1–9.
Josen Limbong, J. A., Sembiring, I., & Dwi Hartomo, K. (2019). Analisis Klasifikasi Sentimen Ulasan Pada E-Commerce Shopee Berbasis Word Cloud dengan Metode Naive Bayes dan K-Nearest Neighbor Analysis of Review Sentiment Classification on E-Commerce Shopee Word Cloud Based With Naïve Bayes and K-Nearest Neighbor. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIIK)*, *9*(2), 347–356. https://doi.org/10.25126/jtiik.202294960
Lilyani Asri Utami. (2017). Analisis Sentimen Opini Publik Berita Kebakaran Hutan Melalui Komparasi Algoritma Support Vector Machine Dan K-Nearest Neighbor Berbasis Particle Swarm Optimization. *Jurnal Pilar Nusa Mandiri*, *13*(1), 103–112.
Ling, J., Kencana, I. P. E. N., & Oka, T. B. (2014). Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square. *E-Jurnal Matematika*, *3*(3), 92–99. https://doi.org/10.24843/mtk.2014.v03.i03.p070
Nirwandani, E. P., Indriati, & Wihandika, R. C. (2021). Analisis Sentimen Pada Ulasan Pengguna Aplikasi Mandiri Online Menggunakan Metode Modified Term Frequency Scheme Dan Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *5*(3), 1039–1047.
Nurulbaiti, F., & Subekti, R. (2018). Analisis Sentimen Terhadap Data Tweet Untuk Badan Penyelenggara Jaminan Sosial (Bpjs) Menggunakan Program R. *Jurnal Pendidikan Matematika Dan Sains UNY*, 1–9.
Oktasari, L., Chrisnanto, Y. H., & Yuniarti, R. (2016). Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Niave Bayes Classifier. *Prosiding SNST*, *7*, 37–42. https://www.publikasiilmiah.unwahas.ac.id/index.php/PROSIDING_SNST_FT/article/view/1506/1589
Pangestika, D. A., & Kusna, I. (2023). Efektivitas Media Sosial Twitter Terhadap kepuasan Followers Akun @fessthai. *Journal of Innovation Research and Knowledge*, *2*(8), 3061–3070.
Pipin, S. J., & Kurniawan, H. (2022). Analisis Sentimen Kebijakan MBKM Berdasarkan Opini Masyarakat di Twitter Menggunakan LSTM. *Jurnal SIFO Mikroskil*, *23*(2), 197–208. https://doi.org/10.55601/jsm.v23i2.900
Prasetyo, H. D., Pramiyati, T., & Isnainiyah, I. N. (2021). Sentimen Analisis Pengguna Twitter Terhadap Kebijakan Merdeka Belajar Menggunakan Algoritma Naive Bayes. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, *April*, 559–568.
Pratiwi, I., Rorong, A. J., & Rares, J. J. (2016). Pengaruh Implementasi Merdeka Belajar Kampus Merdeka Magang Terhadap Kompetensi Mahasiswa Jurusan Ilmu Administrasi Fakultas Ilmu Sosial dan Politik Universitas Sam Ratulangi. *Jurnal Administrasi Publik*, *IX*(2), 1–23.
Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng - Seri Informatika*, *3*(1), 50. https://doi.org/10.35314/isi.v3i1.335
Rochana, R., Darajatun, R. M., & Ramdhany, M. A. (2021). Pengaruh Implementasi Kebijakan Kampus Merdeka terhadap Minat dan Keterlibatan Mahasiswa. *Journal of Business Management Education (JBME)*, *6*(3), 11–21. https://ejournal.upi.edu/index.php/JBME/article/view/40165
Salam, A., Zeniarja, J., & Khasanah, R. S. U. (2018). Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia). *Prosiding SINTAK*, 480–486.
Widowati, T. T., & Sadikin, M. (2021). Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, *11*(2), 626–636. https://doi.org/10.24176/simet.v11i2.4568

Twitter Sentiment Analysis for Exploring Public Opinion on the Merdeka Belajar-Kampus Merdeka (MBKM) 2023 with the Naïve Bayes Classifier Algorithm

ORIGINALITY REPORT

18%	11%	12%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Jember Student Paper	1%
2	Submitted to Universitas Siliwangi Student Paper	1%
3	www.slideshare.net Internet Source	1%
4	Barep Prasetyo, Yaya Sudarya Triana, Rahmat Budiarto, Deris Stiawan. "Intelligent Sentiment Analysis On Cellphone Checking By Police Officer", 2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2022 Publication	1%
5	ojs.uninus.ac.id Internet Source	1%
6	Lecture Notes in Computer Science, 2003. Publication	1%

7	trilogi.ac.id Internet Source	1 %
8	eprints.unram.ac.id Internet Source	1 %
9	Submitted to Universitas Diponegoro Student Paper	1 %
10	Imam Riadi, Sunardi Sunardi, Panggah Widiandana. "Mobile Forensics for Cyberbullying Detection using Term Frequency - Inverse Document Frequency (TF-IDF)", Jurnal Ilmiah Teknik Elektro Komputer dan Informatika, 2020 Publication	1 %
11	journaltest.iums.ac.ir Internet Source	1 %
12	ojs.uho.ac.id Internet Source	1 %
13	Ridwan Suhud, Arafat Febriandirza, Intan Permatasari, Farhan Ramadan. "Recognizing Public Satisfaction Toward Kampus Mengajar Program with Naive Bayes", 2023 International Conference on Computer, Control, Informatics and its Applications (IC3INA), 2023 Publication	<1 %

14	Evi Dewi Sri Mulyani, Dani Rohpandi, Fityan Atqia Rahman. "Analysis Of Twitter Sentiment Using The Classification Of Naive Bayes Method About Television In Indonesia", 2019 1st International Conference on Cybernetics and Intelligent System (ICORIS), 2019 Publication	<1%
15	Arief Fatchul Huda, Asep Solih Awalluddin, Yulia Dwi Lestari, Fahrudin Muhtarulloh. "Gaussian Kernel Naïve Bayes Classifier on Hadith Corpus", 2023 9th International Conference on Wireless and Telematics (ICWT), 2023 Publication	<1%
16	Dwi Andini Putri, Dinar Ajeng Kristiyanti, Elly Indrayuni, Acmad Nurhadi, Denda Rinaldi Hadinata. "Comparison of Naive Bayes Algorithm and Support Vector Machine using PSO Feature Selection for Sentiment Analysis on E-Wallet Review", Journal of Physics: Conference Series, 2020 Publication	<1%
17	Jaouja Maiga, Gilbert Gutabaga Hungilo, Pranowo. "Comparison of Machine Learning Models in Prediction of Cardiovascular Disease Using Health Record Data", 2019 International Conference on Informatics,	<1%

Multimedia, Cyber and Information System (ICIMCIS), 2019
Publication

18	jurnal.unsil.ac.id Internet Source	<1 %
19	Aju Tjatur Nugroho Krisnaningsih, Tri Ida Wahyu Kustyorini, Dimas Pratidina Puriastuti Hadiani, Waluyo Edi Susanto et al. "The level of satisfaction of internal stakeholders on the implementation of the relevance of the MBKM-based learning process at the Faculty of Animal Science, Universitas PGRI Kanjuruhan Malang", Journal of Education Technology and Inovation, 2022 Publication	<1 %
20	doaj.org Internet Source	<1 %
21	Putri Talia Cantika, Gomal Juni Yanris, Mila Nirmala Sari Hasibuan. "Analysis of Public Interest in Telkomsel Cards Using the Decision Tree Method", Sinkron, 2023 Publication	<1 %
22	jurnal.umk.ac.id Internet Source	<1 %
23	Nurhayati, Armanda Eka Putra, Luh Kesuma Wardhani, Busman. "Chi-Square Feature Selection Effect On Naive Bayes Classifier	<1 %

Algorithm Performance For Sentiment Analysis Document", 2019 7th International Conference on Cyber and IT Service Management (CITSM), 2019
Publication

---

24 ejournal.unitomo.ac.id
Internet Source
<1%

---

25 Arif Ridho Lubis, Santi Prayudani, Muharman Lubis, Okvi Nugroho. "Sentiment Analysis on Online Learning During the Covid-19 Pandemic Based on Opinions on Twitter using KNN Method", 2022 1st International Conference on Information System & Information Technology (ICISIT), 2022
Publication
<1%

---

26 Denni Kurniawan, Muhammad Yasir, Farah Chikita Venna. "Optimization of Sentiment Analysis using Naive Bayes with Features Selection Chi-Square and Information Gain for Accuracy Improvement", 2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2022
Publication
<1%

---

27 Kemal Refta Diska, Khairi Budayawan. "Sistem Informasi Prediksi Kelulusan Menggunakan Metode Naive Bayes Classifer (Studi Kasus:
<1%

	Prodi Pendidikan Teknik Informatika)", Jurnal Pendidikan Tambusai, 2023 Publication	
28	repos.ust.edu.sd:8080 Internet Source	<1 %
29	www.mdpi.com Internet Source	<1 %
30	www.polgan.ac.id Internet Source	<1 %
31	www.stmik-budidarma.ac.id Internet Source	<1 %
32	I Made Bhaskara Gautama, I Gusti Ngurah Wikranta Arsa, Ni Putu Virgi Savita. "Visitor Counter and Information Viewer at Photo Exhibitions using Embedded Systems and Web Services", Paradigma - Jurnal Komputer dan Informatika, 2022 Publication	<1 %
33	citeseerx.ist.psu.edu Internet Source	<1 %
34	journal.trunojoyo.ac.id Internet Source	<1 %
35	jurnal.buddhidharma.ac.id Internet Source	<1 %
36	Aditya Hastami Ruger, M Suyanto, Mei P Kurniawan. "Sentimen Analisis Pelanggan	<1 %

	Shopee di Twitter dengan Algoritma Naive Bayes", Journal of Information Technology, 2021 Publication	
37	Arizal Mujibtamala Nanda Imron, Zilvanhisna Emka Fitri, Aji Wahyu Prasetyo, Abdul Madjid et al. "Feature Selection Application to Classify Medicinal Plant Leaves using LVQ", 2022 International Conference on Electrical Engineering, Computer and Information Technology (ICEECIT), 2022 Publication	<1 %
38	Submitted to Winston Churchill Middle School Student Paper	<1 %
39	cdn.repository.uisi.ac.id Internet Source	<1 %
40	eprints.unisbank.ac.id Internet Source	<1 %
41	journal.amikveteran.ac.id Internet Source	<1 %
42	journal.unnes.ac.id Internet Source	<1 %
43	polgan.ac.id Internet Source	<1 %
44	repository.unsri.ac.id Internet Source	<1 %

45	techniumscience.com Internet Source	<1%
46	Adinda Pratiwi Siregar, Deci Irmayani, Mila Nirmala Sari. "Analysis of the Naïve Bayes Method for Determining Social Assistance Eligibility Public", SinkrOn, 2023 Publication	<1%
47	repository.unika.ac.id Internet Source	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		