Sentiment Analysis of Kampus Merdeka Policy on Twitter Using Support Vector Machine and Naïve Bayes Classifier

Mohammad Nashrullah [1,*], Devi Ajeng Efrilianda [1]

[1] Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Semarang, Indonesia
*Corresponding author: nashrul2359@students.unnes.ac.id

ARTICLE INFO

Article history
Received 2 Maret 2022
Revised 6 Maret 2022
Accepted 4 April 2022

Keywords
Sentiment Analysis
Independent Campus
Twitter
Support Vector Machine
Naïve Bayes Classifier

ABSTRACT

Merdeka Belajar Kampus Merdeka (MBKM) is a program that was inaugurated by the Ministry of Education and Culture in 2020 which emphasizes the independence and independence of learning. One of the social media that gives many opinions on this policy is Twitter. The sentiments written by the public about the independent campus policy can be analyzed and categorized as positive or negative sentiments as material for review. In this research, sentiment analysis on the independent campus policy was carried out with the support vector machine algorithm and naïve Bayes classifier. Sentiment analysis begins by crawling data on Twitter in the period from November 20, 2021 to December 19, 2021, with a total of 5980 data. Then preprocessing the data is carried out to normalize and clean the data before data classification is carried out. Data that has gone through preprocessing is then labeled using Vader. Furthermore, word vectorization was carried out with TF-IDF and data classification to test the accuracy of sentiment analysis with the support vector machine algorithm and naïve Bayes classifier. The test results for 20 times show that the highest level of accuracy is obtained by the support vector machine algorithm with an accuracy of 73.12%.

1 Introduction

Merdeka Belajar Kampus Merdeka (MBKM) is a program that was inaugurated by the Indonesian Ministry of Education and Culture in 2020 inspired by K.H. Dewantara emphasizes independence and learning independence (Fuadi & Aswita, 2021). Twitter users often share their opinions in a tweet as a form of their freedom of expression. The opinions or sentiments they write are aimed at various objects, including the new policy of the Ministry of Education and Culture, namely the independent campus (Gunawan et al., 2018). The sentiments written by the public about the independent campus policy can be analyzed and categorized as positive or negative sentiments as material for review. Sentiment analysis is an analytical approach to analyzing texts that aims to determine the subjectivity of opinions based on certain categories (Wongkar & Angdresey, 2019).

Text mining, which is a branch of artificial intelligence, has the goal of extracting useful information from unstructured textual data through the identification and exploration of patterns (Zulfikar et al., 2017). Naïve Bayes classifier and support vector machine is one of the popular text mining algorithms and is suitable for classifying text (Octaviani et al., 2020). The naïve Bayes classifier algorithm consists of a structural model and strong assumptions about the dependence of each random variable. Naïve Bayes is also commonly used for problem classification because of its simplicity and effectiveness (Song et al., 2017). In addition, Naïve Bayes has the advantage of classifying data more flexibly as needed (Gunawan et al., 2018). Support vector machine is a statistical classification approach that is based on maximizing the margin between instances and separation hyper-plane (Al Amrani et al., 2018). The support vector machine is considered accurate in text classification because it is able to solve linear and non-linear problems, besides that this method has the advantage of obtaining the required kernel function by separating linear data from high-dimensional non-linear data input (Santoso et al., 2019). Support vector machines are also considered to easily outperform other machine learning techniques because they perform consistently well across all domains (Rani & Bhatt, 2020).

2 The Proposed Algorithm

2.1 Support Vector Machine (SVM)

SVM is a linear classification based on the principle of maximizing margins to classify two groups of data in a higher dimensional space (Styawati & Mustofa, 2019). The SVM can classify based on cases into different categories by building hyperplanes by training the support vector machine algorithm using training data so that the classification can be categorized effectively (Devi et al., 2016).

The SVM algorithm is an optimization problem that aims to maximize margin (Ren et al., 2019). By maximizing the margin between the training data and the class boundary, the resulting decision function only depends on the training data, thus the support vector machine algorithm can minimize maximum losses and provide good accuracy (Parikh & Shah, 2016).

2.2 Naïve Bayes Classifier

Naïve Bayes classifier is a supervised classification paradigm that considers the assumption of strong independence (Dey et al., 2016). The Naïve Bayes classifier makes it possible to capture uncertainty about the model in a principled manner with probability that helps solve diagnostic and predictive problems by providing a useful perspective for understanding and evaluating many learning algorithms, while being robust against noise in inputs (Parveen, 2016).

Naïve Bayes classifier uses machine learning techniques to build its predictive model. For example, a data is entered into a classifier to be classified, this model calculates the probability that the data will enter each class, then the data is entered into the class with the highest probability (Meti et al., 2017). To perform the classifier, the mixed model concept is used which is able to determine the probability of components consisting of Bayes theorem to appear as a probabilistic classifier, another name that is Naïve Bayes is known as simple Bayes or independent Bayes (Rahat et al., 2020).

3 Method

In this research, sentiment analysis of Kampus Merdeka policy on twitter is performed by applying SVM and Naïve Bayes classifier as text classification method. The desired results in this research are the accuracy of the proposed method. The flowchart of this research method can be seen in Figure 1.

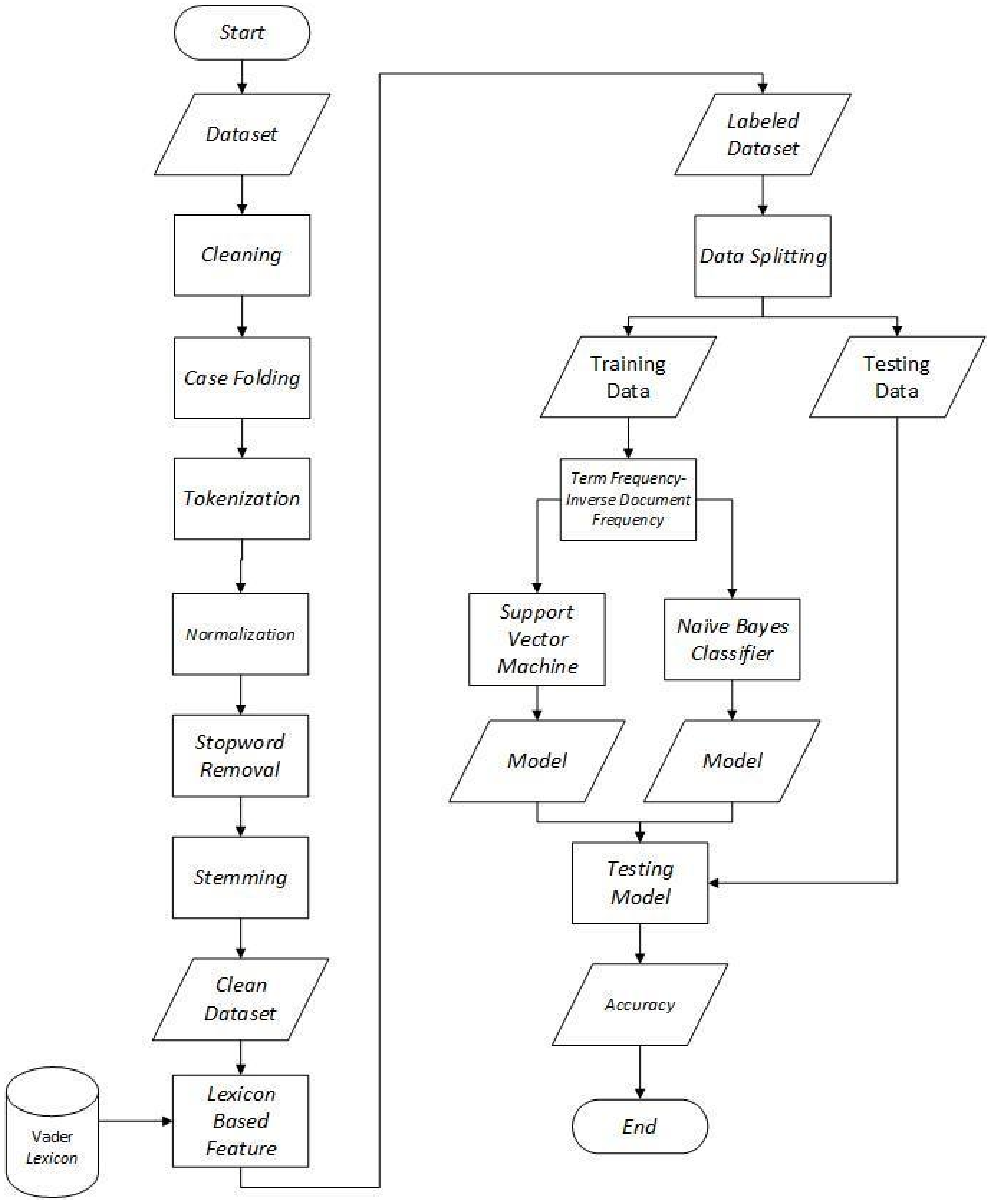

Figure 1. Figure of research steps

In this research, sentiment analysis was carried out using the SWM algorithm and Naïve Bayes classifier with TF-IDF word vectorization and labeling using Vader. The research on analyzing public sentiment about the Kampus Merdeka policy on Twitter with the SVM algorithm and Naïve Bayes classifier was carried out in several stages. The first stage is crawling data on Twitter, then preprocessing the data with stages of cleaning, case folding, tokenization, normalization, stopword removal, and stemming. The next step is to label the data using Vader, then do word vectorization of the data with TF-IDF and finally do the classification using the support vector machine algorithm and Naïve Bayes classifier.

4 Results and Discussion

This section is divided to 2 sections, result and discussion. Result is a description of the data and findings which are obtained using the methods and procedures described in the data collection method. Discussion is a review of the results which answers the research questions more comprehensively.

4.1 Result

This research compares the SVM algorithm and Naïve Bayes classifier in analyzing public sentiment towards the Kampus Merdeka policy on Twitter. The results of the research are as follows.

4.1.1 Data Collection Results

The results of data collection carried out by crawling on Twitter with a time span of 20 November 2021 to 19 December 2021 and obtained 5980 data, all of which are in Indonesian.

Table 1. The result of crawling tweet data

Created at	Tweet	Username
2021-12-19 20:33:04+00:00	b'@rafikurniaputra kalo dr webnya emg dikit, jare kebanyakan daftar lewat perush nya dulu, bru nanti daftar via web kampus merdeka(pas udh muncul)'	Muhammad Irza
2021-12-19 16:37:22+00:00	b'mau tanya dong waktu apply magang kampus merdeka, aku kan sedang menjalani smt 5 ya, jadi di KRS nilainya 0 0 0 0 0 terus jadinya di web kampus merdeka jadi segini:(ada yg tau harus lapor kemana? [cm] https://t.co/T5Fn2zogFd'	COLLE
2021-12-19 16:35:34+00:00	b'-dips! Magang di kampus merdeka itu harus udh punya skill yg dibutuhin company yg dituju ya? Kalau otak kosong gk bisa ya?'	undipmenfess
2021-12-19 16:19:09+00:00	b'@sbyfess Bisa ikut programnya kampus merdeka sampe selesai, alhamdulilah masih itu aja \xf0\x9f\x99\x82'	carbonara udon
2021-12-16 14:03:06+00:00	b'apa kubilang, program rintisan seperti ini bakal banyak bobroknya. which is why I fully avoid kampus merdeka \xf0\x9f\x92\x80\xf0\x9f\x92\x80\xf0\x9f\x92\x80 https://t.co/UyGFX1Lils'	ground on geo

After the data is obtained, data processing can be implement. The preprocessing stage consists of several stages, such as data cleaning, case folding, tokenization, normalization, stopword removal, and stemming, which aims to clean tweets of unwanted words and characters so that it can help the algorithm in classifying text data.

4.1.2 Data Cleaning

At the crawling stage of the tweet data obtained,there are still some special characters, URL links, hashes, and others that are not needed and hinder the classification of data later, for that at the data cleaning stage the tweet is cleaned of numbers, special characters, URL links, and etc. So, the data becomes cleaner from unnecessary characters. The results of data cleaning can be seen in Table 2.

Table 2. The result of data cleaning

Tweet	Data Cleaning
b'@rafikurniaputra kalo dr webnya emg dikit, jare kebanyakan daftar lewat perush nya dulu, bru nanti daftar via web kampus merdeka(pas udh muncul)'	kalo dr webnya emg dikit jare kebanyakan daftar lewat perush nya dulu bru nanti daftar via web kampus merdekapas udh muncul
b'mau tanya dong waktu apply magang kampus merdeka, aku kan sedang menjalani smt 5 ya, jadi di KRS nilainya 0 0 0 0 0 terus	mau tanya dong waktu apply magang kampus merdeka aku kan sedang menjalani smt ya jadi di KRS nilainya terus jadinya di web kampus

jadinya di web kampus merdeka jadi segini:(ada yg tau harus lapor kemana? [cm] https://t.co/T5Fn2zogFd'	merdeka jadi segini ada yg tau harus lapor kemana cm
b'-dips! Magang di kampus merdeka itu harus udh punya skill yg dibutuhin company yg dituju ya? Kalau otak kosong gk bisa ya?'	dips Magang di kampus merdeka itu harus udh punya skill yg dibutuhin company yg dituju ya Kalau otak kosong gk bisa ya
b'@sbyfess Bisa ikut programnya kampus merdeka sampe selesai, alhamdulilah masih itu aja \xf0\x9f\x99\x82'	Bisa ikut programnya kampus merdeka sampe selesai alhamdulilah masih itu aja
b'@rafikurniaputra kalo dr webnya emg dikit, jare kebanyakan daftar lewat perush nya dulu, bru nanti daftar via web kampus merdeka(pas udh muncul)'	kalo dr webnya emg dikit jare kebanyakan daftar lewat perush nya dulu bru nanti daftar via web kampus merdekapas udh muncul

4.1.3 Case Folding

In the case folding stage, the tweets have been cleaned in the data cleaning process are then changed from capital letters to lowercase letters, for example the word "UKT" becomes "ukt". The results of case folding can be seen in Table 3.

Table 3. The result of case folding

Text	Case Folding
kalo dr webnya emg dikit jare kebanyakan daftar lewat perush nya dulu bru nanti daftar via web kampus merdekapas udh muncul	kalo dr webnya emg dikit jare kebanyakan daftar lewat perush nya dulu bru nanti daftar via web kampus merdekapas udh muncul
mau tanya dong waktu apply magang kampus merdeka aku kan sedang menjalani smt ya jadi di KRS nilainya terus jadinya di web kampus merdeka jadi segini ada yg tau harus lapor kemana cm	mau tanya dong waktu apply magang kampus merdeka aku kan sedang menjalani smt ya jadi di krs nilainya terus jadinya di web kampus merdeka jadi segini ada yg tau harus lapor kemana cm
dips Magang di kampus merdeka itu harus udh punya skill yg dibutuhin company yg dituju ya Kalau otak kosong gk bisa ya	dips magang di kampus merdeka itu harus udh punya skill yg dibutuhin company yg dituju ya kalau otak kosong gk bisa ya
Bisa ikut programnya kampus merdeka sampe selesai alhamdulilah masih itu aja	bisa ikut programnya kampus merdeka sampe selesai alhamdulilah masih itu aja
apa kubilang program rintisan seperti ini bakal banyak bobroknya which is why fully avoid kampus merdeka	apa kubilang program rintisan seperti ini bakal banyak bobroknya which is why fully avoid kampus merdeka

4.1.4 Tokenization

At the tokenization stage, the tweets have gone through the data cleaning and case folding stages were then broken down into word for word or commonly referred to as tokens. The results of tokenization can be seen in Table 4.

Table 4. The result of tokenization

Text	Tokenization

kalo dr webnya emg dikit jare kebanyakan daftar lewat perush nya dulu bru nanti daftar via web kampus merdekapas udh muncul	['kalo', 'dr', 'webnya', 'emg', 'dikit', 'jare', 'kebanyakan', 'daftar', 'lewat', 'perush', 'nya', 'dulu', 'bru', 'nanti', 'daftar', 'via', 'web', 'kampus', 'merdekapas', 'udh', 'muncul']
mau tanya dong waktu apply magang kampus merdeka aku kan sedang menjalani smt ya jadi di krs nilainya terus jadinya di web kampus merdeka jadi segini ada yg tau harus lapor kemana cm	['mau', 'tanya', 'dong', 'waktu', 'apply', 'magang', 'kampus', 'merdeka', 'aku', 'kan', 'sedang', 'menjalani', 'smt', 'ya', 'jadi', 'di', 'krs', 'nilainya', 'terus', 'jadinya', 'di', 'web', 'kampus', 'merdeka', 'jadi', 'segini', 'ada', 'yg', 'tau', 'harus', 'lapor', 'kemana', 'cm']
dips magang di kampus merdeka itu harus udh punya skill yg dibutuhin company yg dituju ya kalau otak kosong gk bisa ya	['dips', 'magang', 'di', 'kampus', 'merdeka', 'itu', 'harus', 'udh', 'punya', 'skill', 'yg', 'dibutuhin', 'company', 'yg', 'dituju', 'ya', 'kalau', 'otak', 'kosong', 'gk', 'bisa', 'ya']
bisa ikut programnya kampus merdeka sampe selesai alhamdulilah masih itu aja	['bisa', 'ikut', 'programnya', 'kampus', 'merdeka', 'sampe', 'selesai', 'alhamdulilah', 'masih', 'itu', 'aja']
apa kubilang program rintisan seperti ini bakal banyak bobroknya which is why fully avoid kampus merdeka	['apa', 'kubilang', 'program', 'rintisan', 'seperti', 'ini', 'bakal', 'banyak', 'bobroknya', 'which', 'is', 'why', 'fully', 'avoid', 'kampus', 'merdeka']

4.1.5 Normalization

At the normalization stage, tweet has gone through several previous stages, then the mapping of language that does not match the rules becomes standard writing, words that are considered inappropriate or not standard are filtered to get standard words, for example "blum" becomes "belum". " The results of normalization can be seen in Table 5.

Table 5. The result of normalization

Text	Normalization
['kalo', 'dr', 'webnya', 'emg', 'dikit', 'jare', 'kebanyakan', 'daftar', 'lewat', 'perush', 'nya', 'dulu', 'bru', 'nanti', 'daftar', 'via', 'web', 'kampus', 'merdekapas', 'udh', 'muncul']	['kalau', 'dari', 'websitenya', 'memang', 'sedikit', 'katanya', 'kebanyakan', 'daftar', 'lewat', 'perusahaan', 'nya', 'dulu', 'baru', 'nanti', 'daftar', 'via', 'web', 'kampus', 'merdekapas', 'sudah', 'muncul']
['mau', 'tanya', 'dong', 'waktu', 'apply', 'magang', 'kampus', 'merdeka', 'aku', 'kan', 'sedang', 'menjalani', 'smt', 'ya', 'jadi', 'di', 'krs', 'nilainya', 'terus', 'jadinya', 'di', 'web', 'kampus', 'merdeka', 'jadi', 'segini', 'ada', 'yg', 'tau', 'harus', 'lapor', 'kemana', 'cm']	['mau', 'tanya', 'dong', 'waktu', 'apply', 'magang', 'kampus', 'merdeka', 'saya', 'kan', 'sedang', 'menjalani', 'semester', 'ya', 'jadi', 'di', 'krs', 'nilainya', 'terus', 'jadinya', 'di', 'web', 'kampus', 'merdeka', 'jadi', 'segini', 'ada', 'yang', 'tahu', 'harus', 'lapor', 'kemana', 'cuma']
['dips', 'magang', 'di', 'kampus', 'merdeka', 'itu', 'harus', 'udh', 'punya', 'skill', 'yg', 'dibutuhin', 'company', 'yg', 'dituju', 'ya', 'kalau', 'otak', 'kosong', 'gk', 'bisa', 'ya']	['dips', 'magang', 'di', 'kampus', 'merdeka', 'itu', 'harus', 'sudah', 'punya', 'skill', 'yang', 'dibutuhkan', 'company', 'yang', 'dituju', 'ya', 'kalau', 'otak', 'kosong', 'enggak', 'bisa', 'ya']
['bisa', 'ikut', 'programnya', 'kampus', 'merdeka', 'sampe', 'selesai', 'alhamdulilah', 'masih', 'itu', 'aja']	['bisa', 'ikut', 'programnya', 'kampus', 'merdeka', 'sampai', 'selesai', 'alhamdulillah', 'masih', 'itu', 'saja']

['apa', 'kubilang', 'program', 'rintisan', 'seperti', 'ini', 'bakal', 'banyak', 'bobroknya', 'which', 'is', 'why', 'fully', 'avoid', 'kampus', 'merdeka']	['apa', 'kubilang', 'program', 'rintisan', 'seperti', 'ini', 'bakal', 'banyak', 'bobroknya', 'which', 'is', 'why', 'fully', 'avoid', 'kampus', 'merdeka']

4.1.6 Stopword Removal

At the stopword removal stage, tweets that have gone through the previous stages are then filtered to get words that are considered important and can be processed, one of the goals of this stopword removal is to reduce the dimensions of space that look heavy from a document, one example of words that will be deleted in this stopword removal process are “yang”, “dan”, “sampai”, and others. The results of stopword removal can be seen in Table 6.

Table 6. The result of stopword removal

Text	Stopword Removal
['kalau', 'dari', 'websitenya', 'memang', 'sedikit', 'katanya', 'kebanyakan', 'daftar', 'lewat', 'perusahaan', 'nya', 'dulu', 'baru', 'nanti', 'daftar', 'via', 'web', 'kampus', 'merdekapas', 'sudah', 'muncul']	['websitenya', 'kebanyakan', 'daftar', 'perusahaan', 'daftar', 'via', 'web', 'kampus', 'merdekapas', 'muncul']
['mau', 'tanya', 'dong', 'waktu', 'apply', 'magang', 'kampus', 'merdeka', 'saya', 'kan', 'sedang', 'menjalani', 'semester', 'ya', 'jadi', 'di', 'krs', 'nilainya', 'terus', 'jadinya', 'di', 'web', 'kampus', 'merdeka', 'jadi', 'segini', 'ada', 'yang', 'tahu', 'harus', 'lapor', 'kemana', 'cuma']	['apply', 'magang', 'kampus', 'merdeka', 'menjalani', 'semester', 'krs', 'nilainya', 'web', 'kampus', 'merdeka', 'segini', 'lapor', 'kemana']
['dips', 'magang', 'di', 'kampus', 'merdeka', 'itu', 'harus', 'sudah', 'punya', 'skill', 'yang', 'dibutuhkan', 'company', 'yang', 'dituju', 'ya', 'kalau', 'otak', 'kosong', 'enggak', 'bisa', 'ya']	['dips', 'magang', 'kampus', 'merdeka', 'skill', 'dibutuhkan', 'company', 'dituju', 'otak', 'kosong']
['bisa', 'ikut', 'programnya', 'kampus', 'merdeka', 'sampai', 'selesai', 'alhamdulillah', 'masih', 'itu', 'saja']	['programnya', 'kampus', 'merdeka', 'selesai', 'alhamdulillah']
['apa', 'kubilang', 'program', 'rintisan', 'seperti', 'ini', 'bakal', 'banyak', 'bobroknya', 'which', 'is', 'why', 'fully', 'avoid', 'kampus', 'merdeka']	['kubilang', 'program', 'rintisan', 'bobroknya', 'which', 'is', 'why', 'fully', 'avoid', 'kampus', 'merdeka']

4.1.7 Stemming

At the stemming stage, tweet has gone through all the previous stages then the process of mapping and parsing the word form into its basic word form is carried out, this stemming process is the most important process in the preprocessing stage because it can affect whether or not the text mining application is good, an example of this stemming process is the word “menjalani” to “jalan”. The results of stemming can be seen in Table 7.

Table 7. The result of stemming

Text	Stemming
['websitenya', 'kebanyakan', 'daftar', 'perusahaan', 'daftar', 'via', 'web', 'kampus', 'merdekapas', 'muncul']	['websitenya', 'banyak', 'daftar', 'usaha', 'daftar', 'via', 'web', 'kampus', 'merdekapas', 'muncul']

['apply', 'magang', 'kampus', 'merdeka', 'menjalani', 'semester', 'krs', 'nilainya', 'web', 'kampus', 'merdeka', 'segini', 'lapor', 'kemana']	['apply', 'magang', 'kampus', 'merdeka', 'jalan', 'semester', 'krs', 'nilai', 'web', 'kampus', 'merdeka', 'gin', 'lapor', 'mana']
['dips', 'magang', 'kampus', 'merdeka', 'skill', 'dibutuhkan', 'company', 'dituju', 'otak', 'kosong']	['dips', 'magang', 'kampus', 'merdeka', 'skill', 'butuh', 'company', 'tuju', 'otak', 'kosong']
['programnya', 'kampus', 'merdeka', 'selesai', 'alhamdulillah']	['program', 'kampus', 'merdeka', 'selesai', 'alhamdulillah']
['kubilang', 'program', 'rintisan', 'bobroknya', 'which', 'is', 'why', 'fully', 'avoid', 'kampus', 'merdeka']	['bilang', 'program', 'rintis', 'bobrok', 'which', 'is', 'why', 'fully', 'avoid', 'kampus', 'merdeka']

From all the data preprocessing stages, the initial data amounted to 5980 data, then after preprocessing the data, the data obtained totaled 1730 clean data which was ready for the labeling process.

4.1.8 Labeling

At the labeling stage, data labeled using Vader, but the data must be translated into English because Vader is currently unable to support labeling in Indonesian. Vader groups the polarity of text data based on positive, negative, and neutral categories, namely by looking at the processed polarity score results where if the polarity score is more than or equal to a value of 0.05 then the text will be labeled 1 which means positive, then if the polarity is score is less than or equal to -0.05 then the text will be labeled -1 which means negative, then for a neutral label, 0 is between -0.05 and 0.05. The results of the labeling process are shown in Table 8.

Table 8. The result of labeling

Text	Labeling	Sentiment Result
['websitenya', 'banyak', 'daftar', 'usaha', 'daftar', 'via', 'web', 'kampus', 'merdekapas', 'muncul']	0	*Neutral*
['apply', 'magang', 'kampus', 'merdeka', 'jalan', 'semester', 'krs', 'nilai', 'web', 'kampus', 'merdeka', 'gin', 'lapor', 'mana']	1	*Positive*
['dips', 'magang', 'kampus', 'merdeka', 'skill', 'butuh', 'company', 'tuju', 'otak', 'kosong']	-1	*Negative*
['program', 'kampus', 'merdeka', 'selesai', 'alhamdulillah']	1	*Positive*
['bilang', 'program', 'rintis', 'bobrok', 'which', 'is', 'why', 'fully', 'avoid', 'kampus', 'merdeka']	-1	*Negative*

Then the cumulative results of sentiment based on three categories is neutral amounting to 397, positive amounting to 715, and negative amounting to 618.

4.2 Discussion

This research conducted a sentiment analysis of the Kampus Merdeka policy using the SVM algorithm and Naïve Bayes classifier with word vectorization using TF-IDF and labeling using Vader on datasets that had been collected from Twitter. The data collection on this research using the crawling method on Twitter on November 20, 2021 to December 19, 2021 using the keyword "kampus merdeka" and gets 5980 data from crawling results. The data that has been collected is then preprocessed to clean the data from things that are not needed and can be classified using a classification algorithm.

In this research, a comparison of the results of two text classification algorithm models was carried out where the higher the accuracy of an algorithm model, the better the algorithm used. The research was conducted using k-Fold Cross-Validation, namely by dividing the data into two randomly into training data and testing data by weighting words using TF-IDF. The comparison is made with the ratio on the testing data, which is 20% and iterations are carried out 20 times with a total of 1730 data. After preprocessing and labeling the sentiment analysis on the Kampus Merdeka policy on Twitter, then classification using a SVM and Naïve Bayes classifier, the comparison results are obtained as shown in Table 9.

Table 9. The result of classification

Iteration	Support Vector Machine	Naïve Bayes Classifier
1	69.94%	67.05%
2	71.09%	67.63%
3	68.49%	61.27%
4	66.18%	60.40%
5	71.09%	64.45%
6	70.52%	62.13%
7	64.73%	60.11%
8	69.94%	63.00%
9	67.05%	60.98%
10	69.65%	56.64%
11	69.94%	66.18%
12	70.80%	66.76%
13	69.07%	63.87%
14	68.78%	62.13%
15	65.60%	61.56%
16	69.94%	66.47%
17	67.91%	64.73%
18	73.12%	67.92%
19	67.34%	60.11%
20	68.78%	65.60%

The accuracy results show that the SVM and Naïve Bayes classifier algorithms get the highest accuracy in the 18th iteration with comparison results 73.12% for SVM and 67.92% Naïve Bayes classifier. From the test results, overall SVM algorithm is better overall than the Naïve Bayes classifier algorithm in classifying data about 1730 Kampus Merdeka policies.

The results on the SVM algorithm are influenced by the small size of the testing data as the cost parameter used in the study causes the tolerance value for data errors in the SVM calculation to be higher. Meanwhile, the results of the Naïve Bayes classifier algorithm are influenced by feature selection and laplace smoothing which are used in this study to increase the performance of the

Naïve Bayes classifier algorithm and avoid any probability that is 0 which can reduce the performance of the classification results. In addition, the results of word vectorization using TF-IDF in each different iteration of k-Fold Cross Validation affect the importance of a word which makes the accuracy results change in each iteration.

With the level of accuracy generated by the SVM algorithm and Naïve Bayes classifier, it is expected to be able to analyze data sentiment on the independent campus policy well which has never been done before.

5 Conclusion

Based on the results of research that has been carried out on the Kampus Merdeka policy on Twitter, it can be ascertained that the way to analyze the sentiment about the Kampus Merdeka public policy on Twitter with a SVM algorithm and Naïve Bayes classifier is carried out in several stages. The first stage is crawling data on Twitter, then preprocessing the data with stages of cleaning, case folding, tokenization, normalization, stopword removal, and stemming. The next step is to perform the data labeling process using Vader, then word vectorization of the data is carried out with TF-IDF. The last step is to classify using a SVM algorithm and Naïve Bayes classifier.

This research was conducted to test the accuracy of the SVM algorithm and Naïve Bayes classifier in analyzing the sentiment of the Kampus Merdeka policy. The evaluation process was carried out using the k-Fold Cross Validation method with k as many as 20 times and a testing data ratio of 20% getting the highest accuracy results of 73.12% for the SVM algorithm and 67.92% for the Naïve Bayes classifier algorithm. Based on the classification results, it can be concluded that the accuracy of the SVM algorithm is better than the Naïve Bayes classifier algorithm in analyzing the sentiment of the independent campus policy.

References

Al Amrani, Y., Lazaar, M., & El Kadirp, K. E. (2018). Random forest and support vector machine based hybrid approach to sentiment analysis. *Procedia Computer Science*, *127*, 511–520.

Devi, D. V. N., Kumar, C. K., & Prasad, S. (2016). A Feature Based Approach for Sentiment Analysis by Using Support Vector Machine. *Proceedings - 6th International Advanced Computing Conference, IACC 2016*, 3–8.

Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment Analysis of Review Datasets Using Naïve Bayes‘ and K-NN Classifier. *International Journal of Information Engineering and Electronic Business*, *8*(4), 54–62.

Fuadi, T. M., & Aswita, D. (2021). Merdeka Belajar Kampus Merdeka (MBKM): Bagaimana Penerapan dan Kedala yang Dihadapi oleh Perguruan Tinggi Swasta di Aceh. *Jurnal Dedikasi Pendidikan*, *5*(2), 603–614.

Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *4*(2), 113.

Meti, N., Narayan, D. G., & Baligar, V. P. (2017). *Detection of Distributed Denial of Service Attacks using Machine Learning Algorithms in Software Defined Networks*. 1366–1371.

Octaviani, N. L., Hari Rachmawanto, E., Sari, C. A., & Rosal Ignatius Moses Setiadi, D. (2020). Comparison of multinomial naïve bayes classifier, support vector machine, and recurrent neural network to classify email spams. *Proceedings - 2020 International Seminar on Application for Technology of Information and Communication: IT Challenges for Sustainability, Scalability, and Security in the Age of Digital Disruption, ISemantic 2020*, 17–21.

Parikh, K. S., & Shah, T. P. (2016). Support Vector Machine – A Large Margin Classifier to Diagnose Skin Illnesses. *Procedia Technology*, *23*, 369–375.

Parveen, H. (2016). Sentiment Analysis on Twitter Data-set using Naive Bayes Algorithm. *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (ICATccT)*, 416–419.

Rahat, A. M., Kahir, A., & Masum, A. K. M. (2020). Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset. *Proceedings of the 2019 8th International Conference on System Modeling and Advancement in Research Trends, SMART 2019*, 266–270.

Rani, S., & Bhatt, S. (2020). Sentiment Analysis on twitter data using Machine Learning. *Journal of Xidian University*, *14*(12), 1–4.

Ren, R., Wu, D. D., & Wu, D. D. (2019). Forecasting stock market movement direction using sentiment analysis and support vector machine. *IEEE Systems Journal*, *13*(1), 760–770.

Santoso, I., Gata, W., & Paryanti, A. B. (2019). Penggunaan Feature Selection di Algoritma Support Vectore Machine untuk Sentimen Analisis Komisi Pemilihan Umum. *Jurnal Resti (Rekayasa Sistem Dan Teknologi Informasi)*, *1*(10), 5–11.

Song, J., Kim, K. T., Lee, B., Kim, S., & Youn, H. Y. (2017). A novel classification approach based on Naïve Bayes for Twitter sentiment analysis. *KSII Transactions on Internet and Information Systems*, *11*(6), 2996–3011.

Styawati, S., & Mustofa, K. (2019). A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *13*(3), 219.

Wongkar, M., & Angdresey, A. (2019). Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter. *Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019*, 1–5.

Zulfikar, W. B., Irfan, M., Alam, C. N., & Indra, M. (2017). The comparation of text mining with Naive Bayes classifier, nearest neighbor, and decision tree to detect Indonesian swear words on Twitter. *2017 5th International Conference on Cyber and IT Service Management, CITSM 2017*.