Sentiment Analysis of Online Lectures Tweets using Naïve Bayes Classifier

Analisis Sentimen *Tweet* Kuliah *Online* menggunakan Naïve Bayes Classifier

Jacquline M.S. Waworundeng[1]**, Green A. Sandag**[2]**,**
Reynoldus A. Sahulata[3]**, Godlife Davidson Rellely**[4]
[1,2]Informatika; [3,4]Sistem Informasi, Fakultas Ilmu Komputer, Universitas Klabat
Jl. Arnold Mononutu, Airmadidi, Minahasa Utara, Sulawesi Utara
e-mail: [1]**jacquline.morlav@unklab.ac.id**, [2]greensandag@unklab.ac.id ,
[3]rey_sahulata@unklab.ac.id, [4]s11610104@student.unklab.ac.id

Abstract

Online lecture is an alternative learning method during the Covid-19 pandemic. There are opinions with pro and contra of the learning method. The purpose of this study is to evaluate the tweets of opinion or sentiment retrieved from social media Twitter regarding online lectures among the Indonesian community. Twint is used to collect the data tweet and Jupyter notebook is for text preprocessing and classification. The processes started with scraping data from Twitter, text preprocessing, and text classification. Using the Naïve Bayes classifier shows the performance has a precision value of 100%, an accuracy value of 70.8%, an F-measure of 10.2%, and a recall value of 5.4%. Performance rating can be affected by the dataset used for modeling. This analysis covers the positive sentiment and negative sentiments toward online lectures and the result shows 69% negative sentiments and 31% positive sentiments. The negative sentiments had a higher percentage compared to positive sentiments. The results were also supported by the word cloud which expressed a high frequency of negative words such as sleep problems, bored, tired, dizzy, difficult, and lazy. Therefore, it is concluded that during the Covid-19 pandemic from August 1, 2020, to May 31, 2021, Twitter users in Indonesia had negative sentiments about online lectures.

Keywords—Confusion Matrix, Twint, Twitter, text preprocessing, Covid-19 pandemic

Abstrak

Kuliah online menjadi alternatif metode pembelajaran di masa pandemi Covid-19. Ada pendapat yang pro dan kontra tentang metode pembelajaran. Tujuan penelitian ini adalah untuk mengevaluasi kicauan opini atau sentimen dari media sosial Twitter mengenai kuliah online di kalangan masyarakat Indonesia. Twint digunakan untuk mengumpulkan data tweet dan notebook Jupyter untuk preprocessing dan klasifikasi teks. Proses dimulai dengan scraping data dari Twitter, text preprocessing, dan klasifikasi teks. Dengan Naïve Bayes classifier, kinerja memiliki nilai presisi 100%, nilai akurasi 70,8%, F-measure 10,2%, dan nilai recall 5,4%. Peringkat kinerja dapat dipengaruhi oleh kumpulan data yang digunakan untuk pemodelan. Analisis ini mencakup sentimen positif dan sentimen negatif terhadap kuliah online dan hasilnya menunjukkan 69% sentimen negatif dan 31% sentimen positif. Sentimen negatif memiliki persentase yang lebih tinggi dibandingkan dengan sentimen positif. Hasil tersebut didukung oleh kumpulan kata yang mengungkapkan kata negatif dengan frekuensi lebih banyak seperti masalah tidur, bosan, lelah, pusing, sulit dan malas. Sehingga disimpulkan bahwa selama masa pandemi

Covid-19 sejak 1 Agustus 2020 hingga 31 Mei 2021, pengguna Twitter di Indonesia memiliki sentimen negatif terhadap kuliah online.

Kata kunci—Confusion Matrix, Twint, Twitter, text preprocessing, pandemi Covid-19

1 INTRODUCTION

The Covid-19 pandemic initially known originated in Wuhan, China in December 2019. This pandemic has impacted the political, economic, and social life of the people, including the people of Indonesia. This change also has a major impact on the education world, where the rules made by the government to overcome the problem of the spread of the Covid-19 virus that requires the public to carry out health protocols and restricted community activities make students unable to meet face-to-face directly to continue the learning process [1, 2].

This is a very big change in the learning method, where previously students and teachers met in one room and carried out the teaching and learning process offline, but due to the pandemic, both students and teachers can directly make online or virtual meetings anywhere, without having to meet in person [3]. However, not all educational institutions can adapt to new learning methods due to various influencing factors such as limited infrastructure to support virtual teaching and learning processes, the limited financial ability of students to continue taking online classes, teaching staff who do not have an understanding or knowledge in using existing information technology to carry out the teaching and learning process or do not know teaching techniques and online or virtual material development to be delivered to students [4].

There are many responses from the community to this new learning method. Some communities support this change but some people do not [5]. Public opinion develops on social media in response to actual events that are comparable and dynamically changing. Social networks and opinion expression on social media give insights into dynamic interaction between various social actors [6]. Twitter is a social networking and online news where users communicate in short messages called tweets. Tweeting is the practice of sending quick messages to Twitter followers in the hopes of any informative and entertaining [7]. Tweets can be used to collect public responses as analyzable data for research purposes.

Related research [8] using the Naïve Bayes algorithm to interpret data by mining document-based text on Twitter, found that the sentiment analysis of online learning has 30% positive sentiment, 69% negative sentiment, and 1% neutral in early November 2020. The research [9] uses social network analysis with the technique of Twitter crawling by engine *Drone Emprit Academy* (DEA). The DEA system works with the Application Programming Interface (API) platform from Twitter to collect real-time conversations with the streaming method. Data were collected from 1 October 2021 to 1 December 2021. *Drone Emprit* is a big data system that works to capture and monitor conversations on social media and online platforms. While it doesn't mention the percentage of sentiments, it finds a majority of positive sentiments about online learning on 13 October 2021 but on 28 October 2021, shows a spike in negative tweets. Another study [10] got data from tweets during Covid-19 and it analyzes the satisfaction of university students towards online learning on Twitter using Naïve Bayes, K-NN, and Decision Tree. Using K-NN with TF-IDF has the highest accuracy compared to classification algorithms such as Naïve Bayes and Decision Tree (C4.5). Using K-NN shows percentage results of the positive sentiment of 0.16%, negative sentiments of 21.77%, and neutral sentiments of 78.06%. A study [11] collecting Twitter data and using Naïve Bayes, shows 50.75% positive sentiment and 83.3% negative sentiment. Research [12] using the lexicon-based approach with InSet Lexicon as the dictionary of opinion words in Bahasa Indonesia. The results show 63.4% negative tweets, 27.6% positive tweets, and 8.9% neutral tweets. Research [13] analyze the sentiment of "kuliah daring" on Twitter using the Sentistrength algorithm and found 54.3% neutral, 24.7% positive, and 21.1% negative sentiment. Research [14] using the Naïve Bayes method, collected data during the Covid-19 on October 2020, shows 74% negative, 25% positive, and 1% neutral sentiments. The related research described previously showed a variety of positive, negative, and neutral sentiments. Five

research [8, 9, 11, 12, 14] found a higher percentage of negative sentiment, then followed by two research [10, 13] which show neutral sentiments, and only one [9] with positive sentiment.

The related research mentioned previously discussed the method and the results of the sentiment which can be used to compare with this research. The purpose of this study is to collect a sample of responses from Twitter users in Indonesia at a specified time and analyze the sentiment towards online lectures by using the Naïve Bayes. Naïve Bayes Classifier is the simplest and fastest classification algorithm for large data and is applied successfully in various applications, including spam filtering, text classification, sentiment analysis, and recommendation systems [15]. This research used Naïve Bayes as a method because of its simplicity and good performance in classifying text.

2 LITERATURE REVIEW

2.1 Sentiment Analysis

Sentiment analysis is the process of analyzing or processing text to find out or understand someone's emotions which are poured into a sentence and grouped based on positive, neutral, and negative emotions [16].

In the world of education or academia, it is necessary to evaluate and improve the quality of learning and activities held at certain institutions. Evaluation is done by involving other people's responses to get input from outside parties involved with the institution so that it can be known what is needed, what changes must be made, and what steps must be taken. Sentiment analysis becomes an option in evaluating, by collecting responses from other people and extracting these opinions. It can be seen whether other people's views on the institution are positive or negative and it can be concluded what changes need to be made or what needs to be done based on the results of the sentiment analysis conducted [17].

Sentiment analysis is also an analytical technique used in the business world. In its use, business people also need to monitor social media to find out public and consumer opinions on their business. In addition to monitoring social media, the use of sentiment analysis in the business world, namely, conduct brand monitoring, customer feedback, customer service, and market research [18]. All this is done to get the results of the analysis which can be used in decision-making.

Several types of sentiment analysis are distinguished by the focus of analysis or function:

1. *Fine-grained Sentiment Analysis*: Sentiment analysis of this type uses five separate classes starting from one as the lowest which is very negative, the second is negative, the third is neutral, the fourth is positive and the fifth is very positive [19]. This type of analysis is often used in e-commerce to assess or evaluate customer feedback [18].
2. *Emotion Detection*: Today's social media we often encounter is full of feelings, emotions, and opinions expressed by users from all over the world, so this type of sentiment analysis focuses on collecting emotions that are poured or expressed into social media, whether sad, happy, afraid, angry and sad. feelings or other emotions [20].
3. *Aspect-based Sentiment Analysis*: Based on the name, this type of sentiment analysis will find out what aspects are in a document or sentence and then determine the sentiment for that aspect. Aspect-based sentiment analysis is also called feature-based opinion mining and there are several stages to doing this aspect-based sentiment analysis, namely aspect extraction which extracts and evaluate the opinion, and aspect sentiment classification which categorized the sentiment into positive, neutral or negative [21].
4. *Multilingual Sentiment Analysis*: Sentiment analysis of this type is to conduct sentiment analysis in more than one language, such as conducting sentiment analysis on social media in Indonesian and English. The problem in conducting multilingual sentiment analysis is the lack of significant resources, namely the need for a list of words from various languages or languages that will be used in the sentiment analysis [22].

2.2 Social Media

Social media can be used by anyone, anywhere, and anytime. Its users can provide contributions and feedback openly and provide comments and share information about what is happening around them and what is being thought about quickly and unlimitedly [23]. Social media has become a trend in today's society where people around the world cannot be separated from the habit of using technology, such as computers, smartphones, or tablets. These devices become a mediation to open or access social media to interact with other people virtually both those who are close to them and those who are very far from them [24].

There are so many definitions of social media that have been put forward by researchers, influencers, and even experts, here are some definitions of social media according to communication experts. According to Kotler and Keller, social media is used to share messages and digital information with anyone, while according to Lewis, social media allows people to connect and share messages [25]. Based on these definitions, it can be concluded that social media is a medium to connect people virtually so that they can interact and share information without being limited by space, distance, and time [24].

Twitter is a social media type micro-blogging or small blog founded by Jack Dorsey on July 15, 2006. Twitter is a messaging service that has many characteristics in common with the media or communication tools that we have often used such as email, SMS, blogging, etc. However, several factors make Twitter unique and slightly different [26, 27]. In receiving and sending messages on Twitter, users cannot create messages or tweets longer than 140 characters. Tweets created are also public which allows anyone to read them without having permission from other Twitter users, thus allowing users to meet new people [27]. Based on data uploaded by Databoks in October 2019, Twitter has the highest rating as a news and magazine application, which is 13.68 million and has been downloaded 500 million times [28]. Twitter is a social media that is commonly used in Indonesia and it was in the top five positions in 2020 [29].

2.3 Twint

Twint is a very helpful tool when it comes to scraping data on Twitter with is easy to use but can retrieve a lot of tweet data up to the last 3200 tweets for a single scraping session [30]. Twint itself is written or developed with the Python programming language and the installation of the Twint package can be done via Anaconda or pip [31]. In addition to being able to scrape tweets with a data limit of up to 3200 tweets, Twint also has several other benefits and advantages, namely Twint can be used without the Twitter API to scrape tweets, then Twint can be used without having to register Twitter or not having a Twitter account and the initial settings for Twint are fast also without tariff restrictions [30].

2.4 Online Lectures ("Kuliah Online")

The presence of the internet makes many changes that appear in everyday life when interacting. Today's education has also developed and learning interactions have become flexible with the presence of e-learning. One form of e-learning is online lectures. Online lecture is a teaching and learning process that is supported by information technology and the internet as a facilitator in providing lessons and interactions between students and teachers during learning activities [32].

Online lectures are increasingly being practiced in carrying out the learning process during the pandemic. The purpose of this online lecture is to provide a practical teaching and learning process for lecturers and students without being limited by distance so that the learning process can continue [33].

2.5 Text Preprocessing

There are so many languages spoken by the global population that different writing and grammatical rules make writing in a particular language less regular [34]. In conducting sentiment analysis, text data is often withdrawn on social media networks. Writing texts like this is often

found to be done in the informal and irregular language. For that text preprocessing needs to be done so that the text becomes more grammatical and orderly before running the sentiment analysis process on the data that has been drawn [35].

There are several preprocessing texts, namely case folding, cleansing, tokenizing, stopword removal, and stemming. The case folding stage is to equalize the letters in all sentences so that the data is easier to find and makes the process faster. The cleansing stage is cleaning the data that is not needed. Tokenizing is text preprocessing to cut sentences into single words before proceeding to the next process. The next preprocessing text is stopword where in this process words that often appear in a sentence or paragraph but do not affect their meaning will be removed because they are noise. The last is text preprocessing to remove affixes in a word to get the basic word to produce better accuracy [34].

2.6 Library Sastrawi

Library Sastrawi is an algorithm built based on NA (Nazief and Adriani), where this algorithm follows the rules of Indonesian grammar that regulate word affixes. In the grouping based on the placement of affixes, some prefixes are at the beginning of words, suffixes at the end of words, inserted infixes, and confixes which are a combination of both the beginning and the end. This literary library is continuously updated, the first with the CS (Confix Stripping) algorithm where words are changed into their basic form using a basic word dictionary so that the processed word is searched for the basic word in the dictionary [36]. After that, it was updated again with the ECS (Enhanced Confix Stripping) algorithm. This update addresses an issue where some words could not be stemming. Then the last update was made, namely, Modified ECS where improvements were made to overcome over-stemming and under-stemming with the corpus-based method [36].

2.7 Naïve Bayes Classifier

Naïve Bayes is a simple form of probabilistic classifier. Naïve Bayes is often used as a text classifier because of its simplicity and good performance in classifying documents and text [37]. Naïve Bayes is commonly used for its speed, easy implementation, and effectiveness [38]. To calculate Naïve Bayes, the maximum likelihood is used which is the highest similarity to the formula (1).

$$P(c|x) = \frac{P(x|c)\,P(c)}{P(X)}$$

$$P(c|x) = P(x_1|\,c)\,\times\,P(x_2|\,c)\times ... \times\,P(x_n|\,c)\times P(c) \qquad (1)$$

x: Data with unknown class
c: Hypothesis data is a specific class
P(c|x) is the posterior probability of class (target) given predictor
P(c): the prior probability of a class
P(x|c): likelihood which is the probability of predictor given class
P(x): the prior probability of the predictor

2.8 Research Method

The research method is based on the Naïve Bayes classifier. The stages as shown in Figure 1, describe the initial method of analysis, starting with scraping data using Twint. The next step is text preprocessing and continuing with text classification using the Jupyter notebook.
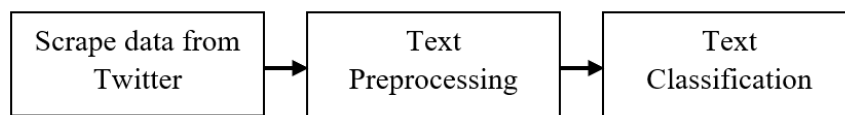
Figure 1 Sentiment analysis stages

The stages are explained as follows.

1. The collection of datasets by scraping data on Twitter using Twint via Google Colaboratory. It is done by taking tweets from the period of August 1, 2020, to May 31, 2021. The scraping process is also carried out by searching for data with the appropriate keywords. The keywords or queries used to retrieve the data are #kuliahonline, #kuliahdaring, and #kulon. After scraping data, it was found that some tweets had duplicated data which caused repetition with the same meaning in the dataset. Irrelevant data and duplicated data obtained from the previous scrape stage had to be removed and cleaned so that the dataset collected is relevant and does not have data loops.
2. Text Preprocessing [39] at the case folding stage is to change the entire text which initially contained capital letters, and changes into lowercase letters so that the data is easier to find. The cleansing [40] stage is carried out to avoid the use of words that are not needed in the dataset such as username, email, URL, hashtag, emoticon, symbol, or RT (retweet). Tokenizing is a preprocessing text that aims to cut a sentence into independent word fragments [41]. The Stopword Removal [39] stage is a preprocessing text that aims to remove words that often appear but are not related to the topic. Stemming [39] is a text preprocessing stage that aims to change a word that is in morphological form or has an affix to be changed to its basic form.
3. Text classification will produce output in the form of data with positive or negative values and determine the sentiments.

Figure 2 shows the flowchart of data analysis starting with scraping data by removing duplicate and irrelevant data, text preprocessing, text classification, evaluation, and ending with the conclusion of sentiment analysis.
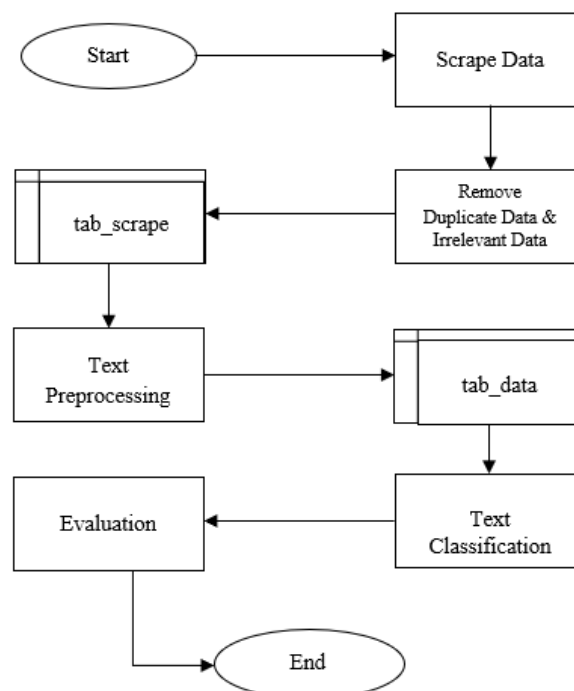

Figure 2 Flowchart of data analysis

3   RESULTS

3.1 Data Collection

Data was collected by scraping data on Twitter social media using Twint with the query or keywords #kuliahonline, #kulon, and #kuliahdaring with 4146 tweet data pulled. Then in the stage of scraped data, it is cleaned for duplicate and irrelevant data. So, the relevant data without repetition and ready to be processed is 597 tweet data.

3.2 Text Preprocessing

The text preprocessing stage is the stage to prepare the data so that it can be processed to the next stage. The data that is usually collected before preprocessing the text is raw data, which is often found noisy and inconsistent [42]. In this research, the preprocessing stage of the text is carried out by case folding, cleansing, tokening, stopword removal, and stemming. After the preprocessing stage is carried out, the data is clean and continues for the sentiment classification stage.

3.3 Word Cloud

Word cloud is a technique or way to present text frequency data with a visual appearance that is more attractive and easier to reach. A visual display containing the words with the most frequently used word frequencies where the larger a word means the word is used more often [43]. The word cloud in this study was created using the TagCrowd web application [44]. The word cloud from tweets as shown in Figure 3 shows the word's frequency of opinion so that it can be seen what words are often used by Indonesian people to communicate and respond to online lectures. The word cloud shows the chosen keywords such as "kuliahonline", "kulon", and "kuliahdaring". Other words highlighted the high frequency of negative words in Bahasa Indonesia such as "tidur" (sleep), "tugas" (assignments), "bosan" (boring), "lelah" (tired), "pusing" (dizzy), "malas" (lazy), "sakit" (sick), "sedih" (sad), "susah" (difficult), etc. Only a few positive words such as "semangat" (spirit) and "pintar" (smart).
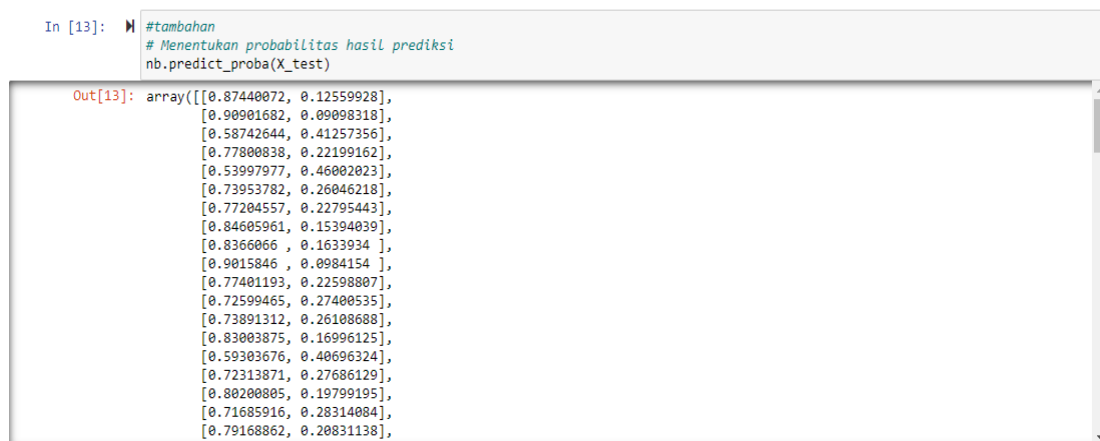

Figure 3 Word cloud of online lectures sentiment

3.4 Data classification

To perform data classification, data splitting is carried out to divide training data and testing data, where training data is to train the classifier model to recognize the data and produce probability models, while testing data is to test sentiment classification. The data ratio for training data and testing data is 80:20. So from 597 tweets, the training data is 477 data and the testing data is 120 data. Classification of data is done by the Naïve Bayes formula which has been discussed in section 2.6.

Figure 4 shows the process classification of testing data to determine the probability of prediction results. It shows some values of testing data but takes only the top five data, classify as data 1, data 2, data 3, data 4, and data 5 as mentioned in Table 1 and Table 2.

```
In [13]: ▶  #tambahan
             # Menentukan probabilitas hasil prediksi
             nb.predict_proba(X_test)

Out[13]: array([[0.87440072, 0.12559928],
                [0.90901682, 0.09098318],
                [0.58742644, 0.41257356],
                [0.77800838, 0.22199162],
                [0.53997977, 0.46002023],
                [0.73953782, 0.26046218],
                [0.77204557, 0.22795443],
                [0.84605961, 0.15394039],
                [0.8366066 , 0.1633934 ],
                [0.9015846 , 0.0984154 ],
                [0.77401193, 0.22598807],
                [0.72599465, 0.27400535],
                [0.73891312, 0.26108688],
                [0.83003875, 0.16996125],
                [0.59303676, 0.40696324],
                [0.72313871, 0.27686129],
                [0.80200805, 0.19799195],
                [0.71685916, 0.28314084],
                [0.79168862, 0.20831138],
```

Figure 4 Process classification of data testing

Table 1 describes the probability values of the top five testing data. Data 1 is the probability of the first testing data and so on until data 5 is the probability of the fifth testing data. The probabilities values of data 1 up to data 5 are in decimals and can be converted to percentage values.

Table 1 Data probability

Probability Data	Negative	Positive
Data 1	0.87440072	0.12559928
Data 2	0.90901682	0.09098318
Data 3	0.58742644	0.41257356
Data 4	0.77800838	0.22199162
Data 5	0.53997977	0.46002023

Data 1 to data 5 show probabilities of data testing successively. Data 1 has a probability of 0.87440072, around 87.4%. Data 2 has a probability of 0.90901682, around 90.9%. Data 3 has a probability of 0.58742644, around 58.7%. Data 4 with 0.77800838, around 77.8%. Data 5 has 0.53997977, around 53.9%. Data with a higher probability value will determine the sentiment classification of whether the data is negative or positive. A higher probability value means the

data is dominant. For example, in Table 1, data 1 shows the probability of 0.87440072 (87.4%), which means the data is classified as negative sentiments.

Table 2 Data classification

Data	Sentiments
Data 1	Negative
Data 2	Negative
Data 3	Negative
Data 4	Negative
Data 5	Negative

Table 2 is the result of testing data classification. The results of the distribution of training data and testing data, where the testing data obtained are 120 tweets. Figure 5 shows that, in 120 tweets as data testing, 83 tweets had negative sentiments and 37 tweets had positive sentiments, with 69% and 31% respectively. This research found that from August 1, 2020, to May 31, 2021, the sentiment toward online lectures has 69% negative sentiment and 31% positive sentiment. It can be related to the word cloud in Figure 3 which shows the frequency of negative words.

As already described earlier in the introduction section that there are several related research with findings about online lecture sentiment. Compare to other related research, this research shows that negative sentiments that are more dominant than positive sentiments.
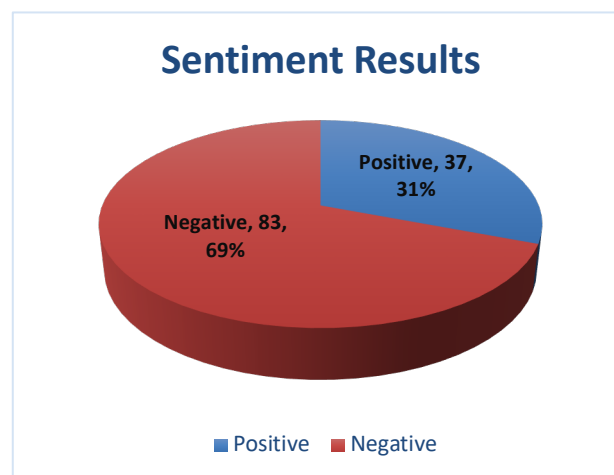
Sentiment Results

Positive, 37, 31%

Negative, 83, 69%

■ Positive ■ Negative

Figure 5 Sentiment results of online lecture tweets

3.5 Evaluation

The evaluation was carried out with the aim of testing and measuring the results of the classification using a confusion matrix table. This classification test is carried out by calculating precision, recall, f-measure, and accuracy so that the ability of this classification and its accuracy can be measured.

Table 3 shows true negative (TN) or negative predictions data with negative facts are 83 tweets. False positive or positive prediction data with negative facts are 0 tweets. False negative (FN) or negative prediction data with positive facts are 35 tweets. True positive (TP) or positive prediction data with positive facts are 2 tweets. Then the percentage value of precision, recall, f-measure, and accuracy will be sought.

Table 3 Confusion Matrix

True	Prediction		Total
	Negative	Positive	
Negative	TN: 83	FP: 0	83
Positive	FN: 35	TP: 2	37
Total	118	2	120

The accuracy between the desired data and the model's predicted outcomes is referred to as precision. Recall (Sensitivity or True Positive Rate) measures how well the model can retrieve data or how frequently the model correctly predicts positive outcomes when the actual class is positive. F-measure (F1-Score) is the harmonic average of Precision and Recall. Accuracy is the overall total of how accurate the model is in classifying correctly.

In table 4 the calculation results from the formula, it is found that the precision value is 100%, the recall value is 5.4%, and the F-measure is 10.2%. The accuracy value is 70.8%.

Table 4 Precision, Recall, F-measure, and Accuracy

Formula	Results	%
$Precission = \dfrac{TP}{TP + FP}$	$\dfrac{2}{2 + 0} = 1$	100%
$Recall = \dfrac{TP}{(TP + FN)}$	$\dfrac{2}{2 + 35} = 0.054$	5.4%
$F - Measure = 2 * \dfrac{(P * R)}{(P + R)}$	$2 * \dfrac{(1 * 0{,}054)}{(1 + 0.054)} = 0.102$	10.2%
$Accuracy = \dfrac{(TP + TN)}{(TP + FP + FN + TN)}$	$\dfrac{(2 + 83)}{(2 + 0 + 35 + 83)}$	70.8%

4 CONCLUSION

The research has several conclusions. The text preprocessing stage greatly affected the results of the Naïve Bayes text classification. With 120 testing data, it shows a precision value of 100% which means the level of precision of the requested data is high as the results of the Naïve Bayes classification. The accuracy describes how accurate the model classifying correctly. The results of accuracy on sentiment analysis with Naïve Bayes are quite good with a value of 70.8%.

From 557 tweets, there were 393 tweets with negative sentiments and 204 tweets with positive sentiments. Taking 120 tweets for training data shows 83 tweets with negative sentiments or 69% and 37 tweets with positive sentiments or 31%. The visual word cloud shows more negative responses related to sleeping disorders, bored, tiredness, and laziness with online lectures. Only a few respondents with a positive opinion. Comparing the results with other related research conducted in Indonesia shows the same trend of negative sentiment toward online lectures.

5   FUTURE RESEARCH

Future research related to the improvement of preprocessing stage so that the data obtained from Twitter is better, such as adding normalization of words from slang or terms in Indonesian to standard Indonesian because it was found that some Twitter users used slang, terms, and also abbreviations when making tweets. In the Twitter scrape process, it is suggested to only take relevant tweets by avoiding queries or keywords and Twitter usernames, to avoid irrelevant tweet data containing conversations from users. Another suggestion is adding data so that the results obtained are more accurate by extending the period of scraping data or adding other keywords related to online lectures. The research found that many responses regarding online lectures were expressed by emojis, emoticons, symbols, stickers, pictures, and videos. Hence future research is to provide a mechanism for converting those categories of responses into words or sentences of sentiment because it can collect more responses or public opinions submitted on Twitter.

REFERENCES

[1]   Setyorini, I., 2020, Pandemi COVID-19 dan online learning: apakah berpengaruh terhadap proses pembelajaran pada kurikulum 13?), *Journal of Industrial Engineering & Management Research*, vol. 1, no. 1, pp. 95. [Online]. Available: https://tinyurl.com/3v4kmuku

[2]   Anhusadar, L., 2020, "PIAUD student perceptions of online lectures in Covid 19 pandemic period (Persepsi Mahasiswa PIAUD terhadap Kuliah Online di Masa Pandemi Covid 19)," *Kindergarten Journal of Islamic Early Childhood Education*, vol. 3, no. 1, pp. 45. [Online]. Available: https://tinyurl.com/yc5ky87x

[3]   Kusnayat, A., Muiz, M.H., Sumarni, N., Mansyur A.S., and Zaqiah, Q.Y., 2020, Pengaruh teknologi pembelajaran kuliah online di era Covid-19 dan dampaknya terhadap mental mahasiswa, *EduTeach*, vol. 1, no. 2, pp. 153–165. [Online]. Available: https://tinyurl.com/35dbbcsk

[4]   Mulawarman, W.G., 2020, Persoalan dosen dan mahasiswa masa pandemik Covid 19: dari gagap teknologi hingga mengeluh boros paket data), *Prosiding Seminar Nasional Hardiknas*, vol. 1, pp. 37–46. [Online]. Available: https://tinyurl.com/2p8pcjeh

[5]   Sulis, M., 2020, Resistance to change mahasiswa psikologi universitas Esa Unggul yang mengikuti perkuliahan online, *UEU Journal*, vol. 18, no. 02, pp. 87. [Online]. Available: https://tinyurl.com/5687bz72

[6]   Zhang, Y., Chen. F., and Rohe, K., 2021, Social media public opinion as flock in a murmuration: conceptualizing and measuring opinion expression on social media, *Journal of Computer-Mediated Communication,* vol. 27, issue 1, pp. 1-22. [Online]. Available: https://academic.oup.com/jcmc/article/27/1/zmab021/6472790

[7]   Gil, P., 2021, What is Twitter & how does it work?, Lifewire. [Online]. Available: https://www.lifewire.com/what-exactly-is-twitter-2483331.

[8]   Samsir., Ambiyar., Verawardina, U., Edi, F., and Watriantros, R., 2021, Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes, *Jurnal Media Informatika Budidarma*, vol. 5, no. 1, pp 157-163. [Online]. Available: http://stmik-budidarma.ac.id/ejurnal/index.php/mib/article/view/2580

[9]   Suharsa, H., Soleh, D.J., and Miftahuddin, A., 2022, Persepsi Publik tentang Pembelajaran Daring dari Jejak Digital Twitter: Analisis Sentiment Positif, Netral, dan Negativ dari Drone Imprit, *Jurnal Aparatur Kementrian Energi dan Sumber Daya Mineral,* vol. 6, no. 1, pp. 33-42. [Online]. Available: https://tinyurl.com/w44ys72v

[10]   Razaq, E.R.M., Jacob, D.W., and Hamami, F., 2021, Analisis Sentimen Kepuasan Mahasiswa terhadap Pembelajaran Online selama Pandemi Covid-19 pada Media Sosial Twitter menggunakan Perbandingan Algoritma Klasifikasi, *e-Proceeding of Engineering¸* vol. 8, no. 5. [Online]. Available: https://tinyurl.com/34swtsvf

[11]   Nursyi'ah, S.Y., Erfina, A., and Warman, C., 2021, Analisis Sentimen Pembelajaran Daring pada Masa pandemi Covid-19 di Twitter menggunakan Naïve Bayes, *Sismatik (Seminar Nasional Sistem Informasi dan Manajemen Informatika)*, vol. 1, no.1. pp 117-123.   [Online].   Available: https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/16

[12]   Musfiroh, D., Khaira, U., Utomo, P.E.P., and Suratno, T., 2021, Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset menggunakan InSet Lexicon, MALCOM: Indonesian Journal of Machine Learning and Computer Science, vol. 1, no.1, pp. 24-33. [Online]. Available: https://journal.irpi.or.id/index.php/malcom/article/view/20

[13]   Hardian, R.Z., Prasetyo, P.E., Khaira, U., and Suratno, T., 2021, Analisis Sentiment Kuliah Daring di Media Sosial Twitter selama Pandemi Covid-19 menggunakan Algoritma Sentistrength, *MALCOM: Indonesian Journal of Machine Learning and Computer Science* vol.   1,   issue.   2   pp.   138-143.   [Online].   Available: https://journal.irpi.or.id/index.php/malcom/article/view/15

[14]   Sahir, S.H., Ramadhana, R.S.A., Marpaung, M.F.R., Munthe, S.R., Watrianthos, R., 2021, Online Learning Sentiment Analysis During the Covid-19 Indonesia Pandemic Using Twitter Data, I*OP   Conf.   Ser.:   Mater.   Sci.   Eng*, vol 1156. [Online]. Available: https://iopscience.iop.org/article/10.1088/1757-899X/1156/1/012011/meta

[15]   Dhaduk, H., 2021, Performing Sentiment Analysis With Naive Bayes Classifier!, Data Science Blogathon. [Online]. Available: https://tinyurl.com/243d8nrh

[16]   Sari, F. V., and Wibowo, A., 2019, Analisis sentimen pelanggan toko online JD.ID menggunakan metode naïve bayes classifier berbasis konversi ikon emosi, Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer, vol. 10, no. 2, pp. 681-686. [Online]. Available: https://jurnal.umk.ac.id/index.php/simet/article/view/3487

[17]   Indurkhya, N., and Damerau, F. J., 2010, Handbook of Natural Language Processing. CRC Press. [E-book] Available: Google Books.

[18]   Arviana, G.N., 2020, Sentiment Analysis: Pengertian, Teknik, dan Penggunaannya, Glints Blog. [Online]. Available: https://glints.com/id/lowongan/sentiment-analysis/ [Accessed Mar 16, 2021].

[19]   Rao, P., 2019, Fine-grained Sentiment Analysis in Python (Part 1), Medium. [Online]. Available: https://towardsdatascience.com/fine-grained-sentiment-analysis-in-python-part-1-2697bb111ed4 [Accessed Mar 16, 2021].

[20]   Gaind, B., Syal, V., and Padgalwar, S., 2019, Emotion Detection and Analysis on Social Media. [Online]. Available: http://arxiv.org/abs/1901.08458 . [Accessed Mar. 16, 2021]

[21]   Aldhi, M. D., 2016, Aspect-Based Sentiment Analysis terhadap Ulasan Produk menggunakan Metode Klasifikasi Naïve Bayes, Universitas Telkom.

[22]   Damayanti, W., 2015, Analisis penggunaan multilingual anak tingkat sekolah dasar di lingkungan gang siti mardiah cibaduyut bandung, studi sosiolinguistik), *Jurnal Gramatika*, vol. 1, no. 1, pp.100-110, Apr. 2015. [Online]. Available: https://tinyurl.com/2p94sazs

[23]   Putri, W. S. R., Nurwati, N., and Budiarti S. M., 2016, Pengaruh media sosial terhadap perilaku remaja), *Prosiding KS*, vol. 3, no. 1, pp.47-51. [Online]. Available: https://jurnal.unpad.ac.id/prosiding/article/view/13625

[24]   Oktaviani, D., 2019., Pengaruh Media Sosial Terhadap Gaya Hidup Mahasiswa IAIN Metro," Thesis (undergraduate), IAIN Metro, 2019. [Online]. Available: https://repository.metrouniv.ac.id/id/eprint/1212/

[25]   Ambar, 2017, 20 Pengertian Media Sosial Menurut Para Ahli, Jun. 08, 2017. [Online]. Available: https://tinyurl.com/5n8y6ty8. [Accessed Mar. 18, 2021].

[26]   Sembodo, J. E., Setiawan, E. B., and Baizal, Z. A., 2016, Data Crawling Otomatis pada Twitter, in INDOSC 2016, pp. 11–16.

[27]   O'Reilly, T., and Milstein, S., 2011, The Twitter Book. O'Reilly Media, Inc. [E-book] Available: Google Books.

[28]   Kusnandar, V. B., 2019, Twitter, Aplikasi Berita dan Majalah Berbasis Android dengan Rating Tertinggi), Databoks. [Online]. Available: https://tinyurl.com/mrxa3cj6 [Accessed Mar. 14, 2021].

[29] Jayani, D. H., 2020, Databoks. [Online]. Available: https://tinyurl.com/t5d8v72w [Accessed Mar. 14, 2021].

[30]   Rahutomo, F., Saputra, P. Y., and Fidyawan, M. A., 2018, Implementasi Twitter sentiment analysis untuk review film menggunakan algoritma support vector machine)," *JIP*, vol. 4, no. 2, pp. 93, Feb. 2018.

[31]   Monarizqa, N., Nugroho, L. E., and Hantono, B. S., 2014, Penerapan analisis sentimen pada Twitter berbahasa indonesia sebagai pemberi rating, *UGM*, vol.1, no.4, pp. 5, Oct. 2014. [Online]. Available: httphttps://tinyurl.com/3b3azywf

[32]   Saputro, F. B., Somantri, M., and Nugroho, A., 2017, Pengembangan sistem kuliah online Universitas Diponegoro untuk antar muka mahasiswa pada perangkat bergerak berbasis Android," *Transmisi*, vol. 19, no. 1, pp. 1-2, Jul. 2017. [Online]. Available: https://tinyurl.com/drabs2d7

[33]   Erin, E., and Maharani, A, 2018, Persepsi Mahasiswa Pendidikan Matematika terhadap Perkuliahan Online)," *Mosharafa*, vol. 7, no. 3, pp. 337–344.

[34]   Palmer, D.D., 2010, Text Preprocessing. Handbook of natural language processing 2. [E-book] Available: Google Scholar.

[35] Dashtipour K., et al., 2016, Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques, *Cognitive Computation*, vol. 8, no. 7, pp. 757-771.

[36] Fatra, A.H.D., Hayatin, N. H., and Aditya, C.S.K., 2020, Analisa Sentimen Tweet Berbahasa Indonesia Dengan Menggunakan Metode Lexicon Pada Topik Perpindahan Ibu Kota Indonesia, *Jurnal Repositor*, vol. 2, no.5, pp. 977-984.

[37] Chakrabarti, S., Roy, S., and Soundalgekar, M. V., 2003, Fast and accurate text classification via multiple linear discriminant projections, *VLDB*, vol. 12, no. 2, pp. 170–185, Aug. 2003.

[38] Domingos, P., and Pazzani, M., 1996, Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier, *Proc. 13th Intl. Conf. Machine Learning*, pp. 105-112

[39] Susilowati, E., Sabariah, M. K., and Gozali, A. A., 2015, Implementasi metode Support Vector Machine untuk melakukan klasifikasi kemacetan lalu lintas pada Twitter, *e-Proceeding of Engineering*, vol. 2, no. 1, pp. 1-7.

[40] Olhang, M., Achmadi, S., and Wibisono, F. X. A., 2020, Analisis sentimen pengguna Twitter terhadap Covid-19 di Indonesia menggunakan metode Naïve Bayes classifier (nbc), *Jati*, vol. 4, no. 2, pp. 214–221, Sep. 2020. [Online]. Available: https://ejournal.itn.ac.id/index.php/jati/article/view/2695

[41] Najjichah, H., Syukur, A., and Subagyo, H., 2019, Pengaruh text preprocessing dan kombinasinya pada peringkas dokumen otomatis teks berbahasa Indonesia, *Jurnal Teknologi Informasi, Jurnal Teknologi Informasi Cyberku*, vol. 15, p. 1-11. [Online]. Available: http://research.pps.dinus.ac.id/index.php/Cyberku/article/view/69

[42] Nurvinda, G., 2021, Pentingnya Preprocessing dalam Pengolahan Data Statistik, DQLab. [Online]. Available: https://tinyurl.com/32nkdfp9 [Accessed Juni 22, 2021].

[43] Januarsjaf, A., 2020, Wordcloud, RPubs. [Online]. Available: https://tinyurl.com/4sbnzexj . [Accessed Juni 30, 2021].

[44] Steinbock, D., Create your own word cloud from any text to visualize word frequency, TagCrowd. [Online]. Available: https://tagcrowd.com/ . [Accessed Juni 30, 2021].