# COMPARISON OF THE ACCURACY OF THE LEXICON-BASED AND NAIVE BAYES CLASSIFIER METHODS TO PUBLIC OPINIONS ABOUT REMOVING MASKS ON SOCIAL MEDIA TWITTER

**ANGGA ADITYA PERMANA[1*], WAHYU ALDHI NOVIYANTO[2] ,**

[1*]Department of Informatic, University of Multimedia Nusantara, Indonesia

[2]Department of Informatic Engineering, University of Muhammadiyah Tangerang, Indonesia

E-mail:  [1*]angga.permana@umn.ac.id,[2]aldinovianto@ft-umt.ac.id

## ABSTRACT

At the beginning of 2020 the world was shocked by the COVID-19 pandemic which paralyzed all aspects of activity for some time. However, over time and with the discovery of a vaccine, the cases caused by COVID-19 began to subside. In 2022, the Indonesian government make a policy that people are allowed to take off their masks when active but are encouraged to maintain health protocols. However, the approach reaped the pros and cons of the Indonesian people. One challenge is to build technology to detect and summarize an overall those pros and cons. So that, we look at Twitter and build models for classifying 'tweets' into positive, negative and neutral sentiment using top two approaches for sentiment analysis, the lexicon-based method and the naive Bayes classifier. This study aimed to analyze public opinion about removing masks through Twitter by comparing the lexicon-based method and the naive Bayes classifier method to find out how the community responded to taking off masks. A total of 639 tweets with the keyword "Lepas Masker" was analyzed include data crawling, text preprocessing, feature extractions and the classification process. The comparison of the results obtained shows the accuracy of 82% for the lexicon-based method and 70% for the naive Bayes classifier method. To the results, the accuracy value of the lexicon-based method is higher than the naive Bayes classifier method.

Keywords: *Analysis Sentiment, Covid-19, Removing Mask, Naïve Bayes, Lexicon.*

## 1. INTRODUCTION

On March 11, 2020, the World Health Organization (WHO), announced that COVID-19 is a global pandemic [1]. WHO reports that there are 52 million Indonesians who have been confirmed positive for COVID-19. A total of 463,000 people have been confirmed positive, and 15,148 people have died in Indonesia. However, over time marked by the discovery of a COVID-19 vaccine, the cases of COVID-19 have gradually decreased. The spokesperson for the COVID-19 Vaccination Ministry of Health stated that the number of cases of COVID-19 has decreased consistently since the end of February 2022. This gives optimism to the efforts to handle COVID-19 currently being carried out by the Indonesian government[2].

With the decline in the number of COVID-19 cases and the improvement in the pandemic situation, the government has eased community activities. Reporting from news.detik.com stated that the government issued a new policy, namely, to provide leeway on the use of masks during the COVID-19 period, the procedure was effective from Wednesday, May 18, 2022. President Jokowi stated that if people are active outdoors or in areas open areas that are not crowded with people, it is permissible not to use masks. If you are indoors and using public transportation, masks are still mandatory to use. For people classified as vulnerable, elderly, or comorbid categories, it is advisable to continue using masks. Likewise, people who experience symptoms of coughs and colds still have to wear masks when doing activities. The easing was carried out considering the development of the COVID-19 situation around the world. According to the Minister of Health of the Republic of Indonesia, based on his observations, the public already has a fairly good resistance to the new variants that are currently circulating, as evidenced by Indonesia's cases which tend to decline and are relatively smaller for the same variant compared to other countries such as China. Taiwan, and the United States. However, this easing must be

carried out with full responsibility considering that COVID-19 worldwide has not been completely destroyed.[2]

Indonesia is one of the countries with the most significant internet usage population in the world. The internet is the most convenient means for people to meet their needs in searching for information.
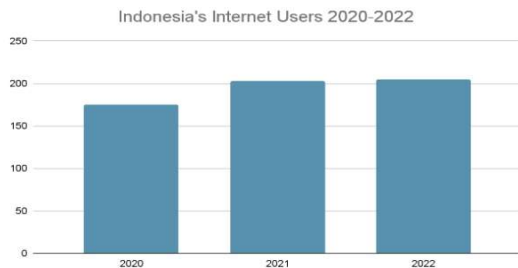


*Figure 1: Internet User Data in Indonesia 2020-2022*

Based on the graph above with data obtained through the website www.hootsuite.com, in 2020 internet users in Indonesia reached 175.4 million internet access. In 2021, there will be a drastic increase compared to 2020 as many as 202.6 million internet access. In 2022, there will be an increase of 204.7 million internet users. This shows that the population growth of internet users has increased every year in the last 3 years. Twitter is one of the social media platforms with the most users, ranked sixth in 2022 [3] [4] with a total user percentage of 58%. One of the features provided by Twitter that users generally use is the tweet feature. In this feature, users can express their ideas through facts, suggestions, and criticism of something. These tweets can be processed into useful information. Data obtained through the Twitter social media platform can be used as a data source for opinion extraction or what is known as Sentiment Analysis [5] [6][7].

According to [8][9] each individual has diverse opinions and opinions. This opinion is critical and can influence human behaviour. Sentiment analysis is a process or method used to classify documents or texts based on the opinions and views of specific communities carried out by the process of understanding, processing and extracting data that are text documents automatically to find out whether an idea or opinion is positive, negative, negative, or is normal to obtain sentimental information about a phenomenon or event being studied [10][11].

This study will discuss the sentiment of public analysis of government policies on leniency in using masks on Twitter social media. Based on the events that are currently busy in the community, many social media users have given their opinions and thoughts on the new policy regarding removing masks on the Twitter social media platform. This is interesting to study to find a public opinion about the current mask removal policy. To support research on this, an algorithm is needed to classify public comments on Twitter social media, both positive and negative comments. Sentiment analysis used in this study is the Lexicon Based method [6] [12][13] and the Naïve Bayes Classifier (NBC) [14] [15] [16] method to compare the performance of the two methods. Vader sentiment is applied to the Lexicon Based analysis method.

Naïve Bayes Classifier (NBC) is an algorithm that uses classification methods in sentiment analysis with large amounts of data and produces fast and precise accuracy.

## 2. RESEARCH METHODS

The research method has a guide in the form of a research flow that serves as a direction so that the research runs as expected. The plot design in this research is as follows:



*Figure 2: Research Stages*

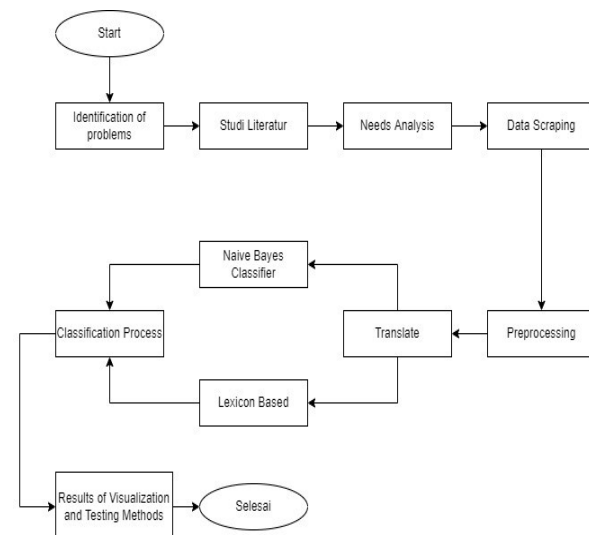### 2.1 Data Scrapping

The data used in this study is commentary data obtained through the Twitter site. The data collected is text data obtained through a crawling technique using the Twitter API as many as 1,000 tweet data with data taken on April 26, 2022, for 639 tweets with the keyword "Lepas Masker", which has been provided on the website https://developer.twitter.com/by the provisions of Twitter.
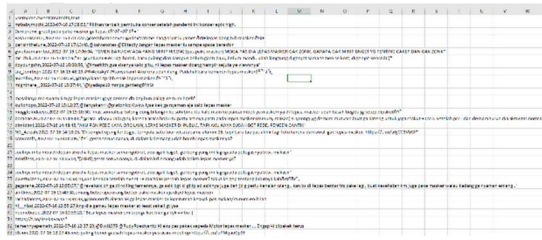
*Figure 3: Data Crawl Results*

The results of this crawl are stored in a CSV file which will later be preprocessed.

## 2.2 Preprocessing

The preprocessing stage is critical because at this stage, the removal of words or symbols that are not needed or words that have no meaning will be carried out. Preprocessing also aims to process raw data into quality data so it can be continued to the following process. The order of the preprocessing stages is as follows:

### A. Cleansing

This stage is carried out to remove punctuation marks and normalise Unicode on comment data. When carrying out the cleaning process, several steps are carried out to obtain maximum results, including:

a. Remove punctuation
b. Deleting numbers
c. Deleting Retweets
d. Remove excess spaces

The following is the implementation of the 4 stages of data cleaning, as shown in Figure 4.



*Figure 4: Data Cleansing Results*

*Table 1: Data Cleansing Detail*

| Before | After |
|---|---|
| @kembarannever @idextratime nangis liat lu jamet 🤣 lepas dong tuh masker🤣 | nangis liat lu jamet lepas dong tuh masker |
| @kaivanotes @Elitedly Jangan lepas masker lu sampe space berakhir | Jangan lepas masker lu sampe space berakhir |
| mas sonnalisa, tolong dong bilangin ke sohibmu itu kalo main ke pantai mbok ya maskernya di lepas. masker udah basah | mas sonnalisa, tolong dong bilangin ke sohibmu itu kalo main ke pantai mbok ya maskernya di lepas. |
| begitu jg tetep dipakeðŸ¥" | masker udah basah begitu jg tetep dipake |
| gesss serius nanya, di dalam krl emang udah boleh lepas masker ya? | gesss serius nanya, di dalam krl emang udah boleh lepas masker ya |

### B. Tokenizing

In the tokenization stage, the NLTK library in this study is used to tokenize tweets. This tokenization is used to reshuffle sentences from tweets into pieces which will later become words. The following is the implementation of the Tokenizing data stage, as shown in Figure 5.

`tweet_df.head(10)` *#tokenization*

| | username | tweetcreatedts | text | clean_tweet | remove_http | remove_hastag | Tweet |
|---|---|---|---|---|---|---|---|
| 0 | mrbabymochi | 2022-07-16 17:38:53 | Pilihan terbaik pembuka konser setelah pandemi... | Pilihan terbaik pembuka konser setelah pandemi... | Pilihan terbaik pembuka konser setelah pandemi... | Pilihan terbaik pembuka konser setelah pandemi... | [pilih, baik, buka, konser, pandemi, konser, e... |
| 1 | Fasa77342812 | 2022-07-16 17:17:42 | @kembarannever @idextratime nangis liat lu jam... | nangis liat lu jamet 😂 lepas dong tuh masker | nangis liat lu jamet lepas dong tuh masker | nangis liat lu jamet lepas dong tuh masker | [nang, liat, jamet, lepas, masker] |
| 2 | parisinthelune | 2022-07-16 17:13:48 | @kaivanotes @Elitedly Jangan lepas masker lu s... | Jangan lepas masker lu sampe space berakhir | Jangan lepas masker lu sampe space berakhir | Jangan lepas masker lu sampe space berakhir | [lepas, masker, sampe, space] |
| 3 | gwultramanribut | 2022-07-16 17:06:04 | TEMEN BARU GW ADA YANG MIRIP HISENG (pas pake ... | TEMEN BARU GW ADA YANG MIRIP HISENG (pas pake ... | TEMEN BARU GW ADA YANG MIRIP HISENG pas pake m... | TEMEN BARU GW ADA YANG MIRIP HISENG pas pake m... | [temen, hiseng, pas, pake, masker, pas, lepas,... |
| 4 | mr_zhik | 2022-07-16 17:01:59 | RT @unkwn6969: Lagi hamil, baru pulang dari ka... | : Lagi hamil, baru pulang dari kampus belum ga... | Lagi hamil baru pulang dari kampus belum ganti... | Lagi hamil baru pulang dari kampus belum ganti... | [hamil, pulang, kampus, ganti, baju, mandi, ud... |
| 5 | doyoungshn | 2022-07-16 16:08:56 | @lmarkith gue sisanya kalo gitu, ni lepas mask... | gue sisanya kalo gitu, ni lepas masker doang ... | gue sisanya kalo gitu ni lepas masker doang ha... | gue sisanya kalo gitu ni lepas masker doang ha... | [sisa, gitu, ni, lepas, masker, doang, juta, v... |
| 6 | Lia_Jenrbyjn | 2022-07-16 15:48:19 | @N4diaAyY @tanyakanrl Aku kira udah ilang. Pad... | Aku kira udah ilang. Padahal baru kemaren lep... | Aku kira udah ilang Padahal baru kemaren lepas... | Aku kira udah ilang Padahal baru kemaren lepas... | [udah, ilang, kemaren, lepas, masker] |
| 7 | awrelli3 | 2022-07-16 15:34:54 | @tanyakanrl tp lbh enak lepas masker 😊👌 | tp lbh enak lepas masker 😊👌 | tp lbh enak lepas masker | tp lbh enak lepas masker | [lepas, masker] |
| 8 | mightnare_ | 2022-07-16 15:27:44 | @lyadapa10 mmps ganteng 🤤\n\nmasalanya mo swami... | mmps ganteng😂\n\nmasalanya mo swami... | mmps ganteng masalanya mo swami km ga lepas ma... | mmps ganteng masalanya mo swami km ga lepas ma... | [mmps, ganteng, masala, mo, swami, lepas, mask... |
| 9 | kuilohopo | 2022-07-16 15:12:27 | @tanyakanrl @platonicc Kwkw iyaa kek ga nyaman... | Kwkw iyaa kek ga nyaman aja kalo lepas masker | Kwkw iyaa kek ga nyaman aja kalo lepas masker | Kwkw iyaa kek ga nyaman aja kalo lepas masker | [kwkw, iyaa, nyaman, aja, lepas,... |

*Figure 5: Data Tokenizing Results*

*Table 2: Data Tokenizing Detail*

| Before | After |
|---|---|
| nangis liat lu jamet lepas dong tuh masker | ['nangis', 'liat', 'lu', 'jamet', 'lepas', 'dong', 'tuh', 'masker'] |
| Jangan lepas masker lu sampe space berakhir | ['Jangan', 'lepas', 'masker', 'lu', 'sampe', 'space', 'berakhir'] |
| sonnalisa tolong bilangin sohib main pantai mbok masker lepas masker udah basah tetep dipake | ['sonnalisa', 'tolong', 'bilangin', 'sohib', 'main', 'pantai', 'mbok', 'masker', 'lepas', 'masker', 'udah', 'basah', 'tetep', 'dipake'] |
| gesss serius nanya krl emang udah lepas masker lepas masker ngobrol kaget gerbong gaada tugas makasii | ['gesss', 'serius', 'nanya', 'krl', 'emang', 'udah', 'lepas', 'masker', 'lepas', 'masker', 'ngobrol', 'kaget', 'gerbong', 'gaada', 'tugas', 'makasii'] |

### C. Case Folding

Case folding is one of the stages in text mining that functions to make words uniform by changing all letters to lowercase.

| | username | tweetcreatedts | text | clean_tweet | remove_http | remove_hastag | Tweet |
|---|---|---|---|---|---|---|---|
| 0 | mrbabymochi | 2022-07-16 17:38:53 | Pilihan terbaik pembuka konser setelah pandemi... | Pilihan terbaik pembuka konser setelah pandemi... | Pilihan terbaik pembuka konser setelah pandemi... | Pilihan terbaik pembuka konser setelah pandemi... | pilih baik buka konser pandemi konser epik hig... |
| 1 | Fasa77342812 | 2022-07-16 17:17:42 | @kembarannever @idextratime nangis liat lu jam... | nangis liat lu jamet 😂 lepas dong tuh masker | nangis liat lu jamet lepas dong tuh masker | nangis liat lu jamet lepas dong tuh masker | nang liat jamet lepas masker |
| 2 | parisinthelune | 2022-07-16 17:13:48 | @kaivanotes @Elitedly Jangan lepas masker lu s... | Jangan lepas masker lu sampe space berakhir | Jangan lepas masker lu sampe space berakhir | Jangan lepas masker lu sampe space berakhir | lepas masker sampe space |
| 3 | gwultramanribut | 2022-07-16 17:06:04 | TEMEN BARU GW ADA YANG MIRIP HISENG (pas pake ... | TEMEN BARU GW ADA YANG MIRIP HISENG (pas pake ... | TEMEN BARU GW ADA YANG MIRIP HISENG pas pake m... | TEMEN BARU GW ADA YANG MIRIP HISENG pas pake m... | temen hiseng pas pake masker pas lepas maskega... |
| 4 | mr_zhik | 2022-07-16 17:01:59 | RT @unkwn6969: Lagi hamil, baru pulang dari ka... | : Lagi hamil, baru pulang dari kampus belum ga... | Lagi hamil baru pulang dari kampus belum ganti... | Lagi hamil baru pulang dari kampus belum ganti... | hamil pulang kampus ganti baju mandi udah genj... |

*Figure 6: Data Case Folding Results*

*Table 3: Case Folding Detail*

| Before | After |
|---|---|
| nangis liat lu jamet lepas dong tuh masker | nangis liat lu jamet lepas dong tuh masker |
| Jangan lepas masker lu sampe space berakhir | jangan lepas masker lu sampe space berakhir |
| sonnalisa tolong bilangin sohib main pantai mbok masker lepas masker udah basah tetep dipake | sonnalisa tolong bilangin sohib main pantai mbok masker lepas masker udah basah tetep dipake |
| gesss serius nanya krl | gesss serius nanya krl |
| emang udah lepas masker lepas masker ngobrol kaget gerbong gaada tugas makasii | emang udah lepas masker lepas masker ngobrol kaget gerbong gaada tugas makasii |

### D. Filtering / Stopword Removal

The Stopword Removal stage serves to remove words that have no meaning.

```
#import stopword
import nltk
nltk.download('stopwords')

from nltk.corpus import stopwords
stopwords_indonesia = stopwords.words('indonesian')

from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory, StopWordRemover, ArrayDictionary

stop_factory = StopWordRemoverFactory().get_stop_words()
more_stopwords = [
    'yg', 'utk', 'cuman', 'deh', 'Btw', 'tapi', 'gua', 'gue', 'lo', 'lu',
    'kalo', 'trs', 'jd', 'nih', 'ntar', 'nya', 'lg', 'gk', 'ecusli', 'dpt',
    'dr', 'kpn', 'kok', 'kyk', 'donk', 'yah', 'u', 'ya', 'ga', 'km', 'eh',
    'sih', 'eh', 'bang', 'br', 'kyk', 'rp', 'jt', 'kan', 'gpp', 'sm', 'usah',
    'mas', 'sob', 'thx', 'ato', 'jg', 'gw', 'wkwk', 'mak', 'haha', 'iy', 'k',
    'tp', 'haha', 'dg', 'dri', 'duh', 'ye', 'wkwkwk', 'syg', 'btw',
    'nerjemahin', 'gaes', 'guys', 'moga', 'kmrn', 'nemu', 'yukkk',
    'wkwkw', 'klas', 'iw', 'ew', 'lho', 'sbnry', 'org', 'gtu', 'bwt',
    'klrga', 'clau', 'lbh', 'cpet', 'ku', 'wke', 'mba', 'mas', 'sdh', 'kmrn',
    'oi', 'spt', 'dlm', 'bs', 'krn', 'jgn', 'sapa', 'spt', 'sh', 'wakakaka',
    'sihhh', 'hehe', 'ih', 'dgn', 'la', 'kl', 'ttg', 'mana', 'kmna', 'kmn',
    'tdk', 'tuh', 'dah', 'kek', 'ko', 'pls', 'bbrp', 'pd', 'mah', 'dhhh',
    'kpd', 'tuh', 'kzl', 'byar', 'si', 'sii', 'cm', 'sy', 'hahahaha', 'weh',
    'dlu', 'tuhh'
]
data = stop_factory + more_stopwords

dictionary = ArrayDictionary(data)
str = StopWordRemover(dictionary)

print(data)
```

*Figure 7: Script Filtering*

*Table 4: Filtering*

| Before | After |
|---|---|
| nangis liat lu jamet lepas dong tuh masker | nangis liat jamet lepas dong tuh masker |
| Jangan lepas masker lu sampe space berakhir | jangan lepas masker lu sampe space akhir |
| sonnalisa tolong bilangin sohib main pantai mbok masker lepas masker udah basah tetep dipake | tolong bilangin main pantai masker lepas masker basah tetep dipake |
| gesss serius nanya krl emang udah lepas masker | serius krl lepas masker lepas masker ngobrol |
| lepas masker ngobrol kaget gerbong gaada tugas makasii | kaget makasii |

### E. Stemming

Stemming is the stage of forming words that have affixes as essential words by removing words with prefixes (prefixes), suffixes behind (suffixes), insertions in the middle of the base word (infixes), front and back affixes (confixes), and repetition of words.

```
['a', 'ada', 'adalah', 'adanya', 'adapun', 'agak', 'agaknya', 'agar', 'akan', 'akankah', 'akhir', 'akhiri', 'akhirnya', 'aku', 'akulah',
'amat', 'amatlah', 'anda', 'andalah', 'antar', 'antara', 'antaranya', 'apa', 'apaan', 'apabila', 'apakah', 'apalagi', 'apatah', 'arti',
'artinya', 'asal', 'asalkan', 'atas', 'atau', 'ataukah', 'ataupun', 'awal', 'awalnya', 'b', 'bagai', 'bagaikan', 'bagaimana', 'bagaimanak
ah', 'bagaimanapun', 'bagainamakah', 'bagi', 'bagian', 'bahkan', 'bahwa', 'bahwasannya', 'bahwasanya', 'baik', 'baiklah', 'bakal', 'bakal
an', 'balik', 'banyak', 'bapak', 'baru', 'bawah', 'beberapa', 'begini', 'beginian', 'beginikah', 'beginilah', 'begitu', 'begitukah', 'beg
itulah', 'begitupun', 'bekerja', 'belakang', 'belakangan', 'belum', 'belumlah', 'benar', 'benarkah', 'benarlah', 'berada', 'berakhir', 'b
erakhirlah', 'berakhirnya', 'berapa', 'berapakah', 'berapalah', 'berapapun', 'berarti', 'berawal', 'berbagai', 'berdatangan', 'beri', 'be
rikan', 'berikut', 'berikutnya', 'berjumlah', 'berkali-kali', 'berkata', 'berkehendak', 'berkeinginan', 'berkenaan', 'berlainan', 'berlal
u', 'berlangsung', 'berlebihan', 'bermacam', 'bermacam-macam', 'bermaksud', 'bermula', 'bersama', 'bersama-sama', 'bersiap', 'bersiap-sia
p', 'bertanya', 'bertanya-tanya', 'berturut', 'berturut-turut', 'bertutur', 'berujar', 'berupa', 'besar', 'betul', 'betulkah', 'biasa',
'biasanya', 'bila', 'bilakah', 'bisa', 'bisakah', 'boleh', 'bolehkah', 'bolehlah', 'buat', 'bukan', 'bukankah', 'bukanlah', 'bukannya',
'bulan', 'bung', 'c', 'cara', 'caranya', 'cukup', 'cukupkah', 'cukuplah', 'cuma', 'd', 'dahulu', 'dalam', 'dan', 'dapat', 'dari', 'daripa
da', 'datang', 'dekat', 'demi', 'demikian', 'demikianlah', 'dengan', 'depan', 'di', 'dia', 'diakhiri', 'diakhirinya', 'dialah', 'diantar
a', 'diantaranya', 'diberi', 'diberikan', 'diberikannya', 'dibuat', 'dibuatnya', 'didapat', 'didatangkan', 'digunakan', 'diibaratkan', 'd
iibaratkannya', 'diingat', 'diingatkan', 'diinginkan', 'dijawab', 'dijelaskan', 'dijelaskannya', 'dikarenakan', 'dikatakan', 'dikatakanny
a', 'dikerjakan', 'diketahui', 'diketahuinya', 'dikira', 'dilakukan', 'dilalui', 'dilihat', 'dimaksud', 'dimaksudkan', 'dimaksudkannya',
'dimaksudnya', 'diminta', 'dimintai', 'dimisalkan', 'dimulai', 'dimulailah', 'dimulainya', 'dimungkinkan', 'dini', 'dipastikan', 'diperbu
at', 'diperbuatnya', 'dipergunakan', 'diperkirakan', 'diperlihatkan', 'diperlukan', 'diperlukannya', 'dipersoalkan', 'dipertanyakan', 'di
punyai', 'diri', 'dirinya', 'disampaikan', 'disebut', 'disebutkan', 'disebutkannya', 'disini', 'disinilah', 'ditambahkan', 'ditandaskan',
'ditanya', 'ditanyai', 'ditanyakan', 'ditegaskan', 'ditujukan', 'ditunjuk', 'ditunjuki', 'ditunjukkan', 'ditunjukkannya', 'ditunjuknya',
'dituturkan', 'dituturkannya', 'diucapkan', 'diucapkannya', 'diungkapkan', 'dong', 'dua', 'dulu', 'e', 'empat', 'enak', 'enggak', 'enggak
nya', 'entah', 'entahlah', 'f', 'g', 'guna', 'gunakan', 'h', 'hadap', 'hai', 'hal', 'halo', 'hallo', 'hampir', 'hanya', 'hanyalah', 'har
i', 'harus', 'haruslah', 'harusnya', 'helo', 'hello', 'hendak', 'hendaklah', 'hendaknya', 'hingga', 'i', 'ia', 'ialah', 'ibarat', 'ibarat
kan', 'ibaratnya', 'ibu', 'ikut', 'ingat', 'ingat-ingat', 'ingin', 'inginkah', 'inginkan', 'ini', 'inikah', 'inilah', 'itu', 'itukah', 'i
tulah', 'j', 'jadi', 'jadilah', 'jadinya', 'jangan', 'jangankan', 'janganlah', 'jauh', 'jawab', 'jawaban', 'jawabnya', 'jelas', 'jelaska
n', 'jelaslah', 'jelasnya', 'jika', 'jikalau', 'juga', 'jumlah', 'jumlahnya', 'justru', 'k', 'kadar', 'kala', 'kalau', 'kalaulah', 'kalau
pun', 'kali', 'kalian', 'kami', 'kamilah', 'kamu', 'kamulah', 'kan', 'kapan', 'kapankah', 'kapanpun', 'karena', 'karenanya', 'kasus', 'ka
ta', 'katakan', 'katakanlah', 'katanya', 'ke', 'keadaan', 'kebetulan', 'kecil', 'kedua', 'keduanya', 'keinginan', 'kelamaan', 'kelihata
n', 'kelihatannya', 'kelima', 'keluar', 'kembali', 'kemudian', 'kemungkinan', 'kemungkinannya', 'kena', 'kenapa', 'kepada', 'kepadanya',
'kerja', 'kesampaian', 'keseluruhan', 'keseluruhannya', 'keterlaluan', 'ketika', 'khusus', 'khususnya', 'kini', 'kinilah', 'kira', 'kira-
kira', 'kiranya', 'kita', 'kitalah', 'kok', 'kurang', 'l', 'lagi', 'lagian', 'lah', 'lain', 'lainnya', 'laku', 'lalu', 'lama', 'lamanya',
'langsung', 'lanjut', 'lanjutnya', 'lebih', 'lewat', 'lihat', 'lima', 'luar', 'm', 'macam', 'maka', 'makanya', 'makin', 'maksud', 'mala
h', 'malahan', 'mampu', 'mampukah', 'mana', 'manakala', 'manalagi', 'masa', 'masalah', 'masalahnya', 'masih', 'masihkah', 'masing', 'masi
ng-masing', 'masuk', 'mata', 'mau', 'maupun', 'melainkan', 'melakukan', 'melalui', 'melihat', 'melihatnya', 'memang', 'memastikan', 'memb
eri', 'memberikan', 'membuat', 'memerlukan', 'memihak', 'meminta', 'memintakan', 'memisalkan', 'memperbuat', 'mempergunakan', 'memperkira
```

*Figure 8: Result Stemming*

*Table 5: Stemming*

| Sebelum | Sesudah |
|---|---|
| nangis liat lu jamet lepas dong tuh masker | nangis liat lu jamet lepas dong tuh masker |
| Jangan lepas masker lu sampe space berakhir | jangan lepas masker lu sampe space akhir |
| sonnalisa tolong bilangin sohib main pantai mbok masker lepas masker udah basah tetep dipake | sonnalisa tolong bilang sohib main pantai mbok masker lepas masker udah basah tetep pake |
| gesss serius nanya krl emang udah lepas masker lepas masker ngobrol kaget gerbong gaada tugas makasii | gesss serius nanya krl emang udah lepas masker lepas masker ngobrol kaget gerbong gaada tugas makasii |

*Table 6: Translate*

| Before | After |
|---|---|
| nangis liat lu jamet lepas dong tuh masker | cry when you see jamet take off the mask |
| Jangan lepas masker lu sampe space berakhir | Don't take off your mask until the space ends |
| sonnalisa tolong bilangin sohib main pantai mbok masker lepas masker udah basah tetep dipake | sonnalisa, please tell your friends to play at the beach, please take the mask off the mask is already wet and still |
| gesss serius nanya krl emang udah lepas masker lepas masker ngobrol kaget gerbong gaada tugas makasii | gesss seriously asking if you have taken off your mask, taken off your mask, you are surprised that the carriage has no work, thank you |

## 2.3 Translate

At the translation stage, it is helpful to translate tweets data taken from Indonesian into English-language tweets.

## 2.4 Sentiment Analysis or Classification

The author uses two methods at this stage: Lexicon Based and Naive Bayes Classifier.

A. Lexicon Based

The Lexicon-based method is a commonly used method for sentiment analysis on social media, and this method is efficient to use. In its implementation, lexicon-based uses a dictionary as a language source or a lexicon used for word selection in data or documents. The Lexicon-based method uses VADER (Valance Dictionary Sentiment Reasoner).

*Table 7: Lexicon Based . Sentiment Results*

| Tweet | Score Polarity | Sentimen |
|-------|----------------|----------|
| cry when you see jamet take off the mask | 0,0 | Netral |
| Don't take off your mask until the space ends | 0,0 | Netral |
| sonnalisa, please tell your friends to play at the beach, please take the mask off the mask is already wet and still | 0,8442 | Positif |
| gesss seriously asking if you have taken off your mask, taken off your mask, you are surprised that the carriage has no work, thank you | -0,4784 | Negatif |

B. Naïve Bayes Classifier

The classification stage by the Naïve Bayes Classifier is used to determine a sentence that is positioned as a collection of positive, neutral or negative sentiments based on an assessment of the probability count of the more powerful Bayes formula.

*Table 8: Naïve Bayes Classifier  Sentiment Results*

| Tweet | Klasifikasi | Klasifikasi Naïve Bayes |
|-------|-------------|-------------------------|
| cry when you see jamet take off the mask | Netral | Netral |
| Don't take off your mask until the space ends | Netral | Netral |
| sonnalisa, please tell your friends to play at the beach, please take the mask off the mask is already wet and still | Negatif | Negatif |
| gesss seriously asking if you have taken off your mask, taken off your mask, you are surprised that the carriage has no | Negatif | Positif |

| work, thank you | | |
|-----------------|--|--|

**2.5 Visualization**

The results were obtained through the visualization stages in this study, namely, histogram and word cloud. The actualization of the histogram displays the percentage results that come from the polarity class of each sentiment.
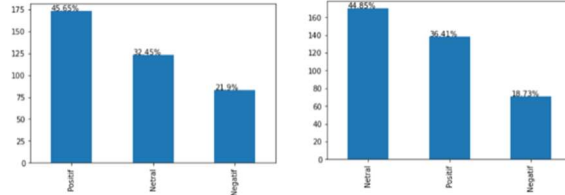


*Figure 9: Histogram Display and Percentage of Each Sentiment Lexicon-Based Method (left) and Naïve Bayes Classifier (right)*

Based on the picture above, it can be seen that public opinion on Twitter social media shows that positive sentiment has a high level of frequency. This indicates that the community has been enthusiastic about removing masks compared to those who are against them.

Then the results of the visualization in the form of a wordcloud with positive sentiment, which shows the words that often appear in the topic of removing the mask are presented in the following image.



*Figure 10: Wordcloud Display Of Positive Sentiment Lexicon-Based (left) Method And Naive Bayes Classifier (Right)*

Figure 11 and Figure 12 show the results of visualization of words that often appear in the neutral and negative sentiment class towards the topic of removing the mask. It can be seen that the negative sentiments that often arise are the words take and mask.

*Figure 11: Wordcloud Display Of Netral Sentiment Lexicon-Based (left) Method And Naive Bayes Classifier (Right)*



*Figure 12: Wordcloud Display Of Negativ Sentiment Lexicon-Based (left) Method And Naive Bayes Classifier (Right)*

**2.6 Performa Matrixs**

The testing stages of the metrics are carried out using the confusion matrix method, which will compare the results of manual and system classification.

*Table 9: Classification Results with Manual Labeling*

| Tweet | Classification |
|---|---|
| cry when you see jamet take off the mask | Netral |
| Don't take off your mask until the space ends | Netral |
| sonnalisa, please tell your friends to play at the beach, please take the mask off the mask is already wet and still | Positif |
| gesss seriously asking if you have taken off your mask, taken off your mask, you are surprised that the carriage has no work, thank you | Negatif |

The accuracy of the output in this study was checked by comparing the actual classification with the classification obtained by the model. The classification report is displayed to test the data classification using precision, recall, f1-score, macro avg, and weighted avg.

*Table 10: Lexicon Based Multiclass Confusion Matrix Results*

| Classification | TP (True Positive) | FP (False Positive) | FN (False Negative) |
|---|---|---|---|
| Positif | 144 | 28 | 16 |
| Netral | 108 | 18 | 21 |
| Negatif | 59 | 22 | 31 |

The table shows the multiclass results from the confusion matrix using the Lexicon Based method to represent the opinion data output that has been converted into numerical data.

*Table 11: Lexicon Based Method Report Classification Results*

| | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Negatif | 0.73% | 0.66% | 0.69% | 90 |
| Netral | 0.86% | 0.84% | 0.85% | 129 |
| Positif | 0.84% | 0.90% | 0.87% | 160 |
| *Micro avg (accuracy)* | **0.82%** | **0.82%** | **0.82%** | 379 |
| *Macro avg* | 0.81% | 0.80% | 0.80% | 379 |
| *Weighted avg.* | 0.82% | 0.82% | 0.82% | 379 |

*Table 12: Naïve Bayes Multiclass Confusion Matrix Results*

| Classification | TP (True Positive) | FP (False Positive) | FN (False Negative) |
|---|---|---|---|
| Positif | 64 | 6 | 74 |
| Netral | 74 | 114 | 2 |
| Negatif | 0 | 0 | 44 |

*Table 13: Naïve Bayes Method Report Classification Results*

| | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Negatif | 0.92% | 0.32% | 0.48% | 71 |
| Netral | 0.64% | 0.98% | 0.77% | 170 |
| Positif | 0.81% | 0.55% | 0.66% | 138 |
| *Micro avg (accuracy)* | **0.70%** | **0.70%** | **0.70%** | 379 |
| *Macro avg* | 0.79% | 0.62% | 0.64% | 379 |
| *Weighted avg.* | 0.75% | 0.70% | 0.67% | 379 |

Previous such works presented lexicon-based method or Naïve Bayes classifier, so that the comparison of those methods needs a lot of attention. Our findings showed that lexicon-based method is more accurate than Naïve Bayes classifier. Naïve Bayes classifier runs relatively fast which makes it suitable for real-time application. We proposed a hybrid classification method consist of lexicon-based method and Naïve Bayes classifier.

**3. CONCLUSION**

Based on the research that has been carried out, it can be concluded that lexicon-based is more accurate than Naïve Bayes classifier. The accuracy result obtained from the lexicon-based is 82%, positive sentiment is 45.65%, neutral 32.45%

negative 21.9%. In comparison, the accuracy result from the nave Bayes classifier model is 70%, and positive sentiment is 36.41% neutral 44.85% negative 18.73%. In further research, the authors suggest adding a collection of Indonesian dictionaries even more, because on social media Twitter there are too many non-standard words and foreign languages. Need some improvements at the preprocessing stage so that the results of the accuracy of the model can be even better. Dataset development can be reproduced even more so that the application runs more optimally. Trials with selection at the preprocessing stage might get even better results, and get a high accuracy value

## ACKNOWLEDGMENTS

## REFERENCE:

[1] W. A. Prabowo and C. Wiguna, "Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 149, 2021, doi: 10.30865/mib.v5i1.2604.

[2] WHO, "COVID-19 Weekly Epidemiological Update 35," *World Heal. Organ.*, no. December, pp. 1–3, 2020, [Online]. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/weekly_epidemiological_update_22.pdf

[3] I. Zulfa and E. Winarko, "Sentimen Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 11, no. 2, p. 187, 2017, doi: 10.22146/ijccs.24716.

[4] A. A. Permana, M. F. Fahrezi, D. Y. Priyanggodo, A. Kristiyanti, and M. Sihotang, "SENTIMEN ANALISIS OPINI MASYARAKAT PADA MEDIA SOSIAL TWITTER TERHADAP VAKSIN BERBAYAR MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER ( NBC ) SENTIMENT ANALYSIS OF PUBLIC OPINION ON SOCIAL MEDIA TWITTER ON PAID VACCINE USING NAÏVE BAYES CLASSIFIER ( NBC )

[5] C. Destitus, W. Wella, and S. Suryasari, "Support Vector Machine VS Information Gain: Analisis Sentimen Cyberbullying di Twitter Indonesia," *Ultim. InfoSys J. Ilmu Sist. Inf.*, vol. 11, no. 2, pp. 107–111, 2020, doi: 10.31937/si.v11i2.1740.

[6] R. S. Patil and S. R. Kolhe, "Resource Creation for Sentiment Analysis of Under-Resourced Language: Marathi," *Communications in Computer and Information Science*, vol. 1380. pp. 445–457, 2021. doi: 10.1007/978-981-16-0507-9_37.

[7] T. L. Maheswari, E. I. Rashmi, M. M. Hasanthi, and R. Elakkiya, "Customer Segmentation Based on Sentimental Analysis," *2022 International Conference on Advanced Computing Technologies and Applications, ICACTA 2022*. 2022. doi: 10.1109/ICACTA54488.2022.9753207.

[8] A. L. Fairuz, R. D. Ramadhani, and N. A. F. Tanjung, "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter," *J. Dinda Data Sci. Inf. Technol. Data Anal.*, vol. 1, no. 1, pp. 42–51, 2021, doi: 10.20895/dinda.v1i1.180.

[9] I. C. Juanatas, R. R. Fajardo, E. T. Manansala, A. A. Pasilan, J. R. Tabor, and H. D. A. Balmeo, "Sentiment Analysis Platform of Customer Product Reviews," *Proceedings of 2019 International Conference on Computational Intelligence and Knowledge Economy, ICCIKE 2019*. pp. 230–234, 2019. doi: 10.1109/ICCIKE47802.2019.9004415.

[10] O. D. B. JURNAL MANAJEMEN, V. 10 N. 02 D. 2021, and I. 1978-8754, "Dampak Pandemi COVID 19 terhadap Keuangan PDAM," vol. 10, pp. 13–21, 2021.

[11] A. Bhatnagar and M. Bhatia, "A Sentiment Analysis Based Approach for Customer Segmentation," *Recent Patents on Engineering*, vol. 16, no. 2. 2021. doi: 10.2174/1872212115666210122161605.

[12] M. Taboada, J. Brooke, and K. Voll, "Lexicon-Based Methods for Sentiment Analysis," no. December 2009, 2011.

[13] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining lexicon-based and learning-based methods for twitter sentiment analysis," *HP Lab. Tech. Rep.*, no. 89, 2011.

[14] C. Steven and W. Wella, "The Right Sentiment Analysis Method of Indonesian

Tourism in Social Media Twitter," *IJNMT (International J. New Media Technol.*, vol. 7, no. 2, pp. 102–110, 2020, doi: 10.31937/ijnmt.v7i2.1732.

[15] W. Yulita *et al.*, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," *Jdmsi*, vol. 2, no. 2, pp. 1–9, 2021.

[16] E. Indrayuni, A. Nurhadi, and D. A. Kristiyanti, "Implementasi Algoritma Naive Bayes , Support Vector Machine , dan K-Nearest Neighbors Untuk Analisa Sentimen," vol. 14, no. 2, pp. 64–71, 2021.