

PENERAPAN ALGORITMA NAÏVE BAYES CLASSIFICATION UNTUK KLASIFIKASI SENTIMENT TWEET TERHADAP PLATFORM STREAMING ILEGAL

Angga Sukma Wahyu P.P¹, Jati Sasongko Wibowo², I Putu Brama Arya Di Putra³
Universitas Stikubank Semarang^{1,2}, Institut Teknologi Bisnis Dan Bahasa Dian Cipta Cendikia³
Jl. Tri Lomba Juang, Mugassari, Kec. Semarang Sel., Kota Semarang, Jawa Tengah
E-mail : anggasukma2057@mhs.unisbank.ac.id¹, jatisw@edu.unisbank.ac.id², iputu@gmail.com³

ABSTRAK

Indonesia termasuk salah satu negara dengan pengguna platform streaming ilegal terbanyak di dunia, menurut survei *Asia Video Industry Association's, Coalition Against Piracy (CAP)* mengungkap bahwa 63% warga Indonesia yang menggunakan streaming online, lebih suka menonton dari platform *streaming* ilegal secara gratis dengan berbagai macam „konsekuensi“, disisi lain tindakan menonton secara ilegal termasuk tindakan melanggar hukum karena tidak memiliki izin siar, dimana hal ini dapat menimbulkan berbagai macam jenis komentar ataupun tanggapan dari *netizen* terhadap maraknya platform streaming ilegal ini, komentar ataupun tanggapan dari *netizen* dapat berjenis sentimen positif dan negatif. Komentar berupa sentimen ini dapat muncul dari berbagai media, salah satunya media sosial twitter, dimana media ini merupakan salah satu tempat yang berguna untuk menyerukan pendapat,tanggapan ataupun reaksi sentimen. Penelitian ini dilakukan untuk mengetahui komentar, tanggapan *netizen* terhadap platform *streaming* ilegal yang akan diklasifikasikan kedalam sentimen positif dan negatif dengan Algoritma *Naïve Bayes Classification* dengan bantuan metode pembobotan kata *TF-IDF(Term Frequency – Invers Dokument Frequency)*. Dalam penelitian ini diperoleh hasil seimbang dimana data dengan sentimen positif sebesar 50% sedangkan data dengan sentiment negatif sebesar 50% dengan nilai rata-rata accuracy 50%, precision 53%, recall 51%, and f-measure 51%.

Kata kunci : Analisis sentimen, *Algoritma naïve bayes classification*, *Streaming ilegal*, *Twitter*.

ABSTRACTS

According to a survey by the Coalition Against Piracy (CAP) of the Asia Video Industry Association, 63% of online streaming service users in Indonesia prefer to watch illegal streaming platforms for free with various "consequences." This makes Indonesia become one of the nations with the highest percentage of users of illegal streaming platforms. In the other side, illegally stream can also violate broadcasting rights rules, which can result some of comments from internet users, including those with both positive and negative feeling about the existence of this illegal streaming platforms. Comments in the form of these sentiments may be found in a variety of media, one of which is the social media site Twitter. Twitter serves as a platform for the expression of thoughts and sentiments as well as for enjoy free time. This research was conducted to find out the comments and opinions of netizens about illegal streaming platforms and will be classified into positive and negative sentiments using the Naïve Bayes Classification Algorithm with the help of the TF-IDF (Term Frequency – Inverse Document Frequency) weighting method. In this study, balanced results were obtained where positive sentiment data was 50% while negative sentiment data was 50% with the value of accuracy 50%, precision 53%, recall 51%, and f-measure 51 %

Keywords: Illegal streaming, Naïve bayes classification algorithm, Sentiment analisys, Twitter

1. PENDAHULUAN

Platform streaming ilegal merupakan situs

penyedia tontonan ilegal yang berisi konten yang tidak resmi dan melanggar hukum hak cipta, tahun 2019 pemerintah Indonesia tercatat telah berhasil memblokir sebanyak 1.946 platform streaming ilegal, namun karena banyaknya peminat di Indonesia, eksistensi platform streaming ilegal ini menjadi tidak mudah dimusnahkan. Survei dilakukan YouGov untuk Coalition Against Privacy (CAP) dari Asia Video Industry Association, menghasilkan data 63% warga Indonesia lebih sering menggunakan platform streaming untuk menonton video premium secara gratis (1), dengan mengunggulkan fasilitas utamanya yaitu memberikan akses secara gratis kepada pengguna, tentunya hal ini memberikan pertimbangan bagi netizen dan menimbulkan berbagai macam komentar, tanggapan serta reaksi sentimen terhadap eksistensi platform streaming ilegal tersebut.

Reaksi sentimen dapat ditemukan di berbagai macam media termasuk twitter. Twitter menjadi salah satu media sosial dengan user terbanyak, tercatat pada kuartal I-2022 total user mencapai sebanyak 229 juta diseluruh dunia (2). Twitter berperan sebagai wadah untuk menyampaikan komentar, pendapat dan juga reaksi sentimen. Banyaknya pengguna twitter dapat dipastikan sumber informasi, tersebar sangat banyak di media sosial tersebut, tentunya hal ini bisa dimanfaatkan untuk sarana dalam mengetahui reaksi sentimen dari netizen terkait platform streaming ilegal apakah termasuk kedalam reaksi sentimen positif atau negatif.

Analisis sentiment adalah proses pengumpulan data dari internet yang bersumber dari berbagai media sosial bertujuan untuk mendapatkan data dalam bentuk teks yang berisi opini, emosi, tingkah laku seseorang terhadap suatu objek dari pengguna media sosial tersebut dengan mengklasifikasikan data kedalam opini positif maupun opini negatif (3), analisis sentiment dapat digunakan untuk menganalisa, menggambarkan serta memprediksi perasaan hati seseorang pada suatu objek tertentu, karena analisis sentiment berfokus pada analisa dan pemahaman emosi dari tinjauan pola pada teks (4).

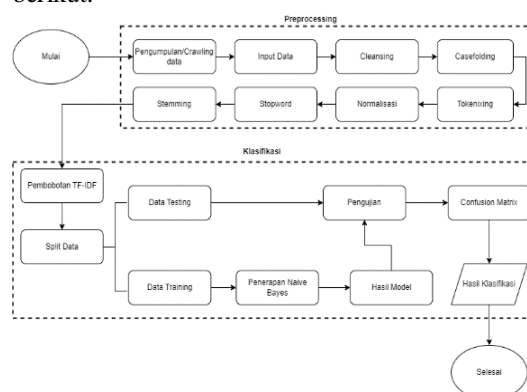
Klasifikasi adalah salah satu metode yang dipergunakan dalam melakukan penelitian di bidang data mining/penambangan data. Kegunaan dari klasifikasi adalah menentukan suatu fungsi atau model dalam membagi data ke dalam kelas yang telah ditentukan kemudian memprediksi kelas dari suatu data yang belum diketahui kelasnya dengan bantuan suatu model algoritma berdasarkan data latih, sehingga data baru dan data uji bisa diklasifikasikan berdasarkan model tersebut (5).

Dalam melakukan proses klasifikasi dibutuhkan sebuah algoritma. *Algoritma Naïve Bayes Classification* adalah metode klasifikasi yang berakar pada teorema *Bayes*. Proses klasifikasi menggunakan metode probabilitas dan statistik, untuk melakukan prediksi peluang berdasarkan hasil model sebelumnya (*Teorema Bayes*) dengan ciri utama yaitu asumsi yang sangat kuat (naif) akan ketergantungan dari masing-masing kondisi/kejadian, (6).

Penelitian ini akan melakukan klasifikasi terhadap data analisis sentimen dalam bentuk teks terhadap platform streaming ilegal dan akan diklasifikasikan kedalam kelas sentimen positif dan negatif. Data pada penelitian ini diperoleh melalui twitter dengan proses crawling data. Algoritma yang dipakai dalam tahap klasifikasi adalah *Algoritma Naïve Bayes Classification*.

2. METODE PENELITIAN

Terdapat beberapa langkah atau tahap pada proses penelitian. Terlampir pada gambar 1 berikut.



Gambar 1. Alur Tahap Penelitian

2.1. Pengumpulan/Crawling data

Sebelum melakukan penelitian tahap pertama pada penelitian ini adalah melakukan proses crawling data, dalam analisis sentimen pengumpulan data juga dapat disebut dengan crawling data, proses ini bertujuan untuk melakukan pengambilan data yang bersifat public secara online, dalam penelitian ini data yang diambil data berbentuk teks dan diperoleh dari media sosial twitter dengan memanfaatkan salah satu fiturnya yaitu *API(Application Programming Interface)*. Data terkumpul sebanyak 480 data, pengumpulan data dilakukan sejak 2 November 2022 hingga 14 Desember 2022. Setiap data akan dilabeli secara manual dengan cara memberi kelas positif dan negatif.

2.2. Pre Processing/ Pra Proses

Setelah diperoleh data dari proses crawling data, kemudian dilakukan pre processing, pre processing/pra proses, pre processing adalah tahap pembersihan data dari berbagai macam noise atau gangguan (7), seperti simbol, tanda baca, emoticon, dsb, hal ini dilakukan dengan tujuan untuk mempermudah sistem dalam proses membacanya. Preprocessing/Pra proses dilakukan dengan melakukan beberapa tahap pembersihan diantaranya dapat disimak dalam tabel 1.

Tabel 1. Pre Processing

No	Tahap	Keterangan
1	Cleansing	Proses pembersihan data dari hal-hal yang tidak diperlukan berupa tanda baca, emoticon, simbol, angka dan sebagainya
2	Casefolding	Proses untuk merubah setiap huruf kapital dalam sebuah kalimat menjadi huruf kecil
3	Tokenizing	Proses yang berfungsi untuk memisahkan sebuah kalimat menjadi beberapa kata/token
4	Normalisasi	Proses yang bertujuan untuk mengubah suatu kata yang tidak baku dalam kalimat menjad kata baku sesuai ketentuan EYD(Ejaan Yang Disempurnakan) kamus

		bahasa Indonesia
5	Stopword	Proses yang bertujuan untuk menghapus ataupun menghilangkan kata yang tidak memiliki makna serta arti yang berpengaruh terhadap proses penelitian analisis sentiment
6	Stemming	proses pengembalian kata berimbuhan yang akan dikembalikan kedalam bentuk kata dasar

2.3. Pembobotan kata dengan TF-IDF

TF-IDF(Term Frequency Invers Document Frequency) merupakan sebuah proses pembobotan atau pemberian *value/nilai* kepada setiap kata dalam suatu dokumen penelitian, hal ini dilakukan untuk merepresentasikan data teks menjadi *numeric*, *TF-IDF* berguna mempermudah algoritma *Naive Bayes Classification* dalam melakukan proses penghitungan, proses pembobotan kata dilakukan dengan bahasa pemrograman *python* menggunakan *google colab* sebagai media praktik. *TF-IDF* menggabungkan dua metode penghitungan, yaitu *TF(Term Frequency)* kata yang terkandung dalam dokumen dan *IDF(Invers Document Frequency)* dokumen yang mengandung kata (8), dapat disimak dalam persamaan 1,2,3 (9).

$$TF(d, t) = \frac{count(d, t)}{count(d)}$$
 (1)

$$IDF(t) = \frac{1}{\log(\frac{N}{count(t)})} + 1$$
 (2)

$$TF-IDF(d, t) = TF(d, t) * IDF(t)$$
 (3)

Keterangan:

TF(d,t) = Term Frequency

F(d,t) = frekuensi kemunculan kata(t) dalam dokumen d

IDF(t) = Invers Dokument Frequency pada setiap term(t)

N = jumlah keseluruhan data

Df(t) = jumlah dokumen yang mengandung term (t)

2.4. Algoritma Naive Bayes Classification

Algoritma Naive Bayes Classification adalah metode klasifikasi berdasarkan

teori kemungkinan dan *teorema Bayesian* dengan asumsi setiap variable atau parameter penentu keputusan bersifat bebas (*independence*) sehingga eksistensi/keberadaan setiap *variable* tidak ada kaitannya dengan keberadaan atribut yang lain. *Teorema bayes* yaitu memperkirakan peluang dengan mengacu hasil model sebelumnya (10). Penghitungan *Naive Bayes Classification* dihitung menggunakan persamaan berikut (11).

$$P(c|X) = \frac{P(X|c)P(c)}{P(X)} \quad (4)$$

Keterangan :

$P(c|X)$ = probabilitas hipotesis berdasarkan kondisi

$P(X|c)$ = probabilitas berdasarkan kondisi pada hipotesis

$P(c)$ = probabilitas hipotesis

$P(X)$ = probabilitas c

Karena data dalam penelitian ini termasuk data berbentuk teks, maka digunakan salah satu model algoritma *Naive Bayes* yang sesuai, yaitu *Multinomial Naive Bayes* (12), dikarenakan *Multinomial Naive Bayes* termasuk dalam metode supervised learning, maka data yang akan digunakan untuk proses penelitian, dilakukan pelabelan sebelum memasuki tahap pelatihan data, proses ini menggunakan persamaan rumus: (13).

$$P(d|c) \propto P(c) \prod_{i=1}^n P(x_i|c) \quad (5)$$

Keterangan :

$P(d|c)$ = Probabilitas/kemungkinan dokumen d pada kelas c

$P(c)$ = Prior probability suatu dokumen pada kelas c $\{c_1, c_2, \dots, c_n\}$ Token dalam dokumen d yang menjadi bagian dari vocabulary dengan jumlah n

$P(x_i|c)$ = Probabilitas bersyarat dari term x_i di dokumen pada kelas c

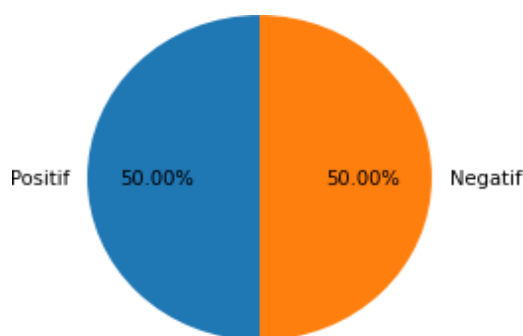
2.5. Uji Validasi Keakuratan

Dalam menguji hasil model dari penghitungan dengan algoritma *Naive Bayes Classification*, metode *K-fold Cross Validation* berfungsi untuk mencari akurasi nilai dari precision, recall, dari model. *K-fold Cross Validation* merupakan suatu jenis pengujian *cross*

validation (validasi silang) dengan tujuan menentukan nilai dan menguji sebuah proses algoritma dengan cara membagi sampel data (*training*) secara acak serta mengelompokkan data sebanyak nilai *k-fold*, dimana sebagian kelompok dijadikan data testing dan kelompok lainnya dijadikan data *training* (14).

3. HASIL DAN PEMBAHASAN

Dalam hasil penelitian, terkumpul sebanyak 266 *tweet* yang diperoleh melalui twitter, dengan metode *crawling* data menggunakan bahasa pemrograman *python* dengan beberapa *library* antara lain *python-csv*, *pandas*, *scikit learn*, *numpy*, *estimator* dan *sastrawi* serta dilakukan proses *import library* agar dapat digunakan. Data diberi kelas positif dan negatif proses pemberian kelas dilakukan secara manual. Selanjutnya dilakukan proses pembersihan data dengan metode *Pre processing*/Pra proses kemudian dilakukan pembobotan kata dengan *TF-IDF* (*Term Frequency Invers Dokument Frequency*) untuk memberikan bobot/nilai disetiap kata, hal ini bertujuan untuk mempermudah proses penghitungan klasifikasi data dengan algoritma *Naive Bayes Classification*. Dilakukan pengujian keakuratan validasi dengan *K-Fold Validation* untuk mengetahui rata-rata nilai *accuracy*, *precision*, *recall* dan *f-measure* dari model. Tahap-tahap proses penelitian dijelaskan pada bagian berikut Grafik pada gambar 1 menjelaskan pembagian data sekaligus akan dilakukan pengelompokan berdasarkan jenis kelas dengan kelas positif berjumlah 50% dan kelas negatif berjumlah 50%.



Gambar 2 Grafik jumlah data setiap kelas

3.1 Pre processing

Proses *Pre processing*/pra proses yang dilakukan dengan data training dapat disimak pada tabel 2 .

Tabel 2. Hasil Preprocessing

No	Tahap	Keterangan
1	<i>Cleansing</i>	Einblauervogel yang sub Indo udah ada di banyak web streaming ilegal
2	<i>Casefolding</i>	einblauervogel yang sub indo udah ada di banyak web streaming ilegal
3	<i>Tokenizing</i>	“einblauervogel”, “yang”, “sub”, “indo”, “udah”, “ada”, “di”, “banyak”, “web”, “streaming”, “ilegal”
4	<i>Normalisasi</i>	einblauervogel yang terjemahan di indonesia sudah ada di banyak website streaming ilegal
5	<i>Stopword</i>	einblauervogel terjemahan indonesia sudah ada banyak website streaming ilegal
6	<i>Stemming</i>	einblauervogel terjemah indonesia ada banyak website streaming ilegal

3.2 Pembobotan TF-IDF

Setelah proses *Pre processing*/pra proses berhasil dilakukan, selanjutnya melakukan pembobotan kata dengan *TF-IDF*(*Term Frequency Invers Dokument Frequency*), proses TF-IDF melakukan perhitungan pada frekuensi kemunculan kata di tiap dokumen data. Hasil dari tahap pembobotan *TF-IDF* dapat dilihat dalam tabel (3) dan (4), dengan menerapkan persamaan rumus (1),(2),(3).

Tabel 3. Data TF-IDF

Dokumen	Isi	Kelas
D1	streaming ilegal salah satu yang lestari film langka bahkan cuma mereka yang punya akses ke film yang sekarang	Positif
D2	jangan sampe nonton film ilegal ya sekarang gampang banget tinggal	Negatif

	langgan	
--	---------	--

Tabel 4. Hasil Proses Perhitungan Pembobotan TF-IDF

No	Term	D1	D2
1	akses	0.326189005	0
2	bahkan	0.326189005	0
3	banget	0	0.288675135
4	cuma	0.326189005	0
5	film	0	0.288675135
6	gampang	0	0.288675135
7	ilegal	0.326189005	0.288675135
8	jangan	0	0.288675135
9	ke	0.326189005	0
10	lestari	0.326189005	0
11	langgan	0	0.288675135
12	langka	0.326189005	0
13	mereka	0.326189005	0
14	nonton	0	0.288675135
15	punya	0.326189005	0
16	streaming	0.326189005	0
17	satu	0.326189005	0
18	sekarang	0.326189005	0.288675135
19	sampai	0	0.288675135
20	tinggal	0	0.288675135
21	ya	0	0.288675135
22	yang	0.326189005	0.288675135

3.3 Penerapan Klasifikasi Algoritma Naive Bayes Classification

Selanjutnya data yang telah diketahui nilai pembobotanya akan dilanjutkan untuk proses klasifikasi menggunakan Algoritma *Naive Bayes Classification*. Data akan dibagi menjadi 2 yaitu data training dan data uji, proses pembagian dilakukan dengan memanfaatkan *K-Fold Cross Validation*. Perlu diketahui, data yang dilakukan pembagian harus sudah melewati proses pembobotan *TF-IDF* (15). Kegunaan data *training/latih* adalah, untuk menciptakan sebuah model, dan kegunaan data *testing* adalah, untuk menguji performa dari model yang tercipta dari data training/latih, sehingga akan diketahui nilai akurasi. Dalam penghitungannya menerapkan rumus (5).

3.4 Uji Validasi Keakuratan Dengan K-Fold Validation

Pengujian validasi dilakukan untuk mengukur atau mengetahui nilai akurasi yang telah dilakukan pada tahap sebelumnya. 266 data

berupa tweet digunakan pada penelitian, dimana jumlah ini termasuk data training dan data testing, akan dilakukan pengujian atau iterasi *K-Fold Validation* dengan 10 kali pengujian dan membagi total 266 buah data menjadi 10 bagian, dimana setiap satu iterasi/pengujian dibagi lagi menjadi 2 kelompok diantaranya, 9 data *training* dan 1 data *testing* pada setiap iterasinya. Terdapat sebanyak 239 data *training* dan 27 data *testing* pada iterasi ke-1, kemudian disetiap tahap iterasi posisi dari data *testing* berubah dengan cara digeser. Terlampir pada tabel 5.

Tabel 5. Pembagian Data Uji

Iteras i 1	2 7	2 7	2 7	2 7	2 7	2 7	2 7	2 6	2 6	2 6	2 6
Iteras i 2	2 7	2 7	2 7	2 7	2 7	2 7	2 6	2 6	2 6	2 6	2 6
Iteras i 3	2 7	2 7	2 7	2 7	2 7	2 7	2 6	2 6	2 6	2 6	2 6
Iteras i 4	2 7	2 7	2 7	2 7	2 7	2 7	2 6	2 6	2 6	2 6	2 6
Iteras i 5	2 7	2 7	2 7	2 7	2 7	2 7	2 6	2 6	2 6	2 6	2 6
Iteras i 6	2 7	2 7	2 7	2 7	2 7	2 7	2 6	2 6	2 6	2 6	2 6
Iteras i 7	2 7	2 7	2 7	2 7	2 7	2 7	2 6	2 6	2 6	2 6	2 6
Iteras i 8	2 7	2 7	2 7	2 7	2 7	2 7	2 6	2 6	2 6	2 6	2 6
Iteras i 9	2 7	2 7	2 7	2 7	2 7	2 7	2 6	2 6	2 6	2 6	2 6
Iteras i 10	2 7	2 7	2 7	2 7	2 7	2 7	2 6	2 6	2 6	2 6	2 6

Dari pengujian validasi keakuratan yang dilakukan dengan *K-Fold Validation* dengan pengulangan sebanyak 10 kali Iterasi, menghasilkan nilai *confusion matrix* yang terlampir dari gambar 2 hingga 11

	precision	recall	f1-score	support
negatif	0.20	0.33	0.25	9
positif	0.50	0.33	0.40	18
accuracy			0.33	27
macro avg	0.35	0.33	0.33	27
weighted avg	0.40	0.33	0.35	27

confusion matrix:
[[3 6]
[12 6]]
=====

Gambar 3. Pengujian ke 1

	precision	recall	f1-score	support
negatif	0.89	0.42	0.57	19
positif	0.39	0.88	0.54	8
accuracy			0.56	27
macro avg	0.64	0.65	0.55	27
weighted avg	0.74	0.56	0.56	27

confusion matrix:
[[8 11]
[1 7]]
=====

Gambar 4. Pengujian ke 2

	precision	recall	f1-score	support
negatif	0.62	0.45	0.53	11
positif	0.68	0.81	0.74	16
accuracy			0.67	27
macro avg	0.65	0.63	0.63	27
weighted avg	0.66	0.67	0.65	27

confusion matrix:
[[5 6]
[3 13]]
=====

Gambar 5. Pengujian ke 3

	precision	recall	f1-score	support
negatif	0.17	0.50	0.25	6
positif	0.67	0.29	0.40	21
accuracy			0.33	27
macro avg	0.42	0.39	0.33	27
weighted avg	0.56	0.33	0.37	27

confusion matrix:
[[3 3]
[15 6]]
=====

Gambar 6. Pengujian ke 4

	precision	recall	f1-score	support
negatif	0.50	0.17	0.25	18
positif	0.25	0.62	0.36	8
accuracy			0.31	26
macro avg	0.38	0.40	0.30	26
weighted avg	0.42	0.31	0.28	26

confusion matrix:
[[3 15]
[3 5]]
=====

Gambar 7. Pengujian ke 5

```

precision    recall    f1-score   support

negatif      0.73     0.50     0.59      16
positif      0.47     0.70     0.56      10

accuracy
macro avg    0.60     0.60     0.58      26
weighted avg 0.63     0.58     0.58      26

confusion matrix:
[[8 8]
 [3 7]]
    
```

Gambar 8. Pengujian ke 6

```

precision    recall    f1-score   support

negatif      0.56     0.75     0.64      12
positif      0.70     0.50     0.58      14

accuracy
macro avg    0.63     0.62     0.62      26
weighted avg 0.64     0.62     0.61      26

confusion matrix:
[[9 3]
 [7 7]]
    
```

Gambar 9. Pengujian ke 7

```

precision    recall    f1-score   support

negatif      0.67     0.59     0.62      17
positif      0.36     0.44     0.40       9

accuracy
macro avg    0.52     0.52     0.51      26
weighted avg 0.56     0.54     0.55      26

confusion matrix:
[[10 7]
 [ 5 4]]
    
```

Gambar 10. Pengujian ke 8

```

precision    recall    f1-score   support

negatif      0.75     0.56     0.64      16
positif      0.50     0.70     0.58      10

accuracy
macro avg    0.62     0.63     0.61      26
weighted avg 0.65     0.62     0.62      26

confusion matrix:
[[9 7]
 [3 7]]
    
```

Gambar 11. Pengujian ke 9

```

precision    recall    f1-score   support

negatif      0.31     0.50     0.38       8
positif      0.69     0.50     0.58      18

accuracy
macro avg    0.50     0.50     0.48      26
weighted avg 0.57     0.50     0.52      26

confusion matrix:
[[4 4]
 [9 9]]
    
```

Gambar 12. Pengujian ke 10

Dengan rata-rata hasil keseluruhan dari iterasi ke 1 hingga 10 dapat disimak pada tabel 6.

Tabel 6. Hasil Pengujian Validasi Keakuratan

Iterasi	Accurac y	Precission	Recall	F-Measure
Ke-1	33%	35%	33%	33%
Ke-2	56%	64%	65%	55%
Ke-3	67%	65%	63%	63%
Ke-4	33%	42%	39%	33%
Ke-5	31%	38%	40%	30%
Ke-6	58%	60%	60%	58%
Ke-7	62%	63%	62%	61%
Ke-8	54%	52%	52%	51%
Ke-9	62%	62%	63%	61%
Ke-10	50%	50%	50%	48%
Rata-rata	50%	53%	51%	51%

4. HASIL DAN PEMBAHASAN

Dalam keseluruhan pengujian yang telah dilakukan dapat disimak pada tabel 5. Pengujian validasi dengan *K-Fold Validation* memperoleh nilai rata-rata *accuracy* 50%, *precission* 53%, *recall* 51% dan *f-measure* 51%, dengan nilai prediksi *accuracy* maksimum terdapat pada pengujian ke-3 dengan hasil 67%, dan nilai prediksi *accuracy* minimum terdapat pada pengujian ke-5 dengan hasil 31%. Nilai prediksi *precission* maksimum terdapat pada pengujian ke-3 dengan hasil 65%, dan nilai prediksi *precission* minimum terdapat pada pengujian ke-1 dengan hasil 35%. Nilai prediksi *recall* maksimum terdapat pada pengujian ke-2 dengan hasil 65%, dan nilai prediksi *recall* minimum terdapat pada pengujian ke-1 dengan hasil 31%. Nilai prediksi *f-measure* maksimum terdapat pada pengujian ke-3 dengan hasil 61%, dan nilai prediksi *f-measure* minimum terdapat pada pengujian ke-5 dengan hasil 30%.

4.1 VISUALISASI DATA

Visualiasi data penelitian ini menggunakan *wordcloud* yang diterapkan pada *google colab* dengan bahasa pemrograman *python*. *WordCloud* merupakan istilah untuk memfvisualisasikan kata yang memiliki frekuensi kemunculan yang tinggi berdasarkan data tweet. Contoh visual dapat disimak pada gambar 13.

Bayes Untuk Klasifikasi Artikel Online Tentang Gempa Di Indonesia,” *J. Teknol. Informasi, Komputer, Dan Apl. (JTika)*, Vol. 2, No. 1, Pp. 89–100, 2020, Doi: 10.29303/Jtika.V2i1.87.

COMMUNITY SENTIMENT ABOUT DEPRESSION ON IMPLEMENTASI ALGORITMA NAIVE BAYES CLASSIFIER UNTUK,” Vol. 3, No. 5, 2022.

- [10] R. Nadia, D. K. M. L, And M. . Fhira Nhita, S.T., “Analisis Dan Implementasi Algoritma Naïve Bayes Classifier Terhadap Pemilihan Gubernur Jawa Barat 2018 Pada Media Online,” *E-Proceeding Eng.*, Vol. 5, No. 1, Pp. 1678–1700, 2018.
- [11] Y. I. Kurniawan, “Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining,” *J. Teknol. Inf. Dan Ilmu Komput.*, Vol. 5, No. 4, P. 455, 2018, Doi: 10.25126/Jtiik.201854803.
- [12] Algorit.Ma, “11,” 2022. <https://Algorit.Ma/Blog/Naive-Bayes-2022/>.
- [13] S. Eka Y. Putri, “Penerapan Model Naive Bayes Untuk Memprediksi Potensi Pendaftaran Siswa Di Smk Taman Siswa Teluk Betung Berbasis Web,” *J. Inform. Dan Rekayasa Perangkat Lunak*, Vol. 1, No. 1, Pp. 93–99, 2020, Doi: 10.33365/Jatika.V1i1.228.
- [14] M. T. Furqon, Indriati, And A. Hutapea, “Penerapan Algoritma Modified K-Nearest Neighbour Pada Pengklasifikasian Penyakit Kejiwaan Skizofrenia,” *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, Vol. 2, No. 10, Pp. 3957–3961, 2018.
- [15] S. Mulyani, R. Novita, P. Studi, And S. Informasi, “IMPLEMENTATION OF THE NAIVE BAYES CLASSIFIER ALGORITHM FOR CLASSIFICATION OF