

ANALISIS SENTIMEN TERHADAP PIALA DUNIA U-20 YANG BATAL DISELENGGARAKAN DI INDONESIA PADA MEDIA YOUTUBE DENGAN MENGUNAKAN METODE NAÏVE BAYES

Aang Aliyudin¹, Rianto²

^{1,2}Universitas Siliwangi

^{1,2}Jl. Siliwangi No.24, Kahuripan, Kec. Tawang, Kab. Tasikmalaya, Kota Tasikmalaya, 46115,
Indonesia

Email : ¹207006009@student.unsil.ac.id, ²rianto@unsil.ac.id

ABSTRAK

Piala Dunia U-20 merupakan turnamen sepak bola internasional yang diselenggarakan oleh FIFA. Penyelenggarannya dilakukan dua tahun sekali. Setiap negara yang mengikuti turnamen tersebut merupakan hasil penyeleksian dari serangkaian proses yang telah dilalui. Indonesia berkesempatan menjadi tuan rumah tersebut di tahun 2023. Begitu banyak keuntungan yang didapatkan seperti timnas Indonesia mendapatkan kesempatan bermain, mengenalkan budaya Indonesia kepada dunia, memajukan bidang pariwisata dan memajukan ekonomi kreatif Indonesia. Tetapi terdapat permasalahan internal sehingga Indonesia batal menjadi tuan rumah Piala Dunia U-20. Mudah-mudahan orang mengakses sosial media membuat banyak yang mengungkapkan *response* dari permasalahan tersebut, terutama pada platform Youtube karena banyak menayangkan video yang membahas hal itu. Dari komentar pada video di Youtube yang membahas hal tersebut dapat dilakukan analisis sentimen untuk mengetahui emosi atau pendapat seseorang yang diklasifikasikan menjadi *positive*, *negative* dan *netral*. Naïve bayes merupakan algoritma untuk melakukan analisis sentimen dengan melakukan pengklasifikasian. Dengan algoritma *Naïve Bayes* menghasilkan akurasi yang cukup tinggi yaitu 0.906 dengan rasio *split* data 80:20 antara data *training* dan data *testing* yang dijadikan pemodelan.

Keywords: *Piala Dunia, Naïve Bayes, Analisis Sentimen*

ABSTRACT

The U-20 World Cup is an international football tournament organized by FIFA. It is held every two years. Every country that takes part in the tournament is the result of selection from a series of processes that have been passed. Indonesia has the opportunity to become the host in 2023. There are so many benefits such as the Indonesian National Team getting the opportunity to play, introducing Indonesian culture to the world, advancing the tourism sector and advancing Indonesia's creative economy. But there were internal problems so that Indonesia was canceled from hosting the U-20 World Cup. The ease with which people access social media makes many express responses to these problems, especially on the Youtube platform because there are lots of videos that discuss it. From comments on videos on Youtube that discuss this matter, sentiment analysis can be carried out to find out someone's emotions or opinions which are classified into positive, negative and neutral. Naïve Bayes is an algorithm for performing sentiment analysis by classifying. The Naïve Bayes algorithm produces a fairly high accuracy of 0.906 with a split data ratio of 80:20 between training data and testing data which is used as modeling.

Keywords: *Piala Dunia, Naïve Bayes, Analisis Sentimen*

1. PENDAHULUAN

Piala Dunia U-20 merupakan turnamen besar sepak bola yang dinaungi oleh FIFA. Acara sepak bola ini biasa diselenggarakan dua tahun sekali. Pada tahun 2023, Indonesia dapat berkesempatan menjadi tuan rumah Piala Dunia U-20. Diselenggarakannya turnamen tersebut di

Indonesia dapat membawa dampak positif seperti halnya timnas Indonesia memiliki kesempatan untuk bermain, pengenalan budaya Indonesia, memperkenalkan destinasi wisata, memajukan ekonomi kreatif dan lain-lain. Namun ada beberapa hal yang menyebabkan turnamen tersebut tidak jadi diselenggarakan di Indonesia. Tentunya hal tersebut menuai pro dan

kontra. Dengan mudah masyarakat Indonesia mengakses sosial media membuat banyak orang yang menanggapi hal tersebut salah satunya pada media *YouTube*.

Youtube merupakan platform digital untuk share video. Terdapat salah satu video yang membahas tidak jadinya Piala Dunia U-20 di Indonesia. Banyak netizen yang menuai ulasan pada video tersebut. oleh karena itu pada penelitian ini akan mengambil data dari ulasannya kemudian akan dilakukan analisis sentimen dengan melakukan serangkaian *natural language processing*[1].

Analisis sentimen merupakan sebuah metode untuk memproses data teks yang sebelumnya tidak terstruktur kemudian akan diproses sehingga data tersebut menjadi terstruktur[2]. Pada analisis sentimen biasa selalu berkaitan dengan ekspresi, emosi ataupun pendapat di dalam komentar pada sosial media[3]. Data teks tersebut akan diambil untuk dilakukan beberapa proses sehingga dapat diklasifikasikan menjadi positif, negatif dan netral[4]. Salah satu algoritma pengklasifikasian yaitu metode *Naïve Bayes* yang akan digunakan pada penelitian ini [5].

Metode *Naïve Bayes* merupakan sebuah algoritma pengklasifikasian data [6]. Pada metode ini prosesnya menggunakan probabilitas Bayes untuk melakukan prediksi sebuah kelas atau sebuah label data. *Naïve Bayes* melakukan perhitungan berdasarkan pada kemunculan fitur data yang digunakan. Kelebihan dari metode ini yaitu mudah diimplementasikan dan cepat dalam melakukan pemrosesan data[7]. Adapun proses pada Metode *Naïve Bayes* yaitu melakukan data yang diklasifikasikan, menghitung probabilitas dan kemudian menentukan kelas dengan probabilitas tertinggi untuk prediksi data yang baru.

2. TINJAUAN PUSTAKA

Pada studi literatur merupakan untuk mengkaji teori dan penelitian terkait sebelumnya. Dari hasil kajian itu akan dijadikan landasan. Salah satu untuk mengkaji penelitian sebelumnya yaitu *state of the art* dengan tujuan untuk melihat bidang atau ruang lingkup penelitian. Dengan *state of the art* dapat menggambarkan kondisi terkini dari teknologi atau penelitian yang paling baru. Berikut ini merupakan studi literatur dengan menggunakan

tabel state of the art yang disajikan pada tabel 1 di bawah ini.

Tabel 1. Penelitian Terkait

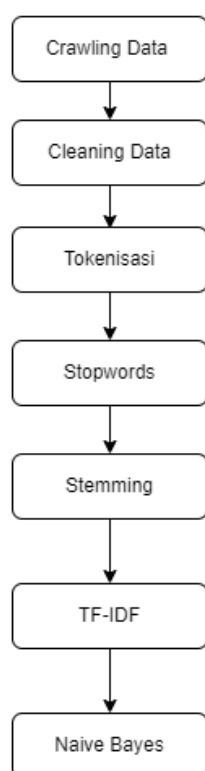
No	Penulis	Algoritma	Permasalahan
1	Fajar dkk (2020). [10]	Latent Dirichlet Allocation	Menganalisis respon Pro dan Kontra Masyarakat Indonesia mengenai Vaksin COVID-19 pada twitter
2	Dwi dkk (2021). [11]	<i>Naïve bayes</i> dan <i>confusion matrix</i>	Analisis sentimen pada twitter yang berbasis teks pada twitter
3	Hennie Tuhuteri (2020).[12]	<i>Support vector machine</i>	Analisis sentimen masyarakat pada pembatasan sosial bereskala besar.
4	Giovani dkk (2020). [13]	<i>Classification algorithm</i>	Analisis sentimen di aplikasi ruang guru pada twitter
5	Darwis dkk (2021). [14]	<i>Naïve bayes</i>	Analisis sentimen review data pada twitter BMKG Nasional
6	Putri dkk (2022). [15]	<i>Naïve bayes</i>	Analisis sentimen kinerja dewan perwakilan rakyat di twitter
7	Prabowo (2021). [16]	<i>Naïve bayes</i>	Analisis sentimen pada pembelajaran daring di masa pandemi COVID-19 pada twitter
8	Debby dkk (2020). [17]	<i>Random forest Classifier</i>	Pendeteksian sarkasme dengan analisis sentimen
9	Rachman dkk (2020). [10]	<i>Support vector machine</i>	Analisis sentimen masyarakat pada program kartu prakerja di twitter.
10	Laurensz dkk (2021). [18]	<i>Support vector machine</i>	Analisis sentimen masyarakat pada tindakan vaksinasi

			dalam upaya mengatasi pandemi COVID-19
--	--	--	--

Dari beberapa penelitian terkait, pada penelitian ini dilakukan analisis sentimen pada berita yang lagi banyak dibahas yaitu mengenai piala dunia U-20. Karena banyaknya tanggapan masyarakat terhadap acara yang batal diselenggarakan di Indonesia.

3. METODE PENELITIAN

Penggunaan metode dibutuhkan dalam penelitian ini supaya lebih terstruktur dalam pelaksanaannya. Metode yang terstruktur, maka hasil yang diperoleh akan sesuai tujuan yaitu analisis sentimen terhadap batalnya piala dunia yang diselenggarakan di Indonesia dengan metode *Naïve Bayes*. Berikut gambar di bawah ini merupakan alur dari penelitian yang dilakukan.



Gambar 1. Alur penelitian

3.1 Data Crawling

Pada proses data *crawling* melakukan pengambilan data yang belum diproses dalam sebuah media. Media yang digunakan pada penelitian ini yaitu *YouTube* sehingga

Application Programming Interface (API) yakni menggunakan API *Youtube* [8]. Data diambil berupa komentar sebuah video di youtube dengan konten mengenai bahasan dibatalkannya piala dunia U-20 di Indonesia sebagai tuan rumah. Dari data tersebut akan disimpan pada file dengan format CSV supaya nanti bisa diproses pada tahapan selanjutnya.

3.2 Data Preprocessing

Pada proses data crawling, data disimpan pada format CSV kemudian pada tahap data *preprocessing* karena data tersebut belum terstruktur agar lebih efektif dalam melakukan pemrosesan data[9]. Tahapan dalam data processing adalah sebagai berikut.

1. *Cleaning* data

Data karena masih tidak terstruktur dan terdapat banyak *noise*. Pada tahapan ini akan menghilangkan *hashtag*, *mentions*, simbol, emotikon, *link*, angka, mengganti garis baru menjadi spasi, menghapus kolom yang tidak diperlukan dan menghapus *double* spasi.

2. Tokenisasi

Teks akan diubah menjadi token yang merepresentasikan kata, frasa ataupun simbol tertentu. Tujuan dari tokenisasi yaitu untuk memudahkan dalam menganalisis teks maupun mengolahnya supaya lebih efektif dan presisi.

3. *Stopwords*

Pada tahap *stopwords* yaitu menghilangkan kata-kata yang dianggap tidak penting dalam melakukan analisis teks karena dianggap tidak memiliki makna yang signifikan. *Stopwords* merupakan kata-kata penghubung sehingga kata-kata yang dihilangkan. Mengenai tahap ini juga di adaptasikan dengan ketikan pada komentar karena terdapat kata-kata yang diketik secara tidak benar berdasarkan ejaannya.

4. *Stemming*

Pada tahapan *Stemming* merupakan proses mengubah bentuk kata menjadi kata aslinya dengan menghilangkan awalan (*prefix*) atau akhiran (*suffix*). Bertujuan untuk mengurangi variasi kata sehingga akan mempercepat komputasi dalam menganalisis teks.

3.3 TF-IDF

Pada *Term Frequency – Inverse Document Frequency* (TF-IDF) akan dilakukannya mengambil informasi dan mengukur seberapa penting

sebuah kata dalam sebuah dokumen. Pada tahapan ini juga data akan dilakukan vektorisasi yang mana teks akan diubah menjadi numerik supaya teks dapat dimengerti oleh komputer.

3.4 Naïve Bayes

Pada proses ini akan dilakukan pengelompokan yaitu positif, netral dan negatif terhadap data yang diambil. Data yang sudah diproses sebelumnya akan pisah menjadi dua yaitu data *training* dan data *testing*. Pemodelan dilakukan sebanyak 4 kali untuk mencari akurasi yang paling tinggi. Pemodelan dengan data *training* dan data *testing* yang berasio 90:10, 80:20, 70:30 dan 60:40. Pada data *training* akan diberikan pembobotan, positif diberi nilai 1, netral diberi nilai 0, dan negative diberi nilai -1. Menggunakan pemodelan dengan *Multinomial Naïve Bayes* yang tersedia di library *python*. Pada data *testing* untuk melakukan klasifikasi dan prediksi berdasarkan pemodelan yang sudah dilakukan oleh data *training* sebelumnya.

4. PEMBAHASAN

3.1 Hasil Data Crawling

Berikut ini merupakan hasil data crawling yang diambil pada platform *youtube* pada video yang membahas batal diselenggarakannya piala dunia U-20 di Indonesia. Mengenai datanya dapat dilihat pada gambar 2 di bawah ini.

	publishedAt	authorDisplayName	textDisplay	likeCount
0	2023-04-06T03:32:28Z	Najwa Shihab	Nonton juga Musyawarah "Piala Dunia U-20 ...	721
1	2023-04-18T07:11:26Z	Dionz Fabaretha	Kita main piala dunia karena jatah tuan rumah...	0
2	2023-04-18T05:18:37Z	otniel kundiman	Pembatalan nnt waktu sdh dekat sekali.	0
3	2023-04-18T05:01:01Z	taufik 59	misalkan Israel nyamar gara gara tidak diperbo...	0
4	2023-04-18T04:52:46Z	allansi_patriot+62	Pemimpin kok kalah sama Partai anccur negara K...	0
...
68058	2023-04-04T05:01:19Z	wakwau	Slankers Jawa Tengah hadir... Semangat terus	3
68059	2023-04-04T05:01:08Z	Heri Saputra	Pertama	0
68060	2023-04-04T05:00:52Z	Scumbag Man	1st min	0
68061	2023-04-04T05:00:50Z	prisa ivaniaap	Akhirnya yang di tunggu tunggu 🙏🙏🙏	8
68062	2023-04-04T05:00:44Z	Lah Kok	Pertama	0

68063 rows x 4 columns

Gambar 2. Hasil Data *Crawling*

Pada gambar 2 data tersebut diambil pada tanggal 19 april 2023. Dapat dilihat data yang diambil sebanyak 68.063 baris dan 4 kolom. Data diambil dari komentar video di youtube tetapi reply pada komentar tidak diambil dikarenakan data yang semakin banyak maka komputasi semakin lama dan akan memakan memory yang banyak. Data yang dihasilkan masih belum terstruktur, terdapat banyak noise seperti adanya link, spesial karakter, emotikon dan lainnya.

3.2 Hasil Cleaning Data

Dari hasil dari *cleaning data* yaitu akan menghilangkan hastag, mentions, simbol, emotikon, link, angka, mengganti garis baru menjadi spasi, menghapus kolom yang tidak diperlukan dan menghapus double spasi. Berikut hasilnya dapat dilihat pada tabel 1 dibawah ini.

Tabel 1. Hasil *Cleaning data*

No	Comments	Comments_Clean
0	nonton juga musyawarah "piala dunia u-20 ..	nonton juga musyawarah quotpiala dunia u gagal...
1	kita main piala dunia karena jatah tuan rumah,...	kita main piala dunia karena jatah tuan rumah ...
2	pembatalan nnt waktu sdh dekat sekali.	pembatalan nnt waktu sdh dekat sekali
....	
68061	akhirnya yang di tunggu tunggu 🙏🙏🙏	akhirnya yang di tunggu tunggu
68062	pertama	Pertama

Dapat dilihat pada tabel 1 perbedaan antara kolom *comments* yang berisi data belum di *cleaning* dan kolom *comments_clean* data yang sudah dilakukan *cleaning*. menghilangkan hastag, mentions, simbol, emotikon, link, angka, mengganti garis baru menjadi spasi, menghapus kolom yang tidak diperlukan dan menghapus double spasi.

3.3 Hasil Preprocessing

Berikut ini merupakan hasil preprocessing data. Pengolahan data yang di proses untuk menjadi lebih terstruktur supaya lebih efisien untuk digunakan. Berikut di bawah ini merupakan hasil data *preprocessing* data.

1. Tokenisasi

Berikut ini merupakan hasil tokenisasi pada tabel 2 dibawah ini.

Tabel 2. Tokenisasi

No	Token
0	[nonton, juga, musyawarah, quotpiala, dunia, u...
1	[kita, main, piala, dunia, karena, jatah, tuan...
2	[pembatalan, nnt, waktu, sdh, dekat, , sekali]
.....

68061	[akhirnya, yang, di, tunggu, tunggu]
68062	[pertama]

2. Stopwords

Hasil *Stopwords* yaitu menghilangkan kata-kata yang dianggap tidak penting dalam melakukan analisis teks karena dianggap tidak memiliki makna yang signifikan. Hasilnya dapat dilihat pada tabel 3 di bawah ini.

Tabel 3. *Stopwords*

No	Filtered_token
0	[nonton, musyawarah, quotpiala, d unia, u, gaga...
1	[main, piala, dunia, jatah, tuan, rumah, prest...
2	[pembatalan, nnt, sdh]
.....
68061	[tunggu, tunggu]
68062	[]

3. Stemming

Stemming merupakan proses mengubah bentuk kata menjadi kata aslinya dengan menghilangkan awalan (*prefix*) atau akhiran (*suffix*). Proses stemming bertujuan untuk mengurangi variasi kata sehingga akan mempercepat komputasi dalam menganalisis teks.

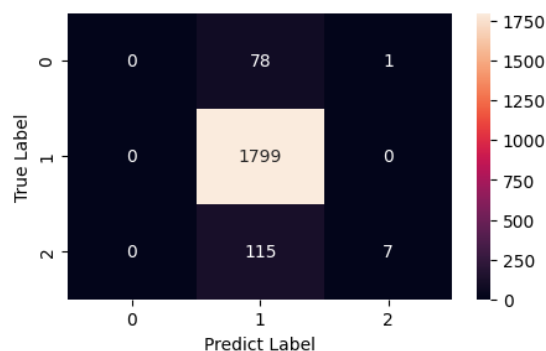
Tabel 4. *Stemming*

No	Stemmed_token
0	[nonton, musyawarah, quotpiala, d unia, u, gaga...
1	[main, piala, dunia, jatah, tuan, rumah, prest...
2	[pembatalan, nnt, sdh]
.....
68061	[tunggu, tunggu]
68062	[]

Dapat dilihat pada proses tokenisasi, *stopwords* dan *stemming* maka terdapat banyak perbedaan dari data *crawling*, pada hasil data *preprocessing* kolom yang tidak dibutuhkan dihilangkan. Kemudian dapat dilihat dua kolom pada tabel 1, data yang lebih terstruktur dan noisnya sudah dihilangkan seperti tidak adanya, spesial karakter, *link*, angka, emotikon dan lainnya. Dengan begitu ukuran data tidak terlalu besar dan lebih efisien untuk digunakan dalam komputasi pemodelan oleh suatu algoritma.

3.4 Hasil Analisis Sentimen Naïve Bayes

Data yang sudah melalui tahap *preprocessing* dan sudah ditranformasi menggunakan TF-IDF untuk kemudian akan melakukan pemodelan. Data akan di bagi menjadi dua yaitu data *training* dan data *testing*. Pengujian model dilakukan sebanyak 4 kali yaitu data training dan data testing dengan rasio 90:10, 80:20, 70:30, dan 60:40 untuk mendapatkan nilai akurasi yang tinggi dan data yang digunakan hanya sebanyak 20000 data karena keterbatasan komputasi komputer yang digunakan. Pada percobaan pertama splitting data yaitu dengan rasio 90:10 untuk melakukan pemodelan dan hasilnya pada gambar 3 sebagai berikut.



Gambar 3. *Confusion Matriks Pemodelan Pertama*

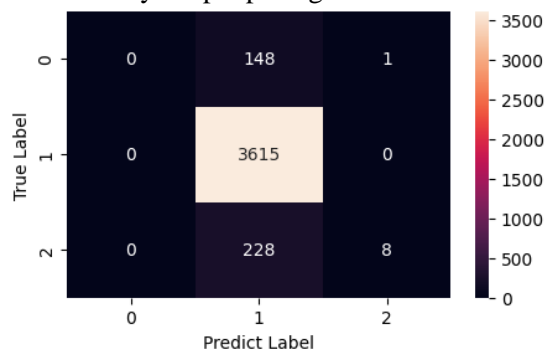
Pada gambar 3 dapat dilihat bahwa terdapat *true negative* 78, *true netral* 1799, dan *true positive* 115. Pada false negative 1, false netral 0 dan false positive 8. Mengenai *classification report* dapat dilihat pada tabel 1 di bawah ini.

Tabel 1. Pemodelan pertama

	Precision	Recall	F1-score
-1	0.00	0.00	0.00
0	0.90	1.00	0.95
1	0.88	0.06	0.11
accuracy			0.90

Pada tabel 1 di atas diketahui dari rasio 90:10 memiliki akurasi yang dihasilkan bernilai 0.90. Pada klasifikasi *negative* (-1) nilai *precision* yang dihasilkan yaitu 0.00, nilai *recall* 0.00, *f1-score* 0.00 dan *support* 79. Pada klasifikasi *netral* (0) nilai *precision* 0.90, *recall* 1.00, dan *f1-score* 0.95. Pada klasifikasi *positive* (1) nilai *precision* 0.88, *recall* 0.06, dan *f1-score* 0.11.

Selanjutnya percobaan kedua dengan melakukan menggunakan rasio data 80: 20 maka hasilnya dapat pada gambar 4.



Gambar 4. Confusion Matriks Pemodelan Kedua

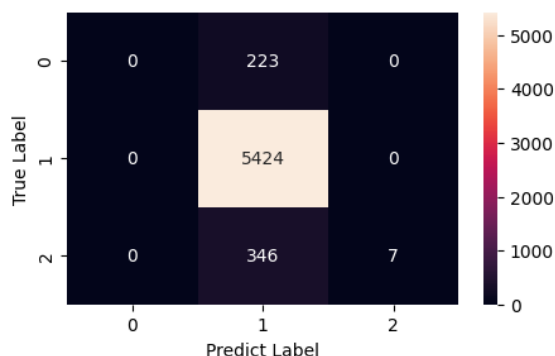
Pada gambar 4 dapat dilihat bahwa *true negative* sebanyak 148, *true netral* 3615, dan *true positive* 228. Diketahui untuk *false negative* 1 dan *false positive* 8. Detail *classification report* dapat dilihat pada tabel 2 di bawah ini.

Tabel 2. Pemodelan kedua

	Precision	recall	F1-score
-1	0.00	0.00	0.00
0	0.91	1.00	0.95
1	0.89	0.03	0.07
accuracy			0.90

Klasifikasi report pada tabel 2. Dapat dilihat dari rasio 80:20 akurasi bernilai 0.91. Pada klasifikasi *negative* (-1) nilai *precision* yang dihasilkan yaitu 0.00, nilai *recall* 0.00, dan *f1-score* 0.00. Pada klasifikasi *netral* (0) nilai *precision* 0.91, *recall* 1.00, dan *f1-score* 0.95. Pada klasifikasi *positive* (1) nilai *precision* 0.89, *recall* 0.03 dan *f1-score* 0.07.

Selanjutnya pemodelan ketiga dengan rasio data 70:30. Data *training* sebanyak 79% dan data *testing* sebanyak 30% dapat dilihat pada gambar 5 di bawah ini.



Gambar 5. Confusion Matriks Pemodelan Ketiga

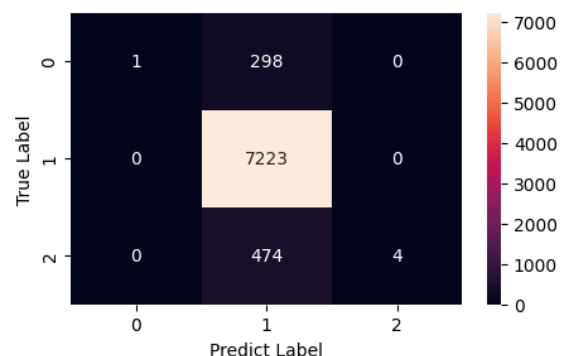
Pada gambar 5 dapat dilihat bahwa terdapat 223 *true negative*, 5424 *true netral*, 346 *true positive*, dan hanya memiliki 7 *false positive*. *Classification report* dalam pemodelan ini dapat dilihat pada tabel 3.

Tabel 3. Pemodelan ketiga

	precision	recall	F1-score
-1	0.00	0.00	0.00
0	0.91	1.00	0.95
1	1.00	0.02	0.04
accuracy			0.91

Pemodelan dengan rasio data 70:30 menghasilkan akurasi 0.91. Klasifikasi *negative* (-1) nilai *precision* yang dihasilkan yaitu 0.00, nilai *recall* 0.00, dan *f1-score* 0.00. Klasifikasi *netral* (0) nilai *precision* 0.91, *recall* 1.00, dan *f1-score* 0.95. Pada klasifikasi *positive* (1) nilai *precision* 1.00, *recall* 0.02, dan *f1-score* 0.04.

Kemudian percobaan keempat dengan melakukan menggunakan rasio data 60:40 maka hasilnya dapat dilihat pada gambar 6.



Gambar 6. Confusion Matriks Rasio 60:40

Gambar 6 menunjukkan hasil dari pemodelan keempat yaitu hanya menghasilkan 4 *false positive*, 298 *true negative*, 7223 *true netral*, dan 474 *true positive*. Pada klasifikasi report dapat dilihat pada tabel 4.

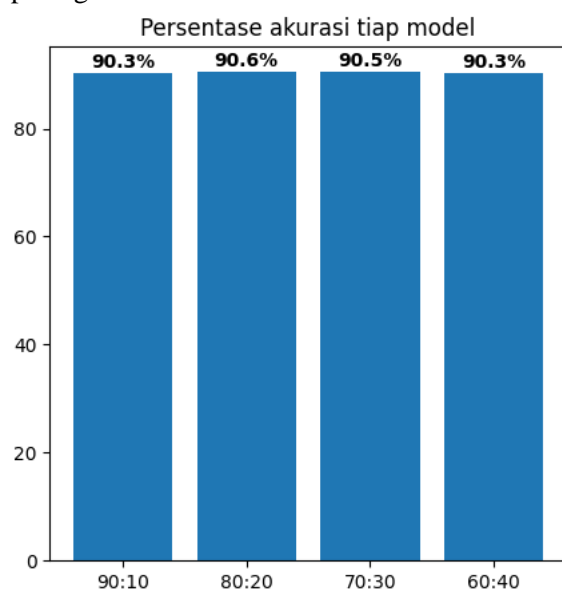
Tabel 4. Pemodelan keempat

	precision	recall	F1-score
-1	0.00	0.00	0.00
0	0.90	1.00	0.95
1	1.00	0.01	0.02
accuracy			0.90

Klasifikasi yang dihasilkan yaitu mendapatkan klasifikasi *negative* (-1) nilai *precision* yang dihasilkan yaitu 0.00, nilai *recall*

0.00, dan *f1-score* 0.00. Pada klasifikasi *netral* (0) nilai *precision* 0.90, *recall* 1.00, dan *f1-score* 0.95. Pada klasifikasi *positive* (1) nilai *precision* 1.00, *recall* 0.01, dan *f1-score* 0.02. Akurasi yang didapatkan yaitu 0.90.

Sebagai perbandingan setiap akurasi yang dihasilkan oleh setiap model dapat dilihat pada gambar 7.



Gambar 7. Grafik Perbandingan Akurasi

Perbandingan akurasi yang dihasilkan dari keempat model tersebut dapat dilihat akurasi yang dihasilkan paling tinggi yaitu pemodelan dengan rasio data 80:20. Nilai akurasi yang dihasilkan yaitu 0.906.

5. KESIMPULAN

Analisis sentimen yang dilakukan dengan mengambil sebanyak 68.063. Data tersebut dilakukan preprocessing sehingga data yang tadinya tidak terstruktur menjadi lebih terstruktur dan minimnya noise. Data training dan data testing yang digunakan hanya 20000 data karena keterbatasan dari komputasi yang besar. Dari pemodelan data *training* dan data *testing* rasio 80:20 yang menggunakan metode Naïve Bayes menghasilkan akurasi prediksi yang paling sebesar 0.906 yang mana nilai tersebut cukup tinggi.

Pada pengembangan penelitian selanjutnya diharapkan dapat menaikkan akurasi prediksi yang tinggi. Akurasi dapat berubah dari pemodelan dan data yang digunakan. Data yang digunakan supaya lebih bersih dari noise. Data yang digunakan untuk data training dan data

testing lebih banyak lagi karena akan sangat berpengaruh terhadap akurasi prediksi yang dihasilkan.

DAFTAR PUSTAKA

- [1] W. Wijaya, I. M. Murwantara, and A. R. Mitra, "A Simplified Method to Identify the Sarcastic Elements of Bahasa Indonesia in Youtube Comments," *2020 8th Int. Conf. Inf. Commun. Technol. ICoICT 2020*, pp. 0–5, 2020, doi: 10.1109/ICoICT49345.2020.9166269.
- [2] I. Verawati and B. S. Audit, "Algoritma Naïve Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By.u," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1411, 2022, doi: 10.30865/mib.v6i3.4132.
- [3] F. Syah, H. Fajrin, A. N. Afif, R. Saeputra, D. Mirranty, and D. D. Saputra, "Analisa Sentimen Terhadap Twitter IndihomeCare Menggunakan Perbandingan Algoritma Smote, Support Vector Machine, AdaBoost dan Particle Swarm Optimization," *J. Teknol. Inf. dan Komunikasi*, vol. 7, no. 1, pp. 0–5, 2023, [Online]. Available: <https://doi.org/10.35870/jti>
- [4] Yuyun, Nurul Hidayah, and Supriadi Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 820–826, 2021, doi: 10.29207/resti.v5i4.3146.
- [5] F. Febri Rahanto and I. Kharisudin, "44683-Article Text-133403-1-10-20210705," *Sentimen Data Ulas. Menggunakan Metod. Naive Bayes Stud. Kasus Wujil Resort Conv. Pada Situs Tripadvisor*, vol. 10, no. 1, pp. 55–62, 2021.
- [6] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, "Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile," *Petir*, vol. 15, no. 2, pp. 264–275, 2022, doi: 10.33322/petir.v15i2.1733.
- [7] M. Ridho Handoko and Neneng, "Sistem Pakar Diagnosa Penyakit Selama Kehamilan Menggunakan Metode Naive Bayes Berbasis Web," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 50–58, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [8] R. Azhar, A. Surahman, and C. Julianne, "Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naïve Bayes," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 1, pp. 267–281, 2022.
- [9] D. Ariyanti and K. Iswardani, "Teks Mining untuk Klasifikasi Keluhan Masyarakat Pada Pemkot Probolinggo Menggunakan Algoritma Naïve Bayes," *J. IKRA-ITH Inform.*, vol. 4, no. 3, pp. 125–132, 2020.
- [10] F. F. Rachman and S. Pramana, "Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter," *Heal. Inf. Manag. J.*, vol. 8, no. 2, pp. 100–109, 2020, [Online]. Available: <https://inohim.esaunggul.ac.id/index.php/INO/ar>

- ticle/view/223/175
- [11] J. S. Komputer, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," vol. 5, no. November 2019, pp. 697–711, 2021.
 - [12] H. Tuhuteru, "Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma Support Vector Machine," *Inf. Syst. Dev.*, vol. 5, no. 2, pp. 7–13, 2020.
 - [13] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
 - [14] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
 - [15] D. D. Putri, G. F. Nama, and W. E. Sulistiono, "ANALISIS SENTIMEN KINERJA DEWAN PERWAKILAN RAKYAT (DPR) PADA TWITTER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER," vol. 10, no. 1, pp. 34–40, 2022.
 - [16] 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.
 - [17] D. Alita and A. R. Isnain, "Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier," *J. Komputasi*, vol. 8, no. 2, pp. 50–58, 2020, doi: 10.23960/komputasi.v8i2.2615.
 - [18] B. Laurensz and Eko Sedyono, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 2, pp. 118–123, 2021, doi: 10.22146/jnteti.v10i2.1421.