

Analisis Sentimen Jersey Erspo Timnas Indonesia Menggunakan Model Klasifikasi dan Algoritma *Naive Bayes*

Syahrul Fauzi^{1,*}, Andika Prayoga Siswono², An-Nisa Firardiansyah Prayitno³

^{1,2,3}Informatika, Sistem Informasi, Institut Teknologi Telkom Purwokerto, Purwokerto, Indonesia
Email: ^{1,*}21103015@ittelkom-pwt.ac.id, ²21103012@ittelkom-pwt.ac.id ³21103043@ittelkom-pwt.ac.id
) Email Penulis Utama

Abstrak—Penelitian ini menganalisis sentimen publik terhadap desain jersey baru Tim Nasional (Timnas) Indonesia menggunakan metode klasifikasi *Naive Bayes* pada platform *Twitter*. Latar belakang penelitian adalah ketidakpuasan penggemar terhadap desain *jersey* yang tidak sesuai dengan ekspektasi, serta desainer yang tidak menerima kritik atau masukan dari berbagai pihak. Data diambil melalui proses *crawling* dari *Twitter* menggunakan kata kunci "timnas Indonesia," menghasilkan 1229 *tweet* yang dikategorikan sebagai *positif*, *negatif*, atau *netral*. Proses *preprocessing* mencakup *cleansing*, *case folding*, *tokenization*, *stopword removal*, *stemming*, dan TF-IDF. Setelah *preprocessing*, data diberi label sentimen. Hasil klasifikasi menggunakan *Naive Bayes* menunjukkan akurasi sebesar 99%, presisi dan *recall* mencapai nilai maksimum 1.00 untuk sebagian besar kelas, dengan *F1-score* 1.00 untuk kelas 1 dan 2, serta 0.98 untuk kelas 2 dengan sebagian besar *tweet* bersentimen *netral*, disusul oleh sentimen *positif* dan *negatif*. Visualisasi data melalui *word cloud* dan grafik batang menunjukkan topik utama yang dibicarakan terkait *jersey*, seperti "*jersey*," "*timnas*," "*Indonesia*," "*harga*," dan "*beli*." Penelitian ini menyimpulkan bahwa jersey Erspo Timnas Indonesia secara umum diterima dengan baik di komunitas *online*, dengan mayoritas sentimen *netral* dan sebagian kecil *positif*. Hal ini menunjukkan bahwa meskipun sebagian besar respon bersifat *netral*, *jersey* tersebut tetap tidak menimbulkan sentimen *negatif* yang signifikan, sehingga secara keseluruhan dianggap diterima dengan baik. Model *Naive Bayes* terbukti efektif dalam mengklasifikasikan sentimen dari *tweet*, memberikan wawasan yang berharga untuk peningkatan produk dan layanan terkait jersey Timnas Indonesia.

Kata Kunci: Sentimen, Timnas, Naive Bayes, Klasifikasi

Abstract—This research analyzes public sentiment towards the new jersey design of the Indonesian National Team (Timnas) using the Naive Bayes classification method on the Twitter platform. The background to the research is fans' dissatisfaction with jersey designs that do not meet expectations, as well as designers who do not accept criticism or input from various parties. Data was taken through a crawling process from Twitter using the keyword "Indonesian national team", resulting in 1229 tweets that were rated positive, negative or neutral. The preprocessing process includes cleaning, case folding, tokenization, stopword removal, stemming, and TF-IDF. After preprocessing, the data is given a sentiment label. Classification results using Naive Bayes show an accuracy of 99%, with the majority of tweets having a neutral sentiment, followed by positive and negative sentiments. Data visualization through word clouds and graphs shows the main topics discussed regarding jerseys, such as "jersey", "national team", "Indonesia", "price", and "buy". This research concludes that the Indonesian National Team's Erspo jersey is generally well received in the online community, with most sentiments being neutral and a small portion being positive. The Naive Bayes model proved effective in classifying sentiment from tweets, providing valuable insights for improving products and services related to the Indonesian National Team jersey.

Keywords: *Sentiment, National Teams, Naive Bayes, Classification*

1. PENDAHULUAN

Twitter telah menjadi salah satu platform media sosial paling populer di dunia, termasuk di Indonesia. Platform ini memungkinkan pengguna untuk berbagi pandangan, perasaan, dan opini dalam bentuk *tweet* yang singkat dan padat [1]. Penggunaan *twitter* di Indonesia menduduki peringkat ke-6, dari survei yang telah dilakukan oleh Cindy Mutiara Annur. *Twitter* selain menjadi media sosial juga telah menjadi sumber data yang kaya untuk berbagai analisis karena kecepatan dan kemudahan dalam berbagi informasi.

Penelitian ini dibuat karena desain *jersey* Tim Nasional (Timnas) Indonesia tidak sesuai dengan ekspektasi yang diharapkan oleh penggemar. Selain itu, desainer yang bertanggung jawab atas pembuatan *jersey* tersebut tidak menerima kritik atau masukan dari berbagai pihak yang mengakibatkan ketidakpuasan lebih lanjut. Informasi mengenai ketidakpuasan ini diambil dari sumber berita yang diterbitkan oleh Kompas, yang melaporkan adanya keluhan dan kritik dari penggemar terkait desain jersey yang tidak memenuhi harapan mereka. Hal ini menimbulkan berbagai permasalahan, termasuk penurunan minat terhadap *jersey* Timnas dan berpotensi terjadi penurunan penjualan. Lebih jauh lagi, ketidakpuasan terhadap desain *jersey* ini dapat mencerminkan masalah yang lebih dalam mengenai bagaimana aspirasi dan harapan penggemar terhadap tim nasional Indonesia tidak terpenuhi. Topik ini dipilih karena relevansinya dengan isu-isu yang sedang dihadapi oleh Timnas Indonesia dan pengemarnya. Penelitian ini juga berfokus pada peluncuran *jersey* terbaru yang dilaksanakan pada tanggal 18 Maret 2024 oleh Erspo, yang memperlihatkan desain yang kontroversial dan menjadi sorotan utama di kalangan penggemar. Dengan memanfaatkan platform *Twitter*, yang merupakan salah satu media sosial paling banyak

digunakan oleh masyarakat Indonesia, penelitian ini bertujuan untuk mengumpulkan data sentimen secara *real-time* dan menganalisis opini publik terhadap desain dari *jersey* baru tim nasional Indonesia.

Analisis sentimen adalah proses mengidentifikasi dan mengklasifikasikan opini yang diekspresikan dalam bentuk teks, terutama untuk menentukan apakah sikap penulis terhadap topik tertentu positif, negatif, atau netral[1]. Dalam konteks ini, analisis sentimen dapat digunakan untuk memahami bagaimana masyarakat Indonesia terutama penggemar tim nasional sepak bola Indonesia, menanggapi *jersey* keluaran terbaru yang dibuat oleh Erspo. Hasil dari analisis sentimen dapat dijadikan acuan dalam peningkatan suatu layanan ataupun kualitas produk[2].

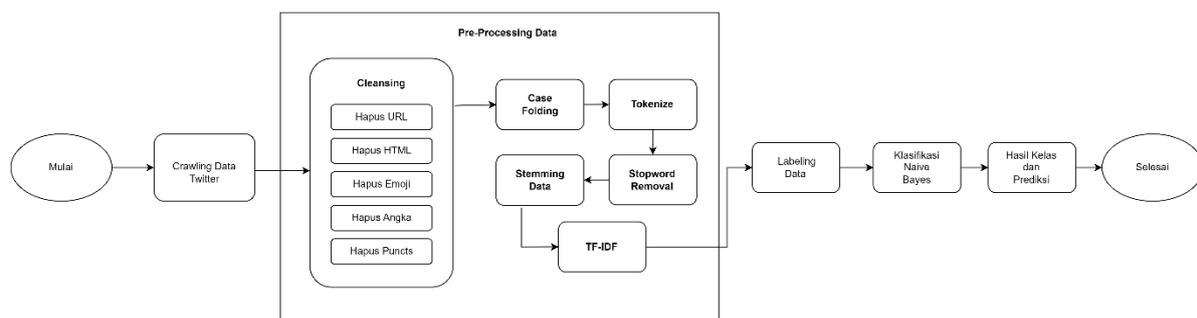
Penelitian sebelumnya dilakukan oleh Tanthy Tawaqila Widowati, dan Mujiono Sadikin membahas tentang analisis sentimen terhadap tokoh publik dengan klasifikasi *Naive Bayes* dan *Support Vector Machine* (SVM) memperoleh hasil sentimen positif sebesar 69,72% dan sentimen negatif sebesar 30,28% dengan akurasi *Naive Bayes* sebesar 91,48% sedangkan SVM sebesar 85,47%. Hasil tersebut mengindikasikan bahwa analisis sentimen terhadap tokoh publik cenderung positif[1]. Penelitian yang dilakukan oleh Elisa Febriyanti, dan Hery Febuariyanti membahas tentang analisis sentimen menggunakan algoritma *Naive Bayes* dan model klasifikasi dengan validasi K-Fold terhadap program kampus merdeka memperoleh hasil sentimen positif sebanyak 272 dan sentimen negatif 229 dengan akurasi 57% menggunakan algoritma *Naive Bayes* dan hasil yang diperoleh dengan model klasifikasi mendapatkan akurasi sebesar 60% yang menandakan bahwa model klasifikasi memperoleh hasil akurasi lebih tinggi[2]. Penelitian yang dilakukan oleh Ade Tiara Susilawati, Nur Anjeni Lestari, dan Puput Alpria Nina membahas tentang analisis sentimen terhadap boikot produk Israel menggunakan metode *Naive Bayes* memperoleh hasil akurasi sebesar 95% menunjukkan tingkat keberhasilan yang tinggi dalam mengidentifikasi sentimen dari *twitter* dengan kesimpulan mayoritas masyarakat mendukung boikot produk Israel[3].

Penelitian ini bertujuan untuk memahami opini dan pendapat masyarakat terhadap *jersey* baru Timnas Indonesia menggunakan analisis sentimen dengan model klasifikasi *Naive Bayes* pada platform *Twitter*. Metode *Naive Bayes* dipilih karena memiliki beberapa keunggulan diantaranya sederhana, cepat dan memiliki akurasi yang tinggi[4]. Platform *twitter* pada penelitian ini dimanfaatkan sebagai sumber data. Data yang telah didapat nantinya akan diklasifikasi ke dalam kategori *positif*, *negatif*, dan *netral*. Proses pengumpulan data dilakukan dengan cara *crawling* untuk membantu mendapatkan data secara *real-time* dari *Twitter*.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Dalam melakukan analisis sentimen dan mengetahui tingkat akurasinya, ada beberapa tahapan yang dilakukan untuk mendapatkan hasil yang optimal. Tahap pertama melakukan *crawling* data dilanjutkan dengan *preprocessing* data, *labeling* data, klasifikasi dengan *Naive Bayes* dan diakhiri dengan hasil kelas dan prediksi.



Gambar 1. Tahapan Penelitian

2.1.1 Pengumpulan Data

Tahap pertama dalam penelitian ini adalah pengumpulan data, proses pengumpulan data dilakukan dengan cara *crawling* menggunakan *tool* Google Collab. *Crawling* dilakukan dengan menggunakan *auth token* yang disediakan oleh *Twitter*. Kata kunci yang digunakan adalah “timnas Indonesia” untuk mengumpulkan *tweet* yang berhubungan dengan Timnas Indonesia. Data yang berhasil diperoleh dari proses *crawling* ini berjumlah 1229 data yang mencakup *tweet* positif, negatif, dan netral. Pengambilan data dilakukan dari bulan Maret 2024, setelah peluncuran *jersey* pada tanggal 18 Maret, hingga bulan Juni 2024. Data yang diperoleh, kemudian akan diolah lebih lanjut melalui tahap *preprocessing* untuk membersihkan data agar siap untuk digunakan.

2.1.2 Preprocessing

Preprocessing adalah proses mengubah data mentah menjadi data yang sesuai dengan prosedur *mining* yang dilakukan dan merupakan tahap yang penting dalam data *mining*[5]. *Preprocessing* dilakukan untuk membersihkan data dari *noise* yang mengganggu sehingga proses analisis dan penghitungan bobotnya jadi lebih mudah, dengan dilakukannya *preprocessing* diharapkan hasil dari klasifikasi akan lebih akurat nantinya[6]. Langkah-langkah yang digunakan pada tahap ini meliputi beberapa cara diantaranya:

- a. *Cleansing*
Cleansing adalah proses dimana karakter dan tanda baca yang tidak dibutuhkan akan dihapus dari teks. Tujuan dari *cleansing* adalah untuk membersihkan data dari *noise* yang mengganggu pada data set. Pembersihan data pada proses ini meliputi penghapusan *html*, *url*, angka, emoji, *da puncts* dari dalam teks[7]. Setelah proses *cleansing* data yang dihasilkan lebih bersih, relevan, dan siap untuk tahap selanjutnya.
- b. *Case Folding*
Case folding adalah proses mengubah setiap kata dalam dataset menjadi huruf kecil menggunakan fungsi *lowercase*[7].
- c. *Tokenize*
Tokenize berfungsi untuk membagi suatu kalimat ke dalam satu kata, sehingga menjadi kata yang selanjutnya dapat dianalisis[8]. Berikut contoh dari proses *tokenize* seperti “jersey timnas” menjadi “jersey”, “timnas”.
- d. *Stopword Removal*
Stopword removal adalah proses menghilangkan kata yang kurang bermakna sesuai dengan kata yang terdapat dalam daftar *stopword*[8]. Contoh kata yang kurang bermakna seperti “di”, “yang”.
- e. *Stemming Data*
Stemming adalah proses menghapus imbuhan dalam kata, diawal ataupun diakhir[8]. Dengan kata lain *stemming* merupakan proses mengubah kata dalam teks menjadi bentuk kata dasar seperti “membeli” menjadi “beli”.
- f. *Term Frequency – Inverse Document Frequency (TF-IDF)*
TF-IDF adalah proses yang digunakan untuk pembobotan kata dalam sebuah dokumen. Tujuan TF-IDF adalah untuk menentukan kata-kata yang paling signifikan dalam sebuah dokumen[9].

2.1.3 Labelling Data

Pada tahap selanjutnya dilakukan proses *labeling*, *labeling* sendiri merupakan proses membagi data menjadi tiga bagian *sentiment* seperti positif, netral, negatif. Berikut contoh dari dataset yang sudah diberi label:

Tabel 1. Jenis Sentimen

<i>Text</i>	<i>Sentiment</i>
<i>feel free</i> mo pake aja	Positif
cek <i>jersey</i> timnas	Netral
mending beli <i>jersey</i> lain	Negatif

2.1.4 Naive Bayes

Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat dalam teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan bentuk model probabilistik dan statistik yang disederhanakan dengan berdasar pada teorema Bayes dengan asumsi bahwa setiap atribut bersifat bebas (*independence*)[5]. Metode *Naive Bayes* digunakan untuk membuat keputusan berdasarkan hasil klasifikasi yang telah diperoleh sebelumnya[10]. Dalam penelitian ini metode *Naive Bayes* dimanfaatkan untuk menentukan sentimen dari *jersey* Ersपो timnas Indonesia.

3. HASIL DAN PEMBAHASAN

3.1 Pengolahan Data

Proses *crawling* dan pengolahan data menggunakan *Google Colab* dengan menggunakan bahasa *Phyton*. Data yang diperoleh dari proses *crawling* berjumlah 1229 data, berikut adalah gambar dari data yang telah diperoleh:

conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name	lang	location	quote_count	reply_count	retweet_count		
0	1.793900e+18	Fri May 24 06:54:07 +0000 2024	Pertama kali beli jersey pabrikan Errea yang n...	1.793900e+18			NaN	art	Indonesia	0	0	0	https
1	1.793900e+18	Fri May 24 06:42:16 +0000 2024	Pertama kali beli jersey pabrikan Errea yang n...	1.793900e+18	https://pbs.twimg.com/media/GOUyO0BfWAAEWVW.jpg		NaN	in	Sakarya, Turtiye	1	3	1	https://x
2	1.793890e+18	Fri May 24 06:38:05 +0000 2024	Tidak di 3 ini sih kabarnya kan cuma kabar tun...	1.793890e+18			NaN	in	Indonesia	1	3	0	https
3	1.793890e+18	Fri May 24 06:33:14 +0000 2024	Tapi isunya persib psis trus ada beberapa lagi...	1.793890e+18			NaN	in	Indonesia	2	4	2	https
4	1.793890e+18	Fri May 24 06:28:05 +0000 2024	Full interview Nunez dengan Por la Jersey: Aku...	1.793890e+18	https://pbs.twimg.com/ext_tw_video_thumb/17933...		NaN	in	Indonesia	0	0	0	https://
...
1224	1.784840e+18	Mon Apr 29 06:52:52 +0000 2024	Seumur hidup ga pernah punya jersey timnas bol...	1.784840e+18			NaN	in	NaN	0	1	0	https://x.c
1225	1.784810e+18	Mon Apr 29 06:43:16 +0000 2024	@MafiaWasit Info sewa jersey timnas buat nobar...	1.784840e+18	https://pbs.twimg.com/media/GMUuCuFabGAAWZGU.jpg		MafiaWasit	in	Indonesia	0	0	0	https://x.c
1226	1.784830e+18	Mon Apr 29 06:27:24 +0000 2024	adakah yang punya info beli jersey timnas ersp...	1.784830e+18			NaN	in	DKI JAKARTA	0	2	0	https://

Gambar 2. Hasil Crawling

Data yang telah dikumpulkan akan melalui tahap pengolahan awal yang dikenal sebagai *preprocessing*. Proses ini melibatkan enam langkah utama: pertama, *cleansing* untuk membersihkan data dari elemen yang tidak relevan atau gangguan; kedua, *case folding* untuk menyamakan format huruf, biasanya dengan mengubah semua teks menjadi huruf kecil; ketiga, *tokenization* untuk memecah teks menjadi unit-unit yang lebih kecil seperti kata atau frasa; keempat, *stopword removal* untuk menghapus kata-kata umum yang tidak memberikan informasi signifikan dalam analisis; dan kelima, *stemming* untuk mengurangi kata-kata ke bentuk dasarnya. Proses ini bertujuan untuk mempersiapkan data agar siap untuk analisis lebih lanjut.

a. *Cleaning*

Tahapan ini dilakukan proses pembersihan teks dari karakter dan tanda baca seperti *url*, angka, dan emoji.

	full_text	cleansing
0	Pertama kali beli jersey pabrikan Errea yang n...	Pertama kali beli jersey pabrikan Errea yang n...
2	Tidak di 3 ini sih kabarnya kan cuma kabar tun...	Tidak di ini sih kabarnya kan cuma kabar tung...
3	Tapi isunya persib psis trus ada beberapa lagi...	Tapi isunya persib psis trus ada beberapa lagi...
4	Full interview Nunez dengan Por la Jersey: Aku...	Full interview Nunez dengan Por la Jersey Aku ...
5	Buat yang mau Jersey Timnas Supporter Version ...	Buat yang mau Jersey Timnas Supporter Version ...
...
1224	Seumur hidup ga pernah punya jersey timnas bol...	Seumur hidup ga pernah punya jersey timnas bol...
1225	@MafiaWasit Info sewa jersey timnas buat nobar...	MafiaWasit Info sewa jersey timnas buat nobar ...
1226	adakah yang punya info beli jersey timnas ersp...	adakah yang punya info beli jersey timnas ersp...
1227	want to buy / wtb / dicari jersey timnas 2022 ...	want to buy wtb dicari jersey timnas home p...
1228	Jadi pengen beli jersey timnas Indonesia for t...	Jadi pengen beli jersey timnas Indonesia for t...

Gambar 3. Hasil *Cleaning*

b. *Case Folding*

Tahap ini dilakukan untuk mengubah setiap kata dalam teks menjadi huruf kecil semua. Berikut adalah hasil dari proses *case folding*.

	full_text	case_folding
0	Pertama kali beli jersey pabrikan Errea yang n...	pertama kali beli jersey pabrikan errea yang n...
2	Tidak di 3 ini sih kabarnya kan cuma kabar tun...	tidak di ini sih kabarnya kan cuma kabar tung...
3	Tapi isunya persib psis trus ada beberapa lagi...	tapi isunya persib psis trus ada beberapa lagi...
4	Full interview Nunez dengan Por la Jersey: Aku...	full interview nunez dengan por la jersey aku ...
5	Buat yang mau Jersey Timnas Supporter Version ...	buat yang mau jersey timnas supporter version ...
...
1224	Seumur hidup ga pernah punya jersey timnas bol...	seumur hidup ga pernah punya jersey timnas bol...
1225	@MafiaWasit Info sewa jersey timnas buat nobar...	mafiawasit info sewa jersey timnas buat nobar ...
1226	adakah yang punya info beli jersey timnas ersp...	adakah yang punya info beli jersey timnas ersp...
1227	want to buy / wtb / dicari jersey timnas 2022 ...	want to buy wtb dicari jersey timnas home p...
1228	Jadi pengen beli jersey timnas Indonesia for t...	jadi pengen beli jersey timnas indonesia for t...

1224 rows x 2 columns

Gambar 4. Hasil Case Folding

c. *Tokenization*

Tahap *Tokenization* mengubah setiap kalimat dalam dataset menjadi bentuk per kata, berikut hasil dari *tokenization*.

	full_text	tokenize
0	Pertama kali beli jersey pabrikan Errea yang n...	[pertama, kali, beli, jersey, pabrikan, errea,...
2	Tidak di 3 ini sih kabarnya kan cuma kabar tun...	[tidak, di, ini, sih, kabarnya, kan, cuma, kab...
3	Tapi isunya persib psis trus ada beberapa lagi...	[tapi, isunya, persib, psis, trus, ada, bebera...
4	Full interview Nunez dengan Por la Jersey: Aku...	[full, interview, nunez, dengan, por, la, jers...
5	Buat yang mau Jersey Timnas Supporter Version ...	[buat, yang, mau, jersey, timnas, supporter, v...
...
1224	Seumur hidup ga pernah punya jersey timnas bol...	[seumur, hidup, ga, pernah, punya, jersey, tim...
1225	@MafiaWasit Info sewa jersey timnas buat nobar...	[mafiawasit, info, sewa, jersey, timnas, buat...
1226	adakah yang punya info beli jersey timnas ersp...	[adakah, yang, punya, info, beli, jersey, timn...
1227	want to buy / wtb / dicari jersey timnas 2022 ...	[want, to, buy, wtb, dicari, jersey, timnas, h...
1228	Jadi pengen beli jersey timnas Indonesia for t...	[jadi, pengen, beli, jersey, timnas, indonesia...

Gambar 5. Hasil Tokenization

d. *Stopword Removal*

Proses ini menghapus kata yang kurang bermakna atau tidak dibutuhkan dalam proses analisis. Berikut hasil dari proses *stopword removal*.

	full_text	Filtering/stopword removal
0	Pertama kali beli jersey pabrikan Errea yang n...	[kali, beli, jersey, pabrikan, errea, ngegarap...
2	Tidak di 3 ini sih kabarnya kan cuma kabar tun...	[sih, kabarnya, kabar, tunggu, aja, juli, kaya...
3	Tapi isunya persib psis trus ada beberapa lagi...	[isunya, persib, psis, trus, pakai, apparel, m...
4	Full interview Nunez dengan Por la Jersey: Aku...	[full, interview, nunez, por, la, jersey, ruma...
5	Buat yang mau Jersey Timnas Supporter Version ...	[jersey, timnas, supporter, version, ori, home...
...
1224	Seumur hidup ga pernah punya jersey timnas bol...	[seumur, hidup, ga, jersey, timnas, bola, indo...
1225	@MafiaWasit Info sewa jersey timnas buat nobar...	[mafiawasit, info, sewa, jersey, timnas, nobar...
1226	adakah yang punya info beli jersey timnas ersp...	[adakah, info, beli, jersey, timnas, erspo, di...
1227	want to buy / wtb / dicari jersey timnas 2022 ...	[want, to, buy, wtb, dicari, jersey, timnas, h...
1228	Jadi pengen beli jersey timnas Indonesia for t...	[pengen, beli, jersey, timnas, indonesia, for...

Gambar 6. Hasil Stopword Removal

e. *Stemming*

Proses ini dilakukan untuk menghilangkan imbuhan yang terdapat di awal dan akhir kata, sehingga hanya menyisakan bentuk dasar dari kata tersebut. Berikut adalah hasil dari tahap *stemming* yang telah dilakukan.

	full_text	stemming_data
0	Pertama kali beli jersey pabrikan Errea yang n...	kali beli jersey pabrikan errea ngegarap jerse...
2	Tidak di 3 ini sih kabarnya kan cuma kabar tun...	sih kabar kabar tunggu aja juli kaya udah dium...
3	Tapi isunya persib psis trus ada beberapa lagi...	isu persib psis trus pakai apparel musim
4	Full interview Nunez dengan Por la Jersey: Aku...	full interview nunez por la jersey rumah timna...
5	Buat yang mau Jersey Timnas Supporter Version ...	jersey timnas supporter version ori home away ...
...
1224	Seumur hidup ga pernah punya jersey timnas bol...	umur hidup ga jersey timnas bola indonesia kal...
1225	@MafiaWasit Info sewa jersey timnas buat nobar...	mafiawasit info sewa jersey timnas nobar malam...
1226	adakah yang punya info beli jersey timnas ersp...	ada info beli jersey timnas erspo mana
1227	want to buy / wtb / dicari jersey timnas 2022 ...	want to buy wtb cari jersey timnas home player...
1228	Jadi pengen beli jersey timnas Indonesia for t...	ken beli jersey timnas indonesia for the first...

Gambar 7. Hasil Stemming

f. TF-IDF

TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah proses pembobotan kata. TF-IDF mengukur seberapa penting sebuah kata dalam sebuah dokumen. Berikut hasil dari proses TF-IDF

	term	idf	tf0	tf1	tf2	tf3	tf4	tf5	tf6	tf7	...	tf1214	tf1215	tf1216	tf1217	tf1218	tf1219	tf1220	tf1221	tf1222	tf1223
0	aaaaaaaaa	6.416732	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
1	aamin	6.011267	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
2	aamiin	6.416732	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
3	abal	6.416732	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
4	abang	6.011267	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
5	abangan	6.011267	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
6	abdi	6.416732	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
7	abdnraa	6.416732	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.524144	...	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
8	aberdeen	6.416732	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
9	abi	6.416732	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
10	abis	4.807294	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
11	aboutmyfeelin	6.416732	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
12	abraham	6.416732	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
13	ac	6.416732	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0

Gambar 8. Hasil TF-IDF

Gambar 8. memperlihatkan hasil dari proses TF-IDF yang menghasilkan nilai idf terbesar yaitu 6.416732.

3.2 Labeling Data

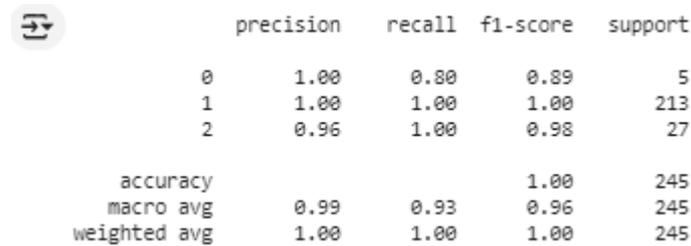
Setelah proses *preprocessing* dilanjutkan dengan proses labeling pada dataset. Proses *labeling* dilakukan untuk mengidentifikasi data berdasarkan sentimennya seperti positif, negatif, dan netral. hasil dari proses labeling dapat dilihat pada Gambar 9.

	stemming_data	sentiment_score	sentiment
0	kali beli jersey pabrikan errea ngegarap jerse...	0.00	netral
1	sih kabar kabar tunggu aja juli kaya udah dium...	0.00	netral
2	isu persib psis trus pakai apparel musim	0.00	netral
3	full interview nunez por la jersey rumah timna...	0.35	positif
4	jersey timnas supporter version ori home away ...	0.00	netral
...
1219	umur hidup ga jersey timnas bola indonesia kal...	0.00	netral
1220	mafiawasit info sewa jersey timnas nobar malam...	0.00	netral
1221	ada info beli jersey timnas erspo mana	0.00	netral
1222	want to buy wtb cari jersey timnas home player...	0.00	netral
1223	ken beli jersey timnas indonesia for the first...	0.25	positif

Gambar 9. Hasil Labeling Data

3.3 Pemodelan *Naive Bayes*

Klasifikasi *Naive Bayes* menggunakan kata yang paling sering muncul dalam dataset [11]. Data yang diklasifikasikan adalah data hasil dari proses *preprocessing* yang sudah di labeling sesuai dengan sentimennya. Hasil dari proses klasifikasi *Naive Bayes* adalah sebagai berikut:



	precision	recall	f1-score	support
0	1.00	0.80	0.89	5
1	1.00	1.00	1.00	213
2	0.96	1.00	0.98	27
accuracy			1.00	245
macro avg	0.99	0.93	0.96	245
weighted avg	1.00	1.00	1.00	245

Gambar 10. Hasil Klasifikasi

Gambar 10 menunjukkan hasil dari proses klasifikasi *Naive Bayes* yang mendapatkan hasil sebesar 5 untuk sentimen negatif, 213 untuk sentimen netral, dan 27 untuk sentimen positif, dengan memperoleh *precision* 1.00 pada sentimen negatif, 1.00 pada sentimen netral, 0.95 pada sentimen positif, *recall* sebesar 0.80 pada sentimen negatif, 1.00 pada sentimen netral, 1.00 pada sentimen positif, *f1-score* sebesar 0.89 pada sentimen negatif, 1.00 pada sentiment netral, dan 0.98 pada sentimen positif. berikut keterangan dari probabilitas hasil klasifikasi :

- Precision*
Mengukur tingkat keakuratan prediksi model positif, berapa persen yang benar-benar positif.
- Recall*
Mengukur kemampuan model untuk menemukan semua instance positif yang benar.
- F1-score*
Merupakan rata-rata harmonis dari *precision* dan *recall*. *F1-Score* memberikan gambaran keseimbangan antara *precision* dan *recall*.
- Support*
Memberikan gambaran tentang seberapa besar atau kecil kelas tersebut dalam dataset.
- Accuracy*
Merupakan metrik umum untuk mengevaluasi kinerja model secara keseluruhan.
- Macro Avg*
Merupakan rata-rata sederhana dari *precision*, *recall*, dan *F1-score* untuk semua kelas tanpa memperhatikan distribusi kelas.
- Weighted Avg*
Merupakan rata-rata sederhana dari *precision*, *recall*, dan *F1-score* untuk semua kelas tanpa dengan memperhitungkan support untuk semua kelas.

Tahap selanjutnya dari proses klasifikasi adalah melakukan *training* dan *testing* pada data. Data akan dibagi menjadi 2 bagian yaitu data untuk *training* sebesar 80% dan *testing* 20%. Tabel 2 menunjukkan pembagian data untuk training dan testing.

Tabel 2. Pembagian Data

Data Training (80%)	Data Testing (20%)
979	245

Dalam pembagian data untuk pelatihan dan pengujian model, data *training* mencakup 80% dari total dataset, yang berjumlah 979 sampel. Data ini digunakan untuk melatih model agar dapat mempelajari pola dan hubungan dalam data. Sementara itu, data *testing* mencakup sisa 20% dari dataset, yaitu sebanyak 245 sampel. Data *testing* ini digunakan untuk mengevaluasi kinerja model yang telah dilatih, memastikan bahwa model dapat menggeneralisasi dan membuat prediksi yang akurat pada data yang belum pernah dilihat sebelumnya. Pembagian ini bertujuan untuk memvalidasi efektivitas model dan menghindari *overfitting* pada data *training*.

```
⇒ x_train = 979
   x_test  = 245
   y_train = 979
   y_test  = 245
```

Gambar 11. Pembagian Data

Klasifikasi *Naive Bayes* yang digunakan pada penelitian ini menghasilkan tingkat akurasi sebesar 99%. hal ini menunjukkan bahwa model *Naive Bayes* memiliki keandalan dan efektivitas dalam mengklasifikasikan data pada penelitian ini. tingkat akurasi ini menandakan bahwa model tersebut mengenali pola dengan baik, sehingga hanya 1% saja dari prediksi yang dibuat oleh model adalah kesalahan.

```
⇒ Tingkat Akurasi: 99 persen
```

Gambar 12. Tingkat Akurasi

3.4 Confusion Matrix

Gambar 3.12 menunjukkan *confusion matrix* yang digunakan untuk mengukur kinerja model *naive bayes* dalam mengklasifikasikan data dalam 3 kelas berbeda yaitu kelas 0, kelas 1, kelas 2. berikut adalah penjelasan dari *confusion matrix* tersebut :

- a. Kelas 0

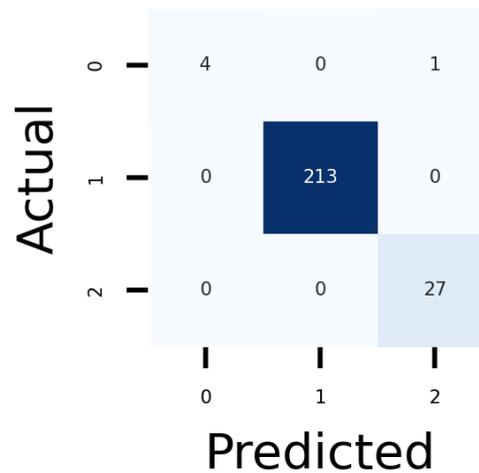
Untuk kelas 0, model berhasil mengidentifikasi 4 *instance* dengan benar, menunjukkan akurasi yang baik dalam mengenali data dari kelas ini. Namun, terdapat satu *instance* yang seharusnya termasuk dalam kelas 2, tetapi diprediksi salah sebagai kelas 0. Hal ini menunjukkan bahwa meskipun model sangat efektif, masih ada sedikit kemungkinan terjadinya kesalahan prediksi untuk kelas 0, tetapi tidak ada *instance* dari kelas lain yang salah diklasifikasikan sebagai kelas 0.
- b. Kelas 1

Dalam hal kelas 1, model mencapai hasil yang sangat memuaskan dengan 213 *instance* yang diklasifikasikan secara benar sebagai kelas 1. Tidak ada kesalahan prediksi yang terdeteksi untuk kelas ini, menunjukkan bahwa model sangat akurat dalam mengenali data kelas 1. Kelas ini memiliki jumlah *instance* terbesar, dan model menunjukkan performa terbaik dengan memprediksi semua *instance* kelas 1 dengan tepat.
- c. Kelas 2

Untuk kelas 2, model juga menunjukkan performa yang solid dengan 27 *instance* yang diprediksi benar sebagai kelas 2. Meski tidak sebesar jumlah *instance* kelas 1, model berhasil mengklasifikasikan semua data kelas 2 dengan akurat tanpa kesalahan prediksi yang signifikan.

Secara keseluruhan *confusion matrix* ini menunjukkan bahwa model *naive bayes* memiliki kinerja yang sangat baik dalam mengidentifikasi data, dengan hanya sedikit kesalahan prediksi. sebagian besar *instance* berhasil diklasifikasikan dengan benar sesuai kelasnya, terutama untuk kelas 1 yang memiliki jumlah *instance* terbesar dan seluruhnya diklasifikasikan dengan tepat.

Confusion Matrix



Gambar 13. Confusion Matrix

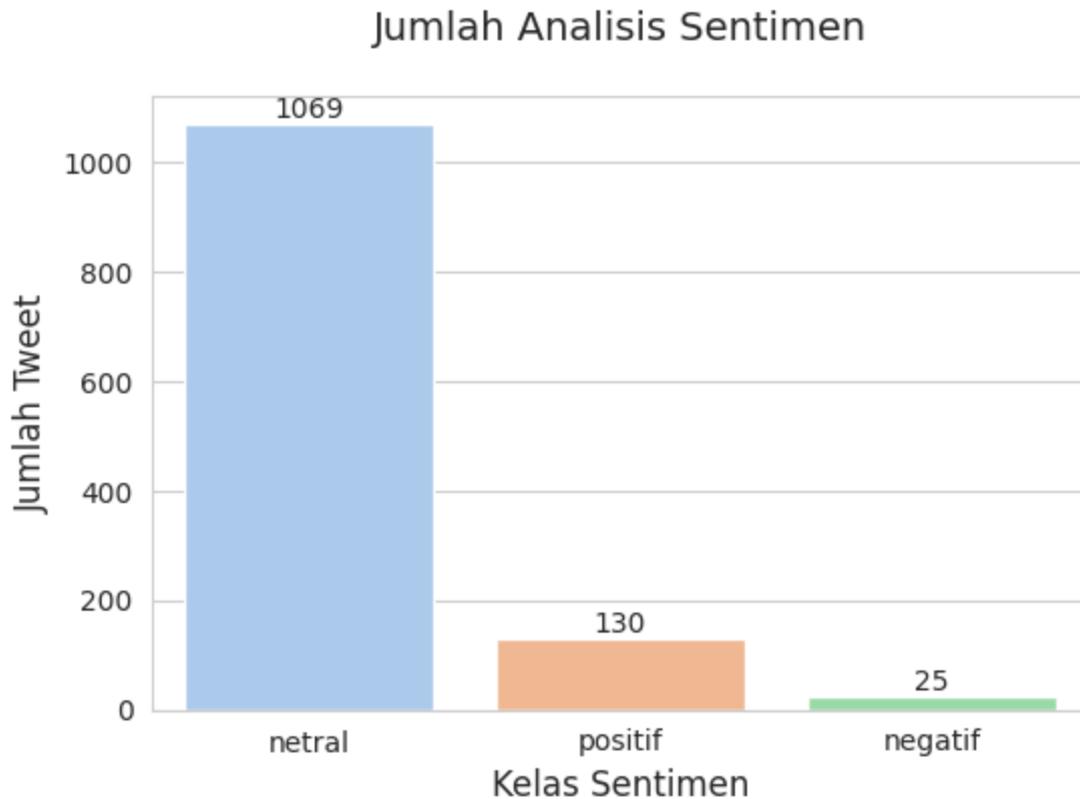
3.5 Visualisasi Data



Gambar 14. Visualisasi Data

Visualisasi data dalam bentuk *word cloud* tersebut menggambarkan frekuensi kemunculan kata-kata berkaitan dengan analisis sentimen jersey Erspo Timnas Indonesia. Beberapa kata yang sering muncul antara lain “jersey”, “timnas”, “indonesia”, “harga”, “beli”, dan “pake”. Kata-kata ini ditampilkan dengan ukuran *font* lebih besar dibandingkan kata-kata lain, menunjukkan bahwa mereka adalah topik utama dalam analisis sentimen jersey Erspo Timnas Indonesia. Selain kata-kata tersebut, terdapat kata lain yang muncul seperti “shopee”, “home”, “third”, “ori”, “grade”, dan “suporter” yang cukup sering disebutkan. Hal ini menunjukkan bahwa banyak pembicaraan mengenai tempat pembelian, jenis jersey home atau third, serta kualitas jersey. visualisasi ini memberikan gambaran umum tentang topik yang dibicarakan dan minat penggemar terkait jersey Timnas Indonesia.

3.6 Jumlah Analisis Sentimen



Gambar 15. Jumlah Analisis Sentimen

Gambar 3.14 merupakan sebuah grafik batang yang menunjukkan jumlah analisis sentimen jersey Erspo timnas Indonesia yang dibagi kedalam 3 kelompok kelas sentimen yaitu netral, positif, dan negatif. berikut penjelasan grafik batang tersebut :

- a. Netral
Terdapat 1069 *tweet* yang dikategorikan netral, hal ini menunjukkan bahwa mayoritas *tweet* tidak menunjukkan sentimen baik positif maupun negatif.
- b. Positif
Terdapat 130 *tweet* yang memiliki positif, menunjukkan adanya sebagian kecil *tweet* memberikan pandangan atau komentar positif tentang jersey Erspo timnas Indonesia.
- c. Negatif
Terdapat 25 sentimen negatif, menunjukkan bahwa hanya sedikit *tweet* yang memberikan pandangan atau komentar negatif tentang jersey Erspo timnas Indonesia.

Secara keseluruhan, grafik ini menunjukkan bahwa sebagian besar *tweet* tentang jersey Erspo timnas Indonesia bersifat netral, dengan proporsi *tweet* positif yang lebih tinggi dibanding *tweet* negatif, meskipun keduanya jauh lebih sedikit dibandingkan *tweet* netral.

4. KESIMPULAN

Berdasarkan penelitian yang dilakukan yaitu analisis sentimen terhadap jersey Erspo Timnas Indonesia menggunakan metode klasifikasi *Naive Bayes*, beberapa kesimpulan dapat ditarik. Pertama, proses pengolahan data dimulai dengan pengambilan data dari Twitter menggunakan proses *crawling* yang menghasilkan 1229 data. Data ini kemudian melalui proses *preprocessing* yang meliputi tahap *cleansing* untuk menghapus karakter dan tanda baca tidak penting, *case folding* untuk membuat semua teks menjadi huruf kecil, *tokenization* untuk memecah teks menjadi kata-kata, *stopword removal* untuk menghilangkan kata-kata tidak bermakna, *stemming* untuk menghapus imbuhan kata, dan TF-IDF untuk pembobotan kata. Tahapan ini membantu mempersiapkan

data agar siap untuk proses analisis lebih lanjut. Kemudian, dilakukan proses *labeling* untuk mengidentifikasi sentimen dari data, yaitu positif, negatif, dan netral. Hasil *labeling* menunjukkan mayoritas data bersentimen netral, dengan sebagian kecil bersentimen positif dan negatif. Proses klasifikasi menggunakan *Naive Bayes* menghasilkan akurasi sebesar 99%, menunjukkan bahwa model mampu mengenali pola dan mengklasifikasikan data dengan baik. Dalam evaluasi klasifikasi, nilai *precision*, *recall*, dan *F1-score* memberikan gambaran yang baik tentang kinerja model dalam mengklasifikasikan masing-masing sentimen. Sentimen netral memiliki nilai tertinggi dalam hal presisi dan *recall*, menunjukkan model dapat dengan baik mengenali dan membedakan sentimen netral dari lainnya. Sedangkan untuk sentimen positif, meskipun memiliki presisi yang tinggi, *recall* sedikit lebih rendah, menunjukkan potensi untuk meningkatkan pengenalan sentimen positif yang lebih baik. Visualisasi data, seperti *word cloud* dan grafik batang, memberikan pemahaman yang lebih mendalam tentang topik yang paling banyak dibicarakan dalam konteks jersey Erspo Timnas Indonesia di media sosial. Kata-kata seperti "jersey", "timnas", "Indonesia", "harga", dan "beli" mendominasi dalam percakapan, menunjukkan minat utama terhadap aspek-aspek tersebut dalam komunitas penggemar. Secara keseluruhan, penelitian ini menunjukkan bahwa jersey Erspo Timnas Indonesia secara umum diterima dengan baik dalam komunitas *online*, dengan mayoritas sentimen netral dan sebagian kecil positif. Hal ini mencerminkan bahwa produk ini memiliki daya tarik di kalangan penggemar, meskipun perlu perhatian lebih lanjut terhadap pengenalan sentimen positif yang lebih baik. Dengan memanfaatkan metode pengolahan data dan klasifikasi yang tepat, penelitian ini memberikan kontribusi dalam memahami dan menginterpretasi pandangan publik terhadap produk spesifik dalam konteks media sosial.

UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini .

REFERENCES

- [1] T. T. Widowati And M. Sadikin, "Analisis Sentimen Twitter Terhadap Tokoh Publik Dengan Algoritma Naive Bayes Dan Support Vector Machine," *Jurnal Simetris*, Vol. 11, No. 2, 2020, [Online]. Available: <https://t.co/xzf91zhk41>
- [2] E. Febriyani And H. Februariyanti, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Algoritma Naive Bayes Classifier Di Twitter," Vol. 17, No. 1.
- [3] A. Tiara Susilawati, A. H. Tiara Susilawati Universitas Muhammadiyah Kalimantan Timur Nur Anjeni Lestari Universitas Muhammadiyah Kalimantan Timur Puput Alpria Nina Universitas Muhammadiyah Kalimantan Timur Jl Ir Juanda No, K. Samarinda Ulu, K. Samarinda, And K. Timur, "Analisis Sentimen Publik Pada Twitter Terhadap Boikot Produk Israel Menggunakan Metode Naïve Bayes," *Jurnal Ilmiah Mahasiswa*, Vol. 2, No. 1, Pp. 26–35, 2024, Doi: 10.59603/Niantanasikka.V2i1.240.
- [4] E. I. Program, S. Sistem, I. A. Kampus, And K. Bogor, "Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes," Vol. Vii, No. 1, 2019.
- [5] M. I. Fikri, T. S. Sabrila, Y. Azhar, And U. M. Malang, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Pada Analisis Sentimen Twitter".
- [6] M. Rangga, A. Nasution, And M. Hayaty, "Perbandingan Akurasi Dan Waktu Proses Algoritma K-Nn Dan Svm Dalam Analisis Sentimen Twitter," *Jurnal Informatika*, Vol. 6, No. 2, Pp. 212–218, 2019, [Online]. Available: <http://ejournal.bsi.ac.id/ejournal/index.php/ji>
- [7] R. Shandy Arischo, "Jurnal Media Informatika Budidarma Analisis Sentimen Pinjaman Online Di Twitter Dengan Metode Naive Bayes Classifier Dan Svm," 2024, Doi: 10.30865/Mib.V8i2.7406.
- [8] Ade Dwi Dayani, Yuhandri, And G. Widi Nurcahyo, "Analisis Sentimen Terhadap Opini Publik Pada Sosial Media Twitter Menggunakan Metode Support Vector Machine," *Jurnal Komtekinfo*, Pp. 1–10, Mar. 2024, Doi: 10.35134/Komtekinfo.V11i1.439.
- [9] R. I. Agustin, "Komparasi Algoritma Naïve Bayes Dan Svm Untuk Analisis Sentimen Twitter Korupsi Bansos Beras Masa Pandemi," *Jurnal Informatika Dan Teknik Elektro Terapan*, Vol. 12, No. 2, Apr. 2024, Doi: 10.23960/Jitet.V12i2.4020.

- [10] W. Ningsih, B. Alfianda, R. Rahmadden, And D. Wulandari, “Perbandingan Algoritma Svm Dan Naïve Bayes Dalam Analisis Sentimen Twitter Pada Penggunaan Mobil Listrik Di Indonesia,” *Malcom: Indonesian Journal Of Machine Learning And Computer Science*, Vol. 4, No. 2, Pp. 556–562, Feb. 2024, Doi: 10.57152/Malcom.V4i2.1253.
- [11] W. A. Prabowo And C. Wiguna, “Analisis Sentimen Pembelajaran Daring Pada Twitter Di Masa Pandemi Covid-19 Menggunakan Metode Naïve Bayes,” *Jurnal Media Informatika Budidarma*, Vol. 5, No. 1, P. 149, Jan. 2021, Doi: 10.30865/Mib.V5i1.2604.