



Sentimen Analisis Masyarakat terhadap Kasus Penembakan Brigadir J Menggunakan Algoritma Naïve Bayes Classifier

Febby Apri Wenando^{*1}, regiolina hayami², Soni³, Ananda Fitria⁴, Deyola Fadwa Shifana⁵

Email: ¹febby.apri@it.unand.ac.id, ²regiolinahayami@umri.ac.id, ³soni@umri.ac.id,

⁴2011522014_ananda@student.unand.ac.id, ⁵2011522032_deyola@student.unand.ac.id

^{1, 4, 5}Sistem Informasi, Fakultas Teknologi Informasi, Universitas Andalas

^{2, 3}Teknik Informatika, Fakultas Ilmu Komputer, Universitas Muhammadiyah Riau

Diterima: 05 Agustus 2023 | Direvisi: - | Disetujui: 30 Agustus 2023

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

Analisis sentimen merupakan riset komputasional dari opini, sentimen dan emosi yang di ekspresikan secara tekstual dengan menggunakan metode pengelompokkan sehingga menghasilkan penilaian bernilai positif atau negatif. Proses analisis ini umumnya dimulai dengan pengumpulan data yang kemudian di olah melalui pendekatan machine learning. Salah satu teknik pengumpulan data tersebut yaitu menggunakan internet dan beragam platform media sosial lainnya. Salah satu jenis platform media sosial yang sangat berkembang saat ini adalah Twitter. Media sosial Twitter mempermudah masyarakat untuk bebas berpendapat melalui cuitan atau biasa disebut dengan tweets. Netizen dapat dengan bebas menyampaikan opini pribadinya untuk topik apapun, termasuk persepsi terhadap kasus kriminal yang terjadi di Indonesia. Salah satu kasus terbaru yang tengah menjadi topik perbincangan hangat saat ini adalah kasus pembunuhan Brigadir Joshua dengan tersangka yaitu seorang Irjen Polri yaitu Ferdy Sambo. Sehingga di dalam penelitian ini opini masyarakat yang terdapat pada platform Twitter tersebut dapat dimanfaatkan sebagai bahan analisis sentimen untuk mengetahui pendapat publik terhadap kasus Ferdy Sambo. Data yang digunakan terdiri dari 234 data tweet dengan persentase opini positive sebesar 51,50% dan negative sebesar 48,50% yang kemudian diklasifikasikan dengan *Algoritma Naive Bayes Classifier* dengan hasil yang didapat nilai f1-score sebesar 75%.

Kata kunci: *Sentiment Analysis, Naïve Bayes, Machine Learning, Twitter*

Analyzing Public Sentiment on the Brigadier J Shooting Case Using the Naïve Bayes Classifier Algorithm

Abstract

Sentiment analysis is computational research of textually expressed opinions, sentiments, and emotions using grouping methods to produce positive or negative assessments. This analysis process generally begins with data collection, which is then processed through a machine learning approach. One of the data collection techniques is using the internet and other social media platforms. One type of social media platform that is currently growing is Twitter. Twitter social media makes it easier for people to express opinions through tweets, commonly called tweets, freely. Netizens can freely express their opinions on any topic, including perceptions of criminal cases in Indonesia. One of the latest cases, which is currently a hot topic of conversation, is the murder case of Brigadier Joshua and the suspect, Inspector General of Police Ferdy Sambo. So, in this research, public opinion on the Twitter platform can be used as material for sentiment analysis to determine public opinion regarding the Ferdy Sambo case. The data used consisted of 234 tweets data with a positive opinion percentage of 51.50% and a negative opinion of 48.50%, which was then classified using the Naive Bayes Classifier Algorithm with the results obtained with an f1-score of 75%.

Keywords: *Sentiment Analysis, Naïve Bayes, Machine Learning, Twitter*

1. PENDAHULUAN

Perkembangan teknologi yang berlangsung dengan cepat telah memudahkan akses dan perolehan informasi, termasuk melalui penggunaan media sosial. Sebagian besar individu dari berbagai usia dan lokasi menghabiskan banyak waktu mereka menggunakan jaringan sosial sebagai platform untuk menyampaikan pendapat mereka, dan salah satunya adalah Twitter. Twitter, sebagai platform media sosial, sangat diminati oleh masyarakat di seluruh dunia [1], dengan catatan memiliki 313 juta pengguna aktif setiap bulannya pada tahun 2016. Pada tahun 2015, jumlah pengguna Twitter dari Indonesia hanya mencapai 72,7 juta. Oleh karena itu, dapat disimpulkan bahwa hampir 1 dari 5 pengguna Twitter di seluruh dunia berasal dari Indonesia.

Pada tahun 2022 diketahui bahwa Twitter merupakan salah satu media sosial paling populer di Indonesia dengan pengguna Twitter kelima teraktif di dunia [2] yakni 79% dari pengguna Twitter Indonesia merupakan pengguna aktif [3]. Lebih detailnya, berdasarkan laporan *We Are Social*, jumlah pengguna Twitter di Indonesia mencapai 18,45 juta pada tahun 2022 [4]. Cara kerja dalam penggunaan Twitter adalah membuat postingan yang disebut dengan tweet, dengan maksimal 140 karakter setiap postingannya. Melalui tweet, pengguna Twitter dapat saling berinteraksi dan berbagi pendapat atau opini. Masyarakat menyampaikan curhatan dan pendapat mereka di media sosial Twitter ini, mereka saling beradu argument terhadap masing-masing mereka bagi yang Pro dan yang Kontra terhadap kasus ini. Hasil dari tweet dan beradu argumen masyarakat ini dapat kita ambil pendapat mereka yang berupa teks, dari hasil teks ini diolah menggunakan pendekatan *machine learning*, Bentuk analisis pendapat atau review dapat dilakukan menggunakan teknik analisis sentimen, dimana proses ini dilakukan untuk mengekstrak, memahami, dan mengolah data tekstual secara otomatis menggunakan pendekatan machine learning untuk mendapatkan informasi baru yang bisa berguna untuk diinformasikan Kembali [5,6,7,8].

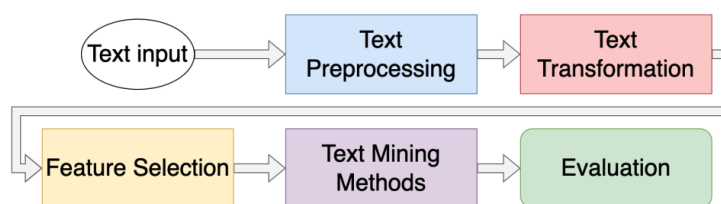
Salah satu contoh kasus yang baru saja terjadi tepatnya pada tanggal 8 Juli 2022 yang sangat menggemparkan negara Indonesia adalah terjadinya kasus pembunuhan yang dilakukan oleh salah seorang inspektur jenderal kepolisian yaitu Irjen Ferdy Sambo kepada bawahannya yaitu seorang brigadir kepolisian yaitu Almarhum Joshua Hutabarat. Permasalahan ini kian memanas ketika banyak terjadi konspirasi dan kebohongan selama berlangsungnya masa persidangan. Hal ini menjadi bahan perbincangan masyarakat umum sampai ke media sosial, ada yang Pro dan ada yang Kontra terkait kasus ini. Media sosial sekarang menjadi tempat bagi masyarakat untuk menyampaikan opini dan pendapat mereka masing-masing. Data yang ada di media sosial ini berbentuk teks, untuk mengekstrak informasi yang ada di media sosial ini digunakan pendekatan *machine learning*, data teks ini diolah dan diklasifikasikan dengan teknik *text mining*. Klasifikasi teks adalah proses mengelompokkan teks berdasarkan kata-kata, frasa, atau kombinasi dari keduanya untuk menentukan kategori yang telah ditetapkan sebelumnya. Pada teks yang berhubungan dengan sentimen, klasifikasi biasanya dilakukan untuk mengidentifikasi apakah teks tersebut mengandung sentimen positif atau negatif. Sebelum melakukan klasifikasi, data teks harus dipersiapkan melalui tahap pre-processing. Preprocessing teks adalah tahap yang sangat penting karena akan memengaruhi kualitas tahap klasifikasi selanjutnya [6,7]. Tahap preprocessing ini bertujuan untuk memilih kata-kata kunci yang signifikan dan menghilangkan kata-kata yang tidak berkontribusi dalam membedakan antara dokumen-dokumen. Tahap preprocessing juga bertujuan untuk mengubah data teks ke dalam struktur data yang cocok untuk analisis data. Tahap-tahap dalam preprocessing teks meliputi membersihkan teks, memberi label, menghapus stopwords, normalisasi, dan stemming [8].

Algoritma Naive Bayes Classifier adalah salah satu algoritma yang terintegrasi dalam data mining yang menonjol karena kemudahannya dalam penerapan serta kinerjanya yang sangat cepat [9]. Algoritma ini memungkinkan implementasi dengan struktur yang simpel, namun tetap memberikan hasil klasifikasi yang efektif. Metode klasifikasi Naive Bayes merupakan pendekatan statistik yang mampu memprediksi probabilitas keanggotaan suatu kategori dengan dasar Teorema Bayes. Keunggulan dari Naive Bayes Classifier terlihat dari kemampuannya yang sebanding dengan decision tree dan neural network, terutama ketika dihadapkan pada dataset besar, di mana algoritma ini terbukti memiliki akurasi tinggi dan waktu pemrosesan yang efisien [10].

Teorema Bayes, yang mendasari pengklasifikasi Bayes, pertama kali ditemukan oleh Thomas Bayes pada abad ke-18. Dalam penelitian perbandingan algoritma klasifikasi, Naive Bayes Classifier, yang juga dikenal sebagai simple Bayesian classifier, terbukti efektif dengan tingkat akurasi dan kecepatan yang tinggi ketika diaplikasikan pada dataset yang besar [11,12]. Oleh karena itu, algoritma ini sering digunakan dalam berbagai aplikasi, termasuk menganalisis opini dan sentimen publik di berbagai platform. Penelitian ini akan memanfaatkan metode Naive Bayes Classifier untuk mengklasifikasikan analisis sentimen terhadap tweet yang berhubungan dengan kasus penembakan Brigadir J yang memungkinkan pemahaman yang lebih dalam terhadap opini publik sehingga bisa menghasilkan informasi baru yang berguna.

2. Metodologi

Tahapan yang dilakukan pada penelitian ini dapat dilihat pada gambar 1 berikut :



Gambar 1. Flowchart Penelitian

2.1 Pengumpulan Data

Crawling atau pengumpulan data merupakan sebuah alat untuk mengindeks dan mengunduh konten dari internet, lalu disimpan ke dalam *database* mesin pencari [12]. Ada berbagai cara untuk melakukan pengumpulan data melalui media sosial Twitter, tentunya dengan kelebihan dan kekurangan masing-masing. Dari *twitter API*, *tweepy*, *TWINT*, dan *snsrape*, penulis memilih *tweepy* yang berbasis bahasa *python* untuk melakukan *crawling data* cuitan Twitter. Pada proses ini (dan proses-proses selanjutnya), data disimpan dalam file berformat CSV.

2.2 Tahap Preprocessing

Proses *preprocessing* data merupakan proses mengubah suatu dokumen yang tidak terstruktur menjadi data yang terstruktur untuk kebutuhan analisis lebih lanjut agar tidak mengurangi performanya sendiri karena pola data yang belum sesuai dengan sistem. Pada tahap *preprocessing* hanya menggunakan variabel data *tweet* opini.

2.2.1. Tahapan Cleaning

Data *cleansing* bisa disebut juga data *cleaning* atau data *scrubbing* [12]. *Cleansing* ini sendiri adalah tahap di mana karakter dan tanda baca yang tidak diperlukan dihapus dari teks [11]. *Cleansing* bertujuan untuk mengurangi gangguan/noise pada dataset. Tahap-tahap pada *cleansing* ini adalah Hapus URL, Hapus *Username*, hapus tanda RT dan hapus Hastag. Proses tersebut menggunakan regular expression (regex) dalam parameter operator RapidMiner. Regex yang digunakan adalah “http\S+|\S-co\S+”, “@S+”, “RT”, “#S+”, dan “#S+[-!#\$%&'()*+,-./:;<=>?@\[\]_`{|}~]””. Contoh karakter yang dihilangkan seperti URL, tag (#), username (@), RT, juga tanda baca seperti titik (.), koma (,) dan tanda baca lainnya

2.2.2. Tahapan Labeling

Dari data *crawling* cuitan Twitter, dipilih 233 data cuitan Twitter untuk merepresentasikan contoh dari data cuitan Twitter dengan sentimen positif 120 data dan negatif 113 data. 233 data ini kemudian yang akan menjadi set data pelatihan (yang sudah dilabeli) dan set data test (data sama yang tidak dilabeli). Untuk pelabelan ini kelas dari setiap data, penulis menggunakan kode angka dengan ketentuan: 1 sebagai sentimen negatif dan 0 sebagai sentimen positif. Pelabelan negatif (kode=1) berarti kalimat tersebut tidak mendukung Ferdy Sambo [13]. Sedangkan, pelabelan positif (kode=0) berarti kalimat tersebut mendukung kasus atau netral terhadap pihak Ferdy Sambo.

2.2.3. Tahapan Tokenisasi

Tokenisasi merupakan proses memecah kalimat menjadi kata-kata menjadi lebih berarti dan bermakna [14,15]. Tokenisasi memisah setiap kata yang menyusun kalimat. Misalkan “kementerian pendidikan dan kebudayaan”, setelah di tokenisasi menjadi “kementerian”, “pendidikan”, “dan”, “kebudayaan”.

2.2.4. Tahapan Stopword

Pada tahapan Stopword, menghapus kata yang dianggap tidak memberikan dampak terhadap informasi yang ada pada suatu data tweet. Pembersihan data stopwords dan pada penelitian ini menggunakan library python pySastrawi. Daftar kata stopwords pada penelitian bersumber dari penelitian [12,16,17]. Pada tahap ini, kumpulan tweet melalui tahap stopwords removal. Setiap kata pada tweet diperiksa. Apabila terdapat kata sambung, kata depan, kata ganti, atau kata yang tidak ada hubungannya dalam analisis sentimen, maka kata tersebut akan dihilangkan. Kata-kata yang terkandung pada daftar stopwords yang terdapat pada daftar kata stopwords Bahasa Indonesia berisi kata-kata yang sering muncul namun tidak memiliki arti. Pada contoh kata “bagaimana”, “juga”, “agar” dan “jadi” terdapat di tabel kata stopwords sehingga kata tersebut harus dihilangkan.

2.2.5. Tahapan Normalisasi

Tahap normalisasi merupakan tahapan pra-pemrosesan yang dilakukan sebagai penyesuaian terhadap kemunculan nilai kontinu dalam fitur dataset yang akan mempengaruhi hasil proses klasifikasi dengan Naïve Bayes.

2.2.6. Tahapan *Steaming*

Stemming merupakan tahapan untuk melakukan proses mengubah kata yang berimbun infix maupun suffix menjadi sebuah kata dasar yang akan lebih mengandung sebuah makna untuk peroleh suatu informasi sehingga komentar akan menjadi lebih spesifik dalam pengkategorian [15]. Contoh stem list yang digunakan adalah: menepatkan; tepat; penyogokan; sogok; ketakutan; takut; menyepakati; sepakat; kesukarelaan; sukarela; strategisnya; strategis; pembetulan; betu; menjengkelkan ; jengkel; dll.

2.3 Tahap Pemodelan

Pada tahapan ini menerapkan metode *Naïve Bayes Classifier* sebagai klasifikasi pada analisis sentimen. *Naïve Bayes Classifier* termasuk dalam teknik *data mining supervised learning* yang merupakan pendekatan terarah, datasetnya sudah memiliki *label* (*class*).

2.4. Klasifikasi

Dalam metode *Naïve Bayes Classifier* dilakukan proses pengklasifikasian teks berdasarkan data latih yang sebelumnya telah ada didalam database sistem. Penelitian terkait naive bayes classifiers melalui beberapa tahap yakni prior probability, conditional probability, pemilihan kategori akan ditentukan melalui nilai maksimum kategori terpilih.

2.5. Evaluasi

Evaluasi performansi dilakukan untuk menguji hasil klasifikasi dengan mengukur nilai kebenaran dari sistem. Parameter yang digunakan untuk mengukur nilai kebenaran yaitu accuracy. Accuracy adalah persentase dokumen yang berhasil diklasifikasikan dengan tepat oleh sistem. Perhitungan accuracy memanfaatkan metode confusion matrix. Perhitungan accuracy menggunakan persamaan 1 sebagai berikut [14,15] :

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (1)$$

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Data yang berhasil diperoleh adalah sebanyak 233 data. Contoh data tweet yang berhasil dikumpulkan pada dapat dilihat pada Gambar 1.

	Tweets
1	Diungkap Kuat Maruf di dalam ruang sidang PN J...
2	Sidang lanjutan obstruction of justice terkait...
3	Ferdy Sambo: Saya Tak Tahu Bagaimana Harus Men...
4	Terdakwa kasus pembunuhan Brigadir Nofriansyah...
5	Kemesraan wagub Emil dengan sang istri Arumi B...
...	...
96	Tidak Ada Jejak DNA Ferdy Sambo di Senjata HS...
97	Pengakuannya Terindikasi Bohong, Ferdy Sambo E...
98	DNA Sambo Disebut Tak Ada di Senjata Glock-17 ...
99	Pengacara Richard Eliezer, mengaitkan keterang...
100	Ferdy Sambo Kecewa pada Saksi Ahli Poligraf, L...

Gambar 1. Hasil *Crawling Data*

3.2. Tahap *Preprocessing*

Setelah dilakukan *crawling*/pengumpulan data, maka dilakukan tahap *pre-processing*. Pada tahap ini, akan dilakukan beberapa hal, yaitu membersihkan data dengan menghilangkan karakter-karakter

yang tidak dibutuhkan (emoji, tanda baca, angka, url), serta tweet duplikat. Kemudian diberikan pelabelan sentimen apakah termasuk positif atau negatif Berikut ini hasil data yang telah dibersihkan pada Gambar 2.

```
tweets_df.head()
```

Out[39]:

	Tweets	Label	polarity
0	terdakwa kasus pembunuhan brigar nofriansyah y...	positif	NaN
1	ferdy sambo gelap mata menghabisi nyawa brigar...	negatif	NaN
2	terdakwa perintangan penyikan pembunuhan brigar j	negatif	NaN
3	kasus pembunuhan nofriansyah yosua hutabarat a...	negatif	NaN
4	blak bl ogah bela putri candrawathi	positif	NaN

Gambar 2. Hasil Data yang Telah Bersih

Selanjutnya, data yang telah bersih akan dilakukan ditokenisasi dimana pada tahap ini tweet akan dipisahkan kata per kata seperti pada Gambar 3 berikut ini.

```
Out[40]: <bound method NDFrame.head of
0    terdakwa kasus pembunuhan brigar nofriansyah y... positif NaN
1    ferdysambo gelap mata menghabisi nyawa brigar... negatif NaN
2    terdakwa perintangan penyikan pembunuhan brigar j negatif NaN
3    kasus pembunuhan nofriansyah yosua hutabarat a... negatif NaN
4    blak bl ogah bela putri candrawathi positif NaN
..
229  ada istilah karungga rumah ferdysambo penje... positif NaN
230  bravo hakim wahyu imam n nlaporkan melanggar... positif NaN
231  serem bak perang koboy ferdysambo putri seb... negatif NaN
232  majelis hakim menggali pengetahuan istri ferd... negatif NaN
233  ferdysambo sebut putri candrawathi perkosa pe... positif NaN

TOKENIZATION
0    [terdakwa, kasus, pembunuhan, brigar, nofrians...
1    [ferdy, sambo, gelap, mata, menghabisi, nyawa,...
2    [terdakwa, perintangan, penyikan, pembunuhan, ...
3    [kasus, pembunuhan, nofriansyah, yosua, hutaba...
4    [blak, bl, ogah, bela, putri, candrawathi]
..
229  [ada, istilah, karungga, rumah, ferdysambo, ...
230  [bravo, hakim, wahyu, imam, n, nlaporkan, mela...
231  [serem, bak, perang, koboy, ferdysambo, putr...
232  [majelis, hakim, menggali, pengetahuan, istri,...
233  [ferdy, sambo, sebut, putri, candrawathi, perk...

[234 rows x 4 columns]>
```

Gambar 3. Hasil Data yang Telah Ditokenisasi

Kemudian selanjutnya sebagai penyesuaian terhadap kemunculan nilai kontinu dalam fitur dataset yang akan mempengaruhi hasil proses klasifikasi dengan Naïve Bayes maka dilanjutkan dengan tahapan normalisasi. Berikut ini hasil gambarnya.

```
Out[42]: 0    [terdakwa, pembunuhan, brigar, nofriansyah, yo...
1    [ferdy, sambo, gelap, menghabisi, nyawa, briga...
2    [terdakwa, perintangan, penyikan, pembunuhan, ...
3    [pembunuhan, nofriansyah, yosua, hutabarat, br...
4    [blak, bl, ogah, bela, putri, candrawathi]
5    [ferdy, sambo, memerintahkan, anak, buahnya, m...
6    [ferdy, sambo, kaget, rekaman, cctv, tunjukan,...
7    [brigar, joshua]
8    [hendra, kurniawan, mengaku, perintah, irfan]
9    [organisasi, efektif, kaum, perempuan, ]
Name: Normalisasi, dtype: object
```

Gambar 4. Hasil Normalisasi

Kemudian akan dibuat fungsi menghapus stopwords pada seluruh tweet di dataset, dimana stopwords ini sebelumnya telah dibuat pada file txt.

Out[41]:

	Tweets	Label	polarity	TOKENIZATION	STOP_REMOVAL
0	terdakwa kasus pembunuhan brigar nofriansyah y...	positif	NaN	[terdakwa, kasus, pembunuhan, brigar, nofrians...	[terdakwa, pembunuhan, brigar, nofriansyah, yo...
1	ferdy sambo gelap mata menghabisi nyawa brigar...	negatif	NaN	[ferdy, sambo, gelap, mata, menghabisi, nyawa,...	[ferdy, sambo, gelap, menghabisi, nyawa, brig...
2	terdakwa perintangan penyikan pembunuhan brigar j	negatif	NaN	[terdakwa, perintangan, penyikan, pembunuhan, ...	[terdakwa, perintangan, penyikan, pembunuhan, ...
3	kasus pembunuhan nofriansyah yosua hutabarat a...	negatif	NaN	[kasus, pembunuhan, nofriansyah, yosua, hutaba...	[pembunuhan, nofriansyah, yosua, hutabarat, br...
4	blak bl ogah bela putri candrawathi	positif	NaN	[blak, bl, ogah, bela, putri, candrawathi]	[blak, bl, ogah, bela, putri, candrawathi]
...
229	ada istilah karungga rumah ferdys sambo penje...	positif	NaN	[ada, istilah, karungga, rumah, ferdys, sambo, ...	[istilah, karungga, rumah, ferdys, sambo, penje...
230	bravo hakim wahyu imam n nlaporkan melanggar ...	positif	NaN	[bravo, hakim, wahyu, imam, n, nlaporkan, mela...	[bravo, hakim, wahyu, imam, nlaporkan, melangg...
231	serem bak perang koboy ferdys sambo putri seb...	negatif	NaN	[serem, bak, perang, koboy, ferdys, sambo, putr...	[serem, perang, koboy, ferdys, sambo, putri, be...
232	majelis hakim menggali pengetahuan istri ferdys...	negatif	NaN	[majelis, hakim, menggali, pengetahuan, istri,...	[majelis, hakim, menggali, pengetahuan, istri,...
233	ferdy sambo sebut putri candrawathi perkosa pe...	positif	NaN	[ferdy, sambo, sebut, putri, candrawathi, perk...	[ferdy, sambo, putri, candrawathi, perkosa, pe...

234 rows x 5 columns

Gambar 5. Hasil *Stopwords Removal*

Dan yang terakhir dibuat fungsi stemming untuk menghilangkan seluruh kata berimbuhan sehingga kata-kata tersebut kembali ke kata dasar. Hasilnya dapat dilihat pada Gambar 6.

	Tweets	Label	polarity	TOKENIZATION	STOP_REMOVAL	Normalisasi	tweet_tokens_stemmed
0	terdakwa kasus pembunuhan brigar nofriansyah y...	positif	NaN	[terdakwa, kasus, pembunuhan, brigar, nofrians...	[terdakwa, pembunuhan, brigar, nofriansyah, yo...	[terdakwa, pembunuhan, brigar, nofriansyah, yo...	[dakwa, bunuh, brigar, nofriansyah, yosua, hut...
1	ferdy sambo gelap mata menghabisi nyawa brigar...	negatif	NaN	[ferdy, sambo, gelap, mata, menghabisi, nyawa,...	[ferdy, sambo, gelap, menghabisi, nyawa, brig...	[ferdy, sambo, gelap, menghabisi, nyawa, brig...	[ferdy, sambo, gelap, habs, nyawa, brigar, de...
2	terdakwa perintangan penyikan pembunuhan brigar j	negatif	NaN	[terdakwa, perintangan, penyikan, pembunuhan, ...	[terdakwa, perintangan, penyikan, pembunuhan, ...	[terdakwa, perintangan, penyikan, pembunuhan, ...	[dakwa, rintang, nyi, bunuh, brigar]
3	kasus pembunuhan nofriansyah yosua hutabarat a...	negatif	NaN	[kasus, pembunuhan, nofriansyah, yosua, hutaba...	[pembunuhan, nofriansyah, yosua, hutabarat, br...	[pembunuhan, nofriansyah, yosua, hutabarat, br...	[bunuh, nofriansyah, yosua, hutabarat, brigar, ...]
4	blak bl ogah bela putri candrawathi	positif	NaN	[blak, bl, ogah, bela, putri, candrawathi]	[blak, bl, ogah, bela, putri, candrawathi]	[blak, bl, ogah, bela, putri, candrawathi]	[blak, bl, ogah, bela, putri, candrawathi]
...
229	ada istilah karungga rumah ferdys sambo penje...	positif	NaN	[ada, istilah, karungga, rumah, ferdys, sambo, ...	[istilah, karungga, rumah, ferdys, sambo, penje...	[istilah, karungga, rumah, ferdys, sambo, penje...	[istilah, karungga, rumah, ferdys, sambo, jelas...
230	bravo hakim wahyu imam n nlaporkan melanggar ...	positif	NaN	[bravo, hakim, wahyu, imam, n, nlaporkan, mela...	[bravo, hakim, wahyu, imam, nlaporkan, melangg...	[bravo, hakim, wahyu, imam, nlaporkan, melangg...	[bravo, hakim, wahyu, imam, nlaporkan, langgar...
231	serem bak perang koboy ferdys sambo putri seb...	negatif	NaN	[serem, bak, perang, koboy, ferdys, sambo, putr...	[serem, perang, koboy, ferdys, sambo, putri, be...	[serem, perang, koboy, ferdys, sambo, putri, be...	[rem, perang, koboy, ferdys, sambo, putri, ante...
232	majelis hakim menggali pengetahuan istri ferdys...	negatif	NaN	[majelis, hakim, menggali, pengetahuan, istri,...	[majelis, hakim, menggali, pengetahuan, istri,...	[majelis, hakim, menggali, pengetahuan, istri,...	[majelis, hakim, gali, tahu, istri, ferdys, sam...
233	ferdy sambo sebut putri candrawathi perkosa pe...	positif	NaN	[ferdy, sambo, sebut, putri, candrawathi, perk...	[ferdy, sambo, putri, candrawathi, perkosa, pe...	[ferdy, sambo, putri, candrawathi, perkosa, pe...	[ferdy, sambo, putri, candrawathi, perkosa, ac...

233 rows x 7 columns

Gambar 6. Hasil *Stemming*

3.3. Tahap Pemodelan

3.3.1. Klasifikasi

Dari pelatihan menggunakan 233 data yang sudah dilabeli, didapatkan hasil Classifier Evaluation Metrics, yaitu hasil Precision, Recall, dan F-measures seperti pada Gambar 4. Untuk sentimen negatif mendapatkan hasil precision 77%, recall 71%, f1-score 74%. Untuk sentimen positif mendapatkan hasil precision 72%, recall 78%, f1-score 75%. Berikut ini hasil Classifier Evaluation Matriks.

	precision	recall	f1-score	support
0	0.77	0.71	0.74	24
1	0.72	0.78	0.75	23
accuracy			0.74	47
macro avg	0.75	0.75	0.74	47
weighted avg	0.75	0.74	0.74	47

Gambar 7. Hasil *Classifier Evaluation Metrics*

PNN tidak cocok digunakan untuk klasifikasi dataset yang besar [16,27]. Hal ini dikarenakan PNN memiliki struktur model yang kompleks dan jumlah parameter yang besar yang dapat menyebabkan overfitting pada data pelatihan dan penurunan kinerja pada data pengujian. PNN juga memerlukan memori yang besar untuk menyimpan semua data yang digunakan selama proses

pelatihan. Dalam kasus data yang sangat besar, memori yang dibutuhkan akan menjadi terlalu besar dan tidak tersedia pada sistem komputasi yang ada [28].

4. KESIMPULAN

Dalam penelitian ini, dilakukan proses klasifikasi sentimen analisis tweets pada berbagai sentimen publik tentang kasus Ferdy Sambo. Dengan persentase opini positive sebesar 51,50% dan negative sebesar 48,50%. Dengan klasifikasi Naive Bayes Classiefier hasilnya diketahui untuk sentimen negatif mendapatkan hasil precision 77%, recall 71%, f1-score 74%. Dan untuk sentimen positif mendapatkan hasil precision 72%, recall 78%, f1-score 85%. Dengan begitu dapat kita ketahui bahwa Algoritma Naive Bayes Classifier dalam penelitian ini bekerja cukup baik dan disarankan untuk penelitian selanjutnya terkait topik sentiment terhadap kasus ini, maka ditambahkan jumlah datanya, sehingga dapat meningkatkan kemampuan machine learning dalam melakukan pelatihan pada tahapan *training data*, karena jika dataset atau data training lebih banyak, maka kemampuan mesin dalam melakukan prediksi maupun klasifikasi akan lebih maksimal. Dan untuk kinerja algoritma naïve bayes sudah cukup baik dengan keterbatasan data pada penelitian ini.

DAFTAR PUSTAKA

- [1] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA," J. Ilm. Edutic, vol. 7, no. 1, pp. 1–11, 2020.
- [2] Kemenkominfo. "Pengguna Internet di Indonesia 63 Juta Orang". from kominfo : https://www.kominfo.go.id/content/detail/3415/kominfo-pengguna-internet-di-indonesia-63-juta-orang/0/berita_satker. 2019. [Accessed:Maret 2019].
- [3] B. Pang, L. Lee, Opinion Mining and Sentiment Analysis, Foundations and Trends in Information Retrieval, Vol. 2, No 1-2 pp. 1–135, 2008.
- [4] Ayu Rizaty, Monavia , Pengguna Twitter di Indonesia Capai 18,45 Juta pada 2022, Available from : <https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022>
- [5] P. Antinasari, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku," vol. 1 No.12, pp. 1733–1741, 2017.
- [6] Wenando, F. A., & Fuad, E. (2019). Detection of Hate Speech in Indonesian Language on Twitter Using Machine Learning Algorithm. Prosiding CELSciTech, 4, 6-8.
- [7] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," Ilk. J. Ilm., vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [8] Han, J., and M.Kamber. 2006. Data Mining Concept and Techniques Second Edition. San Frasisco : Elsevier.
- [9] Tala, Fadillah Z, A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia, Master of Logic ProjectInstitute for Logic, Language and Computation Universiteit van AmsterdamThe Netherlands, 2003
- [10] Wenando, F. A., Hayami, R., & Novermahakim, A. Y. (2020, October). Tweet Sentiment Analysis for 2019 Indonesia Presidential Election Results using Various Classification Algorithms. In 2020 1st International Conference on Information Technology, Advanced Mechanical and Electrical Engineering (ICITAMEE) (pp. 279-282). IEEE.
- [11] D. Alita, Y. Fernando, and H. Sulistiani, "Implementasi Algoritma Multiclass SVM pada Opini Publik Berbahasa Indonesia di Twitter," J. Tekno Kompak, vol. 14, no. 2, pp. 86–91, 2020
- [12] Wenando, F. A., Hayami, R., & Anggrawan, A. J. (2020). Analisis Sentimen Pada Pemerintahan Terpilih Pada Pilpres 2019 Ditwitter Menggunakan Algoritme Naïvebayes. JURTEKSI (Jurnal Teknologi dan Sistem Informasi), 7(1), 101-106.
- [13] Wildan, M. (2023). Sentimen negatif netizen dalam kolom komentar detik. com terhadap pemberitaan kasus Ferdy Sambo. LITERA, 22(1).
- [14] Bayhaqy, A., et al. 2018. "Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naive Bayes". International Conference on Orange Technologies, ICOT 1–6.
- [15] Salam, A., Zeniarja, J., Khasanah, RSU. 2018. "Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia)". Prosiding SINTAK. 480.
- [16] Wenando, F. A., Septiadi, R., Gunawan, R., & Mukhtar, H. (2022). K-Nearest Neighbor (KNN) untuk Menganalisis Sentimen terhadap Kebijakan Merdeka Belajar Kampus Merdeka pada Komentar Twitter. *Jurnal CoSciTech (Computer Science and Information Technology)*, 3(2), 152-158.
- [17] Lisa, R. L. Analisis Sentimen Opini Masyarakat terhadap Institusi Polri pada Media Sosial Twitter menggunakan Metode Support Vector Machine dan Naive Bayes.