

ANALISIS SENTIMEN TERHADAP DAMPAK PERANG ISRAEL - PALESTINA MELALUI DATA TWITTER MENGGUNAKAN NAIVE BAYES

Alfian Noer Halim ¹; Saruni Dwiasnati ²

^{1,2} *Fakultas Ilmu Komputer, Universitas Mercu Buana, Jakarta Barat*

Alfiannoer72@gmail.com ¹, saruni.dwiasnati@mercubuana.ac.id ²

Kata kunci:
Naïve Bayes, Gaussian, Multinomial, Bernoulli.

Abstract

The increasing development of information technology makes it easy for people to get various information only through social media such as Twitter. Twitter is a mainstay social networking application and source of information on world events. With Twitter, people can get a lot of the latest news. One piece of information that is widely discussed and is a trending topic on Twitter is the impact of the Israeli and Palestinian war. It is important to analyze the feelings of the impact of the ceasefire between Israel and Palestine from the amount of information in online media. The data used is Twitter, a social media platform. This research was conducted to analyze people's reactions to data in the form of tweets and group them according to the Naïve Bayes method into positive, neutral or negative opinions. In implementing the Naïve Bayes algorithm which uses 3 models of the Naïve Bayes algorithm, namely Gaussian, Multinomial, and Bernoulli, it shows different results, namely 50% for the Naïve Bayes Gaussian model, 57% for the Naïve Bayes Bernoulli model, and Naïve Bayes Multinomial model is 65 %. This shows that the Multinomial Naïve Bayes model is better than other models in classifying the data in this case.

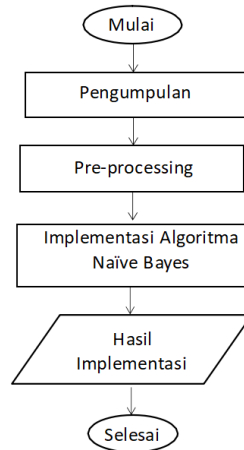
Pendahuluan

A. Latar Belakang

Pada zaman teknologi 4.0 media sosial tidak akan pernah lepas dari kehidupan sehari-hari. Sehingga saat sekarang media sosial sudah setara dengan kebutuhan primer dari kehidupan manusia. Menurut data indeks Tetra pax, sekitar 132 juta orang Indonesia atau lebih dari setengah dari total populasi adalah pengguna internet aktif, 79,5% di antaranya adalah pengguna jejaring sosial. Jadi masyarakat Indonesia sangat membutuhkan media sosial. Bagi masyarakat yang mampu mengendalikan media sosial maka akan menuai hasil yang positif, bahkan juga bisa menjadi bumerang bagi mereka yang membuat kegaduhan dan sebagainya. Sehingga pemerintah untuk menertibkan para pengguna media sosial yang “nakal” dengan menggunakan KUHP No. 19 Tahun 2016 dan UU No. 11 Tahun 2008 tentang Perdagangan Informasi Elektronik (ITE) [1] Konflik Palestina dengan Israel bukan hanya permasalahan agama saja, akan tetapi terkait juga akan tragedi kemanusiaan. Telah banyak korban yang ditunjukkan tentara Israel terhadap rakyat sipil Palestina. Konflik ini merupakan salah satu konflik terpanjang yang masih berlangsung di dunia.[1] Berbagai upaya telah dilakukan untuk menyelesaikan konflik sebagai bagian dari proses perdamaian Israel-Palestina,[2] di mana upaya perdamaian ini juga merupakan upaya lain untuk menyelesaikan konflik Arab-Israel yang lebih luas.[3] Perang Israel-Hamas masih menjadi sorotan utama dunia meski kini sepakat melakukan gencatan senjata selama empat hari dengan imbalan pembebasan tawanan dan sandera oleh kedua pihak. Ini adalah penghentian pertempuran pertama sejak Hamas menyerang Israel selatan pada 7 Oktober 2023 yang menewaskan 1.200 orang.

Menarik dari sejarah beberapa sejarah dari literatur yang telah beredar dan berhubungan langsung dengan aksi kemanusiaan, maka masyarakat dengan menggunakan media sosial sebagai people power untuk berbagai macam yang bisa dilakukan, yakni memberitakan terkait kejadian terkini,

informasi korban berjatuhan, bahkan sampai bisa membuat galangan dana bagi rakyat Palestina yang terkena dampak. Kehadiran platform twitter sehingga dapat kita temui secara mudah berbagai macam opini dan statement yang keluar dari masyarakat baik berupa positif, negatif maupun netral. Terkait hal tersebut kita membutuhkan semua opini tersebut agar bisa mengetahui opini mana yang positif, negatif dan netral.



Gambar 1. Tahapan penelitian

Dari Gambar 1 dapat dijelaskan tahapan penelitian sebagai berikut.

A. Pengumpulan Data

Data yang diterapkan dalam riset ini diperoleh melalui tweet dan retweet yang dibuat oleh pengguna di platform Twitter. Data tersebut hanya diambil dari situs twitter.com dan difokuskan pada kata kunci "Perang Palestina". Pengumpulan data dilakukan melalui metode pengambilan data dengan cara scraping dengan menggunakan bahasa pemrograman Python. Hasil yang didapatkan sebanyak kurang lebih 760 data. Data yang terkumpul dari beragam sumber ini selanjutnya diintegrasikan untuk dianalisis dan diatur sedemikian rupa hingga menjadi dasar yang padu dalam rangka penelitian ini. Selanjutnya, data tersebut disimpan dalam format file CSV, sebelum dilakukan proses pre-processing.

B. Pre-processing

Data yang sudah didapatkan akan dilanjutkan tahapan proses pre-processing agar data dengan melakukan beberapa tahap seperti : case folding, tokenization, filtering dan lain sebagainya.

- *Cleaning* merupakan proses membersihkan review dari kata-kata yang tidak diperlukan untuk mengurangi proses noise pada proses klasifikasi. Kata-kata yang dihilangkan adalah karakter.
- *Case Folding* yaitu tahapan dalam menyamakan case di suatu dokumen hingga mengkonversi seluruh teks dalam huruf kecil atau lowercase dan dilakukan untuk mempermudah pencarian sebuah kata.
- *Tokenizing* merupakan proses pemecahan sebuah kalimat sebagai kata yang dilaksanakan dengan tujuan membuat suatu kalimat lebih bermakna. Tahapan tersebut dimulai dengan melakukan menormalisasikan kata melalui perubahan seluruh karakter huruf menjadi lowercase dan menghilangkan simbol hingga tanda baca yang terdapat sebuah kalimat misalnya '@', '?', '!' dan lain sebagainya. Tahapan ini mengandalkan karakter spasi dalam melakukan pemisahan kalimat menjadi sebuah kumpulan kata.
- *Filtering* yang digunakan yaitu StopWord yang merupakan tahapan pembuangan kata yang tidak penting dari sebuah data misalnya kata hubung 'yang', 'di', 'dan', hingga seterusnya dan menyisakan kata penting.
- *Stemming* bertujuan untuk menghilangkan imbuhan tambahan sehingga kata-kata tersebut dapat diubah menjadi bentuk dasarnya. Contohnya, kata "berlari", "bermain", dan "berbicara" semuanya mengandung imbuhan "ber-" yang menandakan kata kerja dalam bentuk aktif. Dalam proses stemming, imbuhan "ber-" akan dihilangkan sehingga kata-kata tersebut berubah menjadi kata dasar "lari", "main", dan "bicara".

C. Labeling

Pada langkah ini, data yang telah melewati tahap prapemrosesan akan diberi label kelas. Pelabelan kategori data terbagi menjadi dua, yaitu kategori positif, negatif, dan netral. Proses pemberian label kelas dilakukan secara otomatis menggunakan bahasa pemrograman Python dengan dukungan dari platform Google Colab. Data yang akan diberi label kelas adalah data hasil prapemrosesan yang telah dipersiapkan sebelumnya, seperti data tweet yang telah melalui tahap cleaning, tokenisasi, dan stemming.

Metode pemberian label dilakukan berdasarkan skor sentimen dari teks. Skor sentimen mengindikasikan apakah teks memiliki sentimen positif, negatif, atau netral. Data yang memiliki skor sentimen < 0 akan diberi label positif, yang berarti teks tersebut cenderung memiliki sentimen positif. Data yang memiliki skor sentimen > 0 akan diberi label negatif, yang berarti teks tersebut cenderung memiliki sentimen negatif. Sedangkan data yang memiliki skor sentimen $= 0$ akan diberi label netral, yang berarti teks tersebut cenderung memiliki sentimen netral.

D. Splitting Data

Data splitting merupakan pembagian data agar terbagi 2 yakni data dibagi menjadi data training dan data testing.

E. Pembuatan Model

TPemodelan dilakukan menggunakan algoritma Naïve Bayes. Naïve Bayes merupakan salah satu metode dalam data mining yang menggunakan konsep peluang Bayes.[16] Pemodelan akan menggunakan 3 model dari algoritma Naïve Bayes yaitu Gaussian, Multinomial dan Bernoulli

- *Gaussian Naive Bayes* adalah Ini adalah pengklasifikasi Naive Bayes paling sederhana yang memiliki asumsi bahwa data dari masing-masing label diambil dari distribusi Gaussian sederhana. Gaussian Distribution mengasumsikan bahwa setiap feature pada data memiliki pengaruh yang independen dalam memprediksi target. Kombinasi prediksi dari seluruh parameter adalah prediksi akhir dengan probability dari target variable yang diklasifikasikan ke dalam dua kelas. Klasifikasi akhirnya adalah hasil probability yang lebih tinggi dari grup target
- *Bernoulli Naive Bayes*, pembobotan dilakukan dengan menggunakan binary (0 dan 1) dalam pembobotan tiap term, hal ini berbeda dengan perhitungan term frekuensi yang melakukan pembobotan pada setiap term.[17]Persamaan Bernoulli Naive Bayes dapat dilihat pada persamaan 1.[18]
- *Multinomial Naive Bayes* mengasumsikan independensi diantara kemunculan kata-kata dalam dokumen, tanpa memperhitungkan urutan kata dan konteks informasi dalam kalimat. Multinomial Naive Bayes juga memperhitungkan jumlah kemunculan kata dalam dokumen.[17]Persamaan Multinomial dapat dilihat pada persamaan 2.[19]

E. Evaluasi Model

Setelah data di training dengan model maka selanjutnya adalah melakukan evaluasi dan validasi terhadap akurasi dari model algoritma Naïve Bayes yaitu Gaussian, Multinomial dan Bernoulli.

Hasil dan diskusi

Dalam studi kasus kali ini, Dataset yang digunakan adalah data yang diambil hanya dari situs twitter.com di Indonesia Data yang digunakan melakukan penelitian ini yaitu bersumber dari Twitter. Data diperoleh dengan metode scrapping menggunakan library sudo dan nodejs. Data yang diperoleh sejumlah 768 tweet dengan periode data dari 1 Januari 2023 sampai dengan 1 Desember 2023. Tweet yang diambil hanyalah tweet yang mengandung kata “Perang Palestina” didalamnya, kemudian bisa kita export ke dalam bentuk .csv.

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	created_at	id_str	full_text	quote_count	reply_count	retweet_count	favorite_count	lang	user_id	str	conversation_id	username	tweet_url				
2	Thu Nov 30	1,73E+18	@salimafill	0	1	0	4	in	1,05E+18	1,73E+18	roundhatb	https://twitter.com/roundhatboy/status/1730375037918810197					
3	Thu Nov 30	1,73E+18	Perang Pale	0	0	0	0	in	1,71E+18	1,73E+18	hkkhusna	https://twitter.com/hkkhusna/status/1730373315628618226					
4	Thu Nov 30	1,73E+18	@botmenfr	0	0	0	0	in	1,36E+18	1,73E+18	ArifMiqbal2	https://twitter.com/ArifMiqbal2/status/1730371672921026947					
5	Thu Nov 30	1,73E+18	@bejo_sek	0	1	0	0	in	1,72E+18	1,73E+18	Nusantaral	https://twitter.com/Nusantaral46380/status/1730370545781493815					
6	Thu Nov 30	1,73E+18	@Omj_jeN	0	0	0	0	in	102162505	1,73E+18	detosimde	https://twitter.com/detosimde/status/173036894968111123					
7	Thu Nov 30	1,73E+18	Palestina Is	0	0	0	0	in	1,71E+18	1,73E+18	hijberita	https://twitter.com/hijberita/status/1730364356897259584					
8	Thu Nov 30	1,73E+18	@yukitsunc	1	1	0	7	in	1,07E+18	1,73E+18	NadreenShahk	https://twitter.com/NadreenShahk/status/1730361759486341329					
9	Thu Nov 30	1,73E+18	@bejo_sek	0	2	0	0	in	1,72E+18	1,73E+18	Nusantaral	https://twitter.com/Nusantaral46380/status/1730360483860152765					
10	Thu Nov 30	1,73E+18	@DeHoutr	0	2	0	1	in	308526515	1,73E+18	noez_wahyudi	https://twitter.com/noez_wahyudi/status/1730360030137143627					
11	Thu Nov 30	1,73E+18	@Ramadha	0	1	0	22	in	1,53E+18	1,73E+18	janii54	https://twitter.com/janii54/status/1730358864791601153					
12	Thu Nov 30	1,73E+18	@vothban	0	2	0	4	in	7,55E+17	1,73E+18	Rajabunisa	https://twitter.com/Rajabunisa/status/1730349175055995206					
13	Thu Nov 30	1,73E+18	@IATeam_	0	0	0	0	in	160047178	1,73E+18	Ted1_STP	https://twitter.com/Ted1_STP/status/1730345860628983813					
14	Thu Nov 30	1,73E+18	@sahabata	0	0	0	0	in	1,42E+18	1,73E+18	FreeMob	https://twitter.com/FreeMob/status/1730343188194574663					
15	Thu Nov 30	1,73E+18	@Avolanza	0	0	2	15	in	1,68E+18	1,73E+18	balani_sum	https://twitter.com/balani_sum/status/1730332804800753914					
16	Thu Nov 30	1,73E+18	@israel@g	0	0	0	1	in	1,71E+18	1,73E+18	AnakKampun	https://twitter.com/AnakKampun14894/status/1730320218768208171					
17	Thu Nov 30	1,73E+18	@jconricus	0	0	0	1	in	1,71E+18	1,73E+18	AnakKampun	https://twitter.com/AnakKampun14894/status/1730317953336496523					
18	Thu Nov 30	1,73E+18	Sebuah kan	0	0	0	0	in	1,58E+18	1,73E+18	labidhoucin58	https://twitter.com/labidhoucin58/status/173030889968795706					
19	Thu Nov 30	1,73E+18	@AdrySev	0	1	0	0	in	1,30E+18	1,73E+18	ntillee	https://twitter.com/ntillee/status/17302958484368507156					
20	Thu Nov 30	1,73E+18	YERUSALEM	0	1	0	0	in	2,436E+09	1,73E+18	his_hyung	https://twitter.com/his_hyung/status/1730283113690100171					
21	Thu Nov 30	1,73E+18	Kakanwil be	0	0	0	0	in	4,842E+09	1,73E+18	rutan_garu	https://twitter.com/rutan_garut/status/1730280087663288594					
22	Thu Nov 30	1,73E+18	@Omj_jeN	0	1	11	373	in	854511938	1,73E+18	onlyformnt	https://twitter.com/onlyformntkn/status/1730276124322885792					
23	Thu Nov 30	1,73E+18	@NandeBa	0	0	0	0	in	1,58E+18	1,73E+18	kentujibril	https://twitter.com/kentujibril/status/1730262766609707049					

Gambar 2. Dataset

Setelah mengambil dataset, langkah berikutnya adalah memeriksa informasi data tersebut, seperti tipe data dan jumlah data untuk setiap atribut Dataset yang telah dikumpulkan dan dilakukan proses seleksi ini memiliki 12 atribut yaitu `created_dtm`, `id_str`, `full_text`, `quote_count`, `reply_count`, `retweet_count`, `favorite_count`, `lang`, `user_id_str`, `conversation_id_str`, `username`, dan `tweet_url`.

```
data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   created_at            768 non-null   object
1   id_str                768 non-null   int64
2   full_text            768 non-null   object
3   quote_count          768 non-null   int64
4   reply_count          768 non-null   int64
5   retweet_count        768 non-null   int64
6   favorite_count       768 non-null   int64
7   lang                 768 non-null   object
8   user_id_str          768 non-null   int64
9   conversation_id_str  768 non-null   int64
10  username              768 non-null   object
11  tweet_url            768 non-null   object
dtypes: int64(7), object(5)
memory usage: 72.1+ KB
```

Gambar 3. Atribut

Tahap selanjutnya adalah pre-processing dataset mentah sebelum digunakan untuk pelatihan model, tahap pre-processing dilakukan agar dataset yang nantinya akan diuji menjadi lebih baik, yaitu dengan membersihkan hal-hal yang tidak digunakan agar mendapatkan data yang berkualitas dan dapat diproses oleh model. Ada beberapa tahapan dalam melakukan preprocessing yaitu `cleansing`, `case folding`, `tokenization`, `stopword removal`, dan `stemming`.

Proses `cleaning` dilakukan agar dataset yang digunakan menjadi jauh lebih baik dengan cara menghilangkan berbagai hal yang dapat mempengaruhi akurasi dari model. Proses `cleaning` yang dilakukan adalah seperti menghilangkan simbol, nama `username`, `hashtag`, angka, dan `link`. Proses `cleaning` ini dilakukan dengan memanfaatkan library python bernama “`re`”. Berikut adalah perbedaan dataset sebelum dan setelah dilakukan `cleaning`.

	full_text	cleaning
0	@salimafillah Anjir pantesan aja airstrike mel...	salimafillah Anjir pantesan aja airstrike melu...
1	Perang Palestina-Israel: Menanti Respon Politi...	Perang Palestinalrael Menanti Respon Politik ...
2	@botmenfess @neohistoria_id Ancur abis perang ...	botmenfess neohistoriaid Ancur abis perang mah...
3	@bejo_sekali Jadi, yg memulai perang adalah wa...	bejosekali Jadi yg memulai perang adalah warga...
4	@OmJ_JeNggot Teruskan PERANG sampai ISRAHELL m...	OmJJeNggot Teruskan PERANG sampai ISRAHELL mus...

Gambar 4. Cleaning

Setelah dilakukannya `cleaning` tahap selanjutnya adalah `case folding` ini dilakukan untuk mengubah huruf kapital pada dataset menjadi huruf biasa atau huruf kecil seperti “Kangen” menjadi “kangen” atau dapat disebut juga penyeragaman bentuk huruf. . Meskipun memiliki arti yang sama tetapi model menafsirkan bahwa 2 kata tersebut memiliki arti yang berbeda karena memiliki perbedaan pada hurufnya.

	full_text	cleaning	case_folding
0	@salimafillah Anjir pantesan aja airstrike mel...	salimafillah Anjir pantesan aja airstrike melu...	salimafillah anjir pantesan aja airstrike melu...
1	Perang Palestina-Israel: Menanti Respon Politi...	Perang Palestinalrael Menanti Respon Politik ...	perang palestinaisrael menanti respon politik ...
2	@botmenfess @neohistoria_id Ancur abis perang ...	botmenfess neohistoriaid Ancur abis perang mah...	botmenfess neohistoriaid ancur abis perang mah...
3	@bejo_sekali Jadi, yg memulai perang adalah wa...	bejosekali Jadi yg memulai perang adalah warga...	bejosekali jadi yg memulai perang adalah warga...
4	@OmJ_JeNggot Teruskan PERANG sampai ISRAHELL m...	OmJJeNggot Teruskan PERANG sampai ISRAHELL mus...	omjjeNggot teruskan perang sampai israhell mus...

Gambar 5. Case Folding

roses tokenizing atau parsing adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Pada dasarnya proses tokenizing adalah pemenggalan kalimat menjadi kata.

	full_text	cleaning	case_folding	tokenize
0	@salimafillah Anjir pantesan aja airstrike melu...	salimafillah Anjir pantesan aja airstrike melu...	salimafillah anjir pantesan aja airstrike melu...	[salimafillah, anjir, pantesan, aja, airstrike...
1	Perang Palestina-Israel: Menanti Respon Politik ...	Perang Palestinalsrael Menanti Respon Politik ...	perang palestinaisrael menanti respon politik ...	[perang, palestinaisrael, menanti, respon, pol...
2	@botmenfess @neohistoria_id Ancur abis perang ...	botmenfess neohistoriaid Ancur abis perang mah...	botmenfess neohistoriaid ancur abis perang mah...	[botmenfess, neohistoriaid, ancur, abis, peran...
3	@bejo_sekali Jadi, yg memulai perang adalah wa...	bejosekali Jadi yg memulai perang adalah warga...	bejosekali jadi yg memulai perang adalah warga...	[bejosekali, jadi, yg, memulai, perang, adalah...
4	@OmJ_jeNggot Teruskan PERANG sampai ISRAHELL m...	OmJJeNggot Teruskan PERANG sampai ISRAHELL mus...	omjenggot teruskan perang sampai israhell mus...	[omjenggot, teruskan, perang, sampai, israhel...

Gambar 6. Tokenizing

Stopword Removal merupakan proses menghilangkan daftar kata-kata yang tidakmendeskripsikan sesuatu yang semestinya dihilangkan seperti “yang”, “di”, “ke”, “itu” dan lainsebagainya. Contoh proses Stopward Removal data terdapat pada gambar

	full_text	cleaning	case_folding	tokenize	Filtering/stopword removal
0	@salimafillah Anjir pantesan aja airstrike melu...	salimafillah Anjir pantesan aja airstrike melu...	salimafillah anjir pantesan aja airstrike melu...	[salimafillah, anjir, pantesan, aja, airstrike...	[salimafillah, anjir, pantesan, aja, airstrike...
1	Perang Palestina-Israel: Menanti Respon Politik ...	Perang Palestinalsrael Menanti Respon Politik ...	perang palestinaisrael menanti respon politik ...	[perang, palestinaisrael, menanti, respon, pol...	[perang, palestinaisrael, respon, politik, uma...
2	@botmenfess @neohistoria_id Ancur abis perang ...	botmenfess neohistoriaid Ancur abis perang mah...	botmenfess neohistoriaid ancur abis perang mah...	[botmenfess, neohistoriaid, ancur, abis, peran...	[botmenfess, neohistoriaid, ancur, abis, peran...
3	@bejo_sekali Jadi, yg memulai perang adalah wa...	bejosekali Jadi yg memulai perang adalah warga...	bejosekali jadi yg memulai perang adalah warga...	[bejosekali, jadi, yg, memulai, perang, adalah...	[bejosekali, yg, perang, warga, arab, tinggal...
4	@OmJ_jeNggot Teruskan PERANG sampai ISRAHELL m...	OmJJeNggot Teruskan PERANG sampai ISRAHELL mus...	omjenggot teruskan perang sampai israhell mus...	[omjenggot, teruskan, perang, sampai, israhel...	[omjenggot, teruskan, perang, israhell, musna...

Gambar 6. Stopword Removal

Stemming yaitu metode yang bertujuan akan mengganti kata-kata berimbuhan menjadi bentuk dasar yang sesuai dengan sistematis tata bahasa Indonesia yang benar.

	full_text	cleaning	case_folding	tokenize	Filtering/stopword removal	stemming_data
0	@salimafillah Anjir pantesan aja airstrike melu...	salimafillah Anjir pantesan aja airstrike melu...	salimafillah anjir pantesan aja airstrike melu...	[salimafillah, anjir, pantesan, aja, airstrike...	[salimafillah, anjir, pantesan, aja, airstrike...	salimafillah anjir pantesan aja airstrike melu...
1	Perang Palestina-Israel: Menanti Respon Politik ...	Perang Palestinalsrael Menanti Respon Politik ...	perang palestinaisrael menanti respon politik ...	[perang, palestinaisrael, menanti, respon, pol...	[perang, palestinaisrael, respon, politik, uma...	perang palestinaisrael respon politik umat arm...
2	@botmenfess @neohistoria_id Ancur abis perang ...	botmenfess neohistoriaid Ancur abis perang mah...	botmenfess neohistoriaid ancur abis perang mah...	[botmenfess, neohistoriaid, ancur, abis, peran...	[botmenfess, neohistoriaid, ancur, abis, peran...	botmenfess neohistoriaid ancur abis perang mah...
3	@bejo_sekali Jadi, yg memulai perang adalah wa...	bejosekali Jadi yg memulai perang adalah warga...	bejosekali jadi yg memulai perang adalah warga...	[bejosekali, jadi, yg, memulai, perang, adalah...	[bejosekali, yg, perang, warga, arab, tinggal...	bejosekali yg perang warga arab tinggal palest...
4	@OmJ_jeNggot Teruskan PERANG sampai ISRAHELL m...	OmJJeNggot Teruskan PERANG sampai ISRAHELL mus...	omjenggot teruskan perang sampai israhell mus...	[omjenggot, teruskan, perang, sampai, israhel...	[omjenggot, teruskan, perang, israhell, musna...	omjenggot teruskan perang israhell musnah bumi p...

Gambar 7. Stemming

Setelah menyelesaikan pre-processing pada dataset, langkah berikutnya adalah memberikan label pada data agar dapat dibedakan menjadi kelas positif dan negatif. Tujuan dari proses labeling

ini adalah untuk mempermudah pengolahan dan analisis visual pada dataset selanjutnya. Dalam tahap ini, peneliti menggunakan sebuah fungsi yang disebut `determine_sentiment` yang diimplementasikan dalam bahasa pemrograman Python. Setelah itu memasukkan kamus `positive_keywords = ['dunia', 'negara', 'indonesia', 'orang', 'dukung', 'yg', 'warga', 'yahudi', 'tanah', 'damai', 'tentara', 'dukung', 'agama']`, `negative_keywords = ['perang', 'senjata', 'jajah', 'korban', 'bantai', 'mati', 'kalah', 'palestina', 'israel', 'hamas']`, dan untuk netral tidak diantara kata pada keywords positive dan negative.

	stemming_data	Sentiments
0	salimafillah anjir pantesan aja airstrike melu...	Positif
1	perang palestinaisrael respon politik umat arm...	Negatif
2	botmenfess neohistoriaid ancur abis perang mah...	Negatif
3	bejosekali yg perang warga arab tinggal palest...	Positif
4	omjijenggot terus perang israhell musnah bumi p...	Netral
...
495	perang israel vs palestina ribut rebut	Negatif
496	mbuhlah pusing dengar berita perang israel ham...	Negatif
497	sahabatalaqsha ajarabic gencat senjata tukar t...	Negatif
498	wwwowin gaksukekan xoloss gytuxc arcstray bunu...	Negatif
499	palestina israel amerika amp sekutu genosida g...	Negatif

Gambar 7. Labeling Datase

Sebelum memulai pemodelan data, selanjutnya kumpulan data akan dipecah menjadi dua yaitu data pelatihan (training) dan data pengujian (testing), dengan pembagian data dilakukan dengan `rasiotest_size=0.2` (80% : 20%).

A. Klasifikasi menggunakan Algoritma Gaussian Naive Bayes

Dalam penelitian ini, data uji dan data pengujian dibagi menggunakan rasio 80% untuk data pelatihan dan 20% untuk data pengujian. Hasil dari perhitungan yang telah diolah, didapatkan tingkat performa klasifikasi Naive Bayes model Gaussian adalah sebesar 50%. Hasil ini ditunjukkan dalam classification report ditampilkan pada gambar

```

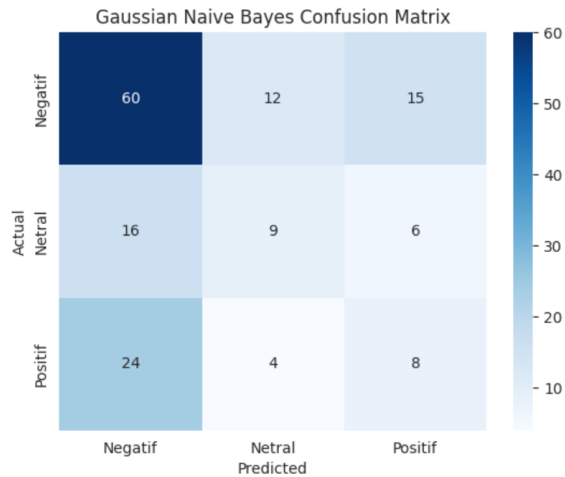
Gaussian Naive Bayes Accuracy: 0.5
Gaussian Naive Bayes Classification Report:

```

	precision	recall	f1-score	support
Negatif	0.60	0.69	0.64	87
Netral	0.36	0.29	0.32	31
Positif	0.28	0.22	0.25	36
accuracy			0.50	154
macro avg	0.41	0.40	0.40	154
weighted avg	0.48	0.50	0.48	154

Gambar 8. Hasil Test Report Gaussian Naive Bayes

Menampilkan pula jumlah hasil sentimen benar dan salah pada setiap kelas sentimen positif dan negatif dengan menggunakan plot confusion matrix dengan hasil berupa plot visualisasi, disajikan pada gambar



Gambar 9. Hasil Confusion Matrix Gaussian Naive Bayes

Berdasarkan hasil confusion matrix pada gambar di atas, untuk klasifikasi Naive Bayes model Gaussian didapatkan bahwa 8 analisis sentimen yang benar pada sentimen positif (True Positive), 15 analisis sentimen yang salah pada sentimen positif (False Positive), 24 analisis sentimen yang salah pada sentimen negatif (False Negative), 60 analisis sentimen yang benar pada sentimen negatif (True Netral). 4 analisis sentimen yang benar pada sentimen positif (True Positive), 12 analisis sentimen yang salah pada sentimen positif (False Positive).

B. Klasifikasi menggunakan Algoritma Bernaouli Naive Bayes

Dalam penelitian ini, data uji dan data pengujian dibagi menggunakan rasio 80% untuk data pelatihan dan 20% untuk data pengujian. Hasil dari perhitungan yang telah diolah, didapatkan tingkat performa klasifikasi Naive Bayes model Bernoulli adalah sebesar 57%. Hasil ini ditunjukkan dalam classification report ditampilkan pada gambar.

```

Bernoulli Naive Bayes Accuracy: 0.5714285714285714
Bernoulli Naive Bayes Classification Report:
      precision    recall  f1-score   support

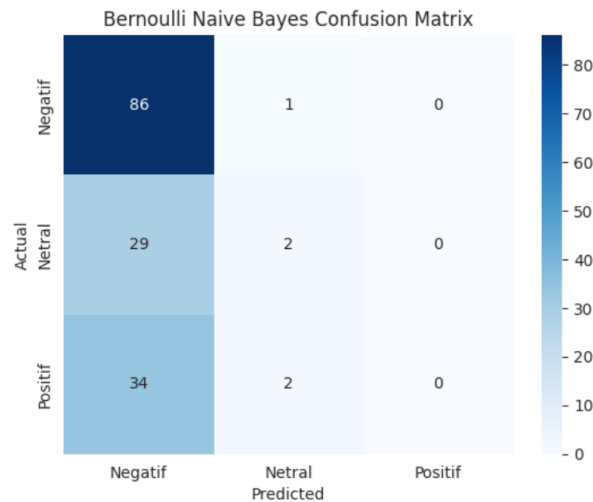
   Negatif      0.58      0.99      0.73      87
    Netral      0.40      0.06      0.11      31
    Positif      0.00      0.00      0.00      36

 accuracy      0.57      0.57      0.57      154
 macro avg      0.33      0.35      0.28      154
 weighted avg      0.41      0.57      0.43      154

```

Gambar 10. Hasil Test Report Bernaouli Naive Bayes

Menampilkan pula jumlah hasil sentimen benar dan salah pada setiap kelas sentimen positif dan negatif dengan menggunakan plot confusion matrix dengan hasil berupa plot visualisasi, disajikan pada gambar



Gambar 11. Hasil Confusion Matrix Bernaouli Naive Bayes

Berdasarkan hasil confusion matrix pada gambar di atas, untuk klasifikasi Naïve Bayes model Bernoulli didapatkan bahwa 0 analisis sentimen yang benar pada sentimen positif (True Positive), 0 analisis sentimen yang salah pada sentimen positif (False Positive). 34 analisis sentimen yang salah pada sentimen negatif (False Negative), 86 analisis sentimen yang benar pada sentimen negatif (True Netral). 2 analisis sentimen yang benar pada sentimen positif (True Positive), 1 analisis sentimen yang salah pada sentimen positif (False Positive).

C. Klasifikasi menggunakan Algoritma Multinomial Naive Bayes

Dalam penelitian ini, data uji dan data pengujian dibagi menggunakan rasio 80% untuk data pelatihan dan 20% untuk data pengujian. Hasil dari perhitungan yang telah diolah, didapatkan tingkat performa klasifikasi Naïve Bayes model Multinomial adalah sebesar 65%. Hasil ini ditunjukkan dalam classification report ditampilkan pada gambar

```

Multinomial Naive Bayes Accuracy: 0.6558441558441559
Multinomial Naive Bayes Classification Report:
      precision    recall  f1-score   support

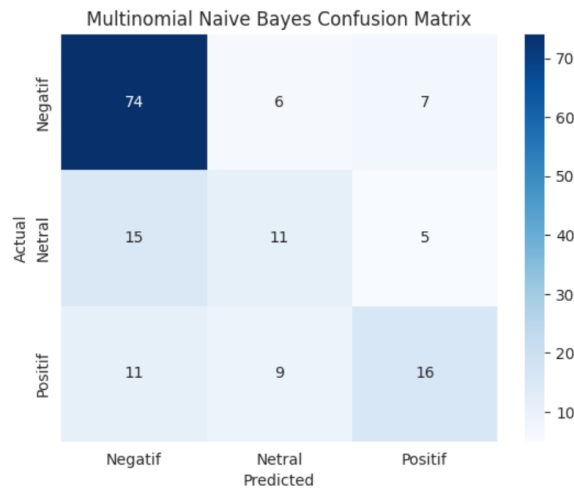
 Negatif      0.74      0.85      0.79         87
  Netral      0.42      0.35      0.39         31
  Positif      0.57      0.44      0.50         36

 accuracy          0.66         154
 macro avg      0.58      0.55      0.56         154
 weighted avg    0.64      0.66      0.64         154

```

Gambar 12. Hasil Test Report Multinomial Naive Bayes

Menampilkan pula jumlah hasil sentimen benar dan salah pada setiap kelas sentimen positif dan negatif dengan menggunakan plot confusion matrix dengan hasil berupa plot visualisasi, disajikan pada gambar



Gambar 13. Hasil Confusion Matrix Multinomial Naive Bayes

Berdasarkan hasil confusion matrix pada gambar di atas, untuk klasifikasi Naive Bayes model Multinomial didapatkan bahwa 16 analisis sentimen yang benar pada sentimen positif (True Positive), 7 analisis sentimen yang salah pada sentimen positif (False Positive). 11 analisis sentimen yang salah pada sentimen negatif (False Negative), 74 analisis sentimen yang benar pada sentimen negatif (True Netral). 9 analisis sentimen yang benar pada sentimen positif (True Positive), 6 analisis sentimen yang salah pada sentimen positif (False Positive)

Dari hasil pengujian dapat kita lihat hasil performa masing-masing model dalam mencari accuracy, precision, f1-score, dan recall dengan masing-masing parameter yang berbeda. Berikut adalah table rinciannya.

Tabel 1. Hasil performa Gaussian Naive Bayes

Model	Sentiment	Presisi	Recall	F1-Score	Akurasi
Gaussian Naive Bayes	Positive	0.28	0.22	0.25	0.5
Gaussian Naive Bayes	Neutral	0.36	0.29	0.32	
Gaussian Naive Bayes	Negative	0.60	0.69	0.64	

Tabel 1. Hasil performa Bernoulli Naive Bayes

Model	Sentiment	Presisi	Recall	F1-Score	Akurasi
Bernoulli Naive Bayes	Positive	0.00	0.00	0.00	0.57
Bernoulli Naive Bayes	Neutral	0.40	0.06	0.11	
Bernoulli Naive Bayes	Negative	0.58	0.99	0.73	

Tabel 2. Hasil performa Multinomial Naive Bayes

Model	Sentiment	Presisi	Recall	F1-Score	Akurasi
Multinomial Naive Bayes	Positive	0.57	0.44	0.50	

Multinomial Naïve Bayes	Neutral	0.42	0.35	0.39	0.65
Multinomial Naïve Bayes	Negative	0.74	0.85	0.79	

Dalam menganalisis perbandingan hasil dari tiga algoritma klasifikasi teks Naive Bayes pada dataset Gencatan Senjata Palestina- Israel , kami telah melakukan evaluasi terhadap performa masing-masing model. Dalam tulisan ini, kami akan memaparkan temuan yang signifikan dari perbandingan kinerja Model Multinomial Naive Bayes (MNB), Bernoulli Naive Bayes (BNB), dan Gaussian Naive Bayes (GNB) berdasarkan beberapa aspek, termasuk akurasi, presisi, dan recall. klasifikasi Naïve Bayes model Multinomial memiliki akurasi tertinggi sebesar 0.65 dengan test_size 0.2.

Kesimpulan

Pada penelitian ini menghasilkan kesimpulan bahwa berdasarkan hasil analisis sentimen tweet pada dataset yang digunakan dalam mengimplementasikan algoritma Naïve Bayes yang menggunakan 3 model dari algoritma Naïve Bayes yaitu Gaussian, Multinomial, dan Bernoulli menunjukkan hasil yang berbeda yaitu sebesar 50% untuk Naïve Bayes model Gaussian, 57% untuk Naïve Bayes model Bernoulli, dan Naïve Bayes model Multinomial adalah sebesar 65%. Hal ini menunjukkan bahwa model Multinomial Naïve Bayes lebih baik dari pada model lain dalam mengklasifikasikan data pada kasus ini.

Referencesi

- Irsyad H, & Taqwiym, A. (2021). Sentimen Analisis Masyarakat Terhadap Rakyat Palestina dengan Klasifikasi Naive Bayes. *Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem & Komputer*, Vol. 1 / No. 2. 167-168
- Astari, N., Divayana, D., & Indrawan, G. (2020). Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier. *JURNAL SISTEM DAN INFORMATIKA (JSI)*, 23.
- Salsabila, (2022). ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP TOKOH GUS DUR MENGGUNAKAN METODE NAÏVE BAYES DAN SUPPORT VECTOR MACHINE (SVM) . 4.
- Ahmad Rifai, “ANALISIS SENTIMEN OPINI MASYARAKAT TENTANG PENYAKIT HEPATITIS AKUT MENGGUNAKAN METODE NAÏVE BAYES,” *J. RESTI*, vol. 6, p. 22, 2022.
- Sukanto, & Shalahuddin, M. (2013). *Rekayasa Perangkat Lunak Terstruktur Dan Berorientasi Objek*. Bandung: Informatika, 1-9.
- Derajad Wijaya, H., & Dwiasnati, S. (2020). Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat. *JURNAL INFORMATIKA*, 7(1). <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>.
- Dwianto, E., & Sadikin, M. (n.d.). Analisis Sentimen Transportasi Online pada Twitter Menggunakan Metode Klasifikasi Naïve Bayes dan Support Vector Machine.

- Khoirul, M., Hayati, U., & Nurdiawan, O. (2023). ANALISIS SENTIMEN APLIKASI BRIMO PADA ULASAN PENGGUNA DI GOOGLE PLAY MENGGUNAKAN ALGORITMA NAIVE BAYES. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 1).
- Miftahusalam, A., Febby Nuraini, A., Khoirunisa, A. A., & Pratiwi, H. (n.d.). Perbandingan Algoritma Random Forest, Naïve Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Penghapusan Tenaga Honorer.
- Naufal, M. F., Arifin, T., & Wirjawan, H. (n.d.). Analisis Perbandingan Tingkat Performa Algoritma SVM, Random Forest, dan Naïve Bayes untuk Klasifikasi Cyberbullying pada Media Sosial. 8, 82. <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>
- Ruger, A. H., Suyanto, M., & Kurniawan, M. P. (2021). Sentimen Analisis Pelanggan Shopee di Twitter dengan Algoritma Naive Bayes. *Journal of Information Technology*, 1(2), 26-29.
- Tanggraeni, A. I., & Sitokdana, M. N. (2022). Analisis Sentimen Aplikasi EGovernment Pada Google Play Menggunakan Algoritma Naïve Bayes. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 9(2), 785-795.
- Das, S., Bhattacharyya, K., & Sarkar, S. (2023). Performance Analysis of Logistic Regression, Naive Bayes, KNN, Decision Tree, Random Forest and SVM on Hate Speech Detection from Twitter. *International Research Journal of Innovations in Engineering and Technology (IRJIET)*, 7(3), 24–28. <https://doi.org/10.47001/IRJIET/2023.703004>
- Duei Putri, D., Nama, G. F., & Sulistiono, W. E. (2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, 10(1). <https://doi.org/10.23960/jitet.v10i1.2262>.

Alfian Noer Halim
Fakultas Ilmu Komputer
Universitas Mercu Buana, Jakarta Barat

Saruni Dwiasnati
Fakultas Ilmu Komputer
Universitas Mercu Buana, Jakarta Barat

