

Klasifikasi Sentimen Masyarakat di Twitter terhadap Ganjar Pranowo dengan Metode K-Nearest Neighbor

¹Sayed Omas Tutus Arifta, ²Yusra, ³Muhammad Fikry

^{1,2,3}Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

¹11950111739@students.uin-suska.ac.id, ²yusra@uin-suska.ac.id, ³muhammad.fikry@uin-suska.ac.id

Article Info

Article history:

Received, 2023-06-23

Revised, 2023-06-27

Accepted, 2023-06-30

Kata Kunci:

Ganjar Pranowo
Klasifikasi Sentimen
K-Nearest Neighbor
Preprocessing
Twitter

Keywords:

Ganjar Pranowo
Sentiment Classification
K-Nearest Neighbor
Preprocessing
Twitter

ABSTRAK

Ada banyak jenis media sosial untuk mengumpulkan informasi, berbagi informasi dan berbagi berita, salah satunya adalah Twitter. Dalam penelitian ini, klasifikasi berbasis sentimen dilakukan dalam dua kategori, yaitu positif dan negatif, dengan mengaplikasikan metode *K-Nearest Neighbor* pada sosok gubernur Jawa Tengah yaitu bapak Ganjar Pranowo. *K-Nearest Neighbor* adalah metode pengklasifikasian objek berdasarkan data *training* yang menggunakan jarak atau kemiripan terkecil dari objek tersebut. Pada tahap pembelajaran, algoritma ini hanya menyimpan vektor karakteristik dan mengklasifikasikan data pembelajaran. Selama tahap klasifikasi, fitur-fitur yang sama dihitung untuk data uji, yang mana kelasnya tidak diketahui. Jarak dari vektor baru ini ke vektor data pelatihan dihitung dan *K* berikutnya diambil. Penelitian ini bertujuan untuk mendapatkan nilai akurasi menggunakan 4.000 data dengan sentimen negatif dan positif yang masing-masing berjumlah 2.000. Setelah data *tweet* berhasil diambil dari Twitter, data tersebut masih berbentuk mentah dan memerlukan tahap *preprocessing* untuk menghasilkan data yang bersih dan siap untuk diproses pada tahap selanjutnya. Perhitungan nilai akurasi dengan mengklasifikasi sentimen masyarakat di Twitter terhadap Ganjar Pranowo menggunakan metode *K-Nearest Neighbor* dalam menguji akurasi menghasilkan nilai akurasi yang cukup baik yaitu 81% *precision* 81% *recall* 81%.

ABSTRACT

There are many types of social media to gather information, share information and share news, one of which is Twitter. In this study, sentiment-based classification is carried out in two categories, namely positive and negative, by applying the *k-Nearest Neighbor* method to the figure of the governor of Central Java, Mr. Ganjar Pranowo. *K-Nearest Neighbor* is a method of classifying objects based on training data that uses the smallest distance or similarity from the object. At the learning stage, this algorithm stores only characteristic vectors and classifies the learning data. During the classification stage, the same features are calculated for the test data, the class of which is unknown. The distance from this new vector to the training data vector is calculated and the next *K* is taken. This study aims to obtain the value of accuracy using 4,000 data with negative and positive sentiment, each of which amounted to 2,000. After the tweet data is successfully retrieved from Twitter, the data is still raw and requires a preprocessing stage to produce clean data and ready for processing at a later stage. Calculation of the value of accuracy by classifying public sentiment on Twitter against Ganjar Pranowo using the *K-Nearest Neighbor* method in testing accuracy produces a pretty good accuracy value of 81% precision 81% recall 81%.

This is an open access article under the [CC BY-NC-ND](https://creativecommons.org/licenses/by-nc-nd/4.0/) license.



Penulis Korespondensi:

Sayed Omas Tutus Arifta,
Program Studi Teknik Informatika,
Universitas Islam Negeri Sultan Syarif Kasim Riau,
Email: 11950111739@students.uin-suska.ac.id

1. PENDAHULUAN

Ada banyak jenis media sosial untuk mengumpulkan informasi, berbagi informasi dan berbagi berita, salah satunya ialah Twitter. Twitter merupakan salah satu media sosial yang paling banyak digunakan diantara media sosial lainnya, dan sejak didirikan pada tahun 2006, jumlah pengguna Twitter terus bertambah. Indonesia menempati urutan kelima pengguna Twitter terbanyak, dengan total 19,5 juta pengguna dari total lebih dari 200 juta total pengguna aktif di seluruh dunia. Menurut studi tahun 2017 oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), penetrasi pengguna internet di Indonesia mencapai 143,26 juta dari total populasi 262 juta di Indonesia. Jumlah itu meningkat 10,56 juta dari tahun 2016 menjadi hanya 132,7 juta. Menurut jajak pendapat yang dilakukan oleh APJII pada tahun yang sama, banyak orang Indonesia menggunakan internet untuk berpartisipasi dalam media sosial, terhitung 87,13% [1].

Ganjar Pranowo, seorang politikus yang memegang jabatan sebagai Gubernur Jawa Tengah selama dua periode, lahir pada tanggal 28 Oktober 1968. Ia berasal dari keluarga sederhana dan dibesarkan di sebuah desa di lereng Gunung Lawu di Karanganya. Ayahnya bernama Sugeng Pramudji dan ibunya bernama Sri Suparmi. Ganjar Pranowo adalah anak kelima dari enam bersaudara. Pendidikan formalnya dimulai di SD Kutoarjo, Purworejo, dan dilanjutkan di SMP Negeri 1 Kutoarjo. Ia kemudian melanjutkan pendidikan menengahnya di SMA BOPKRI 1 Yogyakarta. Setelah lulus SMA, Ganjar melanjutkan pendidikan tinggi di Fakultas Hukum Universitas Gadjah Mada. Selama masa kuliahnya di Universitas Gadjah Mada, ia aktif dalam organisasi Gerakan Mahasiswa Nasional Indonesia (GMNI), yang kemudian berganti nama menjadi GeMiNi, dan juga bergabung dengan Majestic 55, sebuah kelompok mahasiswa pecinta alam di fakultas hukum UGM[2].

Dalam data *mining*, terdapat metode klasifikasi yang bertujuan untuk memprediksi label kategori yang tidak diketahui sebelumnya, sehingga dapat membedakan objek satu dengan yang lain berdasarkan fitur atau atribut yang dimiliki. Salah satu teknik klasifikasi yang populer dan mudah diimplementasikan adalah *K-Nearest Neighbor* (K-NN)[3]. Dalam penelitian ini, diterapkan klasifikasi sentimen pada sosok Ganjar Pranowo, gubernur Jawa Tengah, ke dalam dua kategori, yaitu positif dan negatif, menggunakan metode K-NN[4].

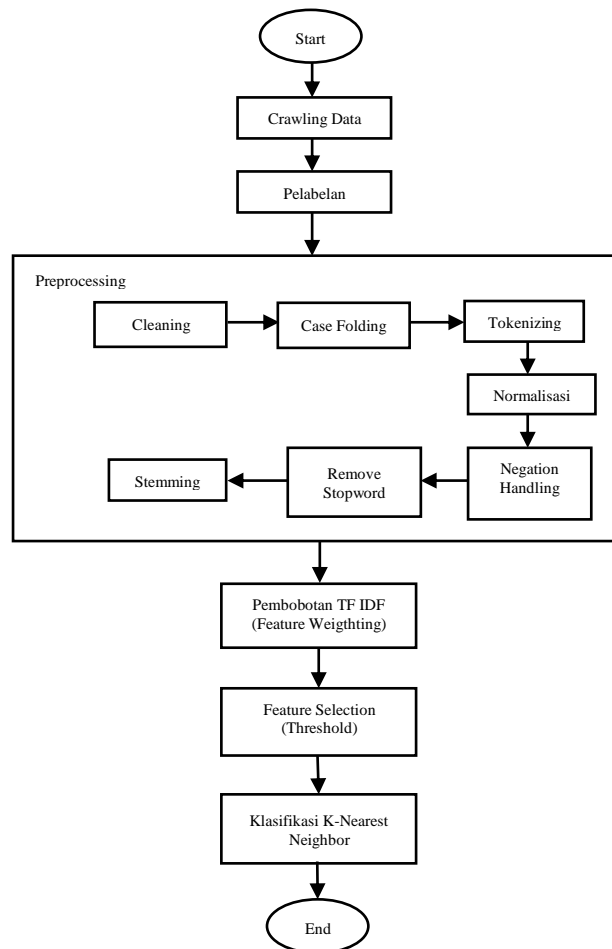
Metode *K-Nearest Neighbor* (KNN) diterapkan untuk mengklasifikasikan objek berdasarkan data pelatihan dengan menggunakan jarak atau kemiripan terkecil antara objek tersebut. Selama tahap pembelajaran, algoritma ini menyimpan vektor karakteristik dari data pelatihan dan melakukan klasifikasi berdasarkan data tersebut. Pada tahap klasifikasi, fitur yang sama diukur untuk data uji yang ingin diklasifikasikan (kelasnya belum diketahui). Selisih antara vektor baru ini dengan vektor data pelatihan dihitung, dan K tetangga terdekat yang memiliki jarak terkecil diambil. Titik yang prediksi klasifikasi barunya adalah yang paling banyak terdapat dalam klasifikasi titik-titik tersebut[5]. Penelitian pada kasus ini menggunakan 3000 data *tweet* dengan kunci "Ganjar Pranowo" yang dikumpulkan dari tanggal 29 November 2022 sampai tanggal 04 Desember 2022.

Penelitian terkait sentiment dengan metode *K-Nearest Neighbor* ini telah dilakukan oleh peneliti-peneliti sebelumnya di antaranya adalah oleh Nurcholis Geofany, Tommy, Risiko Liza yang membahas tentang klasifikasi sentimen *tweet* pada Twitter terhadap pembelajaran e-learning menggunakan metode *K-Nearest Neighbor* mendapatkan hasil akurasi terbaik sebesar 53,03% dengan perbandingan data latih dan data uji 80:20[6]. Selanjutnya penelitian dari Abdul Malik Zuhdi, Ema Utami, Suwanto Raharjo dengan pembahasan analisis sentiment Twitter terhadap capres indonesia 2019 dengan metode K-NN mendapatkan hasil akurasi klasifikasi sentimen mencapai 83,33% [7]. Kemudian penelitian yang dilakukan oleh Meidian Saputra yang membahas klasifikasi sentimen masyarakat terhadap kebijakan pembatasan sosial DKI Jakarta di Twitter menggunakan metode *K-Nearest Neighbor* memiliki akurasi tertinggi sebesar 87,50% [8].

Dalam *supervised learning*, semua objek dalam data memiliki fitur yang merupakan karakteristik unik dari objek tersebut. Fitur-fitur ini digunakan sebagai input untuk mengklasifikasikan objek ke dalam kelas yang sesuai. Dalam metode ini, kelas dari setiap objek sudah diketahui sebelumnya, sehingga tujuan utama adalah memetakan objek ke dalam kelas yang benar berdasarkan fitur-fiturnya. Proses pelatihan melibatkan penggunaan algoritma tertentu untuk mengenali pola-pola dalam data pelatihan dan membangun model yang dapat melakukan prediksi akurat untuk objek-objek baru. Penting untuk memperhatikan kualitas data pelatihan yang digunakan, karena data yang tidak representatif atau label yang salah dapat mempengaruhi kinerja model. *Supervised learning* telah digunakan dalam berbagai bidang seperti pengenalan pola, pengolahan citra, dan analisis sentimen, memungkinkan kita untuk mengklasifikasikan objek dengan akurasi berdasarkan fitur-fitur yang dimilikinya[9].

2. METODE PENELITIAN

Tahapan mengklasifikasikan opini publik di Twitter terhadap Ganjar Pranowo pada Gambar 1 sebagai berikut.



Gambar 1. Desain Penelitian

Dalam penelitian ini, data dikumpulkan menggunakan metode *crawling*. Dalam implementasinya, peneliti menggunakan *Python* sebagai bahasa pemrograman utama untuk mengembangkan program yang dapat melakukan *crawling*[10]. Dengan menggunakan *Python*, peneliti dapat mengakses situs *web* yang dituju dan mengambil data yang relevan sesuai dengan kebutuhan penelitian. Selanjutnya adalah melakukan pelabelan data untuk memberikan klasifikasi atau kategori kepada setiap data dalam *dataset*. Setiap data diidentifikasi dengan label atau *tag* yang menggambarkan kelas atau kategori yang sesuai[11].

Setelah melakukan *crawling* dan pelabelan data *tweet* dari Twitter, data yang diperoleh masih berupa data mentah. Oleh sebab itu, dibutuhkan tahapan *preprocessing* untuk membersihkan data tersebut sehingga siap untuk diproses pada tahap berikutnya[12]. Tahapan pertama yang dilakukan dalam proses *preprocessing* adalah *cleaning*. Dalam hal ini, dilakukan penghapusan kata yang tidak relevan seperti URL, *hashtag* (#), *username* (@*username*), *email*, emotikon (: @,: *,: D), serta tanda baca seperti koma (,), titik (.), dan tanda baca lainnya[13]. Tahapan selanjutnya adalah melakukan *caase folding*, di mana tahapan ini adalah proses mengganti semua huruf dalam setiap dataset menjadi huruf kecil. Hanya huruf dari 'a' hingga 'z' yang diproses, sedangkan karakter selain huruf akan dihapus dan dinyatakan sebagai delimiter[14]. Tahapan ketiga dalam *preprocessing* adalah *tokenizing* yang melibatkan pemisahan string input berdasarkan kata-kata yang ada. Sebagai contoh, karakter spasi, tabulasi, atau baris baru dianggap sebagai jarak antar kata. Namun, karakter seperti tanda petik tunggal ('), titik (.), semikolon (;), titik dua (:), atau karakter lainnya juga bisa berperan sebagai pemisah kata tergantung pada konteksnya. Tahapan keempat adalah melakukan normalisasi, melibatkan konversi data yang tidak baku jadi bentuk kata baku[15]. Hal ini perlu diterapkan karena banyak komentar yang memakai kalimat-kalimat yang tidak baku, sehingga menyulitkan pengujian data. Selanjutnya adalah tahapan *negation handling* yang bertujuan untuk merubah polaritas kata kedalam kalimat yang beda. Tahapan keenam adalah *remove stopwords* di mana ini merupakan proses seleksi kata-kata penting setelah

melakukan tokenisasi. Terdapat dua pendekatan yang dapat digunakan dalam tahap ini, yaitu mengaplikasikan algoritma *stoplist* untuk menghapus kata-kata yang tidak penting atau menggunakan *wordlist* untuk mengambil hanya kata atau token yang dianggap relevan[1]. Dan tahapan terakhir pada *preprocessing* adalah *stemming* yang memiliki peran penting tidak hanya dalam mengurangi jumlah indeks yang berbeda dalam sebuah dokumen, tetapi suatu proses pengelompokan kata-kata dengan akar kata dan makna yang sama, tetapi memiliki bentuk atau tata bahasa yang berbeda karena variasi afiks (imbuhan) dapat dilakukan. Dalam hal ini, kata-kata tersebut memiliki akar kata yang serupa dan arti yang mirip, namun mungkin memiliki perbedaan dalam struktur atau imbuhan yang digunakan[16].

Setelah melakukan *preprocessing*, proses selanjutnya adalah pembobotan TF-IDF. *Term Frequency — Inverse Document Frequency* (TF-IDF) merupakan algoritma yang digunakan untuk mengukur bobot setiap token yang ada dalam dokumen. TF-IDF sering digunakan untuk memberikan karakteristik pada dokumen[4]. Metode ini akan menghitung seberapa sering sebuah token muncul dalam sebuah dokumen. *Term Frequency* (TF) adalah banyaknya i dalam data j , lalu hasilnya dibagi dengan total *term* yang ada dalam data j . Berikut adalah rumus yang diterapkan dalam menghitung nilai IDF.

$$tf_{ij} = \frac{f_{d(i)}}{\max_{j \in d} f_{d(i)}} \quad (1)$$

Inversed Document Frequency (IDF) memiliki tujuan untuk mereduksi bobot pada term jika keberadaannya terdapat pada seluruh dokumen. Berikut adalah rumus yang diterapkan dalam menghitung nilai IDF.

$$idf(t, D) = \log\left(\frac{N}{df(i)}\right) \quad (2)$$

Setelah melakukan pembobotan TF-IDF, selanjutnya dilakukan proses *feature selection* yang tujuan utamanya untuk meningkatkan efisiensi dan efektivitas pengklasifikasi dengan mengurangi jumlah data yang perlu dianalisis. Caranya adalah dengan mengidentifikasi fitur-fitur yang paling relevan dalam proses pengklasifikasian data. Dengan memilih fitur-fitur yang signifikan, kita dapat mengurangi kompleksitas data dan memfokuskan analisis pada informasi yang paling penting[17].

Proses terakhir dalam penelitian ini adalah melakukan klasifikasi dengan menggunakan metode K-NN (*K-Nearest Neighbor*). Metode K-NN merupakan suatu metode klasifikasi yang menggunakan data *training* yang telah terklasifikasikan sebelumnya untuk mengklasifikasikan sekumpulan data[18]. Persamaan yang digunakan untuk K-NN ditunjukkan di bawah ini.

$$d_i = \sum_{i=1}^p (x_{2i} - x_{1i})^2 \quad (3)$$

3. HASIL DAN PEMBAHASAN

Penelitian terhadap klasifikasi sentimen masyarakat di Twitter terhadap Ganjar Pranowo menghasilkan suatu akurasi yang melalui beberapa tahapan yaitu *crawling* data, pelabelan, *preprocessing*, TF-IDF, *feature selection* dan klasifikasi K-NN. Uraian setiap tahapannya sebagai berikut.

3.1 *Crawling* Data

Crawling data merupakan tahapan untuk mengumpulkan data opini publik terhadap Ganjar Pranowo di Twitter dengan mengaplikasikan API yang tersedia. Dataset yang dipakai dalam penelitian sejumlah 4.000 data dan menggunakan dua kelas pelabelan yaitu positif dan negatif. Setiap kelas memiliki 2.000 data yang diambil pada tanggal 29 November 2022 sampai 03 Maret 2023. Tabel 1 contoh data *tweet* yang akan diproses.

Tabel 1. Contoh Data *Tweet*

Label	<i>Tweet</i>
Positif	b'@GanjaranApp @ganjarpranowo Pak Ganjar itu pemimpin yang hebat dan sangat jelas programnya makanya memiliki pendukung yang banyak'
Positif	b'@ganjarpranowo Alhamdulillah semoga bisa membawa manfaat dan bisa selalu menciptakan generasi berprestasi untuk masa depan bangsa. Sukses selalu untuk pemerintah Jawa Tengah'
Negatif	b'@ganjarpranowo gagal utk mengendalikan banjir di jateng'
Negatif	b'@ch_chotimah2 @aniesbaswedan @ganjarpranowo Hehhhhh,,,2 preode jdi gubernur msih ttep termiskin dijdnteng'

Menampilkan semua *data set* yang digunakan dalam proses pengklasifikasian dengan *source code load* data. Gambar 2 tampilan *load* data pada *tools google colab*.

	Text	Kelas
0	b'Jangan Over Apriori bung, \n\nTerkadang\nOra...	POSITIF
1	b'Ganjar Turun Tangan..!!!\n#ganjarpranowo\n#g...	POSITIF
2	b'Sy dukung Ganjar Pranowo sebagai Presiden RI...	POSITIF
3	b'Ganjar Pranowo harapan depan...'	POSITIF
4	b'Tingkat kemiskinan di Jateng turun berkat Pa...	POSITIF
...
3995	b'Kandidat Lain Kudu Wanti-wanti! Survei Voxpo...	NEGATIF
3996	b'Faktanya survei PWS 11 Nov 22 utk\npemilih w...	NEGATIF
3997	b'Jangan karena nafsu terhadap radikal yg keli...	NEGATIF
3998	b'Ayo gass lur..mbok rak kebagen,mengko nyesel...	NEGATIF
3999	b'Ayoooo jangan baperan ya !!! @psi_id @Giring...	NEGATIF

4000 rows × 2 columns

Gambar 2. Tampilan Load Data

3.2 Preprocessing Data

Preprocessing data adalah proses untuk mengganti data awal menjadi data yang lebih mudah untuk dipahami pada tahapan klasifikasi[19]. Penelitian ini menggunakan tujuh tahapan dalam preprocessing data, di antaranya sebagai berikut.

- Cleaning adalah proses untuk membersihkan data dari simbol, link, emoji, tanda baca, dan sebagainya.
- Case Folding adalah proses mengganti huruf kapital menjadi huruf kecil.
- Tokenizing adalah proses membentuk data menjadi sekumpulan token-token.
- Normalisasi adalah tahapan untuk memperbaiki ejaan yang salah.
- Negation Handling merupakan tahapan mengubah kata negasi dan satu kata setelahnya menjadi satu kata baru yang tidak merubah sentimen.
- Remove Stopword adalah proses untuk menghilangkan kata yang kurang penting atau kata penghubung[20].
- Stemming adalah proses mengubah setiap token menjadi kata dasar.

Setelah data yang akan digunakan terkumpul dan telah ditentukan kelas sentimen, selanjutnya mengaplikasikan preprocessing data. berikut Gambar 3 tampilan hasil preprocessing data pada tools google colab.

index	Text	Kelas
0	apriori terkadang benci alas jalan rangkul jatuh kritik gagasannya personinya	POSITIF
1	ganjar turun pranowo	POSITIF
2	dukung ganjar pranowo presiden kiru milu 2024 kerja nasionalisme personalitas pribadi kali kali	POSITIF
3	ganjar pranowo harap	POSITIF
4	tingkat miskin jateng turun berkat ganjar pranowo	POSITIF
5	bagaimana rakyat ganjar jago	POSITIF
6	lelehan wans pilih mas kakak anis kiri antri kakak sumur resap truk ambias deal kakak	POSITIF
7	aku dukung den jateng mem banjir capres duduk mikir	POSITIF
8	pilih ganjar	POSITIF
9	ndhes semarang banjir perum dinar indah kadrun nyalahke ganjar pranowo ganjar duduk gubernur semarang semarang nduwe walikota takon walikotane sisi ngerti	POSITIF
10	prpun ndoro	POSITIF
11	elaah sisi inovasi maju masyarakat jateng gerak ganjar suka kali pancing kali masyarakat kali ikan bukti ganjar tingkat skill ahli masyarakat jalan	POSITIF
12	persis mending data aaja turun miskin jalan unggul provinsi kali mudah kerja keras lindung sejahtera masyarakat nyata	POSITIF
13	dik nasionalis soekarnois marhaenis 1992 terjun dunia politik naung juang memblenger mata mata orba paham karakter pemimpin insya Allah jalan presiden indonesia terimakasih	POSITIF
14	sisi kali lebay cubit	POSITIF
15	hasil kerja nyata prestasi	POSITIF
16	dukung mendo jalan kandidat kuat milu 2024 unggah video kiprah warga indonesia sisi wujud kerja nyata beliau jalan using	POSITIF
17	timpa bencana kunjung ganjar pranowo senang jalan nampak bahagia	POSITIF
18	kali dukung ganjar pranowo capres 2024 yuks pelit kali jempol kamoh kesini yaaak rapat baris gais	POSITIF
19	terima kali	POSITIF
20	sisi erti	POSITIF
21	duit triliun mending warga contoh	POSITIF
22	wajar elektabilitasnya kiri samping keras dukung pema melebihi2kan sebar hoax prestasi beda sisi nganu ikut nyampah henti next	POSITIF
23	ganjar lanjut estafet hasil pemimpin jokowi elektabilitas korelasi tingkat puas masyarakat kerja jokowi bukti ganjar terus jokowi 2024	POSITIF
24	neko neko ngadi ngadi ganjar pranowo erick thohir capres cawapres 2024 idam gas	POSITIF

Show 25 | per page

Gambar 3. Tampilan Hasil Preprocessing

	00	000	01	02	03	04	05	056	06	07	...	ziarah	zikmal	zim	zodiak	zombie	zon	zoom	zubair	zulkifli	Kelas		
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	
...
3995	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF
3996	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF
3997	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF
3998	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF
3999	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF

4000 rows x 5256 columns

Gambar 6. Tampilan Hasil TF-IDF

3.4 Feature Selection

Feature Selection adalah tahapan yang berfungsi untuk mengurangi jumlah fitur yang kurang relevan dalam menentukan nilai kelas target dengan menggunakan nilai batas ambang [24]. Fitur yang kurang relevan akan diabaikan dalam penentuan kelas target. Berikut tampilan hasil feature selection dengan nilai *thresholding* = 0,001 dan hasil yang diperoleh dari feature selection 4.000 rows dan 156 columns seperti Gambar 7.

Hasil Feature Selection																							
	2024	aamin	ajar	alam	alhamdulillah	anak	anies	apresiasi	atas	bagus	...	ubah	udah	ulama	umkm	untung	upaya	urus	usaha	warga	Kelas		
0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	
1	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	
2	0.221937	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	
3	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	
4	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	
...
3995	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.172432	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF
3996	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.132960	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF
3997	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF
3998	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF
3999	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF

4000 rows x 156 columns

Gambar 7. Tampilan Feature Selection

3.5 Klasifikasi

Klasifikasi adalah tahapan untuk melakukan pengujian data yang telah melalui beberapa tahapan sebelumnya [25]. Tahapan klasifikasi menggunakan perbandingan tiga buah perbandingan yaitu 70:30, 80:20, dan 90:10. Didapatkan hasil terbesar pada perbandingan 90:10 dengan menggunakan nilai $K=9$ yaitu 83,0%. Sedangkan untuk 70:30 dengan nilai $K=9$ mendapatkan hasil 81,38% dan 80:20 dengan nilai K yang sama mendapatkan hasil 81,33%. Berikut pada Gambar 8 tampilan hasil dari klasifikasi dengan perbandingan 90:10.

Hasil Klasifikasi																							
	2024	aamin	ajar	alam	alhamdulillah	anak	anies	apresiasi	atas	bagus	...	udah	ulama	umkm	untung	upaya	urus	usaha	warga	Kelas Target	Hasil Klasifikasi		
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	NEGATIF	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.533887	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	POSITIF	
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	POSITIF	
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF	NEGATIF	
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	POSITIF	
...
395	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF	NEGATIF	
396	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF	POSITIF	
397	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	POSITIF	
398	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	NEGATIF	
399	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	POSITIF	

400 rows x 154 columns

Gambar 8. Tampilan Hasil Klasifikasi K-Nearest Neighbor

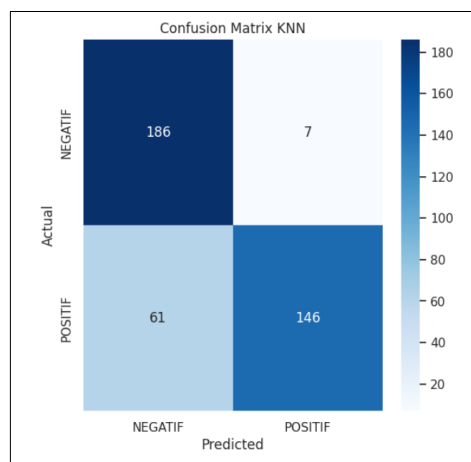
3.6 Hasil

Proses berikutnya adalah pengujian akurasi dari hasil klasifikasi dengan metode *K-Nearest Neighbor*. Berikut *confusion matrix* dari uji coba pada Tabel 2.

Tabel 2. Confusion Matrix

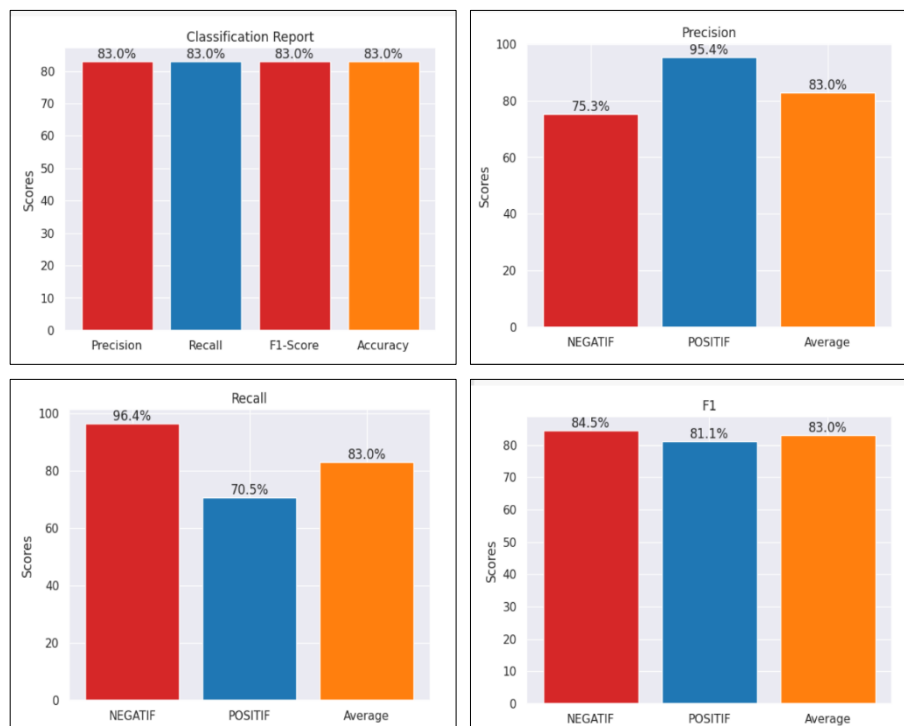
Confusion Matrix		
	Negatif	Positif
Negatif	189	7
Positif	61	146

Dari Tabel 3 dapat disimpulkan *False Positif* (FP) adalah jumlah data positif yang dianggap salah oleh model sedangkan *False Negatif* (FN) adalah jumlah data negatif yang salah pada pengklasifikasian. Di mana *True Positif* (TP) merupakan jumlah data positif bernilai benar dalam klasifikasi, maka *True Negatif* (TN) adalah jumlah data negatif yang bernilai benar dalam klasifikasi. Berikut tampilan grafik *confusion matrix* hasil pengujian pada gambar 7.



Gambar 9. Grafik Confusion Matrix

Hasil dari pengujian memperoleh *classification report*. Berikut Gambar 8 tampilan *classification report*.



Gambar 10. Classification Report

Perhitungan nilai akurasi dengan mengklasifikasikan sentimen masyarakat di Twitter terhadap Ganjar Pranowo menggunakan metode *K-Nearest Neighbor* dengan perbandingan 90:10 dan nilai $K=9$ menghasilkan nilai akurasi yang cukup baik yaitu 83% *precision* 83% *recall* 83%.

4. KESIMPULAN

Berdasarkan analisis dan pembahasan yang dilakukan, dapat diambil kesimpulan bahwa algoritma *K-Nearest Neighbor* lebih baik dalam melakukan klasifikasi sentimen masyarakat terhadap Ganjar Pranowo. Penelitian ini bertujuan untuk mendapatkan nilai akurasi menggunakan 4.000 data dengan sentimen negatif dan positif yang masing-masing berjumlah 2.000. Nilai akurasi yang dengan menggunakan metode KNN $K=9$ perbandingan 90:10 adalah 83%.

UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih kepada bapak Revi Kasman, S.Pd. selaku ahli bahasa yang telah membantu menyumbangkan pengetahuannya sehingga penulis dapat menyelesaikan penelitian ini.

REFERENSI

- [1] R. Arnopalindo, "Klasifikasi Sentimen Terhadap Ustadz Abdul Somad Dengan Metode K-Nearest Neighbor Pada Media Sosial Twitter," 2022.
- [2] A. Probawati, "Kepemimpinan ganjar pranowo," *Yogyakarta, Univ. Muhammadiyah*, no. June, pp. 0–15, 2021.
- [3] A. F. Rahman, "Klasifikasi Tweet di Twitter dengan Menggunakan Metode K-Nearest Neighbor," *J. Sistim Inf. dan Teknol.*, vol. 4, pp. 64–69, 2022.
- [4] R. Aditya, *Klasifikasi Sentimen Pelayanan Maskapai Penerbangan Indonesia Menggunakan Pembobotan Tf-Idf Dan Metode K-Nearest Neighbor (Knn)*, vol. 2, no. 1. 2020.
- [5] Z. U. Siregar, "Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor," vol. 8, pp. 81–92, 2019.
- [6] A. N. Geofany and R. Liza, "Klasifikasi Sentimen Tweet Pada Twitter Terhadap Pembelajaran E-Learning Menggunakan Metode k-Nearest Neighbor Klasifikasi Sentimen Tweet Pada Twitter Terhadap Pembelajaran," vol. 8, p. 380–385, 2021.
- [7] A. M. Zuhdi, E. Utami, and S. Raharjo, "Analisis Sentiment Twitter terhadap Capres Indonesia 2019 dengan Metode K-NN," vol. 5, pp. 1–7, 2019.
- [8] M. Saputra, "Klasifikasi Sentimen Masyarakat Terhadap Kebijakan Pembatasan Sosial Dki Jakarta Di Twitter Menggunakan Metode K-Nearest Neighbor," 2021.
- [9] A. I. Riaddy, Y. Sibaroni, S. Si, A. Aditsania, S. Si, and M. Si, "Ekstraksi Informasi pada Makalah Ilmiah dengan Pendekatan Supervised Learning Information Extraction on Scientific Papers with Supervised Learning Approach," vol. 3, no. 1, pp. 1184–1190, 2016.
- [10] D. A. N. Krisna and U. Salamah, "Perbandingan Algoritma Naïve Bayes Dan K-Nearest Neighbor Untuk Klasifikasi Berita Hoax Kesehatan Di Media Sosial Twitter," *J. Tek. Inform. Kaputama*, vol. 6, no. 2, pp. 836–845, 2022.
- [11] H. Rahman, "Klasifikasi Sentimen Masyarakat Terhadap Layanan Badan Penyelenggara Jaminan Sosial (Bpjs) Kesehatan Di Twitter Menggunakan Metode K-Nearest Neighbor," *J. Ekon. Vol. 18, Nomor 1 Maret201*, vol. 2, no. 1, pp. 41–49, 2020.
- [12] G. K. Pati and E. Umar, "Analisis Sentimen Komentar Pengunjung Terhadap Tempat Wisata Danau Weekuri Menggunakan Metode Naive Bayes Classifier Dan K- Nearest Neighbor," vol. 6, pp. 2309–2315, 2022.
- [13] F. Astuti, R. M. Candra, S. Agustian, and S. Ramadhani, "Klasifikasi Sentimen Masyarakat Terhadap Pemerintah Terkait Penerapan Kebijakan New Normal Menggunakan Metode K-Nearest Neighbor," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 3, pp. 531–538, 2022.
- [14] Irmawati and Sukemi, "Klasifikasi Sentimen Analisis Kartun Dengan Metode Optical Character Recognition (OCR) Dan Algoritma K-Nearest Neighbour (KNN)," *Pros. Annu. Res. Semin.*, vol. 5, no. 1, pp. 978–979, 2019.

- [15] E. W. Sholeha, S. Yunita, R. Hammad, V. C. Hardita, and K. Kaharuddin, "Analisis Sentimen Pada Agen Perjalanan Online Menggunakan Naïve Bayes dan K-Nearest Neighbor," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 3, no. 4, pp. 203–208, 2022.
- [16] H. Sibyan and N. Hasanah, "Analisis Sentimen Pada Wisata Dieng Dengan Algoritma K-Nearest Neighbor (K-NN)," vol. 9, no. 1, pp. 38–47, 2021.
- [17] S. Sahara and R. A. Permana, "Metode KNN Pada Sentiment Analisis Review Produk Game Android," vol. 11, no. 2, 2022.
- [18] V. Novalia, R. Goejantoro, and Sifriyani, "Perbandingan Metode Klasifikasi Naive Bayes dan K-Nearest Neighbor (Studi Kasus : Status Kerja Penduduk Di Kabupaten Kutai Kartanegara Tahun 2018)," *J. EKSPONENSIAL*, vol. 11, pp. 159–166, 2020.
- [19] A. Tangkelayuk, "The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 1109–1119, 2022.
- [20] M. W. Pertiwi, "Analisis Sentimen Opini Publik Mengenai Sarana dan Transportasi Mudik Tahun 2019 Pada Twitter Menggunakan Algoritma Naïve Bayes, Neural Network, K-NN dan SVM," *Inti Nusa Mandiri*, vol. 14, no. 1, pp. 27–32, 2019.
- [21] A. J. J. Limbong, I. Sembiring, and D. K. Hartomo, "Analisis Klasifikasi Sentimen Ulasan Pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes Dan K-Nearest Neighbor Analysis of Review Sentiment Classification on E-Commerce Shopee Word Cloud Based With Naïve Bayes and K-Nearest Neighbor Meth," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, pp. 347–356, 2019.
- [22] S. A. Pratomo, S. Al Faraby, and M. D. Purbolaksono, "Analisis Sentimen Pengaruh Kombinasi Ekstraksi Fitur TF-IDF dan Lexicon Pada Ulasan Film Menggunakan Metode KNN," *eProceedings Eng.*, vol. 8, no. 5, pp. 10116–10126, 2021.
- [23] S. Ernawati and R. Wati, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel," *J. Khatulistiwa Inform.*, vol. 6, no. 1, pp. 64–69, 2018.
- [24] A. Rozaqi, A. Triayudi, and R. T. Aldisa, "Analisis Sentimen Vaksinasi Booster Berdasarkan Twitter Menggunakan Algoritma Naïve Bayes dan K-NN," *J. Sist. Komput. dan Inform.*, vol. 4, no. 1, p. 184, 2022.
- [25] R. Gunawan, R. Septiadi, F. Apri Wenando, H. Mukhtar, and Syahril, "K-Nearest Neighbor (KNN) untuk Menganalisis Sentimen terhadap Kebijakan Merdeka Belajar Kampus Merdeka pada Komentar Twitter," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 3, no. 2, pp. 152–158, 2022.