

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

Yuda Zafitra Fadhlan<sup>1</sup>, Yusra<sup>2</sup>, and Muhammad Fikry<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim, Riau, Indonesia, 28293  
e-mail: <sup>1</sup>11950115235@students.uin-suska.ac.id, <sup>2</sup>yusra@uin-suska.ac.id, <sup>3</sup>muhammad.fikry@uin-suska.ac.id

Submitted Date: 2023-06-13  
Revised Date: 2023-06-28

Reviewed Date: 2023-06-23  
Accepted Date: 2023-06-29

### Abstract

*In welcoming the 2024 democratic party, many politicians have started campaigning in each region which has generated a lot of different positive and negative sentiments in every Indonesian society. Ganjar Pranowo is one of the politicians who will take part as a candidate for the 2024 presidential candidate which has made many netizens on Twitter give their opinions on him. The purpose of this study is to classify public sentiment on Twitter towards Ganjar Pranowo using 4000 tweet data. The classification is divided into two classes, namely positive and negative using the Modified K-Nearest Neighbor method combined with feature weighting, feature selection using a supervised learning approach. The results of this study after going through the stages of retrieval, data labeling, preprocessing, feature weighting, feature selection, MK-NN and evaluating accuracy get the highest accuracy value at 83.8% with a ratio of 90:10 with a value of  $k = 3$ .*

*Keywords: Ganjar Pranowo; Sentiment Classification; Modified K-Nearest Neighbor; Twitter*

### Abstrak

Dalam menyambut pesta demokrasi tahun 2024 banyak politisi mulai melakukan kampanye di setiap daerah yang menimbulkan banyak sentimen positif dan negatif yang berbeda pada setiap masyarakat Indonesia. Ganjar Pranowo merupakan salah satu politisi yang akan ikut andil sebagai bakal calon presiden 2024 yang membuat warganet di Twitter banyak yang memberikan opini terhadapnya. Tujuan penelitian ini adalah untuk mengklasifikasikan sentimen masyarakat di Twitter terhadap Ganjar Pranowo dengan menggunakan 4000 data *tweet*. Klasifikasi dibedakan menjadi dua kelas yaitu positif dan negatif menggunakan metode Modified K-Nearest Neighbor yang dikombinasi dengan *feature weighting*, *feature selection* menggunakan teknik pendekatan *supervised learning*. Hasil dari penelitian ini setelah melewati tahap dari pengambilan, pelabelan data, *preprocessing*, *feature weighting*, *feature selection*, MK-NN dan evaluasi akurasi mendapatkan nilai akurasi tertinggi di 83,8% dengan perbandingan 90:10 dengan nilai  $k=3$ .

Kata kunci: Ganjar Pranowo; Klasifikasi Sentimen; Modified K-Nearest Neighbor; Twitter

### 1. Pendahuluan

Pemilihan umum merupakan komponen penting dalam proses politik di Indonesia. Pada tahun 2024 mendatang, akan dilaksanakan pemilihan DPR, DPD, DPRD Provinsi, DPRD, Presiden, dan Wakil Presiden. Banyak tokoh politik telah mempersiapkan diri untuk berpartisipasi dalam politik tersebut. Para pemimpin politik melakukan berbagai upaya untuk meningkatkan

elektabilitas mereka. Elektabilitas yang merupakan ukuran popularitas seseorang, memiliki peran krusial dalam politik untuk meyakinkan masyarakat untuk memilih tokoh politik yang mereka sukai. Dalam musim pemilihan, elektabilitas menjadi indikator penting untuk mengukur tingkat popularitas seseorang. Menurut data dari Daftar Pemilih Sementara (DPS) Pemilu 2024 yang dikutip dari *website* resmi KPU, terdapat

sebanyak 205.853.518 pemilih. Menurut pengamat politik Ujang Komarudin dari Universitas Al Azhar Indonesia seperti yang dilaporkan oleh <https://news.republika.co.id/>, terdapat prediksi bahwa terdapat 3 calon presiden (capres) yaitu Ganjar Pranowo, Gubernur Jawa Tengah yang diukung oleh PDI Perjuangan (PDIP), Prabowo Subianto, Menteri Pertahanan yang berpotensi diukung oleh Partai Gerindra dan Partai Kebangkitan Bangsa (PKB), serta Anies Baswedan, mantan Gubernur DKI Jakarta yang diukung oleh Partai NasDem, Partai Keadilan Sejahtera (PKS), dan Partai Demokrat.

Opini merujuk pada pandangan, penilaian, atau pendapat subjektif seseorang atau kelompok mengenai suatu masalah, isu, atau topik tertentu. Opini bersifat subjektif karena didasarkan pada pemikiran, pengalaman, keyakinan, atau nilai-nilai pribadi individu atau kelompok tersebut. Pendapat ini mungkin didasarkan pada pengetahuan, informasi atau pengamatan seseorang, tetapi masih bersifat subyektif. Analisis sentimen adalah proses perbandingan yang mengidentifikasi dan memberi peringkat pendapat yang diungkapkan dalam teks untuk menentukan apakah tanggapan terhadap topik tertentu positif, negatif, atau netral (Muhammad, Rizqeya, Kevin, & Nur, 2020). Analisis sentimen juga disebut *opinion mining*, ini adalah proses mengenali polaritas dan membedakan karakteristiknya dalam bentuk klasifikasi dengan dua kategori, positif dan negatif. Menganalisis sentimen dalam teks-teks politik dapat menjadi representasi dari hasil pemilihan, elektabilitas kandidat, atau perwakilan politik nyata di Negara. Metode penelitian yang digunakan adalah Modified k Nearest Neighbour (M-KNN). Modified K-Nearest Neighbour adalah metode klasifikasi hasil dari pengembangan metode k-nearest neighbour untuk klasifikasi objek berdasarkan data latih dengan menggunakan jarak terdekat atau kemiripan dari objek tersebut. Hasil klasifikasi tergantung dari nilai K dan data yang telah dilatihkan. Dengan menggunakan metode M-KNN, sentimen pengguna akan diklasifikasikan menjadi nilai positif dan negatif. Ciri-ciri yang digunakan dalam proses klasifikasi adalah *term* atau kata-kata dalam komentar pengguna. Komentar melewati proses *preprocessing* sebelum didefinisikan sebagai fitur (Ahmad, Indriati, & Lailil, 2018). Keunggulan M-KNN dibandingkan KNN adalah akurasi yang lebih tinggi di mana M-KNN ditambahkan proses menghitung nilai validitas, yang berguna untuk mengatasi masalah

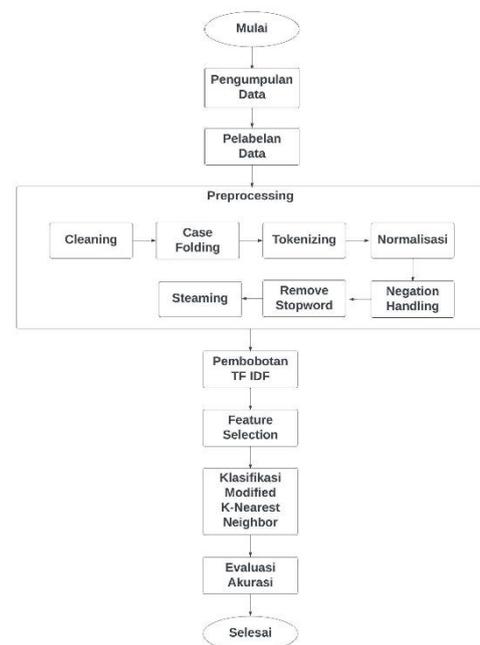
outlier dalam menghitung bobot pada KNN tradisional (Cantya, Yufis, & Nur, 2020).

Pada penelitian yang dilakukan oleh (Adella, Indriati, & Yuita, 2020) dengan objek ulasan pengguna MRT Jakarta mendapatkan nilai akurasi sebesar 0,86769 dan f-measure sebesar 0,86265. Pada penelitian yang dilakukan oleh (Cantya, Yufis, & Nur, 2020) dengan objek topik berita mendapatkan akurasi sebesar 72,5% dengan nilai K=7 pada perbandingan data 80:20. Pada penelitian yang dilakukan (Sofiah, Yusra, & Muhammad, 2023) penggunaan metode MK-NN mendapatkan tingkat akurasi tertinggi pada data opini sebesar 83.33% dengan perbandingan 90:10 dan K=3, Yang artinya baik dalam penelitian mengenai kenaikan BBM.

## 2. Metodologi Penelitian

Penelitian ini menggunakan 4000 *dataset* dengan menggunakan metode Modified K-Nearest Neighbor. Tujuan penelitian ini adalah untuk memberikan pemahaman yang lebih baik tentang bagaimana masyarakat memandang Ganjar Pranowo dan membantu menganalisis sikap masyarakat terhadapnya.

### 2.1. Diagram Alur Penelitian



Gambar 1. Alur dalam Penelitian

Alur penelitian yang meliputi pengumpulan data, pelabelan data, *preprocessing*, pembobotan TF-IDF, *feature selection*, klasifikasi Modified K-Nearest Neighbor, dan penilaian akurasi.

## 2.2. Pengumpulan Data

Data yang digunakan menggunakan 4000 data yang diambil di Twitter menggunakan API Twitter dengan menggunakan bahasa pemrograman *python* diambil pada tanggal 29 November 2022 sampai tanggal 03 Maret 2023. Pengambilan data pada Twitter dibatasi hanya pada *hashtag* Ganjar Pranowo.

## 2.3. Pelabalan Data

Pelabelan dilakukan secara manual di mana memiliki dua kelas yaitu positif dan negatif digunakan untuk melatih model dan mengklasifikasikan data yang belum diberi label. Dengan menggunakan kedua jenis pelabelan ini, MKNN dapat menghasilkan prediksi yang lebih akurat dalam mengklasifikasikan data baru. Pelabelan positif dan negatif memberikan informasi penting kepada algoritma MKNN tentang kelas yang ingin diidentifikasi dan kelas yang bukan target, sehingga membantu dalam proses pengklasifikasian.

## 2.4. Preprocessing

*Preprocessing* merupakan proses pembersihan kata sebelum kata/ data diolah. Proses yang dilakukan pada tahapan ini antara lain:

### 1. Cleaning

Pada Saat melakukan tahap *cleaning* data teks, komentar karakter atau tanda baca dan angka selain huruf akan dihapus (Galih, M, & Putra, 2018).

### 2. Case Folding

Pada tahap *case folding* merubah semua huruf dalam kalimat diubah menjadi huruf kecil dan karakter yang tidak valid seperti angka, tanda baca, dan *Uniform Resources Locator* (URL) dihapus (Alfita, Indriati, & Nurul, 2020).

### 3. Tokenizing

Pada tahap Tokenisasi, kalimat teks komentar dipecah menjadi satu kata (Galih, M, & Putra, 2018).

### 4. Normalisasi

Normalisasi adalah proses mengubah kata-kata yang tidak sesuai ejaan dengan benar (Ahmad F. R., 2022). Merubah ejaan menjadi baku menggunakan Kamus Besar Bahasa Indonesia (KBBI) sebagai pedoman.

### 5. Negation Handling

*Negation handling* adalah proses dalam pemrosesan bahasa alami yang melibatkan identifikasi, pemahaman, dan penanganan

kalimat atau frasa yang menyatakan negasi atau penolakan.

### 6. Remove Stopword

*Remove Stopword* merupakan proses menghilangkan kata-kata umum yang tidak memiliki makna yang dibutuhkan. Mengurangi ukuran kata dalam teks dengan menghapus beberapa kata kerja, kata sifat dan kata keterangan lainnya dimasukkan dalam *stopword list* (Andre, Indriati, & Putra, 2018).

### 7. Steaming

*steaming* mengacu pada proses menghilangkan infleksi kata atau akhiran kata untuk menghasilkan bentuk dasar atau kata dasar yang seragam. Tujuan utama dari stemming adalah mengurangi variasi kata yang berasal dari akar kata yang sama agar dapat dianggap sebagai satu entitas.

## 2.5. Pembobotan TF-IDF

Pada proses ini, dokumen uji dan dokumen latih yang sudah dihitung jumlah kata menggunakan TF-IDF untuk menentukan bobot disetiap dokumen (Faizal & Dwi, 2019). Berikut adalah proses TF-IDF:

### 1. TF

TF mengukur seberapa sering sebuah kata muncul dalam suatu dokumen. Tujuan langkah ini adalah memberi bobot lebih pada kata-kata yang lebih sering muncul di dokumen.

### 2. IDF

IDF mengukur seberapa penting sebuah kata dengan cara mengurangi bobot kata-kata yang muncul dalam banyak dokumen. Rumus IDF:

$$IDF = \log \frac{D}{F} \quad (1)$$

Keterangan:

D =Total dokumen dalam data set

DF =Total dokumen yang mengandung kata

### 3. TF-IDF

Setelah menghitung TF dan IDF, kedua nilai tersebut dikalikan untuk mendapatkan skor TF-IDF untuk setiap kata dalam dokumen. Skor TF-IDF memberi bobot tinggi pada kata-kata yang sering muncul dalam dokumen tetapi jarang di seluruh kumpulan dokumen. Rumus TF-IDF:

$$W_{dt} = t f_t \times I D F_t \quad (2)$$

Keterangan:

$W_{dt}$  =Bobot kata ke t dokumen t

$tf_t$  =Jumlah kemunculan kata t pada dokumen t

## 2.6. Feature Selection

*feature selection* merupakan langkah dalam memilih fitur dari teks. *feature selection* bertujuan untuk menyederhanakan proses klasifikasi dengan mengurangi jumlah token yang sudah diboboti untuk melakukan analisis yang lebih sedikit (Muhammad, Sutrisno, & Marji, 2019).

## 2.7. Klasifikasi Modified K-Nearest Neighbor

Modified K-Nearest Neighbor (M-KNN) dimasukkan dalam metode klasifikasi berdasarkan kedekatan data latih. M-KNN adalah peningkatan dan pengembangan perhitungan dari metode K-Nearest Neighbor (KNN) yang memvalidasi semua data sampel untuk mengatasi data outlier, sehingga hasil yang maksimal dapat dicapai dengan pembobotan pada setiap data (Andre, Indriati, & Putra, 2018). Berikut langkah dalam proses M-KNN:

### 1. Perhitungan Jarak Euclidean

$$d(x_i, y_i) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} \quad (3)$$

Di mana d adalah jarak antara titik data training x dan titik data uji y yang sedang diklasifikasikan, di mana x dan y mewakili nilai atribut dan n adalah dimensi atribut.

### 2. Perhitungan Nilai Validasi

Proses validasi dilakukan untuk semua data yang ada pada data training. Validitas setiap informasi dihitung, nilai validitas digunakan sebagai informasi tambahan untuk informasi tersebut.

$$Validity(x) = \frac{1}{n} \sum_{i=1}^H s(lbl(x), lbl(N_i(x))) \quad (4)$$

Keterangan=

Validitas (x) =Nilai validitas dari data latih x

$lbl(x)$  =Kelas x

$lbl(N_i(x))$  = Label kelas titik terdekat x

### 3. Perhitungan Weighted Voting

*Weighted Voting* memungkinkan tetangga dengan jarak yang lebih dekat atau atribut yang lebih penting memberikan kontribusi yang lebih besar dalam menentukan kelas mayoritas. Dengan cara ini, tetangga dengan karakteristik yang lebih relevan atau serupa dengan data uji memiliki pengaruh yang lebih signifikan dalam proses pengklasifikasian.

$$W(i) = Validity(i) \times \frac{1}{de(i)+0,5} \quad (5)$$

Keterangan=

W(i) =Perhitungan *Weighted Voting*

Validity(i) =Nilai Validitas

de(i) =Jarak *Euclidean*

## 2.8. Evaluasi Akurasi

Evaluasi akurasi adalah metode untuk mengukur sejauh mana model MKNN mampu mengklasifikasikan data dengan benar. Akurasi merupakan salah satu metrik evaluasi yang umum digunakan dalam klasifikasi dan menghitung persentase jumlah prediksi yang benar dibandingkan dengan jumlah total data.

## 3. Hasil dan Pembahasan

Setelah melakukan penelitian yang dilakukan sesuai pada gambar 1, mendapatkan hasil sesuai dengan alur penelitian sebagai berikut:

### 3.1. Pengumpulan Data

Data *tweet* yang didapat melalui *crawling* di Twitter akan melalui seleksi data, atribut yang digunakan untuk membersihkan data dan untuk klasifikasi selanjutnya akan dipilih pada tahapan ini (Tiara, Yusra, Muhammad, & Lola, 2023).

1	Datetime	Tweet Id	Username	Location	Language	Hashtags	Text
2	2023-01-06 23:51:58+00:00	1.61E+23	ganesha4g	Indonesia	in	['Ganesa', 'GanjarU']	b'Selamat Pagi \xf0\x9f\
3	2023-01-06 23:50:19+00:00	1.61E+23	_RismaW	DKI Jakarta, Ind	in	['SahabatGanjar']	b'Berbagai upaya dilakuk
4	2023-01-06 23:49:57+00:00	1.61E+23	JatengTV	Semarang, Jawa	in	['ptperorangan', 'b']	b'(04) Pada PT Peroranga
5	2023-01-06 23:43:44+00:00	1.61E+23	CutSarinal	NANGROE ACEH	in		b'Ini bukan kampenya ya
6	2023-01-06 23:33:04+00:00	1.61E+23	saga_mag	Magelang Utara	in	['capres2024']	b'Sering melakukan kunj
7	2023-01-06 23:29:53+00:00	1.61E+23	abiku_129		in		b'Sepuluh tahun menjab
8	2023-01-06 23:25:36+00:00	1.61E+22	AchmadSutomo3		in		b'Ya Allah perintahkan ai
9	2023-01-06 23:20:05+00:00	1.61E+22	PajaranPe	Maldives	in		b'Apakah ini bukan banji
10	2023-01-06 22:29:00+00:00	1.61E+23	Bantengketan09		in		b'Belum sebulan dari bar

Gambar 2. Data *Tweet*

Pengelolaan data pada penelitian ini menggunakan *tools* Google Collab, hasil *load data* diperlihatkan pada Gambar 3.

	Text	Kelas
0	apriori terkadang benci kiri alas jalan rangku...	POSITIF
1	ganjar turun pranowo	POSITIF
2	dukung ganjar pranowo presiden kiri milu 2024 ...	POSITIF
3	ganjar pranowo harap	POSITIF
4	tingkat miskin jateng turun berkat ganjar pranowo	POSITIF
...	...	...
3995	kandidat duduk wanti wanti survei voxpoll cente...	NEGATIF
3996	fakta survei pws 11 nov pilih klik doang prab...	NEGATIF
3997	nafsu radikal keliru makna didik sisi muslim p...	NEGATIF
3998	gass lur mbok kebagen mengko nyesel jateng	NEGATIF
3999	ayo baperan ganesha juang gibran serang tenang	NEGATIF

4000 rows \* 2 columns

Gambar 3. *Load Data*

### 3.2. Preprocessing

Berikut hasil *preprocessing* yang telah dilakukan:

#### 1. Cleaning

*Cleaning* adalah proses menghapus atau memodifikasi data yang tidak perlu, tidak akurat, atau tidak sesuai dengan tujuan analisis kita seperti URL, tanda baca, *icon*, dan *emot*.

Tabel 1. *Cleaning*

Sebelum	Sesudah
b'Bukan tidak setuju mas @ganjarpranowo tercinta, tapi yang jadi masalah itu klaim anda terhadap dana baznas yang seolah oleh itu punya PDI-P.'	Bukan tidak setuju mas ganjarpranowo tercinta tapi yang jadi masalah itu klaim anda terhadap dana baznas yang seolah oleh itu punya PDIP

#### 2. Case Folding

Seluruh abjad yang terdapat dalam dokumen/ data akan diubah menjadi *lowercase* atau huruf kecil.

Tabel 2. *Case Folding*

Sebelum	Sesudah
Bukan tidak setuju mas ganjarpranowo tercinta tapi yang jadi masalah itu klaim anda terhadap dana baznas yang seolah oleh itu punya PDIP	bukan tidak setuju mas ganjarpranowo tercinta tapi yang jadi masalah itu klaim anda terhadap dana baznas yang seolah oleh itu punya pdip

#### 3. Tokenizing

Membagi suatu kalimat menjadi kumpulan token.

Tabel 3. *Tokenizing*

Sebelum	Sesudah
bukan tidak setuju mas ganjarpranowo tercinta tapi yang jadi masalah itu klaim anda terhadap dana baznas yang seolah oleh itu punya pdip	“bukan”, “tidak”, “setuju”, “mas”, “ganjarpranowo”, “tercinta”, “tapi”, “yang”, “jadi”, “masalah”, “itu”, “klaim”, “anda”, “terhadap”, “dana”, “baznas”, “yang”, “seolah”, “oleh”, “itu”, “punya”, “pdip”

#### 4. Normalisasi

Memperbaiki ejaan yang salah pada kata yang terdapat dalam dokumen.

Tabel 4. Normalisasi

Sebelum	Sesudah
“bukan”, “tidak”, “setuju”, “mas”, “ganjarpranowo”, “tercinta”, “tapi”, “yang”, “jadi”, “masalah”, “itu”, “klaim”, “anda”, “terhadap”, “dana”, “baznas”, “yang”, “seolah”, “oleh”, “itu”, “punya”, “pdip”	“bukan”, “tidak”, “setuju”, “bapak”, “ganjarpranowo”, “tercinta”, “tapi”, “yang”, “jadi”, “masalah”, “itu”, “klaim”, “anda”, “terhadap”, “dana”, “baznas”, “yang”, “seolah”, “oleh”, “itu”, “punya”, “pdip”

#### 5. Negation Handling

*Negation handling* adalah teknik atau langkah-langkah yang dilakukan untuk mengatasi dampak negasi dalam teks. Negasi adalah ketika suatu pernyataan dinyatakan secara negatif, misalnya "tidak", "bukan", "tak", dan sejenisnya.

Tabel 5. *Negation Handling*

Sebelum	Sesudah
“bukan”, “tidak”, “setuju”, “bapak”, “ganjarpranowo”, “tercinta”, “tapi”, “yang”, “jadi”, “masalah”, “itu”, “klaim”, “anda”, “terhadap”, “dana”, “baznas”, “yang”, “seolah”, “oleh”, “itu”, “punya”, “pdip”	Menolak bapak ganjarpranowo tercinta tapi yang jadi masalah itu klaim anda terhadap dana baznas yang seolah oleh itu punya pdip

#### 6. Remove Stopword

*Remove stopwords* adalah penghapusan kata-kata umum yang biasanya tidak memiliki arti khusus atau berkontribusi secara signifikan terhadap pemahaman teks.

Tabel 6. Remove Stopword

Sebelum	Sesudah
Menolok bapak ganjar pranowo tercinta tapi yang jadi masalah itu klaim anda terhadap dana baznas yang seolah oleh itu punya pdip	Menolok bapak Ganjar Pranowo tercinta masalah klaim dana Baznas pdip

### 7. Steaming

Pada proses *stemming*, kata-kata tersebut mengalami perubahan morfologi dengan tujuan untuk menyederhanakan kata ke bentuk dasarnya.

Tabel 7. Steaming

Sebelum	Sesudah
Menolok bapak Ganjar Pranowo tercinta masalah klaim dana Baznas pdip	tolak bapak Ganjar Pranowo cinta masalah klaim dana Baznas pdip

### 3.3. Pembobotan TF-IDF

Setelah melakukan pembobotan TF-IDF menggunakan *tools* Google Collab, maka mendapatkan hasil yang dapat dilihat pada Gambar 4 berikut:

Gambar 4. Pembobotan TF-IDF

### 3.4. Feature Selection

*Feature selection* adalah teknik untuk memilih subset fitur yang paling relevan atau informatif dari kumpulan fitur yang ada. Pada penelitian ini menggunakan *threshold* 0,001.

Gambar 5. Feature Selection

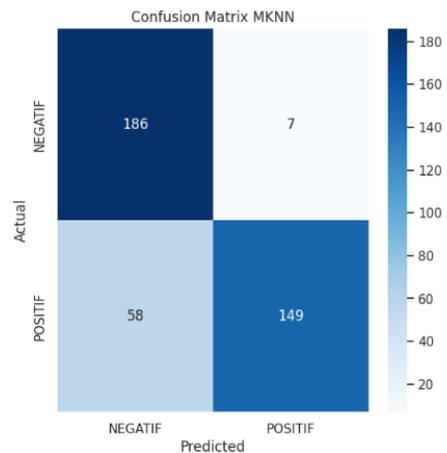
### 3.5. Klasifikasi

Setelah melewati tahap *preprocessing*, maka tahap terakhir yang dilakukan adalah tahap klasifikasi. Dalam pengujian yang dilakukan yaitu menggunakan 10% data uji dari 4.000 data yang digunakan. Hasil dari menggunakan *tools* Google Collab pada tahap ini dapat dilihat pada Gambar 6.

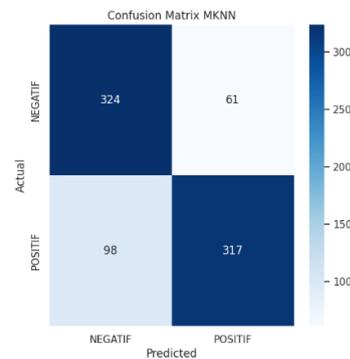
Gambar 6. Klasifikasi

### 3.6. Evaluasi Akurasi

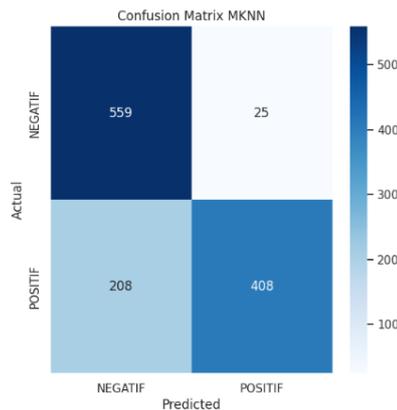
Hasil akurasi pengujian dengan perbandingan data 90:10, 80:20 dan 70:30 dengan menggunakan nilai  $k=3$  mendapatkan *confusion matrix* dapat dilihat pada gambar berikut.



Gambar 7. Confusion Matrix 90:10



Gambar 8. Confusion Matrix 80:20



Gambar 9. Confusion Matrix 70:30

Dari pengujian yang telah dilakukan dengan perbandingan 90:10, 80:20 dan 70:30 dengan nilai  $K=3$ , ditunjukkan pada gambar berikut.

```

Accuracy Score: 83.75
-----
Precision Negatif: 76.22950819672131
Precision Positif: 95.51282051282051
Precision Score: 83.75
-----
Recall Negatif: 96.37305699481865
Recall Positif: 71.98067632850241
Recall Score: 83.75
-----
F1 Negatif: 85.12585812356978
F1 Positif: 82.09366391184572
F1 Score: 83.75
    
```

Gambar 10. Hasil Perbandingan 90:10

```

Accuracy Score: 81.75
-----
Precision Negatif: 73.85229540918164
Precision Positif: 94.98327759197325
Precision Score: 81.75
-----
Recall Negatif: 96.1038961038961
Recall Positif: 68.43373493975903
Recall Score: 81.75
-----
F1 Negatif: 83.52144469525959
F1 Positif: 79.55182072829132
F1 Score: 81.75
    
```

Gambar 11. Hasil Perbandingan 80:20

```

Accuracy Score: 80.58333333333333
-----
Precision Negatif: 72.88135593220339
Precision Positif: 94.22632794457274
Precision Score: 80.58333333333333
-----
Recall Negatif: 95.71917808219177
Recall Positif: 66.23376623376623
Recall Score: 80.58333333333333
-----
F1 Negatif: 82.75351591413768
F1 Positif: 77.78836987607245
F1 Score: 80.58333333333333
    
```

Gambar 12. Hasil Perbandingan 70:30

### 3.7. Wordcloud

Wordcloud adalah visualisasi teks yang menggambarkan frekuensi kemunculan kata-kata dalam sebuah teks. Biasanya, kata-kata yang lebih sering muncul ditampilkan dengan ukuran yang lebih besar, sehingga memungkinkan pengamatan visual yang cepat terhadap kata-kata yang paling penting atau dominan dalam teks tersebut.



Gambar 13. Wordcloud kelas positif



Gambar 13. Wordcloud kelas negatif

## 4. Kesimpulan

Setelah melakukan tahapan penelitian dan pengujian, hasil menunjukkan bahwa metode Modified KNN dengan menggunakan perbandingan data 90:10, 80:20 dan 70:30 dengan nilai  $K=3$  memberikan hasil yang tertinggi dari *Precision*, *recall*, *f1 score*, dan akurasi 83,8% pada perbandingan 90:10. Kesimpulan ini menunjukkan bahwa metode Modified KNN yang diterapkan pada data Twitter dengan parameter yang diatur memiliki kemampuan yang baik dalam mengklasifikasikan data.

## Daftar Pustaka

- Adella, A. P., Indriati, & Yuita, A. S. (2020). Analisis Sentimen Terhadap Ulasan Pengguna MRT Jakarta Menggunakan Information Gain dan Modified K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1125-1132.
- Ahmad, F. R. (2022). Klasifikasi Tweet di Twitter dengan Menggunakan Metode K-Nearest Neighbor. *Jurnal Sistim Informasi dan Teknologi*, 64-69.
- Ahmad, N. R., Indriati, & Lailil, M. (2018). Analisis Sentimen Review Aplikasi Mobile Dengan Menggunakan Metode Modified K Nearest

- Neighbour (MK-NN). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3157-3161.
- Alfita, N., Indriati, & Nurul, H. (2020). Klasifikasi Review Produk Kecantikan Pada Aplikasi Sociolla Menggunakan Algoritme Modified K-Nearest Neighbor (MK-NN) dengan Pembobotan BM25. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3426-3431.
- Andre, R. P., Indriati, & Putra, P. A. (2018). Klasifikasi Hoax Pada Berita Kesehatan Berbahasa Indonesia Dengan Menggunakan Metode Modified K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 7466-7473.
- Cantya, D., Yufis, A., & Nur, H. (2020). *Seminar Nasional Teknologi dan Rekayasa (SENTRA)*, 165-170.
- Faizal, N. R., & Dwi, H. S. (2019). Klasifikasi Berita Hoax Pilpres Menggunakan Metode Modified K-Nearest Neighbor dan Pembobotan Menggunakan TF-IDF. *KONVERGENSI*, 1-10.
- Galih, N. B., M, A. F., & Putra, P. A. (2018). Penerapan Klasifikasi Tweets Pada Berita Twitter Menggunakan Metode K-Nearest Neighbor Dan Query Expansion Berbasis Distributional . *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3849-3855.
- Muhammad, H., Sutrisno, & Marji. (2019). Seleksi Fitur dengan Information Gain pada Identifikasi Jenis Attention Deficit Hyperactivity Disorder Menggunakan Metode Modified K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 10444-10452.
- Muhammad, I. A., Rizqeya, I. P., Kevin, H. U., & Nur, A. R. (2020). Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin COVID-19. *JIEET: Journal Information Engineering and Educational Technology*, 90-92.
- Sofiah, Yusra, & Muhammad, F. (2023). Klasifikasi Sentimen Masyarakat di Twitter Terhadap Kenaikan Harga Bahan Bakar Minyak dengan Metode Modified K-Nearest Neighbor. *SATIN – Sains dan Teknologi Informasi*, 138-148.
- Tiara, D. A., Yusra, Muhammad, F., & Lola, O. (2023). Klasifikasi Sentimen Masyarakat di Twitter terhadap Kenaikan Harga BBM dengan Metode K-NN. *JUKI : Jurnal Komputer dan Informatika*, 140-150.