

# PENERAPAN ALGORITMA *K-NEAREST NEIGHBOR* (KNN) ANALISIS SENTIMEN TWEET PADA ELEKTABILITAS BAKAL CALON PRESIDEN 2024

Rizki Ananda Putra <sup>1\*</sup>, Wahyu Pramusinto<sup>2</sup>

<sup>1,2</sup> Fakultas Teknologi Informasi, Teknik Informatika, Universitas Budi Luhur, Jakarta Selatan, Indonesia

Email: <sup>1\*</sup>1911500419@student.budiluhur.ac.id, <sup>2</sup>wahyu.pramusinto@budiluhur.ac.id  
(\* : corresponding author)

**Abstrak-** Kapasitas yang dirasakan tokoh politik untuk memenangkan pemilihan atau memenangkan dukungan publik disebut sebagai elektabilitas. Itu didasarkan pada berbagai kriteria subyektif, seperti popularitas mereka, kemampuan kepemimpinan, sikap terhadap masalah, latar belakang, karisma, kemampuan komunikasi, dan citra publik. Elektabilitas dapat bervariasi berdasarkan lingkungan politik, pemilu yang ada, dan demografi dan preferensi pemilih. Penelitian ini bertujuan untuk melakukan Analisis Sentimen Tweet Terhadap Elektabilitas Tokoh Politik Bakal Calon Presiden 2024 dengan Menggunakan Algoritma K-Nearest Neighbor (KNN) dan fitur ekstraksi CountVectorizer dengan dataset yang berupa teks kicauan (tweet) yang berasal dari pada media sosial Twitter dengan beberapa kata kunci 'Ganjar Pranowo', 'Prabowo Subianto', 'Anies Baswedan'. Sehingga dapat mengetahui pola dan tren opini publik serta mengikuti perubahan sentimen dari waktu ke waktu dengan menggunakan analisis sentimen untuk percakapan mengenai kebijakan pemerintah di ruang publik. Berdasarkan banyak tweet atau sentimen masyarakat yang diambil pada periode Juni 2023 dengan tiap kata kuncinya yang berbeda mendapatkan, yaitu untuk kata kunci "Ganjar Pranowo" dengan tweet sebanyak 325 memperoleh 241 tweet positif, dan 84 tweet negatif. Lalu dengan kata kunci "Prabowo Subianto" dengan tweet sebanyak 313 memperoleh 77 tweet positif dan 236 tweet negatif, pada kata kunci "Anies Baswedan" dengan tweet sebanyak 322 memperoleh 138 tweet positif dan 184 tweet negatif. Hasil pengujian dan evaluasi dengan menggunakan nilai  $K=3$  dengan tiap kata kunci yang berbeda yaitu "Ganjar Pranowo" mendapatkan akurasi 77 %, presisi 77%, recall 100%, lalu "Prabowo Subianto" mendapatkan akurasi 97%, presisi 87%, recall 100%, dan "Anies Baswedan" mendapatkan akurasi 67%, presisi 44%, recall 42%.

**Kata Kunci:** text mining, analisis sentimen, k-nearest neighbors, tokoh politik.

## IMPLEMENTATION *K-NEAREST NEIGHBOR* (KNN) SENTIMENT ANALYSIS OF 2024 PRESIDENTIAL CANDIDATES' ELECTABILITY TWEETS

**Abstract-** perceived capacity of a political figure to win an election or win public support is referred to as electability. It is based on various subjective criteria, such as their popularity, leadership ability, attitude to problems, background, charisma, communication skills, and public image. Electability can vary based on the political environment, existing elections, and voter demographics and preferences. This study aims to conduct Tweet Sentiment Analysis on the Electability of Political Figures for 2024 Presidential Candidates Using the K-Nearest Neighbor (KNN) Algorithm and the CountVectorizer extraction feature with a dataset in the form of tweet text originating from Twitter social media with several keywords 'Ganjar Pranowo', 'Prabowo Subianto', 'Anies Baswedan'. So that it can find out the patterns and trends of public opinion and follow changes in sentiment over time by using sentiment analysis for conversations about government policies in the public sphere. Based on many tweets or public sentiments taken in the June 2023 period with each different keyword getting, namely for the keyword "Ganjar Pranowo" with 325 tweets getting 241 positive tweets, and 84 negative tweets. Then with the keyword "Prabowo Subianto" with 313 tweets obtained 77 positive tweets and 236 negative tweets, on the keyword "Anies Baswedan" with 322 tweets obtained 138 positive tweets and 184 negative tweets. The results of testing and evaluation using the value  $K = 3$  with each different keyword, namely "Ganjar Pranowo" get 77% accuracy, 77% precision, 100% recall, then "Prabowo Subianto" get 97% accuracy, 87% precision, 100% recall, and "Anies Baswedan" get 67% accuracy, 44% precision, 42% recall.

**Keywords:** text mining, sentiment analysis, k-nearest neighbors, political figures.

## 1. PENDAHULUAN

Kapasitas yang dirasakan tokoh politik untuk memenangkan pemilihan atau memenangkan dukungan publik disebut sebagai elektabilitas. Itu didasarkan pada berbagai kriteria subyektif, seperti popularitas mereka, kemampuan kepemimpinan, sikap terhadap masalah, latar belakang, karisma, kemampuan komunikasi, dan citra

publik. Elektabilitas dapat bervariasi berdasarkan lingkungan politik, pemilu yang ada, dan demografi dan preferensi pemilih [1].

Twitter adalah salah satu *website* yang banyak digunakan oleh masyarakat untuk memberikan opini atau tanggapan terhadap suatu peristiwa ataupun topik yang sedang banyak orang bicarakan, ada banyak metode yang dapat digunakan untuk mengolah pendapat masyarakat tersebut yang dapat kita ketahui yaitu praktik menilai sentimen atau opini yang disampaikan dalam *tweet* di jaringan Twitter dikenal dengan istilah "Analisis Sentimen Twitter". Sentimen *tweet* mungkin netral, negatif, atau menguntungkan. Analisis dapat dilakukan untuk berbagai tujuan, termasuk melacak sentimen kampanye politik atau mengukur opini publik terhadap suatu isu atau produk tertentu [2]. Metode melakukan analisis sentimen Twitter meliputi pengumpulan dan pemrosesan awal *tweet*, mengategorikan setiap sentimen *tweet*, dan menggabungkan skor sentimen untuk memberikan analisis sentimen yang komprehensif dari *tweet* tersebut. Metode *K-Nearest Neighbor* (KNN) untuk pembelajaran mesin adalah contohnya [3].

Pada penelitian sebelumnya, pernah dilakukan penelitian serupa dengan meneliti tokoh politik Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita *Online* (2019) [4] pada penelitian tersebut dilakukan pengumpulan data melalui komentar media berita *online*, dalam penemuan hasil dari penelitian tersebut didapatkan akurasi yang cukup baik, terdapat perbedaan diantaranya dalam metode pengumpulan data. Pada penelitian tersebut menggunakan metode *Support Vector Machine* dan *Naïve Bayes*.

pada penelitian lain juga dengan judul Analisis Sentimen Twitter Terhadap Capres Indonesia 2019 Dengan Metode K-NN oleh [5] mendapatkan tingkat akurasi sentimen mencapai 83,33%. Pada penelitian lain dengan yang metode yang sama dengan judul Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN [6] menunjukkan bahwa Algoritma *K-Nearest Neighbor* (KNN) dapat memperoleh yaitu untuk akurasi, precision dan recall bibit 85,14% , 91,91%, dan 76,44% sedangkan untuk bareksa yaitu 81,70% , 87,15%, 75,73%, dengan perbandingan 60:40.

Sedangkan, Pada penelitian ini, sumber data yang diperoleh melalui Twitter yang di-*scraper* sebagai dataset penelitian, kicauan yang dikirim masyarakat akan terkumpul saat proses *scraper* lalu akan tersimpan pada file excel kemudian di teliti dan diproses dengan algoritma Algoritme *K-Nearest Neighbor* untuk mengetahui elektabilitas tokoh politik.

Penelitian ini bertujuan untuk melakukan Analisis Sentimen *Tweet* Terhadap Elektabilitas Tokoh Politik Bakal Calon Presiden 2024 Menggunakan Algoritma *K-Nearest Neighbor* (KNN) dengan dataset yang berupa teks kicauan (*tweet*) yang berasal dari pada media sosial Twitter dengan beberapa kata kunci ‘Ganjar Pranowo’, ‘Prabowo Subianto’, ‘Anies Baswedan’. Sehingga dapat mengetahui pola dan tren opini publik serta mengikuti perubahan sentimen dari waktu ke waktu dengan menggunakan analisis sentimen untuk percakapan mengenai kebijakan pemerintah di ruang publik. Hal ini dapat membantu ahli strategi politik dan pembuat keputusan untuk lebih memahami bagaimana perasaan publik tentang kebijakan mereka, singkatnya, analisis sentimen dapat menawarkan informasi mendalam tentang bagaimana publik memandang dan bereaksi terhadap kebijakan partai politik dan pemerintah, yang dapat membantu pembuat keputusan dan ahli strategi politik dalam meningkatkan pengambilan keputusan dan teknik komunikasi mereka [7].

## 1.1 Analisis Sentimen

Tujuan dari analisis sentimen, biasanya disebut sebagai penambahan opini, adalah untuk mengidentifikasi sentimen atau nada emosional yang direpresentasikan dalam sebuah teks. Untuk mendeteksi dan mengukur keseluruhan sentimen sebagai menguntungkan, negatif, atau netral, itu termasuk menilai informasi subjektif yang termasuk dalam data teks [8].

## 1.2 Text Mining

Prosedur yang disebut penambahan teks, biasanya disebut sebagai analitik teks, digunakan untuk mengumpulkan pengetahuan dan wawasan penting dari data teks yang tidak terstruktur. Untuk mengevaluasi sejumlah besar dokumen teks dan mengidentifikasi pola, korelasi, dan tren yang signifikan, pemrosesan bahasa alami (NLP), pembelajaran mesin, dan pendekatan statistik digunakan [9].

## 1.3 Algoritma K-Nearest Neighbor (KNN)

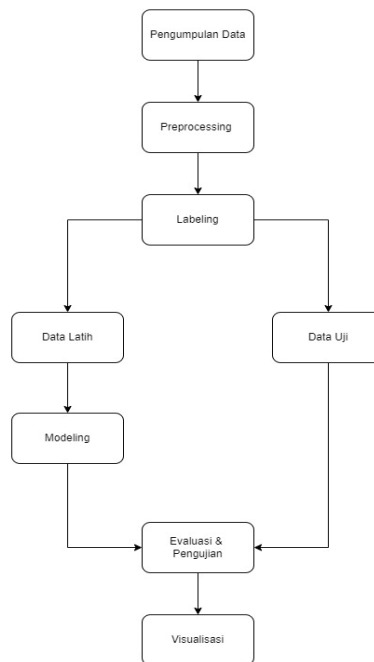
Sejak penemuannya pada tahun 1950-an, *K-Nearest Neighbor* (KNN) telah menjadi teknik pembelajaran mesin yang populer. Evelyn Fix dan Joseph Hodges dari University of California, Berkeley mengemukakan ide tersebut terlebih dahulu, dan banyak sarjana kemudian menyempurnakannya. KNN dibangun berdasarkan prinsip kemiripan atau kedekatan. Menurut algoritme, titik data yang serupa biasanya memiliki label atau nilai yang cocok [10]

## 1.4 Machine Learning

*Artificial intelligence* (AI) memiliki area yang disebut "pembelajaran mesin" yang berfokus pada pembuatan model dan algoritme yang memungkinkan komputer belajar dan membuat prediksi atau penilaian tanpa diprogram secara eksplisit. Ini mencakup pembelajaran pola dan korelasi otomatis dari data menggunakan metode statistik dan algoritme komputasi [11]

## 2. METODE PENELITIAN

### 2.1 Tahapan Penelitian



**Gambar 1.** Tahapan Metode Penelitian

Pengumpulan data dilakukan dengan menggunakan Twitter sebagai scraper media sosial untuk mengumpulkan opini dari pengguna. Data kemudian diimpor ke database untuk preprocessing, yang melibatkan normalisasi kata dan penyesuaian kata. Data latih mengacu pada data yang digunakan untuk model pembelajaran, sedangkan data uji adalah data yang digunakan untuk mengevaluasi kinerja dan model akurasi. Data uji adalah data yang dihasilkan untuk menguji model pelatihan, menguji proses pemodelan, pengujian tingkat akurasi, dan hasil penerapan teknik yang divisualisasikan dalam bentuk proporsi dan grafik.

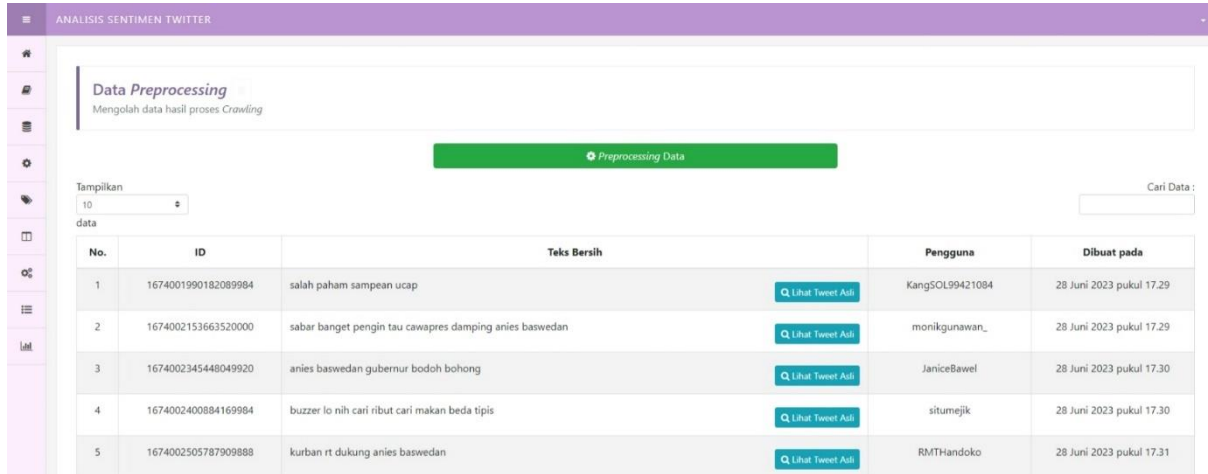
#### 2.1.1 Pengumpulan Data

Pada tahapan yang dilakukan untuk mengumpulkan data yang berasal dari sosial media twitter berupa kicauan (*tweet*) opini masyarakat, hasil dari *scraper* data yang dilakukan pada platform twitter nantinya akan tersimpan kedalam file excel dan kemudian akan di-*import* ke dalam basis data (*database*) *MySQL*.

**Tabel 1.** Data Kicauan Pada *Twitter*

No	Username	<i>Tweet</i>
1	Rimba56365891	Sungguh Mulia sekali Pak Ganjar Pranowo yang Peduli keadaan masyarakat yang tidak peduli dengan sanjungan, cacian dari siapa pun.. beliau ikhlas membantu. insyaallah Bharokah Pak Ganjar Semoga lancar dan sukses selalu aamiin
2	narkosun	Pak @ganjarpranowo diserang lagi dgn kasus e-ktp. Pdhl sudah clear bahwa beliau tak terlibat. Ini yg ngomong Novel baswedan lho, dia menangani kasus e-ktp. Setelah ini, nanti muncul lagi narasi bokep. Trus itu balik lagi ke e-ktp. Gitu aja terus sampe pak Ganjar jd presiden.

### 2.1.2 Preprocessing



Gambar 2. Tampilan Menu Preprocessing

Pemrosesan data adalah fase dalam penambangan data dan analisis data yang mengubah data mentah menjadi format yang dapat dipahami dan dievaluasi oleh komputer dan pembelajaran mesin. Data mentah berupa teks, foto, video, dan sebagainya biasanya campur aduk. Persiapan data sangat penting karena melatih model pembelajaran mesin dengan data yang buruk atau "kotor", biasanya akan mendapatkan model yang buruk dan kurang terlatih yang tidak berguna untuk analisis yang coba dilakukan [12].

#### 1. Case Folding

adalah metode transformasi teks yang mengubah semua karakter dalam *string* tertentu menjadi huruf kecil. Ini juga dikenal sebagai konversi huruf kecil

Tabel 2. Proses Case Folding

Kalimat Asli	Hasil Case Folding
@HidayatNatari Di Tanah Suci aja masih bikin HOAX @Neraka Jahanam lu Nies @aniesbaswedan	@hidayatnatari di tanah suci aja masih bikin hoax @neraka jahanam lu nies @aniesbaswedan

#### 2. Cleansing

Proses pembersihan data adalah proses menghapus atau mengoreksi data yang salah, tidak lengkap, duplikat, atau tidak relevan serta menghapus tanda baca atau simbol (selain huruf) dari kumpulan data.

Tabel 3. Proses Cleansing

Kalimat Asli	Hasil Cleansing
@hidayatnatari di tanah suci aja masih bikin hoax @neraka jahanam lu nies @aniesbaswedan	di tanah suci aja masih bikin hoax neraka jahanam lu nies

#### 3. Merubah Slangword

istilah slang harus diganti dengan bahasa yang lebih umum yang mengkomunikasikan perasaan yang sebanding.

Tabel 4. Proses Merubah Slangword

Kalimat Asli	Hasil Slangword
di tanah suci aja masih bikin hoax neraka jahanam lu nies	di tanah suci saja masih bikin hoax neraka jahanam lu nies

#### 4. Menghapus Stopword

stopword adalah istilah umum yang sering dianggap tidak relevan untuk tugas pemrosesan bahasa tertentu, seperti analisis sentimen atau ekstraksi kata kunci. stopwords dapat dihapus untuk membantu menyederhanakan analisis dan fokus pada istilah yang paling penting dalam teks.

**Tabel 5. Proses Menghapus Stopword**

Kalimat Asli	Hasil Stopword
di tanah suci saja masih bikin hoax neraka jahanam lu nies	tanah suci bikin hoax neraka jahanam lu nies

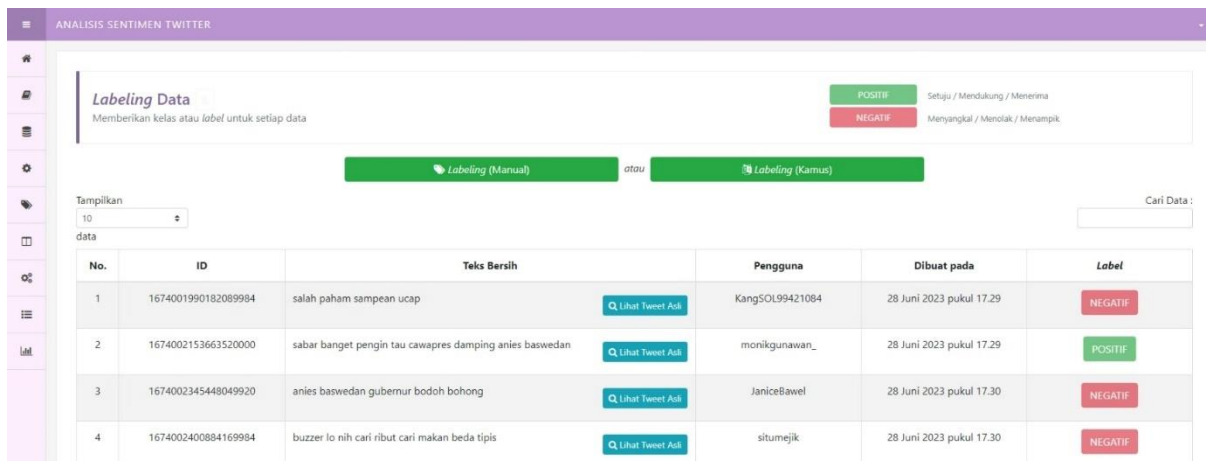
#### 5. Stemming

pendekatan pemrosesan bahasa alami yang mereduksi kata menjadi akar atau bentuk dasarnya, yang dikenal sebagai stem.

**Tabel 6. Proses Stemming**

Kalimat Asli	Hasil Stemming
tanah suci membuat hoax neraka jahanam lu nies	tanah suci buat hoax neraka jahanam lu nies

### 2.1.3 Labeling



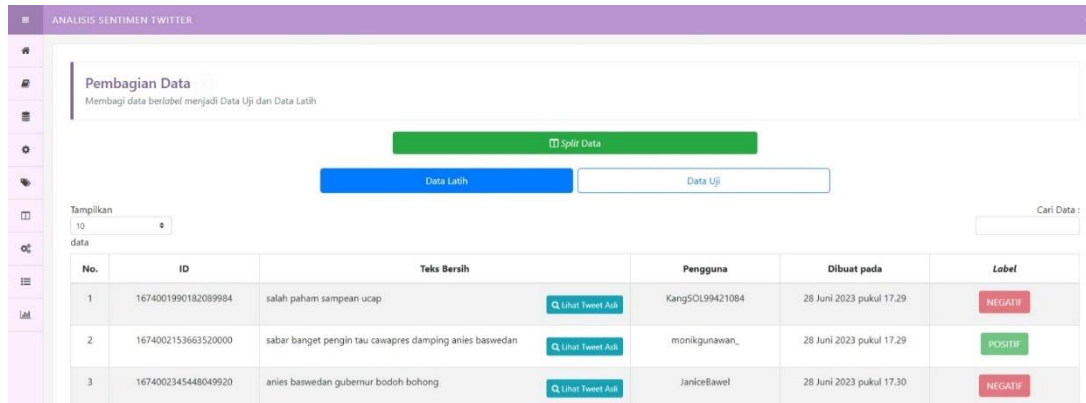
**Gambar 3. Tampilan Menu Proses Labeling**

Pelabelan positif dan negatif mengacu pada proses penerapan label sentimen ke data teks berdasarkan kepositifan atau kenegatifan yang dinyatakan dalam teks. Berikut cara memberi label sentimen positif dan negatif: Saat mengategorikan sentimen positif, artinya menerapkan label ke sampel teks yang menunjukkan pemikiran positif atau penuh harapan

**Tabel 7. Proses Labeling**

No	Teks Bersih	Sentimen
1	tunjuk ketidaksukaannya adminnya partai layak coblos	Positif
2	idiot habis cebong raja ngibul orang solo gue kirim sempel lo idiot	Negatif
3	sabar banget ini ini cawapres damping anies baswedan	Positif
4	anies baswedan gubernur bodoh bohong	Negatif
5	jadi kocak ini ketua rt orang kurban palak perilaku dukung wakil sifat dukung sabda	Positif

### 2.1.4 Pembagian Data



Gambar 4. Tampilan Menu Proses Pembagian Data

Pada proses ini, kicauan yang sudah memiliki label sentimen selanjutnya akan dilakukan tahapan pemisahan data berarti memisahkan dataset menjadi dua bagian: set pelatihan dan set pengujian, dengan sekitar 80% data masuk ke set pelatihan dan 20% sisanya masuk ke set pengujian.

1. Data Latih: Sebagian dari seluruh kumpulan data yang digunakan untuk melatih model pembelajaran mesin disebut sebagai data pelatihan. Itu terdiri dari karakteristik *input* (teks, data numerik, dll.) Dan variabel target (label sentimen, label kelas, dan sebagainya)
2. Data Uji: Set pengujian, juga dikenal sebagai set evaluasi, adalah subset berbeda dari seluruh set data yang digunakan untuk mengevaluasi performa model pembelajaran mesin terlatih. Ini memiliki karakteristik *input* tetapi tidak ada variabel target yang cocok.

### 2.1.5 Modeling

Pada tahapan modeling ini, dilakukan pengubahan satu set dokumen teks menjadi *matriks* hitungan token yang mewakili frekuensi setiap kata (atau n-gram) dalam dokumen, pengubahan teks *tweet* menjadi vektor menggunakan CountVectorizer [13]. Dalam fase ini ada lima proses utama dalam pembuatan sebuah model data latih, yaitu : *men-select* data latih, pembuatan wadah kata, mencari fitur kata, membuat wadah vektor kosong, ini vektor angka sesuai frekuensi.

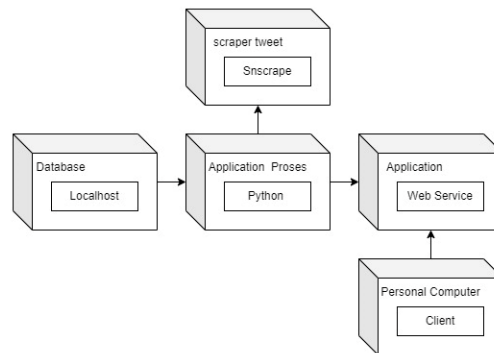
### 2.1.6 Pengujian

Pada tahap ini, klasifikasi dilakukan dengan menggunakan metode *K-Nearest Neighbor* (KNN), yaitu suatu langkah yang mungkin dilakukan setelah satu atau lebih set data model *training* ditemukan. Model pelatihan digunakan untuk melatih data mereka yang telah melalui tahap pemodelan. Serta mengevaluasi metode klasifikasi dengan menggunakan *confusion matrix* pada algoritma k-nearest neighbor.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Prototipe Sistem Analisis Sentimen

*Deployment* menggambarkan komponen-komponen sistem yang sedang diimplementasikan, seperti server, jaringan, internet, dan jaringan lainnya, serta bagaimana komponen-komponen tersebut berinteraksi satu sama lain.



**Gambar 5.** Deployment Diagram

Pada bagian ini, akan memaparkan hasil analisis, hasil implementasi, dan studi kasus, serta pelajaran yang dipetik dari berbagai topik penelitian. Bagian ini juga mencakup informasi seperti teks, gambar, tabel, dan grafik lainnya. Berikut adalah sampel data latih yang, bisa dilihat pada tabel 8 berikut:

**Tabel 8. Sampel Data Latih**

No	Teks Bersih	Sentimen
1	tunjuk ketidaksukaannya adminnya partai layak coblos	Positif
2	idiot habis cebong raja ngibil orang solo gue kirim sempel lo idiot	Negatif
3	sabar banget ini ini cawapres damping anies baswedan	Positif
4	anies baswedan gubernur bodoh bohong	Negatif
5	jadi kocak ini ketua rt orang kurban palak perilaku dukung wakil sifat dukung sabda	Positif

Tabel 8 merupakan atribut yang akan digunakan dalam proses analisis sentimen, yang mana teks bersih adalah tweet yang sudah melewati tahapan preprocessing. Terdapat nilai sentimen yang di dapatkan berdasarkan labeling manual dengan seksama apakah tweet tersebut bersifat sentimen positif atau negatif.

### 3.2 CountVectorizer

Setelah menyelesaikan langkah-langkah preprocessing, *labeling*, dan persiapan data, langkah selanjutnya dalam prosedur ini adalah menggunakan fitur CountVectorizer (pemodelan).

Tujuan dari langkah ini adalah mengembangkan model latih atau memperoleh pengetahuan dengan menggunakan data latih yang ada. Teks bersih pada sampel data latih pada Tabel 9 akan diubah menjadi representasi vektor pada tabel berikut :

**Tabel 9. Representasi Vektor Latih**

No	Representasi Vektor
Latih-1	0 1 0 1 1 1 1 0 1 1 1 1 1 0 1 0 0 0 1 0 0 1 0 0 1
Latih-2	1 0 1 0 0 0 2 0 0 0 0 1 0 1 1 1 1 0 1 1 0 1 1 0 1 1 0
Latih-3	0 1 0 0 0 0 0 0 1 1 1 1 0 1 0 0 0 1 0 0 0 0 0 1
Latih-4	0 1 0 0 1 0 0 0 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 0
Latih-5	0 0 0 0 0 0 0 1 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0

### 3.3 Klasifikasi K-Nearest Neighbor

Setelah tahap pemodelan selesai, langkah selanjutnya adalah melakukan klasifikasi menggunakan *K-Nearest Neighbor* (KNN). Tujuan dari langkah ini adalah untuk memberikan label pada setiap kumpulan data yang masuk berdasarkan model yang dikembangkan selama langkah pemodelan sebelumnya. Setelah model dipilih dan nilai K sudah ditentukan, maka tahapan selanjutnya adalah menganalisis dan mengklasifikasi data uji dengan algoritma KNN, sampel data uji pada tabel 10 berikut:

**Tabel 10. Sampel Data Uji**

No	Teks Bersih
Uji-1	anies baswedan presiden baik dunia
Uji-2	elektabilitas anies baswedan lemah

Setelah mendapatkan sampel data uji pada table 10, maka akan di representasikan menjadi vector, sehingga menjadi seperti tabel 11 berikut:

**Tabel 11. Representasi Vektor Uji**

No	Vektor Representasi
Uji-1	1 1 1 1 0 0 1
Uji-2	1 0 1 0 1 1 0

a. *Euclidean Distance*

Jarak direduksi menggunakan data vektor uji dan model latih, pada tahapan kali ini menggunakan metode *Euclidean Distance*, berikut adalah hasil perhitungan *Euclidean Distance* dengan membandingkan jarak antara vektor uji dan vektor latih:

$$d(\text{uji 1, latih 1}) = (1 - 0)^2 + (1 - 1)^2 + (1 - 0)^2 + (1 - 1)^2 + (0 - 1)^2 + (0 - 1)^2 + (1 - 0)^2$$

$$d(\text{uji 1, latih 1}) = (1 + 0 + 1 + 0 + 1 + 1 + 1)$$

$$d(\text{uji 1, latih 1}) = \sqrt{5}$$

$$d(\text{uji 1, latih 1}) = 2.23606797749979$$

Berdasarkan contoh sebelumnya, Sehingga diperoleh hasil permutasi untuk setiap jarak antara vektor uji dan vektor latih seperti pada tabel 12 berikut:

**Tabel 12. Hasil Jarak *Euclidean Distance***

<i>Euclidean Distance</i>	
$d_{(1-1)}$	2.23606797749979
$d_{(1-2)}$	2
$d_{(1-3)}$	2
$d_{(1-4)}$	1
$d_{(1-5)}$	2
$d_{(1-1)}$	2.449
$d_{(1-2)}$	1
$d_{(1-3)}$	2.236
$d_{(1-4)}$	1.732
$d_{(1-5)}$	2

b. Mencari tetangga terdekat

Pencarian tetangga terdekat memerlukan sebuah nilai K dalam proses klasifikasi dengan *K-Nearest Neighbor* (KNN). Nilai K ini menentukan jumlah tetangga terdekat yang akan dilakukan oleh proses klasifikasi. Selama penelitian ini, ditemukan beberapa nilai K yang dapat digunakan, antara lain K=3, K=5, K=7, K=9, dan K=11. Namun dalam hal ini nilai K yang digunakan adalah K=3. Artinya proses klasifikasi akan menggunakan tiga tetangga yang berdekatan untuk menentukan label data.

1. Mengurutkan nilai jarak

Dalam tahapan ini, hasil dari jarak yang di urutkan berdasarkan nilai jarak, sehingga hasilnya akan menjadi seperti berikut:

**Tabel 13. Pengurutan Hasil Jarak**

Urutan	Jarak ( $d_{(uji-i, latih-i)}$ )	Kicauan
1	1	Uji-1 Latih-4
2	2	Uji-1 Latih-2
3	2	Uji-1 Latih-3



4	2	Uji-1	Latih-5
5	2.236	Uji-1	Latih-1
1	1	Uji-2	Latih-2
2	1.732	Uji-2	Latih-4
3	2	Uji-2	Latih-5
4	2.236	Uji-2	Latih-3
5	2.449	Uji-2	Latih-1

## 2. Mengambil K terdekat

Data dari Tabel 13 setelah proses pengurutan. Nilai pengurang jarak tetangga akan diambil dengan nilai K yang telah ditentukan dan dipilih. Karena nilai K=3, berikut hasil tetangga yang terdistorsi:

**Tabel 14. Data K Tetangga Terdekat**

Urutan	Jarak (d(uji-i, latih-i))	Kicauan	
1	1	Uji-1	Latih-4
2	2	Uji-1	Latih-2
3	2	Uji-1	Latih-3
1	1	Uji-2	Latih-2
2	1.732	Uji-2	Latih-4
3	2	Uji-2	Latih-5

## 3. Menghitung nilai probabilitas

Probabilitas dihitung dengan memeriksa label probabilitas yang muncul pada data K tetangga di atas. Probabilitas bahwa tes *tweet* tertentu akan diberi label positif atau negatif. Ini dapat ditentukan dengan menggunakan `label_list` pada model latih dan Tabel 14 data K terdekat. Tingkat probabilitas yang ditentukan seperti pada Tabel 15 berikut ini:

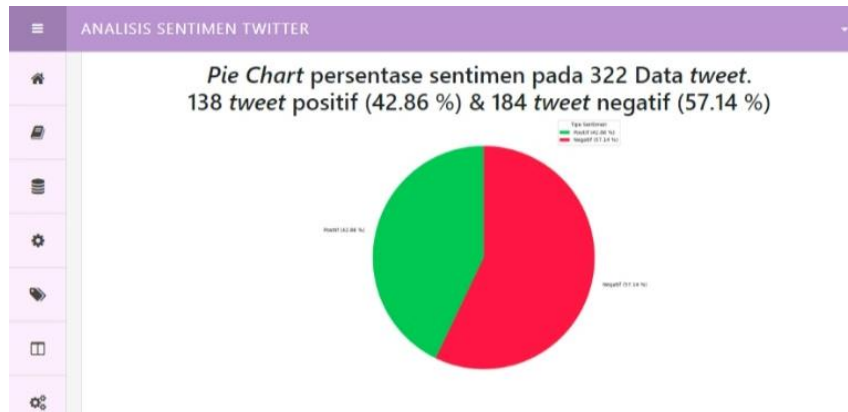
**Tabel 15. Nilai Probabilitas Data Uji**

<i>Tweet</i> (uji -i)	<i>Tweet</i> (latih-i)	Sentimen	Probabilitas positif	Probabilitas negatif
Uji-1	Latih-4	Negatif	0	1
	Latih-2	Negatif	0	1
	Latih-3	Positif	1	0
	Jumlah		1	2
Uji-2	Latih-2	Negatif	0	1
	Latih-4	Negatif	0	1
	Latih-5	Positif	1	0
	Jumlah		1	2

Berdasarkan tabel 15 Nilai probabilitas data uji, dapat diketahui dengan K=3, pada pengujian dengan ulasan uji-1 dan ulasan uji-2 keduanya berlabel negatif. Dengan probabilitas yang sama, yaitu adalah 1: 2.

## 3.4 Visualisasi

Terdapat tampilan visualisasi yang telah di rancang sebelumnya. Hasil dari visualisasi akan muncul berupa frekuensi kata, chart pie, untuk lebih jelasnya bisa dilihat pada gambar dibawah ini.



**Gambar 6. Visualisasi Data Sentimen**

#### 4. KESIMPULAN

Berdasarkan hasil pengujian dan evaluasi menggunakan aplikasi yang di kembangkan Dengan menggunakan ekstraksi fitur *CountVectorizer* dan algoritma *K-Nearest Neighbor* (KNN) dalam melakukan tahap analisis sentimen dapat bekerja dengan baik, hasil pengujian dan evaluasi dengan menggunakan nilai  $K=3$  dengan tiap kata kunci yang berbeda yaitu “Ganjar Pranowo” mendapatkan akurasi 77 %, presisi 77%, recall 100%, lalu “Prabowo Subianto” mendapatkan akurasi 97%, presisi 87%, recall 100%, dan “Anies Baswedan” mendapatkan akurasi 67%, presisi 44%, recall 42%.. Dengan banyak *tweet* atau sentimen masyarakat yang diambil pada periode Juni 2023 dengan tiap kata kuncinya yang berbeda mendapatkan, yaitu untuk kata kunci “Ganjar Pranowo” dengan *tweet* sebanyak 325 memperoleh 241 *tweet* positif (74.15%), dan 84 *tweet* negatif (25.85%). Lalu dengan kata kunci “Prabowo Subianto” dengan *tweet* sebanyak 313 memperoleh 77 *tweet* positif (24.57%) dan 236 *tweet* negatif (75.32%), pada kata kunci “Anies Baswedan” dengan *tweet* sebanyak 322 memperoleh 138 *tweet* positif (42.86%) dan 184 *tweet* negatif (57.14%).

#### DAFTAR PUSTAKA

- [1] M. I. Khatami, “Eksistensi ‘Baliho 2024’ dalam Pertarungan Elektabilitas: Analisis Pencitraan Tokoh Politik,” *JISIP UNJA (Jurnal Ilmu Sosial Ilmu Politik Universitas Jambi)*, vol. 5, no. 2, pp. 14–24, 2021.
- [2] D. Murthy, *Twitter*. Polity Press Cambridge, 2018.
- [3] A. Deviyanto and M. D. R. Wahyudi, “Penerapan analisis sentimen pada pengguna twitter menggunakan metode K-Nearest Neighbor,” *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 3, no. 1, pp. 1–13, 2018.
- [4] S. Kurniawan, W. Gata, D. A. Puspitawati, M. Tabrani, and K. Novel, “Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online,” *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, vol. 3, no. 2, pp. 176–183, 2019.
- [5] A. M. Zuhdi, E. Utami, and S. Raharjo, “Analisis sentiment twitter terhadap capres Indonesia 2019 dengan metode K-NN,” *Jurnal Informa: Jurnal Penelitian dan Pengabdian Masyarakat*, vol. 5, no. 2, pp. 1–7, 2019.
- [6] A. D. A. Putra and S. Juanita, “Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 8, no. 2, pp. 636–646, 2021.
- [7] R. Malonda, “Opini Publik Terhadap Pencitraan Politik Dalam Meningkatkan Tingkat Elektabilitas Politik Pada Pemilu Presiden Tahun 2019 Di Kabupaten Minahasa,” *POLITICO: Jurnal Ilmu Politik*, vol. 8, no. 4, 2019.
- [8] H. Peng, L. Xu, L. Bing, F. Huang, W. Lu, and L. Si, “Knowing what, how and why: A near complete solution for aspect-based sentiment analysis,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, pp. 8600–8607.
- [9] A. Deolika, K. Kusriani, and E. T. Luthfi, “Analisis Pembobotan Kata Pada Klasifikasi Text Mining,” (*JurTI) Jurnal Teknologi Informasi*, vol. 3, no. 2, pp. 179–184, 2019.
- [10] B. Mahesh, “Machine learning algorithms-a review,” *International Journal of Science and Research (IJSR)*. [Internet], vol. 9, no. 1, pp. 381–386, 2020.
- [11] Z.-H. Zhou, *Machine learning*. Springer Nature, 2021.
- [12] M. J. Denny and A. Spirling, “Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it,” *Political Analysis*, vol. 26, no. 2, pp. 168–189, 2018.
- [13] T. Turki and S. S. Roy, “Novel Hate Speech Detection Using Word Cloud Visualization and Ensemble Learning Coupled with Count Vectorizer,” *Applied Sciences*, vol. 12, no. 13, p. 6611, 2022.