

Perbandingan Metode *K-Nearest Neighbors* Dan *Naïve Bayes Classification* Dalam Klasifikasi Sentimen Masyarakat Terhadap Pemilu Presiden Tahun 2024

Miftah Fadil^{1*}, Diana²

Teknik Informatika, Universitas Bina Darma, Palembang, Indonesia

*e-mail *Corresponding Author*: miftahfadil38@gmail.com

Abstract

Sentiment is a subjective expression or assessment of a person's feelings or attitude toward a topic, event, or situation. Therefore, sentiment analysis is important to carry out because it has many positive impacts that can be achieved, one of which is in research and public understanding. In this research, public sentiment was obtained from the social media Twitter. In this research, sentiment classification will use the K-Nearest Neighbors method and the Naïve Bayes Classification method. With the hope of knowing and gaining a deeper understanding of how each method works in classifying sentiment, as well as the public's views on the 2024 presidential election. This study concludes that the findings indicate that public sentiment towards the 2024 presidential election can be said to be positive because the public's positive sentiment has more than half of the total existing data, namely 365 positive, 197 neutral and 82 negative. Then the researchers obtained an accuracy of 58% for the K-Nearest Neighbors Method and 52% for the Naïve Bayes Classification Method.

Keywords: *Presidential Election; Sentiment; Naïve Bayes Classification; K-Nearest Neighbors*

Abstrak

Sentimen merupakan ekspresi atau penilaian subjektif dari perasaan atau sikap seseorang terhadap suatu topik, peristiwa, atau situasi. Oleh karena itu, Analisis sentimen penting untuk dilakukan karena memiliki banyak dampak positif yang dapat di raih salah satunya dalam penelitian dan pemahaman masyarakat. Pada penelitian ini sentimen masyarakat didapatkan dari *media social Twitter*. Di penelitian ini klasifikasi sentimen akan menggunakan metode *K-Nearest Neighbors* dan Metode *Naïve Bayes Classification*. Dengan harapan dapat mengetahui dan memperoleh pemahaman yang lebih mendalam tentang bagaimana setiap metode bekerja dalam mengklasifikasikan sentimen, serta pandangan masyarakat terhadap pemilu presiden tahun 2024. Kesimpulan dari penelitian ini yaitu hasil menunjukkan bahwa sentimen masyarakat terhadap pemilu presiden 2024 bisa dikatakan positif dikarenakan sentimen positif masyarakat sudah melebihi setengah dari total data yang ada yaitu sebanyak 365 positif, 197 netral, dan 82 negatif ini. Kemudian peneliti memperoleh akurasi sebesar 58 % untuk Metode *K-Nearest Neighbors* dan 52% untuk Metode *Naïve Bayes Classification*.

Kata kunci: *Pemilu Presiden; Sentimen; Naïve Bayes Classification; K-Nearest Neighbors*

1. Pendahuluan

Sentimen adalah ekspresi atau penilaian subjektif dari perasaan atau sikap seseorang terhadap suatu topik, peristiwa, atau situasi. Analisis sentimen adalah teknik menafsirkan dan pengolahan data tekstual untuk mengekstrak data opini dan menentukan sentimen yang ada dalam sebuah opini[1]. Analisis sentimen memiliki banyak dampak positif yang dapat di raih salah satunya dalam penelitian dan pemahaman masyarakat. Analisis sentimen penting dalam penelitian sosial untuk memahami pandangan dan sikap masyarakat terhadap isu-isu tertentu[2].

Data yang dikumpulkan untuk menjadi objek analisis sentimen yaitu berjumlah 670 data *tweet* dalam kurun waktu 1 bulan. Perbandingan yang akan dilakukan adalah perbandingan akurasi dalam mengklasifikasikan data dari sentimen masyarakat terhadap Pemilu Presiden Tahun 2024. Dengan membandingkan akurasi diantara kedua metode tersebut diharapkan penelitian ini bisa memberikan informasi yang lebih baik dan akurat dalam klasifikasi sentimen masyarakat terhadap Pemilu Presiden Tahun 2024.

Dalam melakukan analisis sentimen diperlukan metode untuk mengklasifikasikan sentimen tersebut. Di penelitian ini klasifikasi sentimen akan menggunakan metode *K-Nearest Neighbors* serta *NaïveBayes Classification*. *K-Nearest Neighbors* merupakan metode yang biasa digunakan untuk melakukan kategorisasi data berdasarkan jarak terkecil terhadap objek data[3]. Metode *K-Nearest Neighbors* sering digunakan di dalam penelitian sebelumnya karena modelnya cukup mudah dipahami dan sering kali berfungsi cukup baik dengan sedikit modifikasi[4]. Selanjutnya Metode *NaïveBayes Classification* merupakan sebuah metode kategorisasi data yang merupakan turunan dari teorema *Bayes* yang menggunakan teknik statistik serta probabilistik [5].Metode ini memiliki banyak kelebihan seperti dapat digunakan untuk data kualitatif dan kuantitatif, perhitungannya yang cepat dan efisien, serta mudah dipahami.

Tujuan dari penelitian ini adalah melakukan analisis sentimen pada *Twitter* terhadap hasil pemilu Indonesia tahun 2024 dengan menerapkan Metode *K-Nearest Neighbors* dan Metode *NaïveBayes Classification*. Sikap positif, negatif, dan netral dibedakan dan dikelompokkan peneliti menggunakan Metode *K-Nearest Neighbors* dan Metode *NaïveBayes Classification*, serta menggunakan teks *real-time* yang diambil dari *Twitter*. Hasil analisis sentimen ini diharapkan dapat memberikan hasil akurasi tertinggi dari perbandingan metode *K-Nearest Neighbors* dan *NaïveBayes Classification*, mengetahui dan memperoleh pemahaman yang lebih mendalam tentang bagaimana setiap metode bekerja dalam mengklasifikasikan sentimen, serta pandangan masyarakat terhadap pemilu presiden tahun 2024.

2. Tinjauan Pustaka

Sebagai bahan tinjauan dalam melakukan penelitian, peneliti menggunakan penelitian terdahulu yang relevan dengan penelitian saat ini. Penelitian itu digunakan sebagai referensi literatur sebagaimana beberapa penelitian berikut ini.

Penelitian pertama yang dilakukan Slamet Harry Ramadhani dan Muhammad Iwan Wahyudin diberi nama "Analisis Sentimen Terhadap Vaksinasi *Astra Zeneca* pada *Twitter* Menggunakan Metode *Naïve Bayes* dan K-NN"[6]. Setiap pendekatan dalam penelitian ini menghasilkan tingkat akurasi yang berbeda-beda; misalnya metode KNN menghasilkan tingkat akurasi sebesar 74.78% +/- 3.74% (rata-rata mikro: 74.77%), sedangkan metode *Naïve Bayes* menghasilkan tingkat akurasi sebesar 90.71% +/- 4.66% (rata-rata mikro: 90.77%) .

Penelitian kedua Yuyun, Nurul Hidayah, dan Supriadi Sahibu diberi nama "*Algoritma Multinomial Naïve Bayes* Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan *Covid-19* Menggunakan Data *Twitter*"[7]. Penelitian ini menghasilkan *weighted average* untuk *precision*, *recall* dan akurasi sebesar 74%. Penelitian menunjukkan bahwa pendekatan yang disarankan memiliki tingkat akurasi yang sangat tinggi.

Penelitian ketiga yang ditulis oleh Tiara Dwi Arista, Yusra, Muhammad Fikry, dan Lola Oktavia berjudul "Klasifikasi Sentimen Masyarakat di *Twitter* terhadap Kenaikan Harga BBM dengan Metode K-NN"[8]. Metode K-NN diuji pada penelitian ini menggunakan matriks konfusi pada sepuluh nilai K yang berbeda (3,5,7,9,11,13,15,17,19, dan 21) dengan mekanisme perbandingan 70:30, 80:20, dan 90:10. Perbandingan data latih dan uji 90:10 menghasilkan akurasi tertinggi sebesar 83,3% pada K=13 dan K=15.

Pebandingan penelitian ini dengan penelitian terdahulu terletak pada cara melakukan pengklasifikasian dan pencarian hasil akurasi. Dimana penelitian ini sepenuhnya dilakukan oleh perhitungan komputer yaitu dengan menggunakan *Google colab* dan bahasa Pemrograman *Python*. Mulai dari pengumpulan data hingga menghasilkan nilai akurasi yang didapatkan

3. Metodologi

3.1. Jenis Penelitian

Berdasarkan data yang akan diteliti, penelitian ini akan menggunakan jenis penelitian kuantitatif yaitu pendekatan penelitian yang menggunakan informasi serta fakta yang tepat dalam bentuk angka[9]. Penelitian ini akan menghasilkan data berupa angka yang mempresentasikan jumlah sentimen masyarakat yang di klasifikasikan menjadi 3 yaitu Positif, Netral, dan Negatif. Kemudian hasilnya akan berupa persentase akurasi dari metode yang digunakan.

3.2. Data Penelitian

Penelitian ini menggunakan jenis data primer. Data primer adalah data yang bersumber *internal* yang didapatkan peneliti secara langsung[10]. Data penelitian yang telah didapatkan

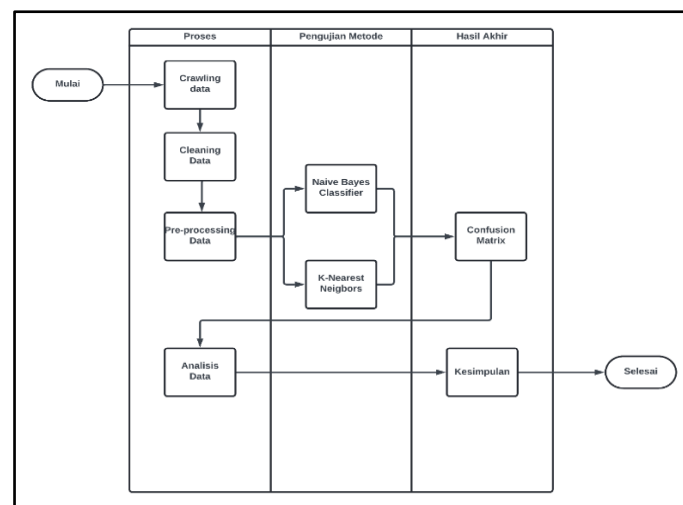
berjumlah 669 data, kemudian berkurang sebanyak 25 data yang merupakan data duplikat. Sehingga data penelitian yang akan digunakan berjumlah 644 data yang berbentuk file csv.

3.3. Metode Pengumpulan Data

Metode pengumpulan data yang digunakan adalah *Social Media Mining*. *Social Media Mining* adalah metode mengumpulkan data dari platform media sosial untuk analisis sentimen atau perilaku pengguna. *Social Media Mining* merupakan bagian dari *Data Mining*. *Data Mining* adalah Proses memperoleh atau mengambil banyak data dan informasi yang sebelumnya tidak diketahui [11]. data yang dikumpulkan menggunakan *Google Colab* yang dapat diakses secara gratis, dengan melakukan *crawling* menggunakan *Google Colab* untuk melakukan pencarian data dengan kata kunci "Pemilu Presiden 2024". Data yang dikumpulkan berjumlah 669 data tweet dalam kurun waktu 1 bulan.

3.4. Tahapan Penelitian

Untuk mencapai tujuan penelitian. Proses penelitian dibagi menjadi beberapa tahapan yang dimulai dari awal sampai akhir. Tahapan penelitian dapat dilihat pada gambar 1 dibawah.



Gambar 1. Tahapan Penelitian

Untuk memperoleh hasil yang diinginkan dari analisis sentimen, pendekatan penelitian memerlukan sejumlah tahapan yang harus diselesaikan. Tahapan ini terdiri dari:

- 1) *Crawling Data*, data yang dikumpulkan menggunakan *Google Colab* yang dapat diakses secara gratis, dengan melakukan *crawling* menggunakan *Google Colab* untuk melakukan pencarian data dengan kata kunci "Pemilu Presiden 2024". Di tahapan ini data dari *tweet* akan dikumpulkan menjadi 1 file yang akan disimpan dalam bentuk file teks yang memiliki format khusus yang memungkinkan data disimpan dalam format tabel terstruktur atau lebih dikenal dengan File CSV (nilai dipisahkan koma). Data yang dikumpulkan berjumlah 669 data tweet dalam kurun waktu 1 bulan.
- 2) *Cleaning dan Pre- Processing* yaitu di tahap ini data yang dikumpulkan akan diklasifikasikan sebagai penentu tingkat keakurasian. Di tahapan ini data akan di proses untuk memudahkan dalam pengujian metode nantinya. Teknik yang digunakan dalam *Pre- Processing* ini adalah *Cleaning* (Menghapus elemen-elemen yang tidak diinginkan atau mengganggu dari teks untuk meningkatkan kualitas data), *Normalisasi* (Mengubah teks menjadi format standar untuk memastikan konsistensi), *Tokenize* (Memecah teks menjadi unit-unit yang lebih kecil seperti kata-kata atau kalimat untuk analisis lebih lanjut), *Stem* (Mengubah kata-kata menjadi bentuk dasarnya dengan menghapus akhiran untuk mengurangi variasi kata), *Tranlate* (Menerjemahkan dataset ke bahasa inggris), dan *Labelling* (memberikan dataset label positif, negatif, netral).
- 3) *Pengujian Metode* yaitu tahap pengujian yang menggunakan Metode *Naive Bayes Classification* dan *K-Nearest Neighbors* untuk menentukan besaran nilai akurasi yang didapatkan. Tahapan ini dilakukan dengan cara terpisah yaitu dengan cara mencari performa

masing-masing metode dengan menggunakan algoritma yang berbeda kemudian akan dibandingkan performa masing-masing metode. Kedua Metode ini akan menghasilkan *Confusion Matrix*. Hasil yang diperoleh dalam tahapan ini akan dibuat dalam bentuk tabel yang akan menjadi perbandingan dalam klasifikasi sentimen. Adapun rumus untuk mencari nilai akurasi yaitu :

$$Akurasi = \frac{AP+ANet+AN}{AP+PP+ANet+PNet+AN+PN} \dots\dots\dots(1)$$

Keterangan:

AP : *Actual* Positif
 PP : *Predicted* Positif
 Anet : *Actual* Netral
 PNet : *Predicted* Netral
 AN : *Actual* Negatif
 PN : *Predicted* Negatif

3.5. Metode *K-Nearest Neighbors*

Metode *K-Nearest Neighbors* merupakan metode yang paling sederhana dan seringkali menjadi langkah pertama dalam mempelajari *machine learning*. *K-Nearest Neighbors* memiliki keuntungan dalam kesederhanaannya dan kemampuannya untuk menangani data non-linier atau data dengan distribusi yang tidak diketahui. Namun, kelemahan utamanya adalah performa yang kurang baik pada *dataset* besar karena memerlukan perhitungan jarak untuk setiap pasangan titik data dalam *dataset* pelatihan. Selain itu, *K-Nearest Neighbors* juga sensitif terhadap skala data dan membutuhkan pengaturan parameter *k* yang tepat.

K-Nearest Neighbors adalah Teknik klasifikasi pengumpulan data berdasarkan data pembelajaran yang telah diklasifikasi sebelumnya[12]. Sudut pandang lain menyatakan bahwa teknik *K-Nearest Neighbors* adalah teknik pembelajaran terawasi yang cocok untuk aplikasi regresi dan klasifikasi. [13]. Untuk memprediksi label kelas atau nilai sampel, pertama-tama cari titik data *K* yang paling dekat dengan sampel yang disediakan. Kemudian, ia menggunakan label kelas atau nilai dari *K* titik data tersebut. Dari kedua definisi di atas terlihat jelas bahwa pendekatan *K-Nearest Neighbors* merupakan suatu cara untuk mengkategorikan objek berdasarkan data pembelajaran yang terdekat dengan item tersebut.

Kategorisasi kedekatan titik digunakan oleh metode *K-Nearest Neighbors* untuk memperkirakan nilai untuk *instance query* baru. Pendekatan *K-Nearest Neighbors* dapat digunakan untuk menghitung jarak kedekatan data dengan cara sebagai berikut:

$$d = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \dots\dots\dots(2)$$

Keterangan :

d : jarak
 a : data uji/testing
 b : sampel data
 i : variabel data
 n : dimensi data

3.6. Metode *NaïveBayes Classification*.

NaïveBayes Classification merupakan metode yang sering digunakan dalam berbagai aplikasi klasifikasi, termasuk klasifikasi teks (misalnya, klasifikasi *email spam* vs. *non-spam*), analisis sentimen, klasifikasi dokumen, dan lain-lain. Kelebihan dari *NaïveBayes Classification* meliputi kemampuan untuk menangani data dengan fitur yang besar, interpretabilitas yang baik, dan efisiensi komputasi yang tinggi. Ciri mendasar klasifikasi *Naïve Bayes* adalah asumsinya yang kuat (naif) mengenai independensi setiap kondisi dan peristiwa.

NaïveBayes Classification merupakan Teknik klasifikasi yang paling banyak digunakan adalah berdasarkan probabilitas dasar, dimaksudkan untuk diterapkan dengan asumsi bahwa faktor-faktor penjas tidak bergantung satu sama lain, dan memiliki tingkat akurasi yang tinggi[14]. Ada juga pendapat lain yang menyatakan bahwa *NaïveBayes Classification (NBC)* merupakan Teknik klasifikasi yang menerapkan teorema *Bayes* untuk meramalkan kemungkinan suatu objek tertentu akan menjadi anggota suatu kelas[15]. Terakhir, metode klasifikasi *Naïve*

Bayes berupaya mengkategorikan data ke dalam kelompok tertentu. Angka akurasi prediksi berfungsi sebagai ukuran efektivitas pengklasifikasi[16]. Dari penjelasan diatas dapat disimpulkan bahwa Metode *NaïveBayes Classification* adalah metode klasifikasi yang digunakan untuk melakukan prediksi data pada kelas tertentu.

Berikut persamaan yang dapat digunakan sebagai panduan untuk memperkirakan nilai probabilitas dalam pengambilan Keputusan yang digunakan dalam *NaïveBayes Classification*:

$$P(A|B) = \frac{P(B|A)*P(A)}{P(B)} \dots\dots\dots (3)$$

Keterangan :

- P(A|B) : Probabilitas bersyarat A yang diberikan oleh B
- P(B|A) : Probabilitas bersyarat B yang diberikan oleh A
- P(A) : Probabilitas A
- P(B) : Probabilitas B

4. Hasil dan Pembahasan

4.1 Pengumpulan Data

Di penelitian ini data tweet dikumpulkan dengan metode *crawling* yang dilakukan menggunakan *Tweet Harvest Tool* yang diinstall menggunakan *Google Colab* dengan kata kunci "Pemilu Presiden 2024". Disini data yang berhasil dikumpulkan berjumlah 669 data. Data tersebut merupakan data *tweet* yang diposting oleh pengguna dalam rentan waktu kurang lebih 1 bulan dalam rician dari tanggal 25 Maret 2024 sampai dengan 20 April 2024. Dibawah ini merupakan hasil dari *Crawling data*.

	conversation_id_str	created_at	favorite_count	full_text
0	1781598947146158204	Sat Apr 20 08:21:04 +0000 2024	14	Area gedung Mahkamah Konstitusi bakal steril s...
1	1781594666720522535	Sat Apr 20 08:04:03 +0000 2024	2	Promo Buku Kuasa Uang karya Prof. @BurhanMuhta...
2	1781215227998326853	Sat Apr 20 05:44:31 +0000 2024	0	@gilang_ahm31272 @Ngadimi08002459 majalah ini ...
3	1781216356282798546	Fri Apr 19 22:33:19 +0000 2024	1	@andiebs @DoankWarto Sengketa tiap pemilu pres...
4	1781276485992743362	Fri Apr 19 10:59:44 +0000 2024	2	Selama kurang lebih satu jam pertemuan berlang...
...
664	1772172093565907109	Mon Mar 25 08:02:07 +0000 2024	0	Partisipasi pemilih Pemilu 2024 sedikit menuru...
665	1772172006156337389	Mon Mar 25 08:01:46 +0000 2024	0	Presiden terpilih Prabowo Subianto menyampaika...
666	1772168641989009601	Mon Mar 25 07:48:38 +0000 2024	0	Ketua Umum Partai Nasdem Surya Paloh menyataka...
667	1772167049134956979	Mon Mar 25 07:42:15 +0000 2024	0	Sebelumnya Prabowo juga mendapatkan ucapan sel...
668	1772166807992058296	Mon Mar 25 07:41:16 +0000 2024	0	Sebelumnya Prabowo juga mendapatkan ucapan sel...

Gambar 2. *Dataset* Hasil Pengumpulan Data

4.2 Pre-Processing Data

Dataset yang telah dikumpulkan dan disimpan selanjutnya akan di proses. Tujuan dari proses ini yaitu menghapus data yang tidak akan digunakan untuk proses selanjutnya nanti. Di proses ini akan dilakukan *Pre-Processing Data* seperti *Cleaning*, *Normalisasi*, *Tokenize*, *Stem*, dan *Tranlate*. Berikut merupakan contoh dari proses *Pre-Processing Data* yang digunakan dalam penelitian, dapat dilihat pada tabel 1.

Tabel 1. Hasil Tahapan *Pre-Processing Data*

Tahapan	Hasil
Data Awal	Tanda tanda bhw MK dlm sidang gugat pemilu curang 2024 hanya berpura pura mengapa 4 saksi menteri tdk di ambil sumpah sblm memberi keterangan? Di ujung sidang sangat mungkin 02 tetap lolos sebagai presiden

Tahapan	Hasil
<i>Cleaning</i>	yg menang cara curang. @officialMKRI @bawaslu_RI https://t.co/vOJ8WksnVx Tanda tanda bhw MK dlm sidang gugat pemilu curang 2024 hanya berpura pura mengapa 4 saksi menteri tdk di ambil sumpah sbm memberi keterangan Di ujung sidang sangat mungkin 02 tetap lolos sebagai presiden yg menang cara curang.
Normalisasi	Tanda tanda bahwa MK dalam sidang gugat pemilu curang 2024 hanya berpura pura mengapa 4 saksi menteri tidak di ambil sumpah sebelum memberi keterangan Di ujung sidang sangat mungkin 02 tetap lolos sebagai presiden yang menang cara curang.
<i>Tokenize</i>	Tanda, tanda, bahwa, MK ,dalam ,sidang, gugat, pemilu, curang ,2024, hanya ,berpura, pura, mengapa ,4 ,saksi ,Menteri, tidak, di, ambil, sumpah, sebelum, memberi, keterangan, Di, ujung, siding,, sangat mungkin, 02, tetap, lolos, sebagai, presiden, yang ,menang, cara, curang
<i>Stem</i>	Tanda, tanda, bahwa, MK ,dalam ,sidang, gugat, pemilu, curang ,2024, hanya ,pura, pura, mengapa ,4 ,saksi ,Menteri, tidak, di, ambil, sumpah, belum, beri, keterangan, Di, ujung, sidang,, sangat mungkin, 02, tetap, lolos, sebagai, presiden, yang ,menang, cara, curang
<i>Translate</i>	<i>There are signs that the Constitutional Court in the 2024 fraudulent election lawsuit trial is just pretending why were the 4 ministerial witnesses not sworn in before giving statements At the end of the trial, it is very possible that 02 will still pass as the president who won by cheating</i>

Setelah melalui proses *Pre-Processing Data*, *dataset* akan diberikan label yang bertujuan untuk memberikan label positif, netral, dan negatif kepada *dataset* dengan cara melihat kolom *full_text* yang berbahasa inggris. Disini *dataset* di berikan label sesuai dengan kata yang terdapat didalam baris data tersebut. Misalkan jika data itu mengandung kata yang memiliki makna “tidak setuju” maka akan diberikan label negatif begitupun sebaliknya. Setelah melakukan tahapan *Labelling*. Peneliti memperoleh hasil analisis sentimen yang dapat dilihat pada tabel 2 dibawah ini.

Tabel 2. Hasil Pelabelan Data

Sentimen	Hasil	Persentase
Positif	365	57 %
Netral	197	30 %
Negatif	82	13 %

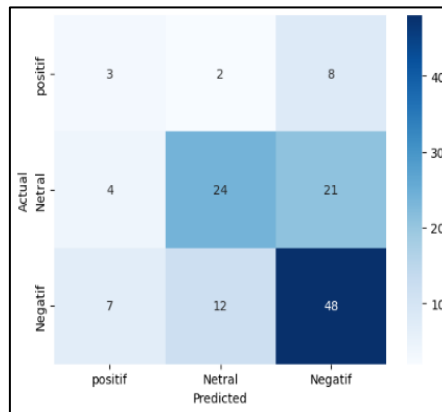
4.3 Pengujian Metode

Dataset yang telah melalui proses *pre-processing* kemudian akan diuji menggunakan 2 metode untuk dicari nilai akurasi dengan menggunakan Metode *K-Nearest Neighbors* dan *NaïveBayes Classification*. Pada pengujian metode ini *dataset* akan diolah menggunakan algoritma masing-masing metode yang dikerjakan secara sepenuhnya dilakukan oleh perhitungan komputer. Berikut adalah Hasil pengujian dari masing-masing metode.

Di metode *K-Nearest Neighbors* peneliti menetapkan nilai k sebesar 5 kemudian peneliti menetapkan 20 % data uji dan 80 % data latih. Disini metode *K-Nearest Neighbors* mendapatkan hasil akurasi sebesar 58 %. Secara detail dapat dilihat pada tabel dibawah yang menyajikan *Confusion Matrix* untuk metode *K-Nearest Neighbors*.

Tabel 3. Hasil Klasifikasi *K-Nearest Neighbors*.

Sentimen	<i>Precision</i>	<i>recall</i>	<i>f1-score</i>
Positif	0.21	0.23	0.22
Netral	0.63	0.49	0.55
Negatif	0.62	0.72	0.67



Gambar 3. *Confusion Matrix* untuk metode *K-Nearest Neighbors*

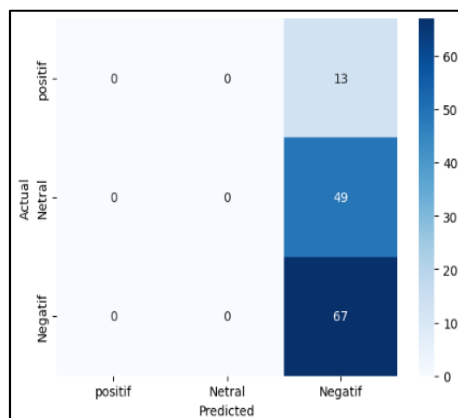
Dimana Hasil Akurasi didapatkan dari :

$$Akurasi = \frac{AP + ANet + AN}{AP + PP + ANet + PNet + AN + PN} = \frac{3 + 24 + 48}{3 + 10 + 24 + 25 + 48 + 19} = \frac{75}{129} = 0.5813 \text{ atau } 58\%$$

Di Metode *NaïveBayes Classification* peneliti menetapkan 20 % data uji dan 80 % data latih. Disini Metode *NaïveBayes Classification* mendapatkan hasil akurasi sebesar 52 %. Secara detail dapat dilihat pada tabel dibawah yang menyajikan *Confusion Matrix* untuk Metode *NaïveBayes Classification*.

Tabel 4. Hasil Klasifikasi *NaïveBayes Classification*.

Sentimen	<i>Precision</i>	<i>recall</i>	<i>f1-score</i>
Positif	0	0	0
Netral	0	0	0
Negatif	0.52	1.00	0.68



Gambar 4. *Confusion Matrix* untuk Metode *NaïveBayes Classification*.

Dimana Hasil Akurasi didapatkan dari :

$$Akurasi = \frac{AP + ANet + AN}{AP + PP + ANet + PNet + AN + PN} = \frac{0 + 0 + 67}{0 + 13 + 0 + 49 + 67 + 0} = \frac{67}{129} = 0.5193 \text{ atau } 52\%$$

Setelah melakukan pengujian metode pada *dataset*, diperoleh hasil Klasifikasi *K-Nearest Neighbors* dapat dilihat pada tabel 3 menunjukkan hasil kinerja yang tidak cukup baik dari metode *K-Nearest Neighbors* begitu juga dengan hasil Klasifikasi *NaïveBayes Classification* yang bisa dilihat pada tabel 4. Dari hasil pengujian kedua metode mendapatkan nilai akurasi yang cukup rendah, hal ini dapat terjadi dikarenakan dalam proses pengklasifikasian sentimen untuk *dataset* yang dilakukan oleh perhitungan komputer tidak dapat berjalan dengan baik yang disebabkan oleh kurangnya performa dalam penerjemahan yang dilakukan oleh komputer. Akan tetapi pengujian metode masih bisa berjalan dengan lancar walaupun dengan kekurangan tadi dikarenakan telah melalui prosedur yang benar. Hal ini selaras dengan penelitian yang dilakukan [7], dan [8] dalam menerapkan metode *K-Nearest Neighbors* dan *NaïveBayes Classification*.

Setelah melihat hasil pengujian metode dimana memperoleh akurasi sebesar 58 % untuk Metode *K-Nearest Neighbors* dan 52% untuk Metode *NaïveBayes Classification*. Hal ini tidak selaras dengan penelitian yang dilakukan [6] dimana di penelitian tersebut metode *NaïveBayes Classification* memiliki angka akurasi yang lebih tinggi daripada metode *K-Nearest Neighbors*. Ini dapat terjadi karena perbedaan dari *tools* yang digunakan dalam pengujian metode.

5. Simpulan

Penelitian ini menerapkan Metode *K-Nearest Neighbors* dan Metode *NaïveBayes Classification* untuk klasifikasi sentimen. Klasifikasi sentimen dalam riset ini menggunakan data tweet Twitter sebanyak 669 data. Kemudian setelah di proses jumlah datanya berkurang sehingga menjadi 644 data. Hasilnya menunjukkan bahwa sentimen masyarakat terhadap pemilu presiden 2024 bisa dikatakan positif dikarenakan sentimen positif masyarakat sudah melebihi setengah dari total data yang ada yaitu sebanyak 365 positif, 197 netral, dan 82 negatif ini. Kemudian dengan menerapkan Metode *K-Nearest Neighbors* dan Metode *NaïveBayes Classification* untuk klasifikasi sentimen, peneliti memperoleh akurasi sebesar 58 % untuk Metode *K-Nearest Neighbors* dan 52% untuk Metode *NaïveBayes Classification*. Sehingga dari hasil diatas dapat dikatakan bahwa metode *K-Nearest Neighbors* mempunyai tingkat akurasi yang lebih baik daripada Metode *NaïveBayes Classification* untuk *dataset* yang diteliti.

Saran untuk penelitian selanjutnya adalah supaya bisa memperoleh hasil akurasi yang lebih tinggi lagi dari kedua metode yang digunakan di penelitian ini dapat dilakukan dengan cara menambah jumlah *dataset* yang lebih banyak serta menerjemahkan *dataset* kedalam bahasa Inggris dengan tata bahasa yang benar agar pengklasifikasian yang dilakukan dapat memberikan hasil yang lebih baik.

Daftar Referensi

- [1] F. V. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online JD.ID Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *Jurnal SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019, doi: 10.52158/jacost.v1i1.9.
- [2] A. Liawati, R. Narasati, D. Solihudin, C. Lukman Rohmat, and S. Eka Permana, "Analisis Sentimen Komentar Politik Di Media Sosial X Dengan Pendekatan Deep Learning," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, pp. 3557–3563, 2024, doi: 10.36040/jati.v7i6.8248.
- [3] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, "Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 6, no. 2, pp. 118–127, 2021, doi: 10.31294/ijcit.v6i2.10438.
- [4] B. Raharjo, *Pembelajaran Mesin (Machine Learning)*. Semarang: Yayasan Prima Agus Teknik, 2021.
- [5] A. F. Watratan, A. Puspita, and D. Moeis, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 di Indonesia," *Journal of Applied Computer Science and Technology*, vol. 1, no. 1, pp. 7–14, 2020, doi: 10.52158/jacost.v1i1.9.
- [6] S. H. Ramadhani and M. I. Wahyudin, "Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naïve Bayes dan K-NN," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 6, no. 4, pp. 526–534, 2022, doi: 10.35870/jti.k.v6i4.530
- [7] Yuyun, Nurul Hidayah, and Supriadi Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 820–826, Aug. 2021, doi: 10.29207/resti.v5i4.3146.

- [8] T. D. Arista, Y. Yusra, M. Fikry, and L. Oktavia, "Klasifikasi Sentimen Masyarakat di Twitter terhadap Kenaikan Harga BBM dengan Metode K-NN," *JUKI: Jurnal Komputer dan Informatika*, vol. 5, no. 1, pp. 140–150, 2023, doi: 10.53842/juki.v5i1.189.
- [9] M. Waruwu, "Pendekatan Penelitian Pendidikan: Metode Penelitian Kualitatif, Metode Penelitian Kuantitatif dan Metode Penelitian Kombinasi (Mixed Method)," *Jurnal Pendidikan Tambusai*, vol. 7, no. 1, pp. 2896–2910, 2023, doi: 10.31004/jptam.v7i1.6187.
- [10] Y. S. Siregar, M. Darwis, R. Baroroh, and W. Andriyani, "Peningkatan Minat Belajar Peserta Didik dengan Menggunakan Media Pembelajaran yang Menarik pada Masa Pandemi Covid 19 di SD Swasta HKBP 1 Padang Sidempuan," *Jurnal Ilmiah Kampus Mengajar*, vol. 2, no. 1, pp. 69–75, Apr. 2022, doi: 10.56972/jikm.v2i1.33.
- [11] C. Zai, "IMPLEMENTASI DATA MINING SEBAGAI PENGOLAHAN DATA," *Jurnal Portal Data*, vol. 2, no. 3, pp. 1–12, 2022.
- [12] D. Cahyanti, A. Rahmayani, and S. A. Husniar, "Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," *Indonesian Journal of Data and Science*, vol. 1, no. 2, pp. 39–43, 2020, doi: 10.33096/ijodas.v1i2.13.
- [13] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *Jurnal KomtekInfo*, vol. 10, no. 1, pp. 1–7, Jan. 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [14] N. R. Siahaan, R. Y. Tiffany, and S. R. E. Sinaga, "Analisis Sentimen Ulasan Aplikasi Media Sosial Whatsapp Menggunakan Metode Naive Bayes Classifier," *Jurnal Ilmiah Betrik*, vol. 14, no. 02, pp. 343–354, 2023, doi: 10.36050/betrik.v14i02%20AGUSTUS.104.
- [15] I. M. B. Adnyana, "Implementasi Naïve Bayes Untuk Memprediksi Waktu Tunggu Alumni Dalam Memperoleh Pekerjaan," in *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 2020, pp. 131–134.
- [16] K. Anwar, "Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 2, no. 4, pp. 148–155, 2022, doi: 10.30865/klik.v2i4.315.