

ANALISIS SENTIMEN TERHADAP PRESIDEN PADA FACEBOOK DENGAN MENGGUNAKAN METODE NAÏVE BAYES

Fania Salsabila¹, Arief Wibowo²

^{1,2}Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, DKI Jakarta, Indonesia

Email: ¹*fania49salsabila@gmail.com, ²arief.wibowo@budiluhur.ac.id

(* : corresponding author)

Abstrak-Facebook merupakan layanan jejaring sosial media yang memungkinkan pengguna untuk saling terhubung dengan pengguna lainnya dari seluruh dunia. Dalam layanan Facebook, para pengguna internet menulis opini dan pendapat tentang berbagai topik. Facebook menempati peringkat ketiga pada website yang banyak dikunjungi orang di Indonesia dengan jumlah total pengunjung 487 juta per bulan. Informasi yang terdapat di Facebook berupa pertanyaan, opini atau komentar, baik positif maupun negatif. Karena pengguna membicarakan opini yang berbeda-beda, hal ini menjadikan Facebook sebagai sumber data yang bagus untuk mengetahui sentiment masyarakat terhadap Presiden Jokowi. Penelitian ini menggunakan Facebook sebagai sumber data karena memiliki beberapa keunggulan, yaitu penggunaan pesan singkat oleh kelompok pengguna yang berbeda, yang berisi opini publik yang berbeda, yang terus bertambah dan menyebarkan berita dengan cepat. Naive Bayes adalah pengklasifikasi probabilistik langsung namun efisien yang menggunakan teorema Bayes dan asumsi "naive" bahwa fitur tidak bergantung satu sama lain. Penelitian ini dilakukan menggunakan algoritma Naive Bayes dengan 300 data yang telah diberi label secara manual. Pada tahap pengujian digunakan Confusion Matrix dan 10 Folds Cross Validation dengan membagi data menjadi beberapa himpunan. Setelah dilakukan pengujian pada 300 data yang dilakukan 10 kali dengan 30 data uji yang berbeda, didapatkan hasil berupa accuracy bernilai 87%, precision bernilai 77% dan recall bernilai 73%.

Kata Kunci: Analisis Sentimen, Facebook, Naive Bayes, Penambangan Teks

SENTIMENT ANALYSIS OF THE PRESIDENT ON FACEBOOK USING NAÏVE BAYES METHOD

Abstract- Facebook is a social media networking service that allows users to connect with other users from around the world. In the Facebook service, internet users write opinions and thoughts on various topics. Facebook ranks third on the website most visited by people in Indonesia with a total number of visitors 487 million per month. Information contained on Facebook is in the form of questions, opinions or comments, both positive and negative. Because users talk about different opinions, this makes Facebook a good source of data to find out people's sentiment towards President Jokowi. This study uses Facebook as a data source because it has several advantages, namely the use of short messages by different groups of users, which contain different public opinions, which continue to grow and spread news quickly. Naive Bayes is a direct but efficient probabilistic classifier that uses Bayes' theorem and the "naive" assumption that features are independent of each other. This research is conducted using the Naive Bayes algorithm with 300 data that have been manually labeled. At the testing stage, the Confusion Matrix and 10 Folds Cross Validation is used by dividing the data into several sets. After testing 300 data 10 times with 30 different test data, the results obtained are accuracy of 87%, precision of 77% and recall of 73%.

Keywords: Sentiment Analysis, Facebook, Naive Bayes, Text Mining

1. PENDAHULUAN

Di era digital saat ini, masyarakat sangat dimudahkan dengan adanya teknologi digital yang dapat menjangkau seluruh lapisan masyarakat. Masyarakat dapat dengan mudah memberikan pendapat, berkeluh kesah, berbagi cerita dan pengalaman, atau apapun yang ingin mereka bagikan kepada masyarakat melalui media sosial. Salah satu jenis media sosial yang populer di kalangan pengguna Internet adalah *microblogging*. Contoh layanan *microblogging* adalah Facebook. Dalam layanan Facebook, para pengguna internet menulis opini dan pendapat tentang berbagai topik. Pengguna dapat berkomunikasi dengan pengguna lain menggunakan Facebook dengan menambahkan teman, mengunggah foto dan video, membuat halaman pribadi (halaman penggemar), membangun grup dan komunitas, dan menggunakan fitur pesan untuk mengirim pesan.

Menurut laporan “Digital 2023 Indonesia” [1], ada sebanyak 83,8% dari jumlah populasi yang menggunakan Facebook di Indonesia pada Januari 2023. Jumlah tersebut meningkat 81,3% dibandingkan tahun lalu. Facebook menempati peringkat ketiga pada website yang banyak dikunjungi orang di Indonesia dengan jumlah total pengunjung 487 juta per bulan. Pengguna dapat berinteraksi dengan orang-orang dari seluruh dunia di Facebook, sebuah *platform* jejaring media sosial. Informasi yang terdapat di Facebook berupa pertanyaan, opini atau komentar, baik positif maupun negatif.

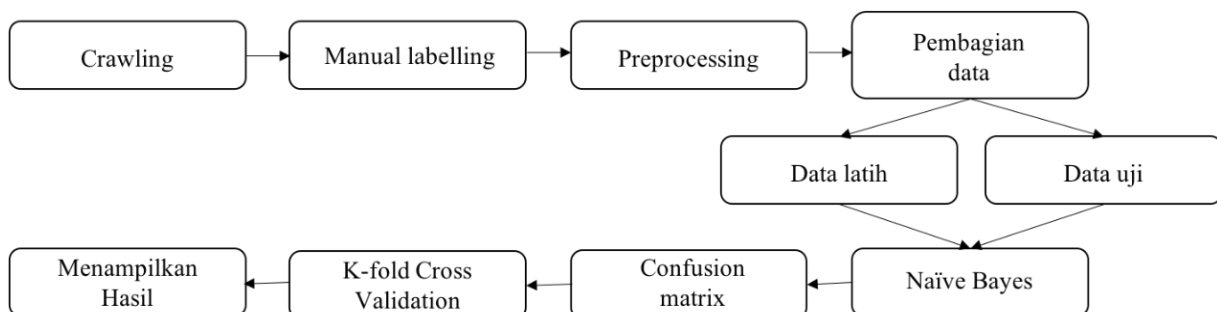
Karena pengguna membicarakan opini yang berbeda-beda, hal ini menjadikan Facebook sebagai sumber data yang bagus untuk mengetahui sentimen masyarakat terhadap Presiden Jokowi. Facebook juga sering digunakan sebagai media kampanye politik yang mendukung pasangan tertentu atau bahkan mencemarkan nama baik pasangan lainnya, digunakan sebagai alat protes, pembelajaran dan alat komunikasi krisis. Penelitian ini menggunakan Facebook sebagai sumber data karena memiliki beberapa keunggulan, yaitu penggunaan pesan singkat oleh kelompok pengguna yang berbeda, yang berisi opini publik yang berbeda, yang terus bertambah dan menyebarkan berita dengan cepat.

Pada penelitian analisis sentimen yang telah dilakukan sebelumnya pada sebuah media sosial yang menerapkan metode *naïve bayes* [2], penelitian tersebut menghasilkan kesimpulan berupa nilai akurasi sebesar 82%, nilai presisi sebesar 93% dan nilai *recall* sebesar 52%. Penelitian selanjutnya yang dilakukan pada komentar media sosial yang sama yaitu Facebook terhadap akun jasa ekspedisi barang menggunakan metode KNN [3] menghasilkan nilai rata-rata akurasi tertinggi sebesar 79.21% sedangkan akurasi terendah bernilai sebesar 70.3%. Adapun penelitian yang dilakukan pada *Twitter* [4] terhadap kinerja Presiden Jokowi 3 Periode menghasilkan akurasi sebesar 94,07% menggunakan *Naïve Bayes*. Berlandaskan latar belakang tersebut, tujuan dari penelitian ini adalah untuk melakukan analisa sentimen dari komentar masyarakat terhadap Presiden Jokowi.

Analisis sentimen merupakan sebuah metode yang digunakan untuk mengekstrak data opini, memahami serta mengolah tekstual data secara otomatis untuk melihat sentimen yang terkandung dalam sebuah opini [5]. Algoritma yang digunakan berupa *Naïve Bayes Classifier* adalah algoritma yang sering digunakan untuk mengklasifikasikan data dengan menggunakan perhitungan pada probabilitas. Dari perhitungan probabilitas pada setiap kata dalam dokumen akan dihasilkan klasifikasi kategori positif, negatif, dan netral yang diperoleh dari proses perhitungan sebelumnya yaitu pembobotan *tf-idf* [6].

2. METODE PENELITIAN

Berdasarkan Gambar 1, tahapan pertama dalam penelitian ini yaitu *crawling* data lalu dilakukan manual labelling positif, negatif atau netral. Tahap selanjutnya dilakukan *preprocessing* setelah itu data tersebut dibagi menjadi data latih dan data uji kemudian dilakukan klasifikasi *Naïve Bayes*. Selanjutnya ada tahap pengujian yang dilakukan menggunakan *confusion matrix* dan *k-fold cross validation* dan tahap terakhir yaitu menampilkan hasil dari metode tersebut.



Gambar 1. Tahapan Metode

2.1 *Crawling* Data

Crawling data pada penelitian ini dilakukan menggunakan *Apify*. *Apify* adalah aplikasi *webscraping* canggih yang memungkinkan pengambilan data dari *website* dengan mudah. Data yang digunakan adalah data yang berasal dari media sosial Facebook dengan *hashtag* tertentu yang diinput.

2.2 Manual Labelling

Pada tahap ini dapat dilihat pada Tabel 1, komentar akan diberi label nilai. Data diberi label -1 untuk nilai negatif, 0 untuk nilai netral, 1 untuk nilai positif. Setelah itu data yang telah diberi label tersebut divalidasi oleh pakar.

Tabel 1. Proses *Manual Labelling*

No	Komentar	Label
1	Pemimpin itu memang sangat menentukan kemajuan/kemunduran suatu negara... Di tangan pemimpin yg tepat negara menjadi lebih cepat maju dan sejahtera rakyatnya... #sudah banyak contohnya dan sejarah panjang Indonesia telah membuktikannya# #Jokowi pemimpin visioner#	1
2	Presiden Joko Widodo (Jokowi) akan langsung meninjau sejumlah jalan di Provinsi Lampung, Jumat 5 Mei 2023. Hal ini disampaikan Jokowi usai mengunjungi Pasar Natar di Lampung Selatan, Jumat pagi. #jokowi #lampung #jalanrusak	0
3	Jalan di Lampung yang akan Ditinjau Jokowi Rusak Lagi Usai 2 Hari Diperbaiki Baru dua hari diperbaiki secara kilat oleh Pemprov Lampung, Jalan Lintas Simpang Randu-Rumbia di Lampung Tengah, Provinsi Lampung, kembali rusak. Lapisan aspal di jalan tersebut mengelupas dan kembali berlubang. Hujan yang turun pada Kamis (4/5/2023) malam diduga jadi sebab jalan berlubang dan bertekstur seperti lumpur. #lampung #jokowi #jalanlintas	0
4	Fakta berbicara, Jokowi masih Presiden RI Terbaik sepanjang sejarah. Pahami ! Kini Duplikat nya Ganjar Pranowo.... #mpgsumut #ganjarpranowo #presiden #relawan #politik #masyarakatpendukungganjar	1
5	Bikin Malu Negara !!! Jokowi Marah Robohkan Pos Lintas Batas Negara Yang Gak Keurus Di Jaman Sby #jokowi #buktikerjanyaajokowi	-1

2.3 Preprocessing

Preprocessing Text merupakan tahapan awal dari *text mining* yang meliputi proses untuk mempersiapkan data tekstual yang akan digunakan agar dapat diproses pada tahapan berikutnya [7]. *Preprocessing* melibatkan serangkaian proses untuk menghilangkan kata-kata atau komentar yang tidak penting dan kata-kata tanpa arti yang ditemukan dalam data dari pengambilan atau *crawling* data di Facebook. Proses ini membutuhkan beberapa langkah yang harus diikuti dengan urutan berikut:

2.3.1 Case Folding

Merupakan tahap perubahan huruf dari huruf kapital menjadi huruf kecil [8]. Dengan tujuan untuk mempermudah komparasi atau perbandingan teks. Pada tahap ini akan ada modifikasi terhadap teks huruf kapital menjadi huruf kecil (*lowercase*). Tanpa *case folding*, kedua *string* akan dikenali sebagai *string* yang berbeda. Namun setelah dilakukan *case folding*, kedua *string* akan dianggap sama karena keduanya sudah dikonversi menjadi huruf kecil atau besar dengan cara yang sama.

2.3.2 Stopword Removal

Yaitu menghilangkan kata umum yang sering muncul tetapi tidak memiliki arti penting dan tidak digunakan [9]. Tujuannya agar mempercepat pemrosesan teks dengan mengurangi jumlah kata yang harus diproses. Terdapat dua metode yang dapat digunakan pada tahap ini. Pertama ada *Wordlist* yaitu dengan cara menyiapkan kumpulan kata penting dan hanya kata yang termasuk ke dalam *wordlist* yang akan digunakan pada proses selanjutnya. Adapun yang digunakan pada penelitian ini adalah metode *Stoplist*, yaitu dengan cara menyiapkan kumpulan kata yang tidak deskriptif (*stoplist*) dan yang termasuk ke dalam kumpulan itu akan dibuang dan tidak digunakan pada proses selanjutnya.

2.3.3 Stemming

Dilakukan untuk mengutrangi variasi kata yang berbeda dalam dokumen yang sama sehingga mempermudah pengolahan dokumen dengan cara membuang imbuhan yang melekat pada kata tertentu agar menjadi kata dasar.

2.3.4 Tokenizing

Adalah proses pemecahan teks dari sebuah kalimat menjadi potongan-potongan [10]. *Tokenization* bertujuan untuk memudahkan pengolahan dan analisis teks dengan mengurangi kompleksitas dan meningkatkan efisiensi pemrosesan data. Pada proses ini juga dilakukan penghilangan angka, tanda baca, dan karakter lain yang dianggap tidak memiliki pengaruh terhadap pemrosesan teks.

2.4 Pembagian Data

Pembagian data dilakukan dengan rasio 9:1 untuk setiap data latih dan data uji. Data latih digunakan untuk melatih algoritma dalam menemukan model yang sesuai maka dari itu biasanya rasio atau jumlah data latih lebih banyak daripada data uji. Sedangkan data uji merupakan data baru yang belum memiliki kelas sehingga diperlukan proses klasifikasi untuk menentukan kelas yang sesuai [11].

2.5 Naïve Bayes

NBC adalah salah satu metode untuk mengklasifikasi data dengan probabilitas sederhana yang mengaplikasikan teorema bayes dengan karakter independen yang tinggi [2]. Metode ini sering digunakan dalam tugas yang melibatkan *machine learning* dan *natural language processing*, termasuk klasifikasi teks, pemfilteran spam, analisis sentimen, dan kategorisasi dokumen. Teorema Bayes, yang mengkuantifikasi kemungkinan hipotesis (label kelas) yang diberikan bukti (fitur), berfungsi sebagai premis dasar *Naive Bayes*. Mengingat label kelas, asumsi "*naive*" adalah bahwa fitur-fitur tersebut secara kondisional tidak bergantung satu sama lain. *Naive Bayes* adalah metode komputasi yang murah karena premis penyederhanaannya, yang membuat komputasi menjadi efisien.

Dengan menggunakan *dataset* training berlabel, pengklasifikasi dilatih dengan memperkirakan probabilitas setiap kelas dan probabilitas bersyarat dari fitur yang diberikan label kelas. Rumus perhitungan *Naive Bayes* yang dapat dilihat pada persamaan (1), digunakan untuk mengklasifikasikan data yang tidak berlabel dengan menghitung kemungkinan setiap label kelas, dan kemudian mengklasifikasikan data ke label kelas dengan kemungkinan tertinggi. Klasifikasi *naive bayes* akan menghasilkan nilai probabilitas pada setiap kategori yaitu "positif", "negatif" dan "netral".

$$P(A|B) = \frac{(P(B|A) \times P(A))}{P(B)} \quad (1)$$

Setelah data sudah bersih dilakukan tahap pembobotan atau *term frequency* yaitu dihitung jumlah kemunculan kata yang terdapat dalam sebuah *post Facebook* pada data latih (*training*). Hasil dari pembobotan akan tersimpan pada sebuah *variable*. Agar lebih mudah dipahami dapat dilihat contoh pada Tabel 2 diambil contoh berupa 3 dokumen yang sudah melalui tahap *preprocessing*.

Tabel 2. *Term Frequency*

Data Bersih	Dokumen
presiden ri baik panjang sejarah	1
bikin malu negara jokowi marah roboh pos lintas batas negara	2
presiden joko widodo jokowi langsung tinjau jumlah jalan provinsi lampung	3

Contoh terdapat 3 dokumen, pada Tabel 3, 4 dan 5 dilakukan perhitungan kemunculan kata pada setiap dokumen.

Tabel 3. TF Dokumen 1

<i>Term</i>	<i>Count</i>
presiden	1
ri	1
baik	1
panjang	1
sejarah	1

Tabel 4. TF Dokumen 2

<i>Term</i>	<i>Count</i>
bikin	1
malu	1
negara	2
jokowi	1
marah	1
robok	1
pos	1
lintas	1
batas	1

Tabel 5. TF Dokumen 3

<i>Term</i>	<i>Count</i>
presiden	1
joko	1
widodo	1
jokowi	1
langsung	1
tinjau	1
jumlah	1
jalan	1
provinsi	1
lampung	1

2.6 Pengujian

Pengujian dalam penelitian ini dilakukan menggunakan *confusion matrix* dan *k-fold cross validation* dengan 10 fold. *Cross-validation* (CV) adalah pendekatan statistik untuk menilai keefektifan model atau algoritma di mana data dibagi menjadi dua subset: data proses pembelajaran dan data validasi atau evaluasi. Dalam mengevaluasi kinerja model, *k-fold cross validation* digunakan untuk mengestimasi kesalahan prediksi. Data dipisahkan menjadi k bagian dengan ukuran yang kira-kira sama. Model kategorisasi dilatih dan dievaluasi k kali. Salah satu subset akan digunakan sebagai data pelatihan dan pengujian di setiap iterasi. CV 10 fold membagi data menjadi 10 folds dengan ukuran yang kira-kira sama, menghasilkan 10 himpunan bagian data untuk menilai keefektifan model atau algoritme.

CV akan menggunakan sembilan *fold* untuk pelatihan dan satu *fold* untuk pengujian untuk masing-masing dari sepuluh pengelompokan data. Pada setiap *fold* dilakukan perhitungan nilai akurasi, presisi dan *recall* menggunakan *confusion matrix* dengan rumus perhitungan yang dapat dilihat pada persamaan (2) untuk akurasi, persamaan (3) untuk presisi dan persamaan (4) untuk *recall*. *Confusion Matrix* biasanya berbentuk tabel yang menampilkan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah. Untuk menghitung nilai akurasi, precision dan *recall* perlu diketahui jumlah TP, TN, FP dan FN. Empat kondisi tersebut dapat dijelaskan sebagai berikut:

- True Positive* (TP), keadaan dimana jumlah dari satu kelas *true* yang dapat diprediksi dengan benar pada kelas *true*.
- True Negative* (TN), adalah jumlah dari satu kelas *false* yang dapat diprediksi dengan benar pada kelas *false*.
- False Positive* (FP), keadaan dimana jumlah kelas *true* yang prediksinya salah pada kelas *false*.
- False Negative* (FN), keadaan pada jumlah kelas *true* yang diprediksi salah pada kelas *true*.

$$Akurasi = \frac{TP + TN}{Total} \quad (2)$$

(3)

$$Presisi = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

3. HASIL DAN PEMBAHASAN

3.1 Tahap *Crawling*

Tahapan pertama yang dilakukan adalah tahap pengumpulan data. Dalam penelitian ini, sumber data yang digunakan berasal dari social media Facebook. Kata kunci yang digunakan untuk mengumpulkan data yang akan dikumpulkan adalah #presiden dan #jokowi yang dipost pada Mei 2023. Proses *crawling* data *post* Facebook menggunakan sebuah *website* Apify. Total *post* Facebook yang didapatkan berjumlah 300 data. Kemudian data yang relevan akan disortir dan dikumpulkan menjadi sebuah *dataset* yang berisi data yang sesuai

3.2 Tahap *Manual Labelling*

Tahapan selanjutnya adalah tahap *labeling* yaitu, tahap melabelkan data yang dilakukan secara manual. Data diberi label oleh peneliti dan kemudian divalidasi oleh pakar. Data diberi label pada excel, -1 untuk *negative*, 0 untuk netral dan 1 untuk positif. Setelah itu data yang sudah diberi label tersebut divalidasi oleh pakar komunikasi politik, Dr. Bayquni, M.Pd.

3.3 Tahap *Preprocessing*

Selanjutnya dilakukan sejumlah tahapan *preprocessing* yang agar data yang akan lebih mudah untuk diolah. Dapat dilihat pada Tabel 6 beberapa *preprocessing* yang dilakukan adalah *Case Folding* yaitu agar semua huruf menjadi *lowercase*, kemudian *Stopword* untuk menghilangkan kata-kata yang sering muncul tetapi tidak berbobot, kemudian *Stemming* dilakukan untuk merubah kata menjadi kata baku dan *Tokenizing* dilakukan untuk memisahkan teks menjadi sebuah 'token'.

Tabel 6. *Preprocessing*

Komentar	Proses
Ayo Viralkan Jalan Jalan di Mamasa yang belum diperbaiki siapa tau di Kunjungi Jokowi. Terutama Messawa sampe Sibanawa. Itu udah bertahun tahun gak diperbaiki #jokowi #presiden	Komen Awal
ayo viralkan jalan jalan di mamasa yang belum diperbaiki siapa tau di kunjungi jokowi. terutama messawa sampe sibanawa. itu udah bertahun tahun gak diperbaiki #jokowi #presiden	<i>Case Folding</i>
ayo viralkan jalan jalan mamasa diperbaiki siapa tau kunjungi jokowi. terutama messawa sampe sibanawa. udah bertahun tahun gak diperbaiki #jokowi #presiden	<i>Stopword Removal</i>
ayo viralkan jalan jalan mamasa baik siapa tau kunjung jokowi utama messawa sampe sibanawa udah tahun tahun gak baik jokowi presiden	<i>Stemming</i>
ayo, viralkan, jalan, jalan, mamasa, baik, siapa, tau, kunjung, jokowi, utama, messawa, sampe, sibanawa, udah, tahun, tahun, gak, baik, jokowi, presiden	<i>Tokenization</i>

3.4 Tahap *Naïve Bayes*

Tahap pertama yang dilakukan adalah menghitung probabilitas masing-masing kategori terhadap keseluruhan data latih. Untuk menjelaskan proses lebih detail, dapat dilihat pada perhitungan persamaan (5) dan (6) di bawah ini.

$$P(\text{Positif}) = \frac{\text{Jumlah kategori Positif}}{\text{Jumlah Data Latih}} \quad (5)$$

(6)

$$P(\text{Positif}) = \frac{61}{300} = 0,2$$

Pertama yang dilakukan adalah perhitungan probabilitas untuk kategori ‘Positif’ dengan menggunakan rumus perhitungan pada persamaan (5). Berdasarkan rumus perhitungan pada persamaan (6) yang telah dihitung, probabilitas untuk kategori ‘Positif’ bernilai 0,2.

$$P(\text{Netral}) = \frac{\text{Jumlah kategori Netral}}{\text{Jumlah Data Latih}} \quad (7)$$

$$P(\text{Netral}) = \frac{224}{300} = 0,75 \quad (8)$$

Kemudian dilakukan perhitungan probabilitas untuk kategori ‘Netral’ dengan menggunakan rumus perhitungan pada persamaan (7). Probabilitas untuk kategori ‘Netral’ yang diperoleh bernilai 0,75 berdasarkan rumus perhitungan pada persamaan (8).

$$P(\text{Negatif}) = \frac{\text{Jumlah kategori Negatif}}{\text{Jumlah Data Latih}} \quad (9)$$

$$P(\text{Negatif}) = \frac{15}{300} = 0,05 \quad (10)$$

Terakhir, dilakukan perhitungan probabilitas untuk kategori ‘Negatif’ dengan rumus perhitungan pada persamaan (9). Probabilitas untuk kategori ‘Negatif’ yang telah dihitung menghasilkan nilai 0,05 berdasarkan hasil dari rumus perhitungan pada persamaan (10). Hal ini mengarah pada kesimpulan bahwa kelima data tersebut masuk dalam kategori “positif”, “negatif”, dan “netral”. Berdasarkan perhitungan ini, probabilitas "Positif" menghasilkan nilai 0,2 dan "Netral" menghasilkan nilai 0,75 dan 0,2 untuk kategori "Negatif" dan dapat disimpulkan bahwa sentiment masyarakat terhadap Presiden Jokowi bernilai netral.

3.5 Tahap Pengujian

Pada penelitian ini, pengujian data latih menggunakan data yang sudah memiliki label dari manual *labeling* dan menggunakan nilai $k = 10$. Metode *10-fold cross validation* ini berarti dibentuk 10 himpunan (*subsets*) dari *dataset* yang ada, 9 himpunan sebagai data *training* dan 1 himpunan sisa digunakan sebagai data *testing* dan dilakukan sebanyak 10 kali iterasi. Pada setiap iterasi dilakukan *confusion matrix* untuk mendapatkan nilai akurasi, presisi dan *recall*. Nilai akurasi didapatkan dengan cara menghitung semua prediksi yang benar dari data yang tersedia. Presisi adalah perhitungan dari prediksi sebagai kelas positif yang sebenarnya positif. *Recall* adalah nilai dari prediksi kelas positif yang diprediksi negatif.

Setelah nilai akurasi, presisi dan *recall* didapatkan dari setiap iterasi, nilai yang didapatkan dibagi sejumlah iterasi yang dilakukan, yaitu 10. Nilai rata-rata dari 10 iterasi yang telah dilakukan menghasilkan Akurasi sebesar 87%, Presisi 77% dan *Recall* 73%.

4. KESIMPULAN

Berdasarkan hasil evaluasi dari aplikasi sentimen *post* Facebook, dapat diambil kesimpulan bahwa system dapat menganalisa sentimen masyarakat yang bernilai positif, negatif dan netral menggunakan metode *Naïve Bayes*, hasil pengujian yang dilakukan menggunakan sebanyak 300 data sampel dengan rasio data 9:1 untuk setiap data latih dan data uji menghasilkan rata-rata nilai akurasi sebesar 87%, rata-rata nilai presisi sebesar 77% dan rata-rata nilai *recall* sebesar 73% dan juga dapat disimpulkan bahwa masyarakat memiliki sentiment netral terhadap Presiden Jokowi. Dengan hasil tersebut dapat disimpulkan bahwa algoritma *Naïve Bayes* cukup baik untuk diimplementasikan pada analisis sentiment. Nilai akurasi dapat ditingkatkan lagi dengan menambah tahapan proses pada preprocessing dan menambah jumlah *dataset* yang digunakan dalam penelitian.

DAFTAR PUSTAKA

- [1] (2023), Data Reportal. [Online]. Available: <https://datareportal.com/reports/digital-2023-indonesia>
- [2] D. Normawati and S. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter", *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, vol. 5, no. 2, pp. 697-711, 2021.
- [3] A. Salam, J. Zeniarja, and R. Khasanah, "ANALISIS SENTIMEN DATA KOMENTAR SOSIAL MEDIA FACEBOOK DENGAN K-NEAREST NEIGHBOR (STUDI KASUS PADA AKUN JASA EKSPEDISI BARANG J&T EKSPRESS INDONESIA)", *SINTAK*, vol. 2, Nov. 2018.
- [4] Luthfanida, "ANALISIS SENTIMEN DATA TWITTER MENGGUNAKAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE (SVM) TENTANG PRESIDEN JOKOWI 3 PERIODE", *Djtechno*, vol. 3, no. 1, pp. 5-11, July 2022.
- [5] F. Sari and A. Wibowo, "Analsis Sentimen Pelanggan Toko Online JD.ID Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi", *Jurnal SIMETRIS*, vol. 10, no. 2, pp. 681-686, Nov. 2019.
- [6] N. Siregar, R. Siregar, and M. Sudirman, "Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ)", *teknologia*, vol. 3, no. 1, Oct. 2020.
- [7] S. Arafah and Fathoni, "Sentiment Analysis Pada Masyarakat Terhadap LRT Kota Palembang Menggunakan Metode Improved K-Nearest Neighbor", *Jurnal Media Informatika Budidarma*, vol. 6, no. 3, pp. 1554-1561, July 2022.
- [8] I. Mawanta, T. S. Gunawan and Wanayumini, "Uji Kemiripan Kalimat Judul Tugas Akhir dengan Metode Cosine Similarity dan Pembobotan TF-IDF", *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, pp. 726-738, April 2021.
- [9] A. Firdaus, "Aplikasi Algoritma K-Nearest Neighbor pada Analisis Sentimen Omicron Covid-19", *Jurnal Riset Statistika (JRS)*, vol. 2, no. 2, pp. 85-92, Dec. 2022
- [10] M. I. Fikri, T. S. Sabrila and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter", *SMATIKA*, vol. 10, no. 2, pp. 71-76, 2020.
- [11] N. Ritha, N. Hayaty, T. Matulatan, A. Uperiati, M. Rathomi, M. Bettiza, F. Farasalsabila, "Sentiment Analysis of Helath Protocol Policy Using K-Nearest Neighbor and Cosine Similarity", *ICSEDIT*, 2022.