

# Klasifikasi Sentimen Masyarakat Terhadap Prabowo Subianto Bakal Calon Presiden 2024 di Twitter Menggunakan Naïve Bayes Classifier

Raja Zaidan Putera Dwitama\*, Yusra, Muhammad Fikry, Febi Yanto, Elvia Budianita

Fakultas Sains dan Teknolgi, Prodi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: <sup>1,\*</sup>11950115176@students.uin-suska.ac.id, <sup>2</sup>Yusra@uin-suska.ac.id, <sup>3</sup>muhammad.fikry@uin-suska.ac.id,

<sup>4</sup>febiyanto@uin-suska.ac.id, <sup>5</sup>elvia.budianita@uin-suska.ac.id

Email Penulis Korespondensi: 11950115176@students.uin-suska.ac.id

Submitted: 04/12/2023; Accepted: 28/12/2023; Published: 29/12/2023

**Abstrak**—Presiden Indonesia yang sudah menjabat 2 priode berturut-turut tidak dapat mencalonkan kembali menjadi Presiden. Sikap publik terhadap tiga bakal calon presiden, Prabowo Subianto, Anies Baswedan, dan Ganjar Pranowo yang digadag-dagang bakal mencalonkan diri di Pilpres 2024 juga menjadi bahan opini warganet yang bisa ditarik kesimpulan. Pengujian akan dilakukan dalam penelitian ini menggunakan informasi yang dikumpulkan dari tweet yang diposting oleh pengguna Twitter. Naïve Bayes Classifier adalah teknik yang akan diterapkan untuk penilaian sentimen. Dalam pemilihan presiden mendatang, penelitian ini sebagai sumber ketika menentukan pilihan presiden. 2100 tweet dengan kata kunci pencarian "Calon Presiden" dan "Prabowo Subianto" merupakan data yang terkumpul dengan pembagian 1050 data positif dan 1050 data negatif. Kemudian dilakukan implementasi menggunakan Google Colab dimulai dari proses data (cleaning, case folding, tokenizing, normalisasi, negation handling, stopword removal, stemming) dilanjutkan pengklasifikasian menggunakan Naïve Bayes Classifier. Menurut temuan pengujian menggunakan Confusion Matrix dengan tiga data uji percobaan 90:10, 80:20 dan 70:30. Mendapatkan hasil akurasi tertinggi 89%, dengan nilai precision 89,7%, 88,6% recall dan 88,9% f1-score pada uji percobaan 90:10.

**Kata Kunci:** Calon Presiden; Klasifikasi Sentimen; Naïve Bayes Classifier; Prabowo Subianto; Twitter

**Abstract**—The Indonesian President who has served for 2 consecutive terms cannot nominate again to become President. The public's attitude towards the three presidential candidates, Prabowo Subianto, Anies Baswedan, and Ganjar Pranowo, who are predicted to run for the 2024 presidential election, is also a matter for netizens' opinions from which conclusions can be drawn. Testing will be carried out in this research using information collected from tweets posted by Twitter users. Naïve Bayes Classifier is a technique that will be applied for sentiment assessment. In the upcoming presidential election, this research will be a source when determining the presidential choice. 2100 tweets with the search keywords "Presidential Candidate" and "Prabowo Subianto" are data collected by dividing 1050 positive data and 1050 negative data. Then implementation was carried out using Google Colab starting from data processing (cleaning, case folding, tokenizing, normalization, negation handling, stopword removal, stemming) followed by classification using the Naïve Bayes Classifier. According to test findings using the Confusion Matrix with three experimental test data 90:10, 80:20 and 70:30. Obtained the highest accuracy results of 89%, with a precision value of 89.7%, 88.6% recall and 88.9% f1-score in the 90:10 trial test.

**Keywords:** Presidential Candidates; Sentiment Classification; Naïve Bayes Classifier; Prabowo Subianto; Twitter

## 1. PENDAHULUAN

Ketika Pemilihan Umum Presiden semakin dekat, setiap tim calon presiden dan wakil presiden bekerja keras untuk menerapkan berbagai strategi. Mereka terlibat dalam pembentukan tim pemenangan, melakukan kampanye langsung di berbagai lokasi, serta mengunjungi wilayah-wilayah yang berbeda. Di samping itu, mereka juga memberikan bantuan sembako kepada warga yang memerlukan, sebagai bagian dari upaya mereka untuk mendapatkan dukungan dari masyarakat [1].

Pemilihan umum Presiden dan Wakil Presiden diadakan setelah masa jabatan mereka berakhir sesuai dengan ketentuan Pasal 7 Konstitusi Republik Indonesia Tahun 1945. Pasal ini menetapkan bahwa "Presiden dan Wakil Presiden menjabat selama lima tahun dan tidak diizinkan mencalonkan diri kembali setelah menjabat selama dua periode atau sepuluh tahun". Di samping itu, pemilihan juga dapat terjadi dalam keadaan khusus sebagaimana dijelaskan dalam Pasal 8 Konstitusi Republik Indonesia Tahun 1945. Oleh karena itu, peran MPR adalah untuk memilih Presiden atau Wakil Presiden sesuai dengan aturan konstitusi yang berlaku [2].

Dalam sebuah negara yang menjunjung tinggi prinsip demokrasi, penyelenggaraan pemilihan umum menjadi salah satu unsur krusial dalam mewujudkan sistem pemerintahan yang demokratis. Pada Pemilihan Umum 2024, pemilihan legislatif dan pemilihan presiden dijadwalkan untuk dilaksanakan secara bersamaan. Keputusan ini didasarkan pada keputusan lanjutan Mahkamah Konstitusi melalui Putusan Nomor 16/PUU11/2013 mengenai pemilihan umum serentak. Langkah ini dimaksudkan untuk mengurangi beban keuangan negara dalam pelaksanaan pemilu dan menyederhanakan struktur kerja pemerintah. Di Indonesia, aspek demokrasi tercermin dalam pelaksanaan pemilihan umum (pemilu) yang diadakan secara berkala setiap lima tahun, untuk menentukan Presiden dan anggota legislatif. Akan tetapi, data yang dirilis oleh KPU menunjukkan bahwa tingkat Partisipasi Pemilih (TPP) di Indonesia cenderung menurun pada setiap periode pemilu setelah era reformasi [3]. Sebelum masa reformasi, pada pemilihan umum 1997 yang merupakan pemilu terakhir pada era Orde Baru, tingkat partisipasi pemilih mencapai 93,6%. Namun, pada pemilihan umum legislatif tahun 2009, terjadi penurunan dalam tingkat partisipasi pemilih menjadi 70,9%, sementara partisipasi pemilih pada Pemilihan Presiden 2009 mencapai

71,7%. Pada pemilihan umum legislatif 2014, tingkat partisipasi pemilih mencapai 72%. Sedangkan pada Pemilihan Presiden 2014, partisipasi pemilih tercatat sebesar 69,58% [4].

Bersamaan dengan pertumbuhan penggunaan media sosial yang pesat di Indonesia, jumlah pengguna Twitter juga mengalami pertumbuhan yang signifikan setiap tahunnya. Pada tahun 2019, jumlah pengguna Twitter di Indonesia mencapai 6,43 juta individu, yang mewakili sekitar 52% dari total pengguna media sosial. Twitter telah menjadi platform media sosial yang sangat populer digunakan oleh berbagai instansi pemerintah, sektor swasta, dan masyarakat sipil [5]. Pada awal tahun 2017, jumlah pengguna Twitter secara global mencapai 328 juta, menunjukkan peningkatan sekitar 14% jika dibandingkan dengan tahun sebelumnya pada periode yang sama. Data tersebut mengungkapkan bahwa 77% pengguna Twitter di Indonesia aktif dan turut berpartisipasi dalam memanfaatkan platform ini. Pada tahun 2016, sekitar 4,1 miliar cuitan dihasilkan, menjadikan Indonesia sebagai salah satu negara dengan jumlah pengguna Twitter terbesar di dunia.

Pendekatan yang diterapkan adalah Metode Klasifikasi Naive Bayes yang menawarkan sejumlah keunggulan, termasuk kemudahan penggunaan, kecepatan eksekusi, dan tingkat akurasi yang tinggi. Metode Klasifikasi Naive Bayes ini dipakai untuk mengelompokkan atau mengklasifikasikan teks berdasarkan kata-kata yang muncul dalam sebuah dokumen. Algoritma klasifikasi Naive Bayes bergantung pada teori probabilitas yang dikembangkan oleh ilmuwan Inggris Thomas Bayes, yang memungkinkan prediksi probabilitas di masa depan berdasarkan pengalaman di masa lampau [6].

Oleh karena itu, dalam menghadapi banyaknya opini yang disuarakan di platform Twitter, penting untuk mengklasifikasikan sentimen tersebut guna memudahkan pemahaman terhadap kecenderungan opini terhadap film, apakah bersifat positif atau negatif. Penelitian ini memanfaatkan Algoritma Naive Bayes. Berdasarkan hasil eksperimen, semakin tinggi tingkat akurasi menunjukkan bahwa sistem berhasil melakukan klasifikasi secara efektif. Kesimpulannya, Naive Bayes dan Support Vector Machine (SVM) dengan pembobotan menggunakan TF-IDF digunakan dalam penelitian ini. Perbandingan kinerja keduanya menunjukkan bahwa Naive Bayes menghasilkan akurasi lebih tinggi, yaitu sebesar 82%. Hasil analisis sentimen menunjukkan bahwa metode SVM dapat digunakan untuk menganalisis sentimen dengan akurasi sekitar 78.3% secara keseluruhan [7].

Untuk melakukan penelitian menggunakan metode Naive Bayes pada berbagai topik, dilakukan pengumpulan data dari hasil berbagai penelitian sebelumnya. Dengan menggunakan kata kunci "Calon Presiden" dan "Prabowo Subianto", berhasil terkumpul sebanyak 2100 cuitan publik yang akan menjadi objek penelitian. Tahap-tahap penting seperti pengumpulan data, pelabelan, preprocessing, klasifikasi sentimen, pembobotan TF-IDF, dan pengujian Confusion Matrix menjadi fokus utama dari penelitian ini. Tujuan penelitian ini adalah memberikan pemahaman tentang pandangan masyarakat terhadap Prabowo Subianto, sebagai calon presiden potensial pada tahun 2024, berdasarkan aktivitas yang tercatat di Twitter.

## 2. METODOLOGI PENELITIAN

### 2.1 Pengumpulan Data

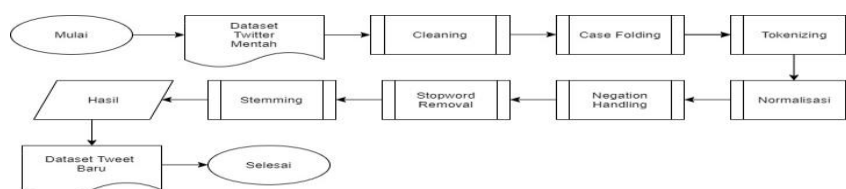
Data didapatkan melalui sebuah komentar masyarakat di Twitter menjadi sumber data penelitian ini. sesuai dengan kata kunci "Calon Presiden" dan "Prabowo Subianto". Mulai 1 Januari 2022, periode Crawling Data berlangsung hingga 1 Januari 2023. Untuk menghasilkan hasil terbaik, studi ini menyeimbangkan data yang digunakannya, yaitu dengan menggunakan 2100 tweet sebagai datasetnya untuk dianalisis. Hasil klasifikasi baik kelas sedikit maupun banyak sangat dipengaruhi oleh keseimbangan pengumpulan data [8].

### 2.2 Pelabelan Data

Metode ini digunakan dalam mengklasifikasikan kata ke dalam kategori yang cocok berdasarkan konten informasi. Label data dalam penelitian ini mengacu pada kategorisasi data dengan kelompok data negatif dan data positif.

### 2.3 Tahapan Text Preprocessing

Dataset masih termasuk dalam kategori data tidak terstruktur setelah pengumpulan data dan pelabelan data selesai. Dataset terlebih dahulu melalui tahap text preprocessing untuk menghapus data yang tidak akurat, memastikan hasil seakurat mungkin [9]. Langkah-langkah yang terlibat dalam preprocessing teks meliputi pembersihan, huruf besar-kecil, tanda baca dalam kalimat tidak memiliki pengaruh pada text preprocessing. Menghapus tanda baca serta emoji yang sering terdapat pada data, tokenisasi, normalisasi, negation handling, stopwords removal, dan steaming. Gambar 1 menunjukkan tahapan text preprocessing.



**Gambar 1.** Tahapan Text Preprocessing

## 2.4 Pembobotan TF IDF

Langkah ini dilakukan guna mengevaluasi tingkat kepentingan suatu kata yang terdapat dalam dokumen. TF atau Term Frequency mengindikasikan bahwa kata yang semakin sering muncul dalam dokumen akan menjadikan bobotnya menjadi semakin tinggi. Di sisi lain, IDF atau Inverse Document Frequency memiliki fungsi yang berkebalikan dari TF yakni bobotnya akan semakin rendah apabila kata tersebut semakin sering muncul [10]. Adapun perhitungan TF-IDF yang dipakai dalam penelitian ini adalah:

$$W_{i,j} = \frac{n_{i,j}}{\sum_{j=1}^p n_{j,i}} \log_2 \frac{D}{d_j} \quad (1)$$

Keterangan :

$W_{i,j}$  = pembobotan tf-idf untuk term ke-j pada dokumen ke-i

$n_{i,j}$  = jumlah kemunculan term ke-j pada dokumen ke-i

$P$  = banyaknya term yang terbentuk

$\sum_{j=1}^p n_{j,i}$  = jumlah kemunculan seluruh term pada dokumen ke-i

$D$  = jumlah keseluruhan dokumen

$d_j$  = banyaknya dokumen yang mengandung term ke-j.

## 2.5 Seleksi Fitur

Setiap fitur yang dipertimbangkan akan dipilih menggunakan proses pemilihan fitur yang sesuai dengan menghilangkan fitur-fitur yang berlebihan atau tidak relevan selama analisis data. Dalam metode ini, ambang batas digunakan untuk menentukan bobot minimum suatu fitur harus dianggap relevan dan dipertahankan. Dengan menerapkan threshold, fitur dengan bobot atau nilai lebih rendah dari batas tersebut akan dihapus dari kumpulan data, sehingga hanya fitur yang dianggap penting saja yang dipertahankan.

## 2.6 Naïve Bayes Classifier

Metode ini merupakan metode klasifikasi yang didasarkan pada teorema Bayesian dan probabilitas yang mengasumsikan variabel  $X$  memiliki sifat bebas atau independence, artinya suatu variabel (atribut) tidak memiliki kaitan dengan atribut lainnya. Meskipun menggunakan asumsi keindependenan atribut, metode ini tetap mempunyai performa yang kompetitif dalam proses pengklasifikasian [11]. Perbedaan utama antara metode ini dengan metode lainnya terdapat pada pembentukan hipotesis. Hipotesis pada metode ini dibentuk secara langsung melalui perhitungan terhadap frekuensi kata yang muncul pada data latih tanpa adanya proses pencarian. Sementara itu, pada metode lain biasanya dilakukan pencarian hipotesis yang sesuai dalam ruang hipotesis [12]. Bayesian classification yang didasarkan pada teorema bayes yang memiliki kemampuan klasifikasi dengan akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar.

Berikut adalah rumusan permasalahan dalam Naïve Bayes Classifier :

$$P(v)_j = \frac{\text{docs}_j}{\text{examples}} \quad (2)$$

$$P(w_k | v_j) = \frac{1+n_k}{n+|\text{kata}|} \quad (3)$$

Keterangan:

$\text{docs}_j$  = kumpulan dokumen dengan nilai target  $v_j$

$\text{examples}$  = jumlah dokumen yang digunakan dalam pelatihan (data latih)

$n$  = total jumlah kata dalam data tekstual dengan nilai target yang sesuai

$n_k$  = jumlah kemunculan kata  $w_k$  dalam semua data tekstual dengan nilai target yang sesuai

$|\text{kata}|$  = jumlah kata yang berbeda yang muncul dalam semua data tekstual yang digunakan

Untuk menjelaskan metode Naïve Bayes Classifier, proses klasifikasi memerlukan sejumlah penunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis.

$$V_{MAP} = \underset{v_j \in V}{\text{argmax}} P(V_j) \prod_i P(X_i | V_j) \quad (4)$$

Keterangan :

$P(V_j)$  : Peluang kemunculan suatu dokumen yang memiliki kategori  $j$

$P(X_i | V_j)$  : Peluang kemunculan  $X_i$  pada kategori  $V_j$

Pada penelitian ini menerapkan model Multinomial Naïve Bayes Classifier karena model tersebut unggul dalam pengklasifikasian teks atau dokumen. Dokumen tidak hanya ditentukan dengan kata yang muncul, tetapi juga jumlah kemunculannya [13].

## 2.7 Pengujian Confusion Matrix

Pengujian dengan Confusion Matrix dilakukan untuk melihat hasil akurasi dari metode yang digunakan yaitu Naïve Bayes Classifier dalam mengklasifikasi. Setelah selesai dilakukan proses klasifikasi, pengujian Confusion

Matrix dapat digunakan untuk menampilkan visualisasi performansi metode Naïve Bayes Classifier dalam melakukan klasifikasi [14]. Tabel 1 merupakan evaluasi dari Confusion Matrix.

**Tabel 1.** Confusion Matrix

Kelas	Prediksi	
	POSITIF	NEGATIF
POSITIF	TP	TN
NEGATIF	FP	FN

Keterangan:

TP (True Positive) = jumlah dokumen dari kelas 1 yang benar diklasifikasikan sebagai kelas 1

TN (True Negative) = jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0

FP (False Positive) = jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1

FN (False Negative) = jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0

Rumus Confusion Matrix untuk menghitung accuracy, precision, Recall dan F1-score seperti berikut :

$$Accuracy = \frac{TP+TN}{Total} \tag{5}$$

$$Precision = \frac{TP}{TP+FP} \tag{6}$$

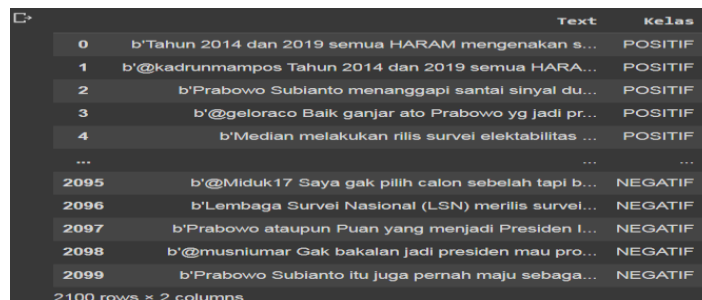
$$Recall = \frac{TP}{TP+FN} \tag{7}$$

$$F1score = \frac{2 \times P \times R}{P+R} \tag{8}$$

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pengumpulan Data

Dalam penelitian ini, data dikumpulkan untuk menganalisis sentimen data Prabowo Subianto, calon presiden 2024. Dengan menggunakan metode Crawling Twitter, topik dan persepsi publik tentang Prabowo Subianto sebagai calon presiden di media sosial menjadi satu-satunya yang digunakan dalam proses pendataan. Menggunakan fitur yang ditawarkan oleh antarmuka aplikasi (API) di aplikasi Twitter. Hingga 2100 data dikumpulkan, dengan 1050 di antaranya positif dan 1050 di antaranya negatif [15]. Data tentang load Google Colab ditunjukkan pada Gambar 2.



	Text	Kelas
0	b'Tahun 2014 dan 2019 semua HARAM mengenakan s...	POSITIF
1	b'@kadrunmampos Tahun 2014 dan 2019 semua HARA...	POSITIF
2	b'Prabowo Subianto menanggapi santai sinyal du...	POSITIF
3	b'@geloraco Baik ganjar ato Prabowo yg jadi pr...	POSITIF
4	b'Median melakukan rilis survei elektabilitas ...	POSITIF
...	...	...
2095	b'@Miduk17 Saya gak pilih calon sebelah tapi b...	NEGATIF
2096	b'Lembaga Survei Nasional (LSN) merilis survei...	NEGATIF
2097	b'Prabowo ataupun Puan yang menjadi Presiden l...	NEGATIF
2098	b'@musnumar Gak bakalan jadi presiden mau pro...	NEGATIF
2099	b'Prabowo Subianto itu juga pernah maju sebaga...	NEGATIF

**Gambar 2.** Load Data pada Google Colab

#### 3.2 Pelabelan Data

Setelah informasi dikumpulkan, itu dibagi menjadi dua kategori, positif dan negatif. Seorang ahli, dosen Bahasa Indonesia dari Fakultas Sains dan Teknologi Roza Afifah S.Pd selaku dosen UIN Sultan Syarif Kasim Riau, secara manual melabeli barang-barang tersebut. Tabel untuk hasil positif dan negatif ditunjukkan pada Tabel 2.

**Tabel 2.** Tweet Positif dan tweet Negatif

tweet Positif	tweet Negatif
b'Pengamat politik sekaligus pendiri Lingkar Madani Ray Rangkuti menduga, Prabowo Subianto tak akan keberatan jika menjadi calon wakil presiden untuk Ganjar Pranowo pada Pemilu Presiden 2024. ^ #Nasional #Pipres2024	b"Dukungan @jokowi ke @prabowo sebagai calon Presiden 2024 sprt yg pernah diucapkan bbrp waktu yg lalu,tak ubahnya seperti sebuah 'prank'. Dia sebenarnya hanya memberikan dukungan kpd @ganjarpranowo, bukan@prabowo.\n\nSeperti itulah bapak prank nasional berperilaku.

tweet Positif	tweet Negatif
<p>b'<a href="https://t.co/Nr6zK0wLNy">https://t.co/Nr6zK0wLNy</a> Kalangan Milenial Kota Bogor Deklarasi Dukung Prabowo-Erick Thohir di Pilpres 2024: Sejumlah anak muda alias milenial di Kota Bogor mendeklarasikan dukungan kepada Prabowo Subianto dan Erick Thohir, sebagai calon presiden</p> <p>b'@prabowo Bp CALON PRESIDEN untuk menggantikan pak Jokowi 2024 insyaallah'</p>	<p>b'Kalau mantan pendukung pada kecewa dapat diprediksi Bapak Prabowo kalah lagi kalau maju sebagai Calon Presiden 2024'</p> <p>b'@KompasTV Iya kasihan pak prabowo blm pernah jd presiden, selama ini calon terus blm pernah menang. Netijen emang ndak ada empatinya. 2009 Mega-Pro kalah, 2014 Prabowo-hatta kalah, 2019 prabowo-sandi kalah jg, masa yg 2024 kalah lg..</p> <p>b'@Miduk17 ASAL JGN DIBALIK, PRABOWO CALON PRESIDEN DAN GANJAR CALON WAKIL PRESIDEN , ITU BLM TENTU MENANG...</p> <p>b'@prabowo Jangan mudah percaya pak, hati2 suara pemilih malah berkurang jika bapak calon presiden nanti'</p>
<p>b'Yusril Ihza Mahendra menyebut bahwa Prabowo Subianto berpotensi memenangkan Pilpres 2024 jika didampingi Joko Widodo sebagai Calon Wakil Presiden. Mending Prabowo <a href="https://t.co/Qpulte6Hxp">https://t.co/Qpulte6Hxp</a> dekade08'</p> <p>b'"Sejumlah anak muda alias milenial di Kota Bogor mendeklarasikan dukungan kepada Prabowo Subianto dan Erick Thohir, sebagai calon presiden dan calon wakil presiden di Pemilu 2024.  "\n#Prabowo\n#IndonesiaMenang\n#14BANTEN'</p>	

### 3.3 Text Preprocessing

Langkah-langkah yang terlibat dalam preprocessing teks adalah:

a. Cleaning

Pembersihan data adalah proses menyiapkan data untuk dianalisis dengan membuang informasi yang tidak akurat dan bahasa yang tidak pantas [16].

**Tabel 3.** Proses Cleaning

Data asli	Data setelah dicleaning
<p>b'<a href="https://t.co/Nr6zK0wLNy">https://t.co/Nr6zK0wLNy</a> Kalangan Milenial Kota Bogor Deklarasi Dukung Prabowo-Erick Thohir di Pilpres 2024: Sejumlah anak muda alias milenial di Kota Bogor mendeklarasikan dukungan kepada Prabowo Subianto dan Erick Thohir, sebagai calon presiden</p>	<p>Kalangan Milenial Kota Bogor Deklarasi Dukung PrabowoErick Thohir di Pilpres 2024 Sejumlah anak muda alias milenial di Kota Bogor mendeklarasikan dukungan kepada Prabowo Subianto dan Erick Thohir sebagai calon presiden</p>

b. Case Folding

Case Folding Case folding merupakan tahapan di mana teks diubah menjadi huruf kecil (lowercase) secara keseluruhan untuk menyamaratakan penggunaan huruf kapital. Tahapan ini dapat mempermudah proses analisis teks yang tidak tergantung pada perbedaan huruf besar dan kecil. Contoh dan penjelasan lebih lanjut tentang tahapan case folding dapat ditemukan pada referensi [17].

**Tabel 4.** Proses Case Folding

Data asli	Data setelah dicase folding
<p>prabowo Bp CALON PRESIDEN untuk menggantikan pak Jokowi 2024 insyaallah</p>	<p>prabowo bp calon presiden untuk menggantikan pak jokowi 2024 insyaallah</p>

c. Tokenizing

Tokenizing merupakan langkah pembagian teks pada dokumen menjadi token yang dimana text yang panjang diubah menjadi fragmen yang kecil[18].

**Tabel 5.** Proses Tokenizing

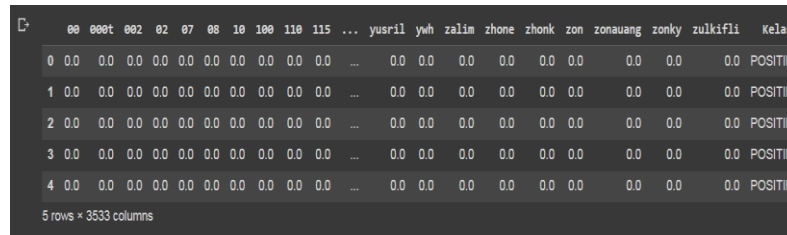
Data asli	Data setelah diTokenizing
<p>kalau mantan pendukung pada kecewa dapat diprediksi bapak prabowo kalah lagi kalau maju sebagai calon presiden 2024</p>	<p>['kalau', 'mantan', 'pendukung', 'pada', 'kecewa', 'dapat', 'diprediksi', 'bapak', 'prabowo', 'kalah', 'lagi', 'kalau', 'maju', 'sebagai', 'calon', 'presiden', '2024']</p>

d. Normalisasi

Normalisasi merupakan langkah mengubah bentuk penulisan masing-masing kata sesuai dengan KBBI(Kamus Besar Bahasa Indonesia) [19].



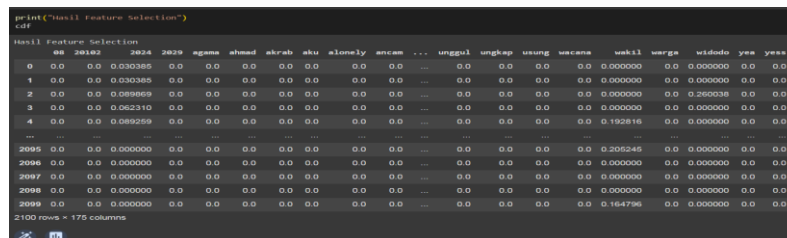
sering sebuah kata muncul dalam sebuah dokumen, digunakan istilah Frekuensi Kemunculan (Term Frequency, TF). Jika sebuah kata jarang muncul dalam berbagai dokumen, bobotnya akan lebih rendah. Setelah menggunakan metode Naive Bayes, langkah berikutnya adalah pemilihan fitur dan klasifikasi. Gambar 3 menampilkan hasil pembobotan TF-IDF untuk Google Colab.



**Gambar 4.** TF-IDF pada Google Colab

### 3.6 Seleksi Fitur

Dalam seleksi fitur, digunakan threshold (ambang batas) nilai untuk menentukan apakah suatu fitur harus dipertahankan atau dieliminasi. Threshold ini biasanya berada dalam rentang 0 hingga 1, dan dapat diatur sesuai dengan kebutuhan dan tujuan analisis. Setiap fitur setelah pembobotan akan diseleksi dengan proses pemilihan fitur yang relevan dengan penghapusan fitur tersebut yang redundan atau tidak relevan dalam analisis data. Dalam metode ini, digunakan sebuah threshold (batas) yang menentukan bobot minimum yang harus dimiliki oleh suatu fitur agar dianggap relevan dan dipertahankan. Menerapkan threshold, fitur-fitur yang memiliki bobot atau nilai di bawah batas tersebut akan dihapus dari dataset, sehingga hanya fitur-fitur yang dianggap signifikan yang tetap dipertahankan. Mengaplikasikan seleksi fitur pada google colab yang ditampilkan oleh gambar 5.



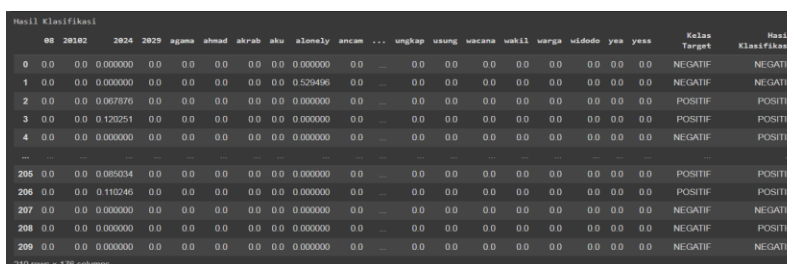
**Gambar 5.** Seleksi Fitur Threshold

### 3.7 Klasifikasi Naïve Bayes Classifier

Uji coba dilakukan untuk mengetahui hasil dari tiga uji coba data, yaitu dengan menggunakan data uji dari 90:10, 80:20 dan 70:30. Nilai akurasi tertinggi atau 89% ditunjukkan oleh 90:10 data uji saat Metode Naïve Bayes Classifier diuji. Hasil klasifikasi positif dan negatif pada tweet dikatakan “True” jika sistem berhasil mengidentifikasi dan mengelompokkan komentar ke dalam kelas data training. Hasil pengklasifikasian akan mendapatkan probabilitas sebagai berikut :

1. Precision atau perbandingan antara banyaknya data yang diprediksi positif dengan true positive (TP).
2. Recall atau perbandingan antara banyaknya data yang sebenarnya positif dengan true positif (TP).
3. F1-score atau perbandingan rata-rata antara recall dan precision.
4. Akurasi atau tingkat keakuratan model pada suatu data.
5. Confusion Matriks atau pengukuran performa untuk model learning dari klasifikasi naïve bayes.

Gambar 6 menunjukkan hasil klasifikasi Naïve Bayes Classifier pada Google Colab.



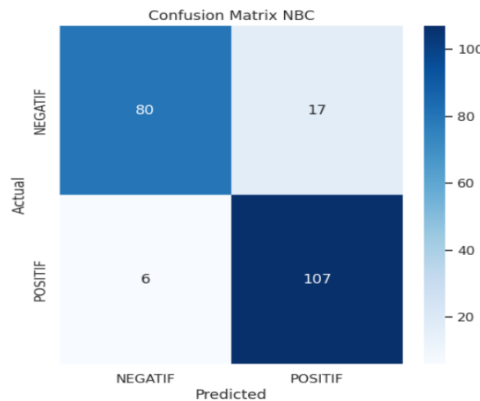
**Gambar 6.** Klasifikasi Naïve Bayes pada Google Colab

### 3.8 Pengujian Confussion Matrix

Untuk menentukan skor akurasi, Precision, Recall, dan F1-score, dilakukan pengujian Confussion Matrix setelah pembobotan kata, pemilihan fitur, dan klasifikasi menggunakan metode Naïve Bayes Classifier. Menggunakan Confussion Matrix untuk menentukan tes untuk data uji 90:10, 80:20 dan 70:30.

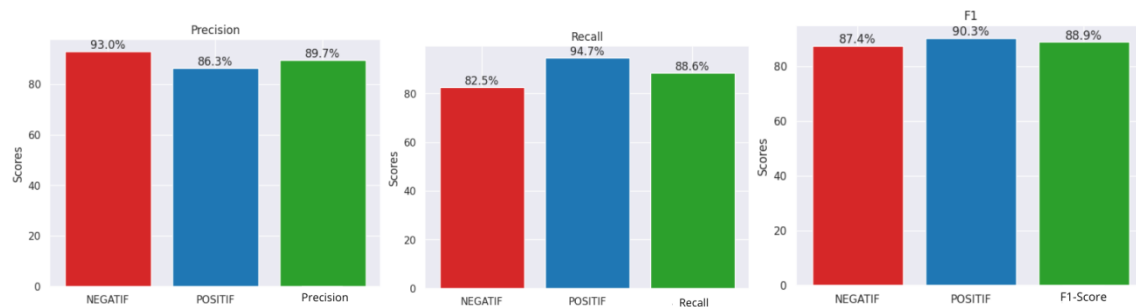
a. Data uji 90:10

Nilai akurasi terbesar 89% ditemukan setelah dilakukan pengujian terhadap 90:10 data uji. Google Colab telah menerapkan uji Confusion Matrix yang ditunjukkan pada Gambar 7.



**Gambar 7.** Confusion Matrix 90:10

Setelah memperoleh nilai akurasi True Negative (TN) 80, False Positive (FP) 17, False Negative (FN) 6, True Positive (TP) 107 dilakukan pengujian skor precision, recall, dan f1-score. Hasil pengujian precision, recall, dan f1-score pada data latih dan data uji dengan perbandingan 90:10 ditampilkan dalam Gambar 8.

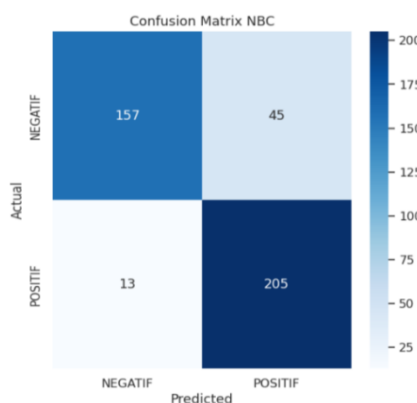


**Gambar 8.** Hasil Precision, Recall, dan F1-score 90:10

Berdasarkan Gambar 8, didapatkan hasil precision akurasi untuk tweet yang bersifat negatif sebesar 93%, sementara untuk tweet yang bersifat positif sebesar 86,3%, dengan nilai precision keseluruhan sebesar 89,7%. Hasil recall untuk tweet yang bersifat negatif adalah 82,5%, sementara untuk tweet yang bersifat positif adalah 94,7%, sehingga mendapatkan nilai recall sebesar 88,6%. Nilai f1-score yang diperoleh untuk tweet yang bersifat negatif adalah 87,4%, sedangkan untuk tweet yang bersifat positif adalah 90,3%, dengan nilai f1-score keseluruhan sebesar 88,9%.

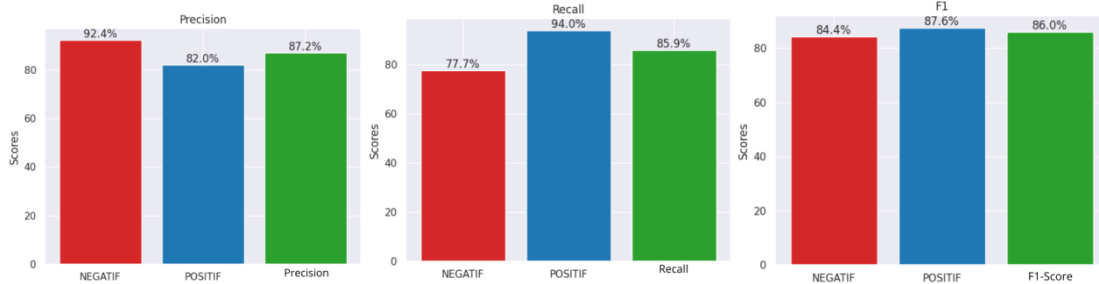
b. Data uji 80:20

Nilai akurasi 86% ditemukan setelah menguji 80:20 data uji. Google Colab telah menerapkan uji Confusion Matrix, yang ditunjukkan pada Gambar 9.



**Gambar 9.** Confusion Matrix 80:20

Setelah memperoleh nilai akurasi TN 157, FP 45, FN 13, TP 205 dilanjutkan dengan pengujian precision, recall, dan f1-score. Hasil pengujian precision, recall, dan f1-score dengan rasio data latih dan data uji 80:20 ditampilkan dalam Gambar 10.

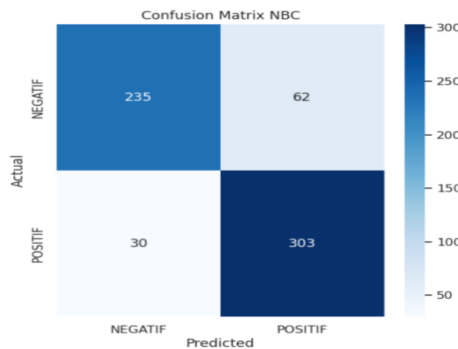


**Gambar 10.** Hasil precision, Recall, dan f1-score 80:20

Berdasarkan Gambar 10, diperoleh hasil precision akurasi untuk tweet yang bersifat negatif sebesar 92,4%, sementara untuk tweet yang bersifat positif sebesar 82%, dengan nilai precision keseluruhan sebesar 87,2%. Hasil recall untuk tweet yang bersifat negatif adalah 77,7%, sementara untuk tweet yang bersifat positif adalah 94%, sehingga mendapatkan nilai recall sebesar 85,9%. Nilai f1-score yang diperoleh untuk tweet yang bersifat negatif adalah 84,4%, sedangkan untuk tweet yang bersifat positif adalah 87,6%, dengan nilai f1-score keseluruhan sebesar 86%.

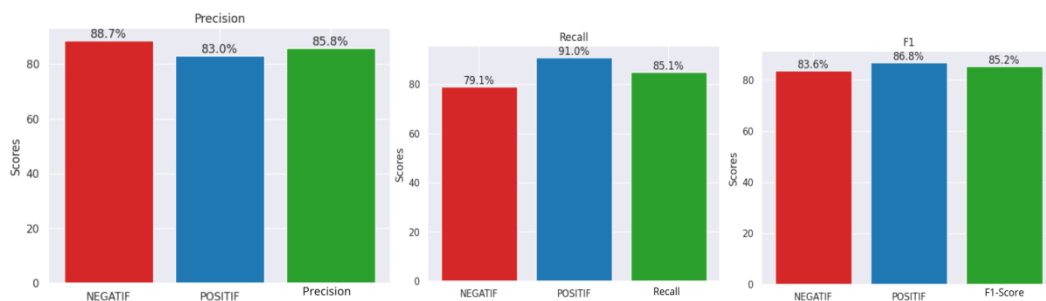
c. Data uji 70:30

Nilai akurasi sebesar 85,4% ditemukan setelah dilakukan pengujian terhadap 70:30 data uji. Confusion Matrix yang telah diimplementasikan pada Google Colab ditunjukkan pada Gambar 11.



**Gambar 11.** Confusion Matrix 70:30

Setelah memperoleh nilai akurasi TN 235, FP 62, FN 30, TP 303 dilakukan pengujian untuk precision, recall, dan f1-score. Hasil pengujian precision, recall, dan f1-score dengan rasio data latih dan data uji 70:30 ditampilkan dalam Gambar 12.



**Gambar 12.** Hasil Precision, Recall, dan F1-score 70:30

Berdasarkan Gambar 12, diperoleh hasil precision akurasi untuk tweet yang bersifat negatif sebesar 88,7%, sementara untuk tweet yang bersifat positif adalah 83%, dengan nilai precision keseluruhan sebesar 85,8%. Hasil recall untuk tweet yang bersifat negatif adalah 79,1%, sementara untuk tweet yang bersifat positif adalah 91%, dengan nilai recall sebesar 85,1%. Nilai f1-score yang diperoleh untuk tweet yang bersifat negatif adalah 83,6%, sedangkan untuk tweet yang bersifat positif adalah 86,8%, dengan nilai f1-score keseluruhan sebesar 85,2%.

#### 4. KESIMPULAN

Berdasarkan hasil pembahasan dari penelitian ini, dapat disimpulkan bahwa metode Naïve Bayes Classifier dapat digunakan untuk mengkategorikan opini terhadap Prabowo Subianto, calon presiden tahun 2024. Hal ini dilakukan melalui beberapa tahapan, seperti pengumpulan data, text preprocessing, pembobotan dengan TF-IDF, pemilihan

fitur dengan threshold, pengklasifikasian dengan metode Naïve Bayes Classifier serta pengujian dengan confusion matrix. Dengan perbandingan yang digunakan 90:10, 80:20, dan 70:30. Dalam penelitian ini menggunakan 2100 data yang telah di klasifikasikan menjadi 2 kelompok yaitu Positif dan Negatif dengan cara melabeli secara manual yang dibantu oleh dosen ahli Bahasa Indonesia. Pengujian 90:10 memiliki nilai akurasi tertinggi, 89%. Hasil dari Precision 89,7% , Recall 88,6%, dan F1-score 88,9%. Dari hasil metode klasifikasi yang digunakan, masyarakat akan mengetahui adanya sentimen terhadap Prabowo Subianto sebagai bakal calon Presiden Indonesia.

## REFERENCES

- [1] S. D. Agusti, S. Y. Dahda, N. Hidayah, and A. T. Tektonika, “Analisis Sentiment Isu Pilpres 2019 Dengan Menggunakan Algoritma Naive Bayes,” *J. Voice Informatics*, vol. 10, no. 1, pp. 1–10, 2021.
- [2] Nadir, W. Y. Wardani, and A. Purwandi, “Urgensi Pendidikan Politik Menyongsong Pemilihan Umum Presiden Dan Wakil Presiden Indonesia Tahun 2024 Untuk Menciptakan Pemilihan Umum Yang Jujur Dan Adil,” *J. Pendidik. dan Konseling*, vol. 4, no. 6, pp. 1349–1358, 2022.
- [3] A. Junaidi, “Peluang dan Tantangan Pesta Demokrasi 2024 dan Jaminan Perlindungan Sosial Ekonomi Lingkungan Hidup dicermati . Beberapa abad lalu , manusia belum memikirkan akibat dari pemanfaatan alam,” *J. Huk. Samudra Keadilan*, vol. 1, no. 1, pp. 23–30, 2024.
- [4] S. Juanita, “Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes,” *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 552, 2020, doi: 10.30865/mib.v4i3.2140.
- [5] N. A. Azmi, A. T. Fathani, D. P. Sadayi, I. Fitriani, and M. R. Adiyaksa, “Social Media Network Analysis (SNA): Identifikasi Komunikasi dan Penyebaran Informasi Melalui Media Sosial Twitter,” *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1422, 2021, doi: 10.30865/mib.v5i4.3257.
- [6] I. Riswanto and R. H. Laluma, “Klasifikasi Kelayakan Pinjaman Pada Koperasi Karyawan Menggunakan Metode Naïve Bayes Classifier Berbasis Web,” *Infotronik J. Teknol. Inf. dan Elektron.*, vol. 5, no. 1, pp. 11–16, 2020, doi: 10.32897/infotronik.2020.5.1.2.
- [7] A. Handayani et al., “Analisis Sentimen Terhadap Bakal Capres RI 2024 di Twitter Menggunakan Algoritma SVM,” *J. Inf. Syst. Res.*, vol. 5, no. 1, pp. 53–63, 2023, doi: 10.47065/josh.v5i1.4379.
- [8] R. Yanuarti, “Analisis Media Sosial Twitter Terhadap Topik Vaksinasi Covid-19,” *JUSTINDO (Jurnal Sist. dan Teknol. Inf. Indones.*, vol. 6, no. 2, pp. 121–130, 2021, doi: 10.32528/justindo.v6i2.5503.
- [9] S. Rahayu and J. J. Purnama, “Klasifikasi Konsumsi Energi Industri Baja Menggunakan Teknik Data Mining,” *J. Teknoinfo*, vol. 16, no. 2, p. 395, 2022, doi: 10.33365/jti.v16i2.1984.
- [10] A. Apriani, H. Zakiyudin, and K. Marzuki, “Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta,” *J. Bumigora Inf. Technol.*, vol. 3, no. 1, pp. 19–27, 2021, doi: 10.30812/bite.v3i1.1110.
- [11] M. K. Maulidina, “Analisis Sentimen Komentar Warganet Terhadap Postingan Instagram Menggunakan Metode Naive Bayes Classifier dan TF-IDF,” *Naskah Publ. Univ. Teknol. Yogyakarta*, pp. 1–15, 2020.
- [12] D. Sandi, “Klasifikasi Opini Dengan Menggunakan Algoritma K-Nearest Neighbor Pada Berita Vaksinasi Di Twitter,” *Nuansa Inform.*, vol. 16, no. 1, pp. 156–160, 2022, doi: 10.25134/nuansa.v16i1.5343.
- [13] A. H. Setianingrum, “Implementasi Algoritma Multinomial Naive Bayes Classifier,” *J. Tek. Inform.*, vol. 10, no. 2, pp. 109–118, 2018, doi 10.15408/jti.v10i2.6822..
- [14] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.
- [15] N. A. Rakhmawati, M. I. Aditama, R. I. Pratama, and K. H. U. Wiwaha, “Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin COVID-19,” *J. Inf. Eng. Educ. Technol.*, vol. 4, no. 2, pp. 90–92, 2020, doi: 10.26740/jieet.v4n2.p90-92.
- [16] A. J. P. Sibarani, “Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 2, pp. 262–276, 2020, doi: 10.35957/jatisi.v7i2.195.
- [17] L. Hermawan and M. B. Ismiati, “Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval,” *J. Transform.*, vol. 17, no. 2, p. 188, 2020, doi: 10.26623/transformatika.v17i2.1705.
- [18] B. H. Hayadi, “Sistem Informasi Geografis,” *Apl. dan Anal. Lit. Fasilkom UI*, vol. m, no. 1998, pp. 7–34, 2018.
- [19] N. M. A. J. Astari, D. G. H. Divayana, and G. Indrawan, “Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier,” *J. Sist. dan Inform.*, vol. 15, no. 1, pp. 27–29, 2020, doi: 10.30864/jsi.v15i1.332.
- [20] A. R. Satria, S. Adinugroho, and Suprpto, “Analisis Sentimen Ulasan Aplikasi Mobile menggunakan Algoritma Gabungan Naïve Bayes dan C4.5 berbasis Normalisasi Kata Levenshtein Distance,” vol. 4, no. 11, pp. 4154–4163, 2020.