

# ANALISIS SENTIMEN AKHIR MASA JABATAN PRESIDEN JOKOWI PADA MEDIA SOSIAL X MENGUNAKAN NAÏVE BAYES

Fadiah Nur Salsabilla<sup>1\*</sup>, Arita Witanti<sup>2</sup>

<sup>1,2</sup>Fakultas Teknologi Informasi, Informatika, Universitas Mercu Buana Yogyakarta,  
Kabupaten Bantul, Daerah Istimewa Yogyakarta, Indonesia  
E-mail: <sup>1\*</sup>fnsalsabill@gmail.com, <sup>2</sup>arita@mercubuana-yogya.ac.id  
(\* : corresponding author)

## Abstrak

Penelitian ini bertujuan untuk mengklasifikasikan sentimen terhadap Presiden Joko Widodo di Media Sosial X/Twitter pada masa akhir jabatannya. Metode yang digunakan adalah Complement Naïve Bayes (CNB) dengan penerapan SMOTE (*Synthetic Minority Over-sampling Technique*) untuk mengatasi masalah ketidakseimbangan data. Evaluasi dilakukan dengan dua variasi rasio data latih dan data uji, yaitu 90:10 dan 80:20. Pada rasio 90:10, model menunjukkan kinerja terbaik dengan mencapai 88% *accuracy*, *precision*, *recall*, dan *f1-score*. Namun, pada rasio 80:20, kinerja model mengalami penurunan dengan nilai 81% untuk semua metrik. Analisis sentimen menunjukkan bahwa sentimen negatif mendominasi, diikuti dengan sentimen netral dan positif, yang mencerminkan ketidakpuasan publik terhadap kebijakan-kebijakan tertentu pada periode akhir masa jabatan Presiden Jokowi.

**Kata kunci:** analisis sentimen, complement naïve bayes, SMOTE, x/twitter, akhir masa jabatan jokowi

## Abstract

*This study aims to classify sentiment towards President Joko Widodo on Social Media X/Twitter during the final period of his term. The method used is Complement Naïve Bayes (CNB) with the application of SMOTE (Synthetic Minority Over-sampling Technique) to address the issue of imbalanced data. Evaluation was conducted with two variations of the training and test data ratio, namely 90:10 and 80:20. In the 90:10 ratio, the model performed the best, achieving 88% accuracy, precision, recall, and F1-score. However, with the 80:20 ratio, the model's performance declined with 81% in all metrics. Sentiment analysis showed that negative sentiment predominated, followed by neutral and positive sentiment, reflecting public dissatisfaction with certain policies during the final period of President Jokowi's term.*

**Keywords:** sentiment analysis, complement naïve bayes, SMOTE, x/twitter, the end of jokowi's term

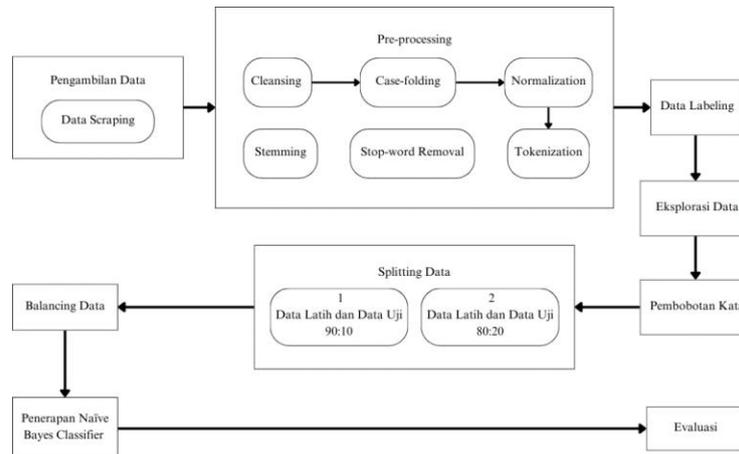
## 1. PENDAHULUAN

Perkembangan teknologi dan meningkatnya penggunaan media sosial telah menciptakan ruang baru untuk masyarakat mengekspresikan opini dan pandangan mereka terhadap berbagai isu sosial, ekonomi, dan politik [1]. Di Indonesia, media sosial seperti X (sebelumnya dikenal sebagai Twitter) menjadi platform yang sering digunakan masyarakat untuk menyuarakan pendapat, termasuk dalam menanggapi isu-isu politik dan kebijakan pemerintah. Salah satu topik yang mendapat perhatian luas adalah akhir masa jabatan Presiden Joko Widodo [2]. Di masa transisi kepemimpinan ini, opini masyarakat terpecah antara yang mendukung, netral, hingga yang mengkritik kinerja pemerintah di bawah kepemimpinan Presiden Joko Widodo.

Analisis sentimen, atau opinion mining, adalah metode yang digunakan untuk secara otomatis memahami, mengekstrak, dan mengolah data teks dari media sosial. Metode ini bertujuan untuk mengidentifikasi kecenderungan opini publik secara kuantitatif berdasarkan teks yang diunggah, sehingga dapat memberikan wawasan tentang sentimen yang terkandung dalam kalimat opini [3]. Dengan algoritma Naïve Bayes, yaitu metode klasifikasi probabilistik sederhana, yang bekerja dengan menghitung probabilitas kata atau frasa dalam teks berdasarkan frekuensi dan kombinasi nilai dari data yang ada [4]. Dalam analisis sentimen, metode ini membantu menentukan kecenderungan sentimen secara keseluruhan dan telah menjadi pendekatan yang populer dan efektif.

## 2. METODE PENELITIAN

Dalam menganalisis sentimen dan mengetahui akurasinya, penelitian yang dilakukan perlu melalui beberapa tahapan sesuai kaidah pengolahan data yang umum diterapkan pada penelitian-penelitian sebelumnya [5] [6]. Proses penelitian dimulai dari awal pengambilan data lalu pre-processing data, labeling data, eksplorasi data, pembobotan kata, *splitting* data, *balancing* data kemudian dilanjutkan dengan proses penerapan Naïve Bayes Classifier, dan evaluasi. Alur tahapan analisis yang dilakukan dapat dilihat pada Gambar 1.



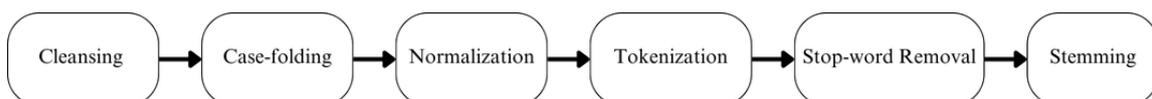
Gambar 1. Alur Analisis Sentimen

### 2.1. Data Scraping

*Data scraping* adalah metode yang digunakan untuk mengambil data dalam jumlah besar dari situs web dan menyimpannya dalam format file lokal atau *database* dalam bentuk tabel [7]. Pada penelitian ini, data dikumpulkan melalui pemanfaatan Application Programming Interface (API) X/Twitter dengan menggunakan *auth token* X/Twitter untuk mengakses tweet berdasarkan *keyword* tertentu lalu disimpan pada file dalam format Comma Separated Values(CSV). Proses pengumpulan data dilakukan dengan bantuan *source code* pihak ketiga yang digunakan secara eksklusif untuk keperluan edukasi [8]. Dikembangkan dengan *Python* dan *node.js*, dataset diperoleh berdasarkan *keyword* “akhir jokowi” dalam rentang waktu setelah pengumuman hasil pemilu pada 20 Maret 2024 hingga sehari sebelum pelantikan Presiden dan Wakil Presiden baru pada 19 Oktober 2024. Sejak bulan Juli tahun 2023, pemilik X/Twitter, Elon Musk, menerapkan batasan harian untuk pengambilan data pada akun yang belum diverifikasi, yaitu maksimal 600 *crawl* data *tweet* per-hari [9]. Oleh karena itu, pengambilan data dilakukan dalam beberapa hari dan didapatkan total 4.409 *tweet*.

### 2.2. Pre-processing

*Pre-processing* atau *text processing* adalah proses pengolahan teks yang bertujuan untuk mengurangi *noise* pada dataset serta mengubahnya menjadi format yang lebih terstruktur [10]. Tahapan pre-processing yang diterapkan seperti Gambar 2, yaitu *cleansing*, *case-folding*, *normalization*, *tokenization*, *stop-word removal*, dan *stemming*.



Gambar 2. Tahapan Pre-Processing Data

- a. *Cleansing*, merupakan tahap pembersihan data dengan cara menghapus data *error*, *null*, dan *duplicate*, serta karakter-karakter yang tidak relevan dalam analisis seperti tanda baca, angka,

- simbol, *link/URL*, tag pengguna (*@username*), dan tagar (*#hashtag*) [5]. Dalam penelitian ini tag pengguna @jokowi dikecualikan.
- b. *Case-folding* adalah proses pengubahan data menjadi huruf kecil (*lowercase*). Proses ini dilakukan untuk memastikan perbedaan kapitalisasi dalam data tidak memengaruhi analisis. [11]. Misalnya, “Jokowi” dan “jokowi” dianggap sama setelah proses *case-folding*.
  - c. *Normalization* adalah proses pengubahan kalimat tidak baku menjadi bentuk standar yang dapat diolah. Dalam konteks Bahasa Indonesia, proses normalisasi penting untuk mengubah bahasa informal atau *slang* menjadi bentuk yang lebih umum yang dapat membantu konsistensi data dan peningkatan akurasi analisis [11]. Pada penelitian ini, digunakan kamus buatan untuk melakukan proses normalisasi. Sebagai contoh, kata “gak” menjadi “tidak” atau “yg” menjadi “yang” atau “gue” menjadi “saya”, dan lain-lain.
  - d. *Tokenization* merupakan pengubahan data menjadi bentuk per kata atau frasa. Unit ini yang disebut token. Proses *tokenization* akan membantu model memahami struktur data serta analisis struktur dan pola menjadi lebih efisien.
  - e. *Stop-word removal* yaitu proses penghapusan kata-kata yang sering muncul dan umumnya tidak memberi makna penting dalam analisis [12], seperti kata “dan”, “yang”, “adalah”, “sana”, “sini”, “begini”, “begitu”, dan lain sebagainya. Pada penelitian ini digunakan kamus *stop-word* dari *library* Sastrawi, dengan tambahan beberapa *stop-word* buatan.
  - f. *Stemming* adalah proses pengubahan kata menjadi bentuk dasarnya untuk mengurangi variasi kata yang memiliki akar yang sama. Proses ini melibatkan penghapusan semua imbuhan, termasuk awalan (prefix), sisipan (infix), dan akhiran (suffix) [12]. *Library* Sastrawi menyediakan metode *stemming* dalam Bahasa Indonesia yang digunakan pada penelitian ini.

Setelah dataset melalui seluruh tahap *pre-processing*, maka dataset pada proses *stemming* akan digabungkan kembali menghasilkan dataset baru yang siap untuk diproses dalam model atau analisis lanjutan. Contoh *pre-processing* data ditunjukkan pada Table 1.

Tabel 1. Tahap *Pre-Processing*

Proses	Hasil
<b>Data Awal</b>	Emg blm sempurna tp setidaknya program-program Pak @jokowi berhasil bawa Indonesia jadi lebih maju dibandingkan 10 thn yg lalu. Terima kasih dan sehat-sehat sampai akhir masa jabatannya ya Pak #lebihbaikjokowi #indonesia #jokowi #indonesiamaju <a href="https://t.co/2WBroEW27g">https://t.co/2WBroEW27g</a>
<b>Cleansing</b>	Emg blm sempurna tp setidaknya program program Pak jokowi berhasil bawa Indonesia jadi lebih maju dibandingkan thn yg lalu Terima kasih dan sehat sehat sampai akhir masa jabatannya ya Pak
<b>Case-folding</b>	emg blm sempurna tp setidaknya program program pak jokowi berhasil bawa indonesia jadi lebih maju dibandingkan thn yg lalu terima kasih dan sehat sehat sampai akhir masa jabatannya ya pak
<b>Normalization</b>	memang belum sempurna tetapi setidaknya program program pak jokowi berhasil bawa indonesia jadi lebih maju dibandingkan tahun yang lalu terima kasih dan sehat sehat sampai akhir masa jabatannya ya pak
<b>Tokenization</b>	['memang', 'belum', 'sempurna', 'tetapi', 'setidaknya', 'program', 'program', 'pak', 'jokowi', 'berhasil', 'bawa', 'indonesia', 'jadi', 'lebih', 'maju', 'dibandingkan', 'tahun', 'yang', 'lalu', 'terima', 'kasih', 'dan', 'sehat', 'sehat', 'sampai', 'akhir', 'masa', 'jabatannya', 'ya', 'pak']
<b>Stop-word Removal</b>	['sempurna', 'program', 'program', 'pak', 'jokowi', 'berhasil', 'bawa', 'indonesia', 'lebih', 'maju', 'dibandingkan', 'tahun', 'lalu', 'terima', 'kasih', 'sehat', 'sehat', 'akhir', 'masa', 'jabatannya', 'pak']

Proses	Hasil
Stemming	['sempurna', 'program', 'program', 'pak', 'jokowi', 'hasil', 'bawa', 'indonesia', 'lebih', 'maju', 'banding', 'tahun', 'lalu', 'terima', 'kasih', 'sehat', 'sehat', 'akhir', 'masa', 'jabat', 'pak']
Data Akhir	sempurna program program pak jokowi hasil bawa indonesia lebih maju banding tahun lalu terima kasih sehat sehat akhir masa jabat pak

### 2.3. Data Labeling

Dataset awal yang terdiri dari 4.409 data melalui proses *pre-processing* untuk menghasilkan data bersih yang siap dianalisis, dan kemudian dihasilkan 3.901 data. Selanjutnya, proses *labeling* dilakukan menggunakan metode *keyword labeling*, yaitu dataset diberi label sentimen berdasarkan *keyword* tertentu di mana dataset dikategorikan menjadi tiga kelas sentimen: negatif, netral, dan positif. Langkah pertama yang harus dilakukan adalah menyusun daftar kata kunci (*keyword list*) yang terdiri dari kata-kata yang mengindikasikan sentimen negatif dan positif. Contoh, kata-kata seperti “serakah”, “bohong”, “busuk”, atau “rezim” termasuk dalam kategori negatif, sedangkan “baik”, “puas”, “hebat”, “dukung” termasuk dalam kategori positif. Setelah daftar kata kunci (*keyword list*) disusun, setiap data dalam dataset diperiksa untuk menemukan kata-kata dari daftar ini. Data yang mengandung kata-kata kunci negatif akan diberi label “negatif” dan proses ini diulang untuk kategori sentimen lainnya. Meskipun metode *keyword labeling* ini sangat menghemat waktu, metode ini memiliki kelemahan. Yaitu, data yang bersifat kompleks, seperti kalimat sarkasme atau ironi yang mengandung kata positif namun memiliki Makna negatif, sering kali sulit diklasifikasikan dengan benar hanya dengan menggunakan kata kunci (*keyword*).

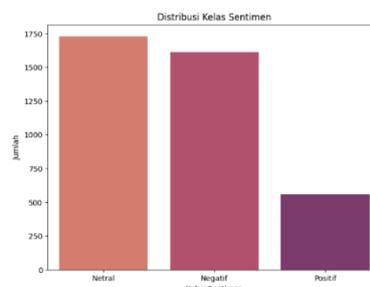
Proses labeling selesai, kemudian pemeriksaan ulang dilakukan secara manual. Hasil akhir dari proses ini didapatkan distribusi jumlah data di setiap kelas sentimen, dengan 1.612 data diberi label negatif, 1.730 data diberi label netral, dan 559 data diberi label positif seperti pada Table 2.

Tabel 2. Distribusi Jumlah Data Sentimen

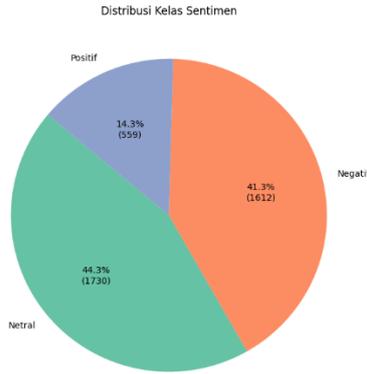
Kelas Sentimen	Jumlah Data
Negatif	1.620
Netral	1.730
Positif	559

### 2.4. Eksplorasi Data

Eksplorasi data bertujuan untuk mendapatkan gambaran awal tentang karakteristik data yang ada dalam setiap kelas sentimen melalui berbagai metode visualisasi. Pada penelitian ini eksplorasi data dilakukan untuk melihat komposisi sentimen, jumlah sentimen, *WordCloud*, dan 10 kata teratas pada dataset. Visualisasi ini membantu memahami pola bahasa dan distribusi sentimen yang dapat memengaruhi tahap analisis selanjutnya. Pada Gambar 3 dan 4 ditampilkan diagram batang dan diagram *pie* komposisi dan jumlah data pada setiap kelas sentimen.



Gambar 3. Diagram Batang Distribusi Data Sentimen



Gambar 4. Diagram Pie Distribusi Data Kelas Sentiment

Pada gambar 5, 6, dan 7 di bawah ini, *WordCloud* masing-masing kelas sentimen, dengan mengecualikan kata “pak”, “jokowi”, “akhir”, “masa”, “jabat”, “tidak”, dan “presiden”. Sedangkan pada gambar 8 ditunjukkan top 10 kata pada semua kelas sentimen.



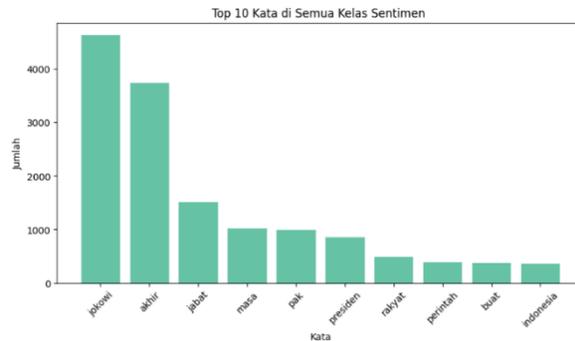
Gambar 5. World Cloud kelas sentimen positif



Gambar 6. World Cloud Kelas Sentimen Negatif



Gambar 7. World Cloud Kelas Sentimen Netral



Gambar 7. Top 10 Kata Di Semua Kelas Sentiment

## 2.5. Pembobotan Kata

Metode TF-IDF adalah metode ekstraksi fitur yang digunakan untuk merepresentasikan teks dokumen menjadi sekumpulan bobot tertentu. Bobot yang merepresentasikan sebuah dokumen dihitung berdasarkan frekuensi setiap kata yang muncul dalam dokumen, yang

kemudian dinormalisasi dengan frekuensi kata yang sama dalam kumpulan dokumen. Tujuan dari proses normalisasi rumus perhitungan bobot TF-IDF adalah untuk menentukan seberapa penting kata dalam dokumen. Dapat dilihat rumus pembobotan untuk setiap kata (term –  $t$ ) dalam dokumen  $d$  pada Persamaan 1 di bawah ini [13],

$$w_{t,d} = t f_{t,d} \times \log \left( \frac{N}{df_t} \right) \quad (1)$$

dengan,

$w_{t,d}$  = bobot TF-IDF

$t f_{t,d}$  = banyaknya  $t$  dalam dokumen  $d$

$N$  = frekuensi dokumen

$df_t$  = frekuensi dokumen yang mengandung kata  $t$

Dalam implementasinya, *TfidfVectorizer* dari *scikit-learn* di Python sering digunakan untuk menghitung bobot TF-IDF. *TfidfVectorizer* menyediakan berbagai opsi, seperti normalisasi dan pengaturan maksimum jumlah fitur, yang memungkinkan analisis lebih akurat dan relevan dengan data yang digunakan. Karena IDF mengurangi bobot kata-kata yang sering muncul di banyak dokumen, proses ekstraksi fitur dengan TF-IDF membantu mengurangi pengaruh kata-kata yang terlalu umum (seperti *stop-words*).

## 2.6. Splitting Data

Langkah penting dalam pengembangan dan evaluasi model adalah membagi dataset menjadi data latih (*train*) dan data uji (*test*) [14]. Pembagian ini memungkinkan model untuk belajar dari subset data (data latih) dan dievaluasi pada subset lain (data uji), memberikan perkiraan tentang kinerja model. Pada penelitian ini digunakan 2 variasi rasio data latih dan data uji yaitu 90:10 dan 80:20. Variasi ini membantu dalam menentukan keseimbangan antara data latih untuk pembelajaran yang lebih baik dan data uji untuk pengukuran performa yang akurat. Fungsi *train\_test\_split* dari *scikit-learn* digunakan dengan pengaturan *random\_state* untuk memastikan hasil yang konsisten.

## 2.7. Balancing Data

Merujuk pada Tabel 2, diketahui bahwa distribusi jumlah data pada kelas sentimen tidak seimbang (*imbalanced*). Ketidakseimbangan data terjadi ketika beberapa kelas sentimen kurang terwakili dalam data, menyebabkan bias prediksi yang signifikan terhadap kelas mayoritas. Meskipun akurasi model tampak tinggi karena dominasi kelas mayoritas, performa nilai prediktif pada kelas minoritas, cenderung rendah [15]. Salah satu cara untuk mengatasi ketidakseimbangan data adalah dengan menggunakan metode *sampling*. *Sampling* mengubah distribusi data di antara kelas mayoritas dan minoritas pada data latih untuk menyeimbangkan jumlah data masing-masing kelas. Salah satu metode *sampling* yang umum digunakan adalah SMOTE (*Synthetic Minority Over-sampling Technique*). Metode ini yang memperbanyak data sintetis dari replikasi data kelas minoritas untuk membuat dataset lebih seimbang [16].

## 2.8. Penerapan Naïve Bayes Classifier

Naïve Bayes adalah algoritma yang umum digunakan dalam pemrosesan bahasa alami (NLP) dan merupakan salah satu metode Machine Learning [17]. Metode ini didasari pada teori yang dikembangkan oleh seorang ilmuwan asal Inggris bernama Thomas Bayes. Bayes memperkenalkan Teorema Bayes sebagai cara untuk menggunakan pengalaman masa lalu untuk memprediksi kemungkinan di masa depan. Dalam konteks analisis sentimen, data masa lalu dianalogikan sebagai data latih (*train*), sementara data masa depan diibaratkan sebagai data uji (*test*) [18]. Pada penelitian ini digunakan varian Complement Naïve Bayes (CNB), yang merupakan adaptasi dari Multinomial Naïve Bayes dan dirancang untuk bekerja lebih baik pada

dataset yang tidak seimbang (*imbalanced*) [19]. Pada algoritma Complement Naïve Bayes (CNB) proses klasifikasi untuk kelas  $c$  pada sebuah dokumen  $d$  dijelaskan pada Persamaan 2 berikut ini.

$$CNB(d) = \operatorname{argmax}_{c \in C} [\log P(c) + \sum_{1 < k < n_d} \log P(x_k | c')] \quad (2)$$

dengan nilai  $P(x_k | c')$  merepresentasikan probabilitas kata ke- $k$  yang diketahui pada kelas bukan  $c$  [13].

## 2.9. Evaluasi

Data selanjutnya akan dievaluasi menggunakan *confusion matrix* untuk mengetahui tingkat keakuratan model dalam analisis data [20]. *Confusion matrix* merupakan representasi tabel jumlah data uji yang benar dan jumlah data uji yang salah dalam klasifikasi [21]. Matriks ini memberikan gambaran rinci tentang kinerja model dengan menunjukkan di mana model membuat klasifikasi yang benar dan di mana terjadi kesalahan, dapat dilihat pada Table 3 memungkinkan perbaikan yang lebih terarah di iterasi model berikutnya.

Table 3. Confusion matrix

		Kelas Prediksi	
		1	0
Kelas Sebenarnya	1	TP	FP
	0	FN	TN

Keterangan:

- TP (True Positive) = jumlah data dari kelas 1 benar diklasifikasikan sebagai kelas 1
- TN (True Negative) = jumlah data dari kelas 0 benar diklasifikasikan sebagai kelas 0
- FP (False Positive) = jumlah data dari kelas 1 salah diklasifikasikan sebagai kelas 0
- FN (False Negative) = jumlah data dari kelas 0 salah diklasifikasikan sebagai kelas 1

Metode evaluasi dalam data mining yang paling sering digunakan adalah mencari nilai *precision*, *recall*, *f1-score* dan *accuracy* [22]. Beberapa nilai yang dihitung pada tahap evaluasi tersebut antara lain adalah

- Precision* adalah hasil perbandingan prediksi benar positif dengan total data yang diprediksi positif, atau dirumuskan dalam Persamaan 3.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

- Recall* adalah hasil perbandingan prediksi benar positif dengan total data yang diprediksi benar positif dan data yang diprediksi salah negatif seperti pada Persamaan 4.

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

- F1-score* adalah metrik evaluasi yang menggabungkan *precision* dan *recall* untuk mengukur keberhasilan *retrieval*. Karena *f1-score* melibatkan informasi *False Positive* (FP) dan *False Negative* (FN), metrik ini cocok untuk diterapkan pada kasus data yang tidak seimbang (*imbalanced*) [23]. Dirumuskan pada Persamaan 5.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (5)$$

- Accuracy* merupakan hasil perbandingan antara data prediksi benar dengan total data. Persamaan 6 ditunjukkan perhitungan *accuracy*.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Dengan menggunakan metrik-metrik ini, kelebihan dan kekurangan model dalam mengklasifikasikan setiap kelas sentimen dapat dianalisis, terutama dalam memahami pola kesalahan klasifikasi yang terjadi.

### 3. HASIL DAN PEMBAHASAN

Berdasarkan hasil pengujian, model yang digunakan menunjukkan kinerja yang baik dalam mengklasifikasikan data sentimen publik terhadap Presiden Joko Widodo pada masa akhir jabatannya. Complement Naïve Bayes (CNB), yang digunakan bersama SMOTE (*Synthetic Minority Over-Sampling Technique*), berhasil memberikan performa yang cukup baik pada dua rasio data latih dan data uji yang diuji, yaitu 90:10 dan 80:20.

Pada rasio 90:10, model mencapai 88% untuk metrik *accuracy*, *precision*, *recall*, dan *f1-score*. Namun, saat rasio diubah menjadi 80:20, performa model mengalami penurunan menjadi 81% untuk semua metrik tersebut. Penurunan ini menunjukkan bahwa rasio data latih (*train*) yang lebih besar memberikan dampak besar pada kemampuan model dalam mengklasifikasikan data.

Tabel 4. Perbandingan Pengujian 2 Variasi Rasio

Rasio Data	Accuracy	Precision	Recall	F1-score
90 : 10	0.88	0.88	0.88	0.88
80 : 20	0.81	0.81	0.81	0.81

Detail dari masing-masing kelas sentimen dengan variasi rasio data latih dan data uji dapat dilihat pada Table 4 dan 5.

Tabel 5. *Classification Report* dengan Rasio Data 90:10

<i>Classification Report</i>	precision	recall	f1-score	support
<i>Negatif</i>	0.96	0.90	0.93	206
<i>Netral</i>	0.85	0.87	0.86	129
<i>Positif</i>	0.70	0.84	0.76	56
<i>accuracy</i>			0.88	391
<i>macro avg</i>	0.84	0.87	0.85	391
<i>weighted avg</i>	0.89	0.88	0.88	391

Tabel 6. *Classification Report* dengan Rasio data 80:20

<i>Classification Report</i>	precision	recall	f1-score	support
<i>Negatif</i>	0.88	0.88	0.88	363
<i>Netral</i>	0.81	0.78	0.79	298
<i>Positif</i>	0.64	0.70	0.67	120
<i>accuracy</i>			0.81	781
<i>macro avg</i>	0.78	0.78	0.78	781
<i>weighted avg</i>	0.81	0.81	0.81	781

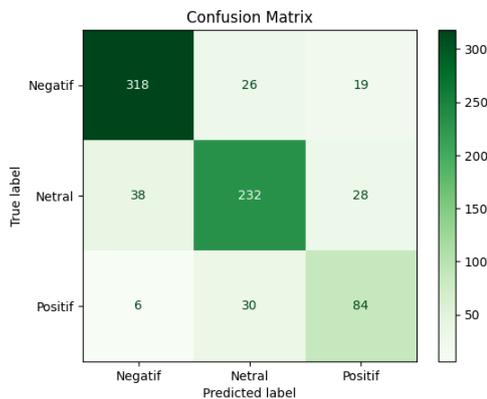
*Confusion matrix* pada variasi rasio data latih dan data uji dipresentasikan pada Tabel 4 dan 5.

Hasil *confusion matrix* menunjukkan distribusi prediksi model terhadap data uji. Pada rasio 90:10, model berhasil mengklasifikasikan mayoritas data negatif dengan benar. Namun, terdapat beberapa kesalahan prediksi pada kelas netral dan positif. Pada rasio 80:20, model menunjukkan lebih banyak kesalahan pada kelas netral dan positif, yang mengindikasikan sensitivitas model terhadap jumlah data latih yang lebih kecil.

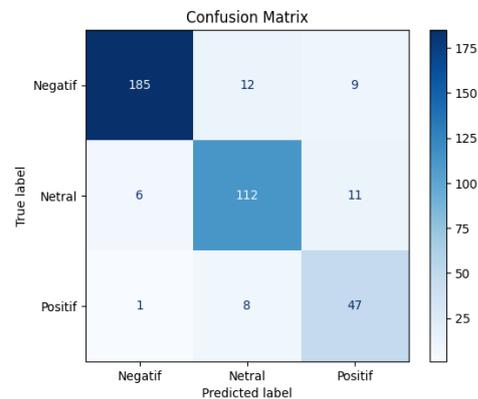
### 4. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengidentifikasi dan mengklasifikasikan sentimen publik terhadap Presiden Joko Widodo pada masa akhir jabatannya melalui analisis teks di media sosial X/Twitter. Complement Naïve Bayes (CNB) yang digunakan bersama dengan SMOTE (*Synthetic Minority Over-sampling Technique*) terbukti efektif, terutama dengan rasio data latih (*train*) dan data uji (*test*) 90:10, yang memberikan performa terbaik dengan 88% *accuracy*, *precision*, *recall*, dan *f1-score*. Namun, model mengalami penurunan performa pada rasio 80:20 menjadi 81%, yang

menunjukkan bahwa jumlah data latih memengaruhi kinerja model secara signifikan, yang terlihat pada Gambar 9 dan Gambar 10 berikut:



Gambar 8. Confusion Matrix Data Rasio 80:20



Gambar 10. Confusion Matrix Data Rasio 80:20

Hasil analisis sentimen menunjukkan bahwa sentimen negatif mendominasi percakapan publik, diikuti oleh sentimen netral dan positif. Dominasi sentimen negatif ini dapat mencerminkan adanya ketidakpuasan masyarakat terhadap beberapa kebijakan yang diambil pada periode akhir masa jabatan Presiden Joko Widodo. Namun, masih terdapat dukungan dalam bentuk sentimen positif dan pandangan netral.

#### DAFTAR PUSTAKA

- [1] R. N. Muhammad, L. W. S, and B. Tanggahma, "Pengaruh Media Sosial Pada Persepsi Publik Terhadap Sistem Peradilan: Analisis Sentimen di Twitter," *Unes LAW Review*, vol. 7, no. 1, pp. 507–516, 2024,.
- [2] I. Sumiarsa, *et al.*, "Evaluasi Kepemimpinan Presiden Joko Widodo Menjelang Akhir Masa Jabatan," *Ministrate: Jurnal Birokrasi & Pemerintah Daerah*, vol. 4, no. 1, pp. 1–23, 2023.
- [3] Y. S. Mahardika, and E. Zuliarso, "Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes Classifier," *Prosiding SINTAK 2018*, vol. 2, 2018, pp. 409–413.
- [4] E. Martantoh, and N. Yanih, "Implementasi Metode Naïve Bayes Untuk Klasifikasi Karakteristik Kepribadian Siswa Di Sekolah MTS Darussa'adah Menggunakan PHP Mysql," *Jurnal Teknologi Sistem Informasi*, vol. 3, no. 2, pp. 166–175, 2022.
- [5] H. Prasetyo, and A. S. Fitriani, "Sentiment Analysis Before Presidential Election 2024 Using Naïve Bayes Classifier Based On Public Opinion In Twitter," *Procedia of Engineering and Life Science*, vol. 4, pp. 1–15, 2023,
- [6] B. Wicaksono, and N. Cahyono, "Analisis Sentimen Komentar Instagram Pada Program Kampus Merdeka Dengan Algoritma Naive Bayes dan Decision Tree," *JATI (Jurnal Mahasiswa Teknologi Informasi)*, vol. 8, no. 2, pp. 2372–2381, 2024.
- [7] A. Z. Rizquina and C. I. Ratnasari, "Implementasi Web Scraping untuk Pengambilan Data Pada Website E-Commerce," *Jurnal Teknologi dan Sistem Informasi Bisnis*, vol. 5, no. 4, pp. 377–383, 2023.
- [8] H. Satria, "Cara Mendapatkan Data(Crawl) Twitter X-Maret 2024," *helmisatria.com*. Accessed: Oct. 20, 2024. [Online]. Available: <https://helmisatria.com/blog/updated-crawl-data-twitter-x-maret-2024/>
- [9] E. Musk, "Temporary Limits on Data Scraping and Post Reading on X/Twitter," X(Twitter). Accessed: Oct. 20, 2024. [Online]. Available: <https://x.com/elonmusk/status/1675187969420828672>
- [10] E. Y. Hidayat, R. W. Hardiansyah, and A. Affandy, "Analisis Sentimen Twitter untuk

- Menilai Opini Terhadap Perusahaan Publik Menggunakan Algoritma Deep Neural Network," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 2, pp. 108–118, 2021.
- [11] A. K. Sari, *et al.*, "Analisis Sentimen Twitter Menggunakan Machine Learning untuk Identifikasi Konten Negatif," *Adopsi Teknologi dan Sistem Informasi*, vol. 3, no. 1, pp. 64–73, 2024.
- [12] A. Q. Surbakti, R. Hayami, and J. A. Amien, "Analisa Tanggapan Terhadap Psbb Di Indonesia Dengan Algoritma Decision Tree Pada Twitter," *Jurnal Computer Science and Information Technology (CoSciTech)*, vol. 2, no. 2, pp. 91–97, 2021.
- [13] M. N. Randhika, J. C. Young, A. Suryadibrata, and H. Mandala, "Implementasi Algoritma Complement dan Multinomial Naïve Bayes Classifier Pada Klasifikasi Kategori Berita Media Online," *Ultimatics Jurnal Teknik Informatika*, vol. 13, no. 1, pp. 19–25, 2021.
- [14] R. Adinugroho, "Perbandingan Rasio Split Data Training dan Data Testing Menggunakan Metode Lstm Dalam Memprediksi Harga Indeks Saham Asia," 2022. [Online]. Available: [https://repository.uinjkt.ac.id/dspace/handle/123456789/67314%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/67314/1/RAHMADHAN ADINUGROHO-FST.pdf](https://repository.uinjkt.ac.id/dspace/handle/123456789/67314%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/67314/1/RAHMADHAN%20ADINUGROHO-FST.pdf)
- [15] M. P. Pulungan, A. Purnomo, and A. Kurniasih, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Kepribadian MBTI Menggunakan Naive Bayes Classifier," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 7, pp. 1493–1502, 2023.
- [16] E. Sutoyo and M. A. Fadlurrahman, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network," *Jurnal Edukasi dan Penelitian Informatika*, vol. 6, no. 3, pp. 379–385, 2020.
- [17] N. Khoirunnisaa, *et al.*, "Klasifikasi Teks Ulasan Aplikasi Netflix Pada Google Play Store Menggunakan Algoritma Naïve Bayes dan SVM," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 7, no. 1, pp. 64–73, 2024.
- [18] F. Nurwanda, and J. R. Rizkiani, "Perbandingan Metode Naive Bayes Classifier dan Support Vector Machine pada Analisis Sentimen Twitter Topik Lifestyle," *Jurnal Ilmu Wahana Pendidikan*, vol. 9, no. 21, pp. 314–323, 2023.
- [19] W. I. Sabilla and C. Bella Vista, "Implementasi SMOTE dan Under Sampling pada Imbalanced Dataset untuk Prediksi Kebangkrutan Perusahaan," *Jurnal Komputer Terapan*, vol. 7, no. 2, pp. 329–339, 2021.
- [20] A. Naufal, W. Zain, and A. Tholib, "Klasifikasi Data Mining di Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Sistem Informasi Fakultas Teknik Universitas Nurul Jadid," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 7, pp. 204–213, 2024.
- [21] D. Normawati, Prayogi Surya Allit, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *Jurnal Sains Komputer dan Informatika*, vol. 5, no. 2, pp. 697–711, 2021.
- [22] N. Arifin, U. Enri, and N. Sulistiyowati, "Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram Untuk Text Classification," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 6, no. 2, pp. 129–136, 2021.
- [23] E. Hokijuliandy, H. Napitupulu, and F. Firdaniza, "Analisis Sentimen Menggunakan Metode Klasifikasi Support Vector Machine (SVM) dan Seleksi Fitur Chi-Square," *SisInfo: Jurnal Sistem Informasi dan Informatika*, vol. 5, no. 2, pp. 40–49, 2023.