

Cek Artikel Hendra

by Mohammad hendra

Submission date: 30-Mar-2023 03:21PM (UTC+1100)

Submission ID: 2050670447

File name: Cek Artikel Hendra.pdf (464.74K)

Word count: 3771

Character count: 23201

Implementasi Metode Naïve Bayes untuk Analisis Sentimen pada Pemilihan Presiden 2024

Mohammad Hendra Ferdiansyah¹⁾, Mochammad Alfian Rosid²⁾, Yulian Findawati³⁾, Ade Eviyanti³⁾

^{1,2,3)} Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi: alfanrosid@umsida.ac.id

Abstract. Twitter is a social media platform that allows users to share photos and videos and has various features such as retweet, photo and video capture, and the ability to share content to other social networks. Opinions and comments from the public can easily be expressed through this application. The 2024 presidential election will be held next year, and topics and hashtags related to the election are already popular and being discussed on Twitter. Many tweets related to the topic are expressed and it is unknown whether they contain positive, negative or neutral sentiment. The purpose of this research is to determine public sentiment towards the 2024 presidential election in Indonesia and to determine the performance of the Naïve Bayes Classifier Algorithm. This research falls under the category of fine-grained sentiment analysis, where the focus is on analyzing sentiment at the sentence or comment level. The data used will be processed using text mining techniques, data obtained through twitter scraping will be classified into 3 classes, namely positive, negative, and neutral with manual labeling. Using the TF-IDF (Term Frequency Inverse Document Frequency) feature extraction method and performance calculation using the confusion matrix technique. The results of this study indicate that the 2024 presidential election in Indonesia has a greater positive sentiment than negative and neutral. Through several testing schemes, the best result was obtained by dividing the test data by 70% and the training data by 30%. The classification using the Naïve Bayes algorithm achieved an accuracy test of 83%.

Keywords :- naïve bayes, sentiment analysis, twitter

Abstrak. Twitter adalah sebuah platform media sosial yang memungkinkan pengguna untuk berbagi foto dan video serta memiliki berbagai fitur seperti retweet, pengambilan foto dan video, dan kemampuan untuk membagikan konten ke jaringan sosial lainnya. Cuitan berupa opini dan komentar dari masyarakat dengan mudah diutarakan melalui aplikasi ini. Pemilihan presiden 2024 akan dilaksanakan tahun depan, topik dan beberapa tagar tentang pilpres sudah ramai dijumpai dan menjadi perbincangan di beranda twitter. Banyak cuitan terkait topik tersebut diutarakan dan belum diketahui apakah mengandung sentimen positif, negatif atau netral. Tujuan penelitian ini adalah untuk mengetahui sentimen masyarakat terhadap pemilihan presiden 2024 di Indonesia dan untuk mengetahui kinerja Algoritma Naïve Bayes Classifier. Penelitian ini termasuk dalam kategori analisis sentimen berdiferensiasi halus (fine-grained sentiment analysis), di mana fokusnya adalah pada analisis sentimen pada level kalimat atau komentar. Data yang digunakan akan diproses menggunakan teknik text mining, data diperoleh melalui proses scraping twitter akan diklasifikasikan menjadi 3 kelas, yakni positif, negatif, dan netral dengan pelabelan secara manual. Menggunakan metode ekstraksi fitur TF-IDF (Term Frequency Inverse Document Frequency) dan perhitungan kinerja menggunakan teknik confusion matrix. Hasil penelitian ini menunjukkan bahwa pemilihan presiden 2024 di Indonesia memiliki Sentimen positif lebih besar daripada negatif dan netral. Melalui beberapa skema pengujian. Didapat hasil terbaik dengan melakukan pembagian data uji 70% dan data latih 30% Klasifikasi menggunakan algoritma Naïve Bayes mendapatkan uji akurasi sebesar 83%.

Kata Kunci :- naïve bayes, Analisis Sentimen, twitter

I. PENDAHULUAN

Media sosial di era modern ini memiliki hubungan erat dengan manusia. Selama beberapa dekade terakhir, pertumbuhan pengguna media sosial meningkat dengan cepat. Pada bulan Januari 2022, jumlah aktif pengguna media sosial di Indonesia telah mencapai 191 juta orang, menjadi meningkat 12,35% dari tahun sebelumnya yang hanya 170 juta orang pada bulan Januari 2021 (dataindonesia.id, 2022). Media sosial memiliki peranan penting dalam kehidupan bermasyarakat, mulai dari penyebaran informasi, komunikasi, pendidikan, ekonomi dan urusan politik. Dalam bidang politik memiliki dampak signifikan, mulai dari sarana membranding diri, mengembangkan pandangan politik di masyarakat, sampai media berkampanye. Kampanye berkedok deklarasi – deklarasi dari bakal calon kandidat yang akan maju untuk pemilihan presiden 2024 mulai ramai. survei elektabilitas juga mulai bermunculan terhadap bakal calon presiden. Hal ini terjadi karena peran media sosial yang semakin penting dalam kehidupan sehari-hari masyarakat.

Media sosial yang populer saat ini adalah Twitter, Instagram, dan TikTok. Twitter merupakan sebuah aplikasi yang memungkinkan pengguna untuk berbagi foto dan video, serta memiliki fitur retweet, pengambilan foto dan video, dan membagikannya ke beberapa jaringan sosial lainnya [1]. Menurut laporan dari We Are Social, jumlah pengguna Twitter yang berada di Indonesia mencapai 18,45 juta pada tahun 2022, yang menempatkan Indonesia sebagai negara

pengguna Twitter terbesar kelima di dunia. Opini dan komentar dari masyarakat dengan mudah diutarakan melalui aplikasi ini. Salah satu elemen konten pada media sosial adalah tagar (hashtag) [2]. fitur utama yang dimiliki oleh aplikasi Twitter. Hashtag memiliki peran besar dalam meramaikan percakapan dan menjadi trending topik media sosial.

Pelaksanaan Pemilu Presiden Indonesia direncanakan oleh Komisi Pemilihan Umum pada tanggal 21 Februari 2024 [3], beberapa tagar tentang pilpres sudah ramai dijumpai di beranda twitter. Tagar capres2024, pilpres2024 menjadi salah satu yang ramai diperbincangkan. Tagar ini dibuat untuk mengkampanyekan nama - nama bakal calon yang digadang akan maju pilpres 2024. Fenomena tersebut mengakibatkan banyaknya informasi, opini, dan berita yang beredar di media sosial. Ketika suatu opini publik diterima secara luas, dapat memengaruhi pandangan dan sikap seseorang terhadap suatu masalah atau isu tertentu.

Beberapa penelitian serupa yang dilakukan peneliti akan ditelusuri keutamaan dan kekurangan dalam penelitian tersebut. Pada penelitian yang berjudul "Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap "Figure Pemimpin" Menggunakan Python" dilakukan penelitian dengan 5 tahapan yaitu pembatasan masalah, pengambilan data (crawling) dengan API twitter, penyimpanan hasil berdasarkan hashtags atau isu, data diolah kembali untuk mencari kata positif, negative, dan netral, dan data dibuat presentase. Diklasifikasikan dengan metode naive bayes. Penelitian tersebut masih perlu ditranslate ke Bahasa Indonesia karena pelabelan textbolb hanya mengenai Bahasa Inggris [4].

Pada penelitian dengan judul "Analisis Sentimen terhadap Ulasan Paris Van Java Resort Lifestyle Place di Kota Bandung Menggunakan Algoritma KNN" dilakukan dengan pelabelan hanya positif negatif data yang didapat dari Google Review. Klasifikasi menggunakan Algoritma KNN dan diuji dengan menggunakan matriks konfusi. Hasil yang diperoleh masih rendah di kisaran 70 %. Menerapkan metode penghitungan jarak dapat berdampak pada akurasi algoritma K-Nearest Neighbor (KNN) dalam analisis sentimen terhadap ulasan Paris Van Java Resort Lifestyle Place di Kota Bandung [5].

Penelitian yang berjudul "Implementasi Twitter Sentiment Analysis untuk Review Film menggunakan Algoritma Support Vector Machine" dilakukan dengan mendaftar API Twitter untuk mendapatkan token terlebih dahulu. Penelitian ini menggunakan kata kunci berjumlah 30 dari 30 judul film populer tahun 2016. Objek yang diteliti adalah review film dan dilakukan pemberian label secara manual ke dalam tiga kelas yaitu positif, netral, dan negatif. Menggunakan algoritma Support Vector Machine (SVM) Hasil yang didapat pada penelitian ini adalah 76 % [6].

penelitian dengan judul "Sentimen Analisis Opini Masyarakat pada media Sosial Twitter terhadap Vaksin Berbayar menggunakan Metode Naive Bayes Classifier (NBC)" dilakukan dengan cara yang sama yakni diawali registrasi pada API twitter didapat dataset dengan jumlah 283 tweet dan objek yang diteliti vaksin berbayar. Klasifikasi model kata positif, netral, dan negatif. Menggunakan Metode Naive Bayes Classifier (NBC). Dari segi emosi, kata "vaksin berbayar" cenderung mendapatkan hasil yang lebih banyak bersifat negatif daripada positif atau netral. [7].

Penelitian yang berjudul "Analisis Sentimen Data Twitter Tentang Pasangan Capres-Cawapres Pemilu 2019 Dengan Metode Lexicon Based Dan Support Vector Machine" dimulai dengan pengambilan dataset oktober 2018 berjumlah 800 dataset, dalam proses pengklasifikasi kata harus dipasangkan sesuai dengan kelasnya. Metode yang digunakan adalah Lexicon Based [8].

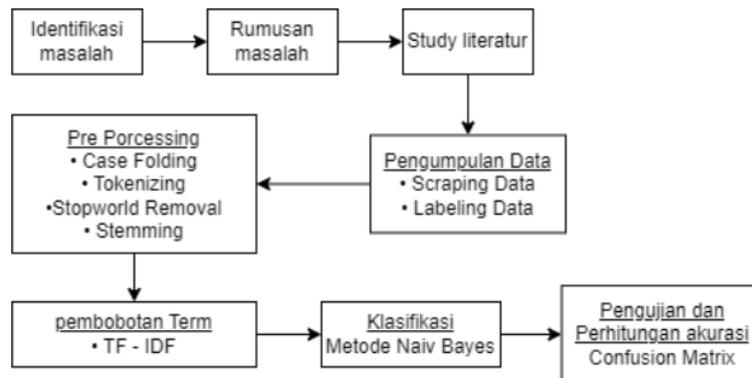
Dalam penelitian yang berjudul "Analisis Sentimen Calon Presiden dan Wakil Presiden periode 2019-2024 Pasca Debat Pilpres di Twitter" dilakukan crawling menggunakan library tweepy, Dataset yang diperoleh adalah opini Data opini yang diambil untuk masing-masing pasangan calon pada periode tanggal 14-17 April 2019 di Twitter dalam bahasa Indonesia sebanyak 100 data untuk setiap pasangan calon dengan menggunakan kata kunci "Jokowi-Ma'ruf" dan "Prabowo-Sandi". dengan pelabelan positif dan negatif. Diklasifikasi menggunakan Metode Naive Bayes Classifier (NBC) [9].

Pada penelitian sebelumnya ditemukan penerapan berbagai macam metode dan penggunaan metode Naive Bayes yang mirip dengan penelitian ini. Peneliti menemukan beberapa kekurangan dari beberapa jurnal referensi, salah satu diantaranya masih adanya nilai akurasi yang rendah, dataset yang sedikit, dan penerapan metode yang kurang simple. Peneliti melakukan pembaruan dengan menggunakan dataset lebih banyak dan terupdate, serta menggunakan tools yang performanya lebih unggul dibanding Twitter API yang tanpa registrasi dan menggunakan tokennya. Algoritma Naive Bayes dipilih karena model yang sederhana dan Implementasinya pun tidak terlalu rumit, cocok untuk mengevaluasi probabilitas bersyarat.

Pada penelitian ini, analisis sentiment twitter dilakukan agar dapat melihat sentiment masyarakat terhadap bakal calon pemilihan presiden 2024. Data tersebut akan diklasifikasikan menjadi 3 kelas, yakni positif, netral, dan negatif. Klasifikasi menggunakan algoritma Naive bayes. Diharapkan dari penelitian ini dapat memberikan gambaran umum dan menyimpulkan opini masyarakat terkait pemilihan presiden 2024. Informasi ini dapat membantu masyarakat untuk membuat keputusan yang lebih baik dalam pemilihan presiden serta dapat membantu mereka memahami dinamika politik yang terjadi dalam konteks pemilihan presiden.

II. METODE

Tahapan Penelitian yang dilakukan pada penelitian ini adalah sebagai berikut.



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1 pada penelitian ini ada 8 tahapan yang akan dilakukan dan dijelaskan lebih rinci pada penulisan berikut ini.

2.1 Identifikasi Masalah

Peneliti memulai dengan langkah identifikasi masalah, di mana mengamati fenomena atau objek yang akan diteliti, yaitu media sosial Twitter sebagai situs microblogging yang digunakan untuk menulis pesan. Data tekstual yang terdapat di Twitter banyak mengandung sentimen yang mengarah pada positif, negatif, maupun netral. Namun, kebanyakan pesan di Twitter disampaikan menggunakan bahasa yang tidak baku, sehingga menyulitkan dalam menemukan orientasi analisis sentimen dari data Twitter. Hal ini menjadi tantangan bagi peneliti untuk mengatasi permasalahan tersebut.

2.2 Rumusan Masalah

Setelah melakukan identifikasi masalah, rumusan masalah dapat dibuat dengan cara mempertimbangkan bagaimana data yang diperoleh dari Twitter dapat diolah menjadi fitur dan bagaimana pesan dari tweet yang mengandung makna sentiment dapat ditemukan. Rumusan masalah ini bertujuan agar penelitian dapat dilakukan dengan lebih terfokus dan efektif.

2.3 Studi Literatur

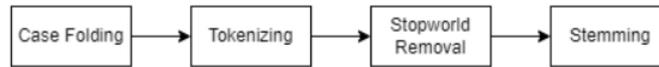
Untuk melakukan penelitian ini, dilakukan studi literatur dengan cara mengumpulkan informasi dan mencari referensi yang terkait dengan topik yang sedang diteliti melalui jurnal ilmiah dan sumber data yang berasal dari internet.

2.4 Pengumpulan Data

Teknik pengumpulan data pada penelitian ini diawali dengan menggunakan library twint, yakni sebuah tools yang digunakan untuk pengambilan data (scraping) dari aplikasi twitter yang khusus menggunakan python. twint mampu mengambil hampir semua tweets dan tidak perlu mendaftar di Twitter layaknya API Twitter yang harus registrasi dan menggunakan tokennya. Data yang didapat dari proses ini disimpan berformat .csv dan dilanjutkan dengan pemberian label positif, negative, dan netral secara manual.

2.5 Text Preprocessing

Setelah proses pengumpulan data, tahapan Text Preprocessing dilakukan. Ditunjukkan dengan pada Gambar 2.



Gambar 2. Tahapan Text Preprocessing

Pada Gambar 2 menunjukkan langkah proses Text Preprocessing yang merupakan tahapan awal dari proses terhadap teks guna mempersiapkan teks menjadi data yang siap diolah lebih lanjut [10]. Beberapa urutan tahapan yang akan dilakukan pada penelitian ini adalah *Case Folding*, *Tokenizing*, *Stopword Removal* dan *Stemming*. Penjelasan singkat akan dijelaskan sebagai berikut :

1. *Case Folding*. Pada tahap ini dilakukan pengecekan awal sampai akhir karakter, penghapusan karakter yang dianggap menjadi noise, jika ditemukan huruf kapital (uppercase) akan dirubah ke huruf kecil (lowercase). Dilakukan juga normalisasi fitur. Tahapan dilanjut dengan
2. *Tokenizing*. Yakni data hasil case folding akan proses dengan cara pemecahan atau pematongan kata (string) terhadap suatu teks yang disebut token. Proses menghilangkan kata-kata yang tidak memiliki kontribusi dalam pengklasifikasian teks.
3. *Stopword Removal* proses ini dilakukan dengan mencari kata – kata yang akan dihilangkan telah disimpan pada ruang stop list [11], proses ini menggunakan Library NLTK (Natural Language ToolKit) yang tersedia di Python.
4. *Stemming* merupakan tahapan terakhir dalam Pre-Processing. Tahap ini merupakan merubah kata menjadi kata dasarnya [12]. Dibantu Library NLP yang dikhususkan untuk Bahasa Indonesia yakni Sastrawi.

2.5 Pembobotan Term

Data yang telah melalui tahap preprocessing harus berbentuk numerik. Untuk mengubah data tersebut menjadi numerik yaitu menggunakan metode pembobotan TF-IDF [13]. Metode Term Frequency - Inverse Document Frequency atau TF-IDF merupakan proses perhitungan bobot kata pada dataset yang telah dilakukan pre-processing, bertujuan untuk mengetahui seberapa sering muncul pada dokumen.

2.6 Klasifikasi

Metode Naïve bayes Classifier atau (NBC), yang merupakan proses klasifikasi didasarkan probabilitas dan Teorima Bayesian. dengan sebutan lain, NBC yaitu salah satu metode machine learning yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, cara kerja Naive Bayes yaitu memprediksi probabilitasdi masa depan berdasarkan pengalaman di masa sebelumnya [14]. Formulasi sederhana dari klasifikasi NBC :

$$\text{posterior probability} = \frac{\text{conditional probability.prior probability}}{\text{evidence}} \quad (3)$$

Notasi yang umum untuk posterior probability dituliskan sebagai berikut :

$$P(Y_j|X_i) = \frac{p(Y_j|X_i)p(y_j)}{p(X_i)} \quad (2)$$

Dim: X_i :

X_i = sampel Vector i , $i \in \{1, 2, \dots, n\}$

Y_j = notasi kelas j , $j \in \{1, 2, \dots, m\}$

$P(Y_j|X_i)$ = probabilitas sampel X_i yang milik kelas W_j .

2.7 Pengujian dan Perhitungan Akurasi

Confusion matrik adalah salah satu metode pengujian untuk mengetahui data yang benar positif (TP), data yang salah positif (FP), data yang benar negatif (TN), dan terakhir data yang salah negatif (FN) [15]. tujuan dari metode pengujian ini untuk mengetahui akurasi dari suatu metode melalui recall, precision, dan f-measure.

Berikut rumus yang digunakan :

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (3)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (4)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (5)$$

$$f - measure = \frac{2 \times (Recall \times Precision)}{(Recall + Precision)} \quad (6)$$

Keterangan :

- True Positive (TP) : data positif diklasifikasi positif
- False Positif (FP) : data negatif diklasifikasi sebagai nilai positif
- True Negatif (TN) : data negative diklasifikasi negative
- False Negatif (FN) : data positif diklasifikasi sebagai nilai negative

III. HASIL DAN PEMBAHASAN

Pada penelitian ini, dataset yang telah diperoleh namun masih belum terstruktur atau data mentah ditunjukkan pada Tabel 1.

Tabel 1. Dataset mentah

Id	Conversation	date	time	timezone	User_id	username	Name
1,58E+18	1,58E+1	2022-10-2	06.58.54	700	1,33E+18	mdy_asm	Maudy a
1,58E+18	1,58E+1	2022-10-2	06.56.16	700	1,39E+17	Purwo_pa	Den-ma
1,58E+18	1,58E+1	2022-10-2	06.55.50	700	1,03E+18	Saharud65	Capres 2
1,58E+18	1,58E+1	2022-10-2	06.53.20	700	1,03E+18	Saharud65	Capres 2
1,58E+18	1,58E+1	2022-10-2	06.52.18	700	1,50E+18	Harmanbu	Eks jorn

Pada Tabel 1, manunjukkan data diperoleh melalui proses scraping menggunakan library Twint berjumlah 1040 data dan kemudian dilanjutkan dengan pelabelan manual, hasil ditunjukkan pada Tabel 2.

Tabel 2. Data hasil Scraping dan labeling

Id	tweet	Ulasan
1,58E+18	Pak Anies Baswedan menghadiri acara Rakernas Go-ANIES 23 Okt	Positif
1,58E+18	Pak Jokowi secara tidak langsung menyindir Pak SP dari Partai @NasDe	Negatif
1,58E+18	@tohir_extrem Anda Keren, Pak Anis... Sumpah! Anda gada lawan untuk	Negatif
1,58E+18	segitu aja belum deklarasi dgn pencitraan membabi buta,apa lagi sdh dek	Negatif
1,58E+18	@_ekokuntadhi @NasDem Apa urusanmu hei banci kaleng...Nasdem ter	Negatif

Pada Tabel 2 data sudah dilabeling dan dihapus data cuitan yang sama (duplicate) dengan total 64 sehingga berjumlah akhir 976 data sehingga dapat dilanjutkan untuk tahap text preprocessing.

Tahap proses text preprocessing diawali dengan tahap case folding diterapkan untuk menghapus karakter (noise) dan menormalisasi fitur yakni dengan penghilangan URL, karakter khusus seperti hastags(#), retweet(RT), tanda baca, emotikon (!.. dll).ditunjukkan pada Tabel 1.

Tabel 1. Case Folding

Sebelum	Sesudah
Pak Anies Baswedan menghadiri acara Rakernas Go-ANIES 23 Oktober 2022 Makin berwisata ya Capres? ðŸŽŹðŸŽŹ... https://t.co/YevgOHH0Iq	pak anies baswedan menghadiri acara rakernas go anies 23 oktober 2022 makin berwisata ya capres

Tokenizing dilakukan untuk mengolah hasil proses sebelumnya dengan memecah kata atau memotong per kata (string) sehingga dapat diproses pada proses berikutnya mendapat hasil pada Tabel 2.

Tabel 2. Tokenizing

Sebelum	Sesudah
pak anies baswedan menghadiri acara rakernas go anies 23 oktober 2022 makin berwibawa ya capres	pak, anies, baswedan, menghadiri, acara, rakernas, go, anies, 23, oktober, 2022, makin, berwibawa, ya, capres,

Stopword removal merujuk pada proses penghapusan kata-kata yang tidak memberikan kontribusi signifikan dalam klasifikasi teks, seperti kata "yang", "oleh", "di", dan sejenisnya, merupakan salah satu teknik yang biasa digunakan dalam pengolahan bahasa alami, terutama dalam analisis teks dan klasifikasi teks. perlu diingat bahwa penghilangan stopwords harus dilakukan dengan hati-hati dan tidak menghilangkan kata-kata yang penting dalam konteks analisis. Misalnya, dalam analisis sentimen, kata-kata seperti "tidak" dan "bukan" memiliki makna yang penting dan tidak boleh dihilangkan. Oleh karena itu, pemilihan kata-kata yang dihilangkan sebagai stopwords harus didasarkan pada konteks analisis yang dilakukan. Ditunjukkan pada Tabel 3.

Tabel 3. Stopword removal

Sebelum	Sesudah
pak, anies, baswedan, menghadiri, acara, rakernas, go, anies, 23, oktober, 2022, makin, berwibawa, ya, capres,	anies, baswedan, menghadiri, acara, rakernas, go, anies, 23, oktober, 2022, makin, berwibawa, capres, 2024,

Stemming dilakukan untuk Menghapus imbuhan kata pada depan dan belakang sebagai contoh, pada kata "penyebaran" diubah menjadi "sebar", "hiraukan" menjadi "hirau", "dukungan" menjadi "dukung" dan lain sebagainya. Ditunjukkan pada Tabel 4.

Tabel 4. Stemming

Sebelum	Sesudah
Aturan	Atur
Mengusung	Usung
Memilih	Pilih
Penyebaran	Sebar
Hiraukan	Hirau

Setelah tahap text preprocessing, mengolah dataset tweet untuk analisis sentimen atau topik tertentu, biasanya dilakukan penghitungan nilai Tf-Idf untuk menentukan kata-kata yang paling penting atau signifikan dalam dataset tersebut. Hal ini dilakukan dengan menghitung frekuensi kemunculan kata dalam dataset (term frequency/Tf) serta frekuensi kemunculan kata tersebut dalam seluruh dataset (inverse document frequency/Idf). Dari hasil perhitungan Tf-Idf ini, kata-kata yang memiliki nilai tinggi dapat dianggap sebagai fitur atau representasi dari dataset dan dapat digunakan dalam analisis lebih lanjut.

Tahap pembobotan TF IDF dilakukan. Dataset dibagi menjadi dua bagian, yaitu data latih dan data uji dengan skema beberapa pengujian rasio 90:10, 80:20, 70:30, dan 60:40. Skema Pengujian dilakukan dan hasil ditunjukkan pada gambar 3 sampai gambar 6.

	precision	recall	f1-score	support
-1	0.82	0.58	0.68	24
0	0.00	0.00	0.00	11
1	0.75	0.95	0.84	63
accuracy			0.76	98
macro avg	0.52	0.51	0.51	98
weighted avg	0.68	0.76	0.71	98

Gambar 3. Hasil Pengujian pertama

	precision	recall	f1-score	support
-1	0.78	0.73	0.75	59
0	0.50	0.04	0.08	23
1	0.77	0.94	0.85	114
accuracy			0.77	196
macro avg	0.68	0.57	0.56	196
weighted avg	0.74	0.77	0.73	196

Gambar 4. Hasil Pengujian kedua

	precision	recall	f1-score	support
-1	0.87	0.84	0.85	110
0	0.00	0.00	0.00	25
1	0.80	0.95	0.87	158
accuracy			0.83	293
macro avg	0.56	0.60	0.57	293
weighted avg	0.76	0.83	0.79	293

Gambar 5. Hasil Pengujian ketiga

	precision	recall	f1-score	support
-1	0.88	0.72	0.80	178
0	0.00	0.00	0.00	24
1	0.74	0.96	0.84	189
accuracy			0.80	391
macro avg	0.54	0.56	0.55	391
weighted avg	0.76	0.80	0.77	391

Gambar 6. Hasil Pengujian 6

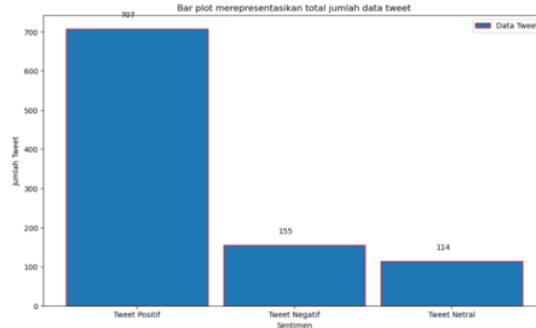
Untuk hasil pengujian keseluruhan dapat dilihat pada table 5.

Tabel 5. Hasil keseluruhan pengujian

Pengujian	Data training	Akurasi
Pertama	90%	76 %
Kedua	80%	77 %
Ketiga	70%	83 %
Keempat	60%	80 %

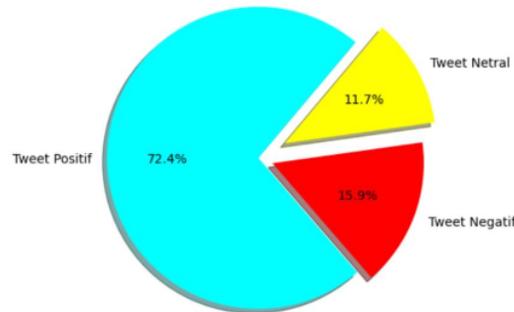
Dari Tabel 5 dapat dilihat bahwa akurasi tertinggi diperoleh dari pengujian ketiga yakni 83 %, kemudian diikuti pengujian ke empat dengan 80%, pengujian kedua 77% dan yang paling terendah pada pengujian pertama yaitu 76%. Hal tersebut menunjukkan bahwa skema pengujian dengan variable berbeda, maka mendapatkan nilai accuracy yang berbeda.

Pada Gambar hasil pengujian tersebut, menunjukkan hasil sentimen dimana nilai 1 adalah positif, 0 adalah netral, dan -1 adalah negatif. Hasil sentimen ini lebih dari 50% menunjukkan bahwa data yang peneliti peroleh menunjukkan hasil sentimen positif yang ditampilkan lebih lanjut melalui bar plot untuk total jumlah data tweet dan pie chart untuk representasi data tweet berdasarkan sentiment. Keseluruhan jumlah total data twitter disajikan pada Gambar 6.



Gambar 6. Jumlah total data tweet

Pada Gambar 6. menunjukkan jumlah total dari 976 data, untuk Sentimen Positif adalah 707, sentimen negatif adalah 155, dan untuk sentimen netral adalah 114. Untuk keseluruhan representasi data tweet berdasarkan sentiment disajikan pada gambar 7.



Gambar 7. Hasil klasifikasi data twitter

Gambar 7 menunjukkan dari total 976 data, diperoleh untuk nilai cuitan positif 72,4 %, cuitan netral 11,7 % dan untuk cuitan negatif 15,9 %.

IV. KESIMPULAN

Metode naïve bayies terbukti efektif mengklasifikasikan sentimen positif, negatif, dan netral terhadap pemilihan presiden 2024. Klasifikasi Algoritma Naive Bayes Classifier diterapkan dengan beberapa skema variable pengujian. Hasil yang terbaik didapat dengan pengujian 70% data latih dan 30% data uji diperoleh akurasi 83%. Algoritma Naive Bayes Classifier digunakan untuk melihat 976 tweet yang telah dibersihkan. Hasilnya adalah sentiment positif dengan jumlah 707 data, netral adalah 114 data, dan negatif adalah 155. Penerapan Algoritma Naive Bayes untuk sentimen analisis pada Pemilihan Presiden 2024 terbukti akurat. Peneliti juga berharap, bagi peneliti lain yang dalam penelitiannya akan menggunakan Algoritma Naïve Bayes, jurnal ini dapat menjadi referensi dalam memilih metode yang mudah dan sederhana untuk pengujian dan pelatihan data.

UCAPAN TERIMAKASIH

Terimakasih kepada Universitas Muhamadiyah Sidoarjo yang telah memberikan fasilitas laboratorium yang baik, sehingga penelitian ini dapat diselesaikan dengan hasil yang sesuai dengan harapan.

REFERENSI

- [1] P. Pasek, O. Mahawardana, G. A. Sasmita, P. Agus, and E. Pratama, "Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap 'Figure Pemimpin' Menggunakan Python," 2022.
- [2] K. Energi, D. Sumber, and D. Mineral, "PERAN TAGAR #ENERGIBERKEADILAN PADA MEDIA SOSIAL KEMENTERIAN ENERGI DAN SUMBER DAYA MINERAL HASHTAG BENEFITS #ENERGIBERKEADILAN ON SOCIAL MEDIA OF THE MINISTRY OF ENERGY AND MINERAL RESOURCES Thoriq Ramadani," 2019.
- [3] P. Pramelani and T. Widyastuti, "Persepsi Milenial terhadap Gaya Kepemimpinan Calon Presiden Tahun 2024," *Nakhoda: Jurnal Ilmu Pemerintahan*, vol. 20, no. 2, pp. 1–13, Dec. 2021, doi: 10.35967/njip.v20i2.196.
- [4] P. Pasek Okta Mahawardana, I. Ayu Putu Febri Imawati, and I. Wayan Dika, "ANALISIS SENTIMEN BERDASARKAN OPINI DARI MEDIA SOSIAL TWITTER TERHADAP 'FIGURE PEMIMPIN' MENGGUNAKAN PYTHON." [Online]. Available: <https://ojs.mahadewa.ac.id/index.php/jmti>
- [5] R. Syafaat Amardita and M. Dwifebri Purbolaksono, "Analisis Sentimen terhadap Ulasan Paris Van Java Resort Lifestyle Place di Kota Bandung Menggunakan Algoritma KNN," *Jurnal Riset Komputer*, vol. 9, no. 1, pp. 2407–389, 2022, doi: 10.30865/jurikom.v9i1.3793.
- [6] R. Fajar, S. Program, P. Rekayasa, N. Lunak, and R. Bengkalis, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," vol. 3, no. 1.
- [7] A. A. Permana *et al.*, "SENTIMEN ANALISIS OPINI MASYARAKAT PADA MEDIA SOSIAL TWITTER TERHADAP VAKSIN BERBAYAR MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER (NBC) SENTIMENT ANALYSIS OF PUBLIC OPINION ON SOCIAL MEDIA TWITTER ON PAID VACCINE USING NAÏVE BAYES CLASSIFIER (NBC) METHOD," vol. 10, no. 2, pp. 84–92, 2021, [Online]. Available: <http://jurnal.umt.ac.id/index.php/jt/index>
- [8] D. W. Seno and A. Wibowo, "Analisis Sentimen Data Twitter Tentang Pasangan Capres-Cawapres Pemilu 2019 Dengan Metode Lexicon Based Dan Support Vector Machine," *Jurnal Ilmiah FIFO*, vol. 11, no. 2, p. 144, Nov. 2019, doi: 10.22441/fifo.2019.v11i2.004.
- [9] S. Juanita, "Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 3, p. 552, Jul. 2020, doi: 10.30865/mib.v4i3.2140.
- [10] O. Y. Findawati, M. M. Muhammad, A. Rosid, S. Kom, and M. Kom, *BUKU AJAR TEXT MINING Diterbitkan oleh UMSIDA PRESS*.
- [11] F. Handayani, D. Feddy, and S. Pribadi, "Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110."
- [12] T. E. Hidayat and A. Rosid, "Analysis of Community Sentiments Regarding Plans to Relocate National Capital Using the Naïve Bayes Method Analisa Sentimen Masyarakat Tentang Rencana Pemindahan Ibukota Negara Dengan Metode Naïve Bayes," 2020.
- [13] A. Deolika and E. Taufiq Luthfi, "ANALISIS PEMBOBOTAN KATA PADA KLASIFIKASI TEXT MINING," *Jurnal Teknologi Informasi*, vol. 3, no. 2, 2019.
- [14] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," vol. 15, no. 1.
- [15] N. Faridhotul Hidayah, K. R. Paranita Kartika, and S. Nur Budiman, "PENERAPAN METODE NAIVE BAYES DALAM ANALISIS SENTIMEN APLIKASI SENTUH TANAHKU PADA GOOGLE PLAY," 2022.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Cek Artikel Hendra

ORIGINALITY REPORT

14%

SIMILARITY INDEX

13%

INTERNET SOURCES

8%

PUBLICATIONS

8%

STUDENT PAPERS

PRIMARY SOURCES

1	media.neliti.com Internet Source	2%
2	ejurnal.teknokrat.ac.id Internet Source	1%
3	jutif.if.unsoed.ac.id Internet Source	1%
4	Submitted to Udayana University Student Paper	1%
5	id.scribd.com Internet Source	1%
6	ejournals.itda.ac.id Internet Source	1%
7	joincs.umsida.ac.id Internet Source	1%
8	ejurnal.stmik-budidarma.ac.id Internet Source	1%
9	Submitted to Universitas Pelita Harapan Student Paper	1%

10	alvindayu.com Internet Source	1 %
11	Submitted to Ho Chi Minh University of Technology and Education Student Paper	1 %
12	repository.ittelkom-pwt.ac.id Internet Source	1 %
13	nakhoda.ejournal.unri.ac.id Internet Source	1 %
14	openlibrary.telkomuniversity.ac.id Internet Source	1 %
15	ojs.uajy.ac.id Internet Source	<1 %
16	repository.uin-suska.ac.id Internet Source	<1 %
17	Submitted to Universitas Brawijaya Student Paper	<1 %
18	jurnal.umt.ac.id Internet Source	<1 %
19	lweb.cfa.harvard.edu Internet Source	<1 %

Exclude bibliography On