

Analisis Sentimen Terhadap Kandidat Calon Presiden Berdasarkan *Tweets* Di Sosial Media Menggunakan *Naive Bayes Classifier*

Sentiment Analysis of Presidential Candidates Based on Tweets on Social Media Using the Naive Bayes Classifier

Allif Rizki Abdillah¹
Firman Noor Hasan^{2*}

^{1,2}Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Indonesia
¹allifrizki02@gmail.com, ²firman.noorhasan@uhamka.ac.id

***Penulis Korespondensi:**
Firman Noor Hasan
firman.noorhasan@uhamka.ac.id

Riwayat Artikel:

Diterima : 26 November 2022
Direview : 11 April 2023
Disetujui : 9 Mei 2023
Terbit : 30 Juni 2023

Abstrak

Penelitian ini untuk menganalisis sentimen rakyat Indonesia tentang para kandidat presiden yang kemungkinan akan maju di pilpres tahun 2024 dari tweet pada aplikasi *twitter*. *Tweet* pada *twitter* ini ditulis, diketik dan dipublish oleh netizen Indonesia tentang para kandidat yang kemungkinan akan maju dalam pilpres tahun 2024. Pada penelitian ini, peneliti menggunakan tools yaitu RapidMiner Studio untuk mengumpulkan data tweet dari netizen Indonesia tentang para kandidat. Selanjutnya peneliti menggunakan algoritma *Naive Bayes Classifier* untuk menentukan apakah sebuah statement atau sentimen itu bernilai positif atau negatif yang dilakukan menggunakan tools *Rapid Miner* juga. Dari keempat kandidat yang telah peneliti teliti, Anies mendapat 74% sentimen positif 26% sentimen negatif, lalu disusul oleh Sandi yaitu 57% sentimen positif 43% sentimen negatif, Ganjar mendapat 53% sentimen positif 47% sentimen negatif dan Prabowo mendapat 32% sentimen positif 68% sentimen negatif. Kesimpulan dari penelitian ini supaya bisa mengetahui mana kandidat yang disukai atau dipandang baik oleh rakyat Indonesia dari hasil analisis sentimen dengan algoritma *Naive Bayes* dan tools yang digunakan yaitu *Rapid Miner*.

Kata Kunci: RapidMiner, Presiden, Data Mining, Analisis Sentimen, Naive Bayes

Abstract

This research is to analyze the sentiments of the Indonesian people about the presidential candidates who are likely to advance in the 2024 presidential election from tweets on the Twitter application. Tweets on Twitter are written, typed and published by Indonesian netizens about the candidates who are likely to advance in the 2024 presidential election. In this study, researchers used tools, namely RapidMiner Studio to collect tweet data from Indonesian netizens about the candidates. Furthermore, the researcher uses the Naive Bayes Classifier algorithm to determine whether a statement or sentiment has a positive or negative value which is carried out using Rapid Miner tools as well. Of the four candidates that the researchers examined, Anies got 74% positive sentiment 26% negative sentiment, then followed by Sandi, namely 57% positive sentiment 43% negative sentiment, Ganjar received 53% positive sentiment 47% negative sentiment and Prabowo received 32% positive sentiment. 68% negative sentiment. The conclusion of this research is to find out which candidates are liked or favored by the Indonesian people from the results of sentiment analysis using the Naive Bayes algorithm and the tools used, namely Rapid Miner.

Keywords: RapidMiner, President, Data Mining, Sentiment Analysis, Naive Bayes

1. Pendahuluan

Pengguna media sosial semakin hari makin meningkat, ini disebabkan karena kemudahan yang diberikan media sosial pada masyarakat [1]. Di tahun 2019, Indonesia mengalami peningkatan pengguna internet apabila dilakukan perbandingan di tahun sebelumnya [2]. Naiknya pengguna internet di Indonesia terjadi karena banyak kemudahan yang ditawarkan dengan menggunakan media sosial atau internet. Dengan menggunakan media sosial orang dapat mengakses informasi dan komunikasi dengan sangat cepat. Mari mengambil contoh seperti menghubungi keluarga atau kerabat yang tinggalnya jauh dari tempat tinggal, dengan adanya media sosial seperti whats app, line, dll. Orang dapat saling bertukar kabar tanpa harus mengunjungi tempat kerabat yang ingin dihubungi, itulah salah satu contoh kegunaan media sosial dibidang komunikasi. Syarat berlangsungnya hubungan antar sesama manusia atau adanya interaksi sosial di antara mereka merupakan pengertian dari komunikasi [3]. Hal ini dapat terjadi karena semakin maju dan semakin canggih perkembangan teknologi di dunia saat ini seperti contoh, orang dapat mengakses informasi apa yang sedang terjadi di luar negeri tanpa harus pergi ke negara tersebut untuk melihatnya. Dengan berkembangnya teknologi di bidang internet sudah membuat penyebaran dari segi informasi dan komunikasi meningkat secara signifikan. Media sosial-lah yang mendukung penyebaran informasi tersebut dengan mudah [4]. Untuk sebagian besar penduduk Indonesia, media sosial menjadi identitas diri dan kebutuhan, bahkan mereka mampu 24 jam untuk tidak lepas dari *gadget* mereka. *Youtube, Facebook, Twitter, Whatsapp, Instagram* merupakan perangkat lunak media sosial yang sampai saat ini banyak digunakan di Indonesia [5]. Pengguna internet memiliki aplikasi media sosial menjadikan itu sebagai media bagi penggunanya untuk mencari informasi dan saling berkomunikasi secara *online* [6]. Twitter menjadi media sosial yang sering digunakan penggunanya untuk mengeluarkan opini atau pendapat dengan bebas [7]. Media sosial dapat memudahkan seseorang untuk melakukan komunikasi dan saling bertukar kabar, hal ini yang menyebabkan kehidupan masyarakat modern menjadi akrab dengan media sosial [8]. Karena media sosial twitter dapat dengan bebas mengutarakan opini atau pendapat melalui tweet, sehingga penelitian ini memanfaatkan twitter sebagai penyedia sentimen yang akan dianalisis nanti. Tweet ini ditulis atau diketik oleh netizen Indonesia sebagai bentuk penilaian atas kinerja para kandidat yang kemungkinan akan maju dalam pilpres 2024 mendatang.

Banyak metode untuk menganalisis pendapat orang yang berada pada internet atau media sosial, salah satunya yaitu metode analisis sentimen yang dapat menganalisis sebuah pendapat atau opini dan diberi nilai dari opini yang sudah dianalisis tadi. Metode analisis sentimen berfungsi untuk mengetahui polaritas dari data yang telah kita dapatkan dari internet [9]. Dengan analisis sentimen, peneliti dapat menganalisis polaritas dari pendapat seseorang atau sentimen yang telah dikumpulkan, sehingga dapat digunakan untuk memprediksi apakah suatu sentimen bernilai positif atau bernilai negatif. Analisis sentimen ialah solusi supaya peneliti dapat mengetahui tingkat kepuasan pengguna. Dari data yang tidak beraturan dapat ditarik sebuah kesimpulan dengan menggunakan metode analisis sentimen [10]. Pemrosesan bahasa natural (NLP) adalah salah satu bidang dari analisis sentimen dan merupakan proses yang saat ini digunakan oleh peneliti untuk mengidentifikasi dan mengklasifikasi isi dataset berupa pendapat atau opini yang berbentuk tulisan atau teks terhadap insiden atau kejadian yang bisa bernilai netral, positif ataupun negatif [11]. Sangat mudah membagikan dan memperoleh informasi saat ini, hal ini disebabkan karena perkembangan internet yang sangat cepat begitu pula dengan munculnya media sosial seperti twitter yang memberikan penggunanya kemudahan untuk membagikan informasi, memperoleh informasi dan memberikan pendapat terhadap topik, isu atau kejadian yang sedang hangat dibicarakan saat ini [12]. Dalam beberapa penelitian, sentimen bisa dikategorikan sebagai *bigdata* yang apabila semakin banyak ukuran data teks tersebut, maka makna dari konteksnya juga makin beragam [13]. *Data mining* adalah ilmu yang cocok diterapkan pada kasus seperti ini [14]. *Teks mining* menghasilkan berkas, yang berguna

untuk mewakili isi pada dokumen dengan mencari sebuah istilah yang berguna sebagai bentuk analisis keterkaitannya [15].

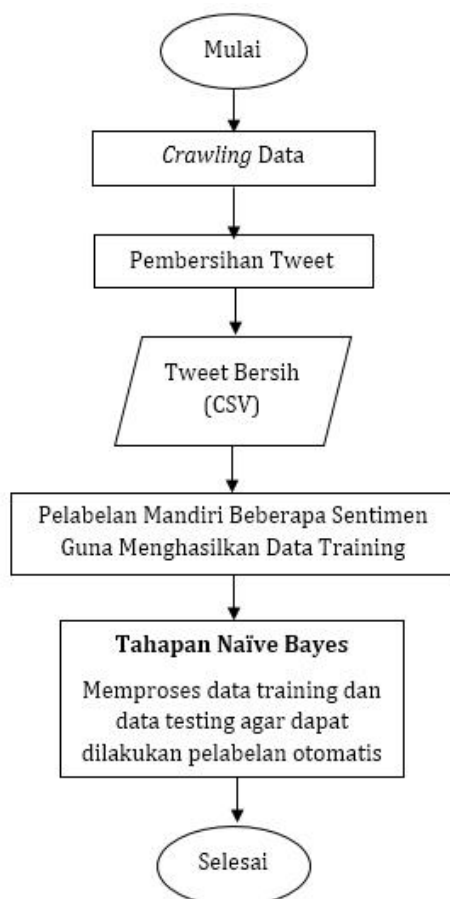
Metode analisis sentimen seperti *naïve bayes classifier*, *Random Forest*, dll merupakan metode yang sering digunakan dalam penerapan analisis sentimen [16]. *Naïve Bayes Classifier* adalah metode *Machine Learning* yang tidak terikat oleh aturan dengan kata lain bebas. *Naïve Bayes* menggunakan probabilitas untuk mencari sebuah kemungkinan dengan melihat frekuensi pada setiap klasifikasi terhadap data latih [17]. *Naïve bayes* mempunyai fungsi untuk melakukan klasifikasi pada data dan mencari nilai akurasi dari data [18]. Konsep dasar *Naive Bayes* menggunakan teorema Bayes, yang di mana sebuah teorema dalam ilmu statistik yang digunakan untuk melakukan perhitungan sebuah peluang. *Naive Bayes* dapat menghitung sebuah peluang pada satu kelas ke atribut-atribut yang tersedia, lalu dapat ditentukan kelas yang paling optimal. Algoritma *naïve bayes* digunakan untuk mencari dan menghitung nilai tertinggi dari probabilitas paling tinggi untuk mengklasifikasikan sebuah data uji dengan kategori yang tepat [19]. Metode ini menggunakan klasifikasi *naïve bayesian* yang memanfaatkan sebuah teorema probabilitas, fungsionalitas *data mining* serta teorema bayes. Penggunaan Multinomial *naïve bayes* yaitu untuk menghitung sebuah frekuensi kemunculan masing-masing token atau kata yang terdapat pada sebuah dokumen [20]. Data tersebut diuji dengan menggunakan *tools RapidMiner* [21]. *RapidMiner* adalah *tools* yang dapat melakukan pengolahan data. Dengan menggunakan algoritma dan prinsip data mining [22]. Analisis prediksi, *text mining*, *data mining* merupakan solusi yang diberikan oleh *tools RapidMiner studio*. Menggunakan berbagai teknik deskriptif supaya dapat mengambil keputusan yang baik [23].

Tujuan yang terdapat pada penelitian kali ini supaya peneliti dapat mengetahui tanggapan masyarakat terhadap para kandidat yang kemungkinan akan maju dalam pilpres 2024 mendatang. Berbeda dengan penelitian sebelumnya, penelitian yang telah kami lakukan yaitu menggunakan *tools RapidMiner* dan menggunakan algoritma *machine learning* yaitu *Naïve Bayes Classifier*. Dengan menggunakan *tools RapidMiner*, peneliti dapat mengambil data atau crawling data dari tweet pada twitter dengan hanya mencari nama dari kandidat tersebut. Twitter merupakan sebuah aplikasi media sosial untuk berbagi video, foto serta memiliki fitur-fitur menarik lainnya. Aplikasi ini juga mengizinkan para penggunanya untuk beropini secara bebas mengenai suatu hal aktual yang sedang hangat diperbincangkan. Salah satu hal yang diperbincangkan akhir-akhir ini mengenai kandidat presiden tahun 2024 [24]. Pada *tools RapidMiner* ini juga dapat menggunakan algoritma *Naïve Bayes Classifier* yang sudah tersedia di dalam aplikasinya. Singkatnya, dengan aplikasi ini peneliti dapat lebih efektif dan efisien dalam melakukan data mining, teks mining, sampai analisis sentimen pada tweet yang telah dikumpulkan.

Penelitian lain yang dilakukan sebelumnya dijelaskan juga mengenai cara crawling data yang dilakukan, yaitu dengan teknik web scraper. Peneliti akan mengekstraksi data tweet pengguna dengan rentang waktu tertentu dan memasukan kata kunci misalnya "vaksin corona". Kemudian hasil dari pencarian akan menampilkan tweet yang memiliki kata "vaksin corona" karena kata kunci yang telah kita masukan adalah "vaksin corona". Semua tweet yang diperoleh akan di simpan pada file csv untuk di analisis [25]. SVM atau kepanjangan dari *Support Vector Machine* merupakan teknik prediksi yang dapat melakukan klasifikasi dan regresi. Metode ini digunakan pada penelitian sebelumnya dan memerlukan pembelajaran yang diawasi di mana perlu fase pelatihan menggunakan SVM berurutan yang diikuti dengan fase pengujian. SVM adalah metode yang dapat memprediksi sebuah kelas dari model atau pola yang didasarkan pada hasil proses pelatihan [26]. Pada penelitian yang dilakukan peneliti kali ini peneliti menggunakan algoritma *naïve bayes* dalam menentukan nilai baik atau buruknya suatu kalimat atau tweet yang telah diambil sebelumnya. Pengambilan data tweet juga dilakukan menggunakan *tools RapidMiner*, dalam *tools RapidMiner* ini juga dilakukan pemberian analisis sentimen di dalam *tools* ini. Jadi, seluruh penelitian ini dilakukan di dalam *tools RapidMiner*.

2. Metode Penelitian

Penelitian yang dilakukan kali ini, peneliti menggunakan algoritma *machine learning* yaitu *naïve bayes classifier*. Algoritma ini berfungsi untuk menganalisa suatu kalimat atau kata apakah kalimat itu bernilai positif atau bernilai negatif. Pada Gambar 1 dijelaskan tahapan proses penelitian yang dilakukan oleh peneliti. Dimulai dari crawling data, pembersihan kalimat pada tweet, pelabelan mandiri kalimat, sampai selesai. Pada Gambar 1 peneliti hanya melakukan penelitiannya menggunakan *tools RapidMiner*, jadi pada Gambar 1 di bawah adalah proses yang dilakukan di dalam aplikasi. Berikut adalah Gambar 1 yaitu proses penelitian yang dilakukan.

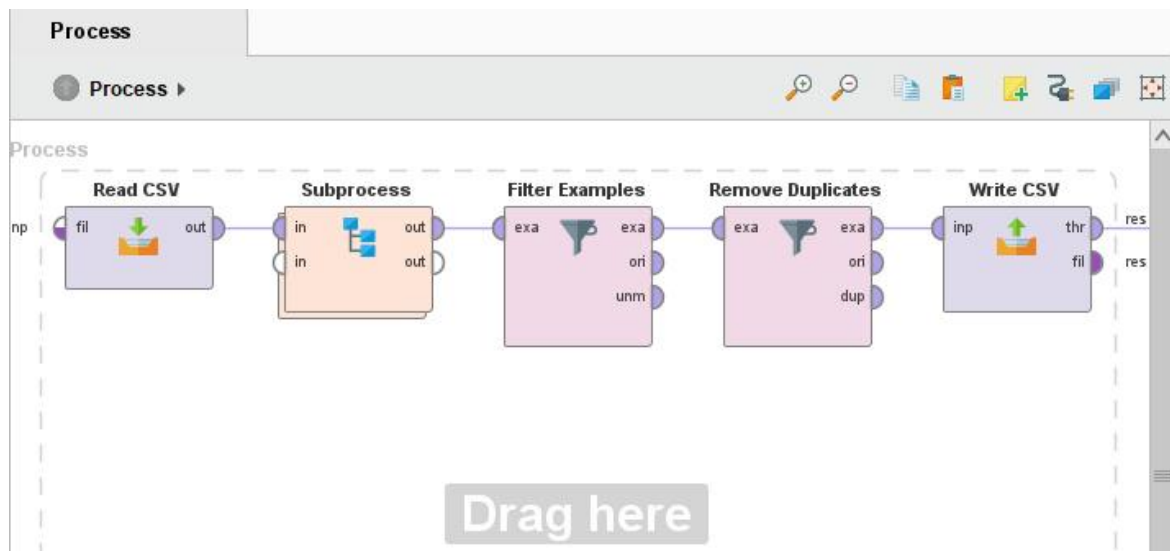


Gambar 1. Tahapan Proses

Gambar 1 merupakan tahapan proses penelitian yang dilakukan, dimulai dari crawling data pada *tools RapidMiner*. Crawling data yang dilakukan yaitu dengan mencari kata kunci yang diperlukan untuk pengambilan data, dalam hal ini kata kunci yang ditulis oleh peneliti yaitu nama dari kandidat yang kemungkinan akan maju pada pilpres 2024 di antaranya adalah Anies Baswedan, Ganjar Pranowo, Prabowo, dan Sandiaga. Dari keempat nama yang diketik pada kolom pencarian di *RapidMiner* munculah tweet yang memuat nama dari kandidat yang telah diketik atau dicari sebelumnya. Setelah itu, peneliti menyimpan tweet yang telah dicari tadi di dalam file csv untuk dilakukan proses selanjutnya.

Tahap selanjutnya yaitu pembersihan tweet yang telah diambil dan disimpan di dalam file csv. Tweet ini perlu dibersihkan karena, pada awalnya tweet ini masih memiliki simbol-simbol seperti [!~@#\$\$*%():"',.-]. Lalu beberapa tweet ada juga yang memiliki link dan kata RT (Retweet) pada kalimatnya. Tentu saja hal ini membuat proses analisis sentimen peneliti terganggu, oleh karena itu diperlukan pembersihan kalimat pada tweet agar kalimatnya bersih, tidak lagi

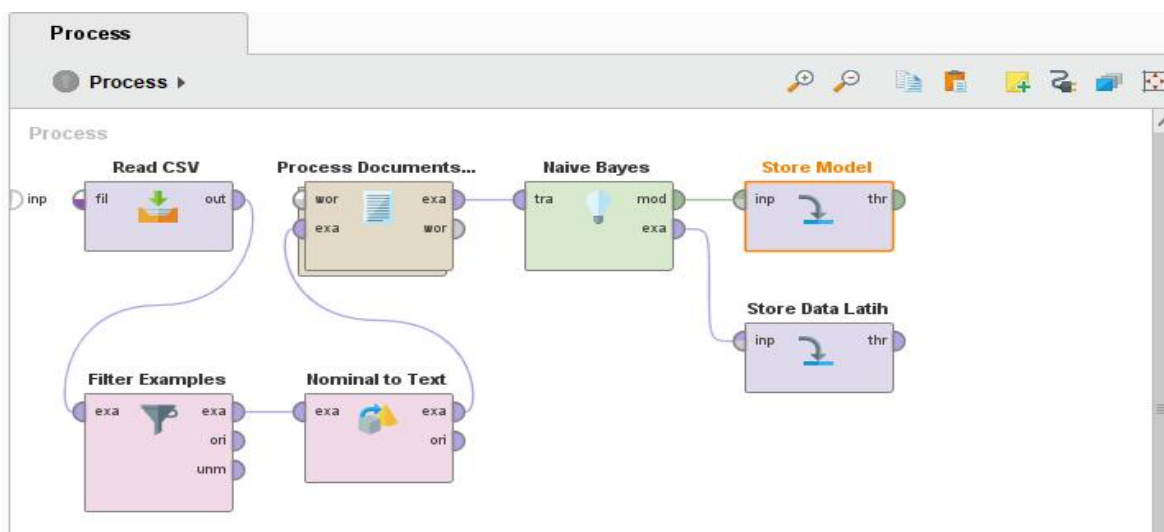
ada simbol-simbol, kata RT, link, dll. Pembersihan ini masih dilakukan dengan menggunakan *tools RapidMiner* dan setelah proses pembersihan selesai, tweet yang sudah dibersihkan tadi akan tersimpan ke dalam file csv. Berikut adalah gambar proses pembersihan tweet yang dilakukan di *RapidMiner* terdapat pada Gambar 2.



Gambar 2. Proses Pembersihan Tweet Pada *RapidMiner*

Setelah pembersihan tweet dari simbol, kata-kata yang tidak diperlukan, dan link sudah selesai, maka tahap selanjutnya adalah pelabelan tweet/kalimat secara mandiri. Pada proses ini, peneliti melabeli suatu kalimat/tweet/sentimen secara manual terlebih dahulu, apakah suatu sentimen itu bernilai positif atau negatif. Proses ini bertujuan untuk memberikan latihan kepada algoritma *naive bayes* supaya algoritma tersebut bisa melakukan pelabelan secara otomatis pada tahap selanjutnya. Pada proses ini, peneliti hanya melabeli 100 dari 300 data yang tersedia pada file csv, sisanya akan dilakukan pelabelan secara otomatis menggunakan algoritma *naive bayes classifier*. Perlu diingat, semakin banyak data yang dilabeli secara manual, maka semakin akurat juga *naive bayes* dalam memberikan label secara otomatis. Karena *naive bayes* akan memberikan label dengan data latih dari file csv yang sudah peneliti labeli secara manual tadi.

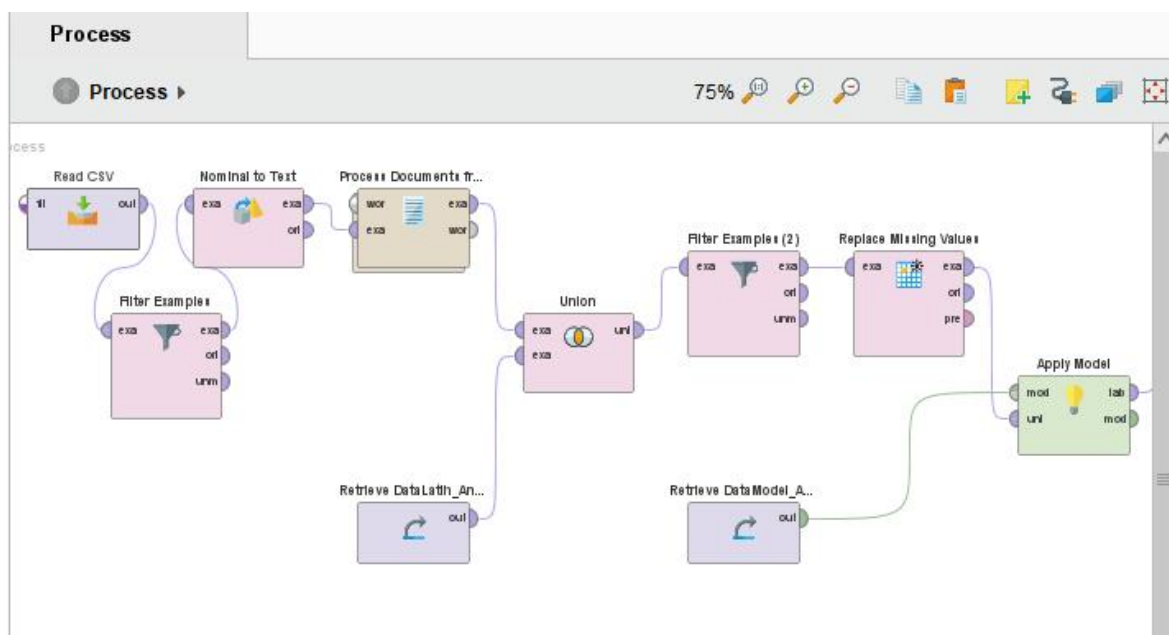
Proses berikutnya adalah memproses data yang telah peneliti labeli secara manual tadi ke dalam *RapidMiner*. Berikut adalah Gambar 3 tentang proses pemberian latihan kepada algoritma *naive bayes* berdasarkan data yang telah peneliti labeli secara manual tadi.



Gambar 3. Pemberian Latihan ke Algoritma *Naive Bayes*

Pada proses ini, peneliti perlu memasukkan data yang telah dilabeli secara manual tadi ke dalam *RapidMiner* karena *naive bayes* akan melatih dirinya untuk mempelajari atau menelusuri pola data yang telah peneliti labeli secara manual tadi. Oleh karena itu, seperti yang sudah dikatakan peneliti pada paragraf sebelumnya, lebih banyak data yang dilabeli secara manual, maka akan lebih akurat *naive bayes* dalam melabeli secara otomatis nantinya. Setelah mempelajari polaritas, algoritma tersebut akan menghasilkan model klasifikasi yang berguna untuk memberikan sentimen yang bernilai positif atau negatif pada tweet lain yang belum dilabeli secara manual. Setelah algoritma *naive bayes* ini mempelajari polaritas dari suatu sentimen, peneliti akan menyimpan model *naive bayes* ini ke dalam operator *store*, terdapat pada Gambar 3 di atas. Pada Gambar 3 di atas terdapat operator data model dan data latih, kedua operator tersebut akan peneliti gunakan untuk melakukan proses selanjutnya. Pada proses ini juga menggunakan parameter *tokenize* dan *stopwords* di dalam operator *Process Document from Data* (proses keempat). *Stopwords* yang digunakan di dalam *RapidMiner* menggunakan *stopwords* bahasa Indonesia yang dapat diunduh pada website *kaggle.com* karena *stopwords* dalam bahasa Indonesia belum tersedia pada *tools RapidMiner* tersebut.

Tahap terakhir adalah pelabelan tweet/sentimen secara otomatis dengan algoritma *naive bayes* yang telah dilatih di tahap sebelumnya tadi. Gambar 4 berikut adalah proses pelabelan otomatis dengan *tools RapidMiner*.



Gambar 4. Proses Pelabelan Secara Otomatis

Proses dari *Read CSV* sampai *Process Document from Data* (tahap 1 sampai 4) disebut dengan memproses data uji, dan pada kotak di bawah ada operator *Retrieve* data latih. Operator *retrieve* adalah operator yang telah peneliti lakukan pada proses sebelumnya (dapat dilihat pada Gambar 4), lalu peneliti menggabungkan data uji tadi dengan operator *retrieve* karena isi dari data uji dan isi dari *retrieve* data latih berbeda. Oleh karena itu, perlu dilakukan penggabungan isi data antara data uji dengan data latih menggunakan operator *Union*. Pada tabel 1 di bawah akan dijelaskan ilustrasi bagaimana cara kerja untuk menggabungkan data latih dan data uji agar mempunyai isi yang sama.

Tabel 1. Ilustrasi Penggabungan Isi Data

Data Latih					Sentimen
A	B	C	D		
0	0.2	0	0.6		Positif
0	0	0.2	0		Negatif

Data Uji (Tanpa Tabel)						Sentimen
C	D	E	F	G		
0	0	0.4	0	0	
0.2	0	0.3	0	0		Negatif

Pada Tabel 1 di atas dijelaskan bahwa antara data latih dengan data uji, jumlah kolom dan jumlah atributnya berbeda. Oleh karena itu, ini tidak bisa diproses langsung dan peneliti perlu menyamakan atribut antara kedua data ini. Caranya adalah, bisa dilihat bahwa di data uji dan data latih mempunyai sebuah irisan yang berarti atribut tersebut ada di data uji dan data latih,

irisian yang terdapat pada tabel 1 adalah kolom C dan kolom D. Kemudian peneliti perlu melakukan 3 langkah untuk mengatasi ini yaitu pertama dengan cara *Union*, pada Gambar 4 di atas terdapat operator *Union* untuk menggabungkan data latih dan data uji. Lalu yang kedua dilakukan *Filter Examples* (pada Gambar 4) untuk membuang data latih tadi. Terakhir peneliti melakukan *Replace Missing Values* karena ketika peneliti ingin menggabungkan data dengan atribut yang berbeda pasti ada atribut yang *missing* atau kosong maka dari itu, memerlukan operator ini. Tabel 2 berikut adalah hasil akhir setelah menggunakan 3 langkah tadi.

Tabel 2. Hasil Setelah Melakukan Penggabungan

A	B	C	D	E	F	G	Sentimen
0	0	0	0	0.4	0	0
0	0	0.2	0	0.3	0	0	Negatif

Pada Tabel 2 di atas dijelaskan bahwa peneliti sudah melakukan penggabungan kolom dan atribut dari data uji dengan data latih pada langkah pertama dengan operator *Union* pada *RapidMiner*. Lalu peneliti juga sudah membuang isi dari data latih tadi sehingga yang peneliti punya saat ini hanyalah data uji yang tidak memiliki label sentimen. Lalu yang terakhir peneliti lakukan adalah *Replace Missing Values* karena sebelumnya pada kolom A dan B terdapat tanda tanya pada *RapidMiner* karena sebelumnya data uji tidak memiliki kolom A dan B. oleh karena itu peneliti menggunakan operator *Replace Missing Values* untuk mengganti tanda tanya tadi dengan angka 0.

Tahap terakhir sekarang peneliti telah siap melakukan sentimen analisis tetapi, pertama-tama peneliti perlu memasukan *retrive data model* yang sebelumnya telah peneliti lakukan untuk melatih *naive bayes* agar dapat melakukan sentimen analisis secara otomatis seperti pada Gambar 4. Setelah itu gunakan operator *Apply Model* pada *tools RapidMiner*. Tarik data model tadi ke dalam mod (model) pada operator *Apply Model* lalu tarik juga *examples* pada *Replace Missing Values* ke unl (unlabel) pada operator *Apply Model*. Setelah itu dijalankan dan selesai, pelabelan otomatis dengan algoritma *machine learning naive bayes classifier* telah selesai.

3. Hasil dan Pembahasan

Peneliti mencari jumlah data masing-masing capres sebanyak 1000 data, tetapi pada akhirnya data tidak sampai 1000 dikarenakan perlu dilakukan pembersihan kalimat atau sentimen pada tweet dari kalimat yang duplikat. Lalu ada juga tweet yang hanya berisi link saja, oleh karena itu peneliti menghapusnya karena tidak berguna untuk proses analisis sentimen ini. Peneliti juga menguji tingkat akurasi dari algoritma *Naive Bayes Classifier* menggunakan *confusion matrix*. Dalam kasus ini, peneliti mengambil 4 kandidat yang kemungkinan akan maju dalam pilpres 2024 oleh karena itu, setiap kandidat mempunyai angka akurasi yang berbeda-beda. Semakin banyak data yang di analisis secara manual, maka akan semakin akurat juga *naive bayes* dalam melakukan analisis secara otomatis.

Sentimen Analisis Anies Baswedan

Pada Gambar 5 di bawah akan disajikan hasil sentimen analisis untuk Anies Baswedan berdasarkan tweet pada aplikasi twitter dengan menggunakan *RapidMiner*

Row No.	prediction(S...	text
1	negatif	anjing teriak cari kesalahan anies salah anjing cirinya suka lompat menjilat kenal cirici anjing
2	positif	dengarkan kebijaksanaan pengusiran sopan
3	negatif	khas buzzers sengaja diblowup atap halte ambruk plafon turun terpasang terinj
4	positif	anies ubah istilah rsud rumah sehat jakarta
5	positif	inshaalla
6	negatif	dasar burung merak
7	positif	orang terbodoh guntur romli orang lihat kenyataan anies kemajuan jakarta tolol bodoh tolol dibuang dipelihara bodoh permanen
8	positif	jafyaya
9	negatif	viral beredar penampakan kampung susun aquarium digusur ahok dibangun ulang anies penampakannya kayak kota
10	positif	pejabat
11	positif	test ombak dukung anies like rewtit koment pemimpin
12	positif	anies ganti nama rumah sakit rumah sehat masyarakat menggarisbawahai sakit sehat perub
13	positif	muara angka hasil karyanya anies
14	negatif	mengangkat angka kemiskinan bilang anies memimpin kemiskinannya pake data data kemiskinan menurun sengaja sembunyikan emar

Gambar 5. Hasil Sentimen Analisis Terhadap Anies Baswedan

Pada Gambar 5 di atas, diperoleh data sentimen analisis terhadap gubernur DKI Jakarta Anies Baswedan sebanyak 273 data. Total data yang diperoleh hanya 273 karena pada saat pembersihan kalimat/sentimen pada proses sebelumnya, dilakukan penghilangan kata-kata yang tidak diperlukan seperti simbol, link, dll. Lalu juga dilakukan penghilangan kata yang duplikat oleh karena itu, total akhir data yang diperoleh hanya sebanyak 273 data.

Langkah berikutnya yaitu melakukan pengujian tingkat akurasi terhadap analisis sentimen dari data Anies Baswedan yang dilakukan pada *tools RapidMiner Studio* menggunakan *confusion matrix*.

Tabel 3. Matrix Confusion Table Anies

	True Negative	True Positive	Class Precision
Pred. Negative	17	20	45.95%
Pred. Positive	10	3	23.08%
Class Recall	62.96%	13.04%	

Sentimen Analisis Ganjar Pranowo

Berikutnya akan disajikan gambar sentimen analisis terhadap gubernur Jawa Tengah yaitu Ganjar Pranowo. Hasil sentimen analisis terdapat pada gambar berikut.

Row ...	prediction(S...	text
12	positif	diakui pengembangan ekowisata diperbaiki jogging track dibanahi material bambu dinilai kuat
13	positif	produksinya bagus dikembangkan wisata orang makan ikan makan kepiting ganjar
14	positif	budidaya sukses memproduksi ikan udang bagus mendukung pengembangan ekowisata pengunjung belajar mangrove berkeliling hutan r
15	positif	coba kembangkan pariwisatanya budidaya ikan kepiting harapkan budidayanya produksinya ganjar
16	positif	ganjar ikhtlar menahan abrasi pantai selatan pohon bakau ditanam ekosistem bagus pohonpohon ditanam menambah fungsi penahan pa
17	positif	lihat sejarahnya mudah usaha membikin hutan mangrove sebgas lihat bentuknya bagus banget ganjar
18	positif	canangkan pengembangan ekowisata mangrove kawasan ekosistem esensial hutan mangrove muara kali dengar ceritanya bagus
19	positif	menurutnya hutan magrove muara kali memiliki nilai sejarah berpotensi ekowisata sumber bahan gula nipah
20	positif	sahabat ganjar jawa mengajak pengembangan ekowisata mangrove kawasan ekosistem esensial hutan mangrove muara kali desa ayah k
21	negatif	thread
22	negatif	seandainya perundangan disembunyikan
23	negatif	mengurangi potensipotensi bullying perundanganperundangan cegah ganjar
24	positif	jogo konco menjaga antarteman menghormati sukunya agamanya golongan sosialnya menghormati
25	positif	ganjar memiliki harapan dirilisnya aplikasi jogo koncom menanamkan kepedulian menjaga pertemanan implementasi peran anak pelopor

Gambar 6. Hasil Sentimen Analisis Terhadap Ganjar Pranowo

Gambar 6 di atas adalah hasil sentimen analisis terhadap gubernur Jawa Tengah saat ini yaitu Ganjar Pranowo. Setelah dilakukan pembersihan kata-kata yang tidak berguna, simbol-simbol, dan kalimat duplikat, peneliti memperoleh data Ganjar Pranowo sebanyak 302 data.

Langkah berikutnya yaitu melakukan pengujian tingkat akurasi terhadap analisis sentimen dari data Ganjar Pranowo yang dilakukan pada *tools RapidMiner Studio* menggunakan *confusion matrix*.

Tabel 4. Matrix Confusion Table Ganjar

	True Negative	True Positive	Class Precision
Pred. Negative	0	0	0.00%
Pred. Positive	7	28	80.00%
Class Recall	0.00%	100.00%	

Sentimen Analisis Prabowo Subianto

Pada Gambar 7 di bawah akan disajikan gambar sentimen analisis terhadap Menteri Pertahanan Indonesia saat ini yaitu Prabowo Subianto.

Row ...	prediction(S...)	text
11	positif	tuntaskan soliditas internal institusi Polri
12	negatif	skrng peran wapres serep pake bocor
13	positif	pembenahan manajemen tata kelola asabri menteri BUMN didukung diapresiasi
14	positif	patut dicontoh menteri lain dg tujua
15	negatif	lawan Anies Prabowo Puan
16	negatif	pengkhianat carilah pengkhianat ditempat maks gubernur diusung Prabowo akibat pengkhianatan skrg rakyat negara menanggung kerugi
17	positif	pidato kebangsaan muhaimin dimeria
18	negatif	elektabilitas Anies menanjak berpeluang menyialip Prabowo Ganjar
19	positif	menteri BUMN Erick Thohir membenahi manajemen tata kelola asabritermasuk menyeret pengemplang penjaras
20	positif	raih medali emas kejuaraan dunia pencak silat pesilat Safira Meliani terima kasih Prabowo Subianto
21	positif	Kapolda Jenderal Iisyo Sigit Prabowo anggota Polri diduga langgar etika terkait tewasnya brigadir Yosua
22	positif	kolaborasi menhan Prabowo Subianto menteri BUMN berhasil berkomitmen bersinergi kema
23	negatif	sampah celana ditemukan kawasan gunung gede Pangrango diduga celana dibuang pendaki mengalami cepiri
24	negatif	satwasatwa unik

Gambar 7. Hasil Sentimen Analisis Prabowo Subianto

Dari keempat data yang telah diperoleh peneliti pada penelitian kali ini, data Prabowo Subianto adalah yang paling banyak yaitu sebanyak 582 data. Tentu saja ini bisa saja terjadi karena jumlah duplikasi data pada tweet tidak terlalu banyak.

Langkah berikutnya yaitu melakukan pengujian tingkat akurasi terhadap analisis sentimen dari data Prabowo Subianto yang dilakukan pada *tools RapidMiner Studio* menggunakan *confusion matrix*.

Tabel 5. Matrix Confusion Table Prabowo

	True Negative	True Positive	Class Precision
Pred. Negative	0	0	0.00%
Pred. Positive	11	13	54.17%
Class Recall	0.00%	100.00%	

Sentimen Analisis Sandiaga

Yang terakhir adalah data sentimen analisis terhadap Menparekraf Indonesia yaitu Sandiaga Uno. Hasil sentimen analisis dari Sandiaga Uno bisa dilihat pada Gambar 8 di bawah ini.

Ro...	predict...	text
1	negatif	kerandoman jumatlan sandiaga semoga kecipratan kayanya
2	positif	sandiaga buka suara wisata pulau komodo tuai protes wisatawan
3	negatif	recovery together recovery stronger
4	positif	acara dihadiri negara iriana joko widodo sekretaris kabinet pramono anung menteri pariwisata ekonomi kreatif sandiaga salahuddin menteri perhu
5	negatif	ricuh tarif komodo susi pudjastuti sentil sandiaga
6	positif	suarasuara aceh sulawesi mendukung menteri pariwisata ekonomi kreatif sandiaga dimajukan capres
7	positif	sandiaga dorong pelaku umkm aceh berinovasi nilai produk
8	positif	sandiaga dorong desa wisata berkelanjutan
9	negatif	reminder
10	positif	sandiaga pengusaha menteri omsset triliunan rajin olahraga
11	positif	sandiaga dorong pengembangan wisata sejarah bengkulu
12	positif	ulee lheue desa miliki wisata
13	positif	acara dihadiri negara iriana joko widodo sekretaris kabinet pramono anung menteri pariwisata ekonomi kreatif sandiaga salahuddin menteri perhu
14	negatif	sandiaga

Gambar 8. Hasil Sentimen Analisis Sandiaga

Data yang diperoleh dari sentimen analisis terhadap Menteri Pariwisata dan Ekonomi Kreatif Indonesia sebanyak 398 data.

Langkah berikutnya yaitu melakukan pengujian tingkat akurasi terhadap analisis sentimen dari data Sandiaga Uno yang dilakukan pada *tools RapidMiner Studio* menggunakan *confusion matrix*. Hasil pengujiannya terdapat pada Tabel 6 berikut.

Tabel 6. Matrix Confusion Table Sandiaga

	True Negative	True Positive	Class Precision
Pred. Negative	3	1	75.00%
Pred. Positive	7	22	75.86%
Class Recall	30.00%	95.65%	

Di atas merupakan hasil daripada penelitian yang telah dilakukan, peneliti memutuskan untuk menggunakan *tools RapidMiner* karena mudah saat diaplikasikan langsung. Berbeda dengan penelitian yang disebutkan di atas yang menggunakan *Support Vector Machine (SVM)*, peneliti menggunakan algoritma *machine learning naive bayes classifier* dalam penelitian ini. Dalam hal crawling data juga berbeda dengan penelitian terdahulu, crawling data yang dilakukan peneliti di sini langsung dilakukan dalam *tools RapidMiner* jadi, semua penelitian ini dilakukan dalam satu *tools* saja, baik dalam crawling data, penggunaan algoritma *machine learning*, bahkan penyajian tingkat akurasi dengan *confusion matrix* di atas juga dilakukan dalam *tools RapidMiner*.

Penelitian ini mempunyai tujuan agar peneliti dapat mengetahui bagaimana pandangan masyarakat Indonesia terhadap calon-calon yang kemungkinan akan maju pada pemilihan presiden di tahun 2024 nanti. Siapapun yang akan menjadi presiden di tahun 2024 nanti akan menanggung beban yang berat dan kita sebagai rakyat Indonesia juga menginginkan sosok pemimpin yang bisa diandalkan, yang bisa diberi amanah untuk menyejahterakan Indonesia baik dalam 1 periode ataupun 2 periode. Oleh karena itulah penelitian ini dibuat agar rakyat Indonesia bisa melihat bayang-bayang mana yang baik dalam memimpin Indonesia.

Kekurangan yang ada pada penelitian ini yaitu pertama dari segi sampel data latih. Peneliti hanya melabeli manual sentimen yang ada sebanyak 100 data, dengan begitu masih ada beberapa data yang kurang akurat dalam pelabelian. Seperti yang sudah dijelaskan di atas, algoritma *naive bayes* akan semakin akurat apabila semakin banyak data yang dilabeli secara manual. Kekurangan yang kedua yaitu dalam sentimen analisis masih ada kalimat atau sentimen yang hanya terdiri dari 2 atau 1 kata saja dan tidak begitu jelas maksudnya apa. Ini bisa terjadi karena pada saat sebelum dilakukan pemberishan, banyak kata atau simbol-simbol yang tidak berguna, karena pada penelitian ini juga menggunakan *stopwords* bahasa Indonesia. Penggunaan *stopwords* yang diperlihatkan Gambar 9 yaitu menghapus kata yang tidak dibutuhkan secara otomatis.

1	ada
2	adalah
3	adanya
4	adapun
5	agak
6	agaknya
7	agar
8	akan
9	akankah
10	akhir
11	akhiri
12	akhirnya

Gambar 9. Contoh *stopwords* Bahasa Indonesia

Dengan menggunakan ini masih ada beberapa kata yang sulit untuk dipahami oleh manusia karena ada beberapa kata yang dihapus secara otomatis dengan menggunakan *stopwords* ini. *Stopwords* pada Gambar 9 didownload pada website [kaggle.com](https://www.kaggle.com), *stopword* berguna untuk menghapus kata yang tidak penting dalam melakukan sentimen analisis, pada Gambar 9 ditampilkan beberapa *stopwords* yang digunakan pada penelitian kali ini.

4. Penutup

Dapat ditarik kesimpulan dari penelitian di atas bahwa *naive bayes* dapat memprediksi sentimen secara otomatis tentang para kandidat yang kemungkinan akan maju dalam pemilihan presiden di tahun 2024 mendatang, baik itu sentimen yang bernilai positif maupun sentimen yang bernilai negatif. Pada *tools RapidMiner*, sebelum dilakukan pelabelan secara otomatis, maka diperlukan pemberian label terlebih dahulu secara manual suatu sentimen tersebut agar *naive bayes* dapat belajar dari pelabelan manual peneliti yang telah dilakukan sebelumnya. Semakin banyak data latih yang peneliti labeli secara manual, semakin akurat juga bagi *naive bayes* dalam melakukan pelabelan secara otomatis nanti. Begitu juga sebaliknya, apabila data yang dilabeli hanya sedikit, maka *naive bayes* tidak akurat dalam memberikan label secara otomatis. Apabila dalam penelitian ini diberi label "negatif" semua secara manual, maka *naive bayes* juga akan melabeli otomatis suatu sentimen/kalimat dengan label negatif. Oleh karena itu, perlu ketelitian dalam memberikan label secara manual juga. Setelah dilakukan pelabelan secara otomatis menggunakan algoritma *machine learning naive bayes classifier* peneliti mendapat kuantitas sentimen baik itu positif ataupun negatif terhadap para kandidat yang kemungkinan akan maju dalam pilpres 2024. Untuk Anies Baswedan mendapatkan sentimen negatif sebanyak 72 dan mendapat sentimen positif sebanyak 201 dari 273 data yang dianalisis, sedangkan Ganjar Pranowo mendapatkan sentimen negatif sebanyak 142 dan sentimen positif sebanyak 160 dari total 302 data. Selanjutnya Prabowo Subianto mendapatkan sentimen negatif sebanyak 394 dan sentimen positif sebanyak 188 dari total 582 data, sedangkan Sandiaga Uno mendapat sentimen

negatif sebanyak 170 dan sentimen positif sebanyak 228 data dari total 398 data yang dianalisis menggunakan *naive bayes*.

5. Referensi

- [1] A. Rahma, S. Delliana, J. Timur, and A. Verbal, "Pengaruh konten youtube ria ricis terhadap agresi verbal netizen," *Translitera*, vol. 11, no. 1, pp. 106–119, 2022, doi: <https://doi.org/10.35457/translitera.v11i1.1668>.
- [2] A. I. Hamid, "Analisis Semiotik Meme Anies Baswedan Banjir Jakarta," *Dialekt. KOMUNIKA J. Kaji. Komun. dan Pembang. Drh.*, vol. 8, no. 1, pp. 36–45, 2020, doi: [10.33592/dk.v8i1.554](https://doi.org/10.33592/dk.v8i1.554).
- [3] N. Anas, "Komunikasi antara Kognitif dan Kemampuan Berbahasa," *EUNOIA (Jurnal Pendidik. Bhs. Indones.*, vol. 1, no. 1, pp. 1–8, 2021, doi: <http://dx.doi.org/10.30829/eunoia.v1i1.997>.
- [4] R. D. Himawan and E. Eliyani, "Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi," *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 1, p. 58, 2021, doi: [10.26418/jp.v7i1.41728](https://doi.org/10.26418/jp.v7i1.41728).
- [5] S. R. I. Rezeki, "Penggunaan Sosial Media Twitter dalam Komunikasi Organisasi (Studi Kasus Pemerintah Provinsi Dki Jakarta Dalam Penanganan Covid-19)," *J. Islam. Law Stud.*, vol. 04, no. 02, pp. 63–78, 2020, doi: <http://dx.doi.org/10.18592/jils.v4i2.3804>.
- [6] T. M.Si., "Analisis Akses Dan Penggunaan Media Sosial Oleh Rumah Tangga Dan Individu Di Kota Batu Jawa Timur," *J. Komunika J. Komunikasi, Media dan Inform.*, vol. 7, no. 2, pp. 72–86, 2018, doi: [10.31504/komunika.v7i2.1627](https://doi.org/10.31504/komunika.v7i2.1627).
- [7] W. Maulana and M. Mulyadi, "Ujaran Kebencian Terhadap Jokowi pada Masa Pandemi Covid-19: Studi Kasus Twitter," *J. Lisnguistik Komputasional*, vol. 4, no. 1, pp. 27–33, 2021, doi: <https://doi.org/10.26418/jlk.v4i1.42>.
- [8] R. Yunanto, A. P. Purfini, and A. Prabuwisesa, "Survei Literatur: Deteksi Berita Palsu Menggunakan Pendekatan Deep Learning," *J. Manaj. Inform.*, vol. 11, no. 2, pp. 118–130, 2021, doi: [10.34010/jamika.v11i2.5362](https://doi.org/10.34010/jamika.v11i2.5362).
- [9] C.- Pandemic, B. Laurensz, and E. Sedyono, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19 (Analysis of Public Sentiment on Vaccination in Efforts to Overcome the," *J. Nas. Tek. Elektro dan Tekonolgi Inf.*, vol. 10, no. 2, pp. 118–123, 2021, doi: <https://doi.org/10.22146/jnteti.v10i2.1421>.
- [10] D. R. Alghifari, M. Edi, and L. Firmansyah, "Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia Bidirectional LSTM Implementation for Sentiment Analysis Against Grab Indonesia Services," *Manaj. Inform.*, vol. 12, no. 2, pp. 89–99, 2022, doi: <https://doi.org/10.34010/jamika.v12i2.7764>.
- [11] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020, doi: [10.32664/smatika.v10i02.455](https://doi.org/10.32664/smatika.v10i02.455).
- [12] L. Ardiani, H. Sujaini, and T. Tursina, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 2, p. 183, 2020, doi: [10.26418/justin.v8i2.36776](https://doi.org/10.26418/justin.v8i2.36776).
- [13] F. Sidik, I. Suhada, A. H. Anwar, and F. N. Hasan, "Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier," *J. Linguist. Komputasional*, vol. 5, no. 1, p. 34, 2022, doi: [10.26418/jlk.v5i1.79](https://doi.org/10.26418/jlk.v5i1.79).
- [14] A. S. Auliadaya, M. Rizki, and M. A. F. Azhary, "Analisa Pola Pembelian Produk Pada Toko Cimahi- Apriori," *Manaj. Inform.*, vol. 9, no. 1, pp. 58–69, 2019, doi: <https://doi.org/10.34010/jamika.v9i1.1668>.
- [15] F. N. Hasan and M. Dwijayanti, "Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naïve Bayes Classifier," *J. Linguist. Komputasional*, vol. 4, no. 2, pp. 52–58, 2021, doi: <https://doi.org/10.26418/jlk.v4i2.61>.
- [16] P. Arsi, R. Wahyudi, and R. Waluyo, "Optimasi SVM Berbasis PSO pada Analisis Sentimen

- Wacana Pindah Ibu Kota Indonesia," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 10, pp. 231–237, 2021, doi: <https://doi.org/10.29207/resti.v5i2.2698>.
- [17] E. Fitriani, "Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan," *Sistemasi*, vol. 9, no. 1, p. 103, 2020, doi: [10.32520/stmsi.v9i1.596](https://doi.org/10.32520/stmsi.v9i1.596).
- [18] N. Legiawati, T. I. Hermanto, and Y. R. Ramadhan, "Analisis Sentimen Opini Pengguna Twitter Terhadap Perusahaan Jasa Ekspedisi Menggunakan Algoritma Naïve Bayes Berbasis PSO," *J. Ris. Komput.*, vol. 9, no. 4, pp. 930–937, 2022, doi: [10.30865/jurikom.v9i4.4629](https://doi.org/10.30865/jurikom.v9i4.4629).
- [19] A. Tangkelayuk and E. Mailoa, "Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes Dan Decision Tree," *JATISI*, vol. 9, no. 2, pp. 1109–1119, 2022, doi: <https://doi.org/10.35957/jatisi.v9i2.2048>.
- [20] J. Winahyu and I. Suharjo, "Aplikasi Web Analisis Sentimen Dengan Algoritma Multinomial Naïve Bayes," *Kumpul. Artik. Mhs. Pendidik. Tek. Inform.*, vol. 10, no. 2, p. 206, 2021, doi: [10.23887/karmapati.v10i2.36609](https://doi.org/10.23887/karmapati.v10i2.36609).
- [21] C. Nas, "Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5," *J. Manaj. Inform.*, vol. 11, no. 2, pp. 131–145, 2021, doi: [10.34010/jamika.v11i2.5506](https://doi.org/10.34010/jamika.v11i2.5506).
- [22] K. F. Irnanda, A. P. Windarto, and I. S. Damanik, "Optimasi Particle Swarm Optimization Pada Peningkatan Prediksi dengan Metode Backpropagation Menggunakan Software RapidMiner," *J. Ris. Komput.*, vol. 9, no. 1, pp. 122–130, 2022, doi: [10.30865/jurikom.v9i1.3836](https://doi.org/10.30865/jurikom.v9i1.3836).
- [23] D. Ardiansyah and W. Walim, "Algoritma c4.5 untuk klasifikasi calon peserta lomba cerdas cermat siswa smp dengan menggunakan aplikasi rapid miner," *J. Inkofar*, vol. 1, no. 2, pp. 5–12, 2018, doi: <https://doi.org/10.46846/jurnalinkofar.v1i2.29>.
- [24] P. Pasek, O. Mahawardana, G. Arya, I. P. Agus, and E. Pratama, "Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap ' Figure Pemimpin ' Menggunakan Python," *JITTER-Jurnal Ilm. Teknol. dan Komput.*, vol. 3, no. 1, pp. 810–820, 2022, doi: <https://doi.org/10.5281/zenodo.7177756>.
- [25] 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](https://doi.org/10.26740/jieet.v4n2.p90-92).
- [26] H. Sulastomo, K. Gibran, and E. Maryansyah, "Analisis Sentimen Pada Twitter @ Ovo _ Id dengan Metode Support Vectore Machine (SVM)," *J. Sains Komput. dan Inform.*, vol. 6, no. September, pp. 1050–1056, 2022, doi: [http://dx.doi.org/10.30645/j-sakti.v6i2.514](https://doi.org/10.30645/j-sakti.v6i2.514).