

PERBANDINGAN ANALISIS SENTIMEN SETELAH PILPRES 2024 DI TWITTER MENGGUNAKAN ALGORITMA MACHINE LEARNING

Imam Syahrohim¹, Septian Dwi Saputra², Rizal Wahyu Saputra³,

Viktor Handrianus Pranatawijaya⁴, Ressa Priskila⁵

^{1,2,3}Universitas Palangkaraya; Jl. Yos Sudarso; Kota Palangkaraya; Kalimantan Tengah;
Telp (0536)3227111

Riwayat artikel:

Received: 26 Maret 2024

Accepted: 30 Maret 2024

Published: 2 April 2024

Keywords:

Analisis Sentimen;
Naive Bayes;
Support Vector Machine;
Logistic Regression.

Corespondent Email:

syahrohim55@mhs.eng.upr.ac.id

Abstrak. Penelitian ini bertujuan untuk menganalisis sentimen opini masyarakat di media sosial Twitter setelah Pemilihan Presiden 2024 menggunakan beberapa algoritma machine learning yaitu Naive Bayes, Support Vector Machine, dan Logistic Regression. Data yang digunakan bersumber dari Twitter dengan total 4.260 data yang terdiri dari 3 dataset calon presiden (Ganjar Pranowo, Anies Baswedan, dan Prabowo Subianto). Dilakukan proses crawling data, pembersihan data, pelabelan data menggunakan Vader, dan pelatihan data dengan TF-IDF sebelum dilakukan klasifikasi sentimen. Hasil eksperimen menunjukkan bahwa algoritma Logistic Regression memiliki performa terbaik pada dataset Ganjar Pranowo dengan akurasi 84,39%, presisi 84,92%, recall 84,39%, dan f-measure 81,52%. Penelitian ini memberikan insight tentang perbandingan performa algoritma klasifikasi dalam kasus analisis sentimen pada media sosial yang dapat digunakan sebagai referensi untuk penelitian serupa di masa mendatang.

Abstract. This study aims to analyze public opinion sentiment in social media on Twitter following the 2024 Presidential Election using many machine learning algorithms, including Naive Bayes, Support Vector Machine, and Logistic Regression. The data used is derived from Twitter and totals 4.260, consisting of three presidential datasets (President Ganjar Pranowo, President Anies Baswedan, and President Prabowo Subianto). Before doing a sentimen classification, the following processes are carried out: data crawling, data pembersihan, data pelabelan using Vader, and data training using TF-IDF. The experiment's results show that the Logistic Regression algorithm performs optimally on the Ganjar Pranowo dataset, with accuracy of 84.39%, prediction of 84.92%, recall of 84.39%, and f-measure of 81.52%. This study provides insight into the performance comparison of classification algorithms in the context of sentiment analysis on social media, which may be used as a reference for future research.

1. PENDAHULUAN

Pemilu merupakan ajang pemilihan setiap lima tahun sekali, yang mana akan dilaksanakan serentak di seluruh indonesia. Media sosial adalah sebuah media online yang bisa digunakan untuk melakukan berbagai hal seperti yang paling populer dilakukan adalah memposting sebuah opini ke media online. Salah satu media yang populer untuk hal tersebut adalah Twitter atau X. Kemudahan

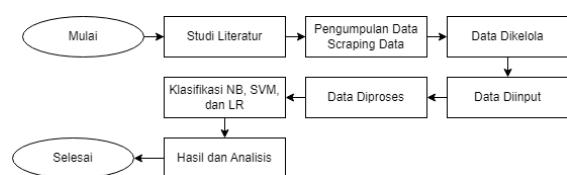
yang diberikan dalam mengakses media sosial dapat memberikan banyak opini yang disampaikan oleh masyarakat, yang dapat memengaruhi stigma atau stereotip yang melekat pada seorang kandidat atau calon. Opini-opini yang ada di laman twitter menarik untuk diklasifikasikan ke dalam sentimen positif ataupun sentimen negative[1]. Implementasi analisis sentimen akan membantu berbagai opini dari tweet tersebut, karena kita

dapat hasil dari tweet tersebut apakah bernilai positif dan negatif[2]. Pada penelitian ini diharapkan dapat menyelesaikan masalah yaitu mempolarisasi opini negatif dan positif yang ada di jejaring media sosial X. Metode yang dapat diimplementasikan yaitu Analisis Sentimen. Analisis Sentimen merupakan teknik yang dapat mendeteksi opini terhadap sebuah subjek dalam suatu kumpulan data[3]. Dengan menggunakan metode Machine Learning maka diharapkan akan memudahkan dalam melakukan polarisasi opini negatif dan positif yang ada cepat.

Penelitian ini dibuat karena pada pemilihan presiden yang ke-8 banyak sekali opini yang bertebaran di media sosial yang menyoroti berbagai paslon setelah dilakukan pencoblosan serentak pada 14 februari 2024. Pada penelitian ini akan membandingkan metode Machine Learning seperti Naive Bayes, Support Vector Machine, dan Logistic Regression yang digunakan untuk membuat pengklasifikasian atau polarisasi opini dengan menggunakan data yang telah dikumpulkan menggunakan crawling data. Hasil yang didapatkan dari penelitian akan membandingkan tingkat accuracy pada setiap metode Machine Learning yang digunakan untuk mempolarisasi.

2. TINJAUAN PUSTAKA

Tahapan penelitian ini dapat diketahui melalui diagram alir yang telah dibuat pada Gambar 2.1. Tahap pertama yang dilakukan adalah studi literatur untuk menemukan sebuah landasan teori dan resfrensi yang dapat mendasari penelitian yang dilakukan ini.



Gambar 2.1 Diagram Alir

2.1 Analisis Sentimen

Analisis sentimen adalah suatu metode yang digunakan untuk mengekstrak opini, memahami, dan mengolah data teks secara otomatis untuk menentukan sentimen yang terdapat dalam suatu opini. Metode ini berguna untuk mengetahui apakah opini tersebut positif, negatif, atau netral[4].

Pada analisis sentimen tingkat dokumen, pendekatan yang diambil adalah memperlakukan keseluruhan isi dokumen sebagai satu kesatuan dan mengategorikannya ke dalam sentimen positif atau negatif secara keseluruhan. Dengan kata lain, analisis ini tidak memecah dokumen ke dalam unit-unit yang lebih kecil seperti kalimat atau frase, melainkan menganalisis keseluruhan teks sebagai satu objek tunggal. Tujuan utamanya adalah menentukan apakah sentimen umum yang disampaikan dalam dokumen tersebut cenderung positif atau negatif. Pendekatan ini sesuai digunakan ketika peneliti atau analis ingin mendapatkan gambaran sentimen umum dari sebuah dokumen panjang, seperti ulasan produk, artikel opini, atau laporan tahunan perusahaan, tanpa harus menyelami setiap detail di dalamnya. Metode ini bertujuan untuk menentukan apakah keseluruhan konten dokumen cenderung menuju sentimen yang positif atau negatif[5]. analisis sentimen juga dapat digunakan untuk memahami emosi yang terkandung dalam cuitan yang disampaikan di Twitter. Dengan menganalisis teks dari cuitan, algoritma analisis sentimen dapat mengidentifikasi apakah cuitan tersebut mencerminkan sentimen positif, negatif, atau netral. Selain itu, dalam beberapa kasus, analisis sentimen juga dapat mencoba mengidentifikasi emosi khusus seperti kegembiraan, kemarahan, kesedihan, atau kecemasan yang terkandung dalam cuitan tersebut. Hal ini dapat memberikan wawasan tambahan tentang bagaimana pengguna Twitter merespons suatu topik atau peristiwa tertentu.

2.2 Crawling Data

Metode crawling data adalah suatu teknik pengambilan data dari suatu website secara otomatis dengan menggunakan program komputer. Metode ini memungkinkan pengambilan data dari berbagai sumber secara efisien dan terstruktur, sehingga dapat digunakan untuk berbagai keperluan seperti penelitian, analisis, dan pengembangan aplikasi[6]. Dengan metode ini dapat mengambil dan mengumpulkan data dari internet secara sistematis dan terorganisir menggunakan program komputer yang disebut *web crawler* atau *bot crawler*.

2.3 Data Cleaning

Disebabkan fakta bahwa data yang dikumpulkan untuk penelitian berasal dari data langsung yang ditemukan di Twitter, peneliti merasa bahwa ada kemungkinan bahwa data tersebut masih kotor dan memerlukan pembersihan tambahan. Untuk melakukan pembersihan ini, peneliti melakukan beberapa langkah yang akan dibahas di bawah ini :

2.3.1 Penghapusan Link, Hashtag(#), dan At(@)

Terkadang tweet warganet memuat link, hashtag, dan mention akun lain. Hal ini dapat mengganggu proses analisis. Dalam hal ini ketiga tersebut dapat mengganggu proses analisis yang dilakukan. Sehingga kita perlu menghapusnya[7].

2.3.2 Perbaikan Ejaan

Banyak data yang tidak menggunakan EYD ditemukan selama proses pembersihan selanjutnya, jadi diperlukan perbaikan untuk mengurangi ambiguitas.

2.3.3 Pelabelan Data

Karena metode klasifikasi diperlukan, penelitian yang akan datang akan membutuhkan label kelas dan data yang dikumpulkan akan dilabeli karena masih bersifat mentah[7].

2.3.4 Penghapusan Data Duplikat

Beberapa akun melakukan retweet atau memposting tweet yang sama lebih dari satu kali. Hal ini dapat menyebabkan data duplikat yang dapat memengaruhi hasil penelitian. Oleh karena itu, data duplikat dihapus[7].

2.3.5 Penghapusan Tanda Baca dan Perubahan Lowercase

Tanda baca dihapus dan semua huruf diubah menjadi lowercase (huruf kecil) untuk menyamakan bentuk huruf dari setiap kalimat[7].

2.3.6 Menghapus Stopword

Dalam proses penambangan data, kata-kata yang tidak penting atau stopword seringkali dihapus karena beberapa kalimat mengandung informasi yang kurang signifikan. Tindakan ini dilakukan untuk meningkatkan kualitas analisis data.

2.3.7 Stemming

Dalam beberapa cuitan, beberapa kata telah ditambahkan afiks, menambah keragaman kamus kata. Metode stemming digunakan untuk mengubah kata menjadi bentuk baku atau dasarnya[8].

2.4 Vader (Valence Aware Dictionary And Sentiment Reasoner)

VADER adalah sistem analisis sentimen berbasis leksikon yang dikhawasukan untuk analisis teks pada media sosial seperti *tweet*, update status, komentar online. Salah satu keuntungan menggunakan pendeksi polaritas VADER adalah tersedianya kamus yang mencakup nilai setiap kata. Apakah hasil preprocessing teks positif, negatif, atau neutral

akan dinilai menggunakan lexicon, dan skor total (compound) akan ditambahkan[9].

2.5 TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF (Term Frequency-Inverse Document Frequency) adalah metrik statistik yang sering digunakan dalam analisis teks dan pencarian informasi untuk menilai signifikansi suatu kata dalam sebuah dokumen atau koleksi dokumen. Term Frequency (TF) mengukur seberapa sering suatu kata muncul dalam sebuah dokumen. Pendekatan umum untuk menghitung TF adalah dengan membagi jumlah kemunculan kata tersebut dengan jumlah total kata dalam dokumen[10]. Inverse Document Frequency (IDF) adalah metode untuk mengevaluasi pentingnya suatu kata dalam konteks keseluruhan koleksi dokumen. Kata-kata yang jarang muncul di seluruh koleksi dokumen cenderung memiliki nilai IDF yang lebih tinggi[11].

2.6 Pattern Evaluation

Setelah proses data mining menggunakan algoritma naive bayes, penting untuk mengevaluasi model yang dihasilkan. Evaluasi ini melibatkan beberapa metrik, seperti akurasi dan Area Under Curve (AUC).

2.6.1 Akurasi

Nilai persentase yang diperoleh dari set tupel data yang telah diklasifikasi secara benar oleh model dikenal sebagai metrik akurasi. Fungsi metrik akurasi ini adalah untuk mengevaluasi kemampuan model untuk membedakan tupel dari berbagai kelas. Persamaan 5 berisi persamaan untuk metrik ini.

$$\text{akurasi} = \frac{TP+TN}{P+N}$$

2.6.2 Area Under Curve (AUC)

Metric ini menghitung sebagian nilai dari luas persegi yang menghasilkan grafik Receiver Operating Characteristic (ROC) Curve, yang hasil AUC-nya selalu berada pada 0,0 dan 1,0. Nilai False Positive (FP) dan True Positive (TP) digunakan pada sumbu x ROC Curve dan pada sumbu y, masing-masing. Persamaan untuk setiap nilai ini ditemukan pada persamaan 6 dan 7.

NILAI AUC	Kualitas
0.90 – 1.00	<i>Excellent classification</i>
0.80 – 0.90	<i>Good classification</i>
0.70 – 0.80	<i>Fair classification</i>
0.60 – 0.70	<i>Poor classification</i>
0.50 – 0.60	<i>Failure</i>

Gambar 2.2 Kualitas Model AUC

$$TP\ Rate = \frac{TP}{TP+FN}$$

$$FP\ Rate = \frac{FP}{FP+TN}$$

2.7 Model Machine Learning

Model algoritma Machine Learning yang digunakan ada 3, yaitu Naïve Bayes Classifier, Support Vector Machine, dan Logistic Regression

- Support Vector Machine (SVM)*: Support Vector Machine (SVM) adalah salah satu algoritma pembelajaran mesin yang digunakan untuk melakukan klasifikasi data. SVM berfungsi dengan menemukan garis pemisah (hyperplane) terbaik yang dapat memisahkan dua kelas data secara optimal. Dalam kasus analisis sentimen, dua kelas tersebut adalah sentimen positif dan sentimen negatif. Hyperplane yang ditemukan oleh SVM akan memaksimalkan jarak antara titik-titik data terdekat dari masing-masing kelas dengan garis pemisah tersebut. Dengan demikian, SVM dapat mengklasifikasikan data baru ke dalam kelas positif atau negatif berdasarkan posisi data tersebut terhadap hyperplane[12].
- Naïve Bayes Classifier (NBC)*: NBC (Naive Bayes Classifier) adalah algoritma yang menggunakan teori probabilitas Bayesian untuk mengklasifikasikan data. Meskipun NBC umumnya digunakan untuk data teks atau kategorikal, namun juga bisa diterapkan pada data numerik[13]
- Logistic Regression (LR)*: LR adalah algoritma yang digunakan untuk menguji probabilitas suatu variabel dependen dengan variabel bebas.[10]

3. METODE PENELITIAN

3.1 Crawling

Data untuk penelitian ini dikumpulkan menggunakan metode crawling dengan menggunakan alat colab.research. Proses pengambilan data memerlukan Token Twitter. Untuk menjalankan alat ini, pengguna diminta untuk memasukkan kata kunci yang ingin dicari dalam tweet. Data dikumpulkan dari rentang waktu tweet antara tanggal 27 Februari hingga 2 Maret 2023. Hasil pengumpulan data ini terdiri dari total 4240 tweet, dengan 1149 tweet berasal dari Ganjar Pranowo, 1553 tweet dari Anies Baswedan, dan 1558 tweet dari Prabowo Subianto. Hasil pengambilan data ini akan ditampilkan dalam Tabel 3.1.

Tabel 3.1 Hasil Crawling Data

Dataset	Jumlah Data
Ganjar Pranowo	1149
Anies Basweda	1553
Prabowo Subianto	1558
Total	4260

Peneliti ini mengambil atribut dari dataset lain yang dianggap penting untuk melakukan peninjauan ulang terhadap data Tweet yang telah dikumpul sebelumnya selama proses Crawling Data. Namun, hanya atribut teks yang digunakan untuk tahap selanjutnya. Jenis atribut yang digunakan ditunjukkan dalam tabel 3.2.

Tabel 3.2 Atribut Dataset Tweet

Nama Atribut	Penjelasan Atribut	Tipe Data Atribut
<i>created_at</i>	Waktu dan tanggal dari pembuatan cuitan tersebut	<i>Date and time</i>
<i>username</i>	Username dari akun twitter yang membuat cuitan tersebut	<i>String</i>
<i>tweet_url</i>	Link dari Tweet warganet Tersebut	<i>String</i>
<i>Interaction</i>	Jumlah Interaksi pada tweet	<i>Integer</i>
<i>full_text</i>	Tweet dari warganet	<i>String</i>

3.2 Data Cleaning

Proses ini dilakukan beberapa tahapan pembersihan yang akan membuat Data diolah lebih mudah. Pada tahapan pembersihan dari Awal hingga pembersihan perbaikan ejaan akan ditampilkan pada table 3.3.

Tabel 3.3 Data cleaning

Nama Atribut	Sebelum	Sesudah
<i>Mengubah semua huruf ke Lowercase</i>	@Belok_dong @miduk17 bikin konser dimana yg datang pemuda pendukung 02, cuma dimanfaatin buat hiburan aja. Serangan fajar dimana-mana sampe ratusan ribu isinya. Juga cuma diterima diutinya, coblosnya tetep 02. G tauanya yg pilin g lebih banyak dr anies Rugi nya double g kira-kira. Kalah terima.	@belok_dong @miduk17 bikin konser dimana yg datang pemuda pendukung 02, cuma dimanfaatin buat hiburan aja, serangan fajar dimana-mana sampe ratusan ribu isinya, juga cuma diterima diutinya, coblosnya tetep 02. g tauanya yg pilin g lebih banyak dr anies rugi nya double g kira-kira. Kalah terima.
<i>Menghapus link, hashtag(#), dan at(@)</i>	@gelarace awalnya minta pemula diulang setelah beberapa tps puu dan makin nyungsep sekaran minta nya baru lagi diulang tpi 02 pak usah ikut wkwkw,, sehat pak? sekalain aja minta langsung anies dilantik jd presiden tanpa pemula	awalnya minta pemula diulang setelah beberapa tps puu dan makin nyungsep sekaran minta nya baru lagi diulang tpi 02 pak usah ikut wkwkw,, sehat pak? sekalain aja minta langsung anies dilantik jd presiden tanpa pemula
<i>Menghapus Stopword</i>	awalnya minta pemula diulang, makin nyungsep, sekaran minta baru lagi diulang tapi gak usah ikut. Sehat pak! Langsung Anies dilantik presiden tanpa pemula	awalnya minta pemula diulang, makin nyungsep, sekaran minta baru lagi diulang tapi gak usah ikut. Sehat pak! Langsung Anies lantik presiden tanpa pemula
<i>Stemming</i>	awalnya minta pemula diulang, makin nyungsep, sekaran minta baru lagi diulang tapi gak usah ikut. Sehat pak! Langsung Anies dilantik presiden tanpa pemula	awalnya minta pemula diulang, makin nyungsep, sekaran minta baru lagi diulang tapi gak usah ikut. Sehat pak! Langsung Anies lantik presiden tanpa pemula

Tahapan Selanjutnya Adalah Menghapus data *double* atau data yang terduplikat untuk mendapat jumlah data real/asli, kemudian hasil dari perubahan kuantitas jumlah data disajikan pada tabel 3.4.

Tabel 3.4 Hasil Penghapusan Data Duplikat

Data	Sebelum	Sesudah	Selisih
Ganjar Pranowo	1149	1104	45
Anies Basweda	1553	1528	25
Prabowo Subianto	1558	1551	7

3.4 Pelabelan Data Vader

Tahapan selanjutnya adalah pemberian label terhadap data dengan menggunakan bantuan *Vader*, dengan bantuan program sehingga mendapat analisis

sentimen yang beragam, dengan hasil data pada tabel 3.5.

Tabel 3.5 Label Dataset Validator

Data	Jumlah Data			Jumlah Data
	Positif	Netral	Negatif	
Ganjar Pranowo	535	133	449	1117
Anies Baswedan	682	184	670	1528
Prabowo Subianto	784	402	365	1551

3.5 Training Data

85% data juga menggunakan proses pelatihan data. Tujuannya adalah agar pembelajaran mesin dapat melatih data yang berupa kata atau teks. Metode pelatihan ini digunakan untuk mengetahui hubungan antara kata dan dokumen dengan memberikan bobot pada setiap kata yang ada di dokumen. Metode ini menggabungkan dua ide: frekuensi kata yang muncul di dokumen dan inverse frekuensi dokumen yang mengandung kata tersebut.

3.6 Data Balancing

Pada Tahap ini Dilakukan Data Balancing untuk memastikan bahwa model yang dibangun tidak cenderung memihak kepada kelas mayoritas karena kelebihan jumlahnya, sehingga meningkatkan kinerja model terhadap kelas minoritas dalam kasus klasifikasi sentimen.

Tabel 3.6 Label Dataset Validator

Data	Sebelum			Sesudah		
	Positif	Netral	Negatif	Positif	Netral	Negatif
Ganjar Pranowo	535	133	449	289	478	350
Anies Baswedan	682	184	670	592	749	569
Prabowo Subianto	784	402	365	667	350	306

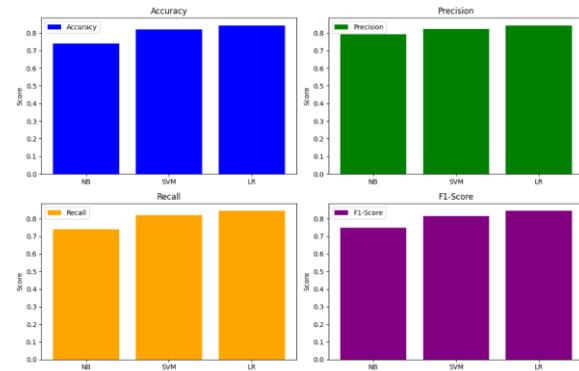
4. HASIL DAN PEMBAHASAN

4.1 Evaluasi Model

Pengujian masing-masing algoritma menggunakan perangkat lunak Colab Research , Machine Learning yang akan digunakan ada 3, yaitu Support Vector Machine Naïve Bayes Classifier, dan Logistic Regression.

Tabel 4.1 Hasil Perhitungan confusion matrix Dataset Ganjar

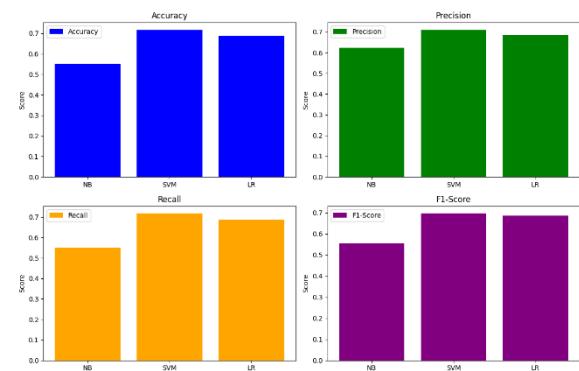
METODE	Jumlah Data			
	Akurasi	Presisi	Recal	F-Measure
Naive Bayes	73,99	79,30	73,99	74,94
Support Vector Machine	82,08	82,50	82,08	81,52
Logistic Regression	84,39	84,92	84,39	81,52



Gambar 4.1 Grafik Perbandingna Peforma Algoritma Data Ganjar

Tabel 4.2 Hasil Perhitungan confusion matrix Dataset Anies

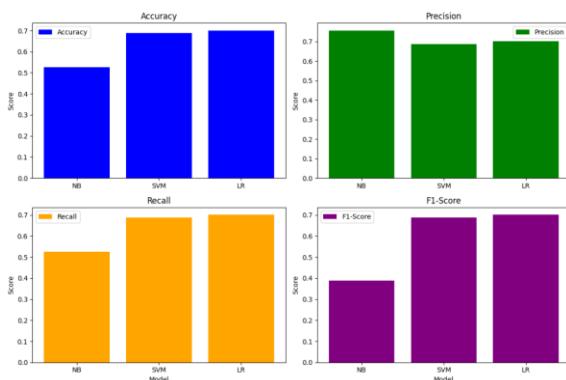
METODE	Jumlah Data			
	Akurasi	Presisi	Recal	F-Measure
Naive Bayes	54,94	62,43	54,94	55,42
Support Vector Machine	68,67	70,54	68,67	68,73
Logistic Regression	71,67	70,99	71,67	69,63



Gambar 4.2 Grafik Perbandingna Peforma Algoritma Data Anies

Tabel 4.3 Hasil Perhitungan confusion matrix Dataset Prabowo

METODE	Jumlah Data			
	Akurasi	Presisi	Recal	F-Measure
Naive Bayes	52,56	75,66	52,56	38,76
Support Vector Machine	68,80	68,71	68,80	68,75
Logistic Regression	70,09	73	70,09	67,86



Gambar 4.3 Grafik Perbandingan Peforma Algoritma Data Prabowo

Tabel dan Grafik Diatas Merupakan Tabel dan Grafik Perbandingan Algoritma dari masing masing dataset.

Pada Dataset Ganjar Algoritma yang menunjukan Nilai akurasi tertinggi terdapat pada algoritma LR(84,39) , Niali Presisi tertinggi terdapat pada algoritma LR (84,92) nilai recall tertinggi terdapat pada algoritma LR(84,39) dan Nilai F-Measure Tertinggi terdapat pada algoritma SVM dan LR (81,52)

Pada Dataset Anies Algoritma yang menunjukan Nilai akurasi tertinggi terdapat pada algoritma LR(71,67) , Niali Presisi tertinggi terdapat pada algoritma LR (70,99) nilai recall tertinggi terdapat pada algoritma LR(71,67) dan Nilai F-Measure Tertinggi terdapat pada algoritma LR (69,63)

Pada Dataset Prabowo Algoritma yang menunjukan Nilai akurasi tertinggi terdapat pada algoritma LR(70,09) , Niali Presisi tertinggi terdapat pada algoritma LR (73) nilai recall tertinggi terdapat pada algoritma LR(70,09) dan Nilai F-Measure Tertinggi terdapat pada algoritma SVM (68,75)

Berdasarkan 3 dataset diatas diperoleh perbandingan nilai akurasi dari masing-masing algoritma adalah sebagai berikut: Naïve Bayes Classifier terbesar terdapat pada dataset Ganjar 73,99% , Support Vector Machine (SVM) terbesar pada data ganjar 82,08% dan Logistic Regression sebesar 84,39%. Algoritma Logistic Regression (LR) menunjukkan nilai akurasi tertinggi di antara algoritma lainnya, menunjukkan kemampuannya dalam pengklasifikasian yang maksimal dibandingkan dengan algoritma lainnya. Sementara itu, nilai presisi tertinggi terdapat pada algoritma Logistic Algoritma juga , menandakan bahwa prediksi label positif dari algoritma ini lebih sering

muncul dibandingkan dengan algoritma lainnya. Selanjutnya, nilai recall tertinggi terdapat pada algoritma LR , menunjukkan kemampuannya dalam melakukan penguatan pada label positif terhadap prediksi label positif. Terakhir, nilai f-measure tertinggi terdapat pada algoritma LR dan juga SVM, yang merupakan nilai rata-rata dari recall dan presisi, menunjukkan keseimbangan antara kedua metrik tersebut dalam algoritma LR. Juga Dapat dilihat Bahwa Algoritma Naïve Bayes Memiliki Nilai Terendah dari semua algoritma menggunakan 3 dataset yang berbeda.

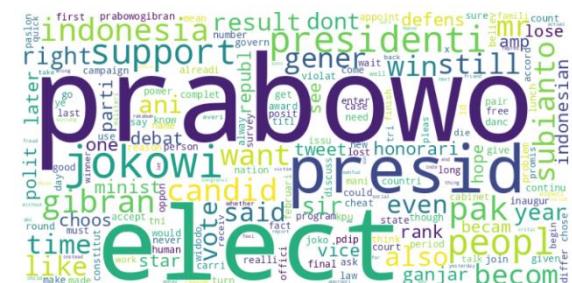
4.2 Visualisasi Data



Gambar 4.4 Kata yang paling Banyak muncul Dari Dataset Ganjar



Gambar 4.5 Kata yang paling Banyak muncul Dari Dataset Anies



Gambar 4.6 Kata yang paling Banyak muncul Dari Dataset Anies

Wordcloud adalah sebuah teknik visualisasi yang merepresentasikan kata-kata dalam bentuk gambar, di mana ukuran font setiap kata menggambarkan frekuensi kemunculannya dalam sebuah korpus teks. Teknik ini berfungsi sebagai alat analisis untuk

mendapatkan gambaran umum atau wawasan holistik terhadap suatu subjek atau topik. Dalam konteks penelitian ini, wordcloud digunakan untuk menganalisis wacana calon presiden Indonesia pada Pemilu 2023. Dengan memvisualisasikan kata-kata yang sering muncul dalam pidato atau pernyataan para calon presiden, wordcloud memungkinkan peneliti untuk mengidentifikasi isu-isu utama dan fokus kampanye masing-masing calon. Hal ini memberikan pemahaman yang lebih mendalam tentang platform dan agenda mereka dalam menyongsong Pilpres 2024.

KESIMPULAN

- a. Dalam penelitian ini, dilakukan uji coba perbandingan kinerja dari 3 algoritma klasifikasi: Naïve Bayes Classifier, Support Vector Machine, dan Logistic Regression. Eksperimen tersebut bertujuan untuk melakukan analisis sentimen terhadap sentimen setelah pilpres 2024, di mana sering kali terdapat perbedaan pendapat walau pilpres itu sendiri sudah selesai. Terdapat perbedaan pemahaman di kalangan masyarakat terkait peraturan tersebut yang menyebabkan kontroversi. Data yang digunakan bersumber dari Twitter dengan total 4260 data yang merupakan gabungan dari 3 dataset setelah dilakukan penghapusan yang sama tersisa 4183 data yang dimana berisi 3 dataset yaitu Ganjar (353 Positif, 478 Netral dan 350 Negatif), Anies (592 Positif, 749 Netral dan 569 Negatif) dan Prabowo (667 Positif, 350 Netral dan 306 Negatif).
- b. Hasil eksperimen menunjukkan bahwa nilai akurasi tertinggi terdapat pada dataset Ganjar Pranowo dengan algoritma Logistic Regression (LR), sebesar 84,39%. Nilai presisi tertinggi juga terdapat pada dataset Ganjar Pranowo dengan algoritma LR, sebesar 84,92%. Selanjutnya, nilai recall tertinggi terdapat pada dataset Ganjar Pranowo dengan algoritma LR, sebesar 84,39%. Terakhir, nilai f-measure tertinggi terdapat pada algoritma LR, sebesar 81,52%. Dapat dikatakan bahwa algoritma LR merupakan algoritma yang paling cocok digunakan untuk analisis sentimen suatu topik, diikuti dengan algoritma

Support Vector Machine (SVM) yang nilainya tidak berbeda jauh.

- c. Hasil perbandingan pada penelitian ini dapat digunakan sebagai sumber untuk menentukan penggunaan algoritma klasifikasi dalam suatu topik tertentu di masa yang akan datang. Dengan mengetahui kinerja masing-masing algoritma pada kasus tertentu, peneliti atau praktisi dapat memilih algoritma yang paling sesuai untuk digunakan dalam analisis sentimen atau klasifikasi lainnya.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

- a. Bapak Viktor Handrianus Pranatawijaya, S.T., M.T. Selaku dosen pengajar mata kuliah Kecerdasan Buatan.
- b. Ibu Ressa Priskila, S.T., M.T. Selaku dosen pengajar mata kuliah Kecerdasan Buatan.

DAFTAR PUSTAKA

- [1] R. Vindua And A. U. Zailani, "Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python," *Jurikom (Jurnal Riset Komputer)*, Vol. 10, No. 2, P. 479, Apr. 2023, Doi: 10.30865/Jurikom.V10i2.5945.
- [2] L. Aji Andika And P. Amalia Nur Azizah, "Analisis Sentimen Masyarakat Terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 Pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," 2019.
- [3] A. N. Hasanah And B. N. Sari, "Analisis Sentimen Ulasan Pengguna Aplikasi Jasa Ojek Online Maxim Pada Google Play Dengan Metode Naïve Bayes Classifier," *Jurnal Informatika Dan Teknik Elektro Terapan*, Vol. 12, No. 1, Jan. 2024, Doi: 10.23960/Jitet.V12i1.3628.
- [4] F. V. Sari And A. Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *Jurnal Simetris*, Vol. 10, No. 2, 2019.
- [5] L. Ardiani, H. Sujaini, And T. Tursina, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan Di Kota Pontianak," *Jurnal Sistem Dan Teknologi Informasi (Justin)*, Vol. 8, No. 2, P. 183, Apr. 2020, Doi: 10.26418/Justin.V8i2.36776.

- [6] S. Algifari Rismawan, Y. Syahidin, P. Piksi Ganesha Bandung, J. Gatot Subroto No, And K. Batununggal, "Implementasi Website Berita Online Menggunakan Metode Crawling Data Dengan Bahasa Pemrograman Python," Vol. 10, No. 3, Pp. 167–178, 2023, [Online]. Available: <Http://Jurnal.Mdp.Ac.Id>
- [7] M. R. Fais Sya' Bani, U. Enri, And T. N. Padilah, "Analisis Sentimen Terhadap Bakal Calon Presiden 2024 Dengan Algoritme Naïve Bayes," *Jurikom (Jurnal Riset Komputer)*, Vol. 9, No. 2, P. 265, Apr. 2022, Doi: 10.30865/Jurikom.V9i2.3989.
- [8] V. Kalra And R. Aggarwal, "Importance Of Text Data Preprocessing & Implementation In Rapidminer," In *Proceedings Of The First International Conference On Information Technology And Knowledge Management*, Pti, Jan. 2018, Pp. 71–75. Doi: 10.15439/2017km46.
- [9] D. Abimanyu, E. Budianita, E. Pandu Cynthia, F. Yanto, P. Studi Teknik Informatika, And F. Sains Dan Teknologi, "Analisis Sentimen Akun Twitter Apex Legends Menggunakan Vader," *Jurnal Nasional Komputasi Dan Teknologi Informasi*, Vol. 5, No. 3, 2022, [Online]. Available: <Https://Techno.Kompas.Com>
- [10] N. C. Lengkong, O. Safitri, S. Machsus, Y. R. Putra, A. Syahadati, And R. Nooraeni, "Analisis Sentimen Penerapan Psbb Di Dki Jakarta Dan Dampaknya Terhadap Pergerakan Ihsg," *Jurnal Teknoinfo*, Vol. 15, No. 1, P. 20, Jan. 2021, Doi: 10.33365/Jti.V15i1.866.
- [11] D. Septiani And I. Isabela, "Sintesia: Jurnal Sistem Dan Teknologi Informasi Indonesia Analisis Term Frequency Inverse Document Frequency (Tf-Idf) Dalam Temu Kembali Informasi Pada Dokumen Teks".
- [12] A. M. Pravina, I. Cholissodin, And P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan Pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (Svm)," 2019. [Online]. Available: <Http://J-Ptiik.Ub.Ac.Id>
- [13] A. Muhammaddin and I. A. Sobari, "Analisis Sentimen Pada Ulasan Aplikasi Kredivo Dengan Algoritma SVM Dan NBC," *Jurnal Rekayasa Perangkat Lunak*, vol. 2, no. 2, 2021, [Online]. Available: <http://jurnal.bsi.ac.id/index.php/reputasi>