

DETEKSI *HATE SPEECH* PADA PEMILU 2024 MENGUNAKAN ALGORITMA *MACHINE LEARNING*

Ahmad Zamsuri¹, Syahtriatna D², Elvira Asril³

¹Program Studi Teknik Informatika, ^{2,3}Program Studi Sistem Informasi Fakultas Ilmu Komputer
Universitas Lancang Kuning

Jl. Yos Sudarso KM. 8 Rumbai, Pekanbaru, Riau, telp. 0811 753 2015

e-mail: ahmadzamsuri@unilak.ac.id, syahtriatna@unilak.ac.id, elvira@unilak.ac.id

Abstrak

Pemilihan umum menjadi salah satu momen krusial dalam suatu negara untuk menentukan arah dan kepemimpinan. Namun, dengan berkembangnya teknologi digital, munculnya *hate speech* atau ujaran kebencian dapat merusak proses demokratis tersebut. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem deteksi *hate speech* pada Pemilihan Umum 2024 menggunakan algoritma *Machine Learning*. Metode yang digunakan melibatkan pengumpulan data teks dari berbagai sumber online yang berkaitan dengan pemilu. Selanjutnya, dilakukan pemrosesan dan pemilihan fitur untuk membangun model deteksi. Algoritma *Machine Learning* seperti *Support Vector Machine* (SVM) dan *Naïve Bayes* akan diterapkan untuk melatih model tersebut dengan *dataset* yang telah diberi label. Pengujian dilakukan dengan menggunakan *dataset* uji yang terdiri dari data teks yang belum pernah dilihat oleh model sebelumnya. Performa model dievaluasi berdasarkan metrik seperti akurasi, presisi, dan *recall*. Hasil eksperimen menunjukkan bahwa sistem deteksi yang diusulkan mampu mengidentifikasi *hate speech* pada konten terkait pemilu dengan tingkat akurasi yang signifikan. Penelitian ini diharapkan dapat memberikan kontribusi dalam menjaga integritas proses pemilihan umum dengan mengurangi dampak negatif dari *hate speech*. Selain itu, penerapan algoritma *machine learning* dalam deteksi *hate speech* juga dapat menjadi dasar untuk pengembangan solusi yang lebih efektif dalam mengatasi tantangan serupa di masa depan.

Kata Kunci: *Hate Speech*, Pemilihan Umum 2024, Algoritma *Machine Learning*, Deteksi Ujaran Kebencian, *Support Vector Machine* dan *Naïve Bayes*

Abstract

Elections are one of the most crucial moments in a country to determine its direction and leadership. However, with the development of digital technology, the emergence of *hate speech* can damage the democratic process. Therefore, this research aims to develop a *hate speech* detection system for the 2024 General Election using *Machine Learning* algorithms. The method used involves collecting text data from various online sources related to the election. Next, processing and feature selection are performed to build a detection model. *Machine Learning* algorithms such as *Support Vector Machine* (SVM) and *Naïve Bayes* will be applied to train the model with the labeled *dataset*. Testing is done using a test *dataset* consisting of text data that has never been seen by the model before. The performance of the model is evaluated based on metrics such as accuracy, precision, and *recall*. The experimental results show that the proposed detection system is able to identify *hate speech* in election-related content with a significant level of accuracy. This research is expected to contribute to maintaining the integrity of the election process by reducing the negative impact of *hate speech*. In addition, the application of machine learning algorithms in *hate speech* detection can also be the basis for developing more effective solutions in addressing similar challenges in the future.

Keywords: *Hate Speech*, General Election 2024, Machine Learning Algorithm, *Hate Speech* Detection, *Support Vector Machine* and *Naïve Bayes*

1. PENDAHULUAN

Teknologi merupakan hal yang tidak asing bagi suatu individu. Berbagai macam teknologi telah tercipta, salah satunya adalah teknologi informasi dimana pada teknologi informasi hadir sebuah

istilah yang dinamakan dengan sosial media. Pada sosial media, pengguna dapat bertukar ataupun mendapatkan informasi secara bebas. Perkembangan sosial media yang pesat dapat mempengaruhi sikap dan pola perilaku masyarakat [1]. Dengan perkembangan sosial media yang begitu pesat membawa tantangan besar, salah satunya adalah penyebaran *hate speech*. *Hate speech* merujuk pada penggunaan bahasa yang agresif, kasar atau bersifat menyerang yang menargetkan kelompok atau orang tertentu. Meskipun Sebagian besar media sosial melarang pernyataan *hate speech*, ukuran jaringan dari sosial media membuatnya hampir tidak mungkin untuk mengendalikan semua kontennya [1].

Twitter cukup populer dan sering digunakan di Indonesia, pengguna twitter dapat berekspresi dan berinspirasi tanpa adanya batasan. Hal tersebut menjadikan twitter sebagai salah satu wadah dalam menempatkan sumber data text [2]. Oleh karena itu diperlukan adanya deteksi *hate speech* dan menyaring konten yang mengandung bahasa kasar atau bahasa yang memprovokasi kebencian yang berada didalam media sosial Twitter menggunakan algoritma *machine learning* yaitu *Support Vector Machine* (SVM) dan *Naïve Bayes*.

Support Vector Machine (SVM) merupakan sebuah metode pembelajaran mesin yang dipopulerkan oleh Boser, Guyon, dan Vapnik pada tahun 1992. Proses pembelajaran pada SVM bertujuan untuk mendapatkan hipotesis berupa bidang pemisah (*hyperplane*) terbaik yang tidak hanya meminimalkan empirical risk yaitu rata-rata error pada data pelatihan, tetapi juga memiliki generalisasi yang baik [3]. Untuk menjamin generalisasi ini, SVM bekerja berdasarkan prinsip *structural risk minimization* (SRM). SRM pada SVM digunakan untuk menjamin batas atas dari generalisasi pada data pengujian dengan cara mengontrol “kapasitas” (fleksibilitas) dari hipotesis hasil pembelajaran [4].

Naïve Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan [5]. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Definisi lain mengatakan *Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya [6] sepiansyah.

Dari penelitian ini penulis menemukan beberapa referensi dari penelitian sebelumnya yang telah menggunakan algoritma *Support Vector Machine* (SVM) adalah penelitian yang membandingkan metode *vectorizer* untuk menganalisis sentimen pada YouTube. Penelitian ini menggunakan 2 algoritma yaitu *Support Vector Machine* (SVM) yang menghasilkan akurasi tertinggi yaitu 97,5% dan algoritma *K-Nearest Neighbor* (KNN) yang menghasilkan akurasi tertinggi yaitu 90,9% (inter svm-knn 3). Penelitian selanjutnya yaitu penelitian yang melakukan analisis sentiment terhadap ulasan dalam bahasa Tionghoa. Penelitian ini menggunakan 2 algoritma yaitu algoritma *Support Vector Machine* (SVM) yang menghasilkan akurasi 78,87% dan menggunakan *Semantic Orientation* menghasilkan akurasi 73,33% (inter svm-semantic). Penelitian lainnya yaitu melakukan analisis sentimen terhadap pemilihan presiden Indonesia pada tahun 2019 menggunakan algoritma *Support Vector Machine* (SVM) yang menghasilkan rata-rata akurasi 69,27% dan algoritma *K-Nearest Neighbor* (K-NN) yang menghasilkan rata-rata akurasi 61,3% (inter svm-knn).

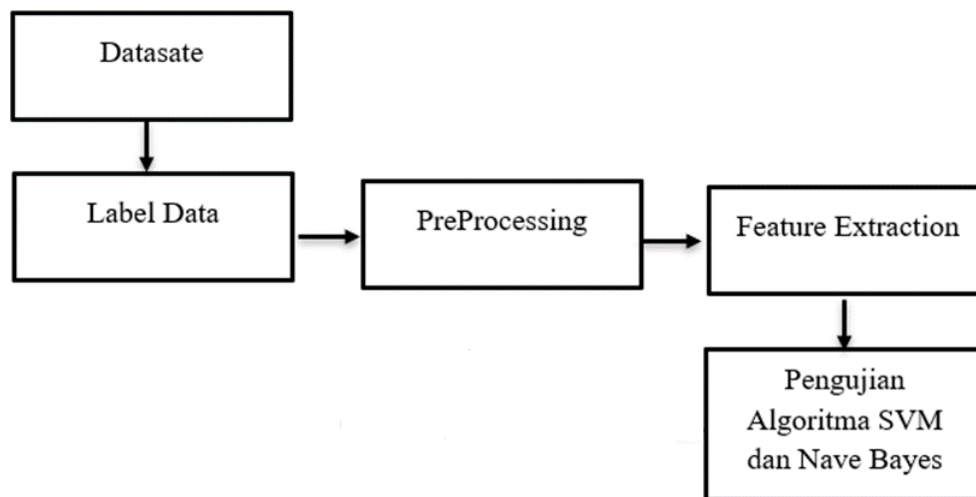
Selain algoritma *Support Vector Machine* (SVM), penelitian ini juga menggunakan algoritma *Naïve Bayes*. Penelitian terdahulu yang menggunakan algoritma *Naïve Bayes* adalah penelitian yang melakukan analisis sentiment terhadap kasus anti LGBT di Indonesia menggunakan algoritma *Naïve Bayes* yang menghasilkan akurasi sebesar 83,43%, dan algoritma *Decision Tree* menghasilkan akurasi sebesar 82,91% yang mana hasil tersebut sama dengan akurasi yang dihasilkan dengan menggunakan algoritma *Random Forest* (inter nb-dt-rf). Penelitian selanjutnya melakukan analisis sentiment mengenai aplikasi ulasan tentang aplikasi PeduliLindungi yang menggunakan algoritma *Naïve Bayes* yang menghasilkan akurasi sebesar 90,12% dengan penambahan Adaboost dan algoritma *K-Nearest Neighbor* (K-NN) yang menghasilkan akurasi 86,15% dengan penambahan Adaboost (inter nb-dt+ab). Penelitian lainnya melakukan analisis sentiment terhadap tweet mengenai E-commerce menggunakan algoritma *Decision Tree* yang menghasilkan presisi 79,96%, algoritma *K-Nearest Neighbor* (K-NN) yang menghasilkan presisi 85,67% dan algoritma *Naïve Bayes* menghasilkan presisi sebesar 88,50% (inter nb-dt-knn). Berdasarkan hasil penelitian terdahulu, algoritma *Support*

Vector Machine (SVM) dan *Naïve Bayes* mendapatkan hasil yang lebih tinggi daripada algoritma lainnya dengan data yang berbeda-beda. Sehingga, penelitian ini akan melihat bagaimana performa algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* dalam mendeteksi hate speech pada tweet mengenai Pemilu 2024.

Dengan demikian penelitian ini menggunakan Algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* untuk mendeteksi *hate speech* terkait Pemilu 2024 di media sosial Twitter. Tujuan penelitian ini adalah untuk mengetahui tingkat akurasi algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* dalam mendeteksi ujaran kebencian pada tweet mengenai Pemilu 2024 serta meningkatkan akurasi deteksi dengan menggabungkan (*hybrid*) algoritma SVM dan *Naïve Bayes*.

2. METODE PENELITIAN

Penelitian ini membutuhkan metodologi penelitian yang terstruktur, mencakup berbagai kegiatan dan proses penelitian secara sistematis. Tahapan penelitian menjelaskan secara rinci bagaimana penelitian dilaksanakan dari awal hingga akhir. Proses ini diatur oleh sebuah kerangka kerja yang diilustrasikan dalam Gambar 1. dengan memanfaatkan kerangka kerja tersebut, penelitian dapat dilaksanakan secara lebih teratur dan sesuai dengan rencana yang telah ditetapkan. Gambar 1. merupakan metodologi penelitian yang digunakan pada penelitian ini.



Gambar 1. Kerangka penelitian yang diusulkan

Kerangka penelitian merupakan diagram yang menjelaskan secara garis besar alur logika berjalannya penelitian. Penelitian ini secara garis besar dibagi ke dalam tiga (Gambar 1.): *Input*, *Proses* dan *Output*. *Input* dari penelitian adalah tipe data sequential, khususnya Dataset Teks. *Proses* menyajikan algoritma SVM dan *Naïve* yang digunakan dalam proses klasifikasi dan pendeteksian *outlier* pada dataset teks. *Output* dari penelitian ini berupa berita/informasi yang menyimpang dari *dataset* teks pada media sosial Twitter.

1.1. Dataset

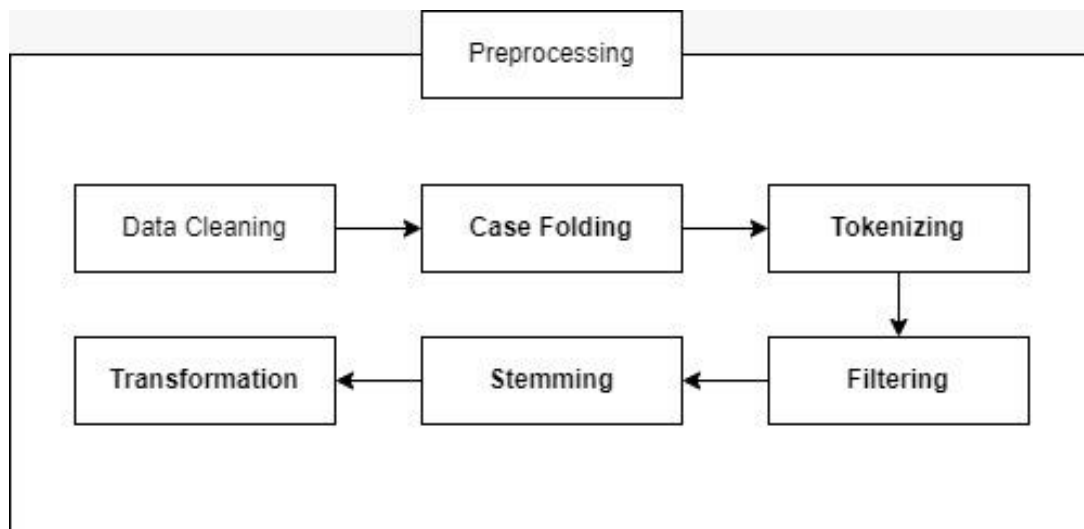
Dataset pada penelitian ini diambil dari Twitter dengan tema pemilu 2024, yang dimulai dari 6 januari sampai 30 april. *Dataset* ini digunakan untuk melihat reaksi seseorang dalam menggambarkan tokoh atau calon yang dikomentari. Reaksi pada penelitian ini merupakan komentar netizen yang berbentuk teks yang diambil dari Tweet kemudian diberi label berdasarkan emosi seperti marah, senang, sedih, dan lain sebagainya [7].

1.2. Label Data

Pelabelan data pada penelitian ini terbagi menjadi 2 yaitu Positif dan Negatif.

1.3. Pre-processing

Proses *Pre-Processing* dalam *Natural Language Processing* (NLP) adalah tahapan awal yang dilakukan untuk membersihkan, mengubah, dan mempersiapkan data teks sebelum diolah lebih lanjut. *Natural Language Processing* (NLP) merupakan teknologi machine learning yang memberikan kemampuan komputer untuk menginterpretasikan, memanipulasi, dan memahami bahasa manusia yang digunakan untuk otomatis memproses data, menganalisis maksud atau sentimen dalam pesan, dan merespons komunikasi manusia secara real-time [8]. Tujuan dari proses ini adalah untuk meningkatkan kualitas data teks agar dapat diolah dengan lebih optimal oleh model NLP. *Pre-Processing* pada penelitian ini menggunakan 6 tahap, yang digambarkan pada gambar 2.



Gambar 2. *Pre-Processing*

1.1.1. Data Cleaning.

Pada proses *data cleaning*, data akan dibersihkan dengan menghapus *mention*, *hashtag*, *retweet*, *symbol*, *link*, angka, mengganti baris baru menjadi spasi, dan menghapus spasi dari kiri dan kanan teks.

1.1.2. Case Folding

Karena data yang dimiliki belum terstruktur dalam penggunaan huruf kapital, maka akan digunakan *case folding* untuk menyamaratakan penggunaan huruf kapital yang mana akan mengubah semua kata menjadi huruf kecil atau *lowercase*.

1.1.3. Tokenizing

Data yang sudah dikumpulkan terdiri dari kalimat. Untuk mempermudah melakukan proses analisis kalimat akan dipecah atau dipisah menjadi kata-kata. Tahap ini disebut dengan *tokenizing*.

1.1.4. Filtering

Dilanjutkan setelah tahap *tokenizing*. Pada tahap *filtering* ini akan dilakukan penghapusan kata-kata yang dianggap tidak penting.

1.1.5. Stemming

Pada proses *stemming* ini akan menghilangkan imbuhan dan akhiran yang ada pada setiap kata, sehingga setiap kata-kata yang ada menjadi kata dasar. *Library* pada proses ini menggunakan *Library* Sastrawi.

1.1.6. Transformation

Tahap terakhir yang dilakukan pada *preprocessing* ini adalah tahap *transformation* yang melakukan proses transformasi label yang sebelumnya dalam bentuk kata menjadi numerik. Pada proses ini dilakukan transformasi data menggunakan *Label Encoder* agar dapat mengubah label secara otomatis.

1.4. Feature Extraction

Feature Extraction adalah teknik pengambilan ciri / feature dari suatu bentuk yang nantinya nilai yang didapatkan akan dianalisis untuk proses selanjutnya. Klasifikasi adalah proses untuk menyatakan suatu objek ke dalam salah satu kategori yang sudah didefinisikan sebelumnya. Tahap ini, menggunakan 2 *feature extraction* yang digunakan untuk melihat akurasi yang dihasilkan, yaitu TF-IDF dan *Bag of Words* (BoW). Keduanya sering digunakan oleh peneliti terdahulu pada *machine learning*.

3. HASIL DAN PEMBAHASAN

Hasil dan pembahasan pada penelitian ini melakukan pemodelan yang terdiri dari dua algoritma. Algoritma yang digunakan yaitu Algoritma *Naïve Bayes* dan Algoritma *Support Vector Machine* (SVM). *Support Vector Machines* (SVM) adalah suatu metode yang handal dalam menyelesaikan masalah klasifikasi data [9] dan Metode *Bayes* merupakan pendekatan *statistic* untuk melakukan inferensi induksi pada persoalan klasifikasi [10]. Berikut ini merupakan hasil yang didapatkan dari penelitian ini berdasarkan dua label yakni positif dan negatif sebagai berikut:

3.1. Data 2 Label

Gambar 3. merupakan sample data yang digunakan untuk pengujian terhadap *datasets*. Pengujian pada penelitian Ini menggunakan *splitting* data 80:20.

	KALIMAT	KATEGORI EMOSI
0	respon cepat macet akibat baik jembatan juwana...	negatif
1	keren gubernur jawa ganjar pranowo hasil bawa ...	negatif
2	baik hasil survei litbang Kompas Januari elekt...	positif
3	temu lembaga survei median kait elektabilitas ...	positif
4	jateng ganjar pranowo ajak insan pers kolabora...	positif
...	...	...
1643	Lambe Waras ke Dedek Prayudi PSI Mangap Kalau...	negatif
1644	Forum Komunikasi antar simpul relawan Anies Ba...	negatif
1645	Anies itu py Value diyakini dpt meyakinkan inv...	negatif
1646	Pemilu di Negara ManaTps mana yg mau dijaga	negatif
1647	IPO dan Voxpol sama mencatat keunggulan Anies-...	positif

1648 rows × 2 columns

Gambar 3. Sample Data 2 label

Gambar 3. merupakan hasil luaran dari data yang telah dilakukan proses selanjutnya, yaitu melakukan *preprocessing*. *Preprocessing* (pra-pemrosesan) adalah langkah penting dalam analisis sentimen menggunakan *Python*. Tujuan dari *preprocessing* adalah membersihkan dan mengubah data teks mentah menjadi bentuk yang lebih cocok untuk analisis.

3.2. Preprocessing 2 Label

Berikut ini merupakan *preprocessing* yang digunakan untuk dua label. *Preprocessing* pada penelitian ini menggunakan *data cleaning*. *Data cleaning* adalah proses memastikan keakuratan, konsistensi, dan kegunaan data dalam kumpulan data. Prosesnya yang dilakukan adalah mendeteksi kesalahan data atau data yang *corrupt* dan memperbaiki atau menghapus data sesuai kebutuhan.

3.2.1. Case Folding

Kemudian *case folding* bertujuan untuk menyamaratakan penggunaan huruf kapital. Gambar 4. merupakan hasil dari *case folding* yang telah dilakukan.

	KALIMAT	KATEGORI EMOSI
0	respon cepat macet akibat baik jembatan juwana...	negatif
1	keren gubernur jawa ganjar pranowo hasil bawa ...	negatif
2	baik hasil survei litbang Kompas Januari elekt...	positif
3	temu lembaga survei median kait elektabilitas ...	positif
4	Jateng Ganjar Pranowo Ajak Insan Pers Kolabora...	positif
...	...	...
1643	Lambe Waras ke Dedek Prayudi PSI Mangap Kalau...	negatif
1644	Forum Komunikasi antar simpul relawan Anies Ba...	negatif
1645	Anies itu py Value diyakini dpt meyakinkan inv...	negatif
1646	Pemilu di Negara ManaTps mana yg mau dijaga	negatif
1647	IPO dan Voxpol sama mencatat keunggulan Anies-...	positif

1648 rows × 2 columns

Gambar 4. Case Folding 2 Label

Pada Gambar 4. terlihat bahwa seluruh huruf terlihat menjadi huruf kecil semua. Konsistensi: Dengan mengubah semua huruf menjadi huruf kecil, menghilangkan perbedaan antara kata-kata yang ditulis dalam huruf besar dan huruf kecil. Ini membantu dalam memastikan konsistensi dalam data teks. *Case folding* memiliki beberapa manfaat seperti pengurangan dimensi, dalam analisis teks, beberapa kata mungkin muncul dalam berbagai bentuk kapitalisasi. Dengan melakukan *case folding*, dapat mengurangi dimensi data karena hanya perlu menganggap setiap kata dalam satu bentuk (huruf kecil). Selanjutnya kemudahan pencarian, ketika mencari kata dalam teks, mengubah semuanya menjadi huruf kecil memungkinkan pencarian yang lebih mudah dan konsisten. Kemudian analisis konsisten, *Case folding* membantu dalam menghindari ambiguitas dalam analisis, di mana kata-kata yang sama tetapi ditulis dengan huruf besar dan kecil dianggap sebagai dua kata yang berbeda.

3.2.2. Tokenizing

Selanjutnya *Tokenizing* digunakan untuk memudahkan proses analisis data kita harus memecah kalimat-kalimat tersebut menjadi kata atau disebut dengan token. Dengan *tokenizing* kita dapat membedakan mana antara pemisah kata atau bukan. Gambar 5. merupakan *tokenizing* yang dihasilkan dari program *python*.

	KALIMAT	KATEGORI EMOSI
0	[respon, cepat, macet, akibat, baik, jembatan, ...]	negatif
1	[keren, gubernur, jawa, ganjar, pranowo, hasil, ...]	negatif
2	[baik, hasil, survei, litbang, Kompas, Januari, ...]	positif
3	[temu, lembaga, survei, median, kait, elektabi, ...]	positif
4	[Jateng, ganjar, pranowo, ajak, insan, pers, k, ...]	positif
...	...	...
1643	[lambe, waras, ke, dedek, prayudi, psi, mangap, ...]	negatif
1644	[forum, komunikasi, antar, simpul, relawan, an, ...]	negatif
1645	[anies, itu, py, value, diyakini, dpt, meyakini, ...]	negatif
1646	[pemilu, di, negara, manatps, mana, yg, mau, d, ...]	negatif
1647	[ipo, dan, voxpol, sama, mencatat, keunggulan, ...]	positif

1648 rows × 2 columns

Gambar 5. Tokenizing 2 label

Dari Gambar 5. terlihat bahwa seluruh kata dipenggal menggunakan koma (.). Tujuan dari *Tokenizing*, yaitu untuk mengeksplorasi kata dalam suatu kalimat. *Tokenizing* dilakukan untuk memecahkan sebuah kata dan melakukan penghapusan terhadap delimiter serta karakter angka bersama entitas *tweet* seperti halnya *hashtag*, *retweet* dan *mention*.

3.2.3. Filtering

Kemudian adalah *filtering* yang digunakan untuk untuk memisahkan dan menyaring data yang relevan atau tidak relevan dalam konteks tugas atau masalah yang sedang dipecahkan. Gambar 6. merupakan hasil *filtering* pada penelitian ini.

	KALIMAT	KATEGORI EMOSI
0	respon cepat macet akibat jembatan juwana ganj...	negatif
1	keren gubernur jawa ganjar pranowo hasil bawa ...	negatif
2	hasil survei litbang Kompas Januari elektabili...	positif
3	temu lembaga survei median kait elektabilitas ...	positif
4	jateng ganjar pranowo ajak insan pers kolabora...	positif
...	...	...
1643	lambe waras dedek prayudi psi mangap anies bag...	negatif
1644	forum komunikasi simpul relawan anies baswedan ...	negatif
1645	anies py value diyakini dpt investor u berinve...	negatif
1646	pemilu negara manatps yg dijaga	negatif
1647	ipo voxpol mencatat keunggulan anies-ahy simul...	positif

1648 rows × 2 columns

Gambar 6. Filtering 2 label

Dari Gambar 4.4 terlihat bahwa penggunaan kata sambung tidak terlihat lagi. Hal ini dilakukan karena penggunaan kata sambung seperti di, ke, dari dan lain sebagainya tidak dianggap penting pada penelitian ini.

3.2.4. Stemming

Tahap ini digunakan untuk untuk mengurangi kata-kata dalam teks menjadi bentuk dasar atau kata dasar. Tujuannya adalah untuk menghapus akhiran atau awalan kata sehingga kata-kata yang memiliki akar yang sama akan direpresentasikan dengan bentuk yang seragam. Berikut merupakan hasil *stemming* yang terlihat pada Gambar 7.

	KALIMAT	KATEGORI EMOSI
0	respon cepat macet akibat jembatan juwana ganj...	negatif
1	keren gubernur jawa ganjar pranowo hasil bawa ...	negatif
2	hasil survei litbang Kompas Januari elektabili...	positif
3	temu lembaga survei median kait elektabilitas ...	positif
4	jateng ganjar pranowo ajak insan pers kolabora...	positif
...	...	...
1643	lambe waras dedek prayudi psi mangap anies bag...	negatif
1644	forum komunikasi simpul rawan anies baswedan b...	negatif
1645	anies py value yakin dpt investor u investasi ...	negatif
1646	milu negara manatps yg jaga	negatif
1647	ipo voxpol catat unggul anies-ahy simulasi pil...	positif

1648 rows × 2 columns

Gambar 7. Stemming 2 label

Dari Gambar 4.5 terlihat bahwa seluruh kata menjadi kata dasar seluruhnya. Alasan dilakukan proses *stemming* ini adalah mengurangi jumlah *noise* dalam data dan pada beberapa kesempatan mengurangi beban kerja dari mesin. Ada dua *library* yang digunakan pada penelitian ini yaitu NLTK (*Natural Language Tool Kit*) dan Sastrawi (Bahasa Indonesia).

3.2.5. Transformation (Label Encoder)

Kemudian tahap dari *preprocessing* yang terakhir adalah *transformation* yang digunakan untuk mengubah *variable* kategori menjadi bentuk numerik. Ini diperlukan karena *variable* besar algoritma machine learning hanya dapat memproses data numerik. *Label Encoder* memberikan label numerik unik untuk setiap kategori dalam suatu *variable*. *Label Encoder* hanya mengubah *variable* kategori menjadi bentuk numerik, dan tidak memberikan interpretasi langsung terhadap arti atau makna dari setiap kategori. Gambar 8. merupakan hasil *encoder* pada penelitian ini.

	KALIMAT	KATEGORI EMOSI
0	respon cepat macet akibat jembatan juwana ganj...	0
1	keren gubernur jawa ganjar pranowo hasil bawa ...	0
2	hasil survei litbang Kompas Januari elektabili...	1
3	temu lembaga survei median kait elektabilitas ...	1
4	jateng ganjar pranowo ajak insan pers kolabora...	1
...	...	...
1643	lambe waras dedek prayudi psi mangap anies bag...	0
1644	forum komunikasi simpul rawan anies baswedan b...	0
1645	anies py value yakin dpt investor u investasi ...	0
1646	milu negara manatps yg jaga	0
1647	ipo voxpol catat unggul anies-ahy simulasi pil...	1

1648 rows × 2 columns

Gambar 8. *Transformation (Label Encoder) 2 Label*

Dari Gambar 4.6 terlihat bahwa seluruh label pada penelitian ini berubah menjadi numerik. positif menjadi 0, negatif menjadi 1.

3.3. Preprocessing 2 label

Proses akurasi pada 2 label yang terlihat pada gambar 9. merupakan *preprocessing* yang dilakukan pada 2 label dengan menggabungkan *preprocessing*.

<table> <tr> <th>KALIMAT</th><th>KATEGORI EMOSI</th></tr> <tr><td>0</td><td>respon cepat macet akibat baik jembatan juwana...</td></tr> <tr><td>1</td><td>keren gubernur jawa ganjar pranowo hasil bawa ...</td></tr> <tr><td>2</td><td>baik hasil survei litbang Kompas Januari elekt...</td></tr> <tr><td>3</td><td>temu lembaga survei median kait elektabilitas ...</td></tr> <tr><td>4</td><td>jateng ganjar pranowo ajak insan pers kolabora...</td></tr> <tr><td>...</td><td>...</td></tr> <tr><td>1643</td><td>lambe waras ke dedek prayudi PSI Mangap Kalau ...</td></tr> <tr><td>1644</td><td>Forum Komunikasi antar simpul melawan Anies Ba...</td></tr> <tr><td>1645</td><td>Anies itu py Value diyakini dpt meyakinkan inv...</td></tr> <tr><td>1646</td><td>Pemilu di Negara ManaTps mana yg mau dijaga</td></tr> <tr><td>1647</td><td>IPO dan Voxpol sama mencatat keunggulan Anies-...</td></tr> </table> 1648 rows × 2 columns Data Cleaning	KALIMAT	KATEGORI EMOSI	0	respon cepat macet akibat baik jembatan juwana...	1	keren gubernur jawa ganjar pranowo hasil bawa ...	2	baik hasil survei litbang Kompas Januari elekt...	3	temu lembaga survei median kait elektabilitas ...	4	jateng ganjar pranowo ajak insan pers kolabora...	...	...	1643	lambe waras ke dedek prayudi PSI Mangap Kalau ...	1644	Forum Komunikasi antar simpul melawan Anies Ba...	1645	Anies itu py Value diyakini dpt meyakinkan inv...	1646	Pemilu di Negara ManaTps mana yg mau dijaga	1647	IPO dan Voxpol sama mencatat keunggulan Anies-...	<table> <tr> <th>KALIMAT</th><th>KATEGORI EMOSI</th></tr> <tr><td>0</td><td>respon cepat macet akibat baik jembatan juwana...</td></tr> <tr><td>1</td><td>keren gubernur jawa ganjar pranowo hasil bawa ...</td></tr> <tr><td>2</td><td>baik hasil survei litbang Kompas Januari elekt...</td></tr> <tr><td>3</td><td>temu lembaga survei median kait elektabilitas ...</td></tr> <tr><td>4</td><td>jateng ganjar pranowo ajak insan pers kolabora...</td></tr> <tr><td>...</td><td>...</td></tr> <tr><td>1643</td><td>lambe waras ke dedek prayudi psi mangap kalau...</td></tr> <tr><td>1644</td><td>forum komunikasi antar simpul melawan anies ba...</td></tr> <tr><td>1645</td><td>anies itu py value diyakini dpt meyakinkan inv...</td></tr> <tr><td>1646</td><td>pemilu di negara manatps mana yg mau dijaga</td></tr> <tr><td>1647</td><td>ipo dan voxpol sama mencatat keunggulan anies-...</td></tr> </table> 1648 rows × 2 columns Case Folding	KALIMAT	KATEGORI EMOSI	0	respon cepat macet akibat baik jembatan juwana...	1	keren gubernur jawa ganjar pranowo hasil bawa ...	2	baik hasil survei litbang Kompas Januari elekt...	3	temu lembaga survei median kait elektabilitas ...	4	jateng ganjar pranowo ajak insan pers kolabora...	...	...	1643	lambe waras ke dedek prayudi psi mangap kalau...	1644	forum komunikasi antar simpul melawan anies ba...	1645	anies itu py value diyakini dpt meyakinkan inv...	1646	pemilu di negara manatps mana yg mau dijaga	1647	ipo dan voxpol sama mencatat keunggulan anies-...
KALIMAT	KATEGORI EMOSI																																																
0	respon cepat macet akibat baik jembatan juwana...																																																
1	keren gubernur jawa ganjar pranowo hasil bawa ...																																																
2	baik hasil survei litbang Kompas Januari elekt...																																																
3	temu lembaga survei median kait elektabilitas ...																																																
4	jateng ganjar pranowo ajak insan pers kolabora...																																																
...	...																																																
1643	lambe waras ke dedek prayudi PSI Mangap Kalau ...																																																
1644	Forum Komunikasi antar simpul melawan Anies Ba...																																																
1645	Anies itu py Value diyakini dpt meyakinkan inv...																																																
1646	Pemilu di Negara ManaTps mana yg mau dijaga																																																
1647	IPO dan Voxpol sama mencatat keunggulan Anies-...																																																
KALIMAT	KATEGORI EMOSI																																																
0	respon cepat macet akibat baik jembatan juwana...																																																
1	keren gubernur jawa ganjar pranowo hasil bawa ...																																																
2	baik hasil survei litbang Kompas Januari elekt...																																																
3	temu lembaga survei median kait elektabilitas ...																																																
4	jateng ganjar pranowo ajak insan pers kolabora...																																																
...	...																																																
1643	lambe waras ke dedek prayudi psi mangap kalau...																																																
1644	forum komunikasi antar simpul melawan anies ba...																																																
1645	anies itu py value diyakini dpt meyakinkan inv...																																																
1646	pemilu di negara manatps mana yg mau dijaga																																																
1647	ipo dan voxpol sama mencatat keunggulan anies-...																																																
<table> <tr> <th>KALIMAT</th><th>KATEGORI EMOSI</th></tr> <tr><td>0</td><td>[respon, cepat, macet, akibat, baik, jembatan,...</td></tr> <tr><td>1</td><td>[keren, gubernur, jawa, ganjar, pranowo, hasil...</td></tr> <tr><td>2</td><td>[baik, hasil, survei, litbang, Kompas, Januari...</td></tr> <tr><td>3</td><td>[temu, lembaga, survei, median, kait, elektabi...</td></tr> <tr><td>4</td><td>[jateng, ganjar, pranowo, ajak, insan, pers, k...</td></tr> <tr><td>...</td><td>...</td></tr> <tr><td>1643</td><td>[lambe, waras, ke, dedek, prayudi, psi, mangap...</td></tr> <tr><td>1644</td><td>[forum, komunikasi, antar, simpul, melawan, an...</td></tr> <tr><td>1645</td><td>[anies, itu, py, value, diyakini, dpt, meyakin...</td></tr> <tr><td>1646</td><td>[pemilu, di, negara, manatps, mana, yg, mau, d...</td></tr> <tr><td>1647</td><td>[ipo, dan, voxpol, sama, mencatat, keunggulan, ...</td></tr> </table> 1648 rows × 2 columns Tokenizing	KALIMAT	KATEGORI EMOSI	0	[respon, cepat, macet, akibat, baik, jembatan,...	1	[keren, gubernur, jawa, ganjar, pranowo, hasil...	2	[baik, hasil, survei, litbang, Kompas, Januari...	3	[temu, lembaga, survei, median, kait, elektabi...	4	[jateng, ganjar, pranowo, ajak, insan, pers, k...	...	...	1643	[lambe, waras, ke, dedek, prayudi, psi, mangap...	1644	[forum, komunikasi, antar, simpul, melawan, an...	1645	[anies, itu, py, value, diyakini, dpt, meyakin...	1646	[pemilu, di, negara, manatps, mana, yg, mau, d...	1647	[ipo, dan, voxpol, sama, mencatat, keunggulan, ...	<table> <tr> <th>KALIMAT</th><th>KATEGORI EMOSI</th></tr> <tr><td>0</td><td>respon cepat macet akibat jembatan juwana ganj...</td></tr> <tr><td>1</td><td>keren gubernur jawa ganjar pranowo hasil bawa ...</td></tr> <tr><td>2</td><td>hasil survei litbang Kompas Januari elektabili...</td></tr> <tr><td>3</td><td>temu lembaga survei median kait elektabilitas ...</td></tr> <tr><td>4</td><td>jateng ganjar pranowo ajak insan pers kolabora...</td></tr> <tr><td>...</td><td>...</td></tr> <tr><td>1643</td><td>lambe waras dedek prayudi psi mangap anies bag...</td></tr> <tr><td>1644</td><td>forum komunikasi simpul melawan anies baswedan ...</td></tr> <tr><td>1645</td><td>anies py value diyakini dpt investor u berinve...</td></tr> <tr><td>1646</td><td>pemilu negara manatps yg jaga</td></tr> <tr><td>1647</td><td>ipo voxpol mencatat keunggulan anies-ahy simul...</td></tr> </table> 1648 rows × 2 columns Filtering	KALIMAT	KATEGORI EMOSI	0	respon cepat macet akibat jembatan juwana ganj...	1	keren gubernur jawa ganjar pranowo hasil bawa ...	2	hasil survei litbang Kompas Januari elektabili...	3	temu lembaga survei median kait elektabilitas ...	4	jateng ganjar pranowo ajak insan pers kolabora...	...	...	1643	lambe waras dedek prayudi psi mangap anies bag...	1644	forum komunikasi simpul melawan anies baswedan ...	1645	anies py value diyakini dpt investor u berinve...	1646	pemilu negara manatps yg jaga	1647	ipo voxpol mencatat keunggulan anies-ahy simul...
KALIMAT	KATEGORI EMOSI																																																
0	[respon, cepat, macet, akibat, baik, jembatan,...																																																
1	[keren, gubernur, jawa, ganjar, pranowo, hasil...																																																
2	[baik, hasil, survei, litbang, Kompas, Januari...																																																
3	[temu, lembaga, survei, median, kait, elektabi...																																																
4	[jateng, ganjar, pranowo, ajak, insan, pers, k...																																																
...	...																																																
1643	[lambe, waras, ke, dedek, prayudi, psi, mangap...																																																
1644	[forum, komunikasi, antar, simpul, melawan, an...																																																
1645	[anies, itu, py, value, diyakini, dpt, meyakin...																																																
1646	[pemilu, di, negara, manatps, mana, yg, mau, d...																																																
1647	[ipo, dan, voxpol, sama, mencatat, keunggulan, ...																																																
KALIMAT	KATEGORI EMOSI																																																
0	respon cepat macet akibat jembatan juwana ganj...																																																
1	keren gubernur jawa ganjar pranowo hasil bawa ...																																																
2	hasil survei litbang Kompas Januari elektabili...																																																
3	temu lembaga survei median kait elektabilitas ...																																																
4	jateng ganjar pranowo ajak insan pers kolabora...																																																
...	...																																																
1643	lambe waras dedek prayudi psi mangap anies bag...																																																
1644	forum komunikasi simpul melawan anies baswedan ...																																																
1645	anies py value diyakini dpt investor u berinve...																																																
1646	pemilu negara manatps yg jaga																																																
1647	ipo voxpol mencatat keunggulan anies-ahy simul...																																																
<table> <tr> <th>KALIMAT</th><th>KATEGORI EMOSI</th></tr> <tr><td>0</td><td>respon cepat macet akibat jembatan juwana ganj...</td></tr> <tr><td>1</td><td>keren gubernur jawa ganjar pranowo hasil bawa ...</td></tr> <tr><td>2</td><td>hasil survei litbang Kompas Januari elektabili...</td></tr> <tr><td>3</td><td>temu lembaga survei median kait elektabilitas ...</td></tr> <tr><td>4</td><td>jateng ganjar pranowo ajak insan pers kolabora...</td></tr> <tr><td>...</td><td>...</td></tr> <tr><td>1643</td><td>lambe waras dedek prayudi psi mangap anies bag...</td></tr> <tr><td>1644</td><td>forum komunikasi simpul melawan anies baswedan b...</td></tr> <tr><td>1645</td><td>anies py value yakin dpt investor u investasi ...</td></tr> <tr><td>1646</td><td>milu negara manatps yg jaga</td></tr> <tr><td>1647</td><td>ipo voxpol catat unggul anies-ahy simulasi pil...</td></tr> </table> 1648 rows × 2 columns Stemming	KALIMAT	KATEGORI EMOSI	0	respon cepat macet akibat jembatan juwana ganj...	1	keren gubernur jawa ganjar pranowo hasil bawa ...	2	hasil survei litbang Kompas Januari elektabili...	3	temu lembaga survei median kait elektabilitas ...	4	jateng ganjar pranowo ajak insan pers kolabora...	...	...	1643	lambe waras dedek prayudi psi mangap anies bag...	1644	forum komunikasi simpul melawan anies baswedan b...	1645	anies py value yakin dpt investor u investasi ...	1646	milu negara manatps yg jaga	1647	ipo voxpol catat unggul anies-ahy simulasi pil...	<table> <tr> <th>KALIMAT</th><th>KATEGORI EMOSI</th></tr> <tr><td>0</td><td>respon cepat macet akibat jembatan juwana ganj... 0</td></tr> <tr><td>1</td><td>keren gubernur jawa ganjar pranowo hasil bawa ... 0</td></tr> <tr><td>2</td><td>hasil survei litbang Kompas Januari elektabili... 1</td></tr> <tr><td>3</td><td>temu lembaga survei median kait elektabilitas ... 1</td></tr> <tr><td>4</td><td>jateng ganjar pranowo ajak insan pers kolabora... 1</td></tr> <tr><td>...</td><td>...</td></tr> <tr><td>1643</td><td>lambe waras dedek prayudi psi mangap anies bag... 0</td></tr> <tr><td>1644</td><td>forum komunikasi simpul melawan anies baswedan b... 0</td></tr> <tr><td>1645</td><td>anies py value yakin dpt investor u investasi ... 0</td></tr> <tr><td>1646</td><td>milu negara manatps yg jaga 0</td></tr> <tr><td>1647</td><td>ipo voxpol catat unggul anies-ahy simulasi pil... 1</td></tr> </table> 1648 rows × 2 columns Transformation	KALIMAT	KATEGORI EMOSI	0	respon cepat macet akibat jembatan juwana ganj... 0	1	keren gubernur jawa ganjar pranowo hasil bawa ... 0	2	hasil survei litbang Kompas Januari elektabili... 1	3	temu lembaga survei median kait elektabilitas ... 1	4	jateng ganjar pranowo ajak insan pers kolabora... 1	...	...	1643	lambe waras dedek prayudi psi mangap anies bag... 0	1644	forum komunikasi simpul melawan anies baswedan b... 0	1645	anies py value yakin dpt investor u investasi ... 0	1646	milu negara manatps yg jaga 0	1647	ipo voxpol catat unggul anies-ahy simulasi pil... 1
KALIMAT	KATEGORI EMOSI																																																
0	respon cepat macet akibat jembatan juwana ganj...																																																
1	keren gubernur jawa ganjar pranowo hasil bawa ...																																																
2	hasil survei litbang Kompas Januari elektabili...																																																
3	temu lembaga survei median kait elektabilitas ...																																																
4	jateng ganjar pranowo ajak insan pers kolabora...																																																
...	...																																																
1643	lambe waras dedek prayudi psi mangap anies bag...																																																
1644	forum komunikasi simpul melawan anies baswedan b...																																																
1645	anies py value yakin dpt investor u investasi ...																																																
1646	milu negara manatps yg jaga																																																
1647	ipo voxpol catat unggul anies-ahy simulasi pil...																																																
KALIMAT	KATEGORI EMOSI																																																
0	respon cepat macet akibat jembatan juwana ganj... 0																																																
1	keren gubernur jawa ganjar pranowo hasil bawa ... 0																																																
2	hasil survei litbang Kompas Januari elektabili... 1																																																
3	temu lembaga survei median kait elektabilitas ... 1																																																
4	jateng ganjar pranowo ajak insan pers kolabora... 1																																																
...	...																																																
1643	lambe waras dedek prayudi psi mangap anies bag... 0																																																
1644	forum komunikasi simpul melawan anies baswedan b... 0																																																
1645	anies py value yakin dpt investor u investasi ... 0																																																
1646	milu negara manatps yg jaga 0																																																
1647	ipo voxpol catat unggul anies-ahy simulasi pil... 1																																																

Gambar 9. Data Preprocessing 2 Label

Sama seperti sebelumnya proses yang dilakukan, namun perbedaannya hanya pada label yang digunakan.

3.4. Pembobotan Kata 2 label

Gambar 10. merupakan pembobotan kata menggunakan TF-IDF maupun BoW dengan 2 label.

1648 Number of reviews has 2723 words	1648 Number of reviews has 2723 words
(0, 1138) 0.22656101381483948	(0, 2045) 1
(0, 2305) 0.2966079434013966	(0, 436) 2
(0, 1931) 0.12454383315798107	(0, 1374) 1
(0, 713) 0.0969134584642297	(0, 61) 1
(0, 1039) 0.28718190554379625	(0, 1002) 1
(0, 1002) 0.2966079434013966	(0, 1039) 1
(0, 61) 0.28718190554379625	(0, 713) 1
(0, 1374) 0.3258878913379022	(0, 1931) 1
(0, 436) 0.6175204395439573	(0, 2305) 1
(0, 2045) 0.30876021977197865	(0, 1138) 1
(1, 324) 0.2006462296183843	(1, 713) 1
(1, 1197) 0.21572241411269344	
(1, 1425) 0.19071580075869568	
TF-IDF	BoW

Gambar 10. Pembobotan kata 2 Label dengan TF-IDF

Gambar 10. merupakan proses pembobotan yang ditampilkan pada 2 label dengan menggunakan TF-IDF.

3.5. Performa Model dengan TF-IDF Terhadap 2 label

Gambar 11. Terlihat bahwa performa model yang dihasilkan 2 label adalah sebagai berikut.

<table><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr><tr><td>0</td><td>0.81</td><td>0.68</td><td>0.74</td><td>207</td></tr><tr><td>1</td><td>0.79</td><td>0.89</td><td>0.84</td><td>288</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.80</td><td>495</td></tr><tr><td>macro avg</td><td>0.80</td><td>0.78</td><td>0.79</td><td>495</td></tr><tr><td>weighted avg</td><td>0.80</td><td>0.80</td><td>0.80</td><td>495</td></tr></table>						precision	recall	f1-score	support	0	0.81	0.68	0.74	207	1	0.79	0.89	0.84	288	accuracy			0.80	495	macro avg	0.80	0.78	0.79	495	weighted avg	0.80	0.80	0.80	495	<table><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr><tr><td>0</td><td>0.79</td><td>0.67</td><td>0.73</td><td>207</td></tr><tr><td>1</td><td>0.79</td><td>0.88</td><td>0.83</td><td>288</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.79</td><td>495</td></tr><tr><td>macro avg</td><td>0.79</td><td>0.77</td><td>0.78</td><td>495</td></tr><tr><td>weighted avg</td><td>0.79</td><td>0.79</td><td>0.79</td><td>495</td></tr></table>						precision	recall	f1-score	support	0	0.79	0.67	0.73	207	1	0.79	0.88	0.83	288	accuracy			0.79	495	macro avg	0.79	0.77	0.78	495	weighted avg	0.79	0.79	0.79	495
	precision	recall	f1-score	support																																																																	
0	0.81	0.68	0.74	207																																																																	
1	0.79	0.89	0.84	288																																																																	
accuracy			0.80	495																																																																	
macro avg	0.80	0.78	0.79	495																																																																	
weighted avg	0.80	0.80	0.80	495																																																																	
	precision	recall	f1-score	support																																																																	
0	0.79	0.67	0.73	207																																																																	
1	0.79	0.88	0.83	288																																																																	
accuracy			0.79	495																																																																	
macro avg	0.79	0.77	0.78	495																																																																	
weighted avg	0.79	0.79	0.79	495																																																																	
Multinomial Bayes					Bernoulli Bayes																																																																
<table><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr><tr><td>0</td><td>0.78</td><td>0.79</td><td>0.78</td><td>207</td></tr><tr><td>1</td><td>0.85</td><td>0.84</td><td>0.84</td><td>288</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.82</td><td>495</td></tr><tr><td>macro avg</td><td>0.81</td><td>0.81</td><td>0.81</td><td>495</td></tr><tr><td>weighted avg</td><td>0.82</td><td>0.82</td><td>0.82</td><td>495</td></tr></table>						precision	recall	f1-score	support	0	0.78	0.79	0.78	207	1	0.85	0.84	0.84	288	accuracy			0.82	495	macro avg	0.81	0.81	0.81	495	weighted avg	0.82	0.82	0.82	495	<table><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr><tr><td>0</td><td>0.75</td><td>0.74</td><td>0.74</td><td>207</td></tr><tr><td>1</td><td>0.81</td><td>0.82</td><td>0.82</td><td>288</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.79</td><td>495</td></tr><tr><td>macro avg</td><td>0.78</td><td>0.78</td><td>0.78</td><td>495</td></tr><tr><td>weighted avg</td><td>0.79</td><td>0.79</td><td>0.79</td><td>495</td></tr></table>						precision	recall	f1-score	support	0	0.75	0.74	0.74	207	1	0.81	0.82	0.82	288	accuracy			0.79	495	macro avg	0.78	0.78	0.78	495	weighted avg	0.79	0.79	0.79	495
	precision	recall	f1-score	support																																																																	
0	0.78	0.79	0.78	207																																																																	
1	0.85	0.84	0.84	288																																																																	
accuracy			0.82	495																																																																	
macro avg	0.81	0.81	0.81	495																																																																	
weighted avg	0.82	0.82	0.82	495																																																																	
	precision	recall	f1-score	support																																																																	
0	0.75	0.74	0.74	207																																																																	
1	0.81	0.82	0.82	288																																																																	
accuracy			0.79	495																																																																	
macro avg	0.78	0.78	0.78	495																																																																	
weighted avg	0.79	0.79	0.79	495																																																																	
Complement Bayes					SVM																																																																

Gambar 11. Akurasi TF-IDF Dengan Algoritma Terhadap 2 Label

Hasil analisis sentimen emosi menunjukkan performa dari beberapa algoritma klasifikasi yang menggunakan metode TF-IDF sebagai representasi fitur. Berikut adalah rincian hasil analisis:

3.5.1. Multinomial Naive Bayes dengan TF-IDF:

Akurasi: 80%

Presisi: 80%

Recall: 78%

F1-Score: 79%

3.5.2. Bernoulli Naive Bayes dengan TF-IDF:

Akurasi: 79%
 Presisi: 79%
 Recall: 77%
 F1-Score: 78%

3.5.3. Complement Naive Bayes dengan TF-IDF:

Akurasi: 82%
 Presisi: 81%
 Recall: 81%
 F1-Score: 81%

3.5.4. Support Vector Machine (SVM) dengan TF-IDF:

Akurasi: 79%
 Presisi: 78%
 Recall: 78%
 F1-Score: 78%

Hasil analisis ini menggambarkan performa algoritma-algoritma dalam mengklasifikasikan sentimen emosi pada data teks. Model *Complement Naïve Bayes* memiliki performa yang paling baik dengan akurasi, presisi, recall, dan F1-Score yang mencapai angka 82%, 81%, 81%, dan 81% secara berurutan. Model *Multinomial Naïve Bayes* juga memiliki performa yang baik dengan nilai di atas 80% pada semua metrik evaluasi. Model *Bernoulli Naïve Bayes*, SVM, memiliki performa yang sedikit lebih rendah, tetapi tetap relatif baik dengan nilai akurasi dan metrik evaluasi lainnya di atas 75%.

Kesimpulannya, hasil analisis menunjukkan bahwa model-model klasifikasi menggunakan metode TF-IDF memiliki performa yang baik dalam mengklasifikasikan sentimen emosi pada data teks. Model *Complement Naïve Bayes* dan *Multinomial Naïve Bayes* menjadi pilihan yang kuat berdasarkan evaluasi akurasi, presisi, recall, dan F1-Score.

3.6. Performa Model dengan BoW terhadap 2 Label

Gambar 12. merupakan hasil akurasi yang dihasilkan dari BoW dengan algoritma yang digunakan.

	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.78	0.76	0.77	207	0	0.79	0.67	0.73	207
1	0.83	0.84	0.84	288	1	0.79	0.88	0.83	288
accuracy			0.81	495	accuracy			0.79	495
macro avg	0.81	0.80	0.80	495	macro avg	0.79	0.77	0.78	495
weighted avg	0.81	0.81	0.81	495	weighted avg	0.79	0.79	0.79	495
Multinomial Bayes					Bernoulli Bayes				
	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.75	0.79	0.77	207	0	0.71	0.72	0.71	207
1	0.84	0.82	0.83	288	1	0.80	0.78	0.79	288
accuracy			0.80	495	accuracy			0.76	495
macro avg	0.80	0.80	0.80	495	macro avg	0.75	0.75	0.75	495
weighted avg	0.81	0.80	0.80	495	weighted avg	0.76	0.76	0.76	495
Complement Bayes					SVM				

Gambar 12. Akurasi BoW Dengan Algoritma Terhadap 6 Label

Hasil analisis sentimen menggunakan 2 label menunjukkan performa dari beberapa algoritma klasifikasi yang menggunakan metode *Bag of Words* (BoW) sebagai representasi fitur. Berikut adalah rincian hasil analisis:

3.6.1. *Multinomial Naive Bayes* dengan BoW:

Akurasi: 81%

Presisi: 81%

Recall: 80%

F1-Score: 80%

3.6.2. *Bernoulli Naive Bayes* dengan BoW:

Akurasi: 79%

Presisi: 79%

Recall: 77%

F1-Score: 78%

3.6.3. *Complement Naive Bayes* dengan BoW:

Akurasi: 80%

Presisi: 80%

Recall: 80%

F1-Score: 80%

3.6.4. *Support Vector Machine* (SVM) dengan BoW:

Akurasi: 76%

Presisi: 75%

Recall: 75%

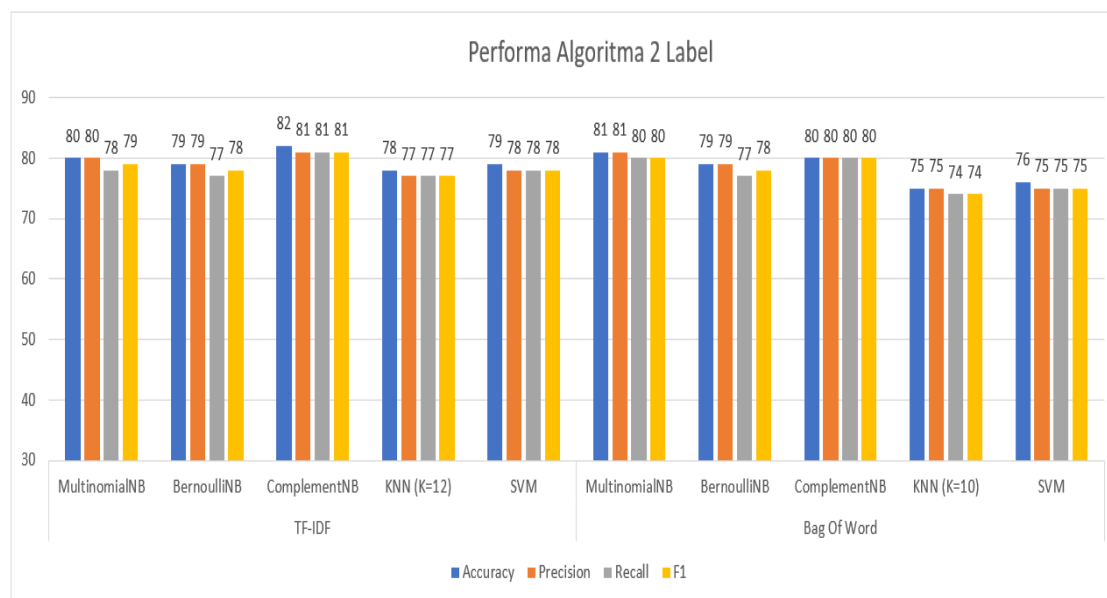
F1-Score: 75%

Hasil analisis menunjukkan performa algoritma-algoritma tersebut dalam mengklasifikasikan sentimen emosi dalam data teks. Model *Multinomial Naive Bayes* dan *Complement Naive Bayes* menunjukkan performa terbaik dengan akurasi, presisi, *recall*, dan F1-Score yang seimbang pada kisaran 80%. Model *Bernoulli Naive Bayes* juga memiliki performa yang baik, meskipun sedikit lebih rendah dibandingkan dengan dua model sebelumnya. Sementara itu, model SVM memiliki performa yang lebih rendah dengan akurasi, presisi, *recall*, dan F1-Score di bawah 80%, meskipun masih memiliki performa yang cukup baik.

Kesimpulannya, hasil analisis menunjukkan bahwa model-model klasifikasi menggunakan metode BoW memiliki performa yang baik dalam mengklasifikasikan sentimen emosi pada data teks. Model *Multinomial Naive Bayes* dan *Complement Naive Bayes* menjadi pilihan yang kuat berdasarkan evaluasi akurasi, presisi, *recall*, dan F1-Score mereka.

3.7. Grafik Performa model Menggunakan 2 label

Gambar 13. merupakan perbandingan akurasi dalam bentuk grafik secara keseluruhan teradap 2 label.



Gambar 13. Akurasi Perbandingan Algoritma dengan 2 Label

Dari Gambar 13. Grafik metode TF-IDF, terlihat bahwa berbagai algoritma klasifikasi memiliki performa yang bervariasi dalam mengklasifikasikan sentimen analisis pada data teks. Model *Complement Naïve Bayes* memiliki akurasi tertinggi (82%), presisi (81%), *recall* (81%), dan F1-Score (81%), diikuti oleh model *Multinomial Naïve Bayes* (80%). Model SVM dan *Bernoulli Naïve Bayes* memiliki performa yang cukup baik dengan akurasi sekitar 79%, sedangkan model KNN memiliki performa sedikit lebih rendah dengan akurasi 78%. Grafik ini membantu visualisasi perbandingan performa relatif antara berbagai algoritma.

Kemudian Grafik dengan Metode BoW, hasil analisis menunjukkan performa algoritma-algoritma klasifikasi dalam mengklasifikasikan sentimen analisis pada data teks dengan menggunakan metode *Bag of Words* (BoW). Model *Complement Naïve Bayes* memiliki akurasi tertinggi (80%), presisi (80%), *recall* (80%), dan F1-Score (80%). Model *Multinomial Naïve Bayes* juga memiliki performa yang baik dengan akurasi, presisi, *recall*, dan F1-Score di atas 80%. Model *Bernoulli Naïve Bayes*, SVM, dan KNN memiliki performa yang relatif baik dengan akurasi di atas 75%. Grafik ini membantu visualisasi perbandingan performa relatif antara berbagai algoritma dalam metode BoW.

Dari analisis ini, dapat disimpulkan bahwa dalam metode TF-IDF, model *Complement Naïve Bayes* memiliki performa paling baik dalam mengklasifikasikan sentimen analisis, diikuti oleh *Multinomial Naïve Bayes*. Dalam metode BoW, model *Complement Naïve Bayes* dan *Multinomial Naïve Bayes* juga menjadi pilihan yang kuat berdasarkan evaluasi akurasi, presisi, *recall*, dan F1-Score. Model-model tersebut dapat menjadi pertimbangan untuk penerapan lebih lanjut dalam analisis sentimen pada data teks.

4. KESIMPULAN

Penelitian ini telah mendapatkan kesimpulan yang dilihat dari beberapa pengujian yang telah dilakukan. Pengujian yang telah dilakukan pemodelan menggunakan 2 algoritma yakni *Naïve Bayes* dan SNM dengan Akurasi 2 label algoritma terbaik adalah *Complement Bayes* TF-IDF dengan nilai akurasi 82% dan nilai ROC 87%.

Daftar Pustaka

- [1] A. Nayla, C. Setianingsih and B. Dirgantoro, "Deteksi Hate Speech Pada Twitter Menggunakan Algoritma BERT," *Journal of computational science*, vol. 10, no. 1, 2023.
- [2] D. Darwis, E. S. Pratiwi and A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *Jurnal Ilmiah Edutic: Pendidikan dan Informatika*, vol. 7, no. 1, pp. 1-11, 2020.

- [3] P. C. Deka, "Support vector machine applications in the field of hydrology: a review," *Applied soft computing*, vol. 19, pp. 372-386, 2014.
- [4] N. Matondang and N. Surantha, "Effects of oversampling SMOTE in the classification of hypertensive dataset," *Adv. Sci. Technol. Eng. Syst*, vol. 5, no. 4, pp. 432-437, 2020.
- [5] I. B. A. Peling, I. N. Arnawan, I. P. A. Arthawan and I. G. N. Janardana, "Implementation of Data Mining To Predict Period of Students Study Using Naive Bayes Algorithm," *Int. J. Eng. Emerg. Technol*, vol. 2, no. 1, p. 53, 2017.
- [6] R. Sepriansyah and S. D. Purnamasari, "Prediction of Student Graduation Using Naïve Bayes," *Budapest International Research And Critics Institute-Journal (Birci-Journal)*, vol. 5, no. 3, pp. 24255-24268, 2022.
- [7] K. Sailunaz and R. Alhajj, "Emotion and sentiment analysis from Twitter text," *Journal of computational science*, vol. 36, pp. 1-18, 2019.
- [8] P. A. Olujimi and A. Ade-Ibijola, "NLP techniques for automating responses to customer queries: a systematic review," *Discover Artificial Intelligence*, vol. 3, no. 1, p. 20, 2023.
- [9] A. N. Ulfah, M. K. Anam, N. Y. S. Munti, S. Yaakub and M. B. Firdaus, "Sentiment Analysis of the Convict Assimilation Program on Handling Covid-19," *JUITA: Jurnal Informatika*, vol. 10, no. 2, pp. 209-215, 2022.
- [10] A. Gelman and C. R. Shalizi, "Philosophy and the practice of Bayesian statistics," *British Journal of Mathematical and Statistical Psychology*, vol. 66, no. 1, pp. 8-38, 2013.



ZONAsi: Jurnal Sistem Informasi

Is licensed under a [Creative Commons Attribution International \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/)