

Analisis Sentimen Bakal Calon Presiden Indonesia 2024 Dengan Algoritma Naïve Bayes

Mohammad Fikri Haikal^{1*}, Jamaludin Indra², Rahmat³

Teknik Informatika, Universitas Buana Perjuangan, Karawang, Indonesia

*e-mail *Corresponding Author*: if19.mohammadhaikal@mhs.ubpkarawang.ac.id

Abstract

Indonesia's current constitutional regulations prohibit a president from running for more than two consecutive terms. Therefore, the presidential election in 2024 will be a very interesting moment, because Jokowi cannot run again. This opens up opportunities for various figures and other political figures to compete and nominate themselves as presidential candidates. Internet residents are also active in expressing their opinions on this matter, one of whom is Twitter user. This opinion from Twitter is used to analyze public sentiment towards presidential candidate figures. This research applies the Naive Bayes algorithm. The hope is that the results of this research can provide insight to the public in the presidential election process in the next presidential election. The conclusion of this research is that the accuracy value is 80% for Ganjar Pranowo, 70.4% for Anies Baswedan, and 75% for Prabowo Subianto. The best Naive Bayes algorithm testing performance is found in the Ganjar Pranowo dataset with an f1-score of 0.88 for the positive class.

Keywords: *Presidential Candidates; Figures; Sentiments; Twitter; Naive Bayes.*

Abstrak

Peraturan konstitusi Indonesia saat ini melarang seorang presiden untuk mencalonkan diri lebih dari dua periode berturut-turut. Oleh karena itu, pemilihan presiden pada tahun 2024 akan menjadi momen yang sangat menarik, karena presiden Jokowi tidak dapat mencalonkan diri lagi. Hal ini membuka peluang bagi berbagai tokoh dan figur politik lainnya untuk bersaing dan mencalonkan diri sebagai bakal calon presiden. Warga internet juga aktif berpendapat mengenai hal ini, salah satunya pengguna *twitter*. Pendapat dari *twitter* ini digunakan untuk menganalisis sentimen masyarakat terhadap tokoh-tokoh calon presiden. Penelitian ini menerapkan algoritma *Naive Bayes*. Harapannya, hasil penelitian ini dapat memberikan wawasan kepada masyarakat dalam proses pemilihan presiden pada pemilihan presiden mendatang. Kesimpulan dari penelitian ini mendapatkan nilai akurasi yaitu 80% untuk Calon Presiden Ganjar Pranowo, 70,4% untuk calon presiden Anies Baswedan, dan 75% untuk calon presiden Prabowo Subianto. Performa pengujian Algoritma *Naive Bayes* terbaik terdapat pada pengujian dataset untuk calon presiden Ganjar Pranowo dengan nilai f1-score sebesar 0,88 untuk kelas positif.

Kata kunci: *Bakal Calon Presiden; Tokoh; Sentimen; Twitter; Naive Bayes.*

1. Pendahuluan

Indonesia telah mengalami perubahan politik yang signifikan dalam beberapa periode terakhir, terutama terkait dengan pemilihan presiden. Pemilihan presiden adalah momen penting dalam kehidupan demokrasi sebuah negara, di mana masyarakat berhak untuk memilih pemimpin yang akan memimpin negara mereka. Pada tahun 2019, Indonesia menyaksikan pemilihan presiden yang memenangkan Joko Widodo (Jokowi) untuk kedua kalinya, yang kini menjabat sebagai presiden selama dua periode. Namun, peraturan konstitusi Indonesia saat ini melarang seorang presiden untuk mencalonkan diri lebih dari dua periode berturut-turut[1]. Oleh karena itu, pemilihan presiden pada tahun 2024 akan menjadi momen yang sangat menarik, karena Jokowi tidak dapat mencalonkan diri lagi. Hal ini membuka peluang bagi berbagai tokoh dan figur politik lainnya untuk bersaing dan mencalonkan diri sebagai bakal calon presiden.

Berdasarkan konteks ini, telah terjadi sejumlah survei elektabilitas yang dilakukan terhadap berbagai tokoh publik yang dipersepsikan memiliki popularitas yang cukup tinggi. Hal ini membuka potensi kemungkinan mereka sebagai calon presiden Indonesia dalam pemilihan presiden (pilpres) tahun 2024. Hasil survei yang beragam tersebut menciptakan variasi dalam

data yang tersedia. Peneliti memutuskan untuk melakukan analisis terhadap tiga data survei yang dilakukan oleh lembaga survei yang berbeda. Data dari survei-survei tersebut telah dipublikasikan dalam sebuah artikel yang berjudul "Zulhas Didorong Nyapres 2024, Hasil Survei Membuktikan..." [2]. Dari ketiga data survei yang dianalisis, peneliti merangkum nilai rata-rata elektabilitas empat tokoh publik utama, yaitu Ganjar Pranowo bernilai (19,2%), Anies Baswedan bernilai (14,2%), Prabowo Subianto bernilai (14%), dan Ridwan Kamil bernilai (9,84%). Selain itu, publik juga aktif berpartisipasi dalam proses demokratis dengan memberikan tanggapan mereka melalui media sosial terkait hasil survei tersebut.

Perkembangan teknologi yang sangat pesat telah menghasilkan berbagai jenis media online yang beragam, termasuk platform seperti *Facebook*, *Twitter*, *Instagram*, *Telegram*, *Tiktok*, *LinkedIn*, dan sebagainya [3]. Saat ini, media sosial telah melampaui peran semula sebagai alat komunikasi dengan teman dan keluarga, dan telah menjadi wadah dalam berbagai aspek kehidupan. Pengguna media sosial tidak hanya berinteraksi sosial, tetapi juga menggunakan platform tersebut untuk keperluan transaksi, pendidikan, serta sebagai alat promosi, termasuk untuk partai politik dan kampanye presiden. Pada tahun 2022, 191 juta warga Indonesia yang aktif menggunakan *platform* media sosial, dan *Twitter* sebagai salah satu *platform* yang paling diminati dengan jumlah pengguna mencapai 18,45 juta[4].

Dalam bidang analisis sentimen, terdapat beragam teknik klasifikasi yang tersedia, dan salah satunya adalah *classifier Naïve Bayes*. Yerik Afrianto dalam penelitiannya yang berjudul "Seleksi Metode dan Algoritma dalam Analisis Sentimen di Media Sosial: Tinjauan Sistematis atas Kajian Literatur" mengungkapkan bahwa algoritma *Naïve Bayes* mendominasi dalam penelitian analisis sentimen di *platform* media sosial[5].

Sejumlah penelitian sebelumnya telah dilaksanakan dalam lingkup analisis sentimen. Sebagai contoh, penelitian yang dilakukan oleh Amril Mutoi dan Maulana Abdur dalam sebuah jurnal berjudul "Pengukuran Tingkat Kepuasan Mahasiswa dalam Proses Pembelajaran dengan Pendekatan *Naïve Bayes*" berhasil mencapai tingkat akurasi sebesar 86,9%. Penelitian lainnya, yang dilakukan oleh peneliti yang tidak disebutkan namanya, memanfaatkan pendekatan *Naïve Bayes* untuk menganalisis sentimen dan menentukan polaritasnya. Hasil penelitian tersebut menghasilkan bahwa metode klasifikasi *Naïve Bayes Classifier* mampu memperoleh tingkat akurasi sebesar 79,6% [6].

Penelitian sebelumnya yang relevan telah dilaksanakan oleh Santoso dan Nugroho, yang mengarah kepada analisis sentimen terkait calon presiden Indonesia pada tahun 2019. Sumber data untuk penelitian tersebut adalah media sosial *Facebook*, dan pengumpulan data dilakukan dalam periode mulai tanggal 17 April 2019 hingga 22 Mei 2019. Hasil dari penelitian tersebut mengindikasikan bahwa mayoritas sentimen mendukung Joko Widodo lebih daripada Prabowo Subianto[7]. Pada penelitian yang dilaksanakan saat ini menggunakan metode klasifikasi *Naïve Bayes* untuk mengukur opini publik yang berfokus pada media sosial *Twitter* terkait 3 kandidat calon presiden pada tahun 2024.

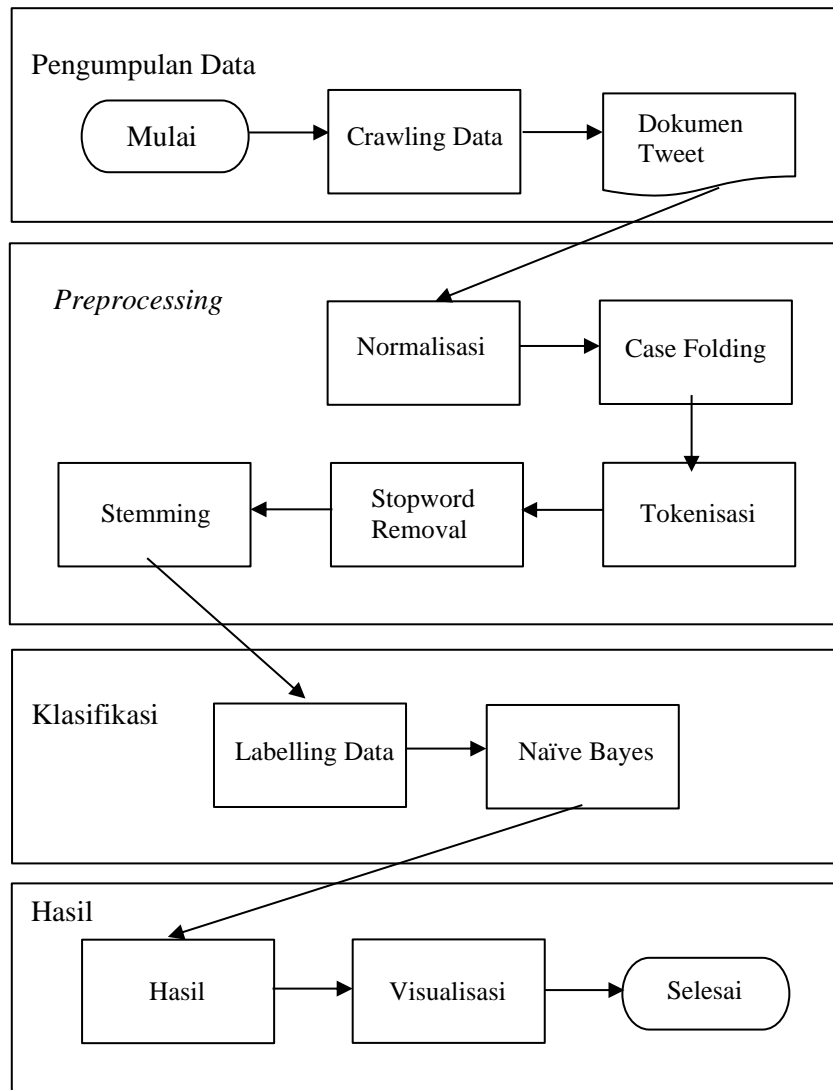
2. Metodologi

Penelitian ini dilakukan dengan beberapa tahap. Langkah langkah penelitian ini dijelaskan pada gambar 1, yakni: pengumpulan data, *preprocessing* data, labelling data, klasifikasi menggunakan *Naïve Bayes*, dan visualisasi.

2.1 Pengumpulan Data

Objek yang diteliti adalah opini masyarakat mengenai bakal calon presiden melalui media sosial *twitter*. Pengumpulan data dilakukan menggunakan API yang telah disediakan oleh *Twitter* dengan bantuan tools "*tweet harvest*" yang dikembangkan oleh seorang YouTuber bernama Helmi Satria. Proses pengumpulan data dilaksanakan dengan mengikuti beberapa pedoman yang telah ditetapkan, termasuk pembatasan periode pencatatan cuitan mulai dari tanggal 2 November 2022 hingga 2 Juli 2023. Selain itu, penelitian ini juga memanfaatkan penggunaan hastag atau tanda pagar tertentu untuk melakukan pencarian data yang diperlukan. Tagar yang digunakan adalah:

(#capres2024, #ganjarpranowo, #aniesbaswedan, #prabowo)



Gambar 1. Prosedur Penelitian

2.2 Preprocessing

Data yang telah dikumpulkan merupakan data mentah dan perlu melalui tahapan *preprocessing* agar siap untuk di olah pada proses berikutnya. *Preprocessing* teks adalah tahap yang penting untuk melakukan segmentasi teks, yang mana karakteristik teks dapat dinilai, dianalisis, dan diklasifikasikan [8]. Proses pre-processing adalah sebagai berikut:

- 1) *Normalisasi*: Normalisasi merujuk pada tahap pengolahan data yang bertujuan untuk mengeliminasi elemen-elemen yang tidak relevan dalam konteks pengambilan data dari platform Twitter. Elemen-elemen ini mencakup, antara lain, tautan URL, identitas pengguna (*user id*), dan unsur-unsur lainnya yang tidak memiliki signifikansi dalam konteks analisis data.
- 2) *Case Folding*: Dalam tweet, penggunaan huruf besar dan kecil sering kali bervariasi. Oleh karena itu, dalam tahap ini, seluruh huruf besar akan diubah menjadi huruf kecil agar teks seragam dalam penggunaan huruf.
- 3) *Tokenizing*: *Tokenizing* adalah tahap dalam pengolahan data yang melibatkan pemisahan sebuah kalimat menjadi unit-unit kecil yang disebut token atau kata-kata. Proses ini dilakukan berdasarkan tanda baca tertentu seperti koma, titik, dan karakter tanda baca lainnya yang digunakan sebagai pemisah antara kata-kata.

- 4) *Stopword Removal*: Eliminasi kata-kata umum dan sering muncul dalam analisis teks adalah langkah penting dalam memproses data teks untuk tujuan analisis sentimen dan tugas pemrosesan bahasa alami lainnya.[9]

2.3 Labelling Data

Langkah ini bertujuan untuk mengelompokkan data ke dalam tiga kategori sentimen yang berbeda, yakni positif, netral, dan negatif. Akan tetapi, penelitian ini hanya berfokus pada sentimen positif dan negatif. Penelitian ini serupa dengan penelitian sebelumnya yang berkaitan dengan analisis sentimen, yang dibuat oleh, Pardamean, Hikmawan, dan Khasanah[10]. Pengelompokan nilai sentimen berlandaskan pada kamus sentimen dari penelitian yang berjudul "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2019 berdasarkan Opini dari *Twitter* dengan Metode Klasifikasi *Naive Bayes*" [11]. Selama tahap ini, data dibagi menjadi dua bagian, yaitu data latih (*training data*), dan data uji (*testing data*). Data latih berfungsi untuk melatih sistem untuk berbagai jenis model dan algoritma, sementara data uji berfungsi untuk mengevaluasi hasil klasifikasi yang telah dilakukan.

2.4 Algoritma Naive Bayes

Dalam penelitian ini, metode klasifikasi yang digunakan adalah yang disebut dengan *Naive Bayes*. Teknik ini adalah suatu metode pengelompokan yang digunakan untuk memproyeksikan peluang munculnya suatu kelas tertentu. *Naive Bayes* berlandaskan pada prinsip-prinsip probabilitas dan konsep statistik yang pertama kali dikembangkan oleh ilmuwan asal Inggris, yakni Thomas Bayes[12]. Salah satu karakteristik utama dari metode *Naive Bayes* adalah asumsi bahwa variabel prediktor yang digunakan dalam klasifikasi adalah independen satu sama lain. Dengan kata lain, keberadaan satu fitur atau prediktor dalam kelas tidak bergantung pada keberadaan prediktor lainnya [13]. Sebagai contoh, jika terdapat tiga fitur yang menggambarkan suatu objek, *Naive Bayes* menganggap bahwa ketiga fitur tersebut memberikan kontribusi terhadap probabilitas yang menggambarkan objek tersebut secara independen, sehingga disebut "*Naive*". Teorema *Bayes* yang mendasari metode ini dapat dirumuskan sebagai berikut:

$$P(H|X) = \frac{P(H).P(X|H)}{P(X)} \dots\dots\dots (1)$$

Dimana:

- X : Data dengan kelas yang belum diketahui
 H : Hipotesis data merupakan suatu kelas spesifik
 $P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (probabilitas posterior)
 $P(H)$: Probabilitas hipotesis H (probabilitas prior)
 $P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H
 $P(X)$: Probabilitas

Langkah selanjutnya adalah melakukan evaluasi model, yang mencakup pengukuran akurasi, presisi, dan sensitivitas pada data pengujian. Evaluasi ini bertujuan untuk menilai sejauh mana ketepatan klasifikasi yang telah dilakukan. Akurasi adalah metrik yang mengukur tingkat ketepatan keseluruhan dari teknik klasifikasi, di mana tingkat akurasi yang lebih tinggi mengindikasikan performa teknik klasifikasi yang lebih baik. Untuk melihat dasar perhitungan evaluasi model, kita dapat merujuk ke tabel matriks (*confusion matrix*) yang tercantum dalam Tabel 1.

Tabel 1. *Confusion Matrix* Hasil Klasifikasi Biner

Kelas	Prediksi Positif	Prediksi Negatif
Positif	tp	fn
Negatif	fp	tn

3. Hasil dan Pembahasan

3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini dikumpulkan dalam rentang waktu tertentu, mulai dari tanggal 2 November 2022 hingga 2 Juli 2023. Hasil dari pengumpulan data ini menghasilkan total 490 cuitan yang terdiri dari 145 cuitan yang terkait dengan Ganjar Pranowo, 154 cuitan yang terkait dengan Anies Baswedan, dan 119 cuitan yang terkait dengan Prabowo Subianto. Informasi lebih rinci mengenai hasil pengumpulan data ini dapat ditemukan dalam Tabel 2.

Dataset	Jumlah Data
Ganjar Pranowo	145
Anies Baswedan	154
Prabowo Subianto	119
Total	490

3.2 Preprocessing

Dalam penelitian ini, terdapat sejumlah teknik yang digunakan untuk melakukan *preprocessing* data pada teks Twitter. Langkah-langkah *preprocessing* tersebut meliputi normalisasi, *case folding*, *tokenizing*, *stopwords removal*, dan *stemming*. Berikut ini adalah contoh dari proses *preprocessing* data Twitter yang diimplementasikan dalam penelitian, seperti yang terlihat dalam Tabel 4.

Proses	Contoh
Data Mentah	Uma @UmaOne Dec 16 Resmi PBB Menangkan Prabowo Subianto Presiden #ganjar #capres2024
Normalisasi	Resmi PBB Menangkan Prabowo Subianto Presiden
<i>Case Folding</i>	resmi pbb menangkan prabowo subianto presiden
Tokenisasi	resmi, pbb, menang, prabowo, subianto, presiden
<i>Stopword Removal</i>	resmi, menang, prabowo, subianto, presiden
<i>Stemming</i>	menang, prabowo, subianto, presiden

Setelah tahap *preprocessing* selesai, langkah selanjutnya adalah melakukan pemisahan data. Dataset akan dibagi menjadi dua bagian, yaitu dataset pelatihan (data latih) dan dataset pengujian (data uji). Proporsi yang digunakan adalah 87,5% dari total data untuk data latih dan 12,5% sisanya untuk data uji pada setiap dataset.

3.3 Pelabelan Kelas Sentimen

Setelah tahap *preprocessing* data selesai, langkah berikutnya adalah melakukan pelabelan pada *tweet* dengan mengkategorikan-kategorikan mereka menjadi sentimen positif, negatif, dan netral. Perlu dicatat bahwa dalam penelitian ini, Tidak ada penggunaan sentimen netral dalam penelitian ini. Langkah selanjutnya adalah melakukan pembobotan kata dengan cara menghitung seberapa sering kata-kata muncul dalam teks. Pendekatan ini didasarkan pada teknik pengelompokan kata yang telah diimplementasikan dalam penelitian ini [11], Pada setiap teks komentar atau *tweet*, perhitungan skor sentimen dilakukan dengan menjumlahkan skor kata-kata positif dan mengurangkannya dengan skor kata-kata negatif dalam kalimat tersebut [14], atau dapat dirumuskan sebagai berikut:

Persamaan 1. Rumus Skor Sentimen

$$Skor = \left(\sum \text{kata positif} \right) - \left(\sum \text{kata negatif} \right)$$

Berdasarkan hasil perhitungan ini, kalimat dengan skor < 0 akan dikategorikan sebagai sentimen negatif. Kalimat dengan skor $= 0$ akan dikategorikan sebagai sentimen netral, sedangkan kalimat dengan skor > 0 akan dimasukkan ke dalam kategori sentimen positif. Penting untuk dicatat bahwa dalam penelitian ini, hanya dua kategori pelabelan yang digunakan, yaitu sentimen positif dan sentimen negative[15].

Tabel 4. Contoh Pelabelan Hasil Sentimen

Tweets	Skor	Sentimen
parah klo prabowo presiden	-2	Negatif
resmi pbb menang prabowo subianto presiden	1	Positif
kayak seru banget acara milad tadi pas prabowo masuk orang orang teriak teriak prabowo presiden pilpres dekade mending prabowo terus maju bersama prabowo	1	Positif
partai bintang resmi dukung prabowo presiden	1	Positif
kalah lagi gak nih	-1	Negatif
bismillah prabowo presidengerindra menang	2	Positif

Adapun hasil sentimen pada setiap variabel adalah sebagai berikut:

Tabel 5. Jumlah Label Dataset

Data	Positif	Negatif	Jumlah Data
Ganjar Pranowo	128	17	145
Anies Baswedan	103	51	154
Prabowo Subianto	150	41	191
Total	381	109	490

3.4 Hasil

Proses pengujian melibatkan dua kelas sentimen, yakni kelas negatif dan positif, yang diaplikasikan pada masing-masing dataset. Ganjar Pranowo memperoleh hasil klasifikasi menggunakan tiga kelas sentimen dengan tingkat akurasi sebesar **80%**. Untuk melihat secara detail, tabel 6 menyajikan *confusion matrix* untuk dataset Ganjar Pranowo.

Tabel 6. Hasil Uji Dataset 1

	precision	recall	f1-score
Negatif	1.00	0.33	0.33
Positif	0.79	1.00	0.88

Anies Baswedan memperoleh hasil klasifikasi menggunakan dua kelas sentimen dengan tingkat akurasi sebesar **70,4%**. Untuk melihat secara detail, tabel 7 menyajikan *confusion matrix* untuk dataset Anies Baswedan.

Tabel 7. Hasil Uji Dataset 2

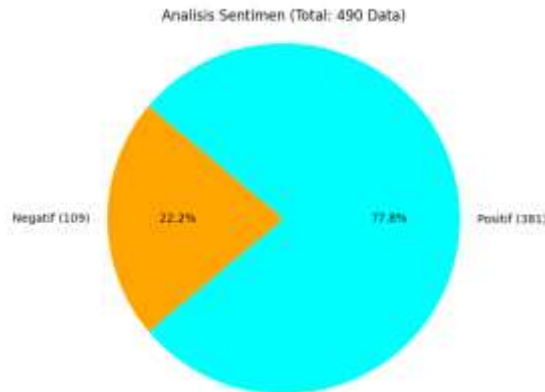
	precision	recall	f1-score
Negatif	0,40	1,00	0.57
Positif	1,00	0,45	0.62

Prabowo Subianto memperoleh hasil klasifikasi menggunakan dua kelas sentimen dengan tingkat akurasi sebesar **75%**. Untuk melihat secara detail, tabel 8 menyajikan *confusion matrix* untuk dataset Prabowo Subianto.

Tabel 8. Hasil Uji Dataset 3

	precision	recall	f1-score
Negatif	0.50	0.80	0.62
Positif	0.92	0.73	0.81

3.5 Visualisasi
1) Pie Chart



Gambar 2. Persentase *Tweets* Berdasarkan Sentimen Positif dan Negatif

Berdasarkan gambar diatas, masyarakat lebih banyak memberikan sentimen/komentar positif terhadap bakal calon presiden 2024. Hal ini diperkuat oleh perbandingan persentase antara kedua jenis sentimen, di mana sentimen positif mendominasi dengan persentase sebesar 77,8%. Perbedaan yang signifikan ini menunjukkan bahwa mayoritas masyarakat yang menggunakan *Twitter* sebagai platform untuk berpendapat secara publik tampaknya masih memiliki perbedaan terhadap para tokoh bakal calon presiden 2024. Perbedaan pendapat tersebut bisa disebabkan karena kinerja buruk dari para tokoh bakal calon presiden 2024. Berdasarkan hal tersebut, sebaiknya mampu menjadi bahan masukan atau evaluasi terhadap kinerja masing masing tokoh.

2) *Wordcloud*



Gambar 3. *Wordcloud* Kelompok Kata Positif



Gambar 4. *Wordcloud* Kelompok Kata Negatif

Dalam analisis sentimen terkait bakal calon presiden 2024 menggunakan data dari *Twitter*, hasil visualisasi *wordcloud* untuk sentimen positif dan negatif menunjukkan adanya keragaman dalam pendapat dan komentar masyarakat terhadap tokoh-tokoh tersebut. *Wordcloud* digunakan sebagai alat visualisasi yang efektif untuk menyoroti kata-kata kunci yang paling sering muncul dalam konteks sentimen positif dan negatif.

Artinya:

- a) Keragaman Sentimen: Dengan melihat *wordcloud* sentimen positif, kita dapat melihat bahwa kata-kata seperti "baik", "dukung", "pilih", dan "jujur" menonjol. Ini mengindikasikan bahwa sebagian besar masyarakat memiliki pandangan positif terhadap beberapa tokoh bakal calon presiden. Namun, penting untuk diingat bahwa kata-kata ini hanya mencerminkan sebagian kecil dari beragam pendapat yang mungkin ada di antara masyarakat.

- b) Kritik dan Keberagaman: Di sisi lain, dalam *wordcloud* sentimen negatif, kata-kata seperti "jelek", "terbengkalai", "tolak", dan "molor" mendominasi visualisasi. Ini menunjukkan bahwa ada juga sebagian masyarakat yang memiliki pandangan kritis dan negatif terhadap beberapa tokoh bakal calon presiden. Kritik ini mungkin berkaitan dengan berbagai masalah atau isu yang dianggap penting oleh masyarakat.
- c) *Wordcloud* sebagai Alat Analisis Visual: Penggunaan *wordcloud* sebagai alat visualisasi membantu untuk dengan cepat mengidentifikasi kata-kata kunci yang mencerminkan sentimen umum masyarakat. Ukuran kata yang berbeda dalam *wordcloud* mencerminkan frekuensi kemunculan kata tersebut dalam dataset, sehingga memudahkan peneliti atau pemangku kepentingan dalam melihat pandangan yang paling dominan.

4. Simpulan

Simpulan Adapun kesimpulan dari penelitian dengan judul analisis sentimen bakal calon presiden 2024 dengan Algoritma *Naive Bayes*, adalah Algoritma *Naive Bayes* mampu melakukan sentimen terkait dengan bakal calon presiden 2024 melalui analisis data yang diperoleh dari platform media sosial Twitter. Analisis sentimen bakal calon presiden 2024 dengan Algoritma *Naive Bayes* mendapatkan nilai akurasi yaitu 80% untuk Ganjar Pranowo, 70,4% untuk Anies Baswedan, dan 75% untuk Prabowo Subianto. Performa pengujian Algoritma *Naive Bayes* terbaik terdapat pada pengujian *dataset* Ganjar Pranowo dengan nilai *f1-score* sebesar 0,88 untuk kelas positif.

Daftar Referensi

- [1] Pemerintah Indonesia, *Undang - Undang Republik Indonesia Nomor 42 Tahun 2008 Tentang Pemilihan Umum Presiden dan Wakil Presiden*. Jakarta: Sekretariat Negara, 2008.
- [2] E. Safitri, "Zulhas Didorong Nyapres 2024, Survey Membuktikan...", *detiknews*. Accessed: Oct. 01, 2023. [Online]. Available: Baca artikel detiknews, "Zulhas Didorong Nyapres 2024, Survei Membuktikan..." selengkapnya <https://news.detik.com/berita/d-5705142/zulhas-didorong-nyapres-2024-survei-membuktikan>. Download Apps Detikcom Sekarang <https://apps.detik.com/detik/>
- [3] A. Rifda, "10 Sosial Media Paling Populer di Indonesia," *Gramedia Blog*. Accessed: Oct. 01, 2023. [Online]. Available: <https://www.gramedia.com/best-seller/sosial-media-paling-populer/>
- [4] M I. Mahdi, "Pengguna Media Sosial di Indonesia Capai 191 Juta pada 2022," *DataIndonesia*. Accessed: Oct. 01, 2023. [Online]. Available: <https://dataindonesia.id/internet/detail/pengguna-media-sosial-di-indonesia-capai-191-juta-pada-2022>].
- [5] Y. A. Singgalen, "Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial: Systematic Literature Review," *Journal of Information Systems and Informatics*, vol. 3, no. 2, pp. 278–302, Jun. 2021, doi: 10.33557/journalisi.v3i2.125.
- [6] A. M. Siregar and M. A. Rofik, "Mengukur Tingkat Kepuasan Mahasiswa Dalam Pembelajaran Dengan Naive Bayes," *Prosiding Konferensi Nasional Penelitian Dan Pengabdian Universitas Buana Perjuangan Karawang*, vol. 1, no. 1, pp. 1–19, 2021.
- [7] A. Muzaki and A. Witanti, "Sentiment Analysis of The Community in The Twitter in The 2020 Election In Pandemic Covid-19 By Method Naive Bayes Classifier," *Jurnal Teknik Informatika (Jutif)*, vol. 2, no. 2, pp. 101–107, Mar. 2021, doi: 10.20884/1.jutif.2021.2.2.51.
- [8] I. M. M. Parwita and D. Siahaan, "Classification of mobile application reviews using word embedding and convolutional neural network," *Lontar Komput. J. Ilm. Teknol. Inf*, vol. 10, no. 1, pp. 1–8, 2019.
- [9] A. F. Anees, A. Shaikh, A. Shaikh, and S. Shaikh, "Survey paper on sentiment analysis: Techniques and challenges," *EasyChair2516-2314*, 2020.
- [10] S. Hikmawan, A. Pardamean, and S. N. Khasanah, "Sentimen Analisis Publik Terhadap Joko Widodo terhadap wabah Covid-19 menggunakan Metode Machine Learning," *Jurnal Kajian Ilmiah*, vol. 20, no. 2, pp. 167-176, 2020.
- [11] A. Deolika, K. Kusriani, and E. T. Luthfi, "Analisis pembobotan kata pada klasifikasi text mining," (*JurTI*) *Jurnal Teknologi Informasi*, vol. 3, no. 2, pp. 179–184, 2019.
- [12] J. J. Aripin, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi pada BPR Pantura," 2021.

- [13] D. A. Agustina and F. Rahmah, "Analisis Sentimen pada Sosial Media Twitter terhadap MRT Jakarta Menggunakan Machine Learning," *Insearch: Information System Research Journal*, vol. 2, no. 01, pp. 1–6, 2022.
- [14] S. Juanita, "Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes," *Jurnal Media Informatika Budidarma*, vol. 4, no. 3, pp. 552-561, Jul. 2020, doi: 10.30865/mib.v4i3.2140.
- [15] A. B. P. Negara, H. Muhandi, and I. M. Putri, "Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes dan Seleksi Fitur Information Gain," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 3, pp. 599-609, 2020, doi: 10.25126/jtiik.2020711947.