



Penerapan Algoritma Naïve Bayes untuk Klasifikasi Sentimen Opini Publik di Media Sosial Twitter terhadap Calon Presiden Anies Baswedan

Siti Arbilah^{1*}, Nurman Herianto², Aditiya Pratama³

¹²³ Program Studi Sistem Informasi, Universitas Nahdlatu Ulama Kalimantan Barat

Email: ^{1*}sitiarbila12345@email.com

Abstrak

Isu-isu politik dan sosial memainkan peran krusial dalam Pemilihan Umum di Indonesia. Perkembangan teknologi dan media sosial, seperti Twitter (X), telah mengubah cara masyarakat menyampaikan opini, menciptakan volume data teks yang sangat besar dan kompleks. Penelitian ini bertujuan untuk menganalisis sentimen opini publik terhadap calon presiden Anies Baswedan menggunakan algoritma Naïve Bayes. Metode penelitian yang digunakan adalah analisis sentimen dengan pendekatan kuantitatif, dengan populasi berupa seluruh *tweet* yang berkaitan dengan #CapresAniesBaswedan. Sampel data dikumpulkan melalui teknik *web crawling* dari tujuh periode berbeda seputar sesi debat Pemilu 2024. Instrumen penelitian menggunakan perangkat lunak Google Colab dan RapidMiner, sementara teknik analisis data melibatkan tahapan *text mining* seperti *Tokenizing*, *Filtering*, *Stemming*, dan TF-IDF. Hasil analisis menunjukkan algoritma Naïve Bayes efektif dalam mengklasifikasikan sentimen, dengan akurasi rata-rata di atas 84% dan tertinggi mencapai 98,40%. Kesimpulannya, sentimen negatif paling dominan ditemukan pada empat dari tujuh dataset, menunjukkan opini publik yang cenderung kurang baik.

Kata Kunci: Analisis Sentimen, Klasifikasi, Naive Bayes, Twitter, Pemilu

PENDAHULUAN

Fenomena Penelitian

Pemilihan umum (Pemilu) merupakan pilar utama dalam sistem demokrasi di Indonesia, memberikan hak suara kepada warga negara untuk memilih pemimpin dan perwakilan mereka secara periodik (Rosyida et al., 2023). Dalam lanskap politik modern, isu-isu sosial dan politik seperti ekonomi, lingkungan, keamanan nasional, dan hak-hak individu, memainkan peran krusial dalam membentuk preferensi pemilih. Perkembangan pesat teknologi dan media sosial telah secara fundamental mengubah cara kampanye politik dilakukan, memungkinkan penyebaran informasi yang lebih cepat dan interaksi langsung antara kandidat dan pemilih (Setiawan et al., 2021). Dahulu, masyarakat menyampaikan opini melalui media cetak yang terbatas, namun kini media sosial seperti Twitter (sekarang X) telah menjadi platform utama bagi publik untuk menyuarakan pandangan, kritik, dan sentimen mereka terhadap tokoh politik dan isu-isu Pemilu. Fenomena ini menciptakan volume data teks yang sangat besar, kompleks, dan tidak terstruktur, yang sulit dianalisis secara manual oleh masyarakat, analis politik, maupun peneliti (Dewi et al., 2023).

Untuk memahami pandangan dan sentimen publik secara efisien, dibutuhkan pendekatan komputasi yang dapat mengklasifikasikan data teks dalam jumlah besar. Analisis sentimen, sebagai salah satu cabang dari *text mining*, menjadi metode yang sangat relevan untuk mengatasi tantangan ini. Dengan menganalisis opini yang diungkapkan dalam *tweet*, kita dapat mengkategorikannya ke dalam sentimen positif, negatif, atau netral, yang pada akhirnya dapat memberikan wawasan berharga mengenai popularitas dan persepsi masyarakat terhadap seorang tokoh politik (Suryadi et al., 2022). Oleh karena itu, penelitian yang berfokus pada analisis sentimen di media sosial menjadi sangat penting untuk memetakan dinamika opini publik dalam konteks Pemilu.

Permasalahan Penelitian

Dalam ranah analisis sentimen, telah banyak penelitian yang menggunakan berbagai algoritma pembelajaran mesin. Algoritma-algoritma seperti Naïve Bayes, Support Vector Machine (SVM), dan *clustering* K-Means umum digunakan untuk tugas klasifikasi sentimen (Budi, 2017). Meskipun demikian, setiap algoritma memiliki karakteristik dan performa yang berbeda, tergantung pada kompleksitas data dan kebutuhan spesifik penelitian. Sebagai contoh, algoritma Naïve Bayes dikenal karena kemudahannya diimplementasikan dan efisiensinya dalam memproses data teks, serta memberikan hasil yang memuaskan dalam kasus-kasus sederhana (Dewi et al., 2023). Algoritma ini memodelkan hubungan antara kata-kata dalam teks dan sentimen yang dikandungnya berdasarkan probabilitas, menjadikannya pilihan yang kuat untuk analisis sentimen di tingkat awal (Putra & Wibowo, 2022).

Meskipun telah banyak studi yang menerapkan analisis sentimen untuk memprediksi opini publik terkait tokoh politik, penelitian yang secara spesifik berfokus pada sentimen terhadap calon presiden Anies Baswedan di media sosial X masih memerlukan eksplorasi yang lebih mendalam. Penting untuk mengidentifikasi bagaimana sentimen publik

berkembang dari waktu ke waktu, terutama dalam konteks momen-momen krusial seperti debat calon presiden. Dengan demikian, permasalahan yang diangkat dalam penelitian ini adalah bagaimana algoritma Naïve Bayes dapat secara efektif mengklasifikasikan sentimen opini publik (positif, negatif, dan netral) terhadap calon presiden Anies Baswedan, khususnya dengan memanfaatkan data *tweet* yang dikumpulkan selama periode debat Pemilu 2024. Penelitian ini juga berupaya mengevaluasi performa algoritma tersebut dengan menghitung nilai akurasinya.

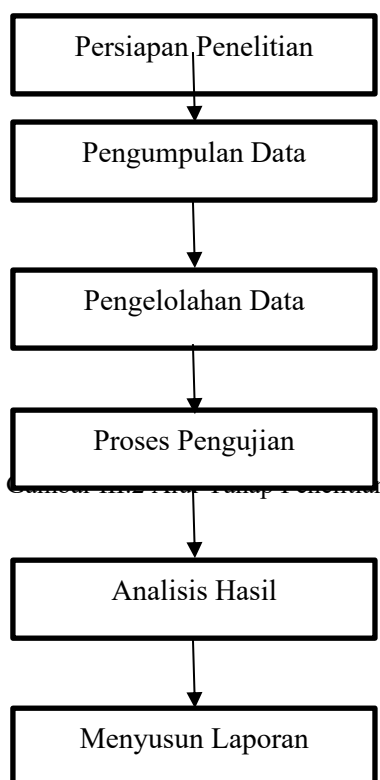
Tujuan, Kebaruan dan Urgensi Penelitian

Berdasarkan permasalahan yang telah diuraikan, penelitian ini memiliki beberapa tujuan utama. Pertama, untuk mengembangkan model klasifikasi sentimen menggunakan algoritma Naïve Bayes yang dapat menganalisis dan mengkategorikan *tweet* mengenai calon presiden Anies Baswedan ke dalam sentimen positif, negatif, atau netral. Kedua, untuk mengevaluasi kinerja model yang dikembangkan dengan menghitung nilai akurasi yang diperoleh dari setiap dataset yang digunakan. Urgensi dari penelitian ini terletak pada pemanfaatannya sebagai alat untuk mendapatkan wawasan kuantitatif dan kualitatif mengenai opini publik, yang dapat menjadi bahan pertimbangan penting bagi analisis politik dan pemangku kepentingan. Kebaruan (*novelty*) penelitian ini terletak pada penggunaan dataset yang diambil secara spesifik dari tujuh periode waktu berbeda seputar sesi debat calon presiden, yang memungkinkan analisis sentimen secara dinamis untuk mengamati perubahan opini publik. Dengan demikian, penelitian ini berkontribusi dalam memperkaya literatur mengenai penerapan algoritma Naïve Bayes dalam analisis sentimen, khususnya dalam konteks politik Indonesia yang terus berkembang.

METODE

Tahapan Penelitian

Adapun tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut :



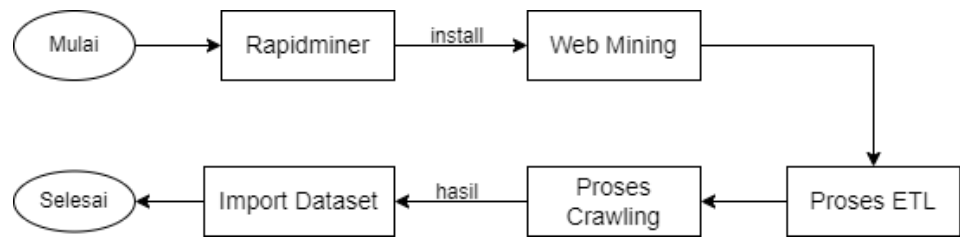
Persiapan Penelitian

Pada tahap persiapan penelitian dilakukan analisa terhadap kebutuhan yang diperlukan dalam penelitian . Persiapan yang dilakukan dengan mempersiapkan pendaftaran akun twitter untuk melakukan. Persiapan alat bantu untuk pengumpulan data tweet yang dibutuhkan menggunakan perangkat lunak rapidminer untuk melakukan crawling data. Untuk mendapatkan data dari twitter menggunakan rapidminer, dapat menggunakan ekstensi twitter API yang telah disediakan oleh rapidminer. Dengan menggunakan ekstensi tersebut bisa langsung terhubung ke API twitter dan mengambil data sesuai kebutuhan.

Pengumpulan Data

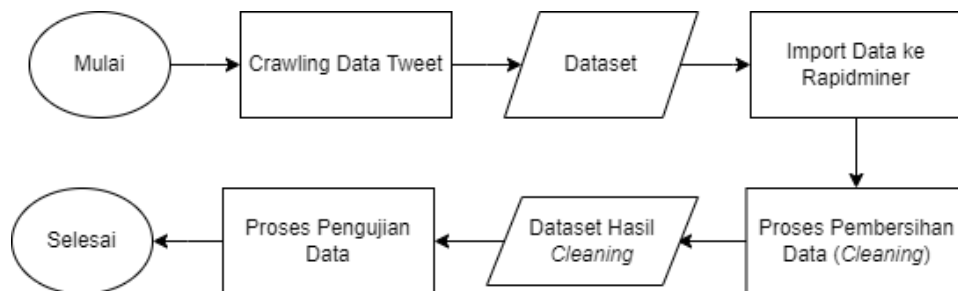
Penelitian ini mengumpulkan dataset dengan rentang waktu persesi debat, dataset yang diambil yaitu satu minggu sebelum debat capres, debat pertama, debat kedua, debat ketiga, debat keempat, debat kelima,debat keenam, dan satu minggu setelah debat, ketujuh dataset ini diambil di X melalui tahapan *crawling* data menggunakan perangkat lunak RapidMiner untuk Tarik data berupa tweet. Pada tahapan pengumpulan dataset ini tweet yang akan diambil hanya dalam bahasa Indonesia. Dengan *search keyword* #Capres Anies Baswedan. Data yang diambil melalui media online X.

Langkah –langkah umum yang dapat dilakukan untuk melakukan crawling data twitter dengan menggunakan perangkat lunak rapidminer :



Gambar 1. Tahapan Crawling Data Tweed

Pengolahan Data



Gambar 2. Tahapan Pengelolaan Data

Text Mining

1. Tokenize

Tahap tokenizing merupakan tahap memisahkan seluruh urutan karakter yang diinputkan menjadi satu bagian kata berdasarkan tiap kata yang menyusunnya dengan cara memecah kata berdasarkan spasi.

2. Filtering

Tabel.III.3 Filtering Data Tweed

username	tweed	result
FSultalaksana	Bukti paslon ini punya konsep tentang environmental ethics.yang sebelah kena MKMK ethnic.	(bukti), (paslon), (konsep), (enviromental), (ethics), (MKMK),(ethnic)

Tahap filtering adalah tahap mengambil kata-kata penting dari hasil tokenizing dan menghilangkan kata-kata yang tidak memiliki makna.

TF-IDF

Dalam perhitungan bobot menggunakan TF-IDF, data tweed yang dihitung untuk mendapatkan jumlah bobot dari masing-masing kata sebelum diklasifikasi. Sehingga nilai IDF di rumuskan pada persamaan berikut :

$$IDF(word) = \log D df$$

Keterangan :

IDF (word) : Nilai IDF dari setiap kata

D : Total dokumen

Df : Total kemunculan kata disemua dokumen

Adapun persamaan yang digunakan untuk menghitung bobot (W) pada masing-masing dokumen Rapidminer. Rapidminer Studio melakukan proses dasar analisis klasifikasi yang bertujuan untuk menemukan pola dalam kumpulan data sehingga dapat dibedakan berdasarkan kategori.

Klasifikasi sentimen terhadap kunci, yaitu :

$$Wdt = tfdt * ID Ft$$

Keterangan : Wdt : Bobot dokumen ke-d terhadap kata ke-t

Tfdt : Banyak term yang dihitung pada seluruh dokumen

ID Ft : Invers dokumen frekuensi

Metode Naïve Bayes dalam proses pengklasifikasian memiliki dua tahapan yaitu tahap pelatihan dan tahap klasifikasi. Tahapan-tahapan pada proses pengklasifikasian Naive Bayes adalah pertama pembuatan *bag of words*. Bag

of words untuk mengumpulkan term pada masing-masing tweet pasangan calon presiden berdasarkan frekuensi kemunculan di tweet tersebut. Setelah *bag of words* dibuat kemudian adalah training.

HASIL DAN PEMBAHASAN

X merupakan platform media sosial yang memungkinkan pengguna untuk berbagi pesan singkat yang disebut "tweet". Setiap tweet memiliki batasan 280 karakter dan dapat mencakup teks, gambar, video, dan tautan. Pengumpulan Data X dalam penelitian ini menggunakan perangkat lunak Google Colab. Data yang diambil dari twitter berupa tweed yang memiliki Hastag #Capres Anis Baswedan.

Tahapan dalam crawling data twitter menggunakan google colab, setelah membuat akun google dan login masuk ke google colab, maka akan ada tambilan menu untuk membuat notebook baru dalam memulai proses awal crawling data, Setelah masuk dalam halaman kerja notebook baru, Tuliskan code codingan python dengan file name format 'capres anis.csv'. Masukkan *search keyword* untuk rentang waktu pengambilan tweed melalui google colab. Pada crawling data dalam pengumpulan data tweed ini di ambil berdasarkan sesi debat yang dilakukan capres dan cawapres . Hasil dari crawling data didapatkan tujuh dataset yang akan digunakan dan disimpan dalam bentuk format csv untuk di lakukan tahapan cleaning data agar data tweet tersebut bisa digunakan sebagai data latih dalam perhitungan algoritma naïve bayes.

Dataset merupakan kumpulan data yang disusun dalam format tertentu, biasanya terdiri dari baris dan kolom, data ini yang akan digunakan untuk analisis sentimen dan pelatihan model mechine learning. dataset yang telah di crawling akan disimpan dalam format CSV.

Cleaning dataset merupakan langkah awal yang harus dilakukan untuk memastikan data yang akan di analisis sentimennya memiliki kualitas.

Gambar 2. Proses *Cleaning Dataset*

Pelebelan Sentimen

Pelebelan sentimen merupakan proses dalam memberikan label sentimen pada suatu text atau data berdasarkan emosi atau opini yang diekspresikan dalam tweet yang terdapat pada X. Pelebellan yang akan digunakan berupa kategori sentimen positif, negatif, dan netral yang dilakukan secara manual pada ketujuh dataset yang telah di cleaning.

2	berlainan hasil	
3	netral	Menika semua saling eysang pok
4	positif	Anies memang nomor 1 dalam debat capres pertama
5	positif	Ngap Di survey 62 unggah di Jalar Kayakng itu Anies go terfalu ngaruh di alur nungtup
6	positif	Mas Anies tolong adnan dibalik layar Soedma ini diberikan data nya mas. Seperitnya orang ini mudah menguqu seperti Prabowo
7	negatif	Untuk Indonesia satu ini memang belum pantas dipanggil alih mas Anies
8	positif	Kerenga kama berulah jadi tufel Dhu dhalang Anies kama kamyen cerdas. Age tala bisa dibongkar pingsir
9	negatif	Milih presiden kaku kaku orangnya dan gik pihk asias memuat saya bawak purnam mbak kafa
10	netral	Ngap pambuko bilang di yg paling saya ingit mbk ekspres Pak Ganjar itu Pak Anies bawak kudu banget ngap pphl lah bisa apa tp knp saya ngakuk ya
11	negatif	Sial kaku kafa Anies Bawakan dibalik layar Soedma
12	netral	Jawaban pok Anies akan statement ini juga paku jadi socran
13	negatif	menakuk 67% sama dengan bodoh bawangan dan Soedma dalam memahamai undang di sini terfalu jelas kaku Anies sama modal orang dan asal beda kaku baw
14	negatif	Anies Cacar Prabowo kaku Putusan MK Hutan Mamanu
15	negatif	Raga kaku pendukung Anies
16	netral	Insosian Pemarah Menyusukan superintec kepada pok Anies amp. Sanja Tidak cukup umur untuk jadi Cawapres
17	negatif	Pak Sanja Anas Tidak Layak Dukungin Perubahan Rangka paku menguqu harsan dan PK bagian dan Perencanaan
18	positif	PAK SANJAWAN Memulau dan cawap secara mental dan wawasan
19	positif	Anies moutap
20	netral	Siapa yang ngaku anasandri baw itu terfalu bentak sama branding jaman sanib program baru paku tidak asari crangry words hengk sandi cabut
21	positif	LURU U SERIKCHU PAK ANES
22	positif	Isak dhalang pok anes dan gik anin baw kulum di Indonesia ini bisa bawar dngakuk
23	positif	Pak Anies gawong pater lagi

Pemodelan Data Latih dan Data Uji

Dalam proses penelitian ini, keseluruhan dataset dibagi menjadi dua bagian utama, yaitu data latih dan data uji. Tahapan ini krusial untuk memberikan label sentimen pada data yang belum terklasifikasi. Data latih, yang merupakan porsi terbesar dari dataset, berfungsi sebagai fondasi untuk membangun sebuah model analitik. Tujuannya agar perangkat lunak seperti Rapidminer dapat mempelajari pola sentimen dari data yang sudah diberi label secara manual. Lebih lanjut, model yang dibangun dari data latih ini akan menyesuaikan parameternya secara optimal, sehingga mampu menghasilkan prediksi yang akurat untuk analisis sentimen di masa mendatang.

Data uji, di sisi lain, merupakan subhimpunan data yang sengaja dipisahkan dari proses pelatihan. Fungsi utamanya adalah untuk memvalidasi dan mengukur kinerja model yang telah dibangun. Berbeda dengan data latih, data uji ini belum pernah 'dilihat' oleh model selama proses pelatihan. Dengan demikian, pengujian menggunakan data ini memungkinkan peneliti untuk mengevaluasi kemampuan model secara objektif dalam membuat prediksi pada data baru yang belum dikenali. Ini penting untuk menilai sejauh mana model dapat digeneralisasi dan memastikan keandalannya dalam skenario dunia nyata.

KESIMPULAN

Berdasarkan analisis sentimen yang telah dilakukan terhadap opini publik di media sosial X terkait calon presiden Anies Baswedan, penelitian ini berhasil mengaplikasikan algoritma Naïve Bayes untuk mengklasifikasikan *tweet* ke dalam kategori sentimen positif, negatif, dan netral. Hasil yang diperoleh dari tujuh dataset menunjukkan bahwa algoritma ini memiliki kinerja yang sangat baik, dengan nilai akurasi yang konsisten di atas 84%. Secara spesifik, akurasi tertinggi mencapai 98,40%, sementara akurasi terendah berada pada 84,60%. Temuan ini secara kuantitatif membuktikan bahwa Naïve Bayes adalah metode yang efektif dan andal untuk tugas klasifikasi sentimen pada data teks yang kompleks. Lebih lanjut, dari analisis yang dilakukan, empat dari tujuh dataset menunjukkan frekuensi sentimen negatif yang paling dominan, sementara sisanya didominasi sentimen positif dan netral. Hal ini mengindikasikan bahwa secara umum, opini publik di X terhadap Anies Baswedan selama periode debat cenderung kurang baik atau lebih banyak memberikan tanggapan negatif.

Meskipun demikian, ada beberapa keterbatasan yang perlu diakui dari penelitian ini. Pertama, dataset yang digunakan terbatas pada *tweet* berbahasa Indonesia dan hanya berfokus pada satu calon presiden, sehingga generalisasi temuan terhadap kandidat lain atau sentimen publik yang lebih luas menjadi terbatas. Kedua, proses pelebelan sentimen (positif, negatif, netral) dilakukan secara manual. Hal ini bisa saja menimbulkan bias subjektif dan tidak sepenuhnya merepresentasikan makna atau emosi yang sesungguhnya dari setiap *tweet*, terutama yang menggunakan bahasa informal, sindiran, atau slang. Selain itu, algoritma Naive Bayes bekerja dengan asumsi independensi fitur, yang mungkin tidak selalu akurat dalam konteks bahasa alami di mana kata-kata seringkali memiliki ketergantungan kontekstual. Keterbatasan-keterbatasan ini dapat mempengaruhi interpretasi hasil dan menjadi ruang untuk perbaikan di masa depan.

Berdasarkan temuan dan keterbatasan tersebut, saran untuk penelitian selanjutnya adalah sebagai berikut: Pertama, peneliti dapat memperluas cakupan dataset dengan menyertakan *tweet* yang lebih beragam, termasuk opini terhadap kandidat lain, untuk melakukan perbandingan dan mendapatkan gambaran yang lebih komprehensif. Kedua, disarankan untuk mengimplementasikan metode pelebelan sentimen semi-otomatis atau menggunakan *annotator*

independen untuk meminimalisir bias subjektif dan meningkatkan kualitas label data. Ketiga, untuk meningkatkan akurasi dan mengatasi keterbatasan Naïve Bayes, peneliti dapat mencoba membandingkan kinerjanya dengan algoritma yang lebih canggih, seperti *Deep Learning* (misalnya, LSTM atau BERT) yang lebih mampu menangani kompleksitas konteks bahasa. Dengan demikian, penelitian di masa depan dapat memberikan wawasan yang lebih akurat dan mendalam mengenai dinamika sentimen politik di media sosial.

DAFTAR PUSTAKA

- Budi, S. (2017). Text mining untuk analisis sentimen review film menggunakan algoritma K-Means Text Mining For Movie Review Sentiment Analysis Using K-Means Algorithm. *Jl. Imam Bonjol No, 16*(1), 1–8. <http://www.cs.cornell.edu/People/pabo/movie-review-data/>
- Dewi, A. R., Diana, S., Fakhrezi, M. A., Awang, N., Ma'arif, H., & Saputra, D. D. (2023). Sentimen Analisis Terhadap Puan Maharani Sebagai Kandidat Calon Presiden 2024 Berdasarkan Opini Twitter Menggunakan Metode Naive Bayes Dan Adaboost. *JSiI (Jurnal Sistem Informasi)*, *10*(1), 75–80. <https://doi.org/10.30656/jsii.v10i1.5785>
- Putra, P. A. W., & Wibowo, W. S. (2022). A systematic literature review on open government data: Challenges and mapped solutions. *Journal of Theoretical and Applied Information Technology*, *101*(5), 1806-1818.
- Rosyida, T., Putro, H. P., & Wahyono, H. (2023). Analisis Sentimen Terhadap Pilpres 2024 Berdasarkan Opini Dari Twitter Menggunakan Naïve Bayes Dan Svm. *Teknokris*, *26*(1), 23–32. <https://doi.org/10.61488/teknokris.v26i1.245>
- Setiawan, R., Alfin, T., Roudhotul, I., Adawiyah, I., & Susilo, B. (2021). Analisis Sentimen Publik Terhadap Peraturan Daerah Tentang Kawasan Tanpa Rokok Menggunakan Metode Naive Bayes Classifier dan Lexicon Based. *Jurnal Inovasi Pendidikan dan Sains*, *2*(2), 115-122.
- Suryadi, A., Wibowo, R., & Handayani, M. (2022). Penerapan Algoritma Naive Bayes dan Support Vector Machine untuk Analisis Sentimen Opini Publik di Media Sosial Mengenai Kebijakan PPKM. *Jurnal Teknologi Informasi dan Ilmu Komputer (JUTIKOM)*, *9*(1), 1-8.