



Comparison of the NBC and C4.5 Algorithms in Sentiment Analysis for The 2024 Presidential Election on Twitter

Perbandingan Algoritma NBC dan C4.5 Dalam Analisa Sentimen Pemilihan Presiden 2024 Pada Twitter

Bayu Delvika^{1*}, Apriana², Naufal Abror³, Umairah Rizky Gurning⁴

^{1,2,3,4} Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

E-Mail: ¹12050313245@students.uin-suska.ac.id, ²12050323247@students.uin-suska.ac.id,
³12050313965@students.uin-suska.ac.id, ⁴11950320687@students.uin-suska.ac.id

Corresponding Author: Bayu Delvika

Abstract

The presidential election is a form of political activity that is people's sovereignty, because at the time of election, the people become the determining party in the implementation of politics in a place through voting which is carried out simultaneously. Problems in the presidential election that often arise in society are differences of opinion on social media networks, especially on Twitter. In this study, we conducted a comparison of the NBC and C4.5 algorithms in sentiment analysis for the 2024 presidential election using the tf-idf method on Twitter. The results showed that the NBC algorithm has a higher accuracy of 99.78% in the experiment with Hold Out, 90% training data and 10% test data, compared to the C4.5 algorithm which has an accuracy of 99.71%.

Keyword: C4.5, Naïve Bayes, Presidential Election 2024, Sentiment Analysis, Twitter

Abstrak

Pemilihan presiden merupakan bentuk kegiatan politik yang bersifat kedaulatan rakyat, dikarenakan pada saat pemilihan, rakyat menjadi pihak penentu dalam pengerjaan politik pada suatu tempat lewat pemungutan suara yang dilaksanakan secara serentak. Permasalahan pada pemilihan presiden yang sering kali muncul di masyarakat merupakan perbedaan pendapat di jejaring media sosial terutama di Twitter. Pada penelitian ini melakukan Perbandingan Algoritma NBC dan C4.5 Dalam Analisa Sentimen Pemilihan Presiden 2024 Menggunakan Metode tf-idf Pada Twitter. Hasil penelitian diketahui bahwa algoritma NBC mempunyai akurasi yang lebih tinggi yaitu 99,78% pada percobaan dengan Hold Out, 90% data latih dan 10% data uji, dibandingkan dengan algoritma C4.5 yang memiliki akurasi 99,71%.

Kata Kunci: Analisis Sentimen, C4.5, Naïve Bayes, Pemilihan Presiden 2024, Twitter

1. PENDAHULUAN

Pemilihan presiden merupakan bentuk kegiatan politik yang bersifat kedaulatan rakyat, dikarenakan pada saat pemilihan rakyat menjadi pihak penentu dalam pengerjaan politik pada suatu tempat lewat pemungutan suara yang dilaksanakan secara serentak. Berdasarkan pasal 1 ayat (22) UU No. 10 Tahun 2008 dan pasal 19 ayat 1 dan 2 UU No. 10 Tahun 2008, syarat keikutsertaan dalam pemilihan Presiden harus mencapai usia 17 tahun atau telah menikah [1]. Sesuai dengan aturan pilpres yang dilaksanakan pertama kali pada tahun 2004 [2]. Ada hal yang menarik saat Pilpres pada tahun 2019 di mana para kandidat memanfaatkan media sosial seperti twitter dalam melakukan kampanye [3].

Permasalahan pada pemilihan presiden yang sering kali muncul di masyarakat merupakan perbedaan pendapat di jejaring media sosial terutama di Twitter. Dengan adanya kampanye menggunakan twitter menjadi hal yang baru dalam dunia politik [4]. Namun tidak dapat dipungkiri bahwa dalam proses penyelenggaraan

pemilu di Indonesia harus di akui masih di warnai oleh berbagai permasalahan, seperti adanya aktivitas Black Campaign [5].

Black Campign ialah sindiran atau rumor yang bisa ditemukan pada sosial media. Analisa sentimen atau diketahui juga dengan opinion mining ialah proses penggalian sudut pandang atau opini dari dokumen terhadap suatu informasi tertentu untuk memahami pandangan publik mengenai informasi yang berkembang di media sosial berhubungan dengan pemilihan presiden [6]. Kegiatan analisa sentimen dilaksanakan untuk mengetahui pendapat publik terhadap suatu peristiwa atau topik, baik itu opini bersifat positif atau negatif. Metode yang umum dipakai ialah Text Mining. Penambangan teks ialah metode untuk mengevaluasi suatu informasi yang bermanfaat dari suatu teks data yang tidak terpola. Penambangan teks mencari kata kunci atau mengekstrak sentimen dan analitik dari analitik teks sehingga bisa membantu Anda memahami sentimen publik dalam data teks [7]. Pengambilan data teks Twitter untuk analisis sentimen karena Twitter merupakan platform media sosial yang sangat populer dan memiliki jumlah pengguna yang besar. Ini membuat data Twitter menjadi sumber yang sangat potensial untuk melakukan analisis sentimen, karena memungkinkan untuk memperoleh umpan balik dan pandangan langsung dari masyarakat pada berbagai topik dan isu.

Dalam analisis sentimen ada sebagian metode seperti Naïve Bayes dan C4.5. NBC adalah algoritma analisis statistik yang melaksanakan pemrosesan data pada data yang bersifat numerik dengan menerapkan probabilitas Bayesian. Sedangkan C4.5 dipakai untuk memahami klasifikasi dan prediksi pola dari data [8]. menurut penelitian dari Firman Tempola, Miftah Muhammad, Amal Khairan dengan judul Perbandingan Klasifikasi Antara KNN dan Naive Bayes Pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation mendapatkan hasil bahwa algoritma NBC mempunyai akurasi yang lebih tinggi [9], sedangkan menurut penelitian dari Eka Fitriani dengan judul Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan menyatakan bahwa algoritma C4.5 mempunyai akurasi yang lebih tinggi dibandingkan NBC [10]. Oleh karena itu, membandingkan kedua metode ini membantu dalam memutuskan metode mana yang lebih sesuai untuk digunakan dalam situasi tertentu. Sistem ini diterapkan untuk menggolongkan data komentar untuk di analisis sentimen. Untuk menggolongkan sentimen, kami akan menerapkan data preprocessing untuk muatan tf-idf [11]. Dalam dunia nyata Naive Bayes adalah algoritma klasifikasi sederhana dalam penerapannya, sedangkan algoritma C.45 pada salah satu penelitian menggunakan klasifikasi pohon keputusan, seperti studi Purushottam, Saxena dan Sharma (2016) dan Kumar dan Umatejaswi (2017) mendapatkan tingkat akurasi yang tinggi [12]. Dalam penelitian ini, peneliti ingin mengetahui metode mana yang memiliki daya kerja terbaik untuk diimplementasikan dalam opini analisis sentimen masyarakat dalam pemilihan presiden 2024.

Berdasarkan penelitian terdahulu yang membedakan penelitian ini yakni mengerjakan perbandingan tingkat akurasi terhadap algoritma NBC dan C4.5. Dimana C4.5 memiliki Struktur yang sederhana dan mudah ditafsirkan sehingga algoritma ini bisa untuk memecahkan masalah atribut multi-type. Dan bukan hanya itu C4.5 juga bisa mengelola data yang noise dan nilai-nilai yang hilang [13]. Sedangkan NBC dikenal dengan metodenya yang sederhana namun memiliki hasil dengan performa yang tinggi dalam pengklasifikasian. Sehingga akan menampilkan algoritma terbaik untuk hasil penelitian ini [14].

Adapun harapan untuk penelitian ini untuk kedepannya dalam analisis data sentimen pada twitter dengan obyek Pilpres ini dapat membangun sebuah aplikasi sehingga dapat menjadi perbandingan antara hasil penelitian ini dengan aplikasi yang dibangun dengan data yang lebih banyak dan dengan pemakaian metode machine learning lainnya supaya mengetahui perbedaannya.

2. METODOLOGI PENELITIAN

Setiap tahapan pada penelitian ini dideskripsikan dan digambarkan melalui flowchart dari awal hingga akhir yang dapat dilihat pada gambar 1.

2.1 Twitter

Twitter adalah layanan jejaring sosial yang sangat populer dikalangan pegiat media sosial, twitter dapat digunakan sebagai media untuk memberikan presepsi tentang kejadian yang sedang populer di bicarakan [13]. Twitter juga manfaat sebagai fasilitas dalam menyampaikan pendapat, seperti memberikan kritik, menjadi sarana kampanye politik, sarana belajar, sarana media komunikasi dan sebagai sarana hiburan [14].

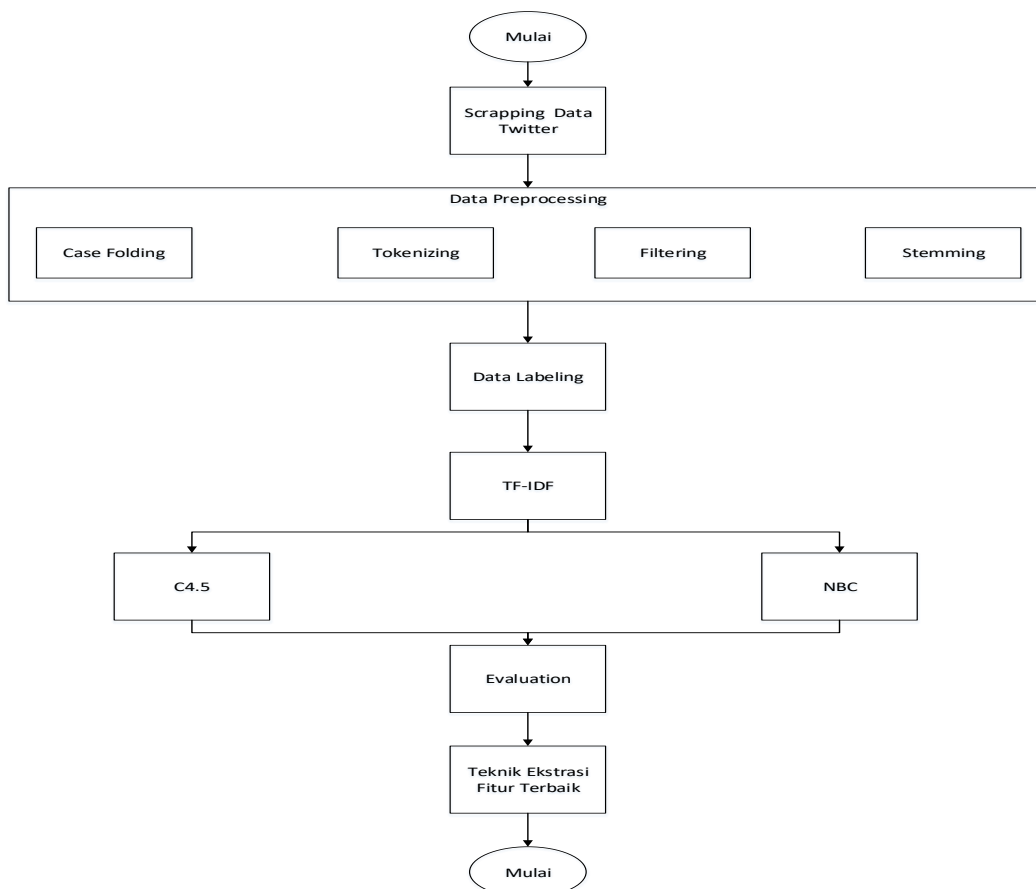
2.2 Crawling Data

Crawling data merupakan suatu proses dalam mengumpulkan data-data ulasan atau komentar menggunakan Bahasa pemograman python. Data dalam penelitian ini didapatkan melalui tweet di twitter berjumlah 1000 data.

2.3 Data Preprocessing

Preprocessing merupakan langkah langkah mengubah teks asli untuk menghapus bagian teks yang tidak diperlukan dalam pengolahan lebih lanjut [15].

- a. Case Folding
Case Folding merupakan proses mengganti huruf besar menjadi kecil
- b. Tokenizing
Tokenizing merupakan proses memotong string input atas tiap kata yang menyusunnya. Setiap karakter yang selain huruf dalam suatu kata akan diapus atau dihilangkan untuk proses memperoleh kata-kata penyusun teks [16].
- c. Filtering
Filtering adalah proses menghilangkan data yang tidak lengkap dan juga data yang mempunyai kesalahan [17]. Filtering juga melakukan proses stopwords removal dengan cara menghapus kalimat yang cukup umum dan sering muncul tetapi tidak memiliki dampak yang berarti untuk definisi kalimat [18].
- d. Stemming
Untuk menemukan kata dasar dalam sebuah teks menggunakan proses stemming [19]. Tujuan dari proses stemming adalah mengubah kata dasar dengan menghapus kata imbuhan pada dokumen [20].



Gambar 1. Metodologi Penelitian

2.4 Data Labeling

Setelah dilakukannya text preprocessing, selanjutnya adalah memberikan labelling untuk setiap data komentar, untuk memberikan label positive, negative, dan neutral. Pelabelan data dalam penelitian ini menggunakan library textblob yang ada pada bahasa pemrograman python.

2.5. TF-IDF

Setelah dilakukannya text preprocessing dan data labelling selanjutnya adalah melakukan pembobotan pada teks. Salah satu metode yang populer dalam melakukan pembobotan data adalah TF-IDF [21]. TF-IDF adalah suatu cara dalam pemberian bobot hubungan suatu kata (term) terhadap dokumen [22].

2.6. NBC

Klasifikasi naïve bayes adalah algoritma dalam teknik data mining yang menggunakan teori Bayes untuk melakukan klasifikasi [23]. NBC adalah algoritma klasifikasi yang digunakan untuk prediski probabilitas keanggotaan suatu kelas [24]. Naïve bayes merupakan pendekatan statistik dasar untuk pengenalan pola. Naive Bayes didasarkan pada asumsi yang disederhanakan bahwa nilai atribut secara kondisional independen ketika diberi nilai keluaran [25].

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

2.7. C4.5

Algoritma C4.5 adalah algoritma yang melakukan pengambilan dataset yang mempunyai label dan membentuk pohon keputusan sebagai output [26]. C4.5 adalah metode klasifikasi data. Decision tree memiliki model seperti pohon yang terdiri dari simpul akar, simpul internal, dan simpul terminal. Sementara simpul akar dan dalam adalah variabel/fitur, simpul daun adalah label kelas [27].

3. HASIL DAN PEMBAHASAN

3.1 Web Scrapping

Web Scrapping digunakan dengan aplikasi web Jupyter Lab dan bahasa pemrograman Python pada twitter didapatkan sebanyak 1000 data ulasan. Instalasi library snsrape.modules.twitter merupakan langkah pertama yang digunakan untuk melakukan scrapping data pada ulasan twitter dengan memasukan kata kunci pada twitter yang ingin diambil ulasannya. Data yang berhasil di scrapping lalu diekspor dalam bentuk excel (.xlsx.). Contoh dokumen dihasilkan dari web scrapping adalah sebagai berikut.

Tabel 1. Contoh Dokumen Hasil Scrapping

Dokumen	Ulasan
Dokumen 1	@2001Tgk @VIVAcoid Antisipasinya sarankan aj partai atau calon yang anda dukung jangan ikut pilpres dari pada sia sia beritau sekarang dari pada nanti kalah betulan trus bilang di zolimi bener g bro ? G ikut akan lebih baik bro saran saya sih
Dokumen 2	@ainurohman Jagoan Deny S pilpres Pak Ganjar Pranowo berarti =
Dokumen 3	@SemutMerah2024 @aniesbaswedan Klo anda ngajak seluruhnya untuk pilih Anis baswedan..Iyah GK usah pilpres LG bro.
...	...
Dokumen x	Anies didukung Grdra di Pilgub&Prabowo didukung pedukung Anies di Pilpres yll walau sll kalah.Agar Grdra tdk kalah terus,mk ganti mencapreskan Anies di 2024 dp musuhan. Anies Dianggap Kacang Lupa Kulit oleh Gerindra, Pengamat Ini Kasih Komentar Menohok https://t.co/zjjGSJjXez

3.2 Text Preprocessing

Sebelum melakukan implementasi algoritma klasifikasi pada dokumen, dilakukannya tahapan preprocessing. Berikut adalah tahapan preprocessing.

a. Case Folding

Merupakan proses yang digunakan untuk mengganti huruf besar menjadi kecil untuk menyamakan hasil teks. Contoh hasil dari penerapan case folding dapat dilihat pada tabel 2.

Tabel 2. Proses Case Folding

antisipasinya sarankan aj partai atau calon yang anda dukung jangan ikut pilpres dari pada sia sia beritau sekarang dari pada nanti kalah betulan trus bilang di zolimi bener bro ikut akan lebih baik bro saran saya sih

b. Tokenizing

Tokenizing merupakan proses untuk membagi kalimat menjadi bagian-bagian tertentu. Tokenizing mempunyai fungsi untuk menghapus beberapa karakter yang dianggap sebagai tanda baca. Contoh hasil dari penerapan tokenizing dilihat pada tabel 3.

Tabel 3. Proses Tokenizing

['antisipasinya', 'saranakan', 'aj', 'partai', 'atau', 'calon', 'yang', 'anda', 'dukung', 'jangan', 'ikut', 'pilpres', 'dari', 'pada', 'sia', 'sia', 'beritau', 'sekarang', 'dari', 'pada', 'nanti', 'kalah', 'betulan', 'trus', 'bilang', 'di', 'zolimi', 'bener', 'bro', 'ikut', 'akan', 'lebih', 'baik', 'bro', 'saran', 'saya', 'sih']
--

c. Filtering

Filtering merupakan proses menghapus data yang tidak lengkap, dan yang mempunyai kesalahan [17]. Filtering mampu menghapus kata-kata yang umum dan sering muncul tetapi tidak mempunyai dampak berarti dalam suatu kalimat [18]. Contoh hasil dari penerapan tahapan filtering dilihat pada tabel 4.

Tabel 4. Proses Filtering

['klo', 'ngajak', 'pilih', 'anis', 'baswedaniyah', 'gk', 'pilpres', 'lg', 'bro']
--

d. Stemming

Stemming adalah metode yang digunakan dalam menghapus kata imbuhan, kata depan, kata ganti, dan kata akhir yang sama dengan KBBI. Penelitian ini menggunakan library sastrawi dalam melangsungkan stemming untuk teks bahasa indonesia. Hasil untuk proses stemming dapat dilihat pada tabel 5.

Tabel 5. Proses Stemming

['antisipasi', 'saran', 'partai', 'calon', 'dukung', 'pilpres', 'sia', 'sia', 'beritau', 'kalah', 'betul', 'trus', 'bilang', 'zolimi', 'bener', 'bro', 'bro', 'saran']
--

3.3 TF-IDF

Metode pembobotan yang digunakan pada penelitian ini adalah TF-IDF yang memberikan hasil berupa vector dengan banyak vector agar dapat dikenali untuk setiap katanya.

a. Term Frequency (TF)

Term Frequency adalah frekuensi kemunculan term i pada dokumen j dibagi jumlah total term pada dokumen j.

$$tf_{ij} = \frac{f_a(i)}{\max f_a(j)} \tag{2}$$

Sebagai contoh melakukan penghitungan nilai TF berdasarkan dokumen 1 yaitu:

['antisipasi', 'sarankan', 'aj', 'partai', 'atau', 'calon', 'yang', 'anda', 'dukung', 'jangan', 'ikut', 'pilpres', 'dari', 'pada', 'sia', 'sia', 'beritau', 'sekarang', 'dari', 'pada', 'nanti', 'kalah', 'betulan', 'trus', 'bilang', 'di', 'zolimi', 'bener', 'bro', 'ikut', 'akan', 'lebih', 'baik', 'bro', 'saran', 'saya', 'sih']

Jika dilakukan penghitungan menggunakan dokumen 1 di atas maka didapatkan hasil sebagai berikut:

{ 'antisipasi': 0.05, 'saran': 0.1, 'partai': 0.05, 'calon': 0.05, 'dukung': 0.05, 'pilpres': 0.05, 'sia': 0.1, 'beritau': 0.05, 'kalah': 0.05, 'betul': 0.05, 'trus': 0.05, 'bilang': 0.05, 'zolimi': 0.05, 'bener': 0.05, 'bro': 0.1 }

b. Inverse Document Frequency (IDF)

Proses ini digunakan supaya dokumen yang mempunyai term yang dicari dapat dilakukan penghitungan dengan memakai persamaan sebagai berikut.

$$idf = \ln \frac{N}{df} + 1 \tag{3}$$

Deskripsi:

- ln : logaritma natural
- N : besaran keseluruhan dokumen
- df : besaran kata dalam dokumen

Melalui proses coding, didapatkan frekuensi kata yang muncul yang muncul dalam satu dataset seperti pada tabel 6.

Tabel 6. Frekuensi Kata yang Muncul

0	[(ikut, 2), (dari, 2), (pada, 2), (sia, 2), (b...
1	[(jagoan, 1), (deny, 1), (pilpres, 1), (pak, 1...
2	[(klo, 1), (anda, 1), (ngajak, 1), (seluruhnya...
3	[(demokrat, 1), (surabaya, 1), (bulatkan, 1), ...
4	[(dari, 1), (verifikasi, 1), (aja, 1), (udah, ...

Selain itu, proses penghitungan nilai bobot TF-IDF dapat digunakan dengan cara pengkalian nilai TF dengan IDF agar diperoleh hasil seperti pada Tabel 7.

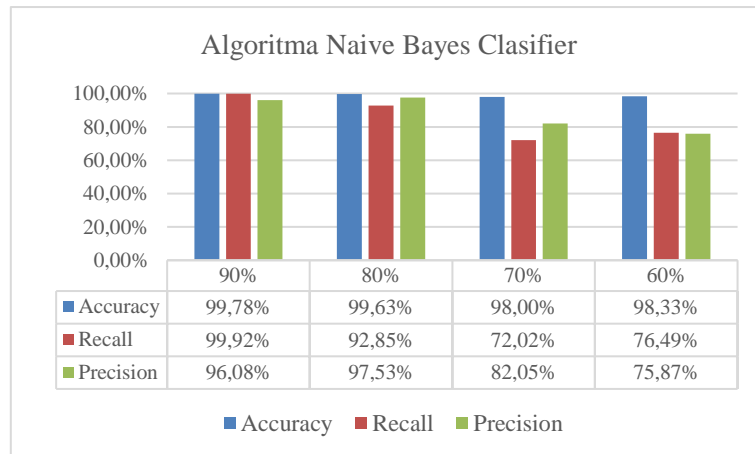
Tabel 7. Hasil TF-IDF

Dokumen	TF-IDF	Label
['antisipasi', 'saran', 'partai', 'calon', 'dukung', 'pilpres', 'sia', 'sia', 'beritau', 'kalah', 'betul', 'trus', 'bilang', 'zolimi', 'bener', 'bro', 'bro', 'saran']	[0.4347640611908164, 0.39745634199512153, 0.3615232419462758, 0.8695281223816328, 0.2612514154945242, 0.19358726486911185, 0.25534785567176116, 0.2882824227017351, 0.07208067036170289, 0.8695281223816328, 1.0543673705309515, 0.4077330539836055]	Neutral
['beliau', 'dukung', 'pilpres', 'ko', 'brutal', 'jawab']	[1.3171744873417197, 0.7259522432591694, 1.4904612824817058, 0.2703025138563858]	Negative
['new', 'post', 'panggil', 'mega', 'gibran', 'pasang', 'puan', 'pilpres']	[1.0497195610913392, 1.0193326349590137, 1.1348571650523378, 0.7403365626970684, 0.1802016759042572, 0.5607434161186888]	Positive

3.4 Penerapan algoritma NBC dan C4.5

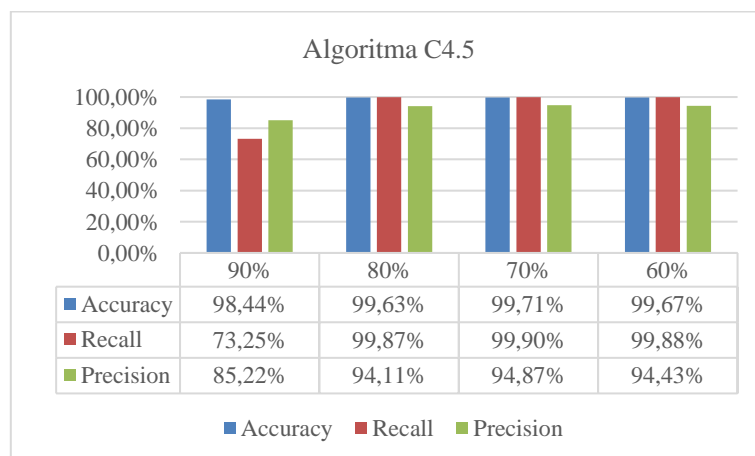
Di penelitian ini pembagian data menggunakan teknik hold out dengan perbandingan 90%:10%, 80%:20%, 70%:30% dan 60%:40%.

Dalam penelitian ini, dilakukan proses klasifikasi menggunakan dua algoritma salah satunya NBC. Berikut adalah hasil dari perbandingan evaluasi pada algoritma NBC dapat dilihat pada gambar 2.



Gambar 2. Algoritma Naive Bayes Clasifier

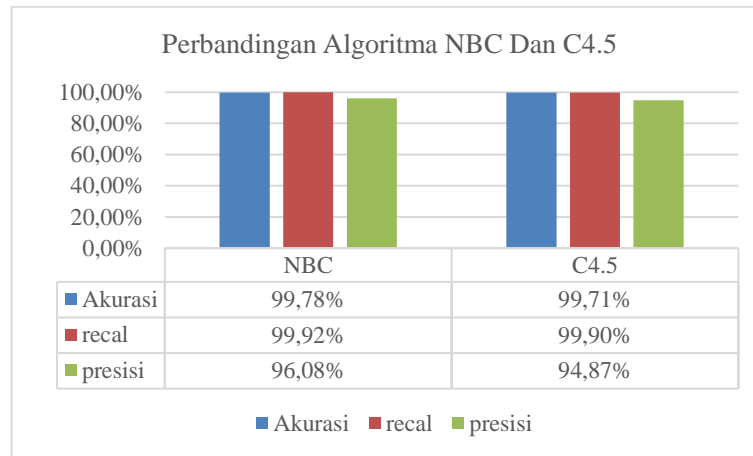
Algoritma lainnya yang digunakan pada penelitian ini adalah C.45 yang juga menggunakan pembagian data latih Hold Out. Hasil perbandingan evaluasi pada algoritma C.45 yang dapat dilihat pada gambar 3 berikut.



Gambar 3. Algoritma C4.5

3.5 Perbandingan Algoritma NBC dan C4.5

Adapun perbandingan algoritma NBC, C.45 dan K-NN dalam mengklasifikasi data Covid-19 dapat dilihat pada gambar 4.



Gambar 4. Perbandingan Algoritma NBC Dan C4.5

Berdasarkan Percobaan Algoritma Klasifikasi pada data sentiment analysis twitter khususnya tentang pemilihan presiden 2024 menghasilkan NBC sebagai algoritma terbaik dengan akurasi 99,78%, recall 99,92% dan presisi 96,08% pada percobaan dengan Hold Out 90% data latih dan 10% data uji. Sedangkan algoritma C4.5 memiliki akurasi 99,71%, recall 99,90% dan presisi 94,87%. Berdasarkan penelitian ini, diketahui bahwa algoritma NBC menghasilkan akurasi yang lebih baik dalam mengklasifikasi data sentiment analysis twitter khususnya tentang pemilihan presiden 2024.

4. Kesimpulan

Berdasarkan hasil penelitian dapat disimpulkan dengan melakukan analisa sentiment pemilihan presiden 2024 pada twitter yang menghasilkan data sebanyak 1000 data setelah dilakukan scrapping data. Setelah dilakukan klasifikasi data pemilihan presiden 2024 dengan menggunakan dua algoritma NBC dan C4.5, didapatkan akurasi 99,78%, recall 99,92% dan presisi 96,08% pada NBC, dan kurasi 99,71%, recall 99,90% dan presisi 94,87% pada C4.5. Dari hasil tersebut dapat dilihat bahwa nilai akurasi tertinggi didapat pada NBC. Dengan demikian algoritma NBC menjadi algoritma terbaik dalam melakukan klasifikasi data sentiment analisis pemilihan presiden 2024.

REFERENSI

- [1] R. Nur, A. Taufik, and M. Tahir, "Perilaku politik pemilih pemula dalam pelaksanaan pemilihan presiden 2014 di Desa Kanaungan Kecamatan Labakkang Kabupaten Pangkep," *Otoritas J. Ilmu Pemerintah.*, vol. 5, no. 1, 2015.
- [2] U. Rauta, "Menggagas pemilihan presiden yang demokratis dan aspiratif," *J. Konstitusi*, vol. 11, no. 3, pp. 600–616, 2016.
- [3] A. Akmaluddin and M. A. R. Wempie, "BUDAYA POLITIK PEMILIH MILENIAL DALAM PEMILU TAHUN 2024 DI KABUPATEN OGAN KOMERING ULU," *J. Ilmu Pemerintah. Unbara*, vol. 1, no. 1, pp. 49–57, 2022.
- [4] M. A. Firmansyah, S. Karlinah, and S. Sumartias, "Kampanye Pilpres 2014 dalam Konstruksi Akun Twitter Pendukung Capres," *J. Messenger*, vol. 9, no. 1, pp. 79–90, 2017.
- [5] L. Januru, "Analisis Wacana Black Campaign (Kampanye Hitam) Pada Pilpres Tahun 2014 di Media Kompas, Jawa Pos dan Kedaulatan Rakyat," *J. Natapraja Kaji. Ilmu Adm. Negara*, vol. 4, no. 2, 2016.
- [6] A. P. R. Nababan, A. S. M. Lumenta, Y. D. Rindengan, F. J. Pontoh, and Y. V. Akay, "Analisis Sentimen Twitter Pasca Pengumuman Hasil Pilpres 2019 Menggunakan Metode Lexicon Analysis," *J. Tek. Inform.*, vol. 15, no. 1, pp. 33–44, 2020.
- [7] I. Kurniawan and A. Susanto, "Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," *J. Eksplora Inform.*, vol. 9, no. 1, pp. 1–10, 2019.
- [8] I. Sutoyo, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik," *J. Pilar Nusa Mandiri*, vol. 14, no. 2, pp. 217–224, 2018.
- [9] F. Tempola, M. Muhammad, and A. Khairan, "Perbandingan Klasifikasi Antara KNN dan Naive Bayes

- pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, p. 577, 2018, doi: 10.25126/jtiik.201855983.
- [10] E. Fitriani, “Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan,” *Sistemasi*, vol. 9, no. 1, p. 103, 2020, doi: 10.32520/stmsi.v9i1.596.
- [11] A. Muhammadin and I. A. Sobari, “Analisis Sentimen Pada Ulasan Aplikasi Kredivo Dengan Algoritma SVM Dan NBC,” *Reputasi J. Rekayasa Perangkat Lunak*, vol. 2, no. 2, pp. 85–91, 2021.
- [12] D. W. Utomo, D. Kurniawan, and Y. P. Astuti, “Teknik pengujian perangkat lunak dalam evaluasi sistem layanan mandiri pemantauan haji pada kementerian agama provinsi jawa tengah,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 2, pp. 731–746, 2018.
- [13] M. A. Sembiring, M. F. L. Sibuea, and A. Sapta, “Analisa Kinerja Algoritma C. 45 Dalam Memprediksi Hasil Belajar,” *J. Sci. Soc. Res.*, vol. 1, no. 1, pp. 73–79, 2018.
- [14] S. R. Elisabet, A. M. Khairul, R. Rahmaddeni, and N. U. Aniq, “Perbandingan Algoritma Svm Dan Nbc Dalam Analisa Sentimen Pilkada Pada Twitter,” *CSRID J.*, vol. 13, no. 3, pp. 169–179, 2021.
- [15] H. Najjichah, A. Syukur, and H. Subagyo, “Pengaruh Text Preprocessing Dan Kombinasinya Pada Peringkat Dokumen Otomatis Teks Berbahasa Indonesia,” *J. Teknol. Inf.*, vol. XV, no. 1, pp. 1–11, 2019.
- [16] A. Riyani, M. Z. Naf’ an, and A. Burhanuddin, “Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen,” *J. Linguist. Komputasional*, vol. 2, no. 1, pp. 23–27, 2019, [Online]. Available: <https://doi.org/10.26418/jlk.v2i1.17>
- [17] N. C. Agustina, D. Herlina Citra, W. Purnama, C. Nisa, and A. Rozi Kurnia, “Implementation of Naïve Bayes Algorithm for Sentiment Analysis of Shopee Reviews on Google Play Store Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Goo,” vol. 2, no. April, pp. 47–54, 2022.
- [18] A. F. Hidayatullah, “Pengaruh Stopword Terhadap Performa Klasifikasi Tweet Berbahasa Indonesia,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 1, no. 1, pp. 1–4, 2016, doi: 10.14421/jiska.2016.11-01.
- [19] M. S. H. Simarankir, “Studi Perbandingan Algoritma - Algoritma Stemming Untuk Dokumen Teks Bahasa Indonesia,” *J. Inkofar*, vol. 1, no. 1, pp. 40–46, 2017, doi: 10.46846/jurnalinkofar.v1i1.2.
- [20] R. E. Aprilliwanto, A. Sanjaya, and D. W. Widodo, “Identifikasi Pola Kalimat Bahasa Indonesia Pada Siswa Sekolah Dasar Menggunakan Metode LALR dan Stemming,” pp. 119–126, 2021.
- [21] M. N. Saadah, R. W. Atmagi, D. S. Rahayu, and A. Z. Arifin, “Sistem Temu Kembali Dokumen Teks Dengan Pembobotan Tf-Idf Dan Lcs,” *JUTI J. Ilm. Teknol. Inf.*, vol. 11, no. 1, p. 19, 2013, doi: 10.12962/j24068535.v11i1.a16.
- [22] R. T. Wahyuni, D. Prastiyanto, and E. Suprpto, “Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi,” *J. Tek. Elektro Univ. Negeri Semarang*, vol. 9, no. 1, pp. 18–23, 2017, [Online]. Available: <https://journal.unnes.ac.id/nju/index.php/jte/article/download/10955/6659>
- [23] B. Delvika, S. Nurhidayarnis, and P. D. Rinada, “Perbandingan Klasifikasi Antara Naive Bayes dan K-Nearest Neighbor Terhadap Resiko Diabetes Pada Ibu Hamil,” vol. 2, no. October, pp. 68–75, 2022.
- [24] Y. E. Fadrial, “Algoritma Naive Bayes Untuk Mencari Perkiraan Waktu Studi Mahasiswa,” *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 4, no. 1, pp. 20–29, 2021, doi: 10.31539/intecomsv4i1.2219.
- [25] M. S. Mustafa, M. R. Ramadhan, and A. P. Thenata, “Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier,” *Creat. Inf. Technol. J.*, vol. 4, no. 2, p. 151, 2018, doi: 10.24076/citec.2017v4i2.106.
- [26] F. Akbar, H. Wira Saputra, A. Karel Maulaya, M. Fikri Hidayat, and Rahmaddeni, “Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke,” vol. 2, no. October, pp. 61–67, 2022.
- [27] N. T. Romadloni, I. Santoso, and S. Budilaksono, “Perbandingan Metode Naïve Bayes, KNN dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line,” *ikraith-informatika*, vol. 3, no. 2, pp. 1–9, 2019.