

Perbandingan Kinerja Algoritma *Decision Tree* Dan *Naïve Bayes* Dalam Analisis Sentimen Publik Terhadap Kebijakan Pemerintah Mengenai Batasan Penggunaan Media Sosial Anak

Amelia Mar'atusholihat ¹, Nuk Ghurroh Setyoningrum ², Dede Rizal Nursamsi ³

¹²³Universitas Cipasung Tasikmalaya, Padakembang, Tasikmalaya, 46466, Indonesia
e-mail: ameliamelly710@gmail.com ^{*1}, nuke@uncip.ac.id ^{*2}, dederizalnursamsi@uncip.ac.id ^{*3}

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi : 08 Oktober 2025

Revisi Akhir : 04 November

Diterbitkan Online : 10 November 2025

Kata Kunci :

Analisis Sentimen, *Decision Tree*, *Naïve Bayes*, Media Sosial, Pembatasan Usia

Korespondensi :

Telepon / Hp : +62 85317163032

E-mail : ameliamelly710@gmail.com

ABSTRAK

Perkembangan teknologi informasi menjadikan media sosial sebagai sarana utama untuk berinteraksi, berbagi informasi, dan mengekspresikan pendapat. Namun, tingginya penggunaan media sosial oleh anak-anak menimbulkan kekhawatiran terhadap dampak negatif yang ditimbulkan, sehingga pemerintah menetapkan kebijakan pembatasan usia anak dalam penggunaan media sosial. Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap kebijakan tersebut dengan menerapkan algoritma *Decision Tree* dan *Naïve Bayes*. Data penelitian diperoleh dari komentar pengguna TikTok terkait isu pembatasan usia. Tahapan penelitian meliputi pengumpulan data dengan web scraping, pra-pemrosesan (*translation, case folding, cleansing, tokenizing, stopword removal, stemming*), pelabelan sentimen (positif, negatif, netral), serta representasi data menggunakan *Bag of Words*. Dataset dibagi menjadi 80% data latih dan 20% data uji. Hasil pengujian menunjukkan bahwa algoritma *Naïve Bayes* memiliki performa lebih baik dibandingkan *Decision Tree*. *Naïve Bayes* mencapai akurasi 97,14%, presisi 97,08%, dan recall 96,92%, sedangkan *Decision Tree* memperoleh akurasi 91,90%, presisi 96,46%, dan recall 79,76%. Temuan ini membuktikan bahwa *Naïve Bayes* lebih efektif dalam analisis sentimen media sosial, terutama pada dataset dengan distribusi kelas beragam. Penelitian ini berkontribusi pada pemahaman penerapan algoritma klasifikasi dalam menganalisis opini publik terhadap kebijakan pemerintah, khususnya pembatasan usia media sosial. Selain itu, hasilnya dapat menjadi rujukan bagi pembuat kebijakan maupun penelitian selanjutnya dengan dataset lebih besar atau algoritma yang lebih kompleks.

1. PENDAHULUAN

Perkembangan teknologi informasi dan internet saat ini telah mengubah cara manusia dalam berkomunikasi [1]. Informasi dapat didapatkan dengan berbagai tempat, mulai dari media sosial (*YouTube*, Instagram, Facebook, Tiktok, X), dan portal berita. Perkembangan dari teknologi ini membawa media sosial sebagai bagian yang tak terpisahkan dari kehidupan kalangan termasuk anak-anak, remaja, orang dewasa dan orang tua [2].

Media sosial merupakan suatu aplikasi yang memungkinkan pengguna dapat melakukan berbagai aktivitas secara *online* seperti berbagi, diskusi, melakukan kreativitas serta dapat berpartisipasi dalam segala bentuk interaksi sosial lainnya [3]. Fenomena lain yang memperkuat peran media sosial dalam kehidupan adalah meningkatnya ketergantungan masyarakat terhadap ponsel pintar. Aktivitas seperti *scrolling* video, membuat konten, hingga mengikuti tren digital sudah menjadi rutinitas sehari-hari.

Berdasarkan laporan Kalodata, jumlah pengguna media sosial di awal tahun 2024 di Indonesia memiliki 185,3 juta pengguna, dengan penetrasi internet sebesar 66,5%. Pada bulan Januari 2024, Indonesia memiliki 139 juta pengguna media sosial, setara dengan 49,9% dari total populasi. Menurut analisis Kepios menunjukkan bahwa di Indonesia, pengguna internet meningkat sebanyak 1,5 juta orang (+0,8 persen) antara Januari 2023 sampai Januari 2024. Sementara itu, data yang

dipublikasikan dalam alat perencanaan iklan di platform media sosial ternama menunjukkan bahwa terdapat 126,8 juta.

Dari jumlah tersebut, sekitar 126,8 juta pengguna media sosial merupakan pengguna yang berusia 18 tahun keatas, yang berarti 12,2 juta pengguna media sosial di Indonesia yang berada di bawah 18 tahun. Angka ini menunjukkan bahwa anak-anak dan remaja memiliki akses terhadap media sosial. Hal ini dapat memicu adanya berbagai resiko bagi anak-anak dan remaja seperti kecanduan, paparan konten yang tidak layak, hingga ancaman keamanan digital.

Menurut (Ariani, 2017), masa remaja adalah masa perkembangan yang sangat sensitif dalam memiliki keinginan yang tinggi untuk bermain berbagai macam permainan dan menghabiskan waktu lebih banyak pada media sosial [4]. Kondisi ini menjadi perhatian serius, dimana di usia remaja dan anak-anak merupakan masa pembentukan identitas diri yang sangat krusial [5]. Saat ini, banyak remaja yang membagikan atau memposting kegiatannya di media sosial, serta tingkah remaja dalam menggunakan media sosial beragam ada yang positif, namun tidak sedikit pula yang bersifat negatif [6]. Tanpa disadari, kebiasaan-kebiasaan ini dapat memberikan pengaruh yang signifikan terhadap perkembangan mental anak-anak dan remaja [7].

Melihat dampak negatif dan permasalahan yang terjadi saat ini, pemerintah Indonesia membuat kebijakan

baru mengenai pembatasan usia anak dalam mengakses media sosial yang dimana tercantum di Peraturan Pemerintah Nomor 17 tahun 2025. Tujuan dari kebijakan ini adalah untuk melindungi anak-anak dari konten berbahaya dan menciptakan ruang digital yang aman dan edukatif. Namun, implementasi dari kebijakan ini tidak mudah, tantangan utamanya terletak pada kurangnya kesadaran masyarakat dan efektivitas pengawasan.

Untuk memahami bagaimana tanggapan masyarakat terhadap kebijakan ini, diperlukan pendekatan analisis sentimen. Analisis sentimen merupakan teknik yang digunakan untuk mengekstraksi, memahami, dan mengolah data opini secara otomatis guna mengidentifikasi sentimen yang terkandung dalam suatu pernyataan atau pendapat [8]. Dengan menggunakan teknik ini, dapat mengetahui apakah kebijakan pembatasan usia mendapatkan dukungan atau penolakan dari masyarakat.

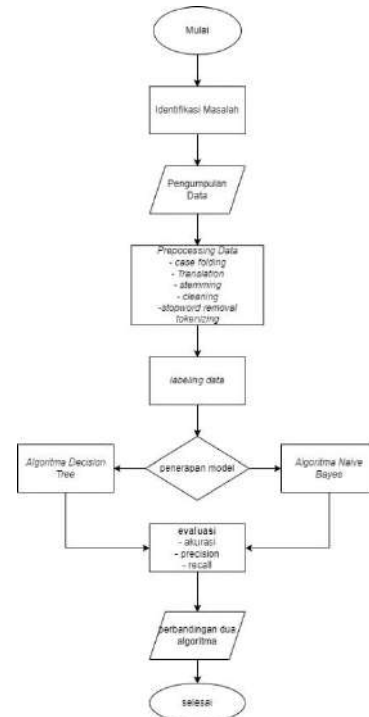
Dalam penelitian ini, akan digunakan dua algoritma dalam proses analisis sentimen, yaitu *Decision Tree* dan *Naïve Bayes*. Kedua algoritma ini termasuk ke dalam teknik klasifikasi dalam *data mining* yang mampu mengelompokkan data berdasarkan pola-pola tertentu. Namun, hasil dan kinerja dari masing-masing algoritma bisa berbeda berdasarkan pada karakteristik data yang digunakan. Oleh karena itu, perlu adanya perbandingan antara kedua algoritma untuk mengetahui manakah algoritma yang lebih efektif dalam konteks analisis sentimen terhadap kebijakan pemerintah ini.

Meskipun penelitian tentang analisis sentimen di media sosial telah banyak dilakukan, masih terdapat kesenjangan yang menjadikan topik ini penting untuk dikaji lebih lanjut. Pertama, belum banyak studi yang secara khusus meneliti respons masyarakat terhadap kebijakan pemerintah mengenai pembatasan usia anak dalam menggunakan media sosial, padahal kebijakan ini memiliki dampak langsung terhadap kehidupan anak-anak. Kedua, masih jarang penelitian yang menggunakan dua algoritma, kebanyakan peneliti sebelumnya hanya menggunakan satu algoritma. Ketiga, mayoritas penelitian sebelumnya hanya fokus pada *platform X (twitter)*, sedangkan *platform* lain seperti Tiktok dan Instagram yang kini lebih populer di kalangan remaja dan anak-anak. Hal ini menjadikan Tiktok dan Instagram sebagai sumber data yang lebih *representative* dalam menganalisis opini publik terkait kebijakan pembatasan usia penggunaan media sosial.

Penelitian ini bertujuan untuk menganalisis sentimen pengguna media sosial Tiktok dan Instagram terhadap kebijakan pemerintah mengenai pembatasan pengguna media sosial oleh anak-anak dengan menggunakan algoritma *Decision Tree* dan *Naïve Bayes* pada *platform RapidMiner*. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam pengembangan literasi digital, memberikan umpan balik bagi pembuat kebijakan, serta meningkatkan pemahaman mengenai efektivitas algoritma dalam analisis sentimen data media sosial.

2. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini melibatkan beberapa tahapan sistematis untuk memperoleh hasil analisis yang akurat.



Gambar 1. Tahapan penelitian

2.1. Identifikasi Masalah

Menentukan isu atau permasalahan yang akan diteliti, yaitu bagaimana tanggapan atau opini masyarakat terhadap kebijakan pembatasan usia anak dalam penggunaan media sosial.

2.2. Pengumpulan Data

Mengumpulkan data berupa komentar publik dari media sosial atau portal berita online yang berkaitan dengan kebijakan tersebut. Data bisa diambil melalui metode web scraping atau dokumentasi digital.

2.3. Pra-pemrosesan Data (Preprocessing)

Melakukan pembersihan dan persiapan data sebelum digunakan dalam proses klasifikasi. Langkah-langkahnya antara lain:

1. *Case folding* (mengubah huruf menjadi huruf kecil semua)
2. *Translation* (Mengubah Bahasa asing menjadi Bahasa Indonesia)
3. *Cleansing* (menghapus karakter khusus, angka, simbol)
4. Tokenisasi (memisahkan kalimat menjadi kata-kata)
5. *Stopword removal* (menghapus kata-kata umum yang tidak memiliki makna penting)

6. *Stemming* (mengubah kata ke bentuk dasar)

2.4. Labelling Data

Menentukan label pada setiap data komentar berdasarkan makna atau opini yang terkandung, yaitu:

1. Positif
2. Negatif
3. Netral

2.5. Penerapan Algoritma

Membangun model klasifikasi menggunakan dua algoritma:

2.5.1. Naïve Bayes

Naïve bayes merupakan suatu metode untuk mengklasifikasikan data dengan menerapkan probabilitas sederhana dengan menggunakan teorema *bayes* dan mengasumsikan angka independen yang tinggi metode ini diterapkan untuk mengklasifikasikan data seperti analisis sentimen, pengenalan teks dan *spam filtering* [8].

Naïve Bayes merupakan salah satu metode klasifikasi yang bersifat sederhana namun cukup efektif dan banyak digunakan, karena kemudahannya dalam penerapan serta kinerjanya yang cukup baik pada berbagai jenis permasalahan [9]. Algoritma ini sangat sesuai untuk diterapkan pada *dataset* berukuran besar, memiliki kemampuan menangani data yang tidak lengkap, serta menggunakan pendekatan probabilistik dan statistik untuk memperkirakan kemungkinan suatu kejadian di masa depan berdasarkan data historis [10]. *Teorema bayes* dinyatakan dalam rumus :

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

Dimana:

$P(H|E)$ = probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis H terjadi jika diberikan bukti (evidence) E terjadi. (probabilitas *posterior*)

$P(E|H)$ = probabilitas sebuah bukti E akan mempengaruhi hipotesis H (*likelihood*)

$P(H)$ = probabilitas awal hipotesis H terjadi tanpa memandang bukti apapun (*prior*).

$P(E)$ = probabilitas awal bukti E terjadi tanpa memandang hipotesis / bukti yang lain (*evidence*).

Algoritma ini mempelajari pola distribusi kata pada setiap kategori berdasarkan data pelatihan (training data) selalu mengklasifikasikan data baru berdasarkan perhitungan probabilitas tersebut.

Untuk mengukur kinerja yang dilakukan oleh algoritma *naïve bayes*, digunakan rumus akurasi sebagai berikut :

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2)$$

Keterangan:

TP (*True Positive*): Data positif yang diprediksi benar

TN (*True Negative*): Data negatif yang diprediksi benar

FP (*False Positive*): Data negatif yang diprediksi salah sebagai positif

FN (*False Negative*): Data positif yang di prediksi salah sebagai negatif

Presisi (*precision*) : untuk mengukur berapa banyak dari prediksi positif yang benar-benar positif.

$$\text{Presisi: } \frac{TP}{TP+FP} \quad (3)$$

Recall (Sensitivitas) : Mengukur seberapa banyak data positif yang berhasil dikenali dengan benar.

$$\text{Recall: } \frac{TP}{TP+FN} \quad (4)$$

Di mana:

TP = *True Positive*

TN = *True Negative*

FP = *False Positive*

FN = *False Negative*

2.5.2. Decision Tree

Decision tree adalah metode prediksi yang memanfaatkan representasi pohon keputusan yang digunakan untuk membuat keputusan dengan mempertimbangkan faktor-faktor tertentu [11]. Algoritma *decision tree* merupakan metode klasifikasi yang menggunakan struktur pohon atau hierarki untuk memfasilitasi proses pengambilan keputusan. Tujuan utama dari pendekatan ini adalah menghasilkan prediksi yang tepat dan konsisten terhadap data, serta mampu mengidentifikasi dan memahami keterkaitan yang kompleks antar variabel dalam dataset [12].

Proses pembentukan *decision tree* dimulai dengan memilih atribut terbaik sebagai akar pohon berdasarkan ukuran tertentu, seperti information gain, gain ratio atau gain index [13]. Dalam konteks analisis sentimen, *decision tree* dapat digunakan untuk mengklasifikasikan komentar atau opini berdasarkan kata-kata atau frasa tertentu yang menjadi atribut dalam pohon keputusan. Sebagai contoh, jika kata "bagus", "baik", atau "senang" sering muncul dalam komentar, maka pohon keputusan akan belajar bahwa komentar tersebut kemungkinan besar termasuk ke dalam kategori positif.

Menurut Y.R. Putri (2016) pada jurnal [14] perhitungan algoritma *decision tree* secara umum memiliki rumus sebagai berikut :

Untuk mencari nilai entropy :

$$\text{Entropy}(S) = \sum_{i=1}^n -p_i * \log_2(p_i) \quad (5)$$

Keterangan:

S: himpunan data

n: jumlah kelas

p_i : proporsi kelas ke-i dalam himpunan data

Untuk mencari nilai gain:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i)$$

(6)

Keterangan:

S: Himpunan data

A: atribut

n: jumlah kelas

|S_i|: jumlah data pada partisi ke i

|S|: jumlah data dalam S

Untuk menghitung nilai akurasi dari model yang dihasilkan, dapat digunakan rumus sebagai berikut:

$$\text{akurasi} = \frac{\text{nilai kecocokan}}{\text{jumlah seluruh kejadian}} * 100\% \quad (7)$$

Presisi (*precision*) : untuk mengukur berapa banyak dari prediksi positif yang benar-benar positif.

$$\text{Presisi: } \frac{TP}{TP+FP} \quad (8)$$

Di mana:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

2.2. Evaluasi Model

Mengukur performa dari masing-masing model dengan metrik evaluasi menggunakan akurasi, *precision* dan *recall*. Tujuannya untuk mengetahui algoritma mana yang memberikan hasil terbaik dalam klasifikasi sentimen.

2.3. Penerapan Algoritma

Menafsirkan hasil evaluasi dan menyimpulkan hasil dari akurasi, *precision* dan *recall* dari dua algoritma. Tujuannya adalah untuk menentukan algoritma mana yang memberikan hasil lebih baik untuk *dataset* dan masalah yang diteliti.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Dari penelitian ini, data diperoleh dari postingan dan video TikTok yang membahas kebijakan pemerintah mengenai pembatasan usia anak dalam penggunaan media sosial. Pengambilan komentar dilakukan menggunakan teknik web *scraping* yang dapat mengekstrak komentar dari postingan atau video TikTok secara otomatis dalam format .xlsx.

Pengumpulan data dilakukan pada bulan Mei–Juli 2025, sehingga data yang terkumpul menggambarkan opini publik terkini terhadap kebijakan pembatasan usia anak dalam penggunaan media sosial. Jumlah data yang berhasil diperoleh adalah 1.096 komentar, yang kemudian digunakan sebagai data mentah sebelum tahap *pre-processing*.

3.2. Pre-processing Data

Adapun tahapan *pre-processing* yang dilakukan adalah:

3.2.1. Translation

Tabel 1. Translation

No	Sebelum Translation	Hasil Translation
1	indonesia strict parents	Orang tua Indonesia yang tegas
2	noooo streakkk gwww	tidak ada goresan aku
3	YouTube be like:welcome back why friend	YouTube seperti: selamat datang kembali mengapa teman

Komentar yang mengandung kata-kata asing diterjemahkan ke dalam Bahasa Indonesia karena proses analisis sentimen pada penelitian ini dilakukan dalam lingkup bahasa Indonesia. Hal ini bertujuan untuk menyeragamkan bahasa pada seluruh data komentar sehingga menghindari perbedaan interpretasi makna kata akibat penggunaan bahasa yang berbeda, serta meningkatkan akurasi model dalam mengenali konteks dan sentimen.

3.2.2. Case Folding

Tahap ini bertujuan untuk menyeragamkan format teks dengan cara mengubah seluruh huruf dalam komentar menjadi huruf kecil (*lowercase*). Dengan demikian, kata yang sama tidak dianggap berbeda hanya karena perbedaan huruf kapital, misalnya kata “SEMUA” dan “semua” akan dianggap sama

Tabel 2. Case Folding

No	Sebelum Case Folding	Hasil Case Folding
1	YAEAAAA, SEMUA AJA GABOLE	yaelaaa semua aja gabole
2	PLS 13 TAUN KEBAWAH AJA..	pls taun dibawah aja
3	NOOOOO,AKU BELAJAR IPA,IPS,MTK, SAMA IT	tidakaku belajar ipaipsmtk sama itk lewat ig sama tiktokk

3.2.3. Cleansing

Pada tahap cleansing, tanda baca, simbol, dan emoji dihapus dari teks komentar. Hal ini dilakukan agar komentar hanya berisi kata-kata yang relevan untuk dianalisis, sehingga tidak ada gangguan dari karakter non-alfabet yang tidak memiliki makna dalam analisis sentimen.

Tabel 3. Cleansing

No	Sebelum Cleansing	Hasil Cleansing
1	 vpn	vpn
2	Manatappppp	mantap
3	Mensiasati para pedo..	Mensiasati para pedo
4	   setuju	setuju

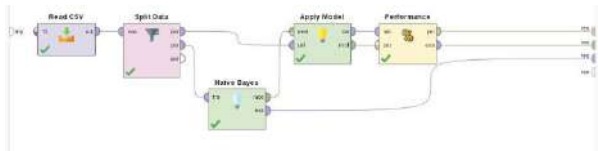
3.2.4. Stopword removal

Fase ini menghapus kata-kata umum (*stopwords*) yang tidak memberikan makna signifikan terhadap penentuan sentimen. Contohnya seperti kata “yang”, “dan”, “atau”. Dengan membuang kata-kata tersebut, teks menjadi lebih ringkas dan hanya menyisakan kata-kata yang penting untuk analisis.

Tabel 4. Stopword removal

No	Sebelum Stopword Removal	Hasil Stopword Removal
1	orang tua indonesia yang tegas	orang tua indonesia tegas
2	tidak ada goresan aku	gores
3	youtube seperti selamat datang kembali meng	youtube selamat datang kembali teman

$$P(\text{Net} \mid \text{Orang}) = \frac{0,0375 * 0,508}{0,0417} = 0,457$$



Gambar 3. Rapidminer Naive Bayes

3.5. Evaluasi Model

Tabel 7. Hasil evaluasi model

	Decision Tree	Naive Bayes
akurasi	91,90%	97,14%
Recall	79,76%	96,92%
Precision	96,46%	97,08%

Berdasarkan tabel 7. hasil pengujian, terlihat bahwa algoritma *Naive Bayes* memiliki performa yang lebih unggul dibandingkan dengan *Decision Tree*. *Naive Bayes* memperoleh nilai akurasi sebesar 97,14%, lebih tinggi dibandingkan *Decision Tree* yang hanya mencapai 91,90%. Dari sisi *recall*, *Naive Bayes* juga menunjukkan hasil yang sangat baik dengan capaian 96,92%, sedangkan *Decision Tree* hanya 79,76%, yang berarti model *Decision Tree* masih cukup banyak melewatkan data positif yang seharusnya terdeteksi. Sementara itu, untuk presisi, *Decision Tree* memiliki nilai 96,46% dan *Naive Bayes* sedikit lebih tinggi yaitu 97,08%.

3.6. Perbandingan Metode

Hasil pengujian menunjukkan bahwa algoritma *Naive Bayes* memiliki performa yang lebih baik dibandingkan *Decision Tree* pada analisis sentimen terhadap kebijakan pembatasan usia penggunaan media sosial. *Naive Bayes* mampu mencapai akurasi sebesar 97,14%, *recall* 96,92%, dan presisi 97,08%, yang menandakan bahwa algoritma ini tidak hanya mampu memprediksi dengan benar secara keseluruhan, tetapi juga seimbang dalam mendeteksi data positif serta menjaga ketepatan hasil prediksi. Sementara itu, *Decision Tree* memperoleh akurasi sebesar 91,90%, presisi 96,46%, dan *recall* 79,76%. Nilai presisi yang tinggi pada *Decision Tree* menunjukkan bahwa ketika model ini memprediksi suatu data sebagai positif, sebagian besar prediksi tersebut benar. Namun, rendahnya *recall* mengindikasikan bahwa masih banyak data positif yang tidak berhasil terklasifikasi dengan baik, sehingga model ini cenderung melewatkan informasi penting.

Keunggulan *Naive Bayes* dalam analisis sentimen tidak terlepas dari sifat dasarnya sebagai algoritma berbasis probabilistik yang menghitung peluang kata-kata dalam dokumen untuk menentukan kelas tertentu. Pada data komentar media sosial yang umumnya terdiri dari teks singkat, informal, dan penuh variasi kata, metode ini sangat efektif karena mampu menangani representasi berbasis kata (*bag-of-words*) dengan baik. Selain itu, *Naive Bayes* juga terkenal sederhana namun sangat efisien, sehingga dapat bekerja optimal meskipun data berukuran besar dan memiliki distribusi kata yang

tidak seimbang. Fakta ini memperkuat bahwa *Naive Bayes* sangat cocok digunakan untuk analisis sentimen, terutama dalam konteks penelitian terhadap kebijakan pembatasan usia penggunaan media sosial, di mana keberagaman bahasa dan gaya penulisan pengguna media sosial menjadi tantangan utama.

Dengan perbandingan ini, dapat disimpulkan bahwa *Naive Bayes* lebih efektif dan konsisten dalam menangani data teks yang bersifat variatif dan kompleks, sebagaimana umumnya komentar di media sosial. Keunggulan *Naive Bayes* juga terletak pada kemampuannya dalam bekerja dengan data berukuran besar serta tetap menjaga keseimbangan antara presisi dan *recall*. Oleh karena itu, pada penelitian ini, *Naive Bayes* dipandang sebagai algoritma yang lebih sesuai digunakan dibandingkan *Decision Tree* untuk analisis sentimen terhadap kebijakan pembatasan usia pada media sosial.

4. KESIMPULAN DAN SARAN

4.1. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan terhadap analisis sentimen kebijakan pembatasan usia anak dalam penggunaan media sosial dengan menggunakan algoritma *Decision Tree* dan *Naive Bayes*, diperoleh beberapa temuan penting.

1. Penerapan algoritma *Naive Bayes*

Algoritma *Naive Bayes* berhasil diterapkan dalam analisis sentimen terhadap komentar masyarakat mengenai kebijakan pembatasan usia penggunaan media sosial anak. Dengan memanfaatkan teknik preprocessing (*translation*, *case folding*, *tokenisasi*, *stopword removal*, *stemming*, dan *Bag of Words*) di *RapidMiner*, *Naive Bayes* dapat mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral dengan hasil evaluasi yang cukup baik.

2. Penerapan Algoritma *Decision Tree*

Algoritma *Decision Tree* juga dapat digunakan untuk menganalisis sentimen komentar masyarakat terhadap kebijakan tersebut. Proses klasifikasi dilakukan dengan membangun pohon keputusan dari data yang sudah melalui tahap preprocessing. Namun, kinerja *Decision Tree* cenderung bias terhadap kelas tertentu, terutama kelas netral, sehingga menghasilkan akurasi yang lebih rendah dibandingkan *Naive Bayes*.

3. Perbandingan Kinerja Algoritma

Hasil pengujian menunjukkan bahwa *Naive Bayes* memiliki performa yang jauh lebih baik dibandingkan dengan *Decision Tree* pada seluruh metrik evaluasi. *Naive Bayes* memperoleh akurasi sebesar 97,14%, presisi 96,92%, dan *recall* 97,08%, sementara *Decision Tree* hanya mencapai akurasi 91,90%, presisi 79,76%, dan *recall* 96,46%.

4.2. Saran

Berdasarkan hasil penelitian ini, terdapat beberapa saran yang dapat dipertimbangkan untuk pengembangan penelitian selanjutnya:

1. Mengatasi ketidakseimbangan data kelas.

Distribusi data komentar dalam penelitian ini tidak seimbang antar kelas sentimen. Oleh karena itu, perlu dilakukan pendekatan seperti *oversampling* (misalnya dengan *SMOTE*), *undersampling*, atau pemberian bobot kelas agar algoritma tidak bias terhadap kelas mayoritas.

2. Menggunakan matrik evaluasi yang lebih beragam. Akurasi saja tidak cukup untuk menilai kualitas model, terutama pada data yang tidak seimbang. Penelitian lanjutan disarankan untuk menambahkan metrik lain seperti *F1-score*, *Macro-average*, *Weighted-average*, dan *Balanced Accuracy* agar hasil evaluasi lebih komprehensif.

DAFTAR PUSTAKA

- [1] M. B. Yel and M. K. M. Nasution, "Keamanan Informasi Data Pribadi Pada Media Sosial," *J. Inform. Kaputama*, vol. 6, no. 1, pp. 92–101, 2022, doi: 10.59697/jik.v6i1.144.
- [2] M. I. Antula, "Pengaruh Media Sosial Terhadap Gaya Hidup dan Pergaulan Remaja di Era Digital," *J. Pendidik. Mosikolah*, vol. 4, no. 2, pp. 53–60, 2024, [Online]. Available: <http://jurnal.ut.ac.id/index.php/jp/search/authors/view?givenName=MeryNoviyanti&familyName=&affiliation=UniversitasTerbuka&country=ID&authorName=MeryNoviyanti>
- [3] Y. Fitriani, "Pemanfaatan Media Sosial Sebagai Media Penyajian Konten Edukasi Atau Pembelajaran Digital," *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 5, no. 4, pp. 1006–1013, 2021, doi: 10.52362/jisamar.v5i4.609.
- [4] M. Nur Cahya, W. Ningsih, and A. Lestari, "Dampak Media Sosial terhadap Kesejahteraan Psikologis Remaja: Tinjauan Pengaruh Penggunaan Media Sosial pada Kecemasan dan Depresi Remaja," *J. Sos. Teknol.*, vol. 3, no. 8, pp. 703–706, 2023, doi: 10.59188/jurnalsostech.v3i8.917.
- [5] F. C. P. L. Shelma Zahira Ardhania, "PENGARUH MEDIA SOSIAL TERHADAP KESEHATAN MENTAL REMAJA," *Triwikrama J. Ilmu Sos.*, vol. 01, no. 11, pp. 40–50, 2023.
- [6] I. Sulistiani, V. W. Sutjipto, M. P. Sary, M. L. Putri, E. N. Prananingrum, and U. N. Jakarta, "PENGARUH MEDIA SOSIAL TERHADAP SIKAP REMAJA PADA ERA DIGITALISASI," *SOSFILKOM*, vol. XVIII, no. 2, pp. 138–146, 2024.
- [7] N. Aulia, S. Hadi, and Nurdyana, "Pengaruh Media Sosial Terhadap Perilaku Sosial Media," *J. Educ. Cult.*, vol. 2, no. 1, pp. 64–70, 2022.
- [8] I. Gustina and A. Yudhistira, "Analisis Sentimen Program Coding Anak SD Menggunakan Metode Naive Bayes," *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 2, pp. 505–514, 2025.
- [9] S. A. Mahmudah, A. Yudhistira, F. Teknik, and U. T. Indonesia, "Analisis Sentimen terhadap Cyberbullying pada Platform Media Sosial X Menggunakan Algoritma Naive Bayes Sentiment Analysis of Cyberbullying on Social Media Platform X Using Naive Bayes Algorithm," *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 1, pp. 189–200, 2025.
- [10] E. M. A. Ernamia and A. Herliana, "Analisis Sentimen Kuliah Daring Dengan Algoritma Naive Bayes, K-NN Dan Decision Tree," *J. Responsif Ris. Sains dan Inform.*, vol. 4, no. 1, pp. 70–80, 2022, doi: 10.51977/jti.v4i1.614.
- [11] H. B. Yolanda Veren, Sunneng Sandino Berutu, "Penerapan metode decision tree pada sentimen media sosial terkait komisi pemilihan umum (kpu) sebelum dan sesudah pilpres 2024," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 4, pp. 2174–2184, 2024, doi: <https://doi.org/10.29100/jipi.v9i4.55321>.
- [12] E. P. Adamansyah, A. Yudhistira, S. Informasi, F. Teknik, and U. T. Indonesia, "Evaluasi Opini Publik di Media Sosial X terhadap Kebijakan Pajak Pertambahan Nilai 12 % di Indonesia Menggunakan Naive Bayes dan Decision Tree Informatika , Fakultas Teknik dan Ilmu Komputer , Universitas Teknokrat Indonesia , Indonesia Evaluation of Publ," *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 3, pp. 831–843, 2025.
- [13] Y. A. Singgalen, "Analisis Sentimen Top 10 Traveler Ranked Hotel di Kota Makassar Menggunakan Algoritma Decision Tree dan Support Vector Machine," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 323–332, 2023, doi: 10.30865/klik.v4i1.1153.
- [14] A. Apriyadi, M. R. Lubis, and B. E. Damanik, "Penerapan Algoritma C5.0 Dalam Menentukan Tingkat Pemahaman Mahasiswa Terhadap Pembelajaran Daring," *Komputa J. Ilm. Komput. dan Inform.*, vol. 11, no. 1, pp. 11–20, 2022, doi: 10.34010/komputa.v11i1.7386.