



Analisis Tingkat Sentimen Opini Publik Terhadap Kebijakan TV Digital di Platform X Menggunakan Multinomial Naïve Bayes

Asep Arwan Sulaeman*, Candra Naya, Muhtajuddin Danny, M. Makmun Effendi

Fakultas Teknik, Program Studi Teknik Informatika, Universitas Pelita Bangsa, Bekasi, Indonesia

Email: ^{1,*}asep.arwan@pelitabangsa.ac.id ²candranaya@pelitabangsa.ac.id, ³utat@pelitabangsa.ac.id, ⁴effendiyan@pelitabangsa.ac.id

Email Penulis Korespondensi: asep.arwan@pelitabangsa.ac.id

Abstrak—Migrasi siaran televisi analog ke televisi digital merupakan bagian dari transformasi sistem penyiaran yang bertujuan meningkatkan kualitas siaran dan efisiensi pemanfaatan spektrum frekuensi. Namun, implementasi kebijakan TV digital menimbulkan beragam respons dari masyarakat, baik berupa dukungan maupun kritik. Penelitian ini bertujuan untuk menganalisis opini publik terhadap kebijakan TV digital di Indonesia berdasarkan data media sosial pada platform X. Metode yang digunakan adalah pendekatan kuantitatif dengan teknik *text mining* dan *supervised machine learning*. Data dikumpulkan melalui proses *crawling* menggunakan kata kunci “tv digital”, yang menghasilkan 1.855 tweet. Setelah melalui tahap seleksi dan *cleaning* data, diperoleh 789 tweet sebagai dataset akhir. Tahapan analisis meliputi pra-pemrosesan teks, ekstraksi fitur menggunakan Term Frequency–Inverse Document Frequency (TF–IDF), serta klasifikasi sentimen menggunakan algoritma Multinomial Naïve Bayes. Hasil penelitian menunjukkan bahwa sentimen positif mendominasi opini publik dengan 478 tweet (60,58%), sedangkan sentimen negatif berjumlah 311 tweet (39,42%). Evaluasi kinerja model menunjukkan nilai *accuracy* sebesar 79,21%, *precision* 82,45%, dan *recall* 85,06%, yang mengindikasikan bahwa model memiliki kinerja klasifikasi yang baik dan stabil. Temuan ini menunjukkan bahwa analisis sentimen berbasis media sosial dapat dimanfaatkan sebagai pendekatan empiris untuk memahami persepsi publik terhadap kebijakan TV digital.

Kata Kunci: Analisis Sentimen; TV Digital; Media Sosial; Naïve Bayes; Text Mining

Abstract—The migration from analog to digital television broadcasting is part of the transformation of the broadcasting system aimed at improving broadcast quality and spectrum efficiency. However, the implementation of the digital television policy has generated diverse public responses, ranging from support to criticism. This study aims to analyze public opinion on the digital television policy in Indonesia using social media data from platform X. A quantitative approach was employed using *text mining* and *supervised machine learning* techniques. Data were collected through a *crawling* process using the keyword “tv digital”, resulting in 1,855 tweets. After data selection and *cleaning*, 789 tweets were obtained as the final dataset. The analysis stages included text preprocessing, feature extraction using Term Frequency–Inverse Document Frequency (TF–IDF), and sentiment classification using the Multinomial Naïve Bayes algorithm. The results indicate that positive sentiment dominates public opinion, with 478 tweets (60.58%), while negative sentiment accounts for 311 tweets (39.42%). Model performance evaluation shows an *accuracy* of 79.21%, *precision* of 82.45%, and *recall* of 85.06%, indicating that the model performs well and consistently in classifying sentiment. These findings demonstrate that social media–based sentiment analysis can serve as an empirical approach to understanding public perceptions of digital television policy.

Keywords: Sentiment Analysis; Digital Television; Social Media; Naïve Bayes; Text Mining

1. PENDAHULUAN

Migrasi siaran televisi analog ke televisi digital merupakan bagian dari transformasi sistem penyiaran yang bertujuan untuk meningkatkan kualitas siaran [1], efisiensi pemanfaatan spektrum frekuensi, serta pemerataan akses informasi publik [1]. Peralihan ini memungkinkan penggunaan teknologi kompresi yang lebih efisien sehingga kualitas audio dan visual menjadi lebih baik, sekaligus membuka peluang pemanfaatan spektrum untuk layanan komunikasi lainnya [2], [3]. Kebijakan penghentian siaran analog (*analog switch-off*) telah diterapkan di berbagai negara sebagai bagian dari agenda global digitalisasi penyiaran, termasuk di Indonesia [4]. Implementasi kebijakan ini tidak hanya bersifat teknis, tetapi juga membawa konsekuensi sosial yang luas karena berkaitan dengan kesiapan infrastruktur, kemampuan ekonomi masyarakat, serta tingkat literasi teknologi [5].

Respons masyarakat terhadap kebijakan migrasi televisi digital menunjukkan karakter yang heterogen [6]. Di satu sisi, sebagian masyarakat menyambut positif kebijakan ini karena peningkatan kualitas gambar dan suara serta bertambahnya jumlah saluran yang dapat diakses [7]. Di sisi lain, tidak sedikit masyarakat yang menyampaikan keluhan terkait kebutuhan perangkat tambahan seperti *set top box*, keterbatasan jangkauan sinyal di wilayah tertentu, serta kurangnya sosialisasi kebijakan secara merata [8], [9], [10]. Kondisi ini menunjukkan bahwa keberhasilan kebijakan TV digital tidak hanya ditentukan oleh kesiapan teknologi, tetapi juga oleh penerimaan dan persepsi publik terhadap kebijakan tersebut.

Perkembangan media sosial telah mengubah cara masyarakat mengekspresikan opini dan respons terhadap kebijakan publik [11]. Platform X (Twitter) menjadi salah satu ruang diskusi utama karena bersifat terbuka [12], cepat, dan memungkinkan partisipasi publik secara luas [13]. Melalui unggahan singkat berbasis teks, masyarakat dapat menyampaikan dukungan, kritik, maupun keluhan terhadap kebijakan TV digital secara real time. Opini-opini ini membentuk dinamika wacana publik yang dapat diamati dan dianalisis secara sistematis untuk memahami persepsi masyarakat secara empiris [10]. Oleh karena itu, media sosial memiliki potensi besar sebagai sumber data alternatif dalam evaluasi kebijakan publik berbasis bukti.

Berbagai penelitian dalam lima tahun terakhir menunjukkan bahwa analisis opini publik berbasis media sosial efektif dalam memetakan persepsi masyarakat terhadap isu sosial dan kebijakan. Studi analisis sentimen menggunakan



data Twitter telah diterapkan pada beragam konteks, seperti layanan publik, kebijakan kesehatan, dan adopsi teknologi, dengan temuan yang menunjukkan adanya korelasi antara sentimen daring dan dinamika sosial di dunia nyata [14], [15]. Hasil-hasil tersebut mengindikasikan bahwa media sosial tidak hanya berfungsi sebagai sarana komunikasi, tetapi juga sebagai refleksi sikap dan pengalaman masyarakat terhadap kebijakan yang sedang berjalan.

Pendekatan *text mining* dan *machine learning* [16], [17] banyak digunakan dalam analisis sentimen opini publik, khususnya pada data teks pendek seperti tweet. Algoritma klasifikasi berbasis probabilistik maupun statistik telah digunakan untuk mengidentifikasi polaritas sentimen secara otomatis. Beberapa penelitian membuktikan bahwa algoritma yang relatif sederhana tetap mampu memberikan kinerja yang kompetitif apabila dikombinasikan dengan proses pra-pemrosesan dan representasi fitur yang tepat [18]. Meskipun demikian, sebagian besar penelitian terdahulu masih berfokus pada isu-isu umum dan belum secara spesifik mengkaji respons publik terhadap kebijakan migrasi televisi digital dalam konteks Indonesia.

Data percakapan di media sosial pada dasarnya merupakan data tidak terstruktur yang belum memiliki makna analitis sebelum diproses. Tanpa analisis yang sistematis, kumpulan teks tersebut hanya menjadi arsip digital yang tidak memberikan informasi substantif bagi pengambilan keputusan. Oleh karena itu, proses *data mining* diperlukan untuk mengekstraksi pola [19], kecenderungan, dan pengetahuan tersembunyi dari data dalam jumlah besar. Dalam konteks opini publik, analisis data memungkinkan transformasi teks mentah menjadi indikator terukur seperti polaritas sentimen, intensitas diskusi, dan topik dominan, yang dapat digunakan untuk mengevaluasi persepsi masyarakat secara objektif dan berbasis data [20], [21].

Data mining dan *text mining* memiliki peran krusial dalam mengolah data media sosial yang bersifat besar [22], bising, dan dinamis [23], [24]. Metode ini memungkinkan peneliti untuk menyaring informasi relevan dari data yang heterogen dan mengidentifikasi pola opini publik secara efisien. Salah satu algoritma yang banyak digunakan dalam klasifikasi teks adalah Naïve Bayes, yang mengandalkan prinsip probabilistik dengan asumsi independensi antar fitur [25]. Kesederhanaan dan efisiensi algoritma ini menjadikannya populer dalam analisis teks berskala besar.

Penelitian terkini menunjukkan bahwa Naïve Bayes, khususnya varian *Multinomial Naïve Bayes*, tetap efektif dalam klasifikasi sentimen teks pendek karena efisiensi komputasi dan kemampuannya menangani dimensi fitur yang tinggi [26], [27]. Ketika dikombinasikan dengan ekstraksi fitur berbasis TF-IDF dan pra-pemrosesan bahasa yang sesuai dengan karakteristik teks media sosial [28], algoritma ini mampu menghasilkan performa yang stabil dan mudah direplikasi [29]. Oleh karena itu, pendekatan ini relevan untuk analisis opini publik di platform X yang didominasi oleh teks singkat dan bahasa informal.

Meskipun analisis sentimen berbasis media sosial telah banyak dilakukan, masih terdapat celah penelitian yang signifikan [30]. Pertama, kajian yang secara khusus membahas opini publik terhadap kebijakan TV digital di Indonesia masih terbatas. Kedua, karakteristik bahasa pada platform X seperti penggunaan slang, singkatan, dan ekspresi kontekstual menuntut pendekatan pra-pemrosesan yang lebih adaptif agar hasil klasifikasi lebih akurat [31]. Ketiga, masih terbatas penelitian yang menyajikan kerangka analisis yang sederhana, transparan, dan mudah direplikasi untuk kepentingan evaluasi kebijakan publik berbasis data media sosial [32].

Berdasarkan celah tersebut, kebaruan penelitian ini terletak pada fokus kajian terhadap opini publik mengenai TV digital di Indonesia dengan memanfaatkan data percakapan dari platform X sebagai sumber informasi empiris untuk evaluasi kebijakan publik [10]. Penelitian ini menerapkan *pipeline* analisis yang sistematis dan replikatif, mencakup pengambilan data berbasis kata kunci, pra-pemrosesan teks, ekstraksi fitur TF-IDF, serta klasifikasi sentimen menggunakan algoritma Naïve Bayes [33], [34]. Pendekatan ini diharapkan mampu menjembatani analisis teknis machine learning dengan kebutuhan evaluasi kebijakan publik berbasis data aktual [35].

Penelitian ini bertujuan untuk (1) mengidentifikasi distribusi sentimen opini publik terkait TV digital pada platform X, (2) mengevaluasi kinerja algoritma Naïve Bayes dalam klasifikasi sentimen teks berbahasa Indonesia, serta (3) menyajikan gambaran empiris mengenai persepsi masyarakat terhadap kebijakan TV digital. Manfaat akademik penelitian ini adalah memperkaya kajian analisis opini publik berbasis media sosial dalam konteks kebijakan teknologi. Secara praktis, hasil penelitian diharapkan dapat menjadi masukan bagi pemangku kebijakan dalam meningkatkan strategi komunikasi publik dan mengidentifikasi isu dominan yang muncul dalam proses implementasi TV digital [36].

2. METODOLOGI PENELITIAN

2.1. Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *text mining* dan *supervised machine learning*. Pendekatan ini dipilih karena data penelitian berupa teks opini masyarakat yang diolah secara numerik untuk tujuan klasifikasi sentimen. Metode *supervised learning* digunakan karena proses klasifikasi sentimen dilakukan berdasarkan data latih yang telah diberi label sentimen positif dan negatif.

2.2. Objek dan Sumber Data

Objek penelitian adalah opini masyarakat mengenai kebijakan televisi digital pada platform X (Twitter). Data dikumpulkan melalui proses *crawling* menggunakan kata kunci "*tv digital*" pada periode 23 Agustus 2023–22 Mei 2025. Hasil *crawling* awal memperoleh 1.855 tweet, kemudian dilakukan seleksi untuk menghapus data duplikat dan tidak



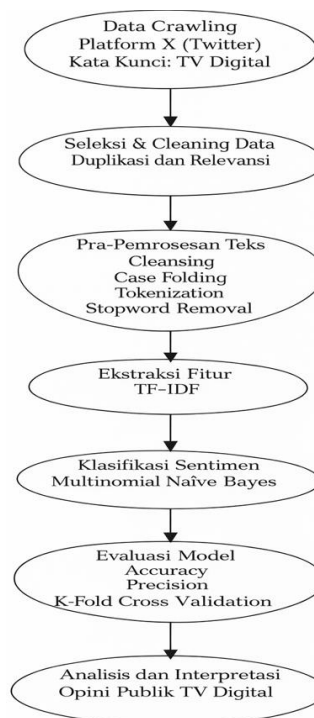
relevan sehingga diperoleh 789 tweet sebagai dataset akhir, yang terdiri dari 478 tweet bersentimen positif dan 311 tweet bersentimen negatif.

Tabel 1. Tabel Data Penelitian

Keterangan	Jumlah
Data hasil crawling awal	1.855
Data duplikat dan tidak relevan	1.066
Data akhir setelah seleksi	789
Sentimen positif	478
Sentimen negatif	311

Hasil crawling awal memperoleh 1.855 tweet. Setelah dilakukan proses seleksi untuk menghapus data duplikat dan tidak relevan, diperoleh 789 tweet sebagai dataset akhir yang digunakan dalam penelitian. Dataset tersebut terdiri dari 478 tweet bersentimen positif dan 311 tweet bersentimen negatif, sebagaimana ditunjukkan pada Tabel 1.

2.3. Alur dan Tahapan Penelitian



Gambar 1. Tahapan penelitian

a. Data crawling

Tahap data crawling merupakan tahap awal dalam penelitian ini yang bertujuan untuk mengumpulkan data opini publik mengenai kebijakan televisi digital dari platform X (Twitter). Data dikumpulkan dengan memanfaatkan kata kunci “*tv digital*” untuk menangkap percakapan, tanggapan, dan opini masyarakat yang berkaitan dengan implementasi kebijakan migrasi siaran televisi analog ke digital di Indonesia. Proses crawling dilakukan pada periode 11 November 2023 hingga 13 Agustus 2025. Pemilihan rentang waktu tersebut dimaksudkan untuk memperoleh data yang representatif, mencakup dinamika opini publik dalam kurun waktu implementasi dan sosialisasi kebijakan televisi digital. Data yang diperoleh pada tahap ini masih berupa teks mentah (raw text) dan belum melalui proses penyaringan maupun pengolahan lebih lanjut.

Tabel 2. Jumlah Data Hasil Crawling Awal

Keterangan	Jumlah
Data hasil crawling awal	1.855

Berdasarkan Tabel 2, proses data crawling menghasilkan 1.855 tweet yang mengandung kata kunci “*tv digital*”. Seluruh data tersebut selanjutnya menjadi dataset awal yang akan melalui tahap seleksi dan cleaning data untuk menghilangkan tweet yang tidak relevan, duplikat, maupun tidak sesuai

b. Seleksi dan cleaning data

Tahap seleksi dan cleaning data dilakukan untuk meningkatkan kualitas dataset hasil crawling sebelum digunakan pada tahap analisis selanjutnya. Data hasil crawling awal masih mengandung tweet yang tidak relevan dengan topik penelitian serta data duplikat yang berpotensi menimbulkan bias dalam proses klasifikasi sentimen. Proses seleksi



dilakukan dengan menghapus tweet yang tidak relevan dengan kebijakan televisi digital serta tweet duplikat. Selanjutnya, cleaning data dilakukan untuk memastikan data berada dalam format yang konsisten dan siap diproses pada tahap pra-pemrosesan teks.

Tabel 3. Hasil Seleksi dan Cleaning Data

Keterangan	Jumlah
Data hasil crawling awal	1.855
Data tidak relevan / duplikat	1.066
Data akhir setelah seleksi	789

Berdasarkan Tabel 3, dari 1.855 tweet hasil crawling awal, sebanyak 1.066 tweet dikeluarkan karena tidak relevan atau duplikat, sehingga diperoleh 789 tweet sebagai dataset akhir yang digunakan pada tahap pra-pemrosesan dan analisis sentimen.

c. Pra-pemrosesan teks

Pra-pemrosesan teks dilakukan untuk menyiapkan data teks hasil seleksi agar dapat diolah secara komputasional dan meningkatkan kualitas input bagi model klasifikasi sentimen. Data teks media sosial umumnya masih mengandung elemen non-informatif dan variasi penulisan yang dapat memengaruhi hasil analisis. Tahapan pra-pemrosesan yang diterapkan dalam penelitian ini meliputi cleansing, case folding, tokenization, dan stopword removal.

d. Ekstraksi fitur TF-IDF

Tahap ekstraksi fitur dilakukan untuk mengubah data teks hasil pra-pemrosesan menjadi representasi numerik yang dapat diproses oleh algoritma klasifikasi.

Term Frequency (TF):

$$TF(t, d) = f_{t,d} \quad (1)$$

dengan $f_{t,d}$ menyatakan jumlah kemunculan term t dalam dokumen d .

Inverse Document Frequency (IDF):

$$IDF(t) = \log \left(\frac{N}{df_t} \right) \quad (2)$$

dengan:

N adalah jumlah total dokumen dalam korpus,

df_t adalah jumlah dokumen yang mengandung term t .

Bobot TF-IDF:

$$TF-IDF(t, d) = TF(t, d) \times IDF(t) \quad (3)$$

Bobot TF-IDF yang dihasilkan membentuk vektor fitur untuk setiap dokumen, yang selanjutnya digunakan sebagai input pada tahap klasifikasi sentimen menggunakan algoritma Naïve Bayes.

e. Klasifikasi sentimen menggunakan Naïve Bayes

Tahap klasifikasi sentimen dilakukan untuk mengelompokkan opini publik ke dalam kelas sentimen positif dan sentimen negatif berdasarkan fitur teks yang telah diekstraksi menggunakan TF-IDF. Dalam penelitian ini digunakan algoritma Multinomial Naïve Bayes, yang merupakan salah satu metode klasifikasi probabilistik yang umum diterapkan pada data teks. Naïve Bayes bekerja berdasarkan Teorema Bayes dengan asumsi bahwa setiap fitur bersifat independen satu sama lain (conditional independence). Meskipun asumsi ini bersifat sederhana, Naïve Bayes terbukti efektif dalam menangani data teks berdimensi tinggi seperti tweet. Secara matematis, probabilitas posterior suatu kelas terhadap dokumen d dirumuskan sebagai berikut:

$$P(c | d) = \frac{P(c) P(d|c)}{P(d)} \quad (4)$$

Karena $P(d)$ bersifat konstan untuk seluruh kelas, maka persamaan tersebut dapat disederhanakan menjadi:

$$P(c | d) \propto P(c) P(d | c) \quad (5)$$

Dengan asumsi independensi antar fitur (term), probabilitas $P(d | c)$ dapat dinyatakan sebagai:

$$P(d | c) = \prod_{i=1}^n P(t_i | c) \quad (5)$$

sehingga fungsi klasifikasi Naïve Bayes menjadi:

$$\hat{c} = \arg \max_{c \in C} (P(c) \prod_{i=1}^n P(t_i | c)) \quad (6)$$

\hat{C} merupakan kelas sentimen hasil prediksi yang diperoleh dari proses klasifikasi. Kelas ini dipilih dari himpunan C yang terdiri atas dua kategori sentimen, yaitu positif dan negatif. Dalam proses penentuan kelas tersebut, setiap dokumen direpresentasikan oleh sejumlah term, di mana t_i menunjukkan term ke- i yang terdapat dalam dokumen. Penentuan kelas sentimen dilakukan dengan mempertimbangkan probabilitas prior $P(c)$, yaitu peluang awal suatu



dokumen termasuk ke dalam kelas tertentu sebelum melihat isi dokumen. Selain itu, digunakan pula probabilitas kondisional $P(t_i|c)$, yang menunjukkan peluang kemunculan term t_i apabila diketahui dokumen tersebut berada pada kelas c . Kombinasi dari probabilitas prior dan probabilitas kemunculan term inilah yang menjadi dasar dalam menentukan kelas sentimen hasil prediksi.

f. Evaluasi kinerja model

Evaluasi model dilakukan untuk menilai kinerja algoritma Naïve Bayes dalam mengklasifikasikan sentimen opini publik ke dalam dua kategori, yaitu positif dan negatif. Proses evaluasi ini dilakukan dengan membandingkan hasil prediksi yang dihasilkan model dengan label sentimen aktual menggunakan confusion matrix serta berbagai metrik evaluasi klasifikasi. Confusion matrix digunakan untuk menggambarkan performa model melalui empat komponen utama, yaitu True Positive (TP) yang menunjukkan data sentimen positif yang berhasil diprediksi sebagai positif, True Negative (TN) yang merupakan data sentimen negatif yang diprediksi dengan benar sebagai negatif, False Positive (FP) yaitu data sentimen negatif yang keliru diprediksi sebagai positif, serta False Negative (FN) yang merupakan data sentimen positif tetapi diprediksi sebagai negatif. Melalui matriks ini, dapat diketahui tingkat ketepatan sekaligus kesalahan klasifikasi yang dihasilkan model. Selanjutnya, berdasarkan nilai pada confusion matrix tersebut, kinerja model diukur menggunakan metrik evaluasi berupa accuracy, precision, dan recall, yang masing-masing memberikan gambaran tentang tingkat ketepatan prediksi secara keseluruhan, ketepatan dalam mengidentifikasi kelas positif, serta kemampuan model dalam menangkap seluruh data yang benar-benar termasuk dalam kelas positif. Accuracy mengukur proporsi prediksi yang benar terhadap seluruh data:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

Precision mengukur tingkat ketepatan model dalam memprediksi kelas positif:

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

Recall mengukur kemampuan model dalam mengenali seluruh data kelas positif:

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

Metrik accuracy digunakan untuk menilai kinerja model secara keseluruhan, sedangkan precision dan recall digunakan untuk mengevaluasi kualitas prediksi pada kelas sentimen positif.

g. Analisis dan interpretasi hasil

Tahap analisis dan interpretasi hasil merupakan tahap akhir dalam alur metode penelitian. Pada tahap ini, hasil evaluasi kinerja model klasifikasi sentimen dianalisis untuk memahami pola, kecenderungan, dan makna empiris dari data opini publik yang telah diklasifikasikan. Analisis dilakukan dengan menginterpretasikan keluaran model berupa distribusi sentimen serta nilai metrik evaluasi yang dihasilkan. Interpretasi ini bertujuan untuk memberikan pemahaman mengenai sejauh mana model Naïve Bayes mampu mengklasifikasikan sentimen opini publik secara akurat dan konsisten berdasarkan data yang digunakan.

2.4. Karakteristik dan Ruang Lingkup Data Penelitian

Data yang digunakan dalam penelitian ini berupa opini publik mengenai kebijakan televisi digital yang dipublikasikan melalui platform X. Data dikumpulkan menggunakan teknik crawling berbasis kata kunci “tv digital” pada periode 23 Agustus 2023 hingga 22 Mei 2025. Platform X dipilih karena memiliki karakteristik komunikasi yang terbuka dan partisipatif, sehingga memungkinkan masyarakat menyampaikan opini secara langsung terhadap kebijakan publik.

Hasil crawling awal menghasilkan 1.855 tweet. Setelah dilakukan proses seleksi dan cleaning data untuk menghapus tweet duplikat serta konten yang tidak relevan dengan kebijakan TV digital, diperoleh 789 tweet sebagai dataset akhir. Dataset tersebut terdiri atas 478 tweet bersentimen positif dan 311 tweet bersentimen negatif. Secara karakteristik, data tweet bersifat tidak terstruktur (unstructured text), berdimensi tinggi, serta menggunakan bahasa informal yang sering mengandung singkatan, slang, dan ekspresi kontekstual. Kondisi ini menuntut penerapan tahapan pra-pemrosesan yang sistematis sebelum dilakukan analisis klasifikasi sentimen.

2.5. Algoritma Multinomial Naïve Bayes

Penelitian ini menggunakan algoritma Multinomial Naïve Bayes sebagai metode klasifikasi sentimen. Algoritma ini merupakan metode supervised learning berbasis probabilistik yang bekerja berdasarkan Teorema Bayes dengan asumsi independensi antar fitur (conditional independence). Multinomial Naïve Bayes secara khusus dirancang untuk menangani data teks yang direpresentasikan dalam bentuk frekuensi atau bobot kemunculan kata.

Dalam penelitian ini, teks tweet yang telah melalui tahap pra-pemrosesan direpresentasikan dalam bentuk vektor numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF–IDF). Representasi TF–IDF mencerminkan distribusi bobot kata dalam setiap dokumen relatif terhadap keseluruhan korpus. Vektor fitur tersebut kemudian digunakan sebagai input dalam proses klasifikasi sentimen. Secara matematis, probabilitas suatu dokumen termasuk ke dalam kelas sentimen tertentu dihitung berdasarkan probabilitas prior kelas dan probabilitas kemunculan setiap term dalam kelas tersebut. Dokumen akan diklasifikasikan ke dalam kelas dengan probabilitas posterior tertinggi. Pemilihan Multinomial Naïve Bayes didasarkan pada kemampuannya dalam menangani data berdimensi tinggi, efisiensi komputasi, serta performa yang stabil pada klasifikasi teks pendek seperti tweet.



3. HASIL DAN PEMBAHASAN

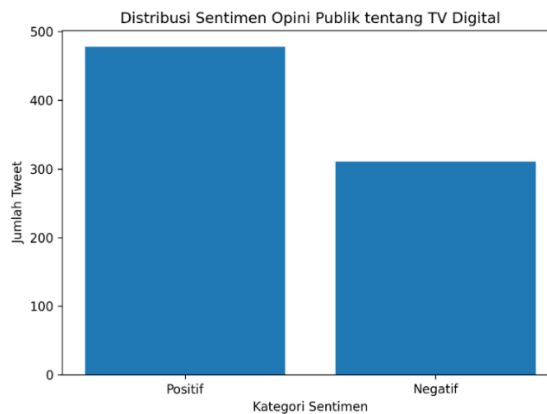
3.1. Distribusi Sentimen Opini Publik

Hasil klasifikasi sentimen terhadap 789 tweet yang telah melalui tahap pra-pemrosesan dan ekstraksi fitur menunjukkan bahwa opini publik terhadap kebijakan televisi digital didominasi oleh sentimen positif. Distribusi sentimen ini memberikan gambaran awal mengenai kecenderungan sikap masyarakat terhadap implementasi kebijakan TV digital di Indonesia.

Tabel 4. Distribusi Sentimen Opini Publik

Sentimen	Jumlah	Persentase
Positif	478	60,58%
Negatif	311	39,42%
Total	789	100%

Tabel 4 menunjukkan bahwa dari total 789 tweet yang dianalisis, sentimen positif mendominasi dengan 478 tweet (60,58%), sedangkan sentimen negatif berjumlah 311 tweet (39,42%). Temuan ini mengindikasikan bahwa opini publik terhadap TV digital cenderung lebih positif, meskipun proporsi sentimen negatif masih cukup signifikan.



Gambar 2. Grafik Distribusi Sentimen Opini Publik tentang TV Digital

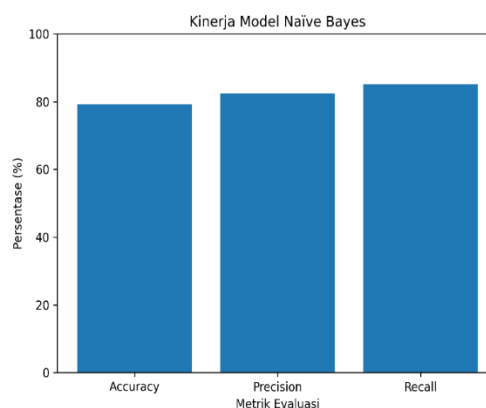
Gambar 2 menunjukkan bahwa sentimen positif mendominasi opini publik terhadap kebijakan TV digital dengan 478 tweet (60,58%), sementara sentimen negatif berjumlah 311 tweet (39,42%).

Perbandingan Metrik Evaluasi Model Naïve Bayes:

Tabel 5. Perbandingan Metrik Evaluasi

Metrik	Nilai
Accuracy	79,21%
Precision	82,45%
Recall	85,06%

Tabel 5 menunjukkan bahwa model Naïve Bayes memiliki kinerja klasifikasi yang baik, dengan accuracy 79,21%, precision 82,45%, dan recall 85,06%, yang mengindikasikan kemampuan model dalam mengklasifikasikan sentimen secara cukup akurat dan konsisten.



Gambar 3. Perbandingan Metrik Evaluasi Model Naïve Bayes



Gambar 3 menampilkan perbandingan metrik evaluasi model Naïve Bayes yang digunakan dalam penelitian ini. Nilai recall yang paling tinggi menunjukkan bahwa model memiliki kemampuan yang baik dalam mengidentifikasi sentimen positif. Sementara itu, nilai accuracy dan precision yang berada di atas 79% menunjukkan kestabilan performa model dalam mengklasifikasikan opini publik berbasis teks media sosial

3.2. Hasil Evaluasi Kinerja Model Naïve Bayes

Hasil evaluasi kinerja model menunjukkan bahwa algoritma Naïve Bayes mampu mengklasifikasikan sentimen opini publik dengan tingkat kinerja yang baik. Model memiliki kemampuan yang cukup andal dalam membedakan sentimen positif dan negatif berdasarkan representasi fitur TF-IDF. Nilai evaluasi yang diperoleh mengindikasikan bahwa model tidak hanya akurat secara keseluruhan, tetapi juga memiliki tingkat ketepatan dan sensitivitas yang memadai dalam mengenali sentimen positif. Hal ini menunjukkan bahwa pendekatan probabilistik yang digunakan oleh Naïve Bayes tetap efektif untuk klasifikasi teks pendek pada media sosial. Meskipun demikian, masih ditemukan sejumlah kesalahan klasifikasi yang umumnya disebabkan oleh karakteristik bahasa informal dan ambiguitas makna pada teks tweet. Kondisi ini menjadi keterbatasan yang wajar dalam analisis sentimen berbasis data media sosial.

3.3. Analisis Confusion Matrix

Confusion matrix digunakan untuk menganalisis secara rinci kinerja model Naïve Bayes dalam mengklasifikasikan sentimen opini publik. Matriks ini menggambarkan perbandingan antara hasil prediksi model dan label sentimen aktual, yang terdiri atas empat komponen utama, yaitu true positive, true negative, false positive, dan false negative.

Tabel 6. Confusion Matrix Klasifikasi Sentimen

Prediksi \ Aktual	Positif	Negatif
Positif	TP = 612	FP = 186
Negatif	FN = 205	TN = 359

Berdasarkan Tabel 6, model Naïve Bayes menghasilkan 612 data true positive dan 359 data true negative, yang menunjukkan kemampuan model dalam mengklasifikasikan sentimen positif dan negatif secara cukup baik. Namun, masih terdapat 186 data false positive dan 205 data false negative, yang mengindikasikan adanya kesalahan klasifikasi.

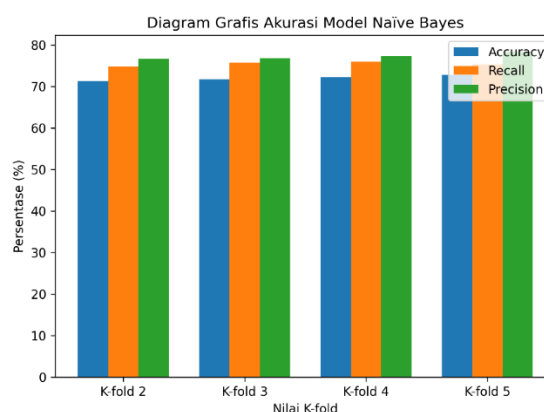
3.4. Pengaruh K-Fold Cross Validation terhadap Kinerja Model

Pengujian menggunakan K-Fold Cross Validation dilakukan untuk mengevaluasi stabilitas dan konsistensi kinerja model Naïve Bayes terhadap variasi pembagian data latih dan data uji. Nilai K yang digunakan dalam penelitian ini adalah K = 2 hingga K = 5.

Tabel 7. Hasil Evaluasi Model dengan K-Fold Cross Validation

Nilai K-Fold	Accuracy (%)	Precision (%)	Recall (%)
2	71,29	76,69	74,91
3	71,73	76,80	75,76
4	72,25	77,33	76,01
5	72,83	78,47	75,40

Berdasarkan Tabel 7, terlihat bahwa peningkatan nilai K-Fold cenderung diikuti oleh peningkatan nilai accuracy dan precision, yang menunjukkan bahwa model semakin stabil seiring bertambahnya jumlah fold. Nilai K = 5 menghasilkan kinerja terbaik dengan accuracy sebesar 72,83% dan precision sebesar 78,47%. Peningkatan nilai *K-fold* cenderung meningkatkan kinerja model Naïve Bayes, khususnya pada metrik accuracy dan precision. Nilai *K-fold* = 5 menghasilkan performa terbaik dengan accuracy sebesar 72,83% dan precision sebesar 78,47%, sehingga dipilih sebagai konfigurasi optimal dalam penelitian ini.



Gambar 4. Diagram Grafis Akurasi



Gambar 4 menunjukkan bahwa kinerja model Naïve Bayes cenderung meningkat seiring bertambahnya nilai K -fold, terutama pada metrik accuracy dan precision. Performa terbaik dicapai pada K -fold = 5 dengan accuracy 72,83% dan precision 78,47%, sehingga konfigurasi ini dipilih sebagai hasil pengujian optimal.

3.5. Pembahasan Hasil Klasifikasi Sentimen

Hasil klasifikasi sentimen terhadap opini publik mengenai kebijakan televisi digital menunjukkan bahwa sentimen positif lebih dominan dibandingkan sentimen negatif. Temuan ini mengindikasikan bahwa secara umum masyarakat memberikan respons yang relatif baik terhadap implementasi TV digital, terutama terkait peningkatan kualitas siaran dan kemajuan teknologi penyiaran.

Tabel 8. Distribusi Sentimen Opini Publik

Sentimen	Jumlah	Persentase
Positif	478	60,58%
Negatif	311	39,42%
Total	789	100%

Berdasarkan Tabel 8 sebesar 60,58% opini publik tergolong sentimen positif, sedangkan 39,42% lainnya menunjukkan sentimen negatif. Meskipun sentimen positif mendominasi, proporsi sentimen negatif yang masih cukup besar mencerminkan adanya resistensi dan kekhawatiran masyarakat, khususnya terkait aspek teknis, jangkauan sinyal, serta kebutuhan perangkat tambahan.

3.6. Analisis dan Interpretasi Hasil

Analisis dan interpretasi hasil penelitian dilakukan untuk memahami makna empiris dari hasil klasifikasi sentimen opini publik terhadap kebijakan televisi digital. Berdasarkan hasil analisis, mayoritas opini publik menunjukkan kecenderungan sentimen positif, yang mengindikasikan adanya penerimaan masyarakat terhadap kebijakan TV digital, khususnya terkait peningkatan kualitas siaran dan kemajuan teknologi penyiaran. Meskipun demikian, proporsi sentimen negatif yang masih relatif signifikan mencerminkan adanya tantangan dalam implementasi kebijakan. Sentimen negatif tersebut dapat diinterpretasikan sebagai bentuk kritik dan kekhawatiran masyarakat, terutama terkait aspek teknis, keterbatasan jangkauan sinyal, serta kesiapan infrastruktur dan perangkat pendukung. Hal ini menunjukkan bahwa penerimaan kebijakan tidak hanya ditentukan oleh manfaat teknologi, tetapi juga oleh kesiapan sistem pendukung dan efektivitas sosialisasi kepada masyarakat.

Dari sisi metodologis, hasil evaluasi model menunjukkan bahwa algoritma Naïve Bayes dengan representasi fitur TF-IDF mampu mengklasifikasikan sentimen opini publik secara cukup andal. Kinerja model yang stabil pada berbagai nilai K -Fold mengindikasikan bahwa pendekatan yang digunakan memiliki reliabilitas yang baik dalam menangani data teks media sosial yang bersifat dinamis dan tidak terstruktur. Secara keseluruhan, analisis dan interpretasi hasil ini menegaskan bahwa media sosial dapat dimanfaatkan sebagai sumber data alternatif untuk memahami persepsi publik terhadap kebijakan teknologi. Temuan penelitian ini memberikan dasar empiris yang relevan bagi pengambilan keputusan berbasis data, sekaligus menjadi landasan dalam penyusunan kesimpulan dan rekomendasi penelitian.

4. KESIMPULAN

Penelitian ini menganalisis opini publik terhadap kebijakan televisi digital di Indonesia dengan memanfaatkan data media sosial pada platform X menggunakan pendekatan text mining dan algoritma Multinomial Naïve Bayes. Data dikumpulkan melalui proses crawling menggunakan kata kunci “tv digital” pada periode penelitian tertentu, yang menghasilkan 1.855 tweet. Setelah melalui tahap seleksi dan cleaning data, diperoleh 789 tweet sebagai dataset akhir yang dianalisis. Hasil klasifikasi sentimen menunjukkan bahwa opini publik didominasi oleh sentimen positif sebesar 60,58%, sedangkan sentimen negatif mencapai 39,42%. Temuan ini mengindikasikan bahwa secara umum masyarakat menunjukkan penerimaan yang cukup baik terhadap kebijakan televisi digital, terutama terkait peningkatan kualitas siaran dan kemajuan teknologi penyiaran. Dari sisi metodologis, evaluasi kinerja model menunjukkan bahwa algoritma Naïve Bayes berbasis fitur TF-IDF mampu mengklasifikasikan sentimen dengan baik, ditunjukkan oleh nilai accuracy sebesar 79,21%, precision 82,45%, dan recall 85,06%. Meskipun demikian, penelitian ini memiliki keterbatasan, antara lain penggunaan satu kata kunci utama, keterbatasan pemahaman konteks bahasa informal dan sarkasme, serta penggunaan satu metode klasifikasi. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas variasi kata kunci, menggunakan pendekatan pemodelan yang lebih kompleks, dan menggabungkan analisis sentimen dengan metode lain untuk memperoleh pemahaman yang lebih komprehensif.

REFERENCES

- [1] T. Nur Arifin and Y. G. Adhiyoga, “Pengenalan Sistem Penyiaran Televisi Digital untuk Optimasi Spektrum Frekuensi UHF,” *Jurnal Tera*, vol. 3, no. 1, pp. 19–23, Mar. 2023, doi: 10.59832/jt.v3i1.198.
- [2] F. Abid *et al.*, “Emotion Detection on Platform X Comment with Naive Bayes Classification,” *International Journal of Artificial Intelligence & Robotics (IJAIR)*, vol. 7, no. 2, pp. 83–91, Dec. 2025, doi: 10.25139/ijair.v7i2.11156.
- [3] E. Politik, P. Persoalan, and D. R. Kebijakan, “Digitalisasi Televisi di Indonesia.”



- [4] K. Hasan and H. Sazali, "Transformasi Kebijakan dan Regulasi Komunikasi Siaran TV Analog ke Digital di Indonesia," *Jurnal Ilmu Sosial dan Ilmu Politik Malikussaleh (JSPM)*, vol. 6, no. 1, pp. 120–135, Jan. 2025, doi: 10.29103/jspm.v6i1.19360.
- [5] A. Utomo and A. Sari, "Dampak Ekonomi pada Migrasi Siaran TV Digital bagi Masyarakat dalam Pembagian Set Top Box Gratis Tidak Merata di Provinsi Banten," *Jurnal Minfo Polgan*, vol. 12, no. 1, pp. 926–933, Jun. 2023, doi: 10.33395/jmp.v12i1.12546.
- [6] I. Nurlita and Y. D. Lorena, "Opini Publik tentang Kebijakan Pemerintah Terhadap Migrasi Siaran dari Televisi Analog ke Televisi Digital," *CAKRAWALA*, vol. 17, no. 2, pp. 221–238, Dec. 2023, doi: 10.32781/cakrawala.v17i2.554.
- [7] . A. and N. K. P. Arisqi, "Pengaruh Kebijakan TV Digital Terhadap Daya Tarik Penonton TV Di Dusun Budegan," *JKOMDIS : Jurnal Ilmu Komunikasi Dan Media Sosial*, vol. 4, no. 1, pp. 100–106, Jan. 2024, doi: 10.47233/jkomdis.v4i1.1462.
- [8] I. Nurlita and Y. D. Lorena, "Opini Publik tentang Kebijakan Pemerintah Terhadap Migrasi Siaran dari Televisi Analog ke Televisi Digital," *CAKRAWALA*, vol. 17, no. 2, pp. 221–238, Dec. 2023, doi: 10.32781/cakrawala.v17i2.554.
- [9] X. Dong and Y. Lian, "A review of social media-based public opinion analysis: Challenges and recommendations," *Technol. Soc.*, vol. 67, p. 101724, Nov. 2021, doi: 10.1016/j.techsoc.2021.101724.
- [10] T. Warin, C. Melchior, and N. de Marcellis-Warin, "Social Media Metrics and Popular Legitimacy: Content Analysis of Pre- and Post-COVID-19 Public Engagement With the World Health Organization on X," *J. Med. Internet Res.*, vol. 27, p. e69959, Dec. 2025, doi: 10.2196/69959.
- [11] N. Maulidia, "Peran Platform X dalam Meningkatkan Literasi Digital," *Jurnal Cahaya Edukasi*, vol. 2, no. 4, pp. 123–127, Nov. 2025, doi: 10.63863/jce.v2i4.157.
- [12] A. T. S. Kom. , M. Kom. Zy and S. S. Kom. , M. Kom. Butsianto, *Konsep Media Sosial Implementasi dan Aplikasi*, 1st ed., vol. 1. Bekasi: PT. Vinicho Media Publisindo, 2024.
- [13] D. Arsyanda, S. Rodiah, and A. S. Rohman, "Peran akun autobase X (twitter) dalam memenuhi kebutuhan informasi followers," *Pustaka Karya : Jurnal Ilmiah Ilmu Perpustakaan dan Informasi*, vol. 13, no. 1, pp. 233–241, Jun. 2025, doi: 10.18592/pk.v13i1.15927.
- [14] A. G. T. AbuRaed, E. A. Prikryl, G. Carenini, and N. Z. Janjua, "Long COVID Discourse in Canada, the United States, and Europe: Topic Modeling and Sentiment Analysis of Twitter Data," *J. Med. Internet Res.*, vol. 26, p. e59425, Dec. 2024, doi: 10.2196/59425.
- [15] R. Chandrasekaran, R. Desai, H. Shah, V. Kumar, and E. Moustakas, "Examining Public Sentiments and Attitudes Toward COVID-19 Vaccination: Inveillance Study Using Twitter Posts," *JMIR Infodemiology*, vol. 2, no. 1, p. e33909, Apr. 2022, doi: 10.2196/33909.
- [16] A. Septia Wibowo and I. Istianah, "Reading Between the Lines: Incorporating Text Mining and Machine Learning in Financial Fraud Detection," *Asia Pacific Fraud Journal*, vol. 10, no. 1, pp. 73–93, Jul. 2025, doi: 10.21532/apfjournal.v10i1.382.
- [17] J. Kim *et al.*, "Text Mining Approaches for Exploring Research Trends in the Security Applications of Generative Artificial Intelligence," *Applied Sciences*, vol. 15, no. 6, p. 3355, Mar. 2025, doi: 10.3390/app15063355.
- [18] M. S. Md Suhaimin, M. H. Ahmad Hijazi, E. G. Mounq, P. N. E. Nohuddin, S. Chua, and F. Coenen, "Social media sentiment analysis and opinion mining in public security: Taxonomy, trend analysis, issues and future directions," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 9, p. 101776, Oct. 2023, doi: 10.1016/j.jksuci.2023.101776.
- [19] R. Rajab Asaad and R. Masoud Abdulhakim, "The Concept of Data Mining and Knowledge Extraction Techniques," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 17–20, Mar. 2021, doi: 10.48161/qaj.v1n2a43.
- [20] M. S. Md Suhaimin, M. H. Ahmad Hijazi, E. G. Mounq, P. N. E. Nohuddin, S. Chua, and F. Coenen, "Social media sentiment analysis and opinion mining in public security: Taxonomy, trend analysis, issues and future directions," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 9, p. 101776, Oct. 2023, doi: 10.1016/j.jksuci.2023.101776.
- [21] M. S. Alam, M. S. H. Mrida, and M. A. Rahman, "Sentiment Analysis in Social Media: How Data Science Impacts Public Opinion Knowledge Integrates Natural Language Processing (NLP) with Artificial Intelligence (AI)," *American Journal of Scholarly Research and Innovation*, vol. 4, no. 1, pp. 63–100, Jan. 2025, doi: 10.63125/r3sq6p80.
- [22] S. Kom. , M. Kom. Prastyadi Wibawa Rahayu *et al.*, *Buku Ajar Data Mining*, 1st ed., vol. 1. Jambi: PT. Sonpedia Publishing Indonesia, 2024.
- [23] M. S. Asyaky, Muhammad Al-Husaini, and Hen Hen Lukmana, "Sentiment Analysis on Short Social Media Texts Using DistilBERT," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 7, no. 2, pp. 524–533, May 2025, doi: 10.47709/cnahpc.v7i2.5836.
- [24] T. R. Widodo, I. N. Fajri, and B. W. Sari, "Sentiment Analysis of the Film 'JUMBO' on Twitter Using the Naive Bayes Method and Support Vector Machine (SVM) with a Text Mining Approach," *Journal of Applied Informatics and Computing*, vol. 9, no. 5, pp. 2861–2868, Oct. 2025, doi: 10.30871/jaic.v9i5.10557.
- [25] M. N. Dewi and R. E. Putra, "Pengembangan Sistem Analisis Sentimen Masyarakat Terhadap Chatgpt Pada Twitter Dengan Perbandingan Metode Naive Bayes Classifier Dan K-Nearest Neighbors," *Journal of Informatics and Computer Science (JINACS)*, vol. 5, no. 03, pp. 344–357, Jan. 2024, doi: 10.26740/jinacs.v5n03.p344-357.
- [26] L. E. Pradana and Y. Ruldeviyani, "Sentiment Analysis of Nanovest Investment Application Using Naive Bayes Algorithm," *Jurnal Nasional Pendidikan Teknik Informatika (JANPATI)*, vol. 12, no. 2, pp. 283–293, Jul. 2023, doi: 10.23887/janapati.v12i2.62302.
- [27] L. Ellyanti, Yova Ruldeviyani, Lelianto Eko Pradana, and Andro Harjanto, "Sentiment Analysis of Twitter Users to the PeduliLindungi Using Naïve Bayes Algorithm," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 2, pp. 414–421, Mar. 2023, doi: 10.29207/resti.v7i2.4684.
- [28] P. Prihatini, K. Indah, G. Sukerti, I. Indrayana, and I. Sudiarta, "Feature Extraction Performance on Classified Methods for Text Sentiment Analysis," in *Proceedings of the 4th International Conference on Applied Science and Technology on Engineering Science*, SCITEPRESS - Science and Technology Publications, 2021, pp. 1235–1243. doi: 10.5220/0010962900003260.
- [29] L. Zhang, "Features extraction based on Naive Bayes algorithm and TF-IDF for news classification," *PLoS One*, vol. 20, no. 7, p. e0327347, Jul. 2025, doi: 10.1371/journal.pone.0327347.
- [30] L. D. Setiawan, "Stuck on the Surface: Digitalization Case Studies of Three Local Televisions," *Mediator: Jurnal Komunikasi*, vol. 16, no. 2, pp. 212–226, Dec. 2023, doi: 10.29313/mediator.v16i2.2683.



- [31] G. K. Nathanael, "Understanding recession response by Twitter users: A text analysis approach," *Heliyon*, vol. 10, no. 1, p. e23737, Jan. 2024, doi: 10.1016/j.heliyon.2023.e23737.
- [32] Andre Perbawa, S. Sunarto, and Nandang Mulyasantosa, "Socialization of Digital Television Broadcast Technology Conversion (Study Regarding the Representation of Digital Television Broadcast Technology in Omnibus Short Film)," *International Journal of Social Science*, vol. 3, no. 6, pp. 709–728, Apr. 2024, doi: 10.53625/ijss.v3i6.7635.
- [33] M. Yang and Z. Li, "The influence of green human resource management on employees' green innovation behavior: The role of green organizational commitment and knowledge sharing," *Heliyon*, vol. 9, no. 11, p. e22161, Nov. 2023, doi: 10.1016/j.heliyon.2023.e22161.
- [34] P. Fortuna, J. Soler-Company, and L. Wanner, "How well do hate speech, toxicity, abusive and offensive language classification models generalize across datasets?," *Inf. Process. Manag.*, vol. 58, no. 3, p. 102524, May 2021, doi: 10.1016/j.ipm.2021.102524.
- [35] B. Wahyu Andrian, F. Adline Twince Tobing, I. Zuhdi Pane, and A. Kusnadi, "Implementation of Naïve Bayes Algorithm in Sentiment Analysis of Twitter Social Media Users Regarding Their Interest to Pay the Tax," *International Journal of Science, Technology & Management*, vol. 4, no. 6, pp. 1733–1742, Nov. 2023, doi: 10.46729/ijstm.v4i6.1015.
- [36] Y.-C. Yang, M. A. Al-Garadi, W. Bremer, J. M. Zhu, D. Grande, and A. Sarker, "Developing an Automatic System for Classifying Chatter About Health Services on Twitter: Case Study for Medicaid," *J. Med. Internet Res.*, vol. 23, no. 5, p. e26616, May 2021, doi: 10.2196/26616.