Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026 Page: 23−31
ISSN 2962-3944 (media online)
DOI: 0.62866/buai.v5i1.238
https://journal.grahamitra.id/index.php/buai

Analisis Sentimen Pengaruh Media Sosial Terhadap Keputusan Pembelian Konsumen Menggunakan Metode K-Nearest Neighbors (K-NN)

Adinda Febiola*, Ratih Manalu, Retno Ajeng Kartika Said, Putrama Alkhairi

[1,2,3,4]Teknik Informatika, STIKOM Tunas Bangsa, Pematang Siantar, Indonesia
Email: [1,*] adindafebiola2005@gmail.com, [2] ratihmanalu01@email.com, [3] retnoajeng750@email.com,
[4] putrama@amiktunasbangsa.ac.id
Email Penulis Korespondensi: adindafebiola2005@gmail.com

Abstrak−Perkembangan media sosial telah mengubah cara konsumen berinteraksi dan mengambil keputusan pembelian. Media sosial menjadi platform utama bagi konsumen untuk berbagi pengalaman, memberikan ulasan, dan mendiskusikan produk atau layanan. Penelitian ini bertujuan untuk menganalisis sentimen konsumen terhadap ulasan produk dimedia sosial serta dampaknya pada keputusan pembelian menggunakan metode K-NN. Data yang digunakan berupa ulasan konsumen yang dikumpulkaan pada platform media sosial. Data tersebut diproses melalui tahapan preprocessing, seperti tokenisasi, penghapusan kata tidak penting, dan stemming, sebelum diklasifikasikan menjadi sentimen positif dan negatif. K-NN dipilih karena keandalannya dalam klasifikasi berbasis jarak dan kemampuannya menangani datasetdengan pola distribusi yang kompleks. Hasil penelitian menunjukkan bahwa metode K-NN mampu mengklasifikasikan sentimen konsumen dengan akurasi 100%. Hasil analisis menunjukkan bahwa sentimen positif berkonstribusi signifikan dalam mempengaruhi keputusan keputusan pembelian, sedangkan sentimen negatif cenderung mengurangi minat konsumen. Hasil yang diperoleh dari tahap modelling dengan menggunakan metode K-NN dan perbandingan 80:20 untuk data training dan testing, maka nilai akurasi yang dihasilkan sebesar 90%, precision 90%, recall 100%. Evaluasi model dilakukan menggunakan matrik akurasi, presisi dan recall, dengan hasil yang menunjukkan kinerja K-NN cukup baik dalam mengklasifikasikan sentimen konsumen. Penelitian ini memberikan wawasan tentang pentingnya memahami sentimen konsumen di media sosial untuk strategi pemasaran yang lebih efektif, serta menunjukkan potensi penerapan machine learning, khususnya K-NN dalam menganalisis data tekstual untuk mendukung pengambilan keputusan bisnis.

Kata Kunci: Analisis Sentimen; Media Sosial; Konsumen; Keputusan Pembelian; K-NN

Abstract−The growth of social media has transformed how consumers interact and make purchasing decisions. Social media has become a primary platform for consumers to share experiences, provide reviews, and discuss products or services. This study aims to analyze consumer sentiment regarding product reviews on social media and its impact on purchasing decisions using the K-NN method. The data consists of consumer reviews collected from social media platforms. The data underwent preprocessing stages—such as tokenization, stop-word removal, and stemming—before being classified into positive and negative sentiments. K-NN was selected due to its reliability in distance-based classification and its ability to handle datasets with complex distribution patterns. The results indicate that the K-NN method is capable of classifying consumer sentiment with 100% accuracy. Analysis shows that positive sentiment significantly influences purchasing decisions, whereas negative sentiment tends to diminish consumer interest. In the modeling phase, using the K-NN method with an 80:20 split for training and testing data, the results yielded 90% accuracy, 90% precision, and 100% recall. Model evaluation was conducted using accuracy, precision, and recall metrics, demonstrating K-NN's strong performance in classifying consumer sentiment. This study offers insights into the importance of understanding consumer sentiment on social media for more effective marketing strategies and highlights the potential of machine learning specifically K-NN in analyzing textual data to support business decision-making.

Keywords: Sentiment Analysis; Social Media; Consumer; Purchasing Decision; K-NN

1. PENDAHULUAN

Di era digital saat ini, perkembangan teknologi informasi telah mengubah cara masyarakat berinteraksi dan mengambil keputusan, termasuk dalam hal membeli produk atau menggunakan layanan tertentu Di era digital saat ini, perkembangan teknologi informasi telah mengubah cara masyarakat berinteraksi dan mengambil keputusan, termasuk dalam hal membeli produk atau menggunakan layanan tertentu[1]. Media sosial merupakan salah satu bentuk perkembangan teknologi yang memberikan pengaruh besar dalam memudahkan masyarakat untuk berkomunikasi dan berinteraksi di era digital[2]. Media sosial, seperti Twitter, Facebook, dan Instagram, tidak hanya berfungsi sebagai sarana komunnikasi, tetapi juga sebagai platform untuk berbagi opini, pengalaman, dan rekomendasi produk[3]. Pengguna dapat dengan mudah mempublikasikan ulasan, memberikan rating, dan menyebarkan pengalaman mereka secara luas dan cepat. Hal ini memberikan konsumen akses langsung ke informasi yang lebih beragam dan autentik, yang berasal dari pengalaman pengguna lain. Sebagai hasilnya, media sosial kini memiliki pengaruh yang besar terhadap keputusan pembelian konsumen[4]. Keputusan pembelian adalah proses dimana konsumen memutuskan apakah akan terus membeli suatu produk atau meninggalkannya. Proses ini mencakup pertimbangan berbagai faktor yang mungkin mempengaruhi keputusan akhir[5].Keputusan pembelian sangat dipengaruhi oleh opini publik yang muncuk di media sosial, karena konsumen cenderung mempercayai ulasan dan pendapat dari pengguna lain yang dianggap lebih objektif dibandingkan iklan tradisional[6]. Ulasan positif atau negatif yang tersebar di media sosial dapat mempengaruhi persepsi dan minat beli konsumen terhadap suatu produk[7]. Sebagai contoh produk yang mendapatkan banyak ulasan positif akan lebih menarik bagi calon pembeli, sedangkan ulasan negatif dapat menurunkan minat beli terhadap produk tersebut. Fenomena ini menciptakan kebutuhan akan pemahaman yang mendalam terhadap sentimen konsumen dalam media sosial[8]. Oleh karena itu, dibutuhkan suatu analisis sentimen untuk membantu konsumen dalam mengambil keputusan apakah ulasan tersebut mengandung opini positif atau negatif. Analisis sentimen, yaitu

Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026 Page: 23−31
ISSN 2962-3944 (media online)
DOI: 0.62866/buai.v5i1.238
https://journal.grahamitra.id/index.php/buai

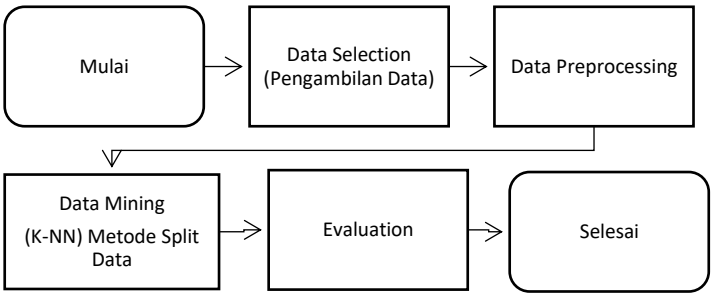
proses mengindentifikasi dan mengklasifikasikan opini publik menjadi kategori positif, negatif, atau netral menjadi alat penting bagi perusahaan untuk memehami bagaimana produk atau layanan mereka diterima oleh konsumen[9]. Tujuan dari analisis sentimen adalah untuk menarik kesimpulan, mengidentifikasi emosi dalam data, dan mengklasifikasikan polaritas[10]. Teknik klasifikasi yang umum digunakan untuk analisis sentimen meliputi Naïve Bayes (NB)[11] dan Decision Tree[12]. Beberapa peneliitian yang menjadi rujukan penelitian ini diantaranya penelitian mengenai analisis pengaruh media sosial terhadap minat beli skincare pada remaja di Indonesia. Data set penelitian diperoleh melalui kuesioner dengan data yang berjumlah 257. Hasil analisis dari penelitian ini adalah berupa diagram sentimen positif dan negatif dengan hasil positif 0,8138 dan yang negatif 0,7526 menunjukkan bahwa sentimen positif atau negatif di media soasial berkorelasi dengan minat beli skincare di kalangan remaja[11]. Penelitian berikutnya menerapkan algoritma decision tree dalam menganalisis keputusan pembelian program pada perkumpulan penggiat proggrammer di Indonesia. Data set penelitian berupa hasil survei kepada para konsumen pengguna program. Hasil analisis pada penelitian ini menhgasilkan tingkat akurasi sebesar 83,33%[12]. Kedua penelitian tersebut mengungkapkan peran algoritma yang berbeda dalam analisis sentimen untuk meningkatkan akurasi dan efektivitas klasfifikasi pada data yang beragam.

Untuk mengatasi permasalahan tersebut, salah satu metode yang efektif untuk analisis sentimen terhadap pengaruh media sosial adalah K-Nearest Neighbors (K-NN)merupakan metode berbasis kesamaan atau jarak antar data. K-NN bekerja dengan mencari “tetangga” terdekat dari data yang baru berdasarkan kemiripan dengan data yang telah dikalsifikasikan. Misalnya, jika sebagian besar ”tetangga” terdekat suatu teks memiliki sentimen positif , maka teks tersebut juga akan diklasifikasikan sebagai positif, dan sebaliknya jika sentimen negatif[13]. Keunggulan metode K-NN terletak pada kemudahan penggunaan dan fleksibilitasnya dalam mengklasifikasikan data dari berbagai jenis dan ukuran[14]. Dengan metode ini, sentimen dalam teks yang diunggah pengguna media sosial dapat diklasifikasikan untuk mengidentifikasi apakah opini mereka terhadap produk bersifat positif, negatif, atau netral[15]. Penelitian ini bertujuan untuk mengeksplorasi pengaruh media sosial terhadap keputusan pembelian konsumen dengan menggunakan metode K-NN dalam analisis sentimen. Dengan metode K-NN, sentimen dari teks di media sosial dapat diklasifikasikan ke dalam kategori tertentu (misalnya positif, negatif, atau netral) berdasarkan kemiripannya dengan data yang sudah diberi label[16]. Informasi ini dapat membantu perusahaan atau pemasar untuk mengidentifikasi pola sentimen konsumen, memprediksi keputusan pembeli, dan merancang strategi pemasaran yang lebih efektif agar sesuai denga preferensi atau persepsi konsumen yang berkembang di media sosial[17]. Dengan kata lain, analisis sentimen dan metode K-NN adalah kombinasi yang baik untuk mendapatkan wawasan mendetail tentang opini dan reaksi masyarakat terhadap keputusan tertentu. Melalui penerapan K-NN, diharapkan dapat diperoleh pemahaman yang lebih mendalam mengenai bagaimana opini publik di media sosial dapat memengaruhi perilaku dan keputusan konsumen.

Berdasarkan latar belakang yang telah diuraikan, maka diusulkan penelitian ini dengan tujuan mengelompokkan dan memahami hubungan antara pengguna di media sosial dan keputusan konsumen untuk membeli suatu produk atau layanan. Hasil dari penelitian ini diharapkan dapat memberi kontribusi bagi perusahaan dalam merancang strategi pemasaran yang lebih efektif, yang disesuaikan dengan sentimen konsumen serta tren opini yang berkembang di media sosial. Selain itu hasil dari penelitian dalam penelitian ini juga dapat menjadi acuan untuk penelitian lain dengan objek yang sama.

2. METODOLOGI PENELITIAN

2.1 Kerangka Dasar Penelitian



Gambar 1. Kerangka Penelitian [18]

Pada Gambar 1 kerangka penelitian, merupakan tahapan dalam penyelesaian penelitian ini yang menggunakan tahapan KDD (Knowledge Discovery in Database)[19]. KDD merupakan cara ataupun metode untuk memperoleh informasi tentang data base yang tersedia. Dalam proses penyelesaiannya, melalui langkah-langkah berikut[18]:

1. Data Selection
   Data selection adalah tahap perolehan dan pemilihan sampel data yang ingin diolah untuk dijadikan suatu informasi penting.
2. Data Preprocessing

Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026 Page: 23−31
ISSN 2962-3944 (media online)
DOI: 0.62866/buai.v5i1.238
https://journal.grahamitra.id/index.php/buai

Data preprocessing adalah tahap pembersihan data dari data yang mengganggu (noise) dan data yang tidak konsisten.

3. Data Mining
   Proses data mining merupakan proses penggalian data dan pola dari data yang telah di proses sebelumnya untuk menajdi sebuah informasi dengan metode K-NN menggunakan metode split data.
4. Evaluation
   Tahap evaluasi merupakan tahap mempresentasikan pengetahuan dari sebelumnya. Selain itu juga dapat dijadikan sebagai bahan untuk mempertimbangkan pada penelitian selanjutnya.

2.3 Data set

Sumber data pada penelitian ini bersumber dari Kaggle https://shorturl.at/g5JBf. Data yang digunakan berasal dari ulasan yang dibuat oleh pengguna media sosial. Pengumpulan data dilakukan menggunakan pustaka dengan teknik scraping. Dalam proses pengumpulan data menggunakan 2 kata kunci yaitu ‘sentimen analysis’ dan ‘consumer behavior’. Selanjutnya, data tersebut melalui tahapan preprocessing agar dapat digunakan sebagai dataset dalam penelitian ini.

Tabel 1. Data Penelitian

NO	TEKS ULASAN	sentimen
1	Saya membeli HD Fire 7 untuk istri saya, saya suka miliknya dan memutuskan untuk membeli Fire HD 8 untuk saya sendiri. Saya menyukainya. Grafiknya bagus dan mudah digunakan.	POSITIF
2	Kakak saya memberi tahu saya tentang produk ini. Saya membelinya untuk anak saya dan dia menyukainya.	POSITIF
3	Speaker ini bagus untuk harganya. Saya suka speakernya yang bisa diputar 360 derajat sehingga Anda bisa mendapatkan suara yang sama di bagian mana pun di dalam ruangan.	POSITIF
4	Mendapatkan diskon Black Friday dan sepadan dengan harganya. Dilengkapi dengan semua fitur penting dan gambar yang bagus untuk video.	POSITIF
5	Cocok untuk menonton film, video, bermain game, dan berbelanja di Amazon. Ukurannya cukup kecil untuk dibawa bepergian, mudah dipasang, dan harganya sangat terjangkau.	POSITIF
6	Ini yang kedua saya beli sekarang. Pengiriman cepat dan pelayanannya bagus seperti biasa	POSITIF
7	Saya suka TV Fire saya, saya lebih menyukainya daripada TV kabel karena harganya lebih murah dan bagus untuk dipasang di setiap ruangan.	POSITIF
8	Saya suka, saya dapat menambahkan daftar belanja, daftar tugas, dan hal-hal lain sejauh ini baik-baik saja tanpa masalah di pihak saya	POSITIF
9	Kindle yang sangat bagus dengan daya tahan baterai yang lama. Kindle lama saya tidak dapat diperbarui, jadi saya membeli Kindle baru dan sangat puas dengan produknya. Barang sampai sesuai dengan yang dijanjikan oleh BESTBUY.COM.	POSITIF
10	Saya membelinya sebagai hadiah untuk putri saya, dia suka bermain	POSITIF
…	…….	…….
100	Saya suka menggunakannya setiap hari. Saya punya 3 cucu yang tinggal bersama saya dan mereka juga suka memainkannya.	POSITIF

2.4 Algoritma K-Neareast Neighbors

Algoritma K-NN merupakan salah satu pendekatan yang populer dalam klasifikasi data[20].Algoritma ini digunakan untuk mengklasifikasikan objek berdasarkan data pelatihan.Objek diklasifikasikan berdasarkan kedekatannya dengan tetangga terdekat atau nilainya sama dengan objek[21].Tujuan dari algoritma ini adalah untuk mengklasifikasikan objek baru menurut dengan mempertimbangkan atribut dan contoh pelatihannya[22].Prinsip inti dari algoritma ini adalah mencari nilai k pada data pelatihan dan menggunakan pengukuran jarak untuk menentukan tetangga terdekat[23].Nilai suara terbanyak diantara k tetangga terdekat selanjutnya dijadikan dasar penentuan kategori sampling berikutnya dan selain itu algoritma ini juga digunakan untuk prediksi data[24].
Prosedur metode tetangga terdekat k dijelaskan sebagai berikut[25]:

1. Tentukan parameter k yang akan digunakan dalam perhitungan K-NN ini
2. Hitunglah jarak antara data pengujian dan data pelatihan K-NN

   Rumus K-NN :

$$d^{(x,y)} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

   Berdasarkan rumus 1, d(x,y) merupakan jarak antara data x ke data y, xi merupakan nilai xi pada data training, yi merupakan nilai yi pada data testing, N merupakan variable data, dan I merupakan dimensi data

3. Mengurutkan berdasarkan nilai Euclideandistance
4. Menentukan hasil dari Nilai K tetangga yang paling terdekatnya

Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026 Page: 23−31
ISSN 2962-3944 (media online)
DOI: 0.62866/buai.v5i1.238
https://journal.grahamitra.id/index.php/buai

5. Target output merupakan kelas yang mayoritas

2.5 Data Mining

Data mining adalah pengolahan pengetahuan yang berguna dan menarik berdasarkan pola dalam data besar. Data yang digunakan berasal dari database, data warehouse, internet dan lainnya yang dapat diolah menjadi sebuah informasi. Data mining dapat digunakan dalam beberapa sektor dan bertujuan untuk berbagai hal yaitu meningkatkan pengetahuan dan lain sebagainya[26]. Salah satu bagian dari data mining adalah klasifiksi. Klasifikasi adalah proses menemukan model melalui analisis terhadap sekumpulan data pelatihan yang menggambarkan dan membedakan kelas label atau konsep data[27].

2.6 Text Mining

Text mining adalah proses pengelolaan data berupa teks yang dijadikan pola untuk menerima suatu informasi penting[28]. Dalam kata lain, text mining seperti hal nya data mining tetapi data yang diolah text mining adalah data berupa tekstual[29]. Sebelum melalukan text mining, data akan melalui tahap text preprocessing. Tahapan dari text preprocessing yaitu case folding, cleansing, tokenizing, stemming, dan stop removal[30].

3. HASIL DAN PEMBAHASAN.

3.1 Pengolahan Data (Preprocessing)

Setelah pengumpulan data dan membagi data menjadi komentar positif dan negatif, selanjutnya dilakukan proses pengolahan data atau preprocessing. Berikut ini adalah tahapan preprocessing:

1. Remove URL, yaitu proses menghilangkan atau membersihkan elemen URL yang terdapat dalam teks pada dataset hasil pengumpulan data. Proses ini merupakan bagian penting dalam tahap praproses data, khususnya ketika bekerja dengan data teks seperti komentar atau ulasan yang diambil dari media sosial atau platform digital. Kehadiran URL dalam data teks sering kali tidak memberikan kontribusi informasi yang relevan terhadap analisis atau pemodelan yang akan dilakukan, sehingga perlu dihapus agar tidak mengganggu hasil analisis maupun performa model. Misalnya, dalam kalimat yang mengandung tautan ke situs seperti YouTube atau situs lainnya, keberadaan URL hanya akan menjadi gangguan karena tidak menyampaikan informasi semantik yang berguna untuk proses seperti klasifikasi teks atau analisis sentimen. Tabel di bawah ini menunjukkan perbedaan antara data sebelum dan sesudah dilakukan proses Remove URL. Pada kolom “Sebelum,” terdapat kalimat yang mengandung URL lengkap ke suatu video YouTube. Setelah dilakukan pembersihan, URL tersebut dihapus, dan yang tersisa hanyalah informasi utama dari kalimat tanpa tautan. Hasil ini membuat data menjadi lebih bersih, fokus, dan siap untuk dianalisis lebih lanjut. Dengan demikian, proses Remove URL sangat penting dalam menjaga kualitas data sebelum memasuki tahap analisis berikutnya.Pada Tabel 1 merupakan contoh perbedaan sebelum dan sesudah proses.

Tabel 1. Jenis jenis database

Sebelum	Sesudah
aku rasa video ini kurang adeganmakan bakso seperti videohttps://www.youtube.com/watch?v=mtAyXJInSUI	akurasavideoinikurangadegan makan bakso

2. Annotation Removal merupakan proses penting dalam praproses data teks, yang bertujuan untuk menghapus tanda mention (simbol “@”) beserta teks yang mengikutinya. Tanda “@” biasanya digunakan dalam media sosial atau platform digital lainnya untuk menyebut akun pengguna tertentu, seperti dalam komentar, ulasan, atau percakapan online. Dalam konteks analisis data teks, informasi berupa mention tersebut umumnya tidak memberikan kontribusi yang berarti terhadap isi pesan utama yang ingin dianalisis. Oleh karena itu, perlu dilakukan penghapusan agar teks menjadi lebih bersih dan fokus pada informasi yang relevan. Proses ini dilakukan secara otomatis menggunakan teknik pemrosesan teks yang dapat mengenali dan menghapus pola karakter khusus tersebut. Contoh perbedaan antara data sebelum dan sesudah dilakukan annotation removal dapat dilihat pada Tabel 2. Pada kolom “Sebelum,” terdapat beberapa kata atau kalimat yang masih mengandung tanda “@” dan teks di belakangnya, seperti “review@indihometetap”. Setelah dilakukan pembersihan, tanda dan teks tersebut dihilangkan sehingga menjadi “reviewtetap”. Hasil ini menunjukkan bahwa proses ini sangat berguna dalam menyederhanakan struktur data teks, sehingga mempermudah proses analisis selanjutnya seperti klasifikasi, klastering, maupun analisis sentimen yang membutuhkan data bebas dari simbol tidak relevan..Pada tabel 2 merupakan contoh perbedaan sebelum dan sesudah proses anatotation removal.

Tabel 2. Hasil Anotation Removal

Sebelum	Sesudah

Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026 Page: 23−31
ISSN 2962-3944 (media online)
DOI: 0.62866/buai.v5i1.238
https://journal.grahamitra.id/index.php/buai

dariduluditontondiulasdi review@indihometetapijaringan tetap lambat	dariduluditontondiulasdi reviewtetapijaringantetaplambat

3. Stemming merupakan salah satu tahapan penting dalam proses praproses data teks, yang berfungsi untuk mengubah kata menjadi bentuk dasarnya. Proses ini dilakukan dengan cara menghilangkan imbuhan pada kata, baik itu awalan, akhiran, maupun sisipan, sehingga menghasilkan bentuk kata dasar atau kata akar. Tujuan dari stemming adalah untuk menyederhanakan variasi kata yang berbeda namun memiliki makna dasar yang sama, agar dapat dikenali sebagai satu entitas yang seragam dalam proses analisis. Misalnya, kata-kata seperti "tercepat", "dipercepat", atau "mempercepat" semuanya akan diubah menjadi bentuk dasar "cepat". Hal ini sangat membantu dalam mengurangi kompleksitas data teks serta meningkatkan akurasi dalam berbagai tugas pemrosesan bahasa alami (NLP) seperti klasifikasi teks, analisis sentimen, atau pencarian informasi. Pada Tabel 3 dapat dilihat perbedaan hasil stemming antara kata sebelum dan sesudah dilakukan proses ini. Misalnya, kata "korupsisipalingcepat" setelah stemming menjadi "korupsipalincepat", yang berarti kata-kata majemuk maupun gabungan kata yang mengandung imbuhan akan dipangkas ke bentuk yang lebih sederhana. Dengan demikian, stemming memberikan kontribusi besar dalam membuat data teks menjadi lebih bersih, seragam, dan lebih mudah dianalisis secara otomatis oleh sistem. Pada tabel 3 merupakan contoh hasil stemming.

Tabel 3. Hasil Stemming

Sebelum	Sesudah
internetdiindonesiapaling lambat, tapikalaukorupsipalingcepat	internetdiindonesiapaling lambattapikalaukorupsipalingcepat

4. Tokenize, memecah sekumpulan karakter atau kalimat menjadi sebuah potongan karakter atau kata-kata sesuai dengan kebutuhan biasa juga disebut tokenisasi.

5. Transforms Cases merupakan salah satu langkah penting dalam proses praproses data teks, yang dilakukan dengan cara mengubah semua huruf kapital yang terdapat dalam data menjadi huruf kecil atau huruf non-kapital. Tujuan dari proses ini adalah untuk menciptakan keseragaman dalam penulisan teks, sehingga dapat meminimalkan risiko kesalahan dalam proses tokenisasi maupun dalam penerapan model klasifikasi. Dalam analisis teks, perbedaan antara huruf besar dan huruf kecil dapat menyebabkan sistem memandang kata yang sama sebagai entitas yang berbeda. Misalnya, kata "Internet" dan "internet" akan dianggap sebagai dua kata berbeda jika tidak dilakukan proses transformasi ini. Oleh karena itu, menyamakan semua huruf menjadi huruf kecil sangat penting agar data menjadi seragam dan konsisten. Proses ini juga membantu meningkatkan akurasi dalam pencocokan kata selama analisis, terutama pada algoritma yang berbasis pembelajaran mesin. Tabel 4 menyajikan hasil dari proses Transforms Cases, yang menunjukkan perbedaan antara teks sebelum dan sesudah transformasi dilakukan. Terlihat bahwa kalimat atau kata yang sebelumnya ditulis dengan kombinasi huruf kapital diubah seluruhnya menjadi huruf kecil pada bagian hasil. Dengan begitu, data teks menjadi lebih terstruktur dan siap untuk diproses pada tahap analisis berikutnya secara lebih efisien dan akurat.Pada tabel 4 merupakan contoh hasil Transforms Cases

Tabel 4. HasilTransforms Cases

Sebelum	Sesudah
Luas negara ikut mempengaruhi cepat lambat pemerataan internet. Hal yang seperti ini harus ditangani	Luas negara ikut mempengaruhi cepat lambat pemerataan internet

6. *Filter Tokens*, menghilangkan kata-kata dengan panjang karakter tertentu, biasanya kata yang memiliki hanya 2 karakter yang tidak memiliki arti.

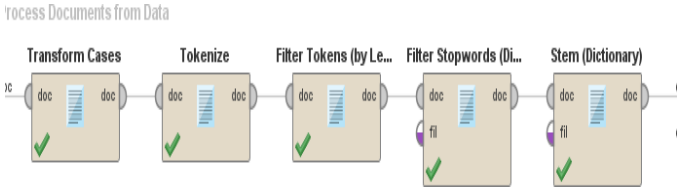
7. Filter Stopword merupakan salah satu tahapan penting dalam proses praproses data teks, khususnya pada analisis sentimen. Tahapan ini bertujuan untuk menghapus kata-kata yang dianggap tidak memiliki kontribusi berarti terhadap makna kalimat atau analisis sentimen. Kata-kata tersebut biasanya berupa kata hubung, kata sambung, atau kata keterangan umum yang sering muncul, seperti "dan", "yang", "adalah", "dengan", "di", "ke", dan sebagainya. Kata-kata semacam ini disebut stopword karena dalam konteks pemrosesan bahasa alami (Natural Language Processing), keberadaannya sering kali justru mengaburkan makna utama atau mengganggu pemodelan algoritma dalam mengidentifikasi informasi penting dari teks. Oleh karena itu, dengan menghapus stopword, model akan lebih fokus pada kata-kata yang mengandung makna utama atau informasi sentimen yang relevan. Proses ini sangat membantu dalam meningkatkan efisiensi analisis dan mengurangi kompleksitas data. Tabel 5 menunjukkan contoh hasil dari proses Filter Stopword, di mana dapat dilihat perbedaan antara data teks sebelum dan sesudah dilakukan penyaringan stopword. Hasilnya menunjukkan bahwa kata-kata yang kurang relevan telah berhasil dihilangkan, dan hanya tersisa kata-kata utama yang lebih signifikan untuk dianalisis lebih lanjut.

Tabel 5. HasilFilter Stopword

Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026 Page: 23−31
ISSN 2962-3944 (media online)
DOI: 0.62866/buai.v5i1.238
https://journal.grahamitra.id/index.php/buai

Sebelum	Sesudah
Semuaproviderinternetindonesiapalinglambat. tetapipalingmahal	memanginternetIndonesiapalinglambattetapipaling mahal

Berikut gambar 1 merupakan proses pengolahan data pada tools *Rapidminer*



Gambar 1. Proses Pengolahan Data Rapidminer.

3.2 Pembagian Data Training dan Data Testing

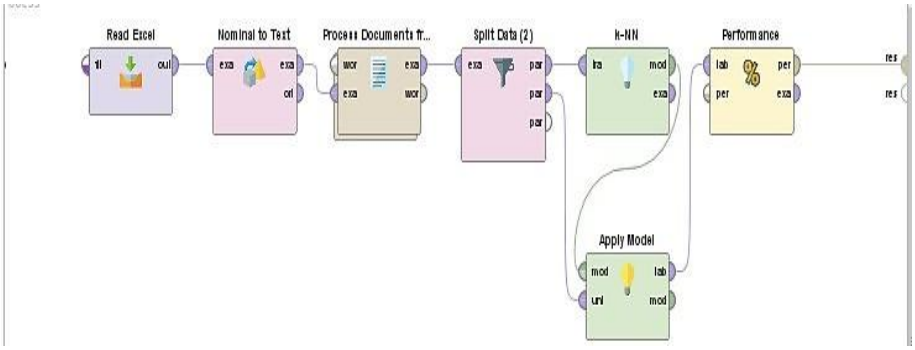
Data training adalah data latih yang akan digunakan untuk proses latihan bagi model klasifikasi. Data testing adalah data yang digunakan untuk uji rule klasifikasi. Berikut ini adalah perbandingan data training dan data testing 60:40, 70:30, dan 80:20 yang terdapat pada Tabel 6 Metode pembagiannya menggunakan metode stratified sampling.

Tabel 6. Sebaran Data Training dan Data Testing pada Dataset

Perbandingan 60:40			
Sentimen	*Training*	*Testing*	*Total*
Positive	42	28	70
Negative	18	12	30
Total	60	40	100
Perbandingan 70:30			
Sentimen	*Training*	*Testing*	*Total*
Positive	49	21	70
Negative	21	9	30
Total	70	30	100
Perbandingan 80:20			
Sentimen	*Training*	*Testing*	*Total*
Positive	64	16	80
Negative	16	4	20
Total	80	20	100

Pada tabel 6 Sebaran Data Training dan Data Testing pada Dataset, data digunakan untuk pelatihan (training) dengan proporsi data positif dan data negatif dari setiap model perbandingan. Data dibagi secara proporsional (stratified sampling) sehingga distribusi kelas tetap seimbang.

3.3 Modelling Menggunakan Split Data



Gambar 2. Metode K-NN Menggunakan Split Data

Berdasarkan Gambar 2, berikut penjelasan detail Metode K-NN menggunakan split data:

a. Read Excel, operator untuk memilih file excel yang akan digunakan.
b. Nominal to Text, berfungsi untuk mengubah jenis atribute nominal yang dipilih menjadi teks. Data teks diproses mengekstrak kata-kata dan mengirim kata menjadi vektor.
c. Process Document from Data, operator untuk preprocessing data yang digunakan untuk memasukkan operator transform cases, tokenize, filter tokenize, filter stamp dan stopwords didalamnya.
d. Split Data, operator yang digunakan untuk membagi data training dan data testing.

Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026 Page: 23−31
ISSN 2962-3944 (media online)
DOI: 0.62866/buai.v5i1.238
https://journal.grahamitra.id/index.php/buai

e. K-NN, operator yang menghasilkan metode yang digunakan untuk klasifikasi.
f. Apply Model, digunakan untuk menerapkan model yang telah di latih sebelumnya menggunakan data training pada data testing.
g. Performance, digunakan untuk mengevaluasi kinerja model yang memberikan daftar nilai kinerja secara otomatis, misalnya accuracy, precision dan recall.

3.4 Hasil Pemodelan Menggunakan Split Data

Setelah melalui proses preprocessing pada dataset, dilakukan pemodelan dengan membagi data ke dalam dua bagian utama, yaitu data training dan data testing. Pembagian ini bertujuan untuk mengevaluasi performa model klasifikasi secara objektif. Tiga variasi rasio data digunakan dalam pembagian data, yaitu 60:40, 70:30, dan 80:20. Hasil evaluasi dari masing-masing model berdasarkan akurasi, precision, dan recall ditampilkan pada Tabel 7. Berdasarkan hasil pada tabel, diketahui bahwa nilai akurasi untuk ketiga rasio berada dalam rentang 86.67% hingga 90%, di mana model dengan rasio 80:20 menghasilkan akurasi tertinggi, yaitu sebesar 90%. Sementara itu, model dengan rasio 70:30 memiliki akurasi terendah, yakni 86.67%. Meski terdapat perbedaan akurasi, pola hasil precision dan recall tetap menunjukkan kecenderungan yang seragam. Precision untuk kelas negatif dan positif adalah 0%, sedangkan recall untuk kelas negatif adalah 0% dan recall untuk kelas positif mencapai 100% di semua skenario. Hal ini mengindikasikan bahwa model hanya mampu mengenali satu kelas (kelas positif) dengan baik, sementara tidak mampu mengenali kelas lainnya (kelas negatif), sehingga terjadi ketidakseimbangan klasifikasi.

Tabel 7. Akurasi menggunakan Split Data.

Model	*Akurasi*	*Precision (Negative)*	*Precision (Positive)*	*Recall (Negative)*	*Recall (Positive)*
60:40:00	87.50%	0%	87.50%	0%	100%
70:30:00	86.67%	0%	86.50%	0%	100%
80:20:00	90%	0%	90%	0%	100%

3.4 Pengujian dan Evaluasi

Penelitian ini akan melakukan pengujian menggunakan confussion matrix yang diperoleh dari tahapan modelling dengan menggunakan metode K-NN pada Tabel 7 akurasi dengan split data Sehingga confussion matrix yang disajikan pada Tabel 8 adalah hasil dari evaluasi dan pengukuran klasifikasi K-NN dengan dataset yang berjumlah 100, terdapat 80 data training dan 20 data testing.

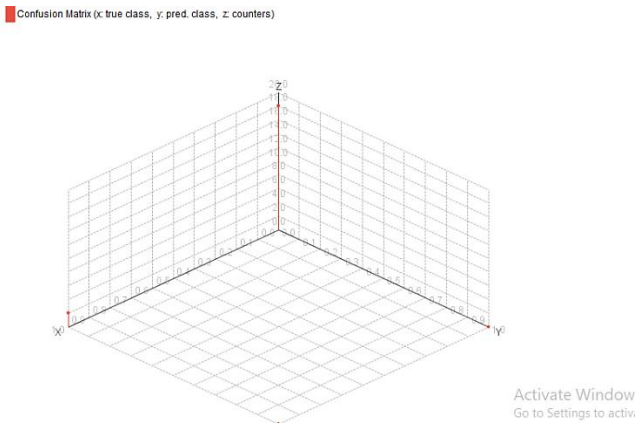
Tabel 8. Confussion Matrix

	True Positive	*True Negative*	*Class precision*
Pred.positive	18	2	90.00%
Pred. negatif	0	0	0.00%
Class recall	100.00%	0.00%	90.00%

Tabel 8 merupakan confussion matrix yang mengevaluasi performa model klasifikasi. Model berhasil memprediksi 88% sampel positif dengan benar (True Positive) dengan precision dan recall sebesar 100%. Namun, model gagal memprediksi kelas negatif, terlihat dari kalkulasi negatif yang 0 % dan precisionnya juga 0%. Hal ini menunjukkan bahwa model cenderung hanya fokus pada prediksi kelas positif sehingga performa untuk kelas negatif sangat rendah. Perhitungan dari Tabel 8 akurasi menggunakan split data adalah sebagai berikut:

a. *Precision* = $\frac{TP}{TP+FP} = \frac{18}{18+2}$ = 0,9% = 90%
b. *Recall* = $\frac{TP}{TP+FP} = \frac{18}{18+0}$ = 100%
c. *Accuracy* = $\frac{TP+TN}{TP+TN+FP+FN} = \frac{18+0}{18+0+2+0}$ = 0,9% = 90%

Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026 Page: 23−31
ISSN 2962-3944 (media online)
DOI: 0.62866/buai.v5i1.238
https://journal.grahamitra.id/index.php/buai



Gambar 3. Confussion Matrix 3D

Gambar 3 menyajikan visualisasi Confusion Matrix dalam format tiga dimensi (3D), yang bertujuan untuk mempermudah pemahaman terhadap hasil klasifikasi model. Pada plot tersebut, sumbu X merepresentasikan kelas sebenarnya (true class), sumbu Y menggambarkan kelas hasil prediksi model (predicted class), sementara sumbu Z menunjukkan jumlah sampel atau frekuensi dari prediksi yang terjadi (counters). Dalam konteks ini, setiap titik yang ditampilkan memiliki nilai tertentu yang menunjukkan seberapa banyak data yang diprediksi untuk kombinasi kelas tertentu. Terlihat bahwa garis vertikal berwarna merah hanya muncul pada koordinat X = 1 dan Y = 0, yang berarti model secara konsisten memprediksi kelas sebagai 0, meskipun kelas sebenarnya adalah 1. Ini menandakan bahwa model memiliki kecenderungan atau bias yang sangat kuat terhadap satu kelas saja. Dengan kata lain, seluruh prediksi model terfokus pada satu kelas, yaitu kelas 0, dan sama sekali tidak mengenali atau mengklasifikasikan data ke dalam kelas lainnya. Selain itu, tidak adanya nilai pada koordinat diagonal yang semestinya menampilkan prediksi benar, memperkuat fakta bahwa model mengalami kegagalan dalam melakukan klasifikasi yang akurat. Hal ini menunjukkan bahwa model tidak mampu mengenali distribusi kelas dengan seimbang, sehingga mengakibatkan kinerja yang sangat rendah. Model hanya mampu memetakan seluruh data ke satu kelas saja, yang secara langsung berdampak negatif terhadap akurasi dan performa keseluruhan dari sistem klasifikasi yang digunakan.

4. KESIMPULAN

Penelitian ini menganalisis pengaruh media sosial terhadap keputusan pembelian konsumen dengan menggunakan metode K-NN. Hasil penelitian menunjukkan bahwa metode K-NN mampu mengklasifikasikan sentimen konsumen dengan akurasi 100%. Hasil analisis menunjukkan bahwa sentimen positif berkonstribusi signifikan dalam mempengaruhi keputusan keputusan pembelian, sedangkan sentimen negatif cenderung mengurangi minat konsumen. Hasil yang diperoleh dari tahap modelling dengan menggunakan metode K-NN dan perbandingan 80:20 untuk data training dan testing, maka nilai akurasi yang dihasilkan sebesar 90%, precision 90%, recall 100%. Hasil analisis menunjukkan bahwa media sosial memiliki peran signifikan dalam membentuk keputusan konsumen, terutam melalui ulasan, rekomendasi, dan interaksi soaial.Sentimen konsumen yang diekstraksi dari komentar dan ulasan di media sosial mencerminkan kecenderungan positif dan negatif terhadap suatu produk atau layanan.Metode K-NN berhasil mengklasifikasikan seentimen dengan tingkat akurasi yang cukup tinggi, mennunjukkan bahwa metode ini efektif untuk menganalisis data teks yang tidak terstruktur.Sentimen positif terbukti memiliki korelasi kuat dengan peningkatan minat beli, sementara sentimen negatif cenderung mengurangi minat konsumen.Selain itu,interaksi yang aktif antara merek dan konsumen di media sosial turut memprkuat leputusan pembelian. Penelitian ini menegaskan pentingnya strategi pemasaran digital yang berfokus pada pengelolaan sentimen positif di media sosial. Perusahaan diharapkan dapat memanfaatkan hasil ini untuk menyusun kampanye pemasaran yang lebih efektif, seperti membangun hubungan baik dengan konsumen dan menanggapi keluhan dengan cepat. Dengan demikian, media sosial dapat menjadi alat yang strategis dalam meningkatkan loyalitas konsumen dan mendorong pertumbuhan bisnis.

REFERENCES

[1] A. Adriyanti and A. H. Abubakar, “Pengaruh Perilaku Konsumen Terhadap Keputusan Pembelian Produk Kecantikan Melalui E-Commerce Shopee,” *J. Pendidik. Ekon.*, vol. 17, pp. 239–248, 2023, doi: 10.19184/jpe.v17i2.42375.
[2] R. N. Shadrina and Y. Sulistyanto, “Analisis Pengaruh Content Marketing, Influencer, Dan Media Sosial Terhadap Keputusan Pembelian Konsumen (Studi Pada Pengguna Instagram Dan Tiktok Di Kota Magelang),” vol. 11, no. 2, pp. 1–11, 2022.
[3] S. C. Mokoginta, “Dampak Media Sosial Terhadap Perilaku Konsumen Dalam Pembelian Produk Fashion Online Di Kota Kotamobagu,” *J. Adm. dan Manaj.*, vol. 14, no. 4, pp. 621–631, 2024, doi: 10.52643/jam.v14i4.5111.
[4] F. W. Umbara, “User Generated Content di Media Sosial Sebagai Strategi Promosi Bisnis,” *J. Manaj. Strateg. dan Apl. Bisnis*, vol. 4, no. 2, pp. 572–581, 2021, doi: 10.36407/jmsab.v4i2.366.
[5] H. Ependi and R. W. Pahlevi, “Keputusan pembelian mahasiswa pada produk online shop shopee dan faktor penentunya,” *J. Competency Bus.*, vol. 5, no. 1, pp. 118–135, 2021, doi: 10.47200/jcob.v5i1.879.

[6] I. Yunus and A. Ariawan, "Keputusan Pembelian Konsumen: Perspektif Online Consumer Review," *PRAGMATIS J. Manaj. dan Bisnis*, vol. 3, no. 1, pp. 36–47, 2022, doi: 10.30742/pragmatis.v3i1.2476.
[7] S. Saepudin, S. Widiastuti, and C. Irawan, "Sentiment Analysis of Social Media Platform Reviews Using the Naïve Bayes Classifier Algorithm," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 2, pp. 236–243, 2023, doi: 10.32736/sisfokom.v12i2.1650.
[8] A. Budiyanto, I. B. Pamungkas, and A. Praditya, "Pengaruh Media Sosial Terhadap Minat Beli Dan Keputusan Pembelian Konsumen: Analisis Bibliometrik," *J. Ekon. Manaj.*, vol. 8, no. 2, pp. 133–142, 2022, doi: 10.37058/jem.v8i2.5468.
[9] S. A. Mahira, I. Sukoco, C. S. Barkah, and N. J. A. Novel, "Teknologi Artificial Intelligence Dalam Analisis Sentimen: Studi Literatur Pada Perusahaan Kata.Ai," *Responsive J. Pemikir. Dan Penelit. Adm. Sos. Hum. Dan Kebijak. Publik*, vol. 6, no. 2, pp. 139–148, 2023, doi: 10.24198/responsive.v6i2.48064.
[10] E. Hokijuliandy, H. Napitupulu, and F. Firdaniza, "Analisis Sentimen Menggunakan Metode Klasifikasi Support Vector Machine (SVM) dan Seleksi Fitur Chi-Square," *SisInfo J. Sist. Inf. dan Inform.*, vol. 5, no. 2, pp. 40–49, 2023, doi: 10.37278/sisinfo.v5i2.670.
[11] Heliyanti Susana, "Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet," *J. Ris. Sist. Inf. dan Teknol. Inf.*, vol. 4, no. 1, pp. 1–8, 2022, doi: 10.52005/jursistekni.v4i1.96.
[12] F. A. Artanto, I. Rosyadi, S. E. Rahmawati, and H. T. B. J. Pangestu, "Decision Tree Dalam Analisis Keputusan Pembelian Program Pada Perkumpulan Penggiat Programmer Indonesia," *J. Fasilkom*, vol. 12, no. 3, pp. 141–144, 2022, doi: 10.37859/jf.v12i3.3948.
[13] A. R. Isnain, J. Supriyanto, and M. P. Kharisma, "Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 15, no. 2, pp. 121–130, 2021, doi: 10.22146/ijccs.65176.
[14] T. A. C. Budianto, H. Fatoni, M. A. Syaharani, and C. Rozikin, "Analisis Sentimen Pengumuman Hasil Pemilu 2024 Di Sosial Media X Menggunakan Knn Dan Naïve Bayes Classifier," *JATI(Jurnal Mhs. Tek. Inform.*, vol. 8, no. 5, pp. 10816–10822, 2024, doi: 10.36040/jati.v8i5.11077.
[15] R. Q. Rohmansa, N. Pratiwi, and M. J. Palepa, "Analisis Sentimen Ulasan Pengguna Aplikasi Discord Menggunakan Metode K-Nearest Neighbor," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 1, pp. 368–378, 2024, doi: 10.29100/jipi.v9i1.4943.
[16] N. Alvionika, S. Faisal, R. Rahmat, and A. F. N. Masruriyah, "Analisis Sentimen Pada Komentar Instagram Provider By.U Menggunakan Metode K-Nearest Neighbors (KNN)," *J. Algoritm.*, vol. 21, no. 2, pp. 50–63, 2024, doi: 10.33364/algoritma/v.21-2.1672.
[17] A. Andirwan, V. Asmilita, M. Zhafran, A. Syaiful, and M. Beddu, "Strategi pemasaran digital: Inovasi untuk maksimalkan penjualan produk konsumen di era digital," *J. Ilm. Multidisiplin Amsir*, vol. 2, no. 1, pp. 155–166, 2023, doi: 10.62861/jimat%20amsir.v2i1.405.
[18] A. Yoga Pratama, Y. Umaidah, and A. Voutama, "Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja)," *J. Sains Komput. Inform.*, vol. 5, no. 2, pp. 897–910, 2021, doi: 10.30645/j-sakti.v5i2.386.
[19] M. A. A. Leza, N. W. Utami, and P. A. C. Dewi, "Prediksi Prestasi Siswa SMAS Katolik Santo Yoseph Denpasar Berdasarkan Kedisiplinan Dan Tingkat Ekonomi Orang Tua Menggunakan Metode Knowledge Discovery In Database Dan Algoritma Regresi Linier Berganda," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 373–379, 2024, doi: 10.36040/jati.v8i1.8754.
[20] F. Sutomo *et al.*, "Optimization of the K-Nearest Neighbors Algorithm Using the Elbow Method on Stroke Prediction," *J. Tek. Inform.*, vol. 4, no. 1, pp. 125–130, 2023, doi: 10.52436/1.jutif.2023.4.1.839.
[21] Sony Simare-mare and Henry Pandia, "Rekomendasi Komoditas Ekspor Menggunakan K-Nearest Neighbor," *JSiI (Jurnal Sist. Informasi)*, vol. 10, no. 2, pp. 150–156, 2023, doi: 10.30656/jsii.v10i2.8205.
[22] Y. Siagian, J. Hutahaean, A. Zikra Syah, J. Efendi Hutagalung, and A. Karim, "Implementasi Metode K-Nearest Neigbours (KNN) Untuk Klasifikasi Penyakit Diabetes," *J. Inform. dan Teknol. Inf.*, vol. 2, no. 3, pp. 253–262, 2024, doi: 10.56854/jt.v2i3.331.
[23] U. Nijunnihayah, S. S. Hilabi, F. Nurapriani, and E. Novalia, "Implementasi Algoritma K-Nearest Neighbor untuk Prediksi Penjualan Alat Kesehatan pada Media Alkes," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 695–701, 2024, doi: 10.57152/malcom.v4i2.1326.
[24] L. N. Aziza, R. Y. Astuti, B. A. Maulana, and N. Hidayati, "Penerapan Algoritma K-Nearest Neighbor untuk Klasifikasi Ketahanan Pangan di Provinsi Jawa Tengah," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 404–412, 2024, doi: 10.57152/malcom.v4i2.1201.
[25] M. F. Gusvino and D. Ahmad, "Algoritma K-Nearest Neighbor Terhadap Peluang Mahasiswa Menjadi Aktivis Kampus Pada Jurusan Matematika Universitas Negeri Padang," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. Dan Stat.*, vol. 4, no. 2, pp. 1202–1210, 2023, doi: 10.46306/lb.v4i2.401.
[26] H. Li and J. Li, "Research and Application of University English Writing Teaching Reform Based on Data Mining," *J. Cases Inf. Technol.*, vol. 26, no. 1, pp. 1–16, 2024, doi: 10.4018/JCIT.347667.
[27] D. I. Purnama, R. L. Islami, L. Sari, and P. R. Sihombing, "Analisis klasifikasi data tracer study dengan support vector machine dan neural network," *J. SISKOM-KB (Sistem Komput. Dan Kecerdasan Buatan)*, vol. 4, no. 2, pp. 46–52, 2021, doi: 10.47970/siskom-kb.v4i2.191.
[28] S. Supriyatna and E. Fahrudin, "Pemanfaatan Algoritma Text Mining dalam Menemukan Pola Risiko Bencana sebagai Pengetahuan Kebencanaan dari Dokumen Kajian Risiko Bencana (KRB)," *J. Inform. Utama*, vol. 2, no. 1, pp. 35–42, 2024, doi: 10.55903/jitu.v2i1.164.
[29] A. I. Rifai, "Analisis Text Clustering Masyarakat Di Twitter Mengenai Omnibus Law Menggunakan Orange Data Mining," *J. Inf. Syst. Informatics*, vol. 3, no. 1, pp. 1–12, 2021, doi: 10.33557/journalisi.v3i1.78.
[30] B. Hakim, "Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning," *JBASE - J. Bus. Audit Inf. Syst.*, vol. 4, no. 2, pp. 16–22, 2021, doi: 10.30813/jbase.v4i2.3000.

