Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2403−2412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6797

Perbandingan Algoritma *Random Forest,* KNN, SVM Untuk Analisis Sentimen Pengalaman Belanja *Thrift* Di X

M Rafi Raihandika*, Ryan Randy Suryono

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia
Email: [1,*]muhammad_rafi_raihandika@teknokrat.ac.id, [2]ryan@teknokrat.ac.id
Email Penulis Korespondensi: muhammad_rafi_raihandika@teknokrat.ac.id
Submitted: **20/01/2025**; Accepted: **26/02/2025**; Published**: 01/03/2025**

Abstrak−Fenomena thrifting semakin diminati, terutama oleh generasi milenial dan Generasi Z. Seiring dengan meningkatnya ketertarikan masyarakat terhadap belanja thrifting, media sosial X pun muncul sebagai salah satu platform utama bagi masyarakat untuk berbagi pengalaman dan opini terkait belanja thrift. Penelitian ini bertujuan menganalisis sentimen masyarakat tentang pengalaman belanja thrift dengan membandingkan kinerja algoritma Random Forest, Support Vector Machine (SVM), dan K-Nearest Neighbors (KNN). Dataset yang digunakan pada penelitian ini diperoleh dari Twitter sebanyak 6.390 tweet yang dikumpulkan melalui teknik crawling dengan rentang waktu 2 Agustus 2024 hingga 4 September 2024. Dataset lalu diproses untuk menghasilkan data bersih. Setelah proses pembersihan, data dibagi 80:20 untuk pelatihan dan pengujian. Dalam pengujian ketiga algoritma, diperoleh tingkat akurasi yang menunjukkan seberapa baik model dalam membuat prediksi. Akurasi ini mengukur sejauh mana model berhasil memprediksi sentimen pengalaman belanja thrifting berdasarkan dataset Twitter. Hasil menunjukkan bahwa algoritma Random Forest memiliki akurasi tertinggi dengan 95%, precision 97%, recall 78%, dan f1-score 85%. SVM mencapai akurasi 93%, precision 93%, recall 72%, dan f1-score 78%. KNN memperoleh akurasi 89%, precision 72%, recall 59%, dan f1-score 61%. Dari hasil yang diperoleh, algoritma *Random Forest* menunjukkan akurasi terbaik untuk analisis sentimen pengalaman *thrifting* di *Twitter* Indonesia. Keunggulannya terletak pada pendekatan *ensemble learning* yang stabil, di mana beberapa *decision tree* digabungkan untuk menghasilkan prediksi yang lebih akurat. Kemampuan ini membuat *Random Forest* efektif menangani data teks *Twitter* yang variatif dan kompleks, menjadikannya algoritma paling andal dalam konteks ini.

Kata Kunci: Analisis Sentimen; KNN; *Random Forest*; SVM; *Thrifting*; *Twitter*

Abstract−The thrifting phenomenon is gaining traction, especially among millennials and Generation Z. Along with the increasing interest in thrifting, X social media has emerged as one of the main platforms for people to share experiences and opinions related to thrift shopping. This research aims to analyze people's sentiments about thrift shopping experiences by comparing the performance of Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) algorithms. The dataset used in this study was obtained from Twitter as many as 6,390 tweets collected through crawling techniques with a time span of August 2, 2024 to September 4, 2024. The dataset is then processed to produce clean data. After the cleaning process, the data is divided 80:20 for training and testing. In testing the three algorithms, an accuracy level is obtained that shows how well the model makes predictions. This accuracy measures the extent to which the model successfully predicts the sentiment of the thrifting shopping experience based on the Twitter dataset. The results show that the Random Forest algorithm has the highest accuracy with 95%, precision 97%, recall 78%, and f1-score 85%. SVM achieved 93% accuracy, 93% precision, 72% recall, and 78% f1-score. KNN obtained 89% accuracy, 72% precision, 59% recall, and 61% f1-score. From the results obtained, the Random Forest algorithm shows the best accuracy for sentiment analysis of thrifting experiences on Twitter Indonesia. Its advantage lies in its stable ensemble learning approach, where multiple decision trees are combined to produce more accurate predictions. This ability makes Random Forest effective in handling varied and complex Twitter text data, making it the most reliable algorithm in this context.

Keywords: Sentiment Analysis; KNN; Random Forest; SVM; Thrifting; Twitter

1. PENDAHULUAN

Bisnis *e-commerce* memiliki peran yang sangat penting dalam mendorong pertumbuhan dunia usaha, terutama di tengah arus globalisasi dan perkembangan teknologi informasi yang pesat. *Platform* toko online menjadi sarana utama untuk memperluas jangkauan distribusi produk sekaligus memenuhi kebutuhan konsumen di era digital. Seiring dengan dinamika bisnis yang terus berubah, berbagai tren belanja baru bermunculan, salah satunya adalah *thrifting*[1]. *Thrifting* adalah aktivitas membeli barang bekas, seperti pakaian, sepatu, aksesori, atau barang lainnya, yang masih memiliki kualitas baik namun dijual dengan harga terjangkau. Tren ini semakin populer di kalangan milenial dan generasi Z di Indonesia karena menawarkan berbagai keuntungan. Selain mendukung gaya hidup hemat, *thrifting* juga memberi peluang bagi konsumen untuk mengekspresikan gaya unik mereka, seperti menemukan pakaian *vintage*, aksesori langka, atau barang-barang unik yang sulit ditemukan di toko konvensional[2]. Namun, beberapa orang masih beranggapan bahwa barang hasil *thrifting* hanyalah barang bekas yang kurang bernilai, sehingga konsumen yang memilih *thrifting* kerap diasosiasikan dengan kelompok masyarakat kelas bawah. Padahal, *thrifting* sebenarnya memberikan pengalaman belanja yang berbeda, di mana konsumen bisa memperoleh barang-barang unik dengan harga yang lebih terjangkau[3].

Fenomena thrifting dalam dunia pemasaran dapat menjadi peluang sekaligus hambatan bagi pelaku usaha. Perusahaan dapat mengambil manfaat dari tren ini dengan menciptakan strategi pemasaran yang lebih tepat sasaran, misalnya dengan menyediakan produk atau layanan yang terjangkau atau mengadakan promosi serta diskon yang menarik perhatian pelanggan[4]. Pendekatan ini sejalan dengan tujuan pembangunan ekonomi Indonesia Sebagaimana diatur oleh ketentuan dalam perundang-undangan yang berlaku RI tahun 2014 tentang Perdagangan, yang bertujuan meningkatkan kesejahteraan masyarakat melalui penerapan demokrasi ekonomi yang berbasis pada prinsip

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2403−2412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6797

kebersamaan, efisiensi, keadilan, keberlanjutan, dan pelestarian lingkungan, sekaligus menjaga kesatuan ekonomi nasional[5]. Namun, Memahami bagaimana masyarakat memandang tren *thrifting* sangat penting untuk menghindari kesalahan dalam strategi pemasaran yang dapat merugikan bisnis. Dalam perdagangan internasional, industri pakaian bekas, yang dikenal sebagai thrifting, telah menjadi salah satu sektor yang menarik perhatian karena pertumbuhannya yang pesat. Meskipun beberapa negara telah memperketat aturan impor untuk melindungi industri lokal, pakaian bekas tetap diminati konsumen dan terus menunjukkan perkembangan yang signifikan di pasar global[6].

Popularitas *thrifting* semakin meningkat seiring bertambahnya kesadaran masyarakat terhadap gaya hidup yang hemat dan ramah lingkungan. Media sosial X berperan besar dalam mendukung tren ini, menjadi platform utama bagi pengguna untuk berbagi pengalaman, pendapat, dan rekomendasi terkait belanja *thrift*. Hal ini turut mendorong pertumbuhan komunitas *thrift* di berbagai kalangan masyarakat. Opini masyarakat mengenai *thrifting* bervariasi, mulai dari yang positif, negatif, hingga netral. Keragaman ini menunjukkan perlunya penelitian lebih lanjut untuk memahami persepsi yang berkembang di media sosial. Salah satu metode yang dapat digunakan adalah analisis sentimen, yaitu proses otomatis yang mengolah data teks untuk menggali informasi tertentu[7]. Metode ini bertujuan mendeteksi opini terhadap suatu subjek atau objek, seperti individu, organisasi, atau produk, dalam kumpulan data yang tersedia[8]. Dengan menggunakan analisis sentimen, bisnis dapat memahami pandangan masyarakat secara lebih mendalam terhadap tren belanja thrift yang semakin populer. Informasi ini menjadi dasar penting untuk merumuskan strategi yang lebih efektif dalam mendukung pertumbuhan industri *thrifting* secara berkelanjutan.

Sebagai dasar kebaruan penelitian ini, sejumlah penelitian terdahulu telah dilakukan untuk memahami analisis sentimen, baik dalam konteks *thrifting* maupun sektor *e-commerce*. Salah satunya adalah penelitian oleh Firmansyah dan Damayanti (2024), yang menganalisis sentimen masyarakat terhadap *thrifting* di media sosial *Twitter* (X.com) menggunakan algoritma *K-Nearest Neighbor* (KNN). Mereka menemukan bahwa algoritma KNN menghasilkan akurasi tertinggi sebesar 76%, namun penelitian ini hanya terbatas pada satu algoritma tanpa perbandingan dengan algoritma lainnya, sehingga belum memberikan wawasan yang lebih luas mengenai performa model lain dalam konteks serupa[4]. Penelitian oleh Wardani, Saepudin, dan Warman (2022) mengkaji komentar masyarakat terkait kegiatan trading di *Twitter*. Hasilnya menunjukkan bahwa algoritma KNN memiliki akurasi tertinggi sebesar 0,999, diikuti oleh *Random Forest* (0,994) dan SVM (0,992), namun fokus mereka berada pada aktivitas *trading*, bukan *thrifting*, sehingga konteksnya berbeda[9]. Mahendrata dkk. (2023) menggunakan *Random Forest* untuk menganalisis sentimen terhadap *marketplace* di Indonesia dan mendapatkan akurasi sebesar 68,56%. Penelitian ini menunjukkan bahwa *Random Forest* dapat digunakan untuk analisis sentimen, namun performanya bergantung pada karakteristik data yang digunakan [10].

Sementara itu, penelitian Watmah (2021) membandingkan KNN, SVM, dan *Random Forest* untuk analisis sentimen di *e-commerce Shopee*. Hasilnya menunjukkan bahwa SVM unggul dengan akurasi 89,4%, *precision* 89,5%, dan *recall* 89,7%, diikuti oleh KNN dengan akurasi 89,0% dan *Random Forest* dengan akurasi 83,0%. Penelitian ini menekankan bahwa pemilihan algoritma sangat bergantung pada data dan konteks spesifik yang dianalisis[11]. Penelitian Safitri dkk. (2024) mengaplikasikan SVM untuk analisis sentimen *tren fashion* di media sosial dan menemukan bahwa SVM mencapai akurasi 80%[12]. Lillah dkk. (2023) juga menggunakan algoritma KNN untuk analisis sentimen ulasan pengguna aplikasi *Tokopedia* di *Play Store*, dengan hasil menunjukkan 89,2% berlabel positif dan 10,8% berlabel negatif, serta 97,0% positif dan 3,0% negatif pada data tertentu[13].

Berdasarkan penelitian-penelitian tersebut, terlihat bahwa algoritma terbaik untuk analisis sentimen bervariasi tergantung pada karakteristik data dan konteksnya. Penelitian sebelumnya juga lebih banyak berfokus pada sektor *e-commerce* atau konteks umum, dengan sedikit perhatian pada analisis sentimen khusus *thrifting* di media sosial. Misalnya, Firmansyah dan Damayanti (2024)[4]. hanya menggunakan algoritma KNN tanpa membandingkan performanya dengan algoritma lain, sementara Wardani dkk. (2022)[9]. dan Watmah (2021)[11]. lebih menitikberatkan pada sektor *trading* dan *e-commerce* tanpa membahas *thrifting* secara spesifik.

Sebagian besar penelitian terdahulu menggunakan tiga kelas sentimen (positif, negatif, dan netral), yang berbeda dengan penelitian ini, yang hanya menggunakan dua kelas sentimen (positif dan negatif) untuk menggali persepsi masyarakat terhadap *thrifting* secara lebih fokus. Oleh karena itu, penelitian ini bertujuan untuk mengisi celah tersebut dengan membandingkan kinerja algoritma *Random Forest*, KNN, dan *Support Vector Machine* (SVM) dalam analisis sentimen *thrifting* di *Twitter*. Hasil penelitian ini diharapkan dapat memberikan wawasan lebih mendalam tentang persepsi masyarakat terhadap *tren thrifting*, serta menjadi acuan bagi strategi pemasaran, kebijakan perdagangan barang bekas, dan penelitian selanjutnya di bidang ini.

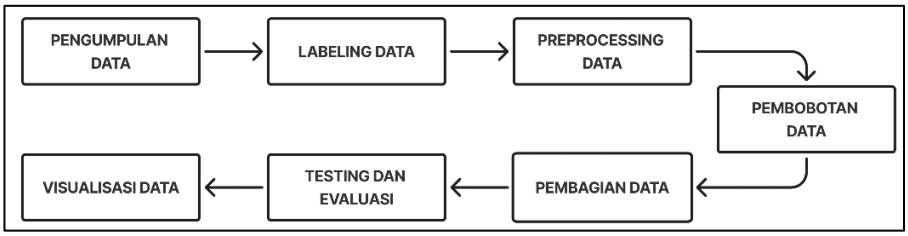
2. METODOLOGI PENELITIAN

2.1 Tahap penelitian

Tahapan penelitian yang dilakukan pada penelitian ini dapat dilihat pada gambar 1. Berdasarkan gambar tersebut, penelitian ini terdiri atas beberapa langkah yang dilakukan secara sistematis. Pertama, pengumpulan data dilakukan menggunakan *library Tweet Harvest*. Selanjutnya, data dilabeli menggunakan *TextBlob*, diikuti oleh tahap *preprocessing* untuk membersihkan dan mempersiapkan data. Setelah itu, dilakukan pembobotan data menggunakan metode *TF-IDF* sebelum data dibagi menjadi data latih dan data uji. Pengujian dilakukan dengan menerapkan tiga algoritma, yaitu *Random Forest*, *K-Nearest Neighbors* (KNN), dan *Support Vector Machine* (SVM). Langkah

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2403−2412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6797

selanjutnya adalah evaluasi hasil pengujian menggunakan *Confusion Matrix* untuk menilai performa model. Akhirnya, hasil penelitian divisualisasikan untuk mempermudah interpretasi. Metode penelitian secara rinci dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

2.2 Pengumpulan Data

Pengumpulan data atau proses *crawling* merupakan langkah awal dalam penelitian ini, yang bertujuan untuk mengumpulkan dataset. Dataset tersebut diperoleh melalui metode *Tweet Harvest*. *Tweet Harvest* adalah sebuah *library* yang dimanfaatkan untuk mengakses data dari platform media sosisal X, dengan memanfaatkan *access token* dari akun yang telah terdaftar pada platform tersebut[14]. Data dikumpulkan dengan rentang waktu 2 Agustus 2024 sampai dengan 4 September 2024.

2.3 Labelling

Pada tahap ini, setiap data dikategorikan untuk menentukan apakah termasuk dalam sentimen positif atau negatif. Proses ini dilakukan dengan memanfaatkan *library TextBlob*, sebuah *library Python* yang dirancang untuk mengolah dan menganalisis teks[15]. *TextBlob* menyediakan API untuk berbagai tugas pemrosesan bahasa alami (NLP), seperti analisis sentimen, klasifikasi teks, dan penerjemahan. Meskipun secara bawaan *TextBlob* lebih optimal untuk bahasa *Inggris*, *library* ini juga dapat digunakan untuk bahasa Indonesia dengan beberapa penyesuaian[16].

2.4 Preprocessing

Preprocessing merupakan rangkaian proses yang bertujuan untuk mengolah dan membersihkan data mentah agar lebih terorganisir serta mudah dipahami. Langkah ini berfokus pada peningkatan kualitas dan akurasi analisis dengan menghilangkan gangguan, inkonsistensi, serta informasi yang tidak penting. Proses ini melibatkan beberapa tahapan yaitu, yaitu *cleansing*, *casefolding*, *tokenizing*, *stopword removal*, dan *stemming*[17]. Tahapan-tahapan tersebut dijelaskan lebih lanjut sebagai berikut:

a. *Cleansing* adalah proses menghapus semua karakter dalam *tweet* yang bukan huruf *alfabet* menggunakan *library re (regular expression)*. Tahapan ini bertujuan untuk mengurangi simbol atau karakter yang tidak relevan dan tidak memiliki makna dalam analisis sentimen[18].
b. *Case Folding* merupakan proses mengubah semua huruf dalam teks menjadi huruf kecil. Selain itu, langkah ini juga melibatkan penghapusan kata-kata yang tidak relevan untuk mengurangi gangguan atau *noise* yang dapat memengaruhi hasil analisis[19].
c. *Tokenizing* adalah metode untuk membagi teks menjadi unit-unit kecil yang disebut token. Token ini dapat berupa kata, frasa, atau kalimat dan berfungsi sebagai dasar dalam proses analisis selanjutnya[20].
d. *Stopword Removal* merupakan proses penghapusan kata-kata umum yang sering muncul namun tidak memberikan kontribusi berarti terhadap isi dokumen. Proses ini dilakukan menggunakan *library* NLTK. Contohnya adalah kata-kata seperti "yang", "untuk", "di", "dan", "atau"[21].
e. *Stemming* adalah proses mengubah kata berimbuhan atau bentuk turunannya menjadi kata dasar dengan menghapus awalan atau akhiran, menggunakan *library Sastrawi* atau NLTK[22].

2.5 Pembobotan Kata

Term Frequency-Inverse Document Frequency (TF-IDF) adalah metode yang digunakan untuk mengukur seberapa efektif suatu istilah merepresentasikan isi sebuah dokumen dengan memberikan bobot pada setiap kata. Teknik ini menilai pentingnya sebuah kata berdasarkan frekuensi kemunculannya dalam dokumen tersebut serta seberapa jarang kata tersebut muncul di seluruh dokumen dalam kumpulan data [23]. Rumus TF-IDF dapat dilihat pada persamaan berikut.

$$\mathrm{TF-IDF\ (t,d)} = \frac{f(t,d)}{max_{t'}(f(t',d))} \times \log\left(\frac{N}{1+df(t)}\right) \quad (1)$$

Term Frequency (TF) adalah ukuran yang menunjukkan seberapa sering sebuah term t muncul dalam dokumen d, yang dihitung sebagai $f(t,d)$ Jumlah kemunculan *term* t dalam dokumen d, $max_{t'}(f(t',d))$ Frekuensi tertinggi dari semua *term* t dalam dokumen d. *Inverse Document Frequency* (IDF) mengukur pentingnya *term* t dalam seluruh koleksi dokumen, N merepresentasikan jumlah total dokumen dalam koleksi, $df(t)$ menunjukkan banyaknya

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2403−2412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6797

dokumen yang mengandung term t, Penambahan 1 dalam penyebut $1 + df(t)$ digunakan untuk menghindari pembagian dengan nol.

2.6 Random Forest

Random Forest adalah metode machine learning yang dirancang untuk mengklasifikasikan dataset berukuran besar. Teknik ini merupakan pengembangan dari metode *Classification* and *Regression Tree* (CART). Dalam implementasinya, *Random Forest* menggunakan pendekatan *bootstrap aggregating (bagging)* dan secara acak memilih fitur saat membangun setiap pohon keputusan[24]. Hasil klasifikasi akhir diperoleh melalui *voting mayoritas* dari semua pohon yang dihasilkan. Dalam penelitian ini, *Random Forest* digunakan untuk memprediksi sentimen teks, apakah termasuk kategori positif, negatif, atau netral. Proses *decision tree* diawali dengan perhitungan nilai *gini impurity* dan rata-rata *gini impurity*. Perhitungan ini digunakan untuk menentukan seberapa informatif suatu *node* atribut serta menghitung nilai *information gain.* Rumus untuk menghitung *gini impurity* adalah sebagai berikut dapat dilihat pada persamaan 2,3 dan 4.

$$Gini \ = 1 - \sum_{i}^{n} = 1 \ (pi)^2 \tag{2}$$

dokumen n merujuk pada jumlah istilah dalam dokumen, sementara Pi adalah *probabilitas* kemunculan sebuah *term*, dihitung berdasarkan frekuensi kemunculannya dibandingkan dengan total *term* dalam dokumen.

$$Average \ gini \ \ impurity = \frac{n}{i} \times gini \tag{3}$$

i merujuk pada jumlah dokumen dalam dataset, sementara n adalah jumlah term dalam suatu kelas, yang bisa berupa kategori seperti positif, negatif, atau netral.

$$Information \ Gain = Gini \ impurity - average \ gini \ impurity \tag{4}$$

2.7 K-Nearest Neighbors (KNN)

K-Nearest Neighbor (KNN) adalah sebuah algoritma yang digunakan untuk mengklasifikasikan objek dengan cara membandingkannya terhadap data pelatihan yang memiliki kedekatan terbesar. Meskipun algoritma ini relatif sederhana untuk diterapkan, KNN terbukti cukup efektif dalam melakukan pengelompokan teks. Namun, algoritma ini memiliki beberapa kelemahan, seperti ketergantungan tinggi pada nilai parameter K yang dipilih, serta kebutuhan akan sumber daya komputasi yang besar, terutama ketika menangani dataset berukuran besar[20]. Berikut ini adalah rumus yang digunakan dalam algoritma KNN. Dapat dilihat pada persamaan 5.

$$d(x_1.x_j) = \sqrt{\sum_{r=1}^{n}((\ a_r(x_i - (a_r(x_i))))^2} \tag{5}$$

$d(x_1.x_j)$ adalah jarak *Euclidean* antara dua titik x_1.dan x_j, di mana $x_{i,r}$ dan $x_{j,r}$ adalah nilai komponen ke-r pada titik x_1. dan x_j, sementara nnn adalah jumlah dimensi data. Rumus ini mengukur kedekatan dua *vektor* dengan menghitung perbedaan kuadrat antara komponen-komponennya, dijumlahkan, dan diakarkan.

2.8 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma *machine learning* yang memanfaatkan *hyperplane* untuk memisahkan area kelas pada data. *Hyperplane* berfungsi sebagai pemisah antara kelas-kelas tersebut. SVM memprediksi kelas sebuah data dengan memberi label berdasarkan area kelas di mana data tersebut berada[25]. Rumus perhitungannya adalah sebagai berikut disajikan pada persamaan 6,7 dan 8.
Pasangan data dan kelas :

$$\{(x_i, y_i)\}_{i=1}^{N} \tag{6}$$

Menghitung nilai w dan b :

$$w = \ \sum_{i=1}^{N} a_i, y_i, x_i \quad b = \ -\frac{1}{2} \ (w.x^+ + w.x^-) \tag{7}$$

Fungsi Keputusan klasifikasi *sign* (f(x)) :

$$f(x) = w.x + b \quad \ atau \quad \ f(x) = \sum_{i=1}^{m} a_1, y_i K(x, x_i) + b \tag{8}$$

N adalah jumlah data dalam dataset, n adalah jumlah fitur, dan m adalah jumlah *support vector* dengan $a_i > 0$, yang menentukan margin optimal. K(x,xi) adalah fungsi kernel yang menghitung kedekatan titik data dalam ruang fitur lebih tinggi, memungkinkan SVM bekerja efektif tanpa menghitung koordinat eksplisit.

2.9 Evaluasi

Pengujian dalam penelitian ini bertujuan untuk menilai performa algoritma yang digunakan. Evaluasi dilakukan menggunakan metode *Confusion Matrix*, yang berfungsi untuk menganalisis kinerja *classifier.* Dalam pengujian ini,

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2403−2412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6797

dihitung nilai akurasi, *recall*, *precision*, dan *f1-score* yang hasilnya akan ditampilkan dalam bentuk persentase[26]. Rumus *Confusion Matrix* untuk menghitung akurasi disajikan pada persamaan 9,10,11 dan 12.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (9)$$

$$precision = \frac{TP}{TP+FP} \quad (10)$$

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

$$f1 - score = \frac{2xPresisixRecall}{PresisixRecall} \quad (12)$$

True Positive (TP) adalah data yang tepat diklasifikasikan sebagai positif, *False Positive* (FP) adalah data yang salah diklasifikasikan sebagai positif, *True Negative* (TN) adalah data yang tepat diklasifikasikan sebagai negatif, dan *False Negative* (FN) adalah data yang salah diklasifikasikan sebagai negatif.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

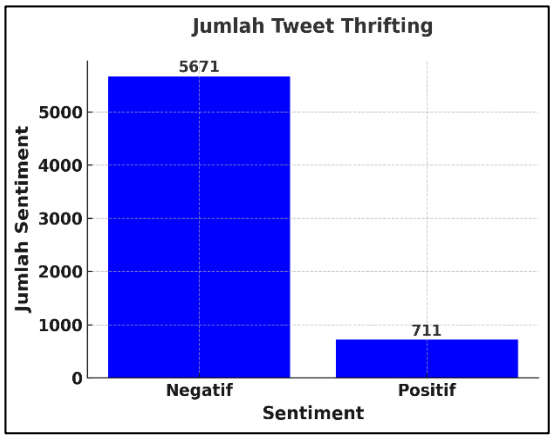
Pada penelitian ini, dilakukan crawling atau pengumpulan data dengan menggunakan *query* '*Thrift*' dan '*Thrifting*' untuk mendapatkan *tweet* dari pengguna *Twitter*. Proses *crawling* dilakukan menggunakan *library Tweet Harvest* dan akses *Auto Token* API *Twitter* untuk mengumpulkan data. Data yang berhasil dikumpulkan pada proses ini berjumlah 6.390 *tweet* berbahasa indonesia. Data dikumpulkan dengan rentang waktu dari 2 Agustus 2024 hingga 4 September 2024. Tabel 1 menampilkan hasil pengumpulan data.

Tabel 1. Hasil Pengumpulan Data

Username	*Tweet*
z4dityaa_	@anyarlfes Beli baju Thrifting tuh murah sih, tapi kadang dapatnya barang udah BULUK, SOBEK, atau ada noda yang nggak bisa hilang. Mending nabung dikit buat beli yang baru deh... 😒
deyaak76	"GUE habis Thrifting seru banget!! Bisa nemu barang-barang unik, vintage, dan berkualitas dengan harga terjangkau. Plus, lebih ramah lingkungan karena kita bantu ngurangin limbah fashion! 🌿💚#$%

3.2 Labelling Data

Setelah melalui tahap pengumpulan data, langkah berikutnya adalah tahap l*abelling*. Pada tahap ini, setiap data yang telah dibersihkan dan dipersiapkan akan diberi label sesuai dengan kategori sentimen positif dan negatif dalam analisis sentimen. Hasil *Labelling* dapat dilihat pada Gambar 2.



Gambar 2. Hasil *Labelling*

Gambar 2 memperlihatkan hasil pelabelan dalam penelitian ini, di mana sebanyak 5.671 data, atau sekitar 88,9%, dikategorikan sebagai *tweet* negatif, sedangkan 711 data, atau 11,1%, dikategorikan sebagai *tweet* positif. Total keseluruhan data yang dianalisis berjumlah 6.390 *tweet.*

3.3 Preprocessing

Setelah tahap pelabelan, langkah selanjutnya adalah tahap *preprocessing.* Pada tahap ini, data yang telah terkumpul sebanyak 6.390 *tweet* akan diproses untuk membersihkan *noise*, sehingga data menjadi lebih terstruktur dan siap untuk diproses pada tahap selanjutnya. Proses *preprocessing* ini sangat penting agar analisis yang dilakukan nantinya

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2403−2412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6797

menghasilkan data yang lebih akurat dan relevan. Dengan demikian, data yang digunakan untuk langkah-langkah berikutnya akan lebih bersih dan terorganisir dengan baik. Proses *preprocessing* ini dapat dilihat pada Tabel 2.

Tabel 2. Hasil *Preprocesing* Data

Tahapan	*Tweet*
Tweet	"GUE habis Thrifting seru banget!! Bisa nemu barang-barang unik, vintage, dan berkualitas dengan harga terjangkau. Plus, lebih ramah lingkungan karena kita bantu ngurangin limbah fashion! 🌿💚#$%
Cleaning	gue habis Thrifting seru banget Bisa nemu barang barang unik vintage dan berkualitas dengan harga terjangkau Plus lebih ramah lingkungan karena kita bantu ngurangin limbah fashion
Case Folding	gue habis thrifting seru banget bisa nemu barang barang unik vintage dan berkualitas dengan harga terjangkau plus lebih ramah lingkungan karena kita bantu ngurangin limbah fashion
Tokenizing	['gue', 'habis', 'thrifting', 'seru', 'banget', 'bisa', 'nemu', 'barang', 'barang', 'unik', 'vintage', 'dan', 'berkualitas', 'dengan', 'harga', 'terjangkau', 'plus', 'lebih', 'ramah', 'lingkungan', 'karena', 'kita', 'bantu', 'ngurangin', 'limbah', 'fashion']
Filtering	['gue', 'habis', 'thrifting', 'seru', 'banget', 'nemu', 'barang', 'barang', 'unik', 'vintage', 'berkualitas', 'harga', 'terjangkau', 'plus', 'bantu', 'ngurangin', 'limbah', 'fashion']
Stemming	gue habis thrifting seru banget nemu barang barang unik vintage berkualitas harga terjangkau plus ramah lingkungan bantu ngurangin limbah fashion

Berdasarkan Tabel 2 yang menjukan Proses *preprocessing* data pada penelitian ini dilakukan melalui beberapa tahap untuk mengolah teks mentah menjadi data yang bersih dan siap digunakan. Tahap pertama adalah *cleaning*, di mana teks dibersihkan dari karakter khusus seperti tanda baca, emoji, dan simbol lainnya sehingga hanya menyisakan huruf dan spasi. Selanjutnya, pada tahap *case folding*, semua huruf dalam teks diubah menjadi huruf kecil *(lowercase)* untuk memastikan konsistensi dalam analisis. Proses dilanjutkan dengan *tokenizing*, yaitu memecah teks menjadi unit-unit kecil berupa kata-kata (token), sehingga teks dapat direpresentasikan sebagai daftar kata. Setelah itu, dilakukan *filtering* untuk menghapus kata-kata yang tidak relevan atau tidak membawa makna penting, seperti kata sambung dan kata umum lainnya *(stopwords)*, guna memfokuskan analisis pada konten utama. Terakhir, pada tahap *stemming*, kata-kata diubah ke bentuk dasarnya *(root word)* agar variasi bentuk kata, seperti "bantu" dan "membantu," disederhanakan menjadi satu bentuk yang sama, yaitu "bantu." Hasil akhir dari seluruh proses ini adalah teks yang lebih bersih, terstruktur, dan siap untuk tahap analisis selanjutnya.

3.4 Pengujian dan Evaluasi

Data yang digunakan dalam penelitian ini diperoleh dari opini atau komentar di platform media sosial X dengan kata kunci "Thrifting" dalam Bahasa Indonesia, yang berjumlah 6.390 data. Sebelum pemodelan dilakukan, data dibagi terlebih dahulu dengan rasio 80:20, di mana 80% digunakan untuk data pelatihan dan 20% sisanya untuk data pengujian. Pembagian ini dilakukan untuk memastikan bahwa model pembelajaran mesin dapat dilatih secara optimal menggunakan sebagian besar data, kemudian diuji pada sisa data untuk mengukur kinerja atau evaluasi dan akurasi prediksinya. Pengujian dalam penelitian ini dilakukan dengan menerapkan klasifikasi menggunakan algoritma *Random Forest*, SVM, dan KNN. Berikut merupakan hasil rincian mengenai pengujian algoritma tersebut pada Tabel 3.

Tabel 3. Hasil Perbandingan Model

Model	*Class*	*Accuracy*	*Recall*	*F1-score*	*Precision*
Random Forest	Negatif	0.95	0.95	1.00	0.97
	Positif		0.99	0.56	0.72
SVM	Negatif	0.93	0.93	1.00	0.96
	Positif		0.92	0.44	0.59
KNN	Negatif	0.89	0.91	0.98	0.94
	Positif		0.54	0.19	0.28

Berdasarkan Tabel 3 yang menunjukan hasil pengujian menunjukkan bahwa *Random Forest* memiliki performa terbaik untuk kelas negatif dengan *Accuracy* 0.95 dan *Recall* 1.00, artinya mampu mengenali semua *instance* negatif dengan sangat baik. Namun, untuk kelas positif, meskipun akurasi keseluruhan tinggi (0.99), *Recall* hanya 0.56, menandakan bahwa hanya 56% *instance* positif yang dikenali dengan benar, dengan *Precision* sebesar 0.72. SVM juga unggul pada kelas negatif dengan *Accuracy* 0.93 dan *Recall* 1.00, tetapi performanya menurun pada kelas positif dengan *Recall* hanya 0.44 dan *Precision* sebesar 0.59, menunjukkan bahwa masih banyak *instance* positif yang tidak terdeteksi. KNN menunjukkan performa baik pada kelas negatif dengan *Accuracy* 0.91 dan *Recall* 0.98, tetapi sangat lemah pada kelas positif, dengan *Recall* hanya 0.54, *Precision* 0.28, dan *F1-score* 0.19, menandakan banyak kesalahan prediksi. Secara keseluruhan, *Random Forest* unggul dalam mengenali kelas negatif, meskipun perlu perbaikan untuk kelas positif, sedangkan SVM cukup baik untuk kelas negatif namun kurang optimal pada kelas

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2403−2412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6797

positif, dan KNN lemah dalam mengenali kelas positif meskipun konsisten pada kelas negatif. Berdasarkan evaluasi menggunakan *matriks* kebingungan, *Random Forest* adalah algoritma terbaik dalam penelitian ini.

Kemudian untuk menentukan algoritma yang paling unggul, peneliti membandingkan kinerja ketiga model tersebut dengan menggunakan matriks kebingungan (*confusion matrix*). Berikut merupakan hasil evaluasi kinerja dari algoritma *Random Forest*, *Support Vector Machine*, dan KNN dapat dilihat pada Tabel 4.

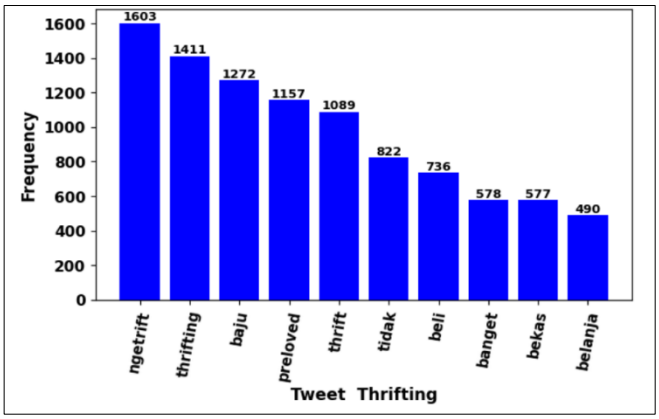
Tabel 4. Hasil *confusion matrix* Model

Model	*Prediction Class*	*Actual Class*	
		Negatif	Positif
Random Forest	Negatif	1136	1
	Positif	61	79
SVM	Negatif	1132	5
	Positif	79	61
KNN	Negatif	1114	23
	Positif	113	27

Berdasarkan Tabel 4 yang menunjukan hasil *Confius matrix* perbandingan performa dari tiga model, *Random Forest*, SVM, dan KNN, dalam mengklasifikasikan kelas negatif dan positif. Model *Random Forest* memiliki kinerja yang sangat baik dalam mengenali kelas negatif, dengan 1136 prediksi benar dan hanya 61 kesalahan prediksi positif (*False Positive*). Untuk kelas positif, model ini juga menunjukkan performa baik dengan 79 prediksi benar dan hanya 1 kesalahan (*False Negative*). SVM juga cukup baik dalam mengenali kelas negatif, dengan 1132 prediksi benar dan 79 kesalahan positif, namun kinerjanya lebih rendah dibandingkan *Random Forest* pada kelas positif, dengan 61 prediksi benar dan 5 kesalahan. Kemudian, KNN memiliki performa yang paling rendah, terutama dalam mengidentifikasi kelas positif, di mana model ini hanya membuat 27 prediksi benar dan melakukan 23 kesalahan. Selain itu, KNN juga memiliki tingkat kesalahan yang lebih tinggi untuk kelas negatif, dengan 113 kesalahan prediksi positif. Secara keseluruhan, *Random Forest* memiliki performa terbaik dalam mengklasifikasikan kedua kelas, diikuti oleh *Support Vector Machine*, sementara KNN menunjukkan performa yang lebih rendah di kedua kelas, terutama pada kelas positif.

3.5 Visualisasi

Visualisasi dalam penelitian ini digunakan dengan mengandalkan frekuensi kata dan *wordcloud* untuk menggambarkan distribusi dan kecenderungan topik yang terkait dengan *thrifting*. Dengan menerapkan metode ini, kata-kata yang sering muncul dalam *tweet* dapat divisualisasikan dengan lebih jelas, sehingga mempermudah dalam mengidentifikasi kata-kata kunci yang berkaitan dengan tren *thrifting*. Hasil Frekuensi Kata tertinggi pada Gambar 3.

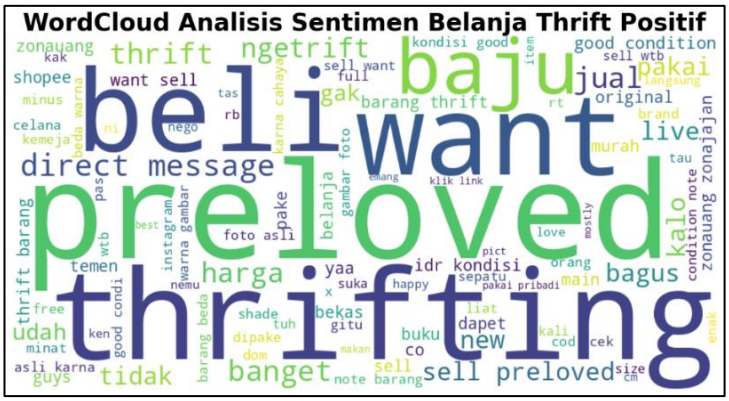


Gambar 3. Frekuensi Kata tertinggi

Gambar 3 tersebut menunjukkan grafik yang menampilkan 10 kata dengan frekuensi tertinggi dalam tweet terkait *thrifting*. Kata "ngethrift" menempati posisi teratas dengan frekuensi 1603, diikuti oleh "thrifting" sebanyak 1411 kali, dan "baju" sebanyak 1272 kali. Kata-kata lain seperti "*preloved*", *thrift"*, "tidak", "beli" ,"banget", "bekas" dan "belanja" yang berhubungan dengan aktivitas belanja juga memiliki frekuensi yang signifikan. Grafik ini memberikan gambaran tentang kata-kata yang paling sering muncul dalam *tweet* tentang *thrifting*, menunjukkan bahwa topik utama berkaitan dengan pakaian bekas dan aktivitas membeli.

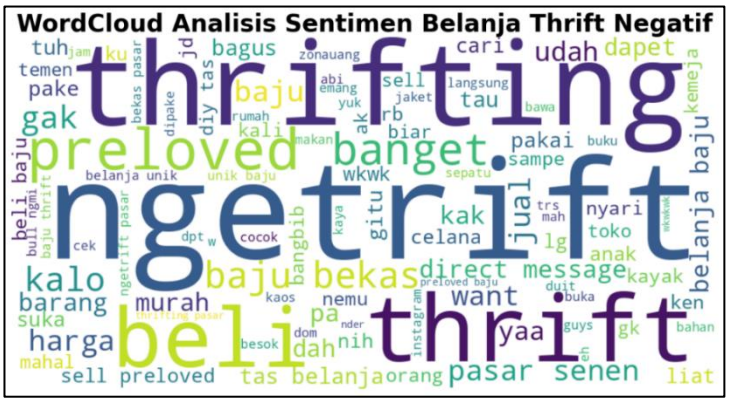
Pada tahap berikutnya, peneliti memanfaatkan Wordcloud untuk mengidentifikasi informasi penting yang terkandung dalam data tweet. Wordcloud digunakan untuk menggambarkan kumpulan kata-kata penting yang paling sering muncul dalam sebuah dataset. Semakin sering suatu kata muncul, semakin besar ukuran *font* kata tersebut. Berikut ini merupakan hasil *wordcloud* sentiment Positif dan Negatif.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2403−2412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6797



Gambar 4. *Wordcloud* Sentimen Positif

Gambar 4 *wordcloud* tersebut menggambarkan hasil analisis sentimen positif terkait dengan aktivitas thrifting. Kata-kata yang paling menonjol seperti "*thrifting*", "*preloved*", dan "beli" mencerminkan tema utama diskusi, yakni mengenai pembelian barang bekas yang berkualitas dengan harga yang terjangkau. Kata "murah", "harga", dan "bagus" menekankan aspek positif dari barang yang dijual, yaitu kondisi yang baik dan harga yang murah. Istilah "*good condition*" juga muncul, menggaris bawahi bahwa meskipun barang-barang ini *preloved* atau bekas, banyak yang masih dalam kondisi yang sangat baik. Kata "*new*" juga muncul, mungkin merujuk pada barang preloved yang kondisinya masih seperti baru. Secara keseluruhan, *wordcloud* ini mencerminkan sikap positif terhadap kegiatan *thrifting*, di mana orang-orang menghargai nilai ekonomis, kualitas barang, serta kesempatan untuk mendapatkan barang berkondisi baik dengan harga murah.



Gambar 5. *Wordcloud* Sentimen Negatif

Gambar 5 *wordcloud* tersebut menunjukkan sentimen negatif terkait *thrifting*. Kata yang paling menonjol adalah "*thrifting*", yang merupakan topik utama pembahasan. Kata-kata seperti "*ngethrift*", "bekas", dan "baju" menunjukkan bahwa diskusi banyak berfokus pada belanja barang bekas, terutama pakaian. Kata "harga" dan "mahal" sering muncul, menandakan adanya ketidakpuasan terkait harga barang-barang bekas yang dianggap terlalu mahal. Selain itu, kata-kata lain seperti "bgt" (banget), "ga" (tidak), dan "yang" juga sering digunakan, mengindikasikan adanya percakapan yang berhubungan dengan pengalaman kurang memuaskan dalam aktivitas *thrifting*. Ada kemungkinan bahwa pengguna membicarakan barang yang seharusnya lebih murah tetapi ditemukan dengan harga yang tidak sesuai harapan. Secara keseluruhan, *wordcloud* ini menggambarkan diskusi dengan sentimen negatif mengenai pengalaman *thrifting*, terutama terkait dengan harga yang dianggap mahal serta kekecewaan atas kualitas barang yang dibeli.

4. KESIMPULAN

Penelitian ini membandingkan performa tiga algoritma klasifikasi, yaitu *Random Forest*, *Support Vector Machine* (SVM), dan *K-Nearest Neighbors* (KNN), terhadap komentar publik yang diambil dari media sosial X dengan pembagian data 80:20 untuk training dan testing. Hasil pengujian menunjukkan bahwa algoritma *Random Forest* memiliki performa terbaik dibandingkan SVM dan KNN, dengan akurasi 95%, *precision* 97%, *recall* 78%, dan *F1-score* 85%. Sementara itu, algoritma SVM menunjukkan akurasi 93%, *precision* 93%, *recall* 72%, dan *F1-score* 78%. Di sisi lain, KNN memiliki performa terendah dengan akurasi 89%, *precision* 72%, *recall* 59%, dan *F1-score* 61%. Penelitian ini menemukan hasil yang berbeda dari studi sebelumnya yang hanya menggunakan KNN untuk analisis sentimen masyarakat terhadap *thrifting* di *Twitter*. Dalam penelitian ini, penggunaan data yang lebih banyak serta perbandingan tiga algoritma memberikan analisis yang lebih komprehensif. Namun, terdapat keterbatasan, seperti

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2403−2412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6797

pembagian data 80:20 yang rentan terhadap *overfitting* pada dataset kecil. Pembagian yang lebih seimbang, seperti 60:40 atau 70:30, dapat meningkatkan generalisasi model. Tantangan lain adalah *imbalanced class*, yang dapat membuat model lebih fokus pada kelas *mayoritas*, serta perbedaan konteks komentar di media sosial X yang memengaruhi analisis. Penelitian selanjutnya disarankan untuk melakukan parameter tuning dan mempertimbangkan algoritma canggih seperti LSTM atau BERT yang lebih efektif dalam memahami teks kompleks. Meski *Random Forest* menunjukkan performa terbaik, penelitian ini masih memiliki ruang untuk pengembangan guna meningkatkan akurasi dan generalisasi model.

REFERENCES

[1] A. Azzahra and D. Gustian, "Penerapan Algoritma Apriori pada Usaha Thrifting," *Jurnal Komisi (Komputer Dan Sistem Informasi)*, vol. 1, no. 2, pp. 68–75, 2024, [Online]. Available: https://komisijournal.indiepress.id/index.php/komisi/article/view/10.

[2] S. Wulandari and F. N. Hasan, "Analisis Sentimen Masyarakat Indonesia Terhadap Pengalaman Belanja Thrifting Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes," *Jurnal Media Informatika Budidarma*, vol. 8, no. 2, p. 768, Apr. 2024, doi: 10.30865/mib.v8i2.7520.

[3] N. Qalbi and Hartini, "Pengaruh Persepsi Harga, Persepsi Kualitas dan Keragaman Produk terhadap Minat Beli Pakaian Bekas (Trifting) di Daerah Sumbawa," *Journal of Nusantara Economic Science (JNES)*, vol. 1, no. 4, pp. 1–7, 2023, [Online]. Available: https://nafatimahpustaka.org/jnes/article/view/161.

[4] D. A. Firmansyah and D. Damayanti, "Analisis Sentimen Masyarakat Terhadap Thrifting Pada Media Sosial Twitter(X.Com) Menggunakan Metode K-Nearest Neighbor," *Jurnal Jaringan Sistem Informasi Robotik (Jsr)*, vol . 8, no. 1, pp. 10 – 14, 2024, doi: 10.58486/jsr.v8i1.322.

[5] Hadisaputra, "Tinjauan Pendapatan Penjual Pakaian Bekas (Trifting) di Pasar Wameo Kelurahan Wameo Kecamatan Batupoaro Kota Baubau," *Jurnal Ilmiah Universitas Muhammadiyah Buton*, volume 10, no 3, pp. 465–475, 2021, doi.org/10.35326/pencerah.v10i3.5518.

[6] A. Amelia, E. Permatasari, P. J. Amanda Z, F. Sinaga, and H. Antoni, "Peningkatan Daya Saing Industri terhadap Kasus Impor Pakaian Bekas (Trifting)," *Jurnal Pendidikan Tambusai*, vol. 7, no. 1, pp. 1959–1964, 2023, [Online]. Available: https://jptam.org/index.php/jptam/article/view/6061/5072

[7] P. Kegiatan, M. Ppkm, K. D. Attarik, and N. Safriadi, "Analisis Sentimen Kebijakan Pemberlakuan Terhadap Pertumbuhan Ekonomi Sektor E- Commerce Di Indonesia Menggunakan Metode K-Nearest Neighbor (KNN)," *Jurnal Informatika dan Teknik Elektro Terapan (JITET)*, vol. 12, no. 3, 2024, doi.org/10.23960/jitet.v12i3S1.5281.

[8] C. F. Hasri and D. Alita, "Penerapan Metode NaãiVe Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, pp. 145–160, 2022, doi: 10.33365/jatika.v3i2.2026.

[9] S. Analisis Kegiatan Trading dengan SVM, K. R. Dan, N. Resti Wardani, S. Saepudin, and C. Warman, "Sentimen Analisis Kegiatan Trading Pada Ap-likasi Twitter dengan Algoritma SVM, KNN Dan Random Forrest," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol 6 no 2, pp. 863-870, 2022, doi.org/10.30645/j-sakti.v6i2.

[10] I. Mahendrata, N. D. H. Sadikin, N. D. H. Sadikin, N. Marco, and Y. Ramdhani, "Sentimen Analisis Marketplace di Indonesia Menggunakan Algoritma Random Forest," *Jurnal PRODUKTIF*, vol. 7, no. 1, pp. 603–609, 2023.

[11] P. E. Shopee and S. Watmah, "Komparasi Metode K-NN, Support Vector Machine, Dan Random Forest Pada E-Commerce Shopee," *INSANtek – Jurnal Inovasi dan Sains Teknik Elektro*, vol. 2, no. 1, pp. 15–21, 2021, doi: 10.31294/instk.v2i1.419.

[12] R. Safitri, I. Ali, and N. Rahaningsih, "Analisis Sentimen Terhadap Tren Fashion Di Media Sosial Dengan Metode Support Vector Machine (Svm)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 1746–1754, 2024, doi: 10.36040/jati.v8i2.9045.

[13] M. R. R. Lillah, D. S. Maylawati, W. B. Zulfikar, W. Uriawan, and A. Wahana, "Implementasi Algoritma K-Nearest Neighbor (KNN) untuk analisis sentimen pengguna aplikasi Tokopedia," *Intellect : Indonesian Journal of Learning and Teachnological Innovation*, vol. 02, no. 02, pp. 171–184, 2023, doi: 10.57255/intellect.v2i2.296

[14] E. Hasibuan and E. A. Heriyanto, "Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier," *Jurnal Teknik dan Science JTS*, vol. 1, no. 3, 2024, doi:10.56127/jts.v1i3.434.

[15] Y. Afrillia, L. Rosnita, and D. Siska, "Analisis Sentimen Ciutan Twitter Terkait Penerapan Permendikbudristek Nomor 30 Tahun 2021 Menggunakan TextBlob dan Support Vector Machine," *G-Tech: Jurnal Teknologi Terapan*, vol. 6, no. 2, pp. 387–394, Oct. 2022, doi: 10.33379/gtech.v6i2.1778.

[16] R. Vindua and A. U. Zailani, "Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python," *JURIKOM (Jurnal Riset Komputer)*, vol. 10, no. 2, p. 479, Apr. 2023, doi: 10.30865/jurikom.v10i2.5945.

[17] H. Ali, N. Hendrastuty, C. Science, and U. T. Indonesia, "Comparison Of Naïve Bayes Classifier, Support Vector Machine, Random Forest Algorithms For Public Sentiment Analysis Of Kip-K Program On Twitter," *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 6, pp. 1701–1712, 2024, doi.org/10.52436/1.jutif.2024.5.6.4030.

[18] H. Syah and A. Witanti, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm)," *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 5, no. 1, pp. 59–67, 2022, doi: 10.47080/simika.v5i1.1411.

[19] H. Harnelia, "Analisis Sentimen Review Skincare Skintific Dengan Algoritma Support Vector Machine (Svm)," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, 2024, doi: 10.23960/jitet.v12i2.4095.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2403−2412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6797

[20] S. Syafrizal, M. Afdal, and R. Novita, "Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 10–19, 2023, doi: 10.57152/malcom.v4i1.983.

[21] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.

[22] J. Supriyanto, D. Alita, and A. R. Isnain, "Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 4, no. 1, pp. 74–80, 2023, doi: 10.33365/jatika.v4i1.2468.

[23] D. S. Ningsih and R. R. Suryono, "Comparison Of Naïve Bayes And Information Gain Algorithms In Cyberbullying Sentiment Analysis On Twitter," *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 4, pp. 1085–1091, 2024, doi: 10.52436/1.jutif.2024.5.4.1908.

[24] T. Fadiyah Basar, D. E. Ratnawati, and I. Arwani, "Analisis Sentimen Pengguna Twitter terhadap Pembayaran Cashless menggunakan Shopeepay dengan Algoritma Random Forest," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 3, pp. 1426-1433, 2022, [Online]. https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/10830.

[25] I. Yunanto and S. Yulianto, "Twitter Sentiment Analysis Pedulilindungi Application Using Naïve Bayes And Support Vector Machine," *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 4, pp. 807–814, Aug. 2022, doi: 10.20884/1.jutif.2022.3.4.292.

[26] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS : Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/tekno.v10i2.419.