



Penerapan Algoritma *Multinomial Naïve Bayes* dengan Penyeimbangan Data *SMOTE* pada KI Asifikasi Sentimen Pengguna Shopee terhadap Produk *Facial Wash Kahf*

Farendika Rezzi

Teknik Informatika, Universitas Muhammadiyah Jember, Jl Gumuk Kerang, Karangrejo,
Kec. Sumbersari, Kabupaten Jember, Jawa Timur Indonesia, 68124

Korespondensi penulis: ulyaanisatur@unmuhjember.ac.id

Abstract. *The rapid growth of e-commerce platforms has significantly transformed the way consumers share and access product feedback. One of the widely used platforms in Indonesia is Shopee, where customers actively provide reviews of various products, including local skincare brands such as Kahf facial wash. Customer reviews on e-commerce platforms contain valuable information that can be analyzed to understand consumer opinions and preferences. Sentiment analysis, as a branch of natural language processing, enables the classification of textual data into categories such as positive, negative, or neutral. This study aims to classify Shopee user sentiments regarding Kahf facial wash products by implementing the Multinomial Naïve Bayes algorithm, a well-known probabilistic classifier suitable for text categorization. The research methodology consisted of several preprocessing stages, including data cleansing, case folding, tokenizing, stopword removal, and stemming, to prepare raw review texts for further analysis. For feature representation, the Term Frequency–Inverse Document Frequency (TF-IDF) method was applied to capture the importance of words across documents. To evaluate the classification performance, K-Fold cross-validation was employed with K values of 4, 5, 6, and 10 to ensure model reliability and robustness. Considering the issue of imbalanced datasets in user-generated reviews, the Synthetic Minority Over-sampling Technique (SMOTE) was utilized to balance the distribution of sentiment classes. Based on the confusion matrix, the Multinomial Naïve Bayes algorithm demonstrated effective performance in classifying sentiments, achieving satisfactory levels of accuracy, precision, and recall across different folds. These results indicate that the algorithm is capable of handling sentiment analysis tasks for local product reviews effectively. The findings of this study are expected to provide meaningful insights for businesses in understanding consumer perceptions, thereby supporting decision-making processes in product development, marketing strategies, and customer engagement for local brands.*

Keywords: *Multinomial Naïve Bayes, Sentiment Analysis, Shopee, SMOTE, TF-IDF*

Abstrak. Pertumbuhan pesat platform e-commerce telah secara signifikan mengubah cara konsumen berbagi dan mengakses ulasan produk. Salah satu platform yang banyak digunakan di Indonesia adalah Shopee, di mana pelanggan aktif memberikan ulasan terhadap berbagai produk, termasuk merek perawatan kulit lokal seperti facial wash Kahf. Ulasan pelanggan di platform e-commerce mengandung informasi berharga yang dapat dianalisis untuk memahami opini serta preferensi konsumen. Analisis sentimen, sebagai cabang dari pemrosesan bahasa alami, memungkinkan pengklasifikasian data teks ke dalam kategori seperti positif, negatif, atau netral. Penelitian ini bertujuan untuk mengklasifikasikan sentimen pengguna Shopee terhadap produk facial wash Kahf dengan menerapkan algoritma Multinomial Naïve Bayes, yang dikenal sebagai salah satu pengklasifikasi probabilistik yang sesuai untuk kategorisasi teks. Metodologi penelitian mencakup beberapa tahapan praproses, antara lain pembersihan data, case folding, tokenizing, stopword removal, dan stemming untuk mempersiapkan teks ulasan mentah sebelum dianalisis lebih lanjut. Representasi fitur dilakukan dengan metode Term Frequency–Inverse Document Frequency (TF-IDF) untuk menangkap tingkat kepentingan kata pada setiap dokumen. Evaluasi performa klasifikasi menggunakan teknik K-Fold cross-validation dengan nilai K 4, 5, 6, dan 10 guna memastikan reliabilitas serta konsistensi model. Mengingat adanya ketidakseimbangan data pada ulasan pengguna, digunakan metode Synthetic Minority Over-sampling Technique (SMOTE) untuk menyeimbangkan distribusi kelas sentimen. Berdasarkan hasil confusion matrix, algoritma Multinomial Naïve Bayes menunjukkan kinerja yang efektif dalam mengklasifikasikan sentimen, dengan capaian akurasi, presisi, dan recall yang memuaskan pada berbagai nilai fold. Temuan penelitian ini diharapkan dapat memberikan wawasan yang bermakna bagi pelaku bisnis dalam memahami persepsi konsumen, sehingga dapat mendukung pengambilan keputusan pada pengembangan produk, strategi pemasaran, serta peningkatan keterlibatan pelanggan terhadap merek lokal.

Kata kunci: Analisis Sentimen, *Multinomial Naïve Bayes*, Shopee, *SMOTE*, *TF-IDF*

1. LATAR BELAKANG

Kemajuan di bidang teknologi informasi telah memacu perkembangan pesat sektor e-commerce di Indonesia. Salah satunya Shopee, yang memudahkan konsumen dalam melakukan pembelian serta memberikan ulasan terhadap produk. Ulasan konsumen, terutama yang berbentuk teks, merupakan sumber data berharga untuk memahami persepsi, pengalaman, dan tingkat kepuasan pelanggan terhadap suatu produk. Salah satu produk lokal yang cukup populer di pasaran adalah *facial wash* merek Kahf. Analisis terhadap ulasan produk ini dapat membantu produsen dalam meningkatkan mutu produk serta menyusun strategi pemasaran yang lebih tepat sasaran.

Analisis sentimen merupakan bagian dari *text mining* yang bertujuan mengidentifikasi *polaritas* opini menjadi kategori positif, negatif, atau netral. Salah satu metode populer untuk klasifikasi teks adalah *Multinomial Naïve Bayes*, yang efektif digunakan pada data berbasis frekuensi kata. Namun, ketidakseimbangan jumlah data pada setiap kelas dapat menurunkan kinerja model karena bias terhadap kelas mayoritas. Untuk mengatasi permasalahan tersebut, digunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) yang menghasilkan data sintetis pada kelas minoritas, sehingga distribusi kelas menjadi seimbang pada tahap pelatihan.

Penelitian terkait analisis sentimen pada ulasan *e-commerce* telah dilakukan sebelumnya, namun sebagian besar belum memanfaatkan kombinasi *Multinomial Naïve Bayes* dengan *SMOTE* pada domain produk perawatan wajah merek lokal. Oleh karena itu, penelitian ini memiliki kebaruan pada penerapan kombinasi metode tersebut untuk mengklasifikasikan sentimen pengguna Shopee terhadap produk *facial wash* Kahf. Dengan pendekatan ini, diharapkan hasil klasifikasi menjadi lebih akurat dan dapat memberikan masukan yang bermanfaat bagi produsen dalam memahami persepsi konsumen.

2. KAJIAN TEORITIS

Analisis Sentimen

Analisis sentimen merupakan salah satu bidang dalam *text mining* yang berfokus pada pengolahan data opini atau ulasan untuk menentukan kecenderungan sentimen, seperti positif, negatif, maupun netral (Sari dkk., 2020). Tahapan umum dalam analisis sentimen mencakup proses pengumpulan data, pra-pemrosesan teks, ekstraksi fitur, hingga penerapan metode klasifikasi (Putri dkk., 2021).

Pra-pemrosesan

Pra-pemrosesan teks bertujuan mengubah data mentah menjadi format yang siap dianalisis dengan menghilangkan elemen yang tidak relevan dan menyamakan bentuk teks. Langkah-langkah yang dilakukan meliputi *cleansing*, *case folding*, *tokenizing*, *stopword removal*, hingga *stemmed* (Prakoso dkk, 2022).

Ekstraksi Fitur

Ekstraksi fitur adalah proses mengonversi teks menjadi representasi angka yang dapat dianalisis oleh model pembelajaran mesin. Salah satu metode yang sering digunakan adalah *Term Frequency–Inverse Document Frequency (TF-IDF)*, yaitu teknik yang memberikan bobot pada setiap kata berdasarkan seberapa sering kata tersebut muncul pada suatu dokumen dibandingkan kemunculannya di seluruh dokumen yang ada (Prakoso dkk., 2022).

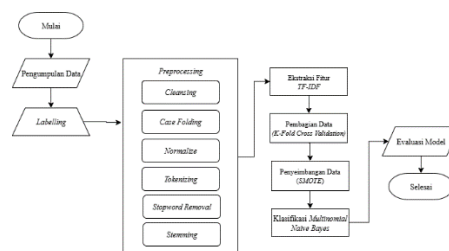
Multinomial Naïve Bayes

Multinomial Naïve Bayes adalah algoritma klasifikasi probabilistik yang banyak dimanfaatkan dalam pengolahan teks karena kemampuannya menangani data frekuensi kata dan asumsi independensi antar fitur (Setiawan & Rachmawati, 2022). Algoritma ini efektif untuk dataset berukuran sedang dan komputasi cepat, namun performanya dapat menurun jika terjadi ketidakseimbangan kelas.

Synthetic Minority Over-sampling Technique (SMOTE)

SMOTE adalah metode *oversampling* yang dikembangkan untuk mengatasi ketidakseimbangan distribusi kelas pada data. Teknik ini menghasilkan data sintesis untuk kelas minoritas dengan mempertimbangkan kedekatan antar sampel dalam ruang fitur (Charte dkk., 2023). Penerapan *SMOTE* dapat membantu mengurangi bias model terhadap kelas mayoritas dan meningkatkan kemampuan prediksi pada kelas minoritas.

3. METODE PENELITIAN



Gambar 1. Alur Tahapan Penelitian

Penelitian ini dilakukan melalui serangkaian tahapan yang mencakup pengumpulan data, pra-pemrosesan teks, ekstraksi fitur, pembagian data, penyeimbangan data, klasifikasi, dan evaluasi model. Pendekatan yang digunakan adalah kuantitatif dengan eksperimen berbasis algoritma pembelajaran mesin, yang bertujuan mengklasifikasikan sentimen ulasan pengguna Shopee terhadap produk *facial wash* Kahf.

Data penelitian berupa 600 ulasan teks yang dikumpulkan dari Shopee pada periode 2024–2025 menggunakan *web scraping* berbasis *Python*. Sebelum masuk ke tahap analisis, data melewati proses pembersihan (*cleansing*), perubahan format huruf menjadi seragam (*case folding*), pemisahan kata (*tokenizing*), penghapusan kata tidak bermakna (*stopword removal*), serta pengembalian kata ke bentuk dasarnya (*stemming*) (Prakoso dkk., 2022). Representasi fitur dilakukan menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)* guna menetapkan pada kata sesuai tingkat kepentingannya (Prakoso dkk., 2022). Penyeimbangan data dilakukan dengan *Synthetic Minority Over-sampling Technique (SMOTE)* (Charte dkk., 2023).

Selanjutnya, data dibagi menggunakan metode *K-Fold cross validation* (Afdal dan Elita, 2022) dengan implementasi *Stratified K-Fold* untuk menjaga distribusi kelas positif dan negatif tetap seimbang di setiap *fold* selama proses pelatihan dan pengujian model. Model klasifikasi yang digunakan adalah *Multinomial Naïve Bayes*, yang menghitung probabilitas suatu dokumen *d* termasuk dalam kelas *c* berdasarkan probabilitas prior kelas dan probabilitas kemunculan kata di dalam kelas tersebut (Atimi & Pratama, 2022). Menurut Atimi and Pratama (2022), untuk menghitung probabilitas sebuah dokumen, yaitu:

$$P(c | \text{term dok } d) = P(c) \times P(t_1 | c) \times P(t_2 | c) \times P(t_n | c)$$

dengan,

$P(c | \text{term dok } d)$ = Probabilitas dokumen pada kelas *c*

$P(c)$ = Probabilitas *prior* pada kelas *c*

$P(t_n | c)$ = Probabilitas kata ke-*n* pada kelas *c*

t_n = Kata ke-*n* pada dokumen

Kinerja model dievaluasi dengan *confusion matrix* untuk menghitung metrik akurasi, presisi, dan *recall* (Bastoni, 2024). Persamaan masing-masing metrik adalah:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Presisi} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

Nilai TP, TN, FP, dan FN digunakan dalam perhitungan ketiga metrik tersebut. Validasi dilakukan pada setiap variasi nilai K, dan hasil dari setiap *fold* dibandingkan untuk menentukan performa terbaik model.

Hasil pengujian validitas menunjukkan bahwa Model *Multinomial Naïve Bayes* dengan pembobotan *TF-IDF* dan teknik *10-Fold cross validation* menghasilkan akurasi 98,33%, presisi 98,15%, dan recall 100%. *Confusion matrix* sebelum *SMOTE* menunjukkan model memprediksi seluruh data sebagai kelas positif, sehingga sentimen negatif tidak terdeteksi. Setelah penerapan *SMOTE*, model mampu mengenali kelas negatif dengan akurat, memperbaiki keseimbangan prediksi kedua kelas. Temuan ini mengindikasikan bahwa *SMOTE* berperan signifikan dalam meningkatkan kinerja klasifikasi pada data yang memiliki ketidakseimbangan kelas.

4. HASIL DAN PEMBAHASAN

Pengumpulan Data

Dataset yang digunakan terdiri dari 600 komentar teks dari pengguna Shopee terhadap produk *facial wash* Kahf, diambil pada periode 2024–2025 melalui *web scraping* menggunakan *Python*. Data dikumpulkan dari laman produk resmi Kahf di platform Shopee. Komentar yang memuat ulasan pengiriman dan layanan logistik dihapus pada tahap pra-pemrosesan.

Labelling Data

Tahap pelabelan dilakukan setelah proses pengumpulan data selesai, dengan tujuan mengelompokkan komentar menjadi dua kategori sentimen, yaitu positif (label “1”) dan negatif (label “0”). Dari total 600 komentar, sebanyak 527 termasuk kategori positif dan 73 kategori negatif. Pelabelan dilakukan secara manual serta divalidasi oleh pakar bahasa, Hardian Oktavianto, S.Si., M.Kom., dosen di Universitas Muhammadiyah Jember.

Hasil Preprocessing

Tahapan pra-pemrosesan meliputi enam langkah utama: *cleansing*, *case folding*, *normalize*, *tokenizing*, *stopword removal*, dan *stemmed*. Komentar yang tidak relevan dengan isi produk dihapus untuk menjaga fokus analisis.

Ekstraksi Fitur *TF-IDF*

Hasil pra-pemrosesan kemudian diubah menjadi bentuk numerik melalui penerapan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Teknik ini menetapkan bobot pada setiap kata dengan mempertimbangkan seberapa sering kata tersebut muncul dalam suatu dokumen serta kelangkaannya di seluruh kumpulan dokumen.

Penyeimbangan Data (*SMOTE*)

Distribusi data yang tidak seimbang diatasi dengan menerapkan *SMOTE* pada data latih di setiap *fold*. Teknik ini menghasilkan sampel sintesis pada kelas minoritas sehingga jumlah data positif dan negatif menjadi seimbang. Data uji tetap menggunakan data asli agar hasil evaluasi tetap valid.

Pembagian dan Pengujian Data

Pengujian dilakukan dengan *Stratified K-Fold cross validation* menggunakan nilai $K = 4, 5, 6,$ dan 10 . Setiap variasi K diuji untuk memperoleh nilai akurasi, presisi, dan *recall*. Untuk memastikan distribusi kelas positif dan negatif tetap seimbang pada setiap *fold*, digunakan metode *Stratified K-Fold cross validation*. Pendekatan ini penting untuk menjaga validitas hasil evaluasi, mengingat data yang digunakan bersifat tidak seimbang.

1. *4-Fold cross validation*

Pada tahap ini, data sebanyak 600 entri dibagi menjadi 4 bagian yang sama besar, sehingga setiap *fold* berisi 150 data. Pengujian menggunakan metode *4-fold cross validation* dilakukan sebanyak 4 kali dengan tujuan memperoleh nilai evaluasi berupa akurasi, presisi, dan *recall* pada masing-masing *fold*.

Tabel 1. Confusion matrix 4-Fold

<i>Fold</i>	TP	TN	FP	FN
Langkah 1	129	15	3	3
Langkah 2	126	18	0	6
Langkah 3	131	15	3	1
Langkah 4	127	16	3	4

Jika nilai-nilai pada *confusion matrix* mengikuti format pada Tabel diatas, maka perhitungan akurasi, presisi, dan *recall* pada langkah ketiga dilakukan menggunakan rumus yang telah ditentukan pada bab 3, sebagai berikut:

Tabel 2. 4-Fold cross validation

Fold	Akurasi	Presisi	Recall
langkah uji 1	96.00%	97.73%	97.73%
langkah uji 2	96.00%	100.00%	95.45%
langkah uji 3	97.33%	97.76%	99.24%
langkah uji 4	95.33%	97.69%	96.95%

2. 5-Fold cross validation

Pada tahap ini, data sebanyak 600 entri dibagi menjadi 5 bagian yang sama besar, sehingga setiap *fold* berisi 120 data. Pengujian menggunakan metode *5-fold cross validation* dilakukan sebanyak 5 kali dengan tujuan memperoleh nilai evaluasi berupa akurasi, presisi, dan recall pada masing-masing *fold*.

Tabel 3. Confusion matrix 5-Fold

Fold	TP	TN	FP	FN
Langkah 1	103	12	2	3
Langkah 2	103	13	1	3
Langkah 3	103	14	1	2
Langkah 4	103	13	2	2
Langkah 5	101	13	2	4

Jika nilai-nilai pada *confusion matrix* mengikuti format pada Tabel diatas, maka perhitungan akurasi, presisi, dan recall pada langkah ketiga dilakukan menggunakan rumus yang telah ditentukan pada bab 3, sebagai berikut:

Tabel 4. 5-Fold cross validation

Fold	Akurasi	Presisi	Recall
langkah uji 1	95.83%	98.10%	97.17%
langkah uji 2	96.67%	99.04%	97.17%
langkah uji 3	97.50%	99.04%	98.10%
langkah uji 4	96.67%	98.10%	98.10%
langkah uji 5	95.00%	98.06%	96.19%

3. 6-Fold cross validation

Pada tahap ini, data sebanyak 600 entri dibagi menjadi 6 bagian yang sama besar, sehingga setiap *fold* berisi 100 data. Pengujian menggunakan metode *6-fold cross validation* dilakukan sebanyak 6 kali dengan tujuan memperoleh nilai evaluasi berupa akurasi, presisi, dan recall pada masing-masing *fold*.

Tabel 5. Confusion matrix 6-Fold

Fold	TP	TN	FP	FN
Langkah 1	84	10	2	4
Langkah 2	87	11	1	1
Langkah 3	85	12	0	3
Langkah 4	88	10	2	0
Langkah 5	86	10	2	2
Langkah 6	86	12	1	1

Jika nilai-nilai pada *confusion matrix* mengikuti format pada Tabel diatas, maka perhitungan akurasi, presisi, dan recall pada langkah ketiga dilakukan menggunakan rumus yang telah ditentukan pada bab 3, sebagai berikut:

Tabel 6. 6-Fold cross validation

<i>Fold</i>	Akurasi	Presisi	Recall
langkah uji 1	94.00%	97.67%	95.45%
langkah uji 2	98.00%	98.86%	98.86%
langkah uji 3	97.00%	100.00%	96.59%
langkah uji 4	98.00%	97.78%	100.00%
langkah uji 5	96.00%	97.73%	97.73%
langkah uji 6	98.00%	98.85%	98.85%

4. 10-Fold cross validation

Pada tahap ini, data sebanyak 600 entri dibagi menjadi 10 bagian yang sama besar, sehingga setiap *fold* berisi 60 data. Pengujian menggunakan metode *10-fold cross validation* dilakukan dengan tujuan memperoleh nilai evaluasi berupa akurasi, presisi, dan recall pada masing-masing *fold*.

Tabel 7. Confusion matrix 10-Fold

<i>Fold</i>	TP	TN	FP	FN
Langkah 1	52	5	2	1
Langkah 2	51	7	0	2
Langkah 3	52	6	1	1
Langkah 4	49	7	0	4
Langkah 5	51	7	0	2
Langkah 6	53	6	1	0
Langkah 7	51	6	1	2
Langkah 8	51	7	1	1

Jika nilai-nilai pada *confusion matrix* mengikuti format pada Tabel diatas, maka perhitungan akurasi, presisi, dan recall pada langkah ketiga dilakukan menggunakan rumus yang telah ditentukan pada bab 3, sebagai berikut:

Tabel 8. 10-Fold cross validation

Fold	Akurasi	Presisi	Recall
langkah uji 1	95.00%	96.30%	98.11%
langkah uji 2	96.67%	100.00%	96.23%
langkah uji 3	96.67%	98.11%	98.11%
langkah uji 4	93.33%	100.00%	92.45%
langkah uji 5	96.67%	100.00%	96.23%
langkah uji 6	98.33%	98.15%	100.00%
langkah uji 7	95.00%	98.08%	96.23%
langkah uji 8	96.67%	98.08%	98.08%
langkah uji 9	98.33%	98.11%	100.00%
langkah uji 10	95.00%	98.04%	96.15%

Hasil evaluasi menunjukkan bahwa metode *10-fold cross validation* mencapai performa tertinggi dengan akurasi tertinggi sebesar 98,33%, presisi sebesar 98,15%, dan recall sebesar 100%. Peningkatan ini menunjukkan kestabilan dan keandalan model dalam mengklasifikasikan data.

Tabel 9. Confusion matrix 10-Fold Sebelum SMOTE

Predict Class	Actual Class		Jumlah
	Positif (+)	Negatif (-)	
Positif (+)	527	73	600
Negatif (-)	0	0	0

Tabel 10. Confusion matrix 10-Fold Sesudah SMOTE

Predict Class	Actual Class		Jumlah
	Positif (+)	Negatif (-)	
Positif (+)	512	8	520
Negatif (-)	15	65	80

Penggunaan *SMOTE* meningkatkan performa klasifikasi dengan menyeimbangkan data latih antara kelas positif dan negatif. Tanpa *SMOTE*, model cenderung memprediksi semua data sebagai kelas mayoritas (positif), sehingga sentimen negatif tidak terdeteksi. Dengan *SMOTE*, model mampu memprediksi kedua kelas secara lebih akurat.

5. KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini menerapkan algoritma *Multinomial Naïve Bayes* dengan pembobotan *TF-IDF* serta validasi *10-Fold cross validation* untuk menganalisis sentimen ulasan produk *facial wash* merek Kahf di Shopee. Model menunjukkan performa baik dengan rata-rata akurasi 98,3%, presisi 98,15%, dan recall 100%, menandakan kemampuan yang kuat dalam membedakan sentimen positif dan negatif.

Saran

Data yang digunakan masih terbatas dan tidak seimbang antar kelas. Penelitian selanjutnya disarankan mengumpulkan data lebih banyak dan beragam dari berbagai periode waktu agar model dapat belajar dari variasi ulasan yang lebih luas dan distribusi kelas yang lebih seimbang.

DAFTAR REFERENSI

- Adha, M., & Hayani, N. (2024). Digital triggers: How flash sales, live streaming, and free shipping fuel impulse buying of Pinkflash cosmetics among Gen Z.
- Afdal, M., & Elita, L. R. (2022). Penerapan text mining pada aplikasi Tokopedia. *Jurnal Teknologi Informasi dan Komunikasi*, 8(1), 78-87.
- Alamin, Z., Missouri, R., Sutriawan, S., Fathir, F., & Khairunnas, K. (2023). Perkembangan e-commerce: Analisis dominasi Shopee sebagai primadona marketplace di Indonesia. *J-ESA (Jurnal Ekonomi Syariah)*, 6(2), 120-131. <https://doi.org/10.52266/jesa.v6i2.2484>
- Ali, M., Jamil, F., Iqbal, N., Rho, S., & Jeon, G. (2023). A deep learning-based decision-making approach for predicting emotional states using physiological signals: A case study of EEG and GSR. *Sensors*, 23(5), 2333. <https://doi.org/10.3390/s23052333>
- Ambar, T., Putra, I., & Sari, D. (2023). Analisis pengaruh ulasan pengguna terhadap keputusan pembelian di e-commerce. *Jurnal Manajemen dan Bisnis*, 15(2), 45-60.
- Arsadhana, M., Efendi, B., & Trihudyatmanto, M. (2025). Analisis kepuasan pelanggan melalui sentimen ulasan menggunakan algoritma Naïve Bayes. *Jurnal Magisma*, 13(1), 1-8. <https://doi.org/10.23960/jitet.v13i1.6002>
- Atimi, R. L., & Pratama, E. E. (2022). Implementasi model klasifikasi sentimen pada review produk Lazada Indonesia. *Jurnal Sains dan Informatika*, 8(1), 88-96. <https://doi.org/10.34128/jsi.v8i1.419>
- Bastoni, R. P. (2024). Analisis sentimen opini publik terhadap calon presiden Indonesia 2024 menggunakan metode Naïve Bayes dan lexicon based pada platform YouTube [Skripsi, UIN Syarif Hidayatullah Jakarta].

- Charte, F., Rivera, A. J., del Jesus, M. J., & Herrera, F. (2023). Addressing imbalanced data in classification: A comprehensive review. *Machine Learning*. <https://doi.org/10.1007/s10994-022-06296-4>
- Firdaus, A., Firdaus, W. I., Studi, P., Informatika, T., Digital, M., & Sriwijaya, P. N. (2021). Text mining. *Jurnal Teknologi Informasi dan Komputer*, 13(1), 66-78.
- Gudiato, C., Sedyono, E., & Sembiring, I. (2022). Analisis sistem e-commerce pada Shopee untuk meningkatkan daya saing. *Jurnal Teknologi Informasi*, 2(1), 6-10. <https://doi.org/10.46229/jifotech.v2i1.294>
- Hadaina, F., & Budiyo, U. (2022). Implementasi metode Multinomial Naïve Bayes untuk sentiment analysis terhadap data ulasan produk Colearn pada Google Play Store. *Jurnal Teknologi dan Sistem Komputer*, 10(3), 660-666.
- Hermawan, A., & Jowensen, I. (2023). Implementasi text-mining untuk analisis sentimen pada Twitter dengan algoritma support vector machine. *Jurnal Ilmu Komputer dan Teknologi Informasi*, 12(1), 129-137. <https://doi.org/10.23887/jstundiksha.v12i1.52358>
- Lubis, F., Ayuningtyas, K. B., Rahmadani, S., Purba, Z., Destria, L., Ginting, P., Syahrani, R. A., Andrew, J., & Prayoga, W. (2023). Penggunaan metode text mining untuk mengekstrak informasi penting dari teks laporan penelitian. *Jurnal Motivasi Pendidikan dan Bahasa*, 1(4). <https://doi.org/10.59581/jmpb-widyakarya.v1i4.1961>
- Nabilla, S., & Hafiar, H. (2024). Analisis media monitoring terhadap brand Kahf melalui tools Brand24 pada bulan Desember 2024. *Journal of Digital Communication Science*, 2(2), 71-77. <https://doi.org/10.56956/jdcs.v2i2.398>
- Pakan, E. D., & Purwanto, S. (2022). Pengaruh beauty influencer dan kualitas produk terhadap keputusan pembelian produk facial wash Garnier di Surabaya. *Scientific Journal of Reflection: Economic, Accounting, Management and Business*, 5(3), 764-772. <https://doi.org/10.37481/sjr.v5i3.535>
- Prakoso, Q. A. N., Muliawati, A., & Isnainiyah, I. N. (2022). Analisis sentimen terhadap produk Skin Game di forum review Female Daily menggunakan metode Multinomial Naïve Bayes dan TF-IDF. *Informatik: Jurnal Ilmu Komputer*, 18(3), 198. <https://doi.org/10.52958/iftk.v18i3.4679>
- Putra, I. E., Asriyanik, & Azzahra, F. F. (2025). Pemanfaatan Multinomial Naïve Bayes untuk analisis sentimen ulasan aplikasi mobile JKN. *JATI - Jurnal Mahasiswa Teknik Informatika*, 9(2), 115-123. <https://doi.org/10.36040/jati.v9i2.13311>
- Sabrani, A., Putu, I. G., Wedashwara, W., & Bimantoro, F. (2020). Metode Multinomial Naïve Bayes untuk klasifikasi artikel online tentang gempa di Indonesia. *Jurnal Sistem dan Informatika*, 2(1), 89-100. <https://doi.org/10.29303/jtika.v2i1.87>

Setiawan, A., & Rachmawati, D. (2022). Penerapan algoritma Naïve Bayes untuk analisis sentimen pada ulasan produk e-commerce. *Jurnal Teknologi Informasi*, 18(1), 45-56.

Setyaningsih, A. F., Septiyani, D., & Widiyari, S. R. (2023). Implementasi algoritma Naïve Bayes untuk analisis sentimen masyarakat pada Twitter mengenai kepopuleran produk skincare di Indonesia. *Jurnal Teknologi Informatika dan Komputer*, 9(1), 224-235. <https://doi.org/10.37012/jtik.v9i1.1409>