

Analisis Kepuasan Pelanggan terhadap Beberapa Produk yang di Jual di E-Commerce Menggunakan Metode *Naïve Bayes* dan *Logistic Regression*

Java Diovanka Alam^{1*}, Musyaffa Ramdhan², Muhammad Yuzakki Raja Rafael³, Muhammad Faiz Hamka⁴, Desmulyati⁵, Imam Budiawan⁶

^{1,2,3,4,6}Teknologi Informasi, ⁵Informatika, Universitas Bina Sarana Informatika, Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat
E-mail: 17230847@bsi.ac.id

* Corresponding Author

 <https://doi.org/10.31004/jerkin.v4i3.4970>

ARTICLE INFO

Article history

Received: 23 Nov 2025

Revised: 05 Dec 2025

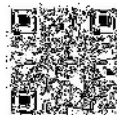
Accepted: 30 Dec 2025

Kata Kunci:

Kepuasan Pelanggan, E-Commerce, *Naive Bayes*, Logistic Regression, Ulasan Pelanggan.

Keywords:

KIP-K, Machine Learning, *Naive Bayes*, Classification, Education.



ABSTRACT

Kepuasan pelanggan adalah elemen penting yang berperan dalam kelangsungan bisnis di sektor e-commerce. Ulasan yang diberikan oleh para konsumen menjadi sumber informasi yang penting untuk menilai seberapa puas mereka terhadap produk yang dibeli. Penelitian ini bertujuan untuk menilai tingkat kepuasan pelanggan dengan menggunakan data ulasan produk melalui dua metode klasifikasi, yaitu Multinomial Naive Bayes dan Logistic Regression. Data yang digunakan berasal dari dataset nyata dalam bahasa Indonesia yang mencakup teks ulasan serta penilaian dari para pembeli. Proses penelitian terdiri dari beberapa tahapan, dimulai dari persiapan teks, ekstraksi fitur menggunakan metode TF-IDF, pengelompokan label kepuasan, pelatihan model, dan evaluasi menggunakan metrik seperti akurasi, presisi, recall, F1-score, serta matriks kebingungan. Temuan dari penelitian ini menunjukkan bahwa kedua metode dapat memprediksi kepuasan pelanggan dengan akurasi yang bersaing. Logistic Regression menunjukkan hasil yang lebih konsisten dibandingkan Naive Bayes dalam konteks teks berbahasa Indonesia. Hasil ini dapat dimanfaatkan oleh perusahaan e-commerce untuk memantau kualitas produk serta terus memperbaiki layanan untuk konsumen.

Customer satisfaction is a crucial element that plays a significant role in the sustainability of businesses in the e-commerce sector. Reviews provided by consumers serve as an important source of information to assess how satisfied they are with the products they purchased. This study aims to evaluate customer satisfaction levels using product review data through two classification methods: Multinomial Naive Bayes and Logistic Regression. The data used comes from a real Indonesian-language dataset that includes review texts and buyer ratings. The research process consists of several stages, starting from text preprocessing, feature extraction using the TF-IDF method, satisfaction label grouping, model training, and evaluation using metrics such as accuracy, precision, recall, F1-score, and confusion matrix. The findings of this study indicate that both methods can predict customer satisfaction with competitive accuracy. Logistic Regression demonstrates more consistent results compared to Naive Bayes in the context of Indonesian-language text. These results can be utilized by e-commerce companies to monitor product quality and continuously improve services for consumers.



This is an open access article under the CC-BY-SA license.

How to Cite: Java Diovanka Alam, et al (2025). Analisis Kepuasan Pelanggan terhadap Beberapa Produk yang di Jual di E-Commerce Menggunakan Metode *Naive Bayes* dan *Logistic Regression*, 4(3) 16606-16613. <https://doi.org/10.31004/jerkin.v4i3.4970>

PENDAHULUAN

Program Kartu Indonesia Pintar Kuliah (KIP-K) merupakan salah satu bentuk dukungan pemerintah dalam meningkatkan akses pendidikan tinggi bagi mahasiswa dari keluarga kurang mampu. Namun, proses seleksi penerima KIP-K sering menghadapi tantangan, terutama dalam menilai kelayakan berdasarkan berbagai variabel sosial, ekonomi, dan akademik. Ketidaktepatan dalam proses

seleksi dapat menyebabkan mahasiswa yang seharusnya layak tidak mendapatkan bantuan, atau sebaliknya, mahasiswa yang kurang memenuhi syarat justru terpilih.

Seiring berkembangnya teknologi, pemanfaatan Machine Learning telah menjadi solusi yang semakin relevan dalam mendukung pengambilan keputusan berbasis data. Algoritma klasifikasi seperti Naive Bayes dapat digunakan untuk memprediksi kelayakan penerima KIP-K berdasarkan faktor-faktor yang memengaruhi, seperti pendapatan keluarga, jumlah tanggungan, pekerjaan orang tua, status tempat tinggal, penerimaan bantuan sosial, IPK, serta tingkat kehadiran mahasiswa. Pendekatan ini memungkinkan proses seleksi menjadi lebih objektif, cepat, dan akurat.

Penelitian ini bertujuan untuk mengembangkan model klasifikasi penerima KIP-K menggunakan algoritma Naive Bayes dengan memanfaatkan dataset yang berisi indikator-indikator kelayakan mahasiswa. Model kemudian diuji menggunakan metrik evaluasi seperti accuracy, precision, recall, dan f1-score guna memastikan performanya dalam memprediksi calon penerima bantuan. Dengan adanya model ini, institusi pendidikan diharapkan dapat terbantu dalam proses penilaian awal terhadap mahasiswa yang berpotensi layak menerima KIP-K.

METODE

Penelitian ini dilaksanakan dengan mengikuti beberapa langkah terencana yang bertujuan untuk mengevaluasi kepuasan konsumen berdasarkan umpan balik produk di platform e-commerce. Langkah pertama dalam penelitian adalah menghimpun data ulasan dari pelanggan, kemudian dilanjutkan dengan preprocessing untuk mempersiapkan teks agar siap untuk dianalisis. Setelah itu, dilaksanakan pemetaan label berdasarkan penilaian yang diberikan oleh pelanggan (Kumar Patra et al., 2023). Sesudah data dipersiapkan, dilakukan pelatihan dengan memanfaatkan dua model pembelajaran mesin, yaitu Multinomial Naive Bayes dan Logistic Regression. Selanjutnya, kedua model tersebut diuji menggunakan berbagai metrik untuk menilai tingkat akurasi yang diperoleh. Setiap tahap dijelaskan secara mendetail sebagai berikut:

Dataset

Dataset yang digunakan dalam studi ini terdiri dari dua komponen utama, yaitu teks ulasan dari pelanggan (review_text) dan penilaian dari pelanggan (rating) (Alzahrani et al., 2022). Dataset ini menggambarkan pandangan dan pengalaman konsumen terhadap barang yang dibeli via platform e-commerce. Sebelum digunakan untuk melatih model, dataset tersebut menjalani tahap pembersihan dan pengolahan, kemudian dibagi menjadi data latih dan data uji agar dapat diterapkan pada fase klasifikasi. Berikut adalah atribut yang diterapkan dalam penelitian ini:

No	Ulasan Pelanggan	Rating Pelanggan
1.	produknya bagus dan sesuai	5
2.	Sangat puas dengan barangnya	4
3.	Kualitas lumayan	3
4.	Barang jelek tidak sesuai	2
5.	Kecewa barangnya	1

Preprocessing

Sebelum data review dipakai dalam pelatihan model, tahap pengolahan awal diperlukan untuk menjamin bahwa teks dalam keadaan bersih, konsisten, dan siap untuk diproses oleh algoritma pembelajaran mesin (Tabassum & Patil, 2020). Pengolahan awal dilakukan untuk meminimalkan gangguan pada data, merampingkan bentuk kata, serta mengkonversi teks menjadi bentuk numerik yang bisa dimengerti oleh model (Alzahrani et al., 2022). Langkah-langkah pengolahan awal yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Lowercase
yaitu mengubah seluruh teks ulasan menjadi huruf kecil agar konsisten.
2. Menghapus tanda baca, angka, dan symbol
untuk membersihkan teks dari karakter yang tidak relevan.
3. Tokenisasi
yaitu memecah kalimat menjadi unit kata.

Analisis Kepuasan Pelanggan terhadap Beberapa Produk yang di Jual di E-Commerce Menggunakan Metode Naïve Bayes dan Logistic Regression, Java Diovanka Alam, Musyaffa Ramdhan, Muhammad Yuzakki Raja Rafael, Muhammad Faiz Hamka, Desmulyati, Imam Budiawan

16608

4. Stopword removal Bahasa Indonesia
untuk menghilangkan kata-kata umum yang tidak memiliki makna penting.
5. Stemming menggunakan Sastrawi
untuk mengubah kata ke bentuk dasarnya.
6. Ekstraksi fitur menggunakan TF-IDF
agar teks dapat diubah menjadi representasi numerik berbobot.

Pemetaan Label

Sebelum data digunakan dalam proses klasifikasi, rating pelanggan perlu dikonversi menjadi label yang lebih sederhana agar model dapat mengenali kategori kepuasan dengan lebih jelas. Pemetaan label ini bertujuan untuk membedakan ulasan yang bernada positif dan negatif berdasarkan nilai rating yang diberikan pelanggan. Dengan demikian, model machine learning dapat dilatih untuk mengidentifikasi pola sentimen dalam teks ulasan berdasarkan kategori kepuasan yang telah ditentukan (Bahtiar et al., 2023b). Adapun aturan pemetaan label dalam penelitian ini adalah sebagai berikut:

Rating	Kategori	Label
5	Sangat puas	1
4	Puas	1
3	Netral	-
2	Tidak Puas	0
1	Sangat Tidak Puas	0

Rating ≥ 4 = Puas (label 1)

Rating ≤ 2 = Tidak Puas (label 0)

Rating = 3 = dihapus (netral)

Pelatihan Model

Tahap pelatihan model dilakukan setelah data teks selesai melalui proses preprocessing dan pemetaan label kepuasan. Pada tahap ini, fitur-fitur hasil ekstraksi TF-IDF digunakan sebagai input untuk melatih algoritma klasifikasi sehingga model dapat mempelajari pola hubungan antara teks ulasan dan tingkat kepuasan pelanggan. Data yang telah diproses kemudian dibagi menjadi dua bagian, yaitu 80% sebagai data latih dan 20% sebagai data uji, untuk memastikan bahwa model dapat dievaluasi secara objektif. Dalam penelitian ini digunakan dua algoritma klasifikasi yang umum digunakan dalam analisis teks (Xiao et al., 2024), yaitu:

Item	Keterangan
Model 1	Multinomial Naïve Bayes
Model 2	Logistic Regression
Fitur Input	TF-IDF
Preprocessing	lowercase, remove punctuation, tokenizing, stopwords removal, stemming
Split data	80% train – 20% test
Catatan	Training dilakukan di python

Pembagian data ini bertujuan untuk memberikan kesempatan bagi model menggeneralisasi pola dari data latih dan menguji performanya pada data yang belum pernah dilihat sebelumnya.

Evaluasi Model

Tahap evaluasi model dilakukan untuk mengukur kemampuan algoritma dalam memprediksi tingkat kepuasan pelanggan secara akurat berdasarkan data uji. Evaluasi ini penting untuk mengetahui seberapa baik model dapat menggeneralisasi data baru yang tidak digunakan dalam proses pelatihan (Bahtiar et al., 2023b). Pada penelitian ini, sejumlah metrik digunakan untuk menilai performa model dari berbagai aspek, baik dari segi ketepatan prediksi maupun kemampuan mendeteksi kelas positif dan negatif dengan benar (Ho et al., 2022). Adapun metrik evaluasi yang digunakan adalah sebagai berikut:

1. Accuracy

Akurasi adalah ukuran yang digunakan untuk menilai persentase prediksi yang benar dari model jika dibandingkan dengan total data uji. Nilai akurasi yang lebih tinggi menunjukkan bahwa model

lebih efektif dalam mengklasifikasikan data secara keseluruhan tanpa melihat distribusi kelas.

Rumus:

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

Penjelasan:

- a. TP = True Positive
- b. TN = True Negative
- c. FP = False Positive
- d. FN = False Negative

2. Precision

Precision digunakan untuk mengukur seberapa akurat model dalam memprediksi kategori puas (label 1). Dengan kata lain, precision mencerminkan bagian dari prediksi positif yang sebenarnya merupakan kelas positif (UC Irvine UC Irvine Electronic Theses and Dissertations Title An Empirical Comparison of Machine Learning Methods for Text-Based Sentiment Analysis of Online Consumer Reviews, 2022).

Rumus:

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

Precision yang tinggi menunjukkan bahwa model jarang menganggap ulasan “tidak puas” sebagai “puas.”

3. Recall

Recall atau sensitivitas mengidentifikasi seberapa baik model dapat menemukan semua data ulasan yang sebenarnya masuk dalam kategori puas. Nilai recall yang tinggi menandakan bahwa model ini tidak sering mengabaikan data positif.

Rumus:

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

4. F1-score

F1-Score merupakan rata-rata harmonis antara Presisi dan Recall. Metrik ini bermanfaat saat diperlukan keseimbangan antara Presisi dan Recall, terutama pada situasi dengan data yang tidak seimbang.

Rumus:

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. Confusion Matrix

Matrix Kebingungan adalah sebuah tabel analisis yang menggambarkan jumlah prediksi yang tepat dan yang keliru untuk setiap kategori. Tabel ini menyajikan informasi rinci tentang kesalahan yang dilakukan model serta kemampuan model dalam membedakan satu kategori dari yang lain.

Struktur Confusion Matrix:

	Prediksi Puas	Prediksi Tidak Puas
Ulasan Puas	True Positive (TP)	False Negative (FN)
Ulasan Tidak Puas	False Positive (FP)	True Negative (TN)

Confusion Matrix mempermudah analisis kesalahan model serta mengetahui apakah model cenderung memberikan prediksi berlebih pada salah satu kelas tertentu.

Hasil Evaluasi Model

Berdasarkan uji yang dilakukan dengan data pengujian, didapatkan hasil untuk mengevaluasi kinerja dua algoritma klasifikasi, yaitu Naïve Bayes dan Regresi Logistik, dengan melibatkan metrik Akurasi, Presisi, Recall, dan F1-score. Hasil secara lengkap dapat dilihat pada tabel berikut:

Model	Accuracy	Precision	Recall	F1-score
Naïve Bayes	0.8375 (83.75%)	0.86 (Weighted Avg)	0.84 (Weighted Avg)	0.83 (Weighted Avg)
Logistic Regression	0.8625 (86.25%)	0.87 (Weighted Avg)	0.86 (Weighted Avg)	0.86 (Weighted Avg)

Interpretasi hasil

Akurasi

1. Naïve Bayes mencapai tingkat ketepatan sebesar 83.75%, yang menunjukkan bahwa model ini dapat dengan tepat memprediksi ulasan pelanggan sebanyak 84 dari setiap 100 data yang diuji.
2. Regresi Logistik menunjukkan tingkat akurasi yang lebih tinggi, yaitu 86,25%, sehingga lebih efektif dalam memberikan prediksi yang akurat.

Precision

1. Naïve Bayes memiliki Precision 0,86, yang berarti dari semua prediksi yang disebut "puas", 86% di antaranya adalah ulasan yang memang diberi label puas.
2. Regresi Logistik menunjukkan tingkat ketepatan yang lebih baik yakni 0.87, sehingga model ini lebih efektif dalam mengenali kategori puas tanpa melakukan kesalahan dalam prediksi.

Recall

1. Model Naïve Bayes menunjukkan tingkat recall sebesar 0.84, yang berarti ada beberapa ulasan positif yang tidak dikenali sebagai positif.
2. Logistic Regression menunjukkan kinerja yang baik dengan tingkat recall 0.86, yang menunjukkan kemampuannya dalam mengenali ulasan yang benar-benar positif.

F1-score

1. F1-score Naïve Bayes: 0.83
2. F1-score Logistic Regression: 0.86

Model Logistic Regression kembali unggul, menunjukkan performa yang lebih seimbang antara precision dan recall.

Kesimpulan Evaluasi Model

Model ini menghasilkan:

1. Akurasi tertinggi (0.8625)
2. Precision tertinggi (0.87)
3. Recall tertinggi (0.86)
4. F1-score tertinggi (0.86)

Ini menunjukkan bahwa Regresi Logistik dapat memberikan prediksi yang lebih konsisten, tepat, dan adil jika dibandingkan dengan Naïve Bayes.

HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil pengkajian yang diperoleh dari proses pelatihan dan penilaian model dengan menggunakan dataset ulasan dari pelanggan. Hasil yang dipresentasikan mencakup statistik dasar dari dataset, kinerja masing-masing model klasifikasi, serta tampilan visual hasil prediksi dalam bentuk matriks kebingungan. Di samping itu, ada diskusi untuk memberikan penjelasan mengenai kinerja model dan menjelaskan elemen-elemen yang mempengaruhi hasil tersebut. Analisis ini bertujuan untuk mengevaluasi seberapa efektif algoritma Multinomial Naive Bayes dan Logistic Regression dalam memprediksi tingkat kepuasan pelanggan berdasarkan teks ulasan dan penilaian dari pelanggan.

Statistik Dataset

Dataset pada penelitian ini terdiri dari 800 ulasan pelanggan. Dari jumlah tersebut, ulasan dibagi menjadi dua kategori utama, yaitu ulasan puas dan ulasan tidak puas, berdasarkan pemetaan rating yang telah ditetapkan. Pembagian ini digunakan sebagai dasar dalam proses pelatihan dan evaluasi model prediksi kepuasan pelanggan.

Kategori	Jumlah
Total ulasan	800
Ulasan Puas	495
Ulasan Tidak Puas	305

Performe Detail

Naïve Bayes mendapatkan akurasi sebesar 0.8375, sementara Logistic Regression menunjukkan hasil lebih baik dengan 0.8625. Selain itu, Logistic Regression mempunyai nilai precision, recall, dan F1-score yang lebih tinggi, sehingga bisa disimpulkan bahwa model ini lebih tepat dan lebih efisien dalam meramalkan kepuasan pelanggan.

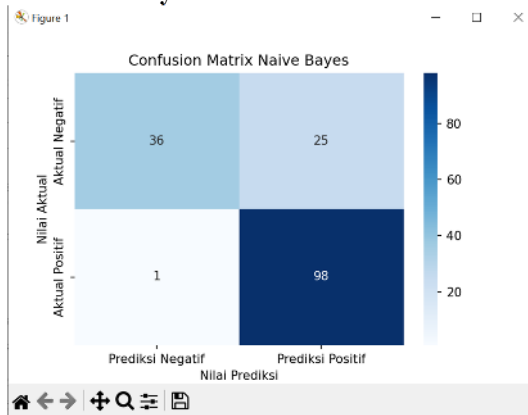
Model	Accuracy	Precision	Recall	F1-score
Naïve Bayes	0.8375 (83.75%)	0.86 (Weighted Avg)	0.84 (Weighted Avg)	0.83 (Weighted Avg)
Logistic Regression	0.8625 (86.25%)	0.87 (Weighted Avg)	0.86 (Weighted Avg)	0.86 (Weighted Avg)

Confusion Matrix

Naïve Bayes

	Prediksi Puas (1)	Prediksi Tidak Puas (0)	TOTAL Ulasan Aktual
Ulasan Aktual: Puas (1)	98 (TP)	1 (FN)	99
Ulasan Aktual: Tidak Puas (0)	25 (FP)	36 (TN)	61
TOTAL Prediksi	123	37	160

Visualisasi Confusion Matrix Naïve Bayes

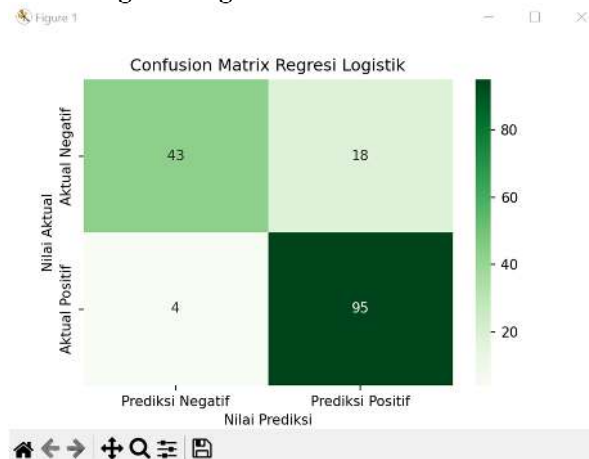


Gambar 1. Visualisasi Confusion Matrix Naïve Bayes

Logistic Regression

	Prediksi Puas (1)	Prediksi Tidak Puas (0)	TOTAL Ulasan Aktual
Ulasan Aktual: Puas (1)	95 (TP)	4 (FN)	99
Ulasan Aktual: Tidak Puas (0)	18 (FP)	43 (TN)	61
TOTAL Prediksi	113	47	160

Visualisasi Confusion Matrix Logistic Regression



Gambar 2. Visualisasi Confusion Matrix Logistic Regression

Pembahasan

Hasil analisis menunjukkan bahwa Logistic Regression memiliki kinerja yang lebih konsisten dibandingkan dengan Naive Bayes, khususnya pada data ulasan dalam bahasa Indonesia. Ini disebabkan karena Logistic Regression mampu mengelola fitur TF-IDF dengan dimensi yang tinggi dan sensitivitas terhadap perubahan pola dalam teks. Di sisi lain, Naive Bayes memiliki pendekatan yang lebih sederhana dan cepat, tetapi kinerjanya lebih terpengaruh oleh ketidaksamaan distribusi kata. Selain itu, data dari confusion matrix menunjukkan bahwa Logistic Regression lebih efektif dalam mengurangi kesalahan klasifikasi untuk kelas puas maupun tidak puas. Oleh karena itu, Logistic Regression bisa dianggap sebagai model yang lebih efisien untuk menganalisis kepuasan pelanggan berdasarkan ulasan e-commerce.

SIMPULAN

Penelitian ini bertujuan untuk mengevaluasi kepuasan konsumen terhadap produk yang ditawarkan di platform e-commerce dengan memanfaatkan dua algoritma klasifikasi, yaitu Multinomial Naive Bayes dan Logistic Regression. Dari hasil pengolahan dan penilaian model, dapat disimpulkan bahwa kedua algoritma tersebut mampu mengklasifikasikan kepuasan konsumen dengan cukup memuaskan. Meskipun demikian, Logistic Regression menunjukkan kinerja yang lebih konsisten dan tepat dalam memprediksi kepuasan konsumen berdasarkan ulasan berbahasa Indonesia. Hal ini terlihat dari tingkat akurasi dan metrik evaluasi lainnya yang lebih baik dibandingkan dengan Naive Bayes. Penelitian ini juga mengindikasikan bahwa pemrosesan teks, pemetaan label, serta pilihan metode ekstraksi fitur seperti TF-IDF sangat berpengaruh dalam meningkatkan kualitas model. Oleh karena itu, Logistic Regression dapat diusulkan sebagai metode yang lebih efisien untuk menganalisis ulasan konsumen di platform e-commerce (Widhiyanti & Sekarini, 2025).

Untuk penelitian mendatang, direkomendasikan agar menggunakan data yang lebih luas dan bervariasi sehingga model dapat memahami pola-pola ulasan dengan lebih mendetail. Di samping itu, penerapan metode deep learning seperti LSTM, Bi-LSTM, atau IndoBERT patut dipertimbangkan untuk mencapai kinerja yang lebih baik dalam analisis teks berbahasa Indonesia. Penambahan fitur seperti analisis berbasis aspek juga dapat memberikan pemahaman yang lebih mendalam mengenai elemen-elemen yang memengaruhi kepuasan pelanggan. Mengintegrasikan hasil analisis ke dalam dashboard pemantauan kepuasan pelanggan secara langsung juga dapat memberi kontribusi yang bermanfaat bagi platform e-commerce dalam meningkatkan layanan dan pengalaman pelanggan.

UCAPAN TERIMA KASIH

Peneliti menyampaikan ucapan terima kasih kepada pihak yang sudah berkontribusi dalam pelaksanaan penelitian dan penyusunan artikel ini.

REFERENSI

- Alzahrani, M. E., Aldhyani, T. H. H., Alsubari, S. N., Althobaiti, M. M., & Fahad, A. (2022). Developing an Intelligent System with Deep Learning Algorithms for Sentiment Analysis of E-Commerce Product Reviews. In *Computational Intelligence and Neuroscience* (Vol. 2022). Hindawi Limited. <https://doi.org/10.1155/2022/3840071>
- Bahtiar, S. A. H., Dewa, C. K., & Luthfi, A. (2023a). Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling. *Journal of Information Systems and Informatics*, 5(3), 915–927. <https://doi.org/10.51519/journalisi.v5i3.539>
- Bahtiar, S. A. H., Dewa, C. K., & Luthfi, A. (2023b). Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling. *Journal of Information Systems and Informatics*, 5(3), 915–927. <https://doi.org/10.51519/journalisi.v5i3.539>
- Ho, I., Goh, H. N., & Tan, Y. F. (2022). Preprocessing Impact on Sentiment Analysis Performance on Malay Social Media Text. *Journal of System and Management Sciences*, 12(5), 73–90.

- <https://doi.org/10.33168/JSMS.2022.0505>
- Kumar Patra, G., Kuraku, C., Konkimalla, S., Boddapati, V. N., Sarisa, M., Rajaram, S. K., Reddy, M. S., Polimetla, K., & Patra, G. K. (2023). A Sentiment Analysis of Customer Product Review Based on Machine Learning Techniques in E-Commerce. 2(4), 1–4.
[https://doi.org/10.47363/JAICC/2023\(2\)389](https://doi.org/10.47363/JAICC/2023(2)389)
- Penyusun, T., Nafi, B., & Mijiarto, dan J. (n.d.). Transisi Kenormalan Baru : Eksistensi BUM Desa, UMKM, dan Ormas Editor: Arimurti Kriswibowo Dandi Darmadi.
- Tabassum, A., & Patil, R. R. (2020). A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing. International Research Journal of Engineering and Technology. www.irjet.net
- UC Irvine UC Irvine Electronic Theses and Dissertations Title An Empirical Comparison of Machine Learning Methods for Text-based Sentiment Analysis of Online Consumer Reviews. (2022). <https://escholarship.org/uc/item/7q62b9b8>
- Widhiyanti, A. A. S., & Sekarini, I. G. A. A. (2025). Naïve Bayes–Based Chatbot with Sentiment Analysis for Culinary Preferences in Bali. Sinkron, 9(4), 2007–2014.
<https://doi.org/10.33395/sinkron.v9i4.15291>
- Xiao, L., Li, Q., Ma, Q., Shen, J., Yang, Y., & Li, D. (2024). Text classification algorithm of tourist attractions subcategories with modified TF-IDF and Word2Vec. PLoS ONE, 19(10 October).
<https://doi.org/10.1371/journal.pone.0305095>