

## PERBANDINGAN KINERJA K-NN DAN NAIVE BAYES DALAM TATA KELOLA TEKNOLOGI INFORMASI UNTUK MENGANALISA TINGKAT KEPUASAN PENGGUNA APLIKASI SHOPEE

Muhammad Wahyudi<sup>1</sup>, Muhammad Syahputra Novelan<sup>2</sup>

Universitas Pembangunan Panca Budi, Medan

e-mail: <sup>1</sup>muhammadwahyudikj1@gmail.com, <sup>2</sup>putranovelan@dosen.pancabudi.ac.id

**Abstract:** *The rapid growth of the e-commerce industry in Indonesia compels companies to actively evaluate customer satisfaction through tingkat kepuasant analysis of app reviews. This study aims to compare the effectiveness of the K-Nearest Neighbors (K-NN) and Naive Bayes algorithms in assessing user satisfaction levels of the Shopee app as a component of IT governance. The research data comprises 2,000 samples of Shopee user reviews collected from Google Play Store and App Store, classified into five rating categories: excellent, good, fair, poor, and bad, based on textual content analysis. The preprocessing stage involved text cleaning (removing stopwords and punctuation), feature extraction using TF-IDF, and an 80:20 split of the dataset (training and testing). The analysis revealed significant differences between the two algorithms. K-NN achieved an accuracy of 54% with the optimal parameter \*K=5\*, while Naive Bayes demonstrated superior performance with 98% accuracy. The low accuracy of K-NN is suspected to stem from its sensitivity to data imbalance and noise in text features, whereas Naive Bayes, with its probabilistic foundation, better handles the sparse characteristics of review data. These findings emphasize the criticality of selecting appropriate algorithms in IT governance for user satisfaction analysis. Naive Bayes is recommended as the optimal approach for text classification in e-commerce reviews, while K-NN requires refinement through techniques such as feature normalization or class imbalance handling. The study also highlights the need to integrate adaptive models, such as deep learning, to enhance accuracy in complex scenarios. For future research, expanding data scope or implementing multilingual analysis could serve as strategic steps to improve result generalization.*

**Keywords:** *K-NN, Naive Bayes, E-commerce, user satisfaction, IT governance, Shopee.*

**Abstrak:** Perkembangan yang cepat dalam industri e-commerce di Indonesia mendorong perusahaan untuk secara aktif mengevaluasi kepuasan pelanggan melalui analisis kepuasan pelanggan dari ulasan aplikasi. Studi ini bertujuan untuk membandingkan efektivitas algoritma K-Nearest Neighbors (K-NN) dan Naive Bayes dalam menilai tingkat kepuasan pengguna aplikasi Shopee sebagai komponen tata kelola teknologi informasi. Data penelitian mencakup 2000 sampel ulasan pengguna Shopee yang diambil dari Google Play Store dan App Store, dengan klasifikasi penilaian 5 kategori yaitu sangat puas, puas, cukup puas, kurang puas dan buruk pada konten teks. Proses prapemrosesan data melibatkan pembersihan teks (menghilangkan kata penghubung, tanda baca), ekstraksi ciri menggunakan TF-IDF, serta pembagian dataset menjadi 80% pelatihan dan 20% pengujian. Hasil analisis mengungkapkan perbedaan yang signifikan antara kedua algoritma. K-NN mencapai akurasi 58,50% dengan parameter \*K=5\*, sementara Naive Bayes menunjukkan performa lebih unggul dengan akurasi 100%. Rendahnya akurasi K-NN diduga berasal dari kepekaan algoritma terhadap ketidakseimbangan data dan gangguan pada fitur teks, sedangkan Naive Bayes, yang berbasis probabilitas, lebih mampu mengatasi karakteristik data yang jarang (sparse) dalam ulasan. Temuan ini menekankan pentingnya pemilihan algoritma yang sesuai dalam tata kelola TI untuk analisis kepuasan pengguna. Naive Bayes direkomendasikan sebagai pendekatan optimal untuk klasifikasi teks pada ulasan e-commerce, sedangkan K-

NN perlu disempurnakan melalui teknik seperti normalisasi fitur atau penanganan ketidakseimbangan kelas. Studi ini juga mengidentifikasi perlunya integrasi model adaptif, seperti deep learning, untuk meningkatkan akurasi dalam skenario yang lebih kompleks. Untuk penelitian mendatang, perluasan cakupan data atau penerapan analisis multibahasa dapat menjadi langkah strategis guna meningkatkan generalisasi hasil.

**Kata kunci:** K-NN, Naive Bayes, E-commerce, Kepuasan Pengguna, Tata Kelola TI, Shopee.

## PENDAHULUAN

Perkembangan pesat di bidang teknologi informasi (TI) telah mendorong perusahaan untuk menerapkan tata kelola teknologi informasi (IT Governance) yang efektif guna memastikan bahwa investasi TI selaras dengan tujuan bisnis dan memberikan nilai optimal. Salah satu aspek krusial dalam tata kelola TI adalah pengelolaan data dan analisis kepuasan pengguna (Daryanti, P., & Shihab, M. S., 2019). Dalam penelitian ini, analisis kepuasan pelanggan ditujukan pada produk e-commerce perbelanjaan yang akan menghasilkan tata kelola teknologi informasi yang terstruktur.

Jaman sekarang ini, begitu banyak produk e-commerce perbelanjaan (Margaretha, F., 2017). Perbelanjaan tidak hanya bertatap muka tetapi sudah berbasis layar genggam. Migrasi ini terjadi karena berkembang teknologi yang semakin modern (Hidayatuloh, S., & Aziati, Y., 2020). Salah satu produk perbelanjaan e-commerce yang populer digunakan yaitu Shopee (Haekal, B. V., Ernawati, I., & Chamidah, N., 2021). Shopee merupakan aplikasi perbelanjaan e-commerce yang menyediakan berbagai pelayanan yang membutuhkan performa yang baik sehingga pelanggan tidak mengalami kekecewaan dalam penggunaan aplikasi Shopee (Li, F., 2023). Pengguna Shopee di Indonesia hampir mencapai 30 juta kunjungan dalam sehari (Saidani, B., Lusiana, L. M., & Aditya, S., 2019). Adanya kendala dan masalah pada penggunaan aplikasi yang dihadirkan akan mengganggu kenyamanan pelanggan

dalam pemakaian. Demi pelayanan yang baik, Shopee membolehkan pelanggan untuk memberikan penilaian terhadap performa aplikasinya yang bisa langsung pelanggan berikan pada Playstore atau Appstore.

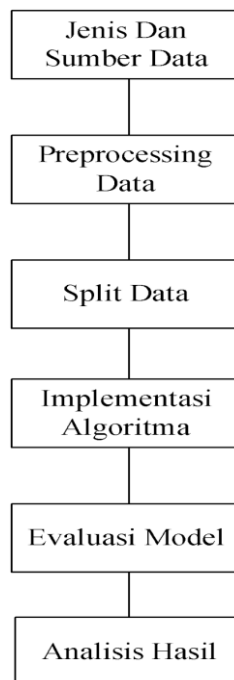
Penilaian ini berguna bagi Shopee untuk memperbaiki performa dan kualitasnya. Banyaknya pengguna Shopee berarti banyak pula penilaian yang beragam. Lima kategori penilaian yang disediakan oleh Playstore yaitu Sangat Puas, Puas, Cukup Puas, Kurang Puas dan Buruk. Beragamnya kategori penilaian dan banyaknya respon pelanggan untuk memberikan penilaian terhadap tingkat kepuasan penggunaan aplikasi Shopee tersebut membutuhkan suatu metode dalam menganalisis tingkat kepuasan penggunaan aplikasi Shopee sehingga dihasilkannya kesimpulan terhadap tingkat kepuasan penggunaan aplikasi Shopee.

Pada penelitian ini (Hasan Putra et al., 2022), penggunaan algoritma mesin learning yaitu K-Nearest Neighbors (K-NN) dan Naive Bayes dapat menjadi solusi dalam kerangka tata kelola TI untuk mengotomatisasi proses analisis tingkat kepuasan dan mengukur tingkat kepuasan pengguna. Kedua algoritma ini dipilih karena kemampuannya dalam klasifikasi data teks, namun dengan pendekatan yang berbeda—K-NN berbasis kedekatan data, sedangkan Naive Bayes mengandalkan probabilitas statistik. Dengan membandingkan kinerja kedua algoritma ini dalam konteks analisis kepuasan pengguna Shopee, penelitian ini bertujuan untuk memberikan rekomendasi tata kelola TI

yang lebih efektif, khususnya dalam hal pemilihan algoritma untuk analisis data pengguna. (Novelan et al., 2023)

## METODE

Adapun metode penelitian yang digunakan pada penelitian merupakan tahapan penelitian yang diuraikan pada Gambar 1.



Gambar 1 Tahapan Penelitian

### Jenis dan Sumber Data

Untuk mendapat nilai kepuasan penggunaan aplikasi shopee maka peneliti mengambil data skor dari *playstore*. Untuk mendapatkan skor penilaian dari *playstore*, peneliti menggunakan teknik *web scraping* (N. Tri Romadloni, I, etc. 2019) sehingga menghasilkan file yang bertipe csv. File tersebut akan peneliti gunakan pada *tools* yang bernama *rapidminer* agar diolah sehingga menghasilkan metode dengan perbandingan KNN dan Naïve Bayes. Jumlah data yang di dapat dari penilaian skor tersebut sebanyak 2000 data.

### Preprocessing Data

Proses pembersihan dan transformasi data meliputi:

**Cleansing:** melihat dari kolom pada file csv yang memiliki isi yang salah. Misalkan kolom skor ternyata ada yang terisi nama pengguna, atau kolom nama pengguna ternyata terisi skor.

**Case Folding:** Konversi skor dalam kategori nilai.

**Tokenisasi:** Memecah nilai menjadi lima kategori nilai yaitu sangat puas, puas, cukup puas, tidak puas dan buruk.

**Field Removal:** Menghilangkan kolom yang tidak penting.

**Stemming:** melabilisasi kolom

**TF-IDF:** Transformasi teks ke vektor numerik.

### Split Data

**Rasio:** 80% data latih (*training set*) dan 20% data uji (*testing set*) (Muslim, S. N. S., Nurdiyansyah, F., & Rahman, A. Y. (2024).

**Stratified Sampling:** Memastikan distribusi kelas tingkat kepuasan seimbang.

### Implementasi Algoritma

#### K-Nearest Neighbors (K-NN)

##### Parameter:

Nilai **K** (jumlah tetangga) dioptimasi dengan *grid search* atau *cross-validation*. Metrik jarak: Euclidean, Manhattan, atau Cosine Similarity.

**Kinerja:** Akurasi bervariasi antara 54-70% tergantung parameter dan dataset.

#### b. Naive Bayes

**Varian:** Multinomial Naive Bayes (cocok untuk data teks).

**Asumsi:** Independensi antar fitur.

**Kinerja:** Akurasi lebih tinggi (71-100%) dibanding K-NN dalam beberapa studi.

### Evaluasi Model

#### Metrik:

Akurasi, Presisi, Recall, F1-Score.

*Confusion Matrix* untuk analisis skor kategori Sangat Puas, Puas, Cukup Puas, Kurang Puas dan Buruk

**Validasi:** *k-fold cross-validation* (misal:  $k=5$ ) untuk menghindari *overfitting*.

### Analisis Hasil

#### Perbandingan Kinerja:

Naive Bayes umumnya lebih unggul dalam kecepatan dan akurasi untuk data teks.

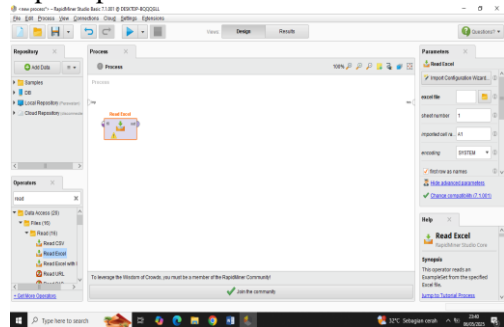
K-NN lebih sensitif terhadap ukuran dataset dan parameter K makanya akurasiya rendah.

#### Relevansi Tata Kelola TI:

Efisiensi komputasi Naive Bayes cocok untuk sistem TI yang memerlukan pemrosesan real-time.

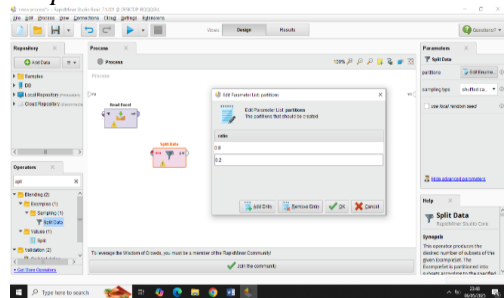
K-NN membutuhkan sumber daya lebih besar untuk dataset besar.

dilakukan pertama kali adalah *read data*. *Read Data* dibuat dalam bentuk bagan seperti pada Gambar 3.



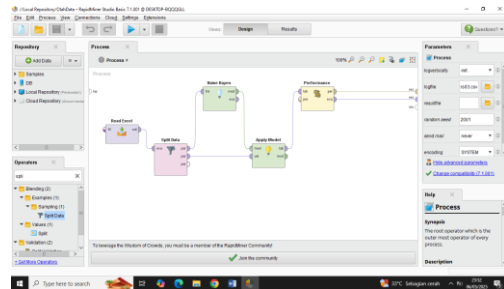
Gambar 3 Read Data

*Read data* diproses dalam bentuk .xls terdiri atas kolom nama pengguna, tanggal, skor, komentar dan konten. Pada kolom content di labelin karena menunjukkan nilai kepuasan pengguna *shopee*.



Gambar 4 Split Data

Gambar 4 menunjukkan bahwa bagan *Split Data* dibuat yang bertujuan membagi data menjadi dua (2) bagian yaitu data latih (*training set*) dan data uji (*testing set*). Jumlah pembagian data yaitu 80% untuk *training set* dan 20% untuk *testing set*. Data latih harus lebih besar dari data uji karena agar membentuk modul yang baik dengan meningkatkan keakurasian data. 80% ditunjukkan dengan 0.8 dan 20% ditunjukkan dengan 0.2.



Gambar 5 Implementasi Naive Bayes

## HASIL DAN PEMBAHASAN

Penelitian ini membandingkan metode KNN dan Naive Bayes dalam menganalisis tingkat kepuasan pengguna aplikasi *shopee* yang bertujuan mengetahui tingkat kepuasan pengguna dari data skor pada *playstore* / *appstore*. Data diambil menggunakan *web scrapping* sejumlah 2000 data. Pengolahan data menggunakan bantuan *tools* yaitu *rapidminer*.

Gambar 2 Dataset penelitian

Gambar 2 merupakan dataset dari penelitian yang diambil dari penilaian aplikasi *shopee* melalui *playstore* menggunakan teknik *scrap web*. Terdapat 3 kolom yaitu *date*, *komentar* dan *content*. *Content* merupakan penilaian dari pengguna terhadap penggunaan aplikasi *shopee* maka *Content* menjadi kolom label yang artinya *output* analisis metode Naive Bayes dan KNN.

### Naive Bayes

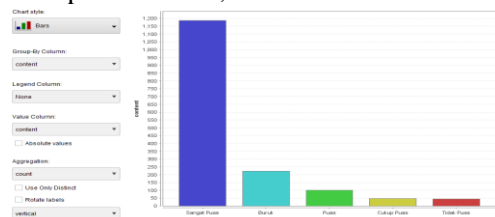
Untuk mengimplementasikan Naive Bayes pada *rapidminer* hal yang

Setelah bagan *Split Data* dibuat maka selanjutnya adalah implementasi metode Naïve Bayes yang menggunakan bagan pada rapidminer serta bagan untuk model yang dihasilkan pada *training set* dan bagan hasil dari olah model yang dihasilkan ditunjukkan pada gambar 5.

Confidence_Sangat Puas	Real	Min	Max	Average
confidence(Sangat Puas)	0	0	1	0.742
Confidence_Buruk	Real	Min	Max	Average
confidence(Buruk)	0	0	1	0.139
Confidence_Tidak Puas	Real	Min	Max	Average
confidence(Tidak Puas)	0	0	1	0.028
Confidence_Puas	Real	Min	Max	Average
confidence(Puas)	0	0	1	0.062
Confidence_Cukup Puas	Real	Min	Max	Average
confidence(Cukup Puas)	0	0	1	0.029

Gambar 6 Stastik Data Naïve Bayes

Gambar 6 menunjukkan dari 1600 jumlah data sebagai data latih (*training set*) yaitu 80% dari total keseluruhan data 2000. Nilai *confidence* dari masing-masing kategori penilaian yaitu: Sangat Puas 0.742, Buruk 0.139, Puas 0.062, Cukup Puas 0.029, dan Tidak Puas 0.028



Gambar 7 Chart Data Naïve Bayes

Kemudian hasil tersebut dapat dilihat dalam bentuk grafik pada gambar 7 dengan perolehan nilai sangat puas mencapai 1600-an warna batang grafik biru disusul penilaian buruk dengan gambar batang grafik warna hijau. Walaupun ada penialain buruk tetapi penilaian sangat puas yang paling tinggi.

Performance Vector (Performance)  
 Result not stored in repository.

PerformanceVector:  
 accuracy: 100.00%

ConfusionMatrix:  
 True: Sangat Puas    Buruk    Tidak Puas    Puas    Cukup Pu

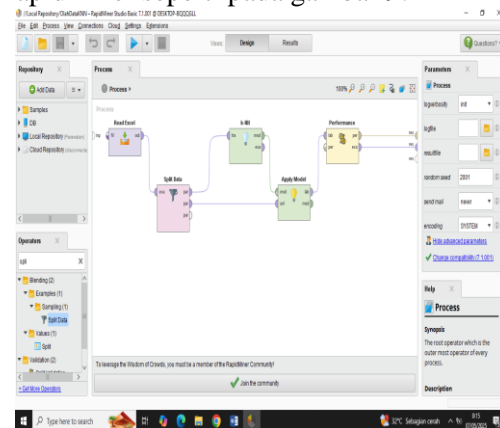
Sangat Puas:	1187	0	0	0	0
Buruk:	0	233	0	0	0
Tidak Puas:	0	0	44	0	0
Puas:	0	0	0	100	0
Cukup Puas:	0	0	0	0	46

Gambar 8 Performa Vektor Naïve Bayes

*Performance Vector* yang dihasilkan mencapai 100% dapat dilihat pada gambar 8. Hal ini menunjukkan penggunaan model Naïve Bayes pada studi kasus analisis kepuasan pengguna *shopee* dapat digunakan dengan sangat baik dengan nilai akurasi yang penuh.

## KNN

Implementasi KNN pada rapidminer hampir sama dengan implementasi Naïve Bayes. Dimulai dari *read data* (gambar 3) kemudian dilanjutkan dengan *split data* yang pembagian data latih (*training set*) dan data uji (*testing set*) adalah 80% dan 20% ditulis dengan 0.8 dan 0.2 (gambar 4). Hal yang berbeda adalah gambar bagan model yang memilih metode KNN pada rapidminer seperti pada gambar 9.



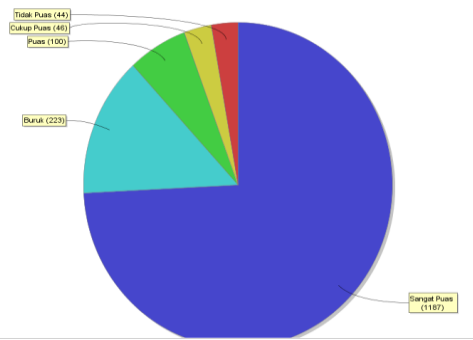
Gambar 9 Bagan KNN pada Rapidminer

Bagan metode KNN pada rapidminer sudah terbentuk maka selanjutnya menjalankan program dari 2000 data sesuai dengan metode yang diinginkan untuk pengolahan data sehingga menghasilkan model yang baik.

Confidence_Sangat Puas	Real	Min	Max	Average
confidence(Sangat Puas)	0	0	1	0.771
Confidence_Buruk	Real	Min	Max	Average
confidence(Buruk)	0	0	1	0.111
Confidence_Tidak Puas	Real	Min	Max	Average
confidence(Tidak Puas)	0	0	1	0.031
Confidence_Puas	Real	Min	Max	Average
confidence(Puas)	0	0	1	0.063
Confidence_Cukup Puas	Real	Min	Max	Average
confidence(Cukup Puas)	0	0	1	0.024

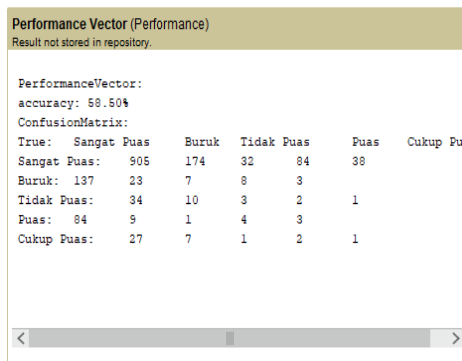
Gambar 10 Statistik Data KNN

Pengolahan data pada rapidminer dengan algoritma KNN menghasilkan statistik data, grafik dan performa vektor yang berbeda dari Naïve Bayes. Gambar 10 menunjukkan Sangat Puas mempunyai nilai 0.771, Buruk mempunyai nilai 0.111, Puas mempunyai nilai 0.063, 0.031 untuk Tidak Puas dan 0.024 untuk Cukup Puas.



**Gambar 11 Grafik Data KNN**

Gambar 11 menunjukkan Grafik untuk Sangat Puas mempunyai 1187 kemudian Buruk mempunyai nilai 223, 100 untuk Puas, 46 untuk Cukup Puas dan 44 untuk Tidak Puas. Warna biru menunjukkan bahwa pengguna aplikasi *shopee* sangat puas dalam penggunaannya. Warna biru hampir mendominasi dari grafik. Terlihat antusiasme masyarakat pengguna *shopee* sangat banyak dan merasa terlayani dengan baik.



**Gambar 12 Performa Vektor Data pada KNN**

Nilai Performa Vektor menunjukkan keakurasian 58.50% pada Gambar 12, keakurasian yang sangat berbeda sekali pada Naïve Bayes. Hal ini disebabkan KNN tidak cocok pada model data berupa penilaian kata sedangkan

Naïve Bayes sangat cocok (Alfaris, S., 2023). Sangat Puas, Puas, Cukup Puas, Tidak Puas dan Buruk merupakan penilaian kata dari level tingkat kepuasan pengguna aplikasi *shopee*

**SIMPULAN**

Perbandingan Naïve Bayes dengan KNN dalam analisis tingkat kepuasan pengguna aplikasi *shopee* mempunyai nilai keakurasian yang berbeda. Naïve Bayes mempunyai keakurasian yang tembus hingga 100% sedangkan KNN mempunyai nilai keakurasian hanya 58.50%. maka dapat disimpulkan penggunaan Naïve Bayes pada kasus ini masih lebih baik dari pada KNN sehingga pengguna masih sangat baik dalam tingkat kepuasan menggunakan aplikasi *shopee*. Hal ini dapat memicu aplikasi sejenis untuk meningkatkan layanan dan tata kelola IT demi tercapainya performa aplikasi yang baik.

Penerapan tata kelola teknologi informasi (TI) yang efektif menuntut pemanfaatan sumber daya komputasi secara optimal, presisi dalam analisis data, dan kecepatan pengolahan informasi. Meskipun algoritma K-NN relatif mudah diimplementasikan, kebutuhan komputasinya dapat meningkat signifikan saat menghadapi dataset berskala besar. Di sisi lain, Naive Bayes dikenal efisien dalam kecepatan pemrosesan, tetapi keakuratannya bergantung pada asumsi independensi antar-fitur yang seringkali tidak realistis dalam data dunia nyata. Dengan membandingkan performa kedua algoritma ini dalam konteks analisis tingkat kepuasan pengguna *Shopee*, penelitian ini bertujuan merumuskan rekomendasi strategis untuk model tata kelola TI yang unggul, terutama dalam pemilihan metode analisis data pengguna yang tepat. Lebih lanjut, hasil penelitian ini dapat dijadikan pedoman bagi profesional TI dalam merancang kebijakan pengelolaan data yang sistematis, sehingga memperkuat praktik pengambilan keputusan berbasis data

(data-driven). Dampaknya, studi ini tidak hanya mendorong peningkatan kualitas layanan aplikasi Shopee, tetapi juga memperkaya kerangka tata kelola TI melalui solusi komputasi yang lebih adaptif, terukur, dan berorientasi pada efisiensi.

## DAFTAR PUSTAKA

- Alfaris, S. (2023). Komparasi Metode KNN dan Naive Bayes Terhadap Analisis Tingkat kepuasan Pengguna Aplikasi Shopee. *The Indonesian Journal of Computer Science*, 12(5).
- Daryanti, P., & Shihab, M. S. (2019). Analisis Tingkat Kepuasan Pelanggan Dengan Menggunakan Konsep E-Servqual (Studi Kasus Pelanggan Shopee). *Journal of Entrepreneurship, Management and Industry (JEMI)*, 2(3), 120-127.
- Hasan Putra, P., Syahputra Novelan, M., & Rizki, M. (2022). ANALYSIS K-NEAREST NEIGHBOR METHOD IN CLASSIFICATION OF VEGETABLE QUALITY BASED ON COLOR. In *Journal of Applied Engineering and Technological Science* (Vol. 3, Issue 2).
- Hidayatuloh, S., & Aziati, Y. (2020). Analisis pengaruh user experience terhadap kepuasan pengguna mobile application e-commerce shopee menggunakan model delone & mclean. *Jurnal ilmiah teknik informatika (TEKINFO)*, 21(1), 73-83.
- Haekal, B. V., Ernawati, I., & Chamidah, N. (2021). Klasifikasi Kepuasan Pengguna Layanan Aplikasi Shopee Menggunakan Metode Decision Tree C4. 5. *Informatik: Jurnal Ilmu Komputer*, 17(3), 188-196.
- Huda, N., Habrizons, F., Satriawan, A., Iranda, M., & Pramuda, T. (2023). Analisis Usability Testing Menggunakan Metode SUS (System Usability Scale) Terhadap Kepuasan Pengguna Aplikasi Shopee. *Jurnal Sistem Informasi dan Sistem Komputer*, 8(2), 208-220.
- Li, F. (2023). Analisis Kepuasan Pengguna Terhadap User Interface Aplikasi E-Commerce Shopee Menggunakan Metode EUCS di Jakarta Barat.
- Margaretha, F. (2017). Analisis Hubungan Antara Motif Dengan Tingkat Kepuasan Pengguna Aplikasi Shopee Sebagai Media Berbelanja Online PAD SHOPEEHOLIC Di Kota Samarinda. *EJournal Ilmu Komunikasi*, 5(4), 26-40.
- Muslim, S. N. S., Nurdiansyah, F., & Rahman, A. Y. (2024). PERBANDINGAN ALGORITMA NAIVE BAYES DAN KNN DALAM ANALISIS TINGKAT KEPUASAN ULASAN PENGGUNA APLIKASI CAPCUT. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(3S1).
- N. Tri Romadloni, etc. Ilmu Komputer STMIK Nusa Mandiri Jakarta, "Perbandingan Metode Naive Bayes, Knn Dan Decision Tree Terhadap Analisis Tingkat kepuasan Transportasi Krl Commuter Line," *J. IKRA-ITH Inform.*, vol. 3, no. 2, pp. 1–9
- Novelan, M. S., Efendi, S., Sihombing, P., & Mawengkang, H. (2023). VEHICLE ROUTING PROBLEM OPTIMIZATION WITH MACHINE LEARNING IN IMBALANCED CLASSIFICATION VEHICLE ROUTE DATA. *Eastern-European Journal of Enterprise Technologies*, 5(3(125)), 49–56. <https://doi.org/10.15587/1729-4061.2023.288280>
- Sabila, A. R., & Kusumaningrum, L. (2020). Analisis kualitas layanan e-commerce SHOPEE dalam meningkatkan kepercayaan dan kepuasan pelanggan (studi kasus: pelanggan shopee di kota tangerang 2020). *Ji-Tech*, 72-80.
- Salim, A., & Alfando, S. (2024). Analisis Tingkat Kepuasan Pelanggan Menggunakan Shopee Paylater. *Jurnal Mahasiswa Eka Prasetya*, 1(1),

21-32.  
Saidani, B., Lusiana, L. M., & Aditya, S.  
(2019). Analisis pengaruh kualitas  
website dan kepercayaan terhadap

kepuasan pelanggan dalam  
membentuk minat pembelian ulang  
pada pelanggan shopee. *Jurnal Riset  
Manajemen Sains Indonesia*, 10(2).