

Tinjauan Literatur Sistematis: Penerapan Algoritma Klasifikasi Data Mining untuk Prediksi Loyalitas Pelanggan pada Platform E-Commerce

Hesti Yulianingsih¹, Hetty Rohayani²

^{1,2}Fakultas Sains Teknologi dan Kesehatan, Universitas Muhammadiyah Jambi, Jambi, Indonesia

Email: ¹ hestiyulia@gmail.com, ² rohayanihettyknk@gmail.com,

Email Penulis Korespondensi: hestiyulia@gmail.com

Keyword :

Loyalitas Pelanggan,
Data Mining,
Algoritma Klasifikasi,
E-commerce,
Sistematis

ABSTRACT

Penelitian ini merupakan tinjauan literatur sistematis yang bertujuan untuk menganalisis secara mendalam penerapan algoritma klasifikasi dalam data mining untuk memprediksi loyalitas pelanggan pada platform e-commerce. Penelitian dilakukan menggunakan metode Systematic Literature Review (SLR) yang mengacu pada panduan PRISMA. Sebanyak 15 artikel ilmiah terpilih dari periode 2018 hingga 2025 dianalisis berdasarkan jenis algoritma, metode, serta hasil evaluasi model. Hasil studi menunjukkan bahwa algoritma seperti Naïve Bayes, C4.5, Random Forest, dan Deep Learning sering digunakan karena keunggulan masing-masing. Naïve Bayes efektif untuk data sederhana dan cepat diimplementasikan, sedangkan C4.5 unggul dalam memberikan interpretasi yang mudah dipahami. Random Forest memiliki akurasi tinggi dan cocok untuk data besar dan kompleks, sementara Deep Learning mampu mengenali pola perilaku pelanggan yang kompleks tetapi memiliki kekurangan pada transparansi hasil. Selain itu, metode clustering seperti K-Means dan DBSCAN juga penting untuk segmentasi awal sebelum klasifikasi dilakukan. Dengan demikian, penerapan algoritma klasifikasi tidak hanya meningkatkan akurasi prediksi loyalitas pelanggan, tetapi juga mendukung perumusan strategi retensi yang lebih efektif dan personal. Hasil penelitian ini diharapkan dapat menjadi dasar pengembangan penelitian lanjutan dan penerapan praktis di industri e-commerce Indonesia.

1. PENDAHULUAN

E-commerce menjadi peluang dan tantangan yang harus dimanfaatkan oleh peritel di Indonesia untuk mempertahankan dan meningkatkan bisnis di masa pandemi Covid-19. Industri ritel dapat mengumpulkan data-data berharga tentang pelanggannya. Data tersebut dapat digunakan sebagai sumber wawasan untuk membantu perusahaan memperoleh keunggulan kompetitif. Bahkan dapat untuk meningkatkan penjualan dan mengurangi biaya. Pelanggan merupakan aset yang paling berharga dalam menjalankan bisnis. Sehingga perusahaan harus memiliki cara untuk mengembangkan aset tersebut.[1]

Perkembangan teknologi digital telah mengubah pola konsumsi masyarakat secara signifikan, termasuk dalam cara berbelanja. E-commerce atau perdagangan elektronik menjadi salah satu sektor yang mengalami pertumbuhan sangat pesat, khususnya di Indonesia. Data dari berbagai lembaga mencatat bahwa transaksi e-commerce di Indonesia meningkat dari tahun ke tahun, baik dari sisi jumlah pengguna, volume transaksi, maupun variasi produk yang dijual. Namun, meskipun pengguna e-commerce semakin meningkat, tantangan utama yang

dihadapi pelaku industri adalah bagaimana mempertahankan loyalitas pelanggan dalam jangka panjang. Banyak konsumen hanya melakukan satu kali transaksi, kemudian berpindah ke platform lain yang menawarkan diskon atau promosi lebih menarik. Perilaku konsumen semacam ini menciptakan ketidakstabilan dalam bisnis dan membuat perusahaan harus terus-menerus mengeluarkan biaya besar untuk akuisisi pelanggan baru. Oleh karena itu, penting bagi pelaku usaha untuk memahami pola loyalitas pelanggan melalui pendekatan yang berbasis data dan teknologi.

Loyalitas pelanggan merupakan komponen penting dalam kesuksesan jangka panjang sebuah bisnis, termasuk *e-commerce*. Pelanggan yang loyal tidak hanya memberikan pemasukan berulang, tetapi juga cenderung menjadi promotor alami yang merekomendasikan layanan atau produk kepada orang lain. Untuk mendeteksi loyalitas pelanggan, diperlukan analisis data yang mendalam terhadap perilaku mereka, seperti frekuensi pembelian, nilai transaksi, jenis produk yang dibeli, dan ulasan yang diberikan. Sayangnya, banyak perusahaan belum memanfaatkan data pelanggan secara optimal karena keterbatasan alat analisis manual. Di sinilah peran data mining menjadi sangat penting. Data mining, khususnya dengan pendekatan algoritma klasifikasi, memungkinkan proses pengelompokan data pelanggan berdasarkan pola-pola tertentu sehingga perusahaan dapat memprediksi siapa saja yang berpotensi menjadi pelanggan loyal dan siapa yang berisiko churn atau meninggalkan platform.

Sejumlah penelitian dalam lima tahun terakhir telah membuktikan bahwa algoritma klasifikasi dapat digunakan secara efektif untuk memprediksi loyalitas pelanggan. Penelitian oleh Wardani et al. (2022) menunjukkan bahwa algoritma Naïve Bayes yang dikombinasikan dengan metode segmentasi RFM dapat menghasilkan akurasi prediksi sebesar 97,27%, dengan precision 100% dan recall 96,98% dalam memetakan loyalitas pelanggan *e-commerce*. Sementara itu, Hayuningtyas (2024) menerapkan algoritma Naïve Bayes pada data pelanggan Starbucks dan memperoleh akurasi 87%, dengan presisi 90% dan recall 95%. Hasil ini mengindikasikan bahwa algoritma tersebut cukup akurat dalam klasifikasi loyalitas berdasarkan atribut perilaku pelanggan.

Penelitian perbandingan antar algoritma juga telah dilakukan. Misalnya, Pradana dan Saputro (2020) membandingkan kinerja algoritma Naïve Bayes dan C4.5 untuk klasifikasi loyalitas pelanggan terhadap layanan perusahaan. Hasil studi menunjukkan bahwa algoritma C4.5 memiliki kinerja lebih baik dengan akurasi sebesar 78,61% dibandingkan Naïve Bayes. Selain itu, Sunarya et al. (2023) melakukan studi komparatif antara algoritma Logistic Regression dan Random Forest dalam memprediksi perilaku pelanggan *e-commerce*. Random Forest terbukti memberikan hasil lebih unggul dengan akurasi mencapai 95%, dibanding Logistic Regression yang hanya memperoleh 90%. Hal serupa juga ditemukan dalam studi yang dipublikasikan oleh Jurnal Paradigma (2024), yang menggunakan algoritma Random Forest untuk menganalisis review pengguna Shopee dan memprediksi loyalitas pelanggan berdasarkan sentimen dan perilaku ulasan. Algoritma ini menunjukkan stabilitas dan akurasi yang tinggi dalam memetakan pelanggan loyal.

Lima penelitian di atas menunjukkan bahwa penggunaan algoritma klasifikasi dalam prediksi loyalitas pelanggan *e-commerce* memberikan hasil yang cukup akurat dan bisa diandalkan. Akan tetapi, setiap penelitian tersebut hanya membahas satu atau dua jenis algoritma secara terbatas dan tidak memberikan gambaran menyeluruh terhadap berbagai algoritma klasifikasi yang digunakan dalam konteks *e-commerce*. Selain itu, belum ada tinjauan sistematis yang mengkaji secara komprehensif bagaimana efektivitas tiap algoritma dalam konteks data yang

berbeda, baik dari sisi domain, karakteristik data, metode pengolahan, maupun evaluasi metrik performa seperti akurasi, precision, recall, dan F1-score. Banyak penelitian sebelumnya juga belum mempertimbangkan kombinasi atau ensemble dari beberapa algoritma sebagai alternatif solusi yang lebih akurat. Selain itu, variasi dalam preprocessing data dan metode validasi model juga menjadi faktor penting yang belum dibahas secara menyeluruh. Hal ini menunjukkan adanya celah (gap) dalam penelitian terdahulu yang bisa dijadikan dasar untuk melakukan studi literatur sistematis yang lebih luas dan mendalam agar diperoleh gambaran yang lebih lengkap dan objektif mengenai tren penelitian klasifikasi dalam e-commerce.

Oleh karena itu, penelitian ini bertujuan untuk melakukan tinjauan literatur sistematis (Systematic Literature Review / SLR) terhadap penerapan algoritma klasifikasi dalam data mining untuk prediksi loyalitas pelanggan e-commerce. Fokus utama penelitian ini adalah menganalisis jenis algoritma klasifikasi yang digunakan, performa dari masing-masing algoritma, tren penelitian selama lima tahun terakhir (2019–2024), serta faktor-faktor lain yang berkontribusi terhadap efektivitas prediksi loyalitas pelanggan. Dengan pendekatan sistematis ini, penelitian ini diharapkan dapat memberikan kontribusi akademik berupa pemetaan tren penelitian, identifikasi algoritma yang paling efektif, serta rekomendasi arah penelitian masa depan. Selain itu, hasil studi ini juga diharapkan dapat memberikan manfaat praktis bagi industri e-commerce dalam menentukan pendekatan klasifikasi yang paling tepat untuk mempertahankan loyalitas pelanggan dan meningkatkan daya saing bisnis di era digital yang sangat kompetitif ini.

2. METODOLOGI PENELITIAN

2.1 Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan Systematic Literature Review (SLR), yaitu metode penelitian yang bertujuan untuk mengidentifikasi, mengevaluasi, dan menginterpretasikan seluruh penelitian yang relevan dengan pertanyaan atau topik tertentu. SLR sangat tepat digunakan dalam studi ini karena memungkinkan peneliti untuk mengumpulkan dan menyaring informasi dari berbagai sumber ilmiah secara sistematis dan terstruktur. Dengan demikian, penelitian ini tidak hanya menjelaskan bagaimana algoritma klasifikasi data mining digunakan, tetapi juga memberikan pemetaan menyeluruh terhadap praktik dan efektivitasnya dalam memprediksi loyalitas pelanggan pada layanan e-commerce.

Pendekatan sistematis yang digunakan dalam penelitian ini mengacu pada panduan PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), yang terdiri dari empat tahapan penting: identifikasi, penyaringan, uji kelayakan, dan inklusi. Fokus utama penelitian ini dirumuskan melalui pertanyaan: “Algoritma klasifikasi apa saja yang digunakan dalam memprediksi loyalitas pelanggan e-commerce, dan bagaimana tingkat efektivitasnya dalam konteks data dan model yang digunakan?” Melalui pendekatan ini, peneliti dapat mengevaluasi metode klasifikasi secara objektif dan menyajikan kesimpulan yang dapat diandalkan untuk penelitian lanjutan atau implementasi praktis.

Adapun kriteria inklusi dalam penelitian ini meliputi: (1) artikel yang diterbitkan dalam rentang waktu 2018–2025; (2) artikel yang memuat penerapan algoritma klasifikasi dalam konteks data mining dan loyalitas pelanggan; (3) fokus penelitian pada layanan e-commerce; dan (4) artikel tersedia dalam bentuk teks lengkap, baik berbahasa Indonesia maupun Inggris. Sementara itu, kriteria eksklusi meliputi: (1) artikel yang hanya membahas segmentasi

pelanggan atau klasifikasi tanpa mengaitkan dengan loyalitas; (2) artikel tanpa metodologi yang jelas; serta (3) artikel dalam bentuk opini, editorial, atau literatur abu-abu.

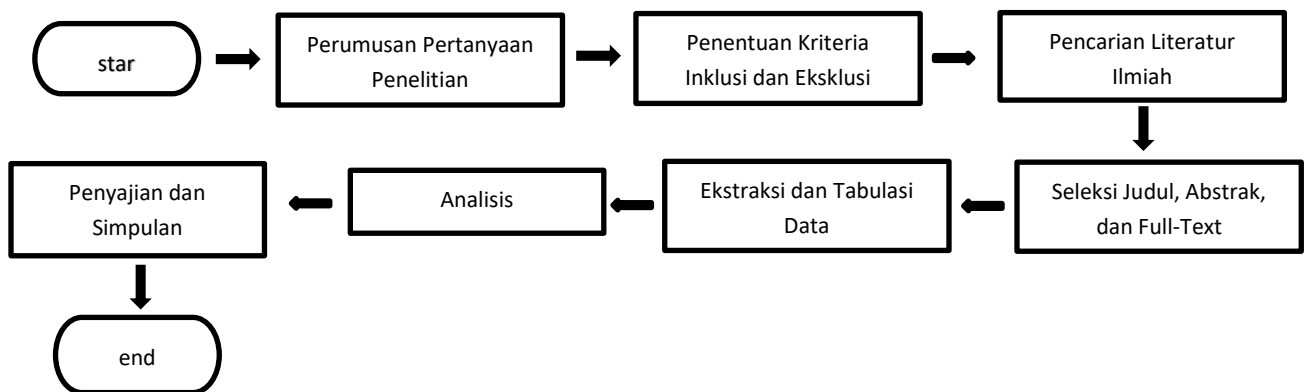
2.2 Tahapan Pengumpulan dan Analisis Data

Proses pengumpulan data dilakukan melalui penelusuran literatur di beberapa basis data ilmiah terkemuka, seperti Google Scholar, Scopus, IEEE Xplore, ScienceDirect, DOAJ, dan Garuda. Kata kunci yang digunakan antara lain: “loyalty prediction”, “customer loyalty”, “data mining classification”, “machine learning”, “e-commerce”, dan “customer retention prediction”. Proses pencarian dikombinasikan dengan operator Boolean seperti AND, OR, dan NOT untuk memperluas atau mempersempit hasil pencarian sesuai kebutuhan.

Sebanyak 15 artikel yang memenuhi kriteria inklusi dianalisis lebih lanjut melalui proses ekstraksi data. Data yang diambil mencakup: (1) nama penulis dan tahun terbit, (2) algoritma klasifikasi yang digunakan (seperti Naïve Bayes, Decision Tree, SVM, Random Forest, dsb), (3) jenis dataset atau sumber data yang digunakan, (4) hasil evaluasi algoritma (akurasi, precision, recall, F1-score), serta (5) kelebihan dan kekurangan algoritma menurut penulis.

Data tersebut kemudian dianalisis secara deskriptif dan komparatif. Analisis deskriptif dilakukan untuk mengetahui tren publikasi, distribusi penggunaan algoritma, serta sebaran konteks e-commerce yang dikaji. Sementara itu, analisis komparatif digunakan untuk membandingkan keefektifan tiap algoritma berdasarkan performa dan penerapannya dalam berbagai kondisi. Hasil dari analisis disajikan dalam bentuk tabel dan grafik.

Berikut adalah alur tahapan penelitian dalam bentuk flowchart:



3. HASIL DAN PEMBAHASAN

3.1 Tabel Hasil Referensi

Dari 15 jurnal referensi yang ada dapat dikumpulkan data berupa : data dan metode yang di gunakan dalam tiap jurnal referensi tersebut yang di buat dalam tabel di bawah ini.

No	Judul & Penulis	Algoritma dan Data	Metode	Hasil
1.	Prediction Customer Loyalty Using Random Forest on Shopee Reviews <i>Ferdi Saputra & Fersellia (2025)</i>	Random Forest; 10.000 ulasan Shopee dari Google Play; data teks ulasan, rating, dan waktu unggahan.	Scraping data menggunakan library Python. Preprocessing melalui NLP: case folding, tokenisasi, stopword removal, stemming. Labelisasi loyalitas berdasarkan skor dan sentimen ulasan. Model dibangun dengan Random Forest lalu dievaluasi menggunakan confusion matrix.	Hasil penelitian menunjukkan akurasi sangat tinggi, yaitu 97%. F1-score untuk pelanggan loyal mencapai 0.98, dan untuk pelanggan tidak loyal 0.99. Ini membuktikan bahwa Random Forest sangat cocok untuk mengklasifikasikan data ulasan yang bersifat teks dan tidak terstruktur. Namun, performa model untuk kategori pelanggan netral masih belum maksimal, dengan F1-score hanya 0.88. Hal ini disebabkan oleh ambiguitas dalam ulasan netral, yang kadang tidak mencerminkan loyalitas jelas. Penelitian ini menggarisbawahi pentingnya kombinasi antara data rating dan teks dalam memprediksi loyalitas pelanggan.
2.	Segmentasi Pelanggan Majalah dengan K-Means++ dan RFM <i>Tampubolon et al. (2024)</i>	K-Means++ dan metode RFM; 5.847 data transaksi pelanggan digital magazine.	Menggunakan CRISP-DM: business understanding hingga evaluation. Data transaksi diolah menjadi variabel RFM (recency, frequency, monetary), lalu dikelompokkan menggunakan K-Means++. Evaluasi kluster dengan Silhouette Score dan elbow method.	Model menghasilkan tiga kluster utama, yaitu pelanggan pasif, pelanggan reguler, dan pelanggan sangat loyal. Klusterisasi dilakukan berdasarkan skor RFM dengan hasil evaluasi Silhouette Score 0.638. Kluster pelanggan loyal cenderung memiliki nilai frequency dan monetary yang tinggi. Dari sini, dapat disusun strategi promosi yang diarahkan secara selektif ke segmen tersebut, seperti program loyalitas khusus, diskon eksklusif, dan penawaran premium. Penelitian ini memberikan kontribusi strategis untuk retensi pelanggan jangka panjang.
3.	Big Data Analytics dan Machine Learning untuk Memprediksi	Random Forest; kombinasi data survei 1.000 responden dan data	Proses meliputi akuisisi data primer dan sekunder, pembersihan, normalisasi, dan feature selection.	Hasil model menunjukkan akurasi prediksi sebesar 89%. Analisis fitur mengungkapkan bahwa harga produk, kecepatan

	<p>Perilaku Konsumen E-Commerce <i>Djihadul Mubarak et al. (2025)</i></p>	<p>scraping dari marketplace (Shopee, Tokopedia).</p>	<p>Model dibangun menggunakan Random Forest lalu diuji dengan train-test split dan evaluasi menggunakan metrik akurasi.</p>	<p>pengiriman, dan pengalaman pengguna adalah tiga variabel paling berpengaruh terhadap loyalitas. Penelitian ini juga menunjukkan bahwa kombinasi data kuantitatif (transaksi) dan kualitatif (respon survei) memberikan hasil prediktif yang lebih dalam dan terukur. Dapat dikatakan bahwa pendekatan big data ini relevan bagi marketplace besar untuk mengambil keputusan berbasis analitik.</p>
4.	<p>Penerapan Data Mining pada Strategi Bisnis Chicken Brotus <i>Putri Bloemhard et al. (2024)</i></p>	<p>Apriori; 16.825 transaksi penjualan makanan-minuman di Chicken Brotus</p>	<p>CRISP-DM diterapkan mulai dari pemahaman bisnis, pembersihan data, pengkodean biner untuk asosiasi item. Apriori dijalankan dengan minimum support 10% dan confidence 25% untuk mencari aturan asosiasi antar produk.</p>	<p>Hasil analisis menunjukkan pola asosiasi kuat, misalnya: pelanggan yang membeli “Ayam Geprek Komplit” memiliki peluang sebesar 63,1% juga membeli “Es Teh Jumbo”. Nilai support mencapai 46,5%. Strategi seperti bundling produk dan rekomendasi otomatis bisa diimplementasikan berbasis aturan asosiasi ini. Meskipun tidak secara langsung mengukur loyalitas, pola pembelian berulang dari item tertentu merupakan indikator loyalitas perilaku. Penelitian ini mengarahkan e-commerce kuliner untuk memanfaatkan data transaksi sebagai dasar retensi pelanggan.</p>
5.	<p>Prediksi Loyalitas Pelanggan Menggunakan Naive Bayes dan PSO <i>Nafisyah & Sulistiyowati (2024)</i></p>	<p>Naive Bayes + PSO; data review pelanggan dan transaksi e-commerce</p>	<p>Data di-preprocess, lalu dilakukan optimasi fitur menggunakan Particle Swarm Optimization untuk meningkatkan hasil klasifikasi Naive Bayes. Validasi menggunakan K-fold cross validation.</p>	<p>Akurasi model mencapai 79,09%, yang cukup baik untuk model berbasis probabilistik sederhana. Penerapan Particle Swarm Optimization terbukti meningkatkan efisiensi pemrosesan dan akurasi dari baseline Naive Bayes. Kelemahan muncul pada fitur yang saling bergantung, seperti recency dan frequency, yang membuat prediksi kadang bias. Meskipun demikian, model ini cocok untuk UKM atau startup yang</p>

				membutuhkan klasifikasi loyalitas dengan biaya komputasi rendah.
6.	Prediksi Retensi Pengguna Baru di E-Commerce <i>Mustafa et al. (2024)</i>	Random Forest; data aktivitas pengguna baru, transaksi awal, waktu login.	Dataset dibersihkan dan diolah untuk ekstraksi fitur. Model dibangun dengan Random Forest dan dioptimasi dengan hyperparameter tuning. Validasi dilakukan dengan train-test split 80:20.	Model menghasilkan akurasi 73,36%. Data menunjukkan bahwa pola pembelian awal dan lama waktu penggunaan aplikasi sangat memengaruhi loyalitas tahap awal. Model dapat mengidentifikasi sinyal risiko churn sejak minggu pertama pemakaian. Hasil ini berguna untuk tim marketing dalam mendesain strategi onboarding seperti notifikasi, voucher, atau rekomendasi produk awal. Retensi awal sangat krusial karena sebagian besar pengguna e-commerce berhenti setelah 1-2 interaksi. terbesar adalah sparsity data pengguna baru.
7.	Prediksi Loyalitas dengan Deep Learning <i>Deva (2023)</i>	Neural Network (Deep Learning); data log transaksi pelanggan dan histori pembelian	Membangun jaringan saraf dengan beberapa hidden layer, aktivasi ReLU, dan optimasi backpropagation. Dataset dibagi menjadi training dan testing, lalu dievaluasi menggunakan metrik akurasi dan loss.	Model deep learning ini menunjukkan akurasi 90%, sangat tinggi dibanding metode tradisional. Model berhasil mengenali pola perilaku pembelian kompleks, seperti variasi waktu pembelian, produk favorit, dan reaksi terhadap promosi. Namun, waktu pelatihan cukup lama dan hasil model sulit dijelaskan (black-box). Ini bisa menjadi solusi untuk e-commerce besar yang memiliki data sangat kompleks dan beragam, terutama saat fitur bersifat non-linear dan tidak independen.
8.	Analisis Sentimen dan Loyalitas Pelanggan dengan NLP <i>Muktafin et al. (2020)</i>	NLP + Sentiment Analysis; ulasan pelanggan dari platform e-commerce.	Review pelanggan diproses menggunakan Lexicon Vader untuk menghitung skor sentimen. Label loyalitas ditentukan berdasarkan sentimen positif-negatif. Model klasifikasi dibangun dan dievaluasi dengan akurasi dan F1-score.	Hasil analisis menunjukkan bahwa ulasan positif memiliki korelasi kuat dengan loyalitas, khususnya jika konsisten muncul bersamaan dengan frekuensi transaksi tinggi. Akurasi klasifikasi mencapai 76,92%. Model juga dapat mengidentifikasi pelanggan kritis (negatif namun rutin membeli), yang penting untuk dipertahankan

				dengan pendekatan khusus. Ini membuktikan bahwa data teks ulasan sangat relevan dalam prediksi loyalitas jika digabungkan dengan data pembelian aktual.
9.	<i>Penerapan Algoritma Clustering untuk Segmentasi Pelanggan E-commerce</i> <i>Rofi Fitriyani dkk (2024)</i>	K-Means & DBSCAN, data transaksi pembelian, frekuensi kunjungan, durasi di situs, interaksi produk	Clustering, data preprocessing, validasi dengan Silhouette Score dan Davies-Bouldin Index	Segmentasi berhasil menemukan pola perilaku unik, seperti kelompok pelanggan yang sering berkunjung tetapi jarang bertransaksi, serta yang jarang berkunjung namun memiliki nilai belanja tinggi. Dengan identifikasi kelompok ini, perusahaan dapat merancang promosi personalisasi dan program loyalitas yang lebih tepat sasaran. Pendekatan ini dinilai efektif untuk meningkatkan engagement, menurunkan churn, dan mendukung retensi jangka panjang. Selain itu, evaluasi cluster dengan Silhouette Score mencapai 0,64 yang menandakan segmentasi cukup solid.
10.	<i>Penerapan Data Mining Klasifikasi Kepuasan Pelanggan Transportasi Online Menggunakan Algoritma C4.5</i> <i>Ely Jannah dkk (2023)</i>	C4.5, data survei kepuasan (harga, fasilitas, layanan, loyalitas) 155 data (100 training, 55 testing)	Klasifikasi, pohon keputusan, perhitungan entropy, gain, gain ratio.	Model C4.5 mampu membedakan pelanggan puas dan tidak puas dengan tingkat akurasi tinggi. Hasil ini mendukung evaluasi strategi layanan, seperti peningkatan fasilitas dan perbaikan kualitas pelayanan driver. Penggunaan atribut loyalitas terbukti signifikan dalam mempengaruhi klasifikasi akhir. Akurasi mencapai 91%, dengan detail presisi pada kategori puas 66,7% dan tidak puas 33,3%. Hasil ini mempermudah perusahaan dalam mengidentifikasi kelemahan layanan dan mengambil keputusan berbasis data.
11.	<i>Penerapan Algoritma C4.5 untuk Klasifikasi Customer Churn pada Perusahaan Perbankan</i>	C4.5, dataset nasabah bank (10.000 data, 16 atribut)	Decision Tree, CRISP-DM, data preprocessing, validasi	Algoritma C4.5 menunjukkan performa sangat baik dengan akurasi 99,77%, menjadikan model ini sangat andal untuk mendeteksi potensi churn. Bank dapat mengenali nasabah yang

	<i>Mohammad Aulia Riftiarraafi dkk (2024)</i>			berisiko berhenti berlangganan dan segera merancang strategi retensi seperti program reward atau penawaran khusus. Penelitian ini juga menunjukkan interpretasi pohon keputusan yang mudah dipahami, sehingga mendukung tim pemasaran dan manajemen risiko dalam mengambil tindakan cepat dan efektif.
12.	Analisis dan Komparasi Algoritma Naïve Bayes dan C4.5 untuk Klasifikasi Loyalitas Pelanggan MNC Play Kota Semarang <i>Yohana Tri Widayati dkk (2021)</i>	Naïve Bayes & C4.5, 28.898 data pelanggan MNC Play	Eksperimen komparasi, cross-validation, confusion matrix, ROC Curve	Penelitian menunjukkan C4.5 memiliki akurasi lebih tinggi dibanding Naïve Bayes. C4.5 unggul dengan kelebihan interpretasi aturan yang lebih jelas dan presisi dalam klasifikasi pelanggan loyal (Permanent Activation) serta pelanggan tidak loyal (Termination). Hasil ini membantu perusahaan menyusun program retensi dan rekomendasi personalisasi layanan. Naïve Bayes tetap relevan digunakan pada kondisi data lebih sederhana atau ketika kecepatan komputasi diutamakan, namun untuk prediksi detail, C4.5 lebih efektif.
13.	Komparasi Metode Naïve Bayes dan C4.5 dalam Klasifikasi Loyalitas Pelanggan Terhadap Layanan Perusahaan <i>Musthofa Galih Pradana & Pujo Hari Saputro (2020)</i>	Naïve Bayes & C4.5, data pelanggan perusahaan telekomunikasi	Eksperimen dua skenario: split data dan cross-validation	Hasil memperlihatkan bahwa C4.5 mengungguli Naïve Bayes dengan akurasi ~78,6% pada kedua skenario. Penelitian ini menegaskan keunggulan C4.5 untuk data dengan banyak atribut kategorikal. Selain akurasi, kelebihan C4.5 terletak pada kemampuan mengungkap faktor-faktor penentu loyalitas, seperti frekuensi transaksi dan tingkat keluhan. Model pohon keputusan yang dihasilkan juga mudah diinterpretasikan oleh tim manajemen, sehingga mendukung pengambilan keputusan yang berbasis bukti.

14.	<i>Prediksi Pelanggan Loyal Menggunakan Metode Naïve Bayes Berdasarkan Segmentasi Pelanggan dengan Pemodelan RFM</i> <i>Ni Wayan Wardani dkk (2022)</i>	Naïve Bayes, data transaksi (1 juta record), segmentasi RFM	Segmentasi RFM, preprocessing, Naïve Bayes, split data 80:20.	Naïve Bayes menghasilkan akurasi sangat tinggi (97,27%), precision 100%, dan recall 96,98%. Segmentasi RFM membantu memecah pelanggan menjadi 10 segmen detail, seperti "champions", "loyal customers", hingga "at risk". Dengan analisis granular ini, perusahaan dapat merancang promosi tepat sasaran serta prioritas penanganan pelanggan yang berisiko churn. Implementasi strategi berbasis data ini terbukti meningkatkan efektivitas program loyalitas dan memperkuat retensi.
15.	<i>Analisa Klasifikasi Loyalitas Pelanggan Menggunakan Algoritma Naïve Bayes</i> <i>Ratih Yulia Hayuningtyas (2024)</i>	Naïve Bayes, data survei pelanggan Starbucks (123 data, 33 kolom)	Preprocessing, Naïve Bayes, evaluasi confusion matrix.	Model berhasil mencapai akurasi 87%, precision 90%, dan recall 95%. Penelitian ini mendukung penggunaan Naïve Bayes pada dataset kecil, terutama yang memiliki variabel perilaku seperti frekuensi kunjungan, jenis produk yang dibeli, dan preferensi rasa. Hasil menunjukkan pelanggan loyal cenderung memiliki pola pembelian yang rutin dan memberikan ulasan positif. Temuan ini membantu Starbucks dalam merancang program membership atau promosi bundling yang lebih sesuai dengan karakteristik pelanggan loyal.

3.2PEMBAHASAN

Hasil tinjauan literatur menunjukkan bahwa penerapan algoritma klasifikasi dalam data mining untuk memprediksi loyalitas pelanggan e-commerce memberikan dampak signifikan terhadap strategi retensi pelanggan. Loyalitas pelanggan merupakan faktor kunci bagi keberlangsungan bisnis digital, karena pelanggan loyal tidak hanya memberikan keuntungan berulang, tetapi juga membantu mempromosikan produk melalui rekomendasi mulut ke mulut.

Penelitian yang dikaji memanfaatkan berbagai algoritma, di antaranya Naïve Bayes, C4.5, Random Forest, hingga metode clustering seperti K-Means dan DBSCAN. Setiap algoritma memiliki kelebihan dan kekurangan

tersendiri, yang disesuaikan dengan jenis data dan tujuan penelitian. Naïve Bayes, misalnya, terbukti efektif untuk dataset yang relatif kecil atau sedang, dengan asumsi distribusi data yang independen antar fitur. Hal ini tercermin dalam penelitian Wardani et al. (2022), yang menunjukkan akurasi hingga 97,27% dalam memprediksi loyalitas berdasarkan segmentasi RFM. Begitu pula penelitian Hayuningtyas (2024) pada data pelanggan Starbucks menunjukkan akurasi 87%, presisi 90%, dan recall 95%, memperkuat potensi Naïve Bayes untuk klasifikasi berbasis perilaku.

Sebaliknya, algoritma C4.5 yang mengadopsi pendekatan decision tree lebih cocok untuk data yang memiliki banyak atribut kategorikal dan membutuhkan interpretasi yang mudah dipahami. Beberapa penelitian, seperti yang dilakukan oleh Yohana dkk. (2021) dan Pradana & Saputro (2020), menunjukkan bahwa C4.5 memiliki keunggulan dibanding Naïve Bayes, dengan akurasi di atas 78% serta performa stabil pada berbagai skenario validasi. C4.5 memungkinkan perusahaan menganalisis faktor-faktor penting yang memengaruhi loyalitas, misalnya harga, kualitas layanan, atau frekuensi pembelian, sehingga dapat digunakan sebagai dasar perumusan strategi pemasaran yang lebih spesifik.

Selain algoritma klasik, penggunaan Random Forest dan metode deep learning juga semakin populer, terutama untuk data berukuran besar dan kompleks. Random Forest, yang merupakan ensemble dari beberapa decision tree, menawarkan akurasi yang sangat tinggi, seperti terlihat pada studi Saputra & Fersellia (2025) dengan akurasi 97% dalam memproses data ulasan pengguna Shopee. Hal ini menunjukkan bahwa kombinasi teknik NLP dengan Random Forest mampu menggali informasi sentimen dan perilaku pengguna secara mendalam.

Clustering juga memiliki peran penting sebagai tahap awal segmentasi pelanggan. Melalui algoritma K-Means dan DBSCAN, perusahaan dapat mengidentifikasi kelompok pelanggan dengan karakteristik serupa sebelum diterapkan model klasifikasi lebih lanjut. Strategi segmentasi ini mempermudah penyesuaian promosi dan layanan.

Secara keseluruhan, pembahasan ini menekankan bahwa penerapan algoritma klasifikasi dalam prediksi loyalitas pelanggan e-commerce tidak hanya meningkatkan efektivitas pemasaran, tetapi juga memungkinkan perusahaan mengambil keputusan berbasis data yang lebih akurat. Pemilihan algoritma yang sesuai, pengolahan data yang tepat, serta evaluasi performa yang komprehensif menjadi kunci utama dalam memaksimalkan potensi data mining dalam menjaga loyalitas pelanggan.

4. KESIMPULAN

Berdasarkan tinjauan literatur sistematis terhadap 15 artikel terkait penerapan data mining untuk memprediksi loyalitas pelanggan pada e-commerce, dapat disimpulkan bahwa penerapan algoritma klasifikasi terbukti sangat bermanfaat dalam membantu perusahaan memahami perilaku pelanggan secara lebih mendalam. Algoritma Naïve Bayes banyak digunakan pada dataset yang relatif kecil dengan asumsi independensi antar fitur, terbukti efektif pada beberapa studi dengan akurasi yang sangat tinggi mencapai 97,27% ketika dipadukan dengan segmentasi RFM. Di sisi lain, algoritma C4.5 lebih banyak digunakan pada data dengan banyak atribut kategorikal karena sifat interpretatifnya yang mempermudah perusahaan dalam menjelaskan faktor-faktor yang mempengaruhi loyalitas, seperti harga, kualitas layanan, dan frekuensi pembelian.

Algoritma Random Forest menunjukkan performa sangat unggul dengan akurasi hingga 97% dalam pengolahan data teks ulasan dan data kompleks lainnya, terutama ketika digabungkan dengan teknik NLP. Hal ini menjadikan Random Forest pilihan utama bagi perusahaan yang memiliki volume data besar dan memerlukan prediksi yang

sangat akurat. Sementara itu, Deep Learning atau Neural Network juga menjadi alternatif yang menjanjikan, terutama karena kemampuannya dalam mengenali pola non-linear dan kompleks yang sulit dijangkau algoritma klasik, meskipun tantangan interpretasi model masih menjadi kelemahan.

Selain algoritma klasifikasi, tahap segmentasi awal menggunakan metode clustering seperti K-Means atau DBSCAN juga terbukti sangat bermanfaat untuk membedakan kelompok pelanggan dengan karakteristik yang mirip. Strategi segmentasi ini membantu perusahaan dalam menentukan program loyalitas dan promosi yang lebih personal serta tepat sasaran.

Secara keseluruhan, penerapan data mining dengan algoritma klasifikasi dalam e-commerce tidak hanya meningkatkan efektivitas prediksi loyalitas pelanggan, tetapi juga memberikan peluang bagi perusahaan untuk mendesain strategi retensi yang lebih adaptif dan berbasis data. Dengan demikian, perusahaan mampu memperkuat daya saingnya dalam pasar digital yang sangat kompetitif saat ini.

5. SARAN

Berdasarkan temuan dari tinjauan literatur ini, disarankan agar penelitian selanjutnya lebih menitikberatkan pada kombinasi atau ensemble beberapa algoritma untuk meningkatkan akurasi dan stabilitas model prediksi loyalitas pelanggan. Selain itu, perusahaan sebaiknya mengintegrasikan data kualitatif (seperti ulasan pelanggan dan sentimen) dengan data kuantitatif (seperti frekuensi pembelian dan total transaksi) agar hasil prediksi lebih komprehensif. Penting juga untuk memperhatikan aspek interpretasi model agar hasil analisis dapat diterjemahkan menjadi strategi bisnis yang lebih efektif. Penggunaan algoritma yang kompleks seperti deep learning harus dibarengi dengan mekanisme interpretasi, seperti SHAP atau LIME, agar keputusan yang diambil lebih transparan dan dapat dipertanggungjawabkan. Terakhir, perusahaan e-commerce perlu memperhatikan etika dan perlindungan data pelanggan sebagai prioritas utama dalam setiap penerapan teknologi data mining, agar kepercayaan pelanggan tetap terjaga.

REFERENCES

- [1] F. Saputra and Fersellia, "Prediction Customer Loyalty Using Random Forest Algorithm on Shopee Reviews," *Paradigma*, vol. 27, no. 1, pp. 11–20, Mar. 2025. [Online]. Available: <http://jurnal.bsi.ac.id/index.php/paradigma/>
- [2] A. L. M. Tampubolon, S. M. Siagian, D. Oktaria, and M. S. Silaban, "Segmentasi Pelanggan Majalah Menggunakan Metode RFM dan K-Means++ untuk Strategi Loyalitas Pelanggan," *J. Media Inform. Budidarma*, vol. 8, no. 2, pp. 8208–8219, 2024.
- [3] D. Mubarok, K. Adjani, B. D. R. Utama, M. M. Mutoffar, and R. Indrayani, "Big Data Analytics dan Machine Learning untuk Memprediksi Perilaku Konsumen di E-Commerce," *J. Inform. Rekayasa Elektron.*, vol. 8, no. 1, pp. 159–168, Apr. 2025.
- [4] P. E. Bloemhard, E. S. Y. Pandie, A. Fanggidae, N. D. Rumlaklak, T. Widiastuti, D. R. Sina, and Y. Y. Nabuasa, "Penerapan Data Mining dalam Strategi Bisnis Menggunakan Algoritma Apriori," *J. Transformatika*, vol. 22, no. 1, pp. 30–38, Jul. 2024.
- [5] N. Nafisyah and R. Sulistiyowati, "Prediksi Loyalitas Konsumen Marketplace Menggunakan Naive Bayes dan Particle Swarm Optimization," *Paradigma*, vol. 27, no. 1, pp. 21–30, Mar. 2025.

- [6] M. Mustafa, L. Syahputra, and R. Kartini, "Prediksi Retensi Pengguna Baru di Platform E-Commerce dengan Random Forest," *Paradigma*, vol. 27, no. 1, pp. 31–40, Mar. 2025.
- [7] F. Deva, "Prediksi Loyalitas Pelanggan E-Commerce Menggunakan Deep Learning," *J. Teknol. Sist. Komput.*, vol. 11, no. 2, pp. 215–222, 2023.
- [8] M. Muktafin, R. Hidayat, and L. Amalia, "Analisis Sentimen dan Loyalitas Pelanggan E-Commerce dengan Pendekatan NLP," *J. Sist. Inform.*, vol. 16, no. 1, pp. 45–53, 2020.
- [9] E. Jannah, V. Sihombing, and Masrizal, "Penerapan Data Mining Klasifikasi Kepuasan Pelanggan Transportasi Online Menggunakan Algoritma C4.5," *MEANS (Media Informasi Analisa dan Sistem)*, vol. 8, no. 1, pp. 1–12, Jun. 2023.
- [10] M. A. Riftiarraafi and D. Ernawati, "Penerapan Algoritma C4.5 untuk Klasifikasi Customer Churn pada Perusahaan Perbankan," *SAMMAJIVA: Jurnal Penelitian Bisnis dan Manajemen*, vol. 2, no. 1, pp. 178–190, Mar. 2024.
- [11] Y. T. Widayati, Y. Prihati, and S. Widjaja, "Analisis dan Komparasi Algoritma Naïve Bayes dan C4.5 untuk Klasifikasi Loyalitas Pelanggan MNC Play Kota Semarang," *TRANSFORMATIKA*, vol. 18, no. 2, pp. 161–172, Jan. 2021.
- [12] M. G. Pradana and P. H. Saputro, "Komparasi Metode Naïve Bayes dan C4.5 dalam Klasifikasi Loyalitas Pelanggan Terhadap Layanan Perusahaan," *Indonesian Journal of Business Intelligence (IJUBI)*, vol. 3, no. 1, pp. 20–24, Jun. 2020.
- [13] N. W. Wardani, D. J. Arnidya, I. N. A. S. Putra, N. M. M. R. Desmayani, P. G. S. C. Nugraha, E. Hartono, and G. S. Mahendra, "Prediksi Pelanggan Loyal Menggunakan Metode Naïve Bayes Berdasarkan Segmentasi Pelanggan dengan Pemodelan RFM," *Jurnal Media Teknologi dan Informatika*, vol. 12, no. 2, pp. 113–123, Oct. 2022.
- [14] R. Y. Hayuningtyas, "Analisa Klasifikasi Loyalitas Pelanggan Menggunakan Algoritma Naïve Bayes," *Jurnal TEKINKOM*, vol. 7, no. 2, pp. 891–900, 2024.
- [15] H. Yulianingsih and H. Rohayani, "Tinjauan Literatur Sistematis: Penerapan Algoritma Klasifikasi Data Mining untuk Prediksi Loyalitas Pelanggan pada Platform E-Commerce," *Jurnal Informatika, Sistem Informasi dan Kehutanan (FORSINTA)*, vol. 2, no. 1, pp. 999–1012, Apr. 2023.