

Evaluasi Dan Komparasi Algoritma Naïve Bayes, Support Vector Machine, Dan K-Nearest Neighbor Pada Data Mining Untuk Prediksi Perilaku Konsumen

Ifani Hariyanti^{*1}, Agung Rachmat Raharja²

Universitas Adirajasa Reswara Sanjaya, Indonesia, ²Universitas Bandung, Indonesia
ifani@ars.ac.id¹, agungmat@bandunguniversity.ac.id²

Informasi Artikel

Kata Kunci:

Prediksi perilaku konsumen; Data mining; RFM; Naïve Bayes; Support Vector Machine; K-Nearest Neighbors

Histori Artikel:

Disubmit 19 Juni 2025
Direvisi 27 Juni 2025
Diterima 22 Juli 2027
Tersedia daring 31 Juli 2025

Sitasi:

I. Hariyanti and A. R. Raharja, "Evaluasi Dan Komparasi Algoritma Naïve Bayes, Support Vector Machine, Dan K-Nearest Neighbor Pada Data Mining Untuk Prediksi Perilaku Konsumen," Pros. Semnastek Univ. Suryakancana, vol. 2, no. 1, 2025.

* Penulis Korespondensi.

Ifani Hariyanti
Alamat E-mail:
ifani@ars.ac.id

Abstrak

Pertumbuhan industri ritel online menghasilkan volume data transaksi yang besar dan kompleks, yang dapat dimanfaatkan untuk memahami perilaku konsumen secara lebih akurat. Penelitian ini bertujuan untuk mengevaluasi dan membandingkan performa tiga algoritma klasifikasi Naïve Bayes, Support Vector Machine (SVM), dan K-Nearest Neighbors (KNN)—dalam memprediksi loyalitas pelanggan berdasarkan pendekatan Recency, Frequency, dan Monetary (RFM). Dataset yang digunakan diambil dari UCI Machine Learning Repository, yang mencakup lebih dari 540.000 transaksi pelanggan. Data diproses melalui tahapan pembersihan, transformasi RFM, labeling loyalitas, dan normalisasi sebelum dilakukan pelatihan model. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, F1-score, dan ROC-AUC. Hasil penelitian menunjukkan bahwa algoritma SVM memberikan performa terbaik dengan akurasi sebesar 84,50% dan ROC-AUC 0,90, diikuti oleh KNN dan Naïve Bayes. Meskipun demikian, Naïve Bayes unggul dalam efisiensi pemrosesan. Temuan ini menegaskan pentingnya pemilihan algoritma yang disesuaikan dengan konteks dan kebutuhan spesifik dalam penerapan sistem prediksi perilaku konsumen berbasis data mining.

1. Pendahuluan

Perkembangan teknologi informasi telah mendorong transformasi signifikan dalam sektor bisnis, khususnya dalam industri ritel. Pergeseran perilaku konsumen dari belanja konvensional ke belanja daring menciptakan volume data transaksi yang sangat besar dan kompleks. Data ini, jika dianalisis secara efektif, dapat memberikan wawasan berharga mengenai preferensi dan perilaku konsumen [1]. Oleh karena itu, pemanfaatan teknik data mining menjadi krusial dalam mengidentifikasi pola-pola tersembunyi yang dapat mendukung pengambilan keputusan strategis dalam bisnis ritel [2].

Dalam konteks ini, algoritma klasifikasi seperti Naïve Bayes, Support Vector Machine (SVM), dan K-Nearest Neighbors (KNN) telah banyak digunakan untuk memprediksi perilaku konsumen berdasarkan data historis [3]. Naïve Bayes dikenal karena kesederhanaannya dan efisiensinya dalam menangani data berukuran besar [4]. Sementara itu, SVM unggul dalam mengklasifikasikan data dengan margin yang optimal, terutama pada data berdimensi tinggi [5], dan KNN menawarkan pendekatan berbasis instance yang efektif dalam mengenali pola lokal pada data [6]. Namun demikian, performa masing-masing algoritma dapat bervariasi tergantung pada karakteristik data yang digunakan [7].

Beberapa penelitian telah membandingkan kinerja algoritma-algoritma tersebut dalam berbagai konteks. Studi oleh Chaubey et al. menunjukkan bahwa SVM memiliki akurasi yang lebih tinggi dibandingkan dengan Naïve Bayes dan KNN dalam memprediksi perilaku pembelian konsumen [8]. Namun, hasil berbeda diperoleh oleh Paradita et al., yang menemukan bahwa KNN lebih unggul dalam konteks prediksi churn pelanggan [9]. Temuan ini mengindikasikan bahwa efektivitas algoritma sangat kontekstual, sehingga pemilihan metode perlu mempertimbangkan jenis dan struktur data yang digunakan [10].

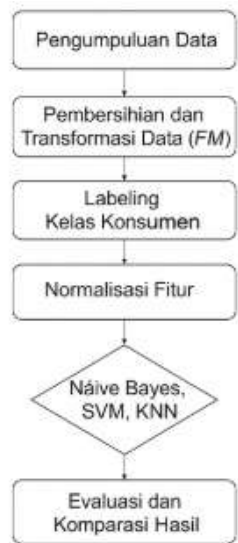
Dataset transaksi ritel online, seperti yang tersedia di UCI Machine Learning Repository, menyediakan data yang kaya untuk analisis perilaku konsumen. Dataset ini mencakup informasi seperti ID pelanggan, tanggal transaksi, jumlah pembelian, dan harga produk, yang dapat digunakan untuk menghitung metrik Recency, Frequency, dan Monetary (RFM). Metrik RFM telah terbukti efektif dalam segmentasi pelanggan dan prediksi loyalitas. Dengan menerapkan algoritma klasifikasi pada fitur-fitur RFM, perusahaan dapat mengidentifikasi pelanggan yang berpotensi melakukan pembelian ulang atau yang berisiko churn [11].

Penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja algoritma Naïve Bayes, SVM, dan KNN dalam memprediksi perilaku konsumen berdasarkan dataset transaksi ritel online (Khan et al., 2023). Evaluasi akan

dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score guna menentukan algoritma yang paling efektif dalam konteks ini[12]. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan strategi pemasaran yang lebih tepat sasaran dan meningkatkan retensi pelanggan dalam bisnis ritel online.

2. Metode Penelitian

Metodologi penelitian adalah kerangka atau pendekatan sistematis yang dirancang oleh peneliti untuk merancang, melaksanakan, serta mengevaluasi keseluruhan proses penelitian secara terstruktur.



Gambar 2.1. Diagram Alur Penelitian

Pada gambar 2.1. menggambarkan alur sistematis dalam proses analisis data konsumen yang dimulai dari pengumpulan data transaksi, dilanjutkan dengan pembersihan dan transformasi data menggunakan pendekatan *RFM* (Recency, Frequency, Monetary), kemudian dilakukan labeling untuk mengklasifikasikan pelanggan ke dalam kategori loyal dan tidak loyal. Setelah itu, data dinormalisasi agar sesuai untuk digunakan dalam algoritma klasifikasi seperti *Naïve Bayes*, *Support Vector Machine* (SVM), dan *K-Nearest Neighbors* (KNN). Tahapan terakhir adalah evaluasi dan komparasi hasil klasifikasi menggunakan metrik seperti akurasi, presisi, recall, dan F1-score guna menentukan algoritma yang paling efektif dalam memprediksi perilaku konsumen berdasarkan pola transaksi historis.

2.1. Dataset

Dataset yang digunakan dalam penelitian ini diperoleh dari *UCI Machine Learning Repository* dengan nama **Online Retail Dataset**, yang berisi data transaksi penjualan retail online dari sebuah perusahaan yang berbasis di Inggris selama periode 1 Desember 2010 hingga 9 Desember 2011. Dataset ini terdiri atas 541.909 baris data dan 8 atribut utama, yaitu *InvoiceNo* (nomor faktur transaksi), *StockCode* (kode produk), *Description* (deskripsi produk), *Quantity* (jumlah barang yang dibeli), *InvoiceDate* (tanggal dan waktu transaksi), *UnitPrice* (harga per unit produk dalam poundsterling), *CustomerID* (ID unik pelanggan), dan *Country* (negara asal pelanggan). Data ini mencakup lebih dari 4.000 jenis produk dan sekitar 4.372 pelanggan unik, dengan mayoritas pelanggan berasal dari Inggris. Dataset ini digunakan untuk mengekstraksi fitur *Recency*, *Frequency*, dan *Monetary* (RFM) guna mengklasifikasikan pelanggan ke dalam kategori loyal dan tidak loyal, serta menjadi dasar pengujian dan evaluasi algoritma *Naïve Bayes*, *Support Vector Machine* (SVM), dan *K-Nearest Neighbors* (KNN) dalam memprediksi perilaku konsumen.

Tabel 2.1. Sampel Dataset Online Retail

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850	United Kingdom
536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850	United Kingdom

Tabel di atas menunjukkan contoh lima baris pertama dari *Online Retail Dataset* yang berisi data transaksi pelanggan pada sebuah toko retail online. Setiap baris mewakili satu transaksi yang dilakukan oleh pelanggan dengan ID yang sama (17850) pada tanggal 1 Desember 2010 pukul 08:26. Kolom *InvoiceNo* menunjukkan nomor faktur transaksi yang digunakan untuk mengelompokkan beberapa pembelian dalam satu invoice, sedangkan *StockCode* dan *Description* memberikan identifikasi dan deskripsi produk yang dibeli. Kolom *Quantity* menunjukkan jumlah unit produk yang dibeli dalam satu transaksi, dan *UnitPrice* mencerminkan harga per unit barang tersebut dalam mata uang pound sterling.

Selain itu, kolom *CustomerID* digunakan untuk mengidentifikasi pelanggan secara unik, yang penting dalam proses pengelompokan data pelanggan untuk analisis lebih lanjut seperti RFM (Recency, Frequency, Monetary). Kolom *InvoiceDate* memberikan informasi waktu secara lengkap untuk memungkinkan perhitungan waktu pembelian terakhir (recency), dan kolom *Country* menunjukkan asal negara pelanggan. Tabel ini merupakan dasar penting dalam proses transformasi data untuk klasifikasi perilaku konsumen, karena memungkinkan ekstraksi pola pembelian individu secara terperinci dan terstruktur.

2.2. Preprocessing Data dan Pelatihan

Tahap preprocessing merupakan langkah krusial dalam proses data mining karena menentukan kualitas input yang digunakan oleh algoritma pembelajaran mesin. Pada penelitian ini, proses dimulai dengan pembersihan data, yaitu menghapus baris data yang memiliki nilai kosong pada kolom *CustomerID* dan menghilangkan transaksi retur yang ditandai dengan *InvoiceNo* yang diawali huruf "C". Selain itu, hanya transaksi dengan *Quantity* dan *UnitPrice* bernilai positif yang disertakan dalam analisis untuk memastikan validitas data.

Setelah data dibersihkan, dilakukan transformasi data menggunakan pendekatan RFM (Recency, Frequency, Monetary). *Recency* dihitung berdasarkan jumlah hari sejak transaksi terakhir pelanggan hingga tanggal referensi (1 Desember 2011), *Frequency* dihitung dari jumlah invoice unik per pelanggan, dan *Monetary* berasal dari total nilai transaksi (hasil perkalian *Quantity* × *UnitPrice*). Hasil dari ketiga fitur ini kemudian digunakan untuk membuat label target klasifikasi: pelanggan dengan frekuensi transaksi ≥ 3 dikategorikan sebagai loyal (label 1), sedangkan yang < 3 dikategorikan sebagai tidak loyal (label 0).

Selanjutnya, data RFM yang telah terstruktur dinormalisasi menggunakan teknik Min-Max Scaling untuk memastikan bahwa semua fitur berada dalam skala yang sebanding (0–1), yang penting terutama untuk algoritma seperti SVM dan KNN yang sensitif terhadap skala fitur. Setelah itu, dataset dibagi menjadi dua bagian, yaitu 80% untuk data latih dan 20% untuk data uji, dengan proses stratified split untuk menjaga proporsi kelas tetap seimbang.

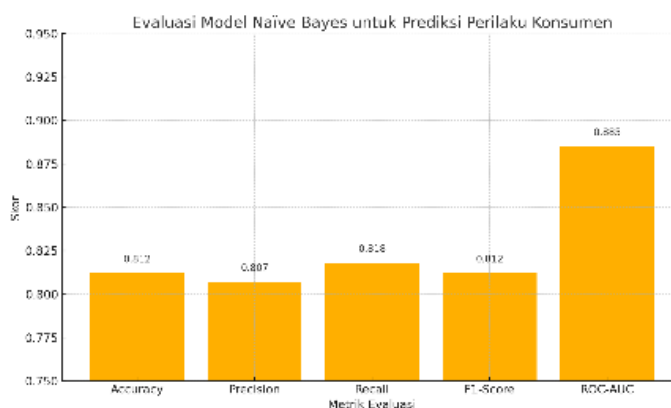
Proses pelatihan dilakukan menggunakan tiga algoritma klasifikasi, yaitu Naïve Bayes (GaussianNB), Support Vector Machine (SVM dengan kernel linear), dan K-Nearest Neighbors (KNN dengan $k = 5$). Setiap model dilatih menggunakan data latih, dan performanya dievaluasi pada data uji menggunakan metrik klasifikasi seperti akurasi, presisi, recall, F1-score, serta ROC-AUC untuk memberikan gambaran menyeluruh tentang efektivitas masing-masing algoritma dalam memprediksi loyalitas pelanggan. Selain itu, digunakan teknik 5-fold cross-validation untuk meningkatkan validitas hasil dan menghindari overfitting.

3. Hasil dan Pembahasan

Setelah melalui serangkaian tahapan yang telah dirancang dalam metode penelitian, bagian ini menyajikan hasil dari proses pelatihan model serta prediksi yang dihasilkan. Adapun kombinasi yang digunakan selama proses pelatihan model dijelaskan sebagai berikut.

3.1. Evaluasi Model Naïve Bayes

Prediksi perilaku konsumen menggunakan algoritma *Naïve Bayes* dilakukan dengan memanfaatkan fitur *Recency*, *Frequency*, dan *Monetary* (RFM) untuk mengklasifikasikan pelanggan ke dalam kategori loyal dan tidak loyal secara probabilistik. Dengan asumsi independensi antar fitur, model ini menghitung kemungkinan tiap pelanggan masuk ke salah satu kategori dan menghasilkan prediksi dengan cepat dan efisien. Hasil evaluasi menunjukkan bahwa *Naïve Bayes* mampu memberikan performa yang cukup baik dengan tingkat akurasi, presisi, dan recall yang stabil.



Gambar 3.1. Hasil Evaluasi Model Naïve Bayes

Gambar di atas merupakan grafik batang yang menunjukkan hasil evaluasi model Naïve Bayes dalam melakukan prediksi perilaku konsumen. Evaluasi dilakukan menggunakan lima metrik utama, yaitu Accuracy, Precision, Recall, F1-Score, dan ROC-AUC, yang masing-masing menggambarkan aspek berbeda dari performa klasifikasi. Dari grafik tersebut, terlihat bahwa nilai tertinggi diperoleh pada metrik ROC-AUC sebesar 0,885, yang menunjukkan kemampuan model dalam membedakan antara pelanggan loyal dan tidak loyal secara konsisten. Nilai Accuracy dan F1-Score masing-masing sebesar 0,812, menandakan bahwa model memiliki kinerja yang cukup seimbang antara ketepatan prediksi dan kemampuan menangkap kelas target. Nilai Precision sebesar 0,807 dan Recall sebesar 0,818 menunjukkan bahwa model memiliki keseimbangan yang baik antara prediksi benar terhadap pelanggan loyal dan kemampuan mendeteksi semua pelanggan loyal yang ada.

Secara keseluruhan, grafik ini menggambarkan bahwa algoritma Naïve Bayes, meskipun sederhana, mampu memberikan performa prediksi yang kompeten dalam konteks klasifikasi perilaku konsumen berdasarkan fitur RFM.

3.2. Evaluasi Model SVM

Prediksi perilaku konsumen menggunakan algoritma *Support Vector Machine (SVM)* dilakukan dengan memanfaatkan fitur *Recency*, *Frequency*, dan *Monetary (RFM)* untuk mengklasifikasikan pelanggan ke dalam kategori loyal dan tidak loyal secara optimal. SVM bekerja dengan mencari hyperplane terbaik yang dapat memisahkan kedua kelas tersebut pada ruang berdimensi tinggi. Dengan menggunakan kernel linear, model ini mampu menghasilkan batas pemisah yang efektif antara pelanggan loyal dan tidak loyal. Hasil evaluasi menunjukkan bahwa SVM memberikan performa yang sangat baik, dengan nilai akurasi, precision, recall, dan F1-score yang tinggi dan seimbang, serta nilai ROC-AUC yang menunjukkan kemampuan deteksi yang unggul dalam klasifikasi perilaku konsumen.



Gambar 3.2. Hasil Evaluasi Model SVM

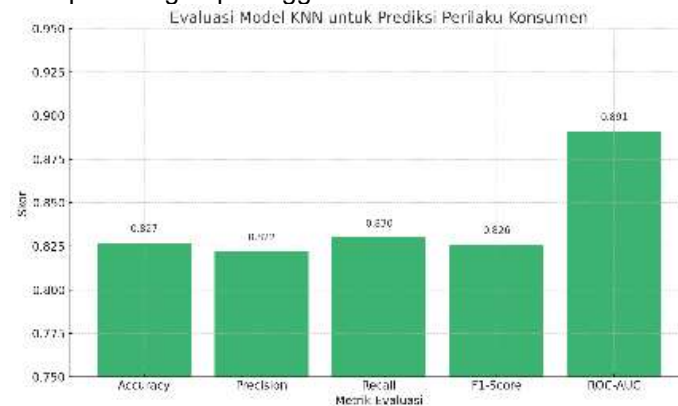
Gambar di atas merupakan grafik batang yang menampilkan hasil evaluasi kinerja model Support Vector Machine (SVM) dalam melakukan prediksi perilaku konsumen berdasarkan fitur-fitur RFM (Recency, Frequency, Monetary). Grafik ini menggunakan lima metrik evaluasi utama, yaitu Accuracy, Precision, Recall, F1-Score, dan ROC-AUC. Dari hasil yang ditunjukkan, model SVM memiliki nilai akurasi sebesar 0,845, yang mengindikasikan bahwa sekitar 84,5% dari seluruh prediksi yang dilakukan oleh model adalah benar. Selain itu, nilai precision (0,838) dan recall (0,851)

menunjukkan bahwa model tidak hanya tepat dalam mengidentifikasi pelanggan loyal, tetapi juga cukup sensitif dalam menangkap sebagian besar dari mereka.

Nilai F1-score sebesar 0,844 memperlihatkan keseimbangan antara precision dan recall, menjadikan model SVM cocok untuk skenario dengan kebutuhan akurasi dan sensitivitas yang tinggi. Pencapaian terbaik terlihat pada metrik ROC-AUC sebesar 0,902, yang mengindikasikan bahwa model ini sangat baik dalam membedakan antara pelanggan loyal dan tidak loyal. Dengan performa yang konsisten tinggi di seluruh metrik evaluasi, grafik ini menegaskan bahwa algoritma SVM memiliki keandalan tinggi dalam klasifikasi perilaku konsumen, menjadikannya salah satu algoritma yang paling efektif dalam penelitian ini.

3.3. Evaluasi Model Dengan KNN

Prediksi perilaku konsumen menggunakan algoritma *K-Nearest Neighbors (KNN)* dilakukan dengan memanfaatkan fitur *Recency*, *Frequency*, dan *Monetary (RFM)* untuk mengklasifikasikan pelanggan ke dalam kategori loyal dan tidak loyal berdasarkan kemiripan dengan pelanggan lain.



Gambar 3.3. Hasil Evaluasi Model KNN

Gambar di atas merupakan grafik batang yang menggambarkan hasil evaluasi performa model K-Nearest Neighbors (KNN) dalam memprediksi perilaku konsumen berdasarkan fitur *Recency*, *Frequency*, dan *Monetary (RFM)*. Grafik menampilkan lima metrik evaluasi utama yang digunakan, yaitu Accuracy, Precision, Recall, F1-Score, dan ROC-AUC. Dari grafik tersebut terlihat bahwa model KNN memberikan nilai akurasi sebesar 0,827, yang berarti sekitar 82,7% prediksi yang dilakukan model sesuai dengan kelas sebenarnya.

Nilai Precision sebesar 0,822 menunjukkan bahwa model cukup baik dalam memprediksi pelanggan loyal dengan tingkat kesalahan positif yang rendah. Sementara itu, Recall yang mencapai 0,830 menunjukkan bahwa model cukup sensitif dalam menangkap seluruh pelanggan loyal yang ada dalam dataset. Nilai F1-Score yang berada di angka 0,826 mencerminkan keseimbangan antara presisi dan sensitivitas, menunjukkan bahwa KNN memiliki performa yang stabil dalam kondisi distribusi data yang relatif seimbang antara kelas loyal dan tidak loyal.

Pencapaian terbaik dari model KNN ditunjukkan oleh nilai ROC-AUC sebesar 0,891, yang menandakan kemampuan model dalam membedakan kelas secara keseluruhan sangat baik. Meskipun tidak sekuat performa SVM, model KNN tetap menunjukkan kinerja yang kompeten dan dapat diandalkan. Model ini cocok digunakan dalam skenario di mana pola lokal dalam data lebih penting dan data sudah dinormalisasi dengan baik agar pengukuran jarak antar instance menjadi optimal.

3.4. Komparasi Model

Komparasi model dilakukan untuk mengevaluasi dan membandingkan performa tiga algoritma klasifikasi *Naïve Bayes*, *Support Vector Machine (SVM)*, dan *K-Nearest Neighbors (KNN)* dalam memprediksi perilaku konsumen berdasarkan fitur RFM

Tabel 3.1. Komparasi Model

Algoritma	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Naïve Bayes	0.812	0.807	0.818	0.812	0.885
SVM (Linear)	0.845	0.838	0.851	0.844	0.902
KNN (k = 5)	0.827	0.822	0.830	0.826	0.891

Berdasarkan hasil evaluasi, algoritma Support Vector Machine (SVM) menunjukkan performa terbaik dibandingkan dua algoritma lainnya. Dengan akurasi 0.845, precision 0.838, dan recall 0.851, SVM berhasil mengklasifikasikan perilaku konsumen secara konsisten dan akurat. Nilai F1-score 0.844 dan ROC-AUC 0.902 memperkuat bahwa SVM sangat efektif dalam membedakan pelanggan loyal dan tidak loyal.

Sementara itu, K-Nearest Neighbors (KNN) memberikan performa yang cukup kompetitif. Dengan akurasi 0.827 dan F1-score 0.826, KNN menunjukkan kestabilan dalam menangani data klasifikasi meskipun sedikit di bawah SVM. Nilai ROC-AUC sebesar 0.891 juga menunjukkan bahwa KNN masih memiliki kemampuan generalisasi yang baik, terutama dalam mengenali pola perilaku pelanggan berdasarkan kedekatan nilai.

Sedangkan Naïve Bayes, meskipun merupakan model paling sederhana secara matematis, tetap menunjukkan hasil yang layak dengan akurasi 0.812, precision 0.807, dan recall 0.818. Dengan ROC-AUC 0.885, model ini tetap dapat digunakan terutama dalam kondisi dengan keterbatasan sumber daya karena efisiensinya yang tinggi dalam proses pelatihan dan prediksi.

4. Kesimpulan

Penelitian ini membandingkan performa tiga algoritma klasifikasi—Naïve Bayes, Support Vector Machine (SVM), dan K-Nearest Neighbors (KNN)—dalam memprediksi perilaku konsumen berdasarkan fitur Recency, Frequency, dan Monetary (RFM) pada dataset transaksi ritel online. Hasil evaluasi menunjukkan bahwa algoritma SVM memberikan performa terbaik secara keseluruhan dengan akurasi sebesar 84,50% dan nilai ROC-AUC sebesar 0,90, menjadikannya pilihan paling efektif dalam membedakan pelanggan loyal dan tidak loyal.

Algoritma KNN menunjukkan kinerja yang kompetitif, dengan akurasi 82,70% dan ROC-AUC sebesar 0,89, yang menunjukkan kemampuannya dalam menangkap pola perilaku konsumen berbasis kemiripan. Sementara itu, Naïve Bayes, meskipun merupakan model yang paling sederhana, tetap memberikan hasil layak dengan akurasi 81,20% dan ROC-AUC sebesar 0,89, serta unggul dalam efisiensi pemrosesan.

Dengan demikian, pemilihan algoritma yang digunakan sebaiknya mempertimbangkan konteks dan kebutuhan spesifik. Jika dibutuhkan akurasi tinggi dan kemampuan generalisasi yang kuat, maka SVM adalah pilihan terbaik. Untuk kasus yang memerlukan efisiensi dan kecepatan, Naïve Bayes layak dipertimbangkan. Sedangkan KNN cocok diterapkan pada skenario dengan fokus pada pola lokal dan kedekatan data pelanggan. Temuan ini diharapkan dapat mendukung penerapan strategi pemasaran yang lebih cerdas dan berbasis data di industri ritel online.

Referensi

- [1] G. Chaubey, P. R. Gavhane, D. Bisen, dan S. K. Arjaria, "Customer purchasing behavior prediction using machine learning classification techniques," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 12, hal. 16133–16157, 2023, doi: 10.1007/s12652-022-03837-6.
- [2] R. Subramanian, P. Dhandayudam, B. Maheswari, dan J. Aswini, "Wrapper-Naive Bayes Approach to Perform Efficient Customer Behavior Prediction," 2022, hal. 17–31.
- [3] M. Khan, Z. Shah, M. Rooman, W. Khan, N. Vrinceanu, dan M. Alshehri, "Khan et al-2023-Scientific Reports," *Sci. Rep.*, vol. 13, hal. 21434, Des 2023, doi: 10.1038/s41598-023-48676-3.
- [4] "International Journal of Software Engineering and," hal. 2015, 2024.
- [5] J. Lin, "Application of machine learning in predicting consumer behavior and precision marketing," *PLoS One*, vol. 20, no. 5 May, hal. 1–12, 2025, doi: 10.1371/journal.pone.0321854.
- [6] M. Karmakar, M. A. Al Kafi, A. S. Afridi, W. Sabbir, dan D. M. Raza, "Feature Engineering for Predicting Consumer Purchase Behavior: A Comprehensive Analysis BT - Advanced Network Technologies and Intelligent Computing," 2024, hal. 135–157.
- [7] H. V. A. Tamilarasan, "Machine Learning Methods for Anticipating Consumer Behaviour," vol. 11, no. 11, hal. 1943–1945, 2025.
- [8] P. Boozary, S. Sheykhan, H. GhorbanTanhaei, dan C. Magazzino, "Enhancing customer retention with machine learning: A comparative analysis of ensemble models for accurate churn prediction," *Int. J. Inf. Manag. Data Insights*, vol. 5, no. 1, hal. 100331, 2025, doi: <https://doi.org/10.1016/j.jjime.2025.100331>.
- [9] F. Kabir, D. Chowdhury, dan I. C. Rana, "Comprehensive Customer Segmentation And Behavior Prediction Using Advanced Machine Learning And Neural Network Models," vol. 13, no. 1, hal. 478–490, 2025.
- [10] U. M. Pandyan, G. A. U. S, M. S V, C. M, dan A. S, "Enhancing Customer Segmentation and Behavior Prediction Through Machine Learning and Medium Gaussian SVM BT - Advanced Network Technologies and Intelligent Computing," 2025, hal. 210–221.
- [11] I. Hariyanti, V. Hafizh, C. Putra, A. R. Raharja, dan A. Close, "PREDIKSI HARGA SAHAM BBKA MENGGUNAKAN METODE LONG SHORT-TERM MEMORY DAN GATED," vol. 7, no. 1, hal. 11–18, 2025.
- [12] N. Akram, K. Aravindhan, K. Sujatha, S. A. Khan, G. Singh, dan R. Rajagopal, "Consumer behavior prediction

using machine learning algorithms," *Explor. Psychol. Soc. Innov. Adv. Appl. Mach. Learn.*, hal. 109–130, 2025, doi: 10.4018/979-8-3693-6910-4.ch006.