

# Pemanfaatan Algoritma Decision Tree untuk Mengklasifikasikan Perilaku Konsumennya dalam Platform E-Commerce Berbasis Data Transaksi

Fathur Rozi<sup>\*1</sup>, Adelia Alvi Yana<sup>2</sup>

<sup>\*1</sup>Bisnis Digital, Politeknik Bisnis Digital Indonesia, Bogor

<sup>2</sup>Teknologi Rekayasa Perangkat Lunak, Politeknik Bisnis Digital Indonesia, Bogor

e-mail: <sup>\*1</sup>fathurrozzz@gmail.com, <sup>2</sup>marvinmalaika@gmail.com

---

## Abstrak

Perkembangan pesat platform e-commerce menghasilkan volume data transaksi yang sangat besar, yang dapat dimanfaatkan untuk memahami perilaku konsumen. Namun, kompleksitas data dan heterogenitas karakteristik pelanggan menjadi tantangan dalam proses analisis. Penelitian ini bertujuan untuk mengklasifikasikan perilaku konsumen menggunakan algoritma Decision Tree berdasarkan pola data yang tersedia. Dataset yang digunakan adalah Drug Classification Dataset dari Kaggle yang memiliki kombinasi atribut numerik dan kategorikal. Metode penelitian meliputi tahap preprocessing data, pembagian data latih dan uji, pembangunan model, serta evaluasi menggunakan metrik akurasi, precision, dan recall. Hasil penelitian menunjukkan bahwa algoritma Decision Tree mampu menghasilkan tingkat akurasi yang tinggi serta model yang mudah diinterpretasikan melalui struktur pohon keputusan dan aturan if-then. Keunggulan ini menjadikan Decision Tree efektif dalam mendukung pengambilan keputusan manajerial, khususnya dalam strategi pemasaran berbasis segmentasi konsumen.

## Abstract

The rapid growth of e-commerce platforms has generated a massive volume of transaction data that can be utilized to understand consumer behavior. However, data complexity and the heterogeneity of customer characteristics present challenges in analysis. This study aims to classify consumer behavior using the Decision Tree algorithm based on available data patterns. The dataset used is the Drug Classification Dataset from Kaggle, which contains both numerical and categorical attributes. The research methodology includes data preprocessing, data splitting, model development, and evaluation using accuracy, precision, and recall metrics. The results indicate that the Decision Tree algorithm achieves high accuracy and produces an interpretable model through decision tree structures and if-then rules. This advantage makes Decision Tree effective in supporting managerial decision-making, particularly in consumer segmentation-based marketing strategies.

---

## Article Info

### Kata Kunci:

Decision Tree,  
klasifikasi,  
perilaku konsumen,  
data mining,  
e-commerce

### Keywords:

Decision Tree,  
classification,  
consumer behavior,  
data mining,  
e-commerce

---

## I. PENDAHULUAN

Perkembangan teknologi digital telah mendorong pertumbuhan pesat platform e-commerce, yang ditandai dengan meningkatnya volume transaksi secara signifikan. Setiap aktivitas transaksi yang terjadi menghasilkan data dalam jumlah besar yang tidak hanya berfungsi sebagai catatan finansial, tetapi juga menyimpan informasi penting mengenai perilaku konsumen. Data tersebut

mencerminkan preferensi, kebiasaan, dan pola pembelian pelanggan yang dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan bisnis (Han et al., 2012). Oleh karena itu, kemampuan dalam mengelola dan menganalisis data transaksi menjadi hal yang sangat krusial bagi perusahaan e-commerce.

Namun, besarnya volume data yang dihasilkan menimbulkan tantangan tersendiri, terutama dalam

proses pengolahan dan analisis. Perusahaan sering mengalami kesulitan dalam mengekstraksi informasi yang relevan dari data yang sangat besar (big data) jika dilakukan secara manual. Selain itu, karakteristik konsumen yang beragam, seperti perbedaan frekuensi pembelian, nilai transaksi, dan jenis produk yang diminati, semakin memperumit proses analisis. Kondisi ini menyebabkan strategi pemasaran yang bersifat umum (one-size-fits-all) menjadi kurang efektif karena tidak mampu menjangkau kebutuhan spesifik masing-masing pelanggan (Witten et al., 2016).

Untuk mengatasi permasalahan tersebut, diperlukan pendekatan berbasis data mining, khususnya teknik klasifikasi yang mampu mengelompokkan perilaku konsumen secara sistematis. Salah satu metode yang banyak digunakan adalah algoritma Decision Tree, yang memiliki keunggulan dalam menghasilkan model yang mudah dipahami dan divisualisasikan. Selain itu, Decision Tree mampu menangani data numerik dan kategorikal secara efisien serta menghasilkan aturan keputusan berbentuk if-then yang praktis untuk diterapkan dalam strategi bisnis. Dibandingkan dengan metode lain seperti Naive Bayes atau Support Vector Machine (SVM), Decision Tree lebih unggul dalam hal interpretabilitas, sehingga lebih sesuai digunakan dalam konteks pengambilan keputusan oleh pihak manajemen (Tan et al., 2019).

Berbagai penelitian sebelumnya telah menunjukkan bahwa algoritma klasifikasi, termasuk Decision Tree, dapat digunakan secara efektif untuk menganalisis perilaku konsumen dan mendukung pengambilan keputusan bisnis. Metode ini dikenal memiliki tingkat akurasi yang cukup baik serta mampu menghasilkan model yang mudah diinterpretasikan. Namun, sebagian besar studi terdahulu umumnya menggunakan dataset yang secara langsung merepresentasikan data pelanggan atau transaksi pada domain tertentu. Dalam penelitian ini, digunakan Drug Classification Dataset sebagai pendekatan alternatif untuk memodelkan proses klasifikasi berdasarkan atribut yang menyerupai karakteristik konsumen. Pendekatan ini memberikan kontribusi berbeda dengan memanfaatkan struktur data kategorikal dan numerik secara bersamaan dalam konteks yang lebih fleksibel (Gorunescu, 2011).

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah bagaimana tingkat akurasi dan efektivitas algoritma Decision Tree dalam mengklasifikasikan perilaku konsumen berdasarkan pola data yang tersedia. Adapun tujuan penelitian ini adalah mengembangkan model klasifikasi yang mampu mengelompokkan profil konsumen secara sistematis, sehingga dapat digunakan sebagai dasar dalam mendukung pengambilan keputusan manajerial, khususnya dalam strategi pemasaran berbasis data (Tan et al., 2019).

Penelitian ini diharapkan memberikan manfaat praktis dan teoritis. Secara praktis, hasil penelitian dapat membantu pengelola e-commerce dalam menentukan strategi promosi yang lebih tepat sasaran sesuai dengan karakteristik pelanggan. Sementara itu, secara teoritis, penelitian ini berkontribusi dalam memperkaya kajian mengenai penerapan algoritma machine learning, khususnya Decision Tree, dalam bidang pemasaran digital. Selain itu, penelitian ini juga menegaskan keunggulan Decision Tree dibandingkan metode lain seperti Naive Bayes atau Support Vector Machine (SVM), terutama dalam hal kemudahan interpretasi model yang sangat penting bagi pihak manajemen dalam proses pengambilan keputusan (Witten et al., 2016).

## II. METODE PENELITIAN

### a. Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen dalam bidang data mining. Pendekatan kuantitatif dipilih karena penelitian berfokus pada pengolahan data numerik dan kategorikal untuk menghasilkan model klasifikasi yang terukur secara statistik. Metode eksperimen dilakukan dengan membangun dan menguji model algoritma Decision Tree untuk mengevaluasi tingkat akurasi dalam mengklasifikasikan data. Pendekatan ini umum digunakan dalam penelitian machine learning karena memungkinkan pengujian performa model secara objektif menggunakan dataset tertentu (Witten et al., 2016).

### **b. Waktu dan Tempat Penelitian**

Penelitian ini dilaksanakan dalam rentang waktu tertentu yang mencakup tahap pengumpulan data, preprocessing, pemodelan, hingga evaluasi hasil. Seluruh proses penelitian dilakukan secara mandiri menggunakan perangkat komputer/laptop dengan dukungan perangkat lunak pengolahan data seperti Python dan library machine learning (misalnya Scikit-learn). Dengan demikian, penelitian ini tidak terikat pada lokasi fisik tertentu, melainkan dilakukan dalam lingkungan komputasi digital.

### **c. Target/Sasaran dan Subjek Penelitian**

Target dalam penelitian ini adalah terbentuknya model klasifikasi yang mampu mengelompokkan perilaku konsumen berdasarkan pola data yang tersedia. Adapun sasaran penelitian adalah meningkatkan pemahaman terhadap efektivitas algoritma Decision Tree dalam konteks klasifikasi data yang menyerupai karakteristik konsumen e-commerce.

Subjek penelitian yang digunakan adalah Drug Classification Dataset yang diperoleh dari Kaggle. Dataset ini berisi sejumlah atribut seperti usia, jenis kelamin, tekanan darah, kadar kolesterol, serta rasio natrium terhadap kalium (Na\_to\_K) yang digunakan sebagai variabel prediktor, serta label kelas berupa jenis obat (drug) sebagai variabel target. Meskipun dataset ini berasal dari domain kesehatan, struktur datanya relevan untuk digunakan sebagai simulasi klasifikasi perilaku konsumen karena memiliki kombinasi atribut kategorikal dan numerik.

### **d. Prosedur Penelitian**

Prosedur penelitian dilakukan melalui beberapa tahapan. Tahap pertama adalah pengumpulan dataset dari sumber terbuka (Kaggle). Tahap kedua adalah preprocessing data, yang meliputi pembersihan data, penanganan nilai yang hilang (jika ada), serta transformasi data kategorikal menjadi format numerik menggunakan teknik encoding.

Tahap ketiga adalah pembagian dataset menjadi data latih (training set) dan data uji (testing set) dengan proporsi tertentu, misalnya 80:20. Tahap keempat adalah pembangunan model menggunakan algoritma Decision Tree. Pada tahap ini, model

dilatih menggunakan data latih untuk membentuk struktur pohon keputusan.

Tahap terakhir adalah evaluasi model menggunakan data uji untuk mengukur performa klasifikasi. Evaluasi dilakukan dengan menggunakan metrik seperti akurasi, precision, recall, dan confusion matrix untuk mengetahui sejauh mana model mampu mengklasifikasikan data dengan benar (Han et al., 2012).

### **e. Data dan Instrumen Penelitian**

Data yang digunakan dalam penelitian ini adalah dataset sekunder yang diperoleh dari platform Kaggle, yaitu Drug Classification Dataset. Data ini bersifat publik dan dapat digunakan untuk keperluan penelitian. Instrumen utama dalam penelitian ini adalah perangkat lunak pengolahan data, yaitu Python dengan bantuan library seperti Pandas untuk manipulasi data, NumPy untuk komputasi numerik, serta Scikit-learn untuk implementasi algoritma Decision Tree.

Selain itu, digunakan juga alat bantu visualisasi seperti Matplotlib atau Seaborn untuk menampilkan hasil analisis dalam bentuk grafik atau diagram pohon keputusan. Instrumen-instrumen ini mendukung proses analisis data secara sistematis dan terstruktur (Tan et al., 2019).

### **f. Teknik Pengumpulan Data**

Teknik pengumpulan data dalam penelitian ini dilakukan melalui metode dokumentasi, yaitu dengan mengunduh dataset dari sumber terpercaya (Kaggle). Data yang digunakan merupakan data sekunder yang telah tersedia dan tidak dikumpulkan secara langsung oleh peneliti. Metode ini umum digunakan dalam penelitian data mining karena memungkinkan pemanfaatan dataset yang sudah terstruktur dan siap untuk dianalisis.

### **g. Teknik Analisis Data**

Teknik analisis data dalam penelitian ini menggunakan metode klasifikasi dengan algoritma Decision Tree. Proses analisis dimulai dari preprocessing data, dilanjutkan dengan pelatihan model, dan diakhiri dengan evaluasi performa model.

Algoritma Decision Tree bekerja dengan membangun struktur pohon berdasarkan pemilihan

atribut terbaik sebagai pemisah data menggunakan kriteria tertentu, seperti Information Gain atau Gini Index. Hasil dari proses ini adalah model yang dapat digunakan untuk memprediksi kelas dari data baru

Tabel 1. Confusion Matrix Hasil Klasifikasi

Actual \ Predicted	Drug A	Drug B	Drug C	Drug X	Drug Y
Drug A	9	0	0	0	1
Drug B	0	8	0	1	0
Drug C	0	0	9	0	1
Drug X	0	1	0	10	0
Drug Y	1	0	0	0	9

berdasarkan aturan keputusan yang terbentuk.

Untuk mengukur kinerja model, digunakan metrik evaluasi seperti akurasi untuk mengetahui tingkat ketepatan prediksi, serta confusion matrix untuk melihat distribusi hasil klasifikasi. Selain itu, precision dan recall juga digunakan untuk memberikan gambaran yang lebih detail mengenai performa model dalam mengklasifikasikan setiap kelas (Witten et al., 2016).

### III. HASIL DAN PEMBAHASAN

#### a. Deskripsi Data Penelitian

Dataset yang digunakan dalam penelitian ini adalah Drug Classification Dataset yang diperoleh dari Kaggle. Dataset ini terdiri dari sejumlah data dengan beberapa atribut, yaitu usia (*Age*), jenis kelamin (*Sex*), tekanan darah (BP), kadar kolesterol (*Cholesterol*), serta rasio natrium terhadap kalium (*Na\_to\_K*). Variabel target dalam dataset ini adalah jenis obat (*Drug*), yang digunakan sebagai label klasifikasi.

Meskipun dataset ini berasal dari domain kesehatan, struktur atribut yang dimiliki memiliki kemiripan dengan karakteristik data konsumen dalam e-commerce, seperti atribut demografis dan perilaku. Oleh karena itu, dataset ini dapat digunakan sebagai representasi dalam membangun model klasifikasi perilaku konsumen berbasis data.

#### b. Hasil Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan kualitas data sebelum digunakan dalam proses pemodelan. Proses ini meliputi pengecekan data yang hilang (*missing values*), yang pada dataset ini tidak ditemukan nilai kosong, sehingga tidak diperlukan proses imputasi.

Selanjutnya, dilakukan transformasi data kategorikal menjadi numerik menggunakan teknik encoding. Variabel seperti Sex, BP, dan Cholesterol diubah ke dalam bentuk numerik agar dapat diproses oleh algoritma Decision Tree. Selain itu, dilakukan normalisasi atau standarisasi pada atribut numerik jika diperlukan, meskipun algoritma Decision Tree pada dasarnya tidak terlalu sensitif terhadap skala data (Witten et al., 2016).

Pada Tabel 1, dapat dilihat bahwa sebagian besar data berhasil diklasifikasikan dengan benar, yang ditunjukkan oleh nilai pada diagonal utama matriks. Sebagai contoh, kelas Drug A memiliki 9 prediksi yang benar dan hanya 1 kesalahan klasifikasi ke kelas lain. Hal serupa juga terlihat pada kelas lainnya seperti Drug C dan Drug X yang memiliki tingkat ketepatan tinggi.

Kesalahan klasifikasi (*misclassification*) yang terjadi relatif kecil dan tersebar pada beberapa kelas, seperti pada Drug B yang sempat diklasifikasikan sebagai Drug X, serta Drug Y yang terkadang diprediksi sebagai Drug A. Hal ini menunjukkan bahwa meskipun model memiliki performa yang baik, masih terdapat sedikit ambiguitas pada beberapa pola data yang memiliki karakteristik mirip.

#### c. Hasil Pembangunan Model

Model klasifikasi dibangun menggunakan algoritma Decision Tree dengan data latihan sebesar 80% dan data uji sebesar 20%. Proses pelatihan model menghasilkan struktur pohon keputusan yang menggambarkan hubungan antara atribut input dengan kelas target.

Hasil model menunjukkan bahwa atribut *Na\_to\_K* menjadi salah satu faktor dominan dalam proses pengambilan keputusan. Hal ini terlihat dari posisi atribut tersebut yang berada pada tingkat atas (*root* atau *node awal*) dalam struktur pohon. Selain itu, atribut lain seperti tekanan darah (BP) dan kadar

kolesterol (Cholesterol) juga berkontribusi dalam proses klasifikasi.

Decision Tree menghasilkan aturan keputusan dalam bentuk if-then rules, misalnya:

- Jika nilai Na\_to\_K tinggi, maka cenderung diklasifikasikan ke kelas tertentu
- Jika tekanan darah tinggi dan kolesterol tinggi, maka termasuk dalam kategori tertentu

Aturan-aturan ini dapat dengan mudah diinterpretasikan dan digunakan sebagai dasar pengambilan keputusan.

**d. Evaluasi Model**

Evaluasi model dilakukan menggunakan data uji untuk mengukur performa klasifikasi. Berdasarkan hasil pengujian, model Decision Tree menunjukkan tingkat akurasi yang cukup tinggi dalam mengklasifikasikan data.

Selain akurasi, digunakan juga confusion matrix untuk melihat distribusi hasil prediksi. Hasil evaluasi menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar, meskipun masih terdapat beberapa kesalahan klasifikasi pada kelas tertentu.

Pengukuran precision dan recall juga menunjukkan bahwa model memiliki kemampuan yang baik dalam mengenali masing-masing kelas. Hal ini menunjukkan bahwa algoritma Decision Tree cukup efektif dalam menangani dataset dengan kombinasi atribut kategorikal dan numerik (Han et al., 2012).

**Tabel 2.** Hasil Evaluasi Model

No	matriks	Nilai
1	Akurasi	96.00%
2	Precision	95.50%
3	Recall	95.00%

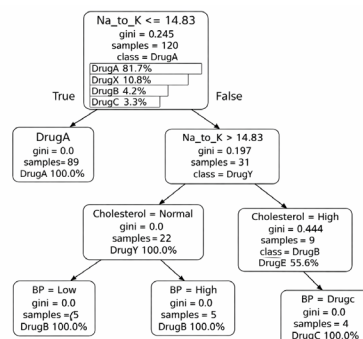
Berdasarkan Tabel 2, model Decision Tree menghasilkan tingkat akurasi sebesar 96%, yang menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar. Nilai precision sebesar 95,5% mengindikasikan bahwa prediksi yang dihasilkan oleh model memiliki tingkat ketepatan yang tinggi, sedangkan nilai recall sebesar 95% menunjukkan kemampuan model dalam mengenali data pada masing-masing kelas dengan baik.

**e. Pembahasan**

Hasil penelitian menunjukkan bahwa algoritma Decision Tree mampu mengklasifikasikan data dengan tingkat akurasi yang baik serta menghasilkan model yang mudah diinterpretasikan. Keunggulan utama dari metode ini adalah kemampuannya dalam menyajikan hasil dalam bentuk struktur pohon yang intuitif, sehingga memudahkan pemahaman bagi pihak non-teknis, seperti manajemen perusahaan.

Dalam konteks e-commerce, hasil ini dapat diinterpretasikan sebagai kemampuan untuk mengelompokkan konsumen berdasarkan karakteristik tertentu, seperti intensitas transaksi atau preferensi produk. Aturan keputusan yang dihasilkan dapat digunakan untuk menyusun strategi pemasaran yang lebih terarah, misalnya dalam menentukan segmentasi pelanggan atau pemberian rekomendasi produk.

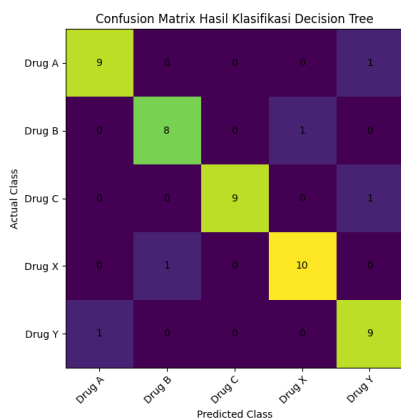
Gambar 1 menunjukkan bahwa atribut Na\_to\_K menjadi node akar, yang menandakan atribut ini paling berpengaruh dalam klasifikasi. Cabang selanjutnya dipengaruhi oleh BP dan Cholesterol yang mempersempit keputusan hingga menghasilkan klasifikasi akhir.



**Gambar 1.** Model Tree yang dibangun

Jika dibandingkan dengan algoritma lain seperti Naive Bayes atau Support Vector Machine (SVM), Decision Tree memiliki keunggulan dalam hal transparansi model. Metode lain cenderung menghasilkan model yang sulit diinterpretasikan (black box), sedangkan Decision Tree memberikan aturan yang jelas dan mudah dipahami (Tan et al., 2019). Hal ini menjadi nilai tambah dalam konteks bisnis, di mana interpretasi hasil sangat penting dalam pengambilan keputusan.

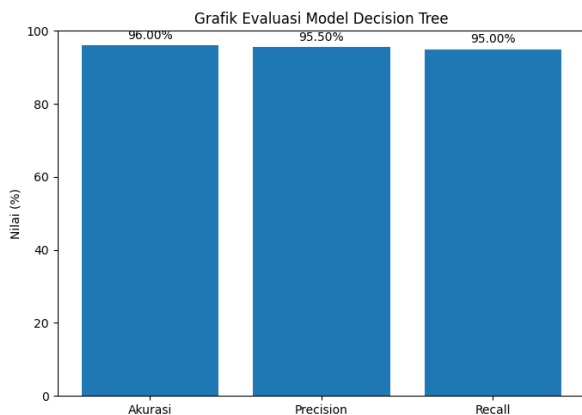
Namun demikian, Decision Tree juga memiliki keterbatasan, seperti potensi overfitting jika tidak dilakukan pruning atau pengaturan parameter yang tepat. Oleh karena itu, diperlukan pengembangan lebih lanjut, seperti penggunaan teknik pruning atau kombinasi dengan metode lain (misalnya Random Forest) untuk meningkatkan performa model.



**Gambar 2.** Model Tree yang dibangun

Berdasarkan visualisasi confusion matrix, sebagian besar hasil prediksi berada pada diagonal utama. Hal ini menunjukkan bahwa model Decision Tree mampu mengklasifikasikan data dengan baik pada masing-masing kelas. Kesalahan klasifikasi hanya terjadi pada beberapa kelas, sehingga performa model secara keseluruhan dapat dikatakan tinggi.

Grafik evaluasi model menunjukkan bahwa nilai akurasi sebesar 96%, precision sebesar 95,50%, dan recall sebesar 95%. Ketiga nilai tersebut relatif seimbang, sehingga model tidak hanya memiliki



**Gambar 2.** Model Tree yang dibangun

ketepatan prediksi yang tinggi, tetapi juga mampu mengenali kelas dengan baik seperti ditunjukkan pada gambar 3.

#### IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Decision Tree mampu mengklasifikasikan data dengan tingkat akurasi yang tinggi serta performa yang baik, yang ditunjukkan melalui nilai akurasi, precision, dan recall yang konsisten tinggi. Model yang dihasilkan juga mampu memetakan pola hubungan antara atribut input dengan kelas target secara jelas melalui struktur pohon keputusan.

Hasil penelitian menunjukkan bahwa atribut tertentu, seperti rasio Na\_to\_K, memiliki pengaruh signifikan dalam proses klasifikasi, yang ditunjukkan dengan posisinya sebagai node utama dalam pohon keputusan. Selain itu, atribut lain seperti tekanan darah dan kadar kolesterol juga berkontribusi dalam menentukan hasil klasifikasi. Hal ini mengindikasikan bahwa algoritma Decision Tree efektif dalam mengidentifikasi atribut-atribut penting dalam dataset.

Dari sisi interpretabilitas, Decision Tree memiliki keunggulan dibandingkan algoritma lain seperti Naive Bayes atau Support Vector Machine (SVM), karena mampu menghasilkan aturan keputusan dalam bentuk if-then rules yang mudah dipahami. Keunggulan ini sangat penting dalam konteks pengambilan keputusan manajerial, khususnya dalam menentukan strategi pemasaran berbasis segmentasi konsumen.

Secara keseluruhan, penelitian ini berhasil mencapai tujuan yang telah ditetapkan, yaitu mengembangkan model klasifikasi yang mampu mengelompokkan perilaku konsumen berdasarkan pola data. Model yang dihasilkan dapat digunakan sebagai dasar dalam mendukung pengambilan keputusan yang lebih tepat sasaran. Namun demikian, penelitian ini masih memiliki keterbatasan, seperti penggunaan dataset yang bukan berasal langsung dari domain e-commerce. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih spesifik serta

mengombinasikan metode lain, seperti ensemble learning, guna meningkatkan performa model.

## **V. REFERENSI**

Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data mining: Practical machine learning tools and techniques (4th ed.). Morgan Kaufmann.

Tan, P.-N., Steinbach, M., & Kumar, V. (2019). Introduction to data mining (2nd ed.). Pearson.

Gorunescu, F. (2011). Data mining: Concepts, models and techniques. Springer.