

PENINGKATAN KLASIFIKASI PENJUALAN PRODUK FASHION DI SABHIRA OFFICIAL DENGAN RANDOM FOREST

Nazwa Arraudhah¹, Ade Irma Purnama Sari², Agus Bahtiar³, Edi Wahyudin⁴

¹Jurusan Teknik Informatika, STMIK IKMI Cirebon

²Rekayasa Perangkat Lunak, STMIK IKMI Cirebon

³Sistem Informasi, STMIK IKMI Cirebon

⁴Komputerisasi Akuntansi, STMIK IKMI Cirebon

e-mail: ¹nzwarrdahah19@gmail.com, ²irma2974@yahoo.com,

³agusbahtiar038@gmail.com, ⁴ediwahyudin@gmail.com

Intisari

Penelitian ini berfokus pada peningkatan akurasi model klasifikasi penjualan produk fashion di Toko Sabhira Official dengan menerapkan algoritma Random Forest. Pendekatan yang digunakan mengikuti tahapan Knowledge Discovery in Database (KDD), yang mencakup pemilihan data, prapemrosesan, transformasi, data mining, dan evaluasi. Data penelitian terdiri dari 1.559 transaksi dalam periode Agustus hingga Oktober 2023, dengan atribut seperti kategori produk, jumlah barang terjual, harga, serta kategori penjualan (rendah, sedang, tinggi). Model dikembangkan menggunakan perangkat lunak RapidMiner, dengan pembagian data sebesar 70% untuk pelatihan dan 30% untuk pengujian. Hasil analisis menunjukkan bahwa algoritma Random Forest mampu mencapai tingkat akurasi sebesar 99,81%, dengan precision untuk kategori “Tinggi” mencapai 100%, sementara kategori lainnya memiliki nilai di atas 99%. Evaluasi menggunakan confusion matrix menunjukkan tingkat kesalahan prediksi yang sangat rendah, sehingga model ini dapat mengklasifikasikan tingkat penjualan secara lebih akurat. Hasil penelitian ini memberikan wawasan yang berguna bagi Toko Sabhira Official dalam pengelolaan stok serta strategi promosi berbasis data.

Kata Kunci— Random Forest, klasifikasi penjualan, produk fashion, KDD.

Abstract

**PENINGKATAN KLASIFIKASI PENJUALAN PRODUK FASHION DI SABHIRA
OFFICIAL DENGAN RANDOM FOREST
(Nazwa Arraudhah, Ade Irma Purnama Sari, Agus Bahtiar, Edi Tohidi)**

This research focuses on improving the accuracy of the fashion product sales classification model at Sabhira Official Store by applying the Random Forest algorithm. The approach used follows the stages of Knowledge Discovery in Database (KDD), which includes data selection, preprocessing, transformation, data mining, and evaluation. The research data consisted of 1,559 transactions in the period August to October 2023, with attributes such as product category, number of items sold, price, and sales category (low, medium, high). The model was developed using RapidMiner software, with a data split of 70% for training and 30% for testing. The analysis results show that the Random Forest algorithm is able to achieve an accuracy rate of 99.81%, with precision for the “High” category reaching 100%, while other categories have values above 99%. Evaluation using confusion matrix shows a very low prediction error rate, so this model can classify sales levels more accurately. The results of this study provide useful insights for Sabhira Official Store in stock management and data-driven promotion strategies.

Keywords— *Random Forest, sales classification, fashion products, KDD.*

PENDAHULUAN

Sektor fashion mengalami perkembangan pesat di era digital, mendorong persaingan yang semakin ketat. Untuk tetap kompetitif, para pelaku bisnis perlu mengadopsi strategi berbasis data guna meningkatkan efisiensi dan efektivitas operasional [1]. Salah satu aspek krusial dalam pengelolaan bisnis fashion adalah kemampuan mengklasifikasikan tingkat penjualan produk dengan akurat. Klasifikasi yang tepat memungkinkan perusahaan mengelola stok, merancang strategi pemasaran, dan merencanakan produksi secara lebih optimal [2].

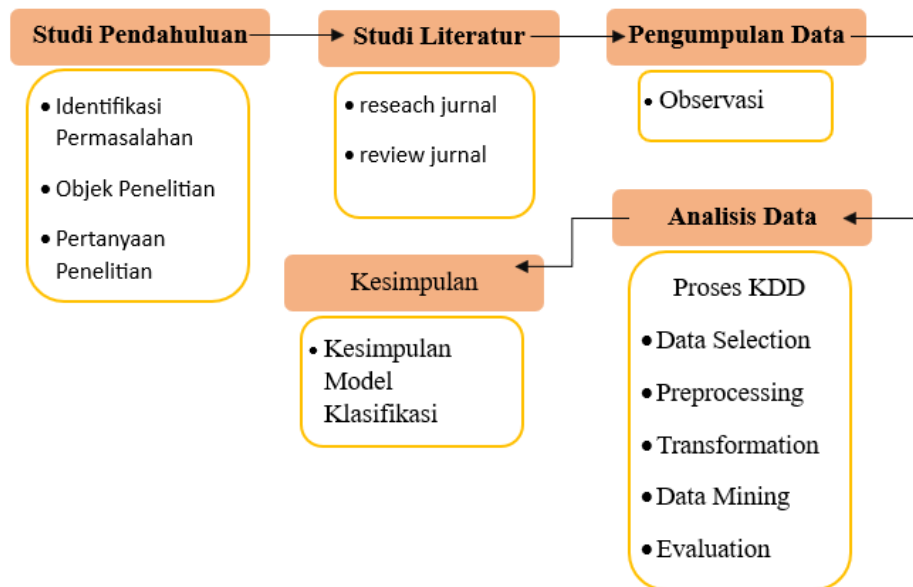
Tantangan utama dalam klasifikasi penjualan terletak pada beragam faktor yang memengaruhi performa produk di pasar, seperti tren konsumen, harga, dan strategi promosi [3]. Penelitian terdahulu menunjukkan bahwa metode pembelajaran mesin, termasuk algoritma Random Forest, mampu memberikan hasil lebih akurat dibandingkan pendekatan tradisional dalam analisis penjualan [4]. Misalnya, penelitian oleh Arumnisaa & Wijayanto (2023) mengungkapkan bahwa Random Forest memiliki kinerja lebih unggul dalam klasifikasi data bisnis dibandingkan metode lain seperti Support Vector Machine (SVM) dan AdaBoost [5]. Sementara itu, studi oleh Ismanto & Novalia (2021) yang membandingkan Random Forest dengan algoritma C4.5 dan Gradient Boosting menunjukkan bahwa Random Forest memiliki tingkat akurasi tertinggi [6].

Penelitian ini bertujuan untuk meningkatkan akurasi model klasifikasi penjualan produk fashion di Toko Sabhira Official dengan menerapkan algoritma Random Forest dalam kerangka Knowledge Discovery in Database (KDD). Dengan menganalisis data transaksi selama tiga bulan, penelitian ini diharapkan dapat mengidentifikasi pola penjualan yang signifikan serta memberikan wawasan strategis dalam pengelolaan stok dan promosi [7]. Selain itu, hasil penelitian ini diharapkan dapat berkontribusi terhadap pengambilan keputusan berbasis data di industri ritel fashion serta memperkaya kajian tentang penerapan pembelajaran mesin dalam analisis bisnis [8].

**PENINGKATAN KLASIFIKASI PENJUALAN PRODUK FASHION DI SABHIRA
OFFICIAL DENGAN RANDOM FOREST**
(Nazwa Arraudhah, Ade Irma Purnama Sari, Agus Bahtiar, Edi Tohidi)

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan menganalisis data transaksi di Toko Sabhira Official menggunakan algoritma Random Forest pada RapidMiner 10.5. Tujuannya adalah mengidentifikasi atribut penting dan meningkatkan akurasi klasifikasi tingkat penjualan. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score. Proses penelitian mencakup pengumpulan data, prapemrosesan, pembangunan model, evaluasi, dan penyusunan rekomendasi.



Gambar 1. Tahapan Metode Penelitian

1. Studi Pendahuluan

- Mengidentifikasi masalah rendahnya akurasi klasifikasi penjualan di Toko Sabhira Official.
- Menentukan tujuan penelitian dan menyusun pertanyaan penelitian terkait penerapan dan evaluasi algoritma Random Forest.

2. Studi Literatur

- Mengkaji penelitian terdahulu tentang metode klasifikasi penjualan dan algoritma Random Forest.
- Meninjau berbagai teknik klasifikasi untuk menentukan metode terbaik.

3. Pengumpulan Data

- Mengumpulkan data transaksi penjualan, termasuk kategori produk, harga, dan jumlah terjual.
4. Analisis Data (Menggunakan pendekatan KDD)
- Data Selection: Memilih data yang relevan.
 - Preprocessing: Membersihkan dan menyiapkan data.
 - Transformation: Mengubah format data sesuai kebutuhan analisis.
 - Data Mining: Menerapkan algoritma Random Forest.
 - Evaluation: Mengevaluasi hasil model dengan metrik akurasi, precision, recall, dan F1-score.
5. Kesimpulan
- Menyimpulkan efektivitas model dan memberikan rekomendasi optimalisasi klasifikasi penjualan di Toko Sabhira Official.

HASIL DAN PEMBAHASAN

1. Hasil Penelitian

Penelitian ini menyoroti pola industri fashion dengan menerapkan pendekatan Knowledge Discovery in Database (KDD) menggunakan teknologi RapidMiner. Proses KDD terdiri dari beberapa tahapan, yaitu seleksi data, pra-pemrosesan, transformasi, penggalian data, dan evaluasi. Studi ini tidak hanya bertujuan untuk meningkatkan akurasi model klasifikasi penjualan tetapi juga untuk memberikan wawasan lebih mendalam mengenai preferensi konsumen dan tren penjualan di Toko Sabhira Official. Temuan ini diharapkan dapat membantu dalam pengembangan strategi bisnis yang lebih efektif dan berkelanjutan untuk memenuhi kebutuhan pasar dengan lebih baik.

1.1 Data Selection

Data yang digunakan dalam penelitian ini merupakan data transaksi penjualan produk fashion di Toko Sabhira Official yang mencakup informasi jumlah produk yang terjual (QYT), jumlah kas masuk, serta kategori penjualan yang menjadi target

**PENINGKATAN KLASIFIKASI PENJUALAN PRODUK FASHION DI SABHIRA
OFFICIAL DENGAN RANDOM FOREST
(Nazwa Arraudhah, Ade Irma Purnama Sari, Agus Bahtiar, Edi Tohidi)**

prediksi (Rendah, Sedang, Tinggi). Data ini dikumpulkan dalam format Excel dan diimpor ke dalam RapidMiner menggunakan operator Read Excel sebagai input awal untuk proses analisis.

Dataset yang digunakan mencakup transaksi selama tiga bulan, dari Agustus hingga Oktober 2023, dengan total 1.559 record yang terdiri dari 10 atribut utama. seperti yang terlihat dalam table 1.

Tabel 1. Deskripsi Atribut Data

No	Atribut	Type Data	Keterangan
1	Nama Barang	Object	Nama produk yang dijual
2	Placement	Object	Jenis pembeli
3	Ukuran	Object	Ukuran produk
4	QYT	int64	Jumlah produk yang dibeli
5	Kategori Penjualan	Object	Kategori tingkat penjualan
6	Harga	int64	Harga satuan produk
7	Ongkir	int64	Biaya pengiriman
8	Jumlah	int64	Jumlah total setelah harga dan diskon
9	Disc	int64	Diskon yang diberikan
10	Kas Masuk	int64	Total uang yang diterima setelah biaya dan diskon

Data ini kemudian diproses lebih lanjut untuk menyaring atribut yang paling relevan melalui tahap Select Attributes, sehingga model dapat lebih fokus pada variabel yang memiliki pengaruh terhadap hasil prediksi.

Row No.	Kategori Pe...	QYT	Harga	Ongkir	Jumlah	Disc	Kas Masuk
1	Rendah	1	120000	0	120000	0	120000
2	Rendah	1	135000	11000	145000	0	146000
3	Rendah	1	115000	0	115000	0	115000
4	Rendah	1	130000	0	130000	0	130000
5	Rendah	1	115000	0	115000	0	115000
6	Rendah	1	115000	0	115000	10000	105000
7	Rendah	1	125000	0	125000	10000	115000
8	Rendah	1	115000	0	115000	10000	105000
9	Rendah	1	120000	0	120000	10000	110000
10	Rendah	1	115000	0	115000	10000	105000
11	Tinggi	8	110000	0	880000	80000	800000
12	Rendah	1	149000	0	149000	10000	139000
13	Rendah	3	110000	0	330000	45000	285000
14	Sedang	5	115000	0	575000	75000	500000
15	Sedang	5	110000	0	550000	75000	475000

Gambar 2. Hasil proses *Select Attribute*

1.2 Preprocessing

Tahap preprocessing bertujuan untuk memastikan kualitas data sebelum digunakan dalam model klasifikasi. Proses ini mencakup penentuan label target menggunakan operator Set Role di RapidMiner, di mana atribut "Kategori Penjualan" ditetapkan sebagai target klasifikasi. Selain itu, dilakukan penanganan nilai yang hilang menggunakan operator Replace Missing Value untuk memastikan integritas data sebelum memasuki tahap berikutnya.

Pembagian data dilakukan dengan rasio 70:30, di mana 70% data digunakan untuk pelatihan dan 30% untuk pengujian. Teknik ini bertujuan untuk mencegah overfitting dan memastikan bahwa model dapat bekerja dengan baik pada data yang belum pernah dilihat sebelumnya.

1.3 Transformation

Transformasi data bertujuan untuk menyesuaikan format dataset agar sesuai dengan kebutuhan algoritma Random Forest. Langkah-langkah utama dalam tahap ini meliputi:

**PENINGKATAN KLASIFIKASI PENJUALAN PRODUK FASHION DI SABHIRA
OFFICIAL DENGAN RANDOM FOREST
(Nazwa Arraudhah, Ade Irma Purnama Sari, Agus Bahtiar, Edi Tohidi)**

1. Penggantian Nilai yang Hilang: Menggunakan metode rata-rata atau median untuk menggantikan nilai kosong dalam dataset.
2. Pemilihan Atribut Relevan: Mengeliminasi atribut yang tidak memiliki kontribusi signifikan terhadap klasifikasi.
3. Penetapan Peran Atribut: Menentukan atribut target sebagai label klasifikasi dan atribut lainnya sebagai prediktor.

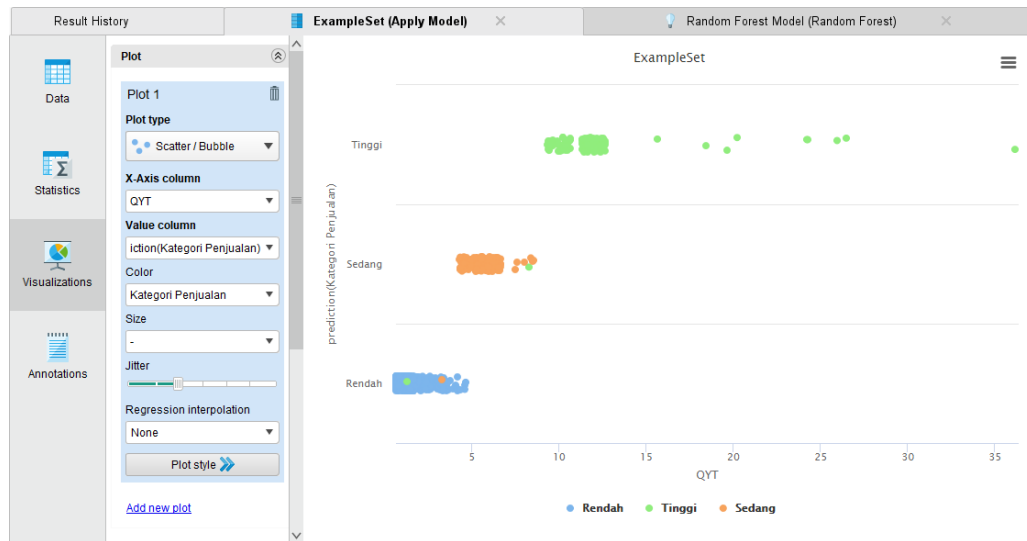
Dengan transformasi yang optimal, dataset menjadi lebih terstruktur sehingga dapat meningkatkan kinerja model dalam memprediksi kategori penjualan dengan lebih akurat.

1.4 Data Mining

Tahap ini mencakup penerapan algoritma Random Forest dalam proses klasifikasi penjualan. Model dikembangkan dengan pengaturan parameter sebagai berikut:

- Number of Trees: 100
- Maximal Depth: 10
- Voting Method: Confidence Vote

Setelah model dikembangkan, data diuji dengan Operator Apply Model untuk mengukur keakuratan prediksi. Visualisasi hasil menunjukkan pola distribusi kategori penjualan berdasarkan jumlah produk yang terjual, dengan pemisahan kategori yang cukup jelas.



Gambar 3. Hasil Visualisasi *Apply model*

1.5 Evaluasi Model

Evaluasi model dilakukan dengan menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil evaluasi menunjukkan bahwa model memiliki akurasi sebesar 99,81%, dengan precision dan recall yang tinggi di setiap kategori. Matriks kebingungan menunjukkan bahwa model mampu mengklasifikasikan kategori dengan tingkat kesalahan yang sangat rendah, mengindikasikan efektivitas algoritma Random Forest dalam menganalisis pola penjualan di Toko Sabhira Official.

accuracy: 99.81%

	true Rendah	true Tinggi	true Sedang	class precision
pred. Rendah	1259	1	1	99.84%
pred. Tinggi	0	138	0	100.00%
pred. Sedang	0	1	159	99.38%
class recall	100.00%	98.57%	99.38%	

Gambar 4. *Confusion Matrix* yang menunjukkan tingkat keberhasilan model

**PENINGKATAN KLASIFIKASI PENJUALAN PRODUK FASHION DI SABHIRA
OFFICIAL DENGAN RANDOM FOREST
(Nazwa Arraudhah, Ade Irma Purnama Sari, Agus Bahtiar, Edi Tohidi)**

2. Hasil Penelitian

Penelitian ini berfokus pada peningkatan model klasifikasi penjualan produk fashion di Toko Sabhira Official dengan menggunakan algoritma Random Forest. Hasil penelitian menunjukkan bahwa metode ini dapat mengelompokkan tingkat penjualan dengan akurasi tinggi berdasarkan berbagai atribut, seperti kategori produk, harga, diskon, dan jumlah barang yang terjual. Model yang dikembangkan bertujuan untuk memprediksi tingkat penjualan berdasarkan data historis, sehingga dapat membantu toko dalam merancang strategi pemasaran dan mengelola stok dengan lebih efektif. Evaluasi model menunjukkan akurasi sebesar 99,81%, di mana sebagian besar data diklasifikasikan dengan benar, meskipun masih terdapat beberapa kesalahan prediksi dalam jumlah kecil. Kendati demikian, model tetap menunjukkan kinerja yang sangat baik dalam mengenali pola penjualan. Hasil klasifikasi ini menjadi dasar untuk menyusun rekomendasi strategis, terutama dalam mengidentifikasi tren penjualan yang dapat mendukung perencanaan promosi yang lebih optimal. Selain itu, wawasan yang diperoleh dari model ini dapat digunakan untuk meningkatkan efisiensi pengelolaan stok serta memperkirakan permintaan pasar dengan lebih akurat, sehingga membantu dalam pengambilan keputusan bisnis yang lebih tepat di Toko Sabhira Official.

KESIMPULAN

Dari hasil penelitian ini, dapat disimpulkan bahwa penerapan algoritma Random Forest berhasil meningkatkan akurasi model klasifikasi penjualan produk fashion di Toko Sabhira Official, dengan tingkat akurasi mencapai 99,81%. Model ini mampu membedakan kategori penjualan (Rendah, Sedang, dan Tinggi) dengan sangat baik, meskipun terdapat beberapa kesalahan klasifikasi minor. Hasil evaluasi menunjukkan bahwa model ini efektif dalam mengidentifikasi pola penjualan dan memberikan wawasan strategis bagi toko dalam pengelolaan stok serta perencanaan promosi. Dengan memahami pola penjualan yang telah teridentifikasi, perusahaan

dapat menyusun strategi pemasaran yang lebih optimal untuk meningkatkan efisiensi bisnis dan penjualan produk fashion di Toko Sabhira Official.

SARAN

Penelitian lanjutan disarankan untuk mengintegrasikan data dari berbagai platform, termasuk media sosial dan marketplace, guna memahami perbedaan pola penjualan antara toko offline dan online. Selain itu, membandingkan algoritma lain seperti Gradient Boosting, XGBoost, atau Neural Networks dapat membantu memastikan metode yang paling efektif. Terakhir, pengembangan dashboard interaktif akan mempermudah visualisasi hasil model, menampilkan tren penjualan, serta mendukung strategi bisnis yang lebih optimal.

UCAPAN TERIMA KASIH

terima kasih kepada Rektor STMIK IKMI Cirebon atas dukungan finansial yang telah diberikan, sehingga penelitian ini dapat terselesaikan dengan baik.

Penulis juga menyampaikan apresiasi yang sebesar-besarnya kepada dosen pembimbing, Ibu Ade Irma Purnama Sari, M.Kom, dan Bapak Agus Bahtiar, M.Kom, atas bimbingan, kesabaran, serta ilmu yang telah diberikan selama proses penelitian ini.

**PENINGKATAN KLASIFIKASI PENJUALAN PRODUK FASHION DI SABHIRA
OFFICIAL DENGAN RANDOM FOREST
(Nazwa Arraudhah, Ade Irma Purnama Sari, Agus Bahtiar, Edi Tohidi)**

DAFTAR PUSTAKA

- [1] R. I. Arumnisaa and A. W. Wijayanto, "Comparison of Ensemble Learning Method: Random Forest, Support Vector Machine, AdaBoost for Classification Human Development Index (HDI)," *Sistemasi*, vol. 12, no. 1, p. 206, 2023, doi: 10.32520/stmsi.v12i1.2501.
- [2] E. Ismanto and M. Novalia, "Komparasi Kinerja Algoritma C4.5, Random Forest, dan Gradient Boosting untuk Klasifikasi Komoditas Performance Comparison Between C4.5 Algorithm, Random Forests, and Gradient Boosting for Commodity Classification," *Techno.COM*, vol. 20, no. 3, pp. 400–410, 2021.
- [3] F. Y. Pamuji and V. P. Ramadhan, "Komparasi Algoritma Random Forest dan Decision Tree untuk Memprediksi Keberhasilan Immunotherapy," *J. Teknol. dan Manaj. Inform.*, vol. 7, no. 1, pp. 46–50, 2021, doi: 10.26905/jtmi.v7i1.5982.
- [4] M. S. Efendi and A. K. Zyen, "Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk," vol. 5, no. 1, pp. 12–20, 2024, doi: 10.30865/resolusi.v5i1.2149.
- [5] H. Hidayat, A. Sunyoto, and H. Al Fatta, "Klasifikasi Penyakit Jantung Menggunakan Random Forest Clasifier," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 7, no. 1, pp. 31–40, 2023, doi: 10.47970/siskom-kb.v7i1.464.
- [6] A. Nugroho and E. Rilvani, "Penerapan Metode Oversampling SMOTE Pada Algoritma Random Forest Untuk Prediksi Kebangkrutan Perusahaan Application of the SMOTE Oversampling Method to the Random Forest Algorithm for Predicting Company Bankruptcy," *Februari*, vol. 22, no. 1, pp. 207–214, 2024.
- [7] R. Rindiyani, A. Primadewi, M. Maimunah, and A. H. Purwantini, "Klasifikasi Penjualan berdasarkan Platform pada UMKM Omah Branded Menggunakan Random Forest," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 5, p. 1520, 2022, doi: 10.30865/jurikom.v9i5.4949.
- [8] M. Rusdi Rahman, A. Febri Diansyah, and H. Hanafi, "Sentiment Analysis on the Shopee Application on Playstore Using the Random Forest Classification Method," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 9, no. 1, pp. 20–24, 2023, doi: 10.25139/inform.v9i1.5465.