



Implementasi Hybrid Data Mining Menggunakan DBSCAN, FP-Growth, dan Random Forest pada Analisis Penjualan Soraya Bedsheet di Shopee

Sandy Mulyanda¹, Nurul Aulya², Gusrianty³

¹Sistem Informasi, Fakultas Ilmu Komputer, Institut Bisnis dan Teknologi Pelita Indonesia

²Informatika, Fakultas Teknologi Industri, Universitas Bung Hatta

³Teknik Informatika, Fakultas Ilmu Komputer, Institut Bisnis dan Teknologi Pelita Indonesia

¹sandy.mulyanda@lecturer.pelitaindonesia.ac.id, ²nurulaulya@bunghatta.co.id, ³gusrianty@lecturer.pelitaindonesia.co.id

Abstract

The rapid growth of e-commerce platforms has generated large volumes of transactional data that can support business decision-making. However, many online sellers still rely on sales reports without further analysis to uncover valuable insights. This study aims to analyze sales patterns and predict sales performance of Soraya Bedsheet products on Shopee using a hybrid data mining approach. The proposed method integrates DBSCAN for product segmentation, FP-Growth for association analysis, and Random Forest for sales prediction. The dataset consists of 120 product records including product stock, number of items sold, and initial product price. Model evaluation was conducted using 10-fold Cross Validation in RapidMiner. Random Forest achieved an average accuracy of 91.67%, indicating good predictive performance while reducing the risk of overfitting. The integration of these methods provides comprehensive insights into product performance and customer purchasing behavior, supporting inventory management, promotional planning, and data-driven decision-making on the Shopee marketplace.

Keywords: Data Mining, DBSCAN, FP-Growth, Random Forest, Shopee

Abstrak

Pertumbuhan marketplace dalam beberapa tahun terakhir telah menghasilkan data transaksi dalam jumlah besar yang berpotensi dimanfaatkan sebagai sumber informasi bagi pelaku usaha. Namun, sebagian besar data penjualan masih digunakan sebatas laporan operasional sehingga informasi yang ada di dalamnya belum dimanfaatkan dengan optimal. Adapun tujuannya dari penelitian ini guna menganalisa karakteristik penjualan dan memprediksi performa penjualan produk Soraya Bedsheet pada Shopee melalui pendekatan hybrid data mining. Metode yang digunakan mengombinasikan DBSCAN untuk segmentasi produk, FP-Growth untuk analisis asosiasi antarproduk, dan Random Forest untuk prediksi penjualan. Dataset penelitian terdiri atas 120 data produk, jumlah stok, jumlah produk terjual, dan harga awal produk. Tahapan penelitian meliputi pengumpulan data, preprocessing, implementasi model, serta evaluasi dengan digunakannya RapidMiner. Temuan ini mengungkapkan bahwasanya DBSCAN mampu mengklasifikasikan produk sebagaimana karakteristik penjualannya, FP-Growth berhasil menemukan pola keterkaitan antarproduk dalam transaksi, dan Random Forest memperoleh rata-rata accuracy sebesar 91,67% melalui 10-fold Cross Validation. Hasil penelitian menunjukkan bahwasanya pendekatan hybrid data mining mampu menghasilkan informasi yang mendukung pengelolaan stok, penyusunan strategi promosi, dan pengambilan keputusan bisnis secara lebih efektif pada marketplace Shopee.

Kata kunci: hybrid data mining, penjualan, soraya bedsheets, shopee, prediksi



1. Pendahuluan

Pemanfaatan marketplace telah mengubah cara pelaku usaha memasarkan dan menjual produk kepada konsumen. Melalui platform digital, transaksi dapat dilakukan tanpa batas wilayah dan waktu sehingga jangkauan pemasaran menjadi lebih luas. Aktivitas tersebut menghasilkan data transaksi yang terus bertambah dan menyimpan berbagai informasi mengenai produk yang diminati, pola pembelian konsumen, serta performa penjualan yang terjadi pada setiap periode [1].

Shopee sebagai satu dari sekian marketplace yang banyak dipergunakan oleh pelaku usaha di Indonesia. Tingginya aktivitas transaksi pada platform ini menghasilkan data penjualan dalam jumlah besar. Meskipun demikian, pemanfaatan data tersebut masih sering terbatas pada pencatatan dan pelaporan penjualan. Padahal, data yang tersedia dapat digunakan dalam mendapatkan informasi yang membantu pelaku usaha dalam menentukan strategi pemasaran, mengelola persediaan produk, dan memahami perilaku konsumen [2].

Soraya Bedsheet merupakan usaha yang memasarkan berbagai perlengkapan tempat tidur melalui Shopee. Produk yang dijual meliputi sprei, bedcover, sarung bantal, dan perlengkapan pendukung lainnya. Banyaknya variasi produk menyebabkan karakteristik penjualan setiap produk tidak selalu sama. Beberapa produk memiliki tingkat penjualan yang tinggi, sedangkan produk lainnya menunjukkan tingkat permintaan yang lebih rendah. Kondisi tersebut menunjukkan perlunya analisis terhadap data penjualan agar informasi yang tersimpan dapat dimanfaatkan secara lebih optimal.

Beberapa penelitian sebelumnya telah memanfaatkan teknik data mining untuk mengolah data transaksi. Saputra et al. [3] menggunakan FP-Growth untuk menemukan pola pembelian konsumen sehingga mampu menghasilkan association rule sebagai dasar penyusunan strategi pemasaran. Amelia et al. [4] juga menunjukkan efektivitas FP-Growth dalam mengidentifikasi frequent itemset pada market basket analysis. Sementara itu, Irawan et al. [5] memanfaatkan DBSCAN untuk melakukan segmentasi data berdasarkan karakteristik transaksi, sedangkan Random Forest banyak digunakan dalam penelitian prediksi karena memiliki kemampuan klasifikasi yang tinggi [6]. Meskipun masing-masing algoritma memiliki keunggulan pada bidang analisis yang berbeda, penelitian-penelitian tersebut masih menerapkan algoritma secara terpisah. FP-Growth hanya berfokus pada analisis pola pembelian,

DBSCAN pada segmentasi data, sedangkan Random Forest pada klasifikasi atau prediksi. Akibatnya, informasi yang dihasilkan masih bersifat parsial dan belum mampu memberikan gambaran yang komprehensif untuk mendukung pengambilan keputusan bisnis pada marketplace.

Dalam konteks analisis penjualan pada marketplace, pelaku usaha memerlukan informasi yang lebih komprehensif, mulai dari mengetahui kelompok produk yang memiliki karakteristik serupa, memahami pola pembelian konsumen, hingga memprediksi performa penjualan pada periode berikutnya. Kebutuhan tersebut tidak dapat dipenuhi secara optimal apabila hanya menggunakan satu algoritma data mining. Oleh karena itu, penelitian ini mengusulkan pendekatan hybrid yang mengintegrasikan DBSCAN, FP-Growth, dan Random Forest agar mampu menghasilkan informasi yang lebih lengkap dan saling melengkapi.

Ketiga algoritma tersebut dipilih karena memiliki fungsi analisis yang berbeda namun saling melengkapi. DBSCAN digunakan untuk mengelompokkan produk berdasarkan karakteristik penjualan sehingga dapat mengidentifikasi produk yang memiliki performa serupa. FP-Growth digunakan untuk menemukan pola keterkaitan antarproduk yang sering dibeli secara bersamaan oleh konsumen, sehingga dapat mendukung penyusunan strategi promosi dan bundling produk. Selanjutnya, Random Forest digunakan untuk memprediksi performa penjualan berdasarkan atribut produk yang tersedia. Apabila ketiga algoritma tersebut digunakan secara terpisah, informasi yang dihasilkan hanya terbatas pada satu aspek analisis. Oleh karena itu, penggabungan DBSCAN, FP-Growth, dan Random Forest dalam penelitian ini diharapkan mampu menghasilkan analisis yang lebih komprehensif, mulai dari segmentasi produk, identifikasi pola pembelian, hingga prediksi penjualan sebagai dasar pengambilan keputusan bisnis yang lebih efektif [17].

Berdasarkan research gap yang telah diidentifikasi, penelitian ini menawarkan pendekatan hybrid data mining yang mengintegrasikan algoritma DBSCAN, FP-Growth, dan Random Forest dalam satu framework analisis. Kebaruan penelitian ini terletak pada kemampuan framework tersebut untuk menghasilkan segmentasi produk, analisis pola pembelian konsumen, dan prediksi penjualan secara terpadu menggunakan dataset yang sama. Berbeda dengan penelitian terdahulu yang umumnya hanya berfokus pada satu jenis analisis, pendekatan yang diusulkan mampu menghasilkan informasi yang

lebih komprehensif sehingga dapat mendukung pengambilan keputusan bisnis secara lebih efektif pada marketplace Shopee.

Tujuan penelitian ini adalah menganalisis karakteristik penjualan dan memprediksi performa penjualan produk Soraya Bedsheet pada Shopee menggunakan pendekatan hybrid data mining. Hasil penelitian diharapkan dapat memberikan informasi yang bermanfaat dalam pengelolaan produk, penyusunan strategi promosi, pengendalian stok, serta mendukung pengambilan keputusan bisnis berbasis data pada marketplace Shopee.

2. Metode Penelitian

Penelitian ini mempergunakan pendekatan kuantitatif dengan mengimplementasikan teknik hybrid data mining guna menganalisa data penjualan produk Soraya Bedsheet pada marketplace Shopee. Dataset penelitian terdiri atas 120 data produk yang diperoleh dari toko Soraya Bedsheet di Shopee dan disusun dalam bentuk dataset terstruktur untuk proses analisis. Keseluruhan proses pengolahan data dilaksanakan dengan digunakannya perangkat lunak RapidMiner.

2.1. Rancangan Penelitian

Tahapan penelitian dimulai dari pengumpulan data produk, dilanjutkan dengan preprocessing data untuk memastikan kualitas data yang digunakan. Data yang telah diproses kemudian dianalisis menggunakan tiga metode yang terintegrasi, yaitu DBSCAN untuk segmentasi produk, FP-Growth untuk analisis asosiasi, dan Random Forest untuk prediksi penjualan. Hasil dari setiap metode selanjutnya dievaluasi dan diinterpretasikan untuk memperoleh informasi yang mendukung pengambilan keputusan bisnis.

Tahapan penelitian dilaksanakan dengan bertahap mulai dari pengumpulan data, preprocessing data, segmentasi produk menggunakan DBSCAN, analisis asosiasi menggunakan FP-Growth, prediksi penjualan menggunakan Random Forest, evaluasi hasil, hingga interpretasi hasil penelitian.

Tahapan penelitian ditunjukkan pada Gambar 1 dan terdiri atas: Pengumpulan data, Preprocessing data, Segmentasi produk menggunakan DBSCAN, Analisis asosiasi menggunakan FP-Growth, Prediksi penjualan menggunakan Random Forest, Evaluasi hasil, dan Interpretasi hasil penelitian.



Gambar 1. Tahapan Penelitian

2.2. Objek Penelitian

Objek ini berupa data penjualan produk Soraya Bedsheet yang dipasarkan melalui marketplace Shopee. Dataset terdiri atas atribut nama produk, jumlah stok, jumlah produk terjual, dan harga awal produk. Atribut tersebut digunakan sebagai dasar dalam proses segmentasi, analisis asosiasi, dan prediksi penjualan.

Variabel yang digunakan ini diperlihatkan pada Tabel 1. Variabel tersebut dipilih karena mampu merepresentasikan kondisi penjualan produk dan digunakan sebagai dasar dalam proses segmentasi, analisis asosiasi, serta prediksi penjualan.

Tabel 1. Variabel Penelitian

Variabel	Pengujian
Nama Produk	Identitas produk Soraya BedSheet
Jumlah Stok	Total stok produk yang tersedia
Jumlah Terjual	Total produk yang berhasil terjual
Harga Awal	Harga produk sebelum penjualan

2.3. Preprocessing Data

Preprocessing dilaksanakan guna mempersiapkan data sebelum memasuki tahap analisis. Tahapan ini bertujuan memberikan kepastian jika data yang dipergunakan mempunyai kualitas yang baik sehingga dapat menghasilkan informasi yang lebih akurat. Pada penelitian ini, preprocessing dilakukan melalui beberapa tahap, yaitu data cleaning, pemeriksaan missing value, transformasi data, normalisasi, dan seleksi atribut.

Data cleaning dilakukan untuk mengidentifikasi serta memperbaiki data yang tidak konsisten maupun data yang berpotensi menimbulkan kesalahan pada proses analisis. Selanjutnya dilakukan pemeriksaan missing value untuk memastikan seluruh atribut memiliki nilai yang lengkap dan dapat digunakan pada tahapan pemodelan.

Tahap transformasi data dilaksanakan guna menyelaraskan format data dengan kebutuhan algoritma yang digunakan. Setelah itu, normalisasi diterapkan untuk menyamakan rentang nilai antar atribut sehingga tidak terjadi dominasi atribut tertentu dalam proses perhitungan. Tahap terakhir adalah seleksi atribut yang bertujuan memilih atribut yang relevan dengan tujuan penelitian, yaitu nama produk, jumlah stok, jumlah produk terjual, dan harga awal produk.

Hasil preprocessing digunakan sebagai data masukan pada proses segmentasi menggunakan DBSCAN, analisis asosiasi menggunakan FP-Growth, serta prediksi penjualan menggunakan Random Forest [18].

2.4. Segmentasi Produk Menggunakan DBSCAN

Segmentasi produk dilakukan menggunakan algoritma DBSCAN dengan memanfaatkan atribut jumlah stok, jumlah produk terjual, dan harga awal produk. Algoritma ini digunakan untuk mengelompokkan produk berdasarkan karakteristik penjualannya sehingga diperoleh kelompok produk yang memiliki tingkat kemiripan tertentu [5].

Proses clustering dilakukan dengan menentukan parameter epsilon (Eps) dan minimum points (MinPts). Hasil segmentasi digunakan untuk mengidentifikasi kelompok produk berdasarkan performa penjualannya, sehingga dapat membantu pelaku usaha dalam menentukan prioritas pengelolaan produk dan strategi pemasaran.

2.5. Analisis Asosiasi Menggunakan FP-Growth

Analisis asosiasi dilakukan dengan digunakannya algoritma FP-Growth dalam menemukan pola keterkaitan antarproduk dalam transaksi penjualan. Data transaksi diolah untuk menghasilkan frequent itemset yang selanjutnya digunakan dalam pembentukan association rule [3][4].

Kekuatan hubungan antarproduk dievaluasi menggunakan nilai support dan confidence. Hasil analisis dipergunakan dalam mengidentifikasi

produk yang seringkali dibeli dengan bersamaan oleh konsumen sehingga mampu dimanfaatkan dalam penyusunan strategi promosi dan bundling produk.

2.6. Prediksi Penjualan Menggunakan Random Forest

Prediksi penjualan dilakukan menggunakan algoritma Random Forest dengan memanfaatkan atribut jumlah stok, jumlah produk terjual, dan harga awal produk sebagai variabel prediktor. Model dibangun menggunakan sejumlah decision tree yang bekerja secara bersama-sama untuk menghasilkan prediksi performa penjualan produk [6].

Evaluasi model dilakukan menggunakan 10-Fold Cross Validation pada RapidMiner. Dataset dibagi menjadi sepuluh subset dengan ukuran yang relatif sama. Pada setiap iterasi, sembilan subset digunakan sebagai data pelatihan dan satu subset digunakan sebagai data pengujian secara bergantian hingga seluruh subset digunakan sebagai data uji. Kinerja model kemudian diukur menggunakan metrik accuracy, precision, recall, dan F1-score untuk memperoleh evaluasi yang lebih representatif terhadap kemampuan generalisasi model.

2.7. Perangkat dan Teknik Analisis

Keseluruhan proses pengolahan data dilaksanakan dengan digunakannya perangkat lunak RapidMiner. Perangkat lunak ini digunakan untuk mendukung proses preprocessing data, segmentasi produk, analisis asosiasi, prediksi penjualan, serta evaluasi model.

Teknik analisis dilakukan secara bertahap, dimulai dari segmentasi produk menggunakan DBSCAN, dilanjutkan dengan analisis asosiasi menggunakan FP-Growth, dan prediksi penjualan menggunakan Random Forest. Hasil dari ketiga metode tersebut kemudian dianalisis secara terpadu untuk memperoleh informasi mengenai karakteristik produk, pola pembelian konsumen, serta kecenderungan penjualan produk Soraya Bedsheet pada marketplace Shopee.

3. Hasil dan Pembahasan

3.1. Hasil Preprocessing Data

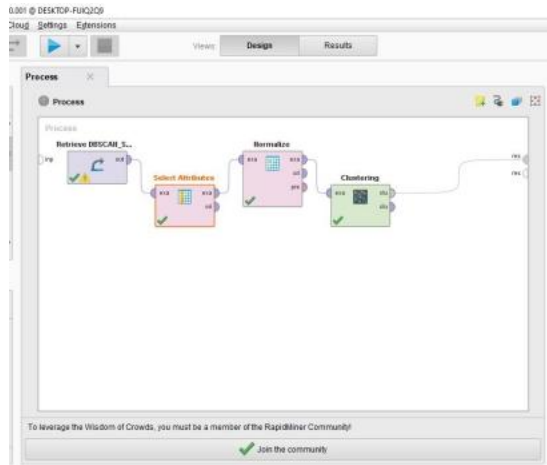
Tahap preprocessing dilaksanakan sebelum implementasi algoritma guna memberikan kepastian kualitas data yang dipergunakan dalam analisis. Dataset penelitian terdiri atas 120 data produk yang diperoleh dari marketplace Shopee. Atribut yang digunakan meliputi nama produk, jumlah stok, jumlah produk terjual, dan harga awal produk.

Hasil pemeriksaan data menunjukkan bahwasanya seluruh data dapat digunakan pada proses analisis tanpa ditemukan atribut yang memiliki nilai kosong. Selain itu, proses transformasi dan normalisasi berhasil dilakukan sehingga data siap digunakan

pada tahap segmentasi, analisis asosiasi, dan prediksi penjualan.

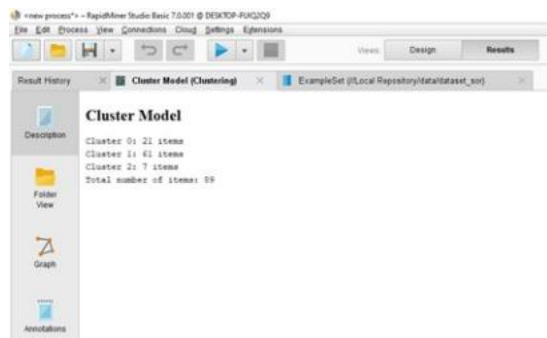
3.2. Hasil Segmentasi Produk Menggunakan DBSCAN

Proses segmentasi produk dilakukan menggunakan algoritma DBSCAN dengan memanfaatkan atribut jumlah stok, jumlah produk terjual, dan harga awal produk. Implementasi DBSCAN pada RapidMiner sebagaimana gambar 2.



Gambar 2. Proses Implementasi DBSCAN pada RapidMiner

Hasil clustering menggunakan DBSCAN ditunjukkan pada gambar 3.



Gambar 3. Hasil Clustering DBSCAN

Berdasarkan hasil clustering, produk Soraya Bedsheet terbagi ke dalam beberapa kelompok berdasarkan karakteristik penjualannya. Contoh hasil segmentasi produk ditunjukkan pada tabel 2.

Tabel 2. Contoh Hasil Segmentasi Produk

Nama Produk	Jumlah Stok	Jumlah Terjual	Harga Awal	Status Penjualan
Soraya Bedsheet Bantal 3Pillow	1	1	0.4	Sedang
Soraya Bedsheet Bedcover Eco	0.4	0.4	1	Kurang Laris
Soraya Bedsheet Sprei Florence	1	1	0.7	Laris

Perhitungan jarak antar data dilakukan dengan mengadopsi Euclidean Distance sebagaimana Persamaan (1).

$$d(x,y) = \sqrt{((x_1-y_1)^2 + (x_2-y_2)^2)} \quad (1)$$

Contoh data yang dipergunakan dalam perhitungan sebagaimana tabel 3.

Tabel 3. Contoh Data Perhitungan DBSCAN

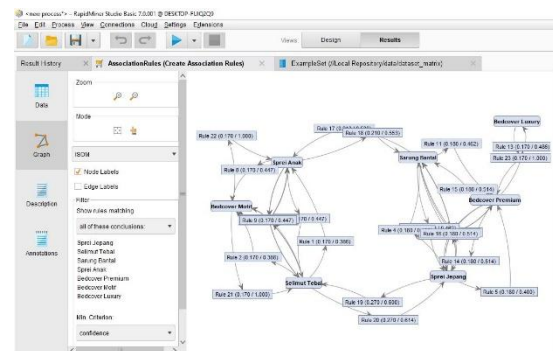
Produk	Jumlah Stok	Jumlah Terjual
P1	1	1
P2	1	0.7
P3	0.4	0.4

Hasil perhitungan menunjukkan bahwasanya jarak antara P1 dan P2 sebesar 0,30, sedangkan jarak antara P1 dan P3 sebesar 0,848.

3.3. Hasil Analisis Asosiasi Menggunakan FP-Growth

Analisis asosiasi dilakukan menggunakan algoritma FP-Growth untuk menemukan pola keterkaitan antarproduk berdasarkan data transaksi penjualan. Visualisasi association rule yang dihasilkan ditunjukkan pada gambar 4.

Perhitungan kekuatan hubungan antarproduk dilakukan menggunakan nilai support dan confidence. Tingkat kemunculan suatu item maupun gabungan beberapa item dalam keseluruhan transaksi diperlihatkan dari nilai *support*, sedangkan tingkat keyakinan pada suatu aturan asosiasi diukur dengan menggunakan nilai *confidence*.



Gambar 4. Visualisasi Association Rule FP-Growth

Support(A)=Jumlah transaksi yang mengandung A / Total transaksi × 100% (2)

Confidence(A→B)=Support(A∩B) / Support(A) × 100% (3)

Hasil association rule ditunjukkan pada gambar 5.



Gambar 5. Hasil Association Rule FP-Growth

Contoh aturan asosiasi yang diperoleh ditunjukkan pada tabel 4.

Tabel 4. Tabel Contoh Hasil Association Rule

Produk A	Produk B	Support	Confidence
Sprei Florence	Bedcover Eco	0,42	0,78
Sprei Katun Gold	Sarung Bantal	0,36	0,71
Sprei Jepang	Selimut Salju	0,31	0,69

Contoh data transaksi yang digunakan dalam perhitungan support dan confidence ditunjukkan pada tabel 5.

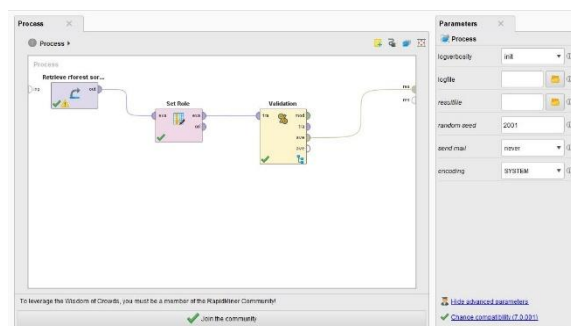
Tabel 5. Contoh Data Transaksi

Transaksi	Produk
T1	Sprei, Bedcover
T2	Sprei, Sarung Bantal
T3	Sprei, Bedcover
T4	Bedcover, Selimut
T5	Sprei, Bedcover

Hasil perhitungan menunjukkan nilai support produk Sprei sebesar 80% dan nilai confidence untuk aturan Sprei → Bedcover sebesar 75%.

3.4 Hasil Prediksi Penjualan Menggunakan Random Forest

Prediksi penjualan dilakukan menggunakan algoritma Random Forest. Implementasi model pada RapidMiner sebagaimana gambar 6.



Gambar 6. Implementasi Random Forest pada RapidMiner

Hasil evaluasi model ditunjukkan pada Gambar 7.

Parameter	Nilai
Accuracy	91,67%
Precision	91,62%
Recall	91,67%
F1-Score	91,50%

Gambar 7. Hasil Evaluasi Random Forest

Nilai evaluasi model Random Forest sebagaimana tabel 6.

Tabel 6. Hasil Evaluasi Model Random Forest

Parameter	Nilai
Accuracy	91,67%
Precision	91,62%
Recall	91,67%
F1-Score	91,50%

Hasil evaluasi menggunakan 10-fold Cross Validation menunjukkan bahwa model Random Forest memperoleh rata-rata accuracy sebesar $91,67\% \pm 5,27\%$. Nilai tersebut menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik dalam memprediksi status penjualan produk pada seluruh kelas.

3.5 Pembahasan

Hasil segmentasi menggunakan DBSCAN menunjukkan bahwasanya produk Soraya Bedsheet memiliki karakteristik penjualan yang berbeda sehingga dapat dikelompokkan berdasarkan performa penjualannya. Informasi tersebut dapat digunakan untuk mengidentifikasi produk yang memiliki tingkat penjualan tinggi maupun produk yang memerlukan perhatian lebih dalam pengelolaan stok dan strategi pemasaran.

Analisis asosiasi menggunakan FP-Growth berhasil mengidentifikasi keterkaitan antarproduk dalam transaksi konsumen. Pola hubungan antarproduk menunjukkan bahwasanya konsumen cenderung membeli beberapa produk secara bersamaan dalam satu transaksi. Informasi tersebut dapat dimanfaatkan sebagai dasar dalam penyusunan strategi bundling produk dan promosi silang [14].

Pada tahap prediksi, algoritma Random Forest dievaluasi menggunakan 10-Fold Cross Validation untuk memperoleh estimasi performa model yang lebih objektif. Hasil evaluasi menunjukkan bahwa model memperoleh rata-rata accuracy sebesar $91,67\% \pm 5,27\%$ (micro accuracy 91,67%), yang menunjukkan kemampuan model dalam mengklasifikasikan status penjualan produk dengan tingkat ketepatan yang baik. Nilai tersebut mengindikasikan bahwa atribut jumlah stok, jumlah produk terjual, harga produk, rating, dan jumlah ulasan memberikan kontribusi terhadap pembentukan model prediksi performa penjualan.

Hasil penelitian ini sejalan dengan penelitian Irawan et al. [5] yang menunjukkan bahwa algoritma Random Forest memiliki kemampuan yang baik dalam melakukan klasifikasi dan prediksi pada data penjualan. Namun, berbeda dengan penelitian terdahulu yang hanya memanfaatkan Random Forest untuk prediksi, penelitian ini mengintegrasikan DBSCAN dan FP-Growth dalam satu pendekatan hybrid data mining. Integrasi tersebut memungkinkan dihasilkannya informasi yang lebih komprehensif, meliputi segmentasi produk, pola pembelian konsumen, serta prediksi status penjualan sebagai dasar pengambilan keputusan bisnis.

Berdasarkan confusion matrix, model mampu mengklasifikasikan seluruh data pada kelas Tidak Laris dengan nilai recall sebesar 100%, sedangkan kelas Laris, Kurang Laris, dan Sangat Laris memperoleh nilai recall masing-masing sebesar 83,33%, 90,00%, dan 93,33%. Kesalahan klasifikasi masih ditemukan pada beberapa data yang memiliki karakteristik penjualan yang serupa, terutama antara kelas Laris dan Sangat Laris. Kondisi tersebut menunjukkan bahwa model telah mampu melakukan generalisasi terhadap data dengan baik, namun masih terdapat sedikit variasi karakteristik antarproduk yang menyebabkan terjadinya salah klasifikasi [16].

Secara keseluruhan, integrasi DBSCAN, FP-Growth, dan Random Forest mampu menghasilkan informasi yang saling melengkapi mengenai segmentasi produk, pola pembelian konsumen, dan prediksi status penjualan. Hasil evaluasi menggunakan 10-Fold Cross Validation menunjukkan bahwa model memiliki kemampuan prediksi yang baik dengan accuracy sebesar 91,67%, sehingga pendekatan hybrid data mining yang diusulkan dapat dimanfaatkan sebagai pendukung pengambilan keputusan bisnis berbasis data pada marketplace Shopee.

4. Kesimpulan

Penelitian ini berhasil mengimplementasikan pendekatan hybrid data mining dengan mengintegrasikan algoritma DBSCAN, FP-Growth, dan Random Forest untuk menganalisis serta memprediksi penjualan produk Soraya Bedsheet pada marketplace Shopee. Pendekatan tersebut mampu menghasilkan informasi yang saling melengkapi, yaitu segmentasi produk berdasarkan karakteristik penjualan, identifikasi pola pembelian konsumen melalui analisis asosiasi, serta prediksi status penjualan produk dalam satu alur analisis yang terpadu.

Hasil penelitian menunjukkan bahwa DBSCAN mampu mengelompokkan produk berdasarkan karakteristik penjualannya sehingga memudahkan identifikasi produk dengan performa tinggi maupun rendah. FP-Growth berhasil menemukan pola

keterkaitan antarproduk yang sering muncul dalam transaksi yang sama sehingga dapat dimanfaatkan untuk strategi bundling dan promosi silang. Sementara itu, Random Forest yang dievaluasi menggunakan 10-fold Cross Validation memperoleh accuracy rata-rata sebesar 91,67%, yang menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik dalam memprediksi status penjualan produk.

Secara keseluruhan, integrasi ketiga algoritma mampu menghasilkan informasi yang lebih komprehensif dibandingkan penggunaan satu algoritma secara terpisah. Informasi tersebut dapat dimanfaatkan sebagai dasar dalam pengelolaan stok, penyusunan strategi promosi, serta pengambilan keputusan bisnis berbasis data pada marketplace Shopee [15].

Penelitian selanjutnya dapat menggunakan dataset dengan jumlah data yang lebih besar, menambahkan atribut penjualan yang lebih beragam, membandingkan performa Random Forest dengan algoritma klasifikasi lain seperti XGBoost, Support Vector Machine (SVM), atau Gradient Boosting, serta menguji pendekatan hybrid ini pada dataset dari marketplace atau kategori produk yang berbeda untuk memperoleh gambaran performa yang lebih luas.

Ucapan Terimakasih

Keberhasilan penyelesaian kajian ini tidak terlepas dari berbagai dukungan yang diberikan oleh Institut Bisnis dan Teknologi Pelita Indonesia dan Universitas Bung Hatta selama proses kajian berlangsung. Penulis juga memberikan sapaian apresiasi kepada seluruh pihak yang telah berpartisipasi dalam pengumpulan data, pengolahan data, serta penyusunan artikel ilmiah ini. Berbagai kontribusi tersebut memberikan bantuan yang sangat berarti sehingga seluruh rangkaian artikel ini mampu terlaksana maupun diselesaikan dengan optimall.

Daftar Rujukan

- [1] Amelia, R., Darmansyah, D., dan Rismadin, A. 2024. Perbandingan Algoritma Apriori dan FP-Growth dalam Pengaplikasian Market Basket Analysis untuk Strategi Bisnis Retail. *Building of Informatics, Technology and Science (BITS)*. 6(1): 279-288.
- [2] Hidayatullah, S., Umam, K., dan Fauziyah, F. 2023. Implementasi Algoritma FP-Growth untuk Analisis Pola Pembelian Konsumen pada Marketplace. *Jurnal Media Informatika Budidarma*. 7(2): 944-952.
- [3] Irawan, F., Amaluddin, F., dan Wijayanti, A. 2025. E-Commerce Customer Segmentation Based on RFM Analysis Using DBSCAN Algorithm to Improve Marketing Strategy. *Jurnal EMT KITA*. 9(4): 1850-1859.
- [4] Kotu, V. dan Deshpande, B. 2022. *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. Amsterdam: Elsevier.
- [5] Kusriani dan Luthfi, E.T. 2021. *Algoritma Data Mining*. Yogyakarta: Andi Publisher.

-
- [6] Larose, D.T. dan Larose, C.D. 2021. *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken: Wiley.
- [7] Prasetyo, E. 2021. *Data Mining Konsep dan Aplikasi Menggunakan RapidMiner*. Yogyakarta: Andi Publisher.
- [8] Saputra, J.P.B., Rahayu, S.A., dan Hariguna, T. 2023. Market Basket Analysis Using FP-Growth Algorithm to Design Marketing Strategy by Determining Consumer Purchasing Patterns. *Journal of Applied Data Sciences*. 4(1): 38-49.
- [9] Tan, P.N., Steinbach, M., dan Kumar, V. 2021. *Introduction to Data Mining*. Boston: Pearson Education.
- [10] Witten, I.H., Frank, E., Hall, M.A., dan Pal, C.J. 2021. *Data Mining: Practical Machine Learning Tools and Techniques*. Cambridge: Morgan Kaufmann.
- [11] Yuliana, S. dan Prasetyo, A. 2022. Product Segmentation Analysis Using DBSCAN Method in E-Commerce Data. *Journal of Soft Computing Exploration*. 3(4): 201-210.
- [12] Wahyudi, R.N., Herwindiati, D.E., dan Hendryli, J. 2023. Market Basket Analysis dengan Perbandingan Metode Apriori dan FP-Growth Pada Data Transaksi XYZ. *Jurnal Ilmu Komputer dan Sistem Informasi*. 11(1).
- [13] Hidayat, T., Nugroho, A., dan Saputra, D. 2022. Customer Segmentation Using DBSCAN Algorithm for Retail Data Analysis. *International Journal of Advanced Computer Science and Applications*. 13(5): 214-221.
- [14] Putra, A., et al. 2024. Hybrid Data Mining for E-Commerce Sales Prediction Using Clustering and Classification. *International Journal of Intelligent Systems*.
- [15] Kumar, S., et al. 2024. Product Recommendation and Customer Segmentation Using Data Mining Techniques. *Applied Soft Computing*.
- [16] Nguyen, T., et al. 2023. Machine Learning Based Sales Prediction in E-Commerce Platforms. *Expert Systems with Applications*.
- [17] Johannes, R., & Alamsyah, A. 2021. Sales Prediction Model Using Classification Decision Tree Approach for Small Medium Enterprise Based on Indonesian E-Commerce Data. *arXiv preprint arXiv:2103.03117*.
- [18] Han, J., Pei, J., Yin, Y., & Mao, R. 2004. Mining Frequent Patterns without Candidate Generation: A Frequent Pattern Tree Approach. *Data Mining and Knowledge Discovery*. 8(1): 53–87.