

BAB III

METODOLOGI PENELITIAN

3.1. Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis eksperimen sebagai metode utama. Pendekatan kuantitatif dipilih karena penelitian ini berfokus pada analisis data numerik yang dapat diolah menggunakan model komputasi. Penelitian ini bertujuan untuk menguji dan membandingkan kinerja dua algoritma *machine learning* yang populer, yaitu *Support Vector Machine (SVM)* dan *Random Forest (RF)*, dalam mengklasifikasikan produk yang paling diminati di Toko Sembako Mentari berdasarkan data transaksi harian. Data yang digunakan dalam penelitian ini adalah data transaksi harian yang memuat informasi mengenai produk, harga, jumlah pembelian, dan waktu transaksi. Dengan demikian, penelitian ini mengandalkan analisis statistik yang berbasis pada data numerik untuk mengidentifikasi pola-pola dalam data yang dapat digunakan untuk mengambil keputusan bisnis yang lebih baik.

3.1.1. Pendekatan Kuantitatif dalam Penelitian

Pendekatan kuantitatif merujuk pada penelitian yang berfokus pada pengumpulan dan analisis data numerik untuk mengevaluasi variabel yang telah ditentukan, dengan tujuan menghasilkan hasil yang objektif, terukur, dan dapat dipertanggungjawabkan. Dalam konteks penelitian ini, pendekatan kuantitatif memungkinkan untuk mengukur kinerja kedua algoritma *machine learning*, *SVM*

dan *RF*, berdasarkan metrik evaluasi yang dapat dihitung dengan jelas seperti akurasi, presisi, recall, dan F1-score.

Penggunaan pendekatan kuantitatif ini sangat relevan dalam penelitian ini karena penelitian bertujuan untuk membandingkan dua algoritma berdasarkan data numerik transaksi harian, yang berisi informasi terstruktur mengenai jumlah pembelian, harga produk, dan waktu transaksi. Data transaksi yang dikumpulkan diharapkan memiliki variasi yang cukup sehingga dapat diuji dengan dua algoritma yang berbeda untuk mengetahui mana yang lebih efektif dalam mengklasifikasikan produk yang diminati pelanggan.

Keuntungan menggunakan pendekatan kuantitatif dalam penelitian ini adalah kemampuan untuk memberikan hasil yang dapat diukur secara objektif dan dapat digunakan untuk menilai kinerja kedua algoritma yang diuji. Sebagai contoh, akurasi dan F1-score adalah metrik yang memberikan ukuran kuantitatif dari seberapa baik masing-masing algoritma dapat mengklasifikasikan produk yang paling diminati berdasarkan data transaksi yang ada.

3.1.2. Eksperimen sebagai Jenis Penelitian

Penelitian ini menggunakan eksperimen sebagai jenis penelitian utama, di mana eksperimen dilakukan untuk menguji efektivitas algoritma *machine learning* dalam klasifikasi produk. Dalam eksperimen ini, peneliti melakukan perbandingan antara dua algoritma, yaitu *Support Vector Machine (SVM)* dan *Random Forest (RF)*, dalam hal keakuratan klasifikasi produk paling diminati di Toko Sembako Mentari. Jenis penelitian eksperimen ini memungkinkan peneliti untuk mengontrol

variabel-variabel tertentu yang dapat mempengaruhi hasil klasifikasi, serta memberikan komparasi objektif antara kedua algoritma.

Eksperimen ini dilakukan dengan membangun model untuk setiap algoritma, yaitu dengan mengimplementasikan *SVM* dan *RF* menggunakan data transaksi harian yang telah disiapkan. Model yang dibangun kemudian diuji untuk mengklasifikasikan produk yang paling diminati oleh konsumen, berdasarkan kriteria yang telah ditetapkan. Dalam eksperimen ini, data transaksi yang terkumpul akan diproses dan dimasukkan ke dalam masing-masing model, yang kemudian menghasilkan output berupa produk yang diklasifikasikan sebagai diminati atau tidak.

Secara keseluruhan, eksperimen ini mengandalkan metode ilmiah yang memadukan algoritma *machine learning* dengan data yang terstruktur untuk menghasilkan kesimpulan yang berbasis pada analisis statistik. Eksperimen ini akan memberikan wawasan yang lebih dalam mengenai kinerja dua algoritma yang diuji, serta membantu dalam menentukan algoritma mana yang lebih efektif untuk mengklasifikasikan produk yang diminati di pasar ritel dengan data yang terbatas.

3.1.3. Kinerja Algoritma *Machine Learning* dalam Klasifikasi Produk

Tujuan utama dari penelitian ini adalah untuk menguji dan membandingkan kinerja algoritma *SVM* dan *RF* dalam mengklasifikasikan produk yang paling diminati di Toko Sembako Mentari. Sebagai bagian dari eksperimen, kedua algoritma ini diuji pada dataset yang mencakup informasi transaksi harian yang meliputi berbagai produk sembako dengan atribut seperti jumlah pembelian, harga satuan, dan frekuensi transaksi. Hasil evaluasi akan mengukur seberapa baik

algoritma tersebut dapat mengidentifikasi produk yang paling diminati oleh pelanggan berdasarkan pola-pola yang ada dalam data transaksi.

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang digunakan untuk menemukan hyperplane terbaik yang memisahkan kelas-kelas dalam data. Dalam konteks penelitian ini, *SVM* akan digunakan untuk mengklasifikasikan produk menjadi kategori produk yang paling diminati dan tidak diminati berdasarkan fitur-fitur seperti jumlah pembelian, harga produk, dan waktu transaksi. *SVM* bekerja dengan memaksimalkan margin antara kelas-kelas data sehingga dapat menghasilkan keputusan yang lebih tepat dan akurasi yang tinggi.

Random Forest (RF) adalah algoritma berbasis *ensemble learning* yang menggunakan banyak pohon keputusan untuk menghasilkan prediksi yang lebih akurat. Pada penelitian ini, *Random Forest* digunakan untuk mengklasifikasikan produk yang paling diminati dengan cara membuat banyak pohon keputusan berdasarkan data transaksi yang ada. Setiap pohon dalam hutan *RF* memberikan hasil prediksi, dan hasil prediksi dari seluruh pohon kemudian digabungkan menggunakan *majority voting* untuk menghasilkan keputusan final.

3.2. Variabel Penelitian

Penelitian ini bertujuan untuk menganalisis produk yang paling diminati berdasarkan data transaksi harian yang tercatat di Toko Sembako Mentari. Dalam upaya mengklasifikasikan produk yang paling diminati, tiga variabel utama yang menjadi fokus analisis adalah Jumlah Pembelian, Harga Satuan, dan Frekuensi Transaksi. Ketiga variabel ini dianggap sebagai faktor penting yang dapat mempengaruhi minat konsumen terhadap produk-produk di toko tersebut. Analisis

terhadap variabel-variabel ini akan dilakukan menggunakan teknik machine learning yang telah dijelaskan pada bagian sebelumnya, yaitu *Support Vector Machine (SVM)* dan *Random Forest (RF)*.

3.2.1. Jumlah Pembelian

Variabel Jumlah Pembelian mengacu pada banyaknya produk yang dibeli oleh konsumen dalam satu transaksi. Data ini diambil langsung dari catatan penjualan harian Toko Sembako Mentari dan mencatat berapa unit produk yang dibeli oleh setiap pelanggan. Variabel ini memainkan peran penting dalam menentukan tingkat minat konsumen terhadap suatu produk. Semakin banyak produk yang dibeli dalam transaksi, semakin tinggi kemungkinan bahwa produk tersebut termasuk dalam kategori produk yang diminati. Oleh karena itu, variabel Jumlah Pembelian menjadi indikator utama dalam analisis klasifikasi produk, karena menunjukkan tingkat permintaan yang nyata di pasar.

Secara statistik, variabel ini akan digunakan untuk menghitung frekuensi pembelian rata-rata per periode tertentu. Misalnya, jika suatu produk dibeli lebih sering dalam bulan tertentu, maka produk tersebut dapat dikategorikan sebagai produk yang memiliki minat konsumen tinggi. Penggunaan data ini dalam model *SVM* dan *RF* memungkinkan analisis yang lebih mendalam terhadap pola pembelian dan hubungan antara jumlah pembelian dengan variabel lain seperti harga dan waktu pembelian.

3.2.2. Harga Satuan

Variabel Harga Satuan merujuk pada harga per unit produk yang dijual di Toko Sembako Mentari. Harga merupakan salah satu faktor yang sangat

mempengaruhi keputusan pembelian konsumen. Pada umumnya, produk dengan harga yang lebih rendah cenderung lebih diminati oleh konsumen, meskipun ada juga produk premium dengan harga lebih tinggi yang tetap memiliki minat yang tinggi. Variabel Harga Satuan digunakan dalam analisis untuk mengevaluasi bagaimana fluktuasi harga dapat memengaruhi volume pembelian.

Dalam konteks penelitian ini, harga produk akan dibandingkan dengan Jumlah Pembelian untuk melihat apakah ada hubungan antara harga dan tingkat permintaan. Produk dengan harga stabil dan rendah mungkin akan lebih sering dibeli oleh konsumen dibandingkan produk dengan harga yang sangat fluktuatif. Oleh karena itu, Harga Satuan menjadi salah satu variabel penting dalam model klasifikasi.

3.2.3. Frekuensi Transaksi

Variabel Frekuensi Transaksi mengukur seberapa sering produk tertentu dibeli oleh konsumen selama periode tertentu. Frekuensi transaksi memberikan gambaran yang jelas tentang tingkat popularitas produk dalam jangka waktu tertentu, seperti per minggu, per bulan, atau bahkan per hari. Produk dengan frekuensi transaksi tinggi menunjukkan bahwa produk tersebut sering dibeli dan memiliki permintaan yang konsisten, yang menunjukkan bahwa produk tersebut adalah salah satu yang paling diminati oleh konsumen.

Variabel ini sangat penting untuk model *Random Forest (RF)*, karena *RF* mampu menangani data dengan atribut yang berinteraksi secara non-linear, seperti hubungan antara jumlah pembelian, harga, dan waktu transaksi. Frekuensi transaksi juga akan menjadi salah satu input dalam model untuk memprediksi produk yang

paling diminati, karena produk yang sering dibeli cenderung lebih menarik bagi pelanggan.

3.3. Teknik Pengumpulan Data

Pengumpulan data merupakan tahap awal yang sangat penting dalam setiap penelitian untuk memastikan bahwa data yang diperoleh relevan, valid, dan dapat digunakan untuk analisis lebih lanjut. Dalam penelitian ini, teknik pengumpulan data dilakukan melalui observasi langsung terhadap transaksi harian di Toko Sembako Mentari. Data yang dikumpulkan mencakup informasi yang terkait langsung dengan produk yang dijual, jumlah pembelian, harga satuan, dan tanggal transaksi. Dengan mengandalkan data transaksi harian yang akurat, penelitian ini bertujuan untuk menganalisis produk mana yang paling diminati oleh konsumen menggunakan metode *machine learning* seperti *Support Vector Machine* (SVM) dan *Random Forest* (RF).

3.3.1. Observasi Langsung Transaksi Harian

Observasi langsung terhadap transaksi harian di Toko Sembako Mentari dilakukan untuk memperoleh data yang mencerminkan perilaku konsumen dalam memilih produk. Toko sembako ini menyediakan berbagai produk kebutuhan pokok yang sering dibeli oleh masyarakat, seperti beras, minyak, gula, dan sebagainya. Data yang diperoleh dari setiap transaksi harian akan mencakup empat informasi utama, yaitu:

1. Nama Produk: Menyebutkan jenis produk yang dijual pada setiap transaksi.
2. Jumlah Pembelian: Banyaknya unit produk yang dibeli oleh pelanggan dalam satu transaksi.

3. Harga Satuan: Harga per unit produk yang dibeli.
4. Tanggal Transaksi: Tanggal dimana transaksi penjualan terjadi.

Observasi ini dilakukan setiap hari selama periode yang telah ditentukan, yaitu enam bulan terakhir. Seluruh data transaksi yang dikumpulkan akan diproses untuk menghasilkan dataset yang terstruktur, yang nantinya akan digunakan dalam tahap analisis dan pemodelan.

Penting untuk diingat bahwa dalam penelitian ini, data transaksi yang diperoleh dari observasi langsung akan menjadi dasar bagi penerapan algoritma *machine learning* untuk mengklasifikasikan produk yang paling diminati. Dengan menggunakan data yang valid dan terstruktur, algoritma SVM dan RF dapat melakukan analisis yang lebih mendalam dan menghasilkan prediksi yang lebih akurat.

3.3.2. Pengolahan dan Penyusunan Data

Setelah data transaksi harian dikumpulkan, langkah berikutnya adalah mengolah dan menyusun data tersebut dalam format yang terstruktur. Pengolahan data ini dilakukan agar data menjadi lebih mudah dipahami dan siap digunakan untuk analisis lebih lanjut. Proses pengolahan data mencakup beberapa tahapan, yaitu:

1. Pembersihan Data: Tahap pertama dalam pengolahan data adalah membersihkan data dari kesalahan atau duplikasi. Hal ini penting untuk memastikan bahwa data yang digunakan tidak mengandung kesalahan penulisan atau duplikasi transaksi, yang dapat mempengaruhi hasil analisis.

Data yang tidak lengkap atau memiliki nilai kosong (*missing values*) juga akan diperiksa dan diatasi pada tahap ini.

2. **Normalisasi Data:** Setelah data dibersihkan, langkah berikutnya adalah normalisasi. Proses ini dilakukan untuk memastikan bahwa setiap atribut dalam data memiliki skala yang seimbang. Normalisasi sangat penting terutama jika data mengandung atribut numerik dengan skala yang berbeda-beda, seperti jumlah pembelian yang bisa memiliki nilai lebih besar dibandingkan harga satuan. Dengan melakukan normalisasi, setiap atribut akan memiliki rentang nilai yang seragam, yang memungkinkan model *machine learning* untuk melakukan analisis dengan lebih baik.
3. **Penggabungan Data:** Selain pembersihan dan normalisasi, penggabungan data dari berbagai sumber yang relevan juga dilakukan. Misalnya, data tentang produk, harga, dan pembelian mungkin berasal dari beberapa sistem atau sumber yang berbeda, seperti sistem *POS* (Point of Sale) dan catatan manual. Oleh karena itu, penggabungan data dilakukan untuk menyatukan informasi ini ke dalam satu dataset yang komprehensif.
4. **Pembuatan Fitur Baru:** Pada tahap ini, peneliti dapat membuat fitur baru yang dapat menambah wawasan dalam analisis. Sebagai contoh, fitur baru seperti frekuensi pembelian rata-rata per bulan dapat ditambahkan untuk memberikan informasi lebih lanjut mengenai pola konsumsi konsumen.

Setelah tahap pengolahan selesai, data yang telah disusun ini akan siap digunakan dalam tahap pemodelan dan evaluasi algoritma *machine learning*. Dataset yang telah terstruktur akan mencakup informasi seperti nama produk,

jumlah pembelian, harga satuan, dan tanggal transaksi yang dapat digunakan untuk mengklasifikasikan produk yang paling diminati.

3.3.3. Keakuratan Data dan Penggunaan untuk Klasifikasi

Keakuratan data yang diperoleh sangat penting untuk memastikan bahwa hasil dari model *machine learning* yang diterapkan akurat dan dapat diandalkan. Setiap langkah pengolahan data harus dilakukan dengan hati-hati untuk menghindari adanya kesalahan atau ketidaktepatan data yang dapat merusak hasil penelitian.

Penggunaan data yang telah diproses dan disusun dengan baik akan memungkinkan algoritma *SVM* dan *RF* untuk mengklasifikasikan produk yang paling diminati dengan akurasi yang tinggi. Pengolahan data yang benar akan meminimalkan kemungkinan kesalahan dalam klasifikasi, seperti *false positives* (produk yang diklasifikasikan sebagai diminati tetapi tidak laris) dan *false negatives* (produk yang diklasifikasikan tidak diminati tetapi sebenarnya diminati oleh konsumen).

Selain itu, penting untuk memastikan bahwa dataset yang digunakan memiliki keberagaman yang representatif terhadap produk yang dijual di Toko Sembako Mentari. Dengan demikian, model *machine learning* dapat belajar dari berbagai pola transaksi yang terjadi, dan tidak hanya fokus pada pola yang terbatas. Hal ini akan memungkinkan model untuk mengklasifikasikan produk dengan lebih efektif, berdasarkan berbagai faktor yang ada, seperti harga, jumlah pembelian, dan frekuensi transaksi.

3.4. Instrumen Penelitian

Instrumen penelitian yang digunakan dalam penelitian ini terdiri dari tiga komponen utama yang mendukung proses klasifikasi dan evaluasi hasil model. Komponen-komponen tersebut meliputi dataset transaksi harian, perangkat lunak Python dengan pustaka *scikit-learn* untuk implementasi algoritma *SVM* dan *RF*, serta metrik evaluasi yang digunakan untuk menilai kinerja model klasifikasi. Setiap instrumen ini memiliki peran yang sangat penting dalam memastikan bahwa penelitian dapat berjalan dengan baik, menghasilkan hasil yang akurat, dan memberikan kontribusi praktis kepada pemilik Toko Sembako Mentari.

3.4.1. Dataset Transaksi Harian

Dataset transaksi harian merupakan sumber utama data untuk proses klasifikasi dalam penelitian ini. Data yang dikumpulkan mencakup informasi transaksi yang terjadi di Toko Sembako Mentari selama enam bulan terakhir, dengan fokus pada produk yang dijual, jumlah pembelian, harga satuan, dan tanggal transaksi. Setiap transaksi yang tercatat menjadi bagian dari dataset yang akan digunakan untuk mengidentifikasi pola-pola yang menunjukkan produk paling diminati oleh konsumen.

Data transaksi harian ini memiliki peran penting dalam membangun model klasifikasi yang akurat. Oleh karena itu, kualitas data menjadi faktor yang sangat krusial. Data harus mencakup semua informasi yang relevan, seperti jenis produk, jumlah pembelian, harga per unit, dan tanggal transaksi, agar model *machine learning* dapat melakukan klasifikasi produk dengan baik. Selain itu, data yang dikumpulkan harus melalui proses pengolahan yang sesuai, seperti pembersihan

data dari duplikasi, pengisian nilai yang hilang, dan normalisasi data, untuk memastikan bahwa model dapat bekerja dengan optimal.

3.4.2. Perangkat Lunak Python

Perangkat lunak Python digunakan untuk implementasi algoritma *SVM* dan *RF* dalam penelitian ini. Python dipilih karena kemampuannya yang kuat dalam pemrosesan data dan *machine learning*, serta didukung oleh pustaka-pustaka seperti *scikit-learn*, *pandas*, *matplotlib*, dan *seaborn* yang memudahkan analisis data dan visualisasi hasil. Pustaka *scikit-learn* adalah pustaka utama yang digunakan untuk implementasi *Support Vector Machine* (SVM) dan *Random Forest* (RF) dalam penelitian ini.

Dengan menggunakan Python dan *scikit-learn*, peneliti dapat mengimplementasikan berbagai algoritma pembelajaran mesin dengan mudah, mulai dari persiapan data, pemodelan, hingga evaluasi hasil. *Scikit-learn* menyediakan berbagai alat untuk membangun dan menguji model klasifikasi, yang memudahkan peneliti dalam memilih algoritma yang sesuai dan mengoptimalkan kinerjanya.

3.4.3. Metrik Evaluasi

Metrik evaluasi digunakan untuk mengukur kinerja model yang dibangun menggunakan algoritma *SVM* dan *RF*. Metrik ini penting untuk menilai sejauh mana model dapat mengklasifikasikan produk dengan benar dan mengidentifikasi produk yang paling diminati berdasarkan data transaksi yang ada. Beberapa metrik evaluasi yang digunakan dalam penelitian ini antara lain akurasi, presisi, recall, dan *F1-score*.

1. Akurasi: Akurasi mengukur seberapa banyak prediksi yang benar dibandingkan dengan total prediksi. Metrik ini memberikan gambaran umum tentang kinerja model, tetapi bisa jadi tidak cukup informatif jika data yang digunakan tidak seimbang.
2. Presisi: Presisi mengukur seberapa banyak dari prediksi positif yang benar-benar positif. Dalam konteks ini, presisi digunakan untuk menilai seberapa tepat model dalam mengklasifikasikan produk yang diminati.
3. Recall: Recall mengukur kemampuan model untuk mendeteksi seluruh produk yang benar-benar diminati. Metrik ini penting untuk memastikan bahwa model tidak melewatkan produk yang seharusnya diklasifikasikan sebagai diminati.
4. F1-score: *F1-score* adalah ukuran yang menggabungkan presisi dan recall untuk memberikan gambaran yang lebih seimbang tentang kinerja model, terutama ketika ada *trade-off* antara keduanya.

3.5. Algoritma Machine Learning

Pada bab ini, penjelasan algoritma Support Vector Machine dan Random Forest disajikan dalam bentuk tahapan metodologis penelitian. Perhitungan manual yang ditampilkan bertujuan untuk memberikan ilustrasi mengenai cara kerja dasar algoritma dalam proses klasifikasi. Perhitungan manual tersebut tidak digunakan sebagai dasar utama penarikan kesimpulan penelitian. Hasil akhir penelitian ditetapkan berdasarkan implementasi Python pada data transaksi harian yang telah diproses secara menyeluruh dan disajikan pada Bab IV.

3.5.1. Persiapan Data

Pembersihan Data, Pastikan data tidak mengandung nilai kosong (missing values) atau duplikat.

Tabel 3. 1 Contoh Data Transaksi Toko Sembako Mentari

Priode Penjualan	Nama Produk	Jumlah Pembelian	Frekuensi Transaksi	Harga Satuan
Bulan_Pertama	Beras	71	2.366666667	Rp 63,000.00
	Minyak Goreng	44	1.466666667	Rp 15,000.00
	Gula Pasir	46	1.533333333	Rp 14,000.00
	Garam	24	0.8	Rp 4,000.00
	Telur	246	8.2	Rp 2,000.00
	Gas Elpiji	21	0.7	Rp 22,000.00
	Rokok	114	3.8	Rp 25,000.00
	Aqua	50	1.666666667	Rp 5,000.00
	Mie Instan	173	5.766666667	Rp 3,000.00
	Sabun Batang	38	1.266666667	Rp 4,000.00
	Tepung	30	1	Rp 10,000.00
	Kopi	30	1	Rp 2,000.00
	Susu	41	1.366666667	Rp 12,000.00
	Detergen	30	1	Rp 15,000.00
	Kecap	132	4.4	Rp 3,000.00
Bulan_Kedua	Beras	70	2.333333333	Rp 63,000.00
	Minyak Goreng	41	1.366666667	Rp 15,000.00
	Gula Pasir	47	1.566666667	Rp 14,000.00
	Garam	25	0.833333333	Rp 4,000.00
	Telur	269	8.966666667	Rp 2,000.00
	Gas Elpiji	22	0.733333333	Rp 22,000.00
	Rokok	113	3.766666667	Rp 25,000.00
	Aqua	52	1.733333333	Rp 5,000.00
	Mie Instan	207	6.9	Rp 3,000.00
	Sabun Batang	36	1.2	Rp 4,000.00
	Tepung	32	1.066666667	Rp 10,000.00
	Kopi	31	1.033333333	Rp 2,000.00
	Susu	45	1.5	Rp 12,000.00
	Detergen	31	1.033333333	Rp 15,000.00

	Kecap	127	4.2333333333	Rp 3,000.00
Bulan_Ketiga	Beras	75	2.5	Rp 63,000.00
	Minyak Goreng	47	1.5666666667	Rp 15,000.00
	Gula Pasir	42	1.4	Rp 14,000.00
	Garam	26	0.8666666667	Rp 4,000.00
	Telur	262	8.7333333333	Rp 2,000.00
	Gas Elpiji	23	0.7666666667	Rp 22,000.00
	Rokok	112	3.7333333333	Rp 25,000.00
	Aqua	51	1.7	Rp 5,000.00
	Mie Instan	192	6.4	Rp 3,000.00
	Sabun Batang	37	1.2333333333	Rp 4,000.00
	Tepung	31	1.0333333333	Rp 10,000.00
	Kopi	31	1.0333333333	Rp 2,000.00
	Susu	44	1.4666666667	Rp 12,000.00
	Detergen	31	1.0333333333	Rp 15,000.00
	Kecap	135	4.5	Rp 3,000.00

3.5.2. Pemisahan Dataset

Pemisahan dataset menjadi 80% untuk pelatihan (training data) dan 20% untuk pengujian (testing data) dengan menggunakan train-test split. Fitur yang Digunakan, Pilih fitur yang relevan untuk algoritma SVM dan RF, seperti Jumlah Pembelian, Frekuensi Transaksi, Harga Satuan, dan Label Diminati.

Tabel 3. 2 Contoh Data Training Transaksi Toko Sembako Mentari

Nama Produk	Jumlah Pembelian	Frekuensi Transaksi	Harga Satuan	Label Diminati
Minyak Goreng	47	1.566667	Rp15.000	0
Garam	24	0.8	Rp4.000	0
Aqua	49	1.633333	Rp5.000	0
Gula Pasir	47	1.566667	Rp14.000	0
Mie Instan	173	5.766667	Rp3.000	1
Rokok	114	3.8	Rp25.000	1
Tepung	31	1.033333	Rp10.000	0
Telur	246	8.2	Rp2.000	1
Detergen	31	1.033333	Rp15.000	0

Telur	269	8.966667	Rp2.000	1
Telur	262	8.733333	Rp2.000	1
Detergen	30	1	Rp15.000	0
Tepung	32	1.066667	Rp10.000	0
Kopi	30	1	Rp2.000	0
Beras	70	2.333333	Rp63.000	1
Susu	45	1.5	Rp12.000	0
Sabun Batang	38	1.266667	Rp4.000	0
Beras	75	2.5	Rp63.000	1
Kopi	31	1.033333	Rp2.000	0
Minyak Goreng	41	1.366667	Rp15.000	0
Sabun Batang	36	1.2	Rp4.000	0
Tepung	33	1.1	Rp10.000	0
Kopi	30	1	Rp2.000	0
Gula Pasir	42	1.4	Rp14.000	0
Mie Instan	176	5.866667	Rp3.000	1
Kopi	31	1.033333	Rp2.000	0
Aqua	51	1.7	Rp5.000	1
Kecap	127	4.233333	Rp3.000	1
Kecap	135	4.5	Rp3.000	1
Minyak Goreng	44	1.466667	Rp15.000	0
Rokok	113	3.766667	Rp25.000	1
Gula Pasir	46	1.533333	Rp14.000	0
Gula Pasir	44	1.466667	Rp14.000	0
Sabun Batang	37	1.233333	Rp4.000	0
Gas Elpiji	23	0.766667	Rp22.000	0
Mie Instan	207	6.9	Rp3.000	1

Tabel ini menyajikan data transaksi penjualan produk di Toko Sembako Mentari, yang mencakup informasi jumlah pembelian (tentang total produk terjual), frekuensi transaksi, harga satuan, dan label minat pelanggan pada produk-produk sembako yang tersedia. Beberapa produk yang tercatat dalam tabel ini adalah Minyak Goreng, Garam, Aqua, Gula Pasir, Mie Instan, Telur, Kopi, dan Kecap. Misalnya, produk Telur tercatat memiliki total terjual sebanyak 246 unit dengan

frekuensi transaksi 8,2 kali, yang menunjukkan tingginya permintaan untuk produk ini. Selain itu, produk Mie Instan juga memiliki penjualan yang signifikan dengan 173 unit terjual dan frekuensi transaksi 5,77 kali. Dalam tabel ini juga tercatat produk dengan permintaan rendah seperti Gas Elpiji yang terjual hanya 23 unit dengan frekuensi transaksi 0,77 kali. Label "Diminati" menunjukkan apakah suatu produk memiliki permintaan tinggi (1) atau rendah (0), yang membantu toko dalam merencanakan stok dan strategi penjualan. Berdasarkan data ini, toko dapat mengidentifikasi produk yang perlu dipromosikan lebih lanjut dan menyesuaikan strategi harga untuk memaksimalkan keuntungan serta memenuhi permintaan pasar secara lebih efisien.

Tabel 3. 3 Contoh Data Testing Transaksi Toko Sembako Mentari

Nama Produk	Jumlah Pembelian	Frekuensi Transaksi	Harga Satuan	Label Diminati
Beras	71	2.366666667	Rp 63,000.00	1
Gas Elpiji	21	0.7	Rp 22,000.00	0
Rokok	112	3.733333333	Rp 25,000.00	1
Beras	78	2.6	Rp 63,000.00	1
Detergen	30	1	Rp 15,000.00	0
Sabun Batang	36	1.2	Rp 4,000.00	0
Garam	26	0.866666667	Rp 4,000.00	0
Garam	24	0.8	Rp 4,000.00	0
Susu	41	1.366666667	Rp 12,000.00	0
Susu	37	1.233333333	Rp 12,000.00	0
Minyak Goreng	49	1.633333333	Rp 15,000.00	0
Gas Elpiji	22	0.733333333	Rp 22,000.00	0

Tabel ini menampilkan data pengujian transaksi penjualan produk di Toko Sembako Mentari, yang mencakup informasi mengenai total produk terjual, frekuensi transaksi, dan label minat pelanggan untuk setiap produk. Produk yang tercatat dalam tabel ini antara lain Beras, Gas Elpiji, Rokok, Sabun Batang, Garam,

Susu, dan Minyak Goreng. Sebagai contoh, produk Beras tercatat terjual sebanyak 71 unit dengan frekuensi transaksi 2,37 kali dan diberi label "Diminati" 1, yang menunjukkan permintaan yang tinggi. Produk Rokok juga tercatat memiliki permintaan yang tinggi dengan total penjualan 112 unit dan frekuensi transaksi 3,73 kali. Sebaliknya, produk seperti Gas Elpiji dan Sabun Batang memiliki angka penjualan yang lebih rendah, masing-masing dengan total terjual 21 dan 36 unit serta frekuensi transaksi 0,7 dan 1,2 kali. Data ini memberikan wawasan yang berguna untuk perencanaan stok, pengelolaan persediaan, dan penetapan strategi harga untuk memaksimalkan keuntungan dan memenuhi permintaan pasar dengan lebih efektif.

3.5.3. Perhitungan Manual Algoritma Support Vector Machine (SVM)

Perhitungan pada bagian ini merupakan contoh ilustratif untuk menunjukkan mekanisme kerja algoritma SVM dan Random Forest terhadap data penelitian. Hasil pada bagian ini tidak digunakan sebagai dasar penarikan kesimpulan akhir, melainkan sebagai simulasi proses klasifikasi.

A. Membangun Model SVM:

Tentukan hyperplane yang memisahkan dua kelas (produk yang diminati vs tidak diminati) menggunakan rumus:

$$w \cdot x + b = 0$$

Di sini w adalah vektor bobot (normal) dari hyperplane dan b adalah bias.

B. Parameter yang digunakan:

- 1) $w_1 = 0.1$ untuk Jumlah Pembelian,
- 2) $w_2 = 0.5$ untuk Frekuensi Transaksi,

3) $w_3 = 0.00001$ untuk Harga Satuan,

4) $b = -6$ (bias),

C. Rumus Fungsi Keputusan:

Fungsi keputusan $f(x)$ akan dihitung dengan rumus:

$$f(x) = w_1 \cdot \text{Jumlah Pembelian} + w_2 \cdot \text{Frekuensi Transaksi} + w_3 \cdot \text{Harga Satuan} + b$$

D. Menghitung Fungsi Keputusan

Contoh Perhitungan Manual:

1. Rokok Jumlah Pembelian = 114

Frekuensi Transaksi = 3.8

Harga Satuan = 25.000

$$f(x) = (0.1 \cdot 114) + (0.5 \cdot 3.8) + (0.00001 \cdot 25.000) - 6$$

$$f(x) = 11.4 + 1.9 + 0.25 - 6 = 7.55$$

Prediksi: Karena $f(x) = 7.55 > 0$, maka prediksi = Positif (Diminati).

2. Aqua Jumlah Pembelian = 50

Frekuensi Transaksi = 1.66

Harga Satuan = 5.000

$$f(x) = (0.1 \cdot 50) + (0.5 \cdot 1.66) + (0.00001 \cdot 5.000) - 6$$

$$f(x) = 5 + 0.833 + 0.05 - 6 = -0.1166$$

Prediksi: Karena $f(x) = -0.1166 < 0$, maka prediksi = Negatif (Tidak Diminati).

Tabel 3. 4 Hasil Perhitungan SVM

Priode Penjualan	Nama Produk	Jumlah Pembelian	Frekuensi Transaksi	Harga Satuan	Label Diminati	w1	w2	w3	Bias (b)	Hasil Klasifikasi	Label Akhir

Bulan _Perta ma	Beras	71	2.36666 667	6300 0	1	7 .1	1.1 83 33 3	0 .6 3	-	2.913 33333 3	1
	Minyak Goreng	44	1.46666 667	1500 0	0	4 .4	0.7 33 3	0 .6 5	-	- 0.716 66666 7	0
	Gula Pasir	46	1.53333 333	1400 0	0	4 .6	0.7 66 7	0 .6 4	-	- 0.493 33333 3	0
	Garam	24	0.8	4000	0	2 .4	0.4	0 .0 4	-	-3.16	0
	Telur	246	8.2	2000	1	2 4 .6	4.1	0 .6 2	-	22.72	1
	Gas Elpiji	21	0.7	2200 0	0	2 .1	0.3 5	0 .2 2	-	-3.33	0
	Rokok	114	3.8	2500 0	1	1 1 .4	1.9	0 .2 5	-	7.55	1
	Aqua	50	1.66666 667	5000	0	5	0.8 33 33 3	0 .6 5	-	- 0.116 66666 7	0
	Mie Instan	173	5.76666 667	3000	1	1 7 .3	2.8 83 33 3	0 .6 3	-	14.21 33333 3	1
	Sabun Bata ng	38	1.26666 667	4000	0	3 .8	0.6 33 3	0 .6 4	-	- 1.526 66666 7	0
	Tepu ng	30	1	1000 0	0	3	0.5	0 .1	-	-2.4	0
	Kopi	30	1	2000	0	3	0.5	0 .0 2	-	-2.48	0
Susu	41	1.36666 667	1200 0	0	4 .1	0.6 83	0 .6	-	- 1.096	0	

							33 3	1 2		66666 7	
	Detergen	30	1	1500 0	0	3	0.5	0	- 6	-2.35	0
	Kecap	132	4.4	3000	1	1 3 2	2.2	0 0 3	- 6	9.43	1
Bulan _Kedu a	Beras	70	2.33333 333	6300 0	1	7	1.1 66 66 7	0 6 6 3	- 6	2.796 66666 7	1
	Minyak Goreng	41	1.36666 667	1500 0	0	4 1	0.6 83 33 3	0 1 5	- 6	- 1.066 66666 7	0
	Gula Pasir	47	1.56666 667	1400 0	0	4 7	0.7 83 33 3	0 1 4	- 6	- 0.376 66666 7	0
	Garam	25	0.83333 333	4000	0	2 5	0.4 16 66 7	0 0 4	- 6	- 3.043 33333 3	0
	Telur	269	8.96666 667	2000	1	2 6 9	4.4 83 33 3	0 6 2	- 6	25.40 33333 3	1
	Gas Elpiji	22	0.73333 333	2200 0	0	2 2	0.3 66 66 7	0 2 2	- 6	- 3.213 33333 3	0
	Rokok	113	3.76666 667	2500 0	1	1 1 3	1.8 83 33 3	0 2 5	- 6	7.433 33333 3	1
	Aqua	52	1.73333 333	5000	1	5 2	0.8 66 66 7	0 0 5	- 6	0.116 66666 7	1
	Mie Instan	207	6.9	3000	1	2 0 7	3.4 5	0 0 3	- 6	18.18	1
	Sabun Bata ng	36	1.2	4000	0	3 6	0.6	0 0 4	- 6	-1.76	0

	Tepung	32	1.06666 667	1000 0	0	3 . 2	0.5 33 3	0 . 1	- 6	- 2.166 66666 7	0
	Kopi	31	1.03333 333	2000	0	3 . 1	0.5 16 66 7	0 . 0 2	- 6	- 2.363 33333 3	0
	Susu	45	1.5	1200 0	0	4 . 5	0.7 5	0 . 1 2	- 6	-0.63	0
	Detergen	31	1.03333 333	1500 0	0	3 . 1	0.5 16 66 7	0 . 1 5	- 6	- 2.233 33333 3	0
	Kecap	127	4.23333 333	3000	1	1 2 . 7	2.1 16 66 7	0 . 0 3	- 6	8.846 66666 7	1
Bulan _Ketiga	Beras	75	2.5	6300 0	1	7 . 5	1.2 5	0 . 6 3	- 6	3.38	1
	Minyak Goreng	47	1.56666 667	1500 0	0	4 . 7	0.7 83 33 3	0 . 1 5	- 6	- 0.366 66666 7	0
	Gula Pasir	42	1.4	1400 0	0	4 . 2	0.7	0 . 1 4	- 6	-0.96	0
	Garam	26	0.86666 667	4000	0	2 . 6	0.4 33 33 3	0 . 0 4	- 6	- 2.926 66666 7	0
	Telur	262	8.73333 333	2000	1	2 6 . 2	4.3 66 66 7	0 . 0 2	- 6	24.58 66666 7	1
	Gas Elpiji	23	0.76666 667	2200 0	0	2 . 3	0.3 83 33 3	0 . 2 2	- 6	- 3.096 66666 7	0
	Rokok	112	3.73333 333	2500 0	1	1 1 . 2	1.8 66 66 7	0 . 2 5	- 6	7.316 66666 7	1

Aqua	51	1.7	5000	1	5 . 1	0.8 5	0 . 0 5	- 6	0	0
Mie Instan	192	6.4	3000	1	1 9 . 2	3.2	0 . 0 3	- 6	16.43	1
Sabun Bata ng	37	1.23333 333	4000	0	3 . 7	0.6 16 66 7	0 . 0 4	- 6	- 1.643 33333 3	0
Tepung	31	1.03333 333	1000 0	0	3 . 1	0.5 16 66 7	0 . 1	- 6	- 2.283 33333 3	0
Kopi	31	1.03333 333	2000	0	3 . 1	0.5 16 66 7	0 . 0 2	- 6	- 2.363 33333 3	0
Susu	44	1.46666 667	1200 0	0	4 . 4	0.7 33 33 3	0 . 1 2	- 6	- 0.746 66666 7	0
Detergen	31	1.03333 333	1500 0	0	3 . 1	0.5 16 66 7	0 . 1 5	- 6	- 2.233 33333 3	0
Kecap	135	4.5	3000	1	1 3 . 5	2.2 5	0 . 0 3	- 6	9.78	1

Tabel ini menyajikan data transaksi produk untuk melatih model klasifikasi dalam memprediksi apakah suatu produk diminati atau tidak. Data terdiri dari beberapa variabel: Nama Produk, Jumlah Pembelian, Frekuensi Transaksi, Harga Satuan, Label Diminati, serta tiga bobot (w_1 , w_2 , w_3) yang digunakan untuk menghitung hasil klasifikasi. Label Diminati mengindikasikan apakah produk tersebut diminati (1) atau tidak diminati (0), yang digunakan sebagai target klasifikasi. Kolom W_1 , W_2 , dan W_3 menunjukkan bobot masing-masing fitur yang

dihasilkan dari analisis sebelumnya, yang dikalikan dengan nilai total terjual dan frekuensi transaksi untuk menghitung hasil yang lebih akurat. Bias (B) digunakan untuk penyesuaian lebih lanjut dalam fungsi keputusan. Kolom Label Akhir memberikan nilai hasil perhitungan dengan penyesuaian bias, yang digunakan untuk menentukan hasil klasifikasi. Hasil Klasifikasi akhir menunjukkan prediksi, apakah produk tersebut "Diminati" atau "Tidak Diminati". Sebagai contoh, produk "Telur" dengan total terjual 246 dan frekuensi transaksi 8.2 menghasilkan Label Akhir sebesar 22.72, yang diprediksi sebagai "Diminati".

E. Evaluasi Model

Tabel 3. 5 Hasil Confusion Matrix SVM

Confusion Matrix	
TP	16
FP	2
TN	28
FN	1

Tabel ini menunjukkan hasil dari matriks kebingungan dalam pengujian klasifikasi pada *SVM*, yang terdiri dari empat elemen utama: True Positives (TP), False Positives (FP), True Negatives (TN), dan False Negatives (FN). Dalam data yang diberikan, terdapat 16 nilai TP, 2 nilai FP, 28 nilai TN, dan 1 nilai FN. Matriks kebingungan ini digunakan untuk mengevaluasi kinerja model klasifikasi dengan mengukur jumlah prediksi yang benar dan salah dalam dua kategori positif dan negatif.

Nilai TP (True Positives) menunjukkan jumlah produk yang benar-benar "Diminati" dan berhasil diklasifikasikan dengan benar sebagai "Diminati" oleh

model, yang dalam hal ini ada sebanyak 16 produk. Sementara itu, FP (False Positives) adalah jumlah produk yang sebenarnya "Tidak Diminati" tetapi salah diklasifikasikan sebagai "Diminati", yang terdapat sebanyak 2 produk. TN (True Negatives) menunjukkan jumlah produk yang benar-benar "Tidak Diminati" dan berhasil diklasifikasikan dengan benar sebagai "Tidak Diminati" sebanyak 28 produk. FN (False Negatives) adalah jumlah produk yang sebenarnya "Diminati" tetapi salah diklasifikasikan sebagai "Tidak Diminati", yang dalam hal ini tidak ditemukan (1 produk).

1. Akurasi:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

Dengan nilai:

$$\text{Accuracy} = \frac{16 + 28}{16 + 2 + 28 + 1} = \frac{44}{47} = 0.9361 \text{ atau } 93.61\%$$

2. Presisi:

$$\text{Presisi} = \frac{TP}{TP + FP}$$

Dengan nilai:

$$\text{Precision} = \frac{16}{16 + 2} = \frac{16}{18} = 0.8888 \text{ atau } 88.88\%$$

3. Recall:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Dengan nilai:

$$\text{Recall} = \frac{16}{16 + 1} = \frac{16}{17} = 0.9411 \text{ atau } 94.11\%$$

4. F1-Score:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Dengan nilai:

$$\text{F1-Score} = 2 \times \frac{0.8888 \times 0.9411}{0.8888 + 0.9411} = 2 \times \frac{0,8364}{1,8299} = 0.9141 \text{ atau } 91.41\%$$

Tabel 3. 6 Hasil Evaluasi

Hasil Evaluasi Svm	
Akurasi	0.9361
Presisi	0.8888
Recall	0.9411
F1-Score	0.9141

Tabel di atas menunjukkan hasil evaluasi model Support Vector Machine (SVM) yang digunakan untuk klasifikasi produk yang diminati dan tidak diminati. Berdasarkan pengujian, model ini menunjukkan akurasi sebesar 0.9361, yang berarti bahwa model ini berhasil memprediksi dengan benar sekitar 93.61% dari total data yang diuji. Presisi tercatat sebesar 0.8888, yang mengindikasikan bahwa dari semua prediksi produk yang dianggap diminati oleh model, 88.88% di antaranya memang benar-benar diminati oleh konsumen. Sementara itu, nilai recall sebesar 94.11% menunjukkan bahwa model ini sangat baik dalam mengidentifikasi seluruh produk yang diminati, tanpa ada yang terlewat. F1-Score yang tercatat

sebesar 91.41% merupakan nilai rata-rata dari presisi dan recall, menandakan bahwa model ini memiliki performa yang sangat baik dalam keseimbangan antara menemukan produk yang diminati dan memberikan prediksi yang akurat. Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa model SVM yang digunakan sangat efektif dalam klasifikasi produk berdasarkan tingkat ketertarikan konsumen.

3.5.4. Perhitungan Manual Algoritma Random Forest (RF)

A. Membangun Pohon Keputusan dan Voting

Tabel 3. 7 Hasil Voting Model Pohon Keputusan RF

Nama_Produk	Jumlah_Pembelian	Frekuensi_Transaksi	Harga_Satuan	Label_Diminati	tr e e 1	tr e e 2	tr e e 3	tr e e 4	tr e e 5	voting
Beras	71	2.36666667	63000	1	1	0	1	0	1	1
Minyak Goreng	44	1.46666667	15000	0	0	0	1	0	0	0
Gula Pasir	46	1.53333333	14000	0	0	0	1	0	0	0
Garam	24	0.8	4000	0	0	0	0	0	0	0
Telur	246	8.2	2000	1	1	1	0	1	0	1
Gas Elpiji	21	0.7	22000	0	0	0	1	0	0	0
Rokok	114	3.8	25000	1	1	1	1	0	0	1
Aqua	50	1.66666667	5000	0	0	0	0	0	0	0
Mie Instan	173	5.76666667	3000	1	1	1	0	1	0	1
Sabun Batang	38	1.26666667	4000	0	0	0	0	0	0	0
Tepung	30	1	10000	0	0	0	0	0	0	0
Kopi	30	1	2000	0	0	0	0	0	0	0
Susu	41	1.36666667	12000	0	0	0	0	0	0	0
Detergen	30	1	15000	0	0	0	1	0	0	0
Kecap	132	4.4	3000	1	1	1	0	0	0	0
Beras	70	2.33333333	63000	1	1	0	1	0	1	1
Minyak Goreng	41	1.36666667	15000	0	0	0	1	0	0	0
Gula Pasir	47	1.56666667	14000	0	0	0	1	0	0	0

Garam	25	0.83333333	4000	0	0	0	0	0	0	0
Telur	269	8.96666667	2000	1	1	1	0	1	0	1
Gas Elpiji	22	0.73333333	22000	0	0	0	1	0	0	0
Rokok	113	3.76666667	25000	1	1	1	1	0	0	1
Aqua	52	1.73333333	5000	1	0	0	0	0	1	0
Mie Instan	207	6.9	3000	1	1	1	0	1	0	1
Sabun Batang	36	1.2	4000	0	0	0	0	0	0	0
Tepung	32	1.06666667	10000	0	0	0	0	0	0	0
Kopi	31	1.03333333	2000	0	0	0	0	0	0	0
Susu	45	1.5	12000	0	0	0	0	0	0	0
Detergen	31	1.03333333	15000	0	0	0	1	0	0	0
Kecap	127	4.23333333	3000	1	1	1	0	0	0	0
Beras	75	2.5	63000	1	1	0	1	0	1	1
Minyak Goreng	47	1.56666667	15000	0	0	0	1	0	0	0
Gula Pasir	42	1.4	14000	0	0	0	1	0	0	0
Garam	26	0.86666667	4000	0	0	0	0	0	0	0
Telur	262	8.73333333	2000	1	1	1	0	1	0	1
Gas Elpiji	23	0.76666667	22000	0	0	0	1	0	0	0
Rokok	112	3.73333333	25000	1	1	1	1	0	0	1
Aqua	51	1.7	5000	1	0	0	0	0	1	0
Mie Instan	192	6.4	3000	1	1	1	0	1	0	1
Sabun Batang	37	1.23333333	4000	0	0	0	0	0	0	0
Tepung	31	1.03333333	10000	0	0	0	0	0	0	0
Kopi	31	1.03333333	2000	0	0	0	0	0	0	0
Susu	44	1.46666667	12000	0	0	0	0	0	0	0
Detergen	31	1.03333333	15000	0	0	0	1	0	0	0
Kecap	135	4.5	3000	1	1	1	0	0	0	0

Tabel di atas menyajikan data penjualan berbagai produk bersama dengan frekuensi transaksi dan hasil evaluasi dari lima model pohon keputusan. Di kolom pertama, tercantum nama produk, dengan total jumlah produk yang terjual

frekuensi transaksi, dan harga satuan. Data ini memberikan gambaran seberapa sering produk tersebut dibeli oleh konsumen. Kolom "Label Diminati" menunjukkan apakah produk tersebut termasuk yang banyak diminati (ditandai dengan angka 1) atau tidak (angka 0). Selanjutnya, ada lima kolom yang mewakili hasil prediksi dari masing-masing pohon keputusan (tree 1 hingga tree 5). Setiap pohon memberikan hasilnya, dan hasil akhir dari evaluasi ini tercatat pada kolom "voting," yang mencerminkan keputusan akhir berdasarkan mayoritas dari hasil kelima pohon keputusan. Sebagai contoh, telur, yang terjual sebanyak 246 unit dengan frekuensi transaksi 8.2, mendapat hasil mayoritas positif (label 1) dari model-model tersebut, yang menunjukkan bahwa telur adalah produk yang banyak diminati konsumen. Voting yang mayoritas 1 menunjukkan bahwa mayoritas pohon keputusan sepakat bahwa produk ini diminati.

B. Confusion Matrix

Tabel 3. 8 Hasil Confusion Matrix RF

Confusion Matrix	
TP	12
FP	0
TN	28
FN	5

Tabel Confusion Matrix di atas menunjukkan hasil evaluasi model Random Forest (RF) dalam mengklasifikasikan produk yang diminati dan tidak diminati oleh konsumen. Dalam hal ini, True Positives (TP) sebanyak 12 menunjukkan jumlah produk yang benar-benar diminati dan terprediksi dengan tepat oleh model sebagai produk yang diminati. False Positives (FP) sebanyak 0 menunjukkan bahwa tidak ada produk yang salah diprediksi sebagai diminati padahal sebenarnya tidak

diminati, yang mengindikasikan bahwa model sangat berhati-hati dalam memprediksi kelas positif. True Negatives (TN) sebanyak 28 menunjukkan bahwa ada 28 produk yang tidak diminati dan terprediksi dengan benar sebagai produk yang tidak diminati. Sedangkan False Negatives (FN) sebanyak 5 menunjukkan jumlah produk yang sebenarnya diminati, namun terprediksi salah sebagai tidak diminati. Meskipun angka FN cukup rendah, hal ini menunjukkan bahwa model masih memiliki beberapa kelemahan dalam menangkap produk yang diminati. Secara keseluruhan, meskipun model RF menunjukkan kinerja yang baik, perbaikan pada pemodelan atau tuning parameter mungkin diperlukan untuk mengurangi jumlah FN dan meningkatkan performa lebih lanjut.

C. Evaluasi Model RF

1. Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

Dengan nilai:

$$\text{Accuracy} = \frac{12 + 28}{12 + 0 + 28 + 5} = \frac{40}{45} = 0.8888 \text{ atau } 88.88\%$$

2. Precision:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Dengan nilai:

$$\text{Precision} = \frac{12}{12 + 0} = \frac{12}{12} = 1 \text{ atau } 100\%$$

3. Recall:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Dengan nilai:

$$\text{Recall} = \frac{12}{12 + 5} = \frac{12}{17} = 0.7058 \text{ atau } 70.58\%$$

4. F1-Score:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Dengan nilai:

$$\text{F1-Score} = 2 \times \frac{1 \times 0.7058}{1 + 0.7058} = 2 \times \frac{0.7058}{1.7058} = 0.8275 \text{ atau } 82.75\%$$

Tabel 3. 9 Hasil Evaluasi RF

Hasil Evaluasi RF	
Akurasi	0.8888
Presisi	1
Recall	0.7058
F1-Score	0.8275

Tabel diatas menunjukkan hasil evaluasi model *Random Forest (RF)* yang digunakan untuk klasifikasi produk yang diminati dan tidak diminati. Berdasarkan hasil pengujian, model ini menunjukkan akurasi sebesar 0.8888, yang berarti bahwa 88.88% dari total data yang diuji dapat diprediksi dengan benar oleh model. Presisi tercatat sebesar 1, yang mengindikasikan bahwa setiap prediksi positif yang dilakukan oleh model benar-benar sesuai dengan kategori yang diinginkan. Ini menunjukkan bahwa model sangat teliti dalam memprediksi kelas positif. Nilai recall yang tercatat sebesar 0.7058 menunjukkan bahwa model ini berhasil menemukan sekitar 70.58% dari data positif yang seharusnya dikenali. F1-Score yang mencapai 0.8275 menunjukkan keseimbangan yang baik antara presisi dan

recall, dengan nilai yang mendekati angka ideal 1. Secara keseluruhan, evaluasi ini mengindikasikan bahwa model RF yang digunakan memiliki performa yang solid dalam mengklasifikasikan data dengan tingkat ketepatan dan kemampuan yang baik dalam menangkap seluruh kelas positif.