

Analisis Kinerja Algoritma Naïve Bayes Dalam Prediksi Tingkat Persediaan Susu Formula Pada Sektor Ritel

DOI: <http://dx.doi.org/10.35889/progresif.v22i3.3946>

Creative Commons License 4.0 (CC BY –NC)



Muhammad Erwan Fuqoha Suryanata¹, Budi Rahmani^{2*}

Teknik Informatika, STMIK Banjarbaru, Banjarbaru, Indonesia

*Email Corresponding Author: budirahmani@gmail.com

Abstract

Formula milk inventory management at Toko Serba Ada has traditionally relied on a manual system that is often inefficient and error-prone. One of the main challenges is the mismatch between inventory levels and market demand, leading to overstocking and the risk of financial losses from product expiration. This study aims to analyze the performance of the Naïve Bayes algorithm in predicting formula milk inventory levels in the retail sector. The variables used in this study include initial stock, demand trends, product price, and sales volume. The prediction model was developed as a web-based application using PHP. Model performance was evaluated using a confusion matrix with accuracy, precision, and recall as evaluation metrics. The experimental results showed that the model achieved 71.43% accuracy, 73.76% precision, and 74.08% recall. These findings indicate that the Naïve Bayes algorithm demonstrates satisfactory decision-making capability for formula milk inventory management in the retail sector.

Keywords: *Naïve Bayes; Inventory Prediction; Formula Milk; Retail Sector; Inventory Management.*

Abstrak

Selama ini, pengelolaan stok susu formula di Toko Serba Ada masih mengandalkan sistem manual yang sering kali kurang efisien dan rentan terhadap kesalahan manusia. Masalah utama yang sering muncul adalah ketidaksesuaian antara jumlah stok dan permintaan pasar sehingga menyebabkan penumpukan produk maupun risiko kerugian akibat kedaluwarsa. Penelitian ini bertujuan untuk menganalisis kinerja algoritma Naïve Bayes dalam memprediksi tingkat persediaan susu formula pada sektor ritel. Variabel yang digunakan meliputi stok awal, tren permintaan, harga, dan tingkat penjualan. Model dikembangkan dalam aplikasi berbasis web menggunakan bahasa pemrograman PHP. Evaluasi dilakukan menggunakan confusion matrix dengan parameter akurasi (*accuracy*), presisi (*precision*), dan *recall*. Hasil pengujian menunjukkan bahwa model menghasilkan tingkat akurasi sebesar 71,43%, *presisi* sebesar 73,76%, dan *recall* sebesar 74,08%. Hasil tersebut menunjukkan bahwa algoritma Naïve Bayes memiliki kemampuan yang cukup baik dalam mendukung proses pengambilan keputusan terkait pengelolaan persediaan susu formula di sektor ritel.

Kata kunci: *Naïve Bayes; Prediksi Stok; Manajemen Persediaan; Toko Serba Ada; Klasifikasi.*

1. Pendahuluan

Manajemen persediaan merupakan salah satu aspek penting dalam kegiatan operasional sektor ritel. Ketepatan dalam menentukan jumlah stok barang akan berpengaruh terhadap tingkat pelayanan kepada konsumen, efisiensi biaya penyimpanan, dan keberlanjutan usaha. Kesalahan dalam pengelolaan persediaan dapat menyebabkan kelebihan stok (*overstock*) yang meningkatkan biaya penyimpanan dan risiko produk kedaluwarsa, serta kekurangan stok (*understock*) yang berpotensi mengurangi tingkat kepuasan pelanggan dan peluang penjualan. Oleh karena itu, diperlukan metode yang mampu membantu pelaku usaha memprediksi tingkat persediaan secara lebih akurat dan terukur [1] [2].

Perkembangan teknologi *data mining* dan *machine learning* telah memberikan berbagai alternatif solusi untuk mendukung proses pengambilan keputusan berbasis data. Salah satu algoritma yang banyak digunakan dalam proses klasifikasi dan prediksi adalah *Naïve Bayes*. Algoritma ini dikenal memiliki tingkat komputasi yang relatif cepat, mudah diimplementasikan, serta mampu menghasilkan performa yang baik pada berbagai jenis data. Beberapa penelitian menunjukkan bahwa *Naïve Bayes* berhasil diterapkan pada proses prediksi persediaan barang, prediksi pengiriman, prediksi penjualan, maupun prediksi pada sektor pendidikan dan kesehatan dengan tingkat akurasi yang cukup tinggi [1][3][4][5][6].

Dalam konteks evaluasi performa algoritma, berbagai penelitian juga telah membandingkan *Naïve Bayes* dengan algoritma lain seperti *K-Nearest Neighbor* (KNN), *Decision Tree*, dan *Support Vector Machine* (SVM). Hasil penelitian menunjukkan bahwa *Naïve Bayes* memiliki keunggulan dalam hal kesederhanaan model, kecepatan proses pelatihan, dan kemampuan menghasilkan tingkat akurasi yang kompetitif pada berbagai kasus klasifikasi dan prediksi [7][8][9][10][11]. Temuan tersebut menunjukkan bahwa *Naïve Bayes* masih menjadi salah satu algoritma yang relevan untuk digunakan sebagai model prediksi dalam berbagai bidang aplikasi.

Meskipun demikian, sebagian besar penelitian terdahulu berfokus pada prediksi persediaan secara umum, prediksi penjualan produk, atau studi komparatif algoritma di berbagai domain aplikasi. Penelitian yang secara khusus membahas prediksi tingkat persediaan susu formula pada sektor ritel menggunakan data primer masih relatif terbatas. Padahal, susu formula merupakan produk dengan tingkat perputaran yang cukup tinggi dan memiliki karakteristik khusus karena memiliki masa simpan tertentu, sehingga membutuhkan pengelolaan persediaan yang lebih hati-hati. Kesalahan dalam menentukan jumlah stok dapat berdampak pada kerugian ekonomi akibat produk kedaluwarsa maupun hilangnya peluang penjualan ketika stok tidak tersedia.

Selain itu, penelitian terdahulu umumnya menitikberatkan pada implementasi sistem atau pencapaian tingkat akurasi tertentu tanpa memberikan pembahasan yang mendalam mengenai kinerja algoritma berdasarkan parameter evaluasi klasifikasi seperti *accuracy*, *precision*, dan *recall*. Padahal, analisis terhadap kinerja model menjadi aspek penting untuk mengetahui tingkat keandalan algoritma dalam menghasilkan prediksi yang dapat digunakan sebagai dasar pengambilan keputusan. Oleh karena itu, diperlukan penelitian yang tidak hanya menerapkan algoritma *Naïve Bayes*, tetapi juga menganalisis secara komprehensif performa algoritma tersebut dalam memprediksi tingkat persediaan produk di lingkungan ritel.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis kinerja algoritma *Naïve Bayes* dalam memprediksi tingkat persediaan susu formula pada sektor ritel menggunakan data penjualan aktual yang diperoleh dari Toko Serba Ada Lintang Jaya. Kinerja model dievaluasi menggunakan *confusion matrix* dengan parameter *accuracy*, *precision*, dan *recall* sehingga dapat diketahui tingkat efektivitas algoritma dalam mendukung pengambilan keputusan terkait pengelolaan persediaan. Hasil penelitian diharapkan dapat memberikan kontribusi bagi pengembangan penerapan *data mining* pada sektor ritel, khususnya dalam mendukung manajemen persediaan yang lebih efektif dan berbasis data.

2. Tinjauan Pustaka

Algoritma *Naïve Bayes* merupakan salah satu metode klasifikasi berbasis probabilitas yang banyak digunakan dalam berbagai penelitian karena memiliki proses komputasi yang sederhana, mudah diimplementasikan, serta mampu menghasilkan tingkat akurasi yang baik pada berbagai jenis data [1][2]. Karakteristik tersebut menjadikan algoritma ini banyak dimanfaatkan dalam berbagai permasalahan prediksi dan klasifikasi yang melibatkan data historis maupun data operasional [3][4].

Penerapan *Naïve Bayes* pada bidang manajemen persediaan menunjukkan bahwa data historis penjualan dapat dimanfaatkan untuk membantu menentukan kebutuhan stok secara lebih terukur. Penggunaan metode ini pada sistem prediksi persediaan mampu menghasilkan tingkat akurasi yang tinggi serta membantu proses pengambilan keputusan terkait pengadaan dan pengelolaan barang [1]. Pada sektor ritel, pendekatan yang sama juga digunakan untuk mengklasifikasikan hasil penjualan produk sehingga dapat membantu menentukan tingkat perputaran barang dan mendukung pengelolaan stok yang lebih efektif [2].

Selain pada bidang persediaan, algoritma *Naïve Bayes* juga banyak digunakan dalam proses prediksi di sektor logistik dan distribusi. Pemanfaatan data historis pengiriman

memungkinkan terbentuknya model yang mampu memprediksi ketepatan maupun keterlambatan pengiriman barang dengan tingkat akurasi yang cukup baik. Hasil tersebut menunjukkan bahwa pendekatan probabilistik masih efektif digunakan untuk mendukung pengambilan keputusan operasional yang melibatkan banyak variabel [3].

Kemampuan algoritma *Naïve Bayes* dalam menghasilkan prediksi juga telah dibuktikan pada berbagai bidang lain, seperti prediksi tingkat penjualan produk, prediksi kinerja mahasiswa, hingga prediksi tingkat penyebaran penyakit. Berbagai penelitian menunjukkan bahwa metode ini mampu memanfaatkan pola pada data historis untuk menghasilkan model prediksi dengan tingkat akurasi yang kompetitif [1][3][4][5][6].

Upaya peningkatan performa algoritma juga telah banyak dilakukan melalui optimasi atribut maupun kombinasi metode tertentu. Hasil penelitian menunjukkan bahwa kualitas atribut yang digunakan dalam proses klasifikasi memiliki pengaruh yang cukup besar terhadap tingkat akurasi model yang dihasilkan. Penggunaan teknik seleksi atribut dan optimasi model terbukti mampu meningkatkan performa klasifikasi dibandingkan dengan penggunaan algoritma standar [7].

Kajian mengenai kinerja algoritma *Naïve Bayes* juga banyak dilakukan melalui pendekatan komparatif. Berbagai penelitian membandingkan *Naïve Bayes* dengan algoritma lain, seperti K-Nearest Neighbors (KNN), *Decision Tree*, *Support Vector Machine* (SVM), dan *jaringan saraf*. Hasil yang diperoleh menunjukkan bahwa *Naïve Bayes* tetap mampu memberikan tingkat akurasi yang baik dengan kompleksitas komputasi yang relatif lebih sederhana dibandingkan dengan beberapa algoritma lainnya [8][9][10][11].

Penelitian lain juga menunjukkan bahwa *Naïve Bayes* dapat digunakan untuk berbagai kasus klasifikasi, seperti klasifikasi data medis, klasifikasi tingkat keparahan kecelakaan lalu lintas, dan klasifikasi berbasis teks. Meskipun objek penelitian yang digunakan berbeda, sebagian besar penelitian tersebut menggunakan *confusion matrix*, *accuracy*, *precision*, dan *recall* sebagai parameter utama untuk mengevaluasi kinerja model klasifikasi yang dibangun [12][13][14].

Berbagai penelitian terdahulu menunjukkan bahwa algoritma *Naïve Bayes* memiliki kemampuan yang baik dalam mendukung proses prediksi maupun klasifikasi di berbagai bidang. Namun demikian, penelitian yang secara khusus membahas analisis kinerja algoritma *Naïve Bayes* dalam memprediksi tingkat persediaan susu formula pada sektor ritel masih relatif terbatas. Sebagian besar penelitian terdahulu menggunakan objek berupa persediaan barang secara umum, data penjualan produk, data logistik, maupun data dari bidang pendidikan dan kesehatan [1][2][3][4][5][6][7].

Selain itu, sebagian besar penelitian terdahulu lebih berfokus pada implementasi metode atau perbandingan antaralgoritma, sedangkan kajian yang secara khusus mengevaluasi performa model berdasarkan parameter *accuracy*, *precision*, dan *recall* dalam kasus prediksi persediaan produk ritel masih belum banyak ditemukan [8][9][10][11][12].

Berdasarkan kondisi tersebut, penelitian ini difokuskan pada analisis kinerja algoritma *Naïve Bayes* dalam memprediksi tingkat persediaan susu formula menggunakan data penjualan aktual di lingkungan ritel. Evaluasi dilakukan melalui pengukuran *accuracy*, *precision*, dan *recall* untuk memperoleh gambaran mengenai tingkat keandalan model dalam mendukung pengambilan keputusan terkait pengelolaan persediaan. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi terhadap penerapan *data mining* pada sektor ritel, khususnya dalam pengendalian dan perencanaan persediaan barang [13][14][15][16][17][18].

3. Metodologi Penelitian

3.1 Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk menganalisis kinerja algoritma *Naïve Bayes* dalam memprediksi tingkat persediaan susu formula di sektor ritel. Fokus penelitian diarahkan pada proses pembentukan model klasifikasi berdasarkan data penjualan historis serta pengukuran performa model menggunakan parameter evaluasi klasifikasi.

3.2 Sumber dan Karakteristik Data

Data yang digunakan merupakan data primer yang diperoleh dari pembukuan Toko Serba Ada Lintang Jaya, Kotabaru, Kalimantan Selatan. Data terdiri atas informasi penjualan dan persediaan susu formula selama periode pengamatan, yang kemudian diolah menjadi dataset untuk klasifikasi. Variabel yang digunakan dalam penelitian meliputi: Penjualan bulan sebelumnya, Harga susu, Stok awal, Tren permintaan, Penjualan. Sementara itu, variabel target

(*class label*) yang diprediksi adalah: Tingkat persediaan stok akhir susu, yang diklasifikasikan ke dalam tiga kategori, yaitu: Banyak, Sedang, Sedikit

3.3 Transformasi dan Kategorisasi Data

Sebelum dilakukan proses klasifikasi, data numerik diubah menjadi data kategorikal agar sesuai dengan karakteristik algoritma *Naïve Bayes* yang digunakan dalam penelitian ini. Kategori penjualan bulan sebelumnya dibagi menjadi:

- Rendah ($\leq$ Rp1.500.000)
- Sedang ($>$ Rp1.500.000 sampai Rp5.000.000)
- Tinggi ($>$ Rp5.000.000)

Kategori harga susu dibagi menjadi: Rendah, Sedang, dan Tinggi. Kategori stok awal dibagi menjadi: Rendah, Sedang, dan Tinggi.

Tren permintaan diperoleh dari hasil perhitungan:

$$\text{Tren Permintaan} = \frac{\text{Penjualan}_{\text{Bulan ini}} - \text{Penjualan}_{\text{Bulan lalu}}}{\text{Penjualan}_{\text{Bulan lalu}}} \times 100\% \quad (1)$$

Hasil perhitungan kemudian dikelompokkan ke dalam kategori: Naik, Tetap, dan Turun.

Batas setiap kategori ditunjukkan pada Tabel 2, yaitu rentang kategori yang telah disusun sebelumnya.

Berikutnya masuk ke tahap analisis data dilakukan melalui dua proses utama:

- 1) Transformasi Data: Data mentah dikonversi ke dalam kategori kualitatif (Rendah, Sedang, Tinggi) berdasarkan rentang yang telah ditentukan agar sesuai dengan karakteristik algoritma *Naïve Bayes*.
- 2) Klasifikasi: Data yang telah dikonversi dimasukkan ke dalam *dataset training* untuk melatih model, kemudian kinerjanya diuji menggunakan *dataset testing* (Tabel 1 hingga Tabel 4).

Tabel 1 Data Prediksi Toko Serba Ada Lintang Jaya

No.	Merek Susu	Penjualan Bulan Sebelumnya	Harga Susu	Stok Awal	Tren Permintaan	Penjualan	Stok Akhir Susu
1	Enfamil A+	Rp3.779.000	Rp380.000	36	-2,78	35	Banyak
2	Dancow	Rp9.539.000	Rp150.000	9	-77,78	2	Sedikit
3	Similac	Rp3.652.000	Rp340.000	28	-57,14	12	Sedikit
4	Aptamil	Rp3.357.000	Rp380.000	50	-16	42	Banyak
5	Hi-Protein	Rp400.000	Rp400.000	3	0	3	Sedang
6	PediaSure Complete	Rp2.916.000	Rp450.000	21	-33,33	14	Sedang
7	Nutribaby Gold	Rp3.029.000	Rp300.000	61	-95,08	3	Sedang
8	S-26 Gold	Rp7.015.000	Rp360.000	90	-60	36	Sedikit
9	Neocate Syneo	Rp7.037.000	Rp600.000	73	-75,34	18	Sedikit
10	Bebelove 1	Rp8.685.000	Rp250.000	31	-29,03	22	Sedang
11	Morinaga BMT	Rp1.365.000	Rp320.000	29	-68,97	9	Sedang
12	Nutrilon 1	Rp7.627.000	Rp350.000	65	-38,46	40	Sedang
13	Morinaga	Rp1.187.000	Rp300.000	45	-22,22	35	Sedikit
14	Nutrilon	Rp9.906.000	Rp350.000	8	0	8	Sedikit
15	Nutramigen	Rp3.819.000	Rp380.000	13	-30,77	9	Sedikit
16	Bebelove	Rp4.998.000	Rp250.000	25	0	25	Sedikit
17	SGM	Rp2.161.000	Rp200.000	41	-2,44	40	Sedang
18	PediaSure	Rp7.221.000	Rp400.000	7	-14,29	6	Banyak
19	Vidoran	Rp5.656.000	Rp220.000	10	0	10	Sedang
20	Nutribaby	Rp1.252.000	Rp270.000	19	-21,05	15	Sedang

Setelah variabel didapat, dilakukan konversi kelas menjadi variabel yang ditentukan dengan melakukan perhitungan pada variabel yang sudah ditentukan, lalu masukkan hasilnya ke range kategori untuk menjadi data training. Proses perhitungan dilakukan berdasarkan kategori penjualan pada bulan sebelumnya. Apabila penjualan kurang dari Rp1.500.000, maka termasuk dalam kategori rendah. Sedangkan kategori sedang apabila rentang penjualan sebelumnya berada pada rentang Rp1.500.000 sampai Rp5.000.000. Dan tingkat penjualan bulan sebelumnya melebihi Rp 5.000.000, masuk ke dalam kategori tinggi.

Harga susu dikategorikan sebagai harga jenis susu formula yang dijual di toko serba ada. Apabila rentang harga susu kurang dari Rp150.000 sampai dengan Rp270.000, termasuk dalam kategori rendah; sedangkan susu dengan rentang harga Rp270.000 sampai dengan Rp380.000 termasuk dalam kategori sedang; dan susu dengan rentang harga lebih dari Rp380.000 sampai dengan Rp600.000 termasuk dalam kategori tinggi.

Stok awal yang dimaksud adalah stok awal susu formula yang tersedia di Toko Serba Ada Lintang Jaya. Apabila jumlah stok susu formula kurang dari 10 unit, termasuk dalam kategori stok awal rendah; stok susu formula 10 sampai 30 unit termasuk dalam kategori stok awal sedang; dan stok awal susu formula lebih dari 30 unit termasuk dalam kategori stok awal tinggi.

Apabila hasil perhitungan menunjukkan kurang lebih 40%, maka dikategorikan 'Naik'; apabila perubahan berada pada rentang 41–69%, dikategorikan 'Tetap'; dan apabila terjadi peningkatan lebih dari 70%, dikategorikan 'Turun'. Data penjualan yang digunakan bersumber dari catatan pembukuan Toko Serba Ada Lintang Jaya tahun 2023.

Tabel 2. Rentang Kategori

	Rendah	Sedang	Tinggi
Penjualans	$\leq \text{Rp}1.500.000$	$>\text{Rp}1.500.000 - \text{Rp}5.000.000$	$>\text{Rp}5.000.000$
Harga Susu	$\leq \text{Rp}150.000 - \text{Rp}270.000$	$>\text{Rp}270.000 - \text{Rp}380.000$	$>\text{Rp}380.000 - \text{Rp}600.000$
Stok Awal	≤ 10 unit	$>10 - 30$ unit	> 30 unit
Penjualan	≤ 10 unit	$>10 - 30$ unit	> 30 unit
Tren Permintaan	Kurang $<40\%$	Perubahan 41- 69%.	Peningkatan $>70\%$.
	Naik	Tetap	Turun

3.4 Algoritma Naïve Bayes

Algoritma *Naïve Bayes* merupakan metode klasifikasi probabilistik yang didasarkan pada Teorema Bayes dengan asumsi bahwa setiap atribut bersifat independen terhadap atribut lainnya. Probabilitas suatu kelas dihitung menggunakan persamaan:

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)} \quad (2)$$

$P(C|X)$ = probabilitas posterior terhadap kelas C

$P(X|C)$ = nilai likelihood atribut X pada kelas C

$P(C)$ = prior probability kelas C

$P(X)$ = probabilitas kemunculan data X

Dalam penelitian ini, proses klasifikasi dilakukan dengan menghitung probabilitas setiap kategori stok akhir (Banyak, Sedang, dan Sedikit) berdasarkan kombinasi atribut pada data uji. Kategori dengan nilai probabilitas terbesar akan dipilih sebagai hasil prediksi.

3.5 Proses Pembentukan Model

Tahapan pembentukan model dilakukan sebagai berikut:

1. Mengumpulkan data historis penjualan dan persediaan susu formula.
2. Melakukan transformasi data numerik menjadi data kategorikal.
3. Menyusun dataset *training*.
4. Menghitung nilai *prior probability* untuk setiap kelas.
5. Menghitung nilai *likelihood* setiap atribut terhadap kelas.

6. Membentuk model klasifikasi *Naïve Bayes*.
7. Menguji model menggunakan data tes.
8. Mengukur performa model menggunakan *matriks konfusi*.

Setelah data variabel dikonversi, selanjutnya masukkan ke dalam tabel data training sebagai data perhitungan untuk melatih algoritma Naive Bayes pada stok susu Toko Serba Ada Lintang Jaya.

Tabel 3. Data Training

No.	Merek Susu	Penjualan Bulan Sebelumnya	Harga Susu	Stok Awal	Tren Permintaan	Penjualan	Stok Akhir Susu
1	Enfamil A+	Sedang	Tinggi	Sedang	Naik	Tinggi	Banyak
2	Dancow	Tinggi	Rendah	Rendah	Naik	Sedang	Sedikit
3	Similac	Sedang	Tinggi	Sedang	Naik	Tinggi	Sedikit
4	Aptamil	Sedang	Sedang	Tinggi	Naik	Tinggi	Banyak
5	Hi-Protein	Rendah	Sedang	Rendah	Turun	Tinggi	Sedang
6	PediaSure Complete	Sedang	Sedang	Sedang	Turun	Sedang	Sedang
7	Nutribaby Gold	Sedang	Tinggi	Tinggi	Tetap	Tinggi	Sedang
8	S-26 Gold	Tinggi	Rendah	Tinggi	Turun	Sedang	Sedikit
9	Neocate Syneo	Tinggi	Sedang	Tinggi	Tetap	Tinggi	Sedikit
10	Bebelove 1	Tinggi	Tinggi	Sedang	Tetap	Sedang	Sedang
11	Morinaga BMT	Rendah	Rendah	Sedang	Tetap	Rendah	Sedang
12	Nutrilon 1	Tinggi	Sedang	Tinggi	Turun	Tinggi	Sedang
13	Morinaga	Rendah	Tinggi	Tinggi	Turun	Rendah	Sedikit
14	Nutrilon	Tinggi	Tinggi	Rendah	Naik	Tinggi	Sedikit
15	Nutramigen	Sedang	Rendah	Rendah	Tetap	Sedang	Sedikit
16	Bebelove	Sedang	Rendah	Sedang	Tetap	Rendah	Sedikit
17	SGM	Sedang	Sedang	Tinggi	Tetap	Tinggi	Sedang
18	PediaSure	Tinggi	Sedang	Rendah	Tetap	Tinggi	Banyak
19	Vidoran	Tinggi	Sedang	Rendah	Naik	Tinggi	Sedang
20	Nutribaby	Rendah	Sedang	Sedang	Turun	Tinggi	Sedang

Data training berisi dataset prediksi yang sudah dalam bentuk klasifikasi variabel kelas, yang memuat faktor-faktor yang menentukan prediksi stok susu, dan akan dilatih menggunakan algoritma Naive Bayes.

3.6 Prior Probability

Berdasarkan data pelatihan yang digunakan, probabilitas awal setiap kelas dihitung menggunakan:

$$P(C) = \frac{\text{Jumlah data pada kelas}}{\text{Jumlah seluruh data}} \quad (3)$$

Nilai probabilitas awal digunakan sebagai dasar dalam proses perhitungan probabilitas posterior untuk setiap kelas stok akhir.

3.7 Likelihood

Nilai *likelihood* menunjukkan peluang kemunculan suatu atribut terhadap kelas tertentu dan dihitung menggunakan persamaan:

$$P(X_i) = \frac{\text{Jumlah Kemunculan Atribut}}{\text{Jumlah data pada kelas}} \quad (4)$$

Perhitungan dilakukan pada seluruh atribut yang digunakan dalam penelitian, yaitu penjualan bulan sebelumnya, harga susu, stok awal, tren permintaan, dan penjualan.

3.8 Pengujian Model

Model yang telah terbentuk kemudian diuji menggunakan data uji untuk mengetahui kemampuan algoritma dalam memprediksi tingkat persediaan susu formula. Setiap data uji diproses dengan menghitung probabilitas posterior untuk masing-masing kelas. Hasil prediksi ditentukan berdasarkan nilai probabilitas tertinggi yang diperoleh. Contohnya diperlihatkan pada Tabel 4.

Tabel 4 Tabel Hasil Uji

No.	Merek Susu	Penjualan Bulan Sebelumnya	Harga Susu	Stok Awal	Tren Permintaan	Penjualan	Stok Akhir Susu
21	SGM 1+	Rendah	Tinggi	Tinggi	Naik	Rendah	?

3.9 Evaluasi Kinerja Model

Kinerja model dievaluasi menggunakan *matriks konfusi*. Evaluasi dilakukan dengan menghitung nilai *akurasi (accuracy)*, *presisi (precision)*, dan *recall*.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (7)$$

Keterangan:

- TP (*True Positive*)
- TN (*True Negative*)
- FP (*False Positive*)
- FN (*False Negative*)

Nilai ketiga parameter tersebut digunakan untuk menilai tingkat keandalan algoritma *Naïve Bayes* dalam memprediksi tingkat persediaan susu formula di sektor ritel.

4. Hasil dan Pembahasan

4.1 Pembentukan Model Naïve Bayes

Model klasifikasi dibentuk menggunakan dataset training yang terdiri dari 20 data historis penjualan susu formula. Data tersebut telah melalui proses transformasi dan kategorisasi sehingga seluruh atribut dapat diproses menggunakan algoritma *Naïve Bayes*. Tahap pertama dalam pembentukan model adalah menghitung nilai *prior probability* untuk masing-masing kelas berdasarkan distribusi kategori stok akhir susu pada data pelatihan. Nilai probabilitas awal ini digunakan sebagai dasar untuk menghitung probabilitas posterior pada tahap prediksi.

Berdasarkan data training yang digunakan, kategori stok akhir terdiri atas kelas Banyak, Sedang, dan Sedikit. Distribusi masing-masing kelas ditunjukkan pada Tabel 5. Hasil perhitungan menunjukkan bahwa kelas Sedang memiliki probabilitas awal terbesar sebesar 0,45, diikuti oleh kelas Sedikit sebesar 0,40 dan kelas Banyak sebesar 0,15. Nilai tersebut menunjukkan bahwa sebagian besar data historis berada pada kategori stok sedang.

Tabel 5. Prior Probability Setiap Kelas

Kelas	Jumlah Data	Prior Probability
Banyak	3	3/20 = 0,15
Sedang	9	9/20 = 0,45
Sedikit	8	8/20 = 0,40

Setelah nilai prior probability diperoleh, dilakukan perhitungan likelihood untuk setiap atribut pada masing-masing kelas. Nilai likelihood menunjukkan peluang kemunculan suatu atribut pada kelas tertentu berdasarkan data pelatihan yang digunakan. Nilai-nilai tersebut kemudian digunakan dalam proses perhitungan probabilitas posterior untuk menentukan prediksi pada data uji.

Tabel 6. Contoh Nilai Likelihood Atribut terhadap Kelas

Atribut	Nilai	Banyak	Sedang	Sedikit
Penjualan Sebelumnya	Rendah	0/3	3/9	2/8
Penjualan Sebelumnya	Sedang	2/3	4/9	2/8
Penjualan Sebelumnya	Tinggi	1/3	2/9	4/8
Harga Susu	Rendah	0/3	2/9	3/8
Harga Susu	Sedang	2/3	5/9	2/8
Harga Susu	Tinggi	1/3	2/9	3/8

Tabel 6 menampilkan sebagian nilai likelihood yang digunakan dalam proses klasifikasi. Perhitungan likelihood dilakukan terhadap seluruh atribut yang digunakan dalam penelitian. Berdasarkan hasil perhitungan probabilitas posterior menggunakan model Naïve Bayes, data uji SGM 1+ diklasifikasikan ke dalam kategori stok akhir Sedang. Kategori tersebut dipilih karena memiliki nilai probabilitas posterior terbesar dibandingkan dengan kategori Banyak dan Sedikit.

4.2 Pengujian Model Menggunakan Data Uji

Setelah model terbentuk, dilakukan pengujian menggunakan data uji sebagaimana ditunjukkan pada Tabel 4. Data uji yang digunakan memiliki atribut penjualan bulan sebelumnya rendah, harga susu tinggi, stok awal tinggi, tren permintaan naik, dan tingkat penjualan rendah. Tujuan pengujian ini adalah untuk mengetahui kategori stok akhir yang diprediksi oleh model Naïve Bayes berdasarkan pola yang dipelajari dari data pelatihan.

Pada proses klasifikasi, dilakukan perhitungan probabilitas posterior untuk setiap kelas menggunakan nilai prior *probability* dan *likelihood* dari masing-masing atribut. Hasil klasifikasi ditentukan berdasarkan nilai probabilitas terbesar pada setiap kelas. Berdasarkan hasil perhitungan probabilitas posterior menggunakan model Naïve Bayes yang telah dibentuk, data uji SGM 1+ diklasifikasikan ke dalam kategori stok akhir Sedang. Hasil tersebut menunjukkan bahwa kombinasi atribut penjualan bulan sebelumnya, harga susu, stok awal, tren permintaan, dan tingkat penjualan pada data uji memiliki kecenderungan terbesar terhadap kelas Sedang dibandingkan kelas lainnya.

4.3 Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan menggunakan *confusion matrix* untuk mengetahui kemampuan algoritma *Naïve Bayes* dalam mengklasifikasikan tingkat persediaan susu formula ke dalam kategori Banyak, Sedang, dan Sedikit. Penggunaan *confusion matrix* memungkinkan pengukuran performa model secara lebih rinci melalui parameter *accuracy*, *precision*, dan *recall*, sehingga tingkat keandalan model dapat dianalisis secara objektif.

Hasil klasifikasi yang diperoleh dari proses pengujian ditunjukkan pada Tabel 6.

Tabel 6. Confusion Matrix			
Prediksi	Rendah	Sedang	Tinggi
Rendah	4	2	0
Sedang	1	5	1
Tinggi	0	2	6

Accuracy dihitung berdasarkan jumlah prediksi yang benar pada diagonal confusion matrix.

$$Accuracy = \frac{4 + 5 + 6}{21} \times 100\% = 71,43\%$$

Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan 71,43% dari seluruh data yang diuji. Dengan demikian, algoritma *Naïve Bayes* dapat digunakan sebagai pendekatan

klasifikasi untuk memprediksi tingkat persediaan susu formula pada sektor ritel dengan tingkat ketepatan yang cukup baik.

Selain akurasi, evaluasi juga dilakukan menggunakan parameter *presisi* dan *recall*. Hasil pengujian menunjukkan nilai *precision* sebesar 73,76% dan *recall* sebesar 74,08%. Nilai tersebut menunjukkan bahwa sebagian besar prediksi yang dihasilkan model berada pada kategori yang sesuai dengan kondisi aktual serta mampu mengenali sebagian besar data pada masing-masing kelas yang tersedia.

4.4 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes* mampu mengidentifikasi pola hubungan antara penjualan bulan sebelumnya, harga susu, stok awal, tren permintaan, dan tingkat penjualan dengan kondisi stok akhir produk. Kemampuan tersebut terlihat dari nilai akurasi yang mencapai 71,43%, yang menunjukkan bahwa model ini dapat digunakan sebagai alat bantu dalam proses pengambilan keputusan terkait pengelolaan persediaan susu formula di sektor ritel.

Secara konseptual, hasil ini sejalan dengan karakteristik algoritma *Naïve Bayes* yang memanfaatkan probabilitas kemunculan atribut untuk menentukan kelas yang paling mungkin. Pendekatan probabilistik tersebut memungkinkan model melakukan klasifikasi meskipun jumlah data yang digunakan relatif terbatas [7][8][9]. Keunggulan tersebut juga menjadi salah satu alasan mengapa *Naïve Bayes* masih banyak digunakan dalam penelitian prediksi dan klasifikasi di berbagai bidang [10][11].

Jika dibandingkan dengan penelitian prediksi persediaan barang yang memperoleh tingkat akurasi sebesar 95,60%, nilai akurasi pada penelitian ini memang lebih rendah [1]. Perbedaan tersebut dapat disebabkan oleh karakteristik data yang digunakan. Data persediaan susu formula memiliki variasi permintaan yang lebih dinamis serta jumlah data yang relatif terbatas, sehingga pola yang terbentuk tidak sekuat data pada penelitian sebelumnya. Selain itu, beberapa atribut memiliki karakteristik yang saling berdekatan sehingga meningkatkan peluang terjadinya kesalahan klasifikasi.

Hasil penelitian ini juga menunjukkan bahwa kategori stok sedang memiliki jumlah data yang paling dominan dibandingkan kategori lainnya. Kondisi tersebut memengaruhi distribusi probabilitas dalam proses pelatihan model, sehingga beberapa data yang berada pada batas kategori berpotensi diklasifikasikan ke kelas yang berbeda. Fenomena ini merupakan salah satu karakteristik yang umum ditemukan dalam penerapan *Naïve Bayes* ketika data antarkelas memiliki distribusi yang tidak seimbang [7][8].

Meskipun demikian, kemampuan model dalam menghasilkan nilai *precision* dan *recall* yang cukup baik menunjukkan bahwa algoritma masih mampu mengenali pola utama pada data persediaan susu formula. Hasil tersebut menunjukkan bahwa metode *Naïve Bayes* layak digunakan sebagai alternatif sistem pendukung keputusan untuk membantu menentukan kondisi persediaan produk di lingkungan ritel. Temuan ini sejalan dengan berbagai penelitian sebelumnya yang menunjukkan bahwa *Naïve Bayes* mampu memberikan performa yang kompetitif pada permasalahan prediksi dan klasifikasi berbasis data historis [3][4][5][6].

Dari sisi praktis, penerapan model ini dapat membantu pengelola toko memantau stok secara lebih terukur dibandingkan dengan pendekatan konvensional yang hanya mengandalkan pengalaman atau perkiraan subjektif. Informasi hasil prediksi dapat digunakan sebagai dasar untuk menentukan prioritas pengadaan produk, sehingga risiko kelebihan maupun kekurangan stok dapat diminimalkan. Dengan demikian, penelitian ini memberikan kontribusi terhadap pemanfaatan teknik *data mining* dalam mendukung pengelolaan persediaan pada sektor ritel berskala mikro dan menengah. ,

4.5 Keterbatasan Penelitian dan Peluang Pengembangan

Penelitian ini masih memiliki beberapa keterbatasan. Pertama, jumlah data yang digunakan dalam proses pelatihan relatif terbatas, sehingga variasi pola yang dapat dipelajari oleh model belum sepenuhnya merepresentasikan kondisi operasional yang sebenarnya. Kedua, variabel yang digunakan hanya terdiri atas faktor internal yang berasal dari data penjualan dan persediaan, sehingga belum mempertimbangkan faktor eksternal seperti promosi, musim penjualan, kondisi ekonomi, dan perubahan perilaku konsumen.

Pada penelitian selanjutnya, jumlah dataset dapat diperluas agar distribusi data antarkelas menjadi lebih representatif dan mampu meningkatkan akurasi model. Selain itu, penambahan

variabel eksternal dan penerapan metode komparatif dengan algoritma lain, seperti *Decision Tree*, *Random Forest*, dan *Support Vector Machine*, dapat dilakukan untuk memperoleh model prediksi yang lebih optimal. Penelitian lanjutan juga dapat memanfaatkan teknik optimasi atribut untuk meningkatkan performa klasifikasi, sebagaimana telah dilakukan dalam beberapa penelitian sebelumnya [7][8][9][10][11].

5. Simpulan

Berdasarkan hasil analisis kinerja yang telah dilakukan, algoritma Naïve Bayes mampu menghasilkan tingkat akurasi sebesar 71,43% dalam memprediksi tingkat persediaan susu formula di sektor ritel. Nilai precision sebesar 73,76% dan recall sebesar 74,08% menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam mengidentifikasi kondisi persediaan berdasarkan data historis yang digunakan.

Meskipun terdapat keterbatasan akurasi pada kategori 'Sedang' akibat tumpang tindih (*overlap*) fitur data, sistem ini secara substansial telah memberikan kontribusi nyata dalam memodernisasi manajemen inventaris toko. Integrasi antara analisis probabilitas algoritma dan kebutuhan manajerial terbukti mampu mengubah pola pengelolaan stok dari yang sebelumnya bersifat reaktif dan subjektif menjadi berbasis data yang terukur.

Sebagai langkah pengembangan di masa mendatang, disarankan untuk memperluas jumlah *dataset* pelatihan guna memperkecil celah kesalahan klasifikasi pada data yang berada di ambang batas (*boundary*). Selain itu, penambahan variabel pendukung seperti promosi musiman atau faktor eksternal lainnya dapat dipertimbangkan untuk meningkatkan performa model agar lebih adaptif terhadap dinamika pasar yang lebih kompleks. Penelitian ini secara keseluruhan memberikan landasan teknis yang kokoh bagi implementasi sistem pendukung keputusan berbasis *data mining* di lingkup ritel mikro.

Referensi

- [1] M. Taufik Hidayat, N. Suarna, and N. Rahaningsih, "Implementasi Algoritma Naïve Bayes Untuk Prediksi Persediaan Barang Pt. Dilmoni Citra Mebel Indonesia," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 693–699, 2023, doi: 10.36040/jati.v7i1.6310.
- [2] Y. Y. L. Yuyun, C. W. Silaban, A. S. Sitepu, and R. K. Telaumbanua, "Classification Analysis of Product Sales Results at Alfamart Using the Naïve Bayes Method," *Electron. Integr. Comput. Algorithm J.*, vol. 1, no. 2, pp. 69–74, 2024, doi: 10.62123/enigma.v1i2.18.
- [3] M. A. Firmansyah and M. M. Alamin, "Sistem Prediksi Pengiriman Pada Dakota Cargo Menggunakan Metode Naive Bayes Berbasis Web," *JUTISI J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 13, no. 1, pp. 324–335, 2024, doi: 10.35889/jutisi.v13i1.1802.
- [4] A. Hasyim, M. Fatchan, and W. Hadikristanto, "Penerapan Algoritma Naïve Bayes Dalam Memprediksi Tingkat Penjualan Mobil Tahun 2022," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 02, pp. 207–215, 2022, doi: 10.46772/intech.v4i02.872.
- [5] A. Karima and T. A. Y. Siswa, "Prediksi Kinerja Mahasiswa Dalam Perkuliahan Berbasis Learning Management System Menggunakan Algoritma Naïve Bayes," *Progresif J. Ilm. Komput.*, vol. 18, no. 2, p. 211, 2022, doi: 10.35889/progresif.v18i2.922.
- [6] Rayuwati, Husna Gemasih, and Irma Nizar, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid," *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 38–46, 2022, doi: 10.55606/jurritek.v1i1.127.
- [7] D. Y. Utami, E. Nurlalah, and N. Hikmah, "Attribute Selection in Naïve Bayes Algorithm Using Genetic Algorithms and Bagging for Prediction of Liver Disease," *J. Informatics Telecommun. Eng.*, vol. 4, no. 1, pp. 76–85, 2020, doi: 10.31289/jite.v4i1.3793.
- [8] B. Asyari, S. Putri, A. Alkadri, and P. Y. Utami, "Uji Akurasi Algoritme K-Nearest Neighbor Dan Naïve Bayes Dalam Klasifikasi Kelayakan Pemberian Kredit Perbankan," *JUTISI J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 12, no. 3, pp. 999–1012, 2023.
- [9] R. Pakan and E. Mailoa, "Komparasi Kinerja Algoritma Machine Learning Untuk Mengklasifikasi Penyakit Kanker Payudara," *Progresif J. Ilm. Komput.*, vol. 20, no. 1, 2024, doi: 10.35889/progresif.v20i1.1812.
- [10] B. Imran, M. N. Karim, and N. I. Ningsih, "Klasifikasi Berita Hoax Terkait Pemilihan Umum Presiden Republik Indonesia Tahun 2024 Menggunakan Naïve Bayes Dan SVM," *J. Ilm. Din. Rekayasa*, vol. 20, no. 1, pp. 1–9, Jan. 2024, doi: 10.20884/1.DINAREK.2024.20.1.27.
- [11] D. Puspa Martiyaningsih, R. Dias Ramadhani, D. Atika, and R. Dewi, "Komparasi Algoritme C4.5 Dan Naïve Bayes Dalam Klasifikasi Produk Zam-Zam Time Berdasarkan Tingkat

- Kepuasan Pelanggan,” *Progresif J. Ilm. Komput.*, vol. 19, no. 2, pp. 578–590, 2023.
- [12] D. A. Tyas, S. Hartati, A. Harjoko, and T. Ratnaningsih, “Erythrocyte Classification using Multi-Layer Perceptron, Naïve Bayes Classifier, RBF Network and SVM,” *Int. J. Eng. Adv. Technol.*, vol. 9, no. 2, pp. 2024–2028, 2019, doi: 10.35940/ijeat.B3231.129219.
- [13] S. N. Abdullah, S. K. Gusti, F. Wulandari, and F. Syafria, “Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas Menggunakan Naive Bayes,” *Progresif J. Ilm. Komputer Jurnal Ilm. Komput.*, vol. 19, no. 2, pp. 858–866, 2023.
- [14] G. G. Warow and H. Pandia, “Analisis Sentimen Aplikasi Dana Menggunakan Naïve Bayes Classifier dan Support Vector Machine,” *JUTISI J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 13, no. 1, pp. 609–621, 2024, doi: 10.35889/jutisi.v13i1.1893.
- [15] F. A. Hizham, C. K. Murni, and M. Qori’atunnadiyah, “Uji Klasifikasi Algoritma Naive Bayes Classification dalam Analisis Sentimen Ulasan Puncak B29 Lumajang,” *Progresif J. Ilm. Komput.*, vol. 20, no. 1, p. 361, 2024, doi: 10.35889/progresif.v20i1.1618.
- [16] E. L. Rara and E. Mailoa, “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Perubahan Piala Dunia U-20,” *Progresif J. Ilm. Komput.*, vol. 20, no. 1, p. 259, 2024, doi: 10.35889/progresif.v20i1.1550.
- [17] F. Kristanto, W. W. Winarno, and A. Nasiri, “Perbandingan Algoritme Naïve Bayes dan Decision Tree Pada Analisis Sentimen Data Komentar Siswa Pada Aplikasi Digital Teacher Assessment,” *JUTISI J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 12, no. 2, pp. 538–548, 2023.
- [18] A. Simanungkalit, J. P. P. Naibaho, and A. De Kweldju, “Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Shopee Menggunakan Algoritma Naïve Bayes,” *JUTISI J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 13, no. 1, pp. 659–670, 2024, doi: 10.35889/jutisi.v13i1.1826.