
IMPLEMENTASI ALGORITMA NAIVE BAYES DALAM MENENTUKAN KATEGORI DISKON PELANGGAN BERDASARKAN PERILAKU PEMBELIAN PADA TOKO RETAIL SEBAGAI PENDUKUNG TRANSFORMASI DIGITAL BERKELANJUTAN

Fikry Dian Anugerah R Siregar¹
Universitas Pembangunan Pancabudi, Medan
e-mail: ¹fikrydiann@gmail.com

Abstract: Digital transformation has encouraged retail companies to leverage information technology and data analytics to improve operational efficiency and the quality of decision-making. The utilization of technology and data analytics enables companies to automate business processes, enhance analytical accuracy, and manage customer data more effectively. This study aims to implement the Naive Bayes algorithm to determine customer discount categories based on purchasing behavior in retail stores as a form of data-driven decision-making that supports digital transformation. The dataset used in this research is the Marketing Campaign Dataset, which consists of the attributes Income, Recency, NumDealsPurchases, NumWebPurchases, NumStorePurchases, and Response. The research stages include data collection, preprocessing, data transformation, classification model development using the Naive Bayes algorithm, and model evaluation using the Confusion Matrix, Accuracy, Precision, Recall, and F1-Score. The results show that the Naive Bayes model achieved an accuracy of 70% in classifying customers into low, medium, and high discount categories based on their purchasing behavior patterns. The implementation of this model enables retail companies to provide discount category recommendations more quickly, accurately, and consistently. In addition to improving the effectiveness of promotional strategies, this study demonstrates that the application of machine learning-based data analytics can support digital transformation through faster, more accurate, and data-driven decision-making, thereby helping retail companies improve service quality and business process effectiveness.

Keywords: Naive Bayes; Digital Transformation; Customer Discount Categories; Purchasing Behavior; Retail Store.

Abstrak: Transformasi digital telah mendorong perusahaan retail untuk memanfaatkan teknologi informasi dan analisis data dalam meningkatkan efisiensi operasional serta kualitas pengambilan keputusan. Pemanfaatan teknologi dan analitik data memungkinkan perusahaan mengotomatisasi proses bisnis, meningkatkan akurasi analisis, serta memanfaatkan data pelanggan secara lebih efektif. Penelitian ini bertujuan mengimplementasikan algoritma Naive Bayes dalam menentukan kategori diskon pelanggan berdasarkan perilaku pembelian pada toko retail sebagai salah satu bentuk pengambilan keputusan berbasis data yang mendukung transformasi digital. Dataset yang digunakan adalah Marketing Campaign Dataset yang terdiri dari atribut Income, Recency, NumDealsPurchases, NumWebPurchases, NumStorePurchases, dan Response. Tahapan penelitian meliputi pengumpulan data, preprocessing, transformasi data, pembentukan model klasifikasi menggunakan algoritma Naive Bayes, serta evaluasi menggunakan Confusion Matrix, Accuracy, Precision, Recall, dan F1-Score. Hasil penelitian menunjukkan bahwa model Naive Bayes memperoleh akurasi sebesar 70% dalam mengklasifikasikan pelanggan ke dalam kategori diskon rendah, sedang, dan tinggi berdasarkan pola perilaku pembelian. Implementasi model ini mampu membantu perusahaan retail dalam memberikan rekomendasi kategori diskon secara lebih cepat, tepat, dan konsisten. Selain meningkatkan efektivitas strategi promosi, penelitian ini

menunjukkan bahwa penerapan analitik data berbasis machine learning dapat mendukung transformasi digital melalui pengambilan keputusan yang lebih cepat, akurat, dan berbasis data sehingga membantu perusahaan retail meningkatkan kualitas layanan serta efektivitas proses bisnis.

Kata kunci: Naive Bayes; Transformasi Digital; Kategori Diskon Pelanggan; Perilaku Pembelian; Toko Retail.

PENDAHULUAN

Transformasi digital telah menjadi salah satu strategi utama yang diterapkan oleh perusahaan retail untuk meningkatkan efisiensi operasional, kualitas pelayanan, dan daya saing bisnis. Pemanfaatan teknologi informasi memungkinkan perusahaan mengolah data pelanggan secara lebih cepat dan akurat sehingga proses pengambilan keputusan dapat dilakukan secara berbasis data (data-driven decision making). Sejalan dengan perkembangan tersebut, konsep transformasi digital berkelanjutan tidak hanya berfokus pada digitalisasi proses bisnis, tetapi juga pada optimalisasi penggunaan teknologi informasi yang efisien, adaptif, dan ramah lingkungan melalui penerapan prinsip Green IT. Pendekatan ini bertujuan mengurangi penggunaan sumber daya secara berlebihan, meminimalkan aktivitas manual, serta meningkatkan efisiensi operasional perusahaan. [1]

Dalam industri retail, setiap aktivitas transaksi pelanggan menghasilkan data yang memiliki nilai strategis untuk mendukung berbagai keputusan bisnis, termasuk dalam penyusunan strategi promosi. Salah satu bentuk promosi yang banyak diterapkan adalah pemberian diskon kepada pelanggan. Namun, pemberian diskon yang dilakukan secara merata sering kali kurang efektif karena setiap pelanggan memiliki karakteristik, tingkat loyalitas, dan perilaku pembelian yang berbeda. Oleh karena itu, perusahaan memerlukan suatu pendekatan yang mampu mengidentifikasi karakteristik pelanggan sehingga pemberian diskon dapat dilakukan secara lebih tepat sasaran dan

memberikan manfaat yang optimal bagi perusahaan maupun pelanggan. [2]

Penerapan analisis data dalam menentukan kategori diskon merupakan salah satu bentuk implementasi transformasi digital pada sektor retail. Melalui pemanfaatan teknologi analitik dan machine learning, proses yang sebelumnya dilakukan secara manual dapat diotomatisasi sehingga mampu mempercepat pengambilan keputusan, meningkatkan akurasi analisis, serta membantu perusahaan dalam mengelola data pelanggan secara lebih efektif. Transformasi digital memungkinkan perusahaan retail memanfaatkan data historis transaksi sebagai dasar dalam menyusun strategi promosi yang lebih tepat sasaran dan berbasis data. [3]

Salah satu pendekatan yang banyak digunakan dalam analisis data adalah data mining, yaitu proses menemukan pola dan pengetahuan yang tersembunyi dari sekumpulan data dalam jumlah besar. Teknik klasifikasi pada data mining memungkinkan sistem mengelompokkan data ke dalam kelas tertentu berdasarkan karakteristik yang dimiliki. Dalam penelitian ini, teknik klasifikasi dimanfaatkan untuk menentukan kategori diskon pelanggan berdasarkan perilaku pembelian sehingga perusahaan dapat menyusun strategi promosi yang lebih efektif dan berbasis data. [4]

Algoritma Naive Bayes merupakan salah satu metode klasifikasi yang banyak digunakan karena memiliki proses komputasi yang sederhana, mampu menangani data dalam jumlah besar, serta memberikan performa yang baik pada berbagai kasus klasifikasi. Algoritma ini bekerja berdasarkan konsep probabilitas dengan menghitung kemungkinan suatu

data termasuk ke dalam kelas tertentu berdasarkan atribut yang dimiliki. Penelitian ini menggunakan atribut *Income*, *Recency*, *NumDealsPurchases*, *NumWebPurchases*, *NumStorePurchases*, dan *Response* untuk membangun model klasifikasi kategori diskon pelanggan. [5]

Dataset yang digunakan dalam penelitian ini adalah Marketing Campaign Dataset yang berisi informasi mengenai karakteristik pelanggan, aktivitas pembelian, serta respons terhadap kampanye pemasaran. Data tersebut diproses melalui tahapan preprocessing, transformasi data, pembentukan model klasifikasi menggunakan algoritma Naive Bayes, serta evaluasi model menggunakan Confusion Matrix, Accuracy, Precision, Recall, dan F1-Score. Melalui tahapan tersebut diharapkan diperoleh model yang mampu mengklasifikasikan pelanggan ke dalam kategori diskon rendah, sedang, dan tinggi berdasarkan perilaku pembelian yang dimiliki. [6]

Penelitian ini bertujuan mengimplementasikan algoritma Naive Bayes dalam menentukan kategori diskon pelanggan berdasarkan perilaku pembelian pada toko retail sebagai bagian dari penerapan transformasi digital dalam pengambilan keputusan berbasis data. Hasil penelitian diharapkan dapat membantu perusahaan retail meningkatkan efektivitas strategi promosi, mempercepat proses pengambilan keputusan, serta mengoptimalkan pemanfaatan data pelanggan dalam menentukan kategori diskon yang sesuai.

METODE

Jenis dan Sumber Data

Pada penelitian ini Pada penelitian ini, jenis data yang digunakan adalah data sekunder. Data sekunder merupakan data yang diperoleh dari sumber yang telah tersedia dan dikumpulkan oleh pihak lain, sehingga peneliti tidak melakukan pengumpulan data secara langsung dari objek penelitian. Penggunaan data

sekunder dipilih karena memiliki cakupan data yang luas, mudah diakses, serta sesuai dengan kebutuhan penelitian dalam melakukan analisis perilaku pembelian pelanggan. Sumber data yang digunakan dalam penelitian ini berasal dari platform Kaggle dengan nama dataset Marketing Campaign Dataset. Dataset tersebut berisi informasi mengenai karakteristik pelanggan dan aktivitas pembelian yang dilakukan pada suatu perusahaan retail. Atribut yang digunakan dalam penelitian ini adalah:

Tabel 1 Atribut

No	Atribut	Keterangan
1	<i>Income</i>	Pendapatan
2	<i>Recency</i>	Waktu Sejak Pembelian Terakhir
3	<i>NumDealsPurchases</i>	Jumlah Pembelian Menggunakan Diskon
4	<i>NumWebPurchases</i>	Jumlah Pembelian Melalui Website
5	<i>NumStorePurchases</i>	Jumlah Pembelian Di Toko
6	<i>Response</i>	Respons Terhadap Kampanye Pemasaran Sebelumnya

Tahapan Penelitian

Proses analisis dilakukan melalui beberapa tahap Berikut:

1. Data Collection
2. Preprocessing
3. Algoritma Naïve Bayes
4. Training
5. Evaluation

Penerapan algoritma Naive Bayes pada penelitian ini tidak hanya bertujuan menghasilkan model klasifikasi kategori diskon pelanggan yang akurat, tetapi juga mendukung implementasi transformasi digital pada sektor retail melalui pemanfaatan analisis data sebagai dasar

pengambilan keputusan. Otomatisasi proses klasifikasi pelanggan berdasarkan perilaku pembelian mampu meningkatkan efisiensi operasional, mempercepat proses analisis, serta mengurangi ketergantungan pada proses manual dalam penyusunan strategi promosi. Dengan demikian, implementasi algoritma Naive Bayes dapat membantu perusahaan retail dalam memanfaatkan data pelanggan secara optimal untuk mendukung proses bisnis yang lebih efektif dan efisien.

HASIL DAN PEMBAHASAN

Hasil Pengumpulan dan Analisis Data (Data Collection)

Dataset yang digunakan dalam penelitian ini diperoleh dari platform Kaggle dengan judul Marketing Campaign Dataset. Dataset ini berisi informasi mengenai profil pelanggan, aktivitas pembelian, serta respons pelanggan terhadap kampanye pemasaran yang dilakukan oleh perusahaan retail. Berdasarkan hasil eksplorasi awal, dataset terdiri dari 2.240 data (baris) dan 29 atribut (kolom).

Beberapa atribut yang digunakan dalam penelitian ini antara lain Income, Recency, NumDealsPurchases, NumWebPurchases, NumCatalogPurchases, NumStorePurchases, NumWebVisitsMonth, serta atribut pengeluaran pelanggan seperti MntWines, MntFruits, MntMeatProducts, MntFishProducts, MntSweetProducts, dan MntGoldProds. Selain itu, terdapat atribut kategorik seperti Education dan Marital_Status yang menggambarkan karakteristik pelanggan. Berdasarkan hasil analisis awal, dataset memiliki kombinasi data numerik dan kategorik yang relevan untuk menggambarkan perilaku pembelian pelanggan. Ditemukan pula beberapa nilai kosong (missing value) pada atribut Income yang akan ditangani pada tahap preprocessing.

Secara umum, dataset yang digunakan memiliki struktur data yang

cukup lengkap dan sesuai dengan tujuan penelitian, yaitu melakukan klasifikasi kategori diskon pelanggan berdasarkan perilaku pembelian menggunakan algoritma Naïve Bayes. Tahap selanjutnya adalah melakukan preprocessing untuk menyiapkan data sebelum digunakan dalam proses pembentukan model klasifikasi.

Data Awal:							
	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome
0	5524	1957	Graduation	Single	58138.0	0	0
1	2174	1954	Graduation	Single	46344.0	1	1
2	4141	1965	Graduation	Together	71613.0	0	0
3	6182	1984	Graduation	Together	26646.0	1	0
4	5324	1981	PhD	Married	58293.0	1	0

Dt_Customer	Recency	MntWines	...	NumWebVisitsMonth	AcceptedCmp3	\
0	04-09-2012	58	635	...	7	0
1	08-03-2014	38	11	...	5	0
2	21-08-2013	26	426	...	4	0
3	10-02-2014	26	11	...	6	0
4	19-01-2014	94	173	...	5	0

AcceptedCmp4	AcceptedCmp5	AcceptedCmp1	AcceptedCmp2	Complain	\
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0

Z_CostContact	Z_Revenue	Response
0	3	11
1	3	11
2	3	11
3	3	11
4	3	11

[5 rows x 29 columns]

Gambar 1 Data Awal

Semua kolom mempunyai jumlah nilai non-null sebanyak 2.216 entri, yang artinya tidak terdapat missing value pada dataset ini. Hal ini menunjukkan bahwa dataset sudah bersih dan siap digunakan untuk proses Machine Learning tanpa perlu imputasi data.

Hasil Data Preprocessing

Tahapan data preprocessing dilakukan untuk memastikan bahwa data yang akan digunakan dalam proses pemodelan Machine Learning berada dalam kondisi yang bersih, terstruktur, dan sesuai dengan kebutuhan algoritma Naive Bayes. Tahapan ini bertujuan untuk meningkatkan kualitas data sehingga proses klasifikasi dapat menghasilkan model yang lebih akurat dan optimal. Proses preprocessing meliputi beberapa langkah, yaitu pemeriksaan kualitas data, pembersihan data, transformasi tipe data, encoding variabel kategorikal, serta pembentukan label target klasifikasi. Langkah awal yang dilakukan adalah pemeriksaan data untuk mengidentifikasi adanya data duplikat dan nilai kosong (missing values). Berdasarkan hasil eksplorasi dataset, tidak ditemukan data duplikat maupun nilai kosong pada atribut

yang digunakan dalam penelitian. Oleh karena itu, tidak diperlukan proses penanganan missing values seperti imputasi atau penghapusan data. Kondisi ini menunjukkan bahwa Marketing Campaign Dataset memiliki kualitas data yang baik dan layak digunakan sebagai sumber data dalam proses analisis.

Index: 2216 entries, 0 to 2239
 Data columns (total 25 columns):

#	Column	Non-Null Count	Dtype
0	Year_Birth	2216 non-null	int64
1	Education	2216 non-null	object
2	Marital_Status	2216 non-null	object
3	Income	2216 non-null	float64
4	Kidhome	2216 non-null	int64
5	Teenhome	2216 non-null	int64
6	Recency	2216 non-null	int64
7	MntWines	2216 non-null	int64
8	MntFruits	2216 non-null	int64
9	MntMeatProducts	2216 non-null	int64
10	MntFishProducts	2216 non-null	int64
11	MntSweetProducts	2216 non-null	int64
12	MntGoldProds	2216 non-null	int64
13	NumDealsPurchases	2216 non-null	int64
14	NumWebPurchases	2216 non-null	int64
15	NumCatalogPurchases	2216 non-null	int64
16	NumStorePurchases	2216 non-null	int64
17	NumWebVisitsMonth	2216 non-null	int64
18	AcceptedCmp3	2216 non-null	int64
19	AcceptedCmp4	2216 non-null	int64
20	AcceptedCmp5	2216 non-null	int64
21	AcceptedCmp1	2216 non-null	int64
22	AcceptedCmp2	2216 non-null	int64
23	Complain	2216 non-null	int64
24	Response	2216 non-null	int64

Gambar 2 Hasil Pembersihan Data

Selanjutnya dilakukan transformasi terhadap beberapa atribut kategorikal menggunakan teknik Label Encoding agar data dapat diproses oleh algoritma Naive Bayes. Proses ini mengubah nilai kategorikal menjadi representasi numerik tanpa menghilangkan informasi yang terkandung di dalamnya. Selain itu, dilakukan pembentukan variabel baru yang merepresentasikan total nilai pembelian pelanggan berdasarkan akumulasi pengeluaran pada berbagai kategori produk. Tahap terakhir adalah pembentukan label target klasifikasi yang digunakan untuk mengelompokkan pelanggan ke dalam kategori diskon rendah, sedang, dan tinggi berdasarkan karakteristik perilaku pembelian yang dimiliki. Melalui tahapan preprocessing ini, data yang semula masih berupa data mentah berhasil diubah menjadi data yang siap digunakan dalam proses pelatihan dan pengujian model klasifikasi Naive Bayes.

Tabel 2 Label Encoding

Kategori	Kode
Rendah	0
Sedang	1
Tinggi	2

Hasil Training Model Naïve Bayes

Setelah tahap preprocessing selesai dilakukan, proses berikutnya adalah membangun model klasifikasi menggunakan algoritma Naive Bayes. Data yang telah dipersiapkan kemudian dibagi menjadi dua bagian, yaitu 80% sebagai data latih (training data) dan 20% sebagai data uji (testing data). Data latih digunakan untuk melatih model agar dapat mengenali pola hubungan antara atribut perilaku pembelian pelanggan dengan kategori diskon, sedangkan data uji digunakan untuk mengukur kemampuan model dalam melakukan klasifikasi terhadap data yang belum pernah dipelajari sebelumnya. Pada penelitian ini, model Naive Bayes digunakan untuk mengklasifikasikan pelanggan ke dalam tiga kategori diskon, yaitu rendah, sedang, dan tinggi, berdasarkan karakteristik perilaku pembelian yang dimiliki. Setelah proses pelatihan selesai, model dievaluasi menggunakan beberapa metrik kinerja, yaitu akurasi (accuracy), presisi (precision), recall, dan F1-score. Metrik-metrik tersebut digunakan untuk menilai tingkat ketepatan model dalam mengklasifikasikan pelanggan ke dalam kategori diskon yang sesuai serta untuk mengetahui performa model secara keseluruhan.

	precision	recall	f1-score	support
0	0.79	0.86	0.82	304
1	0.44	0.40	0.42	111
2	0.33	0.17	0.23	29
accuracy			0.70	444
macro avg	0.52	0.47	0.49	444
weighted avg	0.67	0.70	0.68	444

Gambar 3 Hasil Training

Berdasarkan hasil evaluasi model menggunakan Classification Report, diperoleh nilai akurasi sebesar 70%, yang menunjukkan bahwa model Naive Bayes

mampu mengklasifikasikan kategori diskon pelanggan dengan tingkat ketepatan sebesar 70% dari keseluruhan data uji sebanyak 444 data. Evaluasi dilakukan terhadap tiga kategori diskon, yaitu kategori 0 (Rendah), 1 (Sedang), dan 2 (Tinggi).

Pada kategori 0 (Diskon Rendah), model menghasilkan nilai precision sebesar 0,79, recall sebesar 0,86, dan f1-score sebesar 0,82 dengan jumlah data (support) sebanyak 304 pelanggan. Hasil ini menunjukkan bahwa model memiliki kemampuan yang baik dalam mengenali pelanggan yang termasuk dalam kategori diskon rendah. Tingginya nilai recall mengindikasikan bahwa sebagian besar pelanggan pada kategori ini berhasil diklasifikasikan dengan benar oleh model.

Pada kategori 1 (Diskon Sedang), diperoleh nilai precision sebesar 0,44, recall sebesar 0,40, dan f1-score sebesar 0,42 dengan jumlah data sebanyak 111 pelanggan. Nilai evaluasi yang lebih rendah dibandingkan kategori sebelumnya menunjukkan bahwa model masih mengalami kesulitan dalam membedakan pelanggan yang termasuk kategori diskon sedang dengan kategori lainnya.

Sementara itu, pada kategori 2 (Diskon Tinggi), model memperoleh nilai precision sebesar 0,33, recall sebesar 0,17, dan f1-score sebesar 0,23 dengan jumlah data sebanyak 29 pelanggan. Hasil ini menunjukkan bahwa kemampuan model dalam mengidentifikasi pelanggan yang termasuk kategori diskon tinggi masih relatif rendah. Kondisi tersebut dapat disebabkan oleh jumlah data pada kategori diskon tinggi yang jauh lebih sedikit dibandingkan kategori lainnya sehingga model memiliki keterbatasan dalam mempelajari pola karakteristik pelanggan pada kelas tersebut.

Secara keseluruhan, nilai macro average sebesar 0,52 untuk precision, 0,47 untuk recall, dan 0,49 untuk f1-score menunjukkan performa rata-rata model pada setiap kelas tanpa mempertimbangkan jumlah data pada masing-masing kategori. Sementara itu, nilai weighted average sebesar 0,67 untuk

precision, 0,70 untuk recall, dan 0,68 untuk f1-score menunjukkan performa model dengan mempertimbangkan proporsi jumlah data pada setiap kategori. Hasil evaluasi ini mengindikasikan bahwa algoritma Naive Bayes memiliki kemampuan yang cukup baik dalam mengklasifikasikan pelanggan ke dalam kategori diskon, terutama pada kategori diskon rendah, namun masih memerlukan perbaikan untuk meningkatkan performa klasifikasi pada kategori diskon sedang dan tinggi.

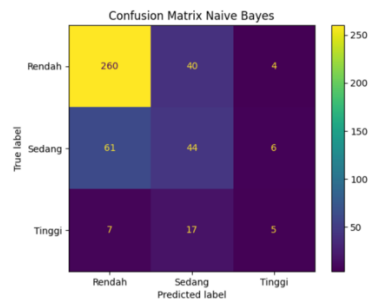
Hasil penelitian menunjukkan bahwa implementasi algoritma Naive Bayes tidak hanya mampu mengklasifikasikan pelanggan berdasarkan perilaku pembelian, tetapi juga mendukung proses pengambilan keputusan secara otomatis dalam menentukan kategori diskon. Melalui pemanfaatan model klasifikasi berbasis data, perusahaan retail dapat mengurangi ketergantungan pada proses analisis manual sehingga penentuan strategi promosi menjadi lebih cepat, konsisten, dan objektif.

Dari perspektif transformasi digital berkelanjutan, penerapan algoritma Naive Bayes menunjukkan bagaimana teknologi analitik dapat dimanfaatkan untuk meningkatkan efisiensi proses bisnis melalui pengambilan keputusan berbasis data (data-driven decision making). Otomatisasi proses klasifikasi pelanggan memungkinkan perusahaan mengelola data dalam jumlah besar secara lebih efektif sehingga mampu meningkatkan kualitas layanan dan daya saing bisnis. Selain itu, penerapan sistem klasifikasi ini juga sejalan dengan konsep Green IT karena membantu mengurangi aktivitas administrasi manual, meminimalkan penggunaan dokumen cetak dalam proses analisis dan pelaporan, serta mengoptimalkan pemanfaatan infrastruktur teknologi informasi. Dengan demikian, implementasi algoritma Naive Bayes tidak hanya memberikan manfaat dalam meningkatkan efektivitas strategi promosi, tetapi juga mendukung penerapan transformasi digital yang

berkelanjutan dan penggunaan teknologi informasi yang lebih efisien.

Hasil Evaluation

Setelah proses pelatihan model menggunakan algoritma Naive Bayes selesai dilakukan, tahap selanjutnya adalah melakukan evaluasi model untuk mengukur kemampuan model dalam mengklasifikasikan pelanggan ke dalam kategori diskon yang sesuai. Evaluasi dilakukan untuk mengetahui seberapa baik model dapat mengenali pola perilaku pembelian pelanggan dan memprediksi kategori diskon rendah, sedang, maupun tinggi secara tepat. Pada penelitian ini, proses evaluasi dilakukan menggunakan Confusion Matrix serta beberapa metrik pengukuran kinerja, yaitu Accuracy, Precision, Recall, dan F1-score. Hasil evaluasi tersebut digunakan untuk menilai tingkat ketepatan klasifikasi yang dihasilkan oleh model Naive Bayes serta mengidentifikasi kemampuan model dalam membedakan setiap kategori diskon pelanggan berdasarkan perilaku pembelian yang dimiliki.



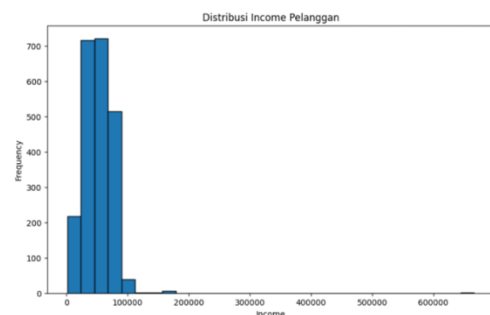
Gambar 4 Confusion Matrix

Berdasarkan Confusion Matrix menunjukkan hasil klasifikasi pelanggan ke dalam tiga kategori diskon, yaitu Rendah, Sedang, dan Tinggi, menggunakan algoritma Naive Bayes. Baris pada matriks menunjukkan kelas sebenarnya (true label), sedangkan kolom menunjukkan hasil prediksi model (predicted label).

Pada kategori Diskon Rendah, terdapat 260 data yang berhasil diklasifikasikan dengan benar sebagai kategori Rendah. Namun, terdapat 40 data yang seharusnya termasuk kategori

Rendah tetapi diprediksi sebagai kategori Sedang, serta 4 data yang diprediksi sebagai kategori Tinggi. Hasil ini menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam mengenali pelanggan yang termasuk kategori diskon rendah. Pada kategori Diskon Sedang, sebanyak 44 data berhasil diklasifikasikan dengan benar sebagai kategori Sedang. Sementara itu, terdapat 61 data yang salah diklasifikasikan menjadi kategori Rendah dan 6 data yang diprediksi sebagai kategori Tinggi. Kondisi ini menunjukkan bahwa model masih mengalami kesulitan dalam membedakan pelanggan kategori Sedang dengan kategori Rendah. Pada kategori Diskon Tinggi, hanya 5 data yang berhasil diprediksi dengan benar sebagai kategori Tinggi. Selain itu, terdapat 7 data yang diprediksi sebagai kategori Rendah dan 17 data yang diprediksi sebagai kategori Sedang. Hasil ini menunjukkan bahwa kemampuan model dalam mengenali pelanggan kategori diskon tinggi masih relatif rendah dibandingkan kategori lainnya.

Secara keseluruhan, nilai diagonal utama pada Confusion Matrix yaitu 260, 44, dan 5 menunjukkan jumlah data yang berhasil diklasifikasikan dengan benar oleh model. Sementara nilai di luar diagonal menunjukkan jumlah kesalahan klasifikasi yang terjadi pada masing-masing kategori. Dominasi prediksi yang benar pada kategori Rendah menunjukkan bahwa model lebih mampu mengenali kelas yang memiliki jumlah data lebih banyak dibandingkan kelas lainnya.



Gambar 5 Distribusi Income

Berdasarkan histogram distribusi Income pelanggan, terlihat bahwa

sebagian besar pelanggan memiliki pendapatan pada rentang Rp20.000 hingga Rp80.000. Frekuensi tertinggi berada pada kelompok pendapatan menengah, yang menunjukkan bahwa mayoritas pelanggan dalam dataset berasal dari segmen ekonomi menengah. Distribusi data cenderung tidak simetris dan membentuk pola miring ke kanan (right-skewed), di mana terdapat sejumlah kecil pelanggan yang memiliki pendapatan sangat tinggi dibandingkan sebagian besar pelanggan lainnya.

Hasil penelitian menunjukkan bahwa implementasi algoritma Naive Bayes mampu mengklasifikasikan pelanggan berdasarkan perilaku pembelian sehingga dapat mendukung proses pengambilan keputusan dalam menentukan kategori diskon. Melalui model klasifikasi berbasis data, perusahaan retail dapat mengurangi ketergantungan pada proses analisis manual sehingga penentuan strategi promosi menjadi lebih cepat, konsisten, dan objektif.

Dari perspektif transformasi digital, penelitian ini menunjukkan bahwa pemanfaatan machine learning dapat membantu perusahaan retail mengelola data pelanggan secara lebih efektif. Otomatisasi proses klasifikasi memungkinkan perusahaan memanfaatkan data historis transaksi sebagai dasar dalam menentukan strategi promosi yang lebih tepat sasaran, sehingga kualitas pelayanan kepada pelanggan dan efektivitas kegiatan pemasaran dapat terus ditingkatkan.

SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Naive Bayes berhasil diimplementasikan untuk menentukan kategori diskon pelanggan berdasarkan perilaku pembelian pada toko retail. Penelitian ini menggunakan Marketing Campaign Dataset yang berisi informasi mengenai karakteristik pelanggan,

aktivitas pembelian, dan respons terhadap kampanye pemasaran. Sebelum proses klasifikasi dilakukan, data terlebih dahulu melalui tahap preprocessing yang meliputi pembersihan data, transformasi data kategorikal ke bentuk numerik, serta pembentukan variabel dan label target kategori diskon. Selanjutnya, data dibagi menjadi data latih sebesar 80% dan data uji sebesar 20% untuk membangun dan menguji model klasifikasi.

Hasil pengujian menunjukkan bahwa model Naive Bayes memperoleh nilai akurasi sebesar 70%, yang menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam mengklasifikasikan pelanggan ke dalam kategori diskon rendah, sedang, dan tinggi. Berdasarkan hasil evaluasi menggunakan Classification Report dan Confusion Matrix, model menunjukkan performa terbaik pada kategori diskon rendah, sedangkan kategori diskon sedang dan tinggi masih memiliki tingkat klasifikasi yang lebih rendah. Hal ini disebabkan oleh ketidakseimbangan jumlah data pada setiap kategori sehingga model lebih mudah mengenali kelas yang memiliki jumlah data lebih banyak. Secara keseluruhan, penelitian ini menunjukkan bahwa algoritma Naive Bayes dapat digunakan sebagai metode klasifikasi yang efektif untuk membantu pihak retail menentukan kategori diskon pelanggan secara lebih cepat, konsisten, dan berbasis data, sehingga dapat mendukung penyusunan strategi promosi yang lebih tepat sasaran serta meningkatkan efektivitas kegiatan pemasaran.

Selain memberikan kontribusi dalam meningkatkan efektivitas penentuan kategori diskon pelanggan, penelitian ini juga mendukung implementasi transformasi digital melalui penerapan pengambilan keputusan berbasis data (data-driven decision making). Pemanfaatan algoritma Naive Bayes memungkinkan proses klasifikasi pelanggan dilakukan secara otomatis, lebih cepat, dan lebih konsisten dibandingkan proses manual. Dengan

demikian, hasil penelitian ini tidak hanya membantu perusahaan dalam menyusun strategi promosi yang lebih tepat sasaran, tetapi juga mendukung pemanfaatan teknologi informasi untuk meningkatkan kualitas layanan dan daya saing bisnis di sektor retail.

DAFTAR PUSTAKA

- Sugiana, N. S., & Musty, B. (2023). Analisis Data Sistem Informasi Monitoring Marketing; Tools Pengambilan Keputusan Strategic. *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, 696-707.
- Tovarel, M. R., Wahidah, N. N., Hidayatullah, W., & Malik, A. (2024). ANALISIS STRATEGI PROMOSI DALAM MENINGKATKAN JUMLAH PELANGGAN LADY FAME SHOP DI BANDAR LAMPUNG. *Al-A. mal Jurnal Manajemen Bisnis Syariah*, 1, 122-132.
- Apriana, D., & Yuliansyah, C. (2024). Optimalisasi Penjualan Online Melalui Teknik Data Mining (Studi Kasus E-Commerce). *AL-MIKRAJ Jurnal Studi Islam dan Humaniora*, 4, 514-527.
- Harahap, F., Fahrozi, W., Adawiyah, R., Siregar, E. T., & Harahap, A. Y. (2023). Implementasi Data Mining dalam Memprediksi Produk AC Terlaris untuk Meningkatkan Penjualan Menggunakan Metode Naive Bayes. *Jurnal Unitek*, 16, 41-5
- Puro, S. I., Hariyan, J., Rafliansyah, R., Aziz, M. R., & Rajagukguk, P. V. (2024). Analisa Data Shopping Trends Menggunakan Algoritma Klasifikasi Dengan Metode Naive Bayes. *Repeater : Publikasi Teknik Informatika dan Jaringan*, 2, 119-134.
- Naya, C., & Rilvani, E. (2025). Prediksi Penjualan Brand di HGVR Store Menggunakan Algoritma C4.5 dan Naive Bayes. *Jurnal Informatika Ekonomi Bisnis*, 7, 646-652.
- Santoso, M. I., Mardiah, N., & Sembiring, A. (2025). ANALISIS SENTIMEN PENGGUNA INSTAGRAM TERHADAP STATEMENT MENTERI KEUANGAN TENTANG "KEBIJAKAN GAJI GURU DAN DOSEN" MENGGUNAKAN NAIVE BAYES. *JOURNAL Of SCIENCE And SOCIAL RESEARCH*, 4, 4049 – 4056.
- Iqbal, Hidayat, N., Gevano, D. P., & Ilahi, A. P. (2025). Segmentasi Pelanggan Menggunakan K-Means Clustering Berdasarkan Data Kepribadian dan Pola Konsumsi. *Jurnal Teknik Informatika (JUTIF)*, 6, 3914-3924.
- Muhammadfadilah, & Tanti, L. (2025). Klasifikasi Jenis Usaha Mikro Kecil Menengah (UMKM) Berdasarkan Skala Usaha Menggunakan Algoritma C4.5. *SENADIMU*, 2, 814-827.
- Purnomo, N., Suri, R. M., Yuliana, D., Rasyid, M., & Lubis, J. R. (2023). Sistem Pakar Identifikasi Penyakit Kulit Melanoma dengan Metode Teorema Bayes. *Jurnal KomtekInfo*, 10, 56-63.
- Oktafiani, R., Hermawan, A., & Avianto, D. (2023). Pengaruh Komposisi Split Data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning. *Jurnal Sains dan Informatika*, 9