Indonesian Journal of Multidisciplinary on Social and Technology
Homepage: https://journal.ilmudata.co.id/index.php/ijmst
Vol. 4 No. 3 (2026) pp: 10323-10334
P-ISSN: 2986-6790, e-ISSN: 2986-6782

Analisis Faktor Minat Beli Generasi Z pada TikTok Shop Menggunakan Naïve Bayes

Lorenza Amanda, Haryanto, Harun Al Rasyid
Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika
lorenzamandaa@gmail.com*, haryanto.hyt@bsi.ac.id, harun.har@bsi.ac.id

Abstrak

Pertumbuhan e-commerce di Indonesia mendorong TikTok Shop menjadi platform belanja sosial yang memengaruhi minat beli Generasi Z. Namun, faktor yang paling berpengaruh kurang jelas sehingga strategi pemasaran digital sulit diarahkan secara tepat. Penelitian ini bertujuan mengklasifikasikan minat beli Generasi Z pada TikTok Shop berdasarkan konten promosi, konten kreator, dan produk, mengukur kinerja model, serta mengidentifikasi faktor yang paling memengaruhi. Penelitian ini merupakan penelitian kuantitatif berbasis data mining dengan kerangka kerja CRISP-DM. Data dikumpulkan melalui survei Google Form dari 150 responden Generasi Z di Kelurahan Ciater. Minat beli dikategorikan menjadi TINGGI dan RENDAH, kemudian diklasifikasikan menggunakan algoritma Gaussian Naïve Bayes dengan 15 atribut prediktor pada RapidMiner dan divalidasi dengan 10-fold cross validation. Pembobotan atribut dilakukan menggunakan Weight by Information Gain. Hasil menunjukkan model memperoleh akurasi 84,00%, precision 76,56%, recall 84,48%, dan AUC 0,889 yang termasuk kategori Baik. Pembobotan menunjukkan rekomendasi konten kreator (X2.4) sebagai atribut paling berpengaruh dengan bobot 0,229, diikuti manfaat produk (X3.4) sebesar 0,218 dan kesesuaian minat produk (X3.3) sebesar 0,205. Rata-rata bobot per variabel menunjukkan konten kreator (0,176) dan produk (0,174) berkontribusi hampir setara, sedangkan konten promosi paling kecil (0,076). Hasil ini menunjukkan pelaku usaha sebaiknya memprioritaskan kerja sama dengan kreator yang kredibel serta produk yang bermanfaat dan sesuai minat konsumen.

Kata kunci: Minat Beli, Generasi Z, TikTok Shop, Naïve Bayes, CRISP-DM

Abstract

The growth of e-commerce in Indonesia has made TikTok Shop a social shopping platform that influences the purchase intention of Generation Z. However, the most influential factors remain unclear, making it difficult to direct digital marketing strategies accurately. This study aims to classify Generation Z's purchase intention on TikTok Shop based on promotional content, creator content, and product, evaluate model performance, and identify the most influential factors. This is a quantitative study based on data mining using the CRISP-DM framework. Data were collected through a Google Form survey of 150 Generation Z respondents in Ciater Subdistrict. Purchase intention was categorized into HIGH and LOW, then classified using the Gaussian Naïve Bayes algorithm with 15 predictor attributes in RapidMiner and validated with 10-fold cross validation. Attribute weighting was performed using Weight by Information Gain. The results show that the model achieved an accuracy of 84.00%, precision of 76.56%, recall of 84.48%, and an AUC of 0.889, which falls into the Good category. The weighting identifies creator recommendation (X2.4) as the most influential attribute with a weight of 0.229, followed by product benefit (X3.4) at 0.218 and product-interest suitability (X3.3) at 0.205. The average weight per variable shows that creator content (0.176) and product (0.174) contribute almost equally, while promotional content contributes the least (0.076). These results suggest that businesses should prioritize collaborating with credible creators and offering products that are beneficial and match consumer interests.

Keywords: Purchase Intention, Generation Z, TikTok Shop, Naïve Bayes, CRISP-DM

1. Pendahuluan

Perkembangan teknologi digital dan akses internet yang semakin baik telah membantu pertumbuhan *e-commerce* di Indonesia. TikTok merupakan salah satu platform yang berkembang besar, dan melalui fitur TikTok Shop platform ini menjadi sarana hiburan sekaligus alat pemasaran daring yang efektif. TikTok Shop dipahami sebagai platform belanja sosial tempat kreator dan pengguna dapat saling menjual dan membeli barang (Nopitasari & Suyatno, 2023), dan platform ini juga membuka peluang bagi UMKM untuk menjangkau audiens yang lebih luas (Maria, 2024). Laporan We Are Social dan Meltwater mencatat bahwa Indonesia memiliki 222.583.000 pengguna akun TikTok pada April 2026 (Maheswara, 2026). Jumlah ini menempatkan Indonesia sebagai salah satu pasar

TikTok terbesar di dunia dan membuka peluang besar bagi platform tersebut untuk memengaruhi perilaku konsumen, terutama minat beli.

Generasi Z merupakan kelompok pengguna aktif media sosial yang memiliki ketertarikan sangat kuat dengan teknologi (Arum et al., 2023). Pola konsumsinya cenderung cepat dan banyak dipengaruhi informasi digital, seperti konten promosi, ulasan pengguna, dan rekomendasi kreator yang mereka percayai (Deanah et al., 2026). Konten promosi yang menarik secara visual dan informatif mampu meningkatkan ketertarikan konsumen dalam waktu singkat, sedangkan kreator atau *influencer* berperan sebagai sarana pemasaran yang membuat produk dikenal dan mendorong konsumen untuk membeli (Suryati et al., 2024). Di sisi lain, karakteristik produk seperti harga dan kualitas turut menjadi pertimbangan utama dalam keputusan pembelian (Khoirunisa & Nurhayati, 2024).

Sejumlah penelitian terdahulu telah memanfaatkan algoritma Naïve Bayes untuk menganalisis minat beli konsumen. Saputra et al. (2025) meneliti faktor dominan minat beli Generasi Z di Shopee dengan 100 responden dan menemukan bahwa harga produk serta kecepatan pengiriman paling berpengaruh, dengan akurasi terbaik 73,75%. Wanny & Iqbal (2026) mengklasifikasikan minat belanja daring pengguna TikTok berdasarkan jenis konten dan menemukan bahwa konten promosi serta ulasan produk berkontribusi terhadap minat belanja. Assegaf et al. (2025) menunjukkan bahwa konten edukasi dan hiburan memengaruhi keputusan pembelian dengan akurasi 68,5%, sedangkan Zein & Suparni (2024) memperoleh nilai AUC sebesar 75% dalam memprediksi minat beli makanan daring pada generasi milenial.

Penelitian-penelitian tersebut memperlihatkan celah yang jelas. Variabel umumnya dianalisis secara terpisah atau dengan pendekatan statistik konvensional, dan sedikit penelitian yang melihat pengaruh konten promosi, konten kreator, dan produk secara bersamaan terhadap minat beli Generasi Z pada TikTok Shop. Selain itu, penelitian terdahulu tidak disertai pembobotan atribut yang secara objektif menunjukkan faktor paling berpengaruh.

Kebaruan penelitian ini terletak pada integrasi ketiga faktor tersebut ke dalam satu model klasifikasi Naïve Bayes pada konteks TikTok Shop, yang dilengkapi pembobotan atribut menggunakan *Weight by Information Gain* untuk menemukan atribut dengan kontribusi terbesar. Penelitian ini bertujuan mengklasifikasikan minat beli Generasi Z pada TikTok Shop berdasarkan konten promosi, konten kreator, dan produk, mengukur kinerja model melalui *accuracy, precision, recall,* dan *Area Under Curve* (AUC), serta mengidentifikasi faktor yang paling memengaruhi.

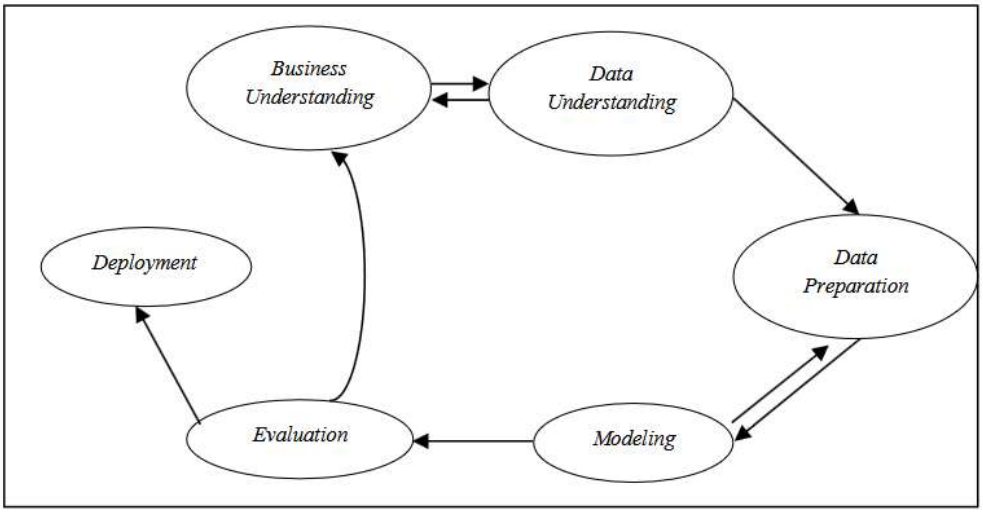
2. Metode Penelitian

Penelitian ini merupakan penelitian kuantitatif yang memanfaatkan teknik *data mining*, yaitu proses menggali informasi bernilai dari data yang sulit dikenali melalui analisis manual (Putra, 2025). Data primer diperoleh dari survei Google Form dan diolah mengikuti kerangka CRISP-DM.

2.1. Tahapan Penelitian

Penelitian ini menggunakan kerangka kerja CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang terdiri atas enam tanapan, yaitu *business understanding, data understanding, data preparation, modeling, evaluation, dan deployment* (Hidayati et al., 2021). Kerangka ini dipilih karena langkah-langkahnya terstruktur mulai dari pemahaman masalah hingga evaluasi model, sehingga analisis dapat dilakukan secara sistematis (Dhewayani et al., 2022). Alur kerja ditunjukkan pada Gambar 1.

DOI: https://doi.org/10.69693/ijmst.v4i3.14411
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



Gambar 1. Tahapan CRISP-DM (Dhewayani et al., 2022)

Pada tahap *business understanding* dirumuskan masalah, serta menentukan tujuan penelitian yang akan diselesikan dengan *data mining*. Tahap *data understanding* meliputi pengumpulan data dan peninjauan distribusi jawaban responden. Tahap *data preparation* mencakup pengecekan data, seleksi atribut, dan transformasi data berupa pembentukan label target, kemudian tahap *modeling* menerapkan algoritma Naïve Bayes menggunakan perangkat lunak RapidMiner, dan pembobotan atribut. Tahap *evaluation*, mengukur kinerja model, dan tahap *deployment* berupa penyusunan laporan hasil penelitian.

2.2. Populasi, Sampel, dan Pengumpulan Data

Populasi adalah keseluruhan objek atau subjek yang memiliki karakteristik tertentu dan menjadi dasar penarikan kesimpulan penelitian (Amin et al., 2023). Populasi penelitian ini adalah Generasi Z di Kelurahan Ciater yang menggunakan aplikasi TikTok, khususnya fitur TikTok Shop, yaitu individu yang lahir antara tahun 1998 hingga 2009. Sampel diambil dengan teknik *non-probability sampling* jenis *purposive sampling*, yaitu pemilihan sampel berdasarkan kriteria tertentu yang berhubungan dengan karakteristik populasi (Hardani, 2023). Jumlah sampel tidak ditentukan dengan rumus probabilitas, tetapi disesuaikan dengan kebutuhan penelitian dan kriteria responden.

Data dikumpulkan melalui kuesioner tertutup berskala Likert 1 sampai 5 yang disebarkan melalui Google Form, dengan rentang Sangat Tidak Setuju (1), hingga Sangat Setuju (5). Penelitian ini memiliki tiga variabel independen, yaitu Konten Promosi (X1), Konten Kreator (X2), dan Produk (X3), serta satu variabel dependen, yaitu Minat Beli (Y). Setiap variabel diukur dengan lima butir pernyataan sehingga terdapat 20 butir.

2.3. Algoritma Naïve Bayes

Naïve Bayes adalah algoritma klasifikasi probabilistik berbasis Teorema Bayes dengan asumsi bahwa setiap prediktor bersifat independen secara kondisional terhadap kelas. Teorema Bayes dinyatakan pada persamaan (1) (Sihananto & Maulana, 2021).

$$P(c|x) = \frac{P(x|c) \cdot P(c)}{P(x)} \quad (1)$$

$P(c|x)$ probabilitas posterior kelas (c) berdasarkan prediktor (x). $P(c)$ probabilitas prior dari kelas (c), $P(x)$ probabilitas dari prediktor (x), dan $P(x|c)$ likelihood, yaitu probabilitas dari prediktor (x) untuk kelas tertentu (c). Probabilitas prior setiap kelas dihitung dengan persamaan (2).

$$P(C) = \frac{n_C}{N} \quad (2)$$

DOI: https://doi.org/10.69693/ijmst.v4i3.14411
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

dengan n_C adalah jumlah data pada kelas C dan N adalah jumlah seluruh data. Karena seluruh atribut prediktor bersifat numerik, penelitian ini menggunakan varian Gaussian Naïve Bayes yang mengasumsikan setiap atribut berdistribusi normal pada tiap kelas. Likelihood dihitung dengan persamaan (3).

$$P(x_i|C) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \tag{3}$$

dengan x_i adalah nilai atribut ke-i, C adalah kelas, μ adalah rata-rata atribut pada kelas tersebut, dan σ adalah standar deviasi sampel atribut pada kelas tersebut. Kelas dengan skor posterior tertinggi, yaitu hasil kali prior dan seluruh likelihood atribut, ditetapkan sebagai kelas prediksi.

2.4. Evaluasi Model

Kinerja model dievaluasi dengan *confusion matrix*, yaitu tabel yang menggambarkan prediksi tepat dan tidak untuk mengukur prediksi tepat dan tidak tepat untuk mengukur keberhasilan model (Nurhidayat & Dewi, 2023; Firdaus & Setiadi, 2023). Matriks ini membuat empat nilai, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Dari nilai tersebut dihitung *accuracy*, *precision*, dan *recall* (Madani & Lubis, 2025; Sholeh et al., 2024) sebagaimana pada persamaan (4) sampai (6).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

Accuracy mengukur proporsi seluruh prediksi yang benar, *Precision* mengukur proporsi prediksi yang benar-benar positif, dan *Recall* mengukur data positif aktual yang berhasil terdeteksi. Selain itu, digunakan luas area di bawah kurva ROC dengan nilai *Area Under Curve* (AUC) yang bernilai 0 sampai 1 (Fakih et al., 2025). Nilai AUC pada rentang 0,90 – 1,00 dikategorikan Sangat Baik, 0,80 – 0,90 kategori Baik, 0,70 – 0,80 kategori Cukup, 0,60 – 0,70 kategori Rendah, dan 0,50 – 0,60 kategori Gagal (Ermanto & Surojudin, 2025).

2.5. *Weight by Information Gain*

Untuk mengetahui faktor yang paling memengaruhi minat beli digunakan operator *Weight by Information Gain* pada RapidMiner. Operator ini menghitung tingkat kepentingan setiap atribut terhadap label berdasarkan nilai *information gain*, sehingga setiap atribut memperoleh bobot yang sebanding dengan kontribusinya dalam membedakan kelas target (Budianita, 2023).

Alur proses pembobotan menggunakan dataset hasil *preprocessing* yang sama dengan proses klasifikasi. Dataset dibaca dengan operator *Read Excel*, atribut Kategori_Minat_Beli ditetapkan sebagi label dengan operator *Set Role*, kemudian dihubungkan ke operator *Weight by Information Gain*. Atribut dengan bobot tertinggi dinilai sebagai faktor yang paling berpengaruh.

3. Hasil dan Pembahasan

3.1. *Bussines Understanding*

Tahap ini, menetapkan tujuan penelitian dari sudut pendang pelaku usaha, lalu menejermahkannya menjadi tujuan *data mining*. Masalah bisnis yang dihadapi adalah kurang jelasnya faktor yang paling memengaruhi minat beli Generasi Z pada TikTok Shop, sehingga strategi pemasaran digital sulit diarahkan secara tepat. Tujuan bisnis adalah mengetahui kelompok faktor yang perlu diprioritaskan pelaku usaha, yaitu konten promosi, konten kreator, dan produk.

DOI: https://doi.org/10.69693/ijmst.v4i3.14411
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Tujuan tersebut diterjemahkan menjadi dua tujuan *data mining*, yaitu membangun model klasifikasi Naïve Bayes yang mengelompokkan responden ke dalam kategori minat beli TINGGI dan RENDAH, serta menghitung bobot atribut dengan *Weight by Information Gain* untuk menentukan atribut yang paling berpengaruh. Keberhasilan model diukur dengan *accuracy*, *precision*, *recall*, dan AUC dengan acuan nilai kategori, sedangkan keberhasilan analisis faktor dilihat dari terbentuknya peringkat bobot seluruh atribut. Data yang tersedia berupa numerik 1 sampai 5, dengan Microsoft Excel, dan RapidMiner sebagai perangkat analisis.

3.2. *Data Understanding*

Dataset dikumpulkan melalui survei Google Form yang berhasil mengumpulkan 150 responden Generasi Z pengguna aktif TikTok Shop di Kelurahan Ciater. Karakteristik responden berdasarkan jenis kelamin, usia, dan lama penggunaan TikTok Shop disajikan pada Tabel 1.

Tabel 1 Karakteristik Responden

Karakteristik	Kategori	Jumlah	%
Jenis Kelamin	Perempuan	87	58,0
	Laki-laki	63	42,0
Usia	17-20 Tahun	26	17,3
	21-24 Tahun	102	68,0
	25-28 Tahun	22	14,7
Lama Penggunaan	1-6 Bulan	13	8,7
	6-12 Bulan	2	1,3
	1-2 Tahun	27	18,0
	>2 Tahun	108	72,0

Statistik deskriptif pada Tabel 2 menunjukkan seluruh variabel memiliki skor rata-rata di atas 3,5 pada skala 1 sampai 5, yang berarti persepsi responden secara umum positif terhadap konten promosi, konten kreator, produk, dan minat beli pada TikTok Shop.

Tabel 2 Statistik Deskriptif Variabel

Variabel	Rata-rata	Skor
Konten Promosi	3,99	1 – 5
Konten Kreator	3,82	1 – 5
Produk	3,95	1 – 5
Minat Beli	3,98	1 – 5

3.3. *Data Preparation*

Pembersihan data dilakukan pada Microsoft Excel dengan memeriksa data hilang, entri ganda, nilai di luar rentang skala 1 sampai 5, serta format data. Tidak ada data yang dihapus sehingga seluruh 150 responden dipakai. Pada seleksi atribut, data identitas seperti alamat email, nama, usia, jenis kelamin, dan lama penggunaan dihapus. Lima belas butir pernyataan konten promosi, konten kreator, dan produk dipertahankan sebagai atribut prediktor, sedangkan lima butir minat beli digunakan untuk membentuk label. Seleksi atribut disajikan pada Tabel 3.

DOI: https://doi.org/10.69693/ijmst.v4i3.14411
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Tabel 3 Hasil Seleksi Atribut

Atribut	Kode	Variabel
Daya tarik visual promosi.	X1.1	Konten Promosi
Kejelasan informasi promosi .	X1.2	Konten Promosi
Kreativitas konten promosi.	X1.3	Konten Promosi
Konten promosi sesuai dengan minat.	X1.4	Konten Promosi
Frekuensi muncul konten promosi.	X1.5	Konten Promosi
Kesesuaian rekomendasi kreator.	X2.1	Konten Kreator
Kejelasan penyampaian kreator.	X2.2	Konten Kreator
Kejujuran kreator.	X2.3	Konten Kreator
Rekomendasi konten kreator.	X2.4	Konten Kreator
Rekam jejak kreator.	X2.5	Konten Kreator
Kesesuaian harga.	X3.1	Produk
Kesesuain harga dan kualitas.	X3.2	Produk
Kesesuain minat produk.	X3.3	Produk
Manfaat produk.	X3.4	Produk
Variasi menarik produk.	X3.5	Produk
Ketertarikan membeli.	Y.1	Minat Beli
Keinginan membeli.	Y.2	Minat Beli
Niat membeli dalam waktu dekat.	Y.3	MinaBeli
Mencoba membeli.	Y.4	Minat Beli
Pilihan utama membeli.	Y.5	Minat Beli

Selanjutnya transformasi data, label Kategori_Minat_Beli dibentuk dari rata-rata lima indikator minat beli (Y) menggunakan fungsi AVERAGE dan IF pada Microsoft Excel. Responden dengan rata-rata 4 atau lebih diberi label TINGGI dan sisanya RENDAH. Kelima butir minat beli tidak dipakai lagi dalam pemodelan karena sudah terwakili oleh label. Transformasi data di sajikan pada Tabel 4

Tabel 4 Transformasi Data

No.	Y.1	Y.2	Y.3	Y.4	Y.5	Rata-rata	Label
1.	5	5	5	5	5	5,0	TINGGI
2.	4	4	4	3	4	3,8	RENDAH
3.	2	4	5	3	5	3,8	RENDAH
4.	2	4	5	2	3	3,2	RENDAH
5.	4	5	5	5	5	4,8	TINGGI
…..	…..	…..	…..	…..	…..	…..	…..
150.	5	4	5	4	5	4,6	TINGGI

3.4. *Modeling*

Menerapkan dua teknik pada dataset hasil *data preparation* yang berisi 15 atribut prediktor, dan satu label, yaitu klasifikasi Naïve Bayes dan pembobotan atribut dengan *Weight by Information Gain*. Perhitungan manual pada Microsoft Excel dilakukan terlebih dahulu untuk memahami tahapan algoritma. Probabilitas prior dihitung sebagai berikut:

$$\text{P (TINGGI)} = \frac{Jumlah\ data\ kelas\ TINGGI}{Jumlah\ seluruh\ data} = \frac{92}{150} = 0{,}6133$$

$$\text{P (RENDAH)} = \frac{Jumlah\ data\ kelas\ RENDAH}{Jumlah\ seluruh\ data} = \frac{58}{150} = 0{,}3867$$

Rata-rata *(mean)* dan standar deviasi dihitung dengan fungsi AVERAGEIF dan STDEV.S untuk 15 atribut pada masing-masing kelas, hasilnya disajikan pada Tabel 5.

DOI: https://doi.org/10.69693/ijmst.v4i3.14411
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Tabel 5 Hasil Perhitungan Mean dan Standar Deviasi

Atribut	Mean TINGGI	SD TINGGI	Mean RENDAH	SD RENDAH
X1.1	4,184	0,769	3,500	0,681
X1.2	4,184	0,694	3,706	0,749
X1.3	4,293	0,703	3,810	0,907
X1.4	4,217	0,767	3,275	0,744
X1.5	4,271	0,771	3,775	1,009
X2.1	4,217	0,608	3,413	0,773
X2.2	4,239	0,731	3,500	0,941
X2.3	3,923	0,841	2,965	0,972
X2.4	4,304	0,691	3,172	0,939
X2.5	4,163	0,842	3,344	0,738
X3.1	4,304	0,886	3,582	0,701
X3.2	4,271	0,727	3,534	0,777
X3.3	4,239	0,716	3,310	0,706
X3.4	4,206	0,703	3,224	0,899
X3.5	4,369	0,690	3,551	0,841

Atribut X2.4 (rekomendari konten kreator) memiliki selisih rata-rata antar kelas terbesar, yaitu 1,132 (4,304 pada kelas TINGGI dan 3,172 pada kelas RENDAH), sedangkan atribut X1.3 (kreativitas konten promosi) memiliki selisih terkecil , yaitu 0,483. Likelihood Gaussian dihitung dengan fungsi NORM.DIST berdasarkan persamaan (3), kemudian dikalikan untuk seluruh atribut dan dikalikan dengan prior untuk memproleh skor posterior.

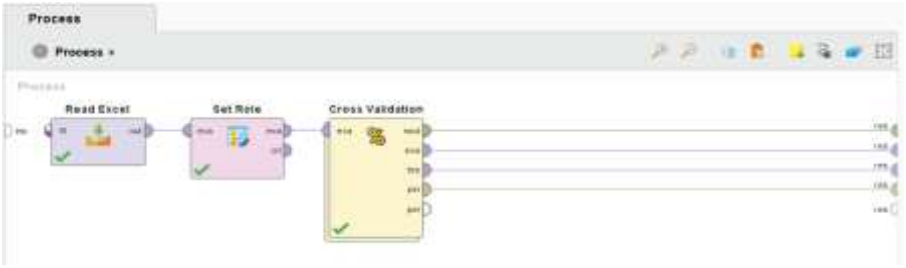
Pada data pertama, skor posterior kelas TINGGI sebesar 1,4263649E-07 jauh melampaui kelas RENDAH sebesar 7,7532780E-16, sehingga data tersebut diprediksi TINGGI sesuai kelas aktualnya. Proses yang sama diterapkan pada seluruh 150 data. Hasilnya, 85 data diprediksi TINGGI dan 65 data diprediksi RENDAH, dengan 129 prediksi benar dan 21 prediksi salah. Hasil perhitungan satu data Likelihood Gaussian dan skor posterior disajikan pada Tabel 6.

Tabel 6 Hasil Perhitungan Likelihood Gaussian dan Skor Posterior

Atribut	Nilai Data	P(X|TINGGI)	P(X|RENDAH)
X1.1	5	0,295783	0,052037
X1.2	4	0,554650	0,493110
X1.3	4	0,519637	0,430237
X1.4	5	0,309071	0,036663
X1.5	5	0,331182	0,189451
X2.1	4	0,615423	0,387100
X2.2	4	0,516783	0,368068
X2.3	5	0,209327	0,046024
X2.4	5	0,347819	0,063955
X2.5	5	0,289096	0,043877
X3.1	5	0,330795	0,074702
X3.2	5	0,332258	0,086755
X3.3	5	0,316835	0,032243
X3.4	5	0,300228	0,063107
X3.5	5	0,380799	0,107761
Total P(X|C)		2,325595E-07	2,0051581E-15
Prior		0,6133	0,3867
Skor Posterior		1,4263649E-07	7,7532780E-16
Posterior		100,00%	0,00%

DOI: https://doi.org/10.69693/ijmst.v4i3.14411
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Pemodelan otomatis dilakukan pada RapidMiner dengan alur prosesnya terdiri atas operator *Read Excel* untuk memuat data, *Set Role* untuk menetapkan Kategori_Minat_Beli sebagai label, dan *Cross Validation* dengan k = 10. Di dalam Cross Validation, bagian training memakai operator Naïve Bayes, sedangkan bagian testing memakai *Apply Model* dan *Performance* seperti pada Gambar 2 dan 3.

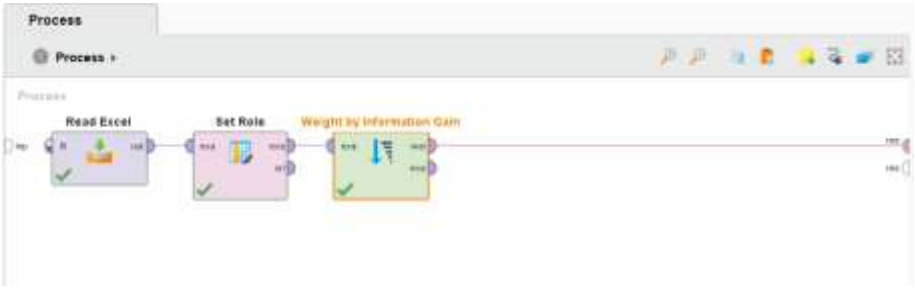


Gambar 2 Alur Proses Naive Bayes pada RapidMiner

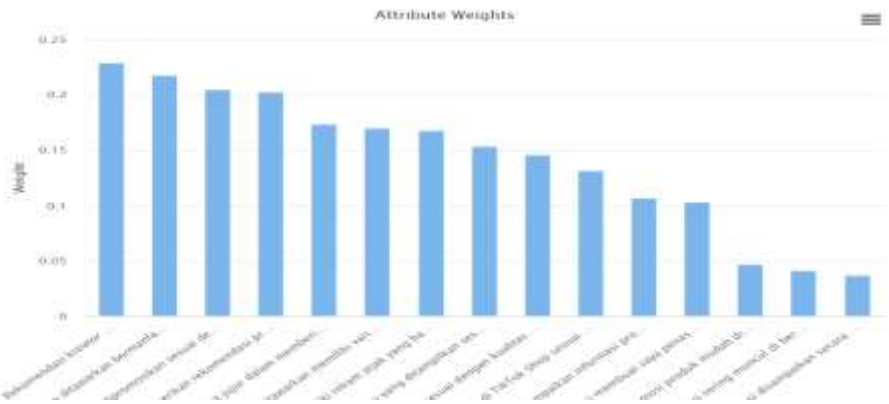


Gambar 3 Alur Proses Naive Bayes pada RapidMiner

Pembobotan atribut dilakukan dengan operator *Weight by Information Gain* pada dataset yang sama dengan proses klasifikasi, dengan label Kategori_Minat_Beli, mengikuti alur pada Gambar 4. Hasil pembobotan seluruh atribut disajikan pada Gambar 5 dan Tabel 6.



Gambar 4 Alur Proses Weight by Information Gain



Gambar 5 Bobot Atribut Hasil Weight by Information Gain

DOI: https://doi.org/10.69693/ijmst.v4i3.14411
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Tabel 6 Bobot Information Gain Seluruh Atribut

Rank	Kode	Atribut	Bobot
1	X2.4	Rekomendasi konten kreator.	0,229
2	X3.4	Manfaat produk.	0,218
3	X3.3	Kesesuain minat produk.	0,205
4	X2.1	Kesesuaian rekomendasi kreator.	0,203
5	X2.3	Kejujuran kreator.	0,174
6	X3.5	Variasi menarik produk.	0,170
7	X2.5	Rekam jejak kreator.	0,168
8	X1.4	Konten promosi sesuai dengan minat.	0,154
9	X3.2	Kesesuain harga dan kualitas.	0,146
10	X3.1	Kesesuaian harga.	0,132
11	X2.2	Kejelasan penyampaian kreator.	0,107
12	X1.1	Daya tarik visual promosi.	0,103
13	X1.2	Kejelasan informasi promosi.	0,047
14	X1.5	Frekuensi muncul konten promosi.	0,041
15	X1.3	Kreativitas konten promosi.	0,037

Atribut rekomendasi konten kreator (X2.4) memperoleh bobot tertinggi sebesar 0,229, sedangkan atribut kreativitas konten promosi (X1.3) memperoleh bobot terendah sebesar 0,037. Jika bobot dirata-ratakan per variabel, konten kreator memperoleh rata-rata 0,176, produk 0,174, dan konten promosi hanya 0,076. Artinya, konten kreator dan produk memiliki kontribusi yang hampir setara dalam membedakan minat beli TINGGI dan RENDAH, sedangkan konten promosi memiliki kontribusi paling kecil. Seluruh atribut kreator kecuali X2.2 berada di peringkat tujuh teratas, dan empat dari lima atribut konten promosi berada di peringkat 12 ke bawah.

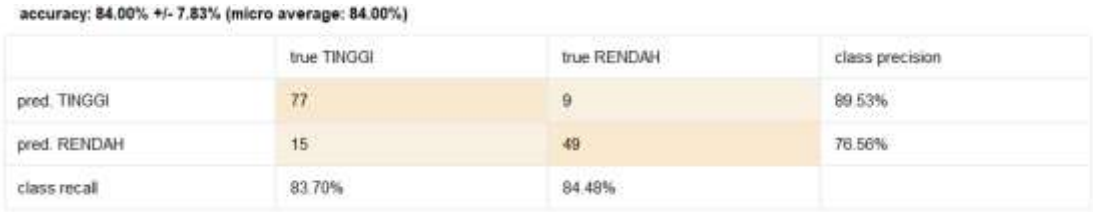
3.5 *Evaluation*

Perhitungan pada Microsoft Excel dihitung pada data yang sama dengan data latih sehingga cenderung lebih optimistis dan hanya digunakan untuk memahami tahapan algoritma. Oleh karena itu, hasil RapidMiner dengan *10-fold cross validation* dijadikan acuan utama kinerja model. AUC tidak dapat dihitung secara manual karena memerlukan variasi *threshold* yang diperoleh dari proses *Cross Validation*. Tabel 7 menunjukkan *confusion matrix* hasil *Cross Validation* pada RapidMiner dengan RENDAH sebagai kelas positif TP 49, TN 77, FP 15, dan FN 9.

Tabel 7 Confusion Matrix

Prediksi	Aktual TINGGI	Aktual RENDAH
TINGGI	77	9
RENDAH	15	49

Berdasarkan persamaan (4) sampai (6), pada Gambar 6 *accuracy* sebesar 84,00%. Pada Gambar 7 diperoleh nilai *precision* sebesar 76,56%, dan pada Gambar 8 diperoleh nilai *recall* sebesar 84,48%. *Precision* berarti bahwa dari seluruh responden yang diprediksi berminat beli RENDAH, benar-benar berminat beli RENDAH. *Recall* berarti bahwa dari seluruh yang sebenarnya berminat RENDAH, berhasil dikenali model.



accuracy: 84.00% +/- 7.83% (micro average: 84.00%)

	true TINGGI	true RENDAH	class precision
pred. TINGGI	77	9	89.53%
pred. RENDAH	15	49	76.56%
class recall	83.70%	84.48%	

Gambar 6 Hasil Accuracy

DOI: https://doi.org/10.69693/ijmst.v4i3.14411
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

precision: 77.68% +/- 11.07% (micro average: 76.56%) (positive class: RENDAH)

	true TINGGI	true RENDAH	class precision
pred. TINGGI	77	9	89.53%
pred. RENDAH	15	49	76.56%
class recall	83.70%	84.48%	

Gambar 7 Hasil Precision

recall: 85.00% +/- 19.95% (micro average: 84.48%) (positive class: RENDAH)

	true TINGGI	true RENDAH	class precision
pred. TINGGI	77	9	89.53%
pred. RENDAH	15	49	76.56%
class recall	83.70%	84.48%	

Gambar 8 Hasil Recall

Nilai *recall* yang lebih tinggi daripada *precision* menunjukkan model cenderung lebih berhati-hati dalam melewatkan responden berminat RENDAH, dengan konsekuensi 15 responden berminat TINGGI ikut diprediksi RENDAH. Model juga memperoleh AUC sebesar 0,889 (± 0,067), yang termasuk kategori Baik, sehingga model mampu membedakan kelas TINGGI dan RENDAH pada berbagai nilai *threshold* seperti terlihat pada Gambar 9.



Gambar 9 Hasil AUC

3.6. *Deployment*

Tahap *deployment* ini diwujudkan dalam bentuk laporan dan artikel ilmiah ini, tanpa pengembangan sistem prediksi yang berjalan secara langsung. Hasil yang disampaikan berupa model klasifikasi Naïve Bayes dengan kinerja yang tergolong baik serta peringkat atribut berdasarkan bobot *information gain*. Berdasarkan peringkat tersebut, pelaku usaha di TikTok Shop dapat memprioritaskan kerja sama dengan kreator yang kredibel dan jujur dalam mengulas produk, serta memastikan produk yang ditawarkan bermanfaat dan sesuai minat konsumen. Konten promosi tetap relevan, tetapi atributnya kurang membedakan minat beli TINGGI dan RENDAH sehingga tidak perlu menjadi satu-satunya fokus. Penggunaan model pada data baru memerlukan pengujian lebih lanjut dengan responden yang lebih banyak dan beragam daerahnya.

DOI: https://doi.org/10.69693/ijmst.v4i3.14411
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

4. Kesimpulan

Penelitian ini berhasil membangun model klasifikasi minat beli Generasi Z pada TikTok Shop berdasarkan konten promosi, konten kreator, dan produk menggunakan algoritma Naïve Bayes. Dari 150 responden, 92 responden (61,33%) tergolong berminat beli TINGGI dan 58 responden (38,67%) tergolong RENDAH. Model pada RapidMiner dengan *10-fold cross validation* memperoleh akurasi 84,00%, *precision* 76,56%, dan *recall* 84,48% dengan kelas positif RENDAH, serta AUC 0,889 yang termasuk kategori Baik, sehingga model mampu membedakan kelas kategori minat beli secara memadai. Berdasarkan pembobotan *Weight by Information Gain*, faktor yang paling memengaruhi minat beli adalah rekomendasi konten kreator (X2.4) dengan bobot 0,229, diikuti manfaat produk (X3.3). Pada tingkat variabel, konten kreator (rata-rata bobot 0,176) dan produk (0,174) memiliki kontribusi yang hampir setara, sedangkan konten promosi memiliki kontribusi terkecil (0,076). Temuan ini berimplikasi bahwa pelaku usaha di TikTok Shop dapat memprioritaskan kerja sama dengan kreator yang kredil dan jujur serta menawarkan produk bermanfaat dan sesuai minat konsumen, tanpa menjadikan konten promosi sebagai satu-satunya fokus. Meskipun demikian, hasil penelitian ini terbatas pada 150 responden di satu wilayah, sehingga kurang dapat digeneralisasi. Penelitian selanjutnya disarankan menggunakan responden yang lebih banyak dan beragam daerahnya, serta membandingkan Naïve Bayes dengan algoritma klasifikasi lain.

Reference

1. Amin, N., Garancang, S., & Abunawas, K. (2023). KONSEP UMUM POPULASI DAN SAMPEL DALAM PENELITIAN. *JURNAL PILAR: Jurnal Kajian Islam Kontemporer*, *14*(1), 15–31. https://doi.org/10.26618/whw41w62
2. Arum, L., Zahrani, A., & Duha, N. (2023). KARAKTERISTIK GENERASI Z DAN KESIAPANNYA DALAM MENGHADAPI BONUS DEMOGRAFI 2030. *Accounting Student Research Journal*, *2*(1), 59–72. https://doi.org/10.62108/asrj.v2i1.5812
3. Assegaf, H., Susanti, R., Wulandari, D., Aisa, S., & Santoso. (2025). IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK MENGANALISIS PENGARUH KONTEN MEDIA SOSIAL TERHADAP KEPUTUSAN PEMBELIAN. *Jurnal Aplikasi Bidang Matematika*, *2*(2), 10–17. https://doi.org/10.36841/jangka.v2i2.7954
4. Budianita, A. (2023). Information Gain Berbasis Algoritma Naive Bayes Classifier Pada Pemodelan Prediksi Kelulusan. *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, *5*(01), 1–10. https://doi.org/10.46772/intech.v5i1.1116
5. Deanah, N., Riorini, S. vandayuli, & Maidita, R. (2026). Pengaruh Ulasan Pelanggan di Media Sosial terhadap Keputusan Pembelian Generasi Z. *Jurnal Ilmu Sosial & Hukum*, *4*(1), 1882–1889. https://doi.org/10.61104/alz.v4i1. 3389
6. Dhewayani, F., Amelia, D., Alifah, D., Sari, B., & Jajuli, M. (2022). Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi (JATI)*, *12*(1), 64–77. https://doi.org/10.34010/jati.v12i1.6674
7. Ermanto, & Surojudin, N. (2025). Perbandingan Metode Klasifikasi dalam Memprediksi Penyakit Ginjal Kronis. *Jurnal Informatika Ekonomi Bisnis*, *7*, 715–723. https://doi.org/10.37034/infeb.v7i3.1263
8. Fakih, A., Hamzami, M. A., Hadianto, M. R., & Shafira, N. I. (2025). Perbandingan Akurasi Algoritma C4 . 5 dan K-NN Untuk Prediksi Kelulusan Mahasiswa Penerima Beasiswa. *Jurnal Komputer Antartika*, *3*(c), 18–25. https://doi.org/10.70052/jka.v3i1.623
9. Firdaus, L., & Setiadi, T. (2023). Perbandingan Algoritma Naive Bayes , Decision Tree , dan KNN untuk Klasifikasi Produk Populer Adidas US dengan Confusion Matrix. *Jurnal Sistem Komputer Dan Informatika (JSON)*, *5*(2), 185–195. https://doi.org/10.30865/json.v5i2.6124
10. Hardani, P. R. (2023). ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI MINAT BELI KONSUMEN TERHADAP PRODUK FAST FASHION DI INDONESIA. *Jurnal Akuntansi Dan Manajemen*, *34*(3), 174–189. https://doi.org/10.53916/jam.v34i3.113
11. Hidayati, N., Suntoro, J., & Setiaji, G. G. (2021). Perbandingan Algoritma Klasifikasi untuk Prediksi Cacat Software dengan Pendekatan CRISP-DM. *Jurnal Sains Dan Informatika*, *7*(2), 117–126. https://doi.org/10.34128/jsi.v7i2.313
12. Khoirunisa, F., & Nurhayati, S. (2024). Pengaruh Customers Online Review , Customers Online Rating , dan Harga Produk terhadap Keputusan Pembelian Melalui Marketplace Shopee. *Jurnal Inovasi Bisnis Dan Manajemen Indonesia*, *07*(04), 456–469. https://doi.org/10.31842/jurnalinobis.v7i4.336
13. Madani, F., & Lubis, A. (2025). CatBoost Algorithm Implementation for Classifying Women's Fashion Products. *JITE (Journal of Informatics and Telecommunication Engineering)*, *9*(1), 249–260. https://doi.org/10.31289/jite.v9i1.15604
14. Maheswara, R. (2026). *Indonesia Jadi Negara Pengguna TikTok Terbanyak April 2026*. Dataloka.Id. https://dataloka.id/humaniora/6567/indonesia-jadi-negara-pengguna-tiktok-terbanyak-april-2026/
15. Maria, V. (2024). Analisis Peran Tiktok Shop dalam Mendorong Ekonomi UMKM di Indonesia. *Jurnal Penelitian Bisnis Dan Manajemen*, *2*(2), 138–146. https://doi.org/10.47861/sammajiva.v2i2.992
16. Nopitasari, N. A., & Suyatno, D. F. (2023). Analisis Kepuasan Pengguna Fitur TikTok Shop pada Aplikasi TikTok Menggunakan Metode End User Computing Satisfaction (EUCS) dan DeLone and McLean. *JEISBI (Journal of Emerging Information Systems and Business Intelligence)*, *04*(03), 9–20. https://doi.org/10.26740/jeisbi.v4i3.54177
17. Nurhidayat, R., & Dewi, E. K. (2023). PENERAPAN ALGORITMA K-NEAREST NEIGHBOR DAN FITUR EKSTRAKSI N-GRAM DALAM ANALISIS SENTIMEN BERBASIS ASPEK. *Komputa : Jurnal Ilmiah Komputer Dan Informatika*, *12*(1), 91–100. https://doi.org/10.34010/komputa.v12i1.9458
18. Putra, A. (2025). PENERAPAN DATA MINING UNTUK PREDIKSI HARGA KAYU MANIS MENGGUNAKAN ALGORITMA REGRESI LINEAR SEDERHANA (Studi Kasus : Usaha Dagang Munang). *Jurnal Sains Informatika Terapan (JSIT)*, *04*(03), 574–580. https://doi.org/10.62357/jsit.v4i3.825
19. Saputra, P. A., Irawan, S. S., & Prianto, R. (2025). Analisis Faktor Dominan Minat Beli Generasi Z di Shopee Menggunakan Algoritma Naïve Bayes. *Jurnal Teknik Informatika*, *4*(3). https://doi.org/10.56211/sudo.v4i3.1137
20. Sholeh, M., Lestari, U., & Andayati, D. (2024). Comparison of Feature Selection with Information Gain Method in Decision Tree, Regression Logistic and Random Forest Algorithms. *Journal of Applied Business and Technology Comparison (JABT)*, *5*(3), 146–153.

DOI: https://doi.org/10.69693/ijmst.v4i3.14411
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

https://doi.org/10.35145/jabt.v5i3.155
21. Sihananto, A., & Maulana, H. (2021). Studi Literatur Tentang Performa Naïve Bayes dalam Klasifikasi Data. *Seminar Nasional Informatika Bela Negara (SANTIKA)*, *2*, 132–135. https://doi.org/10.33005/santika.v2i0.134
22. Suryati, E., Bahtiar, D., & Rohimah Laila, D. (2024). Pengaruh Kredibilitas Influencer Dan Kesadaran Merek Terhadap Minat Beli Produk Hanasui. *Jurnal Minfo Polgan*, *13*(2), 1963–1970. https://doi.org/10.33395/jmp.v13i2.14289
23. Wanny, P., & Iqbal, M. (2026). Klasifikasi Minat Belanja Online pada TikTok Berdasarkan Konten Media Sosial Menggunakan Metode Naive Bayes. *Journal of Informatics Management and Information Technology*, *6*(1), 107–115. https://doi.org/10.47065/jimat.v6i1.967
24. Zein, R., & Suparni. (2024). IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI MINAT BELI MAKANAN SECARA ONLINE PADA GENERASI MILENIAL. *JATI (Jurnal Mahasiswa Teknik Informatika*, *8*(6). https://doi.org/10.36040/jati.v8i6.11768

DOI: https://doi.org/10.69693/ijmst.v4i3.14411
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)