

Analisis Faktor Dominan Minat Beli Generasi Z di Shopee Menggunakan Algoritma Naïve Bayes

Pingki Ans Saputra, Rahmaddeni *, Sandra Septi Irawan, Robi Prianto, Delfi

Teknik Informatika, Universitas Sains dan Teknologi Indonesia, Pekanbaru, Indonesia

INFORMASI ARTIKEL

Diterima Redaksi: 03 Agustus 2025
Revisi Akhir: 17 Oktober 2025
Diterbitkan Online: 10 November 2025

KATA KUNCI

E-Commerce
Generasi Z
Minat Beli
Naïve Bayes
Shopee

KORESPONDENSI (*)

E-mail: rahmaddeni@usti.ac.id

A B S T R A K

Pertumbuhan pesat *e-commerce* di Indonesia telah mengubah perilaku konsumen, khususnya Generasi Z yang dikenal sebagai pengguna digital *native*. Penelitian ini bertujuan untuk menganalisis faktor-faktor dominan yang memengaruhi minat beli Generasi Z di *platform* Shopee dengan menggunakan algoritma Naïve Bayes. Data dikumpulkan melalui survei terhadap 100 responden Generasi Z aktif pengguna Shopee, dengan variabel yang meliputi harga, ulasan produk, kecepatan pengiriman, metode pembayaran, dan pengaruh sosial. Pengolahan data dilakukan menggunakan metode SEMMA (*Sample, Explore, Modify, Model, Assess*) dan diimplementasikan dalam perangkat lunak RapidMiner. Hasil penelitian menunjukkan bahwa faktor harga dan kecepatan pengiriman menjadi dua variabel paling dominan yang memengaruhi minat beli. Model klasifikasi terbaik diperoleh pada skenario *split* data 80:20 dengan akurasi sebesar 73,75%, *recall* kelas "YA" sebesar 88,89%, dan *precision* yang seimbang. Temuan ini memberikan kontribusi strategis bagi pelaku *e-commerce*, khususnya Shopee, untuk mengarahkan strategi pemasaran yang lebih sesuai dengan karakteristik dan preferensi Generasi Z.

PENDAHULUAN

Indonesia mengalami kemajuan pesat di bidang perdagangan *online* seiring dengan meningkatnya pengguna jaringan internet dan perangkat mobile di era digital ini. *Shopee*, sebagai salah satu *platform E-commerce* terbesar di Asia Tenggara, telah memanfaatkan pertumbuhan ini dengan menghadirkan berbagai fitur inovatif seperti *flash sale*, *cashback*, dan interaksi sosial melalui *Shopee Live* [1]. Generasi Z, yang lahir antara tahun 1997 hingga 2012, merupakan kelompok demografis yang paling aktif dalam berbelanja *online* karena mereka adalah digital native tumbuh bersama teknologi dan sangat terhubung dengan media sosial. Karakteristik unik Generasi Z, seperti ketergantungan pada ulasan produk, preferensi terhadap transaksi cepat, dan kepekaan terhadap harga serta diskon, membuat mereka menjadi segmen konsumen yang menarik untuk diteliti. Bagi pelaku pemasaran dan pengelola *platform* belanja online, mengenali pola konsumsi Generasi Z merupakan hal penting dalam menyusun strategi bisnis yang tepat [2].

Dalam konteks *Shopee*, minat beli Generasi Z dipengaruhi oleh berbagai faktor, baik internal maupun eksternal. Faktor-faktor tersebut meliputi harga produk, diskon atau promo, rating dan ulasan dari pembeli sebelumnya, metode pembayaran yang tersedia, serta pengalaman pengguna (*user experience*) dalam aplikasi [3]. Rekomendasi dari tokoh berpengaruh maupun orang terdekat menjadi salah satu aspek sosial yang turut menentukan keputusan belanja konsumen. Untuk menganalisis faktor-faktor dominan yang memengaruhi minat beli Generasi Z, pendekatan data *mining* dapat digunakan karena kemampuannya dalam mengekstrak pola dari dataset besar [4]. Algoritma Naïve Bayes dipilih dalam penelitian ini karena efisiensinya dalam menangani data kategorikal dan kemampuannya untuk memberikan probabilitas prediktif berdasarkan sejumlah variabel independen. Dengan menggunakan RapidMiner sebagai alat analisis, penelitian ini

bertujuan untuk mengidentifikasi faktor mana yang paling berpengaruh terhadap minat beli Generasi Z di Shopee, sehingga dapat memberikan rekomendasi strategis bagi pelaku bisnis [5].

Beberapa penelitian sebelumnya telah mengkaji faktor-faktor yang memengaruhi minat beli konsumen di *platform e-commerce*. Penelitian [6] mengenai peramalan predikat kelulusan mahasiswa di suatu perguruan tinggi (Universitas XYZ) memanfaatkan metode Naive Bayes bersamaan dengan Decision Tree sebagai dasar analisisnya. Pada algoritma naïve bayes, diperoleh tingkat akurasi sebesar 86%. Sementara itu, algoritma decision tree menunjukkan akurasi sebesar 60%. Pada penelitian selanjutnya [7], yang berjudul “Analisis Perbandingan Algoritma Naive Bayes dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan” hasil pengujian dengan penggunaan metode algoritma Naive Bayes nilai keakuratan data untuk menunjukkan keefektifan dataset yang sedang diolah yang diterapkan mencapai 94%. Pada penerapan algoritma C4.5, diperoleh tingkat ketepatan prediksi sebesar 92,60% dalam mengestimasi kelulusan tepat waktu mahasiswa Penelitian [8], mengenai prediksi demensia Alzheimer menemukan bahwa Naive Bayes menghasilkan akurasi prediksi tertinggi sebesar 99%, mengungguli algoritma Decision Tree C4.5. decision tree yaitu 94%.

Merujuk pada pembahasan sebelumnya, studi ini dirancang dengan tiga tujuan inti sebagai landasan penelitian. Pertama, penelitian ini bertujuan mengkaji variabel-variabel kunci yang berdampak pada minat beli Gen Z di *platform* Shopee, meliputi pertimbangan harga, ulasan konsumen, sistem pembayaran, serta pengaruh lingkungan sosial. Kedua, membangun model prediktif berbasis algoritma Naïve Bayes melalui platform RapidMiner untuk mengkuantifikasi kontribusi masing-masing faktor prediktor terhadap variabel minat beli. Ketiga, memberikan rekomendasi praktis bagi Shopee dan pelaku *E-commerce* lainnya dalam merancang strategi pemasaran yang lebih efektif untuk menjangkau Generasi Z. Studi ini diharapkan bermanfaat baik secara teoretis bagi pengembangan ilmu pemasaran digital maupun secara praktis untuk meningkatkan performa penjualan di kalangan Generasi Z. Dengan demikian, penelitian ini menjadi relevan baik dari sisi teoretis maupun praktis dalam konteks perkembangan *E-commerce* yang semakin kompetitif.

TINJAUAN PUSTAKA

Minat Beli Konsumen

Minat beli merupakan salah satu komponen penting dalam teori perilaku konsumen yang menunjukkan adanya keinginan atau kecenderungan seseorang untuk membeli suatu produk atau jasa. Menurut [9], minat beli mencerminkan kesiapan konsumen dalam melakukan tindakan pembelian yang didasari oleh persepsi positif terhadap suatu produk atau merek. Minat beli tidak bersifat spontan, melainkan terbentuk melalui serangkaian proses psikologis dan sosial yang kompleks, meliputi pemahaman terhadap kebutuhan, pencarian informasi, evaluasi alternatif, hingga keputusan pembelian.

Berbeda dengan model perilaku konsumen tradisional, ekosistem digital memperkenalkan dimensi baru dalam pembentukan minat beli dimana faktor-faktor seperti platform credibility dan user review sentiment menjadi sama pentingnya dengan pertimbangan harga dan kualitas, Analisis terkini mengidentifikasi variabel-variabel prediktif baru berupa skor rating produk, riwayat interaksi konsumen (*purchase history*), kualitas visual *merchandising*, dan tingkat *social media virality* sebagai komponen esensial model pembelian digital. Penelitian [10] tidak hanya mengidentifikasi hubungan linier antara variabel-variabel tersebut dengan minat beli, tetapi juga menemukan bahwa efeknya bersifat kumulatif, semakin banyak faktor yang dipersepsikan positif, semakin kuat intensi pembelannya. Dalam konteks *e-commerce*, minat beli (*purchase intention*) secara luas diakui sebagai metrik prediktif utama dan indikator kunci untuk mengestimasi potensi pasar suatu produk sebelum peluncuran komersial.

Minat beli juga bersifat dinamis dan dipengaruhi oleh segmentasi konsumen yang berbeda, seperti usia, pendidikan, gender, hingga generasi. Dalam konteks Generasi Z, minat beli memiliki dimensi yang lebih kompleks karena dipengaruhi oleh faktor emosional dan digital engagement yang tinggi. Mereka lebih tertarik pada pengalaman berbelanja yang cepat, mudah, personal, dan terhubung secara sosial. Maka dari itu, penting untuk mengidentifikasi faktor-faktor spesifik yang membentuk minat beli pada segmen ini agar strategi pemasaran yang dilakukan dapat lebih tepat sasaran dan efektif [11].

Generasi Z

Generasi Z adalah generasi yang lahir antara tahun 1997 hingga 2012, dan dianggap sebagai generasi pertama yang sepenuhnya tumbuh dalam lingkungan digital. Mereka tidak mengalami masa tanpa internet, ponsel pintar, dan media sosial, sehingga gaya hidup mereka sangat dipengaruhi oleh perkembangan teknologi informasi. Karakteristik utama Generasi Z adalah adaptif terhadap teknologi, memiliki kemampuan multitasking yang tinggi, serta mengutamakan efisiensi dan kecepatan dalam melakukan aktivitas sehari-hari, termasuk dalam hal berbelanja [12].

Dalam perspektif perilaku konsumen, Generasi Z memiliki pendekatan yang berbeda dalam proses pengambilan keputusan dibandingkan generasi sebelumnya. Mereka lebih percaya pada ulasan dari pengguna lain daripada iklan konvensional, lebih suka melihat video *unboxing* atau *review* di media sosial sebelum membeli produk, dan seringkali menjadikan pengalaman pengguna lain sebagai acuan dalam memilih produk. Selain itu, mereka lebih cenderung menggunakan smartphone sebagai media utama untuk menjelajahi *e-commerce*, mengakses informasi, dan melakukan transaksi [13].

Penelitian yang dilakukan oleh [14] menyebutkan bahwa Generasi Z sangat responsif terhadap kampanye digital yang interaktif, personal, dan berbasis visual. Mereka juga memiliki kesadaran sosial yang tinggi sehingga mempertimbangkan nilai-nilai etika dan keberlanjutan dalam keputusan pembelian. Oleh karena itu, perusahaan yang ingin menarik perhatian dan loyalitas Generasi Z harus memahami nilai-nilai yang mereka pegang, serta menyediakan pengalaman digital yang menyenangkan, cepat, dan mudah diakses. Mengetahui faktor-faktor apa saja yang paling memengaruhi minat beli Generasi Z, khususnya di *platform E-commerce* seperti Shopee, sangat penting untuk menciptakan strategi pemasaran yang relevan dengan karakteristik mereka.

Shopee

Shopee adalah *platform E-commerce* yang diluncurkan pada tahun 2015 oleh Sea Group, dan telah berkembang pesat menjadi salah satu marketplace terbesar di kawasan Asia Tenggara, termasuk Indonesia. Shopee berhasil membedakan dirinya dari kompetitor melalui berbagai inovasi fitur dan strategi pemasaran digital yang masif. Beberapa fitur unggulan Shopee seperti gratis ongkir, *flash sale*, cashback, Shopee Games, Shopee Live, serta kolaborasi dengan *influencer* membuat *platform* ini sangat populer di kalangan pengguna muda [15].

Sebagai salah satu *marketplace* yang paling banyak digunakan di Indonesia, Shopee telah menjadi bagian penting dalam kehidupan sehari-hari konsumen, terutama Generasi Z. *Platform* ini memberikan kemudahan dalam menemukan berbagai produk dari kategori yang beragam, sistem pencarian yang intuitif, serta sistem pembayaran dan pengiriman yang fleksibel. Selain itu, tampilan antarmuka aplikasi Shopee yang sederhana dan mudah digunakan menjadi daya tarik tersendiri bagi pengguna yang tidak ingin mengalami proses belanja yang rumit [16].

Shopee juga sangat aktif dalam melakukan promosi tematik seperti kampanye 9.9, 11.11, dan 12.12 yang disertai dengan diskon besar-besaran dan voucher menarik. Strategi ini terbukti efektif dalam meningkatkan frekuensi kunjungan dan pembelian pengguna. Selain itu, fitur penilaian dan ulasan pengguna yang tersedia secara terbuka berperan penting dalam membentuk persepsi konsumen terhadap produk. Dalam konteks ini, Shopee tidak hanya menyediakan *platform* transaksi, tetapi juga menjadi medium interaksi sosial dan pertukaran informasi antar konsumen. Oleh karena itu, Shopee menjadi objek studi yang relevan untuk menganalisis perilaku minat beli Generasi Z, khususnya dalam memahami faktor apa yang dominan memengaruhi keputusan pembelian mereka [17].

Naïve Bayes

Naïve Bayes adalah salah satu algoritma klasifikasi dalam *machine learning* yang berbasis pada prinsip probabilitas. Algoritma ini menggunakan Teorema Bayes untuk menghitung probabilitas suatu data termasuk ke dalam kelas tertentu, dengan asumsi bahwa setiap fitur bersifat independen terhadap fitur lainnya. Meskipun asumsi independensinya sering dianggap terlalu sederhana, algoritma Naïve Bayes terbukti cukup andal dan efisien dalam berbagai tugas klasifikasi, terutama dalam kasus-kasus di mana jumlah data besar dan kompleksitas model perlu dikendalikan [18].

Keunggulan utama dari Naïve Bayes adalah kemampuannya dalam memproses data dalam jumlah besar secara cepat, serta memberikan hasil klasifikasi yang akurat meskipun data tidak sepenuhnya memenuhi asumsi independensi. Algoritma ini sering digunakan dalam *text classification* (seperti *spam detection* dan *sentiment analysis*), prediksi perilaku pengguna, serta pengambilan keputusan berbasis data. Dalam penggunaannya di bidang *e-commerce*, Naïve Bayes dapat membantu dalam mengklasifikasikan konsumen berdasarkan preferensi, minat beli, dan kecenderungan pembelian [19].

Dalam konteks penelitian ini, algoritma Naïve Bayes digunakan untuk menganalisis dan mengidentifikasi faktor-faktor yang paling dominan memengaruhi minat beli Generasi Z di *platform* Shopee. Data yang diperoleh dari kuesioner diolah menggunakan perangkat lunak RapidMiner, yang menyediakan antarmuka visual untuk pemrosesan data dan penerapan algoritma *machine learning*. Proses klasifikasi dengan Naïve Bayes akan membantu peneliti dalam memahami hubungan antara berbagai faktor (seperti harga, ulasan, promosi, dan kepercayaan) dengan intensi beli, sehingga dapat disimpulkan faktor mana yang paling berkontribusi terhadap keputusan pembelian konsumen Generasi Z [20].

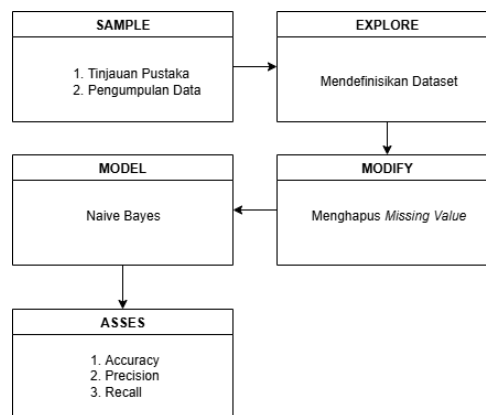
SEMMA

Metode SEMMA merupakan akronim dari lima tahapan utama, yaitu *Sample*, *Explore*, *Modify*, *Model*, dan *Assess*. Pendekatan ini dirancang sebagai standar dalam proses data *mining* yang berfungsi sebagai strategi sistematis dalam menyelesaikan berbagai permasalahan pada sektor bisnis maupun penelitian [21]. Adapun rincian tahapan tersebut meliputi :

1. *Sample* (Pengumpulan Data)
Tahapan ini berfokus pada proses pengumpulan data yang akan dijadikan bahan analisis dalam penelitian.
2. *Explore* (Eksplorasi Data)
Langkah ini mencakup proses pemilahan data untuk keperluan analisis, dikenal juga sebagai tahap *preprocessing*. Aktivitas dalam fase ini meliputi pembersihan data, pemberian label, serta normalisasi.
3. *Modify* (Transformasi Data)
Sebuah proses data *mining* yang digunakan untuk memodifikasi data dan mengubah variabel-variabel untuk memfokuskan proses pemilihan model. Tahap *modify* melakukan proses pembersihan, nilai yang hilang dapat dihilangkan, data yang tidak konsisten dapat diperiksa, dan data dapat diperbaiki.
4. *Model* (Pemodelan Data)
Tahap ini bertujuan membangun model berdasarkan data yang telah diproses. Salah satu teknik yang dapat digunakan adalah *elbow method* untuk menentukan nilai optimal seperti jumlah kluster dalam proses *clustering*.
5. *Assess* (Evaluasi)
Fase terakhir adalah penilaian terhadap performa model yang telah dibangun. Evaluasi ini menjadi pedoman penting dalam menilai efektivitas proses data *mining* yang dilakukan secara menyeluruh.

METODOLOGI

Didalam implementasi dan pengujian ini terdapat beberapa tahapan yaitu : Pada penelitian ini menggunakan Naïve Bayes. Data dikumpulkan melalui survei *Google Form* yang disebarakan kepada responden generasi Z yang aktif menggunakan aplikasi Shopee. Data tersebut akan di proses dengan penerapan metode SEMMA yang mengacu pada lima tahap yaitu :



Gambar 1. Metodologi Penelitian

Sample

Langkah ini diawali dengan penelusuran literatur yang relevan, baik dari jurnal ilmiah, buku, maupun sumber daring lainnya. Pengumpulan data dilakukan melalui survei *Google Form* yang disebarakan kepada responden generasi Z yang aktif menggunakan aplikasi Shopee.

TARGET	Harga produ...	Kualitas pro...	Saya sering ...	Harga produ...	Kemudahan ...	Saya lebih c...	Pengalaman ...	Saya lebih m...	Kecepatan p...	Saya lebih c...	Rekomenda...	Saya sangat ...	Saya akan m...	Shopee pert...
YA	5	4	5	5	4	4	5	4	4	5	4	5	5	5
TIDAK	1	1	4	2	1	2	4	3	4	4	3	2	3	2
TIDAK	1	4	3	2	3	3	5	3	2	1	5	3	2	4
TIDAK	1	3	4	4	1	3	4	2	5	4	4	3	1	2
YA	1	5	4	5	4	1	1	4	3	4	5	5	2	2
TIDAK	2	1	2	2	3	3	3	5	2	2	2	5	1	5
YA	3	4	5	5	3	1	2	1	5	4	2	1	2	4
YA	5	3	4	5	4	2	5	2	4	4	1	3	1	5
YA	5	4	2	1	1	5	4	5	5	1	1	3	3	5
YA	4	2	3	5	4	5	4	1	5	2	5	3	1	1
TIDAK	2	2	1	2	4	1	1	1	3	5	3	1	4	3
YA	2	3	5	1	5	5	5	2	5	3	4	1	2	5
TIDAK	5	1	1	2	2	2	4	2	5	3	4	1	5	2
YA	2	3	5	5	3	3	4	5	1	1	3	1	5	3
TIDAK	2	4	1	5	1	1	4	2	5	2	2	5	3	2
YA	5	2	3	2	4	3	4	2	3	2	3	5	5	4

Gambar 2. Data *Sample*

Tahap awal dalam penelitian ini dimulai dengan proses pengumpulan data, jumlah responden yang diharapkan adalah antara 100 orang. Data yang dikumpulkan mencakup berbagai atribut, seperti harga, kualitas produk, ulasan, pengalaman belanja, rating produk, kemudahan pembayaran, dan faktor-faktor lain yang mempengaruhi keputusan pembelian.

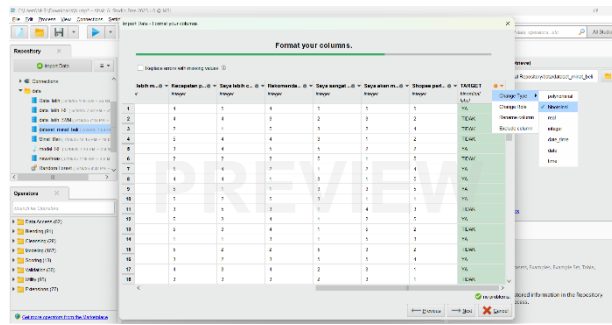
Explore

Eksplorasi data akan dilakukan pada tahap *explore* dengan memilih *trend* atau anomali yang tidak terduga. Setelah dataset berhasil diperoleh pada tahap ini, digunakan untuk pemahaman dan pemilihan data. Setelah itu dilakukan data *Understanding* yang terdapat 100 data *Example*, 14 atribut reguler dan 1 atribut label yaitu “TARGET” yang berisi dua label yakni “YA” dan “TIDAK”.

Label	Target	Ya	Tidak	Rata-rata
TARGET	Ya	56	44	50
Harga produk sangat mempen...	Ya	1	5	3,070
Kualitas produk sangat memp...	Ya	1	5	2,880
Saya sering melihat ulasan da...	Ya	1	5	2,830
Harga produk lebih penting da...	Ya	1	5	3,100
Kemudahan pembayaran di Sh...	Ya	1	5	2,990
Saya lebih cenderung membel...	Ya	1	5	3,010
Pengalaman belanja sebatam...	Ya	1	5	3,410
Saya lebih memilih berbelanja...	Ya	1	5	2,830
Kecepatan pengiriman produk ...	Ya	1	5	3,270
Saya lebih cenderung membel...	Ya	1	5	2,680
Rekomendasi dari teman atau ...	Ya	1	5	3,310
Saya sangat puas dengan pen...	Ya	1	5	3,130
Saya akan merekomendasikan...	Ya	1	5	2,930
Shopee perlu meningkatkan k...	Ya	1	5	3,040

Gambar 3. Data *Understanding*

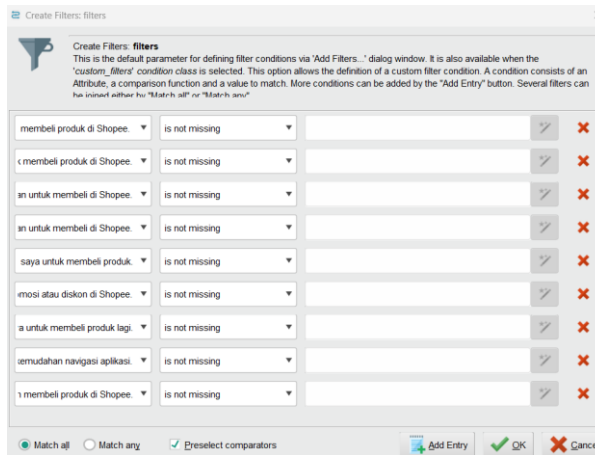
Berdasarkan Gambar 3, seluruh atribut input berupa pertanyaan persepsi konsumen terhadap layanan Shopee telah diatur dalam format numerik (*integer*) dengan skala penilaian *Likert* 1 sampai 5. Sementara itu, atribut TARGET yang menjadi variabel dependen diklasifikasikan sebagai atribut bertipe *binomial* dengan dua kategori, yaitu “YA” dan “TIDAK”, yang merepresentasikan preferensi konsumen untuk tetap menggunakan atau tidak menggunakan layanan Shopee di masa mendatang. Dari total data yang digunakan, sebanyak 56 responden menjawab YA dan 44 responden menjawab TIDAK, sebagaimana ditampilkan dalam Gambar 3. Setiap atribut independen memiliki nilai minimum 1 dan maksimum 5, menunjukkan kesesuaian skala penilaian. Nilai rata-rata pada masing-masing atribut juga dianalisis untuk melihat kecenderungan sikap responden terhadap setiap aspek layanan. Misalnya, atribut "Harga produk lebih penting dari kualitas" memiliki rata-rata tertinggi sebesar 3,41, menunjukkan bahwa faktor harga menjadi salah satu pertimbangan utama dalam keputusan pembelian konsumen. Selain itu, atribut "Kecepatan pengiriman produk Shopee" juga memiliki rata-rata tinggi yaitu 3,27, menunjukkan bahwa aspek logistik menjadi perhatian penting bagi pengguna. Sementara itu, atribut dengan rata-rata terendah yaitu "Saya lebih cenderung membeli produk mahal dibandingkan produk murah" dengan nilai 2,68, menandakan bahwa konsumen cenderung sensitif terhadap harga.



Gambar 4. Tampilan Memilih Atribut dan Variabel

Gambar 4, menunjukkan proses awal pengolahan data pada tahap *formatting column* menggunakan perangkat lunak RapidMiner. Dalam tahap ini, setiap atribut atau kolom data hasil survei yang terdiri dari pernyataan-pernyataan responden terkait pengalaman dan persepsi mereka terhadap layanan Shopee, diformat dengan tipe data numerik (*integer*) berdasarkan skala *Likert* 1–5. Atribut-atribut tersebut meliputi indikator seperti “Saya lebih memilih berbelanja *online*,” “Kecepatan pengiriman produk,” “Rekomendasi teman atau influencer,” hingga “Shopee perlu meningkatkan kualitas pelayanan.” Setiap responden memberikan penilaian terhadap pernyataan tersebut dengan skala dari 1 (sangat tidak setuju) hingga 5 (sangat setuju). Sementara itu, kolom TARGET yang merupakan variabel dependen dalam penelitian ini diklasifikasikan sebagai atribut bertipe binomial, yang terdiri dari dua label yaitu “YA” dan “TIDAK.” Label ini merepresentasikan keputusan responden apakah mereka bersedia untuk tetap menggunakan Shopee di masa mendatang (YA) atau tidak (TIDAK). RapidMiner memungkinkan pengguna untuk mengubah tipe atribut ini menjadi “label,” yang penting untuk mempersiapkan data dalam proses klasifikasi menggunakan algoritma *machine learning*. Tahap ini sangat krusial karena memastikan bahwa data telah ditetapkan secara benar sebelum dilanjutkan ke proses pelatihan model klasifikasi Naïve Bayes.

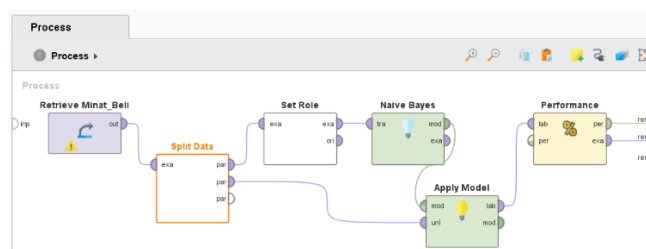
Modify



Gambar 5. Filter Example

Pengolahan data merupakan proses yang dilakukan pada tahap perubahan. Pada tahapan ini proses *cleaning* dilakukan untuk penghapusan data pada semua atribut termasuk atribut target untuk mengecek nilai yang missing value dan diperoleh dari total 100 baris data.

Model



Gambar 6. Proses Naïve Bayes Menggunakan Rapidminer

Dalam penelitian ini, pemilihan rasio split data 80:20 didasarkan pada pertimbangan keseimbangan antara jumlah data latih yang cukup untuk membangun model yang stabil dan jumlah data uji yang memadai untuk menilai performa model secara objektif. Rasio 80:20 juga merupakan konfigurasi yang paling umum digunakan dalam praktik machine learning karena memberikan kompromi optimal antara bias dan variance. Selain itu, hasil pengujian menunjukkan bahwa pada rasio ini model Naïve Bayes menghasilkan akurasi tertinggi dibandingkan rasio lain (90:10, 70:30, dan 60:40), sehingga justifikasi metodologisnya semakin kuat. Rasio ini memungkinkan model memperoleh cukup banyak pola dari data latih tanpa mengorbankan validitas hasil pengujian, sehingga menghasilkan kinerja yang paling representatif terhadap keseluruhan data Generasi Z pengguna Shopee.

Pada gambar 6, menunjukkan alur proses implementasi algoritma Naïve Bayes dalam perangkat lunak RapidMiner untuk menganalisis dan mengklasifikasikan minat beli Generasi Z di platform Shopee. Proses dimulai dengan operator “Retrieve Minat_Beli” yang berfungsi untuk memanggil dataset utama dari repositori proyek. Selanjutnya, data dibagi menggunakan operator “Split Data” ke dalam data latih dan data uji dengan rasio yang telah ditentukan, seperti 90:10, 80:20, 70:30, dan 60:40. Data latih kemudian diarahkan ke operator “Set Role” untuk menetapkan atribut target, yaitu variabel yang akan diprediksi (minat beli), sehingga algoritma dapat membedakan antara atribut fitur dan label. Setelah itu, data latih diproses melalui operator “Naïve Bayes” untuk membangun model klasifikasi berbasis probabilistik. Model yang telah terbentuk kemudian diterapkan ke data uji menggunakan operator “Apply Model”, di mana hasil prediksi (label) akan dibandingkan dengan label asli. Akhirnya, operator “Performance” digunakan untuk mengevaluasi hasil klasifikasi berdasarkan metrik evaluasi seperti akurasi, precision, dan recall, yang menjadi dasar dalam menilai efektivitas model dalam mengklasifikasikan minat beli konsumen.

Asses

Tahap assess digunakan untuk mengevaluasi kinerja dari model yang telah dibangun. Tahapan ini bertujuan untuk menilai performa model yang telah dibangun dengan cara membandingkan hasil prediksi terhadap data uji dengan label sentimen yang telah ditentukan sebelumnya. Proses evaluasi dilakukan menggunakan metrik precision, recall, dan accuracy sebagai indikator kinerjanya.

HASIL DAN PEMBAHASAN

accuracy: 65.56%

	true YA	true TIDAK	class precision
pred. YA	43	24	64.18%
pred. TIDAK	7	16	69.57%
class recall	86.00%	40.00%	

Gambar 7. Splitting Data 90:10

Pada gambar 7, menggunakan skenario 90:10 diperoleh akurasi sebesar 65,56%, dengan recall untuk kelas YA sebesar 86,00% dan kelas TIDAK hanya 40,00%. Precision untuk kelas YA sebesar 64,18%, sedangkan kelas TIDAK sebesar 69,57%. Nilai recall yang tinggi pada kelas YA menunjukkan bahwa model mampu mengenali sebagian besar data positif dengan baik, tetapi kesulitan mengenali data negatif (kelas TIDAK), sehingga berdampak pada akurasi keseluruhan yang masih tergolong rendah.

accuracy: 73.75%

	true YA	true TIDAK	class precision
pred. YA	40	16	71.43%
pred. TIDAK	5	19	79.17%
class recall	88.89%	54.29%	

Gambar 8. Splitting Data 80:20

Pengujian pada rasio ini menghasilkan akurasi sebesar 73,75%, yang merupakan akurasi tertinggi dari seluruh skenario. recall kelas YA mencapai 88,89%, dan kelas TIDAK sebesar 54,29%, dengan precision masing-masing 71,43% dan 79,17%. Kombinasi antara nilai recall dan precision yang seimbang membuat performa model cukup optimal, terutama dalam mengenali kedua kelas dengan cukup baik.

accuracy: 72.86%

	true YA	true TIDAK	class precision
pred. YA	34	14	70.83%
pred. TIDAK	5	17	77.27%
class recall	87.18%	54.84%	

Gambar 9. Splitting Data 70:30

Akurasi model sebesar 72,86%, dengan *recall* kelas YA sebesar 87,18% dan kelas TIDAK sebesar 54,84%. *Precision* kelas YA tercatat 70,83% dan untuk kelas TIDAK sebesar 77,27%. Hasil ini menunjukkan bahwa walaupun akurasi sedikit lebih rendah dari skenario 80:20, model tetap stabil dalam mengenali kedua kelas, khususnya pada identifikasi minat beli (kelas YA).

accuracy: 70.00%

	true YA	true TIDAK	class precision
pred. YA	29	13	69.05%
pred. TIDAK	5	13	72.22%
class recall	85.29%	50.00%	

Gambar 10. Splitting Data 60:40

Dengan akurasi 70,00%, *recall* kelas YA adalah 85,29%, sedangkan kelas TIDAK sebesar 50,00%. *Precision* untuk masing-masing kelas adalah 69,05% (YA) dan 72,22% (TIDAK). Meskipun performa masih dapat diterima, peningkatan proporsi data uji pada skenario ini mulai menunjukkan penurunan ketepatan model dalam klasifikasi, terutama dari segi *recall* kelas negatif.

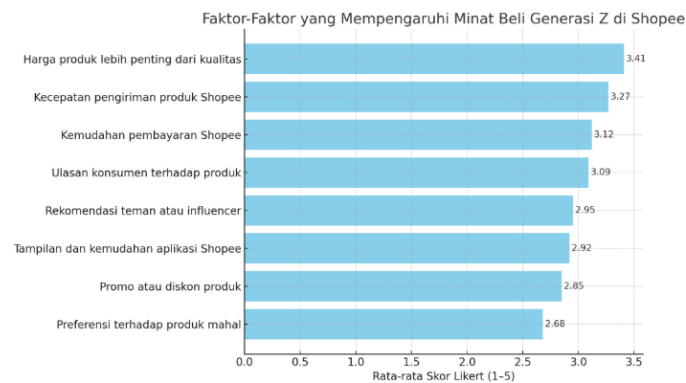
Split Data 80:20

Tabel 1. Hasil Evaluasi Model Naïve Bayes Berdasarkan Split Data

Split Data	Accuracy	recall (YA)	recall (Tidak)	Precision (YA)	Precision (TIDAK)
90:10	65,56%	86%	40%	64,18%	69,57%
80:20	73,75%	88,89%	54,29%	71,43%	79,17%
70:30	72,86%	87,18%	54,84%	70,83%	77,27%
60:40	70%	85,29%	50%	69,05%	72,22%

Hasil analisis juga menunjukkan bahwa nilai recall untuk kelas “TIDAK” (tidak berminat beli) secara konsisten lebih rendah dibandingkan kelas “YA”. Hal ini menunjukkan bahwa model Naïve Bayes lebih sensitif terhadap pola-pola yang mencerminkan minat beli (kelas positif) dibandingkan dengan pola ketidaktertarikan (kelas negatif). Fenomena ini kemungkinan disebabkan oleh distribusi data yang tidak seimbang, di mana responden dengan keputusan “YA” sedikit lebih banyak (56 responden) dibandingkan dengan “TIDAK” (44 responden), sehingga model lebih terlatih mengenali karakteristik pembeli yang berminat. Implikasinya bagi Shopee adalah perlunya perhatian terhadap segmen pengguna yang tidak berminat beli — misalnya dengan memperbaiki pengalaman pengguna, meningkatkan transparansi ulasan, atau menyesuaikan strategi promosi agar lebih relevan bagi kelompok yang cenderung pasif. Dengan memahami alasan rendahnya recall pada kelas “TIDAK”, Shopee dapat mengembangkan strategi retensi yang lebih tepat sasaran untuk mengurangi potensi kehilangan pelanggan dari kalangan Generasi Z.

Berdasarkan hasil evaluasi terhadap empat skenario pembagian data (split data 90:10, 80:20, 70:30, dan 60:40) pada tabel 1, dapat disimpulkan bahwa model Naïve Bayes memberikan performa terbaik pada skenario split 80:20 dengan akurasi tertinggi sebesar 73,75%, *recall* kelas YA sebesar 88,89%, *recall* kelas TIDAK sebesar 54,29%, serta *precision* masing-masing sebesar 71,43% dan 79,17%, yang menunjukkan keseimbangan prediksi antara kedua kelas; sementara itu, secara konsisten model lebih baik dalam mengenali kelas YA (minat beli) dibandingkan kelas TIDAK (tidak berminat beli), dan performa model cenderung menurun ketika porsi data uji diperbesar, sebagaimana terlihat pada skenario 60:40, sehingga dapat disimpulkan bahwa proporsi data latih sebesar 80% dan data uji 20% merupakan kombinasi optimal dalam klasifikasi minat beli Generasi Z di platform Shopee menggunakan algoritma Naïve Bayes.



Gambar 11. Faktor-Faktor yang Mempengaruhi Minat Beli

Pada Gambar 11 menggambarkan delapan faktor utama yang memengaruhi minat beli Generasi Z di Shopee, berdasarkan rata-rata penilaian responden menggunakan skala *Likert*. Faktor dengan pengaruh tertinggi adalah harga produk lebih penting dari kualitas dengan skor rata-rata 3,41, diikuti oleh kecepatan pengiriman produk sebesar 3,27. Kedua faktor ini menandakan bahwa Generasi Z sangat mempertimbangkan aspek harga dan efisiensi pengiriman dalam keputusan pembelian. Faktor lain seperti kemudahan pembayaran (3,12), ulasan konsumen (3,09), dan rekomendasi teman atau *influencer* (2,95) juga berkontribusi terhadap minat beli, meskipun dengan pengaruh yang sedikit lebih rendah. Sementara itu, faktor preferensi terhadap produk mahal menempati skor terendah yaitu 2,68, menunjukkan bahwa Generasi Z cenderung memilih produk yang terjangkau. Temuan ini menegaskan bahwa perilaku belanja Generasi Z didominasi oleh pertimbangan rasional, dengan fokus utama pada harga dan waktu pengiriman.

KESIMPULAN DAN SARAN

Berdasarkan hasil analisis yang dilakukan menggunakan algoritma *Naïve Bayes* melalui perangkat lunak RapidMiner terhadap 100 data responden Generasi Z pengguna Shopee, ditemukan bahwa terdapat beberapa faktor utama yang secara signifikan memengaruhi minat beli. Dari hasil *preprocessing* dan eksplorasi data, dua atribut dengan nilai rata-rata tertinggi pada skala *Likert* adalah pernyataan "Harga produk lebih penting dari kualitas" dengan nilai rata-rata sebesar 3,41, dan "Kecepatan pengiriman produk Shopee" dengan nilai rata-rata 3,27. Hal ini menunjukkan bahwa responden cenderung sangat mempertimbangkan aspek harga serta efisiensi pengiriman dalam menentukan keputusan pembelian mereka. Kedua variabel ini kemudian teridentifikasi sebagai faktor dominan setelah dilakukan proses klasifikasi dan evaluasi model, yang menghasilkan performa optimal pada rasio split data 80:20, dengan akurasi sebesar 73,75% dan *recall* tertinggi untuk kelas "YA" sebesar 88,89%. Temuan ini memperkuat bahwa minat beli Generasi Z lebih dipengaruhi oleh faktor-faktor praktis dan fungsional seperti harga yang kompetitif dan layanan logistik yang cepat. Di sisi lain, variabel-variabel lain seperti ulasan konsumen, kemudahan pembayaran, dan pengaruh sosial seperti rekomendasi dari teman atau *influencer*, juga berkontribusi dalam model tetapi tidak sebesar dua faktor utama tersebut. Dengan demikian, strategi pemasaran yang berfokus pada efisiensi harga dan pengiriman berpeluang besar dalam meningkatkan minat beli Generasi Z di *platform* Shopee.

DAFTAR PUSTAKA

- [1] Nofiwaty, Iisnawati, and A. Nazaruddin, "Consumer's Shopping Experience in Shopee," pp. 200–203, May 2020, doi: 10.2991/AEBMR.K.200520.034.
- [2] R. A. Sutarno and S. Purwanto, "Perilaku Konsumtif Generasi Z terhadap Keputusan Pembelian Produk Silverqueen di Kota Sidoarjo," *J-MAS (Jurnal Manajemen dan Sains)*, vol. 7, no. 1, pp. 309–313, Apr. 2022, doi: 10.33087/JMAS.V7I1.400.
- [3] M. Rosmiati, Fandhilah, and L. Zain, "Analisis Penerimaan Aplikasi Shopee Pada Generasi Milenial Dengan Technology Acceptance Model," *SATIN - Sains dan Teknologi Informasi*, vol. 8, no. 1, pp. 82–91, Jun. 2022, doi: 10.33372/STN.V8I1.832.
- [4] Syaparudin, Arizal, and R. Handayani, "Analisis Pengaruh Content Marketing, Influencer, dan Media Sosial Terhadap Keputusan Pembelian Konsumen," *Jurnal Manajemen dan Bisnis Terapan*, vol. 5, no. 2, pp. 78–83, Dec. 2023, doi: 10.31849/JMBT.V5I2.14231.
- [5] I. W. Dharmas, I. Putra, N. Mandenni, and Yusliza, "The Impact Of Iklanku Shopee Feature Through Brand Awareness On Purchase Decisions Using SEMPLS, Naive Bayes and K-NN Method," *Jurnal Ilmiah Merpati*

- (*Menara Penelitian Akademika Teknologi Informasi*), vol. 10, no. 1, pp. 69–80, Jan. 2022, doi: 10.24843/JIM.2022.V10.I01.P07.
- [6] A. Wibowo and A. Rohman, “Prediksi Predikat Kelulusan Mahasiswa Menggunakan Naive Bayes dan Decision Tree pada Universitas XYZ,” *Desember*, vol. 12, no. 2, Dec. 2022, doi: 10.36448/expert.v12i2.2810.
- [7] S. Setyaning, M. Febianah, F. Solikhah, A. Kamil, and W. Arifin, “Analisis Perbandingan Algoritma Naive Bayes dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan,” *TEMATIK - Jurnal Teknologi Informasi Dan Komunikasi*, vol. 8, no. 1, 2021.
- [8] O. Prasetia and S. Machfud, “Analisis Prediksi Penyakit Demensia Alzheimer Menggunakan Metode Decision Tree C4.5 Dan Naive Bayes,” *Jitu: Jurnal Informatika Utama*, vol. 2, no. 1, pp. 14–25, 2024, Accessed: Jul. 28, 2025. [Online]. Available: <https://jurnal.astinamandiri.com/index.php/jitu/article/view/162/118>
- [9] N. Anggraini, C. Yolandika, B. Utoyo, and L. Irawati, “Proses Pengambilan Keputusan Konsumen Dalam Pembelian Produk Lada Di Provinsi Lampung,” *Jurnal Agrisepe*, vol. 23, no. 1, pp. 43–51, Jul. 2022, doi: 10.17969/AGRISEP.V23I1.26152.
- [10] L. L. Hakim and K. Keni, “Pengaruh Brand Awareness, Brand Image dan Customer Perceived Value Terhadap Purchase Intention,” *Jurnal Manajemen Bisnis dan Kewirausahaan*, vol. 4, no. 3, pp. 81–86, May 2020, doi: 10.24912/JMBK.V4I3.7921.
- [11] F. M. Hapsari, S. Sudarwati, and F. S. Marwati, “Pengaruh Brand Trust, Media Sosial Dan *Online Consumer Review* Terhadap Minat Beli,” *JURNAL MANAJEMEN*, vol. 14, no. 1, pp. 91–97, Mar. 2022, doi: 10.30872/JMMN.V14I1.10747.
- [12] S. Gomes, J. M. Lopes, and S. Nogueira, “Willingness to pay more for green products: A critical challenge for Gen Z,” *J Clean Prod*, vol. 390, Mar. 2023, doi: 10.1016/J.JCLEPRO.2023.136092.
- [13] D. Fatrina, I. Kamil, and D. A. Hasan, “Pengaruh Live-Chat dan E-WOM Terhadap Pengambilan Keputusan Transaksi *Online* Pada Matketplace E-Commerce,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 3, pp. 121–129, Jan. 2019, doi: 10.25077/TEKNOSI.V5I3.2019.121-129.
- [14] T. Tu and N. Binh, “Impact of gamification on GenZ consumers’ *online* impulse buying behavior intention: evidence from Shopee application in the Vietnamese market,” *Proceedings The International Conference on Business Based on Digital Platform (BDP-3)*, pp. 321–334, 2023, Accessed: Jul. 28, 2025. [Online]. Available: <https://hdl.handle.net/10419/279147>
- [15] E. Asih, “Analisis pada Shopee sebagai *E-commerce* Terpopuler di Indonesia | Jurnal Ekonomi Bisnis Antartika,” *Jurnal Ekonomi Bisnis Antartika*, vol. 2, no. 1, 2024, Accessed: Jul. 28, 2025. [Online]. Available: <https://ejournal.mediaantartika.id/index.php/jeba/article/view/299>
- [16] M. I. P. Hant and H. Hendry, “Data Mining Technique Using Naïve Bayes Algorithm To Predict Shopee Consumer Satisfaction Among Millennial Generation,” *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 4, pp. 829–838, Aug. 2022, doi: 10.20884/1.JUTIF.2022.3.4.295.
- [17] F. Dulay, R. Liban, J. Rustia, M. Tayag, and M. Castaño, “Generation Y and Z Filipino Consumers’ Purchasing Behavior on Electronic Products in Shopee and Lazada,” *Journal of Business and Management Studies*, vol. 4, no. 1, pp. 294–305, Mar. 2022, doi: 10.32996/JBMS.2022.4.1.30.
- [18] S. G. Fitri, R. Selsi, Z. Rustam, and J. Pandelaki, “Naïve bayes classifier models for cerebral infarction classification,” *J Phys Conf Ser*, vol. 1490, no. 1, Mar. 2020, doi: 10.1088/1742-6596/1490/1/012019.
- [19] M. Haroon, Z. Alam, R. Kousar, J. Ahmad, and F. Nasim, “Sentiment Analysis of Customer Reviews on *E-commerce Platforms*: A Machine Learning Approach,” *Bulletin of Business and Economics (BBE)*, vol. 13, no. 3, pp. 230–238, Aug. 2024, doi: 10.61506/01.00480.
- [20] A. Yulianto and Firmansyah, “Prediksi Customer Churn Pada Bisnis Retail Menggunakan Algoritma Naïve Bayes,” *REMIK: Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 6, no. 1, pp. 41–47, Nov. 2021, doi: 10.33395/REMIK.V6I1.11196.
- [21] N. Noperahila and F. Sulianta, “Analisis Pola Perilaku Konsumen dan Kebiasaan Belanja menggunakan Metode K-Means Clustering,” Jul. 2024.