

IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI MINAT BELI MAKANAN SECARA ONLINE PADA GENERASI MILENIAL

Rani Febriani Zein, Suparni

Teknologi Informasi, Universitas Bina Sarana Informatika
Jl. Kramat No. 98, Jakarta Pusat
ranifebrnzein@gmail.com

ABSTRAK

Generasi milenial dikenal sebagai generasi yang sangat akrab dengan teknologi dan cenderung lebih nyaman berbelanja secara *online*. Namun, tingkat minat pembelian secara *online* di kalangan milenial, terutama yang berdomisili di Provinsi Jakarta tidak seragam dan bisa bervariasi tergantung pada berbagai aspek, seperti keputusan pembelian. Penelitian ini dimaksudkan untuk memprediksi daya minat beli makanan secara *online* pada generasi milenial dengan melihat keterkaitannya terhadap aspek keputusan pembelian, seperti frekuensi pembelian, kualitas produk, bentuk produk, harga produk, pengalaman merk, kemudahan mendapatkan produk serta cara pembayaran dengan algoritma naïve bayes menggunakan program RapidMiner. Prediksi ini dilakukan dengan membagi data menjadi kelas Tinggi, Sedang, Rendah dan Tidak Minat serta menerapkan teknik *resampling* dengan metode *downsampling* dikarenakan distribusi data yang akan digunakan untuk memprediksi kelas klasifikasi tidak seimbang. Dengan nilai *Area Under Curve* (AUC) sebesar 75%, model yang dihasilkan oleh algoritma naïve bayes dan diuji dengan *resampling* ini dinilai cukup baik untuk memprediksi minat beli makanan secara *online* pada generasi milenial. Berdasarkan hasil pengolahan dan analisa, diketahui terdapat adanya tingkat minat beli sedang dari generasi milenial yang lahir pada tahun 1992 hingga 1995 dan bertempat tinggal di Jakarta Timur.

Kata kunci: *Prediksi, Algoritma Naive Bayes, Minat Beli Makanan, Keputusan Pembelian*

1. PENDAHULUAN

Pada perkembangan era industri 5.0, dimana masyarakat Indonesia telah menjadi bagian integral dari kemajuan teknologi saat ini. Penggunaan teknologi yang maju membuat segala sesuatu menjadi lebih mudah. Peran teknologi semakin penting seiring berjalannya waktu karena membantu dan mempermudah manusia dalam melakukan banyak hal di kehidupan. Pengusaha harus memahami dan memanfaatkan perkembangan ini dalam kegiatan bisnis mereka untuk meningkatkan keunggulan mereka di dunia teknologi informasi yang berkembang pesat. Dengan kemajuan teknologi, banyak bisnis kuliner saat ini menggunakan proses penjualan secara *online* untuk meningkatkan kinerja penjualan. Setiap pelanggan pun memiliki preferensi minatnya sendiri dalam memesan makanan baik secara langsung di outlet maupun pemesanan *online*.

Pengalaman yang dialami pelanggan selama berbelanja *online* melalui pelayanan yang diberikan oleh penjual *online* dapat memengaruhi cara mereka melihat barang yang mereka beli [1]. Dengan adanya teknologi ini semakin memudahkan orang-orang untuk menjangkau kebutuhan pangannya serta memunculkan daya minat pembelian yang beragam, salah satunya bagi generasi milenial. Haroviz mendefinisikan, generasi Y atau yang dikenal sebagai generasi *millennial* adalah sekelompok anak muda yang lahir dari tahun 1980 hingga 2000an. Mereka cenderung nyaman dengan keberagaman teknologi dan *platform* komunikasi *online* untuk tetap terhubung dengan teman-temannya [2]. Berdasarkan pada penelitian sebelumnya yang sudah cukup beragam dengan studi serta metodenya dalam mengolah data

minat pembelian, adapun pemanfaatan data *mining* dapat juga diterapkan untuk memprediksi minat generasi milenial untuk membeli makanan secara *online* pada periode tahun 2024 ini. Data *mining* adalah disiplin ilmu yang digunakan untuk mengatasi tantangan pengambilan informasi dari basis data besar dengan mengintegrasikan teknik-teknik dari statistik, *machine learning*, visualisasi data, pengenalan pola, dan *database* [3].

Dengan permasalahan yang sudah penulis jelaskan di atas, penulis tertarik untuk melakukan penelitian terkait prediksi terhadap daya minat beli ini lebih dalam dan terfokuskan pada suatu generasi tertentu di wilayah Provinsi Jakarta dengan menilai dari aspek keputusan pembelian yang dapat menyebabkan adanya tingkat minat pembelian makanan secara *online* tersebut menggunakan algoritma naïve bayes. Kemudian, penulis juga mengukur keakuratan dan ketepatan model algoritma yang dihasilkan dalam memprediksi tingkat daya minat beli makanan *online* tersebut. Berdasarkan topik penelitian yang telah dibahas sebelumnya, penulis pun mengangkat judul “Implementasi Algoritma Naïve Bayes Untuk Memprediksi Minat Beli Makanan Secara *Online* Pada Generasi Milenial” untuk skripsi ini.

2. TINJAUAN PUSTAKA

Tinjauan pustaka adalah kumpulan temuan penelitian yang dilakukan oleh peneliti lain atau para ahli. Tinjauan pustaka ini meliputi identifikasi, penemuan, dan analisis dokumen yang berisi informasi tentang masalah penelitian.

2.1. Keputusan Pembelian

Keputusan pembelian merupakan proses konsumen mengidentifikasi masalahnya dan mencari informasi tentang produk atau merek tertentu [4]. Indikator-indikator utama yang mempengaruhi keputusan pembelian pelanggan, yaitu [5]:

- a. Keputusan Tentang Jenis Produk
Seringkali proses pengambilan keputusan membutuhkan banyak keputusan lain. Suatu keputusan adalah membuat pilihan antara dua atau lebih pilihan tindakan.
- b. Keputusan Tentang Bentuk Produk
Setelah konsumen memilih suatu produk, mereka akan melihat berbagai jenis produk yang ditawarkan oleh perusahaan dan memilih produk yang paling sesuai dengan kebutuhan dan karakteristik masing-masing.
- c. Keputusan Tentang Merek
Ketika konsumen yakin bahwa mereka telah mengetahui semua opsi yang tersedia, mereka cenderung tidak akan mencari opsi yang lainnya. Salah satunya faktor yang dapat mempengaruhi keputusan pelanggan adalah pengalaman pelanggan saat menggunakan merek tersebut, jumlah pembelian, kecenderungan pelanggan untuk mengingat atau mengaktifkan merek yang telah mereka gunakan sebelumnya
- d. Keputusan Tentang Penjualnya
Perusahaan tidak hanya dapat memikirkan dan membuat strategi untuk membuat pelanggan merasa senang sebelum mereka memutuskan untuk membeli produknya.
- e. Keputusan Tentang Waktu Pembelian
Mereka yang telah membuat keputusan untuk membeli sebuah produk tentunya akan mendorong pelanggan untuk menunggu cukup lama untuk membeli produk tersebut.
- f. Keputusan Tentang Cara Pembayaran
Faktor-faktor yang memengaruhi pemilihan produk oleh konsumen tidak hanya berhenti pada tahap penentuan dan alokasi waktu, tetapi juga mencakup kemudahan proses transaksi sebagai prioritas dalam memilih produk.

2.2. Minat Beli

Aspek-aspek dari minat beli dapat diidentifikasi sebagai berikut [6]:

- a. Minat Transaksional
Kebutuhan seseorang untuk membeli barang. Dalam hal ini, konsumen telah memiliki keinginan untuk membeli barang tertentu yang mereka inginkan.
- b. Minat Transaksional
Kecenderungan seseorang untuk merekomendasikan produk kepada orang lain. Dalam hal ini, seorang pembeli yang tertarik untuk membeli produk tersebut akan menyarankan orang-orang di sekitarnya untuk membelinya.

- c. Minat Preferensial

Menggambarkan bagaimana seseorang bertindak setelah memiliki preferensi utama terhadap suatu produk. Preferensi ini hanya dapat diubah jika terjadi sesuatu dengan produk tersebut.

- d. Minat Eksploratif

Perilaku seseorang yang selalu mencari tahu informasi tentang produk atau barang yang disukainya dan mendukung fitur positifnya.

2.3. Layanan Makanan Online

Menurut Akbar, layanan pesan antar bukan sesuatu yang baru, terutama bagi bisnis yang berkembang di sektor kuliner. Layanan pesan antar makanan juga harus berubah di era yang serba digital dan terhubung melalui internet. Aplikasi pesan antar makanan berbasis ponsel marak muncul dan menjadi tren baru di masyarakat [7]. Jogiyo mengatakan bahwa, "Layanan aplikasi pemesanan makanan *online* pertama kali diluncurkan pada April 2015 dengan nama GoJek, dan diikuti oleh aplikasi Grab pada Mei 2016. Beberapa restoran besar seperti Pizza Hut, McDonald, dan lainnya saat ini memberikan kesempatan kepada pelanggan untuk melakukan pemesanan makanan secara *online*. Pada layanan pesan makanan *online*, pelanggan dapat membayar dengan kartu debit atau kredit, sistem digital, atau pembayaran elektronik [8].

2.4. Generasi Milenial

Generasi milenial merupakan mereka yang lahir dan tumbuh di masa dimana teknologi informasi berkembang dengan sangat cepat. Kehidupan sehari-hari mereka sangat dipengaruhi oleh penggunaan dan dampak teknologi digital, dari saat bangun pagi hingga kembali tidur di malam hari [9]. Menurut beberapa pendapat para ahli dari berbagai negara dan profesi, dapat disimpulkan bahwa generasi milenial adalah kelompok individu yang lahir antara tahun 1980 dan 2000 [10].

2.5. Data Mining Prediksi

Teknik *forecasting* mengambil sekumpulan angka yang menunjukkan nilai-nilai yang berubah seiring waktu, kemudian dengan menggunakan berbagai teknik *machine learning* dan statistik untuk memprediksi nilai masa depan. Teknik ini mempertimbangkan faktor-faktor seperti musim, tren dan *noise* pada data [11]. Untuk teknik prediksi dan estimasi, ada banyak algoritma yang dapat digunakan, seperti regresi linear, jaringan saraf tiruan (*neural network*), mesin vektor pendukung (*support vector machine*), dan *k-nearest neighbor*. Algoritma yang paling cocok dipilih berdasarkan kebutuhan kasus yang dihadapi dan jenis data yang digunakan [12].

2.6. Algoritma Naïve Bayes

Metode probalistik sederhana *naive bayes* digunakan untuk menghitung probabilitas dengan menambahkan frekuensi dan kombinasi nilai dari *dataset* yang dimasukkan. Algoritma ini menerapkan

semua karakteristik yang independen atau tidak saling bergantung yang diberikan oleh nilai variabel kelas dengan menerapkan teorema Bayes [11].

Tahapan-tahapan pada algoritma naïve bayes adalah sebagai berikut [13]:

- Persiapkan data *set*.
- Hitung jumlah kelas dalam data *training*.
- Hitung jumlah kasus yang cocok untuk kelas yang sama.
- Kalikan hasil-hasil tersebut sesuai dengan data testing yang akan dicari kelasnya.
- Bandingkan hasil perkelas dan tentukan nilai tertinggi yang ditetapkan sebagai kelas baru.

2.7. Confusion Matrix

Confusion matrix digunakan sebagai pengukur kinerja. Pada dasarnya, pengukuran ini digunakan untuk memberikan informasi tentang perbandingan hasil klasifikasi yang dilakukan oleh algoritma dengan hasil sebenarnya [14]. Dalam penelitian ini, penulis menggunakan 4 kategori variabel untuk diprediksi maka dari itu tabel *confusion matrix* yang dihasilkan berukuran 4x4.

Dengan rumus untuk menghitung nilai *accuracy*, *precision*, *recall* (sensivitas), dan *specificity* dari tiap kelas (i) sebagai berikut:

$$a. \text{Accuracy} = \frac{TP_{all}}{Total} \times 100\% \quad [15]$$

$$b. \text{Precision} = \frac{TPP}{TPP_{all} + TFP_i} \quad [15]$$

$$c. \text{Recall} = \frac{TPP_{all}}{TPP_{all} + TFP_i} \quad [15]$$

$$d. \text{Specificity} = \frac{TN_{all}}{TN_{all} + TFP_i} \quad [15]$$

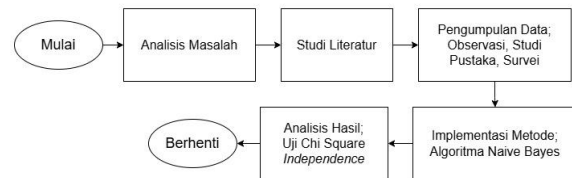
Dengan keterangan sebagai berikut:

- Total True Positive (TP) = Ini menunjukkan berapa banyak jumlah data kelas yang positif, dan juga menunjukkan bahwa model memprediksi positif.
- Total True Negative (TN) = Ini menunjukkan seberapa banyak jumlah data kelas yang sebenarnya negatif, dan model memprediksi negatif.
- Total False Positive (FP) = Ini menunjukkan seberapa banyak jumlah data kelas yang sebenarnya negatif, tetapi model memprediksi positif.
- Total False Negative (FN) = Ini menunjukkan seberapa banyak jumlah data kelas aktual yang positif, tetapi model memperkirakan bahwa itu negatif.

3. METODE PENELITIAN

3.1. Tahapan Penelitian

Penelitian ini dilakukan secara bertahap dengan melewati proses-proses berikut:



Gambar 1. Tahapan penelitian

- Analisis Masalah dan Studi Literatur**
Masalah terkait prediksi minat beli, khususnya bagi generasi milenial di Provinsi Jakarta dianalisis untuk menemukan solusi dan menetapkan cakupan masalah yang akan diselidiki. Penulis juga mempelajari dasar teori tentang penerapan algoritma naïve bayes, konsep dan teori data *mining*, serta prediksi minat beli generasi milenial melalui studi literatur dan penelitian terkait.
- Pengumpulan Data**
Dalam penelitian ini, data primer didapatkan melalui observasi aktivitas dan perilaku pembelian makanan generasi milenial serta penyebaran kuesioner berbentuk Google Form. Sementara itu, data sekunder didapatkan dari artikel literatur dan studi penelitian sebelumnya.
- Implementasi Metode**
Semua data yang telah dikumpulkan dari berbagai sumber diproses menggunakan metode algoritma data *mining*. Penulis menggunakan RapidMiner sebagai *tools* untuk menerapkan metode algoritma data *mining* ini. Sebelum mengimplementasi algoritma, penulis melakukan *preprocessing* dengan *cleansing* data (pembersihan data) agar dalam proses pengujian dapat menghasilkan data prediksi yang cukup akurat.
- Analisis Hasil**
Setelah pengujian dengan algoritma Naive Bayes di RapidMiner dan memperoleh hasil, data diorganisir oleh penulis berdasarkan pola, kategori, dan satuan dasar untuk menemukan tema dan merumuskan hipotesis kerja sesuai rekomendasi dari data tersebut. Penulis menguji hipotesa penelitian ini dengan menggunakan program SPSS 25 dan uji chi square untuk mengetahui apakah hasil pengolahan data cukup akurat untuk menentukan jawaban dari rumusan masalah sebelumnya.

3.2. Metode Pengumpulan Data

Beberapa metode yang digunakan penulis dalam proses pengumpulan data penelitian ini, antara lain:

- Observasi**
Observasi ini dilakukan untuk mengamati frekuensi pembelian di suatu restoran atau *outlet* dan penggunaan aplikasi pesan makanan di lingkungan sekitar penulis serta melakukan pencatatan terhadap jenis restoran atau *outlet* yang sering dikunjungi atau digunakan oleh subjek.
- Studi Pustaka**
Penulis mengumpulkan data serta informasi tambahan melalui penelitian, analisis literatur

terkait data *mining* dan algoritma naïve bayes serta referensi lainnya berhubungan dengan permasalahan yang penulis angkat dalam skripsi ini.

c. Survei

Penulis melakukan survei dengan menyebarkan kuesioner dalam bentuk Google Form kepada responden yang merupakan penduduk Jakarta yang lahir dari tahun 1980 hingga 2000, serta memiliki riwayat pembelian makanan melalui aplikasi. Pengumpulan data ini dilakukan menggunakan platform media sosial.

3.3. Populasi dan Sampel Penelitian

Penelitian ini mencakup populasi penduduk Provinsi Jakarta yang berkelahiran dari tahun 1980 hingga 2000 serta pernah atau aktif membeli makanan melalui aplikasi. Namun, karena jumlah populasi yang besar serta jumlah generasi milenial yang pernah atau aktif membeli makanan melalui aplikasi tidak diketahui secara pasti, maka penulis menggunakan rumus Cochran dan teknik *Purposive Sampling* untuk menentukan jumlah sampel. Berdasarkan rumus tersebut, sampel yang didapatkan berjumlah 96 dan kemudian dibulatkan menjadi 100 responden.

3.4. Uji Validitas

Uji validitas ini adalah proses untuk menilai seberapa baik alat pengukur yang digunakan dalam mengukur variabel yang dimaksud [16]. Salah satu kriteria pengujian ini adalah [17]:

- Ketika nilai $r_{hitung} > r_{tabel}$ (uji 2 sisi dengan sig. 0,05), maka instrumen atau item pertanyaan dinyatakan valid.
- Ketika $r_{hitung} < r_{tabel}$ (uji 2 sisi dengan sig. 0,05) atau nilai r_{hitung} negatif, maka instrumen atau item pertanyaan dinyatakan tidak valid.

3.5. Uji Reliabilitas

Dalam penelitian, reliabilitas merujuk pada sejauh mana hasil pengukuran tetap konsisten saat diulang terhadap subjek yang sama dan dalam kondisi yang serupa [16]. Metode *Cronbach's Alpha* adalah metode yang paling sering digunakan dalam penelitian. Nilai ini ditulis sebagai r_x , di mana x adalah index kasus yang dicari [17]:

Tabel 1. Kategori koefisien reliabilitas

Nilai Alpha Cronbach's	Keterangan
$0,80 < \alpha$	Reliabilitas sangat tinggi
$0,60 < \alpha$	Reliabilitas tinggi
$0,40 < \alpha$	Reliabilitas sedang
$0,20 < \alpha$	Reliabilitas rendah
$-1,00 < \alpha < 0,20$	Reliabilitas sangat rendah (tidak reliabel)

3.6. Data Preprocessing

Preprocessing akan dilakukan terlebih dahulu terhadap *dataset* yang telah dikumpulkan ini untuk menyiapkan data agar lebih terorganisir dan dapat

dianalisa dengan lebih efisien. Tahap *preprocessing* dilakukan dengan menggunakan program RapidMiner.

Adapun tahapan-tahapan dalam *preprocessing* pada penelitian ini adalah [12]:

a. Pembersihan Data

Tahap penting dalam proses analisis data adalah pembersihan data yang melibatkan identifikasi, penanganan, dan/atau penghapusan nilai yang hilang (*missing values*), pencilan (*outliers*), dan anomali dalam kumpulan data.

b. Transformasi Data

Transformasi ini dapat meningkatkan distribusi data, mengurangi heteroskedastisitas, atau mempermudah interpretasi. Transformasi data yang paling umum digunakan adalah normalisasi, transformasi, dan lainnya.

c. Standarisasi

Standarisasi adalah proses pembuatan dan penerapan standar untuk mencapai tujuan tertentu dalam bidang tertentu, seperti industri, teknologi, atau layanan.

3.7. Pembagian Data

Pada penelitian ini, data *training* telah dikelompokkan berdasarkan kelas yang akan menjadi target klasifikasi, yaitu kelas tidak minat, rendah, sedang, tinggi sebanyak 80 data atau 80% dari keseluruhan data kuesioner. Serta, penulis menggunakan 20 data atau 20% dari total keseluruhan data kuesioner.

3.8. Implementasi Algoritma Naïve Bayes

Dataset yang dinyatakan sudah siap akan diolah untuk memprediksi kelas objek pada program RapidMiner. Dalam proses ini, algoritma naïve bayes perlu diberikan pengetahuan awal untuk dijadikan acuan agar dapat melakukan klasifikasi terhadap masing-masing kelas, kemudian menghitung nilai probabilitas dari masing-masing kelasnya.

3.9. Pengukuran Performa Algoritma

Ada beberapa cara untuk mengukur performa prediksi dan cara yang paling umum digunakan adalah dengan menghitung nilai *accuracy*, *precision*, *recall*, dan *specificity* pada acuan *confusion matrix* serta nilai *Area Under the Curve* (AUC). Nilai AUC memberikan gambaran tentang seberapa baik model yang digunakan dalam memprediksi atau mengukur akurasi keseluruhan. Semakin tinggi nilai AUC, semakin baik kemampuan variabel yang sedang diteliti dalam memprediksi kejadian tertentu [18]. Perhitungan AUC ditunjukkan dalam persamaan berikut:

$$AUC = \frac{1}{2} (Sensivitas + Specifitas) \quad [19]$$

Nilai *Area Under the Curve* (AUC) dalam klasifikasi data *mining* sendiri dibagi menjadi beberapa kelompok, yaitu [20]:

Tabel 2. Kriteria nilai *area under the curve* (AUC)

Rentang Nilai AUC	Keterangan
0.90 – 1.00	Klasifikasi Sangat Baik
0.80 – 0.89	Klasifikasi Baik
0.70 – 0.79	Klasifikasi Cukup
0.60 – 0.69	Klasifikasi Buruk
< 0.60	Klasifikasi Gagal

Setelah mengolah data *mining* menggunakan model klasifikasi, *confusion matrix* digunakan sebagai pengukur kinerja.

3.10. Uji Hipotesis

Untuk hipotesis dalam penelitian ini menggunakan teknik uji komparatif yang merumuskan tentang dugaan ada tidaknya keterkaitan antara dua variabel dengan komparasi 2 (dua) kelompok sampel. Penelitian ini menggunakan 100 data responden kuesioner berskala ordinal. Oleh karena itu, disini penulis akan menerapkan uji chi square dengan metode uji *Chi Square of Independence* untuk menguji hipotesis yang sebelumnya telah dirumuskan menggunakan program SPSS 25.

Untuk mengetahui taraf signifikansi hasil analisis dapat dihitung dengan persamaan seperti dibawah ini:

$$d_f = (k - 1)(m - 1) \quad [21]$$

Dimana:

k = jumlah kategori

m = jumlah golongan

Pedoman kriteria pengukuran yang dipakai untuk memutuskan akan menerima atau menolak jika hipotesis nol (H_0) yang diusulkan sebagai berikut [21]:

- Nilai *p-value* pada kolom asymp. sig (2-tailed) > 0,05 (α) dan nilai chi square < chi square tabel (d_f) maka H_0 diterima.
- Nilai *p-value* pada kolom asymp. sig (2-tailed) < 0,05 (α) dan nilai chi square > chi square tabel (d_f) maka H_0 ditolak.

4. HASIL DAN PEMBAHASAN

4.1. Deskripsi Data

Data yang diolah dan dianalisa oleh penulis ini terkumpul melalui kuesioner Google Form dengan total sebanyak 100 responden. Data yang terkumpul melalui penyebaran kuesioner ini terdiri dari 28 atribut yang dimana 5 atribut berisikan data identitas responden, 19 atribut berisikan variabel yang mencakup keputusan pembelian, 3 atribut berisikan variabel yang mencakup minat beli serta 1 atribut berisikan jenis aplikasi pesan makanan yang biasa digunakan oleh responden.

4.2. Transformasi Data

Disini penulis melakukan *attribute construction* dengan menentukan kategori dari skor variabel keputusan pembelian yang diberikan oleh masing-masing responden menjadi 4 kelompok yaitu Sangat

Tidak Setuju, Tidak Setuju, Setuju, Sangat Setuju, serta variabel minat beli menjadi 4 kelompok yang terdiri dari tingkat minat Rendah, Sedang, Tinggi, Sangat Tinggi. Dengan hanya menggunakan 25 atribut atau variabel dari keseluruhan data, yaitu atribut tahun lahir, domisili, frekuensi pembelian, kualitas produk, bentuk produk, harga produk, pengalaman merek, kemudahan mendapatkan produk, cara pembayaran, minat beli serta aplikasi.

Dalam penentuan kategorinya, penulis menghitung rentang skala dari masing-masing skor penilaian responden terhadap variabel keputusan pembelian dan minat beli dengan rumus interval sebagai berikut ini:

$$P = \frac{\text{rentang}}{\text{banyaknya kelas}} \quad [22]$$

Keterangan:

P = panjang kelas interval

Rentang = data tertinggi – data terendah

Berdasarkan rumus tersebut maka perhitungan rentang skala untuk menentukan kategori dari setiap variabel di penelitian ini, yaitu:

$$p = \frac{4-1}{4} = \frac{3}{4} = 0,75$$

Dengan demikian jarak antar rentang untuk masing-masing kategori keputusan pembelian adalah 0,75, sehingga kategori secara keseluruhannya sebagai berikut.

Tabel 3. Rentang skala keputusan pembelian

Rentang Skala	Kategori
1,00 – 1,75	Sangat Tidak Setuju
1,75 – 2,5	Tidak Setuju
2,5 – 3,25	Setuju
3,25 – 4,00	Sangat Setuju

Sementara jarak antar rentang untuk masing-masing kategori minat beli adalah 0,75 dengan penjelasan sebagai berikut:

Tabel 4. Rentang skala minat beli

Rentang Skala	Kategori
1,00 – 1,75	Tidak Minat
1,75 – 2,5	Rendah
2,5 – 3,25	Sedang
3,25 – 4,00	Tinggi

Selain itu nama atribut Kualitas produk diinisialkan menjadi “Kualitas” atau (KU), atribut Bentuk Produk menjadi “Bentuk” atau (BE), atribut Harga Produk menjadi “Harga” atau (HA), atribut Pengalaman Merek menjadi “Merek” atau (ME), atribut Kemudahan Mendapatkan Produk menjadi “Mudah” atau (MU), atribut Cara Pembayaran menjadi “Pay” atau (PE) serta atribut Minat Beli menjadi “Minat” atau (MI).

Tabel 5. Transformasi data

No.	Kualitas (KU)	Bentuk (BE)	Harga (HA)	Merk (ME)	Mudah (MU)	Pay (HA)	Minat (MI)	Aplikasi
1	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Tinggi	Shopeefood
2	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Rendah	GoFood
3	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Setuju	Sangat Setuju	Sedang	ShopeeFood
4	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Rendah	GrabFood
5	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Tidak Setuju	Sangat Setuju	Tinggi	ShopeeFood
...	...	...	...	...	...	...	...	...
95	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Tinggi	GoFood
96	Setuju	Setuju	Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Tinggi	GoFood
97	Sangat Setuju	Sangat Setuju	Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sedang	GrabFood
98	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Setuju	Sedang	GoFood
99	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Setuju	Sangat Setuju	Sedang	GoFood
100	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Tinggi	GrabFood

4.3. Pengujian Dengan Algoritma Naïve Bayes

Berdasarkan pengujian pertama yang penulis lakukan, model yang dihasilkan oleh algoritma naïve bayes tidak bekerja secara maksimal untuk memprediksi minat beli makanan secara *online* pada generasi milenial. Terdapat adanya perbedaan tingkatan nilai seperti pada Gambar 2.

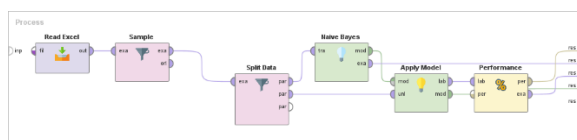
accuracy: 75.00%

	true Tinggi	true Rendah	true Sedang	true Tidak Minat	class precision
pred. Tinggi	9	1	1	0	81.82%
pred. Rendah	0	0	0	0	0.00%
pred. Sedang	2	1	6	0	66.67%
pred. Tidak Minat	0	0	0	0	0.00%
class recall	81.82%	0.00%	85.71%	0.00%	

Gambar 2. Performance algoritma naïve bayes

Hal ini bisa disebabkan karena *imbalance data* atau ketidakseimbangan data. *Dataset* yang digunakan memiliki distribusi data yang tidak seimbang untuk klasifikasi kelas minat, di mana sebagian besar nilai berada di kelas sedang dan tinggi, sehingga algoritma hanya memprediksi probabilitas dari kelas mayoritas saja.

Untuk menyeimbangkan label data pada masing-masing kelas, penulis akan melakukan *resampling* dengan metode *downsampling*. Metode *downsampling* mengurangi jumlah kelas data mayoritas sehingga kedua kelompok klasifikasi memiliki jumlah data yang seimbang. Berikut dibawah ini adalah gambaran pengujian ulang *resampling* dengan menambahkan operator “Sample” pada RapidMiner.



Gambar 3. Resampling data

4.4. Evaluasi dan Validasi Hasil Analisis

Berdasarkan pengujian *resampling* yang telah dilakukan maka didapatkan hasil evaluasi kinerja algoritma seperti *confusion matrix* dibawah ini.

accuracy: 75.00%

	true Tinggi	true Rendah	true Sedang	true Tidak Minat	class precision
pred. Tinggi	3	0	1	0	75.00%
pred. Rendah	1	2	0	0	66.67%
pred. Sedang	1	0	4	0	80.00%
pred. Tidak Minat	0	0	0	0	0.00%
class recall	60.00%	100.00%	80.00%	0.00%	

Gambar 4. Performance algoritma naïve bayes dengan *resampling*

Model algoritma naïve bayes dinilai efektif dikarenakan berhasil mengidentifikasi kasus nilai kelas negatif aktual secara tepat pada masing-masing kelas, seperti yang ditunjukkan oleh gambaran *confusion matrix* diatas, bersama dengan hasil perhitungan *specificity* yang dihasilkan oleh kelas tinggi sebesar 75%, kelas sedang 85%, kelas rendah 10% dan kelas tidak minat 1%.

Sementara itu, nilai AUC yang didapat untuk kinerja algoritma naïve bayes terhadap prediksi minat beli makanan *online* pada generasi milenial yaitu:

$$AUC = \frac{1}{2} (Sensivitas + Specifitas)$$

$$= \frac{0,6 + 0,9}{2} = 0,75$$

Ini menunjukkan bahwa kinerja algoritma naïve bayes cukup baik dalam memprediksi minat beli makanan *online* pada generasi milenial untuk kondisi target variabel *multi class* dengan sampel acak.

Berdasarkan plot diagram, prediksi yang dihasilkan oleh algoritma naïve bayes terhadap masing-masing klasifikasi kelas dan peminatan pemakaian aplikasi adalah sebagai berikut:

Tabel 6. Hasil prediksi model

Tingkat Minat Beli	Tahun Lahir	Domisili	Nilai Confidence	Aplikasi
Tinggi	1988 – 1991	Jakarta Pusat	0,991	Shopee Food
Sedang	1992 – 1995	Jakarta Timur	0,999	GoFood
Rendah	1996 – 2000	Kepulauan Seribu	0,644	Happy Fresh
Tidak Minat	1996 - 2000	Jakarta Selatan	0,00	-

Model memprediksi adanya tingkat minat beli sedang dari generasi milenial yang berkelahiran dari tahun 1992 hingga 1995 dan bertempat tinggal di Jakarta Timur dengan aplikasi GoFood sebagai aplikasi pesan makanan yang biasa digunakan oleh generasi milenial dalam membeli makanan secara *online*.

4.5. Tahap Analisis Data

Pada tahap analisis data ini menggunakan teknik uji komparatif untuk menguji hipotesis yang telah dirumuskan sebelumnya. Penulis menerapkan uji hipotesis chi square dengan metode uji *Independency* yang diujikan kepada variabel minat dan keputusan pembelian. Penulis menetapkan tingkat kepercayaan sebesar 95% ($\alpha = 0,05$).

Untuk menganalisis data, variabel keputusan pembelian seperti kualitas, bentuk, harga, merek, mudah, pay dikelompokkan menjadi kategori ordinal, yaitu:

- 1 = “Tidak Puas” dan 2 = “Puas”
- Variabel frekuensi dikategorikan menjadi 1 = “1 – 3 kali”, 2 = “4 – 5 kali” dan 3 = “Lebih dari 5”
- Variabel minat beli yang menjadi acuan yaitu minat dikelompokkan menjadi kategori 1 = “Tidak Minat”, 2 = “Rendah”, 3 = “Sedang” dan 4 = “Tinggi”

Hipotesis nol (H_0) yang diusulkan adalah terdapat daya minat beli dari generasi milenial untuk membeli makanan secara *online*.

Tabel 7. Hasil uji chi square terhadap variabel keputusan pembelian

Variabel	Nilai Chi Square	Nilai Tabel Chi Square	Nilai p-value	Nilai α	Keterangan
Frekuensi	3,008	12,591	0,808	0,05	H0 diterima
Kualitas	5,526	7,814	0,137	0,05	H0 diterima
Bentuk	7,842	7,814	0,049	0,05	H1 diterima
Harga	7,845	7,814	0,049	0,05	H1 diterima
Merk	3,453	7,814	0,327	0,05	H0 diterima
Mudah	2,719	7,814	0,437	0,05	H0 diterima
Pay	5,040	7,814	0,169	0,05	H0 diterima

Dapat disimpulkan bahwa terdapat hubungan antara daya minat beli dengan keputusan pembelian makanan secara *online* pada generasi milenial. Aspek-aspek keputusan pembelian yang saling berhubungan dengan minat beli tersebut adalah frekuensi pembelian, kualitas produk, pengalaman merek, kemudahan mendapatkan produk dan cara pembayaran, pengalaman merek, kemudahan mendapatkan produk dan cara pembayaran.

5. KESIMPULAN DAN SARAN

Dengan melakukan teknik *resampling*; metode *downsampling* menggunakan program RapidMiner, model algoritma naive bayes meraih nilai *Area Under Curve* (AUC) sebesar 75%. Ini menunjukkan bahwa kinerja dari model cukup baik untuk memprediksi kelas positif dan negatif dari masalah tingkat minat beli makanan *online* pada generasi milenial dengan tepat dan akurat. Dengan tingkat akurasi probabilitas sebesar 0,999, model memprediksi bahwa generasi milenial yang lahir dari tahun 1992 hingga 1995 dan tinggal di Jakarta Timur memiliki tingkat minat beli sedang serta menggunakan aplikasi Gofood untuk membeli makanan secara *online*. Adanya minat beli makanan secara *online* pada generasi milenial ini juga berhubungan dengan aspek keputusan pembelian terkait variabel frekuensi pembelian, kualitas produk, pengalaman merek makanan, kemudahan mendapatkan dan cara pembayaran. Berdasarkan hasil penelitian, penulis menyarankan untuk penelitian selanjutnya dapat melakukan pengukuran performa algoritma prediksi dengan teknik yang berbeda serta diharapkan dapat menerapkan algoritma lain yang dapat menghasilkan model prediksi lebih akurat terkait permasalahan minat beli seperti algoritma *decision tree*, algoritma *random forest*, algoritma *neural network* ataupun membandingkan beberapa algoritma sehingga dapat mengembangkan penelitian-penelitian di masa yang akan datang.

DAFTAR PUSTAKA

- [1] Taswiyah and A. Imron, “Persepsi dan Preferensi Ibu Rumah Tangga Terhadap Sistem Berbelanja Secara Online/online shop,” *Jurnal Abdimas BSI: Jurnal Pengabdian Kepada Masyarakat*, vol. 4, no. 1, pp. 128–137, Feb. 2021, doi: 10.31294/JABDIMAS.V4I1.9169.
- [2] Erdawati, Endarwita, and R. Widiyari, “PENGARUH ELECTRONIC WORD OF MOUTH (e-WOM) DAN KUALITAS PRODUK TERHADAP KEPUTUSAN PEMBELIAN SKINCARE,” *Journal of Social and Economics Research*, vol. 5, no. 1, pp. 184–190, Jul. 2023, doi: 10.54783/JSER.V5I1.82.
- [3] I. Werdiningsih, B. Nuqoba, and Muhammadun, *DATA MINING MENGGUNAKAN ANDROID, WEKA, DAN SPSS*. Airlangga University Press, 2020.
- [4] D. Gunawan, *Keputusan Pembelian Konsumen Marketplace Shopee Berbasis Social Media Marketing*, 1st ed. PT Inovasi Pratama Internasional, 2022.
- [5] D. K. Sawlani, *Keputusan Pembelian Online: Kualitas Website, Keamanan dan Kepercayaan*, 1st ed. Surabaya: Scopindo Media Pustaka, 2021.

- [6] S. Seftila, O. Saryono, F. Haris, and E. Prabowo, "KEPERCAYAAN DAN PENGALAMAN BERBELANJA SERTA PENGARUHNYA PADA MINAT BELI KONSUMEN SECARA ONLINE DALAM APLIKASI E-COMMERCE," *Value : Jurnal Manajemen dan Akuntansi*, vol. 16, no. 2, pp. 499–511, Dec. 2021, doi: 10.32534/JV.V16I2.2018.
- [7] H. N. Az-zahra, V. A. Tanya, and N. C. Apsari, "LAYANAN ONLINE FOOD DELIVERY DALAM MEMBANTU MENINGKATKAN PENJUALAN PADA USAHA MIKRO," *Jurnal Penelitian dan Pengabdian Kepada Masyarakat (JPPM)*, vol. 2, no. 2, pp. 156–165, Aug. 2021, doi: 10.24198/JPPM.V2I2.33513.
- [8] V. Vincentius and L. S. Putranto, "ANALISIS PENGARUH LAYANAN PESAN MAKANAN ONLINE TERHADAP PERJALANAN BERBASIS RUMAH DAN BERBASIS TEMPAT AKTIVITAS," *JMTS: Jurnal Mitra Teknik Sipil*, vol. 3, no. 4, pp. 1003–1016, Nov. 2020, doi: 10.24912/JMTS.V3I4.8391.
- [9] Wasiyem, H. Purba, and M. K. Karima, *Pendidikan Kewarganegaraan Untuk Perguruan Tinggi (Upaya Mewujudkan Good Citizenship)*, 1st ed., vol. 1. April: CV Merdeka Kreasi Group, 2021.
- [10] M. Arif, *Generasi Millenial Dalam Internalisasi Karakter Nusantara*, 1st ed. Kediri: IAIN Kediri Press, 2021.
- [11] Mustika *et al.*, *Data Mining dan Aplikasinya*, 1st ed. Bandung: Widina Bhakti Persada Bandung, 2021.
- [12] P. W. Rahayu *et al.*, *Buku Ajar Data Mining*, 1st ed. Jambi: PT Sonpedia Publishing Indonesia, 2024.
- [13] N. L. W. S. R. Ginantra *et al.*, *Data Mining dan Penerapan Algoritma*, 1st ed. Bali: Yayasan Kita Menulis, 2021.
- [14] P. Romadloni, B. A. Kusuma, and W. Maulana Baihaqi, "KOMPARASI METODE PEMBELAJARAN MESIN UNTUK IMPLEMENTASI PENGAMBILAN KEPUTUSAN DALAM MENENTUKAN PROMOSI JABATAN KARYAWAN," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 622–628, Sep. 2022, doi: 10.36040/JATI.V6I2.5238.
- [15] L. Anggraini Susanti, A. M. Soleh, B. Sartono, I. Pertanian Bogor, and P. Korespondensi, "Deep Learning Image Classification Rontgen Dada pada Kasus Covid-19 Menggunakan Algoritma Convolutional Neural Network," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 5, pp. 973–982, Oct. 2023, doi: 10.25126/JTIK.20231057142.
- [16] M. M. Sanaky, "ANALISIS FAKTOR-FAKTOR KETERLAMBATAN PADA PROYEK PEMBANGUNAN GEDUNG ASRAMA MAN 1 TULEHU MALUKU TENGAH," *JURNAL SIMETRIK*, vol. 11, no. 1, pp. 432–439, Aug. 2021, doi: 10.31959/JS.V11I1.615.
- [17] R. Slamet and S. Wahyuningsih, "VALIDITAS DAN RELIABILITAS TERHADAP INSTRUMEN KEPUASAN KERJA," *Aliansi : Jurnal Manajemen dan Bisnis*, vol. 17, no. 2, Dec. 2022, doi: 10.46975/ALIANSI.V17I2.428.
- [18] V. R. Sari, F. Firdausi, and Y. Azhar, "Perbandingan Prediksi Kualitas Kopi Arabika dengan Menggunakan Algoritma SGD, Random Forest dan Naive Bayes," *Edumatic: Jurnal Pendidikan Informatika*, vol. 4, no. 2, pp. 1–9, Dec. 2020, doi: 10.29408/EDUMATIC.V4I2.2202.
- [19] Z. Annisa and B. S. S. Ulama, "Analisis Sentimen Data Ulasan Pengguna Aplikasi 'PeduliLindungi' pada Google Play Store Menggunakan Metode Naïve Bayes Classifier Model Multinomial," *Jurnal Sains dan Seni ITS*, vol. 11, no. 6, pp. D464–D471, May 2023, doi: 10.12962/j23373520.v11i6.94064.
- [20] G. Budiyanto and D. Suryadi, "A Machine Learning Prediction Model to a Scholarship Program for New Undergraduate Students at a Private University," *Jurnal Rekayasa Sistem Industri*, vol. 12, no. 2, pp. 187–200, Oct. 2023, doi: 10.26593/JRSI.V12I2.6595.187-200.
- [21] M. Naek Pane, Mardiaman, and M. Azhar, "PENENTUAN FAKTOR KETERLAMBATAN WAKTU PELAKSANAAN PROYEK APBN/APBD DENGAN ANGGARAN TAHUN TUNGGAL UNTUK MENGURANGI KLAIM," *Menara: Jurnal Teknik Sipil*, vol. 18, no. 2, pp. 96–109, Jun. 2023, doi: 10.21009/JMENARA.V18I2.35422.
- [22] D. Sunardi, "Pengaruh Gaya Kepemimpinan dan Jabatan Fungsional Perekayasa Terhadap Kinerja Pegawai Balai Besar Aerodinamika, Aeroelastika, dan Aeroakustika di Kawasan Puspiptek Setu, Tangerang Selatan," *Jurnal Madani: Ilmu Pengetahuan, Teknologi, dan Humaniora*, vol. 3, no. 1, pp. 151–163, Mar. 2020, doi: 10.33753/MADANI.V3I1.105