

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 lebih mudah. Peran teknologi semakin penting seiring berjalannya waktu karena membantu dan mempermudah manusia dalam melakukan banyak hal dalam 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]. Keputusan pembelian pelanggan adalah etika seorang pembeli menemukan masalah mereka, mereka mencari informasi tentang produk atau merek tertentu, dan menilai seberapa baik alternatif yang dapat menyelesaikan masalah mereka sebelum membuat keputusan pembelian [2]. 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 [3]. 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* [4].

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 ini terkait dengan apa yang dibeli, kapan dibeli, mengapa dibeli, bagaimana pembelian dilakukan, dimana pembeliannya dan siapa yang melakukan pembelian. Pemasar mempelajari keputusan pembelian konsumen untuk mengetahui kebutuhan dan keinginan pelanggan dalam waktu, jumlah, dan harga yang tepat [5]. Keputusan pembelian merupakan proses konsumen mengidentifikasi masalahnya dan mencari informasi tentang produk atau merek tertentu [6]. Sebelum tahap perilaku pasca pembelian, salah satu langkah dalam proses keputusan pembelian produk adalah pengambilan keputusan untuk membeli. Pada tahap ini, pelanggan dihadapkan pada beberapa pilihan dan mereka kemudian melakukan tindakan untuk memilih produk mana yang akan mereka beli [7].

Indikator-indikator utama yang mempengaruhi keputusan pembelian pelanggan, yaitu [8]:

1. Keputusan Tentang Jenis Produk

Seringkali proses pengambilan keputusan membutuhkan banyak keputusan lain. Suatu keputusan adalah membuat pilihan antara dua atau lebih pilihan tindakan. Dalam hal ini, konsumen akan memilih produk yang iklannya dapat meyakinkan mereka untuk membeli.

2. 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.

3. Keputusan Tentang Merek

Ketika konsumen yakin bahwa mereka telah mengetahui semua opsi yang tersedia, mereka cenderung tidak akan mencari opsi yang lainnya. Ada sejumlah faktor yang dapat memengaruhi keputusan pelanggan untuk menggunakan merek tertentu. Salah satunya adalah pengalaman pelanggan saat menggunakan merek tersebut, jumlah pembelian, kecenderungan pelanggan untuk mengingat atau mengaktifkan merek yang telah mereka gunakan sebelumnya

4. Keputusan Tentang Penjualnya

Perusahaan tidak hanya dapat memikirkan dan membuat strategi untuk membuat pelanggan merasa senang sebelum mereka memutuskan untuk membeli produknya. Distribusi yang mudah bagi pelanggan untuk mengakses produk perusahaan juga menjadi pertimbangan utama bagi pelanggan.

5. 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.

6. 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

Menurut Assael, minat beli merujuk pada keinginan pelanggan untuk membeli produk tertentu serta jumlah unit produk yang diinginkan dalam periode waktu tertentu. Dengan ini, minat beli mencerminkan niat mental pelanggan untuk membeli sejumlah barang dengan merek tertentu [9].

Aspek-aspek dari minat beli dapat diidentifikasi sebagai berikut [10]:

1. Minat Transaksional

Kebutuhan seseorang untuk membeli barang. Dalam hal ini, konsumen telah memiliki keinginan untuk membeli barang tertentu yang mereka inginkan.

2. 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.

3. Minat Preferensial

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

4. 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 biasanya melibatkan pemesanan telepon. Jasa ini kurang digunakan karena harga pulsa yang mahal. 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 [11]. Jogiyanto 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*. Proses pemesanan makanan *online* mencakup serangkaian langkah termasuk memilih restoran, memilih menu, melakukan konfirmasi pemesanan, menerima slip pesanan, dan pengaturan pengiriman. Pada layanan pesan makanan *online*, pelanggan dapat membayar

dengan kartu debit atau kredit, sistem digital, atau pembayaran elektronik [12].

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 [13]. Generasi milenial dikenal sebagai generasi Y yang lahir setelah generasi X dan juga dikenal sebagai *The Me Me Me Generation* atau *The Net Generation*, didefinisikan antara lain dengan berbagai faktor, termasuk usia, gaya hidup dan kepribadian yang berpusat pada diri sendiri [14]. Dengan mempertimbangkan berbagai pendapat dari para ahli yang membahas definisi generasi tersebut, kesimpulannya adalah mereka yang dilahirkan antara tahun 1980 dan 2000 dikelompokkan sebagai generasi milenial. Di Indonesia, generasi milenial mencakup orang-orang yang lahir dalam rentang waktu yang sama, yaitu antara tahun 1980 hingga 2000. Namun, peran generasi sebelumnya tidak terlepas pada pembentukan Generasi Y. Sebaliknya, Generasi Y terbentuk dari kemajuan dan perkembangan ilmu teknologi dan informasi [15]. 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 [16].

2.5. Data Mining Prediksi

Forecasting atau prediksi merupakan proses memproyeksikan nilai atau perilaku di masa depan berdasarkan pola dan tren yang ditemukan dalam data historis. 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 [17]. Metode prediktif berguna untuk membangun dan mengembangkan model pengetahuan yang dapat digunakan untuk membuat prediksi. Variabel yang tidak diketahui jenisnya dapat diprediksi dengan menggunakan pola yang dibuat dari data yang diketahui dari variabel yang terdapat pada data. Klasifikasi dan regresi adalah dua teknik yang sering digunakan dalam metode prediktif [18]. 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 [19].

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 [17]. Berdasarkan teorema Bayes atau aturan Bayes dengan asumsi independensi (ketidaktergantungan) yang kuat (naïve), metode prediksi Bayes adalah berbasis probalistik sederhana. Dalam kasus naïve bayes, independensi yang kuat pada fitur berarti keberadaan atau tidaknya fitur lain dalam data yang sama tidak berpengaruh terhadap fitur tersebut [20].

Teorema Bayes adalah dasar dari prediksi Bayes yang memiliki formula umum dengan persamaan berikut [20]:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

Penjelasan formula diatas adalah:

$P(H|E)$: Probabilitas bebas bersyarat (*conditional probability*) suatu hipotesis H terjadi jika bukti E (*evidence*) diberikan.

$P(E|H)$: Hipotesis H akan dipengaruhi oleh kemungkinan bahwa bukti E akan terjadi.

$P(H)$: Probabilitas awal (priori) hipotesis H terjadi tanpa mempertimbangkan bukti apapun.

$P(E)$: Probabilitas awal (apriori) bukti E terjadi tanpa mempertimbangkan hipotesis atau bukti lain

Teori dasar aturan Bayes adalah bahwa hasil hipotesis peristiwa (H) dapat diprediksi dari bukti (E) tertentu. Aturan Bayes mencakup beberapa hal penting, antara lain:

1. Probabilitas hipotesis sebelum adanya bukti adalah sebuah probabilitas awal/prior H atau $P(H)$.
2. Probabilitas akhir H atau $P(H|E)$ adalah sebuah probabilitas akhir dari suatu hipotesis setelah bukti diamati.

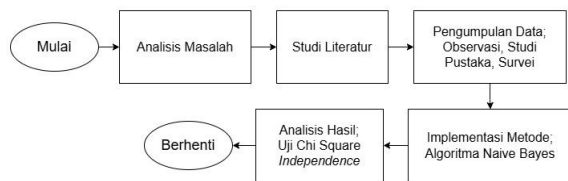
Tahapan-tahapan pada algoritma naïve bayes adalah sebagai berikut [18]:

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

3. METODE PENELITIAN

3.1. Tahapan Penelitian

Tahapan penelitian adalah prosedur ilmiah untuk mengumpulkan data dan informasi yang digunakan untuk kebutuhan penelitian. Penelitian ini dilakukan secara bertahap dengan melewati proses-proses berikut:



Gambar 1. Tahapan penelitian

1. Analisis Masalah dan Studi Literatur

Disini penulis meneliti isu-isu yang terkait dengan prediksi minat beli, terutama bagi generasi milenial di Provinsi Jakarta. Kemudian, masalah yang terjadi 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.

2. Pengumpulan Data

Penulis melakukan prosedur sistematis untuk mengumpulkan data penelitian yang menunjukkan daya minat beli dengan pendekatan penelitian kuantitatif. Subyek dalam skripsi ini ialah penduduk Provinsi Jakarta lahir dari tahun 1980 hingga 2000 serta pernah atau aktif membeli makanan melalui aplikasi. Dalam penelitian ini, data primer didapatkan melalui observasi aktivitas dan perilaku pembelian makanan generasi milenial serta penyebaran kuesioner berbentuk Google Form yang ditujukan kepada responden terkait. Sementara itu, data sekunder didapatkan dari artikel literatur dan studi penelitian sebelumnya.

3. 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) untuk menghilangkan dari data-data yang tidak valid, duplikasi, anomali, seleksi data dan transform data agar dalam proses pengujian dapat menghasilkan data prediksi yang cukup akurat. Kemudian, dilakukan pengujian untuk menentukan sejauh mana penelitian telah memenuhi tujuan yang diharapkan.

4. 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. Untuk menguji hipotesa penelitian ini penulis menggunakan program SPSS 25 dan selanjutnya, mengukur 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:

1. 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.

2. 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.

3. 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 seperti WhatsApp grup dan Twitter (X). Kuesioner yang disebarkan ini mencakup pernyataan terkait aspek-aspek keputusan pembelian seperti frekuensi pembelian, kualitas produk, bentuk produk, harga produk, pengalaman merek, kemudahan mendapatkan produk, cara pembayaran, aspek minat beli serta penulis menambahkan parameter pernyataan jenis aplikasi pesan makanan yang biasa digunakan. Selain itu, responden mengisi pernyataan kuesioner berdasarkan skala Likert serta memberikan tanggapan pribadi sesuai dengan penilaian masing-masing terhadap faktor keputusan mereka untuk membeli makanan secara *online*.

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. Metode Analisis Data

3.5.1. Uji Validitas

Validitas penelitian menggambarkan seberapa akurat alat pengukur penelitian terhadap informasi yang sesungguhnya diukur. Uji validitas ini adalah proses untuk menilai seberapa baik alat pengukur yang digunakan dalam mengukur variabel yang dimaksud. Sebuah kuesioner dianggap valid hanya jika dapat menjelaskan tujuan kuesioner [21]. Beberapa alat uji validitas menggunakan program SPSS [22]:

1. Korelasi Person (*Bivariate Pearson*) atau disebut *Product Moment*

Semua jawaban responden untuk masing-masing pertanyaan disebut sebagai skor total. Kriteria pengujian ialah sebagai berikut:

- a. Instrumen atau item pertanyaan dinyatakan valid jika terdapat korelasi yang signifikan dengan skor total, yaitu ketika nilai $r_{hitung} > r_{tabel}$ (uji 2 sisi dengan sig. 0,05).
- b. Instrumen atau item pertanyaan dinyatakan tidak valid jika tidak ada korelasi signifikan dengan skor total, yaitu ketika $r_{hitung} < r_{tabel}$ (uji 2 sisi dengan sig. 0,05) atau nilai r_{hitung} negatif.

2. Corrected Item to Total Correlation

Banyak orang menganggap metode ini lebih akurat dalam mengukur validitas daripada *Product Moment* karena metode ini hanya mengurangi efek *Overlaps Spurious*. Konsepnya sederhana adalah bahwa sebuah indikator diukur untuk menunjukkan hubungannya dengan skor total, dimana elemen skor indikator tersebut juga termasuk dalam skor total. Jika diukur dua kali, hasilnya akan cenderung lebih baik dari yang sebenarnya. Kriteria pengujian yang sama dengan *Product Moment*.

3.5.2. Uji Reliabilitas

Dalam penelitian, reliabilitas merujuk pada sejauh mana hasil pengukuran tetap konsisten saat diulang terhadap subjek yang sama dan dalam kondisi yang serupa. Jika hasil pengukuran konsisten dalam situasi yang sama, maka penelitian dianggap dapat diandalkan. Sebaliknya, jika tidak konsisten, maka penelitian dianggap tidak dapat diandalkan [21].

Metode *Cronbach's Alpha* adalah yang paling sering digunakan dalam penelitian. Koefisien reliabilitas, yang berkisar antara 0 dan 1, menunjukkan tingkat reliabilitas yang tinggi atau rendah. Nilai ini ditulis sebagai α , di mana α adalah index kasus yang dicari [22]:

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)

3.5.3. Data Preprocessing

Penelitian ini menggunakan *dataset* kuesioner minat beli makanan *online* pada generasi milenial yang terkumpul dalam bentuk dokumen .xlsx dan telah diisi oleh 100 orang responden. Dari *dataset* yang telah diperoleh ini akan dilakukan *preprocessing* terlebih dahulu untuk menyiapkan data agar lebih terstruktur dan efisien untuk dianalisa. Tahap *preprocessing* dilakukan dengan menggunakan program RapidMiner.

Adapun tahapan-tahapan dalam *preprocessing* pada penelitian ini adalah [19]:

1. 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.

2. Transformasi Data

Analisis data menggunakan transformasi data untuk mengubah nilai variabel untuk memenuhi asumsi atau persyaratan tertentu. Transformasi ini dapat meningkatkan distribusi data, mengurangi heteroskedastisitas, atau mempermudah interpretasi. Transformasi data yang paling umum digunakan adalah normalisasi, transformasi, dan lainnya.

3. Standarisasi

Standarisasi adalah proses pembuatan dan penerapan standar untuk mencapai tujuan tertentu dalam bidang tertentu, seperti industri, teknologi, atau layanan. Standar adalah dokumen yang berisi kriteria, pedoman, atau spesifikasi teknis yang digunakan sebagai referensi umum untuk memastikan bahwa produk atau layanan konsisten, *interoperable*, aman, dan berkualitas.

3.5.4. Pembagian Data

Proses selanjutnya, yaitu masuk ke tahap pembagian *dataset* menjadi data latih dan data uji. Pembagian data dalam penelitian ini dijabarkan sebagai berikut:

1. Data Training (Data Latih)

Pada penelitian ini, data *training* telah dikelompokkan berdasarkan kelas yang akan menjadi target klasifikasi, yaitu kelas tidak minat, rendah, sedang, tinggi. "Dalam banyak kasus, set pelatihan mencakup sekitar 70-80% dari total data" [19]. Jumlah data yang digunakan pada data training penelitian ini adalah sebanyak 80 data atau 80% dari keseluruhan data kuesioner yang memuat minat beli makanan *online* pada generasi milenial berdasarkan keputusan pembelian.

2. Data Testing (Data Uji)

Untuk melakukan pengujian dengan algoritma, penulis menggunakan 20 data atau 20% dari total keseluruhan data kuesioner yang minat beli makanan *online* pada generasi milenial berdasarkan keputusan pembelian sebagai data testing. Setelah model dilatih, set pengujian digunakan untuk menentukan seberapa baik model dapat menerapkan polanya ke data baru.

3.5.5. Implementasi Algoritma Naïve Bayes

Proses menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data untuk memprediksi kelas objek yang tidak diketahui labelnya. *Dataset* yang dinyatakan sudah siap digunakan untuk pengujian dan analisa akan diolah pada program RapidMiner. Dalam proses ini, algoritma naïve bayes perlu diberikan pengetahuan awal untuk dijadikan acuan agar dapat melakukan klasifikasi terhadap kelas tinggi, rendah, sedang dan

tidak minat, kemudian menghitung nilai probabilitas dari masing-masing kelasnya.

3.5.6. Pengukuran Performa Algoritma

Ada beberapa cara untuk mengukur performa prediksi dan cara yang paling umum digunakan adalah dengan menghitung *accuracy*, *precision*, *recall*, dan *specificity* pada acuan *confusion matrix* serta nilai *Area Under the Curve* (AUC). Maskoen dan Purnama menjelaskan, AUC (*Area Under the Curve*) adalah metrik yang menghitung area di bawah kurva *Receiver Operating Characteristic* (ROC). Secara teoritis, nilai AUC berada di rentang antara 0 dan 1. 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 [23]. Perhitungan AUC ditunjukkan dalam persamaan berikut [24]:

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

Nilai *Area Under the Curve* (AUC) dalam klasifikasi data *mining* sendiri dibagi menjadi beberapa kelompok, yaitu [25]:

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. Pada dasarnya, pengukuran ini digunakan untuk memberikan informasi tentang perbandingan hasil klasifikasi yang dilakukan oleh algoritma dengan hasil klasifikasi sebenarnya [26]. *Confusion matrix* juga dapat merepresentasikan prediksi dan kondisi aktual dari data yang telah diolah oleh algoritma data *mining*, salah satunya algoritma *naïve bayes*. Adapun kriteria model klasifikasi diantaranya adalah *accuracy*, *precision*, *recall* dan *specificity*.

Dengan rumus untuk menghitung nilai *accuracy*, *precision*, *recall* (sensivitas), dan *specificity* dari tiap kelas (i) sebagai berikut [27]:

$$1. \text{ Accuracy} = \frac{TP_{all}}{Total} \times 100\%$$

$$2. \text{ Precision} = \frac{TPP}{TPP_{all} + TFP_i}$$

$$3. \text{ Recall} = \frac{TPP_{all}}{TPP_{all} + TFP_i}$$

$$4. \text{ Specificity} = \frac{TN_{all}}{TN_{all} + TFP_i}$$

Dengan keterangan sebagai berikut:

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

3.5.6. Uji Hipotesis

Setelah melalui tahap evaluasi performa serta beberapa serangkaian proses pengolahan yang sebelumnya telah dilakukan, tahap ini merupakan tahap terakhir untuk menentukan hasil hipotesa yang didapat berdasarkan analisa dari proses pengujian. 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 yang 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.

Uji *Chi Square of Independence* atau *Chi Square Test of Association* sangat erat kaitannya dengan *table contingency* atau analisis *cross tabulasi*. Setelah mengetahui distribusi respons di setiap level kelompok, analisis penelitian dapat dimulai untuk menentukan apakah ada independensi atau dependensi antara variabel baris dan variabel kolom yang berskala nominal. Arti indenpensi atau dependensi menunjukkan hubungan antara variabel baris dan kolom [28]. Untuk mengetahui taraf signifikansi hasil analisis, pertama-tama hitung nilai derajat kebebasan (d_f) dari data yang dianalisis dengan persamaan seperti dibawah ini [29]:

$$d_f = (k - 1)(m - 1)$$

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:

1. 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.

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

4. HASIL DAN PEMBAHASAN

4.1. Uji Validitas

Pada penelitian ini, penulis menggunakan 30 responden sebagai sampel untuk pengujian validitas dan reliabilitas. Tabel 3 di bawah ini menunjukkan hasil dari uji validitas terhadap parameter-parameter pernyataan terkait keputusan pembelian dan minat beli.

Tabel 3. Hasil uji validitas

Parameter	Indikator	r_{hitung}	r_{tabel}	Keterangan
Kualitas Produk (KU)	Pernyataan 1	0,836	0,196	Valid
	Pernyataan 2	0,830		Valid
	Pernyataan 3	0,765		Valid
Bentuk Produk (BE)	Pernyataan 1	0,753	0,196	Valid
	Pernyataan 2	0,911		Valid
	Pernyataan 3	0,749		Valid
Harga Produk (HA)	Pernyataan 1	0,930	0,196	Valid
	Pernyataan 2	0,869		Valid
	Pernyataan 3	0,775		Valid
Pengalaman Merek (ME)	Pernyataan 1	0,841	0,196	Valid
	Pernyataan 2	0,841		Valid
	Pernyataan 3	0,856		Valid
Kemudahan Mendapatkan Produk (MU)	Pernyataan 1	0,863	0,196	Valid
	Pernyataan 2	0,821		Valid
	Pernyataan 3	0,884		Valid
Cara Pembayaran (PA)	Pernyataan 1	0,926	0,196	Valid
	Pernyataan 2	0,851		Valid
	Pernyataan 3	0,821		Valid
Minat Beli (MI)	Pernyataan 1	0,927	0,196	Valid
	Pernyataan 2	0,830		Valid

	Pernyataan 3	0,865		Valid
--	--------------	-------	--	-------

Berdasarkan Tabel 3 diketahui bahwa secara keseluruhan r hitung memiliki nilai koefisien korelasi yang lebih besar jika dibandingkan dengan r_{tabel} . Ini menunjukkan bahwa seluruh butir pernyataan kuesioner dapat dinyatakan valid.

4.2. Uji Reliabilitas

Pengujian ini bertujuan untuk membuktikan bahwa alat ukur dapat mengukur gejala yang sama pada berbagai kesempatan. Hasil pengujian reliabel untuk masing-masing parameter keputusan pembelian dan minat beli ditunjukkan pada tabel dibawah ini.

Tabel 3. Hasil uji reliabilitas

Parameter	Reliabilitas		Keterangan
	Koefisien α	Batas Standar	
KU	0,739	$0,60 < \alpha$	Reliabel tinggi
BE	0,733	$0,60 < \alpha$	Reliabel tinggi
HA	0,817	$0,80 < \alpha$	Reliabel sangat tinggi
ME	0,824	$0,80 < \alpha$	Reliabel sangat tinggi
MU	0,792	$0,60 < \alpha$	Reliabel tinggi
PE	0,826	$0,80 < \alpha$	Reliabel sangat tinggi
MI	0,846	$0,80 < \alpha$	Reliabel sangat tinggi

Hasil uji reliabilitas menunjukkan bahwa setiap parameter memiliki koefisien Alpha yang sangat tinggi di atas skor standar, sehingga dapat disimpulkan bahwa semua pernyataan dalam kuesioner tersebut adalah reliabel. Oleh karena itu, indikator untuk masing-masing parameter dapat digunakan sebagai indikator untuk mengukur tingkat minat beli makanan generasi milenial.

4.3. Deskripsi Data

Data yang diolah dan analisa oleh penulis ini terkumpul melalui kuesioner Google Form dengan total sebanyak 100 responden yang disebarakan melalui media sosial seperti WhatsApp grup dan Twitter (X). 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.

No.	Nama	Jenis Kelamin	Tahun_Lahir	Domisili	Pekerjaan	Frekuensi	Kualitas Produk (X1)			Bentuk Produk (X2)			Harga Produk(X3)			Pengalaman Merk (X4)			Kemudahan Mendapatkan Produk (X5)			Cara Pembayaran (X6)			Minat Beli (Y1)			Aplikasi	
							X1.1	X1.2	X1.3	X2.1	X2.2	X2.3	X3.1	X3.2	X3.3	X4.1	X4.2	X4.3	X5.1	X5.2	X5.3	X6.1	X6.2	X6.3	Y1.1	Y1.2	Y1.3		
1	Danikisnaa	Wanita	1992 - 1995	Jakarta Selatan	Ibu Rumah Tangga	4 - 5 kali	2	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	4	3	3	4	3	ShopeeFood	
2	Desya	Wanita	1996 - 2000	Jakarta Selatan	Karyawan Swasta	1 - 3 kali	4	3	3	4	4	3	3	3	4	3	4	4	4	4	4	3	4	4	1	3	3	GoFood	
3	Vio	Wanita	1988 - 1991	Jakarta Timur	Karyawan Swasta	1 - 3 kali	4	4	4	4	4	3	4	4	4	4	3	4	3	2	3	4	4	4	3	4	2	ShopeeFood	
4	Ica	Wanita	1996 - 2000	Jakarta Selatan	Karyawan Swasta	1 - 3 kali	4	4	4	3	4	4	4	4	4	4	4	4	4	3	3	4	4	4	2	2	3	GrabFood	
5	Undrey	Wanita	1996 - 2000	Jakarta Barat	Mahasiswa/i	1 - 3 kali	4	4	4	3	4	4	4	3	4	4	4	4	1	3	3	4	4	3	4	4	4	ShopeeFood	
6	Michelle	Wanita	1996 - 2000	Jakarta Selatan	Karyawan Swasta	1 - 3 kali	2	3	3	3	4	2	2	3	3	3	3	4	3	4	3	3	2	4	4	4	4	GoFood	
7	Farhan	Pria	1996 - 2000	Jakarta Selatan	Wirasaha	4 - 5 kali	4	4	4	3	3	3	3	3	3	4	4	4	2	1	2	4	4	4	4	4	4	ShopeeFood	
8	Yaamin	Pria	1992 - 1995	Jakarta Barat	Pegawai Negeri Sipil	4 - 5 kali	3	2	4	3	2	2	4	4	3	2	3	1	4	4	4	3	4	2	3	2	3	GrabFood	
9	Cila	Wanita	1996 - 2000	Jakarta Selatan	Mahasiswa/i	1 - 3 kali	4	4	3	4	3	3	4	3	4	3	4	4	3	3	3	4	4	4	4	4	3	ShopeeFood	
10	Sofi	Wanita	1996 - 2000	Jakarta Pusat	Karyawan Swasta	Lebih dari 5	3	4	2	3	4	4	4	4	4	3	3	3	4	4	4	3	3	2	4	4	4	ShopeeFood	
90	Muhamad Rafli Ar	Pria	1996 - 2000	Jakarta Barat	Mahasiswa/i	4 - 5 kali	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	Lainnya	
91	Kazuma	Pria	1996 - 2000	Jakarta Selatan	Mahasiswa/i	Lebih dari 5	4	4	4	3	4	3	4	4	3	4	3	4	4	4	3	3	4	4	4	3	3	GoFood	
92	Endah Rizka	Wanita	1992 - 1995	Jakarta Utara	Karyawan Swasta	1 - 3 kali	4	3	4	3	3	3	2	4	4	3	3	3	3	4	3	3	4	3	4	3	3	GoFood	
93	Jane Evelyn	Wanita	1992 - 1995	Jakarta Barat	Mahasiswa/i	Lebih dari 5	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	GrabFood
94	adeyana visaka	Wanita	1992 - 1995	Jakarta Barat	Wiraswasta	1 - 3 kali	4	3	4	2	4	2	4	4	4	4	4	4	3	4	4	3	4	3	3	4	3	GoFood	
95	chintya	Wanita	1992 - 1995	Jakarta Barat	Mahasiswa/i	1 - 3 kali	3	4	4	3	3	4	3	4	3	4	4	3	4	4	3	4	3	4	2	3	3	GoFood	
96	natasya	Wanita	1992 - 1995	Jakarta Selatan	Karyawan Swasta	Lebih dari 5	4	3	3	3	4	3	4	4	3	4	4	4	3	4	2	4	4	4	3	4	3	4	GoFood
97	Putri Filtria Indah H	Wanita	1996 - 2000	Jakarta Timur	Karyawan Swasta	Lebih dari 5	3	3	2	2	3	3	1	4	4	3	4	3	4	4	4	3	4	4	3	4	3	4	GrabFood
98	Laili	Wanita	1992 - 1995	Jakarta Selatan	Mahasiswa/i	4 - 5 kali	3	3	4	3	3	2	4	3	4	4	3	4	3	3	3	4	4	2	2	4	3	GoFood	
99	Elite	Wanita	1992 - 1995	Jakarta Pusat	Mahasiswa/i	1 - 3 kali	4	3	3	3	4	4	3	4	4	4	4	3	4	4	4	2	3	3	3	3	3	GoFood	
100	Sechan Ariyatama	Pria	1996 - 2000	Jakarta Timur	Mahasiswa/i	1 - 3 kali	3	4	3	3	4	3	2	4	4	3	4	4	4	1	4	4	4	4	3	3	3	GrabFood	

Gambar 2. Dataset kuesioner

4.4. Preprocessing Data

Sebelum masuk ke tahap analisa menggunakan algoritma naive bayes, penulis akan melakukan beberapa tahapan *preprocessing* terlebih dahulu sesuai yang ada pada proses *Knowledge Discovery Data* (KDD). Tahapan yang penulis lakukan untuk mempersiapkan data sebelum memasuki proses analisa adalah seperti berikut ini:

4.4.1. Data Selection (Seleksi Data)

Data yang telah dikumpulkan melalui kuesioner ini kemudian dilakukan proses penyeleksian terhadap

atribut atau variabel data. Pada tahap ini dilakukan seleksi data dengan atribut, variabel-variabel prediktor dan satu target variabel. Penulis hanya akan menggunakan 25 (dua puluh lima) 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. Berikut Gambar 3 adalah gambaran data yang digunakan sesuai dengan penentuan seleksi.

No.	Tahun Lahir	Domisili	Frekuensi	Kualitas Produk (X1)			Bentuk Produk (X2)			Harga Produk (X3)			Pengalaman Merk (X4)			Kemudahan Mendapatkan (X5)			Cara Pembayaran (X6)			Minat Beli (Y1)		
				X1.1	X1.2	X1.3	X2.1	X2.2	X2.3	X3.1	X3.2	X3.3	X4.1	X4.2	X4.3	X5.1	X5.2	X5.3	X6.1	X6.2	X6.3	Y1.1	Y1.2	Y1.3
1	1992 - 1995	Jakarta Selatan	4 - 5 kali	2	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	4	3	4	3
2	1996 - 2000	Jakarta Selatan	1 - 3 kali	4	3	3	4	4	3	3	3	4	3	4	4	4	4	4	3	4	4	1	3	3
3	1988 - 1991	Jakarta Timur	1 - 3 kali	4	4	4	4	4	3	4	4	4	4	3	4	3	2	3	4	4	4	3	4	2
4	1996 - 2000	Jakarta Selatan	1 - 3 kali	4	4	4	3	4	4	4	4	4	4	4	4	3	4	3	3	4	4	2	2	3
5	1996 - 2000	Jakarta Barat	1 - 3 kali	4	4	4	3	4	4	4	4	3	4	4	4	1	3	3	4	4	3	4	4	4
6	1996 - 2000	Jakarta Selatan	1 - 3 kali	2	3	3	3	4	4	2	2	3	3	3	3	4	3	4	3	3	2	4	4	4
7	1996 - 2000	Jakarta Selatan	4 - 5 kali	4	4	4	3	3	3	3	3	3	4	4	4	2	1	2	4	4	4	4	4	4
8	1992 - 1995	Jakarta Barat	4 - 5 kali	3	2	4	3	2	2	4	4	3	2	3	1	4	4	4	3	4	2	3	2	3
9	1996 - 2000	Jakarta Selatan	1 - 3 kali	4	4	3	4	3	3	4	3	4	3	4	4	3	3	3	4	4	4	4	4	3
10	1996 - 2000	Jakarta Pusat	Lebih dari 5	3	4	2	3	4	4	4	4	4	3	3	3	4	4	4	3	3	2	4	4	4
85	1996 - 2000	Jakarta Selatan	4 - 5 kali	3	3	4	3	4	3	4	4	2	4	4	4	4	4	4	4	3	4	4	4	4
86	1996 - 2000	Jakarta Barat	1 - 3 kali	3	3	2	2	3	3	1	4	4	3	4	3	2	2	4	2	3	1	3	3	3
87	1992 - 1995	Jakarta Selatan	1 - 3 kali	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2
88	1996 - 2000	Jakarta Selatan	1 - 3 kali	4	4	4	4	4	3	4	4	4	4	3	3	4	4	4	4	4	4	4	4	3
89	1996 - 2000	Jakarta Selatan	Lebih dari 5	4	4	3	3	4	3	2	4	4	4	4	4	4	3	4	4	4	4	3	3	3
90	1996 - 2000	Jakarta Barat	4 - 5 kali	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
91	1996 - 2000	Jakarta Selatan	Lebih dari 5	4	4	4	3	4	3	4	4	3	4	3	4	4	4	3	3	4	4	4	3	3
92	1992 - 1995	Jakarta Utara	1 - 3 kali	4	3	4	3	3	3	2	4	4	3	3	3	3	4	3	3	4	3	4	3	3
93	1992 - 1995	Jakarta Barat	Lebih dari 5	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3
94	1992 - 1995	Jakarta Barat	1 - 3 kali	4	3	4	2	4	2	4	4	4	4	4	4	3	4	4	3	4	3	3	4	3
95	1992 - 1995	Jakarta Barat	1 - 3 kali	3	4	4	3	3	4	3	4	3	4	4	3	4	4	3	4	3	4	2	3	3
96	1992 - 1995	Jakarta Selatan	Lebih dari 5	4	3	3	3	4	4	3	4	4	4	4	3	4	2	4	4	4	3	4	3	4
97	1996 - 2000	Jakarta Timur	Lebih dari 5	3	3	2	2	3	3	1	4	4	3	4	3	4	3	4	4	3	4	4	3	4
98	1992 - 1995	Jakarta Selatan	4 - 5 kali	3	3	4	3	4	3	2	4	3	4	4	4	4	3	3	4	4	2	2	4	3
99	1992 - 1995	Jakarta Pusat	1 - 3 kali	4	3	3	3	4	4	3	4	4	4	3	4	4	4	2	3	3	3	3	3	3
100	1996 - 2000	Jakarta Timur	1 - 3 kali	3	4	3	3	4	3	2	4	4	3	4	4	4	1	4	4	4	3	3	3	3

Gambar 3. Data kuesioner siap olah

4.4.2. Pembersihan Data (Data Cleansing)

Setelah tahap pengumpulan data dan filter data selesai, tahap selanjutnya adalah membersihkan data untuk menghindari duplikat, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data. Dengan demikian, data dapat diolah dan diproses untuk data *mining*. Setelah tahap pembersihan selesai, data disimpan dalam *dataset* baru dalam format *.xlsx*.

4.4.3. Transformasi Data

Transformasi ini dilakukan untuk meningkatkan kualitas data sehingga analisis menjadi lebih efektif

dan hasil yang diperoleh lebih akurat. Transformasi ini akan mengubah skala pengukuran data asli menjadi bentuk lain sehingga data dapat memenuhi ketentuan pengujian. 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.

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 [30]:

$$P = \frac{\text{rentang}}{\text{banyaknya kelas}}$$

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 4. 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:

No.	Kualitas Produk (KU)		Bentuk Produk (BE)		Harga Produk (HA)		Pengalaman Merek (ME)		Kemudahan Mendapatkan Produk (MU)		Cara Pembayaran (PE)		Minat Beli (MI)	
	Mean	Kategori (KU)	Mean	Kategori (BE)	Mean	Kategori (HA)	Mean	Kategori (ME)	Mean	Kategori (MU)	Mean	Kategori (PE)	Mean	Kategori (MI)
1	3,3	Sangat Setuju	4,0	Sangat Setuju	4,0	Sangat Setuju	4,0	Sangat Setuju	3,7	Sangat Setuju	3,3	Sangat Setuju	3,3	Tinggi
2	3,3	Sangat Setuju	3,7	Sangat Setuju	3,3	Sangat Setuju	3,7	Sangat Setuju	4,0	Sangat Setuju	3,7	Sangat Setuju	2,3	Rendah
3	4,0	Sangat Setuju	3,7	Sangat Setuju	4,0	Sangat Setuju	3,7	Sangat Setuju	2,7	Setuju	4,0	Sangat Setuju	3,0	Sedang
4	4,0	Sangat Setuju	3,7	Sangat Setuju	4,0	Sangat Setuju	4,0	Sangat Setuju	3,3	Sangat Setuju	3,7	Sangat Setuju	2,3	Rendah
5	4,0	Sangat Setuju	3,7	Sangat Setuju	3,7	Sangat Setuju	4,0	Sangat Setuju	2,3	Tidak Setuju	3,7	Sangat Setuju	4,0	Tinggi
6	2,7	Setuju	3,7	Sangat Setuju	2,3	Tidak Setuju	3,0	Setuju	3,7	Sangat Setuju	2,7	Setuju	4,0	Tinggi
7	4,0	Sangat Setuju	3,0	Setuju	3,0	Setuju	4,0	Sangat Setuju	1,7	Sangat Tidak Setuju	4,0	Sangat Setuju	4,0	Tinggi
8	3,0	Setuju	2,3	Tidak Setuju	3,7	Sangat Setuju	2,0	Tidak Setuju	4,0	Sangat Setuju	3,0	Setuju	2,7	Sedang
9	3,7	Sangat Setuju	3,3	Sangat Setuju	3,7	Sangat Setuju	3,7	Sangat Setuju	3,0	Setuju	4,0	Sangat Setuju	3,7	Tinggi
10	3,0	Setuju	3,7	Sangat Setuju	4,0	Sangat Setuju	3,0	Setuju	4,0	Sangat Setuju	2,7	Setuju	4,0	Tinggi
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
85	3,3	Sangat Setuju	3,3	Sangat Setuju	3,3	Sangat Setuju	4,0	Sangat Setuju	4,0	Sangat Setuju	3,7	Sangat Setuju	4,0	Tinggi
86	2,7	Setuju	2,7	Setuju	3,0	Setuju	3,3	Sangat Setuju	2,7	Tidak Setuju	2,0	Sangat Setuju	3,0	Sedang
87	3,0	Setuju	3,0	Setuju	3,0	Setuju	3,0	Setuju	3,0	Setuju	3,0	Setuju	2,0	Rendah
88	4,0	Sangat Setuju	3,7	Sangat Setuju	4,0	Sangat Setuju	3,3	Sangat Setuju	4,0	Sangat Setuju	4,0	Sangat Setuju	3,7	Tinggi
89	3,7	Sangat Setuju	3,3	Sangat Setuju	3,3	Sangat Setuju	4,0	Sangat Setuju	3,7	Sangat Setuju	4,0	Sangat Setuju	3,0	Sedang
90	4,0	Sangat Setuju	4,0	Sangat Setuju	4,0	Sangat Setuju	4,0	Sangat Setuju	4,0	Sangat Setuju	4,0	Sangat Setuju	4,0	Tinggi
91	4,0	Sangat Setuju	3,3	Sangat Setuju	3,7	Sangat Setuju	3,7	Sangat Setuju	3,7	Sangat Setuju	3,7	Sangat Setuju	3,3	Tinggi
92	3,7	Sangat Setuju	3,0	Setuju	3,3	Sangat Setuju	3,0	Setuju	3,3	Sangat Setuju	3,3	Sangat Setuju	3,3	Tinggi
93	4,0	Sangat Setuju	4,0	Sangat Setuju	4,0	Sangat Setuju	4,0	Sangat Setuju	4,0	Sangat Setuju	4,0	Sangat Setuju	3,3	Tinggi
94	3,7	Sangat Setuju	2,7	Setuju	4,0	Sangat Setuju	4,0	Sangat Setuju	3,7	Sangat Setuju	3,3	Sangat Setuju	3,3	Tinggi
95	3,7	Sangat Setuju	3,3	Sangat Setuju	3,3	Sangat Setuju	3,7	Sangat Setuju	3,7	Sangat Setuju	3,7	Sangat Setuju	2,7	Sedang
96	3,3	Sangat Setuju	3,7	Sangat Setuju	3,7	Sangat Setuju	3,7	Sangat Setuju	3,3	Sangat Setuju	3,7	Sangat Setuju	3,7	Tinggi
97	2,7	Setuju	2,7	Setuju	3,0	Setuju	3,3	Sangat Setuju	3,7	Sangat Setuju	3,7	Sangat Setuju	3,7	Tinggi
98	3,3	Sangat Setuju	3,3	Sangat Setuju	3,0	Setuju	3,7	Sangat Setuju	3,3	Sangat Setuju	3,3	Sangat Setuju	3,0	Sedang
99	3,3	Sangat Setuju	3,7	Sangat Setuju	3,7	Sangat Setuju	3,7	Sangat Setuju	4,0	Sangat Setuju	2,7	Setuju	3,0	Sedang
100	3,3	Sangat Setuju	3,3	Sangat Setuju	3,3	Sangat Setuju	3,7	Sangat Setuju	3,0	Setuju	3,7	Sangat Setuju	3,0	Sedang

Gambar 4. Transformasi kategori variabel

4.5. Pengujian Dengan Algoritma Naïve Bayes

Dalam proses ini melibatkan penggunaan metode prediksi dengan algoritma naïve bayes untuk menemukan informasi tentang tingkat minat beli makanan secara *online* pada generasi milenial. Gambar 5 disamping merupakan model yang dibangun pada RapidMiner dengan menggunakan *dataset* minat beli makanan *online* dalam bentuk .xlsx.

Tabel 5. 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

Penulis juga menggunakan rumus mean (rata-rata hitung) untuk menghitung nilai rata-rata dari setiap skor yang diberikan oleh masing-masing responden terhadap variabel, rumusnya sebagai berikut:

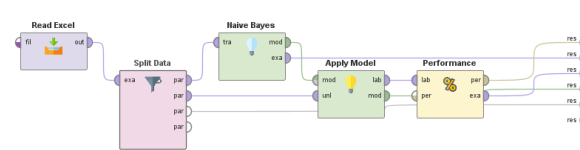
$$\text{Mean} = \frac{\sum x}{N}$$

Keterangan:

$\sum x$ = jumlah nilai kuesioner

N = jumlah responden

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).



Gambar 5. Proses pengujian dengan algoritma naïve bayes

Berdasarkan pengujian pertama yang penulis lakukan tersebut, model yang dihasilkan oleh algoritma naive bayes tidak bekerja secara maksimal untuk memprediksi minat beli makanan secara *online* pada generasi milenial. Terdapat adanya perbedaan nilai tingkat akurasi, *recall* dan *precision* yang cukup jauh diantara kelasnya, terutama kelas rendah dan tidak minat yang bernilai 0,00%, seperti pada Gambar 6.

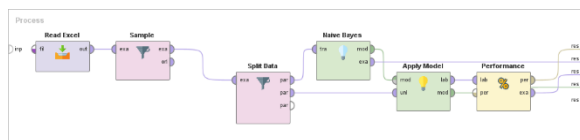
accuracy: 75.00%

	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 6. Performance algoritma naïve bayes

Hal ini bisa disebabkan karena *imbalance data* atau ketidakseimbangan data. Ketika *dataset* memiliki jumlah data yang lebih kecil daripada kelompok klasifikasi lainnya, masalah ini disebut *imbalance dataset*. Kondisi ini dapat menyebabkan salah klasifikasi model karena data kelompok minoritas lebih penting daripada data mayoritas [31]. *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 kelas tinggi, sedang, rendah dan tidak minat, penulis akan melakukan *resampling* dengan metode *downsampling*. Metode *downsampling* mengurangi jumlah kelas data mayoritas sehingga kedua kelompok klasifikasi memiliki jumlah data yang seimbang [31]. Berikut dibawah ini adalah gambaran pengujian ulang dengan *resampling*.



Gambar 7. Resampling data

4.6. Evaluasi dan Validasi Hasil Analisis

Evaluasi model dilakukan dengan menggunakan *confusion matrix* dan pengujian AUC untuk mengetahui tingkat keakuratannya. Berdasarkan pengujian *resampling* yang telah dilakukan maka didapatkan hasil evaluasi kinerja algoritma dalam memprediksi minat beli makanan secara *online* pada generasi milenial seperti di gambar 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 7. Performance algoritma naïve bayes dengan *resampling*

Berikut hasil pengujian akurasi dengan mengikuti kaidah aturan dari *confusion matrix* dalam kondisi *multi class* tabel 4x4.

$$1. \text{ Accuracy} = \frac{TTP}{total} \times 100\% = \frac{3+2+4+0}{12} \times 100\% = 75\%$$

Dapat diketahui jika tingkat kedekatan antara keseluruhan nilai prediksi dengan nilai aktual dari masing-masing kelas variabel termasuk cukup tinggi yaitu dengan proporsi mencapai 75%.

$$2. \text{ Precision} = \frac{TTP}{TTP+TFP}$$

- Tinggi = $\frac{3}{3+1} = 0,75$
- Rendah = $\frac{2}{2+1} = 0,66$
- Sedang = $\frac{4}{4+1} = 0,8$
- Tidak minat = $\frac{0}{0+0} = 0$

Dengan ini, model algoritma naive bayes dapat memprediksi nilai *True* yang benar-benar positif pada kelas tinggi, sedang dan rendah dengan baik, sementara gagal dalam memprediksi nilai *True* yang benar-benar positif pada kelas tidak minat.

$$3. \text{ Recall} = \frac{TTP}{TTP+TFN}$$

- Tinggi = $\frac{3}{3+2} = 0,60$
- Rendah = $\frac{2}{2+0} = 1$
- Sedang = $\frac{4}{4+1} = 0,8$
- Tidak minat = $\frac{0}{0+0} = 0$

Dengan ini, model algoritma naive bayes dapat mengidentifikasi proporsi nilai positif aktual pada kelas tinggi, sedang, dan rendah dengan baik sementara gagal dalam mengidentifikasi proporsi nilai positif aktual pada kelas tidak minat.

$$4. \text{ Specificity} = \frac{TTN}{TTN+TFP}$$

- Tinggi = $\frac{6}{6+2} = 0,75$
- Rendah = $\frac{9}{9+0} = 1$
- Sedang = $\frac{6}{6+1} = 0,85$
- Tidak minat = $\frac{12}{12+0} = 1$

Berdasarkan perhitungan di atas, model yang dihasilkan oleh algoritma naïve bayes efektif untuk berhasil mengidentifikasi kasus nilai kelas negatif aktual secara tepat pada kelas tinggi, sedang, rendah dan tidak minat.

Sementara itu untuk memastikan lebih dalam mengenai keakuratan *dataset*, nilai AUC yang didapat untuk kinerja algoritma naive bayes terhadap prediksi minat beli makanan *online* pada generasi milenial yaitu:

$$\begin{aligned} \text{AUC} &= \frac{1}{2} (\text{Sensivitas} + \text{Specifitas}) \\ &= \frac{0,6 + 0,9}{2} = 0,75 \end{aligned}$$

Ini menunjukkan bahwa kinerja algoritma naive 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 naive 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	-

Dengan melihat tingkat *confidence* (kepercayaan) atau prediksi tertinggi yang dihasilkan oleh model dengan melakukan *resampling* menggunakan metode *downsampling* pada Tabel 6, 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*.

4.7. 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. Uji hipotesis ini tidak mengharuskan data yang akan diuji terdistribusi normal, setiap variabel memuat data kategorik berupa nominal dan ordinal serta dilakukan hanya pada sampel berukuran besar (sebaiknya diatas 40). 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”
2. Variabel frekuensi dikategorikan menjadi 1 = “1 – 3 kali”, 2 = “4 – 5 kali” dan 3 = “Lebih dari 5”
3. 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	H_0 diterima
Kualitas	5,526	7,814	0,137	0,05	H_0 diterima
Bentuk	7,842	7,814	0,049	0,05	H_1 diterima
Harga	7,845	7,814	0,049	0,05	H_1 diterima
Merk	3,453	7,814	0,327	0,05	H_0 diterima
Mudah	2,719	7,814	0,437	0,05	H_0 diterima
Pay	5,040	7,814	0,169	0,05	H_0 diterima

Berdasarkan data-data yang telah diolah dan dianalisa pada tabel diatas, 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.

5. KESIMPULAN DAN SARAN

Dengan melakukan teknik *resampling*; metode *downsampling* menggunakan program RapidMiner, model algoritma naive bayes menghasilkan tingkat akurasi sebesar 75% dengan nilai *recall* untuk kelas Tinggi sebesar 60%, kelas Sedang sebesar 80%, kelas Rendah sebesar 100%, kelas Tidak Minat sebesar 0%. Nilai *precision* untuk kelas tinggi sebesar 75%, kelas sedang sebesar 80%, kelas rendah sebesar 66.67%, kelas tidak minat sebesar 0%. Kemudian, nilai *Area Under Curve* (AUC) untuk model algoritma yang dihasilkan adalah sebesar 75%. Ini menunjukkan bahwa kinerja dari model algoritma naive bayes cukup baik untuk memprediksi kelas positif dan negatif dari masalah tingkat minat beli makanan *online* pada generasi milenial dengan tepat dan akurat. Berdasarkan hasil penelitian dan analisis yang telah dilakukan, model memprediksi adanya tingkat minat beli sedang dari generasi milenial yang lahir pada tahun 1992 hingga 1995 dan bertempat tinggal di

Jakarta Timur dengan tingkat akurasi probabilitas sebesar 0,999. Adapun aplikasi GoFood sebagai aplikasi pesan makanan yang biasanya digunakan oleh generasi milenial dalam membeli makanan secara *online* dengan tingkat akurasi probabilitas sebesar 0,999. 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 yang telah dibahas pada bab sebelumnya, 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 yaitu algoritma *decision tree*, algoritma *random forest*, algoritma *neural network* ataupun membandingkan beberapa algoritma sehingga dapat mengembangkan penelitian-penelitian di masa yang akan datang. Penulis juga menyarankan untuk mengoptimalkan pembagian parameter dan distribusi data kelas klasifikasi atau prediksi yang akan digunakan untuk penelitian selanjutnya.

DAFTAR PUSTAKA

- [1] Taswiyah and A. Imron, "Persepsi dan Preferensi Ibu Rumah Tangga Terhadap Sistem Berbelanja Secara Online/online shop," *J. Abdimas BSI J. Pengabd. Kpd. Masy.*, vol. 4, no. 1, pp. 128–137, Feb. 2021, doi: 10.31294/JABDIMAS.V4I1.9169.
- [2] F. Tjiptono and A. Diana, *Kepuasan Pelanggan : Konsep, Pengukuran, Dan Strategi*, i. Yogyakarta: Andi Offset, 2019.
- [3] Erdawati, Endarwita, and R. Widiarsari, "PENGARUH ELECTRONIC WORD OF MOUTH (e-WOM) DAN KUALITAS PRODUK TERHADAP KEPUTUSAN PEMBELIAN SKINCARE," *J. Soc. Econ. Res.*, vol. 5, no. 1, pp. 184–190, Jul. 2023, doi: 10.54783/JSER.V5I1.82.
- [4] I. Werdiningsih, B. Nuqoba, and Muhammadun, *DATA MINING MENGGUNAKAN ANDROID, WEKA, DAN SPSS*. Airlangga University Press, 2020.
- [5] R. Irwansyah *et al.*, *PERILAKU KONSUMEN*, 1st ed. Bandung: WIDINA BHAKTI PERSADA BANDUNG, 2021.
- [6] D. Gunawan, *Keputusan Pembelian Konsumen Marketplace Shopee Berbasis Social Media Marketing*, 1st ed. PT Inovasi Pratama Internasional, 2022.
- [7] Y. Arfah, *Keputusan Pembelian Produk*, Padangsid. vol. 1. Sumatera Utara: PT Inovasi Pratama Internasional, 2022.
- [8] D. K. Sawlani, *Keputusan Pembelian Online: Kualitas Website, Keamanan dan Kepercayaan*, 1st ed. Surabaya: Scopindo Media Pustaka, 2021.
- [9] M. R. Stansyah, R. Choirunnisa, M. Tegar, and S. S. A. Putri, "ANALISIS PENGARUH MINAT BELI KONSUMEN TERHADAP PEMBELIAN MAKANAN DAN MINUMAN MELALUI APLIKASI GO FOOD," *J. Pendidik. Tata Niaga*, vol. 11, no. 1, pp. 43–49, Apr. 2023, doi: 10.26740/JPTN.V11N1.P43-49.
- [10] 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 J. Manaj. dan Akunt.*, vol. 16, no. 2, pp. 499–511, Dec. 2021, doi: 10.32534/JV.V16I2.2018.
- [11] H. N. Az-zahra, V. A. Tanya, and N. C. Apsari, "LAYANAN ONLINE FOOD DELIVERY DALAM MEMBANTU MENINGKATKAN PENJUALAN PADA USAHA MIKRO," *J. Penelit. dan Pengabd. Kpd. Masy.*, vol. 2, no. 2, pp. 156–165, Aug. 2021, doi: 10.24198/JPPM.V2I2.33513.
- [12] V. Vincentius and L. S. Putranto, "ANALISIS PENGARUH LAYANAN PESAN MAKANAN ONLINE TERHADAP PERJALANAN BERBASIS RUMAH DAN BERBASIS TEMPAT AKTIVITAS," *JMTS J. Mitra Tek. Sipil*, vol. 3, no. 4, pp. 1003–1016, Nov. 2020, doi: 10.24912/JMTS.V3I4.8391.
- [13] 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.
- [14] J. M. Tulung, A. Syahid, Y. Janis, and Y. O. Kalampung, *Generasi Milenial: Diskursus Teologi, Pendidikan, Dinamika Psikologis dan Kelekatan pada Agama di Era Banjir Informasi*, 1st ed., vol. 1. Depok: RAJAWALI PERS, 2019.
- [15] H. Madiistriyatno and D. Hadiwijaya, *Generasi Milenial Tantangan Membangun Komitmen Kerja/Bisnis dan Adversity Quotient (AQ)*, 1st ed. Tangerang: Indigo Media, 2019.
- [16] M. Arif, *Generasi Millennial Dalam Internalisasi Karakter Nusantara*, 1st ed. Kediri: IAIN Kediri Press, 2021.
- [17] Mustika *et al.*, *Data Mining dan Aplikasinya*, 1st ed. Bandung: Widina Bhakti Persada Bandung, 2021.
- [18] N. L. W. S. R. Ginantra *et al.*, *Data Mining dan Penerapan Algoritma*, 1st ed. Bali: Yayasan Kita Menulis, 2021.
- [19] P. W. Rahayu *et al.*, *Buku Ajar Data Mining*, 1st ed. Jambi: PT Sonpedia Publishing Indonesia, 2024.
- [20] M. A. Muslim *et al.*, *Data Mining Algoritma C4.5 Disertai contoh kasus dan penerapannya dengan program computer*, 1st ed. Semarang:

- CV Harian Jateng Network, 2019.
- [21] M. M. Sanaky, "ANALISIS FAKTOR-FAKTOR KETERLAMBATAN PADA PROYEK PEMBANGUNAN GEDUNG ASRAMA MAN 1 TULEHU MALUKU TENGAH," *J. SIMETRIK*, vol. 11, no. 1, pp. 432–439, Aug. 2021, doi: 10.31959/JS.V11I1.615.
- [22] R. Slamet and S. Wahyuningsih, "VALIDITAS DAN RELIABILITAS TERHADAP INSTRUMEN KEPUASAN KERJA," *Aliansi J. Manaj. dan Bisnis*, vol. 17, no. 2, Dec. 2022, doi: 10.46975/ALIANSI.V17I2.428.
- [23] V. R. Sari, F. Firdausi, and Y. Azhar, "Perbandingan Prediksi Kualitas Kopi Arabika dengan Menggunakan Algoritma SGD, Random Forest dan Naive Bayes," *Edumatic J. Pendidik. Inform.*, vol. 4, no. 2, pp. 1–9, Dec. 2020, doi: 10.29408/EDUMATIC.V4I2.2202.
- [24] Z. Annisa and B. S. S. Ulama, "Analisis Sentimen Data Ulasan Pengguna Aplikasi 'PeduliLindungi' pada Google Play Store Menggunakan Metode Naive Bayes Classifier Model Multinomial," *J. Sains dan Seni ITS*, vol. 11, no. 6, pp. D464–D471, May 2023, doi: 10.12962/j23373520.v11i6.94064.
- [25] G. Budiyo and D. Suryadi, "A Machine Learning Prediction Model to a Scholarship Program for New Undergraduate Students at a Private University," *J. Rekayasa Sist. Ind.*, vol. 12, no. 2, pp. 187–200, Oct. 2023, doi: 10.26593/JRSI.V12I2.6595.187-200.
- [26] P. Romadloni, B. A. Kusuma, and W. Maulana Baihaqi, "KOMPARASI METODE PEMBELAJARAN MESIN UNTUK IMPLEMENTASI PENGAMBILAN KEPUTUSAN DALAM MENENTUKAN PROMOSI JABATAN KARYAWAN," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 622–628, Sep. 2022, doi: 10.36040/JATI.V6I2.5238.
- [27] 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," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 5, pp. 973–982, Oct. 2023, doi: 10.25126/JTIK.20231057142.
- [28] S. Yamin, *TUTORIAL STATISTIK SPSS, LISREL, WARPLS, & JASP (MUDAH & APLIKATIF)*, 1st ed. Depok: PT Dewangga Energi Internasional, 2021.
- [29] M. Naek Pane, Mardiaman, and M. Azhar, "PENENTUAN FAKTOR KETERLAMBATAN WAKTU PELAKSANAAN PROYEK APBN/APBD DENGAN ANGGARAN TAHUN TUNGGAL UNTUK MENGURANGI KLAIM," *Menara J. Tek. Sipil*, vol. 18, no. 2, pp. 96–109, Jun. 2023, doi: 10.21009/JMENARA.V18I2.35422.
- [30] D. Sunardi, "Pengaruh Gaya Kepemimpinan dan Jabatan Fungsional Perekayasa Terhadap Kinerja Pegawai Balai Besar Aerodinamika, Aeroelastika, dan Aeroakustika di Kawasan Puspipetek Setu, Tangerang Selatan," *J. Madani Ilmu Pengetahuan, Teknol. dan Hum.*, vol. 3, no. 1, pp. 151–163, Mar. 2020, doi: 10.33753/MADANI.V3I1.105.
- [31] A. Zahra Hafizhah, S. Suakanto, R. Yanu Fa'rifah, and E. Triono Nuryatno, "Klasifikasi Laju Pernapasan dan Saturasi Oksigen Menggunakan Regresi Logistik," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 8, no. 2, pp. 448–458, May 2023, doi: 10.29100/JUPI.V8I2.3481.