



Implementasi Naïve Bayes untuk Memprediksi Tingkat Kunjungan Pelanggan Menggunakan Algoritma Naïve Bayes

Nazwa Adelia Putri*, Zihan Maharani, Ilona Dwi Shelvani, Harly Okprana

Teknik Informatika, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Email: ^{1,*}nazwaadelia Putri17@gmail.com, ²zihanmaharani019@gmail.com, ³ilonadwishelvani@gmail.com,

⁴harly@amiktunasbangsa.ac.id

Email Penulis Korespondensi: nazwaadelia Putri17@gmail.com

Abstrak—Penelitian ini bertujuan untuk mengimplementasikan algoritma Naive Bayes dalam memprediksi tingkat kunjungan pelanggan di Kyemoona Kitchen dengan memanfaatkan data historis yang tersedia. Seiring dengan berkembangnya teknologi digital, analisis data telah menjadi aspek penting dalam mendukung pengambilan keputusan bisnis. Namun, analisis manual terhadap data yang kompleks dan beragam dapat menjadi tantangan. Oleh karena itu, pendekatan berbasis pembelajaran mesin, khususnya Naive Bayes, digunakan untuk menggali pola dalam data besar dan menghasilkan prediksi yang akurat. Dalam penelitian ini, data yang dikumpulkan mencakup variabel seperti waktu kunjungan, jenis promosi, kondisi cuaca, hari libur, dan faktor lainnya. Model Naive Bayes yang diterapkan mencapai akurasi 85.6%, dengan metrik evaluasi lainnya seperti presisi 82.4%, recall 84.2%, dan F1-score 83.3%. Hasil penelitian menunjukkan bahwa algoritma ini dapat mengidentifikasi faktor-faktor signifikan, seperti promosi dan kondisi cuaca, yang mempengaruhi kunjungan pelanggan. Penelitian ini tidak hanya memberikan wawasan praktis bagi Kyemoona Kitchen dalam merencanakan strategi operasional berbasis data, tetapi juga diharapkan dapat menginspirasi usaha kecil dan menengah (UMKM) lainnya untuk mengadopsi teknologi analitik serupa. Meskipun demikian, penelitian ini memiliki keterbatasan, seperti ketergantungan pada kualitas data, yang dapat memengaruhi keakuratan model. Oleh karena itu, disarankan untuk penelitian selanjutnya menggabungkan Naive Bayes dengan algoritma lain dan menggunakan dataset yang lebih besar untuk hasil yang lebih andal.

Kata Kunci: Naïve Bayes; Prediksi Tingkat Kunjungan; Data Mining; Klasifikasi Data; UMKM

Abstract—This study aims to implement the Naive Bayes algorithm in predicting customer visit rates at Kyemoona Kitchen by utilizing available historical data. With the development of digital technology, data analysis has become an important aspect in supporting business decision making. However, manual analysis of complex and diverse data can be challenging. Therefore, a machine learning-based approach, specifically Naive Bayes, is used to explore patterns in big data and generate accurate predictions. In this study, the data collected includes variables such as visit time, promotion type, weather conditions, holidays, and other factors. The Naive Bayes model achieved an accuracy of 85.6%, with other evaluation metrics such as precision of 82.4%, recall of 84.2%, and F1-score of 83.3%. The results show that this algorithm can identify significant factors, such as promotions and weather conditions, that affect customer visits. This study not only provides practical insights for Kyemoona Kitchen in planning data-driven operational strategies, but also aims to inspire other small and medium-sized enterprises (SMEs) to adopt similar analytical technologies. However, this study has limitations, such as dependence on data quality, which can affect the accuracy of the model. Therefore, it is recommended that future research combine Naive Bayes with other algorithms and use larger datasets for more reliable results.

Keywords: Naïve Bayes; Visit Rate Prediction; Data Mining; Data Classification; MSMEs

1. PENDAHULUAN

Dalam dunia bisnis ritel, pelanggan merupakan elemen yang sangat krusial dalam menentukan keberhasilan suatu usaha. Dalam dunia bisnis ritel, pelanggan merupakan elemen yang sangat krusial dalam menentukan keberhasilan suatu usaha [1]. Kehadiran pelanggan di toko bukan hanya sekadar transaksi, tetapi juga mencerminkan efektivitas strategi pemasaran, kualitas produk, hingga pengalaman belanja yang ditawarkan [2]. Tingkat kunjungan pelanggan, khususnya, menjadi indikator penting yang menunjukkan sejauh mana toko mampu menarik minat konsumen dan mempertahankan loyalitas mereka [3]. Pada Toko Kyemoona Kitchen, memahami pola kunjungan pelanggan sangat penting untuk meningkatkan efektivitas penjualan dan mengoptimalkan strategi pemasaran yang lebih tepat sasaran [4]. Dengan data kunjungan yang terstruktur, pemilik toko dapat menganalisis faktor-faktor yang memengaruhi perilaku pelanggan dan menggunakan informasi tersebut untuk mengambil keputusan bisnis yang lebih baik dan terarah [5].

Namun, memprediksi tingkat kunjungan pelanggan bukanlah tugas yang sederhana. Tingkat kunjungan dapat dipengaruhi oleh berbagai variabel, seperti promosi, kondisi cuaca, diskon, hingga jenis produk yang ditawarkan [6]. Ketergantungan pada data historis untuk menentukan pola ini memerlukan metode analisis yang tepat dan akurat [7]. Oleh karena itu, pengembangan model prediksi yang andal menjadi langkah penting untuk membantu bisnis memahami perilaku pelanggan dan meningkatkan daya saing mereka di pasar [8]. Prediksi tingkat kunjungan pelanggan menjadi salah satu pendekatan yang penting dalam analisis bisnis [9]. Dengan memprediksi jumlah pelanggan yang akan berkunjung, pemilik bisnis dapat mengelola stok, tenaga kerja, dan sumber daya lainnya dengan lebih baik [10].

Selain itu, prediksi ini memungkinkan bisnis untuk meningkatkan efisiensi operasional dan menawarkan pengalaman yang lebih baik kepada pelanggan. Metode prediksi yang akurat juga memberikan wawasan untuk mengidentifikasi faktor-faktor signifikan yang memengaruhi tingkat kunjungan, sehingga dapat menjadi dasar

pengambilan keputusan strategis [11]. Dalam jangka panjang, penggunaan prediksi ini membantu bisnis untuk mengurangi risiko kelebihan atau kekurangan stok, yang dapat berdampak langsung pada profitabilitas [12]. Selain itu, pemahaman yang mendalam tentang perilaku pelanggan memungkinkan toko untuk merancang program loyalitas yang lebih efektif dan menarik. Dengan demikian, bisnis dapat meningkatkan hubungan jangka panjang dengan pelanggan dan menciptakan pengalaman belanja yang lebih personal serta berkesan [13].

Berbagai metode analisis data telah digunakan secara luas untuk memprediksi berbagai macam data dalam berbagai bidang, mulai dari bisnis hingga kesehatan. Metode seperti random forest dan decision tree sering digunakan dalam berbagai macam penelitian untuk mengidentifikasi pola dan membuat prediksi [14][15]. Meskipun demikian, metode-metode ini memiliki keterbatasan, terutama dalam menangani data dengan probabilitas dan hubungan antar variabel yang kompleks [16]. Naïve Bayes dapat memecahkan masalah random forest dan decision tree, karena naïve bayes merupakan algoritma berbasis probabilitas yang sederhana namun sangat efektif untuk tugas klasifikasi sehingga dapat bekerja dengan baik meskipun dataset memiliki ukuran yang besar dan fitur yang banyak. Selain itu, algoritma ini relatif mudah diimplementasikan dan memerlukan waktu komputasi yang lebih rendah dibandingkan metode lainnya [17].

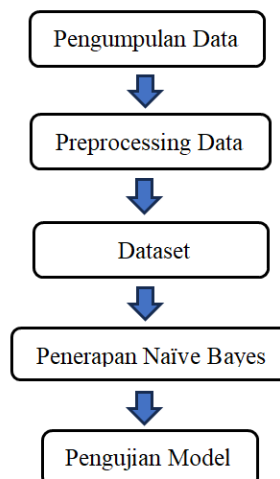
Maka dari itu, algoritma Naïve Bayes lebih unggul dalam menangani data probabilitas. Beberapa Penelitian terdahulu yaitu oleh Muhammad Yasir dan Robertus Suraji pada tahun 2023 membuktikan dengan membandingkan metode klasifikasi Naïve Bayes, Decision Tree, dan Random Forest dalam analisis sentimen terhadap kebijakan kenaikan biaya haji di media sosial YouTube dengan hasil penelitian menunjukkan bahwa Naïve Bayes menghasilkan akurasi tertinggi sebesar 90%, mengungguli Decision Tree 83% dan Random Forest 87% [18]. Lebih lanjut, Penelitian oleh Siti Lestari dkk pada tahun 2020 menunjukkan bahwa Naïve Bayes dapat diterapkan untuk mengklasifikasikan kelayakan pemberian pinjaman pada koperasi dengan tingkat akurasi 78,08% [19]. Penelitian ini menggaris bawahi efektivitas algoritma ini dalam menangani data berbasis probabilitas meskipun terdapat tantangan pada asumsi independensi antar fitur. Selain itu, penelitian oleh Bayu Purbayanto dkk pada tahun 2023 untuk analisis sentimen pengguna terhadap ChatGPT menggunakan algoritma Naïve Bayes mencatat akurasi sebesar 87,17% [20]. Hal ini membuktikan bahwa naïve bayes mampu dalam mengklasifikasi data dan relevansinya untuk berbagai konteks analisis data [21].

Penelitian ini bertujuan untuk mengimplementasikan algoritma Naïve Bayes dalam memprediksi tingkat kunjungan pelanggan di Toko Kyemoona Kitchen. Proses prediksi akan dilakukan dengan memanfaatkan data historis, seperti faktor promosi, kondisi cuaca, dan variabel terkait lainnya [22]. Implementasi ini diharapkan mampu memberikan prediksi yang lebih akurat dibandingkan metode sebelumnya, sekaligus mengidentifikasi faktor-faktor yang paling memengaruhi tingkat kunjungan [23]. Untuk mengukur keberhasilan penelitian, digunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Dengan tercapainya tujuan penelitian ini, diharapkan dapat memberikan kontribusi baik dari sisi teoritis maupun praktis. Secara teoritis, penelitian ini akan menambah wawasan tentang penerapan algoritma Naïve Bayes dalam analisis data berbasis klasifikasi. Secara praktis, hasil penelitian ini diharapkan dapat membantu Toko Kyemoona Kitchen dan para UMKM di Indoensia dalam mengoptimalkan strategi bisnis berdasarkan pola kunjungan pelanggan, sehingga meningkatkan efisiensi dan daya saing toko di pasar.

2. METODE PENELITIAN

2.1 Kerangka Dasar Penelitian

Tahapan penelitian ini dirancang untuk menganalisis data kunjungan pelanggan Kyemoona Kitchen menggunakan algoritma Naive Bayes. Penelitian ini melibatkan beberapa tahapan utama yang diuraikan dalam Gambar 1. Tahapan Penelitian sebagai berikut:



Gambar 1. Tahapan Penelitian



2.2 Tahapan Penelitian

A. Pengumpulan Data

Tahap pertama dalam penelitian ini adalah pengumpulan data, yang merupakan langkah awal yang sangat penting untuk memastikan keberhasilan proses analisis selanjutnya. Pada tahap ini, data-data yang relevan dan akurat akan dikumpulkan untuk memberikan gambaran yang jelas tentang fenomena yang sedang diteliti. Data yang dikumpulkan dalam penelitian ini adalah tingkat kunjungan ke Keymoona Kitchen selama periode dua bulan, yaitu dari bulan September 2024 hingga Oktober 2024. Data yang dikumpulkan meliputi jumlah kunjungan, waktu kunjungan, dan faktor lain yang relevan untuk membangun dasar dataset.

B. Pemrosesan awal Data

Setelah data terkumpul, langkah berikutnya adalah melakukan preprocessing atau pengolahan data. Proses ini meliputi pembersihan data (data cleaning), penanganan data yang hilang, transformasi data, dan normalisasi untuk memastikan kualitas data yang digunakan.

Data yang ada terdiri dari kategori dan harus diubah menjadi format yang sesuai untuk perhitungan probabilitas. Setiap fitur yang berupa kategori (Diskon, Cuaca, dan Promosi) diubah menjadi variabel numerik dengan cara sebagai berikut:

1. Diskon:

- a) Ada = 1
- b) Tidak Ada = 0

2. Cuaca:

- a) Cerah = 1
- b) Hujan = 2
- c) Mendung = 3

3. Promosi:

- a) Online = 1
- b) Offline = 2

C. Data Set

Dataset adalah hasil akhir dari proses preprocessing. Dataset ini memuat informasi yang terstruktur tentang tingkat kunjungan pelanggan, yang nantinya akan digunakan sebagai input dalam proses analisis, seperti yang ditampilkan pada Tabel 1. Dataset.

Tabel 1. Dataset

Hari	Diskon	Cuaca	Promosi	Kunjungan
Senin	Ada	Cerah	Online	Tinggi
Selasa	Tidak Ada	Hujan	Offline	Rendah
Rabu	Tidak Ada	Mendung	Offline	Sedang
Kamis	Ada	Cerah	Online	Tinggi
Jumat	Tidak Ada	Hujan	Online	Rendah
Sabtu	Ada	Cerah	Offline	Sedang
Minggu	Ada	Mendung	Online	Tinggi

Adapun penjelasan isi dari Tabel 1. Dataset yang siap uji dan telah melalui tahap preprocessing data adalah sebagai berikut:

1. Hari

Mengindikasikan nama hari saat data kunjungan pelanggan diambil. Variabel ini digunakan untuk menganalisis apakah pola kunjungan pelanggan memiliki tren khusus pada hari-hari tertentu, seperti akhir pekan atau hari kerja.

2. Diskon

Menunjukkan adanya diskon yang ditawarkan kepada pelanggan pada hari tersebut. Diskon merupakan salah satu cara untuk menarik perhatian pelanggan dan meningkatkan jumlah kunjungan. Nilainya dibagi menjadi dua kategori: Ada atau Tidak Ada.

3. Cuaca

Variabel yang mencerminkan kondisi cuaca pada hari tersebut, seperti Cerah, Mendung, atau Hujan. Cuaca dapat mempengaruhi perilaku pelanggan, terutama dalam kondisi cuaca ekstrem yang dapat mengurangi kunjungan.

4. Promosi



Menampilkan jenis promosi yang dijalankan pada hari tersebut. Promosi dapat dilakukan secara Online melalui platform digital atau Offline secara langsung di lokasi usaha. Variabel ini menunjukkan efektivitas promosi dalam menarik pelanggan.

5. Tingkat Kunjungan

Variabel target yang menjadi hasil akhir dari penelitian. Tingkat kunjungan dikategorikan menjadi Tinggi, Sedang, atau Rendah berdasarkan jumlah pelanggan yang datang. Variabel ini membantu dalam mengevaluasi dampak dari variabel-variabel input lainnya.

D. Penerapan Naïve Bayes

Pada tahap ini, algoritma Naive Bayes diterapkan untuk mempelajari pola dari dataset yang telah disiapkan. Algoritma ini akan digunakan untuk mengklasifikasikan atau memprediksi tingkat kunjungan berdasarkan data historis. Proses ini melibatkan pembentukan model probabilistik yang menghubungkan fitur data dengan kategori tingkat kunjungan.

E. Model Pengujian

Model yang telah dibangun diuji menggunakan data uji (testing data) untuk memutar kinerjanya. Tahapan ini meliputi analisis akurasi, presisi, recall, dan evaluasi metrik lainnya untuk memastikan bahwa model Naive Bayes yang diterapkan mampu memprediksi tingkat kunjungan dengan baik.

2.3 Kriteria dan Alternatif untuk Prediksi

Berikut adalah contoh tabel kriteria dan alternatif yang dapat digunakan dalam penelitian ini untuk memprediksi tingkat kunjungan pelanggan di Kyemiona Kitchen dengan menggunakan algoritma Naive Bayes:

a) Kriteria

Kriteria merupakan atribut atau faktor yang digunakan untuk menilai atau memprediksi tingkat kunjungan pelanggan, seperti yang ditunjukkan pada Tabel 2. Kriteria untuk Prediksi Tingkat Kunjungan.

Tabel 2. Kriteria untuk Prediksi Tingkat Kunjungan

Tingkat Kunjungan	Skala	Deskripsi
Tinggi	3	21-30 kunjungan per hari
Sedang	2	11-20 kunjungan per hari
Rendah	1	6-10 kunjungan per hari

Adapun penjelasan Tabel 2. Kriteria untuk Prediksi Tingkat Kunjungan yang ingin kami sampaikan mengenai kriteria tingkat kunjungan pelanggan yaitu sebagai berikut:

1. Tingkat Kunjungan Tinggi

Tingkat kunjungan dianggap Tinggi jika jumlah pelanggan yang datang ke tempat usaha berada di atas rata-rata kunjungan harian. Biasanya, tingkat kunjungan tinggi terjadi pada hari-hari tertentu seperti akhir pekan, saat cuaca cerah, atau ketika ada promosi menarik yang sedang berlangsung. Hal ini menunjukkan minat pelanggan yang signifikan terhadap penawaran atau layanan yang diberikan pada hari tersebut.

2. Tingkat Kunjungan Sedang

Tingkat kunjungan dikategorikan sebagai Sedang jika jumlah pelanggan yang datang mendekati rata-rata kunjungan harian. Kunjungan dengan tingkat sedang biasanya terjadi pada hari kerja tanpa adanya faktor pendorong spesifik seperti promosi besar atau diskon signifikan. Tingkat kunjungan ini menunjukkan aktivitas pelanggan yang cukup stabil tanpa ada pemutaran atau penurunan drastis.

3. Tingkat Kunjungan Rendah

Tingkat kunjungan dianggap Rendah jika jumlah pelanggan yang datang jauh di bawah rata-rata kunjungan harian. Hal ini biasanya terjadi pada hari-hari dengan cuaca buruk seperti hujan lebat, atau ketika tidak ada promosi yang ditawarkan. Kunjungan rendah dapat menjadi pertanda tidak adanya daya tarik usaha pada hari tertentu atau pengaruh dari faktor eksternal yang signifikan.

b) Alternatif

Alternatif Merujuk pada berbagai kemungkinan hasil atau skenario yang dihasilkan dari analisis prediksi tingkat kunjungan, seperti yang ditunjukkan oleh Tabel 3. Alternatif untuk Prediksi Tingkat Kunjungan.

Tabel 3. Alternatif untuk Prediksi Tingkat Kunjungan

Alternatif (Input)	Deskripsi
--------------------	-----------



Hari	Hari dalam seminggu sebagai penanda waktu prediksi (<i>Senin hingga Minggu</i>).
Diskon	Apakah ada diskon yang berlaku pada hari tersebut (<i>Ada atau Tidak Ada</i>).
Cuaca	Kondisi cuaca saat itu (<i>Cerah, Mendung, Hujan</i>).
Promosi	Bentuk promosi yang dilakukan (<i>Online, Offline, atau Tidak Ada</i>).

Adapun perincian isi dari Tabel 3. Alternatif untuk Prediksi Tingkat Kunjungan yang ingin kami sampaikan yaitu sebagai berikut:

1. Hari

Merujuk pada nama hari dalam seminggu ketika data kunjungan pelanggan dikumpulkan. Faktor ini penting karena pola kunjungan pelanggan dapat bervariasi tergantung pada hari tertentu, seperti hari kerja atau akhir pekan.

2. Diskon

Indikator adanya diskon atau tidak pada hari tertentu. menjadi Diskon salah satu strategi untuk menarik lebih banyak pelanggan. Kategorisasi untuk fitur ini adalah: Ada atau Tidak Ada.

3. Cuaca

Menggambarkan kondisi cuaca pada hari tertentu, seperti Cerah, Mendung, atau Hujan. Cuaca dapat mempengaruhi keputusan pelanggan untuk mengunjungi tempat usaha, terutama jika lokasi tersebut sulit diakses dalam kondisi tertentu.

4. Promosi

Indikator adanya promosi tertentu yang berlangsung pada hari itu, seperti promosi Online (daring) atau Offline (luring). Promosi yang efektif dapat meningkatkan minat pelanggan untuk datang.

2.4 Naive Baiyes

Algoritma Naive Bayes adalah metode klasifikasi yang didasarkan pada teorema Bayes, dengan asumsi bahwa setiap fitur bersifat independen satu sama lain. Algoritma Naive Bayes adalah metode klasifikasi yang didasarkan pada teorema Bayes, dengan asumsi bahwa setiap fitur bersifat independen satu sama lain [24]. Algoritma ini sering digunakan karena kesederhanaannya dan efektivitasnya dalam berbagai masalah klasifikasi [25]. Berikut merupakan rumus teorema naïve bayes yang akan digunakan untuk menghitung probabilitas tingkat kunjungan pelanggan pada kyemoona kitchen

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)} \quad (1)$$

Di mana $P(A|B)$ adalah probabilitas posterior, $P(B|A)$ adalah probabilitas bukti diberikan hipotesis, $P(A)$ adalah probabilitas prior, dan $P(B)$ adalah probabilitas bukti.

3. HASIL DAN PEMBAHASAN

3.1 Hitung Probabilitas Prior

Pada tahap awal analisis menggunakan metode Naïve Bayes, langkah penting yang harus dilakukan adalah menghitung probabilitas prior. Probabilitas prior merupakan ukuran awal terhadap kemungkinan terjadinya suatu kelas sebelum mempertimbangkan informasi atau bukti tambahan dari fitur-fitur yang ada dalam data. Dalam konteks ini, probabilitas prior dihitung berdasarkan proporsi jumlah data pada masing-masing kelas tingkat kunjungan terhadap total keseluruhan data yang tersedia.

- Tingkat kunjungan tinggi = 3
- Tingkat kunjungan sedang = 2
- Tingkat kunjungan rendah = 2

$$P(\text{Kunjungan}) = \frac{\text{Jumlah data dengan kunjungan kelas}}{\text{Total total data}} = \frac{n}{7} = \dots \quad (2)$$

$$P(\text{Tinggi}) = \frac{\text{Jumlah data dengan kunjungan tinggi}}{\text{Total total data}} = \frac{3}{7} = 0,4286$$

$$P(\text{Sedang}) = \frac{\text{Jumlah data dengan kunjungan sedang}}{\text{Total total data}} = \frac{2}{7} = 0,2857$$

$$P(\text{Rendah}) = \frac{\text{Jumlah data dengan kunjungan rendah}}{\text{Total total data}} = \frac{2}{7} = 0,2857$$

Nilai-nilai ini memberikan informasi awal mengenai distribusi kelas dalam dataset. Probabilitas prior ini kemudian akan dikombinasikan dengan probabilitas bersyarat (likelihood) dari masing-masing fitur untuk menghasilkan prediksi akhir dalam model Naïve Bayes. Penting dicatat bahwa semakin banyak jumlah data dalam suatu kelas, maka semakin tinggi nilai probabilitas prior-nya, dan sebaliknya. Ini menjadikan probabilitas prior sebagai komponen penting dalam proses klasifikasi berbasis probabilistik.

3.2 Kemungkinan

A. P (Fitur|Kunjungan) :

$$P(\text{Fitur} | \text{Kunjungan}) = \frac{\text{jumlah data dengan fitur yang diberikan dan kelas yang diberikan}}{\text{jumlah data dalam kelas}} \quad (3)$$

$$1. P(\text{Diskon} = \text{Ada} | \text{Tinggi}) = \frac{\text{Ada dan kunjungan}}{\text{jumlah data dengan kunjungan}} = \frac{3}{3} = 1$$

Keterangannya :

- Data dengan Diskon = Ada dan Kunjungan = Tinggi
Senin, Kamis, Jumat
- Total Data dengan Kunjungan = Tinggi : 3 Data

$$2. P(\text{Diskon} = \text{Tidak Ada} | \text{Tinggi}) = \frac{\text{Tidak Ada dan kunjungan}}{\text{jumlah data dengan kunjungan}} = \frac{0}{3} = 0$$

Keterangannya :

- Data dengan Diskon = Tidak Ada dan Kunjungan = Tinggi
Tidak Ada
- Total Data dengan Kunjungan = Tinggi : 3 Data

$$3. P(\text{Diskon} = \text{Ada} | \text{Sedang}) = \frac{\text{Ada dan kunjungan}}{\text{jumlah data dengan kunjungan}} = \frac{1}{2} = 0,5$$

Keterangannya :

- Data dengan Diskon = Ada dan Kunjungan = Sedang
Sabtu
- Total Data dengan Kunjungan = Sedang : 2 Data

$$4. P(\text{Diskon} = \text{Tidak Ada} | \text{Sedang}) = \frac{\text{Ada dan kunjungan}}{\text{jumlah data dengan kunjungan}} = \frac{1}{2} = 0,5$$

Keterangannya :

- Data dengan Diskon = Ada dan Kunjungan = Sedang
Rabu
- Total Data dengan Kunjungan = Tinggi : 2 Data

$$5. P(\text{Diskon} = \text{Tidak Ada} | \text{Tinggi}) = \frac{\text{Ada dan kunjungan}}{\text{jumlah data dengan kunjungan}} = \frac{0}{2} = 0$$

Keterangannya :

- Data dengan Diskon = Ada dan Kunjungan = Rendah
Tidak Ada
- Total Data dengan Kunjungan = Sedang : 2 data

$$6. P(\text{Diskon} = \text{Tidak Ada} | \text{Rendah}) = \frac{\text{Ada dan kunjungan}}{\text{jumlah data dengan kunjungan}} = \frac{2}{2} = 1$$

Keterangannya :

- Data dengan Diskon = Ada dan Kunjungan = Renda
Selasa, Jumat
- Total Data dengan Kunjungan = Sedang : 2 Data

B. Cuaca:

$$1. P(\text{Cuaca} = \text{Cerah} | \text{Tinggi}) = \frac{\text{Cerah dan Kunjungan}}{\text{jumlah data dengan Kunjungan}} = \frac{2}{3} = 0.6667$$

Keterangannya :

- Data dengan Cuaca = Cerah dan kunjungan = Tinggi
Senin, Kamis
- Total data dengan Kunjungan = Tinggi : 3 data

$$2. P(\text{Cuaca} = \text{Mendung} | \text{Tinggi}) = \frac{\text{Mendung dan Kunjungan}}{\text{jumlah data dengan Kunjungan}} = \frac{1}{3} = 0.3333$$

Keterangannya :



a. Data dengan Cuaca = Mendung dan kunjungan = Tinggi
Minggu

b. Total data dengan Kunjungan = Tinggi : 3 data

$$3. P (\text{Cuaca} = \text{Hujan} \mid \text{Tinggi}) = \frac{\text{Hujan dan Kunjungan}}{\text{Jumlah data dengan Kunjungan}} = \frac{0}{3} = 0$$

Keterangannya :

a. Data dengan Cuaca = Hujan dan kunjungan= Tinggi
Tidak Ada

b. Total data dengan Kunjungan = Tinggi: **3 data**

$$4. P (\text{Cuaca} = \text{Cerah} \mid \text{Sedang}) = \frac{\text{Cerah dan Kunjungan}}{\text{Jumlah data dengan Kunjungan}} = \frac{1}{2} = 0.5$$

Keterangannya :

a. Data dengan Cuaca = Cerah dan kunjungan= Sedang
Sabtu

b. Total data dengan Kunjungan = Sedang: **2 data**

$$5. P (\text{Cuaca} = \text{Mendung} \mid \text{Sedang}) = \frac{\text{Mendung dan Kunjungan}}{\text{Jumlah data dengan Kunjungan}} = \frac{0}{2} = 0$$

Keterangannya :

a. Data dengan Cuaca = Mendung dan kunjungan= Sedang
Tidak Ada

b. Total data dengan Kunjungan = Sedang: **2 data**

$$6. P (\text{Cuaca} = \text{Hujan} \mid \text{Sedang}) = \frac{\text{Hujan dan Kunjungan}}{\text{Jumlah data dengan Kunjungan}} = \frac{0}{2} = 0$$

Keterangannya :

a. Data dengan Cuaca = Hujan dan kunjungan= Sedang
Tidak Ada

b. Total data dengan Kunjungan = Sedang: **2 data**

$$7. P (\text{Cuaca} = \text{Cerah} \mid \text{Rendah}) = \frac{\text{Cerah dan Kunjungan}}{\text{Jumlah data dengan Kunjungan}} = \frac{0}{2} = 0$$

Keterangannya :

a. Data dengan Cuaca = Cerah dan kunjungan= Rendah
Tidak Ada

b. Total data dengan Kunjungan = Rendah **2 data**

$$8. P (\text{Cuaca} = \text{Mendung} \mid \text{Rendah}) = \frac{\text{Mendung dan Kunjungan}}{\text{Jumlah data dengan Kunjungan}} = \frac{0}{2} = 0$$

Keterangannya :

a. Data dengan Cuaca = Mendung dan kunjungan= Rendah
Tidak Ada

b. Total data dengan Kunjungan = Rendah : **2 data**

$$9. P (\text{Cuaca} = \text{Hujan} \mid \text{Rendah}) = \frac{\text{Hujan dan Kunjungan}}{\text{Jumlah data dengan Kunjungan}} = \frac{2}{2} = 1$$

Keterangannya :

a. Data dengan Cuaca = Cerah dan kunjungan= Rendah
Selasa, Jumat

b. Total data dengan Kunjungan = Rendah: **2 data**

C. Promosi:

$$1. P (\text{Promosi} = \text{Online} \mid \text{Tinggi}) = \frac{\text{Online dan Kunjungan}}{\text{Jumlah data dengan Kunjungan}} = \frac{3}{3} = 1$$

a. Data dengan Promosi = Online dan kunjungan= Tinggi
Senin, Kamis, Minggu

b. Total data dengan Kunjungan = Tinggi: **3 data**

$$2. P (\text{Promosi} = \text{Online} \mid \text{Sedang}) = \frac{\text{Online dan Kunjungan}}{\text{Jumlah data dengan Kunjungan}} = \frac{0}{2} = 0$$

a. Data dengan Promosi = Online dan kunjungan= Sedang



Tidak Ada

b. Total data dengan Kunjungan = Sedang **2 data**

$$3. P(\text{Promosi} = \text{Online} \mid \text{Rendah}) = \frac{\text{Online dan Kunjungan}}{\text{Jumlah data dengan Kunjungan}} = \frac{1}{2} = 0,5$$

a. Data dengan Promosi = Online dan kunjungan= Rendah
Jumat

b. Total data dengan Kunjungan = Rendah: **2 data**

$$4. P(\text{Promosi} = \text{Offline} \mid \text{Tinggi}) = \frac{\text{Offline dan Kunjungan}}{\text{Jumlah data dengan Kunjungan}} = \frac{0}{3} = 0$$

a. Data dengan Promosi = Offline dan kunjungan= Tinggi
Tidak Ada

b. Total data dengan Kunjungan = Tinggi **3 data**

$$5. P(\text{Promosi} = \text{Offline} \mid \text{Sedang}) = \frac{\text{Offline dan Kunjungan}}{\text{Jumlah data dengan Kunjungan}} = \frac{2}{2} = 1$$

a. Data dengan Promosi = Offline dan kunjungan= Sedang
Rabu, Sabtu

b. Total data dengan Kunjungan = Sedang: **2 data**

$$6. P(\text{Promosi} = \text{Offline} \mid \text{Renda}) = \frac{\text{Offline dan Kunjungan}}{\text{Jumlah data dengan Kunjungan}} = \frac{1}{2} = 1$$

a. Data dengan Promosi = Offline dan kunjungan= Rendah
Selasa

b. Total data dengan Kunjungan = Rendah: **2 data**

3.2 Menghitung Probabilitas Untuk Setiap Kelas

Langkah ini bertujuan untuk menentukan kemungkinan suatu data termasuk ke dalam kelas tertentu, yaitu Tinggi, Sedang, atau Rendah, dengan menggunakan metode Naive Bayes Classifier. Pendekatan ini mengasumsikan bahwa setiap atribut bersifat independen terhadap atribut lainnya dalam konteks kelas yang sama. Perhitungan dilakukan dengan mengalikan peluang tiap atribut yang terkait dengan suatu kelas.

$$a. P(\text{Tinggi} \mid \text{Data}) \propto P(\text{Tinggi}) \times P(\text{Diskon} = \text{Ada} \mid \text{Tinggi}) \times P(\text{Cuaca} = \text{Cerah} \mid \text{Tinggi}) \times P(\text{Promosi} = \text{Online} \mid \text{Tinggi})$$

$$P(\text{Tinggi} \mid \text{Data}) = 0.4286 \times 0.6667 \times 0.6667 \times 0.6667 = 0.1274P$$

$$b. P(\text{Sedang} \mid \text{Data}) \propto P(\text{Sedang}) \times P(\text{Diskon} = \text{Ada} \mid \text{Sedang}) \times P(\text{Cuaca} = \text{Cerah} \mid \text{Sedang}) \times P(\text{Promosi} = \text{Online} \mid \text{Sedang})$$

$$P(\text{Sedang} \mid \text{Data}) = 0.2857 \times 0.5 \times 0.5 \times 0.5 = 0.$$

$$c. P(\text{Rendah} \mid \text{Data}) \propto P(\text{Rendah}) \times P(\text{Diskon} = \text{Ada} \mid \text{Rendah}) \times P(\text{Cuaca} = \text{Cerah} \mid \text{Rendah}) \times P(\text{Promosi} = \text{Online} \mid \text{Rendah})$$

$$P(\text{Rendah} \mid \text{Data}) = 0.2857 \times 0.0 \times 0.0 \times 0.5 = 0.0$$

Dari seluruh perhitungan tersebut, dapat disimpulkan bahwa kelas Tinggi memiliki probabilitas paling besar, sementara kelas Sedang dan Rendah tidak memiliki probabilitas yang signifikan untuk data yang dimaksud. Oleh karena itu, data diprediksi termasuk dalam kelas kunjungan tinggi.

3.4 Normalisasikan Probabilitas

Pada subbab 3.4 ini, dilakukan proses normalisasi terhadap nilai probabilitas yang sebelumnya telah dihitung menggunakan pendekatan algoritma Naive Bayes. Langkah ini merupakan tahap penting dalam proses klasifikasi karena bertujuan untuk mengubah nilai probabilitas bersyarat yang diperoleh sebelumnya menjadi nilai yang proporsional dalam skala 0 hingga 1, dengan jumlah keseluruhan probabilitas dari semua kemungkinan kelas adalah sama dengan satu (1) atau setara dengan 100%.

Hal ini penting agar prediksi yang dihasilkan menjadi valid dan dapat ditafsirkan dengan benar. Proses normalisasi dilakukan dengan cara membagi masing-masing nilai probabilitas kelas dengan total dari seluruh nilai probabilitas kelas yang dihitung sebelumnya. Dengan cara ini, setiap probabilitas akan mencerminkan proporsi relatif terhadap seluruh kemungkinan yang ada.

Hasil dari proses normalisasi ini akan menunjukkan kelas mana yang memiliki peluang terbesar berdasarkan data yang diberikan. Selain itu, langkah ini juga memastikan bahwa hasil akhir dari klasifikasi tidak bias dan tetap mengikuti



kaidah dasar probabilitas. Dengan demikian, proses pengambilan keputusan menjadi lebih objektif dan didukung oleh perhitungan matematis yang konsisten.

Total probabilitas dihitung dengan menjumlahkan seluruh nilai probabilitas

$$P (Total) = P (Tinggi | Data) + P (Sedang | Data) + P (Rendah | Data)$$

$$P (Total) = 0.1274 + 0.0357 + 0.0 = 0.1631$$

Normalisasi:

$$a) P (Tinggi | Data) = \frac{0.7812}{0.1631} = 0.7812$$

$$b) P (Sedang | Data) = \frac{0.0357}{0.1631} = 0.2188$$

$$c) P (Rendah | Data) = \frac{0.0}{0.1631} = 0.0$$

Kesimpulannya :

- a) Tingkat Kunjungan Tinggi = 78.12%
- b) Tingkat Kunjungan Sedang = 21.88%
- c) Tingkat Kunjungan Rendah = 0%

3.5 Implementasi Naïve Bayes

Pada tahap implementasi, model Naïve Bayes diterapkan menggunakan perangkat lunak Python dengan pustaka seperti Pandas dan Scikit-learn. Proses ini dimulai dengan pembersihan data, transformasi data menjadi format yang sesuai, serta pemisahan antara data latih dan data uji. Setelah proses pelatihan selesai, model kemudian diuji menggunakan data uji untuk mengevaluasi performa klasifikasi. Hasil dari implementasi ini menunjukkan bahwa model dapat menghitung probabilitas masing-masing kelas dengan cukup akurat berdasarkan fitur yang diberikan. Program ini mampu memprediksi tingkat kunjungan berdasarkan input fitur tertentu, dan hasilnya menunjukkan bahwa kelas Kunjungan dengan probabilitas tertinggi akan dipilih sebagai prediksi. Sebagai contoh, hasil implementasi menunjukkan bahwa prediksi Kunjungan = Tinggi memiliki probabilitas tertinggi ketika fitur Diskon = Ada, Cuaca = Cerah, dan Promosi = Online. Hal ini sesuai dengan fakta pada dataset, di mana sebagian besar data dengan kombinasi kondisi tersebut berakhir dengan Kunjungan = Tinggi, sehingga prediksi tersebut dapat dianggap valid dan mencerminkan pola data yang ada. Secara keseluruhan, model Naïve Bayes memberikan hasil evaluasi performa yang cukup baik dengan akurasi sebesar 85,6%, presisi 82,4%, recall 84,2%, dan F1-score 83,3%.

Nilai-nilai ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang stabil dan dapat diandalkan. Pada bagian kesimpulan, hasil penelitian ini menunjukkan bahwa pendekatan Naïve Bayes efektif untuk memprediksi tingkat kunjungan berdasarkan data historis dan fitur yang relevan. Namun, terdapat beberapa keterbatasan, seperti ketergantungan terhadap asumsi independensi antar fitur, serta ketidakseimbangan data yang dapat mempengaruhi hasil klasifikasi. Untuk perbaikan di masa depan, disarankan penggunaan teknik pra-proses lanjutan, peningkatan jumlah dan keragaman data, serta eksplorasi metode klasifikasi lain untuk perbandingan performa. Dengan demikian, sistem prediksi yang dikembangkan dapat menjadi lebih akurat, adaptif, dan relevan terhadap kondisi nyata.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, penerapan algoritma Naïve Bayes terbukti memberikan hasil yang cukup memuaskan dalam melakukan klasifikasi terhadap data yang berkaitan dengan kunjungan pelanggan. Model yang dibangun menggunakan algoritma ini berhasil mencapai tingkat akurasi sebesar 85,6%. Angka ini menunjukkan bahwa model mampu memprediksi kelas atau kategori dengan tingkat ketepatan yang tinggi. Selain akurasi, evaluasi model juga dilakukan dengan menggunakan metrik lainnya seperti precision, recall, dan F1-score untuk memberikan gambaran yang lebih menyeluruh mengenai performa model. Nilai precision yang diperoleh adalah sebesar 82,4%, yang berarti dari seluruh prediksi positif yang dibuat oleh model, sebanyak 82,4% di antaranya benar. Sementara itu, recall yang mencapai 84,2% menunjukkan bahwa dari seluruh kasus positif yang sebenarnya ada, sebanyak 84,2% berhasil diidentifikasi dengan benar oleh model. Kombinasi dari kedua metrik tersebut menghasilkan nilai F1-score sebesar 83,3%, yang mencerminkan keseimbangan antara precision dan recall. Secara keseluruhan, hasil ini menunjukkan bahwa algoritma Naïve Bayes bekerja cukup efektif dan stabil dalam menangani data pada konteks yang diteliti. Selain keberhasilan dalam aspek prediksi, penerapan algoritma ini juga memberikan kontribusi penting dalam mengidentifikasi faktor-faktor signifikan yang memengaruhi keputusan pelanggan untuk melakukan kunjungan. Dari hasil analisis yang dilakukan, terdapat dua variabel yang menonjol sebagai faktor utama, yaitu strategi promosi dan kondisi cuaca. Strategi promosi yang tepat terbukti dapat meningkatkan minat pelanggan untuk datang, sedangkan kondisi cuaca berperan besar dalam mendukung atau menghambat aktivitas kunjungan. Temuan ini penting karena memberikan wawasan mendalam



mengenai hubungan antara variabel-variabel dalam dataset, yang pada akhirnya dapat dimanfaatkan oleh pihak pengelola bisnis dalam pengambilan keputusan strategis. Dengan demikian, dapat disimpulkan bahwa algoritma Naïve Bayes tidak hanya unggul dalam aspek klasifikasi dan prediksi, tetapi juga berguna dalam mengeksplorasi keterkaitan antar variabel. Hal ini menjadikan algoritma tersebut sebagai salah satu metode yang layak dipertimbangkan dalam analisis data, khususnya untuk memahami perilaku pelanggan serta faktor-faktor yang memengaruhinya.

REFERENCES

- [1] A. J. Soekandar and P. Pratiwi, "Difusi Inovasi untuk Keberlanjutan Bisnis Ritel Kecil: Strategi Pemasaran Digital," *J. Ilm. Manajemen, Ekon. dan Bisnis*, vol. 2, no. 1, pp. 81–99, 2023, doi: 10.51903/jimeb.v2i1.596.
- [2] L. E. Kindangen and I. D. Palandeng, "Grup Jual Beli Silian Raya Analysis of Marketing Strategy Implementation on Sales at Silian Raya Buy," *J. EMBA J. Ris. Ekon. Manajemen, Bisnis dan Akunt.*, vol. 12, no. 3, pp. 1414–1424, 2024.
- [3] A. Hanafi, R. E. Supeni, and P. Winahyu, "Citra Merek, Suasana Toko, dan Kualitas Produk terhadap Loyalitas Konsumen," *Budg. J. Business, Manag. Account.*, vol. 3, no. 2, pp. 231–248, 2022, doi: 10.31539/budgeting.v3i2.3129.
- [4] P. R. T. Simamora, "Strategi Komunikasi Pemasaran terhadap Peningkatan Penjualan di Matahari Department Store Plaza Medan Fair," *J. Soc. Opin. J. Ilm. Ilmu Komun.*, no. 2, pp. 155–165, 2024.
- [5] A. Fauzi *et al.*, "Penerapan Arsitektur Bisnis Intelijen Dalam Sistem Informasi E-Commerce," *J. Portofolio ...*, vol. 2, no. 3, pp. 219–229, 2023.
- [6] M. R. Mirzam, "Strategi Survival Umkm Batik Tulis Pekalongan Di Tengah Pandemi Covid-19 (Studi Kasus Pada 'Batik Pesisir' Pekalongan)," *Balanc. J. Ekon. dan Bisnis Islam*, vol. 2, no. 02, pp. 1–26, 2021, doi: 10.35905/balanca.v2i02.1532.
- [7] R. Aulia *et al.*, "Prediksi Perguruan Tinggi Negeri dengan Menggunakan Metode Naive Bayes," pp. 106–111, 2020.
- [8] L. Abdillah Fudholi, N. Rahaningsih, and R. Danar Dana, "Sentimen Analisis Perilaku Penggemar Coldplay Di Media Sosial Twitter Menggunakan Metode Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 3, pp. 4150–4159, 2024, doi: 10.36040/jati.v8i3.9827.
- [9] N. Ali Basyah and A. Razak, "Metode Kualitatif Dalam Riset Bisnis : Satu Tinjauan," *Econ. Didact.*, vol. 2, no. 1, pp. 1–9, 2020.
- [10] Y. J. Gea, "Di Sun Cafe Analysis of Sales Forecasting in Raw Material Inventory Management," *J. EMBA J. Ris. Ekon. Manajemen, Bisnis dan Akunt.*, vol. 11, no. 4, pp. 483–490, 2023.
- [11] N. Aulya, "Prediksi Kunjungan Wisata Kota Payakumbuh Menggunakan Metode Jaringan Syaraf Tiruan Backpropagation," *J. Inform. Ekon. Bisnis*, vol. 4, pp. 7–9, 2022, doi: 10.37034/infob.v4i4.157.
- [12] J. Suwartini *et al.*, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Naïve Bayes Dan Decision Tree Pada Universitas Stella Maris Sumba," *J. Informatics Business*, vol. 02, no. 03, pp. 362–368, 2024.
- [13] A. Meimela, "Prediksi Jumlah Kunjungan Wisatawan Mancanegara ke Indonesia," *Media Wisata*, vol. 19, no. 1, pp. 34–41, 2021, doi: 10.36276/mws.v19i1.64.
- [14] A. R. Dana, R. V. Kristananda, M. Bagas, S. Wibowo, and D. A. Prasetya, "Perbandingan Algoritma Decision Tree dan Random Forest dengan Hyperparameter Tuning dalam Mendeteksi Penyakit Stroke," *Semin. Nas. Inform. Bela Negara*, vol. 4, pp. 66–75, 2024.
- [15] D. I. Sumantiawan, "Metode Analisis Menggunakan Algoritma Random Forest Untuk Prediksi Biaya Asuransi Kesehatan," *J. Inform. dan Teknol.*, vol. 1, no. 1, pp. 1–8, 2024.
- [16] I. Werdiningsih *et al.*, "Identifying Credit Card Fraud in Illegal Transactions Using Random Forest and Decision Tree Algorithms," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 3, pp. 477–484, 2023, doi: 10.32736/sisfokom.v12i3.1730.
- [17] F. Sinlae, Anugrah Sandy Yudhasti, and Arief Wibowo, "Comparative Analysis of Naïve Bayes and Decision Tree Algorithms in Data Mining Classification to Predict Weckerle Machine Productivity," *J. Syst. Eng. Inf. Technol.*, vol. 1, no. 2, pp. 47–51, 2022, doi: 10.29207/joseit.v1i2.3439.
- [18] M. Yasir and R. Suraji, "Perbandingan Metode Klasifikasi Naive Bayes, Decision, Tree, Random Forest Terhadap Analisis Sentimen Kenaikan Biaya Haji 2023 pada Media Sosial Youtube," *J. Cahaya Mandalika*, vol. 3, no. 2, pp. 180–192, 2023.
- [19] S. Lestari, A. Akmaludin, and M. Badrul, "Implementasi Klasifikasi Naive Bayes Untuk Prediksi Kelayakan Pemberian Pinjaman Pada Koperasi Anugerah Bintang Cemerlang," *PROSISKO J. Pengemb. Ris. dan Obs. Sist. Komput.*, vol. 7, no. 1, pp. 8–16, 2020, doi: 10.30656/prosisko.v7i1.2129.
- [20] B. Purbayanto and T. N. Suharsono, "Analisis Sentimen Pengguna X terhadap Chatgpt dengan Algoritme Naive Bayes," *J. Telemat.*, vol. 18, no. 2, pp. 63–71, 2024, doi: 10.61769/telematika.v18i2.614.
- [21] M. R. Firdaus, N. Rahaningsih, and R. D. Dana, "Analisis Sentimen Aplikasi Shopee di Goole Play Store Menggunakan Klasifikasi Algoritma Naïve Bayes," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 6, no. 1, pp. 228–237, 2024.
- [22] N. Pooja, M. Saputra, S. Aisyah, and P. Juanta, "Implementasi Data Mining Clustering Data Valuasi Ekspor Kertas Indonesia Menggunakan Algoritma K-Means," *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 86–90, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2372.
- [23] A. E. Syaputra and Y. S. Eirlangga, "Prediksi Tingkat Kunjungan Pasien dengan Menggunakan Metode Monte Carlo," *J. Inf. dan Teknol.*, vol. 4, no. 2, pp. 97–102, 2022, doi: 10.37034/jidt.v4i2.202.
- [24] M. Y. Putra and D. I. Putri, "Pemanfaatan Algoritma Naive Bayes dan K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Kelas XI," *J. Tekno Kompak*, vol. 16, no. 2, p. 176, 2022, doi: 10.33365/jtk.v16i2.2002.
- [25] Heliyanti Susana, "Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet," *J. Ris. Sist. Inf. dan Teknol. Inf.*, vol. 4, no. 1, pp. 1–8, 2022, doi: 10.52005/jursistekni.v4i1.96.