



## **KLASIFIKASI SENTIMEN PELANGGAN RESTORAN KOKI SUNDA MENGUNAKAN ALGORITMA NAIVE BAYES**

**Dedi Irawan<sup>1</sup>, Yayu Sri Rahayu<sup>2</sup>, Mohamad Farozi<sup>3</sup>**

<sup>1</sup> Program Studi Teknik Informatika, Universitas Deztron Indonesia · <sup>2</sup> Program Studi Teknik Komputer, Universitas Muhammadiyah Karanganyar, <sup>3</sup> Program Studi Sistem Informasi, Universitas Bina Darma

Jln. Perintis Kemerdekaan No.9, Perintis, Kecamatan. Medan Timur., Kota Medan, Sumatera Utara

<sup>1</sup> [dediirawan8223@gmail.com](mailto:dediirawan8223@gmail.com), <sup>2</sup> [rahayupink2024@gmail.com](mailto:rahayupink2024@gmail.com) <sup>3</sup> [mohamad.farozi@binadarma.ac.id](mailto:mohamad.farozi@binadarma.ac.id)

### **Abstract**

*Koki Sunda is a culinary business that has received numerous customer reviews online. These reviews reflect consumer experiences; however, there is no automated system to analyze these opinions efficiently. This study aims to classify customer sentiment toward food at Koki Sunda Restaurant using the Naïve Bayes algorithm. Data were obtained through web scraping techniques from online review platforms, as well as through direct surveys. The pre-processing process includes text cleaning, normalization, removal of stopwords, stemming, and tokenization. Features were extracted using the TF-IDF method, and the data was divided into training data (80%) and test data (20%). The evaluation results show that the Multinomial Naïve Bayes model has good performance, achieving 88.33% accuracy, 87.10% precision, 90.00% recall, and 88.54% F1-score. These findings indicate that the algorithm is effective for automatically identifying customer opinions. The developed classification system has the potential to help restaurant managers better understand customer needs and improve service quality based on live review data..*

**Keywords:** *Customer Sentiment, Naive Bayes, TF-IDF, Koki sunda Restaurant, Text Classification*

### **Abstrak**

Koki Sunda adalah bisnis kuliner yang telah menerima banyak ulasan pelanggan secara online. Ulasan-ulasan ini mencerminkan pengalaman konsumen; namun, tidak ada sistem otomatis untuk menganalisis opini-opini ini secara efisien. Penelitian ini bertujuan untuk mengklasifikasikan sentimen pelanggan terhadap makanan di Restoran Koki Sunda dengan menggunakan algoritma Naïve Bayes. Data diperoleh melalui teknik web scraping dari platform ulasan online, serta melalui survei langsung. Proses pra-pemrosesan meliputi pembersihan teks, normalisasi, penghilangan stopwords, stemming, dan tokenisasi. Fitur diekstraksi menggunakan metode TF-IDF, dan data dibagi menjadi data latih (80%) dan data uji (20%). Hasil evaluasi menunjukkan bahwa model Multinomial Naïve Bayes memiliki kinerja yang baik, mencapai akurasi 88,33%, presisi 87,10%, recall 90,00%, dan F1-score 88,54%. Temuan ini menunjukkan bahwa algoritma ini efektif untuk mengidentifikasi opini pelanggan secara otomatis. Sistem klasifikasi yang dikembangkan memiliki potensi untuk membantu manajer restoran lebih memahami kebutuhan pelanggan dan meningkatkan kualitas layanan berdasarkan data ulasan langsung.

**Kata kunci :** *Sentimen Pelanggan, Naive Bayes, TF-IDF, Restoran Koki sunda, Klasifikasi Teks*



## 1. PENDAHULUAN

Pola interaksi antara pelanggan dan bisnis sedang mengalami transformasi yang signifikan di era digitalisasi yang cepat. Salah satu perubahan terpenting yang terjadi di industri kuliner adalah penggunaan ulasan pelanggan secara online sebagai alat penting untuk menilai kualitas restoran atau produk makanan. Pelanggan dapat mengekspresikan pendapat mereka secara terbuka dan real-time dengan adanya platform digital seperti Google Review, TripAdvisor, Facebook, dan Instagram. Fenomena ini telah menghasilkan sumber informasi yang sangat berharga bagi pemilik bisnis untuk memahami bagaimana konsumen memandang barang dan jasa yang mereka tawarkan [1].

Restoran Koki Sunda merupakan salah satu bisnis kuliner lokal yang menyajikan makanan khas daerah Sunda. Dengan popularitasnya yang terus meningkat, restoran ini secara teratur menerima berbagai macam ulasan pelanggan yang mencakup aspek rasa makanan, harga, layanan, dan suasana. Ulasan-ulasan ini tidak hanya mencerminkan kepuasan atau ketidakpuasan pelanggan, tetapi juga menyimpan informasi mendalam tentang preferensi konsumen. Sayangnya, data ulasan teks yang tidak terstruktur sering kali rumit untuk dianalisis secara manual karena banyaknya jumlah dan variasi bahasa yang digunakan. Proses menganalisis data semacam ini secara manual memakan waktu dan tenaga serta sangat rentan terhadap subjektivitas [2].

Tantangan lain dalam menafsirkan ulasan pelanggan adalah ketidaksesuaian antara peringkat numerik (seperti 1 hingga 5 bintang) dan konten naratif. Sebagai contoh, seorang pelanggan mungkin memberikan peringkat bintang 4 tetapi menyisipkan keluhan dalam teks ulasan. Sebaliknya, beberapa pelanggan memberikan peringkat rendah tetapi menulis komentar positif. Hal ini menunjukkan bahwa analisis berbasis teks atau narasi diperlukan untuk menangkap esensi sebenarnya dari persepsi pelanggan, yang tidak selalu tercermin dalam peringkat numerik [3].

Untuk menjawab tantangan tersebut, pendekatan berbasis teknologi seperti text mining dan analisis sentimen menjadi sangat relevan. Analisis sentimen merupakan komponen dari data mining yang mengkategorikan opini ke dalam kategori positif, negatif, atau netral berdasarkan isi

teks. Dengan pendekatan ini, sistem dapat secara otomatis mengenali emosi atau sikap seseorang terhadap sebuah entitas, dalam hal ini restoran dan produknya. Teknologi ini dapat membantu bisnis seperti Restoran Koki Sunda untuk memahami opini pelanggan secara lebih sistematis dan obyektif, sehingga dapat mengambil keputusan yang lebih tepat dan strategis [4].

Salah satu metode yang banyak digunakan dalam analisis sentimen adalah algoritma Naïve Bayes. Algoritma ini memiliki keunggulan dalam hal kecepatan, efisiensi, dan kemudahan implementasi, terutama dalam memproses data teks. Naïve Bayes beroperasi dengan prinsip probabilitas bersyarat untuk menentukan kemungkinan suatu teks termasuk ke dalam kategori tertentu (misalnya, positif atau negatif) berdasarkan kata-kata yang terdapat di dalam teks tersebut. Dalam konteks klasifikasi sentimen, algoritma ini telah terbukti memberikan hasil yang akurat meskipun beroperasi di bawah asumsi independensi antar fitur (kata) [5].

Penelitian ini dilakukan untuk mengembangkan model klasifikasi sentimen pelanggan terhadap makanan di Restoran Koki Sunda dengan menggunakan algoritma Naïve Bayes. Data ulasan dikumpulkan melalui dua metode: web scraping dari platform digital dan survei langsung terhadap pelanggan restoran. Setelah dilakukan preprocessing data, termasuk pembersihan teks, tokenisasi, penghilangan stopword, dan stemming, fitur-fitur teks diekstraksi dengan metode TF-IDF. Selanjutnya, data dibagi menjadi dua bagian: data latih dan data uji. Pelatihan model kemudian dilakukan dengan menggunakan pendekatan Multinomial Naïve Bayes, yang cocok untuk data berbasis frekuensi kata [6].

Dengan memanfaatkan teknologi klasifikasi teks berbasis machine learning ini, restoran dapat memperoleh gambaran yang lebih akurat mengenai kepuasan pelanggan secara otomatis. Selain itu, hasil penelitian ini juga memberikan kontribusi akademis terhadap pengembangan metode klasifikasi teks bahasa Indonesia di bidang kuliner yang selama ini masih relatif terbatas [7].

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini dapat dirangkum sebagai berikut:

1. Bagaimana mengotomatisasi klasifikasi sentimen pelanggan terhadap makanan di Restoran Koki Sunda dengan algoritma Naïve Bayes?
2. Seberapa akurat dan efektif model dalam membedakan sentimen positif dan negatif dari ulasan pelanggan?



3. Faktor-faktor ulasan apa saja yang paling dominan memengaruhi hasil klasifikasi sentimen?

Melalui penelitian ini, sebuah sistem dapat dikembangkan untuk mendukung pengambilan keputusan manajemen berbasis data dan memberikan landasan ilmiah untuk penelitian lebih lanjut di bidang analisis sentimen dan klasifikasi teks bahasa Indonesia, khususnya dalam konteks bisnis kuliner lokal [8].

## 2. TINJAUAN PUSTAKA

### 2.1 Teori Sentimen

Analisis sentimen adalah proses sistematis untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini atau perasaan dalam sebuah teks, baik positif, negatif, maupun netral. Dalam konteks bisnis, khususnya di industri kuliner, seperti Restoran Koki Sunda, analisis sentimen dapat digunakan untuk mengetahui kepuasan konsumen terhadap kualitas makanan.

Sentimen pelanggan biasanya muncul dalam bentuk ulasan yang ditulis secara alami oleh konsumen di berbagai platform digital. Setiap ulasan mencerminkan pengalaman subjektif pelanggan yang dapat dianalisis untuk mengevaluasi produk atau layanan secara keseluruhan. Analisis sentimen dapat dilakukan pada tingkat dokumen, kalimat, atau aspek, tergantung pada tujuan akhir analisis. Pada penelitian ini, pendekatan yang digunakan adalah klasifikasi sentimen pada level dokumen ulasan pelanggan [9].

Beberapa pendekatan dalam analisis sentimen telah diterapkan dalam berbagai domain. Misalnya, penelitian oleh Elisa et al. (2022) menunjukkan tingkat akurasi tinggi (88%) pada data penjualan UMKM menggunakan algoritma Naïve Bayes. Meskipun demikian, studi tersebut berfokus pada klasifikasi produk, bukan opini berbasis teks [10].

Studi oleh Aji et al. (2023) menggunakan data mahasiswa dan mencapai akurasi 85%. Namun, domain pendidikan memiliki struktur data dan bahasa yang berbeda dibanding ulasan pelanggan restoran yang lebih variatif dan informal [11].

Lanjut dengan Salsha & Kurniawan (2024) meneliti ulasan McDonald's dengan akurasi 90%, namun tidak menyertakan sentimen netral. Penelitian kami lebih spesifik pada restoran lokal dengan data Bahasa Indonesia serta

mempertimbangkan distribusi label yang seimbang (positif dan negatif), yang menjadikannya relevan untuk pengambilan keputusan praktis di bisnis kecil menengah [12].

Dengan demikian, dibandingkan studi sebelumnya yang banyak berfokus pada e-commerce atau pendidikan, penelitian ini memperluas penerapan Naïve Bayes ke dalam ranah kuliner lokal, dengan pendekatan berbasis teks ulasan Bahasa Indonesia secara langsung [13].

### 2.2. Naive Bayes

Naïve Bayes adalah metode klasifikasi yang sering digunakan dalam analisis teks dan pemrosesan bahasa. Metode ini bergantung pada prinsip Teorema Bayes dengan asumsi bahwa setiap kata dalam teks tidak bergantung pada kata lainnya. Meskipun asumsi ini tidak selalu sesuai dengan kenyataan, pendekatan ini masih praktis, terutama ketika menggunakan representasi berbasis frekuensi kata seperti TF-IDF. Rumus dasarnya dituliskan sebagai berikut [14]:

$$P(C | X) = \frac{P(X | C) \cdot P(C)}{P(X)} \quad (1)$$

Keterangan:

- $P(C|X)$ : Probabilitas dokumen X termasuk dalam kelas C (misalnya positif atau negatif),
- $P(X|C)$ : Probabilitas fitur X muncul dalam kelas C,
- $P(C)$ : Probabilitas awal suatu kelas (prior),
- $P(X)$ : Probabilitas keseluruhan fitur X.

Namun dalam konteks klasifikasi teks, nilai  $P(X)$  konstan untuk semua kelas sehingga dapat diabaikan, dan rumus disederhanakan menjadi:

$$P(C | X) \propto P(C) \cdot \prod_{i=1}^n P(x_i | C) \quad (2)$$

Keterangan :

- $x_i$ : Kata ke -i dalam dokumen
- $n$ : Jumlah total kata dalam dokumen
- $P(x_i | C)$  Probabilitas kata  $x_i$  muncul dalam kelas C,
- $\prod$ : Perkalian dari seluruh probabilitas kata dalam dokumen

Naïve Bayes memiliki beberapa keunggulan yang membuatnya populer dalam tugas

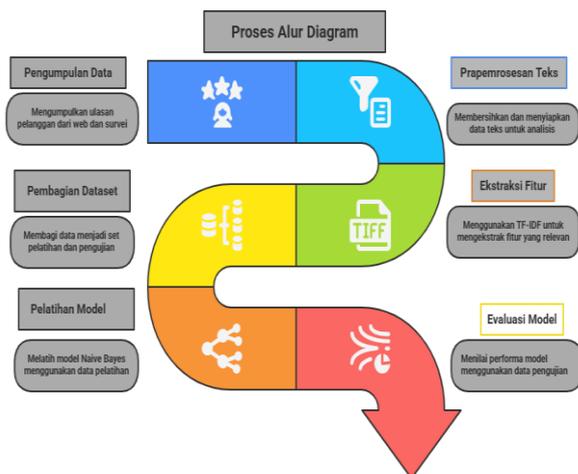
klasifikasi, terutama pada data teks. Naïve Bayes dikenal dengan kesederhanaan dan kecepatannya dalam proses pelatihan dan prediksi, bahkan ketika diterapkan pada kumpulan data yang besar. Algoritma ini bekerja dengan sangat baik dalam klasifikasi dokumen berbasis kata, seperti analisis sentimen pelanggan, dan relatif tahan terhadap overfitting dibandingkan dengan model yang lebih kompleks [15].

Meskipun cukup efektif, algoritma Naïve Bayes memiliki beberapa kelemahan. Asumsi independensi antar fitur sering kali tidak mencerminkan struktur bahasa alami, di mana kata-kata saling berhubungan. Selain itu, algoritma ini dapat mengalami masalah frekuensi nol jika sebuah kata tidak muncul dalam data pelatihan untuk kelas tertentu, yang biasanya diatasi dengan teknik seperti penghalusan Laplace. Naïve Bayes juga kurang mampu mengenali konteks emosional yang kompleks, seperti ironi atau sarkasme, karena tidak secara menyeluruh mempertimbangkan hubungan antar kata [16].

### 3. METODOLOGI PENELITIAN

#### 3.1. Tahapan Penelitian

Tahapan penelitian secara umum dapat dijelaskan melalui diagram alir pada Gambar 1 sebagai berikut:



**Gambar 1.** Alur diagram penelitian

#### 3.2. Teknik Pengumpulan Data

Data dalam penelitian ini diperoleh dari dua sumber utama:

##### a. Metode Web Scraping

Teknik ini digunakan untuk mengambil data ulasan dari platform online, seperti Google Review, Facebook Page, dan Instagram resmi Restoran Koki Sunda. Ulasan yang dipertimbangkan hanya ulasan yang memiliki konten teks dan berasal dari akun publik.

##### b. Metode Survei Langsung

Selain web scraping, penulis juga membagikan kuesioner singkat kepada pelanggan yang mengunjungi restoran untuk memberikan ulasan dalam bentuk komentar dan penilaian. Tujuan dari survei ini adalah untuk mengurangi variasi data dan meningkatkan validitas dataset.

#### 3.3. Prapemrosesan Data

Data yang diperoleh umumnya berupa teks bebas yang tidak terstruktur. Oleh karena itu, beberapa tahap preprocessing dilakukan untuk mempersiapkan data menjadi input yang dapat diproses oleh algoritma klasifikasi

- Pembersihan teks: Menghapus karakter tidak relevan seperti tanda baca berlebih, angka, emotikon, dan URL
- Tokenizing: Memecah kalimat ulasan menjadi unit kata atau token
- Penghapusan Stopwords: Menghilangkan kata-kata umum yang tidak memiliki makna sentimen, seperti "dan", "yang", "itu".
- Pembendaan (stemming): Mengubah kata ke bentuk dasar menggunakan pustaka Sastrawi untuk Bahasa Indonesia.
- Normalisasi: Menstandarkan kata-kata tidak baku atau slang menjadi bentuk formal

#### 3.4. Pelabelan dan Pembagian Dataset

Setiap ulasan yang terkumpul dilabeli secara manual ke dalam dua kategori sentimen -positif dan negatif- berdasarkan makna keseluruhan dari konten teks. Sentimen netral tidak digunakan dalam penelitian ini karena dianggap tidak cukup representatif untuk pengambilan keputusan manajerial. Setelah proses pelabelan, dataset yang terdiri dari 300 ulasan dibagi menggunakan teknik train-test split dari pustaka Scikit-learn, dengan proporsi 80% data latih (240



ulasan) dan 20% data uji (60 ulasan). Pembagian ini dilakukan secara acak untuk menjaga distribusi label yang seimbang [17].

### 3.5. Ekstraksi Fitur dan Pemodelan

Ekstraksi fitur dari data teks dilakukan dengan menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency), yang menilai pentingnya sebuah kata dalam sebuah dokumen dibandingkan dengan seluruh kumpulan dokumen. Pendekatan ini membantu menekan pengaruh kata-kata umum dan menyoroti kata-kata yang lebih relevan dengan sentimen. Algoritma ini menghitung probabilitas setiap kelas, baik positif maupun negatif, dan memberikan label berdasarkan probabilitas tertinggi yang diperoleh.

### 3.6. Pelatihan dan Evaluasi Model

Model yang digunakan dalam penelitian ini adalah Multinomial Naïve Bayes yang cocok untuk data berbasis frekuensi kata seperti hasil dari metode TF-IDF. Setelah proses pelatihan model, evaluasi dilakukan dengan menggunakan beberapa metrik, antara lain Confusion Matrix, akurasi, presisi, recall, dan F1-score. Metrik-metrik ini menilai kemampuan model untuk mengidentifikasi sentimen pelanggan secara otomatis dan tepat.

## 4. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengklasifikasikan sentimen pelanggan terhadap Restoran Koki Sunda berdasarkan ulasan online dengan menggunakan algoritma Multinomial Naïve Bayes. Total dataset yang digunakan terdiri dari 300 ulasan yang dibagi menjadi 150 ulasan positif dan 150 ulasan negatif. Data diperoleh melalui metode web scraping dan survei langsung. Setelah dilakukan pra-pemrosesan dan ekstraksi fitur TF-IDF, dilakukan pelatihan model dan evaluasi.

### 4.1. Data Ulasan Pelanggan

Contoh data ulasan yang digunakan dalam proses klasifikasi pada tabel 1 sebagai berikut:

**Tabel 1.** Data Ulasan Pelanggan

No	Ulasan Pelanggan	Sumber
1	Makanannya sangat enak, dan porsinya pas	Google Review
2	Pelayanannya lama dan makanan kurang hangat	Google Review
3	Tempatnya nyaman, tapi makanannya biasa saja	Google Review
4	Sambal khasnya luar biasa, saya akan datang lagi	Google Review
5	Harga mahal dan tidak sesuai dengan rasa.	Google Review
6	Porsinya besar dan rasanya mantap!	Google Review

Tabel 1 menerangkan Ulasan tersebut mencerminkan variasi opini dari pelanggan, baik dalam bentuk pujian maupun keluhan.

### 4.2. Proses Pembersihan Teks

Pembersihan dilakukan untuk menghapus karakter tidak penting seperti angka, simbol, dan tanda baca berlebih yang tidak relevan untuk analisis sentimen.

**Tabel 2.** Proses Pembersihan Teks

No	Teks Asli	Pembersihan
1	Makanannya sangat enak, dan porsinya pas	Makanannya sangat enak dan porsinya pas
2	Pelayanannya lama dan makanan kurang hangat	Pelayanannya lama dan makanan kurang hangat
6	Porsinya besar dan rasanya mantap!	Porsinya besar dan rasanya mantap

Tabel 2 menerangkan teks dibersihkan dari tanda baca seperti titik, koma, dan tanda seru (!) serta karakter yang tidak mendukung proses klasifikasi.

#### 4.2.1 Tokenisasi

Tokenisasi adalah proses memecah kalimat menjadi unit kata (token), sehingga dapat diolah lebih lanjut.



**Tabel 3.** Hasil Tokenisasi

No	Teks Dibersihkan	Tokenisasi
1	Makanannya sangat enak dan porsinya pas	[makanannya, sangat, enak, dan, porsinya, pas]
2	Pelayanannya lama dan makanan kurang hangat	[pelayanannya, lama, dan, makanan, kurang, hangat]
3	Harga mahal dan tidak sesuai dengan rasa	["Harga", "mahal", "dan", "tidak", "sesuai", "dengan", "rasa"]

Tabel 3 menerangkan tokenisasi yang bertujuan mengisolasi kata agar bisa dianalisis frekuensi dan bobot kemunculannya

**4.2.2 Penghapusan Stopwords**

Stopwords adalah kata umum seperti "yang", "dan", "itu" yang tidak berkontribusi terhadap makna sentimen. Kata-kata ini dihapus agar fokus pada kata-kata bermakna kuat.

**Tabel 4.** Penghapusan Stopwords

No	Sebelum Stopwords Removal	Setelah Stopwords Removal
1	[makanannya, sangat, enak, dan, porsinya, pas]	[makanannya, enak, porsinya, pas]
2	[pelayanannya, lama, dan, makanan, kurang, hangat]	[pelayanannya, lama, dan, makanan, kurang, hangat]

Tabel 4 menerangkan penghapusan kata sambung yang tidak memberi nilai sentimen membuat analisis lebih fokus pada opini utama.

**4.2.3 Stemming (Pembendaan)**

Stemming mengubah kata ke bentuk dasar. Bahasa Indonesia menggunakan library Sastrawi untuk proses ini.

**Tabel 5.** Hasil Stemming

No	Sebelum Stemming	Setelah Stemming
1	["Makanannya", "enak", "porsinya", "pas"]	["makan", "enak", "porsi", "pas"]
2	["Pelayanannya", "lama", "makanan", "kurang", "hangat"]	["layan", "lama", "makan", "kurang", "hangat"]

Tabel 5 menerangkan transformasi "Makanannya" menjadi "makan, dan "Porsinya" menjadi "porsi"

**4.2.4 Normalisasi**

Normalisasi adalah mengubah kata tidak baku menjadi kata baku. Dalam ulasan informal, sering muncul kata seperti "bgt", "gk", "maknyus".

**Tabel 6.** Contoh Normalisasi

No	Kata Tidak Baku	Hasil Normalisasi
1	"bgt"	"banget"
2	"maknyus"	"enak"
3	"gk" / "ga"	"tidak"
4	"pgn"	"ingin"

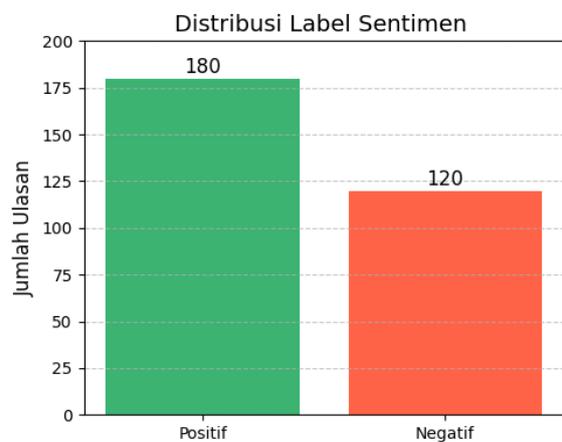
Tabel 6 menjelaskan tidak semua data memiliki kata tidak baku, tetapi normalisasi tetap penting karena data ulasan digital sering menggunakan slang.

**4.3 Pelabelan Sentimen**

**Tabel 7.** Pelabelan Sentimen

No	Ulasan Pelanggan	Label Asli
1	Makanannya sangat enak dan porsinya pas	Positif
2	Pelayanannya lama dan makanan kurang hangat	Negatif
3	Tempatnya nyaman, tapi makanannya biasa saja	Negatif
4	Sambal khasnya luar biasa, saya akan datang lagi	Positif
5	Harga mahal dan tidak sesuai dengan rasa.	Negatif
6	Porsinya besar dan rasanya mantap!	Negatif

Tabel 7 menerangkan proses pelabelan untuk mengklasifikasikan data ke dalam dua kategori (positif atau negatif) untuk memudahkan pelatihan model klasifikasi sentimen. Sentimen netral diabaikan karena tidak mendukung pengambilan keputusan manajerial.



Gambar 2 . Pelabelan Sentimen

#### 4.4 Ekstraksi Fitur

Setelah pra-pemrosesan, setiap ulasan pelanggan diubah menjadi representasi numerik menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). Teknik ini mengevaluasi pentingnya sebuah kata dalam satu dokumen dibandingkan dengan seluruh koleksi dokumen, sehingga menyortir kata kunci dan mengabaikan kata yang terlalu umum. Rumus TF-IDF sebagai berikut:

a. TF (Term Frequency)

Mengukur seberapa sering kata  $t$  muncul dalam

$$TF(t,d) = \frac{f_{t,d}}{\sum_k f_{k,d}} \quad (3)$$

Keterangan:

a)  $f_{t,d}$ : jumlah kemunculan kata  $t$  dalam dokumen  $d$

b)  $\sum_k f_{k,d}$ : jumlah seluruh kata dalam dokumen  $d$

b. IDF (Inverse Document Frequency)

Mengukur seberapa unik kata tersebut dalam seluruh korpus dokumen.

$$IDF(t) = \log \left( \frac{N}{df(t)} \right) \quad (4)$$

Keterangan:

a)  $N$ : jumlah total dokumen

b)  $df(t)$ : jumlah dokumen yang mengandung kata  $t$

c. TF-IDF Score

$$TF-IDF(t,d) = TF(t,d) \times IDF(t) \quad (5)$$

Keterangan :

a)  $TF-IDF(t,d)$  : Nilai bobot TF-IDF untuk kata  $t$  dalam dokumen  $d$ , digunakan untuk mengukur seberapa penting kata tersebut dalam dokumen.

b)  $TF(t,d)$ : Frekuensi relatif kemunculan kata  $t$  dalam dokumen  $d$  (Term Frequency).

c)  $IDF(t)$ : Ukuran seberapa jarang atau unik kata  $t$  dalam seluruh koleksi dokumen (Inverse Document Frequency). Kata yang jarang muncul dalam banyak dokumen akan memiliki nilai IDF yang tinggi.

Berikut contoh perhitungan TF-IDF yang terdapat 5 dokumen (D1–D5), dan kata yang akan dianalisis adalah: makan, enak, tidak, mahal, rasa.

Dimana:

a)  $N=5$

b) makan muncul di 3 dokumen  
 $IDF = \log(5/3) \approx 0,222$

c) enak muncul di 1 dokumen  $\Rightarrow IDF = \log(5/1) \approx 0,699$

d) tidak, mahal, rasa hanya di D5  $\Rightarrow IDF$  sama:  $\log(5/1) \approx 0,699$

Tabel 8. TF-IDF Hasil perhitungan

Ulasan	TF (makan)	TF (enak)	TF-IDF (makan)	TF-IDF (enak)
D1	0,25	0,25	$0.25 \times 0.222 = 0.056$	$0.25 \times 0.699 = 0.175$
D2	0,20	0	$0.20 \times 0.222 = 0.044$	0
D5	0	0	0	0

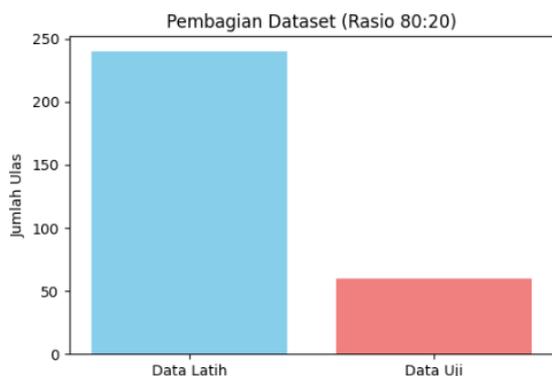
Tabel 8 menerangkan kata “enak” memiliki bobot yang tinggi pada dokumen sehingga dapat dianggap sebagai indikator sentimen positif. Sebaliknya, kata tidak, mahal, dan rasa hanya muncul di D5 dan menghasilkan skor tinggi di sana, yang merepresentasikan sentimen negatif.

#### 4.5 Pembagian Data Set

Setelah proses pelabelan, preprocessing, dan ekstraksi fitur dilakukan, data dibagi menjadi dua



bagian: data latih untuk membangun model dan data uji untuk mengevaluasi kinerjanya. Dalam penelitian ini, rasio yang digunakan adalah 80:20, yaitu 240 dari 300 ulasan digunakan sebagai data latih dan 60 ulasan sebagai data uji. Rasio ini dipilih karena cukup untuk melatih model sekaligus memberikan evaluasi objektif yang praktis, dan merupakan pendekatan yang umum digunakan dalam supervised learning.



Gambar 3 . Pembagian data set

#### 4.6 Pelatihan Model

Penelitian ini menggunakan model Multinomial Naïve Bayes, yang sangat cocok untuk data teks karena penggunaan frekuensi kata sebagai fitur diskrit. Sebanyak 240 data pelatihan (80% dari total data) direpresentasikan sebagai matriks TF-IDF dengan dua label: positif dan negatif. Model ini kemudian menghitung probabilitas kemunculan kata di setiap kelas untuk membentuk distribusi kata. Ketika menerima ulasan baru, model menentukan kelas berdasarkan probabilitas tertinggi antara P (ulasan positif) dan P (ulasan negatif) menggunakan

#### 4.7 Evaluasi Model

Setelah proses pelatihan model klasifikasi menggunakan algoritma Multinomial Naïve Bayes selesai, langkah selanjutnya adalah melakukan evaluasi terhadap 60 titik data uji (20% dari total 300 titik data). Evaluasi ini bertujuan untuk menilai performa model dalam mendeteksi dan mengklasifikasikan sentimen pelanggan, khususnya apakah ulasan yang diberikan bersifat positif atau negatif.

Tabel 9. Evaluasi Model

	Prediksi Positif	Prediksi Negatif
Aktual Positif	27 (True Positive)	3 (False Negative)
Aktual Negatif	4 (False Positive)	26 (True Negative)

Berdasarkan perhitungan confusion matrix tersebut dapat di perhitungan sebagai berikut:

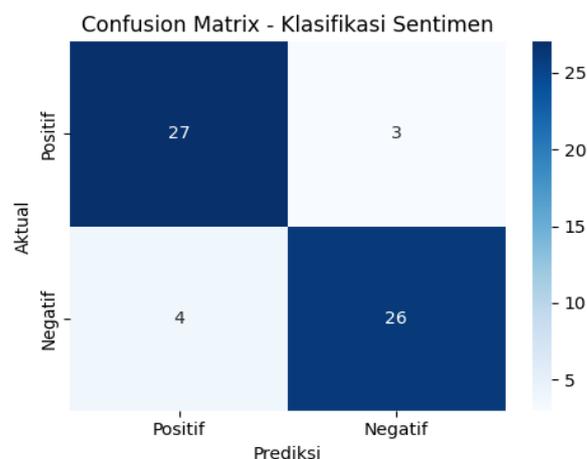
$$d. \text{ Akurasi} = \frac{TP+TN}{\text{Total}} = \frac{27+26}{60} = \frac{53}{60} = 88.33\%$$

$$e. \text{ Presisi} = \frac{TP}{TP+FN} = \frac{27}{27+4} = \frac{27}{31} = 87.10\%$$

$$f. \text{ Recall} = \frac{TP}{TP+FN} = \frac{27}{27+3} = \frac{27}{30} = 90.00\%$$

$$g. \text{ F1-score} = \frac{2 \times \text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 88.54\%$$

Menurut interpretasi hasil, model mencapai akurasi 88,33% pada data pengujian. Nilai recall sebesar 90% mencerminkan kemampuan model untuk mengenali hampir semua ulasan positif, sedangkan presisi sebesar 87,10% menunjukkan bahwa sebagian besar prediksi optimis benar. Nilai F1 sebesar 88,54% menunjukkan keseimbangan yang kuat antara presisi dan recall.



Gambar 4 . Confusion Matrix - Sentimen



Berikut keterangan hasil dalam visualisasi Confusion Matrix hasil evaluasi model klasifikasi sentimen:

- Aktual Positif – Prediksi Positif: 27 (True Positive)
- Aktual Positif – Prediksi Negatif: 3 (False Negative)
- Aktual Negatif – Prediksi Positif: 4 (False Positive)
- Aktual Negatif – Prediksi Negatif: 26 (True Negative)

#### 4.8 Interpretasi dan Pembahasan

Model ini menunjukkan performa yang stabil dan akurat, dengan tingkat akurasi di atas 88%. Nilai recall yang tinggi menunjukkan bahwa model ini sangat responsif terhadap ulasan positif, yang bermanfaat untuk mempertahankan loyalitas pelanggan. Namun, terdapat beberapa false positive dan false negative yang mengindikasikan adanya kesalahan dalam prediksi. Hal ini dapat disebabkan oleh kalimat yang ambigu, ironi, atau penggunaan kata-kata yang bermakna ganda, yang merupakan tantangan umum dalam analisis sentimen berbasis teks pendek.

#### 4.9 Alat Pemrograman Python

Dalam penelitian ini, bahasa pemrograman Python digunakan sebagai alat utama dalam proses klasifikasi sentimen pelanggan. Python dipilih karena memiliki library yang lengkap dan mendukung pemrosesan teks yang efisien, termasuk Sastrawi untuk stemming bahasa Indonesia, Scikit-learn untuk klasifikasi Naïve Bayes, dan TfidfVectorizer untuk ekstraksi fitur. Prosesnya dimulai dengan membaca data, melakukan prapemrosesan teks, lalu melatih dan mengevaluasi model klasifikasi. Python juga digunakan untuk memvisualisasikan hasil evaluasi seperti confusion matrix dengan bantuan Matplotlib dan Seaborn, sehingga memudahkan untuk menganalisa kinerja model secara intuitif dan akurat.

Laporan Evaluasi Model:				
	precision	recall	f1-score	support
positif	0.90	0.87	0.88	30
negatif	0.87	0.90	0.89	30
accuracy			0.88	60
macro avg	0.88	0.88	0.88	60
weighted avg	0.88	0.88	0.88	60

Gambar 5 . Laporan Evaluasi Model

```
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import confusion_matrix, classification_report

# Data aktual dan prediksi
# Total = 60 (30 positif, 30 negatif)
# 27 TP, 3 FN, 4 FP, 26 TN
y_true = ['positif'] * 30 + ['negatif'] * 30
y_pred = ['positif'] * 27 + ['negatif'] * 3 + ['positif'] * 4 + ['negatif'] * 26

# Hitung confusion matrix
labels = ['positif', 'negatif']
cm = confusion_matrix(y_true, y_pred, labels=labels)
```

Gambar 6 . Perintah Program Python

## 5. KESIMPULAN DAN SARAN

### 5.1 KESIMPULAN

Penelitian ini berhasil mengembangkan model klasifikasi sentimen untuk menganalisis opini pelanggan terhadap Restoran Koki Sunda dengan menggunakan algoritma Multinomial Naïve Bayes. Dengan total 300 ulasan pelanggan yang telah melalui tahap pra-pemrosesan dan ekstraksi fitur menggunakan metode TF-IDF, model mencapai performa yang baik.

Hasil evaluasi pada data uji menunjukkan akurasi sebesar 88,33%, presisi 87,10%, recall 90,00%, dan F1-score 88,54%. Pencapaian tersebut menunjukkan bahwa algoritma Naïve Bayes cukup efektif dalam mengenali sentimen pelanggan dari data teks yang tidak terstruktur.

Penelitian ini memberikan kontribusi ilmiah pada penerapan analisis sentimen berbasis teks bahasa Indonesia, khususnya dalam konteks masakan lokal. Penelitian ini memberikan manfaat praktis untuk pengambilan keputusan berbasis data oleh manajer restoran.

### 5.2 Saran

Penelitian ini masih dapat dikembangkan lebih lanjut.

- Diversifikasi Metode: Integrasikan algoritme lain seperti Support Vector Machine (SVM), Random Forest, atau model berbasis deep learning untuk membandingkan kinerja dan meningkatkan akurasi klasifikasi.



- b. Penambahan Dimensi Sentimen: Memperluas klasifikasi multi-label dengan menambahkan kategori netral atau menerapkan analisis sentimen berbasis aspek untuk hasil yang lebih kaya dan lebih kontekstual.
- c. Integrasi ke dalam Industri Nyata: Sistem klasifikasi sentimen ini memiliki potensi untuk diintegrasikan secara langsung ke dalam:
- 1) Aplikasi manajemen restoran (misalnya, dasbor umpan balik pelanggan).
  - 2) Layanan chatbot pelanggan untuk secara otomatis mendeteksi dan menanggapi keluhan.
  - 3) Platform pemantauan ulasan online yang menganalisis opini pelanggan dari berbagai sumber (Ulasan Google, media sosial, dll.).
- d. Validasi Lapangan: Uji coba langsung sistem di lingkungan restoran diperlukan untuk mengukur dampak nyata terhadap peningkatan layanan, efisiensi respons pelanggan, dan loyalitas pelanggan.

#### 6. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada seluruh pihak Responden dan Pelanggan Restoran Koki Sunda yang telah memberikan dukungan dan bantuan selama proses penyusunan penelitian ini.

#### DAFTAR PUSTAKA

- [1] N. Nurzaman, N. Suarna, and W. Prihartono, "Analisis Sentimen Ulasan Aplikasi Threads Di Google Playstore Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 1, pp. 967–974, 2024, doi: 10.36040/jati.v8i1.8708.*
- [2] D. R. Firmansyah and E. Lestariningsih, "Analisis Sentimen Ulasan Aplikasi Smart Campus Unisbank di Google Playstore Menggunakan Algoritma Naive Bayes," *J. JTik (Jurnal Teknol. Inf. dan Komunikasi), vol. 8, no. 2, pp. 498–507, 2024, doi: 10.35870/jtik.v8i2.1882.*
- [3] V. Fazrian, T. Suprapti, and R. Narasati, "Penerapan Algoritma Naive Bayes Terhadap Analisis Sentimen Aplikasi Game Multiplayer Online Battle Arena," *JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 1, pp. 1005–1012, 2024, doi: 10.36040/jati.v8i1.8432.*
- [4] S. A. Prasetyo and W. T. Atmojo, "Analisis Sentimen Terhadap Penggunaan Aplikasi Astro Menggunakan Algoritma Naïve Bayes," *J. Media Inform. Budidarma, vol. 7, no. 4, p. 2149, 2023, doi: 10.30865/mib.v7i4.6750.*
- [5] M. Yusuf Rismanda Gaja, I. Maulana, and O. Komarudin, "Analisis Sentimen Opini Pengguna Aplikasi Vidio Pada Ulasan Playstore Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform., vol. 7, no. 4, pp. 2767–2774, 2024, doi: 10.36040/jati.v7i4.7197.*
- [6] F. Tri Sanudin, B. Irawan, and A. Bahtiar, "Analisis Sentimen Ulasan Aplikasi Sapawarga Di Playstore Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 1, pp. 170–175, 2024, doi: 10.36040/jati.v8i1.8300.*
- [7] A. Nurian, M. S. Ma'arif, I. N. Amalia, and C. Rozikin, "Analisis Sentimen Pengguna Aplikasi Shopee Pada Situs Google Play Menggunakan Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap., vol. 12, no. 1, 2024, doi: 10.23960/jitet.v12i1.3631.*
- [8] S. Sisnawati, R. Astuti, and F. Muhamad basyiar, "ANALISIS SENTIMEN PADA APLIKASI KFCKU Di GOOGLE PLAYSTORE MENGGUNAKAN NAÏVE BAYES," *JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 3, pp. 3010–3016, 2024, doi: 10.36040/jati.v8i3.8382.*
- [9] T. Arifqi, N. Suarna, and W. Prihartono, "Penggunaan Naive Bayes Dalam Menganalisis Sentimen Ulasan Aplikasi Mcdonald'S Di Indonesia," *JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 2, pp. 1949–1956, 2024, doi: 10.36040/jati.v8i2.8740.*
- [10] Erlin elisa, Y. Mardiansyah, and R. Fauzi, "Pola Pembelian Konsumen Terhadap Product Umkm Martista Ikhsan Dengan Algoritma Naive Bayes," *J. Manaj. Inform. dan Sist. Inf., vol. 5, no. 2, pp. 159–167, 2022, doi: 10.36595/misi.v5i2.661.*
- [11] N. M. A. Mahar, Vihi Atina, and Nugroho Arif Sudiby, "Pemodelan Prediksi Kelulusan Mahasiswa Dengan Metode Naive Bayes Di Uniba," *J. Manaj. Inform. dan Sist. Inf., vol. 6, no. 2, pp. 148–158, 2023, doi: 10.36595/misi.v6i2.875.*
- [12] S. D. S. Kurniawan and A. Fauzy,



- “Penggunaan Naïve Bayes Classifier dalam Analisis Sentimen Ulasan Aplikasi McDonald’s: Perspektif Pengguna di Indonesia,” *J. Media Inform. Budidarma*, vol. 8, no. 3, p. 1545, 2024, doi: 10.30865/mib.v8i3.7765.
- [13] Ernianti Hasibuan and Elmo Allistair Heriyanto, “Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier,” *J. Tek. dan Sci.*, vol. 1, no. 3, pp. 13–24, 2022, doi: 10.56127/jts.v1i3.434.
- [14] D. R. Fathwa Daud, B. Irawan, and A. Bahtiar, “Penerapan Metode Naive Bayes Pada Analisis Sentimen Aplikasi Mcdonalds Di Google Play Store,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 759–766, 2024, doi: 10.36040/jati.v8i1.8784.
- [15] Permana A, Taufiq R, and Wijaya M, “Implementasi Algoritma Naïve Bayes Terhadap Review Aplikasi KFCKU,” *J. Tek.*, vol. Vol. 12 No. 02, no. 02, pp. 128–137, 2023.
- [16] M. R. A. F. Zain and M. Kamayani, “Analisis Sentimen Ulasan Pelanggan Online Ubi Madu Cilembu Abah Nana Menggunakan Algoritma Naïve Bayes,” *J. Sist. Komput. dan Inform.*, vol. 5, no. 1, p. 11, 2023, doi: 10.30865/json.v5i1.6646.
- [17] W. O. Astri Ahmad and S. Supangat, “Sentimen Analisis Menggunakan Naïve Bayes Untuk Rekomendasi Restoran Wakatobi,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 4, pp. 7704–7711, 2024, doi: 10.36040/jati.v8i4.10432.