

## IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN ULASAN PRODUK MAKANAN DAN MINUMAN DI TOKOPEDIA

Hajaroh, Tati Suprapti, Riri Narasati

Teknik Informatika, STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135

*hajaroh15@gmail.com*

### ABSTRAK

Analisis sentimen adalah proses evaluasi dan pemahaman terhadap pendapat, perasaan, atau sentiment yang terkandung dalam teks, ulasan, atau komunikasi yang diungkapkan oleh individu atau kelompok. Penelitian ini bertujuan untuk menerapkan algoritma *Naive Bayes* dalam analisis sentimen terhadap ulasan produk makanan dan minuman di platform e-commerce Tokopedia. Melalui metode *Knowledge Discovery in Databases (KDD)*, langkah pertama penelitian adalah memahami dan memproses data yang diperoleh dari sumber data terbuka yaitu Kaggle Data ulasan pelanggan dikumpulkan dan dieksplorasi untuk memahami distribusi sentimen serta dilakukan *preprocessing* data untuk mempersiapkan dataset pelatihan. Algoritma *Naive Bayes* diterapkan untuk mengklasifikasikan ulasan produk ke dalam tiga kategori sentimen: positif, negatif, dan netral. Hasil evaluasi kinerja model menunjukkan tingkat Accuracy sebesar 75.90%. Sementara itu, Presisi sebesar 66.45% dan *recall* sebesar 97.42% turut menjadi sorotan penting. Penelitian ini memberikan pemahaman yang lebih mendalam terkait distribusi sentimen dalam ulasan makanan dan minuman di Tokopedia, menunjukkan mayoritas sentimen yang positif. Hasil ini mengindikasikan bahwa algoritma *Naive Bayes* dapat menjadi instrumen efektif untuk menganalisis sentimen produk di platform *e-commerce*, memberikan landasan bagi pengembangan strategi bisnis yang lebih cermat dan responsif terhadap preferensi pelanggan.

*Kata Kunci: Analisis Sentimen, KDD, Naive Bayes, Tokopedia, Ulasan produk*

### 1. PENDAHULUAN

Pesatnya perkembangan di bidang Informatika telah mengubah secara signifikan cara kita berinteraksi dengan dunia sekitar. Inovasi teknologi informasi telah mendominasi berbagai aspek kehidupan, mengubah cara kita bekerja, berbelanja, berkomunikasi, dan bahkan memberikan umpan balik. Terlebih lagi, di era *digital* ini, e-commerce atau perdagangan elektronik telah menjadi salah satu pilar utama dalam ekonomi global. Platform-platform e-commerce seperti Tokopedia telah memberikan pelanggan akses tak terbatas ke berbagai produk, termasuk makanan dan minuman, dengan ulasan pelanggan menjadi salah satu sumber informasi utama dalam pengambilan keputusan. Oleh karena itu, analisis sentimen ulasan produk makanan dan minuman di Tokopedia dengan menggunakan algoritma *Naive Bayes* menjadi semakin relevan dalam konteks ini. Penelitian ini bertujuan untuk memahami bagaimana algoritma ini dapat diterapkan dalam menginterpretasikan ulasan pelanggan dan memberikan wawasan yang berharga dalam meningkatkan kualitas layanan dan produk di platform e-commerce.

Tujuan utama dari penelitian ini adalah menerapkan algoritma *Naive Bayes* untuk analisis sentimen ulasan produk makanan dan minuman di Tokopedia. Dengan demikian, penelitian ini bertujuan untuk mengembangkan pemahaman yang lebih mendalam tentang bagaimana teknik analisis sentimen dapat diterapkan dalam konteks *e-commerce* khususnya untuk produk makanan dan minuman. Dengan memahami sentimen pelanggan terhadap produk tersebut, penelitian ini berkontribusi pada

upaya meningkatkan kualitas layanan dan produk yang ditawarkan oleh penjual di platform e-commerce, serta memberikan panduan yang berharga dalam pengambilan keputusan bisnis. Selain itu, penelitian ini juga memberikan wawasan tentang penggunaan algoritma *Naive Bayes* dalam analisis sentimen yang dapat bermanfaat secara praktis untuk penjual dan produsen di industri e-commerce. Hasil penelitian ini diharapkan dapat membantu dalam meningkatkan kepuasan pelanggan, memahami preferensi pelanggan, serta mengidentifikasi area perbaikan dalam produk dan layanan. Dengan demikian, penelitian ini memberikan kontribusi pada bidang Informatika dengan menggabungkan analisis sentimen dan e-commerce untuk hasil yang bermanfaat secara bisnis dan praktis.

### 2. TINJAUAN PUSTAKA

#### 2.1. Analisis Sentimen

Dalam subjek text mining, analisis sentimen mengkaji pandangan, perasaan, penilaian sikap, dan emosi masyarakat terhadap berbagai entitas, termasuk orang, organisasi, barang, jasa, isu, peristiwa, tema, dan karakteristiknya. Teknik memahami, mengekstraksi, dan memproses data teks secara mekanis untuk mendapatkan informasi sentimen yang terkandung dalam kalimat opini dikenal sebagai analisis sentimen, atau penambangan opini. Analisis sentimen berupaya memastikan kecenderungan individu terhadap sudut pandang atau opini tertentu. Analisis sentimen adalah gabungan dari pemrosesan bahasa alami dan penambangan teks.[1]

## 2.2. Algoritma Naive Bayes

Teknik klasifikasi yang disebut Naive Bayes didasarkan pada teorema Bayes. Alasan mengapa pendekatan kategorisasi ini disebut teorema Bayes adalah karena pendekatan ini menggunakan metode statistik dan probabilitas yang dikemukakan oleh ilmuwan Inggris *Thomas Bayes*. Secara khusus, pendekatan ini memprediksi peluang masa depan dengan memanfaatkan data historis. Ciri utama pengklasifikasi Bayesian yang naif adalah asumsinya yang ekstrem (naif) mengenai independensi setiap kondisi dan peristiwa.[2]

Model Naive Bayes:

1. Model Gaussian Naive Bayes yang dapat mengasumsikan distribusi nilai yang kontinu dengan nilai numerik untuk setiap fitur (distribusi Gaussian) merupakan jenis model Naive Bayes yang pertama.
2. Multinomial Naive Bayes: Model yang sering digunakan untuk klasifikasi dokumen atau pemrosesan bahasa alami (NLP), menentukan frekuensi kemunculan istilah dalam suatu dokumen (term frekuensi).
3. Model Bernoulli Naive Bayes, yang menggunakan prediktor variabel boolean untuk mengkategorikan data sebagai Ya atau Tidak.

Secara umum rumus teorema Bayes adalah sebagai berikut:

$$P(A|B) = P(B|A)P(A)P(B) \quad (1)$$

Detail:

$P(A|B)$  : kemungkinan terjadinya A jika ada bukti terjadinya B (kemungkinan lebih tinggi)

$P(B|A)$  : Peluang terjadinya B berdasarkan bukti terjadinya A

$c(x)$  : Peluang Terjadinya A

## 2.3. Multinomial Naive Bayes

Dalam pemrosesan bahasa alami (NLP), algoritma multinomial naïve Bayes adalah teknik pembelajaran probabilistik berdasarkan teorema Bayes. Dasar dari teknik ini adalah frekuensi kata, atau frekuensi kemunculan sebuah kata dalam sebuah dokumen. Model ini memberikan penjelasan untuk dua fakta: apakah sebuah kata muncul dalam dokumen dan seberapa sering kata tersebut muncul.[3]

Proses dimulai dengan memasukkan data latih yang digunakan untuk pembelajaran kemudian dilakukan perhitungan peluang kemunculan suatu kelas pada data latih yang dilakukan menggunakan:

$$P(c) = \frac{N_c}{N_{doc}}$$

Keterangan:

$C$  : Kategori atau kelas

$Doc$  : Dokumen

$N_c$  : Banyaknya kategori c pada dokumen latih

$N_{doc}$  : Banyaknya keseluruhan dokumen latih yang digunakan

Perhitungan selanjutnya dari probabilitas bahwa kata  $i$  termasuk dalam kategori atau kelas tertentu dapat dilakukan dengan menggunakan:

$$P(w_i, c) = \frac{\text{count}(w_i, c) + 1}{\sum \text{count}(w, c) + |V|}$$

Keterangan:

$w_i$  : Kata ke- $i$  dalam seluruh dokumen yang berkategori

$\text{count}(w_i, c)$  : Jumlah kata tertentu yang muncul dalam suatu kategori atau kelas

$\sum \text{count}(w, c)$  : Jumlah seluruh kata pada kelas

$|V|$  : merupakan jumlah seluruh kata unik pada kelas

## 2.4. Representasi Teks

Proses menerjemahkan teks ke dalam format yang dapat dipahami oleh komputer atau algoritma pembelajaran mesin dikenal sebagai representasi teks. Agar komputer dapat menafsirkan, mengevaluasi, dan membuat prediksi berdasarkan teks, data teks harus diubah menjadi representasi numerik, seperti vektor.

## 2.5. Ulasan Produk di Tokopedia

Tokopedia adalah platform e-commerce yang menyediakan berbagai produk, termasuk makanan dan minuman. Ulasan produk di Tokopedia mencakup umpan balik dari pelanggan terkait pengalaman mereka dengan produk tertentu.

## 2.6. Evaluasi Model

Evaluasi model melibatkan penggunaan metrik seperti akurasi, presisi, dan recall. Metrik ini membantu mengukur sejauh mana model Naive Bayes dapat memprediksi sentimen dengan akurat berdasarkan ulasan pelanggan.

### a. Accuracy

Ketepatan kategorisasi model ditunjukkan oleh keakuratannya. Didefinisikan sebagai tingkat kesesuaian antara hasil yang diharapkan dan hasil aktual untuk menilai keakuratan persiapan data pengujian atau kinerja model.[4]

$$\text{Accuracy} = \frac{\sum_{i=1}^n \frac{TPI + TNI}{TPI + TNI + FPI + FNI}}{n} \times 100\% \quad (1)$$

Keterangan:

$Tpi$  = True Positive (jumlah data positif yang terklasifikasi benar oleh sistem untuk kelas ke- $i$ )

$Tni$  = True Negative (jumlah data negatif yang terklasifikasi benar oleh sistem untuk kelas ke- $i$ )

$Fni$  = False Negative (jumlah data negatif terklasifikasi salah oleh sistem untuk kelas ke- $i$ )

$Fpi$  = False Positive ( jumlah data positif terklasifikasi benar oleh sistem untuk kelas ke- $i$ )

$i$  = jumlah kelas

## b. Precision

Precision mengacu pada kualitas hasil yang diharapkan model dan keakuratan data, atau kecepatan sistem memberikan hasil sebagai respons terhadap permintaan informasi dari pengguna. Rumus perhitungan precision:

$$Precision = \frac{\sum n_i * TP_i}{\sum n_i * (FP_i + TP_i)} \times 100\% \quad (2)$$

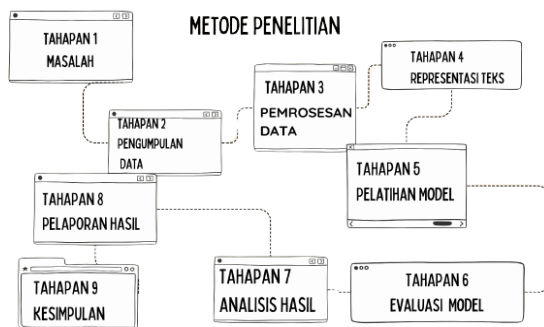
## c. Recall

Recall, yang dapat juga dicirikan sebagai pengukuran kuantitas dokumen teks penting, menjelaskan pencapaian model dalam inovasi informasi. Rumus perhitungan recall:

$$Recall = \frac{n \sum n_i * TP_i}{\sum n_i * (TP_i + FN_i)} \times 100\% \quad (3)$$

### 3. METODE PENELITIAN

Metode penelitian untuk analisis sentimen ulasan produk makanan dan minuman dengan algoritma Naive Bayes melibatkan serangkaian langkah yang sistematis untuk mengumpulkan data, memproses data, melatih model, dan mengevaluasi hasil.



Gambar 1. Tahapan Metode Penelitian

Tahapan Metode Penelitian:

1. Masalah yaitu menentukan permasalahan penelitian
2. Pengumpulan Data yaitu Pada tahap ini, peneliti akan mengidentifikasi sumber data ulasan produk makanan dan minuman, seperti situs web ulasan produk atau platform e-commerce yang relevan. Kemudian, peneliti akan mengunduh atau mengumpulkan data ulasan, termasuk teks ulasan dan label sentimen yang sesuai.
3. Pemrosesan Data Langkah ini melibatkan pembersihan data untuk menghilangkan karakteristik yang tidak relevan, seperti tanda baca, karakter khusus, dan kata-kata yang tidak diperlukan. Peneliti juga akan melakukan normalisasi teks, seperti konversi ke huruf kecil, penghilangan spasi ekstra, dan penghapusan stop words. Selain itu, peneliti akan membagi data menjadi dua set, yaitu data pelatihan dan data pengujian.
4. Representasi Teks yaitu Pada tahap ini, peneliti akan menerapkan metode representasi teks, seperti Bag-of-Words (BoW) atau Term

Frequency-Inverse Document Frequency (TF-IDF), untuk mengubah teks ulasan menjadi vektor fitur. Ini akan menghasilkan matriks fitur yang akan digunakan sebagai masukan untuk model Naive Bayes.

## 5. Pelatihan Model

6. Evaluasi Model yaitu Peneliti akan mengevaluasi model yang telah dilatih menggunakan data pengujian untuk mengukur kinerja model dalam mengklasifikasikan sentimen. Ini termasuk menghitung metrik evaluasi seperti akurasi, presisi, recall.

7. Analisis Hasil yaitu Peneliti akan menganalisis hasil untuk memahami pandangan konsumen terhadap produk makanan dan minuman. Peneliti akan mencari pola, tren, atau temuan penting dalam data ulasan yang bisa menjadi wawasan berharga.

8. Pelaporan Hasil yaitu Pada tahap ini, peneliti akan menulis laporan penelitian yang mencakup deskripsi metode, hasil analisis, dan interpretasi temuan. Peneliti juga akan menyajikan visualisasi data, grafik, atau tabel yang relevan.

9. Kesimpulan dan Saran yaitu Peneliti akan membuat kesimpulan berdasarkan hasil penelitian dan memberikan saran untuk penelitian lanjutan atau penggunaan hasil analisis sentimen dalam konteks tertentu.

#### 3.1. Sumber Data

Portal open source Kaggle, yang menawarkan beragam kumpulan data, menjadi sumber data yang digunakan dalam penelitian ini. Dataset yang digunakan dalam penelitian ini mencakup ulasan kondisi bersih yang diposting pengguna dengan label sentimen barang makanan dan minuman di situs e-commerce Tokopedia. Informasi ini terdiri dari serangkaian ulasan yang menunjukkan bagaimana perasaan pelanggan terhadap makanan dan minuman yang mereka beli. Dengan penggunaan dataset ini, metode Naive Bayes dapat digunakan dalam analisis sentimen untuk menganalisis sentimen konsumen dan menawarkan informasi mendalam mengenai produk makanan dan minuman yang tersedia di platform Tokopedia.

#### 3.2. Teknik Analisis Data

Metode analisis data yang digunakan dalam penelitian. Hal ini menggunakan pendekatan Knowledge Discovery In Database (KDD), yang memiliki banyak langkah yang *Data Selection*, *Data Preprocessing*, *Data Transformation*, *datamining*, hingga *evaluation*. [5]

Penjelasan mendalam mengenai fase-fase tersebut disajikan di bawah ini:

##### 3.2.1. Data Selection

Pemilihan atribut dan pengumpulan data akan ditempatkan pada langkah KDD ini. Data ulasan pada halaman ini disampaikan oleh pengguna aplikasi

Tokopedia. Kumpulan data ini diperoleh dari data sumber terbuka Kaggle. Kumpulan datanya berupa file CSV dengan keseluruhan 4.060 baris data, dua kolom atribut, dan satu label sentimen yang menampilkan hasil ulasan positif, negatif, dan netral.

### 3.2.2. Preprocessing

Tahap pembersihan dan penyempurnaan data dikenal sebagai preprocessing. Sebagian besar data yang dikumpulkan tidak terstruktur dan memiliki jumlah karakter yang besar. Preprocessing dilakukan untuk menghilangkan noise.[6] preprocessing membantu mempersiapkan data sehingga algoritma naïve bayes dapat bekerja lebih baik dan menghasilkan hasil yang lebih akurat. Tahap ini memainkan peran penting dalam menyederhanakan data serta meminimalkan efek negatif dari kecacatan atau ketidakseimbangan yang mungkin ada dalam dataset.

### 3.2.3. Tokenize

*Tokenize* adalah prosedur yang melibatkan pembagian string karakter menurut karakter spasi. Ini mungkin juga melibatkan penghapusan karakter tertentu, seperti tanda baca.[7]

### 3.2.4. Tranformation

*Transformation* data menggunakan operator nominal *to teks* merupakan contoh bagaimana data ditransformasikan pada tahap transformasi sehingga dapat diproses pada tahap data mining.

### 3.2.5. Data Mining

Data mining menggunakan Algoritma Naïve Bayes bertujuan untuk mengklasifikasikan data berdasarkan probabilitas terjadinya suatu kejadian, sehingga penggunaan Algoritma Naïve Bayes untuk identifikasi pola sentimen dari ulasan tersebut.

### 3.2.6. Evaluation

Evaluasi membantu mengukur kinerja dari model atau hasil yang dihasilkan dari proses data mining. dalam hal ini evaluasi membantu memvalidasi keandalan dan keakuratan dari informasi atau pola yang ditemukan.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Data

Tahap pengumpulan data ini dilakukan dengan mendownload data dari situs resmi Kaggle dengan keyword Ulasan Makanan dan Minuman di Tokopedia. Data yang berisi ulasan makanan dan minuman di aplikasi Tokopedia berisi 4.061 data dengan bantuan bahasa program python di google colab. Terdapat tiga atribut yaitu Review, Rating, dan Sentiment. Berikut adalah dataset hasil mendownload dari kaggle.

Review	Rating	Sentiment
1. Review		
2. enak banget	5	positive
3. pengiriman cepat packing bagus sesuai pesanan belum di coba terima kasih	5	positive
4. pengiriman kurir lama tapi tidak ada masalah saya lebih senang ke rasa original atau rasa asli kayaknya mas x	4	negative
5. terimakasih mas	5	neutral
6. udah order untuk keesokan hari ini	5	positive
7. sesuai dan grafik bagus	5	positive
8. respon cepat harga terjangkau	5	positive
9. mantap banget recomended emangnya ada	5	positive
10. mantap banget mas baik grafik bagus emm mas jajan lg terimakasih	5	positive
11. kondisi baik kurir baik mas terimakasih	5	positive
12. rekomended	5	positive
13. pengiriman cepat packing rapi kurir baik mas order lg tm klu seler dan tokopedia	5	positive
14. cepat n fresh	4	positive
15. excellent	5	neutral
16. lumayan ada ada sedikit di kemasnya hehe	4	negative
17. lagi ngejeng banget mas terimakasih mas pengiriman juga cepat mas mas post buat di order mas	5	positive
18. pesan ini mulai sekarang mau di order lagi walaupun bisa mending dan lebih cepat terima yg baik akan nyoba	5	negative
19. terimakasih sudah sangat memuaskan dan pengiriman sangat cepat terimakasih	5	positive

Gambar 2. Data

### 4.2. Selection

Set Role sangat berguna dalam analisis sentimen untuk menetapkan peran dari setiap fitur atau variabel dalam dataset, memungkinkan identifikasi variabel yang menjadi faktor penting dalam memahami dan memprediksi sentimen dari teks atau data yang dianalisis



Gambar 3. Set Role

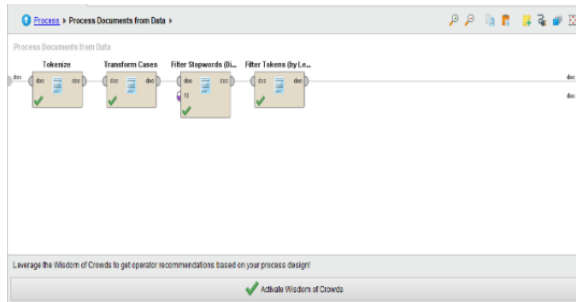
Parameter pada operator Set Role dapat dilihat pada gambar dibawah ini:

attribute name	target role
Sentiment	label
Rating	regular
Review	regular

Gambar 4. Parameter set role

### 4.3. Process Document From Data

Proses dokument from data merupakan tahapan penting dalam manajemen data yang melibatkan pengumpulan, pengorganisasian, dan dokumentasi informasi terkait dengan data yang dikumpulkan atau digunakan dalam suatu proyek atau analisis. Proses ini dirancang untuk memberikan pemahaman yang jelas dan komprehensif tentang asal, struktur, dan sifat data kepada pengguna atau pemroses data lainnya. Process Document From data dapat dilihat pada gambar dibawah ini:



Gambar 5. Process Document From Data

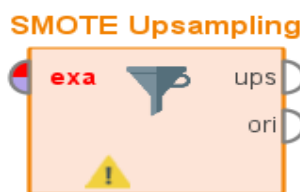
Tahap Process Document From Data ini dilakukan dengan langkah awal seperti Tokenize, Transform cases, Stopword, dan filter Tokens. dapat dilihat hasil dari process document from data pada gambar dibawah ini:

Word	Attribute	Total Count	Document	positive	negative	neutral
aaaa	?	1	1	1	0	0
aaaaaaa	?	1	1	0	0	0
aaaaaa	?	1	1	0	0	0
aaaaa	?	1	1	0	0	1
aaaa	?	7	7	4	0	3
aaaa	?	1	1	1	0	0
aaaa	?	4	4	4	0	0
aaa	?	19	19	17	2	0
aaaa	?	3	3	3	0	0
aaa	?	1	1	1	0	0
aaa	?	13	13	9	4	0
aaaaaa	?	6	6	5	1	0
aaaa	?	3	3	3	0	0

Gambar 6. Hasil Process Document Data

#### 4.4. SMOTE upsampling

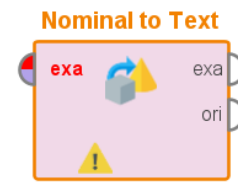
**SMOTE** (*Synthetic Minority Over-sampling Technique*) adalah teknik untuk menangani ketidakseimbangan kelas dalam dataset dengan membuat sampel sintetis dari kelas minoritas. Ini digunakan dalam tugas klasifikasi untuk meningkatkan kinerja model pada kelas yang kurang mewakili. Operator *SMOTE* dapat dilihat pada gambar dibawah ini:



Gambar 7. *SMOTE Upsampling*

## 4.5. Transformation

Tahap transformasi adalah tahap yang mengubah data menjadi format yang dapat diolah oleh tahap data mining yaitu tahap yang mengubah data menjadi teks nominal dan nominal menjadi teks. berikut operator Nominal to Text yang dapat dilihat pada gambar dibawah ini:



Gambar 8. Nominal to Text

#### 4.6. Data mining

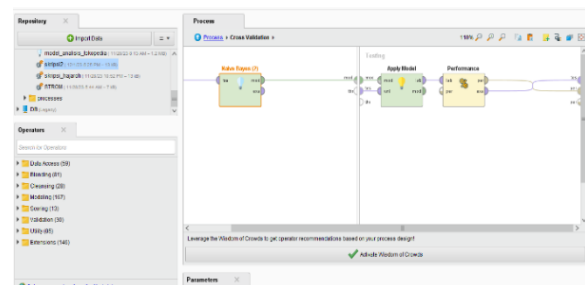
Dalam data mining, data digabungkan atau diubah ke dalam format yang dapat diproses[8]. Pada tahap ini dilakukan model menggunakan naive bayes yang mana terdapat beberapa operator diantaranya Read CSV yang berfungsi menampung data train lalu dihubungkan dengan Set Role yang mana untuk mengubah target role nya menjadi label kemudian dihubungkan dengan Nominal to Text untuk mengubah nominal ke text lalu ke proses document from data yang didalamnya terdapat sub proses seperti Tokenize, Transform cases, Stopword, dan filter Tokens. Kemudian SMOTE upsampling untuk menyeimbangkan data kemudian cross validation untuk menguji model naive bayes.



Gambar 9. Proses model *Naïve Bayes*

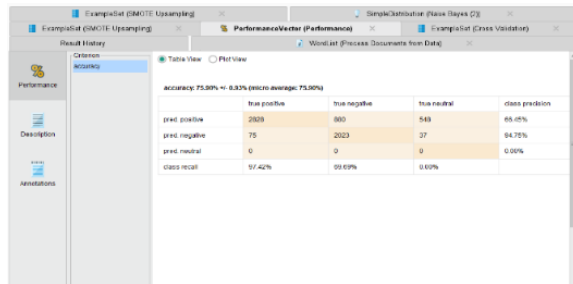
## 4.7. Evaluation

Tahap evaluasi adalah salah satu tahap penting dalam proses penelitian, pengembangan, atau analisis di berbagai bidang, termasuk ilmu pengetahuan, teknologi, bisnis, dan pendidikan. Evaluasi bertujuan untuk menilai atau mengukur efektivitas, kinerja, atau kualitas suatu program, produk, atau proses berdasarkan pada kriteria tertentu. Tahap ini dilakukan untuk menghitung evaluasi dari model naive bayes dengan *cross validation* yang dihubungkan dengan apply model dan performance untuk dihitung terhadap hasil dari *naive bayes*.



Gambar 10. proses *Validation Naive Bayes*

Berdasarkan hasil evaluasi model, bahwa hasil klasifikasi terhadap model naive bayes didapatkan hasil accuracy sebesar 75.90%, Presisi 66,45%, dan Recall 97,42%. bisa dilihat pada gambar dibawah ini.

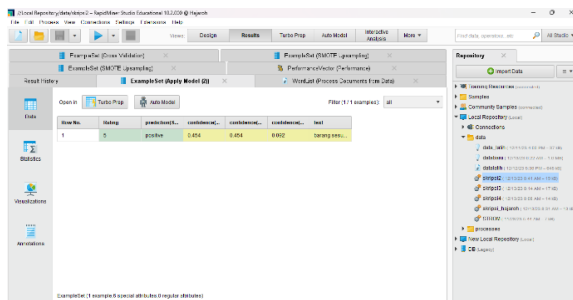


	Accuracy	Precision	Recall	F1 Score
Overall	75.90%	66.45%	97.42%	80.00%
Class				
positive	75.90%	66.45%	97.42%	80.00%
negative	75.90%	66.45%	97.42%	80.00%
neutral	75.90%	66.45%	97.42%	80.00%

Gambar 11. Hasil Evaluasi

#### 4.8. Implementasi Model Naive Bayes

Implementasi model *Naive Bayes* adalah proses menerapkan model yang telah dilatih menggunakan algoritma Naive Bayes pada data baru untuk melakukan prediksi atau klasifikasi. Model *Naive Bayes* adalah salah satu metode klasifikasi yang digunakan dalam pembelajaran mesin untuk memprediksi probabilitas kelas atau label berdasarkan fitur-fitur yang ada dalam data. Hasil implementasi dapat dilihat pada gambar dibawah ini:



Case	Class	Label	Score	Confidence	Probability	Posterior
1	positive	0.854	0.854	0.854	0.854	0.854

Gambar 12. Hasil implementasi Data Baru

#### 4.9. Penerapan Algoritma Naive Bayes

Algoritma *Naive Bayes* terbukti menjadi instrumen yang efektif dalam menganalisis sentimen ulasan produk makanan dan minuman di platform Tokopedia. Proses dimulai dengan pra-pemrosesan data yang melibatkan pengumpulan data dari ulasan, diikuti dengan langkah-langkah seperti tokenisasi, *transform case*, penghapusan *stopwords*, dan filterisasi token. Data yang sudah siap kemudian dibagi menjadi data pelatihan dan pengujian. Model *Naive Bayes* dilatih menggunakan data pelatihan dengan menghitung probabilitas munculnya kata-kata dalam kategori sentimen (positif, negatif, netral) berdasarkan data tersebut. Setelahnya, data pengujian digunakan untuk menguji kinerja model dengan mengklasifikasikan ulasan ke dalam kategori sentimen yang diperkirakan. Algoritma *Naive Bayes* memberikan hasil evaluasi dengan tingkat akurasi sebesar 75,90%, presisi sekitar 66,45%, dan recall mencapai 97,42%. Evaluasi ini memberikan

pemahaman tentang distribusi sentimen pada ulasan, yang mayoritasnya adalah sentimen positif. Hasil ini memberikan wawasan berharga bagi pemilik bisnis atau pemasar dalam mengambil langkah-langkah selanjutnya untuk meningkatkan produk atau layanan, memperbaiki area dengan sentimen negatif, dan mempertahankan aspek yang mendapat respons positif dari pelanggan. Dengan demikian, algoritma *Naive Bayes* dapat digunakan secara efektif untuk menganalisis sentimen ulasan produk makanan dan minuman di Tokopedia dan memberikan dasar yang kokoh untuk pengembangan strategi bisnis yang lebih baik.

#### 4.10. Hasil Evaluasi kinerja model Naive Bayes

Hasil evaluasi kinerja model *Naive Bayes* adalah langkah penting dalam memahami seberapa baik model dapat mengklasifikasikan sentimen dari ulasan yang ada. Evaluasi ini memberikan wawasan tentang seberapa efektif model dalam melakukan tugasnya dan seberapa dapat diandalkannya untuk digunakan dalam mengklasifikasikan data baru. Berikut adalah beberapa poin penting dalam membahas hasil evaluasi kinerja model *Naive Bayes*:

##### 4.10.1. Akurasi (Accuracy)

Akurasi model *Naive Bayes* sebesar 75.90% menunjukkan bahwa dari total prediksi yang dilakukan oleh model, sekitar 75.90% di antaranya adalah prediksi yang benar. Meskipun akurasi ini memberikan gambaran keseluruhan performa, tidak selalu mencerminkan kualitas sebenarnya dari model jika data memiliki ketidakseimbangan kelas yang signifikan.

##### 4.10.2. Presisi (Precision)

Presisi mengukur seberapa banyak prediksi positif yang benar dibandingkan dengan total prediksi positif yang dilakukan oleh model. Dengan presisi sebesar 66.45%, dari semua prediksi yang dianggap positif oleh model, sekitar 66.45% di antaranya adalah prediksi yang benar. Presisi yang rendah menunjukkan kemungkinan terjadinya kesalahan dalam memprediksi ulasan positif dan negatif.

##### 4.10.3. Recall (Sensitivitas)

*Recall* mengukur seberapa banyak kelas positif yang sebenarnya diidentifikasi dengan benar oleh model. *Recall* model Naive Bayes sebesar 97.42% menunjukkan bahwa model dapat mengidentifikasi sekitar 97.42% dari keseluruhan ulasan yang seharusnya diklasifikasikan sebagai positif. Meskipun recall yang tinggi menunjukkan bahwa model jarang melewatkan ulasan yang seharusnya diklasifikasikan sebagai positif, namun presisi yang rendah bisa berarti adanya banyak ulasan yang sebenarnya negatif tetapi diprediksi sebagai positif.



#### 4.11. Evaluasi Keseluruhan

Model *Naive Bayes* menunjukkan kemampuan yang relatif baik dalam mengidentifikasi sentimen positif dari ulasan, ditunjukkan dengan recall yang tinggi. Meskipun *recall* tinggi memberikan keyakinan bahwa model jarang melewatkan ulasan positif yang seharusnya, presisi yang rendah menunjukkan adanya kesalahan dalam memprediksi ulasan sebagai positif yang sebenarnya negatif. Akurasi yang cukup baik memberikan indikasi umum bahwa model memiliki kemampuan untuk melakukan klasifikasi dengan benar.

Menurut [9] dalam penelitiannya menyatakan Klasifikasi ini menghasilkan nilai akurasi *Naive Bayes* sebesar 75%, presisi sebesar 73%, recall sebesar 75%, dan nilai skor f1 sebesar 74%. Hasil data tweet yang digunakan pada penelitian ini menunjukkan bahwa algoritma *Naive Bayes* mempunyai nilai akurasi yang cukup baik. Menurut [10] dalam penelitiannya mendapatkan hasil yaitu nilai akurasi dengan Algoritma *Naive Bayes* dihasilkan sebesar 75%, ini menunjukkan keakuratan metode *Naive Bayes* dalam memprediksi hasil klasifikasi dalam kategori baik.

#### 4.12. Hasil Implementasikan model *Naive Bayes*

Dalam konteks RapidMiner, penggunaan operator *Apply Model* untuk menerapkan model *Naive Bayes* pada data baru biasanya dilakukan setelah model tersebut telah dilatih (dibuat). Dengan model yang sudah siap, operator *Apply Model* digunakan untuk membuat prediksi atau mengklasifikasikan data baru berdasarkan informasi yang diperoleh dari model tersebut. Langkah-langkah untuk menerapkan model *Naive Bayes* pada data baru menggunakan operator *Apply Model* di RapidMiner memerlukan serangkaian langkah yang terstruktur. Pertama, dilakukan pelatihan model *Naive Bayes* dengan menggunakan data latihan yang ada. Proses ini melibatkan penggunaan operator seperti "*Naive Bayes*" atau serangkaian operator lainnya untuk mengolah data dan melatih model sesuai dengan informasi yang tersedia. Selanjutnya, data baru yang akan diprediksi atau diklasifikasikan harus diproses agar formatnya sesuai dengan yang digunakan saat melatih model. Setelah model *Naive Bayes* dilatih, langkah berikutnya adalah menerapkan operator *Apply Model*. Dalam tahap ini, model yang telah dilatih akan digunakan pada data baru untuk membuat prediksi atau mengklasifikasikan data tersebut. Setelah operator *Apply Model* dijalankan, hasilnya akan berupa prediksi atau klasifikasi dari data baru berdasarkan model *Naive Bayes* yang telah diterapkan.

Table 1 Data Baru

Text	Rating	category
Barang sesuai pesanan dan cepat sampai	5	pertukangan

Table 2 Implementasi Data Baru

Rating	Predict ion (Sentiment)	Convidence (Positive)	Convidence (Negatif)	Convidence (netral)	Text
5	Positive	0.454	0.454	0.092	Barang sesuai pesanan cepat

Hasil implementasi data baru dengan prediksi sentimen, ditemukan bahwa prediksi sentimennya adalah positif dengan tingkat keyakinan yang cukup tinggi. Prediksi positif memiliki nilai kepercayaan (confidence) sebesar 0.454, menunjukkan bahwa model cenderung yakin dengan prediksi positifnya. Sementara itu, hasil implementasi juga menunjukkan adanya sejumlah prediksi sentimen negatif dengan tingkat kepercayaan yang juga signifikan, meskipun lebih rendah dibandingkan dengan prediksi positif. Prediksi negatif ini memiliki nilai kepercayaan sebesar 0.454. Selain itu, terdapat sejumlah kecil prediksi yang diklasifikasikan sebagai netral dengan tingkat kepercayaan sebesar 0.092. Meskipun demikian, jumlah prediksi netral ini terlihat lebih rendah dibandingkan dengan prediksi positif dan negatif. Secara keseluruhan, implementasi data baru menghasilkan prediksi sentimen yang didominasi oleh sentimen positif dengan tingkat kepercayaan yang relatif tinggi, disusul dengan adanya beberapa prediksi sentimen negatif dan hanya sedikit prediksi netral. Hasil implementasi menggunakan *apply model* ini bisa diterapkan untuk analisis sentimen pada data ulasan baru dengan model yang sudah dibuat.

## 5. KESIMPULAN DAN SARAN

Berdasarkan Hasil dan pembahasan dapat diambil kesimpulan sebagai berikut: Algoritma *Naive Bayes* memberikan hasil evaluasi dengan tingkat akurasi sebesar 75,90%, presisi sekitar 66,45%, dan recall mencapai 97,42%. Evaluasi ini memberikan pemahaman tentang distribusi sentimen pada ulasan, yang mayoritasnya adalah sentimen positif. Hasil ini memberikan wawasan berharga bagi pemilik bisnis atau pemasar dalam mengambil langkah-langkah selanjutnya untuk meningkatkan produk atau layanan, memperbaiki area dengan sentimen negatif, dan mempertahankan aspek yang mendapat respons positif dari pelanggan. Dengan demikian, algoritma *Naive Bayes* dapat digunakan secara efektif untuk menganalisis sentimen ulasan produk makanan dan minuman di Tokopedia dan memberikan dasar yang kokoh untuk pengembangan strategi bisnis yang lebih baik. Evaluasi Kinerja Model *Naive Bayes* dengan Hasil evaluasi model *Naive Bayes* menunjukkan akurasi sebesar 75,90%, presisi sebesar 66,45%, dan *recall* sebesar 97,42%. Meskipun akurasi yang relatif baik, presisi yang sedikit lebih rendah menunjukkan potensi untuk meningkatkan kemampuan model dalam mengklasifikasikan ulasan dengan lebih tepat, terutama dalam mengidentifikasi sentimen negatif dan

netral. Hasil implementasi data baru dengan Prediksi positif memiliki nilai kepercayaan (confidence) sebesar 0.454, Prediksi negatif ini memiliki nilai kepercayaan sebesar 0.454. Selain itu, terdapat sejumlah kecil prediksi yang diklasifikasikan sebagai netral dengan tingkat kepercayaan sebesar 0.092. Secara keseluruhan, implementasi data baru menghasilkan prediksi sentimen yang didominasi oleh sentimen positif dengan tingkat kepercayaan yang relatif tinggi, disusul dengan adanya beberapa prediksi sentimen negatif dan hanya sedikit prediksi netral. Hasil implementasi menggunakan *apply model* ini bisa diterapkan untuk analisis sentimen pada data ulasan baru dengan model yang sudah dibuat.

Berdasarkan hasil penelitian ada beberapa saran yang diberikan untuk penelitian selanjutnya diantaranya sebagai berikut: Penelitian ini bisa membahas lebih mendalam dengan melibatkan sentimen netral yang ada dalam ulasan pengguna. Sehingga bisa memberi kita pandangan lebih komplit tentang bagaimana pengguna menanggapi produk makanan dan minuman di Tokopedia. Pada penelitian berikutnya dapat menggunakan algoritma yang berbeda untuk dapat memberikan hasil perbandingan serta model dan mendapatkan hasil yang lebih baik.

#### DAFTAR PUSTAKA

- [1] R. L. Atimi and E. E. Pratama, "Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia," *J. Sains dan Inform.*, vol. 8, no. 1, pp. 88–96, 2022, doi: 10.34128/jsi.v8i1.419.
- [2] A. F. Watratan, A. P. Puspita, and D. Moeis, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid," *Jurnal Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 38–46, 2022, doi: 10.55606/jurritek.v1i1.127.
- [3] H. Yuyun, S. Nurul, and Supriadi, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 820–826, 2021, doi: 10.29207/resti.v5i4.3146.
- [4] V. Fitriyana, L. Hakim, D. C. R. Novitasari, and A. H. Asyhar, "Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine," *J. Buana Inform.*, vol. 14, no. 01, pp. 40–49, 2023, doi: 10.24002/jbi.v14i01.6909.
- [5] K. Akmal, A. Faqih, and F. Dikananda, "Perbandingan Metode Algoritma Naïve Bayes Dan K-Nearest Neighbors Untuk Klasifikasi Penyakit Stroke," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 470–477, 2023, doi: 10.36040/jati.v7i1.6367.
- [6] M. K. Insan, U. Hayati, and O. Nurdianan, "Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 478–483, 2023, doi: 10.36040/jati.v7i1.6373.
- [7] I. R. Padiku, "Penerapan Metode Naive Bayes Classifier (Nbc) Untuk Klasifikasi Kondisi Internal Program Studi," *J. Tek.*, vol. 19, no. 1, pp. 65–74, 2021, doi: 10.37031/jt.v19i1.118.
- [8] I. W. Saputro and B. W. Sari, "Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa," *Creat. Inf. Technol. J.*, vol. 6, no. 1, p. 1, 2020, doi: 10.24076/citec.2019v6i1.178.
- [9] I. S. Thalib, S. K. Gusti, F. Yanto, and M. Affandes, "Klasifikasi Sentimen Tragedi Kanjuruhan Pada Twitter Menggunakan Algoritma Naïve Bayes," *J. Sist. Komput. dan Inform.*, vol. 4, no. 3, p. 467, 2023, doi: 10.30865/json.v4i3.5852.
- [10] S. A. Harahap, R. Aprilia, and R. S. Lubis, "Penerapan Metode Naïve Bayes Dalam Klasifikasi Kabupaten/Kota Di Sumatera Utara Berdasarkan Produktivitas Pangan Padi," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 4, no. 1, pp. 79–90, 2023, doi: 10.46306/lb.v4i1.202.