

## ***Naïve Bayes Classifier untuk Analisis Sentimen Ulasan Pelanggan pada Domo Coffee and Resto***

**Puji Hartini<sup>1\*</sup>, Nana Suarna<sup>2</sup>, Willy Prihartono<sup>3</sup>**

<sup>1,2</sup>Jurusan Teknik Informatika, STMIK IKMI CIREBON

<sup>3</sup>Jurusan Komputerisasi Akuntansi, STMIK IKMI CIREBON

\*pujihartini016@gmail.com

### **Abstrak**

Domo *Coffee and resto* adalah salah satu café ternama yang terletak di Jl. DR Sudarsono No.45 Kesambi, Kecamatan Kesambi, Kota Cirebon. Domo Coffe and Resto memiliki beragam makanan dan minuman yang disajikan serta tempat yang dirancang sedemikian rupa agar indah dan nyaman untuk untuk dikunjungi dengan berbagai tujuan. Tentulah ada banyak ragam masalah yang terkait pelayanan yang kurang memuaskan, suasana yang tidak nyaman atau rasa makanan yang kurang enak serta beberapa kekecewaan dan ketidakpuasan lainnya yang memunculkan komentar atau ulasan negatif. *Café Domo* sering menerima beragam ulasan dari pelanggan di platform *Google review*. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pelanggan pada Domo *Coffee and resto* dan akan diselesaikan dengan menggunakan metode *Naïve Bayes Classifier* yaitu metode klasifikasi berdasarkan teorema Bayes. Dalam penelitian ini Berdasarkan pemahaman penulis terhadap kalimat mengenai analisis sentimen, penulis mendapatkan 374 ulasan Positif dan 58 ulasan Negatif terhadap makanan. 469 ulasan positif dan 40 ulasan negatif terhadap suasana dan mendapatkan 253 ulasan positif dan 99 ulasan negatif terhadap pelayanan. Jumlah komentar positif tertinggi didapatkan oleh aspek suasana dengan 469 ulasan dan komentar negatif tertinggi didapatkan oleh aspek pelayanan dengan 99 ulasan. Pada pengujian nilai *split data* 0.8 dan 0.2 akurasi tertinggi didapat oleh aspek pelayanan dengan akurasi sebesar 98.22%, presisi 97.58%, Recall 100% dan nilai *F1-score* 98.78%. Hasil penelitian ini memberikan wawasan mendalam tentang pandangan pelanggan terhadap *cafe Domo*. Pemilik dan pemangku kepentingan kafe dapat menggunakan temuan ini untuk memahami aspek-aspek yang perlu diperbaiki atau ditingkatkan.

**Kata kunci:** Analisis Sentiment, Google Review, Naïve Bayes

### **Abstract**

*Domo Coffee and resto is one of the well-known cafes located on Jl. DR Sudarsono No.45 Kesambi, Kesambi District, Cirebon City. Domo Coffee and Resto has a variety of food and drinks served and the place is designed to be beautiful and comfortable to visit for various purposes. Of course, there are many kinds of problems related to unsatisfactory service, uncomfortable atmosphere or bad taste of food as well as several other disappointments and dissatisfaction that give rise to negative comments or reviews. Café Domo often receives mixed reviews from customers on the Google review platform. This research aims to analyze the sentiment of customer reviews on Domo Coffee and restaurant and will be completed using the Naïve Bayes Classifier method, namely a classification method based on Bayes' theorem. In this research, based on the author's understanding of sentences regarding sentiment analysis, the author received 374 positive reviews and 58 negative reviews regarding food. 469 positive reviews and 40 negative reviews regarding the atmosphere and 253 positive reviews and 99 negative reviews regarding the service. The highest number of positive comments was obtained by the atmosphere aspect with 469 reviews and the highest negative comments were obtained by the service aspect with 99 reviews. In testing the split data values of 0.8 and 0.2, the highest accuracy was obtained by the service technician with an accuracy of 98.22%, precision of 97.58%, recall of 100% and an F1-score value of 98.78%. The results of this research provide in-depth insight into customers' views of Domo cafe. Cafe owners and stakeholders can use these findings to understand aspects that need to be improved or improved.*

**Keywords:** Sentiment Analysis, Google Review, Naïve Bayes

## PENDAHULUAN

Perkembangan pesat dalam teknologi informasi telah mengubah paradigma berbagai aspek kehidupan. Dalam era digital ini, ulasan pelanggan di *platform* media sosial dan situs ulasan seperti *Google review* memainkan peran penting dalam memengaruhi reputasi bisnis, terutama dalam industri kuliner. Kafe dan restoran, seperti Domo *Coffee and resto* tidak terkecuali dari dampak ulasan pelanggan ini. Melihat pelanggan berdatangan di berbagai kalangan akan memunculkan beberapa pandangan terhadap tempat yang didatangi tersebut terkait makanan, pelayanan atau suasana. *Review* dan rating yang diberikan membantu pelanggan lain untuk mengevaluasi usaha atau servis yang di berikan dan mereka dapat menentukan pilihan. Terdapat beberapa hal yang perlu diidentifikasi. Pertama, dengan banyaknya ulasan pelanggan yang masuk, pemilik kafe mungkin mengalami kesulitan untuk memahami sentimen keseluruhan dari ulasan-ulasan tersebut secara manual. Kedua, ulasan pelanggan seringkali berisi bahasa yang kompleks dan subyektif, sehingga memerlukan pendekatan analisis yang cermat. Ketiga, pemilik kafe ingin dapat merespons ulasan pelanggan dengan cepat, terutama jika terdapat ulasan negatif yang dapat merusak reputasi bisnis mereka.

Berdasarkan hasil penelitian terdahulu oleh (Sari, Sinduningrum and Noor Hasan, 2023) dengan judul “ Analisis Sentimen Ulasan Pelanggan Pada Aplikasi *Fore Coffee* Menggunakan Metode *Naïve Bayes* ”. Bertujuan mengidentifikasi dan mengklasifikasi sentimen pelanggan pada aplikasi tersebut. Sebanyak 1801 data ulasan pelanggan *Fore Coffee* berhasil diproses mayoritas sentimen yang tercatat adalah positif (1163 ulasan), sedangkan sentimen negatif hanya sebanyak 315 ulasan. Diperoleh skor nilai akurasi sebesar 74.28%, *precision* sebesar 24.46% dan *Recall* sebesar 45.33%. Temuan menunjukkan bahwa sebagian besar pengguna *Fore Coffee* memberikan sentimen positif terhadap pelayanan yang diberikan oleh barista.

Penelitian terdahulu juga dilakukan oleh (Irvandi, Irawan and Nurdiawan, 2023) dengan judul “ *Naive bayes* Dan *Wordcloud* Untuk Analisis Sentimen Wisata Halal Pulau Lombok ”. Penelitian ini dilakukan dengan tujuan ingin mengetahui sejauh mana asumsi ini tercermin dalam opini dan ulasan wisatawan. Setelah

menggunakan analisis sentimen ini hasilnya kata-kata yang sering muncul dalam ulasan wisatawan, seperti "indah," "wisata," "pantai," "alam," "gunung," dan "masjid," mencerminkan aspek positif dari destinasi wisata Lombok. Tingkat akurasi model *Naive bayes*: 74,75%.

Keterbaruan dalam penelitian ini selain populasi, sampel dan objek yang berbeda pada penelitian ini terdapat tiga aspek atau tiga penilaian yang digunakan dalam analisis sentimen terhadap Domo coffee and resto yang menjadikan informasi yang didapatkan lebih spesifik atau lebih terarah tentang ulasan yang dilontarkan oleh pelanggan.

Tujuan penelitian ini untuk menganalisis ulasan pelanggan yang telah mengunjungi Domo *Coffee and resto* dengan pendekatan *Algoritma Naïve Bayes*. Studi ini mengidentifikasi pola sentimen umum yang terkait dengan ulasan pelanggan pada domo *Coffee and resto*. Melalui penelitian ini penulis dapat memperoleh pemahaman yang lebih dalam tentang persepsi pelanggan sebagai konsumen Domo *Coffee and resto* dan wawasan sebagai bahan evaluasi untuk pihak domo perihal apa yang kurang dan harus diperbaiki.

Metode dalam penelitian ini menggunakan pendekatan *Algoritma Naïve Bayes*. Dalam metode ini dilakukan dua proses yaitu proses pelatihan data dengan membuat model klasifikasi, yang kedua proses pengujian dengan menggunakan data uji yang dimasukkan kedalam model klasifikasi yang telah dibuat menggunakan proses pelatihan data. Metode *Naïve Bayes* memiliki keunggulan dalam mengklasifikasikan sentimen positif, negatif, atau netral dari ulasan pelanggan dengan akurasi yang tinggi.

Hasil penelitian ini dapat memberikan pemahaman yang lebih baik tentang sentimen pelanggan terhadap Domo *Coffee and Resto*, yang dapat digunakan oleh pemilik dan pengelola kafe untuk meningkatkan kualitas layanan mereka.

## TINJAUAN PUSTAKA

### 2.1 Analisis Sentimen

Analisis sentimen adalah teknik penggalian informasi berupa sikap seseorang terhadap suatu topik atau peristiwa dengan mengklasifikasikan polaritas suatu teks. Pengelompokan dilakukan untuk melihat apakah teks tersebut positif, negatif, atau netral.

Analisis sentimen dapat digunakan untuk menentukan opini publik tentang suatu topik. (Zusrotun, Murti and Fati, 2022).

## 2.2 Google review

*Google review* merupakan salah satu dari fitur *Google maps*. Fitur sederhana yang mampu berdampak besar ini digunakan untuk memberi nilai angka dan ulasan berupa teks dan foto pada sebuah tempat. Di era *industry 4.0* yang ditandai dengan tersedianya *big data*, pengguna dengan mudah memberikan nilai sesuai dengan faktor-faktor kepuasan pengunjung. *Domo Coffee and resto* yang dicuplik dengan cara *web scrapping* merupakan data yang akan dianalisis berasal dari data ulasan pada situs *Google review*. *Domo Coffee and resto* merupakan salah satu *Café* atau *Restourant* di kota Cirebon. Aplikasi tersebut ditambahkan sebagai *extension* pada *web browser chrome* dengan nama *Data Miner*. (Herlawati *et al.*, 2021)

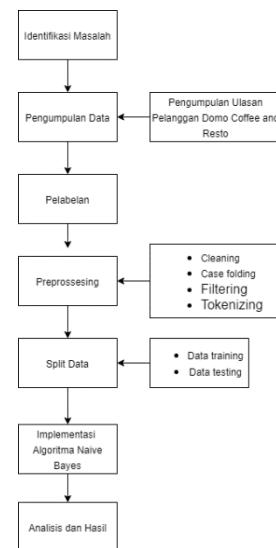
## 2.3 Teori Naïve Bayes

*Algoritma Naïve bayes* merupakan salah satu *Algoritma* yang sering digunakan pada penelitian dalam upaya mengatasi klasifikasi pada teks dengan menggunakan perhitungan probabilitas. *Naïve bayes* adalah salah satu *Algoritma* yang banyak digunakan pada penelitian dengan jumlah data yang banyak dan data yang kosong dengan menggunakan perhitungan probabilitas, serta *statistic* dalam prediksi suatu data terkait dari pengujian yang telah dilakukan sebelumnya istilah tersebut biasa dikenal dengan *missing value*. (Putri, Triayudi and Aldisa, 2023) .

## METODE PENELITIAN

Penelitian yang dipergunakan dalam penelitian ini yaitu melalui pendekatan penelitian kuantitatif. Penelitian kuantitatif yaitu penelitian yang memakai data berupa angka sebagai alat untuk menganalisis keterangan tentang apa yang ingin diketahui. Penelitian ini menggunakan pendekatan kuantitatif. Pendekatan kuantitatif, biasanya menekankan analisisnya pada data-data numerik atau angka yang digunakan untuk meneliti pada populasi atau sampel tertentu sesuai dengan banyaknya populasi atau sample

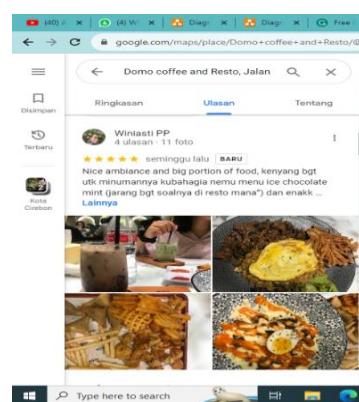
yang akan di uji. Pada Gambar 1 merupakan alur dari penelitian.



Gambar 1. Diagram Alir

## 3.1 Sumber Data

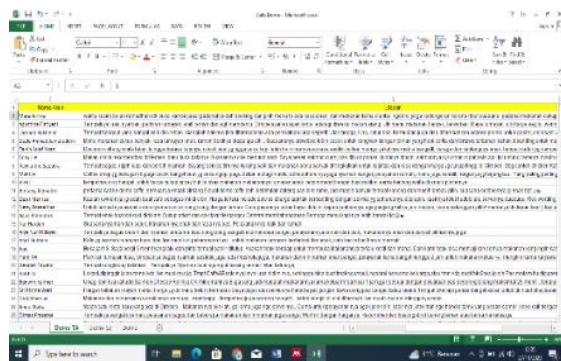
Penelitian ini menggunakan data sekunder yang didapat melalui website yang mencakup ulasan-ulasan pelanggan yang telah diposting di *Google review cafe* Domo oleh pengguna platform tersebut. Data ini diakses melalui *Google review*, yang merupakan sumber data publik yang dapat diakses secara *online* seperti pada Gambar 2.



Gambar 2. Ulasan Google Review

Data yang diambil dari ulasan pelanggan di *Google review* dianggap relevan dengan tujuan penelitian, yaitu untuk melakukan analisis sentimen terhadap *cafe* Domo. Ulasan pelanggan mencerminkan pandangan dan pengalaman pelanggan terhadap kafe tersebut.

Sumber data yang dipilih akan menjadi dasar utama untuk analisis sentimen menggunakan metode naïve bayes dalam penelitian ini. Data ini akan dianalisis untuk memahami sentimen pelanggan terhadap *cafe Domo* dan memberikan wawasan yang dapat digunakan untuk tujuan penelitian. Penting untuk menjaga integritas data dan memastikan bahwa data yang digunakan adalah representatif dari pengalaman pelanggan sebenarnya.



Gambar 3. Web Scraping manual

### 3.2 Populasi dan Sampel

Dalam penelitian ini, populasi yang akan diteliti adalah semua ulasan pelanggan yang terdapat pada *Google review cafe Domo* yang memiliki 1000 ulasan dan dua atribut yang akan diambil pada pemrosesan datanya yaitu atribut ulasan dan atribut sentimen. Populasi dalam penelitian ini adalah semua ulasan yang diberikan oleh pelanggan *cafe Domo* di *Google review*. Ini mencakup ulasan-ulasan yang mencakup berbagai pengalaman dan pendapat pelanggan tentang kafe tersebut. Ulasan yang berisi teks pada *google review Domo Coffee and resto* hanya terdapat 610 ulasan yang akan dibagi lagi kedalam tiga aspek yaitu ulasan terhadap makanan, suasana dan pelayanan.

### 3.3 Teknik Pengumpulan Data

#### 1. Google review

Data yang digunakan dalam penelitian ini adalah data ulasan mengenai ulasan pelanggan yang diambil dari *platform google review* pada *Domo Coffee and resto*.

#### 2. Web scraping

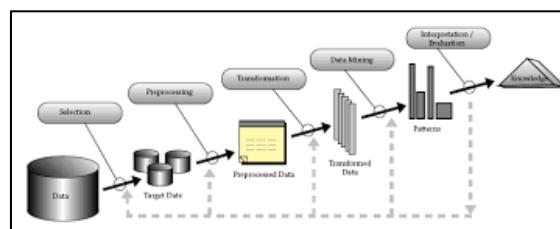
Pengumpulan data dilakukan dengan cara *input* manual dan disimpan di Microsoft Excel.

### 3. Studi Pustaka

Penelitian ini dilakukan berdasarkan beberapa kajian literature, jurnal dan sumber internet lainnya yang berkaitan dengan metode penelitian, tujuan serta rumusan masalah yang berkaitan dengan penelitian serupa yang dilakukan.

### 3.5 Teknik Analisis Data

Metode analisa data yang digunakan pada penelitian ini menggunakan metode *Knowledge Discovery in Database* (KDD), dengan beberapa tahap mulai dari *Selection*, *Preprocessing*, *Transformation*, *Data Mining*, hingga *Evaluation*.



Gambar 4. Tahapan dalam KDD

Sumber: [www.research.net](http://www.research.net)

Penjelasan rinci tentang tahapan tersebut sebagai berikut:

#### 1. Selection

Data hasil seleksi yang digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

#### 2. Preprocessing

Proses ini bertujuan untuk menyeleksi data yang benar-benar akan digunakan pada proses pembentukan model. Beberapa hal yang dilakukan pada proses ini seperti membuang atribut yang tidak perlu, membuang duplikasi data, memeriksa data mana yang inkonsisten, dan memperbaiki kesalahan yang ada pada data, seperti kesalahan cetak (tipografi).

#### 3. Transformation

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Pada tahap ini yang telah siap untuk diolah kemudian disimpan ke dalam format yang bisa dibaca oleh aplikasi *data mining* seperti disimpan ke dalam format CSV, MySQL, dll. (Azhari, Situmorang and Rosnelly, 2021) Pada

penelitian ini penulis menyimpannya dengan format *Excel*.

#### 4. Data Mining

Pada tahap ini, data yang sudah didapatkan akan digunakan pada metode tertentu untuk membentuk pola klasifikasi. Pola dibentuk dengan menggunakan data training dan metode klasifikasi. Hasil yang berupa pola klasifikasi selanjutnya digunakan untuk data testing agar mengetahui performa dari metode klasifikasi yang digunakan.

#### 5. Evaluation

Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan berlawanan dengan fakta atau hipotesis yang ada sebelumnya. Ada tiga nilai yang digunakan untuk mengukur kemampuan sistem klasifikasi yang dibangun, yakni *precision*, *Recall* dan *accuracy*. Dalam pengukuran performa klasifikasi ada beberapa cara, namun cara yang digunakan dalam evaluasi hasil klasifikasi program studi dengan menghitung akurasi, *precision*, *Recall*. Akurasi adalah persentase dari hasil semua klasifikasi yang benar dikenali. Performa untuk perhitungan ini akan disajikan dengan *Confusion Matrix*. Perhitungan akurasi dilakukan dengan cara membagi jumlah data yang diprediksi benar dengan total data uji. Menghitung nilai akurasi dengan menggunakan persamaan. Pencarian nilai akurasi dapat dihitung dengan persamaan 1.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Metode yang digunakan selain perhitungan akurasi untuk mengukur kinerja pengklasifikasi adalah perhitungan *Precision*, dan *Recall*. Precision berarti hasil yang mendekati kejadian sebenarnya. Presisi ini memastikan ukuran hasil yang tepat ditentukan selama proses ekstraksi. Presisi juga merupakan metrik standar untuk menentukan apakah dokumen atau kumpulan data relevan dengan tujuan yang dimaksudkan

*Precision* (Presisi) adalah perbandingan antara jumlah data relevan yang ditemukan dan jumlah data yang ditemukan. Presisi dihitung dengan membagi data benar dengan nilai positif dengan data benar dengan nilai positif dan

data salah dengan nilai positif. Nilai data positif palsu untuk diambil dari jumlah nilai yang berbeda dari kolom positif benar yang sesuai dengan di setiap kelas. Untuk menghitung nilai *Precision* dapat menggunakan persamaan 2.

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (2)$$

Perhitungan lain yang memberikan hasil yang baik adalah *callback*. *Recall* adalah ukuran pentingnya dokumen yang ada untuk hasil tertentu. *Recall* adalah perbandingan dari jumlah materi relevan yang ditemukan terhadap jumlah materi yang relevan. Perhitungan *Recall* dilakukan dengan cara membagi data benar bernilai positif dengan hasil penjumlahan dari data benar yang bernilai positif dan data salah yang bernilai negatif. Nilai dari data salah yang bernilai negatif diambil dari jumlah nilai selain true positive baris yang sesuai tiap kelasnya. Perhitungan *Recall* dapat menggunakan Persamaan 3.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

*F1-score* merupakan nilai rata-rata dari penjumlahan presisi dan *Recall*. *F1-score* digunakan untuk menyeimbangkan nilai presisi dan juga *Recall*. Persamaan 4 dapat digunakan untuk menghitung *F1-score*.

$$2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Keterangan:

1. TP adalah *True Positive*, yaitu jumlah data yang nilai aktualnya adalah positif dan terkласifikasi sebagai positif.
2. TN adalah *True Negative*, yaitu jumlah data yang nilai aktualnya adalah negatif dan terkласifikasi sebagai negatif.
3. FN adalah *False Negative*, yaitu jumlah data yang nilai aktualnya adalah negatif namun terklasifikasi salah.
4. FP adalah *False Positive*, yaitu jumlah data yang nilai aktualnya adalah positif namun terklasifikasi salah.

## HASIL DAN PEMBAHASAN

### 4.1 Data Selection

Dataset ulasan dari pelanggan diperoleh dari *review* pada Situs *Google review*. Domo

*Coffee and resto* mendapatkan 1000 data ulasan dari pelanggan yang sudah dan sempat mengunjunginya, namun ulasan yang memiliki teks dalam penelitian ini hanya mendapatkan 610 data dimana didalamnya terdapat ulasan yang memiliki sentiment Positif dan Negatif kemudian di *input* dan diberikan label secara manual. Setiap ulasan memiliki tiga pandangan dan penilaian oleh sebab itu ulasan akan dibagi menjadi tiga aspek yaitu ulasan terhadap makanan, suasana dan pelayanan. Namun perlu diingat bahwa tidak semua ulasan juga memiliki tiga penilaian atau komentar sekaligus, beberapa ulasan hanya tertuju pada satu aspek saja. Data ulasan yang sudah di *input*, dilakukan pembagian aspek dan dilakukan pelabelan kemudian di simpan didalam *Microsoft excel*. Pada Tabel 1 merupakan hasil dari penginputan data secara manual dari *google review*.

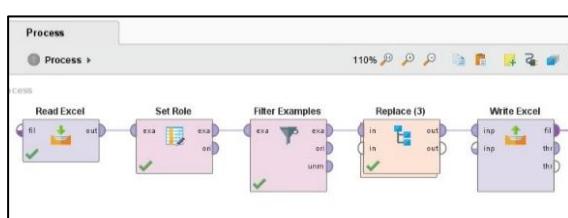
Tabel 1. Hasil Input

No	Text
1	Tempatnya bagus banget, makannya banyak, enak enaaaak
2	tempat ngopi dan ngobrol yg cozy. makanan dan minumannya jd enak. mungkin agak hectic karena tempat baru dan berbarengan dgn bulan puasa (bukber) orderan ada yg agak lama keluar. tapi semuanya bagusss 🌟⭐
3	makanannya enak banget .. porsinya banyak.. kenyang 😊😊😊
4	Pesan bento, beef teriyakinya keras buang. Paraahhhh
5	Tempat nya nyaman makanan minuman enak 🙌👍 ...
610	Kafe terbaik di kota cirebon

## 4.2 Preprocessing

### 1. Data Cleansing

Pada tahap ini digunakan fitur replace untuk menghilangkan karakter, tanda baca, angka dan emoji yang tidak diperlukan.



Gambar 5. Data Cleansing

Hasil dari penggunaan fitur replace untuk menghilangkan karakter, tanda baca, angka dan emoji yang tidak diperlukan dapat dilihat pada Tabel 2.

Tabel 2. Hasil Replace

Input	Output
servicenya JELEK bgt. dateng2 dibiarin ngang ngong sndiri, ga ada yg nunjukin meja kosong. boro2 deh, even greetings "selamat datang" jd ga ada. cuma ditanya "udh reservasi kak?" pas dijawab "engga, kak", terus ditinggal	servicenya jelek bgt dateng dibiarin ngang ngong sndiri ga ada yg nunjukin meja kosong boro deh even greetings selamat datang jd ga ada cuma ditanya udh reservasi kak pas dijawab engga kak terus ditinggal wow makasih?? 🌟⭐

### 2. Labeling Manual

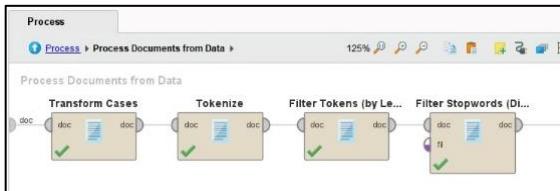
Setelah data terkumpul tahap selanjutnya yaitu dengan melakukan pelabelan manual. Pelabelan dilakukan dengan mengklasifikasikan komentar kedalam dua kelas, yaitu positif dan negatif. Cara kerja dari pelabelan ini yaitu dengan memahami makna dari kalimat, bukan dilakukan per kata. (Apriliani *et al.*, 2023) Proses pelabelan data berjumlah 619 dibagi menjadi tiga aspek dan dilakukan secara manual dengan kategori sentimen positif dan negatif. (Irvandi, Irawan and Nurdianwan, 2023)

Tabel 3. Labeling Manual

No	Text	Sentiment
1	Tempatnya bagus banget, makannya banyak, enak enaaaak	Positif
2	tempat ngopi dan ngobrol yg cozy. makanan dan minumannya jd enak. mungkin agak hectic karena tempat baru dan berbarengan dgn bulan puasa (bukber) orderan ada yg agak lama keluar. tapi semuanya bagusss 🌟⭐	Positif
3	makanannya enak banget .. porsinya banyak.. kenyang 😊😊😊	Positif
4	Pesan bento, beef teriyakinya keras buang. Paraahhhh	Negatif
5	Tempat nya nyaman makanan minuman enak 🙌👍 ...	Positif
610	Kafe terbaik di kota cirebon	Negatif

### 3. Preprocessing

Tahapan *text preprocessing* dengan menggunakan *Rapid Miner* dapat dilihat pada Gambar 6.



Gambar 6. Preprocessing

Keterangan dari Gambar 6 sebagai berikut:

- Transform Cases* : mengubah semua jenis huruf menjadi huruf kecil.
- Tokenize* : memisahkan atau memenggal kata dalam suatu kalimat.
- Filter Tokens (by Length)*: Mengondisikan kata yang akan tersaring hanya memiliki panjang karakter tertentu dengan minimal 4 karakter dan maksimal 22 karakter.
- Filter Stopwords* : menghapus kata tertentu yang tidak dibutuhkan.

Hasil *preprocessing* dapat dilihat pada Tabel 4.

Tabel 4. Hasil Preprocessing

No	Text	Sentiment
a	servicenya JELEK bgt. dateng2 dibiarin ngang ngong sndiri, ga ada yg nunjukin meja kosong. boro2 deh, even greetings "selamat datang" jg ga ada. cuma ditanya "udh reservasi kak?" pas dijawab "engga, kak", terus ditinggal. wow, makasih??	servicenya jelek bgt dateng dibiarin ngang ngong sndiri ga ada yg nunjukin meja kosong boro deh even greetings selamat datang jg ga ada cuma ditanya udh reservasi kak pas dijawab engga kak terus ditinggal wow makasih
b	servicenya JELEK bgt. dateng2 dibiarin ngang ngong sndiri, ga ada yg nunjukin meja kosong. boro2 deh, even greetings "selamat datang" jg ga ada. cuma ditanya "udh reservasi kak?" pas dijawab "engga, kak", terus ditinggal. wow, makasih??	'servicenya' 'jelek' 'bgt' 'dateng' 'dibiarin' 'ngang' 'ngong' 'sndiri' 'ga' 'ada' 'yg' 'nunjukin' 'meja' 'kosong' 'boro' 'deh' 'even' 'greetings' 'selamat' 'datang' 'jg' 'ga' 'ada' 'cuma' 'ditanya'

c	servicenya JELEK bgt. dateng2 dibiarin ngang ngong sndiri, ga ada yg nunjukin meja kosong. boro2 deh, even greetings "selamat datang" jg ga ada. cuma ditanya "udh reservasi kak?" pas dijawab "engga, kak", terus ditinggal. wow, makasih??	'udh' 'reservasi' 'kak' 'pas' 'dijawab' 'engga' 'kak' 'terus' 'dinggal' 'wow' 'makasih'
d	servicenya JELEK bgt. dateng2 dibiarin ngang ngong sndiri, ga ada yg nunjukin meja kosong. boro2 deh, even greetings "selamat datang" jg ga ada. cuma ditanya "udh reservasi kak?" pas dijawab "engga, kak", terus ditinggal. wow, makasih??	'servicenya' 'jelek' 'dateng' 'dibiarin' 'ngang' 'ngong' 'sndiri' 'nunjukin' 'meja' 'kosong' 'boro' 'even' 'greetings' 'selamat' 'datang' 'jg' 'ga' 'ada' 'cuma' 'ditanya'

### 4.3 Transformation

Pada tahap ini dilakukan *Term Weighting* atau mengimplementasikan metode *TF-IDF*. Pada Gambar 7 contoh mengitung *tf-idf* secara manual.

Kosakata	TF			df	N/df	Idf
	d1	d2	d3			
makan	1	1	1	3	1	0
enak	1	0	1	2	1,5	0,176
banget	1	1	0	2	1,5	0,176
porsi	1	0	0	1	3,0	0,477
banyak	1	0	0	1	3,0	0,477
kenyang	1	0	0	1	3,0	0,477
Mantap	0	1	0	1	3,0	0,477
minum	0	1	1	2	1,5	0,176
pas	0	1	0	1	3,0	0,477
suasana	0	1	0	1	3,0	0,477
satai	0	1	0	1	3,0	0,477
denger	0	1	0	1	3,0	0,477
musik	0	1	0	1	3,0	0,477
live	0	1	0	1	3,0	0,477
tempat	0	0	1	1	3,0	0,477
cozy	0	0	1	1	3,0	0,477
asik	0	0	1	1	3,0	0,477
nyaman	0	0	1	1	3,0	0,477
rekomeded	0	0	1	1	3,0	0,477

Gambar 7. Menghitung TF-IDF manual

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Setiap kata akan diberikan bobot guna mengubah data ke dalam bentuk numerik. Sehingga diketahui nilai kepentingan dari suatu kata pada sebuah dokumen berdasarkan intensitas kemunculan kata.

#### 4.4 Data Mining

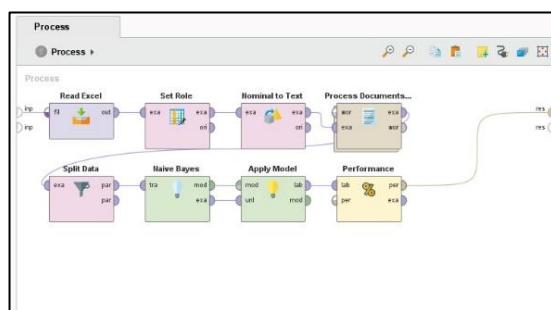
Dalam tahap ini data yang sudah di *preprocessing* atau dibersihkan akan dibagi menjadi data training atau data latih dan data testing atau data uji dengan perbandingan nilai ratio 0.8:0.2. Data latih digunakan untuk membuat model machine learning, sedangkan data uji digunakan untuk mengevaluasi kinerja model tersebut. Pembagian data training (data latih) dan data testing (data uji) tertera pada Tabel 8.

Tabel 8. Pembagian Split Data

Aspek	Pengujian 0.8 dan 0.2		Total
	Train	Test	
Makanan	345	87	432
Suasana	407	102	509
Pelayanan	281	71	352

#### 4.5 Evaluation

Model Klasifikasi *Algoritma Naive bayes*. Analisis pengklasifikasian menggunakan *Algoritma Naive bayes* menggunakan RapidMiner yang operatornya sudah disesuaikan dapat dilihat pada Gambar 8.



Gambar 8. Tahapan Evaluasi

#### Keterangan Gambar 8:

- Read Excel*: melakukan pembacaan data berextensi file xlsx yang merupakan dataset hasil pemrosesan data cleansing yang sudah

diberikan keterangan sentimen positif dan negatif.

- Set Role*: mengondisikan kolom sentimen dijadikan sebagai label pada pemrosesan *Algoritma Naive bayes*
- Nominal to Text*: mengubah jenis data nominal menjadi text yang akan diteruskan pada operator berikutnya.
- Process Documents from Data*: operator ini mengonversi data text yang digunakan dalam dengan pembobotan *TF-IDF*, memiliki subproses yang diisi sebagaimana tahap text preprocessing.
- Split data*: *Split data* dilakukan untuk membagi dataset menjadi partisi data latih dan data uji sesuai porsi yang ditentukan
- Implementasi Naïve Bayes*: Menerapkan *Algoritma* yang dipakai
- Apply Model*: Operator *apply model* digunakan untuk menerapkan model yang telah dilatih sebelumnya menggunakan data training pada data testing.
- Performance*: digunakan untuk mengevaluasi kinerja model yang memberikan daftar nilai kriteria kinerja secara otomatis sesuai dengan tugas yang diberikan.

Tabel 9. Confusion Matrix

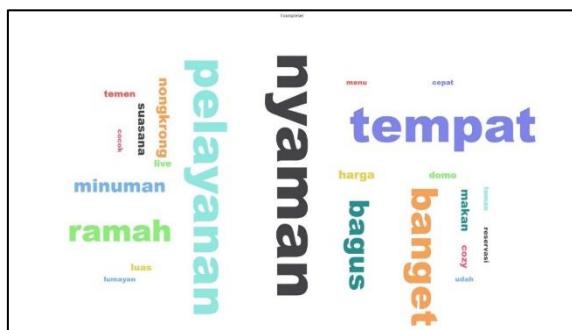
Aspek	Sentimen		TN	TP	FN	FP
	P	N				
Makanan	374	58	46	241	0	58
Suasana	469	40	32	339	0	36
Pelayanan	253	99	74	202	5	0

Berdasarkan penelitian pada pengujian nilai 0.8 dan 0.2 diperoleh ulasan Domo Coffee and resto pada kelas *sentiment* positif dan diklasifikasikan dengan benar (*True Positive* / TP) sebanyak 241 untuk aspek makanan, 339 untuk aspek suasana dan 202 untuk aspek pelayanan. Sedangkan hasil klasifikasi salah (*False Positive* / FP) pada kelas *sentiment* positif adalah sebanyak 58 ulasan untuk aspek makanan, 36 untuk aspek suasana dan 0 untuk aspek pelayanan. Untuk hasil klasifikasi benar pada kelas *sentiment* negatif (*True Negative* / TN) terdiri dari 46 ulasan untuk aspek makanan, 32 untuk aspek suasana dan 74 untuk aspek pelayanan, sedangkan hasil klasifikasi salah pada kelas *sentiment* negatif (*False Negative*) adalah 0 ulasan untuk aspek

makanan, 0 untuk aspek suasana dan 5 untuk aspek pelayanan. Dari hasil *Confusion Matrix* maka dilakukan perhitungan nilai akurasi dengan persamaan 1 sehingga diperoleh akurasi 83,19% untuk makanan, 91,15% untuk suasana dan 98,22% untuk pelayanan.

#### 4.6 Visualization

Visualisasi ditampilkan dengan tools *wordcloud* pada *rapidminer*. Visualisasi ini adalah gambaran ulasan apa saja yang banyak didapatkan dalam keseluruhan komentar dari pelanggan terhadap Domo *Coffee and resto*. Berikut adalah hasil *wordcloud* keseluruhan data dari teks yang sering muncul dapat dilihat pada Gambar 9.



Gambar 9. Wordcloud

#### 4.7 Pembahasan

Analisis sentimen terhadap Domo *Coffee and resto* memiliki keseluruhan 1000 ulasan yang berdasarkan rating dan teks. Ulasan berisikan teks hanya terdapat 610 data dimana dalam satu ulasan terdapat tiga komentar atau pandangan yaitu terhadap makanan, suasana dan pelayanan. Berdasarkan pemahaman penulis terhadap kalimat mengenai analisis ini penulis mendapatkan 374 ulasan Positif dan 58 ulasan Negatif terhadap makanan. 469 ulasan positif dan 40 ulasan negatif terhadap suasana dan mendapatkan 253 ulasan positif dan 99 ulasan negatif terhadap pelayanan. Jumlah komentar positif tertinggi didapatkan oleh aspek suasana dengan 469 ulasan dan komentar negatif tertinggi didapatkan oleh aspek pelayanan dengan 99 ulasan. Pada pengujian nilai *split data* 0.8 dan 0.2 akurasi tertinggi didapat oleh aspek pelayanan dengan akurasi sebesar 98.22%, presisi 97.58%, Recall 100% dan nilai *F1-score* 98.78%. Sedangkan untuk *wordcloud* menghasilkan kata yang sering muncul dalam keseluruhan ulasan seperti " Nyaman "," Tempat "," Nongkrong ","Cozy "," Cocok " " Luas ", "Suasana"," Pelayanan ", " Bagus ", " Makan " , " Ramah " , " live ", " domo ". Melihat dari kata yang sering muncul mayoritas merupakan kata yang ditujukan kepada suasana atau tempat dan hasil ini sesuai dengan sentimen positif tertinggi yang didapatkan oleh aspek suasana.

Tabel 10. Hasil Perhitungan pemodelan

Aspek	Akurasi	Presisi	Recall	F1-Score
Makanan	83,19%	44,23%	100%	61,33%
Suasana	91,15%	47,06%	100%	64,00%
Pelayanan	98,22%	97,58%	100%	98,88%

#### SIMPULAN

Model yang dihasilkan oleh naïve bayes dalam mengolah data banyak atau sedikit memiliki akurasi yang baik dan dapat dikatakan cukup tinggi yang berarti pemodelan dalam analisis sentimen ini dianggap sesuai untuk pengklasifikasian aspek dan kelas terhadap sentimen. Dalam penelitian ini berdasarkan pemahaman penulis terhadap kalimat mengenai analisis ini penulis mendapatkan 374 ulasan Positif dan 58 ulasan Negatif terhadap makanan. 469 ulasan positif dan 40 ulasan negatif terhadap suasana dan mendapatkan 253 ulasan positif dan 99 ulasan negatif terhadap pelayanan. Jumlah komentar positif tertinggi didapatkan oleh aspek suasana dengan 469 ulasan dan komentar negatif tertinggi didapatkan oleh aspek pelayanan dengan 99 ulasan. Pada pengujian nilai *split data* 0.8 dan 0.2 akurasi tertinggi didapat oleh aspek pelayanan dengan akurasi sebesar 98.22%, presisi 97.58%, Recall 100% dan nilai *F1-score* 98.78%. Sedangkan untuk *wordcloud* menghasilkan kata yang sering muncul dalam keseluruhan ulasan seperti " Nyaman "," Tempat "," Nongkrong ","Cozy "," Cocok " " Luas ", "Suasana"," Pelayanan ", " Bagus ", " Makan " , " Ramah " , " live ", " domo ". Melihat dari kata yang sering muncul mayoritas merupakan kata yang ditujukan kepada suasana atau tempat dan hasil ini sesuai dengan sentimen positif tertinggi yang didapatkan oleh aspek suasana.

Untuk penelitian selanjutnya pada pelabelan untuk mencari efektivitas waktu lebih baik menggunakan library python. Selain itu juga mencoba algoritma lain selain naïve bayes untuk analisis sentimen seperti *Support Machine Vectorn (SVM)*, *K Nearest Neighboor (K-NN)* dan lain sebagainya.

**DAFTAR PUSTAKA**

- Apriliani, D. *et al.* (2023) ‘Sentimen Analisis Pandangan Masyarakat Terhadap Vaksinasi Covid 19 Menggunakan K-Nearest Neighbors’, *Jurnal Informatika: Jurnal Pengembangan IT*, 8(1), pp. 34–37. Available at: <https://doi.org/10.30591/jpit.v8i1.4759>.
- Azhari, M., Situmorang, Z. and Rosnelly, R. (2021) ‘Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes’, *Jurnal Media Informatika Budidarma*, 5(2), p. 640. Available at: <https://doi.org/10.30865/mib.v5i2.2937>.
- Herlawati, H. *et al.* (2021) ‘Analisis Sentimen Pada Situs Google Review dengan Naïve Bayes dan Support Vector Machine’, *Jurnal Komtika (Komputasi dan Informatika)*, 5(2), pp. 153–163. Available at: <https://doi.org/10.31603/komtika.v5i2.6280>.
- Irvandi, Irawan, B. and Nurdiawan, O. (2023) ‘Naïve Bayes Dan Wordcloud Untuk Analisis Sentimen Wisata Halal Pulau Lombok’, *INFOTECH journal*, 9(1), pp. 236–242. Available at: <https://doi.org/10.31949/infotech.v9i1.5322>.
- Putri, T.A.Q., Triayudi, A. and Aldisa, R.T. (2023) ‘Implementasi Algoritma Decision Tree dan Naïve Bayes Untuk Klasifikasi Sentimen Terhadap Kepuasan Pelanggan Starbucks’, *Journal of Information System Research (JOSH)*, 4(2), pp. 641–649. Available at: <https://doi.org/10.47065/josh.v4i2.2949>.
- Sari, T.A., Sinduningrum, E. and Noor Hasan, F. (2023) ‘KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Sentimen Ulasan Pelanggan Pada Aplikasi Fore Coffee Menggunakan Metode Naïve Bayes’, *Media Online*, 3(6), pp. 773–779. Available at: <https://doi.org/10.30865/klik.v3i6.884>.
- Zusrotun, O.P., Murti, A.C. and Fati, R. (2022) ‘Analisis Sentimen Terhadap Belajar Online pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes’, *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 11(3), pp. 310–319. Available at: <https://doi.org/10.23887/janapati.v11i3.49160>.