

# PERANCANGAN APLIKASI ANALISIS SENTIMEN PADA ULASAN MASKAPAI PENERBANGAN DI TRIPADVISOR MENGGUNAKAN METODE NAÏVE BAYES

Rio Christian<sup>1</sup>, Indah Fenriana<sup>2</sup>

<sup>1,2</sup>Fakultas Sains dan Teknologi, Universitas Buddhi Dharma  
Jalan Imam Bonjol No. 41, Tangerang, Indonesia  
Email: [1ricev7@gmail.com](mailto:ricev7@gmail.com), [2indah.f88@gmail.com](mailto:indah.f88@gmail.com)

## Abstrak

Pesatnya kemajuan teknologi informasi, terutama di bidang *internet* telah mengubah cara orang mencari informasi, termasuk dalam konteks maskapai penerbangan. Maskapai penerbangan termasuk salah satu transportasi yang sangat populer dan penting di dunia. Sebelum melakukan perjalanan, sering kali masyarakat mencari informasi dan ulasan mengenai maskapai penerbangan, salah satunya melalui *platform* Tripadvisor. Karena jumlah ulasan yang banyak di Tripadvisor membuat sulit pengguna untuk cepat memahami sentimennya. Untuk mengatasi masalah tersebut diperlukan analisis sentimen untuk memahami sentimen secara sistematis. Maka dari itu dirancang aplikasi analisis sentimen dengan mengimplementasikan metode Naïve Bayes dalam mengklasifikasikan sentimen menjadi kategori positif dan negatif. Aplikasi ini menampilkan hasil analisis dalam bentuk grafik yang memberikan gambaran visual tentang sentimen positif dan negatif yang terkandung dalam ulasannya. Dengan melakukan pengujian menggunakan metode *confusion matrix* dengan tiga skenario yaitu 70:30, 80:20, dan 90:10. Berdasarkan hasil perbandingan antara tiga skenario tersebut, skenario terbaik untuk model klasifikasi naïve bayes adalah 90:10 yang menunjukkan nilai akurasi sebesar 77%, presisi sebesar 77%, *recall* sebesar 100%, dan F1-Score sebesar 87,1%. Dengan adanya aplikasi ini, diharapkan dapat membantu masyarakat dalam mencari maskapai penerbangan dan membuat keputusan perjalanan yang lebih terinformasi berdasarkan ulasan-ulasan yang telah dianalisis secara efektif.

## Kata Kunci

Maskapai Penerbangan, Klasifikasi, Naïve Bayes

## Latar Belakang

Teknologi informasi telah berkembang dengan cepat sejalan dengan perkembangan zaman, khususnya di bidang *internet*. Setiap tahun, semakin banyak orang yang menggunakan *internet*, mencerminkan pentingnya *internet* dalam kehidupan sehari-hari. Menurut hasil survei yang dilakukan oleh APJII (Asosiasi Penyelenggara Jasa *Internet* Indonesia) pada tahun 2022 hingga 2023, populasi pengguna *internet* Indonesia mencapai 215,63 juta [1]. Angka ini menempatkan Indonesia pada peringkat keempat di dunia dalam hal jumlah pengguna *internet*. Peningkatan jumlah pengguna ini mencerminkan peran yang semakin penting dari *internet* sebagai sumber informasi dalam berbagai aspek kehidupan.

Pengguna *internet* memiliki kecenderungan untuk mencari informasi mengenai berbagai produk atau layanan sebelum mereka mengambil keputusan. Hal ini terkait dengan keinginan mereka untuk mendapatkan pengetahuan yang lebih mendalam tentang produk atau layanan tersebut, termasuk dalam konteks maskapai penerbangan. Maskapai penerbangan termasuk salah satu transportasi yang sangat populer dan penting di dunia. Dianggap sebagai sektor yang paling efisien dalam menghubungkan satu wilayah ke wilayah lain, termasuk dari satu pulau ke pulau lain. Pada tahun 2017, pesawat terbang menjadi moda transportasi terbanyak

kedua dengan 9,3 juta penumpang setelah kereta api sebesar 30,9 juta, berdasarkan data BPS (Badan Pusat Statistik) dari katadata.id [2].

Sebelum melakukan perjalanan, sering kali masyarakat mencari terkait informasi dan ulasan-ulasan mengenai maskapai penerbangan yang ada. Salah satu *platform online* yang banyak digunakan oleh masyarakat dan sangat dikenal di seluruh dunia adalah Tripadvisor [3]. Tripadvisor merupakan salah satu *platform online* yang menyediakan informasi mengenai wisata, restoran, hotel, maskapai, dan berbagai destinasi lainnya di seluruh dunia. *Platform* ini memungkinkan seseorang untuk memberikan ulasan dan penilaian mengenai berbagai tempat dan layanan perjalanan, termasuk maskapai penerbangan. Di sini, seseorang dapat membaca ulasan dan pengalaman perjalanan orang lain tentang berbagai maskapai penerbangan.

Namun, melalui ulasan yang diberikan oleh pengguna di Tripadvisor, karena kuantitasnya yang banyak, membuat sulit bagi pengguna lainnya untuk dengan cepat memahami sentimen yang ada dalam ulasan tersebut. Oleh karena itu, diperlukan metode analisis sentimen yang efektif untuk mengolah dan memahami sentimen pengguna secara sistematis.

Analisis sentimen adalah suatu proses untuk mengekstraksi, mengidentifikasi dan mengevaluasi sentimen yang terkandung dalam teks. Dalam penelitian ini mengimplementasikan metode naïve bayes dengan tujuan untuk mengklasifikasi sentimen menjadi kategori positif dan negatif. Hasil analisis ini dapat memberikan wawasan bagi pihak maskapai penerbangan dalam meningkatkan kualitas pelayanan serta bagi masyarakat dalam membuat keputusan perjalanan yang lebih informasi dan terinformasi.

## Metode Penelitian

### 1. *Data Mining*

*Data Mining* adalah proses penggalian informasi dari sekumpulan data yang besar. Tujuan mendasar dari data mining adalah untuk menemukan pola yang tersembunyi di dalam data untuk mendapatkan informasi atau pengetahuan yang berharga, itu bermanfaat untuk membuat keputusan yang lebih baik [4]. Dalam data mining, terdapat beragam metode dan teknik yang digunakan untuk menganalisis dan menggali informasi dari data. Setiap metode dan teknik tersebut memiliki berbagai fungsi dan tujuan yang berbeda. Berikut adalah pengelompokan data mining berdasarkan fungsi dan tujuannya [5]:

- a. Deskripsi  
Menemukan pola tersembunyi dalam data dan mengubahnya menjadi aturan yang dapat digunakan untuk memudahkan aktivitas atau pengambilan keputusan.
- b. Klasifikasi  
Proses pengklasifikasian data berdasarkan keterkaitan antara variabel kriteria dengan variabel sasaran.
- c. Prediksi  
Mirip dengan klasifikasi, melibatkan upaya untuk memperkirakan atau meramalkan data di masa yang akan datang berdasarkan data sebelumnya.
- d. Estimasi  
Mirip seperti prediksi, tetapi memiliki perbedaan dalam pengelompokan data. Estimasi dilakukan dengan mengelompokkan data secara numerik.
- e. Pengklasteran  
Proses pengelompokan data ke dalam kelompok-kelompok yang memiliki nilai yang mirip berdasarkan karakteristik tertentu.
- f. Asosiasi

Berfungsi untuk mencari hubungan antara atribut yang muncul dalam waktu yang bersamaan.

## 2. *Text Mining*

*Text Mining* adalah teknik yang digunakan untuk menggali informasi dari teks atau dokumen dengan tujuan mengidentifikasi pola, tren, dan hubungan yang tersembunyi dalam data teks. Teknik ini memungkinkan eksplorasi dan analisis yang komprehensif terhadap teks guna mendapatkan pemahaman yang mendalam, yang pada akhirnya dapat digunakan untuk pengambilan keputusan yang lebih baik [6].

## 3. Pengumpulan Data

Pengumpulan data menggunakan ekstensi google chrome yaitu "*Data Scraper*" untuk mengambil data mengenai ulasan terhadap tiga maskapai penerbangan yaitu Citilink, Lion Air, dan Garuda Indonesia pada website Tripadvisor.co.id.

## 4. *Text Preprocessing*

*Text Preprocessing* adalah serangkaian langkah atau proses yang dilakukan pada data teks mentah untuk membersihkannya sebelum dilakukan pemrosesan lebih lanjut. Data teks mentah yang diperoleh umumnya tidak terstruktur dan sering kali mengandung banyak *noise* seperti tanda baca, angka, karakter khusus, dan sebagainya. Oleh karena itu, dilakukan pemrosesan untuk membersihkan, mengubah format teks, dan mengorganisirnya agar dapat diolah dengan lebih efisien dan menghasilkan hasil yang lebih baik. Tahapan *text preprocessing* terdiri dari tahapan *case folding*, *cleaning*, *tokenizing*, *filtering*, dan *stemming* [7].

## 5. Klasifikasi

Klasifikasi adalah proses mengidentifikasi pola dan memisahkan data ke dalam kelas-kelas yang telah ditentukan sebelumnya, dengan tujuan menyatakan objek tersebut termasuk dalam kategori tertentu yang sudah ditentukan sebelumnya. Metode klasifikasi terdiri dari dua proses, yaitu proses pelatihan dan proses pengujian. Pada proses pelatihan, algoritma klasifikasi menganalisis atau belajar dari kumpulan data latihan yang telah diberi label sebelumnya. Kemudian, pada proses pengujian dilakukan evaluasi terhadap data latihan tersebut menggunakan data uji yang merupakan kumpulan data baru yang dipilih secara acak dan independen dari data latihan [8].

## 6. Naïve Bayes

Naïve Bayes adalah metode klasifikasi probabilistik sederhana yang menggunakan teorema Bayes untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Algoritma ini menghitung sejumlah probabilitas dengan menggabungkan frekuensi dan kombinasi nilai dari dataset yang diberikan [9]. Secara sistematis, teorema Bayes dirumuskan sebagai berikut :

$$P(A|B) = (P(B|A) * P(A)) / P(B)$$

Keterangan :

B = data dengan kelas yang belum diketahui

A = hipotesis bahwa data merupakan suatu kelas spesifik

$P(A|B)$  = probabilitas hipotesis A berdasarkan kondisi data B (*posterior probability*)

$P(A)$  = probabilitas hipotesis A (*prior probability*)

$P(B|A)$  = probabilitas data B berdasarkan kondisi pada hipotesis A

$P(B)$  = probabilitas data B

## 7. Confusion Matrix

*Confusion Matrix* merupakan alat yang berguna untuk mengukur kinerja model seberapa baik pengklasifikasiannya dapat mengenali fitur dari kelas yang berbeda [10]. *Matrix* ini terdiri dari empat sel, yaitu sebagai berikut:

- True Positive* (TP): Menunjukkan total data yang benar positif dan diprediksi benar oleh model sebagai positif.
- True Negative* (TN): Menunjukkan total data yang benar negatif dan diprediksi benar oleh model sebagai negatif.
- False Positive* (FP): Menunjukkan total data yang sebenarnya negatif tetapi salah diprediksi oleh model menjadi positif.
- False Negative* (FN): Menunjukkan total data yang sebenarnya positif tetapi salah diprediksi oleh model menjadi negatif.

Tabel 1 Tabel *Confusion Matrix*

Label		Nilai Sebenarnya	
		Benar	Salah
Nilai Prediksi	Benar	TP	FP
	Salah	FN	TN

Setelah dibuat *confusion matrix*, terdapat beberapa rumus yang dapat digunakan untuk menghitung *matrix* evaluasi kinerja model klasifikasi. Berikut adalah beberapa rumus yang umum digunakan:

- Akurasi  

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN}$$
- Presisi  

$$\text{Presisi} = \frac{TP}{TP + FP}$$
- Recall*  

$$\text{Recall} = \frac{TP}{TP + FN}$$
- Spesifisitas  

$$\text{Spesifisitas} = \frac{TN}{TN + FP}$$
- F1-Score  

$$\text{F1-Score} = 2 * \frac{\text{Presisi} * \text{Recall}}{\text{Presisi} + \text{Recall}}$$

## Hasil

### 1. Pengambilan Data

Data yang dikumpulkan berupa kumpulan kalimat yang didapat melalui proses *scraping* dengan menggunakan ekstensi google chrome bernama "*Data Scraper*". *Data scraper* digunakan untuk secara otomatis mengambil data atau konten dari situs *web*. Dengan menggunakan teknik ekstraksi data dari halaman *web* HTML, *data scraper* dapat mengambil informasi yang diinginkan dan mengorganisirnya dalam format yang dapat digunakan, seperti *spreadsheet Microsoft Excel*. Proses *scraping* data menghasilkan sebanyak 3356 data mentah terhadap ulasan tiga maskapai penerbangan. Berikut adalah hasil pengambilan data, sebagai berikut:

id	nama	ulasan	maskapai	tahun
1	Yunita A	"Pelayanan citilink buruk banget. Orang sudah tua di atas 90 tahun tidak dibantu untuk turun tangga dari pesawat, disuruh menunggu di		
2	marnitamunthe89@gmail.com	"Sangat kecewa utk kemunduran jam kbrgktan maskapai ini, sgt buruk, awal pesen tiket 12.45 undur jdi 13.45 und		
3	surya r	Kalau review kursi luas dll sudah banyak yg review. Untuk pelayanan bagi yg traveling membawa bayi kurang baik. Untuk extra safety belt		
4	eRyS	"Se 20 QG-14545 QG-79545 ke 16:55 ke 17:05), juga menjadi sangat jauh (hrsnya HLP jadi ke CGK), serta ditambahin drama 'salah parkir'. Bar		
5	JLes M	Percuma melakukan check in online karena pada akhirnya dipindahkan juga saat drop bagasi. Tidak profesional, citilink, 2023		
6	Andre	"Cancelled flight, input data whatsapp salah, input data credit card salah, pengembalian dana berbelit, harus minta ke pihak ketiga. Yang s		
7	Vio briyant	Tolong di training lagi itu pramugarinya atas nama VIVI LUMBANMerasa sok cantik tapi pelayanan BURUK ! , citilink, 2023		
8	Jane doe	"Seumur hidup cmn pernah beli tiket citilink 2x dan 2â€ nya kecewa bgt. Customer service rude bgt bgt sama skali gak membantu yang a		
9	Quwina N	"Penerbangan agak delay saat itu ,well tapi saat masuk pesawat di sambut oleh pramugari yang sangat ramah, sangat membantu ,apal		
10	Billy	"Jelek banget pelayanannya, sangat sangat kecewa dengan kelakuan kru nya, mulai dari kru check in counter yg saya ucapkan twrima kasih		

Gambar 1 Hasil Scraping Data

## 2. Text Preprocessing

Setelah melakukan pengumpulan data, selanjutnya dilakukan pemrosesan teks dengan tujuan untuk menghapus kata-kata yang tidak relevan atau tidak memiliki makna yang signifikan. Adapun urutan dari tahapan *text preprocessing* yang dimulai dari *cleaning*, *case folding*, *tokenizing*, *filtering*, dan *stemming*.

- Cleaning* adalah proses untuk menghilangkan simbol-simbol, tanda baca dan angka selain teks yang terdapat dalam sebuah kalimat atau ulasan tersebut.
- Case Folding* adalah proses untuk mengubah semua data yang digunakan menjadi huruf kecil. Hal ini penting karena sistem dapat membedakan antara huruf kapital dan huruf kecil. Dalam proses ini, data yang digunakan akan diubah menjadi huruf kecil secara keseluruhan. Tujuannya adalah untuk menyamakan representasi kata-kata, sehingga jika terdapat data dengan huruf kapital, tidak akan terjadi perbedaan saat dibandingkan dengan data training yang sudah diubah menjadi huruf kecil.
- Tokenizing* adalah proses untuk membagi sebuah teks menjadi unit-unit yang lebih kecil atau sederhana yang biasanya disebut token. Setiap token mewakili satu kata atau bagian penting lainnya didalam teks.
- Filtering* adalah proses untuk menghapus kata-kata yang tidak memiliki makna atau arti sehingga data menjadi berkualitas dan proses menjadi lebih cepat. Untuk menghilangkan *stopword* dalam bahasa Indonesia menggunakan daftar *stopword* dari *Indonesian stoplist* yang bersumber dari kaggle.com. Daftar *stopword* yang digunakan berjumlah 756 kata.
- Stemming* adalah proses untuk mengubah kata-kata dalam teks menjadi bentuk kata dasar dengan menghapus imbuhan yang terdapat di depan atau di belakang kata.

Contoh hasil dari proses *text preprocessing* tersebut ditampilkan pada Tabel 2 dibawah ini:

Tabel 2 Contoh Hasil Text Preprocessing

No.	Sebelum	Sesudah
1	Tiba-tiba jadwal penerbangan diubah 2 kali dan pemberitahuannya mepet.	jadwal terbang ubah kali pemberitahuan mepet
2	Penerbangan lancar, tidak ada delay.	terbang lancar tidak delay

## 3. Pelabelan

Setelah melakukan pemrosesan teks, langkah selanjutnya adalah memberi label dengan menggunakan kamus berupa kata-kata yang mengandung sentimen. Berikut adalah hasil data yang sudah diberikan label:

id	nama	ulasan	maskapai	tahun	label
1	Yunita A	Pelayanan citilink buruk banget. Orang sudah tua di atas 90 tahun tidak dibantu unt	citilink	2023	negatif
2	marnitamunthe89@gmail.com	"Sangat kecewa utk kemunduran jam kbrgktan maskapai ini, sgt buruk, awal pesen tik	citilink	2023	negatif
3	surya r	Kalau review kursi luas dll sudah banyak yg review. Untuk pelayanan bagi yg travelin	citilink	2023	positif
4	eRyS	Seharusnya saya kemarin tgl 06/03/23 menggunakan QG-145 jurusan HLP krn dekat r	citilink	2023	negatif
5	JLes M	Percuma melakukan check in online karena pada akhirnya dipindahkan juga saat dro	citilink	2023	negatif

Gambar 2 Pelabelan Dataset

## Pembahasan

### 1. Perhitungan Algoritma Naïve Bayes

Dalam tahap ini, akan dilakukan perhitungan menggunakan metode Naïve Bayes terhadap 4 sampel dokumen, sebagai berikut:

D1 = "lion air emang maskapai buruk dunia tiket murah layan"

D2 = "masuk kabin take off landing struktur berantak gitu tumpang berisik tunggu bagasi buruk"

D3 = "kali terbang lion air kesan bagus jadwal time kendala delay lancar jalan"

D4 = "suka terbang lion air pagi time sesuai schedule"

Ke-empat dokumen tersebut sudah melalui proses *text preprocessing*, selanjutnya akan dilakukan pemisahan kata dan menentukan nilai bobot dari setiap kata-katanya menggunakan *Term Frequency* agar dapat diproses ke tahap klasifikasi.

Tabel 3 Pembobotan Kata

<b>Term</b>	<b>D1</b>	<b>D2</b>	<b>D3</b>	<b>D4</b>
Air	1	0	1	1
Bagasi	0	1	0	0
Bagus	0	0	1	0
Berantak	0	1	0	0
Berisik	0	1	0	0
Buruk	1	1	0	0
Delay	0	0	1	0
Dunia	1	0	0	0
Emang	1	0	0	0
Gitu	0	1	0	0
Jadwal	0	0	1	0
Jalan	0	0	1	0
Kabin	0	1	0	0
Kali	0	0	1	0
Kendala	0	0	1	0
Kesan	0	0	1	0
Lancar	0	0	1	0
Landing	0	1	0	0
Layan	1	0	0	0
Lion	1	0	1	1
Maskapai	1	0	0	0
Masuk	0	1	0	0
Murah	1	0	0	0
Off	0	1	0	0
Pagi	0	0	0	1
Schedule	0	0	0	1
Sesuai	0	0	0	1
Struktur	0	1	0	0
Suka	0	0	0	1
Take	0	1	0	0
Terbang	0	0	1	1
Tiket	1	0	0	0
Time	0	0	1	1
Tumpang	0	1	0	0
Tunggu	0	1	0	0
<b>Total Term</b>	<b>9</b>	<b>13</b>	<b>12</b>	<b>8</b>

Setelah melakukan perhitungan bobot untuk setiap term, langkah selanjutnya adalah menghitung nilai probabilitas untuk setiap term.

a. Menghitung Nilai Probabilitas *Prior*

Pada tahap ini akan dilakukan perhitungan terhadap atribut kelas, terdapat dua jenis kelas yaitu positif dan negatif. Kedua kelas tersebut akan dicari nilai probabilitasnya dengan menggunakan 4 dokumen *training*.

$$P(c) = \frac{N_c}{N}$$

$N_c$  = merupakan jumlah dokumen pada data latih yang masuk kategori  $c$

$N$  = merupakan dokumen pada data latih

$$P(\text{positif}) = \frac{2}{4} = 0,5$$

$$P(\text{negatif}) = \frac{2}{4} = 0,5$$

b. Menghitung Conditional Probability

Pada tahap ini dilakukan perhitungan *conditional probability* yaitu mencari nilai probabilitas dari setiap kata atau term. Perhitungan probabilitas ini menggunakan model *Multinomial*.

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

Untuk menghindari adanya nilai probabilitas nol, biasanya dilakukan penambahan +1 pada pembilang dan  $|V|$  pada penyebut.

$\text{Count}(w,c)$  = merupakan total munculnya kata  $w$  di kategori  $c$ .

$\text{Count}(c)$  = merupakan total setiap kata di kategori  $c$ .

$|V|$  = merupakan jumlah term.

Berikut contoh perhitungan menggunakan model *multinomial*:

Dokumen 1 = "lion air emang maskapai buruk dunia tiket murah layan"

1. Probabilitas Kata "lion"

$$P(\text{lion}|\text{negatif}) = \frac{(1+0)+1}{(9+13)+35} = 0,03508$$

$$P(\text{lion}|\text{positif}) = \frac{(1+1)+1}{(12+8)+35} = 0,05454$$

2. Probabilitas Kata "air"

$$P(\text{air}|\text{negatif}) = \frac{(1+0)+1}{(9+13)+35} = 0,03508$$

$$P(\text{air}|\text{positif}) =$$

$$\frac{(1+1)+1}{(12+8)+35} = 0,05454$$

3. Probabilitas Kata "emang"

$$P(\text{emang}|\text{negatif}) = \frac{(1+0)+1}{(9+13)+35} = 0,03508$$

$$P(\text{emang}|\text{positif}) = \frac{(0+0)+1}{(12+8)+35} = 0,01818$$

4. Probabilitas Kata "maskapai"

$$P(\text{maskapai}|\text{negatif}) = \frac{(1+0)+1}{(9+13)+35} = 0,03508$$

$$P(\text{maskapai}|\text{positif}) = \frac{(0+0)+1}{(12+8)+35} = 0,01818$$

5. Probabilitas Kata "buruk"

$$P(\text{buruk}|\text{negatif}) = \frac{(1+1)+1}{(9+13)+35} = 0,05263$$

$$P(\text{buruk}|\text{positif}) = \frac{(0+0)+1}{(12+8)+35} = 0,01818$$

6. Probabilitas Kata "dunia"

$$P(\text{dunia}|\text{negatif}) = \frac{(1+0)+1}{(9+13)+35} = 0,03508$$

$$P(\text{dunia}|\text{positif}) = \frac{(0+0)+1}{(12+8)+35} = 0,01818$$

7. Probabilitas Kata "tiket"

$$P(\text{tiket}|\text{negatif}) = \frac{(1+0)+1}{(9+13)+35} = 0,03508$$

$$P(\text{tiket}|\text{positif}) = \frac{(0+0)+1}{(12+8)+35} = 0,01818$$

8. Probabilitas Kata "murah"

$$P(\text{murah}|\text{negatif}) =$$



$$\frac{(1+0)+1}{(9+13)+35} = 0,03508$$

$$P(\text{murah}|\text{positif}) = \frac{(0+0)+1}{(12+8)+35} = 0,01818$$

9. Probabilitas Kata “layan”

$$P(\text{layan}|\text{negatif}) = \frac{(1+0)+1}{(9+13)+35} = 0,03508$$

$$P(\text{layan}|\text{positif}) = \frac{(0+0)+1}{(12+8)+35} = 0,01818$$

Berikut adalah hasil perhitungan *conditional probability* semua dokumen:

Tabel 4 Hasil *Conditional Probability*

<b>Term</b>	<b>P(w negatif)</b>	<b>P(w positif)</b>
Air	0,035088	0,054545
Bagasi	0,035088	0,018182
Bagus	0,017544	0,036364
Berantak	0,035088	0,018182
Berisik	0,035088	0,018182
Buruk	0,052632	0,018182
Delay	0,017544	0,036364
Dunia	0,035088	0,018182
Emang	0,035088	0,018182
Gitu	0,035088	0,018182
Jadwal	0,017544	0,036364
Jalan	0,017544	0,036364
Kabin	0,035088	0,018182
Kali	0,017544	0,036364
Kendala	0,017544	0,036364
Kesan	0,017544	0,036364
Lancar	0,017544	0,036364
Landing	0,035088	0,018182
Layan	0,035088	0,018182
Lion	0,035088	0,054545
Maskapai	0,035088	0,018182
Masuk	0,035088	0,018182
Murah	0,035088	0,018182
Off	0,035088	0,018182
Pagi	0,017544	0,036364
Schedule	0,017544	0,036364
Sesuai	0,017544	0,036364
Struktur	0,035088	0,018182
Suka	0,017544	0,036364
Take	0,035088	0,018182
Terbang	0,017544	0,054545
Tiket	0,035088	0,018182
Time	0,017544	0,054545

Tumpang	0,035088	0,018182
Tunggu	0,035088	0,018182

Setelah menghitung *conditional probability*, selanjutnya dilakukan perhitungan untuk mencari nilai *Posterior* pada masing-masing kategori. Hasil nilai dari *Posterior* yang terbesar akan menjadi kelas baru dari data tersebut.

$$P(c|d) = P(c) * \prod_{w \in d} P(w|c)$$

$P(c|d)$  disebut peluang (*posterior*) kelas  $c$  pada dokumen  $d$ .

$P(c)$  merujuk pada probabilitas awal (*prior*) dari munculnya kategori  $c$ .

$\prod_{w \in d} P(w|c)$  adalah hasil perkalian dari *conditional probability* masing-masing kata atau *term* yang terdapat pada dokumen  $d$ . Berikut contoh perhitungan klasifikasi data *testing* pada empat dokumen sebelumnya.

Dokumen 1 = "lion air emang maskapai buruk dunia tiket murah layan"

Negatif:

$$P(\text{negatif}|d1) = P(\text{negatif}) * P(\text{lion}|\text{negatif}) * P(\text{air}|\text{negatif}) * P(\text{emang}|\text{negatif}) * P(\text{maskapai}|\text{negatif}) * P(\text{buruk}|\text{negatif}) * P(\text{dunia}|\text{negatif}) * P(\text{tiket}|\text{negatif}) * P(\text{murah}|\text{negatif}) * P(\text{layan}|\text{negatif})$$

$$P(\text{negatif}|d1) = 0,5 * 0,035088 * 0,035088 * 0,035088 * 0,035088 * 0,052632 * 0,035088 * 0,035088 * 0,035088 * 0,035088$$

$$P(\text{negatif}|d1) = 6,04585216281564E-14$$

Positif:

$$P(\text{positif}|d1) = P(\text{positif}) * P(\text{lion}|\text{positif}) * P(\text{air}|\text{positif}) * P(\text{emang}|\text{positif}) * P(\text{maskapai}|\text{positif}) * P(\text{buruk}|\text{positif}) * P(\text{dunia}|\text{positif}) * P(\text{tiket}|\text{positif}) * P(\text{murah}|\text{positif}) * P(\text{layan}|\text{positif})$$

$$P(\text{positif}|d1) = 0,5 * 0,054545 * 0,054545 * 0,018182 * 0,018182 * 0,018182 * 0,018182 * 0,018182 * 0,018182 * 0,018182$$

$$P(\text{positif}|d1) = 9,77120912730205E-16$$

Berikut ini adalah hasil perhitungan klasifikasi data terhadap empat dokumen sebelumnya:

Tabel 5 Hasil Klasifikasi

Dok1			
Teks	Peluang Kelas	Skor Peluang	Kelas Baru
lion air emang maskapai buruk dunia tiket murah layan	P(negatif)	<b>6,0459E-14</b>	Negatif
	P(positif)	9,7712E-16	
Dok2			
masuk kabin take off landing struktur berantak gitu tumpang berisik tunggu bagasi buruk	P(negatif)	<b>9,1639E-20</b>	Negatif
	P(positif)	1,1865E-23	
Dok3			
kali terbang lion air kesan bagus jadwal time kendala delay lancar jalan	P(negatif)	1,7003E-21	Positif

	P(positif)	<b>1,3531E-17</b>	
Dok4			
suka terbang lion air pagi time sesuai schedule	P(negatif)	1,7949E-14	Positif
	P(positif)	<b>7,7388E-12</b>	

Berdasarkan perhitungan diatas yang memiliki nilai peluang *posterior* terbesar maka akan masuk pada kategori yang memiliki nilai terbesar, contohnya pada dokumen 1 dimana nilai peluang negatif lebih tinggi dari pada nilai peluang positif maka dokumen 1 dimasukkan ke kategori negatif.

## 2. Evaluasi

Pengujian dilakukan dengan metode *confusion matrix* untuk mengukur kinerja dan akurasi model klasifikasi. Pengujian dilakukan dengan tiga skenario yang terdiri dari 90:10, 80:20, dan 70:30. Berikut ini adalah pembagian *data training* dan *data testing* pada Tabel 6 dibawah ini:

Tabel 6 Pembagian Data

<b>Training</b>	<b>Testing</b>	<b>Training</b>	<b>Testing</b>
90%	10%	3020	336
80%	20%	2888	468
70%	30%	2349	1007
Total		3356	

Berikut ini adalah hasil dari *confusion matrix* setiap skenario, sebagai berikut:

Tabel 7 *Confusion Matrix* 70:30

Label		Prediksi		Total
		Negatif	Positif	
Aktual	Negatif	1	253	254
	Positif	0	753	753

Tabel 8 *Confusion Matrix* 80:20

Label		Prediksi		Total
		Negatif	Positif	
Aktual	Negatif	1	173	174
	Positif	0	498	294

Tabel 9 *Confusion Matrix* 90:10

Label		Prediksi		Total
		Negatif	Positif	
Aktual	Negatif	1	77	78
	Positif	0	258	258

Berdasarkan pada *confusion matrix* tersebut, didapatkan hasil akurasi, presisi, *recall*, dan F1-Score. Berikut hasil perhitungan evaluasi dari setiap skenario, dapat dilihat pada Tabel 9 berikut:

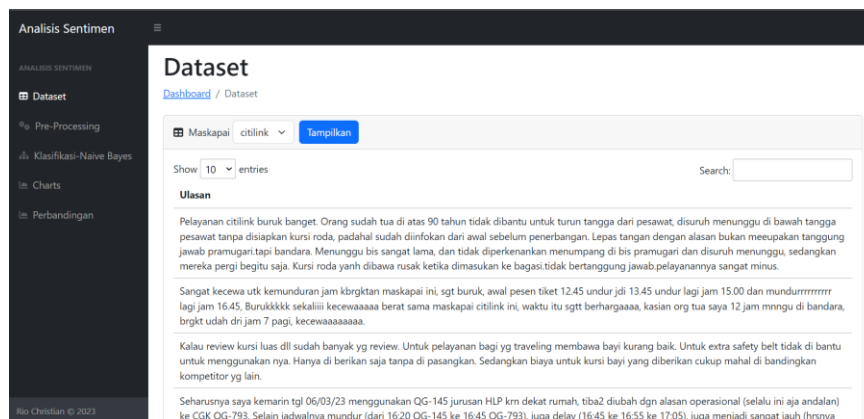
Tabel 10 Hasil Perhitungan Evaluasi

Skenario	Akurasi	Presisi	Recall	F1-Score
70:30	74,8%	74,8%	100%	85,7%
80:20	74,3%	74,2%	100%	85,2%
90:10	77%	77%	100%	87,1%

Berdasarkan hasil perhitungan setiap skenario, skenario terbaik untuk model klasifikasi naïve bayes yaitu pada skenario pembagian data 90:10 (90% sebagai *data training* dan 10% sebagai *data testing*) dengan nilai akurasi sebesar 77%, presisi sebesar 77%, *recall* sebesar 100% dan F1-Score sebesar 87,1%. Skenario ini dinilai baik karena mempunyai persentase terbesar dari skenario lainnya.

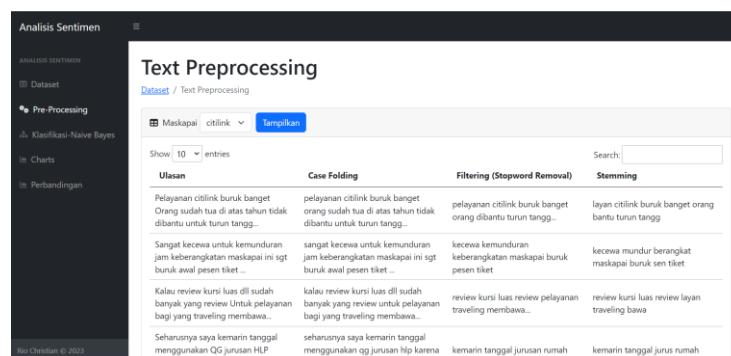
### 3. Tampilan Program

Berikut ini merupakan tampilan program yang telah dibuat:



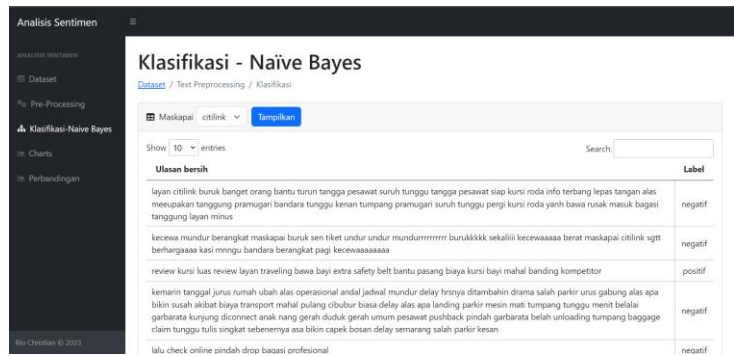
Gambar 3 Tampilan Dataset

Pada halaman ini berisi dataset maskapai penerbangan yang diperoleh menggunakan ekstensi Google Chrome bernama *Data Scraper* terhadap tiga maskapai penerbangan, yaitu Garuda Indonesia, Citilink, dan Lion Air.



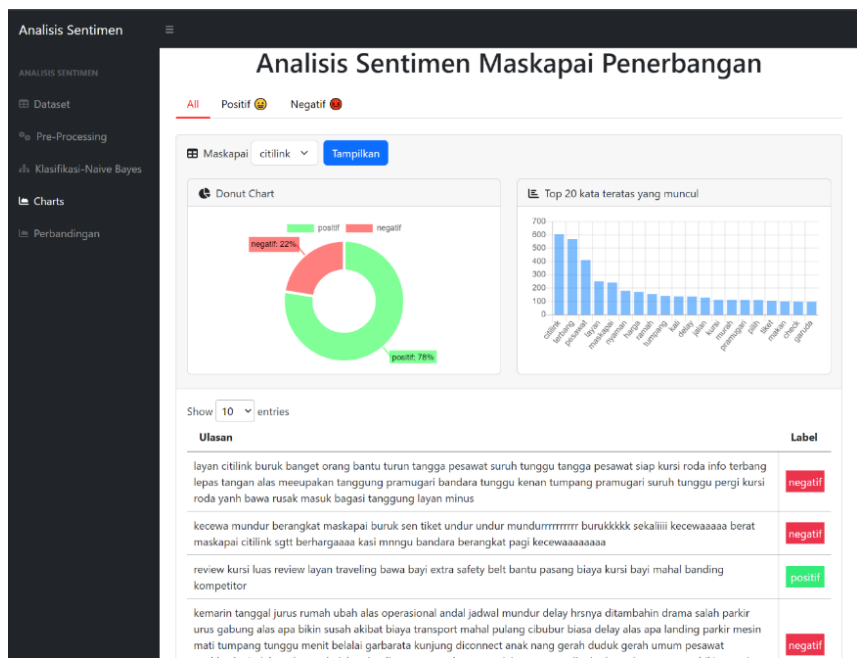
Gambar 4 Tampilan *Text Preprocessing*

Dalam halaman ini berfungsi untuk melakukan proses *text preprocessing* terhadap dataset yang telah diperoleh sebelumnya, dengan tujuan untuk mempersiapkan dan membersihkan data sebelum diolah lebih lanjut.



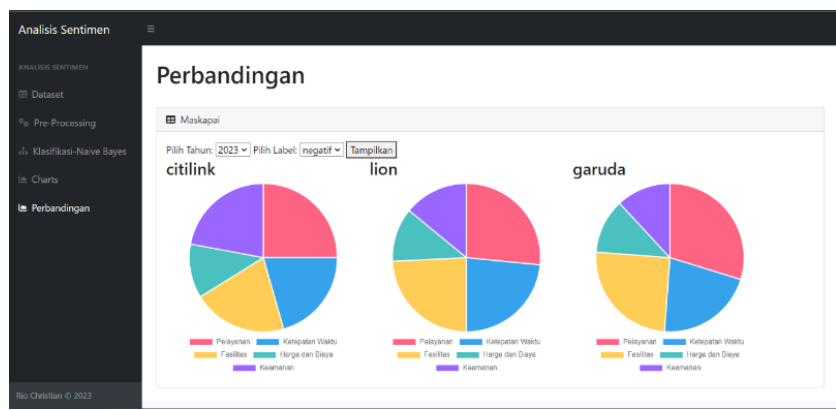
Gambar 5 Tampilan Klasifikasi Naïve Bayes

Pada halaman ini, hasil dari proses *text preprocessing* sebelumnya akan diklasifikasikan menggunakan metode Naïve Bayes dengan dua kategori, yaitu positif dan negatif.



Gambar 6 Tampilan Grafik

Dalam halaman ini, setelah dilakukan pengklasifikasian menggunakan metode Naïve Bayes, hasilnya akan divisualisasikan dalam bentuk grafik donut dan grafik batang.



Gambar 7 Tampilan Perbandingan

Dalam halaman ini, dapat melihat perbandingan antara maskapai penerbangan berdasarkan tahun dan label positif atau negatif, yang diwujudkan dalam grafik pie dengan 5 kategori utama, yaitu pelayanan, ketepatan waktu, fasilitas, harga dan biaya, serta keamanan.

### Kesimpulan

Berdasarkan pada hasil penelitian yang telah dilakukan, maka dapat disimpulkan sebagai berikut:

1. Penggunaan metode Naïve Bayes berhasil diterapkan dalam pembuatan aplikasi analisis sentimen terhadap ulasan maskapai penerbangan di TripAdvisor. Metode ini mampu mengklasifikasikan ulasan menjadi kategori sentimen positif dan negatif dengan cepat.
2. Aplikasi analisis sentimen yang dibuat dapat memberikan informasi yang berguna kepada masyarakat dalam mencari maskapai penerbangan. Dengan menggunakan analisis sentimen, dapat memperoleh gambaran umum tentang pengalaman pengguna lain, sehingga memudahkan dalam mengambil keputusan yang tepat.
3. Hasil evaluasi dengan membagi *data training* dan *data testing* menjadi 90:10 memperoleh hasil akurasi sebesar 77%, presisi sebesar 77%, *Recall* sebesar 100% dan *F1-Score* sebesar 87,1%. Hal ini menunjukkan bahwa model analisis sentimen dengan metode Naïve Bayes pada penelitian ini mampu memberikan prediksi yang cukup akurat terhadap sentimen ulasan maskapai penerbangan di TripAdvisor.

### Ucapan Terima Kasih

Terima Kasih kepada ibu Indah Fenriana selaku dosen pembimbing yang selalu membantu dan memberikan masukan dalam proses penyusunan penelitian ini dan Fakultas Sains dan Teknologi Universitas Buddhi Dharma.

### Referensi :

- [1] E. Mayasari, "Pengaruh Kepuasan Kerja terhadap Keinginan Berpindah Bekerja pada Perusahaan Startup," vol. 11, no. 02, pp. 76–91, 2023.
- [2] F. Z. Hidayat and P. L. Samputra, "Evaluasi Tingkat Kepuasan Masyarakat Pengguna Maskapai Penerbangan Kelas Ekonomi Sebelum dan Setelah Kebijakan Kenaikan Tarif Tiket," *Akuntabel*, vol. 17, no. 1, pp. 122–132, 2020.
- [3] A. Setiawan and F. H. Sukmana, "Mengurai Pengalaman Positif Tamu Saat Menginap di Sheraton Senggigi Beach Resort: Bukti Dari Ulasan TripAdvisor," *J. Kepariwisata Indonesia. J. Penelit. dan Pengemb. Kepariwisata Indonesia.*, vol. 17, no. 1, pp. 64–84, 2023, doi: 10.47608/jki.v17i12023.64-84.
- [4] M. Arhami and M. Nasir, *Data Mining - Algoritma dan Implementasi*. Andi Offset, 2020. [Online]. Available: <https://books.google.co.id/books?id=AtcCEAAAQBAJ>
- [5] Z. Munawar *et al.*, *Big Data Analytics: Konsep, Implementasi, dan Aplikasi Terkini*. Kaizen Media Publishing, 2023. [Online]. Available: <https://books.google.co.id/books?id=fWmxEAAAQBAJ>
- [6] G. Urva *et al.*, *PENERAPAN DATA MINING DI BERBAGAI BIDANG: Konsep, Metode, dan Studi Kasus*. PT. Sonpedia Publishing Indonesia, 2023. [Online]. Available: <https://books.google.co.id/books?id=uq6-EAAAQBAJ>
- [7] F. A. Nugraha, N. H. Harani, and R. Habibi, *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*. Kreatif, 2020. [Online]. Available: <https://books.google.co.id/books?id=f738DwAAQBAJ>
- [8] A. Wulandari and A. Sifaunajah, *Perbandingan Efektivitas Klasifikasi Algoritma C.4.5 dan Algoritma Naive Bayes dengan Menggunakan Pihak ke 3 (WEKA)*. Lembaga

- Penelitian dan Pengabdian kepada Masyarakat Universitas KH. A. Wahab Hasbullah, 2019. [Online]. Available: <https://books.google.co.id/books?id=s6t-EAAAQBAJ>
- [9] D. A. Pratiwi, R. M. Awangga, and M. Y. H. Setyawan, *SELEKSI CALON KELULUSAN TEPAT WAKTU MAHASISWA TEKNIK INFORMATIKA MENGGUNAKAN METODE NAÏVE BAYES*. Kreatif, 2020. [Online]. Available: <https://books.google.co.id/books?id=d6wGEAAAQBAJ>
- [10] A. Saifudin, *LEVEL DATA DAN ALGORITMA UNTUK PENANGANAN KETIDAKSEIMBANGAN KELAS*. Pascal Books, 2022. [Online]. Available: <https://books.google.co.id/books?id=MG6dEAAAQBAJ>