

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

SKRIPSI



Disusun Oleh:

NAMA : Rio Christian

NIM : 20191000060

TEKNIK INFORMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS BUDDHI DHARMA

TANGERANG

2023

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

SKRIPSI

**Diajukan sebagai salah satu syarat untuk kelengkapan gelar kesarjanaan pada
Program Studi Teknik Informatika
Jenjang Pendidikan Strata 1**



Disusun Oleh:

NAMA : Rio Christian

NIM : 20191000060

TEKNIK INFORMATIKA

**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS BUDDHI DHARMA
TANGERANG**

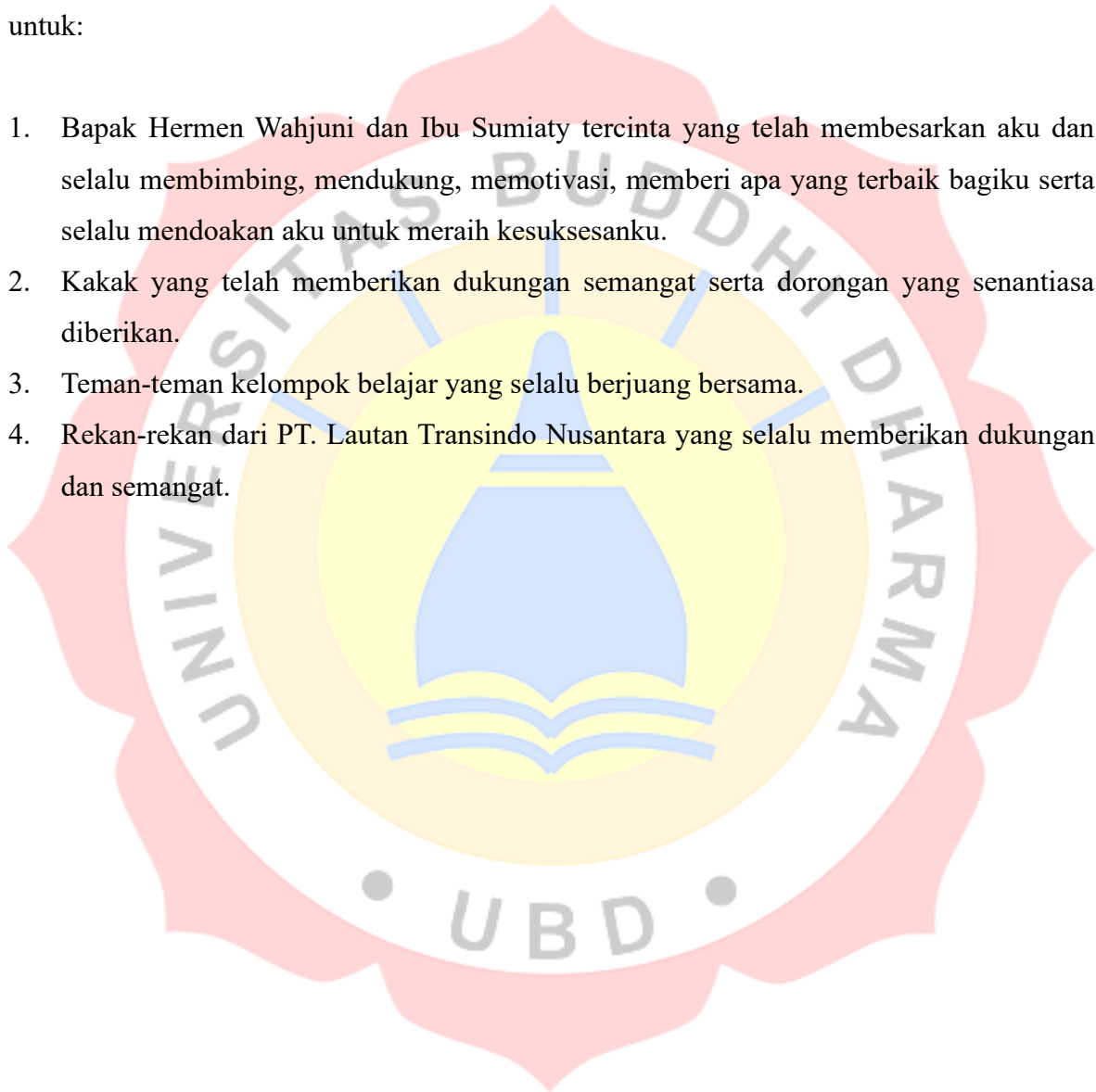
2023

LEMBAR PERSEMBAHAN

“Whatever the mind of man can conceive and believe, it can achieve.”
(Napoleon Hill)

Dengan mengucap puji syukur kepada Tuhan Yang Maha Esa, Skripsi ini kupersembahkan untuk:

1. Bapak Hermen Wahjuni dan Ibu Sumiaty tercinta yang telah membesarkan aku dan selalu membimbing, mendukung, memotivasi, memberi apa yang terbaik bagiku serta selalu mendoakan aku untuk meraih kesuksesanku.
2. Kakak yang telah memberikan dukungan semangat serta dorongan yang senantiasa diberikan.
3. Teman-teman kelompok belajar yang selalu berjuang bersama.
4. Rekan-rekan dari PT. Lautan Transindo Nusantara yang selalu memberikan dukungan dan semangat.



UNIVERSITAS BUDDHI DHARMA
LEMBAR PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan di bawah ini,

NIM : 20191000060
Nama : Rio Christian
Jenjang Studi : Strata 1
Program Studi : Teknik Informatika
Peminatan : *Database*

Dengan ini saya menyatakan bahwa:

1. Skripsi ini adalah asli dan belum pernah diajukan untuk mendapat gelar akademik Sarjana atau kelengkapan studi, baik di Universitas Buddhi Dharma maupun di Perguruan Tinggi lainnya.
2. Skripsi ini saya buat sendiri tanpa bantuan dari pihak lain, kecuali arahan dosen pembimbing.
3. Dalam Skripsi ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dan dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan daftar pustaka.
4. Dalam Skripsi ini tidak terdapat pemalsuan (kebohongan), seperti buku, artikel, jurnal, data sekunder, pengolahan data, dan pemalsuan tanda tangan dosen atau Ketua Program Studi Universitas Buddhi Dharma yang dibuktikan dengan keasliannya.
5. Lembar pernyataan ini saya buat dengan sesungguhnya, tanpa paksaan dan apabila dikemudian hari atau pada waktu lainnya terdapat penyimpangan dan ketidak benaran dalam pernyataan ini, saya bersedia menerima sanksi akademik berupa pencabutan gelar akademik yang telah saya peroleh karena Skripsi ini serta sanksi lainnya sesuai dengan peraturan dan norma yang berlaku.

Tangerang, 08 Agustus 2023

Penulis,



Rio Christian

UNIVERSITAS BUDDHI DHARMA
LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini,

NIM : 20191000060
Nama : Rio Christian
Jenjang Studi : Strata 1
Program Studi : Teknik Informatika
Peminatan : *Database*

Dengan ini menyetujui untuk memberikan izin kepada pihak Universitas Buddhi Dharma, Hak Bebas Royalti Non – Eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah kami yang berjudul: “PERANCANGAN APLIKASI ANALISIS SENTIMEN PADA ULASAN MASKAPAI PENERBANGAN DI TRIPADVISOR MENGGUNAKAN METODE NAÏVE BAYES”, beserta alat yang diperlukan (apabila ada).

Dengan Hak Bebas Royalti Non – Eksklusif ini pihak Universitas Buddhi Dharma berhak menyimpan, mengalih-media atau format-kan, mengelolanya dalam pangkalan data (*database*), mendistribusikannya, dan menampilkan atau mempublikasikannya di *internet* atau media lain untuk kepentingan akademis tanpa perlu meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis atau pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Universitas Buddhi Dharma, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 08 Agustus 2023

Penulis,



Rio Christian

20191000060

UNIVERSITAS BUDDHI DHARMA
LEMBAR PENGESAHAN PEMBIMBING
PERANCANGAN APLIKASI ANALISIS SENTIMEN PADA ULASAN
MASKAPAI PENERBANGAN DI TRIPADVISOR MENGGUNAKAN
METODE NAÏVE BAYES

Dibuat Oleh:

NIM : 20191000060

Nama : Rio Christian

Telah disetujui untuk dipertahankan di hadapan Tim Penguji Ujian
Komprehensif

Program Studi Teknik Informatika

Peminatan *Database Development*

Tahun Akademik 2022/2023

Disahkan oleh,

Tangerang, 18 Juli 2023

Pembimbing,



Indah Fenriana, S.Kom., M.Kom.

NIDN. 0406028801

UNIVERSITAS BUDDHI DHARMA
LEMBAR PENGESAHAN SKRIPSI
PERANCANGAN APLIKASI ANALISIS SENTIMEN PADA ULASAN
MASKAPAI PENERBANGAN DI TRIPADVISOR MENGGUNAKAN
METODE NAÏVE BAYES

Dibuat Oleh:

NIM : 20191000060

Nama : Rio Christian

Telah disetujui untuk dipertahankan di hadapan Tim Penguji Ujian
Komprehensif

Program Studi Teknik Informatika

Peminatan *Database Development*

Tahun Akademik 2022/2023

Disahkan oleh,

Tangerang, 08 Agustus 2023

Dekan,



Dr. Eng, Ir. Amin Suyitno, M.Eng.

NIDK : 8826333420

Ketua Program Studi,



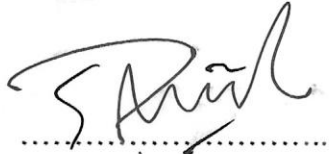


Hartana Wijaya, S.Kom., M.Kom.

NIDN : 0412058102

LEMBAR PENGESAHAN TIM PENGUJI

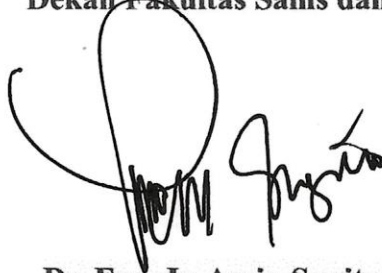
Nama : Rio Christian
NIM : 20191000060
Fakultas : Sains dan Teknologi
Judul Skripsi : Perancangan Aplikasi Analisis Sentimen Pada Ulasan Maskapai Penerbangan Di Tripadvisor Menggunakan Metode Naïve Bayes

Dinyatakan LULUS setelah mempertahankan di depan Tim Penguji pada hari Selasa, 08 Agustus 2023.

	Nama Penguji :	Tanda Tangan :
Ketua Sidang	: Rino, M.Kom NIDN. 0420058502	
Penguji I	: Benny Daniawan, M.Kom NIDN. 0424049006	
Penguji II	: Indah Fenriana, S.Kom., M.Kom NIDN. 0406028801	

Mengetahui,

Dekan Fakultas Sains dan Teknologi



Dr. Eng, Ir. Amin Suyitno, M.Eng.

NIDK : 8826333420

KATA PENGANTAR

Dengan mengucapkan Puji Syukur kepada Tuhan Yang Maha Esa, yang telah memberikan Rahmat dan karunia-Nya kepada penulis sehingga dapat menyusun dan menyelesaikan Skripsi ini dengan judul **“Perancangan Aplikasi Analisis Sentimen Pada Ulasan Maskapai Penerbangan Di Tripadvisor Menggunakan Metode Naïve Bayes”**. Tujuan utama dari pembuatan Skripsi ini adalah sebagai salah satu syarat kelengkapan dalam menyelesaikan program pendidikan Strata 1 Program Studi Teknik Informatika di Universitas Buddhi Dharma. Dalam penyusunan Skripsi ini penulis banyak menerima bantuan dan dorongan baik moril maupun materiil dari berbagai pihak, maka pada kesempatan ini penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Ibu Dr. Limajatini, S.E., M.M., B.K.P., sebagai Rektor Universitas Buddhi Dharma.
2. Bapak Dr. Eng, Ir. Amin Suyitno, M.Eng., sebagai Dekan Fakultas Sains dan Teknologi.
3. Bapak Rudy Arijanto, S.Kom., M.Kom., sebagai Wakil Dekan Fakultas Sains dan Teknologi.
4. Bapak Hartana Wijaya, S.Kom., M.Kom., sebagai Ketua Program Studi Teknik Informatika.
5. Ibu Indah Fenriana, S.Kom., M.Kom., sebagai pembimbing yang telah membantu dan memberikan dukungan serta harapan untuk menyelesaikan penulisan Skripsi ini.
6. Orang tua dan keluarga yang selalu memberikan dukungan baik moril dan materiil.
7. Teman-teman dan sahabat yang selalu membantu dan memberikan semangat.

Serta semua pihak yang terlalu banyak untuk disebutkan satu-persatu sehingga terwujudnya penulisan ini. Penulis menyadari bahwa penulisan Skripsi ini masih belum sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan di masa yang akan datang.

Akhir kata semoga Skripsi ini dapat berguna bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Tangerang, 08 Agustus 2023

Penulis

ABSTRAK

Pesatnya kemajuan teknologi informasi, terutama di bidang *internet* telah mengubah cara orang mencari informasi, termasuk dalam konteks maskapai penerbangan. Di Indonesia, industri penerbangan memainkan peran penting dalam transportasi dan pariwisata. Namun selama masa pandemi COVID-19, industri ini mengalami penurunan yang signifikan. Setelah pemulihan dari pandemi minat masyarakat untuk melakukan perjalanan meningkat, ini mencerminkan keinginan masyarakat untuk menjelajahi destinasi wisata baru, dan lainnya. 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, Tripadvisor, Analisis Sentimen*

ABSTRACT

The rapid advancement of information technology, especially the internet, has changed the way people search for information, including in the context of airlines. In Indonesia, the airline industry plays an important role in transportation and tourism. However, during the COVID-19 pandemic, the industry has experienced a significant decline. After recovering from the pandemic, people's interest in traveling has increased, reflecting people's desire to explore new tourist destinations, and more. Before traveling, people often look for information and reviews about airlines, one of which is through the Tripadvisor platform. Due to the large number of reviews on Tripadvisor, it is difficult for users to quickly understand the sentiment. To overcome this problem, sentiment analysis is needed to understand sentiment systematically. Therefore, a sentiment analysis application was designed by implementing the Naïve Bayes method in classifying sentiments into positive and negative categories. This application displays the results of the analysis in the form of graphs that provide a visual representation of the positive and negative sentiments contained in the reviews. By testing using the confusion matrix method with three scenarios namely 70:30, 80:20, and 90:10. Based on the comparison between the three scenarios, the best scenario for the naïve bayes classification model is 90:10 which shows an accuracy value of 77%, precision of 77%, recall of 100%, and F1-Score of 87.1%. With this application, it is expected to help people in finding airlines and making more informed travel decisions based on reviews that have been analyzed effectively.

Key word: Airlines, Tripadvisor, Sentiment Analysis

DAFTAR ISI

LEMBAR JUDUL DALAM	
LEMBAR PERSEMBAHAN	
LEMBAR PERNYATAAN KEASLIAN SKRIPSI	
LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH	
LEMBAR PENGESAHAN PEMBIMBING	
LEMBAR PENGESAHAN SKRIPSI	
LEMBAR PENGESAHAN TIM PENGUJI	
KATA PENGANTAR	i
ABSTRAK.....	ii
<i>ABSTRACT</i>	iii
DAFTAR ISI	iv
DAFTAR TABEL.....	vii
DAFTAR GAMBAR.....	ix
DAFTAR LAMPIRAN.....	x
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Identifikasi Masalah.....	4
1.3. Rumusan Masalah.....	5
1.4. Ruang Lingkup	5
1.5. Tujuan dan Manfaat	5
1.5.1. Tujuan.....	5
1.5.2. Manfaat.....	6
1.6. Metodologi Penelitian.....	6
1.6.1. Teknik Pengumpulan Data.....	6
1.6.2. Metode <i>Text Mining</i>	7
1.7. Sistematika Penulisan	8
BAB II LANDASAN TEORI.....	10
2.1. Teori Umum.....	10
2.1.1. Data.....	10
2.1.2. Informasi.....	11
2.1.3. Aplikasi.....	13

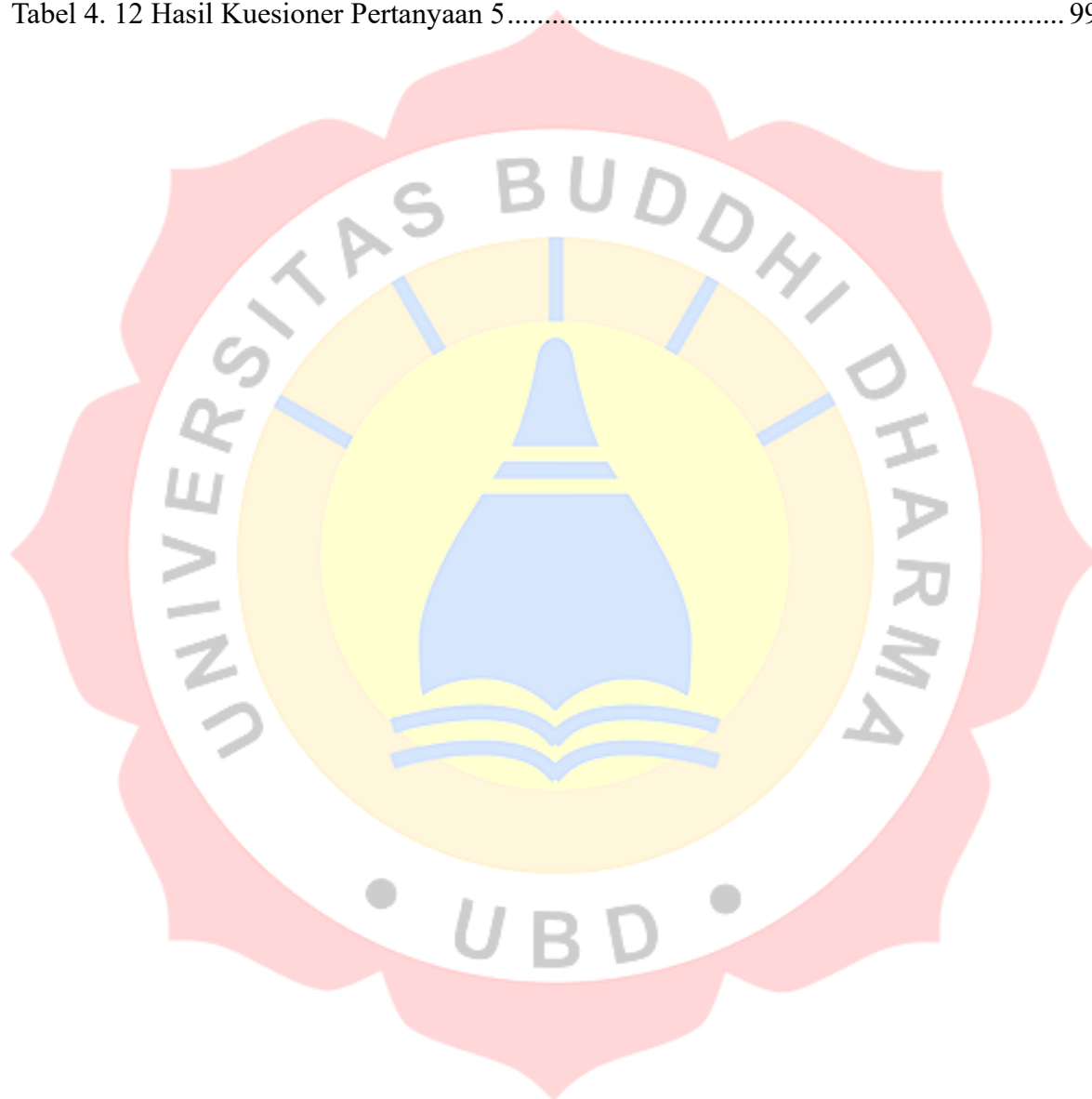
2.1.4.	<i>Website</i>	13
2.1.5.	Tripadvisor.....	14
2.1.6.	Analisis Sentimen.....	14
2.2.	Teori Khusus	14
2.2.1.	<i>Data Mining</i>	14
2.2.2.	<i>Machine Learning</i>	16
2.2.3.	Klasifikasi.....	17
2.2.4.	Naïve Bayes.....	17
2.2.5.	<i>Text Mining</i>	18
2.2.6.	<i>Text Preprocessing</i>	19
2.2.7.	<i>Confusion Matrix</i>	20
2.2.8.	<i>Flowchart</i>	22
2.3.	Teori Perancangan.....	25
2.3.1.	<i>Hypertext Markup Language (HTML)</i>	25
2.3.2.	<i>Hypertext Preprocessor (PHP)</i>	26
2.3.3.	<i>JavaScript (JS)</i>	26
2.3.4.	<i>Cascading Style Sheets (CSS)</i>	26
2.3.5.	<i>Bootstrap</i>	27
2.3.6.	XAMPP	27
2.3.7.	<i>Visual Studio Code</i>	28
2.3.8.	<i>MySQL</i>	28
2.3.9.	<i>ERD (Entity Relationship Diagram)</i>	29
2.3.10.	Python.....	29
2.4.	Tinjauan Studi	29
2.5.	Kerangka Pemikiran	55
BAB III ANALISA MASALAH & PERANCANGAN APLIKASI		56
3.1.	Identifikasi Kebutuhan Sistem.....	56
3.1.1.	<i>Requirement Elicitation Tahap I</i>	56
3.1.2.	<i>Requirement Elicitation Tahap II</i>	57
3.1.3.	<i>Requirement Elicitation Tahap III</i>	58
3.1.4.	<i>Requirement Elicitation Final</i>	60
3.2.	Pengambilan Data	61
3.3.	<i>Text Preprocessing</i>	62
3.4.	Pelabelan Leksikon	65

3.5.	Klasifikasi Manual Naïve Bayes.....	67
3.5.1.	Perhitungan Manual Naïve Bayes	67
3.5.2.	Proses Klasifikasi Data <i>Training</i>	71
3.5.3.	Proses Klasifikasi Data <i>Testing</i>	75
3.6.	<i>Flowchart</i>	77
3.6.1.	<i>Flowchart</i> Pengumpulan Data	77
3.6.2.	<i>Flowchart Text Preprocessing</i>	78
3.6.3.	<i>Flowchart</i> Proses Klasifikasi Naïve Bayes	79
3.7.	Perancangan <i>Database</i>	80
3.8.	Perancangan Layar.....	81
BAB IV PENGUJIAN DAN IMPLEMENTASI.....		85
4.1.	Tampilan Aplikasi	85
4.1.1.	Tampilan Antarmuka <i>Dataset</i>	85
4.1.2.	Tampilan Antarmuka <i>Text Preprocessing</i>	85
4.1.3.	Tampilan Antarmuka Klasifikasi Naïve Bayes.....	86
4.1.4.	Tampilan Antarmuka Grafik.....	87
4.1.5.	Tampilan Antarmuka Perbandingan	88
4.2.	Spesifikasi <i>Hardware</i> dan <i>Software</i>	88
4.2.1.	Hardware	88
4.2.2.	Software.....	89
4.3.	Pengujian <i>Black Box Functional Testing</i>	89
4.4.	Evaluasi Model Klasifikasi.....	91
4.5.	Hasil Pengolahan Data Kuesioner	94
BAB V SIMPULAN DAN SARAN		101
5.1.	Simpulan	101
5.2.	Saran	102
DAFTAR PUSTAKA		103
DAFTAR RIWAYAT HIDUP.....		Error! Bookmark not defined.
LAMPIRAN		L1

DAFTAR TABEL

Tabel 2. 1 <i>Confusion Matrix</i>	21
Tabel 2. 2 Simbol-simbol <i>Flowchart</i>	22
Tabel 2. 3 Penelitian Winda Yulita, Eko Dwi Nugroho, Muhammad Habib Algifari	29
Tabel 2. 4 Penelitian Fransiska Vina Sari, Arief Wibowo	32
Tabel 2. 5 Penelitian Retno Sari, Ratih Yulia Hayuningtyas	34
Tabel 2. 6 Penelitian Rizki Wahyudi, Gilang Kusumawardhana.....	36
Tabel 2. 7 Penelitian Yerik Afrianto Singgalen	39
Tabel 2. 8 Penelitian Titan Nugraha, Purwantoro, Yuyun Umaidah.....	42
Tabel 2. 9 Ringkasan Model Penelitian	45
Tabel 3. 1 <i>Requirement Elicitation</i> Tahap I	56
Tabel 3. 2 <i>Requirement Elicitation</i> Tahap II.....	57
Tabel 3. 3 <i>Requirement Elicitation</i> Tahap III	59
Tabel 3. 4 <i>Requirement Elicitation Final</i>	60
Tabel 3. 5 Contoh Hasil Proses <i>Cleaning</i>	62
Tabel 3. 6 Contoh Hasil Proses <i>Case Folding</i>	63
Tabel 3. 7 Contoh Hasil Proses <i>Tokenizing</i>	63
Tabel 3. 8 Daftar Kata <i>Stopword</i>	64
Tabel 3. 9 Contoh Daftar Kata Sentimen Positif dan Negatif	66
Tabel 3. 10 Contoh Proses <i>Text Preprocessing</i>	67
Tabel 3. 11 Proses <i>Tokenizing</i>	68
Tabel 3. 12 Pembobotan Kata.....	69
Tabel 3. 13 Hasil <i>Conditional Probability</i>	73
Tabel 3. 14 Hasil Klasifikasi Data <i>Testing</i>	76
Tabel 3. 15 Rincian Tabel dataset.....	80
Tabel 3. 16 Rincian Tabel <i>text_preprocessing</i>	81
Tabel 3. 17 Rincian Tabel klasifikasi.....	81
Tabel 4. 1 Spesifikasi <i>Hardware</i>	88
Tabel 4. 2 Spesifikasi <i>Software</i>	89
Tabel 4. 3 Pengujian <i>Black Box Functional Testing</i>	89
Tabel 4. 4 <i>Confusion Matrix</i> 70:30.....	91
Tabel 4. 5 <i>Confusion Matrix</i> 80:20	91

Tabel 4. 6 <i>Confusion Matrix</i> 90:10	92
Tabel 4. 7 Hasil Perhitungan Setiap Skenario	93
Tabel 4. 8 Hasil Kuesioner Pertanyaan 1	95
Tabel 4. 9 Hasil Kuesioner Pertanyaan 2	96
Tabel 4. 10 Hasil Kuesioner Pertanyaan 3	97
Tabel 4. 11 Hasil Kuesioner Pertanyaan 4	98
Tabel 4. 12 Hasil Kuesioner Pertanyaan 5	99



DAFTAR GAMBAR

Gambar 2. 1 Diagram Kerangka Pemikiran	55
Gambar 3. 1 Hasil <i>Scraping</i> Data Ulasan.....	61
Gambar 3. 2 Hasil Pelabelan Menggunakan Leksikon.....	67
Gambar 3. 3 <i>Flowchart</i> Pengumpulan Data	77
Gambar 3. 4 <i>Flowchart Text Preprocessing</i>	78
Gambar 3. 5 <i>Flowchart</i> Proses Klasifikasi Naïve Bayes.....	79
Gambar 3. 6 <i>Entity Relationship Diagram</i>	80
Gambar 3. 7 Rancangan Antarmuka Dataset.....	82
Gambar 3. 8 Rancangan Antarmuka <i>Text Preprocessing</i>	82
Gambar 3. 9 Rancangan Antarmuka Klasifikasi	83
Gambar 3. 10 Rancangan Antarmuka Grafik	83
Gambar 3. 11 Rancangan Antarmuka Perbandingan.....	84
Gambar 4. 1 Tampilan Antarmuka <i>Dataset</i>	85
Gambar 4. 2 Tampilan Antarmuka <i>Text Preprocessing</i>	86
Gambar 4. 3 Tampilan Antarmuka Klasifikasi Naïve Bayes.....	86
Gambar 4. 4 Tampilan Antarmuka Grafik	87
Gambar 4. 5 Tampilan Antarmuka Perbandingan.....	88
Gambar 4. 6 Diagram Usia Responden	94
Gambar 4. 7 Diagram Pekerjaan Responden.....	94
Gambar 4. 8 Grafik Jawaban Pertanyaan 1	95
Gambar 4. 9 Grafik Jawaban Pertanyaan 2	96
Gambar 4. 10 Grafik Jawaban Pertanyaan 3	97
Gambar 4. 11 Grafik Jawaban Pertanyaan 4	98
Gambar 4. 12 Grafik Jawaban Pertanyaan 5	99
Gambar 4. 13 Grafik Hasil Keseluruhan Kuesioner.....	100

DAFTAR LAMPIRAN

Lampiran 1: Hasil Kuesioner <i>Requirement Elicitation</i>	L1
Lampiran 2: Hasil Kuesioner Evaluasi Aplikasi.....	L5



BAB I

PENDAHULUAN

1.1. 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 (Mayasari, 2023, p. 76). 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 memainkan peran penting dalam industri transportasi dan pariwisata di Indonesia. Dengan jumlah populasi yang sangat besar dan perekonomian yang signifikan sebagai negara kepulauan terbesar di dunia, Indonesia memiliki kebutuhan tinggi akan konektivitas udara yang efisien dan handal. 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 (Hidayat & Samputra, 2020, p. 123).

Di Indonesia, industri penerbangan mengalami kemajuan yang signifikan setiap tahunnya. Seiring dengan pertumbuhan ekonomi yang pesat dan meningkatnya mobilitas masyarakat, permintaan akan perjalanan udara telah meningkat secara substansial. Hal ini mendorong lahirnya berbagai maskapai penerbangan di Indonesia yang beroperasi baik dalam skala nasional maupun internasional.

Namun, selama masa pandemi COVID-19, industri penerbangan mengalami penurunan signifikan baik dalam jumlah penerbangan maupun jumlah penumpang. Berdasarkan data dari Badan Pusat Statistik (BPS) mencatat bahwa jumlah penumpang pesawat domestik mengalami penurunan drastis sepanjang tahun 2020. Jumlah penumpang pesawat domestik turun sebesar 57,76% menjadi 32,4 juta orang dibandingkan dengan periode yang sama pada tahun 2019. Tidak hanya itu, penurunan jumlah penumpang internasional juga sangat signifikan. BPS mencatat bahwa jumlah penumpang pesawat internasional turun sebesar 80,61% menjadi hanya sebanyak 3,7 juta orang pada tahun 2020 dibandingkan dengan tahun sebelumnya (Putra, 2021). Penurunan jumlah penumpang ini disebabkan oleh berbagai faktor yang terkait dengan pandemi COVID-19. Pembatasan perjalanan, penutupan perbatasan, dan kekhawatiran terhadap kesehatan masyarakat untuk melakukan perjalanan udara sebagai langkah pencegahan penyebaran virus telah memberikan dampak yang signifikan bagi industri penerbangan di Indonesia.

Setelah meredanya pandemi COVID-19 dan penerapan langkah-langkah pemulihan yang efektif, industri penerbangan di Indonesia mulai mengalami

pemulihan dan pertumbuhan yang signifikan. Banyak masyarakat yang kini semakin antusias untuk melakukan perjalanan, baik itu dalam negeri maupun luar negeri. Hal ini mencerminkan keinginan masyarakat untuk menjelajahi destinasi wisata baru, mengunjungi keluarga dan kerabat, serta melakukan perjalanan bisnis. 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 (Setiawan & Sukmana, 2023, p. 66).

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.

Berdasarkan penjelasan latar belakang yang telah disampaikan sebelumnya, maka diperlukan sebuah sistem untuk menganalisis sentimen ulasan maskapai yang diberi judul, “PERANCANGAN APLIKASI ANALISIS SENTIMEN PADA ULASAN MASKAPAI PENERBANGAN DI TRIPADVISOR MENGGUNAKAN METODE NAÏVE BAYES”.

1.2. Identifikasi Masalah

Didasarkan pada paparan latar belakang sebelumnya, terdapat beberapa permasalahan yang dapat diidentifikasi, yaitu sebagai berikut:

1. Karena banyaknya ulasan dalam maskapai penerbangan di Tripadvisor, masyarakat mengalami kesulitan dalam memproses dan memahami sentimen yang ada di dalam ulasan-ulasan tersebut dengan cepat dan akan memakan banyak waktu jika dilakukan secara manual dengan membaca dan memahami ulasan tersebut.
2. Tidak adanya sistem otomatis dalam menganalisa dan mengklasifikasi ulasan-ulasan tersebut berdasarkan sentimen positif dan negatif secara otomatis.
3. Bagaimana memastikan bahwa metode klasifikasi Naïve Bayes mampu mengklasifikasi sentimen positif dan negatif dengan tingkat akurat yang tinggi.

1.3. Rumusan Masalah

Berlandaskan dari permasalahan yang telah dipaparkan sebelumnya, rumusan masalah dalam penelitian ini adalah bagaimana menganalisis sentimen masyarakat terhadap ulasan maskapai penerbangan di Indonesia dengan algoritma naïve bayes dengan cepat dan akurat?

1.4. Ruang Lingkup

Batasan masalah dalam penelitian ini mencakup ruang lingkup sebagai berikut:

1. Data yang digunakan diambil dari *website* Tripadvisor.co.id dengan tiga maskapai yaitu garuda indonesia, lion air, dan citilink.
2. Ulasan yang diambil adalah ulasan dalam bahasa Indonesia.
3. Penelitian ini menggunakan pendekatan *Supervised Learning*
4. Pengambilan data dengan cara *scraping* menggunakan *extension browser* yaitu data *scraper*.
5. Jumlah *dataset* yang digunakan adalah sebanyak 3356 ulasan.
6. Mengklasifikasikan sentimen menjadi positif dan negatif.

1.5. Tujuan dan Manfaat

1.5.1. Tujuan

Dalam penelitian ini terdapat beberapa tujuan meliputi:

- a. Menerapkan algoritma naïve bayes pada *text mining*.
- b. Mempermudah dan mempercepat masyarakat dalam mencari maskapai penerbangan berdasarkan ulasan dalam bahasa Indonesia sehingga menghemat waktu.

- c. Mengevaluasi performa model menggunakan metode naïve bayes dengan membagi data menjadi data *training* dan data *testing*, serta mengukur akurasi, presisi, *recall*, dan *F1-Score* untuk mengukur keberhasilan prediksi sentimen ulasan maskapai penerbangan.

1.5.2. Manfaat

Terdapat beberapa manfaat dari penelitian ini antara lain:

- a. Dapat menganalisis dan memahami sentimen positif dan negatif dengan cepat.
- b. Memberikan kemudahan bagi masyarakat dalam mengumpulkan atau mencari informasi mengenai maskapai penerbangan secara efisien, tanpa perlu membaca satu per satu ulasan secara manual.
- c. Dapat memberikan tingkat akurasi, presisi, *recall* dan *F1-Score* yang tinggi yang mencerminkan kualitas yang unggul dalam memprediksi sentimen ulasan maskapai penerbangan.

1.6. Metodologi Penelitian

1.6.1. Teknik Pengumpulan Data

Didalam penelitian ini menggunakan data *primer* yang diperoleh dari *website* Tripadvisor.co.id. Pengumpulan data dilakukan menggunakan ekstensi *browser* yaitu “Data *Scraper*” untuk mengambil data mengenai ulasan maskapai penerbangan. Setelah mengumpulkan data, langkah selanjutnya adalah melakukan

pemrosesan teks terhadap data tersebut dan melabelinya dengan menggunakan kamus atau leksikon berupa kata-kata besar. Kamus yang digunakan didapatkan dari sumber yang terdapat di <https://github.com/masdevid/ID-OpinionWords>.

1.6.2. Metode *Text Mining*

a. *Text Pre-processing*

Dalam tahap ini, setelah data dikumpulkan akan diproses dengan membersihkan data yang tidak penting atau tidak relevan.

b. *Text Transformation*

Dalam tahap ini, sesudah teks yang di bersihkan akan diubah kedalam bentuk pecahan dalam sebuah kalimat yang biasanya disebut sebagai token.

c. Klasifikasi Model

Didalam proses ini, dilakukan pengelompokkan dengan menggunakan metode klasifikasi Naïve Bayes. Model klasifikasi akan dibuat dengan memanfaatkan skor probabilitas masing-masing kata untuk menentukan kelas sentimen (positif atau negatif) dari ulasan. Metode Naïve Bayes memperhitungkan kemunculan kata-kata dalam kelas sentimen yang berbeda untuk melakukan klasifikasi.

d. Evaluasi

Hasil ulasan yang sudah dianalisis akan diklasifikasikan menjadi sentimen positif atau negatif.

1.7. Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini menguraikan informasi mencakup dasar dan latar belakang penelitian yang dilakukan, identifikasi masalah yang menjadi fokus utama dalam penelitian yang dilakukan, rumusan masalah, ruang lingkup, tujuan serta manfaat yang diharapkan dari hasil penelitian, metodologi penelitian yang digunakan, dan sistematika penulisan.

BAB II LANDASAN TEORI

Dalam bab ini berfokus pada berbagai teori-teori mendasar yang menjadi dasar dalam penelitian yang dilakukan. Teori yang dijelaskan mencakup teori umum yang relevan dengan topik penelitian, teori khusus, teori analisis dan perancangan yang akan digunakan dalam menganalisa dan merancang suatu sistem, tinjauan studi untuk meninjau atau mengevaluasi referensi yang relevan dengan topik penelitian, dan kerangka pemikiran.

BAB III ANALISA MASALAH DAN PERANCANGAN APLIKASI

Bab ini menjelaskan analisa kebutuhan, identifikasi kebutuhan sistem, perancangan UML, perancangan layar, *menu*, *database*, serta perancangan *prototype* dengan mendemokan aplikasi.

BAB IV PENGUJIAN DAN IMPLEMENTASI

Pada bab ini, akan membahas tentang metode dan algoritma yang digunakan dalam perancangan aplikasi, serta spesifikasi perangkat keras dan perangkat lunak yang digunakan selama proses pembuatan aplikasi. Selain itu, akan dijelaskan tentang tampilan antarmuka program yang dibuat, dan pengujian aplikasi untuk memastikan kinerjanya sesuai dengan yang diharapkan.

BAB V SIMPULAN DAN SARAN

Dalam bab ini berfokus pada kesimpulan yang diperoleh dari hasil penelitian yang telah dilakukan dan saran yang digunakan sebagai masukan untuk pengembangan lebih lanjut di masa mendatang.



BAB II

LANDASAN TEORI

2.1. Teori Umum

2.1.1. Data

Menurut (Jauhari et al., 2020, p. 1) data merupakan suatu fakta atau kejadian yang akan diproses atau diolah untuk menghasilkan informasi. Data dapat berupa angka, simbol, gambar, suara, karakter, ataupun tanda-tanda yang dapat digunakan untuk dijadikan sebuah informasi. Data dibedakan menjadi dua macam, yaitu Data Internal dan Data Eksternal. Data Internal merupakan sebuah data yang menggambarkan kondisi suatu organisasi secara internal, sedangkan Data Eksternal merupakan sebuah data yang menggambarkan kondisi yang berada diluar organisasi. Selain kedua macam jenis data diatas, data juga terbagi berdasarkan klasifikasi dan cara memperolehnya, antara lain :

- a. Data Kuantitatif adalah data yang dapat dikuantifikasikan atau diukur secara langsung, biasanya disajikan dalam bentuk angka.
- b. Data Kualitatif adalah jenis data yang mencirikan atau menjelaskan sesuatu, biasanya diberikan sebagai kata-kata, gambar, atau suara yang bermakna.
- c. Data Primer adalah data yang dihimpun atau diperoleh langsung dalam suatu organisasi atau perorangan untuk kepentingan penelitian.

- d. Data Sekunder adalah data yang diperoleh secara tidak langsung, dikatakan secara tidak langsung karena datanya bisa diperoleh atau dikumpulkan lewat dokumen-dokumen yang sudah ada sebelumnya atau yang diterbitkan oleh berbagai instansi lain.

2.1.2. Informasi

Informasi merupakan sekumpulan data yang telah diolah dengan cara atau metode tertentu sehingga dapat dipahami dan digunakan sebagai dasar pengambilan keputusan (Bagaskoro, 2019, p. 35). Informasi dapat dianggap berguna jika memenuhi syarat-syarat sebagai berikut:

- a. Ketersediaan data yang memadai: Informasi yang berguna harus didukung oleh data yang memadai. Ini berarti bahwa data yang digunakan untuk menghasilkan informasi tersebut harus lengkap dan terpercaya.
- b. Mudah dipahami: Informasi yang berguna harus dapat dipahami oleh target audiensnya. Bahasa dan presentasi informasi harus jelas, sederhana, dan sesuai dengan tingkat pemahaman pengguna yang dimaksud. Pengguna harus dapat mengerti informasi tersebut tanpa kesulitan atau kebingungan yang berlebihan.
- c. Relevan: Informasi yang berguna harus terkait dengan kebutuhan atau pertanyaan yang ingin dijawab. Informasi tersebut harus memiliki hubungan langsung dengan topik atau masalah yang sedang dibahas, sehingga pengguna dapat mengambil keputusan atau tindakan yang tepat berdasarkan informasi tersebut.

- d. Bermanfaat: Informasi yang berguna harus memberikan manfaat nyata kepada pengguna. Informasi tersebut harus memberikan wawasan baru, pemahaman yang lebih baik, atau solusi untuk masalah yang dihadapi.
- e. Tepat waktu: Informasi yang berguna harus tersedia dalam waktu yang sesuai dengan kebutuhan pengguna. Keterlambatan dalam penyediaan informasi dapat membuatnya menjadi tidak berguna atau bahkan tidak relevan lagi.
- f. Andal: Informasi yang berguna harus diperoleh dari sumber yang andal dan terpercaya. Sumber informasi harus memiliki kredibilitas yang baik dan dipercaya oleh pengguna. Informasi yang tidak dapat dipercaya atau berasal dari sumber yang meragukan dapat mengakibatkan kesalahan atau keputusan yang salah.
- g. Akurat dan konsisten: Informasi yang berguna harus akurat dan konsisten dengan fakta atau kejadian yang ada. Informasi yang tidak akurat atau bertentangan dengan data lain dapat menyebabkan kesalahan dalam pengambilan keputusan.

Dengan memenuhi syarat-syarat di atas, informasi dapat dianggap berguna karena dapat memberikan nilai tambah bagi pengguna dalam konteks tertentu.

2.1.3. Aplikasi

Menurut (Apsiswanto & Muharni, 2022, p. 66) Aplikasi merupakan *program* komputer yang dirancang khusus untuk menjalankan tugas tertentu bagi pengguna. Aplikasi seringkali dikembangkan untuk memenuhi kebutuhan atau menyelesaikan masalah tertentu yang dihadapi oleh pengguna. Jenis-jenis aplikasi sangat beragam dan masing-masing memiliki fungsi yang berbeda-beda.

2.1.4. Website

Menurut (Widia & Asriningtias, 2021, p. 3) *Website* merupakan kumpulan-kumpulan halaman *web* yang berisi teks berformat HTML (*Hyper Text Markup Language*). Tidak hanya berisikan teks saja, halaman *web* umumnya juga berisi gambar, video dan elemen lainnya yang disajikan dalam format digital. Dengan menggunakan alamat unik yaitu URL (*Uniform Resource Locator*), *website* dapat dikunjungi menggunakan *browser* yang terhubung dengan *internet*. *Website* terdiri dari 2 macam, antara lain:

a. *Website* Statis

Situs *web* yang kontennya statis atau tidak berubah. Jika ingin melakukan perubahan, harus dilakukan secara manual pada kode sumber situs *web* tersebut.

b. *Website* Dinamis

Website yang kontennya dapat berubah secara dinamis memungkinkan pembaruan konten dengan mudah dan teratur tanpa perlu mengedit *source code* secara langsung. Hal ini disebabkan karena konten dalam *website* disimpan dalam *database*.

2.1.5. Tripadvisor

Tripadvisor merupakan sebuah *platform* daring yang memungkinkan pengguna untuk mencari dan menemukan berbagai jenis akomodasi, tempat wisata, restoran, dan pengalaman perjalanan di berbagai tempat di seluruh dunia. Setiap bulan, lebih dari 350 juta orang mengunjungi situs Tripadvisor, yang merupakan komunitas wisata terbesar di dunia dengan lebih dari 290 juta ulasan dan komentar tentang tempat wisata (Utama & Junaedi, 2018, p. 160).

2.1.6. Analisis Sentimen

Opinion mining atau sering disebut Analisis sentimen merupakan sebuah bidang penelitian yang bertujuan untuk menganalisis dan mengidentifikasi opini, sentimen, dan emosi yang terkandung dalam sebuah teks. Proses analisis sentimen melibatkan pengelompokan polaritas dari teks yang dianalisis, baik itu dalam dokumen, kalimat, komentar, atau ulasan untuk menentukan apakah pendapat yang diungkapkan bersifat positif, negatif, atau netral (Nugraha et al., 2020, p. 25).

2.2. Teori Khusus

2.2.1. Data Mining

Menurut (Arhami & Nasir, 2020, p. 1) 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.

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 (Munawar et al., 2023, pp. 62–63):

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.

2.2.2. *Machine Learning*

Menurut (Erwin et al., 2023, p. 109) *machine learning* adalah sebuah bagian dari kecerdasan buatan dimana komputer dapat untuk belajar dari data dan mengambil keputusan atau melakukan tugas tertentu tanpa perlu di kode secara eksplisit. Dengan kata lain, komputer dapat memahami data yang diberikan dan mengidentifikasi pola-pola yang tersembunyi didalamnya dan menggunakan pengetahuan tersebut untuk melakukan prediksi, klasifikasi, atau pengambilan keputusan secara akurat dan efisien. Dalam *machine learning* terdapat tiga pendekatan utama diantara sebagai berikut:

a. *Supervised Learning*

Pendekatan di mana data yang telah diberikan kategori (label atau kelas) ke dalam algoritma untuk mempelajari pola-pola dalam data tersebut dan melakukan prediksi atau mengklasifikasikan pada data yang belum dilabeli dengan akurat.

b. *Unsupervised Learning*

Pada pendekatan ini, tidak ada label yang diberikan pada data dan algoritma harus berusaha menemukan pola-pola yang tersembunyi dalam kumpulan data itu sendiri.

c. *Reinforcement Learning*

Pada pendekatan ini, komputer belajar melalui pengalaman dengan melakukan serangkaian tindakan percobaan dan menerima umpan balik positif atau negatif berdasarkan hasil tindakan tersebut.

2.2.3. Klasifikasi

Menurut (Wulandari & Sifaunajah, 2019, p. 1) 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.

2.2.4. Naïve Bayes

Menurut (Pratiwi et al., 2020, p. 20) 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.

Tahapan dari proses naive bayes meliputi langkah-langkah sebagai berikut (Isa et al., 2022, p. 82):

- a. Menjumlahkan kelas/label.
- b. Menjumlahkan kasus per kelas.
- c. Mengalikan semua variabel kelas.
- d. Membandingkan hasil per kelas.

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

2.2.5. Text Mining

Menurut (Urva et al., 2023, p. 122) *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.

2.2.6. *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 (Nugraha et al., 2020, p. 4). Adapun tahapan dalam *text processing* diantaranya adalah sebagai berikut (Nugraha et al., 2020, p. 5):

a. *Case Folding*

Pada tahap ini, dilakukan perubahan terhadap keseluruhan teks menjadi huruf kecil semua.

b. *Cleaning*

Pada tahap ini, pembersihan data dilakukan yang melibatkan penghapusan karakter khusus seperti tanda baca, URL, dan huruf selain abjad A-Z, serta penghapusan tanda pagar (#) dan nama pengguna (@). Dalam tahap *cleaning* dapat dilakukan dengan cara yang berbeda-beda tergantung pada kebutuhan dari teks yang akan dibersihkan.

c. *Tokenizing*

Pada tahap ini, dilakukan pemisahan kalimat menjadi satuan kata. Langkah ini bertujuan untuk memudahkan dalam membedakan karakter tertentu yang dapat dianggap sebagai pemisah kata atau bukan.

d. *Filtering*

Pada tahap ini, dilakukan proses penghapusan kata yang dianggap tidak penting atau tidak memiliki sebuah makna menggunakan algoritma *stoplist*.

e. *Stemming*

Pada tahap ini, dilakukan perubahan kata-kata yang berimbuhan menjadi bentuk kata dasar.

2.2.7. *Confusion Matrix*

Confusion matrix merupakan alat yang berguna untuk mengukur kinerja model seberapa baik pengklasifikasiannya dapat mengenali fitur dari kelas yang berbeda (Saifudin, 2022, p. 64). *Matrix* ini terdiri dari empat sel, yaitu sebagai berikut:

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

Tabel 2. 1 *Confusion Matrix*

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

Sumber: (Saifudin, 2022, p. 64)

Setelah dibuat *confusion matrix*, terdapat beberapa rumus yang dapat digunakan untuk menghitung matrik evaluasi kinerja model klasifikasi.

Berikut adalah beberapa rumus yang umum digunakan (Saifudin, 2022, p. 65):

a. Akurasi

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

b. Presisi

$$\text{Presisi} = \frac{TP}{TP + FP}$$

c. Recall

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

d. Spesifisitas

$$\text{Spesifisitas} = \frac{TN}{TN + FP}$$

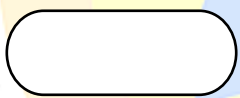
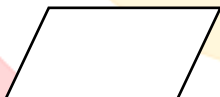

e. F1-Score

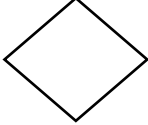



$$\text{F1-Score} = 2 * \frac{\text{Presisi} * \text{Recall}}{\text{Presisi} + \text{Recall}}$$




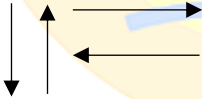
2.2.8. Flowchart

Flowchart adalah penggambaran sebuah algoritma yang terstruktur dan mudah dipahami oleh orang lain, dengan tujuan mengilustrasikan urutan logis dari suatu prosedur pemecahan masalah secara sederhana, terurai, rapi dan jelas. *Flowchart* menyajikan langkah-langkah penyelesaian masalah dalam bentuk simbol-simbol tertentu dan dapat dimengerti secara umum (Sitorus, 2015, p. 14). Dengan menggunakan simbol-simbol tersebut, *flowchart* membantu dalam menganalisis, merencanakan, dan mengkomunikasikan serangkaian langkah-langkah atau proses secara efektif. Dengan demikian, *flowchart* dapat menyajikan solusi secara visual yang lebih jelas dan mudah dipahami. Berikut simbol-simbol dari *flowchart*:

Tabel 2. 2 Simbol-simbol *Flowchart*

Simbol	Nama	Fungsi
	<i>Terminal</i>	Digunakan untuk memulai dan mengakhiri proses.
	<i>Input/Output</i>	Berfungsi untuk menyatakan proses <i>input</i> dan <i>output</i> .
	<i>Process</i>	Berfungsi untuk setiap tindakan atau tahap yang dilakukan oleh komputer.

	<p><i>Decision</i></p>	<p>Digunakan untuk menunjukkan suatu kondisi tertentu yang akan menghasilkan dua kemungkinan jawaban yaitu ya atau tidak.</p>
	<p><i>Connector</i></p>	<p>Digunakan untuk menghubungkan proses ke proses lain yang berada di halaman yang sama.</p>
	<p><i>Offline Connector</i></p>	<p>Digunakan untuk menghubungkan proses ke proses lainnya yang berada di halaman yang berbeda.</p>
	<p><i>Predefined Process</i></p>	<p>Digunakan untuk menyediakan tempat penyimpanan suatu pengolahan untuk memberikan nilai awal.</p>

	<p><i>Punched Card</i></p>	<p>Digunakan untuk menyatakan <i>input</i> yang berasal dari kartu atau <i>output</i> ditulis ke kartu.</p>
	<p><i>Punch Tape</i></p>	<p>Digunakan untuk menunjukkan <i>input</i> atau <i>output</i> menggunakan pita kertas berlubang.</p>
	<p><i>Document</i></p>	<p>Digunakan untuk menunjukkan bahwa <i>input</i> berasal dari dokumen dalam bentuk fisik, atau mencetak <i>output</i>.</p>
	<p><i>Flow</i></p>	<p>Digunakan untuk menyambungkan antara setiap simbol dan menunjukkan aliran proses.</p>

Sumber: (Sitorus, 2015, p. 15)

2.3. Teori Perancangan

2.3.1. *Hypertext Markup Language (HTML)*

Hypertext Markup Language (HTML) adalah sebuah bahasa *markup* yang berfungsi untuk membuat dan mengatur konten pada halaman *web*. Dalam pembuatan halaman *web*, HTML tidak berdiri sendiri, melainkan dapat digunakan bersama dengan CSS dan JavaScript untuk menghasilkan tampilan *web* yang terstruktur, menarik dan interaktif (Anamisa & Mufarroha, 2022, p. 21).

HTML menggunakan sebuah *tag* yang ditempatkan di sekitar teks untuk memberikan instruksi kepada *browser web* tentang bagaimana menampilkan konten tersebut. Setiap *tag* HTML memiliki tanda "<" dan ">" yang menunjukkan awal dan akhir. Terdapat beberapa *tag* yang penting dalam dokumen HTML (Anamisa & Mufarroha, 2022, p. 26), antara lain:

- a. <html> dan </html> : Tag ini menunjukkan awal dan akhir dari dokumen HTML.
- b. <head> dan </head> : Tag ini digunakan untuk membuat bagian header yang berisi informasi meta dan judul halaman web.
- c. <title> dan </title> : Pada Tag ini digunakan untuk memberi judul pada tab browser.
- d. <body> </body> : Tag ini berfungsi sebagai pembuka dan penutup bagian isi atau body, yang berisi konten yang akan ditampilkan pada halaman web.

2.3.2. *Hypertext Preprocessor (PHP)*

Menurut (Fauzan & Nurhidayah, 2020, p. 1) PHP merupakan bahasa pemrograman *server-side* yang berarti kode PHP akan diproses di sisi *server* sebelum hasilnya dikirim ke *browser*. PHP bersifat *software open source*, yang berarti dapat diakses, digunakan, dan dimodifikasi secara gratis oleh siapa pun, serta dapat berjalan pada berbagai macam sistem operasi. Dalam pengembangan *web*, memungkinkan pengembang untuk menyisipkan kode PHP ke dalam dokumen *Hypertext Markup Language (HTML)* guna menghasilkan konten yang dinamis sesuai dengan permintaan pengguna.

2.3.3. *JavaScript (JS)*

Menurut (Azis et al., 2019, p. 10) *JavaScript* merupakan bahasa pemrograman yang ringan dan mudah digunakan untuk pengembangan aplikasi *web* dengan kemampuan interaktif yang menarik. Dengan *JavaScript*, halaman *web* dapat memiliki beberapa fungsionalitas dan antarmuka yang interaktif, sehingga lebih dari sekadar tampilan statis yang menampilkan data dan informasi saja.

2.3.4. *Cascading Style Sheets (CSS)*

CSS (*Cascading Style Sheets*) adalah bahasa *stylesheet* yang digunakan untuk mengatur dan mendesain tampilan pada *website* yang ditulis dengan menggunakan penanda bahasa markup (Assyifa & Andarsyah, 2020, p. 20). Dengan menggunakan CSS, pengembang dapat menentukan dan mengatur berbagai elemen seperti warna, jenis huruf, ukuran tulisan, latar belakang, dan banyak lagi. CSS memberikan fleksibilitas yang besar dalam

mendesain tampilan sebuah *website*, sehingga pengembang dapat menciptakan tampilan yang menarik dan konsisten di berbagai perangkat dan ukuran layar.

2.3.5. *Bootstrap*

Bootstrap adalah sebuah *library framework* CSS yang dibuat khusus dalam pengembangan *front-end web*. *Bootstrap* juga menggunakan kombinasi HTML, CSS, dan *JavaScript* yang digunakan untuk menciptakan situs *website* yang responsif. Artinya, halaman *website* yang dikembangkan dengan *Bootstrap* dapat menyesuaikan tampilannya dengan baik pada berbagai perangkat dan ukuran layar, mulai dari desktop, tablet, dan ponsel (Sunarya & Bahit, 2020, p. 16).

Bootstrap menyediakan berbagai komponen yang siap pakai, seperti tombol, formulir, navigasi, tabel dan banyak lagi. Dengan menggunakan *Bootstrap*, pengembang dapat menghemat waktu dalam membangun tampilan *website* yang menarik.

2.3.6. XAMPP

XAMPP adalah perangkat lunak *open source* yang sangat populer untuk pengembangan *web* lokal. Dengan menginstal XAMPP, maka tidak perlu lagi melakukan instalasi dan konfigurasi manual untuk *web server Apache*, PHP, dan MySQL, karena XAMPP menyediakan paket perangkat lunak tersebut dalam satu paket yang siap digunakan (Aprilian et al., 2020, p. 50).

Fungsi utama XAMPP adalah sebagai *server web* lokal yang mandiri (*localhost*) yang terdiri dari berbagai *program*, seperti *Apache server*, MySQL, dan penerjemahan bahasa menggunakan PHP dan Perl. XAMPP berasal dari singkatan X (mengacu pada sistem operasi apa pun), *Apache*, MySQL, PHP, dan Perl.

2.3.7. *Visual Studio Code*

Menurut (Roza et al., 2020, p. 78) *Visual Studio Code* adalah sebuah teks editor yang ringan dan handal yang dikembangkan oleh *Microsoft*. Teks editor ini bersifat *open source* dan dapat digunakan di berbagai sistem operasi, termasuk Windows, macOS, dan Linux. *Visual Studio Code* mendukung berbagai bahasa pemrograman seperti *JavaScript*, *TypeScript*, dan banyak lagi. *Visual Studio Code* menawarkan berbagai fungsionalitas, seperti fitur *Intellisense*, *Git Integration*, *Debugging* dan fitur ekstensi. Fitur-fitur tersebut akan terus berkembang seiring dengan peningkatan versi *Visual Studio Code*.

2.3.8. *MySQL*

MySQL adalah sebuah sistem manajemen basis data (*Database Management System/DBMS*) yang beroperasi dengan perintah dasar SQL (*Structured Query Language*) sebagai bahasa yang menghubungkan perangkat lunak dengan *server database* (Aprilian et al., 2020, p. 55). *MySQL* dirancang untuk menyediakan penyimpanan, pengelolaan, dan akses data yang efisien dalam struktur tabel dengan kolom dan baris, menjadikannya pilihan yang populer di kalangan pengembang perangkat lunak.

2.3.9. ERD (*Entity Relationship Diagram*)

Menurut (Zein et al., 2023, p. 67) ERD adalah alat visual yang digunakan untuk menggambarkan struktur data dan hubungan antar entitas. Tujuannya untuk memberikan gambaran yang mudah dipahami mengenai hubungan antar data yang saling terhubung satu sama lain.

2.3.10. Python

Menurut (Wardana, 2019, p. 1) Python adalah bahasa pemrograman yang bersifat *open source* yang dibuat oleh Guido van Rossum dari Amsterdam, Belanda. Python merupakan bahasa pemrograman tingkat tinggi yang sangat populer dan sering digunakan untuk pembuatan perangkat lunak, pengembangan web, analisis data, IoT, pembelajaran mesin, dan sebagainya.

2.4. Tinjauan Studi

a. Penelitian Winda Yulita, Eko Dwi Nugroho, Muhammad Habib Algifari

Tabel 2. 3 Penelitian Winda Yulita, Eko Dwi Nugroho, Muhammad Habib Algifari

No	Data Jurnal / Makalah	Keterangan
1	Judul	Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier
2	Jurnal	Jurnal Data Mining dan Sistem Informasi
3	Volume dan Halaman	Volume 2 No. 2, 1-9
4	Tanggal dan Tahun	2021

5	Penulis	Winda Yulita, Eko Dwi Nugroho, Muhammad Habib Algifari
6	Penerbit	Universitas Teknokrat Indonesia
7	Tujuan Penelitian	Untuk menganalisis pendapat publik terkait proses vaksinasi dalam kasus COVID-19 dengan mempertimbangkan pesan yang diposting di Twitter.
8	Lokasi dan Subjek Penelitian	Penelitian ini menggunakan data yang diambil menggunakan API pada twitter sebanyak 3780 <i>tweet</i>
9	Perancangan Sistem	-
10	Hasil Penelitian	Hasil analisis menggunakan metode Naïve Bayes <i>Classifier</i> menunjukkan bahwa mayoritas masyarakat Indonesia memberikan respons positif terhadap kebijakan vaksinasi Covid-19, yang mencapai persentase sebesar 60.3%. Respon negatif terhadap vaksinasi hanya sebesar 5.4%, sedangkan respon netral mencapai 34.4%. Tingkat akurasi keseluruhan penelitian mencapai 93%.
11	Kekuatan Penelitian	Kekuatan penelitian adalah: a. Metode yang digunakan dapat digunakan sebagai dasar untuk penelitian yang akan datang.

		<p>b. Hasil penelitian ini menunjukkan keakuratan yang sangat tinggi sebesar 93%.</p>
12	Kelemahan Penelitian	<p>Kelemahan penelitian adalah:</p> <p>Tidak adanya perancangan sistem untuk menampilkan analisis sentimennya dan keterbatasan dalam memperoleh data yang hanya bisa diperoleh dalam tujuh hari terakhir</p>
13	Kesimpulan	<p>Kesimpulan:</p> <p>Analisis sentimen terhadap opini masyarakat tentang vaksinasi Covid-19 pada twitter dapat menghasilkan tingkat akurasi sebesar 93% dengan persentase data terbanyak pada respon positif sebesar 60.3%, dengan persentase sebanyak itu membuktikan bahwa kebijakan vaksinasi Covid-19 akan membawa dampak positif bagi masyarakat</p>

b. Penelitian Fransiska Vina Sari, Arief Wibowo

Tabel 2. 4 Penelitian Fransiska Vina Sari, Arief Wibowo

No	Data Jurnal / Makalah	Keterangan
1	Judul	Analisis Sentimen Pelanggan Toko Online JD.ID Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi
2	Jurnal	Jurnal Simetris
3	Volume dan Halaman	Volume 10 No. 2, 681-686
4	Tanggal dan Tahun	2 November 2019
5	Penulis	Fransiska Vina Sari, Arief Wibowo
6	Penerbit	Universitas Budi Luhur
7	Tujuan Penelitian	Menganalisis opini masyarakat pada twitter terhadap kinerja JD.id serta untuk mengetahui layanan maupun produk dari <i>marketplace</i> yang bermanfaat bagi pelanggan maupun penjual dan menambahkan fitur konversi ikon emosi
8	Lokasi dan Subjek Penelitian	Penelitian ini menggunakan data yang diambil menggunakan RStudio dengan memasukkan settingan twitter API dan mengetikkan kata kunci JD.id sehingga memperoleh sampel data sebanyak 900 tweet
9	Perancangan Sistem	a. Bahasa Pemrograman RStudio, PHP dan HTML

		b. RapidMiner
10	Hasil Penelitian	Hasil dari penelitian ini adalah merancang aplikasi program analisis sentimen yang dikategorikan menjadi sentimen positif, negatif dan netral menggunakan metode naïve bayes <i>classifier</i> dan konversi ikon emosi sehingga mendapatkan akurasi sebesar 98%
11	Kekuatan Penelitian	Kekuatan penelitian adalah: <ul style="list-style-type: none"> a. Metode yang digunakan dapat menjadi acuan untuk penelitian-penelitian berikutnya b. Hasil penelitian ini menunjukkan keakuratan yang sangat tinggi sebesar 98%
12	Kelemahan Penelitian	Kelemahan penelitian adalah: Pengambilan sampel berdasarkan pertimbangannya saja, besar dan kriteria sampel yang ditentukan.
13	Kesimpulan	Kesimpulan: Analisis sentimen opini masyarakat terhadap kinerja, layanan, produk pada JD.id di twitter dengan mengkategorikan menjadi sentimen positif, negatif dan netral menggunakan Naïve Bayes <i>Classifier</i> dan konversi ikon

		emosi sehingga menghasilkan tingkat akurasi sebesar 98%
--	--	---

c. Penelitian Retno Sari, Ratih Yulia Hayuningtyas

Tabel 2. 5 Penelitian Retno Sari, Ratih Yulia Hayuningtyas

No	Data Jurnal / Makalah	Keterangan
1	Judul	Penerapan Algoritma Naïve Bayes Untuk Analisis Sentimen Pada Wisata TMII
2	Jurnal	<i>IJSE – Indonesian Journal on Software Engineering</i>
3	Volume dan Halaman	Volume 5 No. 2, 51-60
4	Tanggal dan Tahun	2 Desember 2019
5	Penulis	Retno Sari, Ratih Yulia Hayuningtyas
6	Penerbit	Universitas Bina Sarana Informatika
7	Tujuan Penelitian	Melakukan analisis terhadap ulasan para wisatawan yang telah mengunjungi Taman Mini Indonesia Indah. Hal ini bertujuan untuk memahami kondisi dari objek wisata tersebut sebelum calon wisatawan mengunjunginya.
8	Lokasi dan Subjek Penelitian	Penelitian ini menggunakan data ulasan pada objek wisata Taman Mini Indonesia Indah yang didapat dari Tripadvisor sebanyak 100 ulasan

9	Perancangan Sistem	<p>a. Bahasa Pemrograman PHP dan HTML</p> <p>b. Database <i>MySQL</i></p>
10	Hasil Penelitian	<p>Hasil dari penelitian ini adalah membuat <i>website</i> menggunakan metode naïve bayes untuk analisis sentimen ulasan wisatawan Taman Mini Indonesia Indah dengan pengklasifikasiannya menjadi positif atau negatif dan menghasilkan akurasi sebesar 70%</p>
11	Kekuatan Penelitian	<p>Kekuatan penelitian adalah:</p> <p>a. Metode yang digunakan dapat menjadi landasan atau dasar yang baik untuk penelitian-penelitian mendatang</p> <p>b. Hasil penelitian ini menunjukkan keakuratan yang cukup tinggi sebesar 70%</p>
12	Kelemahan Penelitian	<p>Kelemahan penelitian adalah:</p> <p>Data yang digunakan dalam penelitian ini dapat terbilang cukup sedikit sehingga mempengaruhi tingkat akurasiannya</p>
13	Kesimpulan	<p>Kesimpulan:</p> <p>Membuat <i>website</i> untuk menganalisis ulasan wisatawan Taman Mini Indonesia Indah</p>

		dengan menggunakan algoritma Naive Bayes dengan pengklasifikasiannya menjadi positif dan negatif guna mengetahui terlebih dahulu kondisi dari objek wisata tersebut. Dengan algoritma Naive Bayes dapat menghasilkan tingkat akurasi sebesar 70%
--	--	--

d. Penelitian Rizki Wahyudi, Gilang Kusumawardhana

Tabel 2. 6 Penelitian Rizki Wahyudi, Gilang Kusumawardhana

No	Data Jurnal / Makalah	Keterangan
1	Judul	Analisis Sentimen Pada <i>Review</i> Aplikasi Grab di <i>Google Play Store</i> Menggunakan <i>Support Vector Machine</i>
2	Jurnal	Jurnal Informatika
3	Volume dan Halaman	Volume 8 No. 2, 200-207
4	Tanggal dan Tahun	2 September 2021
5	Penulis	Rizki Wahyudi, Gilang Kusumawardhana
6	Penerbit	Universitas Amikom Purwokerto
7	Tujuan Penelitian	Tujuan dari penelitian ini adalah untuk melakukan analisis terhadap ulasan pengguna pada aplikasi Grab yang terdapat di situs <i>Google Play</i> , dengan menggunakan metode klasifikasi <i>Support Vector Machine</i> (SVM). Penelitian ini bertujuan untuk

		<p>mengidentifikasi sentimen yang terkandung dalam ulasan pengguna, sekaligus memahami pandangan pengguna terhadap layanan transportasi <i>online</i> Grab. Dengan mendapatkan informasi yang lebih akurat dan komprehensif dari hasil analisis, diharapkan informasi tersebut dapat menjadi pedoman bagi Grab dalam mempertahankan popularitas, meningkatkan kualitas pelayanan, serta melakukan perbaikan dan evaluasi untuk meningkatkan pengalaman pengguna secara keseluruhan</p>
8	Lokasi dan Subjek Penelitian	<p>Penelitian ini menggunakan data sebanyak 1000 <i>review</i> pengguna Grab yang diperoleh melalui teknik <i>web scraping</i> dari situs <i>Google Play</i></p>
9	Perancangan Sistem	-
10	Hasil Penelitian	<p>Hasil dari penelitian ini adalah visualisasi data <i>review</i> pengguna aplikasi Grab di situs <i>Google Play</i> menggunakan metode klasifikasi <i>Support Vector Machine</i> (SVM). Analisis dilakukan untuk mengidentifikasi kata-kata yang paling sering muncul dalam <i>review</i> positif dan negatif. Dalam <i>review</i> positif, kata-kata yang paling sering muncul</p>

		adalah "ovo", "sangat", dan "grabfood", sedangkan dalam <i>review</i> negatif adalah "driver", "aplikasi", dan "tolong". Dalam penelitian ini menghasilkan akurasi sebesar 85,54%
11	Kekuatan Penelitian	Kekuatan penelitian adalah: <ul style="list-style-type: none"> a. Metode yang digunakan dapat dipertimbangkan sebagai dasar penelitian selanjutnya b. Hasil penelitian ini menunjukkan tingkat keakurasian sebesar 85,54%
12	Kelemahan Penelitian	Kelemahan penelitian adalah: Tidak adanya perancangan sistem untuk menampilkan analisis sentimennya
13	Kesimpulan	Kesimpulan: Berdasarkan analisis sentimen pada <i>review</i> aplikasi grab di <i>Google Play Store</i> menggunakan <i>Support Vector Machine</i> , diperoleh tingkat akurasi sebesar 85,54% dengan kata-kata yang sering muncul dalam <i>review</i> positif adalah "ovo", "sangat" dan "grabfood" sedangkan <i>review</i> negatif adalah "driver, "aplikasi", dan "tolong"

e. Penelitian Yerik Afrianto Singgalen

Tabel 2. 7 Penelitian Yerik Afrianto Singgalen

No	Data Jurnal / Makalah	Keterangan
1	Judul	Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Tripadvisor Menggunakan Algoritma Naive Bayes Classifier
2	Jurnal	<i>Building of Informatics, Technology and Science (BITS)</i>
3	Volume dan Halaman	Volume 4 No. 3, 1343-1352
4	Tanggal dan Tahun	Desember 2022
5	Penulis	Yerik Afrianto Singgalen
6	Penerbit	Forum Kerjasama Pendidikan Tinggi (FKPT)
7	Tujuan Penelitian	Tujuan dari penelitian ini adalah melakukan analisis sentimen pengunjung di Candi Borobudur, sebuah destinasi wisata di Indonesia. Penelitian akan menggunakan metode klasifikasi Naive Bayes berdasarkan data ulasan pengguna dari aplikasi <i>website</i> Tripadvisor. Dengan demikian, penelitian ini bertujuan untuk memberikan rekomendasi yang tepat untuk mengakomodasi perubahan preferensi pengunjung dan meningkatkan citra pariwisata Indonesia. Melalui analisis

		sentimen yang komprehensif terhadap produk dan jasa di Candi Borobudur, diharapkan hasil penelitian ini dapat memberikan kontribusi positif terhadap pengembangan dan promosi pariwisata di Indonesia
8	Lokasi dan Subjek Penelitian	Penelitian ini menggunakan data ulasan pengguna aplikasi <i>website</i> Tripadvisor untuk destinasi wisata Candi Borobudur. Proses pengambilan data dilakukan menggunakan aplikasi <i>Webharvy</i> , dan berhasil mengumpulkan sebanyak 7.603 data ulasan.
9	Perancangan Sistem	-
10	Hasil Penelitian	Hasil penelitian ini merupakan hasil dari pengklasifikasian data teks ulasan pengguna dari <i>website</i> Tripadvisor terhadap destinasi wisata Candi Borobudur menggunakan algoritma Naïve Bayes Classifier. Dalam penelitian ini, tingkat akurasi yang dicapai mencapai sebesar 96,36%. Hasil pengklasifikasian ini memberikan rekomendasi yang dapat meningkatkan kinerja layanan produk dan jasa yang terkait dengan keaslian warisan budaya serta menerapkan nilai-nilai Sapta Pesona untuk

		meningkatkan kepuasan pengunjung dan citra destinasi wisata Candi Borobudur.
11	Kekuatan Penelitian	<p>Kekuatan penelitian adalah:</p> <ol style="list-style-type: none"> a. Metode yang digunakan dalam penelitian ini dapat menjadi landasan atau acuan bagi penelitian-penelitian yang akan dilakukan di masa mendatang b. Hasil penelitian ini menunjukkan tingkat akurasi yang sangat tinggi sebesar 96,36%
12	Kelemahan Penelitian	<p>Kelemahan penelitian adalah:</p> <p>Tidak adanya perancangan sistem untuk menampilkan hasil analisis sentimennya</p>
13	Kesimpulan	<p>Kesimpulan:</p> <p>Menganalisis sentimen wisatawan terhadap ulasan Candi Borobudur di Tripadvisor menggunakan algoritma Naïve Bayes <i>Classifier</i> menjadi sentimen positif dan negatif dengan tingkat akurasi sebesar 96,36%. Penelitian ini juga mengimplementasikan metode <i>SMOTE Upsampling</i> untuk menyelesaikan masalah data yang tidak seimbang. Hasil penelitian ini memberikan rekomendasi untuk</p>

		meningkatkan kinerja layanan produk dan jasa yang terkait dengan keaslian warisan budaya, serta menerapkan nilai-nilai Sapta Pesona guna meningkatkan kepuasan pengunjung dan citra destinasi wisata Candi Borobudur.
--	--	---

f. Penelitian Titan Nugraha, Purwantoro, Yuyun Umaidah

Tabel 2. 8 Penelitian Titan Nugraha, Purwantoro, Yuyun Umaidah

No	Data Jurnal / Makalah	Keterangan
1	Judul	Analisis Sentimen terhadap Perpanjangan Masa Jabatan Presiden Indonesia Menggunakan Algoritma Naïve Bayes
2	Jurnal	Jurnal Pendidikan dan Konseling
3	Volume dan Halaman	Volume 4 No. 4, 4625-4635
4	Tanggal dan Tahun	2022
5	Penulis	Titan Nugraha, Purwantoro, Yuyun Umaidah
6	Penerbit	Universitas Pahlawan Tuanku Tambusai
7	Tujuan Penelitian	Tujuan penelitian ini adalah untuk menggunakan algoritma Naïve Bayes dalam menganalisis sentimen masyarakat Indonesia terhadap usulan perpanjangan masa jabatan Presiden Indonesia. Selain itu, penelitian ini juga bertujuan untuk mengidentifikasi opini

		positif dan negatif yang terdapat dalam cuitan di media sosial Twitter yang berkaitan dengan wacana perpanjangan masa jabatan Presiden tersebut
8	Lokasi dan Subjek Penelitian	Penelitian ini menggunakan metode crawling data dengan menggunakan Twitter API. Data diambil dalam rentang waktu dari 1 Juni 2022 hingga 30 Juni 2022, dan sebanyak 1.000 tweet berhasil dikumpulkan
9	Perancangan Sistem	-
10	Hasil Penelitian	Hasil penelitian ini adalah analisis sentimen masyarakat di Twitter terhadap usulan perpanjangan masa jabatan Presiden Indonesia menggunakan algoritma Naïve Bayes. Data yang dikumpulkan dari 1.000 tweet kemudian diproses dan dilabeli menjadi 641 data, dengan 422 tweet positif dan 192 tweet negatif. Berdasarkan analisis, ditemukan bahwa sebanyak 69% sentimen masyarakat cenderung positif terhadap usulan perpanjangan masa jabatan Presiden, sementara 31% sentimen cenderung negatif. Hasil ini menunjukkan mayoritas masyarakat Indonesia memiliki pendapat yang positif terhadap usulan tersebut

11	Kekuatan Penelitian	<p>Kekuatan penelitian adalah:</p> <ol style="list-style-type: none"> a. Metode yang digunakan dalam penelitian ini dapat menjadi landasan atau acuan bagi penelitian-penelitian yang akan dilakukan di masa mendatang b. Hasil penelitian ini menunjukkan tingkat akurasi yang sangat tinggi sebesar 95% dengan perbandingan 90% data <i>training</i> dan 10% data <i>testing</i>
12	Kelemahan Penelitian	<p>Kelemahan penelitian adalah:</p> <p>Tidak adanya perancangan sistem untuk menampilkan hasil analisis sentimennya</p>
13	Kesimpulan	<p>Kesimpulan:</p> <p>Menganalisis sentimen masyarakat di twitter terhadap usulan perpanjangan masa jabatan Presiden Indonesia dengan menggunakan algoritma Naïve Bayes, penelitian ini menggunakan data sebanyak 1.000 data kemudian diproses dan dilabeli menjadi 641 data, dengan 422 tweet positif dan 192 tweet negatif. Berdasarkan analisis, ditemukan bahwa sebanyak 69% sentimen masyarakat cenderung positif terhadap usulan</p>

		<p>perpanjangan masa jabatan Presiden, sementara 31% sentimen cenderung negatif. Hasil ini menunjukkan mayoritas masyarakat Indonesia memiliki pendapat yang positif terhadap usulan tersebut. Penelitian ini mendapatkan akurasi sebesar 95% dengan membagi 90% data <i>training</i> dan 10% data <i>testing</i></p>
--	--	---

g. Ringkasan Dari Model Penelitian

Tabel 2. 9 Ringkasan Model Penelitian

Peneliti	Nama Jurnal	Tahun	Institusi	Judul dan Metode yang digunakan	Kesimpulan
Winda Yulita, Eko Dwi Nugroho, Muhammad Habib Algifari	Jurnal Data Mining dan Sistem Informasi Vol. 2 No. 2 ISSN: 2745-8458	2021	Universitas Teknokrat Indonesia	Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes <i>Classifier</i>	Analisis sentimen terhadap opini masyarakat tentang vaksinasi Covid-19 pada twitter dapat menghasilkan tingkat akurasi sebesar 93%

					dengan persentase data terbanyak pada respon positif sebesar 60.3%, dengan persentase sebanyak itu membuktikan bahwa kebijakan vaksinasi Covid-19 akan membawa dampak positif bagi masyarakat
Fransiska Vina Sari, Arief Wibowo	Jurnal Simetris Vol. 10 No. 2 P-ISSN: 2252-4983 E-ISSN: 2549-3108	2019	Universitas Budi Luhur	Analisis Sentimen Pelanggan Toko Online JD.ID Menggunakan Metode Naïve Bayes	Analisis sentimen opini masyarakat terhadap kinerja, layanan, produk pada JD.id di twitter

				<p><i>Classifier</i> Berbasis Konversi Ikon Emosi</p>	<p>dengan dikategorikan menjadi sentimen positif, negatif dan netral menggunakan Naïve Bayes <i>Classifier</i> dan konversi ikon emosi sehingga menghasilkan tingkat akurasi sebesar 98%</p>
Retno Sari, Ratih Yulia Hayuningtyas	<p>IJSE – <i>Indonesian Journal on Software Engineering</i> Vol. 5 No. 2 p-ISSN: 2461-0690 e-ISSN: 2714-9935</p>	2019	Universitas Bina Sarana Informatika	<p>Penerapan Algoritma Naïve Bayes Untuk Analisis Sentimen Pada Wisata TMII</p>	<p>Membuat <i>website</i> untuk menganalisis ulasan wisatawan Taman Mini Indonesia Indah dengan menggunakan algoritma</p>

					Naive Bayes dengan klasifikasi menjadi positif dan negatif guna mengetahui terlebih dahulu kondisi dari objek wisata tersebut. Dengan algoritma Naive Bayes dapat menghasilkan tingkat akurasi sebesar 70%
Rizki Wahyudi, Gilang Kusumawardhana	Jurnal Informatika Vol. 8 No. 2 ISSN: 2355-6579 E-ISSN: 2528-2247	2021	Universitas Amikom Purwokerto	Analisis Sentimen Pada <i>Review</i> Aplikasi Grab di <i>Google Play Store</i> Menggunakan	Berdasarkan analisis sentimen pada <i>review</i> aplikasi grab di <i>Google Play Store</i> menggunakan

				<i>Support Vector Machine</i>	<i>Support Vector Machine</i> , diperoleh tingkat akurasi sebesar 85,54% dengan kata-kata yang sering muncul dalam <i>review</i> positif adalah “ovo”, “sangat” dan “grabfood” sedangkan <i>review</i> negatif adalah “driver”, “aplikasi”, dan “tolong”
Yerik Afrianto Singgalen	<i>Building of Informatics, Technology and Science (BITS)</i> Vol. 4 No. 3	2022	Forum Kerjasama Pendidikan Tinggi (FKPT)	Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di	Menganalisis sentimen wisatawan terhadap ulasan Candi Borobudur di

	<p>p-ISSN: 2684-8910</p> <p>e-ISSN: 2685-3310</p>			<p>Tripadvisor</p> <p>Menggunakan</p> <p>Algoritma</p> <p>Naive Bayes</p> <p><i>Classifier</i></p>	<p>Tripadvisor</p> <p>menggunakan</p> <p>algoritma</p> <p>Naïve Bayes</p> <p><i>Classifier</i></p> <p>menjadi</p> <p>sentimen</p> <p>positif dan</p> <p>negatif dengan</p> <p>tingkat akurasi</p> <p>sebesar</p> <p>96,36%.</p> <p>Penelitian ini</p> <p>juga</p> <p>menggunakan</p> <p>metode</p> <p><i>SMOTE</i></p> <p><i>Upsampling</i></p> <p>untuk</p> <p>menyelesaikan</p> <p>masalah data</p> <p>yang tidak</p> <p>seimbang.</p> <p>Hasil</p> <p>penelitian ini</p>
--	---	--	--	--	---

					<p>memberikan rekomendasi untuk meningkatkan kinerja layanan produk dan jasa yang terkait dengan keaslian warisan budaya, serta menerapkan nilai-nilai Sapta Pesona guna meningkatkan kepuasan pengunjung dan citra destinasi wisata Candi Borobudur</p>
Titan Nugraha, Purwantoro,	Jurnal Pendidikan	2022	Universitas Pahlawan	Analisis Sentimen terhadap	Menganalisis sentimen masyarakat di

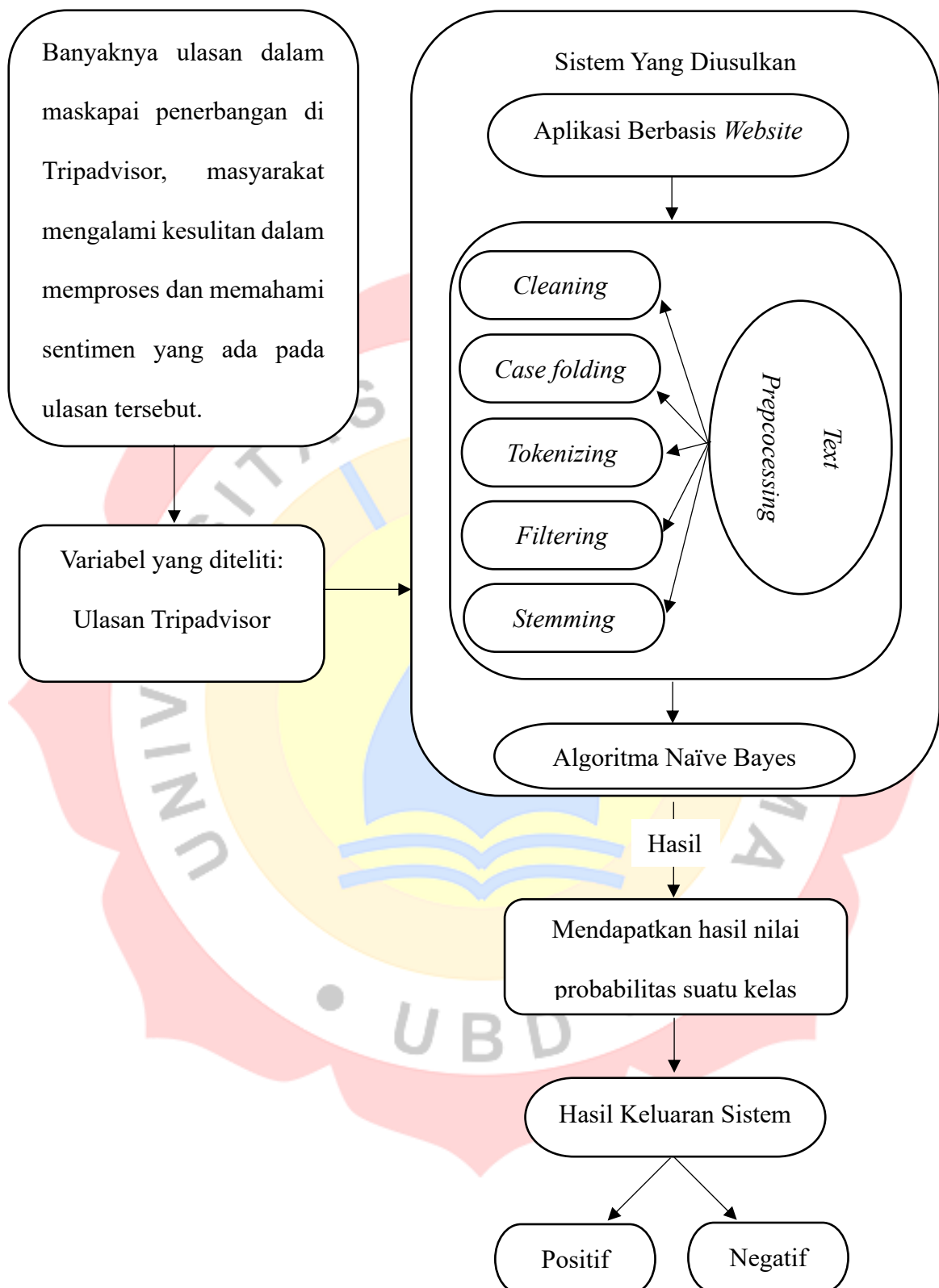
<p>Yuyun Umaidah</p>	<p>dan Konseling Vol. 4 No. 4 E-ISSN: 2685-936X P-ISSN: 2685-9351</p>		<p>TuanKu Tambusai</p>	<p>Perpanjangan Masa Jabatan Presiden Indonesia Menggunakan Algoritma Naïve Bayes</p>	<p>twitter terhadap usulan perpanjangan masa jabatan Presiden Indonesia dengan menggunakan algoritma Naïve Bayes, penelitian ini menggunakan data sebanyak 1.000 data kemudian diproses dan dilabeli menjadi 641 data, dengan 422 tweet positif dan 192 tweet negatif. Berdasarkan analisis,</p>
--------------------------	---	--	----------------------------	---	--

					<p>ditemukan bahwa sebanyak 69% sentimen masyarakat cenderung positif terhadap usulan perpanjangan masa jabatan Presiden, sementara 31% sentimen cenderung negatif. Hasil ini menunjukkan mayoritas masyarakat Indonesia memiliki pendapat yang positif terhadap</p>
--	--	--	--	--	--

					<p>usulan tersebut.</p> <p>Penelitian ini mendapatkan akurasi sebesar 95% dengan membagi 90% data <i>training</i> dan 10% data <i>testing</i></p>
--	--	--	--	--	---

Berdasarkan tinjauan jurnal yang telah dilakukan, dapat disimpulkan bahwa setelah membandingkan berbagai jurnal, maka memutuskan untuk menggunakan metode Naïve Bayes. Metode ini terbukti memiliki performa yang baik dalam analisis sentimen dan tingkat akurasi yang tinggi.

2.5. Kerangka Pemikiran



Gambar 2. 1 Diagram Kerangka Pemikiran

BAB III

ANALISA MASALAH & PERANCANGAN APLIKASI

3.1. Identifikasi Kebutuhan Sistem

Untuk memenuhi kebutuhan sistem, dilakukan pengadaaan data melalui survei yang melibatkan masyarakat berbagai kalangan sebagai responden. Jawaban yang diberikan oleh responden dalam survei ini digunakan sebagai acuan dalam perancangan aplikasi analisis sentimen yang akan dikembangkan.

3.1.1. *Requirement Elicitation* Tahap I

Hasil dari survei yang dilakukan oleh responden mendapatkan beberapa kebutuhan yang diperlukan untuk membangun sistem analisis sentimen ini. Berikut ini merupakan uraian hasil survei yang telah dilakukan:

Tabel 3. 1 *Requirement Elicitation* Tahap I

Fungsional	
Analisa Kebutuhan Sistem	
Saya ingin sistem ini dapat:	
1	Menampilkan hasil klasifikasi sentimen (positif dan negatif)
2	Menampilkan hasil analisa dalam bentuk grafik
3	Kemampuan untuk melakukan filter ulasan berdasarkan sentimen (positif dan negatif)
4	Menampilkan nama maskapai yang dipilih dan ulasannya
5	Menampilkan perbandingan antar maskapai
6	Menampilkan proses dari analisa data dan dapat memilih maskapai lain
7	Memberikan informasi tentang jadwal penerbangan dan harga

Non Fungsional	
Saya ingin sistem ini dapat:	
1	Memberikan tampilan yang mudah dipahami
2	Memberikan desain yang menarik
3	Dapat diakses melalui berbagai perangkat (<i>responsive</i>)
4	Menampilkan ulasan dan hasil grafik dengan cepat

3.1.2. Requirement Elicitation Tahap II

Berdasarkan hasil *requirement elicitation* tahap I, langkah selanjutnya melibatkan klasifikasi menggunakan metode MDI yang bertujuan untuk memisahkan antara rancangan sistem yang penting dengan rancangan yang disanggupi. Pada metode ini, tanda “M” yaitu *Mandatory* (Tidak diperbolehkan untuk dihilangkan), “D” yaitu *Desirable* (Boleh dihilangkan), dan “I” yaitu *Inessential* (kebutuhan yang bukanlah bagian dari sistem yang dibahas). Berikut hasil pengklasifikasiannya:

Tabel 3. 2 *Requirement Elicitation* Tahap II

Fungsional		M	D	I
Analisa Kebutuhan Sistem				
Saya ingin sistem ini dapat:				
1	Menampilkan hasil klasifikasi sentimen (positif dan negatif)	✓		
2	Menampilkan hasil analisa dalam bentuk grafik	✓		
3	Kemampuan untuk melakukan filter ulasan berdasarkan sentimen (positif dan negatif)		✓	

4	Menampilkan nama maskapai yang dipilih dan ulasannya	✓		
5	Menampilkan perbandingan antar maskapai	✓		
6	Menampilkan proses dari analisa data dan dapat memilih maskapai lain	✓		
7	Memberikan informasi tentang jadwal penerbangan dan harga			✓
Non Fungsional				
Saya ingin sistem ini dapat:				
1	Memberikan tampilan yang mudah dipahami		✓	
2	Memberikan desain yang menarik		✓	
3	Dapat diakses melalui berbagai perangkat (<i>responsive</i>)		✓	
4	Menampilkan ulasan dan hasil grafik dengan cepat	✓		

3.1.3. Requirement Elicitation Tahap III

Pada tahap ini, dilakukan penyusutan dengan mengeliminasi semua kebutuhan pada pilihan “I” yaitu *Inessential* berdasarkan *requirement elicitation* tahap II diatas, kemudian diklasifikasikan kembali menggunakan metode TOE. Pada metode ini, tanda “T” yaitu *Technical* adalah tata cara pembuatan *requirement*, tanda “O” yaitu *Operational* adalah tata cara penggunaan *requirement*, dan tanda “E” yaitu *Economy* adalah biaya yang diperlukan. Berikut hasil pengklasifikasiannya:

Tabel 3. 3 *Requirement Elicitation* Tahap III

Fungsional										
Analisa Kebutuhan Sistem										
Saya ingin sistem ini dapat:										
<i>Feasibility</i>		T			O			E		
<i>Risk</i>		L	M	H	L	M	H	L	M	H
1	Menampilkan hasil klasifikasi sentimen (positif dan negatif)	✓				✓		✓		
2	Menampilkan hasil analisa dalam bentuk grafik		✓			✓			✓	
3	Kemampuan untuk melakukan filter ulasan berdasarkan sentimen (positif dan negatif)		✓			✓			✓	
4	Menampilkan nama maskapai yang dipilih dan ulasannya	✓			✓			✓		
5	Menampilkan perbandingan antar maskapai			✓		✓			✓	
6	Menampilkan proses dari analisa data dan dapat memilih maskapai lain		✓			✓		✓		

Non Fungsional										
Analisa Kebutuhan Sistem										
Saya ingin sistem ini dapat:										
<i>Feasibility</i>		T			O			E		
<i>Risk</i>		L	M	H	L	M	H	L	M	H
1	Memberikan tampilan yang mudah dipahami		✓			✓		✓		
2	Memberikan desain yang menarik	✓			✓			✓		
3	Dapat diakses melalui berbagai perangkat (<i>responsive</i>)		✓			✓			✓	
4	Menampilkan ulasan dan hasil grafik dengan cepat		✓			✓			✓	

3.1.4. Requirement Elicitation Final

Setelah melakukan tahap III, pada tahap ini merupakan tahap akhir dari proses elisitasi yang digunakan dalam pembuatan sistem berdasarkan kebutuhan yang diperlukan. Berikut adalah hasil akhirnya:

Tabel 3. 4 Requirement Elicitation Final

Fungsional	
Analisa Kebutuhan Sistem	
No.	Saya ingin sistem ini dapat:
1	Menampilkan hasil klasifikasi sentimen (positif dan negatif)
2	Menampilkan hasil analisa dalam bentuk grafik
3	Kemampuan untuk melakukan filter ulasan berdasarkan sentimen (positif dan negatif)

4	Menampilkan nama maskapai yang dipilih dan ulasannya
5	Menampilkan perbandingan antar maskapai
6	Menampilkan proses dari analisa data dan dapat memilih maskapai lain
Non Fungsional	
No	Saya ingin sistem ini dapat:
1	Memberikan tampilan yang mudah dipahami
2	Memberikan desain yang menarik
3	Dapat diakses melalui berbagai perangkat (<i>responsive</i>)
4	Menampilkan ulasan dan hasil grafik dengan cepat

3.2. Pengambilan Data

Dalam tahap ini, 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*.

Gambar 3. 1 Hasil *Scraping* Data Ulasan

3.3. 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*.

a. *Cleaning*

Di situs *web* Tripadvisor, simbol, tanda baca, dan angka sering muncul di setiap ulasan, sehingga data menjadi tidak berguna. Pada tahap ini dilakukan untuk menghilangkan simbol-simbol, tanda baca dan angka selain teks yang terdapat dalam sebuah kalimat atau ulasan tersebut dengan menggunakan *regex*. Berikut ini adalah contoh hasil *cleaning*:

Tabel 3. 5 Contoh Hasil Proses *Cleaning*

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

b. *Case Folding*

Case folding dilakukan 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. Proses ini dapat dilakukan menggunakan fungsi bawaan yang tersedia, seperti *strtolower()* dalam PHP, yang berfungsi untuk mengubah karakter dalam sebuah kalimat menjadi huruf kecil. Berikut ini adalah contoh hasil *case folding*:

Tabel 3. 6 Contoh Hasil Proses *Case Folding*

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

c. *Tokenizing*

Tokenizing merupakan proses 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. *Tokenizing* ini dilakukan agar setiap bagian teks yang ada pada ulasan Tripadvisor dapat dibersihkan. *Tokenizing* dilakukan pada saat akan melakukan proses *Stopword* atau *Filtering*.

Berikut ini adalah contoh hasil proses *tokenizing*:

Tabel 3. 7 Contoh Hasil Proses *Tokenizing*

Sebelum	Sesudah
tiba tiba jadwal penerbangan diubah kali dan pemberitahuannya mepet	["tiba", "tiba", "jadwal", "penerbangan", "diubah", "kali",

	“dan”, “pemberitahuannya”, “mepet”]
penerbangan lancar tidak ada delay	[“penerbangan”, “lancar”, “tidak”, “ada”, “delay”]

d. *Filtering/Stopword*

Setelah dilakukan proses *Cleaning*, *Case Folding* dan *Tokenizing*, kemudian dilakukan proses penghapusan *stopword* bertujuan 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. Proses penghapusan *stopword* dilakukan dengan menggunakan token-token berupa teks. Setiap kata dalam token akan diperiksa satu per satu untuk melihat apakah kata-kata tersebut terdapat dalam daftar *stopword* atau tidak. Jika tidak ada kata tersebut didalam daftar maka kata tersebut akan digunakan kembali kemudian digabungkan menjadi teks utuh. Berikut ini adalah contoh daftar kata-kata *stopword*:

Tabel 3. 8 Daftar Kata *Stopword*

Daftar Kata <i>Stopword</i>	
adalah	adanya
adapun	agak
agakny	agar
akan	akankah
akhir	akhiri

aku	amat
anda	antara
apa	atas
atau	bahwa
balik	bapak

e. *Stemming*

Proses *stemming* dilakukan setelah menghapus *stopword*, dengan tujuan mengubah kata-kata dalam teks menjadi bentuk kata dasar dengan menghapus imbuhan yang terdapat di depan atau di belakang kata. Dalam proses ini, digunakan *library* Sastrawi, yang merupakan sebuah *library* untuk melakukan pemrosesan *Natural Language Processing* (NLP) dalam bahasa Indonesia yang dikembangkan oleh Andy Librian. *Library* ini bersumber pada <https://github.com/sastrawi/sastrawi/tree/master/src/Sastrawi/Stemmer>.

3.4. Pelabelan Leksikon

Setelah melakukan pemrosesan teks, langkah selanjutnya adalah memberi label dengan menggunakan kamus atau leksikon berupa kata-kata yang mengandung sentimen. Kamus ini akan digunakan untuk menentukan sentimen positif atau negatif dari setiap ulasan. Kamus yang digunakan dapat ditemukan di <https://github.com/masdevid/ID-OpinionWords>. Kamus tersebut awalnya dikembangkan oleh Bing Liu dan Minqing Hu, kemudian dimodifikasi atau diterjemahkan oleh Wahid, D.H., dan Azhari, S.N. ke dalam bahasa Indonesia. Kamus ini terdiri dari 2402 kata negatif dan 1182 kata positif. Berikut ini merupakan contoh dari kumpulan kata-kata positif dan negatif dalam kamus tersebut:

Tabel 3. 9 Contoh Daftar Kata Sentimen Positif dan Negatif

Kata Bersentimen Positif		Kata Bersentimen Negatif	
acungan jempol	adaptif	abnormal	absurd
adil	afinitas	acak	acak-acakan
afirmasi	agilely	acuh	adiktif
agung	ahli	agresi	agresif
ahlinya	ajaib	aib	alergi
akurat	aman	ambigu	amoral
amanah	amat	anarki	anarkis
ambisius	andal	ancaman	aneh
anggun	antusias	angkuh	anjlok
apik	apresiasi	apatis	apatis

Setelah memiliki kamus, langkah selanjutnya adalah mencocokkan setiap kata dalam ulasan dengan kata-kata dalam kamus positif dan kamus negatif. Jika sebuah kata dalam ulasan cocok dengan kata-kata dalam kamus positif, jumlah kata yang cocok dengan kamus positif akan dihitung. Begitu pula jika kata dalam ulasan cocok dengan kata-kata dalam kamus negatif, jumlah kata yang cocok dengan kamus negatif akan dihitung. Setelah menghitung jumlah kata yang cocok dengan kamus positif dan negatif, langkah berikutnya adalah membandingkan jumlah kata positif dengan jumlah kata negatif. Jika jumlah kata positif lebih besar daripada jumlah kata negatif, ulasan tersebut akan diberi label positif. Sebaliknya, jika jumlah kata negatif lebih besar daripada jumlah kata positif, ulasan tersebut akan diberi label negatif. Setelah melakukan pelabelan menggunakan kamus atau leksikon, data dibagi menjadi 90% sebagai data *training* dan 10% sebagai data *testing*. Pembagian ini dilakukan untuk

melatih model menggunakan data *training* dan menguji kinerja model menggunakan data *testing*.

A	B	C	D
1 maskapai	ulasan_bersih	label	
2 citilink	layanan citilink buruk banget orang bantu turun tangga pesawat suruh tunggu tangga pesawat siap kursi roda info terbang lepas tangan alas meeupakan tanggung pramug; negatif		
3 citilink	kecewa mundur berangkat maskapai buruk sen tiket undur undur mundurrrrrrrrrr burukkkkk sekali kecewaaaaa berat maskapai citilink sggt berhargaaaa kasi mnngu b; negatif		
4 citilink	review kursi luas review layanan traveling bawa bayi extra safety belt bantu pasang biaya kursi bayi mahal banding kompetitor	positif	
5 citilink	kemarin tanggal jurus rumah ubah alas operasional andal jadwal mundur delay hrsnya ditambahin drama salah parkir urus gabung alas apa bikin susah akibat biaya trar; negatif		
6 citilink	lalu check online pindah drop bagasi profesional	negatif	
7 citilink	cancelled flight input data whatsapp salah input data credit card salah kembali dana belit tiga salah citilink lempar tanggung poor customer service langsung tutup bicara; negatif		
8 citilink	tolong training pramugari nama vivi lumbanmerasa cantik layanan buruk	negatif	
9 citilink	umur hidup beli tiket citilink kecewa customer service rude sekali bantu sulit trip kali pilih citilink karna terbang jadwal ambil salah hospital sana karna willingly citilink com; negatif		
10 citilink	terbang delay well masuk pesawat sambut pramugari ramah bantu beda kursi suami syukur temu pramugari nama claudia moga salah spell nama hehe langsung info tu; positif		
11 citilink	jelek banget layanan kecewa laku check counter ucap twirima kasih diam jalan gate hand carry sampe foto gate masuk pesawat komplek hebat layanan maaf sopan terima ka; positif		
12 citilink	kali citilink delay kacau pemberitahuan kali lambat delay lah delay kali check ruang tunggu delay delay announcement berangkat pesawat citilink announcementnya palsu a; negatif		
13 citilink	terbang juni terbang didelay citilink hari berangkat malam berangkat berangkat bandara terbang delay menit berangkat protes kerja mgkin pepet pesawat jakarta tuju j; negatif		
14 citilink	boarding menit take masuk alas customer servicenya ngasih jelas kecewa	negatif	
15 citilink	cancel citilink coba customer service angkat chat email balas uang refund anggap payah citilink hrsnya tanggung	positif	
16 citilink	pesawat delay tulis sesuai jadwal terbang citilink delay lambat kecewa	negatif	
17 citilink	menit take boleh masuk tugas citilink bilang lambat beli tiket terbang beli tiket sampe boarding tugas bilang sebentar boarding jadwal take kecewa jadwal bilang lambat; negatif		
18 citilink	berangkat reschedule bayar ribu biaya admin selisih hargabandingkan tetangga batik indonesia full service reschedule dibatasin berangkat biaya administrasi kali lipatov; positif		
19 citilink	memikmati malam medan terbang bandara soekarno hatta masuk pesawat terbang darat kuala namu layanan crew citilink ramah malam terbang murah murah terima ka; positif		
20 citilink	jakarta semarang pilih maskapai Garuda Indonesia grup citilink kali jalan wisata kerja pilih maskapai kesan terbang citilink landing asa	negatif	
21 citilink	pesawat banget panas bagus bagasi negri beli	positif	
22 citilink	terbang bagus Denpasar di pramugari pramugara ramah ramah makan enak harga last minutes booking baik mudah promo shock sale hadir bagus pake website apps r; positif		
23 citilink	standar full service terbang pagi cocok bisnis time pilih seat free minum mineral	positif	
24 citilink	beli tiket citilink terbang Surabaya Palembang langsung alhamdulillah makan enak beli beli online enak tambah miles Garuda bagasi gratis maskapai saing citilink tarik uar; positif		
25 citilink	kecewa temu senior flight attendant skill komunikasi mes beli seat online terbang Surabaya Banjarmasin seat nomer sesuai info beli seat daftar beli seat serius peleh buk; positif		
26 citilink	pertamakali coba maskapai buka rute Perth Bandara Yogyakarta mes rute tanggal Februari ganti keluarga bawa anak jalan lelah tambah bagasi Bandara Denpasar bal am; negatif		
27 citilink	nyaman terbang citilink awak kabin rapi ramah salam pisah unik pantun jenaka maju citilink bangga maskapai nasional idola negri	positif	
28 citilink	ontime check mudah bagasi dapet meals enak pesawat harga tiket sahabat	positif	
29 citilink	citilink enak take landingnya smooth stewardessnya ramah ramah pantun selalu ngenengin terbang citilink	positif	

Gambar 3. 2 Hasil Pelabelan Menggunakan Leksikon

3.5. Klasifikasi Manual Naïve Bayes

3.5.1. Perhitungan Manual Naïve Bayes

Dalam tahap ini, akan dilakukan contoh perhitungan manual menggunakan metode Naïve Bayes. Sebelum melakukan perhitungan manual, data akan dilakukan tahap *Text Preprocessing* terlebih dahulu agar datanya menjadi bersih dan bernilai. Berikut ini merupakan proses *preprocessing* dan perhitungan manual menggunakan metode naïve bayes.

Tabel 3. 10 Contoh Proses *Text Preprocessing*

Sebelum <i>Text Preprocessing</i>	Sesudah <i>Text Preprocessing</i>
Lion air emang bener bener maskapai terburuk sedunia, udah tiket gak murah, tapi pelayanannya sangat buruk	lion air emang maskapai buruk dunia tiket murah layanan

Masuk kabin sebelum take off dan keluar setelah landing sangat tidak terstruktur, berantakan, udah gitu penumpang berisik. nunggu bagasi sampai satu jam. sangat buruk.	masuk kabin take off landing struktur berantak gitu tumpang berisik tunggu bagasi buruk
Pertama kali saya melakukan penerbangan dengan Lion Air dan untuk kesannya bagus kok. Jadwal penerbangan on time tidak terkendala delay, dan lancar selama perjalanan meskipun sedikit was was.	kali terbang lion air kesan bagus jadwal time kendala delay lancar jalan
Satu hal yang disukai dari Penerbangan Lion Air di pagi hari adalah selalu On time sesuai schedule awal	suka terbang lion air pagi time sesuai schedule

Setelah dilakukan tahap *Text Preprocessing*, selanjutnya akan dilakukan pemisahan kata-kata (*tokenizing/tokenisasi*) untuk menentukan nilai bobot dari setiap kata-katanya agar dapat diproses ke tahap klasifikasi.

Tabel 3. 11 Proses *Tokenizing*

<i>Term</i>				
air	bagasi	bagus	berantak	berisik
buruk	delay	dunia	emang	gitu
jadwal	jalan	kabin	kali	kendala
kesan	lancar	landing	layan	lion
maskapai	masuk	murah	off	pagi

schedule	sesuai	struktur	suka	take
terbang	tiket	time	tumpang	tunggu

Setelah memisahkan kalimat menjadi per kata maka selanjutnya akan dilakukan pembobotan kata menggunakan *Term Frequency* pada masing-masing dokumen.

Tabel 3. 12 Pembobotan Kata

<i>Term</i>	D1(negatif)	D2(negatif)	D3(positif)	D3(positif)
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
Total Term	9	13	12	8
Total Bobot	42			

3.5.2. Proses Klasifikasi Data *Training*

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 layanan”

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”

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

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

Setelah dihitung nilai probabilitas setiap *term* maka selanjutnya masuk ke tahap perhitungan data *testing*. Berikut adalah hasil perhitungan *conditional probability*:

Tabel 3. 13 Hasil *Conditional Probability*

<i>Term</i>	P(w negatif)	P(w positif)
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

3.5.3. Proses Klasifikasi Data *Testing*

Pada tahap ini, 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.

“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 *testing*:

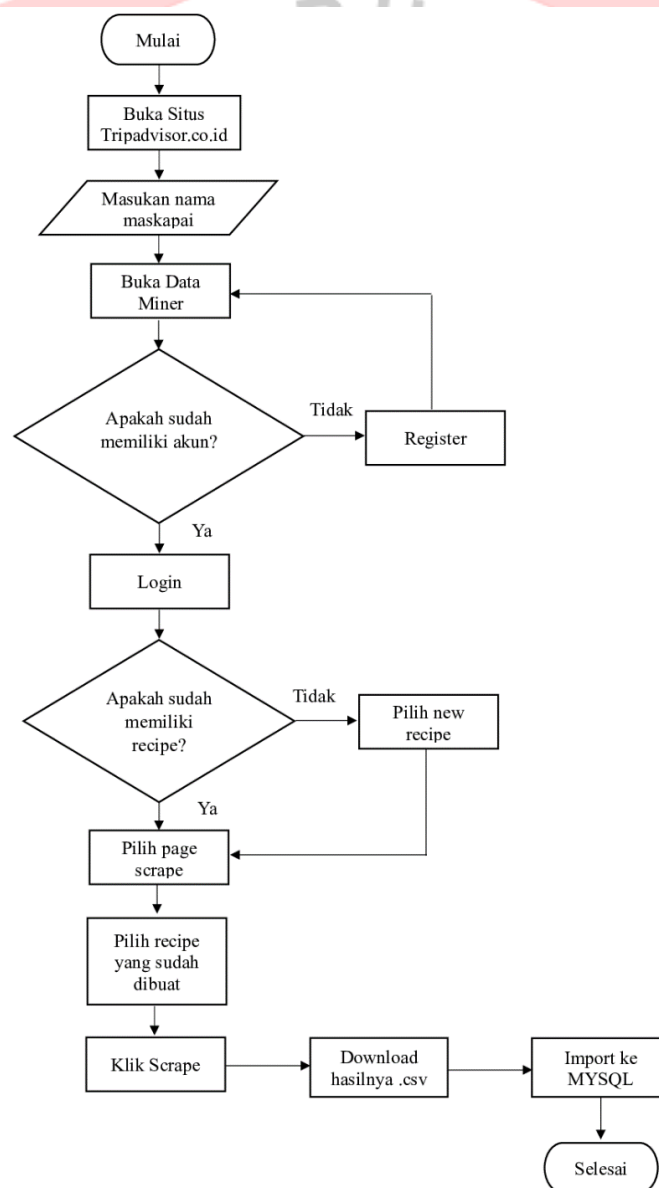
Tabel 3. 14 Hasil Klasifikasi Data *Testing*

Doc1			
Teks	Peluang kelas	Skor Peluang	Kelas Baru
lion air emang maskapai buruk dunia tiket murah layan	P(negatif)	6,0459E-14	Negatif
	P(positif)	9,7712E-16	
Doc2			
masuk kabin take off landing struktur berantak gitu tumpang berisik tunggu bagasi buruk	P(negatif)	9,1639E-20	Negatif
	P(positif)	1,1865E-23	
Doc3			
kali terbang lion air kesan bagus jadwal time kendala delay lancar jalan	P(negatif)	1,7003E-21	Positif
	P(positif)	1,3531E-17	
Doc4			
suka terbang lion air pagi time sesuai schedule	P(negatif)	1,7949E-14	Positif
	P(positif)	7,7388E-12	

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.

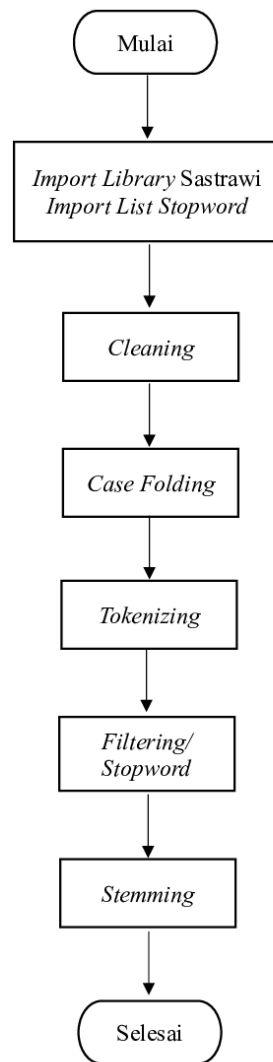
3.6. Flowchart

3.6.1. Flowchart Pengumpulan Data



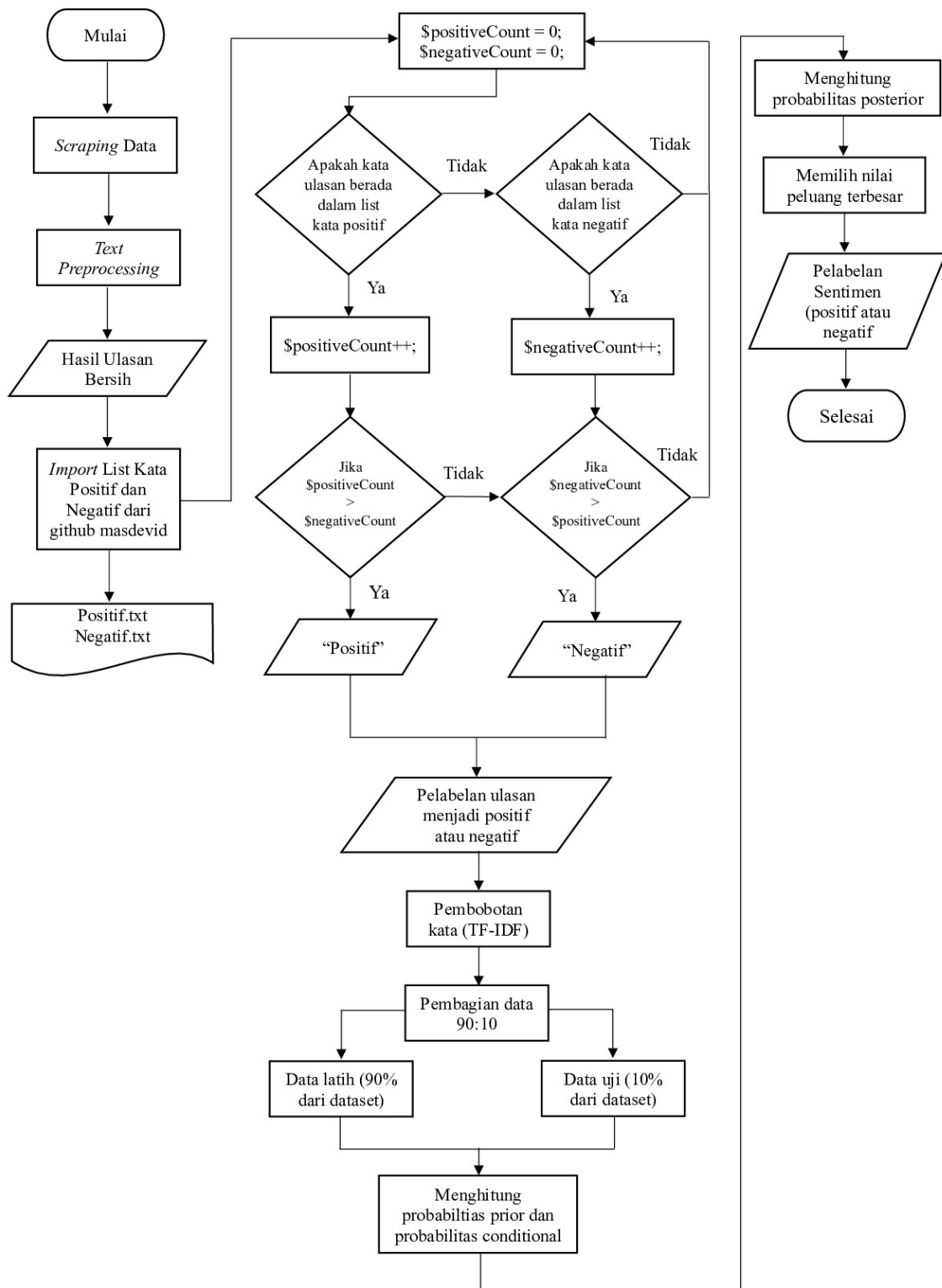
Gambar 3. 3 Flowchart Pengumpulan Data

3.6.2. *Flowchart Text Preprocessing*



Gambar 3. 4 *Flowchart Text Preprocessing*

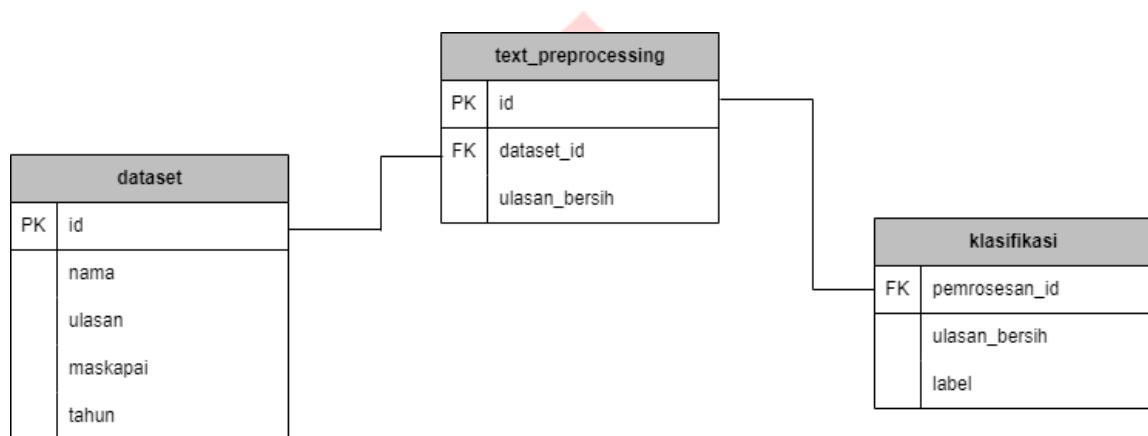
3.6.3. Flowchart Proses Klasifikasi Naïve Bayes



Gambar 3. 5 Flowchart Proses Klasifikasi Naïve Bayes

3.7. Perancangan Database

Pada tahap ini akan menjelaskan mengenai *database* yang dibuat dan mendesain struktur *database* yang digunakan untuk menyimpan dan mengolah data pada aplikasi yang akan dibuat. Berikut ini adalah diagram perancangan *database* atau *Entity Relationship Diagram* (ERD):



Gambar 3. 6 Entity Relationship Diagram

Tabel 3. 15 Rincian Tabel dataset

No.	Atribut	Jenis Data	Keterangan
1	id	int(4)	primary key dataset_maskapai
2	nama	varchar(32)	nama dari yang memberikan ulasan
3	ulasan	text	ulasan kotor hasil pengambilan dari extension google chrome
4	maskapai	varchar(8)	nama maskapai dari setiap ulasan
5	tahun	int(4)	tahun ulasan

Tabel 3. 16 Rincian Tabel *text_preprocessing*

No.	Atribut	Jenis Data	Keterangan
1	Id	int(4)	<i>primary key text_preprocessing</i>
2	dataset_id	int(4)	<i>Foreign key id pada tabel dataset_maskapai</i>
3	ulasan_bersih	text	ulasan yang sudah dibersihkan menggunakan tahap <i>preprocessing</i>

Tabel 3. 17 Rincian Tabel klasifikasi

No.	Atribut	Jenis Data	Keterangan
1	<i>text_preprocessing_id</i>	int(4)	<i>Foreign key id pada tabel text_preprocessing</i>
2	ulasan_bersih	text	ulasan yang sudah dibersihkan
3	label	varchar(15)	hasil pengkategorian menggunakan algoritma naïve bayes

3.8. Perancangan Layar

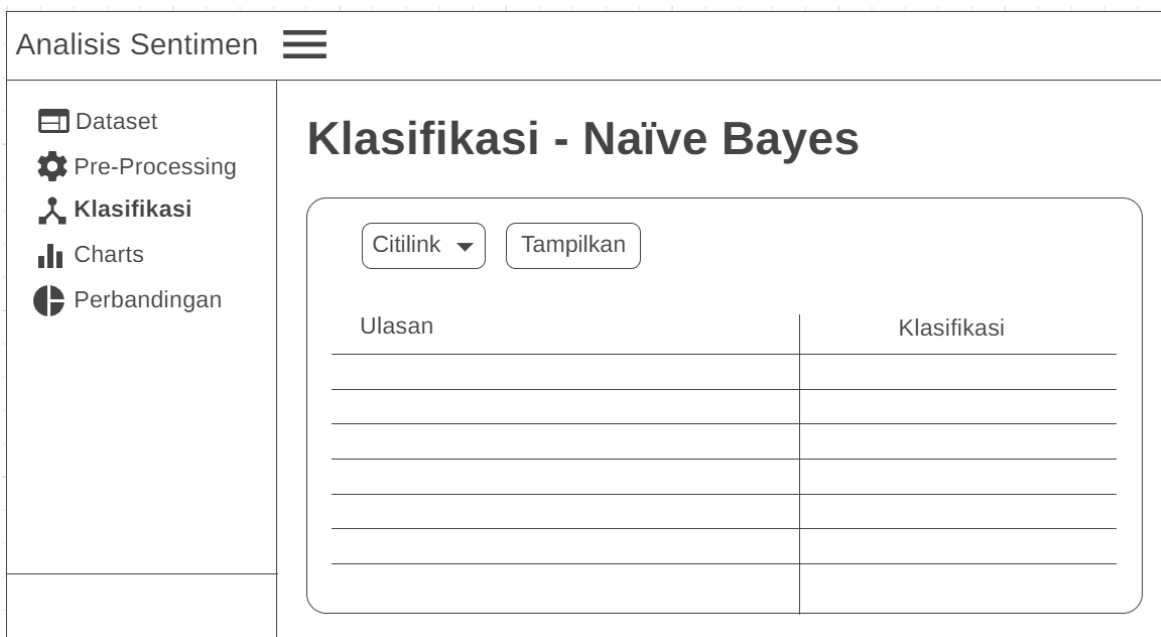
Aplikasi ini terdapat empat halaman, halaman pertama yaitu berisikan data yang sudah dikumpulkan dari situs Tripadvisor menggunakan data *scraper* dan disimpan kedalam database, halaman kedua berupa proses *text preprocessing*, data yang sudah dikumpulkan akan di lakukan pemrosesan teks mencakup *cleaning, case folding, tokenizing, filtering, dan stemming*. Setelah dilakukan proses tersebut pada halaman ketiga berupa proses pengklasifikasian menggunakan algoritma naïve bayes berbasis *lexicon based* dengan mengklasifikasikan sentimen menjadi positif dan

negatif. Kemudian hasil klasifikasi tersebut akan ditampilkan dalam sebuah grafik di halaman empat. Berikut ini adalah rancangan tampilan antarmuka pada aplikasi yang akan dibuat:

Gambar 3. 7 Rancangan Antarmuka Dataset

Ulasan	Case Folding	Filtering	Stemming

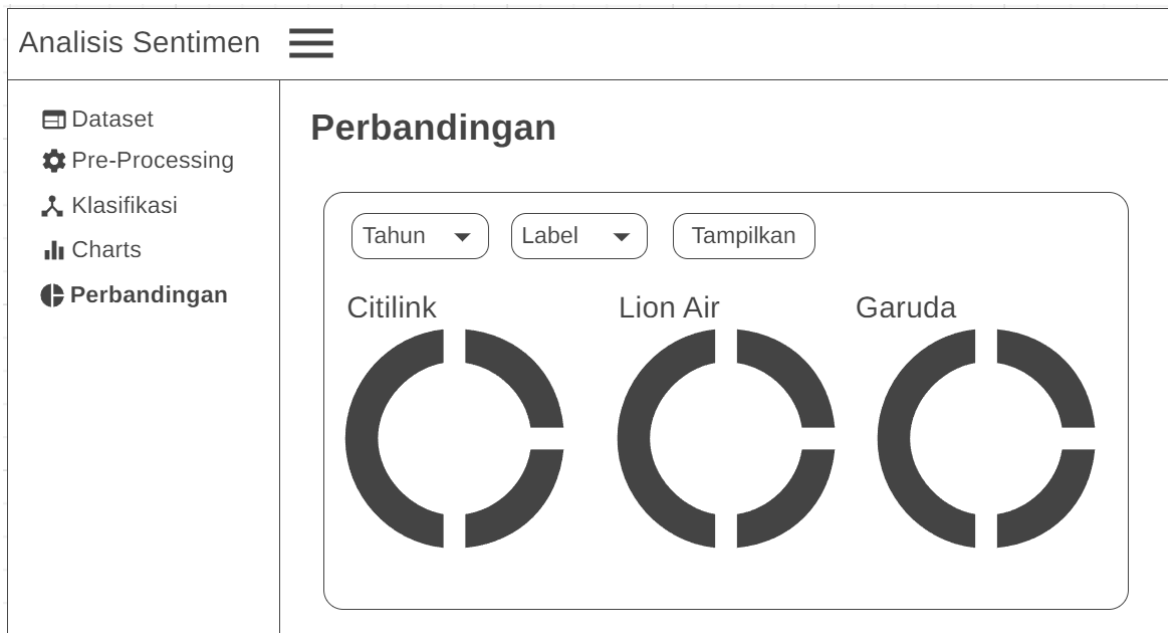
Gambar 3. 8 Rancangan Antarmuka *Text Preprocessing*



Gambar 3. 9 Rancangan Antarmuka Klasifikasi



Gambar 3. 10 Rancangan Antarmuka Grafik



Gambar 3. 11 Rancangan Antarmuka Perbandingan

