

**ANALISIS SENTIMEN WISATAWAN TERHADAP DESTINASI PANTAI
TANJUNG BIRA PADA ULASAN DI *GOOGLE MAPS* MENGGUNAKAN
ALGORITMA *NAIVE BAYES***

SKRIPSI

*Diajukan sebagai Salah Satu Syarat untuk Menyusun Skripsi Program
Studi Informatika*



**FAJAR MAULANA RAHMAN
105841107419**

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MUHAMMADIYAH MAKASSAR

2024

**ANALISIS SENTIMEN WISATAWAN TERHADAP DESTINASI PANTAI
TANJUNG BIRA PADA ULASAN DI *GOOGLE MAPS* MENGGUNAKAN
ALGORITMA *NAIVE BAYES***

**Diajukan Untuk Memenuhi Salah Satu Syarat Guna Memperoleh Gelar
Sarjana Komputer Prodi Informatika Fakultas Teknik Universitas
Muhammadiyah Makassar**



Disusun Dan Diajukan Oleh:

**FAJAR MAULANA RAHMAN
105841107419**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MAKASSAR**

2024



UNIVERSITAS MUHAMMADIYAH MAKASSAR
FAKULTAS TEKNIK

GEDUNG MENARA IQRA LT. 3

Jl. Sultan Alauddin No. 259 Telp. (0411) 866 972 Fax (0411) 865 588 Makassar 90221

Website: www.unismuh.ac.id, e_mail: unismuh@gmail.com

Website: <http://teknik.unismuh.makassar.ac.id>

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

PENGESAHAN

Skripsi atas nama **Fajar Maulana Rahman** dengan nomor induk Mahasiswa **105841107419**, dinyatakan diterima dan disahkan oleh Panitia Ujian Tugas Akhir/Skripsi sesuai dengan Surat Keputusan Dekan Fakultas Teknik Universitas Muhammadiyah Makassar Nomor : 001/05/A.5-III/1/45/2024, sebagai salah satu syarat guna memperoleh gelar Sarjana Komputer pada Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar pada hari kamis tanggal 31 Januari 2024.

Panitia Ujian :

Makassar, 19 Rajab 1445 H
 31 Januari 2024 M

1. Pengawas Umum

- a. Rektor Universitas Muhammadiyah Makassar
 Prof. Dr. H. AMBO ASSE, M.Ag
- b. Dekan Fakultas Teknik Universitas Hasanuddin
 Prof. Dr. Eng. MUHAMMAD ISRAN RAMLI, ST., MT

2. Penguji

- a. Ketua : Dr. Ir. Zahir Zainuddin, M.Sc
- b. Sekretaris : Muhyiddin A. M. Hayat, S.Kom., M.T

3. Anggota

- 1. Lukman Anas, S.Kom., M.T
- 2. Desi Anggreani, S.Kom., M.T
- 3. Titin Wahyuni, S.Pd., MT

Mengetahui :

Pembimbing I

Fahrim Irhamna Rachman S.Kom., M.T

Pembimbing II

Lukman, S.Kom., M.T



Dekan Fakultas Teknik

Dr. Ir. H. Nurmawaty, ST., MT., IPM

NBM : 795 108



UNIVERSITAS MUHAMMADIYAH MAKASSAR

FAKULTAS TEKNIK

GEDUNG MENARA IQRA LT. 3

Jl. Sultan Alauddin No. 259 Telp. (0411) 866 972 Fax (0411) 865 588 Makassar 90221
Website: www.unismuh.ac.id, e_mail: unismuh@gmail.com
Website: <http://teknik.unismuh.makassar.ac.id>

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan untuk memenuhi syarat ujian guna memperoleh gelar Sarjana Komputer (S.Kom) Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar.

Judul Skripsi : **ANALISIS SENTIMEN WISATAWAN TERHADAP DESTINASI PANTAI TANJUNG BIRA PADA ULASAN DI GOOGLE MAPS MENGGUNAKAN ALGORITMA NAIVE BAYES**

Nama : FAJAR MAULANA RAHMAN

Stambuk : 105841107419

Makassar, 31 Januari 2024

Telah Diperiksa dan Disetujui
Oleh Dosen Pembimbing:

Pembimbing I

Pembimbing II

Fahrim Irhamna Rachman, S.Kom., M.T

Lukman, S.Kom., M.T

Mengetahui,

Ketua Program Studi Informatika

Muhyidin A. M. Hayat, S.kom., M.T

NBM 1604577

ABSTRAK

Analisis sentimen merupakan sebuah penggambaran polaritas pada suatu teks atau kata. Jadi bisa disimpulkan bahwa analisis sentimen adalah pengolahan bahasa dari suatu teks atau kata yang dapat dianalisa dari pendapat, sikap atau penilaian dari seseorang terhadap suatu topik atau pembicaraan. Penelitian ini dilakukan dengan tujuan untuk mengetahui apakah analisis sentimen ulasan wisatawan terhadap destinasi Pantai Tanjung Bira dapat dilakukan menggunakan algoritma *Naive Bayes* dan untuk mengetahui sentimen umum wisatawan terhadap Pantai Tanjung Bira pada ulasan di *Google Maps* cenderung positif atau negatif. Penelitian ini membutuhkan data ulasan wisatawan di Pantai Tanjung Bira pada *Google Maps* yang bernilai Positif, Negatif dan Netral. Data ulasan yang digunakan diperoleh melalui ulasan pada *Google Maps* dengan teknik *web scraping* pada *Google Maps*. Pada penelitian ini, data dibagi menjadi dua yaitu data latih dan data uji. Jadi data sebanyak 720 data latih dan 180 data uji. Berdasarkan hasil yang telah didapatkan dari analisis sentimen wisatawan terhadap destinasi Pantai Tanjung Bira dengan menggunakan *Naive Bayes* diperoleh hasil yang cukup tinggi dengan *accuracy* sebesar 88%. Berdasarkan hasil yang telah didapatkan menunjukkan bahwa sentimen ulasan wisatawan terhadap Pantai Tanjung Bira pada *google maps* cenderung bersentimen negatif dengan persentase 37,2%. Sedangkan sentimen positif persentasenya 28,9% dan netral 33,9%.

Kata kunci: Analisis sentimen, Wisatawan, Tanjung Bira, *Naive Bayes*

ABSTRACT

Sentiment analysis is a depiction of polarity in a text or word. So it can be concluded that sentiment analysis is the language processing of a text or word that can be analyzed from someone's opinion, attitude or assessment of a topic or discussion. This research was conducted with the aim of finding out whether sentiment analysis of tourist reviews of the Tanjung Bira Beach destination can be carried out using the Naive Bayes algorithm and to find out whether the general sentiment of tourists towards Tanjung Bira Beach in reviews on Google Maps tends to be positive or negative. This research requires data on reviews of tourists at Tanjung Bira Beach on Google Maps with Positive, Negative and Neutral values. The review data used was obtained through reviews on Google Maps using web scraping techniques on Google Maps. In this research, the data is divided into two, namely training data and test data. So the data is 720 training data and 180 test data. Based on the results obtained from the analysis of tourist sentiment towards Tanjung Bira Beach destinations using Naive Bayes, quite high results were obtained with an accuracy of 88%. Based on the results obtained, it shows that the sentiment of tourist reviews of Tanjung Bira beach on Google Maps tends to be negative with a percentage of 37.2%. Meanwhile, the percentage of positive sentiment was 28.9% and neutral 33.9%.

Keyword: *Sentiment analysis, Travelers, Tanjung Bira, Naive Bayes*

KATA PENGANTAR

Assalamualaikum Wr. Wb

Segala puji bagi Allah SWT yang telah memberikan rahmat dan hidayahnya kepada penulis, sholawat beserta salam semoga tercurahkan kepada nabi Muhammad SAW sang revolusioner sejati yang menjadi suri tauladan seluruh ummat, yang telah menyebarluaskan Islam hingga sampai detik ini kita masih merasakan nikmat berislam. Sehingga penulis dapat menyelesaikan skripsi dengan judul "ANALISIS SENTIMEN WISATAWAN TERHADAP DESTINASI PANTAI TANJUNG BIRA PADA ULASAN DI *GOOGLE MAPS* MENGGUNAKAN ALGORITMA *NAIVE BAYES*" serta dapat menyelesaikan tepat waktu.

Penulis dapat menyelesaikan proposal skripsi ini karena bantuan dan dukungan dari berbagai pihak. Oleh karena itu dalam kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Terima kasih kepada Allah SWT
2. Kedua orangtua penulis, Ayahanda Abdul Rahman dan Ibunda Fatimah Endang yang selalu memberikan dukungan dan doa dalam setiap penulisan skripsi. Terima kasih yang tak terhingga atas do'a yang tulus, segala bentuk perhatian dan motivasi serta semua hal yang diberikan yang tidak dapat saya sebutkan satu satu.
3. Ibu Dr.Hj.Ir. Nurnawaty,ST.,MT Selaku Dekan Fakultas Teknik Universitas Muhammadiyah Makassar
4. Bapak Muh Syafaat S.Kuba S.T.,MT Selaku Dekan 1 Fakultas Teknik Universitas Muhammadiyah Makassar
5. Bapak Muhyiddin AM Hayat S.T.,MT Selaku Ketua Prodi Informatika
6. Bapak Fahrin Irhamna Rachman S.Kom.,MT Selaku Dosen Pembimbing 1
7. Bapak Lukman S.Kom,MT Selaku Dosen Pembimbing 2
8. Dosen dan Staf Fakultas Teknik Universitas Muhammadiyah Makassar
9. Paramita yang selalu membantu dan memotivasi penulis ketika

merasa jenuh dan malas pada penyelesaian proposal skripsi ini.

10. Rekan- rekan di Program Studi Informatika Angkatan 2019 kelas C terkhusus Muh. Ilham Kurniawan, Kambran, Dian Olivia yang selalu memotivasi selama penulisan proposal skripsi ini.

11. Semua pihak yang tidak bisa saya sebutkan satu persatu yang telah memberikan dukungan, motivasi, inspirasi dan membantu dalam proses penulisan proposal ini.

Semoga Tuhan yang maha Esa memberikan balasan yang lebih besar kepadabeliau beliau, akhir kata semoga proposal skripsi ini dapat memberikan manfaat bagi pembaca umumnya dan bagi penulis pada khususnya

Makassar, 03 Februari 2024

Fajar Maulana Rahman



DAFTAR ISI

HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
ABSTRAK.....	v
<i>ABSTARCT</i>	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xii
DAFTAR ISTILAH.....	xiii
BAB I PENDAHULUAN	1
A. Latar Belakang	1
B. Rumusan Masalah	2
C. Tujuan Penelitian.....	2
D. Manfaat Penelitian.....	2
E. Ruang Lingkup Penelitian.....	2
F. Sistematika Penelitian	3
BAB II TINJAUAN PUSTAKA.....	4
A. Landasan Teori.....	4
1. Analisis Sentimen.....	4
2. <i>Naive Bayes</i>	6
3. Ulasan <i>Google Maps</i>	7
4. <i>Python</i>	8
5. Pelabelan Kelas Sentimen	9
6. <i>Preprocessing</i> Data	9
7. Pembobotan TDF-IF	10
8. Confusion Matrix	10
B. Penelitian Terkait	12

C. Kerangka Pikir.....	22
BAB III METODE PENELITIAN.....	23
A. Tempat dan Waktu Penelitian	23
B. Alat dan Bahan	23
C. Perancangan Sistem.....	24
D. Teknik Pengujian Sistem.....	26
E. Teknik Analisis Data	28
BAB IV HASIL DAN PEMBAHASAN.....	30
A. Pengumpulan Data	30
B. <i>Labelling</i>	31
C. <i>Preprocessing Data</i>	32
D. <i>Splitting Data</i>	35
E. Pembobotan TDF-IF	36
F. Klasifikasi <i>Naive Bayes</i>	36
G. Evaluasi Model <i>Naive Bayes</i>	38
H. Pengujian <i>Confusion Matrix</i>	39
I. Visualisasi Sentimen	43
BAB IV PENUTUP.....	45
A. Kesimpulan.....	45
B. Saran.....	45
DAFTAR PUSTAKA	46
LAMPIRAN	45

DAFTAR GAMBAR

Gambar 1. <i>Confusion Matrix</i>	11
Gambar 2. Kerangka Pikir	22
Gambar 3. <i>Flowchart</i>	24
Gambar 4. Proses Menentukan <i>Query</i> Saat <i>Crawling</i>	30
Gambar 5. Hasil Data <i>Scraping</i>	31
Gambar 6. Hasil Pelabelan.....	31
Gambar 7. Setelah Proses <i>Cleaning</i>	32
Gambar 8. Sebelum Proses <i>Cleaning</i>	32
Gambar 9. Hasil Tokenisasi.....	33
Gambar 10. Hasil Proses <i>Stopword</i>	34
Gambar 11. Hasil Proses <i>Splitting Data</i>	35
Gambar 12. Proses Pembobotan TDF-IF.....	36
Gambar 13. Proses Pembuatan Model <i>Naive Bayes</i>	37
Gambar 14. Hasil Pengujian Data.....	39
Gambar 15. Hasil Pengujian <i>Confusion Matrix</i>	39
Gambar 16. <i>Confusion Matrix</i>	40
Gambar 17. Distribusi Data hasil prediksi <i>Naive Bayes</i>	43
Gambar 18. <i>Pie Chart</i> Hasil Prediksi <i>Naive Bayes</i>	44

DAFTAR TABEL

Table 1. Spesifikasi Komputer	23
Table 2. Hasil Proses <i>Stemming</i>	34
Table 3. Hasil Pengujian Data	38
Table 4. Hasil <i>Confusion Matrix</i>	40
Table 5. Tabel <i>Confusion Matrix</i>	41
Table 6. Hasil Klasifikasi <i>Labelling</i> Manual Dan <i>Confusion Matrix</i>	42



DAFTAR ISTILAH

- Crawling Data*** Merupakan proses mengumpulkan informasi dari internet dengan bantuan program komputer yang menjelajahi situs web dan mengambil data untuk keperluan analisis atau penyimpanan.
- Flowchart*** Merupakan representasi visual dari langkah-langkah atau proses dalam bentuk diagram, menggunakan simbol-simbol grafis untuk menggambarkan urutan tindakan atau keputusan dalam suatu sistem atau program.
- Library Sklearn*** Merupakan perpustakaan perangkat lunak sumber terbuka untuk pemrosesan data dan pembelajaran mesin dalam bahasa pemrograman Python
- Machine Learning*** Merupakan cabang dari kecerdasan buatan yang memungkinkan komputer untuk belajar dari data dan pengalaman tanpa secara eksplisit diprogram menggunakan algoritma untuk mengidentifikasi pola dalam data dan membuat prediksi atau keputusan tanpa intervensi manusia.
- Naive Bayes*** Merupakan algoritma klasifikasi dalam machine learning yang mengandalkan teorema probabilitas Bayes, dengan asumsi bahwa fitur-fitur yang digunakan untuk klasifikasi adalah independen.

Natural Language Processing Merupakan adalah cabang dari kecerdasan buatan yang berkaitan dengan kemampuan komputer untuk memahami, memproses, dan berinteraksi dengan bahasa manusia secara alami.

Scrapping Merupakan proses ekstraksi data otomatis dari situs web menggunakan program atau bot untuk mendapatkan informasi dari halaman web dan mengubahnya menjadi format yang dapat dipahami atau disimpan.

Training Yaitu proses melatih model atau algoritma dalam machine learning menggunakan data pelatihan untuk menyesuaikan parameter-model sehingga model dapat mengenali pola atau hubungan dalam data dan melakukan tugas tertentu dengan akurasi yang tinggi pada data yang tidak pernah dilihat sebelumnya.

Testing Yaitu tahap dalam machine learning di mana model yang telah dilatih diuji coba dengan menggunakan data yang belum pernah dilihat sebelumnya (testing data) untuk mengevaluasi kinerja dan akurasi prediksi model. Tujuan testing adalah untuk mengukur sejauh mana model mampu menggeneralisasi dan memberikan hasil yang dapat diandalkan pada situasi dunia nyata.

BAB 1

PENDAHULUAN

A. Latar Belakang

Salah satu objek wisata yang mempunyai daya tarik keindahan alam yaitu pantai, merupakan salah satu tempat wisata populer yang digemari semua wisatawan terutama wisata pantai yang ada di Indonesia bagian timur kota Bulukumba, ada beberapa wisata yang terdapat pada kota Bulukumba salah satunya yaitu Pantai Tanjung Bira.

Sebelum para wisatawan berkunjung ke suatu objek wisata, sebaiknya wisatawan mengetahui terlebih dahulu kondisi dari objek wisata tersebut. Hal ini dapat dilihat dari opini wisatawan yang pernah mengunjungi. Tempat wisata yang dijadikan tujuan destinasi para wisatawan sangat berpengaruh terhadap banyak tidaknya wisatawan yang datang ke tempat tersebut. Setiap tempat mempunyai daya tarik sendiri.

Dengan kemajuan teknologi saat ini, pengunjung dapat berkomentar dan mengulas tempat yang mereka kunjungi, termasuk pada objek wisata. Untuk mengetahui ulasan dari wisatawan di *Google Maps*, hal yang harus dilakukan adalah mengumpulkan data ulasan tersebut dengan metode analisis sentimen. Dalam analisis sentimen diperlukan sebuah metode, salah satunya *Naive Bayes*. Dalam metode ini klasifikasi dilakukan dengan menghitung probabilitas.

Dari latar belakang masalah diatas, kami sebagai penulis ingin membuat sebuah aplikasi tugas akhir dengan judul **“ANALISIS SENTIMEN WISATAWAN TERHADAP DESTINASI PANTAI TANJUNG BIRA PADA ULASAN DI *GOOGLE MAPS* MENGGUNAKAN ALGORITMA *NAIVE BAYES*”**. Diharapkan bahwa aplikasi ini dapat memberikan manfaat bagi pengembangan destinasi wisata, serta berkontribusi pada pemahaman lebih lanjut tentang analisis sentimen dalam konteks pariwisata dengan menggunakan algoritma *Naive Bayes*.

B. Rumusan Masalah

Berdasarkan uraian latar belakang di atas, maka penulis dapat merumuskan masalah sebagai berikut:

1. Bagaimana analisis sentimen ulasan wisatawan terhadap destinasi Pantai Tanjung Bira dapat dilakukan menggunakan algoritma *Naive Bayes*?
2. Apakah sentimen umum wisatawan terhadap Pantai Tanjung Bira pada ulasan di *Google Maps* cenderung positif, negatif, atau netral?

C. Tujuan Penelitian

Adapun yang menjadi tujuan dari penelitian ini adalah:

1. Untuk mengetahui apakah analisis sentimen ulasan wisatawan terhadap destinasi Pantai Tanjung Bira dapat dilakukan menggunakan algoritma *Naive Bayes*
2. Untuk mengetahui sentimen umum wisatawan terhadap Pantai Tanjung Bira pada ulasan di *Google Maps* cenderung positif atau negatif

D. Manfaat Penelitian

Manfaat penelitian ini adalah sebagai berikut:

1. Memberikan informasi tentang sentimen wisatawan terhadap destinasi pantai Tanjung Bira. Dari hasil analisis sentimen dapat memberikan gambaran tentang bagaimana wisatawan merespons destinasi pantai Tanjung Bira.
2. Menjadi referensi bagi wisatawan yang ingin berkunjung ke destinasi pantai Tanjung Bira.

E. Ruang Lingkup Penelitian

Ruang lingkup dari penelitian ini adalah:

1. Data ulasan yang digunakan dalam penelitian ini akan diambil dari platform *Google Maps*. Ulasan ini dapat memberikan gambaran tentang pengalaman wisatawan yang telah mengunjungi destinasi pantai Tanjung Bira.
2. Metode yang digunakan dalam analisis sentimen adalah algoritma *Naive*

Bayes

3. Penelitian ini akan menggunakan skala sentimen yang terdiri dari positif, negatif, dan netral.

F. Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini menjelaskan mengenai latar belakang, pokok permasalahan, tujuan penulisan, batasan masalah, metoda penelitian dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini membahas landasan-landasan teori sebagai hasil dari studi pustaka yang berhubungan dalam analisis sentimen wisatawan dengan Algoritma *Naive Bayes* serta buku-buku, jurnal, makalah, internet, atau referensi lainnya dipergunakan sebagai acuan atas sistem pakar yang di buat.

BAB III METODE PENELITIAN

Bab ini berisikan tentang permasalahan yang dihadapi, penyelesaian, masalah, metode pengembangan, *flowchart*, algoritma program, dan struktur datayang digunakan.

BAB II

TINJAUAN PUSTAKA

A. Landasan Teori

Dalam pembuatan aplikasi ini, tidak terlepas dari beberapa teori yang menjadi dasar. Adapun teori yang dijadikan landasan dalam pembuatan aplikasi ini sebagai berikut :

1. Analisis Sentimen

Kata analisis diadaptasi dari bahasa Inggris “*Analysis*” yang secara etimologis dari bahasa Yunani kuno “*Analusis*”, terdiri dari dua suku kata yaitu “*ana*” yang artinya kembali dan “*luein*” yang artinya melepas atau mengurai. Bila digabungkan memiliki arti menguraikan kembali. Secara umum analisis adalah aktivitas yang terdiri dari serangkaian kegiatan seperti: mengurai, membedakan, dan memilah sesuatu untuk dikelompokkan kembali menurut kriteria tertentu kemudian dicari keitannya lalu ditafsirkan maknanya. (Khofifah et al., 2022)

Sentimen adalah pendapat atau pandangan yang didasarkan pada perasaan yang berlebih – lebih terhadap sesuatu *Sentiment analysis* atau *opinion mining* mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan *text mining* yang bertujuan menganalisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenaan dengan suatu topik , produk, layanan, organisasi, individu, ataupun kegiatan tertentu. Analisis sentimen merupakan sebuah penggambaran polaritas pada suatu teks atau kata. Jadi bisa disimpulkan bahwa analisis sentimen adalah pengolahan bahasa dari suatu teks atau kata yang dapat dianalisa dari pendapat, sikap atau penilaian dari seseorang terhadap suatu topik atau pembicaraan.(Khofifah et al., 2022)

Analisis sentimen terhadap produk dan jasa di destinasi wisata perlu dikaji secara komprehensif untuk menghasilkan rekomendasi yang tepat dalam menyesuaikan perubahan preferensi pengunjung serta

meningkatkan citra pariwisata Indonesia. Terdapat kesenjangan dalam studi tentang analisis sentimen wisatawan yang terbatas pada implementasi metode klasifikasi tanpa menghubungkan hasil klasifikasi untuk menghasilkan sejumlah rekomendasi berdasarkan kata populer dalam kelas positif maupun negatif untuk mengoptimalkan manajemen destinasi wisata. (Singgalen, 2022)

Dalam perkembangannya, pendekatan *machine learning* digunakan untuk memprediksi sentimen publik berdasarkan ulasan wisatawan di berbagai media digital, data ulasan pengguna media sosial terhadap atraksi wisata dapat diproses menggunakan algoritma metode klasifikasi untuk memperoleh informasi terkait ulasan yang mengandung makna positif, netral dan negatif lebih jauh. (Singgalen, 2022)

Secara umum, analisis sentimen terbagi menjadi 2 kategori :

- a. *Coarse – grained* sentiment analysis: Proses analisis dan klasifikasi orientasi sebuah dokumen secara keseluruhan. Orientasi ini dibagi menjadi 3 jenis yaitu positif, netral, dan negatif, akan tetapi ada juga yang menjadikan nilai orientasi bersifat kontinyu / tidak diskrit.
- b. *Fined – grained* sentiment analysis: Obyek yang diklasifikasi tidak pada level dokumen melainkan pada level kalimat dalam sebuah dokumen. Analisis sentimen terdiri dari 3 subprocess yaitu
 - 1) *Subjectivity Classification*: menentukan kalimat yang merupakan opini.
 - 2) *Orientation Detection*: Pengklasifikasian opini ke dalam kelas positif, negatif, atau netral.
 - 3) *Opinion Holder and Target Detection*: menentukan bagian yang merupakan *opinion holder* (pemberi opini) dan bagian yang merupakan target. (Aponno, 2022)

Perkembangan studi tentang analisis sentimen di sektor pariwisata berkontribusi penting bagi optimalisasi sistem manajemen destinasi pariwisata di Indonesia. Perkembangan studi tentang analisis sentimen

tidak terbatas pada text dan emojis, melainkan juga gambar yang dilampirkan oleh pengulas berdasarkan konteks dan konten dalam ulasan di website maupun aplikasi digital lainnya, hal ini menunjukkan bahwa kajian tentang analisis sentimen mengalami perkembangan, demikian juga algoritma dan metode yang diadopsi oleh peneliti untuk meningkatkan akurasi dari hasil analisis sentimen. Dengan demikian dapat diketahui bahwa kajian tentang analisis sentimen penting untuk dilakukan untuk berkontribusi secara empiris dan teoretis. (Singgalen, 2022)

2. *Naive Bayes*

Algoritma *Naive Bayes* termasuk pada sepuluh klasifikasi metode data mining terpopuler diantara algoritma-algoritma lainnya. Algoritma *Naive Bayes* pun tergolong memiliki nilai yang berpotensi baik dalam klasifikasi teks dokumen dalam hal akurasi maupun efisiensi komputasi diantara algoritma lain.

Algoritma *Naive Bayes* merupakan salah satu algoritma yang termasuk dalam klasifikasi. Metode ini merupakan metode terbaik dalam pengklasifikasian yang dikemukakan oleh *Thomas Bayes* dengan memprediksi peluang di masa depan berdasarkan data di masa sebelumnya. *Naive Bayes* menerapkan fungsi statistik sederhana berdasarkan teorema bayes dengan asumsi keberadaan dari suatu fitur tertentu terhadap suatu kelas yang tidak berhubungan dengan fitur lainnya. Klasifikasi-klasifikasi Bayes adalah klasifikasi statistik yang dapat memprediksi kelas suatu anggota probabilitas. Untuk klasifikasi Bayes sederhana yang lebih dikenal sebagai *Naive Bayesian Classifier* dapat diasumsikan bahwa efek dari suatu nilai atribut sebuah kelas tidak dipengaruhi atau mempengaruhi nilai dari atribut lainnya. Asumsi ini disebut *class conditional independence* yang diciptakan untuk memudahkan perhitungan, pengertian ini dianggap “*naive*”, dalam bahasa lebih sederhana *naive* itu mengasumsikan bahwa kemunculan suatu term kata dalam suatu kalimat tidak dipengaruhi kata-kata yang lain, sehingga dalam analisis sentimen kata yang muncul memiliki bobot masing-masing

yang kemudian dihitung total bobot seluruhnya apakah kalimat tersebut termasuk positif ataupun negatif. Dalam penelitian ini metode *Naive Bayes* digunakan untuk menghitung nilai *precision*, *Recall*, dan akurasi klasifikasi data. (Aponno, 2022)

Algoritma *naive bayes classifier* merupakan teknik klasifikasi berdasarkan *teorema bayes* dengan asumsi independensi di antara para prediktor. Klasifikasi *Naive bayes* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Tahapan algoritma *naive bayes* yaitu (Khofifah et al., 2022) :

a. Hitung probabilitas bersyarat / *likelihood* :

$$P(x | C) = P(x_1, x_2, \dots, x_n | C)$$

C = class

x = vektor dari nilai atribut

$P(x_i | C)$ = proporsi dokumen dari class C yang mengandung nilai atribut x_i .

b. Hitung probabilitas prior untuk tiap class:

$$P(C) = N_j / N$$

N_j = jumlah dokumen pada suatu class

N = jumlah total dokumen

3. Ulasan *Google Maps*

Google maps adalah aplikasi peta online gratis dari *google*. *Google maps* dapat diakses melalui *browser web* atau melalui perangkat *mobile*, *google maps* bisa diakses di *maps.google.com*. Di *google maps* juga dapat memberikan berbagai informasi bukan hanya sekedar untuk petunjuk jalan tapi bisa memberikan informasi tentang tempat wisata, tempat makan, kondisi jalan yang akan di lalui, jalan alternatif. Di *google maps* juga bisa memberikan ulasan tentang tempat yang telah dikunjungi. (Khofifah et al., 2022)

Proses pengambilan data pada *Google Maps* untuk analisis sentimen menggunakan teknik *web scraping*. *Web scraping* adalah proses yang dilakukan untuk mengambil data khusus dari sebuah halaman

website secara semi-terstruktur. Cara kerja *web scraping* adalah dengan mengakses halaman *web*, memilih elemen data yang ada dalam halaman tersebut, melakukan ekstraksi dan transformasi bila diperlukan, dan terakhir menyimpan data tersebut menjadi dataset terstruktur. (Nurkholis et al., 2023)

Dari google Maps kita mendapatkan data *review* masyarakat analisa komentar masyarakat terhadap tempat wisata terutama tempat wisata pantai. Pada ulasan masyarakat tersebut dapat dilihat *review* penilaian masyarakat apakah positif atau negatif terhadap tempat wisata pantainya. Untuk *me-review* dari masyarakat apakah positif atau negatif dapat dilakukan dengan menggunakan analisis sentimen.

4. *Python*

Python adalah bahasa pemrograman interpretatif dengan banyak aplikasi yang berfokus pada keterbacaan kode. Faktanya, *Python* juga diklaim sebagai bahasa dengan sintaks kode yang jelas, standar fungsional, dan pustaka yang komprehensif selain menggabungkan kemampuan. *Python* memiliki tipe sistem dinamis dan manajemen memori otomatis. *Python* mendukung beberapa paradigma pemrograman, diantaranya pemrograman berorientasi objek, imperatif, fungsional dan prosedural, serta memiliki *library* yang besar dan komprehensif. *Python* memiliki sejumlah karakteristik, salah satunya adalah bahasa pemrograman dinamis dengan manajemen memori otomatis bawaan. *Python* biasanya digunakan sebagai bahasa *scripting*, namun dalam penggunaan sebenarnya, *python* juga digunakan dalam situasi di mana ia biasanya tidak digunakan dalam bahasa pemrograman. *Python* bisa digunakan untuk berbagai kebutuhan untuk pengembangan perangkat lunak dan dapat berjalan pada berbagai platform sistem operasi. Beberapa diantaranya adalah sebagai berikut: *Windows, Linux/Unix, OS/2, Mac OS X, Amiga, Palm, Symbian, Java Virtual Machine*. (Suryadi et al., 2023)

5. Pelabelan Kelas Sentimen

Pada tahap ini pelabelan kelas sentimen atau opini pada data ulasan pengguna dikelompokkan menjadi dua kelas sentimen yaitu kelas positif dan kelas negatif. Pelabelan tersebut dilakukan secara manual untuk memudahkan dalam mengelola data dengan membaca maksud dari kalimat yang ada dalam ulasan tersebut, sehingga dapat diberi label sentimen bahwa ulasan tersebut termasuk kedalam sentimen positif atau negatif. Namun karena data yang perlu diberi label banyak, hal ini membutuhkan waktu yang tidak singkat. Pengelompokkan kelas positif berdasarkan komentar yang mengandung kata bermakna positif, saran, mendukung dan menyatakan persetujuan. Sedangkan untuk kriteria kelas negatif adalah berdasarkan komentar yang mengandung kata bermakna negatif, ejekan, dan keluhan atau ketidakpuasan pengguna terhadap aplikasi. Komentar yang bersifat netral dan juga komentar komentar yang tidak ada hubungannya dengan opini terhadap aplikasi akan dihapus dari dataset. Tahap ini bertujuan untuk memberikan pembelajaran pada model yang akan dibentuk di tahap pelatihan data. (Suryati et al., 2023)

6. *Preprocessing* Data

Preprocessing Data merupakan tahapan dari proses awal terhadap teks untuk mempersiapkan teks menjadi data yang akan diolah lebih lanjut. Suatu teks tidak dapat diproses langsung oleh algoritma pencarian, oleh karena itu dibutuhkan *preprocessing* untuk mengubah teks menjadi data numerik. Dalam tahap *preprocessing* data pada dataset akan diolah agar data dapat menjadi lebih baik atau maksimal. (Atmadja, 2022)

a. *Cleaning*

Pada tahap ini akan menghapus simbol, bilangan angka, dan tanda baca tersebut. Dan proses ini dilakukan menggunakan program, sehingga dilakukan secara otomatis sebelum menyimpan hasil *decode* ke dalam bentuk *xlsx*.

b. Tokenisasi

Tokenisasi adalah tahap memecah kalimat menjadi bagian-bagian dinamakan token. Sebuah token dianggap sebagai satu bentuk sebuah kata, frasa, atau suatu elemen yang berarti.

c. Stopword

Pada tahap ini kata-kata yang kurang bermakna atau tidak memiliki arti dilakukan penghapusan seperti kata: saya, dan, atau.

d. Stemming

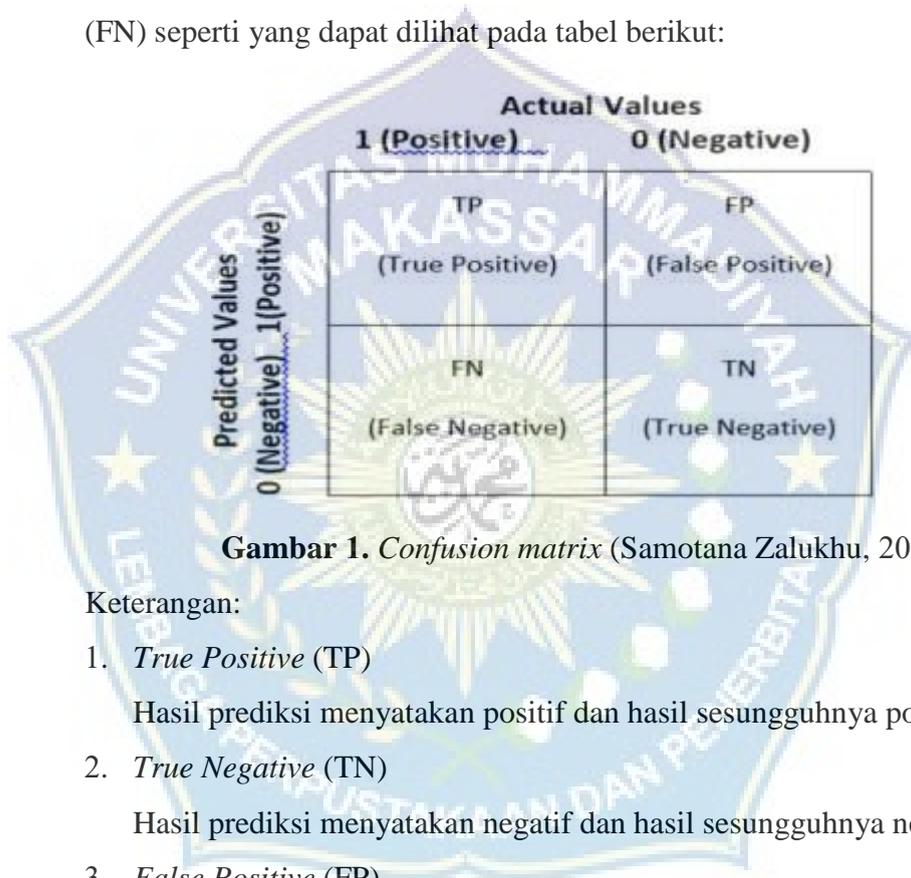
Stemming adalah tahap mengubah sebuah kata ke dalam bentuk kata dasarnya dengan menghapus kata imbuhan di depan maupun imbuhan di belakang kata. (Imron Ali, 2019)

7. Pembobotan TDF-IDF

Metode TF-IDF merupakan suatu cara untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan *inverse* frekuensi dokumen yang mengandung kata tersebut. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Pada rancangan ini TF-IDF berfungsi sebagai metode dalam pembobotan kata sehingga nantinya dapat diolah oleh *naive bayes*. Pada dasarnya, TF-IDF bekerja dalam menentukan frekuensi relatif suatu kata kemudian dibandingkan dengan proporsi kata tersebut pada seluruh dokumen (TF). Intuisinya bahwa kata yang muncul di banyak dokumen bukanlah pembeda yang baik, dan harus diberi bobot kurang dari satu yang terjadi dalam beberapa dokumen (IDF). Pembobotan TF-IDF adalah hasil perkalian dari pembobotan *term frequency* dan *inverse document frequency* dari suatu *term*. (Samotana Zalukhu et al., 2023)

8. Confusion Matrix

Confusion matrix adalah suatu metode yang digunakan untuk melakukan perhitungan tingkat akurasi pada *data mining*. *Confusion matrix* memuat informasi tentang klasifikasi yang diprediksi dengan benar oleh sebuah sistem klasifikasi. Terdapat empat istilah sebagai representasi hasil proses klasifikasi pada *confusion matrix* diantaranya adalah *True Positif* (TP), *True Negatif* (TN), *False Positif* (FP) dan *False Negative* (FN) seperti yang dapat dilihat pada tabel berikut:



		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	TP (True Positive)	FP (False Positive)
	0 (Negative)	FN (False Negative)	TN (True Negative)

Gambar 1. *Confusion matrix* (Samotana Zalukhu, 2023)

Keterangan:

1. *True Positive* (TP)
Hasil prediksi menyatakan positif dan hasil sesungguhnya positif.
2. *True Negative* (TN)
Hasil prediksi menyatakan negatif dan hasil sesungguhnya negatif.
3. *False Positive* (FP)
Hasil prediksi menyatakan positif dan hasil sesungguhnya negatif.
4. *False Negative* (FN)
Hasil prediksi menyatakan negatif dan hasil sesungguhnya positif.
Setelah melalui proses tersebut, hasil yang didapat kemudian akan dihitung nilai *accuracy* dengan rumus sebagai berikut: (Samotana Zalukhu et al., 2023)

Berdasarkan dari data Akurasi, presisi, *recall* dan *f-score* adalah

ketentuan yang digunakan dalam *Confusion matrix*. Ketentuan adalah kinerja model dalam seberapa tepat model dalam mengklasifikasi data secara keseluruhan.

Akurasi adalah seberapa tepat model dalam memprediksi data *Positive* dan *negative* dengan benar dengan keseluruhan jumlah data.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \dots\dots\dots(1)$$

Presisi adalah seberapa tepat model dalam membandingkan data secara keseluruhan dengan jumlah *Positive* yang di klasifikasi secara benar.

$$Precision = \frac{TP}{TP + FP} \dots\dots\dots(2)$$

Recall adalah sistem melakukan perbandingan jumlah data benar *Positive* dengan jumlah data yang sebenarnya *Positive* secara keseluruhan.

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

F-Score didapatkan dari penggabungan hasil presisi dan *recall* dengan menggunakan rata-rata *harmonic* presisi dan *recall*.

$$F-Score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \dots\dots\dots(4)$$

B. Penelitian Terkait

Penelitian terdahulu adalah upaya peneliti untuk mencari perbandingan dan selanjutnya untuk menemukan inspirasi baru untuk penelitian selanjutnya di samping itu kajian terdahulu membantu penelitian dalam memosisikan penelitian serta menunjukkan orsionalitas dari penelitian. Pada bagian ini peneliti mencamtumkan berbagai hasil penelitian terdahulu yang terkait dengan penelitian yang hendak dilakukan, kemudian membuat ringkasannya, baik penelitian yang sudah terpublikasi atau belum terpublikasi (skripsi, tesis, disertasi dan sebagainya).

1. Nova Rosalina Siahaan, Dkk 2023 “Analisis Sentimen Ulasan Aplikasi

Media Sosian *Whatsapp* Menggunakan Metode *Naive Bayes Classifier*”

Dengan mengacu pada penelitian yang dilakukan tentang Analisis Sentimen pada ulasan aplikasi media sosial *Whatsapp*, menggunakan metode klasifikasi *Naive Bayes Classifier*, Pada Pelabelan, untuk mengklasifikasi ulasan ulasan dari suatu produk ke dalam kategori Positif, Negatif, dan Netral. Pada penelitian ini, tim peneliti telah melakukan pelabelan kedalam dataset. Pelabelan pada dataset ini dapat berupa : Rating 3 (lebih besar dari angka -3) adalah Positif. Dan disertai penambahan kolom yaitu 'Label' yang berisi Sentimen Positif, Negatif, dan Netral. Pada *splitting* data, tim peneliti membagi data dan memecah dataset 20% dari keseluruhan data dan dibagi menjadi *training* dan *testing* dengan *testsize* = 0.20 dan *random statenya* = 0. Dan pada Evaluasi Model dapat kita lihat bahwa pada analisis dataset ini cenderung Negatif. (Siahaan et al., n.d.)

2. Yerik Afrianto Singalen 2022. “Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur Di Tripadvisor Menggunakan Algoritma *Naive Bayes Classifier*”

Hasil penelitian ini menunjukkan bahwa proses klasifikasi data teks berdasarkan sentimen negatif dan positif menggunakan algoritma *Naive Bayes Classifier* terhadap data ulasan pengguna *website Tripadvisor* dapat diproses menggunakan aplikasi *Rapidminer*. Hasil pengolahan data ulasan terhadap destinasi wisata Candi Borobudur menunjukkan adanya performa yang baik. Hal tersebut dapat dilihat dari 3850 data ulasan yang telah diproses menghasilkan nilai akurasi sebesar 96,36%, nilai presisi 93,23% dan *recall* sebesar 100%. Adapun, nilai *classification error* sebesar 3,64% dengan *Area Under Curve (AUC)* sebesar 0,714 (71,4%). Disisi lain, hasil klasifikasi sepuluh kata yang terdeteksi paling sering muncul di setiap ulasan pengunjung ialah sebagai berikut: *temple, visit, borobudur, sunrise, place, time, guide, tour, take, people*. Dengan

demikian dapat diketahui bahwa hal-hal yang berkesan bagi pengunjung di Candi Borobudur lebih dominan pada aspek atraksi. Dengan demikian, rekomendasi bagi pengelola destinasi wisata Candi Borobudur agar mempertahankan dan meningkatkan performa layanan produk dan jasa yang berhubungan dengan otentikasi warisan budaya serta mengimplementasikan nilai-nilai sapta pesona untuk meningkatkan kepuasan pengunjung dan citra destinasi wisata Candi Borobudur sebagai destinasi wisata super prioritas di Indonesia. (Singgalen, 2022)

3. Bobby Rizki Atmadja 2022. “Analisis Sentimen Bahasa Indonesia Pada Tempat Wisata Di Kabupaten Sukabumi Dengan *Naive Bayes*”

Setelah melakukan pada penelitian yang telah dilaksanakan, adapun kesimpulan dari penelitian ini adalah pengetahuan dari analisis sentimen wisatawan pada objek wisata di kabupaten Sukabumi pada data ulasan wisatawan dengan kelas positif terdapat presentase paling besar dibandingkan kelas negatif dan netral dengan sebesar 47,81% yang menunjukkan bahwa wisatawan dimayoritaskan bersentimen positif. Performa nilai akurasi pada penggunaan algoritma *naive bayes* dalam analisis sentimen pada ulasan wisatawan objek wisata sukabumi terdapat nilai sebesar 0.78. Pengujian dengan *k-fold cross validation* pada model klasifikasi agar mendapat nilai yang pasti terdapat nilai akurasi sebesar 0.78. Frekuensi kata yang telah diolah dengan *word cloud* dari dokumen dataset ulasan wisatawan terdapat kata-kata yang paling banyak muncul berdasarkan sentimennya. (Atmadja, 2022)

4. Wawah Khofifah Dkk 2022. “Analisis Sentimen Menggunakan *Naive Bayes* Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Kerawang Pada Ulasan *Google Maps*”

Berdasarkan hasil dari analisis sentimen yang telah dilakukan memiliki beberapa kesimpulan diantaranya penulisan ini adalah menganalisa *review* masyarakat pada ulasan google maps untuk tempat wisata pantai diantaranya pantai Cibendo, pantai Pakis, pantai Samudera

Baru, pantai Sedari, dan pantai Tanjung Baru yang ada di Kabupaten Karawang. Hasil perhitungan *naive bayes* yang telah dilakukan bahwa dari 5 Pantai yang sudah dianalisis terdapat 2 Pantai yang memiliki nilai negatif yaitu Pantai Cibendo dengan hasil dari *review* negatif mendapat 0,550 dan Pantai Tanjung Baru mendapat hasil *review* negatif 0,650 dan 3 pantai memiliki nilai positif yaitu pantai Samudera Baru mendapatkan hasil positif 0,850, pantai Sedari 0,700 dan pantai Pakis mendapatkan hasil 0,650. Setelah diuji menggunakan aplikasi *RapidMiner* pantai Cibendo mendapat *accuracy* 55%, pantai Pakis mendapat *accuracy* 65%, Pantai Samudera Baru mendapat *accuracy* 85%, Pantai Sedari mendapat *accuracy* 70%, dan Pantai Tanjung Baru mendapat *accuracy* 65%. Hasil *word cloud* dari pantai Cibendo kata yang sering muncul adalah “kurang”, “tidak”, “terawat”, “sampah”, Pantai Pakis adalah “kurang”, “bagus”, “kotor”, “cukup”, pantai Samudera Baru adalah “bersih”, “bagus”, “murah”, “mudah”, pantai Sedari adalah “bagus”, “cukup”, “jalan”, “keruh” dan pantai Tanjung Baru adalah “jalan”, “kurang”, “menyesal”, “tidak”. (Khofifah et al., 2022)

5. Julius Chrisostomus Aponno 2022. “Penerapan Algoritma *Sentiment Analysis* Dan *Naive Bayes* Terhadap Opini Pengunjung Di Tempat Wisata Pantai Pintu Kota, Kota Ambon”

Berdasarkan hasil pengujian dan analisis yang telah dilakukan pada bahasan sebelumnya, maka dapat diambil kesimpulan sebagai berikut: Analisis Sentimen pada data ulasan *Google Maps* terhadap wisata Pantai Pintu Kota Ambon dengan metode *Naive Bayes* dengan akurasi 90.65% Ini menunjukkan bahwa nilai akurasi dengan menggunakan metode *Naive Bayes* baik ini dikarenakan pada *Naive Bayes*, TP dengan TN memiliki nilai tidak terlampau jauh dengan nilai TP 83.33% dan TN 96.43% Nilai *Recall* pada *Naive Bayes* positif memiliki nilai yang tinggi dikarenakan nilai *True* positif nilainya juga sangat tinggi. Karena nilai TP sangat tinggi maka mengakibatkan *recall*nya juga semakin tinggi. Dan nilai FN nya

rendah. Dari hasil penelitian ini terdapat berbagai komentar yang diberikan pengunjung untuk bagaimana pengelola wisata pantai dan dinas terkait dapat memperbaiki dan meningkatkan infrastruktur dan mutu tempat wisata yang ada di kota Ambon agar menjadi tempat wisata unggulan yang dapat di bangakan. (Aponno, 2022)

6. Anisa Falasari dkk 2022. “Optimize naive bayes classifier using chi square and term frequency inverse document frequency for amazon review sentiment analysis”

Pada penelitian ini, penerapan chi square dan TF-IDF dapat meningkatkan akurasi algoritma klasifikasi *naive bayes* dalam analisis sentimen ulasan Amazon. Pada tahap pra-pemrosesan dan pembobotan fitur menggunakan TF-IDF menghasilkan 1375 fitur. Kemudian dilakukan seleksi fitur menggunakan chi square untuk menentukan tingkat optimasi akurasi pada algoritma klasifikasi *naive bayes*. Tingkat akurasi tertinggi ditunjukkan pada K 450 teratas sebesar 83%. Sedangkan tingkat akurasi klasifikasi *naive bayes* sebelum menerapkan chi square dan TF-IDF adalah sebesar 82%. (Falasari & Muslim, 2022)

7. Ni Luh Putu Merawati, Dkk 2021. “Analisis Sentimen Dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes Dan Latent Dirichlet Allocation”

Penelitian ini berhasil melakukan analisis sentimen menggunakan metode *Naive Bayes* serta pemodelan topik dengan metode *Latent Dirichlet Allocation* (LDA) untuk topik pariwisata Lombok. Berdasarkan hasil analisis sentimen diperoleh 9.496 tweet dengan pembagian 8.996 tweet sentimen positif dan 500 tweet sentimen negatif, sehingga dapat disimpulkan bahwa lebih banyak wisatawan memberikan respon positif daripada negatif terhadap pariwisata Lombok. Hasil pengujian menunjukkan bahwa algoritma *Naive Bayes* mampu mengklasifikasi sentimen wisatawan dengan baik yang ditunjukkan dengan pengukuran hasil kinerja algoritma melalui 4 parameter yaitu nilai akurasi sebesar

92%, presisi 100%, recall 83,84% dan Spesifisitas 100%. Sedangkan pemodelan topik menggunakan algoritma LDA menghasilkan topik terbaik untuk kelas positif pada 8 topik dengan nilai koherensi 0,613 kemudian untuk kelas negatif topik terbaik pada 12 topik dengan nilai koherensi 0,528. Hasil interpretasi topik topik yang sering diperbincangkan oleh masyarakat pada kelas positif yaitu tentang keindahan pantai di Pulau Lombok khususnya gili trawangan dan pantai senggigi. Sedangkan untuk kelas negatif topik yang banyak dibicarakan adalah mengenai sampah. (Putu et al., 2021)

8. Rahmadya Trias Handayanto 2021. “Analisis Sentimen Pada situs Google Review Dengan Naive Bayes Dan Support Vector Machine”

Berdasarkan hasil pembahasan maka dapat disimpulkan bahwa teknik *web scraping* dapat digunakan untuk mendapatkan data yang besar dari halaman *website* secara otomatis sehingga mempermudah proses pengambilan data. Data ulasan mengenai *Summarecon Mall* Bekasi pada situs *Google Review* cenderung memuaskan. Hasil pengujian menunjukkan *Support Vector Machine* layak diimplementasikan mengingat akurasi yang lebih baik dari *Naive Bayes*. Untuk riset berikutnya pada bidang *Natural Language Processing* (NLP), khususnya *sentiment analysis*, perlu menerapkan *Deep Learning* baik untuk *review* lokasi wisata maupun bidang lain seperti kesehatan, pendidikan, pemerintahan, dan lain-lain. (Herlawati et al., 2021)

9. Rifan Ferryawan, Dkk 2019. “Analisis Sentimen Wisata Jawa Tengah Menggunakan Naive Bayes”

Berdasarkan hasil analisis uji validitas data, algoritma yang digunakan sudah cukup memiliki akurasi yang baik dan dapat diterima sebagai suatu metode perhitungan pemeringkatan. Khususnya dalam domain pariwisata, komponen-komponen penilaian yang digunakan sudah dapat memberikan penilaian yang lebih objektif yang dihasilkan berdasarkan konten opini atau ulasan dari media sosial maupun situs

pariwisata lainnya. Dengan mengambil data uji terhadap 10 point of interest (POI) tujuan wisata di Jawa Tengah, hasil ekstraksi dan klasifikasi opini maupun ulasan yang berkaitan dengan destinasi wisata di Jawa Tengah tersebut menghasilkan persepsi masyarakat yang dapat menjadi bahan pertimbangan bagi pemerintah untuk memperbaiki fasilitas dan pelayanan pada obyek wisata. (Ferryawan & Wahyu Wibowo, 2019)

10. Amar P. Natasuwarna 2019. “Analisis Sentimen Keputusan Pemindahan Ibukota Negara Menggunakan Klasifikasi *Naive Bayes*”

Penggunaan metode klasifikasi *Naive Bayes* cukup baik pada analisis sentimen mengenai keputusan pemerintah Indonesia akan pindah ibukota dari Jakarta ke kota lain di luar pulau Jawa. Analisis sentimen pada penelitian ini menggunakan data dari *Twitter* yang berisi opini negatif dan opini positif dari masyarakat. Opini yang disampaikan oleh masyarakat baik opini positif ataupun opini negatif melalui *Twitter* setelah pengumuman pemindahan ibukota memiliki kesamaan dengan pendapat peneliti yang melakukan pengkajian sebelum pengumuman pemindahan ibukota tersebut. Untuk merubah data tidak terstruktur dari *Twitter* menjadi terstruktur dilakukan pre-processing supaya dapat diolah menggunakan algoritma *Naive Bayes*. Ketika dilakukan proses manual pada pre-processing diperlukan pemahaman penggunaan bahasa Indonesia terutama penggunaan kalimat dan kata dasar yang termasuk opini positif atau opini negatif. Nilai akurasi tertinggi diperoleh 92,00% pada rasio 90:10 dan akurasi rata-rata dari lima rasio yang diukur adalah 89,86%. (Natasuwarna STMIK Pontianak Jurusan Sistem Informasi, n.d.)

11. Bong-Hyun back 2019. “*Comparison of Sentiment Analysis from Large Twitter Datasets by Naive Bayes and Natural Language Processing Methods*”

Dalam penelitian ini, kami mengusulkan dan mengimplementasikan sebuah sistem yang dapat mengekstrak informasi sentimen manusia dari sejumlah besar jumlah data besar SNS yang tidak terstruktur

menggunakan algoritma naïve Bayes algoritma *naive bayes* dan NLP. Untuk mengevaluasi kinerja analisis pola dalam sistem yang diusulkan, beberapa percobaan dilakukan. Mengenai percobaan akurasi, akurasi dari analisis pola menunjukkan hasil yang berbeda tergantung pada metode analisis yang digunakan. Keakuratan analisis pola dengan metode pembelajaran mesin adalah 63,50%, dan oleh NLP adalah 72,28% rata-rata. Keakuratan analisis pola oleh pembelajaran mesin menggunakan naïve Bayes lebih rendah dibandingkan dengan NLP. Namun, Namun, berdasarkan percobaan kecepatan pemrosesan data, kecepatan analisis pola dengan *machine learning* menggunakan *naive bayes* adalah sekitar 5,45 kali lebih cepat dibandingkan dengan NLP. Oleh karena itu, metode analisis pola dengan machine learning menggunakan *naive bayes* kurang akurat dibandingkan dengan NLP. Namun, ini mungkin menguntungkan dalam pemrosesan data besar lingkungan di mana sejumlah besar data harus diproses dengan cepat. Kontribusi dari penelitian ini dapat diringkas sebagai berikut. Pertama, metode analisis pola data yang efektif yang dapat mengekstrak informasi sentimen seperti positif, negatif, dan netral dengan menganalisis pola dalam data besar SNS yang tidak terstruktur yang tidak terstruktur telah diusulkan. (Back & Ha, 2019)

12. Normah 2019. "Naive Bayes Algorithm For Sentiment Analysis Windows Phone Store Application Reviews"

Berdasarkan hasil pengujian model yang dilakukan yang dilakukan, diperoleh akurasi sebesar 84,50%, dan dapat dikatakan bahwa *Naive Bayes* merupakan metode yang metode yang baik dalam mengklasifikasikan teks terutama dalam dalam kasus analisis sentimen seperti pada penelitian ini. Hasil dari penelitian ini dapat membantu pihak perusahaan, pengembang, dan pengguna aplikasi dalam menganalisis sentimen produk *Windows Phone Store* pada ulasan ulasan aplikasi *Windows Phone Store* dengan mengklasifikasikan secara otomatis ulasan berbahasa Inggris ke dalam dua kategori yaitu positif dan negatif secara

otomatis dengan waktu yang singkat. (Normah, 2019)

13. Veny Amilia Fitri dkk 2019. “Sentiment analysis of social media twitter with case of anti-LGBT campaign in Indonesia using naive bayes, decision tree, and random forest algorithm”

Berdasarkan data komentar yang diperoleh dari *twitter* tentang kampanye Anti LGBT kecenderungan komentar yang disampaikan berisi komentar yang netral. Pengguna *twitter* bersikap netral terhadap isu tersebut. Dari 936 data pengujian, terdapat 703 komentar dengan sentimen netral, kemudian 102 komentar dengan sentimen positif, dan 4 komentar dengan sentimen negatif. Sentimen negatif. Disini dapat disimpulkan bahwa pengguna media sosial di Indonesia bersikap netral terhadap kampanye anti-LGBT, namun lebih banyak yang mendukung kampanye anti-LGBT daripada yang menolak. Berdasarkan hasil akurasi algoritma *Naive Bayes* sebesar 83.43%, yang lebih tinggi dari akurasi Algoritma *Decision Tree* dan Algoritma *Random Forest* dapat disimpulkan bahwa algoritma *Naive Bayes* sangat baik digunakan untuk melakukan analisis sentimen pada kasus kampanye Anti-LGBT di media sosial *Twitter*. (Fitri et al., 2019)

14. Meylan Wongkar dkk 2019. “Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler : Twitter”

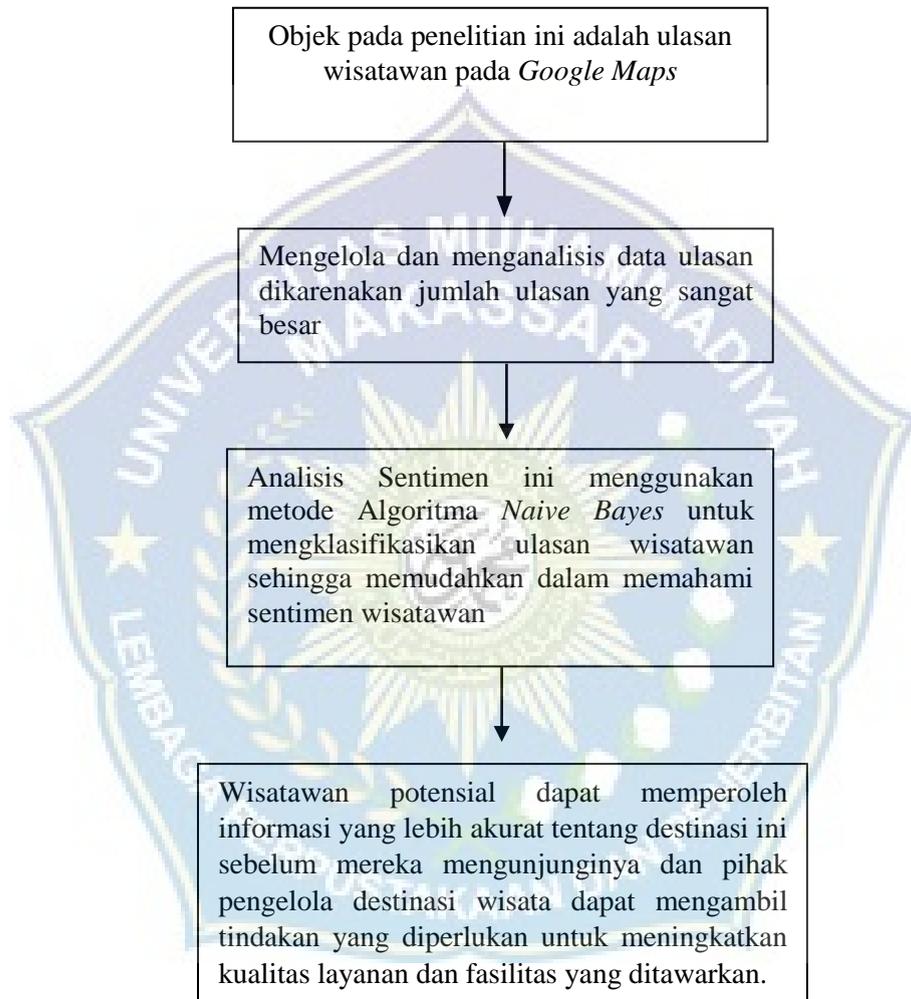
Dalam makalah ini, kami membahas analisis sentimen masyarakat terhadap calon presiden Republik Indonesia periode 2019-2024. Periode 2019-2024, dengan menggunakan data *tweet* yang diperoleh dari media sosial *Twitter*, dengan menggunakan *crawler*. Selain itu, kami melakukan teks dari data yang diperoleh dan menggunakan metode *Naive Bayes* untuk memprediksi kelas. Setelah itu, bandingkan dengan metode lain seperti seperti SVM dan KNN. Kami mengklasifikasikan dengan dua kelas yaitu positif positif dan negatif. Dari hasil percobaan yang kami lakukan, dapat terlihat bahwa metode *Naive Bayes* memiliki tingkat akurasi yang lebih baik (yaitu 80.90%) dibandingkan dengan

menggunakan metode lain, seperti KNN yang hanya memiliki tingkat akurasi sebesar 75,58% dan tingkat akurasi menggunakan tingkat akurasi menggunakan SVM yaitu 63.99%. Untuk penelitian selanjutnya, kami berencana untuk menganalisis sentimen kepuasan masyarakat terhadap kinerja presiden terpilih Republik Indonesia, dengan menggunakan data Indonesia dengan menggunakan data dari media sosial lain, seperti Facebook dan Instagram. (Wongkar & Angdresey, 2019)



C. Kerangka Pikir

Untuk menyelesaikan masalah pada bab sebelumnya maka penulis membuat gambaran singkat sebagai alur penyusunan pada laporan ini dengan kerangka pemikir sebagai berikut:



Gambar 2. Kerangka Pikir

BAB III

METODE PENELITIAN

A. Tempat dan Waktu Penelitian

Pada penelitian ini penulis akan melakukan penelitian di Pantai Tanjung Bira Kecamatan Bonto Bahari, Kabupaten Bulukumba, Sulawesi Selatan. Waktu penelitian ini ≤ 3 bulan, terhitung pada bulan Oktober-Desember 2023.

B. Alat dan Bahan Penelitian

Penelitian ini membutuhkan data ulasan wisatawan di Pantai Tanjung Bira pada *Google Maps* yang bernilai Positif, Negatif dan Netral. Data ulasan yang digunakan diperoleh melalui ulasan pada *Google Maps* dengan teknik *web scraping* pada *Google Maps*.

Alat yang dibutuhkan dalam penelitian ini adalah sebagai berikut:

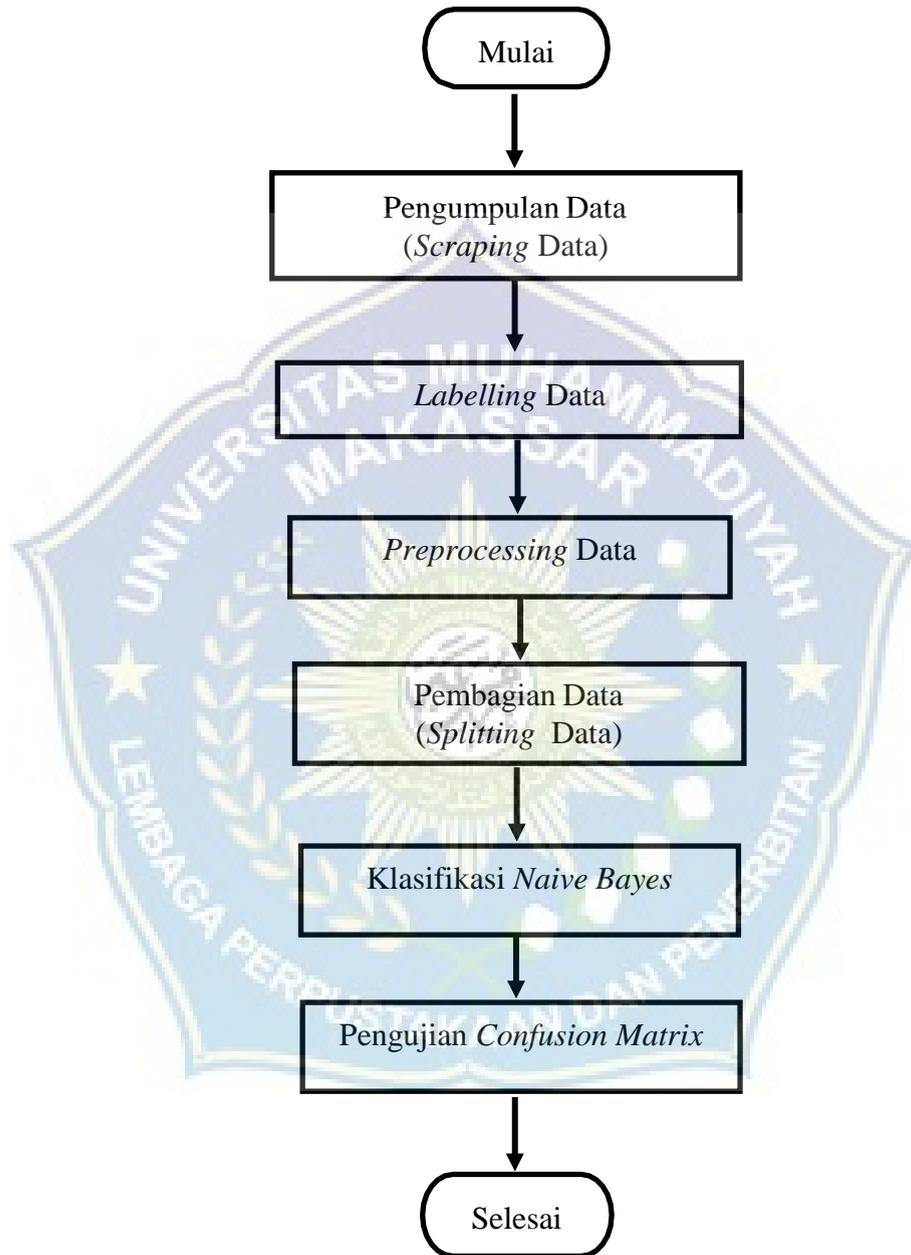
1. Perangkat Lunak
 - *Microsoft Windows 10 64 bit*
 - *Microsoft Office word 10 64 bit*
 - *Python*
 - *Google Maps*
2. Perangkat Keras (*Hardware*)

Table 1. Spesifikasi Komputer

Spesifikasi	Keterangan
<i>Processor</i>	<i>Intel(R) Core(TM) i3-4030U CPU @</i>
1.90GHz RAM	4 GB <i>onboard</i> DDR4 + 8 BG (opsional)
<i>Storage</i>	1 TB
<i>Graphic Card</i>	<i>Intel® UHD Graphic 600</i>
<i>Operating System</i>	<i>Windows 10</i>

C. Perancangan Sistem

Berikut merupakan diagram alir dari tahapan-tahapan penelitian yang akan dilakukan untuk analisis sentimen ulasan wisatawan menggunakan Algoritma *Naive Bayes*:



Gambar 3. *flowchart*

Ada beberapa langkah perancangan sistem analisis sentimen yang dilakukan dalam penelitian ini yaitu :

1. Pengumpulan Data (*Scraping Data*) :

Pengambilan data ulasan dari *Google Maps* tentang destinasi wisata pantai Tanjung Bira. Tahap ini dilakukan dengan menggunakan teknik *web scraping*. *Web scraping* adalah metode yang memungkinkan kita untuk mengambil informasi dari halaman *web* secara otomatis. Dalam konteks analisis sentimen, data yang diambil berupa ulasan dan komentar yang ditinggalkan oleh pengguna di *Google Maps*. Teknik ini memungkinkan untuk mengumpulkan sejumlah besar ulasan dan teks yang diperlukan untuk analisis lebih lanjut.

2. *Labelling data*

Tahap ini dilakukan untuk memberikan label positif atau negatif pada setiap data berdasarkan sentimen yang terkandung di dalamnya.

3. *Preprocessing Data*

Tahap ini dilakukan untuk melakukan *preprocessing* data, seperti menghilangkan tanda baca, mengubah huruf menjadi *lowercase*, dan menghapus kata-kata yang tidak penting. Ada empat tahapan *text preprocessing* yang meliputi *cleaning*, *tokenizing*, *stopword*, dan *stemming*.

- a. Tahapan *Cleaning* adalah tahapan untuk penghapusan karakter seperti tanda baca, menghapus emoticon, menghilangkan simbol yang tidak relevan serta mengkonversi keseluruhan teks dalam dokumen menjadi seluruhnya huruf kecil (*lowercase*).
- b. Tahapan *Tokenizing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya.
- c. Tahapan *Stopword* adalah tahap mengambil kata-kata penting dari hasil token dengan menghapus sekelompok kata umum yang seringkali dianggap tidak memberikan kontribusi signifikan terhadap makna utama dalam teks dan cenderung muncul dengan frekuensi tinggi.
- d. Teknik *Stemming* diperlukan selain untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen, juga untuk melakukan

pengelompokan kata-kata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk atau form yang berbeda karena mendapatkan imbuhan yang berbeda.

4. Pembagian Data (Splitting Data)

Data yang telah di *preprocessing* kemudian dibagi menjadi dua bagian utama yaitu data *training* dan data *testing*. Data *training* digunakan untuk melatih model sentimen, sementara data *testing* digunakan untuk menguji sejauh mana model tersebut akurat dalam mengklasifikasikan sentimen.

5. Klasifikasi *Naive Bayes*

Tahapan ini melibatkan penggunaan algoritma klasifikasi *Naive Bayes* untuk membangun model analisis sentimen. *Naive Bayes* adalah salah satu algoritma klasifikasi yang sering digunakan dalam analisis sentimen dan tugas klasifikasi teks.

6. Pengujian *Confusion Matrix*

Tahapan ini merupakan langkah untuk mengukur kinerja model analisis sentimen yang telah dibangun, dalam hal ini menggunakan *confusion matrix*. *Confusion matrix* adalah alat evaluasi yang memberikan wawasan mendalam tentang sejauh mana model mampu mengklasifikasikan sentimen dengan benar. Hasil dari pengujian ini akan memberikan pemahaman tentang kinerja model, dan apakah model dapat digunakan dalam analisis sentimen dengan hasil yang dapat diandalkan.

D. Teknik Pengujian Sistem

Pada penelitian ini, teknik pengujian yang akan digunakan adalah *Confusion Matrix*. *Confusion Matrix* akan digunakan untuk mengevaluasi kinerja algoritma *Naive Bayes* dalam memprediksi sentimen pada data *testing*. *Confusion matrix* akan membantu dalam menghitung akurasi, presisi, recall, dan *F1-score* dari algoritma *Naive Bayes*. Pada tahap pengujian sistem, data yang telah dilabeli akan dibagi menjadi data *training* dan data *testing*. Data *training* digunakan untuk melatih algoritma *Naive Bayes*, sedangkan data *testing* digunakan untuk mengevaluasi kinerja

algoritma. Algoritma *Naive Bayes* akan dilatih pada data training. Algoritma akan belajar distribusi probabilitas kata-kata pada ulasan wisatawan dan label sentimen yang sesuai. Kinerja algoritma *Naive Bayes* akan diuji pada *data testing*. Akurasi algoritma dapat diukur dengan menggunakan teknik pengujian sistem seperti *confusion matrix*. *Confusion matrix* akan menunjukkan jumlah data yang diklasifikasikan dengan benar dan salah berdasarkan label sebenarnya dan label prediksi. *Confusion matrix* terdiri dari empat bagian yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Akurasi, presisi, recall, dan *F1-score* dari algoritma *Naive Bayes* akan dihitung menggunakan *confusion matrix*.

Berdasarkan dari data Akurasi, presisi, *recall* dan *f-score* adalah Ketentuan yang digunakan dalam *Confusion matrix*. Ketentuan adalah kinerja model dalam seberapa tepat model dalam mengklasifikasi data secara keseluruhan.

Akurasi adalah seberapa tepat model dalam memprediksi data *Positive* dan *negative* dengan benar dengan keseluruhan jumlah data.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \dots\dots\dots(1)$$

Presisi adalah seberapa tepat model dalam membandingkan data secara keseluruhan dengan jumlah *Positive* yang di klasifikasi secara benar

$$Precision = \frac{TP}{TP + FP} \dots\dots\dots(2)$$

Recall adalah sistem melakukan perbandingan jumlah data benar *Positive* dengan jumlah data yang sebenarnya *Positive* secara keseluruhan.

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

F-Score didapatkan dari penggabungan hasil presisi dan recall dengan menggunakan rata-rata *harmonic presisi* dan *recall*.

$$F-Score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \dots\dots\dots(4)$$

E. Teknik Analisis Data

Teknik analisis data pada proposal adalah suatu proses pengolahan data ulasan wisatawan di *Google Maps* mengenai destinasi Pantai Tanjung Bira menjadi informasi baru. Proses ini dilakukan dengan menggunakan teknik analisis sentimen menggunakan algoritma *Naive Bayes*. Berikut adalah alur teknik analisis data yang digunakan pada proposal ini:

1. Pengumpulan data merupakan tahapan awal dalam analisis sentimen. Dalam konteks proposal ini, pengumpulan data dilakukan dengan teknik *web scraping* atau penggunaan *API Google Maps* untuk mengambil ulasan dari platform tersebut. Data yang digunakan adalah ulasan wisatawan mengenai destinasi Pantai Tanjung Bira.
2. Proses selanjutnya adalah memberikan label sentimen pada setiap ulasan. Proses ini melibatkan pemberian label atau kategori sentimen pada setiap ulasan atau teks yang akan dianalisis. Sentimen positif mengindikasikan ulasan yang memberikan pendapat positif tentang destinasi, sentimen negatif mencerminkan ulasan kritis atau negatif.
3. *Preprocessing* data adalah tahap kritis dalam analisis sentimen yang melibatkan pembersihan dan persiapan data sebelum analisis dilakukan. Tujuannya adalah untuk memastikan bahwa data yang digunakan dalam analisis adalah data yang bersih, konsisten, dan siap untuk dievaluasi. Ada empat tahapan *preprocessing* data yang meliputi *case folding*, *tokenizing*, *filtering*, dan *stemming*.
4. Setelah *preprocessing* data selesai, *dataset* akan dibagi menjadi dua bagian utama yaitu data pelatihan (*training data*) dan data pengujian (*testing data*). Data ulasan wisatawan dibagi menjadi dua bagian yaitu 80% data *training* dengan jumlah 1520 ulasan dan 20% data *testing* dengan jumlah 380 ulasan. Data pelatihan digunakan untuk melatih model analisis sentimen, memungkinkan model untuk belajar pola-pola sentimen dari ulasan yang ada. Data pengujian digunakan untuk menguji kinerja model yang telah dilatih pada data yang belum pernah dilihat sebelumnya, memberikan pemahaman tentang seberapa baik model dapat memprediksi

sentimen ulasan dalam situasi dunia nyata.

5. Algoritma *Naive Bayes* dilatih pada data *training*. Algoritma akan belajar distribusi probabilitas kata-kata pada ulasan wisatawan dan label sentimen yang sesuai. Dalam klasifikasi *Naive Bayes*, data pelatihan yang telah dilabeli dengan sentimen digunakan untuk melatih model. Model ini belajar mengenali pola sentimen dengan menghitung probabilitas kemunculan kata-kata dalam setiap kategori sentimen.
6. Setelah model *Naive Bayes* dilatih, langkah selanjutnya adalah menguji kinerjanya menggunakan data pengujian. *Confusion matrix* digunakan sebagai teknik pengujian sistem untuk mengevaluasi kinerja algoritma *Naive Bayes* dalam memprediksi sentimen pada data *testing*. Akurasi, presisi, *recall*, dan *F1-score* dari algoritma *Naive Bayes* akan dihitung menggunakan *confusion matrix*.
7. Hasil dari teknik pengujian sistem akan digunakan untuk mengevaluasi kinerja algoritma *Naive Bayes* dan menentukan apakah algoritma tersebut cocok digunakan untuk analisis sentimen wisatawan terhadap destinasi Pantai Tanjung Bira pada ulasan di *Google Maps*.

BAB IV

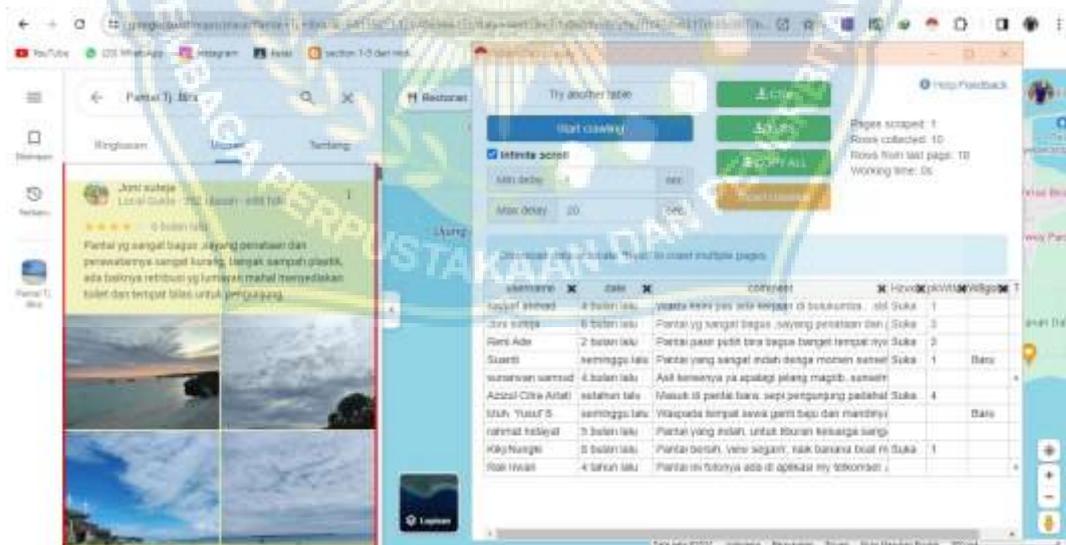
HASIL PENELITIAN DAN PEMBAHASAN

Pada penelitian analisis sentimen mengenai ulasan wisatawan terhadap destinasi pantai tanjung bira pada *google maps* menggunakan metode *naive bayes*. Tujuannya untuk mengetahui apakah ulasan wisatawan terhadap destinasi pantai tanjung bira pada *google maps* tergolong positif, negatif atau netral.

Untuk penelitian ini, data di dapatkan dengan *crawling* atau *scrapping* data ulasan wisatawan terhadap destinasi pantai tanjung bira pada *google maps*.

A. Pengumpulan Data

Pengumpulan data dari ulasan wisatawan terhadap destinasi pantai tanjung bira pada *google maps* menggunakan teknik *web scrapping*. Proses pengambilan data menggunakan *scrapper data* dimulai dengan mengaktifkan ekstensi *Instant Data Scrapper* di browser *Google Chrome*, kemudian menentukan kueri untuk mengambil data pada bagian yang ditandai saat kueri ditentukan.



Gambar 4. Proses Menentukan Query Saat Crawling.

Setelah memutuskan data mana yang akan diambil, selanjutnya untuk memulai proses *crawling* yaitu menekan tombol *start crawling* pada *Instant Data*

Scraper untuk memulai mengambil data. Ketika proses *crawling data* selesai, maka semua data komentar wisatawan akan terkumpul dan kemudian data disimpan dalam file dengan ekstensi *excel*.

	user	date	comment
1	Andi Muldatunnisa	setahun lalu	pantai tercantik yang pernah aku lat
2	Aldal Reckers	8 bulan lalu	pantainya indah
3	Rosadhy Wahyudhy	3 tahun lalu	Sangat menyukai pantai, apalagi pasinya putih dan bersih
4	Juani Agung	3 tahun lalu	Tempatnya bagus
5	Toni Wardono	4 tahun lalu	Sangat indah pantai dengan pasir putihnya semoga destinasi wisata ini terjaga kebersihannya...
6	atang handy	5 tahun lalu	Kalo disini enakya bermalam.. karena mandi dipantai lebih asyik di pagi hari. Jangan lupa bawa peralatan snorkling..
7	Arthy Jupie	3 tahun lalu	Pemandangan sangat indah... keren buat liburan sama teman teman dan keluarga
8	Yuni Antoko	5 tahun lalu	Pasir putihnya lembut... ada permainan berbayar bagi penggemar kegiatan pantai...
9	Nur Alam	9 bulan lalu	Sering dibicarakan orang inyaalah bakal kesini
10	Anneke Mutiarsari	3 tahun lalu	Pantai pasir putih, banyak kegiatan yang bisa di lakukan di pantai bisa bersama keluarga.
11	Early Carolina	2 tahun lalu	Tempat terkeren... pasir putih laut biru banyak tempat wisata baru pokoknya paling favorit
12	Diman Sinergy Channel	10 bulan lalu	Pemandangan yang bagus
13	Darsty Eskobra	4 tahun lalu	Pasir pantainya bersih dan halus sangat memanjakan mata
14	Ganendra Cahyo	5 tahun lalu	Pantainya biru dan jernih dengan pasir putih yang memukau
15	Syarifa Agriah Assegaf	3 tahun lalu	Pasir putihnya cantik pemandangan lautnya biru
16	Taufik Lubis	4 tahun lalu	Keseluruhan masih alamii keren
17	resmi idang tabe	3 tahun lalu	Wisata pantai wisata bisa yang keren
18	Cici Karlina	4 tahun lalu	Tempat-citra favorit wisata pantai terbaik terbayakan dengan pemandangannya

Gambar 5. Hasil Data Scraping

B. Labelling

Labelling dalam konteks analisis sentimen adalah proses memberikan klasifikasi tertentu pada teks atau data berdasarkan sentimen yang terkandung di dalamnya. Ulasan atau teks diberi label "positif" jika menyuarakan kepuasan, "negatif" jika mengekspresikan ketidakpuasan, dan "netral" jika tidak ada sentimen kuat. *Labelling* dapat dilakukan secara manual oleh manusia atau otomatis menggunakan metode komputasional seperti model *machine learning*. Akurasi *labelling* penting dalam membangun model analisis sentimen yang efektif.

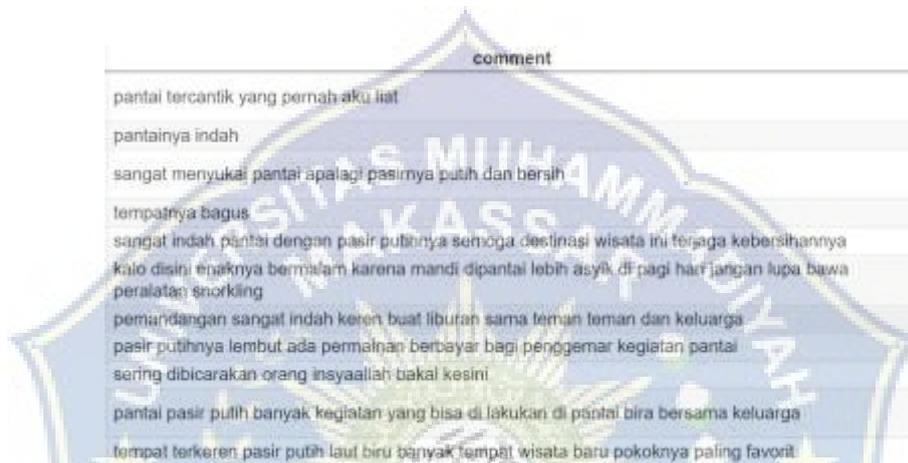
index	user	date	comment	label
0	Andi Muldatunnisa	setahun lalu	pantai tercantik yang pernah aku lat	Positif
1	Aldal Reckers	8 bulan lalu	pantainya indah	Positif
2	Rosadhy Wahyudhy	3 tahun lalu	Sangat menyukai pantai, apalagi pasinya putih dan bersih	Positif
3	Juani Agung	3 tahun lalu	Tempatnya bagus	Positif
4	Toni Wardono	4 tahun lalu	Sangat indah pantai dengan pasir putihnya semoga destinasi wisata ini terjaga kebersihannya...	Positif
5	atang handy	5 tahun lalu	Kalo disini enakya bermalam.. karena mandi dipantai lebih asyik di pagi hari. Jangan lupa bawa peralatan snorkling..	Positif
6	Arthy Jupie	3 tahun lalu	Pemandangan sangat indah... keren buat liburan sama teman teman dan keluarga	Positif
7	Yuni Antoko	5 tahun lalu	Pasir putihnya lembut... ada permainan berbayar bagi penggemar kegiatan pantai...	Positif
8	Nur Alam	9 bulan lalu	Sering dibicarakan orang inyaalah bakal kesini	Positif
9	Anneke Mutiarsari	3 tahun lalu	Pantai pasir putih, banyak kegiatan yang bisa di lakukan di pantai bisa bersama keluarga.	Positif
10	Early Carolina	2 tahun lalu	Tempat terkeren... pasir putih laut biru banyak tempat wisata baru pokoknya paling favorit	Positif

Gambar 6. Hasil Pelabelan

C. Preprocessing Data

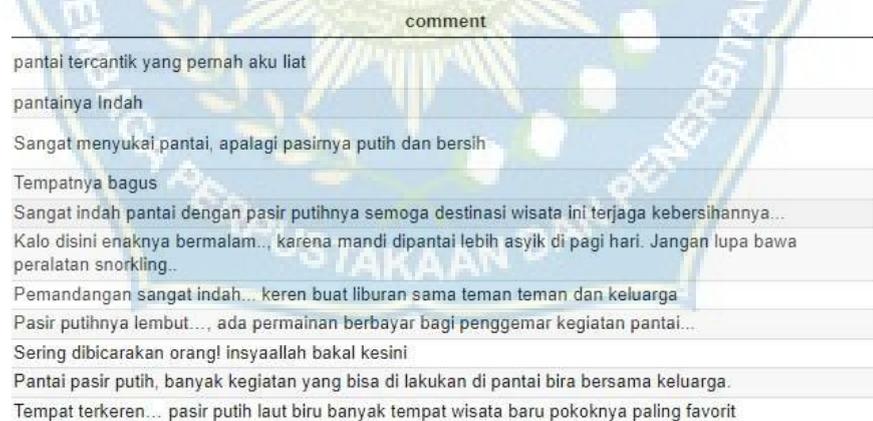
1. Cleaning data

Di tahapan *cleaning data* yang sudah di *scraping* akan dilakukan penghapusan karakter seperti tanda baca, menghapus *emoticon*, mengubah semua huruf ke huruf kecil, serta menghilangkan simbol yang tidak releval. Gambar berikut merupakan hasil dari tahap *cleaning* :



comment
pantai tercantik yang pernah aku liat
pantainya indah
sangat menyukai pantai apalagi pasirmya putih dan bersih
tempatny bagus
sangat indah pantai dengan pasir putihnya semoga destinasi wisata ini terjaga kebersihannya
Kalo disini enakny bermalam karena mandi dipantai lebih asyik di pagi hari jangan lupa bawa peralatan snorkling
pemandangan sangat indah keren buat liburan sama teman teman dan keluarga
pasir putihnya lembut ada permainan berbayar bagi penggemar kegiatan pantai
sering dibicarakan orang insyaallah bakal kesini
pantai pasir putih banyak kegiatan yang bisa di lakukan di pantai bira bersama keluarga
tempat terkeren pasir putih laut biru banyak tempat wisata baru pokoknya paling favorit

Gambar 7. Setelah Proses *Cleaning*



comment
pantai tercantik yang pernah aku liat
pantainya Indah
Sangat menyukai pantai, apalagi pasirmya putih dan bersih
Tempatnya bagus
Sangat indah pantai dengan pasir putihnya semoga destinasi wisata ini terjaga kebersihannya...
Kalo disini enakny bermalam... karena mandi dipantai lebih asyik di pagi hari. Jangan lupa bawa peralatan snorkling..
Pemandangan sangat indah... keren buat liburan sama teman teman dan keluarga
Pasir putihnya lembut..., ada permainan berbayar bagi penggemar kegiatan pantai...
Sering dibicarakan orang! insyaallah bakal kesini
Pantai pasir putih, banyak kegiatan yang bisa di lakukan di pantai bira bersama keluarga.
Tempat terkeren... pasir putih laut biru banyak tempat wisata baru pokoknya paling favorit

Gambar 8. Sebelum Proses *Cleaning*

2. Tokenizing

Tokenizing adalah proses memecah sebuah teks menjadi unit-unit yang lebih kecil, yang disebut dengan token. Token bisa berupa kata, frasa, atau karakter tergantung pada kebutuhan analisis. Tujuan utama dari tokenisasi adalah untuk

mempersiapkan teks sehingga dapat diolah lebih lanjut dalam analisis teks atau pemrosesan bahasa alami. Dalam banyak kasus, tokenisasi dilakukan dengan memisahkan kalimat menjadi kata-kata, memisahkan kata-kata tersebut dari tanda baca, atau bahkan memisahkan teks menjadi kalimat-kalimat. Gambar berikut merupakan hasil dari tahap *tokenizing* :

tokenisasi
pantai,tercantik,yang,pehah,aku,lihat
pantai,nya,indah
sangat,menyukai,pantai,apalagi,pasir,nya,putih,dan,bersih
tempat,nya,bagus
sangat,indah,pantai,dengan,pasir,putih,nya,samoga,destinasi,wisata,ini,terjaga,kebersihannya
kalo,dishi,enaknya,bermalam,karena,mandi,dipantai,lebih,asyik,di,pagi,hari,jangan,lupa,bawa,peralatan,snorkling
pemandangan,sangat,indah,keren,buat,liburan,sama,teman,teman,dan,keluarga
pasir,putih,nya,lembut,ada,permainan,berbayar,bagi,peggemar,kegiatan,pantai
sering,dibicarakan,orang,insyaallah,baka,kesini
pantai,pasir,putih,banyak,kegiatan,yang,bisa,di,lakukan,di,pantai,bira,bersama,keluarga
tempat,terkeren,pasir,putih,laut,biru,banyak,tempat,wisata,baru,pokoknya,paling,favorit

Gambar 9. Hasil *Tokenizing*

3. *Stopword Removal*

Stopword Removal adalah sekelompok kata umum yang seringkali dianggap tidak memberikan kontribusi signifikan terhadap makna utama dalam teks dan cenderung muncul dengan frekuensi tinggi. Tahapan Penghapusan *stopwords* merupakan langkah umum dalam pra-pemrosesan teks dalam pemrosesan bahasa alami (NLP), bertujuan untuk meningkatkan efisiensi pemrosesan dan meningkatkan fokus pada kata-kata yang dianggap lebih informatif. Daftar *stopwords* bervariasi sesuai dengan bahasa, dan penghapusan mereka dapat bermanfaat dalam analisis teks besar-besaran, seperti dalam klasifikasi teks atau analisis sentimen. Meskipun penghapusan *stopwords* umumnya meningkatkan kualitas analisis, keberadaan *stopwords* tertentu dapat memiliki arti penting dalam beberapa konteks dan perlu dipertimbangkan secara kontekstual. Gambar berikut merupakan hasil dari tahap *stopword* :

stopword
pantai tercantik liat
pantainya indah
menyukai pantai pasirnya putih bersih
tempatya bagus
indah pantai pasir putihnya semoga destinasi wisata terjaga kebersihannya
kalo enakya bermalam mandi dipantai asyik pagi lupa bawa peralatan snorkling
pemandangan indah keren liburan teman teman keluarga
pasir putihnya lembut permainan berbayar penggemar kegiatan pantai
dibicarakan orang insyaallah kesini
pantai pasir putih kegiatan lakukan pantai bira keluarga

Gambar 10. Hasil Proses *Stopword*

4. *Stemming*

Stemming adalah proses pemangkasan atau penghapusan awalan atau akhiran kata dengan tujuan untuk menghasilkan bentuk dasar kata, yang disebut sebagai "akar kata" atau "*stems*". Tujuan utama dari *stemming* adalah untuk mengurangi kata-kata ke bentuk dasarnya sehingga kata-kata yang memiliki akar kata yang sama akan direpresentasikan secara seragam.

Tabel 2. Hasil Proses *Stemming*

Sebelum	Sesudah
pantai tercantik liat	pantai cantik liat
pantainya indah	pantai indah
menyukai pantai pasirnya putih bersih	suka pantai pasir putih bersih
tempatya bagus	tempat bagus
indah pantai pasir putihnya semoga destinasi wisata terjaga kebersihannya	indah pantai pasir putih moga destinasi wisata jaga bersih
kalo enakya bermalam mandi dipantai asyik pagi lupa bawa peralatan snorkling	kalo enak malam mandi pantai asyik pagi lupa bawa alat snorkling

D. Splitting Data

Setelah dilakukan proses *preprocessing* perlu dilakukan proses pemilihan dan pembagian data yang digunakan untuk proses *training* dan *testing* algoritma *machine learning*. Pembagian data, atau *splitting data*, memainkan peran kunci dalam mengembangkan model *machine learning* untuk memastikan generalisasi yang baik pada data yang belum pernah dilihat. Data latih digunakan untuk melatih model, sementara data uji memberikan gambaran objektif tentang sejauh mana model dapat menggeneralisasi pada situasi dunia nyata. Proporsi pembagian yang tepat, seperti 80% data latih dan 20% data uji, sering digunakan untuk menjaga keseimbangan antara melatih model dengan data yang cukup dan menguji model pada variasi yang mencakup sebagian besar dataset. Pada penelitian ini, data dibagi menjadi dua yaitu data latih dan data uji. Jadi data sebanyak 720 data latih dan 180 data uji.

```
Ukuran Set Pelatihan: 720 data
Ukuran Set Pengujian: 180 data

Sampel dari Set Pengujian:
['objek wisata indah cocok libur keluarga' 'sih baguz' 'pasir putih'
 'susah umat muslim sholat mushollah' 'indah pasir putih']

Jumlah Label pada Set Pelatihan:
(array(['Negatif', 'Netral', 'Positif'], dtype=object), array([234, 204, 282]))

Jumlah Label pada Set Pengujian:
(array(['Negatif', 'Netral', 'Positif'], dtype=object), array([66, 32, 82]))
```

Gambar 11. Hasil Proses *Splitting Data*

E. Pembobotan TF-IDF

Dalam pembobotan *term*, dijalankan suatu proses untuk memberikan nilai atau bobot pada setiap dokumen kata komentar dan pada tahap ini digunakan metode TF-IDF. Untuk tujuan memberikan nilai pada term yang nilainya akan digunakan pada saat entri dalam proses klasifikasi. Pada proses implementasi pembobotan dengan menggunakan TF-IDF terdapat menggunakan *library* dari *scikit-learn*.

```
from sklearn.feature_extraction.text import TfidfVectorizer
tfidf_vectorizer = TfidfVectorizer()
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)
X_test_tfidf = tfidf_vectorizer.transform(X_test)

print("Ukuran matriks fitur Tf-idf Set Pelatihan:", X_train_tfidf.shape)
print("Ukuran matriks fitur Tf-idf Set Pengujian:", X_test_tfidf.shape)
```

Ukuran matriks fitur Tf-idf Set Pelatihan: (720, 968)
Ukuran matriks fitur Tf-idf Set Pengujian: (180, 968)

Gambar 12. Proses Pembobotan TDF-IF

F. Klasifikasi *Naive Bayes*

Klasifikasi merupakan teknik komputasi untuk mengelompokkan data berdasarkan keterikatan data. Pengelompokkan data dimana data tersebut mempunyai kelas label atau target. Dalam penelitian ini menggunakan metode *naive bayes* untuk mengimport *machine learning naive bayes* menggunakan *library sklearn*. Metode ini hanya memerlukan data yang kecil untuk menghasilkan akurasi. Cara kerja *naive bayes* yaitu dengan menghitung jumlah label, menghitung jumlah kasus perkelas, kalikan semua variabel kelas dan bandingkan hasil perkelas. Metode ini untuk menghitung berapa nilai *accuracy*, *precision*, *recall*, *f1_score*. Pada tahap klasifikasi penelitian ini, data di *split* menjadi 20% untuk data *test* dan 80% untuk data *train*. Setelah sebelumnya telah dilakukan pembobotan kata dengan menggunakan TF-IDF, lalu untuk proses klasifikasi menggunakan *library Multinomial Naive Bayes*. Setelah melakukan semua proses pada tahap klasifikasi dengan menghitung probabilitas antar kelas sentimen pada kalimatnya, maka selanjutnya dapat mendapatkan hasil prediksi. Maka selanjutnya baru dapat dilakukan tahap untuk menghitung performa dari algoritma klasifikasi.

```
[40] nb_classifier = MultinomialNB()
      nb_classifier.fit(X_train_tfidf, y_train)
```

▾ MultinomialNB
MultinomialNB()

Gambar 13. Pembuatan Model *Naive Bayes*

G.Evaluasi Model Naive Bayes

Pada tahap evaluasi model ini bertujuan untuk mengetahui nilai performa algoritma yang digunakan pada penelitian ini. Hasil pengujian data ulasan wisatawan pantai tanjung bira dengan menggunakan *naive bayes* mendapatkan nilai *accuracy* sebesar 88%.

Tabel 3. Hasil Pengujian Data

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
Positif	0.88	0.94	0.91	82
Negatif	0.89	0.94	0.91	66
Netral	0.91	0.62	0.74	32
Macro Average	0.89	0.83	0.85	180
Weighted Average	0.88	0.88	0.88	180

Dengan demikian menunjukkan bahwa analisis sentimen ulasan wisatawan pantai tanjung bira pada *google maps* dengan *metode naive bayes* mempunyai tingkat kinerja yang cukup bagus karena memiliki nilai *accuracy* diatas 50%. *Macro average* adalah salah satu metode atau cara yang digunakan untuk menghitung rata-rata dari metrik evaluasi pada setiap kelas. *Weighted average* adalah metode penghitungan rata-rata yang memberikan bobot berbeda pada setiap

nilai yang akan dihitung berdasarkan faktor tertentu.

```
Classification Report:
              precision    recall  f1-score   support

   Negatif      0.89      0.94      0.91         66
    Netral      0.91      0.62      0.74         32
    Positif      0.88      0.94      0.91         82

 accuracy              0.88         180
 macro avg              0.89      0.83      0.85         180
 weighted avg           0.88      0.88      0.88         180
```

Gambar 14. Hasil Pengujian Data

H. Pengujian Confusion Matriks

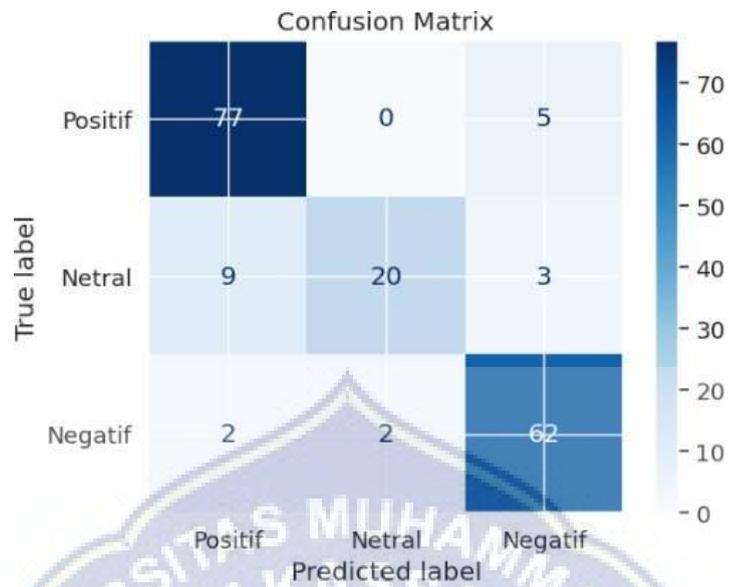
Untuk menghitung dan mengetahui performa dari algoritma klasifikasi yang digunakan yaitu algoritma *Naive Bayes*, maka perlu dilakukan pengujian model klasifikasi. Hasil performa dari algoritma ditampilkan dalam bentuk nilai *accuracy* dan bentuk tabel *confusion matrix*, salah satu metode yang dapat digunakan untuk mengukur performa suatu metode klasifikasi. *Confusion matrix* berisi informasi yang membandingkan hasil dari klasifikasi yang dilakukan oleh sistem dengan *default* hasil klasifikasi. Berikut hasil nilai *accuracy* dan tabel *confusion matrix* terdapat nilai sebagai berikut :

Evaluasi Model Naive Bayes pada Data Pengujian:
Accuracy: 0.88

Confusion Matrix:

```
[[62  2  2]
 [ 3 20  9]
 [ 5  0 77]]
```

Gambar 15. Hasil Pengujian *Confusion Matrix*



Gambar 16. Confusion Matrix

Tabel 4. Hasil Confusion Matrix

		True Label		
		Positif	Netral	Negatif
Predicted Label	Class			
	Positif	77	9	2
	Netral	0	20	2
	Negatif	5	3	62

Tabel 5. Tabel Confusion Matrix

		True Label		
		Positif	Netral	Negatif
Predected Label	Class			
	Positif	TP	FP	FP
	Netral	FN	TP	FN
	Negatif	FN	FN	TN

Keterangan :

<i>True Positive (TP)</i>	97
<i>False Negative (FN)</i>	10
<i>False positive (FP)</i>	11
<i>True Negative (TN)</i>	62

Dalam analisis sentimen, *confusion matrix* digunakan untuk menghitung *accuracy*, *presisi*, *recall*, dan *f1-score* dari model.

Accuracy adalah proporsi data yang diklasifikasikan dengan benar dari keseluruhan data. Berikut adalah hasil *accuracy* dari hasil *confusion matrix* untuk data ulasan wisatawan :

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% = \frac{97+62}{97+62+11+10} = 88\%$$

Presisi adalah seberapa tepat model dalam membandingkan data secara keseluruhan dengan jumlah *Positive* yang di klasifikasi secara benar.

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP + FP} \times 100\% = \frac{97}{97+11} \\ &= 90\% \end{aligned}$$

Recall adalah sistem melakukan perbandingan jumlah data benar *Positive* dengan jumlah data yang sebenarnya *Positive* secara keseluruhan.

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP + FN} \times 100\% = \frac{97}{97+10} \\ &= 91\% \end{aligned}$$

F-Score didapatkan dari penggabungan hasil presisi dan *recall* dengan menggunakan rata-rata *harmonic* presisi dan *recall*.

$$\begin{aligned} \text{F-Score} &= 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = 2 \cdot \frac{90 \cdot 91}{90 + 91} = \frac{16.380}{181} \\ &= 90\% \end{aligned}$$

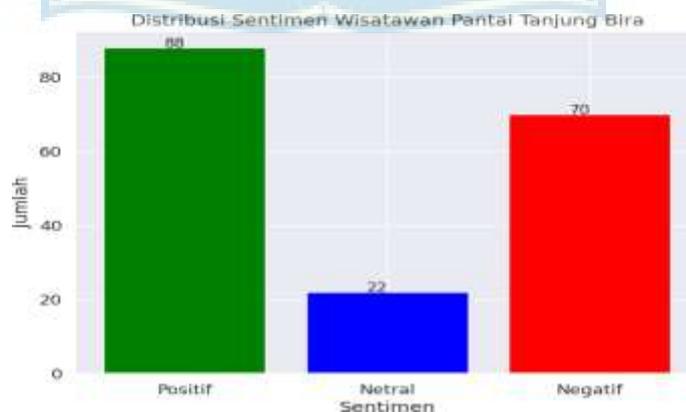
Terdapat 4 istilah sebagai representasi hasil proses klasifikasi pada *confusion matrix*. Keempat istilah tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN).

Tabel 6.Hasil Klasifikasi Dengan Labelling Manual Dan Confusion Matrix

Ulasan	Label	Label Prediksi <i>Naive Bayes</i>	Hasil <i>Confusion Matrix</i>
pantai indah bersih	Positif	Positif	<i>True Positif</i>
tempat kotor sampah	Negatif	Negatif	<i>True Negatif</i>
bira bonto bahari kabupaten bulukumba sulawesi selatan	Netral	Positif	<i>False Positive</i>
bersih	Positif	Negatif	<i>False Negative</i>

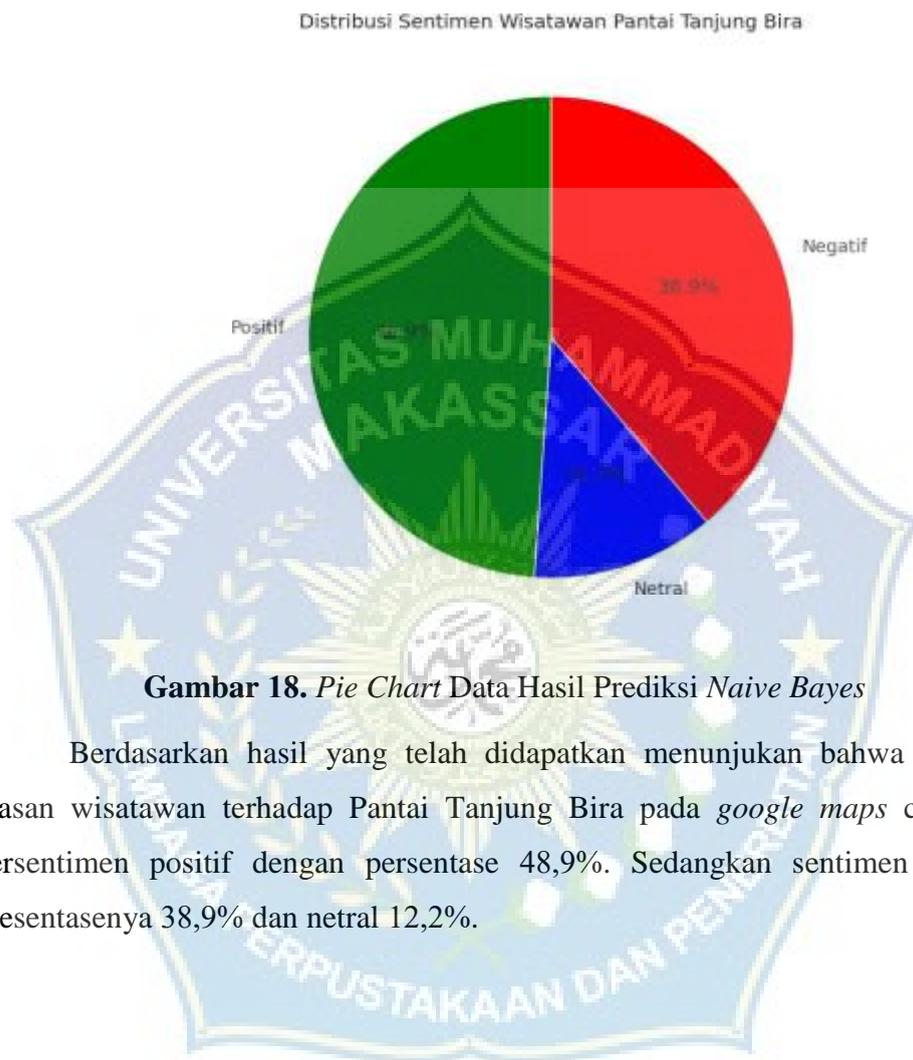
I. Visualisasi Sentimen

Berdasarkan data yang sudah diolah menggunakan teknik analisis sentimen, opini positif wisatawan destinasi Pantai Tanjung Bira masih terlihat lebih tinggi dibandingkan dengan opini negatif dan netral.



Gambar 17. Distribusi Data Hasil Prediksi *Naive Bayes*

Dari 180 jumlah *data testing* yang sudah diklasifikasikan menjadi sentimen positif, negatif dan netral, 48,9% diantaranya berkomentar positif, 12,2% berkomentar netral dan 38,9% berkomentar negatif.



Gambar 18. *Pie Chart* Data Hasil Prediksi *Naive Bayes*

Berdasarkan hasil yang telah didapatkan menunjukkan bahwa sentimen ulasan wisatawan terhadap Pantai Tanjung Bira pada *google maps* cenderung bersentimen positif dengan persentase 48,9%. Sedangkan sentimen negative persentasenya 38,9% dan netral 12,2%.

BAB V

PENUTUP

A. Kesimpulan

Berdasarkan penelitian analisis sentimen yang telah peneliti lakukan maka dapat diambil kesimpulan sebagai berikut :

1. Penelitian ini merupakan analisis sentimen terhadap ulasan wisatawan terhadap destinasi wisata pantai tanjung bira pada ulasan di google maps dengan metode *naive bayes*. Data *scrapping* dari ulasan wisatawan pada *google maps* disimpan dan selanjutnya akan diklasifikasikan menjadi tiga polaritas, yaitu positif, negatif dan netral. Dari total 1000 data, setelah melewati proses *preprocessing* didapatkan data bersih 900. Selanjutnya menggunakan model *splitting data* dengan perbandingan 80:20 terhadap dataset ulasan wisatawan dengan ukuran dataset 720 untuk *data training* dan 180 *data testing*.
2. Berdasarkan hasil yang telah didapatkan dari analisis sentimen wisatawan terhadap destinasi pantai tanjung bira dengan menggunakan *Naive Bayes* diperoleh hasil yang cukup tinggi dengan *accuracy* sebesar 88%.
3. Dari analisis sentimen wisatawan terhadap destinasi pantai tanjung bira pada ulasan wisatawan di *google maps* dengan algoritma *naive bayes* mendapatkan hasil dengan presentase kelas positif sebesar 48,9%, kelas negatif 38,9% dan netral dengan sebesar 12,2%. Ini menunjukkan bahwa sentimen ulasan wisatawan terhadap pantai tanjung bira pada *google maps* cenderung bersentimen negatif dengan persentase 48,9%.

B. Saran

Berdasarkan kesimpulan yang telah dijabarkan, maka disarankan pada penelitian selanjutnya menggunakan data yang lebih banyak lagi dan dapat menggunakan algoritma lain atau menggabungkan dua metode agar meningkatkan tingkat *accuracy* serta menggunakan model algoritma yang berbeda.

DAFTAR PUSTAKA

- ANALISIS SENTIMEN TERHADAP TEMPAT WISATA DI KABUPATEN REMBANG MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER.* (n.d.).
- Aponno, J. C. (2022). *Penerapan Algoritma Sentiment Analysis Dan Naïve Bayes Terhadap Opini Pengunjung Di Tempat Wisata Pantai Pintu Kota, Kota Ambon.* 9(4), 3180–3188. <http://jurnal.mdp.ac.id>
- Atmadja, B. R. (2022). *Analisis Sentimen Bahasa Indonesia Pada Tempat Wisata di Kabupaten Sukabumi Dengan Naïve Bayes.* 15(2), 371–382. <http://journal.stekom.ac.id/index.php/elkom/page371>
- Back, B. H., & Ha, I. K. (2019). Comparison of sentiment analysis from large twitter datasets by naive bayes and natural language processing methods. *Journal of Information and Communication Convergence Engineering*, 17(4), 239–245. <https://doi.org/10.6109/jicce.2019.17.4.239>
- Falasari, A., & Muslim, M. A. (2022). Optimize Naïve Bayes Classifier Using Chi Square and Term Frequency Inverse Document Frequency For Amazon Review Sentiment Analysis. *Journal of Soft Computing Exploration*, 3(1), 31–36. <https://doi.org/10.52465/josce.v3i1.68>
- Ferryawan, R., & Wahyu Wibowo, F. (2019). *Analisis Sentimen Wisata Jawa Tengah Menggunakan Naïve Bayes* (Vol. 5).
- Fitri, V. A., Andreswari, R., & Hasibuan, M. A. (2019). Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm. *Procedia Computer Science*, 161, 765–772. <https://doi.org/10.1016/j.procs.2019.11.181>
- Herlawati, H., Handayanto, R. T., Atika, P. D., Khasanah, F. N., Yusuf, A. Y. P., & Septia, D. Y. (2021). Analisis Sentimen Pada Situs Google Review dengan Naïve Bayes dan Support Vector Machine. *Jurnal Komtika (Komputasi Dan Informatika)*, 5(2), 153–163. <https://doi.org/10.31603/komtika.v5i2.6280>
- Khofifah, W., Rahayu, D. N., & Yusuf, A. M. (2022). Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps. *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi Dan*

- Komunikasi*, 16(4), 28–38. <https://doi.org/10.35969/interkom.v16i4.192>
- Natasuwarna STMIK Pontianak Jurusan Sistem Informasi, A. P. (n.d.). *ANALISIS SENTIMEN KEPUTUSAN PEMINDAHAN IBUKOTA NEGARA MENGGUNAKAN KLASIFIKASI NAIVE BAYES*.
- Normah, N. (2019). Naïve Bayes Algorithm For Sentiment Analysis Windows Phone Store Application Reviews. *Sinkron*, 3(2), 13. <https://doi.org/10.33395/sinkron.v3i2.242>
- Nurkholis, A., Fernando, Y., & Ans, F. A. (2023). METODE VECTOR SPACE MODEL UNTUK WEB SCRAPING PADA WEBSITE FREELANCE. *INTI Nusa Mandiri*, 18(1), 52–58. <https://doi.org/10.33480/inti.v18i1.4266>
- Putu, N. L. P. M., Ahmad Zuli Amrullah, & Ismarmiaty. (2021). Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 123–131. <https://doi.org/10.29207/resti.v5i1.2587>
- Samotana Zalukhu, P., Handhayani, T., & Sitorus, M. (2023). *ANALISIS SENTIMEN TERHADAP KENAIKAN BBM DI INDONESIA PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE NAÏVE BAYES*. 8(1).
- Siahaan, N. R., Tiffany, R. Y., Roland, S., Sinaga, E., Nauli, E. V., Fahmi, M. I., Vio, E., & Naibaho, N. B. (n.d.). *ANALISIS SENTIMEN ULASAN APLIKASI MEDIA SOSIAL WHATSAPP MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER*.
- Singgalen, Y. A. (2022). Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Tripadvisor Menggunakan Algoritma Naïve Bayes Classifier. *Building of Informatics, Technology and Science (BITS)*, 4(3). <https://doi.org/10.47065/bits.v4i3.2486>
- Suryadi, A., Syb'an, W. A., Alfa'inna, N., Hermaliani, E. H., & Mandiri, U. N. (2023). Implementasi Web Scraping dan Sentiment Analysis Terhadap Berita Menggunakan Machine Learning. *JURNAL SWABUMI*, 11(1), 2023.
- Suryati, E., Ari Aldino, A., Penulis Korespondensi, N., & Suryati Submitted, E. (2023). *Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine*

(SVM). 4(1), 96–106. <https://doi.org/10.33365/jtsi.v4i1.2445>

Wongkar, M., & Angdresey, A. (2019, October 1). Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter. *Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019*. <https://doi.org/10.1109/ICIC47613.2019.8985884>



LAMPIRAN

Lampiran 1. Sourcode

Install library

```
!pip install Sastrawi

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import re
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
import pandas as pd
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics import accuracy_score, confusion_matrix,
classification_report
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import ConfusionMatrixDisplay

# Download NLTK resources
import nltk
nltk.download('punkt')
nltk.download('stopwords')
```

Data understanding

```
df = pd.read_excel('/content/dataset ulasan wisatawan tanjung
bira.xlsx')
df.head(11)
df.describe()
df.isnull().sum()
df.dtypes
df.shape
df.duplicated().sum()
```

Pre processing data

```
#CLEANING DATA
df['comment'] = df['comment'].apply(lambda x: str(x).lower() if
isinstance(x, str) else x)

comment = 'comment'
df[comment] = df[comment].apply(lambda x: re.sub(r'^a-zA-Z0-
9\s]', '', str(x)))
df[comment] = df[comment].apply(lambda x: re.sub(r'\d', ' ',
str(x)))
df[comment] = df[comment].apply(lambda x: re.sub(r'\.{2,}',
' .', str(x)))

df.head(11)

#TOKENIZING
comment = 'comment'
df['tokenisasi'] = df['comment'].apply(lambda x:
word_tokenize(str(x)) if isinstance(x, str) else [])
df.head(11)

#STOPWORD REMOVAL
# Unduh set stopwords Bahasa Indonesia dari NLTK
stop_words = set(stopwords.words('indonesian'))
# Stopword removal pada kolom teks
df['stopword'] = df['tokenisasi'].apply(lambda tokens: '
'.join([word for word in tokens if word.lower() not in
stop_words]) if isinstance(tokens, list) else '')
# Tampilkan hasil
df.head()

#STEMMING
# Membuat objek stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()
# Kolom teks yang ingin di-stem
stemming = 'stopword'
# Melakukan stemming pada seluruh kolom teks
df['stemming'] = df[stemming].apply(lambda x: '
'.join([stemmer.stem(word) for word in x.split()]) if
isinstance(x, str) else '')

# Menyimpan hasil ke Excel jika diperlukan
```

```
df.to_excel('output_stemming.xlsx', index=False,
engine='openpyxl')
df.head(6)
```

Splitting Data

```
# Bagi data menjadi set pelatihan dan pengujian (80% pelatihan,
20% pengujian)
X_train, X_test, y_train, y_test =
train_test_split(stemming['stemming'], data_stemming['label'],
test_size=0.2, random_state=42)

# Cetak ukuran set pelatihan dan pengujian
print("Ukuran Set Pelatihan:", len(X_train), len(y_train))
print("Ukuran Set Pengujian:", len(X_test), len(y_test))
```

Vektorisasi TF-IDF

```
tfidf_vectorizer = TfidfVectorizer()
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)
X_test_tfidf = tfidf_vectorizer.transform(X_test)
```

Klasifikasi Model Naive Bayes

```
#latih model naive bayes
nb_classifier = MultinomialNB()
nb_classifier.fit(X_train_tfidf, y_train)

# Evaluasi performa pada data uji
accuracy_test = accuracy_score(y_test, y_test_pred)
conf_matrix_test = confusion_matrix(y_test, y_test_pred)
classification_rep_test = classification_report(y_test,
y_test_pred)

# menampilkan hasil evaluasi pada data uji
print("\nEvaluasi Model Naive Bayes pada Data Pengujian:")
print(f"Accuracy: {accuracy_test:.2f}")
print("\nConfusion Matrix:")
print(conf_matrix_test)
print("\nClassification Report:")
print(classification_rep_test)

# Tampilkan presentase akurasi pada data uji
```

```

print(f"Presentase Akurasi pada Data Pengujian: {accuracy_test
* 100:.2f}%")

# Membuat DataFrame dari X_test, y_test, dan y_test_pred
df_results = pd.DataFrame({'ulasan': X_test, 'label': y_test,
'predict_label': y_test_pred})

# Menyimpan DataFrame ke file Excel
df_results.to_excel('hasil_prediksi_naive_bayes.xlsx',
index=False)

```

Pengujian Confusion Matrix

```

# Uji model pada data pengujian
X_test_tfidf = tfidf_vectorizer.transform(X_test)
y_test_pred = nb_classifier.predict(X_test_tfidf)

# Hitung confusion matrix
conf_matrix = confusion_matrix(y_test, y_test_pred,
labels=['Positif', 'Netral', 'Negatif'])
# Evaluasi model dan mendapatkan confusion matrix
cm = confusion_matrix(y_test, y_pred)

# Menampilkan hasil prediksi dari confusion matrix
print(f'True Positive (TP): {cm[1, 1]}')
print(f'False Positive (FP): {cm[0, 1]}')
print(f'True Negative (TN): {cm[0, 0]}')
print(f'False Negative (FN): {cm[1, 0]}')

# Plot confusion matrix
sns.set(font_scale=1.2)
disp = ConfusionMatrixDisplay(conf_matrix,
display_labels=['Positif', 'Netral', 'Negatif'])
disp.plot(cmap='Blues', values_format='d')

plt.title('Confusion Matrix')
plt.show()

```

Visualisasi Sentimen dengan Pie chart

```

# Menghitung jumlah sentimen positif, netral, dan negatif
positive_count = (y_pred == 'Positif').sum()
neutral_count = (y_pred == 'Netral').sum()
negative_count = (y_pred == 'Negatif').sum()

```

```
# Labels dan data untuk plotting
labels = ['Positif', 'Netral', 'Negatif']
counts = [positive_count, neutral_count, negative_count]

# Plot pie chart
plt.figure(figsize=(8, 8))
sns.set_palette("viridis") # Ganti warna plot sesuai keinginan
plt.pie(counts, labels=labels, autopct='%1.1f%%',
startangle=90)
plt.title('Distribusi Sentimen Wisatawan Pantai Tanjung Bira')

# Tampilkan plot
plt.show()
```



Lampiran 2. Hasil Pengumpulan Data

1	user	date	comment
2	Andi Mufidatunnisa	setahun lalu	pantai tercantik yang pernah aku liat
3	Aldal Reckers	8 bulan lalu	pantainya indah
4	Rosadhy Wahyudhy	3 tahun lalu	Sangat menyukai pantai, apalagi pasirnya putih dan bersih
5	Juani Agung	3 tahun lalu	Tempatnya bagus
6	Toni Wardono	4 tahun lalu	Sangat indah pantai dengan pasir putihnya semoga destinasi wisata ini terjaga kebersihannya...
7	ateng handy	5 tahun lalu	Kalo disini enakny bermalam... karena mandi dipantai lebih asyik di pagi hari. jangan lupa bawa peralatan snorkling...
8	Ardhy Jupie	3 tahun lalu	Pemandangan sangat indah... keren buat liburan sama teman teman dan keluarga
9	Yuni Antoko	5 tahun lalu	Pasir putihnya lembut... ada permainan berbayar bagi penggemar kegiatan pantai...
10	Nur Alam	9 bulan lalu	Sering dibicarakan orang insyaallah bakal kesini
11	Anneke Mutiarsani	3 tahun lalu	Pantai pasir putih, banyak kegiatan yang bisa di lakukan di pantai bisa bersama keluarga.
12	Early Carolina	2 tahun lalu	Tempat terkeren... pasir putih laut biru banyak tempat wisata baru pokoknya galing favorit
11	Diman Sinergy Channel	10 bulan lalu	Pemandangan yang bagus
14	Denny Eskobra	4 tahun lalu	Pasir pantainya bersih dan halus, sangat memanjakan mata
15	Gianendra Cahyo	5 tahun lalu	pantai nya biru dan jernih dengan pasir putih yang memukau
16	Syarita Agriana Assegaf	3 tahun lalu	Pasir putihnya cantik pemandangan lautnya biru
17	Taufiq Lubis	4 tahun lalu	Keseluruhan masih alami keren
18	resmi daeng tabe	3 tahun lalu	Wisata pantai langsung bisa yang keren
19	Clai Karina	4 tahun lalu	Tempat sangat terkeren walaupun jauh tapi terbayarkan dengan pemandangannya

Lampiran 3. Hasil Pelabelan

index	user	date	comment	label
0	Andi Mufidatunnisa	setahun lalu	pantai tercantik yang pernah aku liat	Postif
1	Aldal Reckers	8 bulan lalu	pantainya indah	Postif
2	Rosadhy Wahyudhy	3 tahun lalu	Sangat menyukai pantai, apalagi pasirnya putih dan bersih	Postif
3	Juani Agung	3 tahun lalu	Tempatnya bagus	Postif
4	Toni Wardono	4 tahun lalu	Sangat indah pantai dengan pasir putihnya semoga destinasi wisata ini terjaga kebersihannya...	Postif
5	ateng handy	5 tahun lalu	Kalo disini enakny bermalam... karena mandi dipantai lebih asyik di pagi hari. jangan lupa bawa peralatan snorkling...	Postif
6	Ardhy Jupie	3 tahun lalu	Pemandangan sangat indah... keren buat liburan sama teman teman dan keluarga	Postif

Lampiran 4. Hasil Proses Cleaning

pantai tercantik yang pernah aku liat
pantainya indah
sangat menyukai pantai apalagi pasirnya putih dan bersih
tempatnya bagus
sangat indah pantai dengan pasir putihnya semoga destinasi wisata ini terjaga kebersihannya
kalo disini enakny bermalam karena mandi dipantai lebih asyik di pagi hari jangan lupa bawa peralatan snorkling
pemandangan sangat indah keren buat liburan sama teman teman dan keluarga
pasir putihnya lembut ada permainan berbayar bagi penggemar kegiatan pantai
sering dibicarakan orang insyaallah bakal kesini

Lampiran 5. Hasil Proses Tokenizing

tokenisasi
pantai, tercantik, yang, pernah, aku, liat
pantainya, indah
sangat, menyukai, pantai, apalagi, pasirnya, putih, dan, bersih
tempatnya, bagus
sangat, indah, pantai, dengan, pasir, putihnya, semoga, destinasi, wisata, ini, terjaga, kebersihannya
kalo, disini, enakny, bermalam, karena, mandi, dipantai, lebih, asyik, di, pagi, hari, jangan, lupa, bawa, peralatan, snorkling
pemandangan, sangat, indah, keren, buat, liburan, sama, teman, teman, dan, keluarga

Lampiran 6. Hasil Proses Stopword

stopword
pantai tercantik liat
pantainya indah
menyukai pantai pasirnya putih bersih
tempatny bagus
indah pantai pasir putihnya semoga destinasi wisata terjaga kebersihannya
kalo enakny bermalam mandi dipantai asyik pagi lupa bawa peralatan snorkling
pemandangan indah keren liburan teman teman keluarga

Lampiran 7. Hasil Proses Stemming

pantai cantik liat
pantai indah
suka pantai pasir putih bersih
tempat bagus
indah pantai pasir putih moga destinasi wisata jaga bersih
kalo enak malam mandi pantai asyik pagi lupa bawa alat snorkling

Lampiran 8. Hasil Splitting data

Ukuran Set Pelatihan: 720 data
Ukuran Set Pengujian: 180 data

Sampel dari Set Pengujian:

```
['objek wisata indah cocok libur keluarga' 'sih baguz' 'pasir putih' 'susah umat muslim sholat mushollah' 'indah pasir putih']
```

Jumlah Label pada Set Pelatihan:

```
(array(['Negatif', 'Netral', 'Positif'], dtype=object), array([234, 204, 282]))
```

Jumlah Label pada Set Pengujian:

```
(array(['Negatif', 'Netral', 'Positif'], dtype=object), array([66, 32, 82]))
```

Lampiran 9. Hasil Pembobotan TF-IDF

```
from sklearn.feature_extraction.text import TfidfVectorizer
tfidf_vectorizer = TfidfVectorizer()
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)
X_test_tfidf = tfidf_vectorizer.transform(X_test)

print("Ukuran matriks fitur Tf-idf Set Pelatihan:", X_train_tfidf.shape)
print("Ukuran matriks fitur Tf-idf Set Pengujian:", X_test_tfidf.shape)
```

Ukuran matriks fitur Tf-idf Set Pelatihan: (720, 968)
Ukuran matriks fitur Tf-idf Set Pengujian: (180, 968)

Lampiran 10. Hasil Pengujian Confusion Matrix

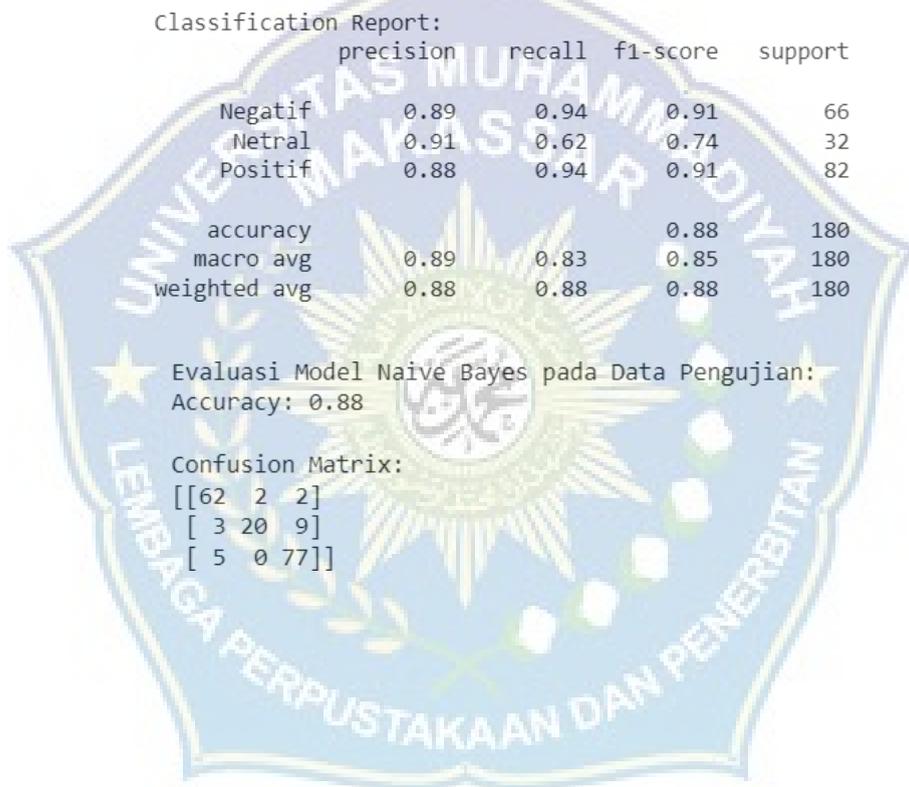
Classification Report:

	precision	recall	f1-score	support
Negatif	0.89	0.94	0.91	66
Netral	0.91	0.62	0.74	32
Positif	0.88	0.94	0.91	82
accuracy			0.88	180
macro avg	0.89	0.83	0.85	180
weighted avg	0.88	0.88	0.88	180

Evaluasi Model Naive Bayes pada Data Pengujian:
Accuracy: 0.88

Confusion Matrix:

```
[[62  2  2]
 [ 3 20  9]
 [ 5  0 77]]
```





Lampiran 11. Surat Keterangan Bebas Plagiasi
MAJELIS PENDIDIKAN TINGGI PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH MAKASSAR
UPT PERPUSTAKAAN DAN PENERBITAN

Alamat kantor: Jl. Sultan Alauddin NO.259 Makassar 90221 Tlp.(0411) 866972,881593, Fax.(0411) 865588

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

SURAT KETERANGAN BEBAS PLAGIAT

UPT Perpustakaan dan Penerbitan Universitas Muhammadiyah Makassar,
 Menerangkan bahwa mahasiswa yang tersebut namanya di bawah ini:

Nama : Fajar Maulana Rahman
 Nim : 105841107419
 Program Studi : Teknik Informatika

Dengan nilai:

No	Bab	Nilai	Ambang Batas
1	Bab 1	10 %	10 %
2	Bab 2	19 %	25 %
3	Bab 3	10 %	10 %
4	Bab 4	7 %	10 %
5	Bab 5	5 %	5 %

Dinyatakan telah lulus cek plagiat yang dilakukan oleh UPT- Perpustakaan dan Penerbitan Universitas Muhammadiyah Makassar Menggunakan Aplikasi Turnitin.

Demikian surat keterangan ini diberikan kepada yang bersangkutan untuk dipergunakan seperlunya.

Makassar, 27 Januari 2024
 Mengetahui,

Kepala UPT Perpustakaan dan Penerbitan,



Jl. Sultan Alauddin no 259 makassar 90222
 Telepon (0411)866972,881 593,fax (0411)865 588
 Website: www.library.unismuh.ac.id
 E-mail : perpustakaan@unismuh.ac.id

Lampiran 12. Scan Plagiasi Per Bab



Fajar Maulana Rahman

105841107419 BAB I

by Tahap Tutup

Submission date: 26-Jan-2024 02:10PM (UTC+0700)

Submission ID: 2278821406

File name: BAB_I_-_2024-01-26T151024.805.docx (56.72K)

Word count: 461

Character count: 3038

Fajar Maulana Rahman 105841107419 BAB I

ORIGINALITY REPORT

10 **10%** **2%** **3%**
SIMILARITY INTERNET SOURCES PUBLICATIONS STUDENT PAPERS

PRIMARY SOURCE

1	repositori.uin-alauddin.ac.id Internet Source	3%
2	risetsdmstieasia.blogspot.com Internet Source	3%
3	theclearvoyantlife.com Internet Source	2%
4	123dok.com Internet Source	2%

Exclude quotes Exclude matches
Exclude bibliography

Fajar Maulana Rahman
105841107419 BAB II
by Tahap Tutup



Submission date: 26-Jan-2024 02:11PM (UTC+0700)

Submission ID: 2278821587

File name: BAB_II_-_2024-01-26T151032.884.docx (175.2K)

Word count: 2906

Character count: 19452

rajar Maulana Rahman 105841107419 BAB II

ORIGINALITY REPORT

19%



10%

SIMILARITY INDEX

INTERNET SOURCES

8%

PUBLICATIONS

7%

STUDENT PAPERS

PRIMARY SOURCES

Rank	Source	Percentage
1	ejurnal.seminar-id.com Internet Source	3%
2	123dok.com Internet Source	2%
3	repository.iainkudus.ac.id Internet Source	2%
4	lecturer.ukdw.ac.id Internet Source	1%
5	Submitted to Udayana University Student Paper	1%
6	Submitted to Universitas Brawijaya Student Paper	1%
7	Herlawati Herlawati, Rahmadya Trias Handayanto, Prima Dina Atika, Fata Nidaul Khasanah et al. "Analisis Sentimen Pada Situs Google Review dengan Naïve Bayes dan Support Vector Machine", Jurnal Komtika (Komputasi dan Informatika), 2021 Publication	1%

Fajar Maulana Rahman

105841107419 BAB III

by Tahap Tutup



Submission date: 26-Jan-2024 02:11PM (UTC+0700)

Submission ID: 2278821761

File name: BAB_III_-_2024-01-26T151053.056.docx (138.52K)

Word count: 1282

Character count: 8317

Fajar Maulana Rahman 105841107419 BAB III

ORIGINALITY REPORT

10%

SIMILARITY INDEX



INTERNET SOURCES

5%

PUBLICATIONS

8%

STUDENT PAPERS

PRIMARY SOURCE

turnitin

1	jurnal-ppi.kominfo.go.id Internet Source	3%
2	Submitted to UPN Veteran Yogyakarta Student Paper	1%
3	unsada.e-journal.id Internet Source	1%
4	docplayer.info Internet Source	1%
5	digilibadmin.unismuh.ac.id Internet Source	1%
6	e-journal.unipma.ac.id Internet Source	1%
7	Submitted to Forum Komunikasi Perpustakaan Perguruan Tinggi Kristen Indonesia (FKPPTKI) Student Paper	1%

Exclude quotes

Off

Exclude matches

Off

Fajar Maulana Rahman
105841107419 BAB IV
by Tahap Tutup



Submission date: 26-Jan-2024 02:15PM (UTC+0700)

Submission ID: 2278823001

File name: BAB_IV_-_2024-01-26T151118.230.docx (1.55M)

Word count: 1143

Character count: 7356

rajar Maulana Rahman 105841107419 BAB IV

ORIGINALITY REPORT

7 %
SIMILARITY INDEX



3 %
PUBLICATIONS

4 %
STUDENT PAPERS

PRIMARY SOURCES

- | | | |
|---|--|----|
| 1 | Submitted to Universitas Pendidikan Ganesha
Student Paper | 2% |
| 2 | Submitted to Sriwijaya University
Student Paper | 2% |
| 3 | digilib.uin-suka.ac.id
Internet Source | 1% |
| 4 | zoelonline.wordpress.com
Internet Source | 1% |
| 5 | docplayer.info
Internet Source | 1% |
| 6 | eprints.uns.ac.id
Internet Source | 1% |

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off

Fajar Maulana Rahman
105841107419 BAB V
by Tahap Tutup



Submission date: 26-Jan-2024 02:15PM (UTC+0700)

Submission ID: 2278823279

File name: BAB_V_-_2024-01-26T151519.554.docx (53.75K)

Word count: 184

Character count: 1206

ORIGINALITY REPORT

5%

SIMILARITY INDEX

PRIMARY SOURCES



5%

INTERNET SOURCES

0%

PUBLICATIONS

0%

STUDENT PAPERS

1

repository.its.ac.id
Internet Source

5%

Exclude quotes

Off

Exclude bibliography

Off

Exclude matches

Off



