

**ANALISIS SENTIMEN WISATAWAN TERHADAP
DESTINASI PANTAI TANJUNG BIRA MENGGUNAKAN
ALGORITMA *REGRESI LOGISTIK***

SKRIPSI

Diajukan sebagai salah satu syarat untuk menyusun skripsi Program Studi
Informatika



SADDAN

105841107719

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MUHAMMADIYAH MAKASSAR

2024

**ANALISIS SENTIMEN WISATAWAN TERHADAP
DESTINASI PANTAI TANJUNG BIRA MENGGUNAKAN
ALGORITMA *REGRESI LOGISTIK***

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

Disusun Dan Di Ajukan Oleh :

SADDAN

105841107719

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

2024

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 **Saddan** dengan nomor induk Mahasiswa **105 84 11077 19**, dinyatakan diterima dan disahkan oleh Panitia Ujian Tugas Akhir/Skripsi sesuai dengan Surat Keputusan Dekan Fakultas Teknik Universitas Muhammadiyah Makassar Nomor : 459/05/A.5-VI/X/45/2023, sebagai salah satu syarat guna memperoleh gelar Sarjana Komputer pada Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar pada hari Sabtu tanggal 06 Januari 2024.

Panitia Ujian : Makassar, 21 Muharram 1445 H
27 Juli 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. Sekertaris : Lukman, S.Kom., M.T

3. Anggota

- 1. Muhyidin A.M Hayat, S.kom, M.T
- 2. Rizki Yusliana Bakti, ST., MT
- 3. Lukman Anas, S.Kom., MT.

Mengetahui :

Pembimbing I

Pembimbing II

Fahrim Irhamna Rahman, S.Kom., MT.

Titin Wahyuni, S.Pd., MT.



Dekan Fakultas Teknik

Dr. Ir. Hj. Nurnawaty, ST., MT., IPM

NBM : 795 108



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 PERSETUJUAN

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 MENGGUNAKAN ALGORITMA REGRESI LOGISTIK**

Nama : SADDAN

Stambuk : 105841107719



Makassar, 27 Juli 2024

Telah Diperiksa dan Disetujui Oleh Dosen Pembimbing;

Pembimbing I

Pembimbing II

Fahrir Irhamna Rahman, S.Kom., MT.

Titin Wahyuni, S.Pd., MT.

Mengetahui,

Ketua Program Studi Informatika



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

NBM : 1504 577

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh

Puji syukur senantiasa kita panjatkan kepada Allah SWT karena atas limpahan rahmat, ridha, dan karunianya penulis dapat menyelesaikan penelitian ini tepat pada waktunya. Shalawat serta salam tak lupa saya haturkan kepada Nabi Muhammad SAW yang selalu menjadi teladan bagi umatnya. Sehingga dapat menyelesaikan skripsi ini dengan judul “ANALISIS SENTIMEN WISATAWAN TERHADAP DESTINASI PANTAI TANJUNG BIRA MENGGUNAKAN ALGORITMA *REGRESI LOGISTIK*”.

Pada kesempatan ini penulis ingin mengucapkan terima kasih kepada semua pihak yang telah memberikan bantuan, baik berupa moril maupun spiritual selama pelaksanaan penelitian ini. Ucapan terima kasih ini penulis sampaikan kepada :

1. Allah Subhanahu Wa Ta'ala yang telah memberikan nikmat sehat dan nikmat kesempatan untuk bisa menyelesaikan skripsi ini.
2. Bapak **Mustari** dan Ibu **Rukiati** tercinta yang selalu memberikan dukungan dan selalu mendoakan anaknya pada setiap pengerjaan skripsi
3. Ibu **Dr.Hj.Ir Nurnawaty, S.T., M.T** Selaku dekan fakultas Teknik Universitas Muhammadiyah Makassar
4. Bapak **Muh Syafaat S.Kuba S.T., M.T** Selaku Wakil Dekan 1 Fakultas Teknik Universitas Muhammadiyah Makassar
5. Bapak **Muhyiddin A.M Hayat, S.Kom, M.T.** Selaku Ketua Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar
6. Bapak **Fahrim Irhamna Rachman, S.Kom., M.T** Selaku Pembimbing Pertama
7. Ibu **Titin Wahyuni, S.Pd., M.T** Selaku Pembimbing Kedua
8. Dosen dan Staf Fakultas Teknik Universitas Muhammadiyah Makassar
9. Saudara/saudari di Fakultas Teknik, KOORDINAT 19 yang selalu belajar dan berjuang bersama atas nama persaudaraan, yang banyak membantu serta selalu memberikan dukungan dalam pengerjaan skripsi.

10. Seluruh pihak terkait yang tidak dapat saya sebutkan satu persatu selama proses pengerjaan skripsi.

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

Makassar , 25 Mei 2024

Penulis

Saddan



ABSTRAK

SADDAN. Analisis Sentimen Wisatawan Terhadap Destinasi Pantai Tanjung Bira Menggunakan Algoritma *Regresi logistik* (Dibimbing oleh Fahrin Irhamna Rachman, S.Kom., M.T dan Titin Wahyuni, S.Pd., M.T).

Penelitian ini bertujuan untuk menganalisis sentimen wisatawan terhadap destinasi Pantai Tanjung Bira yang diungkapkan dalam ulasan di *Google maps* menggunakan algoritma *regresi logistik*. Metode analisis sentimen yang digunakan melibatkan pengumpulan data ulasan dari *Google maps*, pra-pemrosesan teks, dan klasifikasi sentimen ke dalam kategori positif, negatif, atau netral. *Regresi logistik* dipilih sebagai algoritma untuk membangun model prediktif yang mampu mengidentifikasi pola sentimen dari teks ulasan.

Dalam penelitian ini, kami mengumpulkan 1.000 ulasan wisatawan mengenai Pantai Tanjung Bira. Data ulasan tersebut kemudian diproses melalui beberapa tahapan, termasuk *cleaning*, *case folding*, *stopword*, dan *tokenisasi*. Model *regresi logistik* dilatih dan dievaluasi menggunakan dataset yang telah dibagi menjadi data latih dan data uji dengan 3 perbandingan 90:10, 80:20, 70:30.

Hasil penelitian menunjukkan bahwa model *regresi logistik* yang dibangun memiliki akurasi rata-rata 70% dalam mengklasifikasikan sentimen ulasan wisatawan. Faktor-faktor yang paling mempengaruhi sentimen positif termasuk keindahan alam, kebersihan pantai, dan keramahan penduduk setempat, sedangkan sentimen negatif banyak disebabkan oleh keluhan tentang fasilitas umum dan aksesibilitas.

Kata kunci: Analisis Sentimen, Pantai Tanjung Bira, Ulasan Wisatawan, *Google maps*, *Regresi Rogistik*.

ABSTRACT

SADDAN. *Analysis of Tourist Sentiment towards the Tanjung Bira Beach Destination Using the Logistic Regression Algorithm (Supervised by Fahrir Irhamna Rachman, S.Kom., M.T and Titin Wahyuni, S.Pd., M.T).*

This research aims to analyze tourist sentiment towards the Tanjung Bira Beach destination expressed in reviews on Google maps using a logistic regression algorithm. The sentiment analysis method used involves collecting review data from Google maps, pre-processing the text, and classifying sentiment into Positive, negative, or neutral categories. Logistic regression was chosen as an algorithm to build a predictive model capable of identifying sentiment patterns from review text. In this research, we collected 1,000 tourist reviews about Tanjung Bira Beach. The review data is then processed through several stages, including cleaning, case folding, stopwords, and tokenization. The logistic regression model is trained and evaluated using a dataset that has been divided into training data and test data with 3 ratios 90:10, 80:20, 70:30.

The research results show that the logistic regression model built has an average accuracy of 70% in classifying tourist review sentiment. The factors that most influence Positive sentiment include natural beauty, cleanliness of beaches, and the friendliness of local residents, while negative sentiment is mostly caused by complaints about public facilities and accessibility.

Keywords: *Sentiment Analysis, Tanjung Bira Beach, Tourist Reviews, Google maps, Logistic Regression*

DAFTAR ISI

SAMPUL.....	i
HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
KATA PENGANTAR	v
ABSTRAK.....	vii
<i>ABSTRACT</i>	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xii
DAFTAR LAMPIRAN.....	xiii
DAFTAR ISTILAH	xiv
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 Kepenulisan.....	3
BAB II TINJAUAN PUSTAKA.....	4
A. Landasan Teori.....	4
1. Analisis Sentimen.....	4
2. Klasifikasi.....	5
3. Pantai Tanjung Bira.....	6
4. <i>Google maps</i>	6

5. Algoritma <i>Regresi logistik</i>	7
6. <i>Prepossessing Data</i>	8
B. Penelitian Terkait	9
C. Kerangka Fikir	11
BAB III METODE PENELITIAN.....	12
A. Tempat dan Waktu Penelitian	12
B. Alat Dan Bahan	12
C. Perancangan Sistem	13
D. Teknik Pengujian Sistem	16
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	19
A. Pengumpulan Data	19
B. Pelabelan Data.....	21
C. <i>Prepossessing Data</i>	22
D. Klasifikasi <i>Regresi logistik</i>	28
E. Evaluasi Model <i>Regresi logistik</i>	29
F. Pengujian <i>Confusion Matrix</i>	32
BAB V PENUTUP.....	41
A. Kesimpulan	41
B. Saran.....	41
DAFTAR PUSTAKA	42
LAMPIRAN.....	45

DAFTAR GAMBAR

Gambar 1. Kerangka Berfikir.....	11
Gambar 2. <i>Flowchart System</i>	13
Gambar 3. Tahapan Pengumpulan Data.....	14
Gambar 4. Tahapan <i>Prepossessing</i> Data	15
Gambar 5. <i>Confusion Matrix</i>	16
Gambar 6. Proses <i>Instal</i> Ekstensi <i>Instant Data Scraper</i>	19
Gambar 7. Proses <i>Scrapping</i> Data	20
Gambar 8. Proses Pembuatan Model <i>Regresi Logistik</i>	29
Gambar 9. Hasil Pengujian Data (90% data <i>training</i> 10% data <i>test</i>).....	31
Gambar 10. Hasil Pengujian Data (80% data <i>training</i> 20% data <i>test</i>).....	32
Gambar 11. Hasil Pengujian Data (70% data <i>training</i> 30% data <i>test</i>).....	32
Gambar 12. Hasil Pengujian <i>Confusion Matrix</i> (90% data <i>training</i> 10% data <i>test</i>)	33
Gambar 13. <i>Confusion Matrix</i> (90% data <i>training</i> 10% data <i>test</i>)	33
Gambar 14. Hasil Pengujian <i>Confusion Matrix</i> (80% data <i>training</i> 20% data <i>test</i>)	33
Gambar 15. <i>Confusion Matrix</i> (80% data <i>training</i> 20% data <i>test</i>)	34
Gambar 16. Hasil Pengujian <i>Confusion Matrix</i> (70% data <i>training</i> 30% data <i>test</i>)	34
Gambar 17. <i>Confusion Matrix</i> (70% data <i>training</i> 30% data <i>test</i>)	34

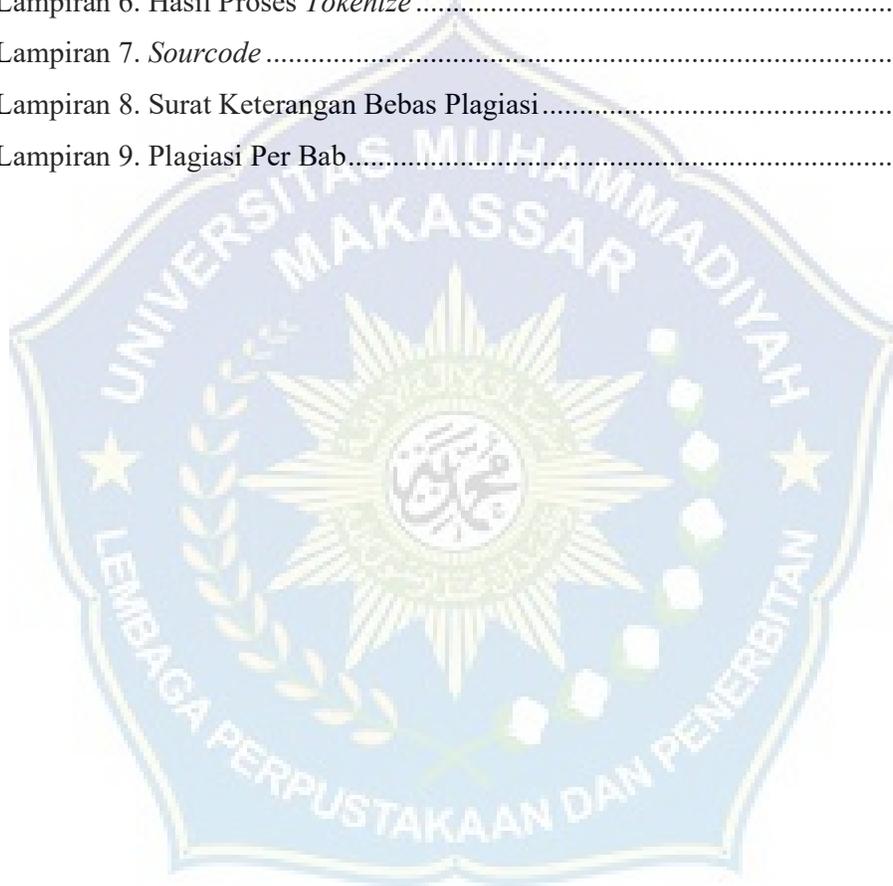
DAFTAR TABEL

Tabel 1. Spesifikasi Komputer.....	12
Tabel 2. <i>Dataset Hasil Scapping</i>	20
Tabel 3. Tahap Pelabelan Data.....	22
Tabel 4. <i>Proses Cleaning Data</i>	22
Tabel 5. <i>Proses Case Folding</i>	24
Tabel 6. <i>Proses Stopword</i>	26
Tabel 7. <i>Proses Tokenize</i>	27
Tabel 8. Hasil Pengujian Data (90% data <i>training</i> 10% data <i>test</i>).....	30
Tabel 9. Hasil Pengujian Data (80% data <i>training</i> 20% data <i>test</i>).....	30
Tabel 10. Hasil Pengujian Data (70% data <i>training</i> 30% data <i>test</i>).....	31
Tabel 11. <i>Confusion Matrix</i>	35
Tabel 12. Hasil <i>Confusion Matrix</i> (90% data <i>training</i> 10% data <i>test</i>).....	35
Tabel 13. Hasil <i>Confusion Matrix</i> (80% data <i>training</i> 20% data <i>test</i>).....	37
Tabel 14. Hasil <i>Confusion Matrix</i> (70 data <i>training</i> 30% data <i>test</i>)	38
Tabel 15. Hasil Klasifikasi Dengan <i>Labelling</i> Manual Dan <i>Confusion Matrix</i>	39

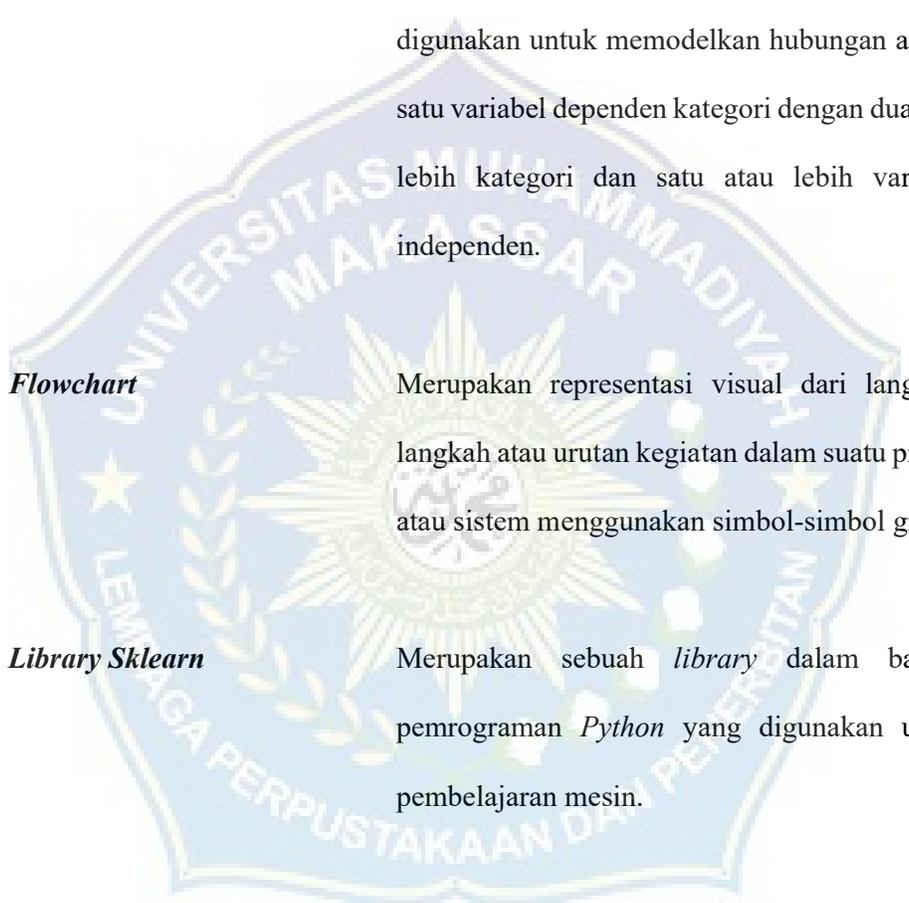


DAFTAR LAMPIRAN

Lampiran 1. <i>Dataset</i> Ulasan.....	45
Lampiran 2. Hasil Pelabelan	49
Lampiran 3. Hasil Proses <i>Cleaning</i> data.....	49
Lampiran 4. Hasil Proses <i>Case Folding</i>	49
Lampiran 5. Hasil Proses <i>Stopword</i>	50
Lampiran 6. Hasil Proses <i>Tokenize</i>	50
Lampiran 7. <i>Sourcode</i>	51
Lampiran 8. Surat Keterangan Bebas Plagiasi.....	56
Lampiran 9. Plagiasi Per Bab.....	57



DAFTAR ISTILAH



<i>Analisis Sentimen</i>	Proses evaluasi opini, perasaan, dan emosi yang di ungkapkan dalam teks.
<i>Algoritma Regresi Logistik</i>	Merupakan teknik analisis statistik yang digunakan untuk memodelkan hubungan antara satu variabel dependen kategori dengan dua atau lebih kategori dan satu atau lebih variabel independen.
<i>Flowchart</i>	Merupakan representasi visual dari langkah-langkah atau urutan kegiatan dalam suatu proses atau sistem menggunakan simbol-simbol grafis.
<i>Library Sklearn</i>	Merupakan sebuah <i>library</i> dalam bahasa pemrograman <i>Python</i> yang digunakan untuk pembelajaran mesin.
<i>Machine Learning</i>	Merupakan cabang dari kecerdasan buatan AI yang berfokus pada pengembangan algoritma dan teknik yang memungkinkan komputer untuk belajar dari dan membuat prediksi.

Preprocessing Data

Yaitu serangkaian langkah awal dalam pengolahan data mentah untuk mempersiapkannya menjadi format yang lebih bersih, terstruktur, dan siap untuk dianalisis atau digunakan dalam model *machine learning*.

Scrapping

Proses otomatisasi untuk mengambil konten dan data dari situs web. Data yang diekstrak bisa berupa teks, gambar, video, atau informasi lainnya yang terdapat pada halaman web.

Training

Merupakan sekumpulan data yang digunakan untuk melatih model *machine learning* atau algoritma prediksi.

Testing

Yaitu untuk mengukur seberapa baik model yang telah dilatih dapat menggeneralisasi pola dari data yang belum pernah dilihat sebelumnya



BAB I

PENDAHULUAN

A. Latar Belakang

Pantai tanjung bira adalah destinasi wisata utama di tanjung bira yang berupa pantai pasir putih dengan butiran yang halus. Pantai bira adalah tempat yang sangat bagus untuk bermain air serta berenang di laut. Ada berbagai macam wahana permainan untuk berekreasi di pantai misalnya *banana boat* dan donat *boat*. Dataran pantai yang luas sangat cocok untuk ditempati bermain seperti *volley* pantai dan sepak bola. Pemberian ulasan oleh pengguna di platform *google maps* dapat menjadi sumber informasi yang sangat bermanfaat bagi calon wisatawan yang hendak mengunjungi destinasi pantai tanjung bira. *Google maps* merupakan salah satu platform yang paling sering digunakan oleh pengguna untuk mencari berbagai informasi seperti lokasi, ulasan, dan *rating* suatu destinasi wisata. *Google maps* memberikan kesempatan untuk pengunjung dan pengguna umum untuk berbagi pengalaman saat mengunjungi Pantai Tanjung Bira. Ulasan, komentar, dan *rating* yang diberikan oleh pengguna di *Google maps* menjadi cerminan umum tentang pengalaman mereka selama berkunjung ke destinasi tersebut.

Analisis sentimen dan *Opinion mining* adalah bidang studi yang menganalisis pendapat seseorang, sentimen seseorang, evaluasi seseorang, sikap seseorang dan emosi seseorang ke dalam bahasa tertulis . Analisis sentimen telah banyak menerima perhatian semenjak penelitian Pang, Turney, Goldberg dan Zhu (Wardani, 2015). Analisis sentimen atau *Opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini (Sabaruddin, 2016).

Dengan teknologi canggih saat ini, pengunjung dapat berkomentar ataupun mengulas tempat atau destinasi yang mereka kunjungi, termasuk pada objek wisata. Jadi untuk mengetahui ulasan dari wisatawan di *Google maps*, hal yang harus dilakukan adalah mengumpulkan data ulasan tersebut menggunakan metode analisis sentimen. Dalam analisis sentimen diperlukan sebuah metode, salah satu yang sering digunakan adalah *Regresi logistik*.

Berdasarkan latar belakang yang dijelaskan di atas membuat peneliti tertarik untuk meneliti mengenai ulasan-ulasan dari pengguna aplikasi *google maps*. Maka peneliti memfokuskan dengan judul “**ANALISIS SENTIMEN WISATAWAN TERHADAP DESTINASI PANTAI TANJUNG BIRA MENGGUNAKAN ALGORITMA *REGRESI LOGISTIK***”.

B. Rumusan Masalah

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

1. Bagaimana melakukan analisis sentimen berdasarkan ulasan terhadap destinasi wisata pantai tanjung bira di *google maps* menggunakan metode *regrasi logistik*?
2. Bagaimana tingkat keakuratan *regresi logistik* dalam menganalisis sentimen wisatawan terhadap Pantai Tanjung Bira?

C. Tujuan Penelitian

Adapun tujuan penelitian yaitu:

1. Menganalisis ulasan destinasi pantai tanjung bira pada aplikasi *google maps* menggunakan *regresi logistik*
2. Menganalisis tingkat keakuratan sentimen wisatawan terhadap destinasi pantai tanjung bira menggunakan algoritma *regresi logistik*

D. Manfaat Penelitian

1. Dapat merekomendasikan kepada wisatawan yang akan memulai perjalanan dengan melihat nilai akurasi yang telah diteliti pada ulasan di *Google maps*.
2. Penelitian ini dapat memberikan pengetahuan tentang bagaimana algoritma *regresi logistik* dalam analisis sentimen wisatawan terhadap pantai tanjung bira pada ulasan di *google maps*.
3. Menjadi tolak ukur peneliti selanjutnya bagi yang akan melakukan penelitian terkait

E. Ruang Lingkup Penelitian

Adapun ruang lingkup penelitian ini ditentukan di antaranya:

1. Menganalisis ulasan wisatawan yang ada di *google maps* khususnya di pantai tanjung bira

2. Data diambil berasal dari ulasan para pengguna aplikasi *google maps* khususnya di pantai tanjung bira
3. Klasifikasi kelas sentimen yang dipakai hanya ulasan positif ulasan netral dan ulasan negatif.

F. Sistematika Kepenulisan

Struktur sistematika penulisan penelitian ini terbagi menjadi beberapa BAB, masing-masing diuraikan sebagai berikut :

BAB I PENDAHULUAN

Bab ini meliputi latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan pada penelitian yang dilakukan.

BAB II TINJAUAN PUSTAKA

Bab ini menguraikan pembahasan tentang penelitian yang terkait, landasan teori, serta kerangka pikir.

BAB III METODE PENELITIAN

Bab ini memberikan gambaran tentang penelitian yang dilakukan dan penjelasan tentang apa yang dilakukan pada penelitian ini.

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

Bab ini berisi tentang hasil penelitian, pembahasan dan hasil implementasi sistem.

BAB V PENUTUP

Bab ini berisi kesimpulan dan saran berdasarkan hasil penelitian.

BAB II

TINJAUAN PUSTAKA

A. Landasan Teori

1. Analisis Sentimen

Analisis sentimen adalah studi komputasi dari opini-opini, sentimen, dan emosi yang di ungkapkan dalam teks. Tugas utama analisis sentimen adalah mengelompokkan polaritas dari kalimat atau dokumen untuk mengetahui pendapat yang di ungkapkan dalam apakah bersifat positif, negatif, atau netral. Polaritas mengacu pada apakah teks dalam pendapat, dokumen, atau kalimat memiliki unsur positif atau negatif (Amrullah et al., 2020).

Analisis sentimen, juga dikenal sebagai pemrosesan pendapat, adalah bidang penelitian di bidang teks *mining* yang mulai mendapat perhatian besar pada tahun 2013. Analisis sentimen adalah penelitian komputasional tentang opini, sentimen, dan emosi yang di ungkapkan dalam teks. *Opinion mining* dilakukan ketika sekumpulan dokumen teks berisi opini (atau sentimen) tentang suatu subjek. Tujuan *opinion mining* adalah untuk mengekstrak fitur dan komponen dari objek yang dikomentari pada setiap dokumen dan untuk menentukan apakah komentar tersebut positif, negatif, atau netral. *Opinion mining* juga merupakan bagian dari pekerjaan yang melakukan *review* dan berkaitan dengan perlakuan opini, sentimen, dan *subjektifitas* teks (Alita & Isnain, 2020).

Tujuan utama dari analisis sentimen adalah untuk mengklasifikasikan teks dalam sebuah kalimat atau dokumen dan menentukan apakah pendapat yang di ungkapkan dalam kalimat atau dokumen tersebut positif, negatif, atau netral. Analisis emosi juga dapat digunakan untuk mengekspresikan keadaan emosi seperti kesedihan, kegembiraan, dan kemarahan. Kita dapat mencari opini di internet tentang produk, merek, dan orang dan melihat apakah opini tersebut dinilai positif atau negatif (Asri et al., 2022).

Menurut Medhat et al dalam Wati Analisis sentimen adalah suatu bidang yang sedang berlangsung dalam penelitian berbasis teks. Analisis sentimen atau opini *mining* adalah kajian tentang cara untuk memecahkan masalah dari opini masyarakat, sikap dan emosi suatu entitas, di mana entitas tersebut dapat mewakili individu, peristiwa atau topik (Wati, 2016).

Dapat disimpulkan secara singkat bahwa analisis sentimen bertujuan untuk menganalisis suatu ulasan atau pendapat memiliki opini positif atau negatif terhadap suatu produk, layanan, ataupun aplikasi.

2. Klasifikasi

Metode klasifikasi banyak digunakan dalam berbagai bidang, seperti pendidikan, pemerintahan, kesehatan, teknologi, maupun sosial. Klasifikasi didefinisikan sebagai pekerjaan mengelompokkan suatu objek ke dalam kategori tertentu. Klasifikasi dapat dilakukan pada data kategori maupun bukan. Jika data bukan kategori maka harus diubah dalam bentuk kategori terlebih dahulu (Lutfiah & Habibah, 2022). Klasifikasi merupakan salah satu metode statistika yang mengelompokkan data sesuai karakteristiknya secara sistematis dalam kelas yang telah ditentukan. Klasifikasi merupakan metode penelitian ketika suatu objek data dinilai untuk dimasukkan ke dalam suatu kelas tertentu dari sejumlah kelas yang ada atau tersedia (Innassuraiya et al., 2022).

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari jumlah kelas yang tersedia. Klasifikasi melakukan pembangunan model berdasarkan data latih yang ada, kemudian menggunakan model tersebut untuk mengklasifikasikan pada data yang baru. Klasifikasi dapat didefinisikan sebagai pekerjaan yang melakukan pelatihan/pembelajaran terhadap fungsi target yang memetakan setiap set atribut (fitur) ke satu jumlah label kelas yang tersedia. Sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua data set dengan benar, tetapi tidak dapat dipungkiri bahwa kinerja sistem tidak bisa 100% benar sehingga sebuah sistem klasifikasi juga harus diukur kinerjanya. Umumnya,

pengukuran kinerja klasifikasi dilakukan dengan *confusion matrix* (Utomo & Mesran, 2020).

3. Pantai Tanjung Bira

Bulukumba terdapat pada koordinat antara 5°20" sampai 5°40" Lintang Selatan dan 119°50" sampai 120°28" Bujur Timur. Kabupaten ini berdekatan dengan Kabupaten Sinjai di sebelah utara, sebelah timur berdekatan dengan teluk Bone, sebelah selatan berdekatan dengan laut Flores, dan di sebelah barat dengan Kabupaten Bantaeng. Kabupaten Bulukumba terdiri dalam 10 kecamatan dan 24 kelurahan. Adapun kecamatan yang ada di Kabupaten Bulukumba yakni kecamatan Gantarang, kecamatan Ujungbulu, Kecamatan Ujung Loe, Kecamatan Bontobahari, kecamatan Bontotiro, Kecamatan Kajang dan kecamatan Herlang. Tanjung Bira merupakan salah satu lokasi pariwisata yang cukup terkenal di kabupaten Bulukumba Sulawesi Selatan. Selain mempunyai Keindahan Panorama alam berupa pantai pasir putih yang halus dan juga mempunyai keindahan ekosistem terumbu karang dan padalamun. Penelitian ini mengambil lokasi di kabupaten Bulukumba karena ketersediaan data yang ada di lokasi kajian (Anto & Makbul Syamsuri, 2023).

Total kunjungan wisatawan di Pantai Tanjung Bira terus mengalami peningkatan yang laju pertumbuhan tertinggi terjadi pada tahun 2018, yaitu sebesar 27,58%. Wisatawan di wilayah Pantai Tanjung Bira didominasi oleh wisatawan nusantara. Pada tahun 2018, jumlah wisatawan nusantara mencapai sebanyak 238.810 orang, sedangkan wisatawan mancanegara hanya sebanyak 3.555 orang. Perubahan jumlah wisatawan setiap tahun dapat dipengaruhi oleh pola keluar-masuk wisatawan, yaitu pada kondisi *low visit* (hari biasa), *condenset visit* (Sabtu), dan *peak visit* (minggu), ataupun pada hari libur (Inayah & Istiqomah, 2020).

4. Google maps

Google maps adalah layanan *mapping online* yang disediakan oleh *google*. Layanan ini dapat diakses melalui situs <http://maps.google.com>. Pada situs

tersebut kita dapat melihat informasi geografis pada hampir semua wilayah di bumi. Layanan ini interaktif, karena di dalamnya peta dapat digeser sesuai keinginan pengguna, mengubah tingkat *zoom*, serta mengubah tampilan peta. *Google maps* menyediakan peta yang sangat akurat, sistem pemetaannya juga sudah menyediakan pilihan peta biasa dan peta *satellite* (Sholeh et al., 2011).

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).

5. Algoritma *Regresi logistik*

Regresi logistik merupakan suatu metode analisis statistika yang digunakan untuk mendeskripsikan hubungan antara variabel terikat yang memiliki dua kategori atau lebih dengan satu atau lebih peubah bebas berskala kategori atau kontinu. Adapun *regresi logistik* dapat dibagi menjadi *regresi logistik* biner, *regresi logistik* multinomial dan *regresi logistik* ordinal (Tampil et al., 2017)

Menurut (Salim & Alfian, 2019) *Regresi logistik* merupakan salah satu metode klasifikasi yang sering digunakan. *Regresi logistik* biner digunakan saat variabel dependen merupakan variabel *dikotomus*. Menurut Hosmer, D.W. dan Lemeshow, S. 2000 di dalam Purwa, 2019 bahwa secara umum model *regresi logistik* adalah :

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}} \dots\dots\dots(1)$$

Di mana $\pi(x)$ merupakan nilai probabilitas dari $0 \leq \pi(x) \leq 1$, yang berarti bahwa *regresi logistik* menggambarkan suatu probabilitas. Dengan mentransformasikan $\pi(x)$ pada persamaan di atas dengan transformasi *logit* $g(x)$, di mana :

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) \dots\dots\dots(2)$$

Maka diperoleh bentuk *logit* :

$$g(x) = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k} \dots\dots\dots(3)$$

6. *Preprocessing* Data

Preprocessing mengacu metode standar yang digunakan dalam studi analisis sentimen pada teks bahasa Indonesia. Teknisnya terdiri dari empat langkah, yaitu penghapusan *cleaning*, *case folding*, *stopword* dan *tokenize*. Proses-proses ini telah diverifikasi dan menjadi standar umum dalam analisis sentimen (Khomsah & Aribowo, 2017). Secara umum, *preprocessing* dalam penelitian ini terdiri beberapa tahapan:

- a) *Cleaning*, tahapan untuk penghapusan karakter seperti tanda baca, menghapus *emoticon*, menghilangkan simbol yang tidak *releval* serta mengkonversi keseluruhan teks dalam dokumen menjadi seluruhnya huruf kecil (*lowercase*).
- b) *Casefolding*, 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.
- c) *Stopwords*, mengambil kata-kata penting dari hasil token dengan menghapus sekelompok kata umum yang sering kali dianggap tidak memberikan kontribusi signifikan terhadap makna utama dalam teks dan cenderung muncul dengan frekuensi tinggi.
- d) *Tokenize*, pemotongan *string input* berdasarkan tiap kata yang menyusunnya.

Algoritma dalam *pseudocode* di atas akan bekerja mengkonversi kosakata slang dalam satuan pengecekan per-kalimat. Kalimat yang mengandung kosakata slang akan dikirim ke fungsi *conv_slang* bersama dengan kamus *slang dict*. Pada fungsi *conv_slang*, kalimat yang dikirim akan dipisahkan per-kata, kemudian dicocokkan apakah terdaftar dalam kamus slang tersebut. Jika kalimat yang dikirim menemukan padanan kata dalam kamus slang maka akan diganti

dengan kata baku sesuai dalam kamus slang. Kalimat baru yang sudah bersih akan dirangkai kembali di bagian akhir fungsi.

B. Penelitian Terkait

Peneliti memperoleh banyak inspirasi dan referensi untuk penyusunan skripsi ini dari peneliti sebelumnya, terkait dengan latar belakang masalah pada skripsi ini. Penelitian sebelumnya yang terkait meliputi :

Pertama, Skripsi yang ditulis oleh (Rahmaningrum, 2019) yang berjudul “Klasifikasi Sentimen Terhadap Review Layanan Hotel Bintang Tiga Di Surabaya Pada Situs Traveloka Menggunakan Naïve Bayes Classifier (Nbc) Dan *Regresi logistik* Biner” Pada Hasil Penelitian Menunjukkan bahwa Perbandingan metode antara Naïve Bayes Classifier (NBC) dan *Regresi logistik* Biner (RLB) menunjukkan bahwa metode RLB dengan SMOTE lebih baik jika dibandingkan dengan metode NBC yang mana pada data *testing* untuk nilai AUC Favehotel sebesar 0,84 dan AUC Gunawangsa sebesar 0,82.

Kedua, jurnal yang ditulis oleh (Lengkong et al., 2021) yang berjudul “Analisis Sentimen Penerapan Psbb Di Dki Jakarta Dan Dampaknya Terhadap Pergerakan Ihsg” Pada penelitian ini menunjukan bahwa IHSG akan cenderung turun karena para investor di pasar modal lebih memilih untuk menarik sahamnya. Kemudian, hasil perbandingan klasifikasi menunjukkan dalam kasus ini metode *regresi logistik* memiliki akurasi tertinggi. Dengan perbedaan yang tidak besar, posisi selanjutnya ditempati metode k-nearest neighbor dan naive bayes.

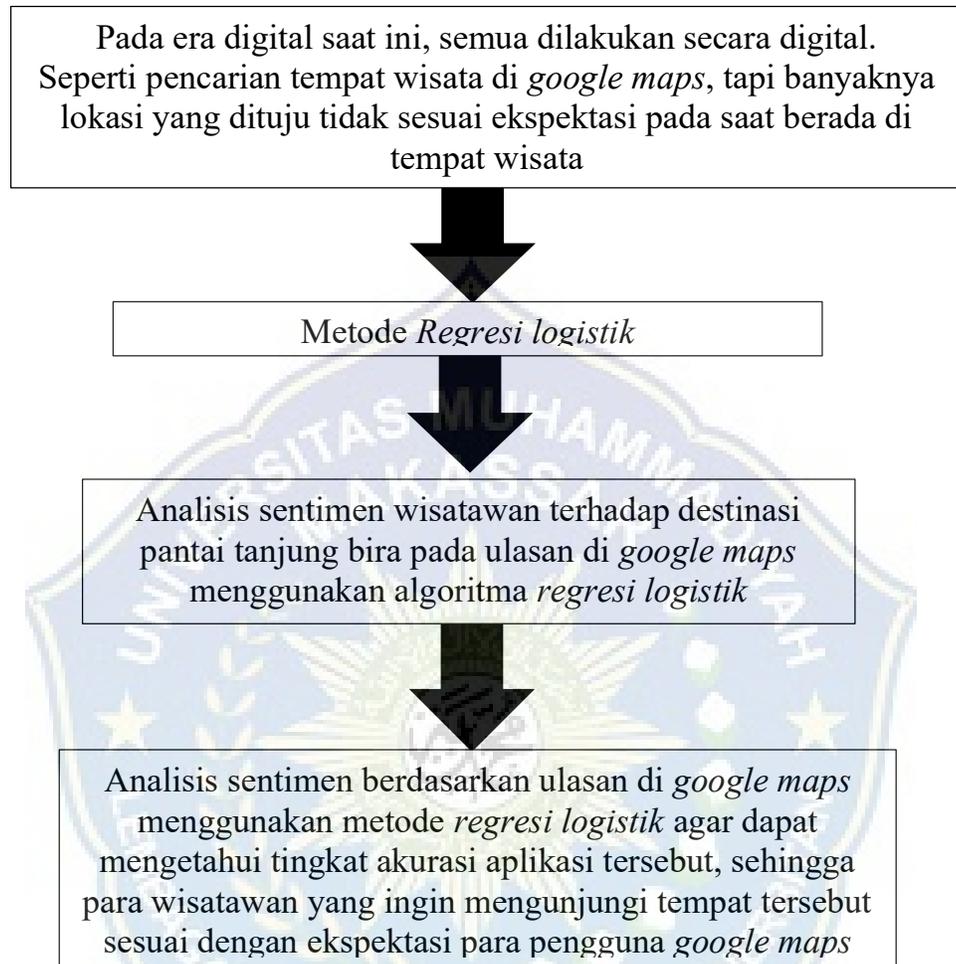
Ketiga, jurnal yang ditulis oleh (Averina et al., 2022) yang berjudul “Analisis Sentimen Multi-Kelas Untuk Film Berbasis Teks Ulasan Menggunakan Model *Regresi logistik*” Pada penelitian ini menunjukkan bahwa prediksi sentimen dari suatu ulasan model *Regresi logistik* menghasilkan akurasi klasifikasi terbaik sebesar 83% menggunakan CountVectorizer sebagai input fitur dengan menerapkan dimensionality reduction.

Keempat, Skripsi yang ditulis oleh (Shiddicky, 2022) yang berjudul “Analisis Sentimen Masyarakat Terhadap Kebijakan Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Metode Logistic Regression” Pada

Penelitian ini menunjukkan bahwa sentimen terhadap vaksinasi Covid Twitter telah berhasil dibangun dengan menerapkan algoritma Logistic Regression, di mana akurasi tertinggi didapatkan di angka 82% data dev artinya algoritma ini cukup baik dalam melakukan klasifikasi, tapi tidak diimbangi dengan F1-Score yang cukup baik yakni 42% terhadap data dev, hal ini terjadi karna terjadinya tidak keseimbangan data atau imbalance pada data train atau data pembelajaran pada saat membangun modelnya, maka dilakukan proses menyeimbangkan data serta dalam penelitian ini dan juga melakukan Tuning Hyperparameter yang bertujuan untuk meningkatkan tingkat akurasi dari F1-Score dari proses sebelumnya, setelah dilakukan penyeimbangan data dan menerapkan parameter terbaik yang diperoleh ke dalam model, metode Logistic Regression menghasilkan nilai akurasi yang cukup baik dari pada penelitian sebelumnya di mana F1-score 60% dan akurasi 67% terhadap data *test* hasil ini membuktikan bahwa model yang dibangun cukup handal dalam melakukan klasifikasi.

Kelima, Jurnal yang ditulis oleh (Prabowo et al., 2023) yang berjudul “Analisis Sentimen Pengguna Twitter Terhadap Kasus COVID-19 di Indonesia Menggunakan Metode *Regresi logistik* Multinomial” Dari hasil penelitian yang dilakukan dapat disimpulkan bahwa sistem dapat mengklasifikasi tweet terhadap kasus COVID-19 di Indonesia secara otomatis ke dalam sentimen positif, netral, dan negatif. Adapun data latih yang digunakan sebanyak 870 yang telah dilabel dengan masing-masing data sentimen positif 290, sentimen netral 290, dan sentimen negatif 290 yang diuji menggunakan threshold 10-cross validation. Sistem ini memiliki akurasi sebesar 64% dengan precision positif sebesar 85%, netral sebesar 56%, dan negatif sebesar 53% serta recall positif sebesar 74%, netral sebesar 67%, dan negatif sebesar 50%.

C. Kerangka Fikir



Gambar 1. Kerangka Berfikir

BAB III

METODE PENELITIAN

A. Tempat dan Waktu Penelitian

1. Tempat Penelitian

Penelitian ini dilakukan di laboratorium Fakultas Teknik Prodi Informatika Universitas Muhammadiyah Makassar yang dalam penelitian ini pusat datanya adalah ulasan aplikasi di *google maps*. Waktu penelitian ini ≤ 3 bulan, terhitung mulai pada bulan November-Januari 2024.

B. Alat Dan Bahan

Alat yang dipakai dalam penelitian ini adalah seperangkat komputer yang dilengkapi dengan perangkat keras dan perangkat lunak pendukung. Sementara itu, bahan yang digunakan ialah data yang diambil berasal dari ulasan di *google maps*. Oleh karena itu, untuk melakukan analisis sentimen dalam penelitian ini, beberapa alat dan bahan menjadi kebutuhan untuk melakukan analisis sentimen yaitu :

1. Kebutuhan *Hardware* (Perangkat Keras)

Tabel 1. Spesifikasi Komputer

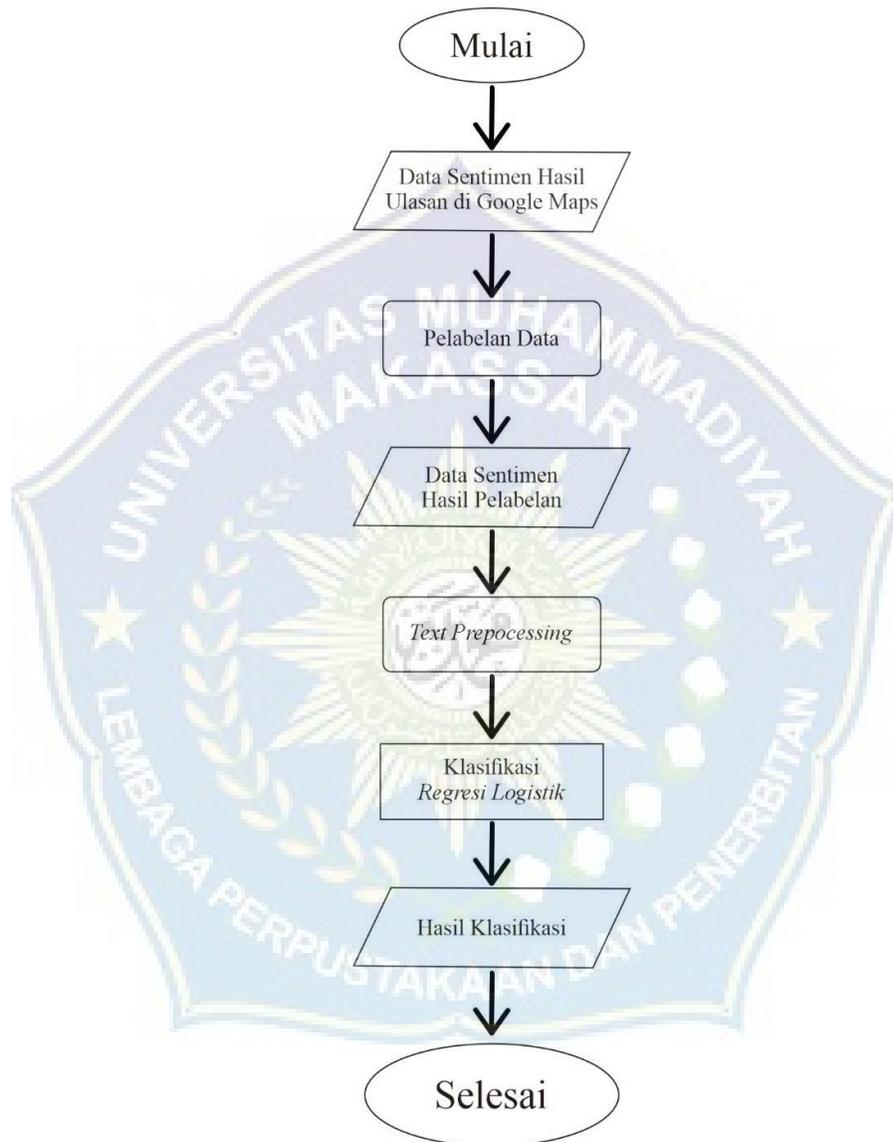
Spesifikasi	Keterangan
<i>Processor</i>	<i>Intel® Core i5-3320M</i>
<i>Graphics Card</i>	<i>Intel® UHD Graphics 4000</i>
<i>Operation System</i>	<i>Windows 10 Pro</i>
<i>Storage</i>	<i>128GB SSD + 300GB HDD</i>
<i>RAM</i>	<i>6GB</i>

2. Kebutuhan *Software* (Perangkat Lunak)

- *Microsoft Windows 10 64-bit*
- *Microsoft Office*
- *Phyton*
- *Google maps*
- *Google Colab*

C. Perancangan Sistem

Untuk mempermudah dalam proses perancangan, peneliti membuat perancang sistem dengan sebuah *Flowchart*, seperti gambar di bawah ini :



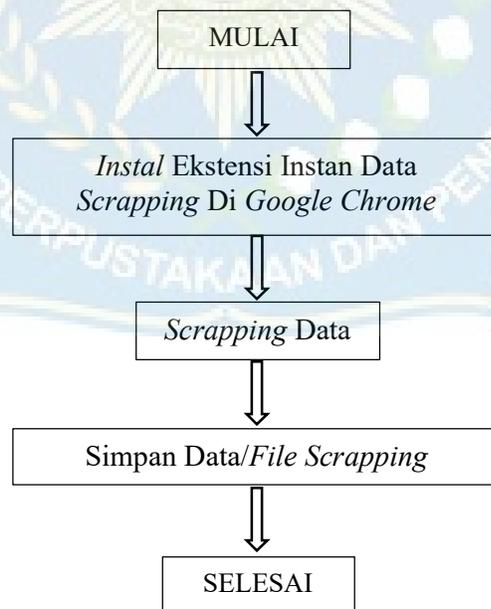
Gambar 2. *Flowchart System*

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

1. Pengumpulan Data

Proses pengumpulan data pada penelitian adalah langkah penting dalam melakukan sebuah penelitian, karena data yang dihasilkan akan digunakan sebagai dasar untuk menjawab pertanyaan penelitian dan menguji hipotesis. Proses pengumpulan data pada penelitian ini yaitu mengambil data primer. Sumber data yang diperoleh secara langsung dari sumber asli disebut sebagai data primer. Penggunaan data primer umumnya untuk kebutuhan menghasilkan informasi yang mencerminkan kebenaran sesuai dengan kondisi faktual, sehingga informasi yang dihasilkan dapat berguna dalam pengambilan keputusan.

Pada tahapan ini dilakukan pengambilan data dari ulasan pengguna aplikasi *google maps* dengan objek pantai tanjung bira dengan metode *scrapping* data dengan menggunakan ekstensi *instans data scrapping* di *google chrome*.



Gambar 3. Tahapan Pengumpulan Data

2. Pelabelan Data

Setelah data berhasil dikumpulkan menjadi sebuah *dataset*, tahap selanjutnya adalah pelabelan. Pelabelan di sini ditujukan untuk membagi data ke dalam beberapa kelas sentimen yang akan digunakan dalam penelitian. Jumlah kelas sentimen yang banyak digunakan adalah dua dan tiga kelas, yaitu negatif, netral dan positif. Tujuan dari proses pelabelan ini adalah membagi *dataset* menjadi 2 bagian, yaitu menjadi *training* data dan *testing* data. *Training* data adalah data yang digunakan untuk melatih sistem agar mampu mengenali pola yang sedang dicari, sedangkan *testing* data adalah data yang digunakan untuk menguji hasil pelatihan yang sudah dilakukan.

3. Tahap *Preprocessing*

Setelah dilakukan pelabelan data, tahap selanjutnya yang harus dilalui adalah *preprocessing*. Tahap ini adalah tahapan di mana data disiapkan agar menjadi data yang siap untuk dianalisis. Ada beberapa tahap dalam *preprocessing* ini, antara lain *cleaning*, *case folding*, *stopword*, *tokenize*.



Gambar 4. Tahapan *Prepossessing* Data

4. Pengklasifikasian

Metode yang digunakan pada tahap ini adalah metode *regresi logistik*. *Regresi logistik* merupakan metode yang menghubungkan antara variabel dependen yang bersifat *kategorik* dengan variabel independen. Model *regresi logistik* dengan variabel dependen yang terdiri dari dua kategori disebut model *regresi logistik biner (dikotomus)*. Model *regresi logistik biner* digunakan untuk menganalisis hubungan antara satu variabel dependen (Y) dengan satu atau lebih variabel independen (X). Variabel dependennya berupa data kualitatif dikotomi yang bernilai 1 untuk menyatakan keberadaan sebuah karakteristik dan bernilai 0 untuk menyatakan ke tidak beradaan sebuah karakteristik.

D. Teknik Pengujian Sistem

1. Pengujian *Confusion Matrix*

Confusion matrix merupakan alat pengukuran kinerja metode prediksi dengan menghitung tingkat kebenaran proses klasifikasi. Hasil pengujian ini berupa perbandingan akurasi prediksi kemunculan hari baik keluaran ketiga metode *fuzzy* yang diimplementasikan pada aplikasi pengujian. Akurasi prediksi kemunculan hari baik dihitung menggunakan *Confusion matrix*. Jumlah data prediksi dari sistem dan jumlah data prediksi dari pakar sebagai variabel masukan *Confusion matrix*, selanjutnya oleh sistem memberikan masing-masing metode *fuzzy* nilai *Accuracy*, *Precision*, *Recall* dan *F-1 Score*.

		<i>Observed</i>	
		<i>True</i>	<i>False</i>
<i>Predicted Class</i>	<i>True</i>	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
	<i>False</i>	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Gambar 5. *Confusion Matrix*

Dimana :

1. TP adalah *True* Positif, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.
2. TN adalah *True* Negatif, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
3. FN adalah *False* Negatif, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.
4. FP adalah *False* Positif, yaitu jumlah data positif namun terklasifikasi salah oleh sistem.

Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh dengan persamaan :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \dots\dots\dots(1)$$

Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasi secara benar dibagi dengan total data yang diklasifikasi positif, Presisi dapat diperoleh dengan persamaan

$$Presisi = \frac{TP}{TP+F} \times 100\% \dots\dots\dots(2)$$

Sementara itu, *recall* menunjukkan beberapa persen data kategori positif yang terklasifikasi dengan benar oleh sistem

$$Recall = \frac{TP}{TP+FN} \times 100\% \dots\dots\dots(3)$$

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

$$F1_Score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \dots\dots\dots(4)$$

E. Teknik Analisis Data

Adapun langkah-langkah untuk menganalisis data dalam penelitian ini adalah sebagai berikut :

1. Pengumpulan Data

Pengumpulan data adalah pekerjaan penting dalam meneliti. Istilah “pengumpulan” juga merujuk pada pengertian “perolehan” dan “pengolahan awal” yang berhubungan dengan konsep sumber atau asal. Pengumpulan data menentukan tingkat keberhasilan pengolahan data selanjutnya. Pengumpulan data adalah suatu usaha sistematis dengan prosedur terstandar untuk memperoleh ukuran tentang variabel dan jawaban atas pertanyaan penelitian.

Data yang dikumpulkan dapat berupa teks atau dokumen atau ulasan yang mengandung opini atau sentimen. Sumber data dapat berasal dari berbagai media sosial atau platform digital atau wawancara.

2. Klasifikasi Sentimen

Dalam analisis sentimen, teks atau dokumen diklasifikasikan ke dalam kategori sentimen positif, negatif, atau netral. Hal ini dapat dilakukan dengan menggunakan metode pengolahan bahasa alami atau menggunakan perangkat lunak analisis sentimen.

3. Perhitungan frekuensi

Setelah teks atau dokumen diklasifikasikan ke dalam kategori sentimen, frekuensi masing-masing kategori dapat dihitung. Frekuensi dapat dihitung dalam bentuk absolut (jumlah sentimen positif, negatif, atau netral) atau relatif (persentase sentimen positif, negatif, atau netral dari total dokumen).

4. Interpretasi hasil

Hasil analisis dapat diinterpretasikan untuk memberikan pemahaman tentang sentimen atau opini yang dikemukakan dalam dokumen atau teks. Hasil analisis dapat memberikan wawasan tentang pandangan masyarakat terhadap suatu topik atau produk tertentu.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

Pada penelitian analisis sentimen mengenai ulasan wisatawan terhadap destinasi pantai tanjung bira pada ulasan di *google maps* menggunakan algoritma *regresi logistik*. Tujuannya adalah untuk mengetahui apakah ulasan wisatawan terhadap destinasi pantai tanjung bira pada ulasan *google maps* tergolong positif, negatif, atau netral.

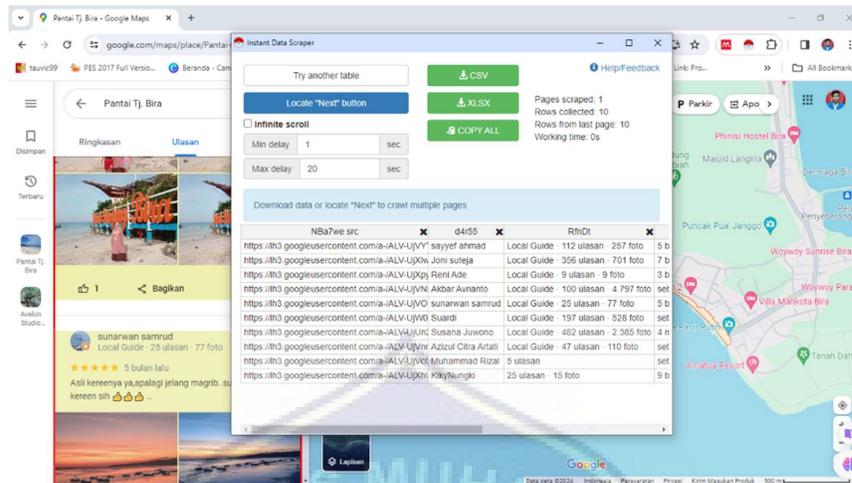
Untuk Penelitian ini data di dapatkan dengan *scrapping* data ulasan wisatawan terhadap destinasi pantai tanjung bira pada *google maps*.

A. Pengumpulan Data

Pengumpulan data ulasan wisatawan terhadap destinasi pantai tanjung bira pada *google maps* menggunakan teknik web *scrapping*. Proses pengambilan data ulasan di *google maps* menggunakan scrapper data dimulai dengan *install* ekstensi *Instant Data Scraper* di *Browser Google Chrome*.



Gambar 6. Proses *Instal* Ekstensi *Instant Data Scraper*



Gambar 7. Proses *Scrapping* Data

Setelah memutuskan data yang akan di ambil, selanjutnya untuk memulai proses *crawling* yaitu menekan tombol *start crawling* pada *Instant Data Scrapper* untuk memulai proses pengambilan data ulasan. Ketika proses *crawling* selesai, maka data ulasan wisatawan akan terkumpul menjadi file Excel.

Tabel 2. *Dataset* Hasil *Scapping*

No	Ulasan
1	pantai tercantik yang pernah aku liat
2	pantainya Indah
3	Sangat menyukai pantai, apalagi pasirnya putih dan bersih
4	Tempatnya bagus
5	Sangat indah pantai dengan pasir putihnya semoga destinasi wisata ini terjaga kebersihannya...
6	Kalo disini enaknya bermalam..., karena mandi dipantai lebih asyik di pagi hari. Jangan lupa bawa peralatan snorkling..
7	Pemandangan sangat indah... keren buat liburan sama teman teman dan keluarga

8	Pasir putihnya lembut..., ada permainan berbayar bagi penggemar kegiatan pantai...
9	Sering dibicarakan orang! insyaallah bakal kesini
10	Pantai pasir putih, banyak kegiatan yang bisa di lakukan di pantai bira bersama keluarga.
...
991	semoga ya hari ini
992	tahun depan ke sini ramaiiiiiiiii
993	Biasa
994	biasa saja
995	jalan menuju tempat ini cukup berbatu tapi bagus viewnya banyak pohon dan rindang sejuk banget .
996	kemarin pa liburan datangnya ke sini
997	teman aku tdk dapat deh jadinya haru gantin tapi aku rekomendasikanlah kapan klau liburan bisa ke sini
998	harga masuk terjangkau pelayanan ok secara umum baguslah
999	seruuu dan puass tetapi yg jual makanan kebet mau umroh kek na
1000	seruuu dan puass tetapi yg jual makanan kebet mau umroh kek na

B. Pelabelan Data

Proses pelabelan data sentimen dilakukan secara manual dengan tujuan mengidentifikasi pola dan ciri-ciri dalam teks yang mencerminkan sentimen positif, negatif, netral. Ulasan atau teks diberi label "positif" jika menyuarakan kepuasan, "negatif" jika mengekspresikan ketidakpuasan, dan "netral" jika tidak ada sentimen kuat.

Proses pelabelan data adalah sebagai berikut :

Tabel 3. Tahap Pelabelan Data

Ulasan	Label
Kalo disini enakya bermalam.., karena mandi dipantai lebih asyik di pagi hari. Jangan lupa bawa peralatan snorkling..	Positif
tdk ada tempat sampah yg disediakan jd pantas saja orang2 asal buang sampah sembarang saja di pinggir pantai.	Negatif
Pantai bira sdh buka dg penerapan new normal dan saat ini masih gratis masuk sampai bln akhir bulan	Netral

C. Preprocessing Data

1. Cleaning

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*. Tabel berikut merupakan hasil dari tahap *cleaning* :

Tabel 4. Proses *Cleaning* Data

No	Sebelum	Sesudah
1	yang bikin jelek disini ada banyak orang jualan dipinggiran pantai dan banyaknya sampah berserakan disepanjang pantai yang tidak segera dibersihkan begitu juga	yang bikin jelek disini ada banyak orang jualan dipinggiran pantai dan banyaknya sampah berserakan disepanjang pantai yang tidak segera dibersihkan begitu juga
...		

2	kurang bersih karena terlihat beberapa sampah yang berserakan. Ingat, kita satu bumi, jangan buang sampah sembarangan! Cintai bumi. ♡🌍😊👍 ...	kurang bersih karena terlihat beberapa sampah yang berserakan. Ingat kita satu bumi jangan buang sampah sembarangan. Cintai bumi
3	pasirnya sekarang sudah habis,,, berubah menjadi batu karang ...	pasirnya sekarang sudah habis berubah menjadi batu karang
4	Coba diving deh, bagus banget!!	coba diving deh bagus banget
5	The best lah pasir nya putihCocok untuk bulan madu	the best lah pasir nya putih cocok untuk bulan madu
6	pantai yang indah love to you pantai bira	pantai yang indah sepanjang masa love to you pantai bira
7	Indah...landai... hotel juga ok...	indahlandai hotel juga ok
8	keren	keren
9	harganya mahal...tiket masuk dewasa 15k...anak 6k....msh tergolong mahal. Parkir di area bayar lg....masuk ke pantai sebelahnya bayar lg padahal 1 ...	harganya mahal...tiket masuk dewasa kanak kmsk tergolong mahal parkir di area bayar lgmasuk ke pantai sebelahnya bayar lg padahal 1 ...

10	Pantai Tanjung Bira... pasirnya putih..., semoga tetap bersih...	pantai tanjung bira pasirnya putih semoga tetap bersih
----	--	--

2. Case Folding

Case folding adalah proses konversi teks dalam dokumen atau data menjadi format yang seragam dengan mengubah semua huruf menjadi huruf kecil atau besar, tergantung pada aturan tertentu. Tujuan dari *case folding* adalah untuk menyederhanakan perbandingan teks dalam proses pemrosesan dan pencarian data. Dalam beberapa kasus, *case folding* dilakukan dengan mengubah semua huruf dalam teks menjadi huruf kecil (*lowercase*). Tabel berikut merupakan proses dari *Case folding* :

Tabel 5. Proses Case Folding

No	Sebelum	Sesudah
1	Bulukumba keren	bulukumba keren
2	Kecil area pasirnya dan sedikit kotor sampahnya	kecil area pasirnya dan sedikit kotor sampahnya
3	Tempat terpesona jauh dari pariwisata massal Orang-orang yang ramah dan membantu mereka selalu membuat kami merasa aman Pasir putih dan keheningan surga.	tempat terpesona jauh dari pariwisata massal orang-orang yang ramah dan membantu mereka selalu membuat kami merasa aman pasir putih dan keheningan surga.
4	Banyak sekali tempat indah di kawasan ini	banyak sekali tempat indah di kawasan ini

5	Pantai berpasir putih yang indah cocok untuk berkemah	pantai berpasir putih yang indah cocok untuk berkemah
6	Pantai yang indah	pantai yang indah
7	fasilitas MCK nya sangat kurangjadi tolong pengelola sediakan MCK nya lebih banyak lagi	fasilitas mck nya sangat kurangjadi tolong pengelola sediakan mck nya lebih banyak lagi
8	Pantainya banyak sampah	pantainya banyak sampah
9	Pelayanan hotel jelek Ac tidak dingintidak ada air	pelayanan hotel jelek ac tidak dingintidak ada air
10	Transpotasi kendaraan umum belum ada	transpotasi kendaraan umum belum ada

3. *Stopword*

Stopword adalah sekelompok kata umum yang sering kali dianggap tidak memberikan kontribusi signifikan terhadap makna utama dalam teks dan cenderung muncul dengan frekuensi tinggi. Tahapan Penghapusan *stopword* merupakan langkah umum dalam *pra-pemrosesan* teks dalam pemrosesan *Natural Language Toolkit* (NLTK), bertujuan untuk meningkatkan efisiensi pemrosesan dan meningkatkan fokus pada kata-kata yang dianggap lebih informatif. Meskipun penghapusan *stopwords* umumnya meningkatkan kualitas analisis, keberadaan *stopwords* tertentu dapat memiliki arti penting dalam beberapa konteks dan perlu dipertimbangkan secara kontekstual. Tabel berikut merupakan hasil dari tahap *stopword* :

Tabel 6. *Proses Stopword*

No	Sebelum	Sesudah
1	pantai tercantik yang pernah aku liat	pantai tercantik liat
2	sangat menyukai pantai apalagi pasirnya putih dan bersih	menyukai pantai pasirnya putih bersih
3	sangat indah pantai dengan pasir putihnya semoga destinasi wisata ini terjaga kebersihannya	indah pantai pasir putihnya semoga destinasi wisata terjaga kebersihannya
4	kalo disini enakya bermalam karena mandi dipantai lebih asyik di pagi hari jangan lupa bawa peralatan snorkling	kalo enakya bermalam mandi dipantai asyik pagi lupa bawa peralatan snorkling
5	pemandangan sangat indah keren buat liburan sama teman dan keluarga	pemandangan indah keren liburan teman teman keluarga
6	pasir putihnya lembut ada permainan berbayar bagi penggemar kegiatan pantai	pasir putihnya lembut permainan berbayar penggemar kegiatan pantai
7	sering dibicarakan orang insyaallah bakal kesini	dibicarakan orang insyaallah kesini

8	pantai pasir putih banyak kegiatan yang bisa di lakukan di pantai bira bersama keluarga	pantai pasir putih kegiatan lakukan pantai bira keluarga
9	tempat terkeren pasir putih laut biru banyak tempat wisata baru pokoknya paling favorit	terkeren pasir putih laut biru wisata pokoknya favorit
10	pasir pantainya bersih dan halussangat memanjakan mata	pasir pantainya bersih halussangat memanjakan mata

4. Tokenize

Tokenisasi 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. Tabel berikut merupakan hasil dari tahap tokenisasi :

Tabel 7. Proses Tokenize

No	Sebelum	Sesudah
1	pantai tercantik liat	['pantai', 'tercantik', 'liat']
2	pantainya indah	['pantainya', 'indah']
3	menyukai pantai pasirnya putih bersih	['menyukai', 'pantai', 'pasirnya', 'putih', 'bersih']
4	tempatnya bagus	['tempatnya', 'bagus']

5	indah pantai pasir putihnya semoga destinasi wisata terjaga kebersihannya	['indah', 'pantai', 'pasir', 'putihnya', 'semoga', 'destinasi', 'wisata', 'terjaga', 'kebersihannya']
6	kalo enakya bermalam mandi dipantai asyik pagi lupa bawa peralatan snorkling	['kalo', 'enaknya', 'bermalam', 'mandi', 'dipantai', 'asyik', 'pagi', 'lupa', 'bawa', 'peralatan', 'snorkling']
7	pemandangan indah keren liburan teman teman keluarga	['pemandangan', 'indah', 'keren', 'liburan', 'teman', 'teman', 'keluarga']
8	pasir putihnya lembut permainan berbayar penggemar kegiatan pantai	['pasir', 'putihnya', 'lembut', 'permainan', 'berbayar', 'penggemar', 'kegiatan', 'pantai']
9	dibicarakan orang insyaallah kesini	['dibicarakan', 'orang', 'insyaallah', 'kesini']
10	pantai pasir putih kegiatan lakukan pantai bira keluarga	['pantai', 'pasir', 'putih', 'kegiatan', 'lakukan', 'pantai', 'bira', 'keluarga']

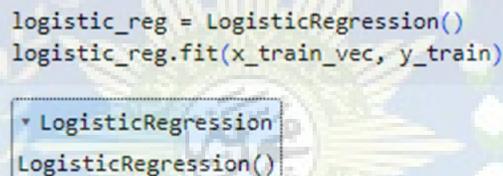
D. Klasifikasi *Regresi logistik*

Klasifikasi merupakan Teknik komputasi yang mengelompokkan data berdasarkan keterikatan data. Pengelompokan data di mana data tersebut mempunyai kelas label atau target. Dan dalam penelitian ini menggunakan metode *regresi logistik* untuk analisis atau klasifikasi data. Dan metode ini di menggunakan *library scikit-learn (sklearn)* di *Python*. Cara kerja metode *regresi logistik* yaitu dengan menghitung jumlah label, menghitung jumlah kasus per kelas, kalikan semua variabel kelas dan bandingkan hasil per kelas. Metode ini untuk menghitung berapa nilai *Accuracy, precision, recall, f1_score*.

Pada tahap klasifikasi pada penelitian ini, data *split* dibagi menjadi tiga yaitu :

1. 10% data *test* 90% data *training*
2. 20% data *test* 80% data *training*
3. 30% data *test* 70% data *training*

Lalu untuk proses klasifikasi menggunakan *library Regresi logistik*, proses selanjutnya melatih klasifikasi dengan data latih yang kemudian data modelnya disimpan dengan nama dan format "*regresi_logistic_model.joblib*". Lalu *load* kembali data model, yang di lanjut dengan dilakukan *test* klasifikasi dengan menggunakan data *test*. 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.



```
logistic_reg = LogisticRegression()  
logistic_reg.fit(x_train_vec, y_train)
```

▼ LogisticRegression
LogisticRegression()

Gambar 8. Proses Pembuatan Model Regresi Logistik

E. Evaluasi Model Regresi logistik

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 *regresi logistik* adalah sebagai berikut :

1. *Split* data 90% data *training* 10% data *test*

Dengan *split* data 10% data *test* dan 90% data *training* mendapatkan nilai *Accuracy* sebesar 70%.

Tabel 8. Hasil Pengujian Data (90% data *training* 10% data *test*)

	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>	<i>Support</i>
Negatif	0.77	0.74	0.75	31
Netral	0.64	0.55	0.59	29
Positif	0.69	0.80	0.74	30
<i>Accuracy</i>			0.70	90
<i>Macro Average</i>	0.70	0.70	0.70	90
<i>Weighted Average</i>	0.70	0.70	0.70	90

Dengan *split* data 10% data *test* dan 90% data *training* mendapatkan nilai *Accuracy* sebesar 70%.

2. *Split* data 80% data *training* 20% data *test*

Dengan *split* data 20% data *test* dan 80% data *training* mendapatkan nilai *Accuracy* sebesar 70%.

Tabel 9. Hasil Pengujian Data (80% data *training* 20% data *test*)

	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>	<i>Support</i>
Negatif	0.75	0.72	0.74	58
Netral	0.64	0.55	0.59	55
Positif	0.69	0.80	0.74	67
<i>Accuracy</i>			0.70	180
<i>Macro Average</i>	0.70	0.69	0.68	180
<i>Weighted Average</i>	0.70	0.70	0.69	180

3. *Split* data 70% data *training* 30% data *test*

Dengan *split* data 30% data *test* dan 70% data *training* mendapatkan nilai *Accuracy* sebesar 68%.

Tabel 10. Hasil Pengujian Data (70% data *training* 30% data *test*)

	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>	<i>Support</i>
Negatif	0.69	0.72	0.73	89
Netral	0.64	0.55	0.52	77
Positif	0.71	0.81	0.75	104
<i>Accuracy</i>			0.69	270
<i>Macro Average</i>	0.68	0.67	0.67	270
<i>Weighted Average</i>	0.68	0.69	0.68	270

Dengan demikian menunjukkan bahwa analisis sentimen ulasan wisatawan pantai tanjung bira pada *google maps* dengan *regresi logistik* mempunyai tingkat kinerja yang cukup bagus karena memiliki nilai *accuracy* di atas 50%. *Macro average* adalah salah satu metode atau cara yang digunakan untuk menghitung rata-rata dari metrik evaluasi pada setiap kelas. Sedangkan *weighted average* adalah metode perhitungan rata-rata yang memberikan bobot pada setiap nilai yang akan dihitung berdasarkan faktor tertentu.

```

Accuracy: 0.7
Precision: 0.6974603174603174
Recall: 0.6978865406006675
F1 Score: 0.6950508305699562
Classification Report pada Data Uji REGRESI LOGISTIK:
      precision    recall  f1-score   support

   Negatif         0.77     0.74     0.75         31
    Netral         0.64     0.55     0.59         29
    Positif         0.69     0.80     0.74         30

 accuracy         0.70
 macro avg         0.70     0.70     0.70         90
weighted avg         0.70     0.70     0.70         90
Accuracy: 0.7
    
```

Gambar 9. Hasil Pengujian Data (90% data *training* 10% data *test*)

```

Accuracy: 0.7
Precision: 0.6996732026143789
Recall: 0.6875122818509335
F1 Score: 0.6843971631205674
Classification Report pada Data Uji REGRESI LOGISTIK:
      precision    recall  f1-score   support

   Negatif      0.75      0.72      0.74        58
    Netral      0.67      0.47      0.55        55
    Positif      0.68      0.87      0.76        67

 accuracy          0.70
 macro avg         0.70      0.69      0.68
 weighted avg     0.70      0.70      0.69

```

Accuracy: 0.7

Gambar 10. Hasil Pengujian Data (80% data *training* 20% data *test*)

```

Accuracy: 0.6888888888888889
Precision: 0.6804231126412829
Recall: 0.6710985643569914
F1 Score: 0.6679042930164009
Classification Report pada Data Uji REGRESI LOGISTIK:
      precision    recall  f1-score   support

   Negatif      0.69      0.76      0.73        89
    Netral      0.64      0.44      0.52        77
    Positif      0.71      0.81      0.75       104

 accuracy          0.68
 macro avg         0.68      0.67      0.67
 weighted avg     0.68      0.69      0.68

```

Accuracy: 0.6888888888888889

Gambar 11. Hasil Pengujian Data (70% data *training* 30% data *test*)

F. Pengujian *Confusion Matrix*

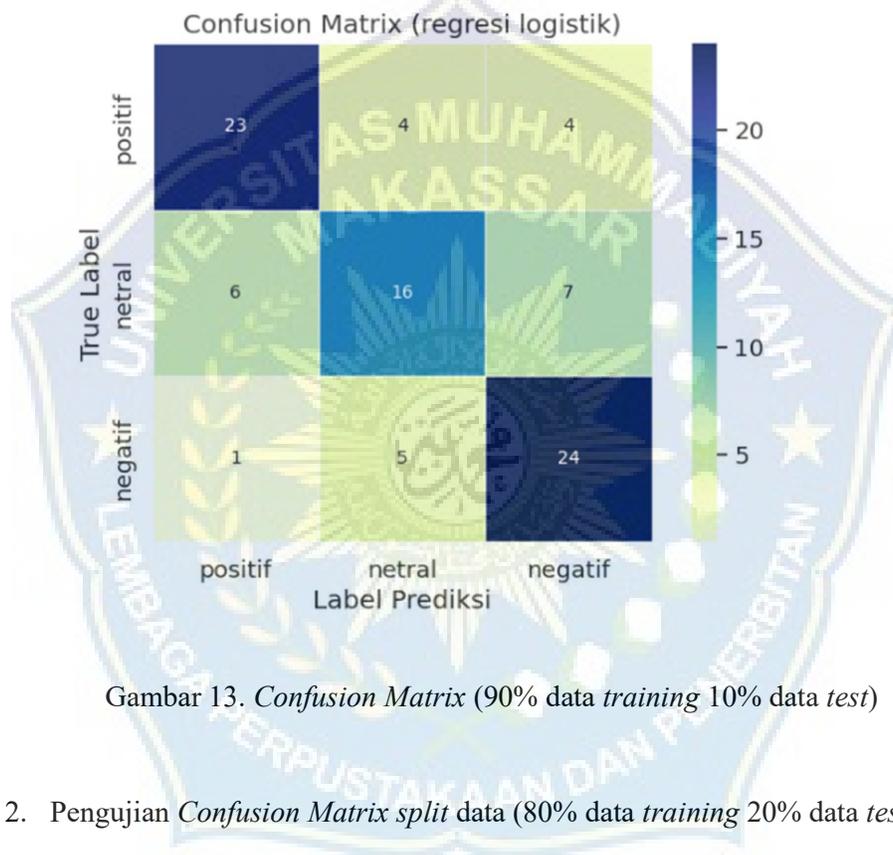
Untuk menghitung dan mengetahui performa dari algoritma klasifikasi yang digunakan yaitu algoritma *regresi logistik*, maka perlu dilakukan pengujian model klasifikasi. Hasil performa dari algoritma ditampilkan dalam bentuk nilai *accuracy* dan bentuk tabel *confusion matrix*, salah satu metode yang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

1. Pengujian *Confusion matrix split data (90% data training 10% data test)*

Confusion Matrix pada Data Uji regresi logistik:

```
[[23 4 4]
 [ 6 16 7]
 [ 1 5 24]]
```

Gambar 12. Hasil Pengujian *Confusion Matrix (90% data training 10% data test)*



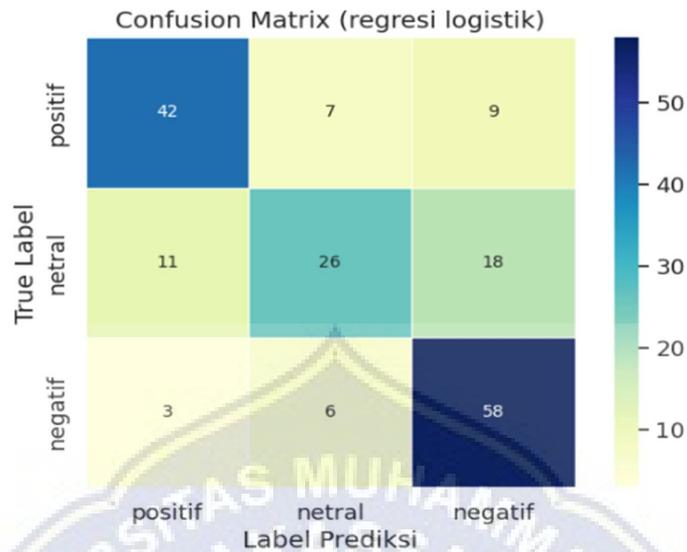
Gambar 13. *Confusion Matrix (90% data training 10% data test)*

2. Pengujian *Confusion Matrix split data (80% data training 20% data test)*

Confusion Matrix pada Data Uji regresi logistik:

```
[[42 7 9]
 [11 26 18]
 [ 3 6 58]]
```

Gambar 14. Hasil Pengujian *Confusion Matrix (80% data training 20% data test)*

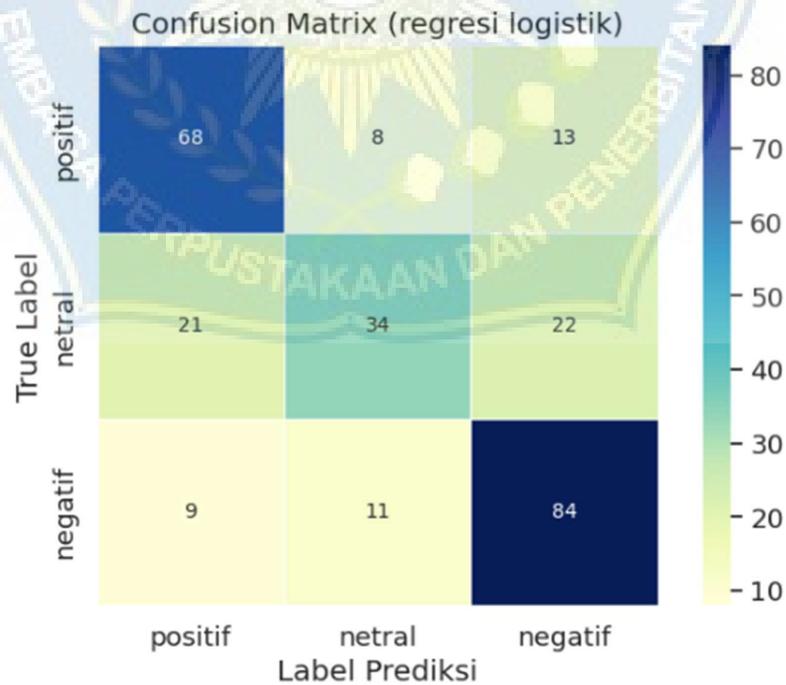


Gambar 15. *Confusion Matrix* (80% data training 20% data test)

3. Pengujian *Confusion matrix split data* (70% data training 30% data test)

Confusion Matrix pada Data Uji regresi logistik:
[[68 8 13]
[21 34 22]
[9 11 84]]

Gambar 16. Hasil Pengujian *Confusion Matrix* (70% data training 30% data test)



Gambar 17. *Confusion Matrix* (70% data training 30% data test)

Tabel 11. *Confusion Matrix*

		<i>True Label</i>		
		Class	Positif	Netral
<i>Predected Label</i>	Positif	TP	FN	FN
	Netral	FP	TP	FN
	Negatif	FP	FP	TN

Tabel 12. Hasil *Confusion Matrix* (90% data training 10% data test)

		<i>True Label</i>		
		Class	Positif	Netral
<i>Predected Label</i>	Positif	23	4	4
	Netral	6	16	7
	Negatif	1	5	24

Keterangan :

- True Positive* (TP) : 39
- False Positive* (FP) : 12
- True Negative* (TN) : 24
- False Negative* (FN) : 15

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

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

$$\begin{aligned} Accuracy &= \frac{TP+TN}{TP+TN+FP+F} \times 100\% = \frac{39+24}{39+24+12+1} \times 100\% \\ &= 70\% \end{aligned}$$

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

$$\begin{aligned} Presisi &= \frac{TP}{TP+FP} \times 100\% = \frac{39}{39+12} \times 100\% \\ &= 76\% \end{aligned}$$

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

$$\begin{aligned} Recall &= \frac{TP}{TP+FN} \times 100\% = \frac{39}{39+15} \times 100\% \\ &= 72\% \end{aligned}$$

F1-Score didapatkan dari penggabungan hasil presisi dan *recall* dengan menggunakan rata-rata *harmonic* presisi dan *recall*.

$$\begin{aligned} F1_{Score} &= 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = 2 * \frac{76*72}{76+72} \\ &= 2 * \frac{5472}{148} \\ &= 74\% \end{aligned}$$

Tabel 13. Hasil *Confusion Matrix* (80% data training 20% data test)

		True Label		
		Class	Positif	Netral
Predicted Label	Positif	42	7	9
	Netral	11	26	18
	Negatif	3	6	58

Keterangan :

True Positive (TP) : 68

False Positive (FP) : 20

True Negative (TN) : 58

False Negative (FN) : 34

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

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

$$Accuracy = \frac{TP+TN}{TP+TN+FP+F} \times 100\% = \frac{68+58}{68+58+20+3} \times 100\% = 70\%$$

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

$$Presisi = \frac{TP}{TP+FP} \times 100\% = \frac{68}{68+2} \times 100\% = 77\%$$

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

$$Recall = \frac{TP}{TP+F} \times 100\% = \frac{68}{68+34} \times 100\% = 66\%$$

F1-Score didapatkan dari penggabungan hasil presisi dan *recall* dengan menggunakan rata-rata *harmonic* presisi dan *recall*.

$$F1_Score = 2 \cdot \frac{precision \cdot recall}{precision + recall} = 2 \cdot \frac{77 \cdot 66}{77+66} = 2 \cdot \frac{5082}{143} = 71\%$$

Tabel 14. Hasil *Confusion Matrix* (70 data training 30% data test)

		True Label		
		Positif	Netral	Negatif
Predicted Label	Class			
	Positif	68	8	13
	Netral	21	34	22
	Negatif	9	11	84

Keterangan :

- True Positive* (TP) : 102
- False Positive* (FP) : 41
- True Negative* (TN) : 84
- False Negative* (FN) : 43

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

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

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% = \frac{102+}{102+84+41+43} \times 100\% = 68\%$$

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

$$Presisi = \frac{TP}{TP+F} \times 100\% = \frac{102}{102+4} \times 100\% = 71\%$$

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

$$Recall = \frac{TP}{TP+FN} \times 100\% = \frac{102}{102+43} \times 100\% = 70\%$$

F1-Score didapatkan dari penggabungan hasil presisi dan *recall* dengan menggunakan rata-rata *harmonic* presisi dan *recall*.

$$F1_Score = 2 \cdot \frac{precision \cdot recall}{precision + recall} = 2 * \frac{71*70}{71+70} = 2 * \frac{4970}{141} = 70\%$$

Tabel 15. Hasil Klasifikasi Dengan *Labelling* Manual Dan *Confusion Matrix*

Ulasan	Label	Label Prediksi	Hasil <i>Confusion Matrix</i>
Pantai indah love to you	Positif	Positif	<i>True Positive</i>
pantai bira			

Pantai kotor	Negatif	Negatif	<i>True Negative</i>
Sulawesi dikenal pulau kaya potensi alam budayanya	Netral	Positif	<i>False Positive</i>
Bersih	Positif	Negatif	<i>False Negative</i>



BAB V

PENUTUP

A. Kesimpulan

Berdasarkan penelitian analisis sentimen yang telah peneliti lakukan maka dapat di ambil 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 *regresi logistik*. Data *scrapping* yang dari ulasan wisatawan pada *google maps* akan di klasifikasikan menjadi 3 polaritas, yaitu positif, negatif, dan netral. Dari total 1000 data, setelah melewati proses *preprocessing* didapatkan data bersih 900. Selanjutnya menggunakan model spitting data dengan perbandingan 90% data *training* 10% data *test*, 80% data *training* 20% data *test*, 70% data *training* 30% data *test*.
2. Berdasarkan hasil yang telah didapatkan dari analisis sentimen wisatawan terhadap destinasi pantai tanjung bira dengan menggunakan Algoritma *regresi logistik* menunjukkan kinerja yang baik dalam mengklasifikasikan sentimen dan diperoleh hasil yang cukup tinggi. Pengujian pertama dengan *split* data 90% data *training* 10% data *test* memperoleh *accuracy* 70%, pengujian kedua dengan *split* data 80% data *training* 20% data *test* memperoleh *accuracy* 70%, pengujian ketiga dengan *split* data 70% data *training* 30% data *test* memperoleh *accuracy* 68%.

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

- Alita, D., & Isnain, A. R. (2020). Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *Jurnal Komputasi*, 8(2), 50–58. <https://doi.org/10.23960/komputasi.v8i2.2615>
- Amrullah, A. Z., Sofyan Anas, A., & Hidayat, M. A. J. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal*, 2(1), 40–44. <https://doi.org/10.30812/bite.v2i1.804>
- Anto, S., & Makbul Syamsuri, A. (2023). *Analysis Of Changes In The Tanjung Bira Coastline, Bontobahari District, Bulukumba District*. 08(01), 60–68. <http://ejurnal.unmuhjember.ac.id/index.php/HEXAGON>
- Asri, Y., Suliyanti, W. N., Kuswardani, D., & Fajri, M. (2022). Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile. *Petir*, 15(2), 264–275. <https://doi.org/10.33322/petir.v15i2.1733>
- Averina, A., Hadi, H., & Siswantoro, J. (2022). Analisis Sentimen Multi-Kelas Untuk Film Berbasis Teks Ulasan Menggunakan Model *Regresi logistik*. *Teknika*, 11(2), 123–128. <https://doi.org/10.34148/teknika.v11i2.461>
- Inayah, H., & Istiqomah, A. (2020). Nilai Ekonomi Sampah di Kawasan Wisata Pantai Tanjung Bira, Sulawesi Selatan. *Jurnal Ilmu Pertanian Indonesia*, 26(1), 159–166. <https://doi.org/10.18343/jipi.26.1.169>
- Innassuraiya, S., Widiharih, T., & Utami, I. T. (2022). ANALISIS KLASIFIKASI MENGGUNAKAN METODE *REGRESI LOGISTIK* BINER DAN BOOTSTRAP AGGREGATING CLASSIFICATION AND REGRESSION TREES (BAGGING CART) (Studi Kasus: Nasabah Koperasi Simpan Pinjam Dan Pembiayaan Syariah (KSPPS)). *Jurnal Gaussian*, 11(2), 183–194. <https://doi.org/10.14710/j.gauss.v11i2.35458>
- 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>

- Khomsah, S., & Aribowo, A. S. (2017). Text-Preprocessing Komentar Youtube Dalam Bahasa Indonesia. *Rekayasa Sistem Dan Teknologi Informasi*, 1(3), 648–654.
- Lengkong, N. C., Safitri, O., Machsus, S., Putra, Y. R., Syhadati, A., & Nooraeni, R. (2021). Analisis Sentimen Penerapan Psbb Di Dki Jakarta Dan Dampaknya Terhadap Pergerakan Ihsg. *Jurnal Teknoinfo*, 15(1), 20. <https://doi.org/10.33365/jti.v15i1.866>
- Lutfiah, A., & Habibah, N. A. (2022). Indonesian Journal of Mathematics and Natural Sciences. *Indonesian Journal of Mathematics and Natural Sciences*, 45(1), 1–8.
- Prabowo, R., Sujaini, H., & Rismawan, T. (2023). Analisis Sentimen Pengguna Twitter Terhadap Kasus COVID-19 di Indonesia Menggunakan Metode Regresi logistik Multinomial. *Jurnal Sistem Dan Teknologi Informasi (JustIN)*, 11(1), 85. <https://doi.org/10.26418/justin.v11i1.57450>
- Rahmaningrum, A. (2019). *KLASIFIKASI SENTIMEN TERHADAP REVIEW LAYANAN HOTEL BINTANG TIGA DI SURABAYA PADA SITUS TRAVELOKA MENGGUNAKAN NAÏVE BAYES CLASSIFIER (NBC) DAN REGRESI LOGISTIK PADA SITUS TRAVELOKA MENGGUNAKAN NAÏVE BAYES CLASSIFIER (NBC) DAN REGRESI LOGISTIK.*
- Sabaruddin, R. (2016). ANALISIS SENTIMEN HATESPEECH PADA TWITTER DENGAN METODE NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE. *Jurnal Dinamika Informatika Volume 5, Nomor 2, September 2016 ISSN 1978-1660 ANALISIS, May*, 31–48.
- Salim, A., & Alfian, M. R. (2019). Optimalisasi Regresi logistik Menggunakan Algoritma Genetika Pada Data Klasifikasi. *Jurnal Teknologi Informasi Dan Terapan*, 6(2), 50–55. <https://doi.org/10.25047/jtit.v6i2.109>
- Shiddicky, A. S. H. (2022). *Ash shiddicky.*
- Sholeh, M., Widyastuti, Naniek, dan A. M., & Mashuri. (2011). Aplikasi Google maps API untuk Sistem Informasi Geografis(Google maps API Applications for Geograpaphic Informaation System). *Juita*, 97–103.
- Tampil, Y., Komaliq, H., & Langi, Y. (2017). Analisis Regresi logistik Untuk

Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado. *D’CARTESIAN*, 6(2), 56. <https://doi.org/10.35799/dc.6.2.2017.17023>

Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437. <https://doi.org/10.30865/mib.v4i2.2080>

Wardani, S. (2015). Analisis Sentimen Data Presiden Jokowi Dengan Preprocessing Normalisasi Dan Stemming Menggunakan Metode Naive Bayes Dan Svm. *Jurnal Dinamika Informatika*, 5(November), 1–13.

Wati, R. (2016). Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Bayes Risa. 4, 25–31.



LAMPIRAN

Lampiran 1. *Dataset* Ulasan

No	Ulasan
1	pantai tercantik yang pernah aku liat
2	pantainya Indah
3	Sangat menyukai pantai, apalagi pasirnya putih dan bersih
4	Tempatnya bagus
5	Sangat indah pantai dengan pasir putihnya semoga destinasi wisata ini terjaga kebersihannya...
6	Kalo disini enaknya bermalam.., karena mandi dipantai lebih asyik di pagi hari. Jangan lupa bawa peralatan snorkling..
7	Pemandangan sangat indah... keren buat liburan sama teman teman dan keluarga
8	Pasir putihnya lembut..., ada permainan berbayar bagi penggemar kegiatan pantai...
9	Sering dibicarakan orang! insyaallah bakal kesini
10	Pantai pasir putih, banyak kegiatan yang bisa di lakukan di pantai bira bersama keluarga.
11	Tempat terkeren... pasir putih laut biru banyak tempat wisata baru pokoknya paling favorit
12	Pemandangan yang bagus
13	Pasir pantainya bersih dan halus,sangat memanjakan mata
14	pantai nya biru dan jernih dengan pasir putih yang memukau
15	Pasir putihnya cantik pemandangan lautnya biru
16	Keseluruhan masih alami keren
17	Wisata pantai tanjung bira yang keren
18	Tempat camp favorit walaupun jauh tapi terbayarkan dengan pemandangannya
19	Pantai dengan pasir putih. Instagramable.

20	Tempatnya sangat indah pantai ini mirip dengan Bali indah asri dan nyaman
21	Indah pasirnya putih seperti tepung terigu
22	Pasirnya lembut
23	Sangat indah namun ramai
24	Mantap wisata menakjubkan...
25	pantai pasir putih yang indah dan mengasikkan serta pasir putihnya halus.
26	Liburan yang enak ketika pantai ini sepi sekali yaitu di weekday
27	Coba diving deh, bagus banget!!
28	Sangat menarik karena bisa menenangkan hati
29	Tempat ini indah sekali, minggu lalu saya berkunjung kesana.
30	Salah satu destinasi alam di provinsi sulse
31	Surga dunia di bulukumba ..
32	Pasir pantai paling putih yang pernah kami kunjungi
33	Siapa yang kenal dengan tempat ini harus segera kesini. Pencinta pantai pasti suka
34	Luas dan sepi bagus banget
35	Tempat buat keluarga berakhir pekan atau berlibur yang sangat tepat
36	Keren banget cuk
37	layak untuk dikunjungi
38	Salah satu destinasi wisata yang wajib di kunjungi... Surga dunia yang tersembunyi.
39	Alhamdulillah luar Biasa Tempatnya Tapi masih perlu di perbaiki Tempat Jualannya atur biar indah
40	Indah pasir putihnya
41	Yang suka pantai wajib deh kesini...buktiin sendiri
42	Cukup bersih dan di senangi turis manca negara untuk berjemur di bibir pantai
43	pantai yang indah dan suasana yang nyaman untuk pilihan berlibur
44	Tempat yang luar biasa

45	Indah banget
46	Pantai impian nasional dan internasional
47	Pantai dengan pasir putih terbaik
48	Tempat yang paling indah tempatnya
49	Paling suka sama pantai ini,tempatnya nyaman dan adem..
50	Pasirnya halus
...
850	okelah
851	ayo kesini
852	tempatnya lumayan bagus
853	bagu kesini sama keluarga atau teman
854	sebagai pengunjung tolong jaga kebersihan dan fasilita jangan dirusak
855	rekomendasi tempat heal keluarga
856	salah satu tempat wisata yang rekomend
857	di musim seperti bulan ini banyak yang berkunjung
858	suka banget ngga nyesel deh jlanan masuknya aja yg perlu diperhatikan lagi
859	bagu hanya perlu diperbaiki aks jalan yg masih berlubang dan fasilita umum yg kurang memadai
860	pemandangan yang asri
861	banyak anak agak mengganggu keasyikan berenang
862	biaya masuk mahal tapi lumayan bagus sih
863	segitu dulu bintangnya belum sempat berkunjung kesitu kalau udah entar d tambah
864	pemandangannya asyik sudah terkesan modern
865	sere dibicarakan org insyaallah bakal kesini
866	tempatnya bagus hanya aks ke sana yg haru di perbaiki lagi dan tdk bisa bawa makanan dari luar
867	cissssss
868	keren cuma sayang waktunya kurang pa

869	destinasi wisata ya sih bagus dn indah cuman agak mahal
870	sayang sekali pa kesini kurang pengunjung
871	daerah wisata yg menjanjikan tapi jalan masuknya masih perlu banyak perhatian lagi
872	tempat hiburan keluarga
873	tempatny dekat rumah aku
874	tempat yang bagus namun minim fasilita kurang berkembang
875	keren cuma aks jalan masih kurang baik
876	keseluruhan masih alami
877	sangat menarik karena bisa menenangkan hati
878	lua dan sejuk cuma jalan menuju kesini agak kurang mumpuni
879	alhamdulillah luar biasa tempatnya tapi masih perlu di perbaiki tempat jualannya atur biar indah
880	bagu n masih perlu dikelola dgn maksim oleh pemerintah setempat
881	apaka lebaran tetap buka
882	kurang banyak yang tau t ini
883	nanti kesini lagi
884	belum kesampaian baru mau pergi menikmatinya
885	ndak lgi deh yang tahu tahu aja
886	alhamdulillah akhirnya bisa kesini
887	sangat mengesankan bagi saya karna teman teman saya ingin berpisah
888	ayo ajak keluarga atau teman ke sini
889	aks kesana spi dipermulu kalau perlu diperlebar
890	klw ada waktu nanti kesini lagi
891	semoga ya hari ini
892	tahun depan ke sini ramaiiiiiiiii
893	biasa
894	biasa saja
895	jalan menuju tempat ini cukup berbatu tapi bagus viewnya banyak pohon dan rindang sejuk banget .

- 896 kemarin pa liburan datangnya ke sini
- 897 teman aku tdk dapat deh jadinya haru gantin tapi aku rekomendasikanlah kapan klaw liburan bisa ke sini
- 898 harga masuk terjangkau pelayanan ok secara umuk baguslah
- 899 seruuu dan puass tetapi yg jual makanan kebelet mau umroh kek na
- 900 bagus sih cuman banyak sampahnya

Lampiran 2. Hasil Pelabelan

comment	label
1 pantai cantik yang pernah aku lat	Positif
2 pantainya indah	Positif
3	Positif
4 Sangat menyukai pantai, apalagi pasainya putih dan bersih	Positif
5 Tempatnya bagus	Positif
6 Sangat indah pantai dengan pasir putihnya semoga destinasi wisata ini terjaga kebersihannya.	Positif
7 Kalo disini endunya bermalam... karena mandi dipantai lebih ayuk di pagi hari. Jangan lupa bawa peralatan snorkeling.	Positif
8 Pemandangan sangat indah... karena saat liburan sama teman teman dan keluarga	Positif
9 Pasir putihnya lembut... ada permainan kayak kayak punggamaan kayak kayak pantai...	Positif
10 Sering dibicarakan orang! inayallah bakal lebih	Positif
11 Pantai pasir putih, banyak kegiatan yang bisa di lakukan di pantai bisa bersama keluarga.	Positif
12 Tempat tenang... pasir putih laut biru banyak tempat wisata baru-pokoknya paling favorit	Positif
13 Pemandangan yang bagus	Positif
14 Pasir pantainya bersih dan halus, sangat memanjakan mata	Positif
15 pantai nya biru dan putih dengan pasir putih yang memukau	Positif
16 Pasir putihnya cantik pemandangan lautnya biru	Positif
17 Kebersihan masih alami keren	Positif
18 Wisata pantai sangat bira yang keren	Positif
19 Tempat camping di pantai yang tapi terlayarkan dengan pemandangannya	Positif
20 Pantai dengan pasir putih. Instagramable.	Positif
21 Tempat nya sangat indah pantai ini mirip dengan Bali indah ari dan nyaman	Positif
22 Indah pasainya putih seperti tepung terigu	Positif
23 Pasainya lembut	Positif
24 Sangat indah namun ramai	Positif
25 Manjur wisata menakutkan.	Positif
26 pantai pasir putih yang indah dan mengasikan serta pasir putihnya halus.	Positif
27 Liburan yang enak ketika pantai ini sepi sekali yaitu di weekday	Positif
28 Coba diving lah, bagus banget!	Netral
29 Sangat menarik karena bisa memenangkan hati	Positif
30 Tempat ini indah sekali, minggu lalu saya berkunjung kesana.	Positif
31 Indah ari destinasi alam di provinsi sulawesi	Positif
32 Surga dunia di Sulawesi	Positif
33 Pasir pantai paling putih yang pernah kami kunjungi	Positif

Lampiran 3. Hasil Proses Cleaning data

379	Jamban bau	10 bulan lalu	sampah berserakan disepanjang pantai yang tidak segera dibersihkan begitu juga	Negatif	banyaknya sampah berserakan disepanjang pantai yang tidak segera dibersihkan begitu juga
380	Darwis	3 bulan lalu	lumayan banyak sampah di bibir pantainya	Negatif	lumayan banyak sampah di bibir pantainya
381	Masih Mau	2 minggu lalu	infrastruktur & bangunan di sana tdk tertata dengan baik jd kelihatan kumuh	Negatif	infrastruktur bangunan di sana tdk tertata dengan baik jd kelihatan kumuh
382	Rangga Pramana	5 tahun lalu	dekat pantainya agak banyak karangnya	Negatif	dekat pantainya agak banyak karangnya
383	Fandi Hakim	5 tahun lalu	kurang bersih karena terlihat beberapa sampah yang berserakan. ingat, kita satu bumi, jangan buang sampah sembarangan! Cinta bumi!	Negatif	kurang bersih karena terlihat beberapa sampah yang berserakan. ingat kita satu bumi jangan buang sampah sembarangan. Cinta bumi
384	Supriadi Uppie	setahun lalu	klu bisa adakan petugas kebersihan	Negatif	klu bisa adakan petugas kebersihan
385	Sabirin Andi	3 minggu lalu	masuk tiap spot mesti bayar, tarif parkir juga tidak masuk akal.	Negatif	masuk tiap spot mesti bayar tarif parkir juga tidak masuk akal
386	agung Triatnojo	4 tahun lalu	infrastruktur yang kurang memadai sehingga aktivitas di lokasi kurang maksimal. Harapan kami kedepan ada fasilitas wisata yang pro pada "wisata halal"...	Negatif	infrastruktur yang kurang memadai sehingga aktivitas di lokasi kurang maksimal. Harapan kami kedepan ada fasilitas wisata yang pro pada wisata halal
387	Man Dar	6 tahun lalu	para penjual membuat lapak persis di bibir pantai. Menurut Kepala Dinas Pariwisata, segera akan ditata dan lapak yang sekarang terkesan kurang rapi akan diganti dengan yang sedikit modis.	Negatif	para penjual membuat lapak persis di bibir pantai. Menurut Kepala Dinas Pariwisata segera akan ditata dan lapak yang sekarang terkesan kurang rapi akan diganti dengan yang sedikit modis
388	Brahmana Muhammad Iman	5 tahun lalu	Pantainya berpasir putih cantik sekali, tapi sayang mungkin karena kurangnya perhatian dari warga, wisatawan dan pemerintah terkait. Pantai ini mulai kotor.	Negatif	Pantainya berpasir putih cantik sekali tapi sayang mungkin karena kurangnya perhatian dari warga wisatawan dan pemerintah terkait Pantai ini mulai kotor
389	dian nita	4 tahun lalu	jalan yg menuju pantai sebelahnya pantai bira memang jalannya masih berbatu tdk rata.	Negatif	jalan yg menuju pantai sebelahnya pantai bira memang jalannya masih berbatu tdk rata
390	Rizz Fozz	6 tahun lalu	keramaian warung, pedagang, hotel dan resort di sekitar pantai, dan ditambah pengunjung yang membludak di hari liburan, harus ...	Netral	keramaian warung pedagang hotel dan resort di sekitar pantai dan ditambah pengunjung yang membludak di hari liburan harus
391	CRS Explore Channel	4 tahun lalu	masih perlu banyak pembenahan.	Negatif	masih perlu banyak pembenahan
392	Awal Sewang	setahun lalu	kebersihan masih minim butuh di tingkatkan	Negatif	kebersihan masih minim butuh di tingkatkan
393	Indi maharani	5 bulan lalu	pantainya sudah mulai kotor dan susah air bersih untuk bilas	Negatif	pantainya sudah mulai kotor dan susah air bersih untuk bilas

Lampiran 4. Hasil Proses Case Folding

Index	user	date	comment	label
386	agung Triatmojo	4 tahun lalu	infrastruktur yang kurang memadai sehingga aktivitas di lokasi kurang maksimal Harapan kami kedepan ada fasilitas wisata yang pro pada wisata halal	Negatif
387	Man Dar	6 tahun lalu	para penjual membuat lapak persis di bibir pantai Menurut Kepala Dinas Pariwisata segera akan ditata dan lapak yang sekarang terkesan kurang rapi akan diganti dengan yang sedikit modis	Negatif
388	Brahmana Muhammad Ilman	5 tahun lalu	Pantainya berpasir putih cantik sekali tapi sayang mungkin karena kurangnya perhatian dari warga wisatawan dan pemerintah terkait Pantai ini mulai kotor	Negatif
389	dian nita	4 tahun lalu	jalan yg menuju pantai sebelahnya pantai bira memang jalannya masih berbatu tdk rata	Negatif
390	Rizz Fozz	6 tahun lalu	keramaian warung pedagang hotel dan resort di sekitar pantai dan ditambah pengunjung yang membludak di hari liburan harus	Netral
391	CRS Explore Channel	4 tahun lalu	masih perlu banyak pembenahan	Negatif
392	Awal Sewang	setahun lalu	kebersihan masih minim butuh di tingkatkan	Negatif
393	Indri Maharani	5 bulan lalu	pantainya sudah mulai kotor dan susah air bersih untuk bilas	Negatif
394	Bara Syarif	10 bulan lalu	Kurang terawat	Negatif
395	EDDY susilo	5 tahun lalu	PANTAI YG INDAH PUTIH BERSIH cuma kurang pengelolannya seandainya dikelola dengan baik oleh pemerintah daerah yakin akan menjadi wisata yg menarik	Negatif
396	Sri Mulyati	3 tahun lalu	Penataan Kebersihan Keramatahman Masih perlu ditingkatkan	Negatif
397	Icha Chan	9 bulan lalu	masih sering ada sampah yg dibawa ombak yuk jangan buang sampah sembarangan	Negatif
398	Ikhwan Mahmud	setahun lalu	Sekarang semakin banyak bangunan kesannya jd kurang tertata dengan baik alias semrawut Mungkin site plan pengembangan kawasan	Netral
399	ana service	2 tahun lalu	Pantai kotor Tdk ada sarana tempat sampah di pantai	Negatif

Lampiran 5. Hasil Proses *Stopword*

Index	user	date	comment	label
0	Andi Mufidatunnisa	setahun lalu	pantai tercantik yang pernah aku liat	Positif
1	Afdal Reckers	8 bulan lalu	pantainya indah	Positif
2	Rosadhy Wahyudhy	3 tahun lalu	sangat menyukai pantai apalagi pasirmya putih dan bersih	Positif
3	Jusri 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	ateng handy	5 tahun lalu	kalo disini enakny bermalam karena mandi dipantai lebih asyik di pagi hari jangan lupa peralatan snorkling	Positif
6	Ardry Juple	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 insyaallah kesini	Positif
9	Anneke Mutiarsari	3 tahun lalu	pantai pasir putih banyak kegiatan yang bisa di lakukan di pantai bira bersama keluarga	Positif
10	Early Carolina	2 tahun lalu	tempat terkeren pasir putih laut biru banyak tempat wisata baru pokoknya paling favorit	Positif
11	Diman Sinergy Channel	10 bulan lalu	pemandangan yang bagus	Positif
12	Danny Eskobra	4 tahun lalu	pasir pantainya bersih dan halusangat memanjakan mata	Positif
13	Ganendra cahyo	5 tahun lalu	pantai nya biru dan jernih dengan pasir putih yang memukau	Positif
14	Syarifa Ajinah Assegaf	3 tahun lalu	pasir putihnya cantik pemandangan lautnya biru	Positif

Lampiran 6. Hasil Proses *Tokenize*

Index	user	date	comment	label
0	Andi Mufidatunnisa	setahun lalu	pantai tercantik liat	Positif
1	Afdal Reckers	8 bulan lalu	pantainya indah	Positif
2	Rosadhy Wahyudhy	3 tahun lalu	menyukai pantai pasirmya putih bersih	Positif
3	Jusri Agung	3 tahun lalu	tempatnya bagus	Positif
4	Toni Wardono	4 tahun lalu	indah pantai pasir putihnya semoga destinasi wisata terjaga kebersihannya	Positif
5	ateng handy	5 tahun lalu	kalo enakny bermalam mandi dipantai asyik pagi lupa bawa peralatan snorkling	Positif
6	Ardry Juple	3 tahun lalu	pemandangan indah keren liburan teman teman keluarga	Positif
7	Yuni Antoko	5 tahun lalu	pasir putihnya lembut permainan berbayar penggemar kegiatan pantai	Positif
8	Nur Alam	9 bulan lalu	dibicarakan orang insyaallah kesini	Positif
9	Anneke Mutiarsari	3 tahun lalu	pantai pasir putih kegiatan lakukan pantai bira keluarga	Positif
10	Early Carolina	2 tahun lalu	terkeren pasir putih laut biru wisata pokoknya favorit	Positif
11	Diman Sinergy Channel	10 bulan lalu	pemandangan bagus	Positif
12	Danny Eskobra	4 tahun lalu	pasir pantainya bersih halusangat memanjakan mata	Positif
13	Ganendra cahyo	5 tahun lalu	pantai nya biru jernih pasir putih memukau	Positif
14	Syarifa Ajinah Assegaf	3 tahun lalu	pasir putihnya cantik pemandangan lautnya biru	Positif
15	Taufiq Lubis	4 tahun lalu	alami keren	Positif
16	resmi daeng tabe	3 tahun lalu	wisata pantai tanjung bira keren	Positif
17	Cici Karlina	4 tahun lalu	camp favorit terbayarkan pemandangannya	Positif
18	Eka Kurniawan	setahun lalu	pantai pasir putih instagramable	Positif
19	Mul Tazam	5 tahun lalu	nya indah pantai balik indah asri nyaman	Positif
20	John Silentnu	2 tahun lalu	indah pasirmya putih tepung terigu	Positif
21	Pnji Media	5 tahun lalu	pasirmya lembut	Positif
22	Erwinda Giastuti	9 bulan lalu	indah ramai	Positif
23	INCE ILUHAMMAD	4 tahun lalu	mantap wisata menakjubkan	Positif
24	Sur I am	5 tahun lalu	nanti nasi putih indah memantapkan nasi putihnya halus	Positif

Lampiran 7. Sourcode

1. Instal and import Library

```
# Install dan import library
!pip install nltk
import nltk
nltk.download('stopwords')
nltk.download('punkt')
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
import sklearn
warnings.simplefilter("ignore")
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics import confusion_matrix, classification_report
```

2. Preprocessing Data

- Cleaning

```
import pandas as pd
import re
from google.colab import files
uploaded = files.upload()
df = pd.read_excel('dataset ulasan.xlsx')
df
# Fungsi untuk membersihkan teks dari emoji, tanda baca, titik, koma, dan angka
def clean_text(text):
    # Menghapus emoji
    text = text.encode('ascii', 'ignore').decode('ascii')
    # Menghapus tanda baca, titik, koma, dan angka
    text = re.sub(r'[^\w\s]', '', text)
    text = re.sub(r'\d', '', text)
```

```

return text

# Membersihkan kolom ulasan (ganti 'column_name' dengan nama kolom yang sesuai)
df['comment'] = df['comment'].apply(clean_text)
df
df.to_excel('Cleaning.xlsx', index=False)

```

- *Case Folding*

```

import pandas as pd
import re
df = pd.read_excel('Cleaning.xlsx')
df
df['comment'] = df['comment'].str.lower()
df
df.to_excel('Casefolding.xlsx', index=False)

```

- *Stopword*

```

# Install dan import library pandas
!pip install pandas
import pandas as pd
# Membaca dataset (ganti 'nama_file.csv' dengan nama file dataset Anda)
df = pd.read_excel('Casefolding.xlsx')
df
# Tokenisasi dan hapus stopwords
stop_words = set(stopwords.words('indonesian'))
# Fungsi untuk menghapus stopwords dari teks
def remove_stopwords(text):
    words = word_tokenize(text)
    filtered_words = [word for word in words if word.lower() not in stop_words]
    return ' '.join(filtered_words)
# Menerapkan fungsi pada kolom teks dataset
df['stopword'] = df['comment'].apply(remove_stopwords)
df
df.to_excel('Stopword.xlsx', index=False)

```

- *Tokenize*

```
import pandas as pd

# Ganti 'nama_file.xlsx' dengan nama file Excel Anda
path_to_excel = 'Stopword.xlsx'

# Baca dataset Excel
df = pd.read_excel(path_to_excel)
# Tampilkan beberapa baris pertama dataset
df.head()
kolom_teks = 'comment'
teks = df[kolom_teks]
import nltk
from nltk.tokenize import word_tokenize

nltk.download('punkt')

# Tokenisasi teks dalam kolom 'teks'
df['comment'] = df['comment'].apply(lambda x: word_tokenize(str(x)))
df
df.to_excel('Tokenize.xlsx', index=False)
```

3. *Splitting Data*

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x,y,random_state=0,test_size=0.2)
```

4. *Klasifikasi Regresi logistik*

```
logistic_reg = LogisticRegression()
logistic_reg.fit(x_train_vec, y_train)
y_pred = logistic_reg.predict(x_test_vec)
y_pred
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score

#accuracy = accuracy_score(y_test, y_pred)
#print("Akurasi: {:.2f}%".format(accuracy * 100))

# Hitung metrik evaluasi
accuracy = accuracy_score(y_test, y_pred)
```

```

precision = precision_score(y_test, y_pred, average='macro')
recall = recall_score(y_test, y_pred, average='macro')
f1 = f1_score(y_test, y_pred, average='macro')

# Tampilkan hasil evaluasi
print(f'Accuracy: {accuracy}')
print(f'Precision: {precision}')
print(f'Recall: {recall}')
print(f'F1 Score: {f1}')
# Hitung metrik evaluasi
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average='macro')
recall = recall_score(y_test, y_pred, average='macro')
f1 = f1_score(y_test, y_pred, average='macro')

# Tampilkan hasil evaluasi
print(f'Accuracy: {accuracy}')
print(f'Precision: {precision}')
print(f'Recall: {recall}')
print(f'F1 Score: {f1}')

# Tampilkan classification report pada data uji
class_report_test = classification_report(y_test, y_pred)
print('Classification Report pada Data Uji REGRESI LOGISTIK:')
print(class_report_test)

print(f'Accuracy: {accuracy}')

```

5. Pengujian *Confusion matrix*

```

# Tampilkan confusion matrix pada data uji
conf_matrix_test = confusion_matrix(y_test, y_pred)
print('Confusion matrix pada Data Uji regresi logistik:')
print(conf_matrix_test)

# Define your class labels (assuming you have 'Positive', 'negative', and 'neutral')
class_labels = ['positif', 'netral', 'negatif']

# Display confusion matrix as a heatmap with class annotations
conf_matrix_test = confusion_matrix(y_test, y_pred)

```

```

plt.figure(figsize=(8, 5))
sns.set(font_scale=1.2) # Adjust font size
sns.heatmap(conf_matrix_test, annot=True, fmt='d', cmap='YlGnBu', cbar=True,
             xticklabels=class_labels, yticklabels=class_labels,
             linewidths=.5, square=True, annot_kws={"size": 10})
plt.title('Confusion matrix (regresi logistik)')
plt.xlabel('Label Prediksi')
plt.ylabel('True Label')
plt.show()

# Mengekstrak nilai True Positive (TP), False Positive (FP), True Negative (TN), dan False
# Negative (FN)
TP = 40
FP = 11
TN = 23
FN = 16

# Menghitung akurasi, presisi, recall, dan F1-score
accuracy = (TP + TN) / (TP + FP + TN + FN)
precision = TP / (TP + FP)
recall = TP / (TP + FN)
f1 = 2 * (precision * recall) / (precision + recall)

# Menampilkan hasil
print("Akurasi:", accuracy)
print("Presisi:", precision)
print("Recall:", recall)
print("F1 Score:", f1)

```

Lampiran 8. 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 : Saddam
Nim : 105841107719
Program Studi : Teknik Informatika

Dengan nilai:

No	Bab	Nilai	Ambang Batas
1	Bab 1	10 %	10 %
2	Bab 2	22 %	25 %
3	Bab 3	8 %	10 %
4	Bab 4	4 %	10 %
5	Bab 5	5 %	5 %

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

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

Makassar, 25 Juni 2024
Mengetahui,
Kepala UPT- Perpustakaan dan Penerbitan,


F. M. M. I. P.
NIP. 19640404199104591



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

Lampiran 9. Plagiasi Per Bab



addan 105841107719 Bab I

ORIGINALITY REPORT

10%

SIMILARITY INDEX

10%

INTERNET SOURCES

0%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1	smartlib.umri.ac.id Internet Source	2%
2	docplayer.info Internet Source	2%
3	123dok.com Internet Source	2%
4	digilib.unila.ac.id Internet Source	2%
5	www.penemuanterbaru.com Internet Source	2%

Exclude quotes Off

Exclude bibliography Off

Exclude matches < 2%

Saddan 105841107719 Bab II

by Tahap Tutup

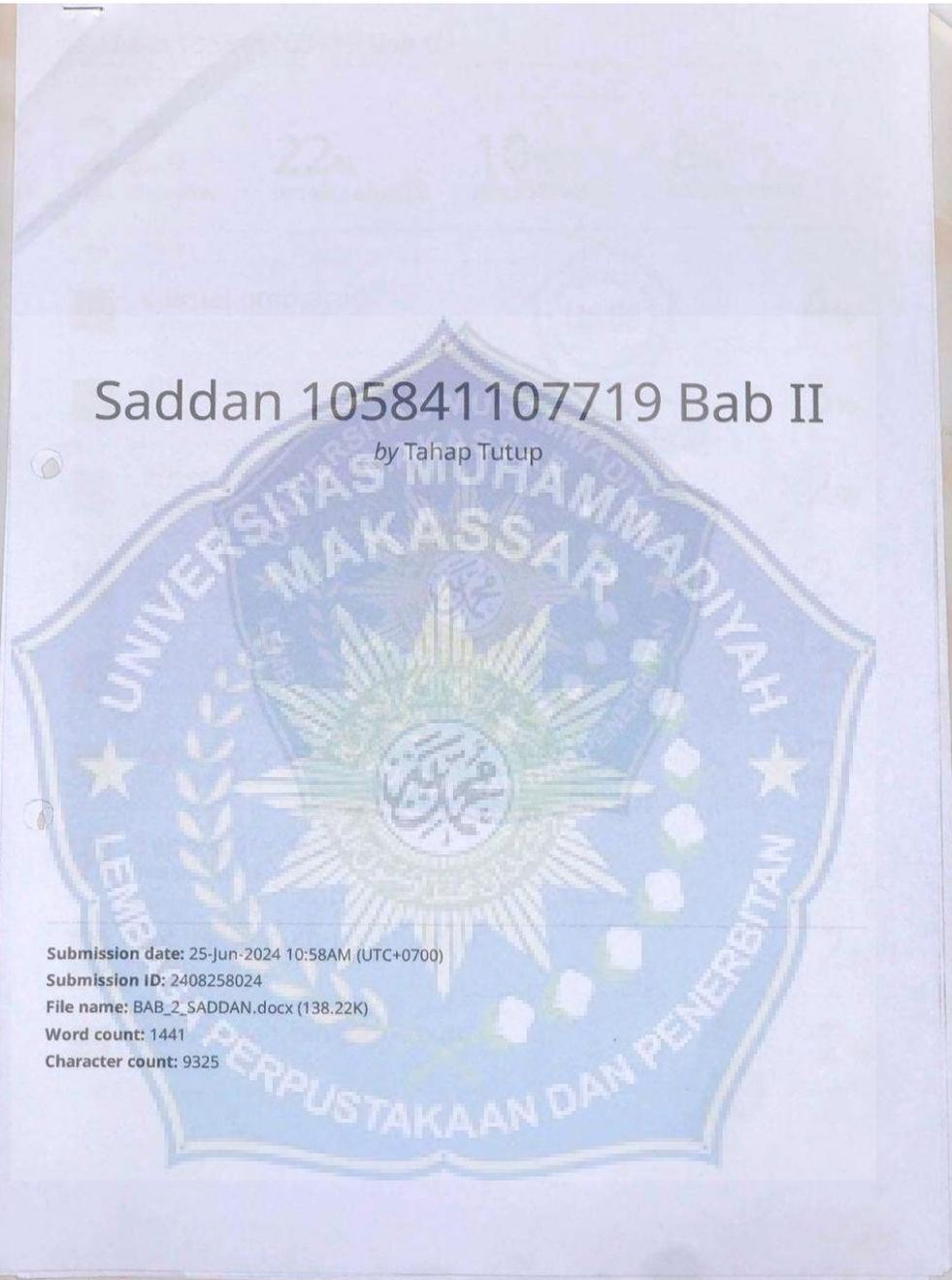
Submission date: 25-Jun-2024 10:58AM (UTC+0700)

Submission ID: 2408258024

File name: BAB_2_SADDAN.docx (138.22K)

Word count: 1441

Character count: 9325



Saddan 105841107719 Bab II

ORIGINALITY REPORT

22%
SIMILARITY INDEX

22%
INTERNET SOURCES

10%
PUBLICATIONS

8%
STUDENT PAPERS

PRIMARY SOURCES

1	ejurnal.umri.ac.id Internet Source		4%
2	pdfs.semanticscholar.org Internet Source		3%
3	id.scribd.com Internet Source		2%
4	jurnal.fmipa.unila.ac.id Internet Source		2%
5	123dok.com Internet Source		2%
6	repository.ubaya.ac.id Internet Source		2%
7	repository.its.ac.id Internet Source		2%
8	repository.uin-suska.ac.id Internet Source		2%
9	peta-peta1.blogspot.com Internet Source		2%

Exclude quotes Off

Exclude bibliography Off

Exclude matches < 2%



Saddan 105841107719 Bab III

by Tahap Tutup

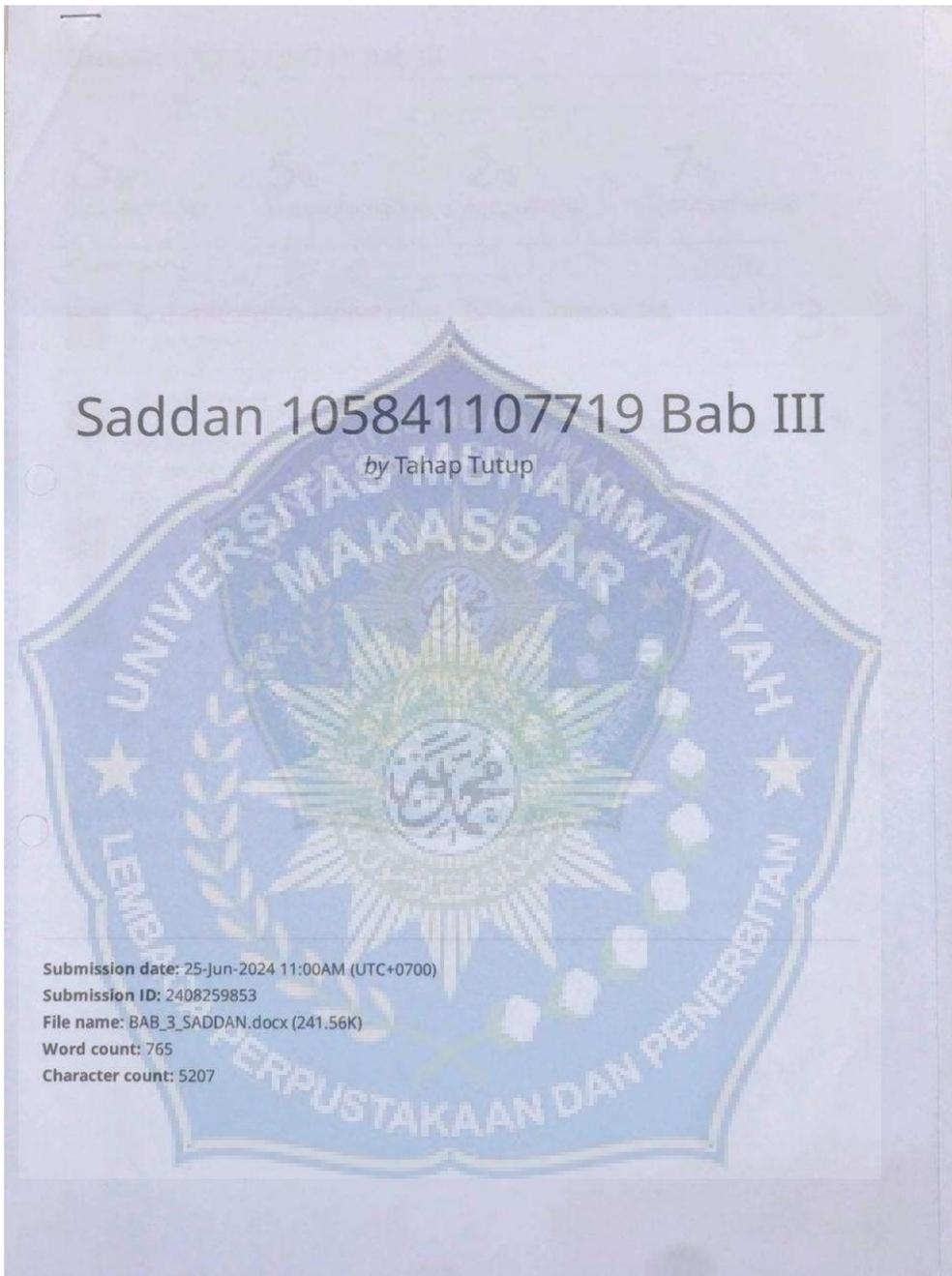
Submission date: 25-Jun-2024 11:00AM (UTC+0700)

Submission ID: 2408259853

File name: BAB_3_SADDAN.docx (241.56K)

Word count: 765

Character count: 5207



ORIGINALITY REPORT

8%

SIMILARITY INDEX

5%

INTERNET SOURCES

2%

PUBLICATIONS

7%

STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Islam Indonesia Student Paper	3%
2	Submitted to Universitas Muhammadiyah Makassar Student Paper	2%
3	docplayer.info Internet Source	2%
4	es.scribd.com Internet Source	2%



Exclude quotes Off
Exclude bibliography Off

Exclude matches < 2%

Exclude bibliography Off

Saddan 105841107719 Bab IV
by Tahap Tutup



Submission date: 25-Jun-2024 11:01AM (UTC+0700)
Submission ID: 2408260375
File name: BAB_4_SADDAN.docx (1.29M)
Word count: 2224
Character count: 13301

ORIGINALITY REPORT

4%	2%	1%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Dr. B R Ambedkar National Institute of Technology, Jalandhar Student Paper	1%
2	aditpermana.wordpress.com Internet Source	1%
3	Submitted to Singapore Institute of Technology Student Paper	1%
4	Submitted to Universite Saint Joseph Student Paper	<1%
5	core.ac.uk Internet Source	<1%
6	repository.stei.ac.id Internet Source	<1%
7	tempatku-wisata.blogspot.com Internet Source	<1%
8	balarmedan.wordpress.com Internet Source	<1%

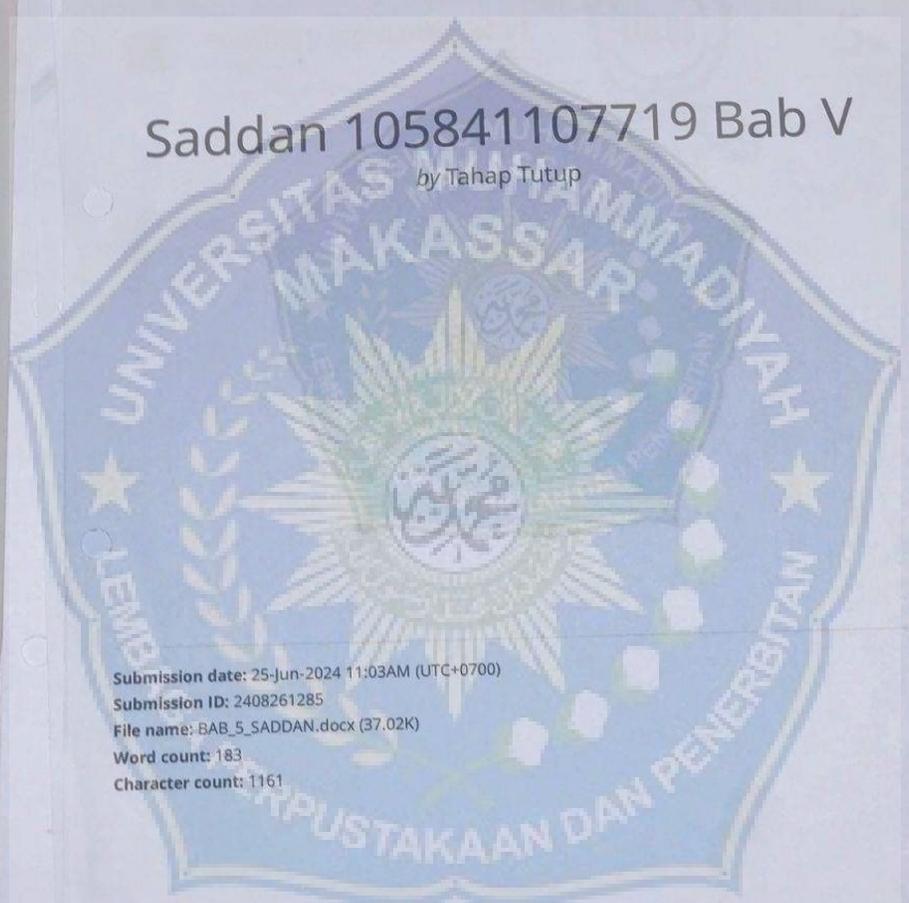


Exclude quotes Off
Exclude bibliography Off

Exclude matches Off



Saddan 105841107719 Bab V
by Tahap Tutup



Submission date: 25-Jun-2024 11:03AM (UTC+0700)
Submission ID: 2408261285
File name: BAB_5_SADDAN.docx (37.02K)
Word count: 183
Character count: 1161

Saddan 105841107719 Bab V

ORIGINALITY REPORT

5%
SIMILARITY INDEX

5%
INTERNET SOURCES

5%
PUBLICATIONS

0%
STUDENT PAPERS

PRIMARY SOURCES

1 repository.lppm.unila.ac.id
Internet Source



5%

Exclude quotes

Off

Exclude matches

Off

Exclude bibliography

Off



