

**ANALISIS SENTIMEN REVIEW E-COMMERCE APLIKASI
SHOPEE DI GOOGLE PLAY STORE MENGGUNAKAN
KOMBINASI ALGORITMA NAIVE BAYES DAN BERT**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

BRIAN BINTANG CHAHYA

21.11.4195

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

**ANALISIS SENTIMEN REVIEW E-COMMERCE APLIKASI
SHOPEE DI GOOGLE PLAY STORE MENGGUNAKAN
KOMBINASI ALGORITMA NAIVE BAYES DAN BERT**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

BRIAN BINTANG CHAHYA

21.11.4195

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN REVIEW E-COMMERCE APLIKASI SHOPEE DI
GOOGLE PLAY STORE MENGGUNAKAN KOMBINASI ALGORITMA
NAIVE BAYES DAN BERT**

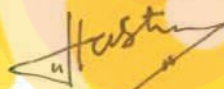
yang disusun dan diajukan oleh

Brian Bintang Chahya

21.11.4195

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Januari 2025

Dosen Pembimbing,



Hastari Utama, S.Kom., M.Cs

NIK. 190302230

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN REVIEW E-COMMERCE APLIKASI SHOPEE DI
GOOGLE PLAY STORE MENGGUNAKAN KOMBINASI ALGORITMA
NAIVE BAYES DAN BERT**

yang disusun dan diajukan oleh

Brian Bintang Chahya

21.11.4195

Telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Januari 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Moch Farid Fauzi S.Kom., M.Kom.
NIK. 190303384

Sharazita Dyah Anggita, M.Kom.
NIK. 190302285

Hastari Utama, S.Kom., M.Cs.
NIK. 190302230



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Januari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Brian Bintang Chahya
NIM : 21.11.4195

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Review E-Commerce Aplikasi Shopee di Google Play Store Menggunakan Kombinasi Algoritma Naive Bayes dan BERT

Dosen Pembimbing : Hastari Utama, S.Kom., M.Cs

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 22 Januari 2025

Yang Menyatakan,



Brian Bintang Chahya

HALAMAN PERSEMBAHAN

Dengan mengucapkan syukur kehadirat Alloh SWT, kupersembahkan skripsi ini untuk:

1. Orang Tua dan Keluarga

Terima kasih atas doa, dukungan, dan kasih sayang yang tak terhingga. Semua pencapaian ini adalah hasil dari perjuangan dan doa restu kalian.

2. Dosen Pembimbing

Terima kasih atas bimbingan, arahan, dan ilmu yang telah diberikan. Kontribusi dan dorongan yang diberikan sangat berarti dalam penyelesaian skripsi ini.

3. Tim Dosen Penguji

Saya mengucapkan terima kasih kepada Tim Dosen Penguji yang telah memberikan penilaian kritis, masukan, dan saran konstruktif sehingga penelitian ini dapat disempurnakan.

4. Dosen Pengajar

Terima kasih kepada seluruh dosen di Program Studi S1 Informatika Universitas Amikom Yogyakarta yang telah berbagi ilmu, pengalaman, dan wawasan selama masa studi. Ilmu yang diberikan menjadi bekal berharga dalam menyelesaikan skripsi ini dan dalam perjalanan akademik serta profesional ke depan.

5. Teman, Sahabat, dan Rekan Sejawat

Terima kasih atas semangat, motivasi, serta dukungan moral dan intelektual yang diberikan selama perjalanan penelitian ini.

6. Semua Pihak yang Turut Mendukung

Terima kasih kepada semua pihak yang turut serta dalam menyukseskan penelitian ini. Kontribusi dan dukungan kalian sangat berarti.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT, yang telah memberikan rahmat, petunjuk, dan kekuatan sehingga penulis dapat menyelesaikan skripsi ini dengan judul “Analisis Sentimen Review E-Commerce Aplikasi Shopee di Google Play Store Menggunakan Kombinasi Algoritma Naive Bayes dan BERT”. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Penulis ingin menyampaikan penghargaan dan terima kasih yang sebesar-besarnya kepada:

1. Orang Tua dan Keluarga

Keluarga tercinta, terima kasih atas doa, dukungan, dan cinta kasih yang tak terhingga. Segala keberhasilan ini adalah berkat doa restu dan semangat dari kalian.

2. Dosen Pembimbing

Bapak Hastari Utama, S.Kom., M.Cs, yang telah memberikan bimbingan, arahan, dan dorongan positif dalam setiap tahap penulisan skripsi ini.

3. Tim Dosen Penguji

Bapak Moch Farid Fauzi S.Kom., M.Kom, Ibu Sharazita Dyah Anggita, M.Kom, dan Bapak Hastari Utama, S.Kom. M.Cs., yang telah memberikan waktu dan penilaian kritis yang membangun untuk pengembangan skripsi ini.

4. Para Dosen Pengajar

Penghargaan setinggi-tingginya saya sampaikan kepada para dosen yang telah membagikan ilmu, pengalaman, dan inspirasi selama masa studi di Program Studi S1 Informatika Universitas Amikom Yogyakarta. Dedikasi serta komitmen mereka telah memberikan bekal pengetahuan yang mendalam, yang sangat membantu dalam penyusunan skripsi ini serta dalam perjalanan akademik dan profesional saya ke depan.

5. Rekan Sejawat dan Teman-teman

Teman-teman seperjuangan, terima kasih atas semangat, ide-ide konstruktif, dan dukungan yang selalu diberikan sepanjang perjalanan penelitian.

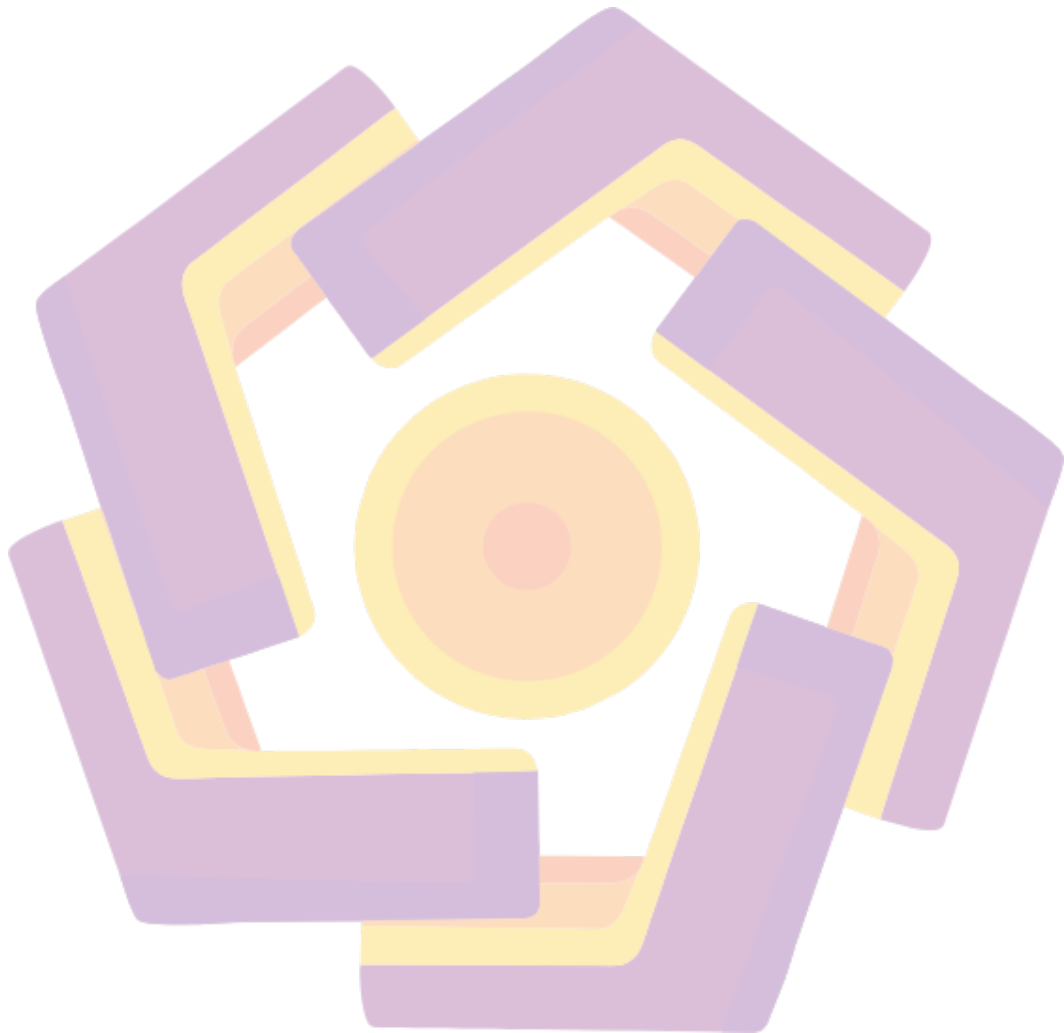
6. Pihak-pihak Terkait

Instansi atau perusahaan yang memberikan izin dan dukungan dalam pelaksanaan penelitian, terima kasih atas kerjasama yang baik. Semua bantuan dan dukungan ini

memberikan inspirasi dan motivasi bagi penulis dalam menyelesaikan skripsi ini. Semoga hasil penelitian ini dapat memberikan manfaat dan kontribusi positif bagi perkembangan ilmu pengetahuan dan teknologi.

Yogyakarta, 31 Januari 2025

Penulis



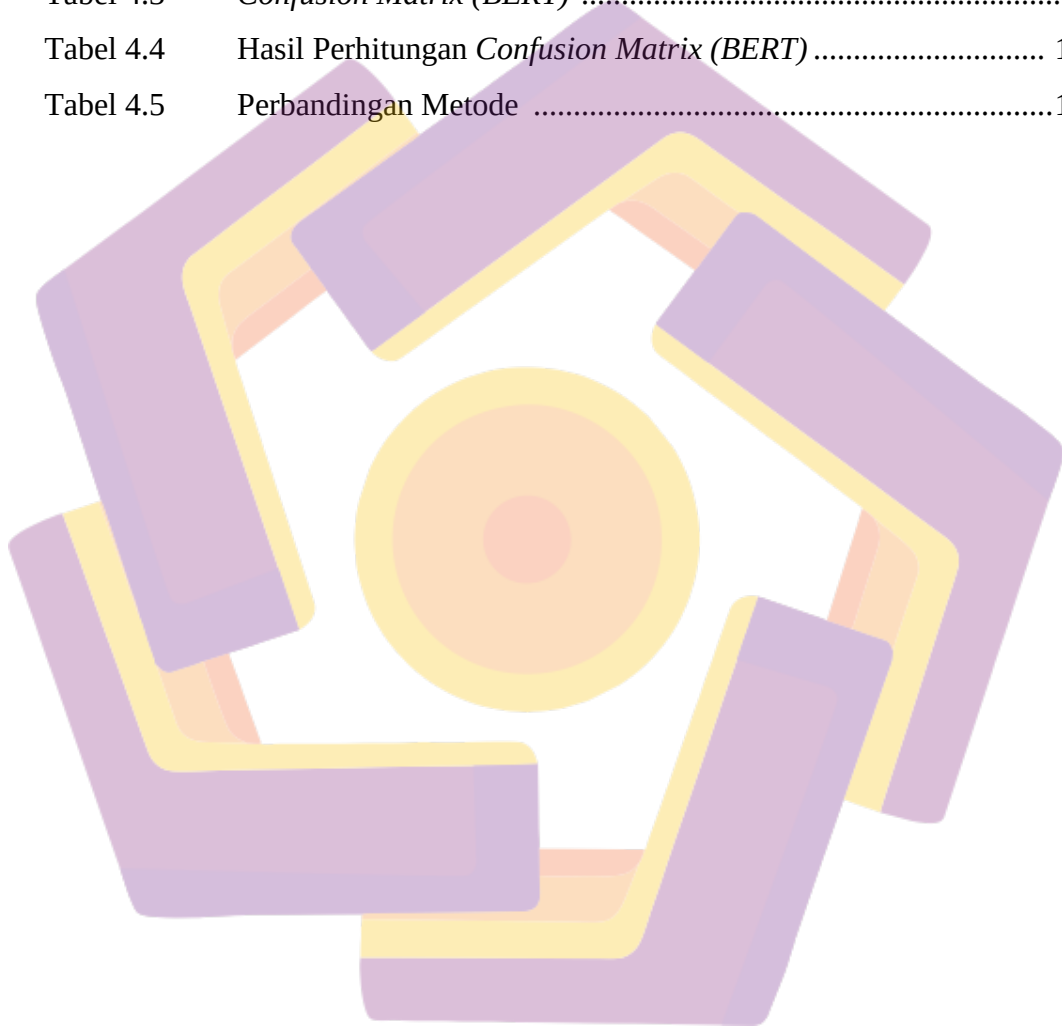
DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xvii
INTISARI	xxiii
ABSTRACT.....	xxiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1 Studi Literatur	7
2.2 Dasar Teori.....	19

BAB III METODE PENELITIAN	25
3.1 Objek Penelitian.....	25
3.2 Alur Penelitian	28
3.3 Alat dan Bahan.....	31
BAB IV HASIL DAN PEMBAHASAN	33
4.1 Deskripsi Data.....	33
4.2 Instalasi Library	33
4.3 Data Ulasan Shopee	34
4.4 Menyimpan Data yang Telah Diproses ke dalam File CSV.....	36
4.5 <i>Exploratory Data Analysis (EDA)</i>	38
4.6 Konversi Score Ratings ke Sentimen.....	45
4.7 Pelabelan Sentimen.....	46
4.8 <i>Preprocessing</i> Teks untuk Analisis Sentimen	48
4.9 Membuat <i>Pie Chart</i> untuk Distribusi Sentimen.....	50
4.10 Visualisasi <i>Word Cloud</i> Berdasarkan Sentimen	53
4.11 Implementasi Model <i>Naive Bayes</i> dan <i>BERT</i>	59
4.12 Evaluasi Model	85
4.13 Hasil Pengujian Model.....	85
4.14 Analisis Hasil	101
BAB V PENUTUP	106
5.1 Kesimpulan	106
5.2 Saran	106
REFERENSI	108
LAMPIRAN.....	112

DAFTAR TABEL

Tabel 2.1	Keaslian Penelitian	9
Tabel 4.1	<i>Confusion Matrix (Naive Bayes)</i>	89
Tabel 4.2	Hasil Perhitungan <i>Confusion Matrix (Naive Bayes)</i>	92
Tabel 4.3	<i>Confusion Matrix (BERT)</i>	97
Tabel 4.4	Hasil Perhitungan <i>Confusion Matrix (BERT)</i>	101
Tabel 4.5	Perbandingan Metode	105

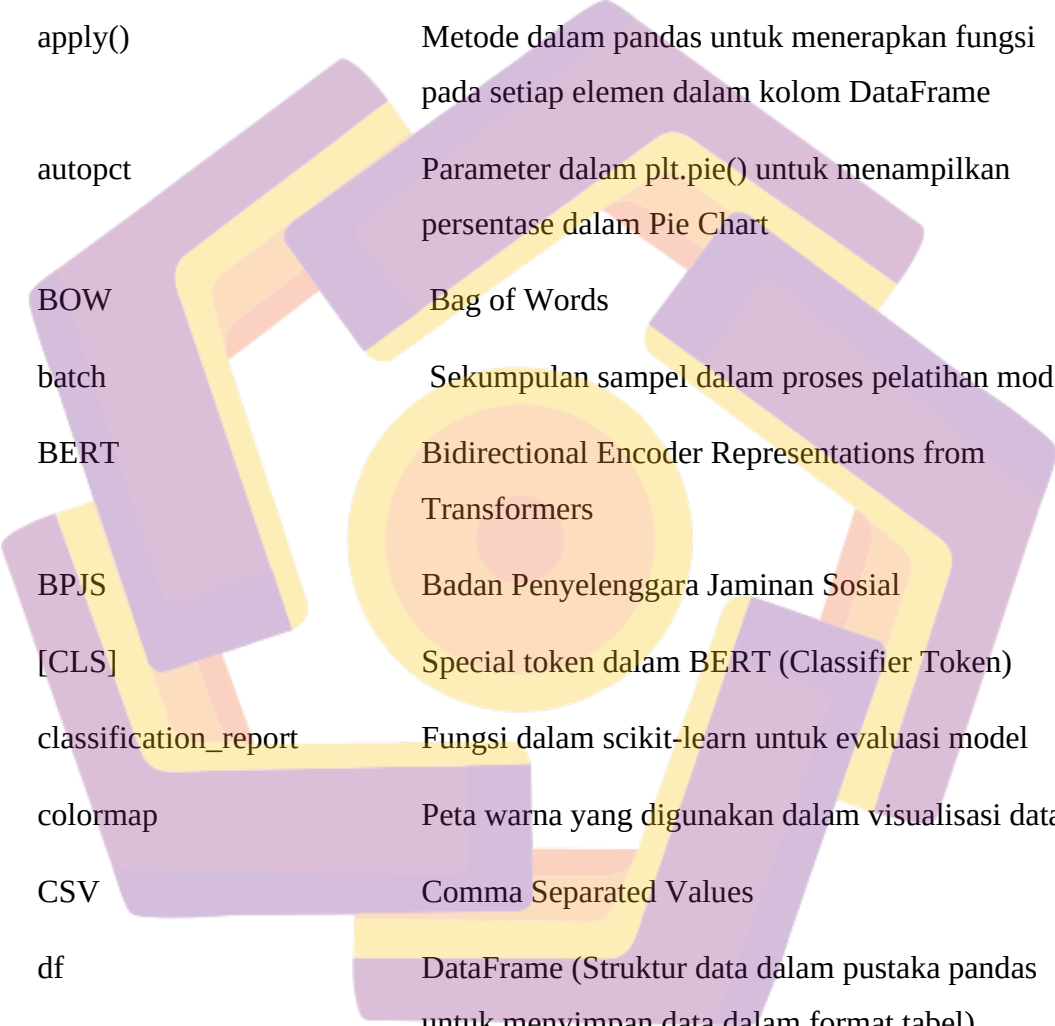


DAFTAR GAMBAR

Gambar 2.1	<i>NLP</i>	20
Gambar 2.2	<i>Naive Bayes</i>	21
Gambar 2.3	<i>BERT</i>	22
Gambar 3.1	Alur Penelitian	31
Gambar 4.1	Kode intalasi library	33
Gambar 4.2	Kode data ulasan shopee	35
Gambar 4.3	Kode menyimpan data	37
Gambar 4.4	<i>shopee_reviews_processed.csv</i>	37
Gambar 4.5	Kode EDA 1	38
Gambar 4.6	Kode EDA 2	39
Gambar 4.7	<i>Distribution of Ratings</i>	41
Gambar 4.8	<i>Distribution of Review Length</i>	42
Gambar 4.9	Kata yang sering muncul pada ulasan	43
Gambar 4.10	<i>Relationship Between Rating and Review Length</i>	44
Gambar 4.11	Kode konversi score ke sentimen	45
Gambar 4.12	Hasil konversi score ke sentimen	46
Gambar 4.13	Kode pelabelan sentimen	47
Gambar 4.14	Hasil output pelabelan sentimen	48
Gambar 4.15	Kode <i>preprocessing</i> teks	49
Gambar 4.16	Hasil output <i>preprocessing</i> teks	50
Gambar 4.17	Kode membuat <i>Pie Chart</i>	51
Gambar 4.18	<i>Pie Chart</i> persentase distribusi sentimen	52
Gambar 4.19	Kode Visualisasi Word Cloud	54
Gambar 4.20	Word Cloud	56
Gambar 4.21	Word Cloud Sentimen Positif	57
Gambar 4.22	Word Cloud Sentimen Netral	58
Gambar 4.23	Word Cloud Sentimen Negatif	58
Gambar 4.24	Membagi Data latih dan data uji <i>Naïve Bayes</i>	59
Gambar 4.25	Potongan data dari <i>Cleaned Review</i>	60

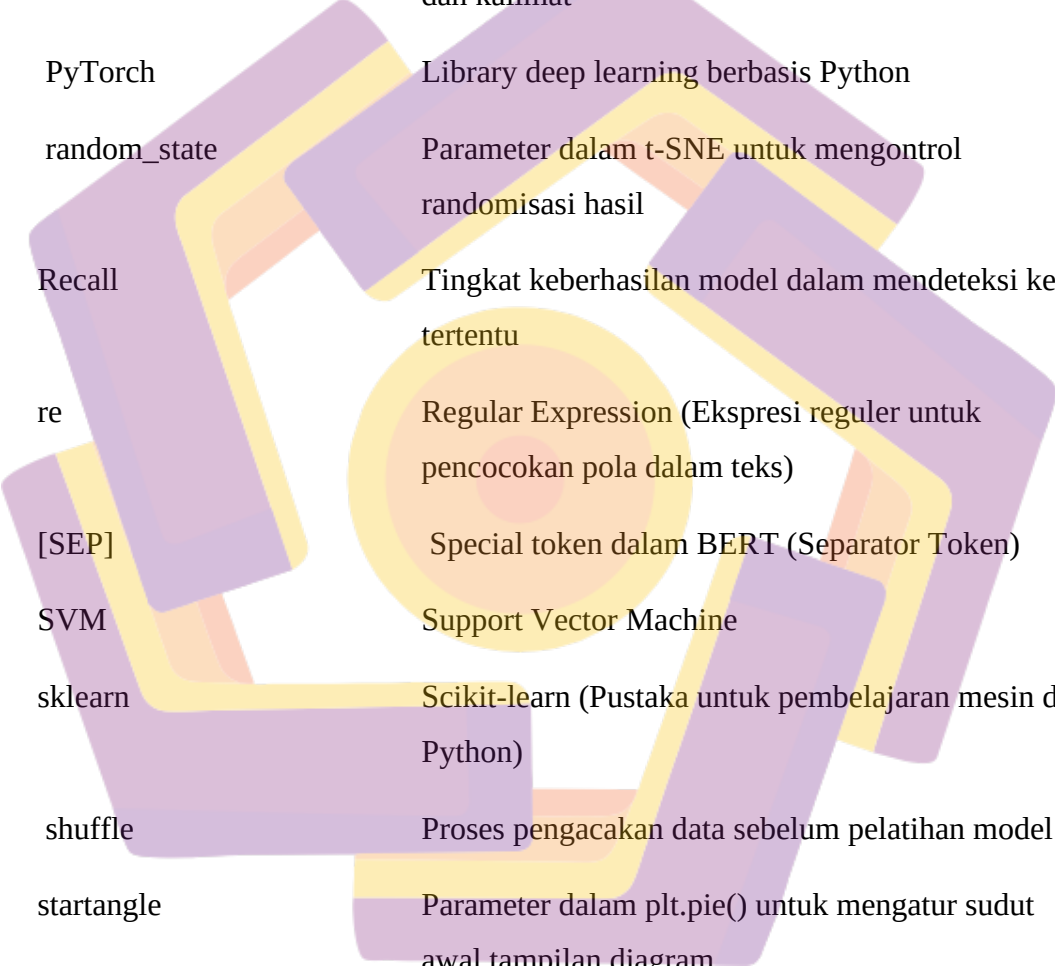
Gambar 4.26	Kode Perhitungan <i>TF-IDF</i>	61
Gambar 4.27	Hasil Perhitungan <i>TF-IDF</i>	62
Gambar 4.28	Kode Model <i>Naïve Bayes</i>	63
Gambar 4.29	<i>Classification Report Naive Bayes</i>	64
Gambar 4.30	Akurasi Model <i>Naive Bayes</i>	64
Gambar 4.31	Kode Visualisasi <i>Confusion Matrix Model Naive Bayes</i>	65
Gambar 4.32	<i>Confusion Matrix Naïve Bayes</i>	67
Gambar 4.33	Kode Model <i>BERT</i> 1	69
Gambar 4.34	Kode Model <i>BERT</i> 2	70
Gambar 4.35	Kode Model <i>BERT</i> 3	71
Gambar 4.36	Melatih dan mengevaluasi Model <i>BERT</i>	72
Gambar 4.37	Unduh file <i>BERT</i>	73
Gambar 4.38	Peforma Pelatihan Model <i>BERT</i>	74
Gambar 4.39	Kode Menghitung Embedding	75
Gambar 4.40	Hasil embedding	76
Gambar 4.41	Kode <i>T-SNE</i>	78
Gambar 4.42	<i>t-SNE</i> Visual.....	79
Gambar 4.43	Kode Predikasi dan Evaluasi Model <i>BERT</i>	80
Gambar 4.44	<i>Classification Report BERT</i>	81
Gambar 4.45	Akurasi <i>BERT</i>	82
Gambar 4.46	Kode Visualisasi <i>Confusion Matrix Model BERT</i>	82
Gambar 4.47	<i>Confusion Matrix BERT</i>	84
Gambar 4.48	Hasil <i>Naïve Bayes</i>	85
Gambar 4.49	Hasil <i>Confusion Matrix Naïve Bayes</i>	88
Gambar 4.50	Hasil <i>BERT</i>	93
Gambar 4.51	Hasil <i>Confusion Matrix BERT</i>	96

DAFTAR LAMBANG DAN SINGKATAN



Accuracy	Tingkat ketepatan keseluruhan prediksi model
AI	Artificial Intelligence (Kecerdasan Buatan)
App Store	Platform toko aplikasi untuk perangkat Apple
apply()	Metode dalam pandas untuk menerapkan fungsi pada setiap elemen dalam kolom DataFrame
autopct	Parameter dalam <code>plt.pie()</code> untuk menampilkan persentase dalam Pie Chart
BOW	Bag of Words
batch	Sekumpulan sampel dalam proses pelatihan model
BERT	Bidirectional Encoder Representations from Transformers
BPJS	Badan Penyelenggara Jaminan Sosial
[CLS]	Special token dalam BERT (Classifier Token)
classification_report	Fungsi dalam scikit-learn untuk evaluasi model
colormap	Peta warna yang digunakan dalam visualisasi data
CSV	Comma Separated Values
df	DataFrame (Struktur data dalam pustaka pandas untuk menyimpan data dalam format tabel)
EDA	Exploratory Data Analysis
F1-Score	Harmonic mean dari precision dan recall
FN	False Negatives
FP	False Positives

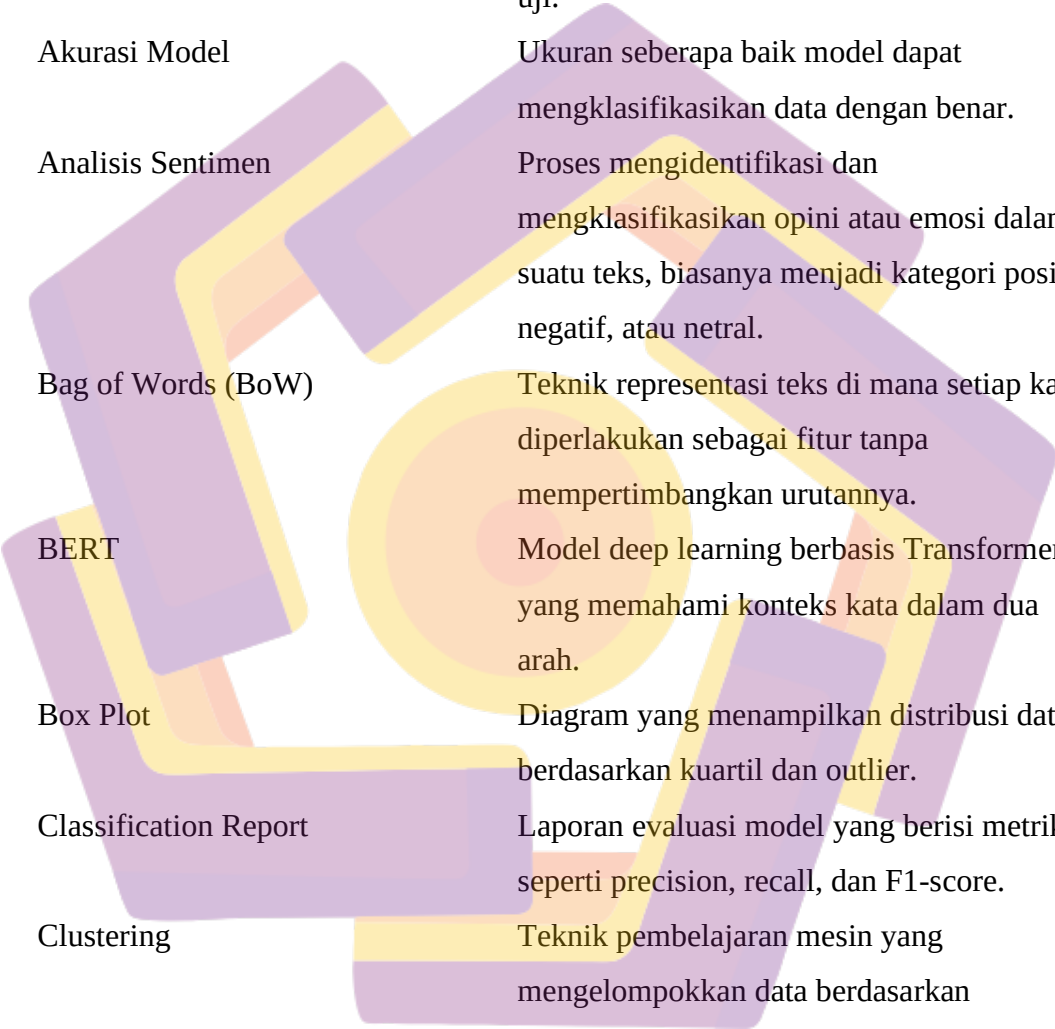
GPU	Graphics Processing Unit
Greens, Blues, Reds	Skema warna dalam pustaka Matplotlib untuk menunjukkan sentimen positif, netral, dan negatif
ID	Identifier (dalam konteks ID unik aplikasi)
interpolation='bilinear'	Parameter dalam Matplotlib untuk menghaluskan tampilan gambar pada grafik
JSON	JavaScript Object Notation
KDE	Kernel Density Estimation
KNN	K-Nearest Neighbor
logits	Output dari model neural network sebelum aktivasi
Macro Average	Rata-rata metrik tanpa mempertimbangkan jumlah sampel di setiap kelas
Naïve Bayes	Metode dalam machine learning untuk klasifikasi probabilistik
NBC	Naïve Bayes Classifier
nltk	Natural Language Toolkit (Pustaka Python untuk pemrosesan bahasa alami)
NLP	Natural Language Processing (Pengolahan Bahasa Alami)
perplexity	Parameter dalam t-SNE yang mengatur keseimbangan antara kepadatan lokal dan global dalam visualisasi
pip	Pip Installs Packages (alat untuk mengelola pustaka Python)



plt	Matplotlib.pyplot (Pustaka Python untuk membuat grafik dan visualisasi data)
Precision	Tingkat ketepatan model dalam klasifikasi
[PAD]	Special token dalam BERT (Padding Token)
punkt	Tokenizer dalam pustaka nltk untuk pemisahan kata dan kalimat
PyTorch	Library deep learning berbasis Python
random_state	Parameter dalam t-SNE untuk mengontrol randomisasi hasil
Recall	Tingkat keberhasilan model dalam mendeteksi kelas tertentu
re	Regular Expression (Ekspresi reguler untuk pencocokan pola dalam teks)
[SEP]	Special token dalam BERT (Separator Token)
SVM	Support Vector Machine
sklearn	Scikit-learn (Pustaka untuk pembelajaran mesin di Python)
shuffle	Proses pengacakan data sebelum pelatihan model
startangle	Parameter dalam plt.pie() untuk mengatur sudut awal tampilan diagram
stopwords	Kata-kata umum yang diabaikan dalam analisis teks
subplot()	Fungsi dalam Matplotlib untuk membuat beberapa grafik dalam satu tampilan
T-SNE	t-distributed Stochastic Neighbor Embedding
TF-IDF	Term Frequency-Inverse Document Frequency

<code>tight_layout()</code>	Metode dalam Matplotlib untuk mengatur tata letak agar tidak bertumpukan
TN	True Negatives
TP	True Positives
<code>train_test_split</code>	Fungsi dalam scikit-learn untuk membagi dataset
Trainer	Kelas dari library transformers Hugging Face
<code>value_counts()</code>	Metode dalam <code>pandas</code> untuk menghitung jumlah kemunculan nilai unik dalam suatu kolom
Weighted Average	Rata-rata metrik dengan mempertimbangkan jumlah sampel di setiap kelas
WandB	Weights & Biases
<code>word_tokenize</code>	Fungsi dalam <code>nltk</code> untuk membagi teks menjadi kata-kata individu
WordCloud	Pustaka Python untuk membuat visualisasi kata berdasarkan frekuensi kemunculan
<code>X_test</code>	Data uji untuk fitur input
<code>X_train</code>	Data latih untuk fitur input
<code>y_test</code>	Data uji untuk label output
<code>y_train</code>	Data latih untuk label output
<code>axis('off')</code>	Perintah dalam Matplotlib untuk menyembunyikan sumbu dalam grafik

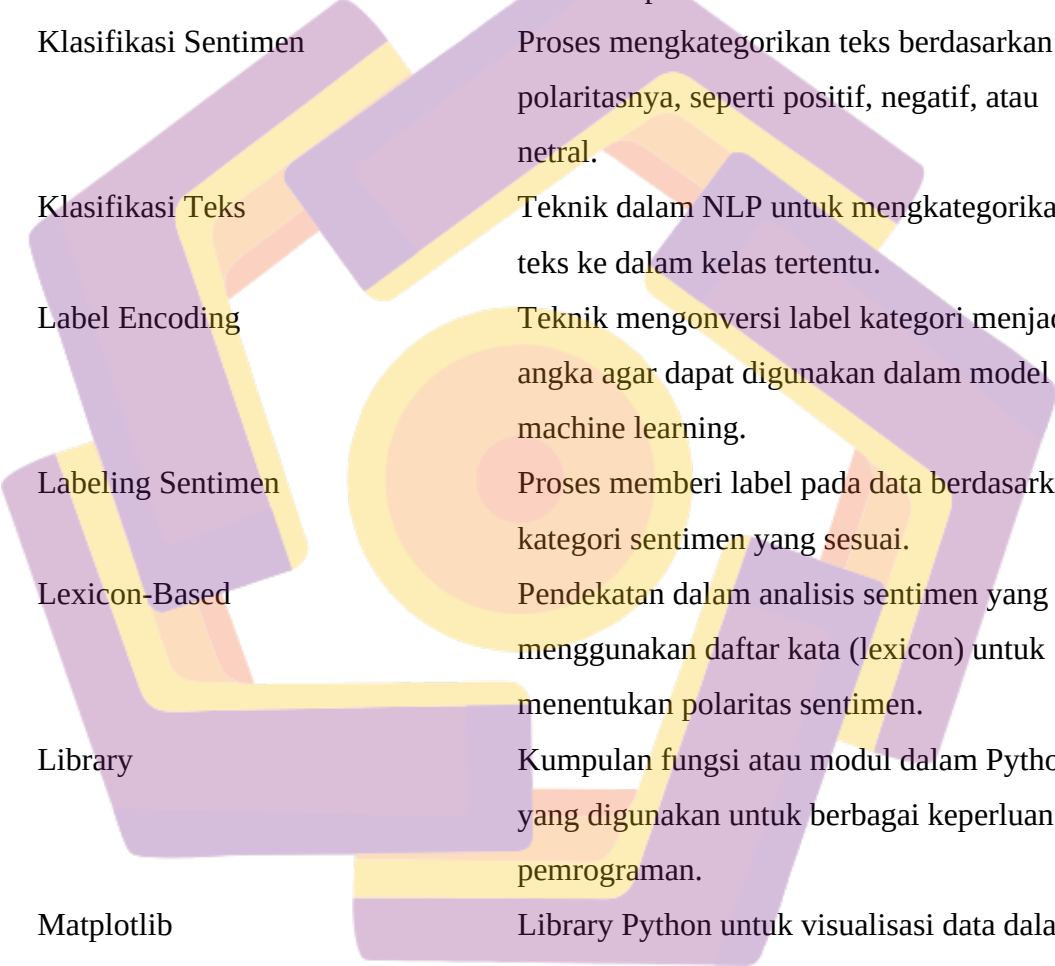
DAFTAR ISTILAH



Accuracy Score	Metrik yang menunjukkan persentase prediksi yang benar dibandingkan total data uji.
Akurasi Model	Ukuran seberapa baik model dapat mengklasifikasikan data dengan benar.
Analisis Sentimen	Proses mengidentifikasi dan mengklasifikasikan opini atau emosi dalam suatu teks, biasanya menjadi kategori positif, negatif, atau netral.
Bag of Words (BoW)	Teknik representasi teks di mana setiap kata diperlakukan sebagai fitur tanpa mempertimbangkan urutannya.
BERT	Model deep learning berbasis Transformer yang memahami konteks kata dalam dua arah.
Box Plot	Diagram yang menampilkan distribusi data berdasarkan kuartil dan outlier.
Classification Report	Laporan evaluasi model yang berisi metrik seperti precision, recall, dan F1-score.
Clustering	Teknik pembelajaran mesin yang mengelompokkan data berdasarkan kemiripan fitur.
Confusion Matrix	Tabel yang menunjukkan performa model klasifikasi dengan membandingkan prediksi dan label asli.
Counter	Teknik sederhana untuk menghitung frekuensi kemunculan kata dalam teks.



Crawling	Proses otomatis dalam mengumpulkan data dari berbagai sumber di internet.
CSV	Format file untuk menyimpan data tabular dalam bentuk teks dengan pemisah koma.
DataFrame	Struktur data dalam Pandas yang menyerupai tabel dan digunakan untuk menyimpan serta memproses data.
Ekspresi Reguler (Regex)	Pola yang digunakan untuk mencocokkan dan memanipulasi teks, seperti menghapus tanda baca dan angka.
E-commerce	Platform digital yang digunakan untuk melakukan transaksi jual beli secara online.
Evaluasi Model	Proses penilaian performa model menggunakan metrik seperti akurasi, precision, recall, dan F1-score.
Exploratory Data Analysis	Proses eksplorasi data untuk memahami karakteristik dan pola dalam dataset.
False Negative (FN)	Kasus di mana model memprediksi negatif tetapi seharusnya positif.
False Positive (FP)	Kasus di mana model memprediksi positif tetapi seharusnya negatif.
F1 Score	Ukuran kinerja model yang merupakan rata-rata harmonis dari precision dan recall.
Google Play Scraper	Pustaka Python yang digunakan untuk mengambil data ulasan dari Google Play Store.
Google Play Store	Layanan distribusi digital milik Google yang menyediakan aplikasi, game, dan ulasan dari pengguna.



Histogram	Grafik yang menunjukkan distribusi data berdasarkan frekuensi kemunculan nilai tertentu.
K-Nearest Neighbor	Algoritma berbasis kedekatan atau jarak antara data untuk klasifikasi dan regresi.
KDE	Metode statistik untuk memperkirakan distribusi probabilitas suatu variabel.
Klasifikasi Sentimen	Proses mengkategorikan teks berdasarkan polaritasnya, seperti positif, negatif, atau netral.
Klasifikasi Teks	Teknik dalam NLP untuk mengkategorikan teks ke dalam kelas tertentu.
Label Encoding	Teknik mengonversi label kategori menjadi angka agar dapat digunakan dalam model machine learning.
Labeling Sentimen	Proses memberi label pada data berdasarkan kategori sentimen yang sesuai.
Lexicon-Based	Pendekatan dalam analisis sentimen yang menggunakan daftar kata (lexicon) untuk menentukan polaritas sentimen.
Library	Kumpulan fungsi atau modul dalam Python yang digunakan untuk berbagai keperluan pemrograman.
Matplotlib	Library Python untuk visualisasi data dalam bentuk grafik dan diagram.
Deep Learning	Model kecerdasan buatan yang menggunakan jaringan saraf buatan untuk memahami pola dalam data.
Multinomial Naive Bayes	Jenis Naive Bayes yang cocok untuk data dengan distribusi frekuensi kata, seperti teks.

Naive Bayes

Algoritma klasifikasi berbasis probabilitas yang sering digunakan dalam pemrosesan teks.

NLP

Cabang kecerdasan buatan (AI) yang berkaitan dengan pemrosesan dan pemahaman bahasa manusia oleh komputer.

Normalisasi Teks

Proses mengubah teks menjadi format standar, seperti menghilangkan huruf kapital dan tanda baca.

NumPy

Library Python untuk komputasi numerik dan operasi matriks atau array.

Outlier

Data yang memiliki nilai ekstrem dan berbeda jauh dari mayoritas data lainnya.

Overfitting

Kondisi ketika model terlalu menyesuaikan dengan data latih sehingga kurang mampu menggeneralisasi data baru.

Oversampling

Teknik dalam pemrosesan data untuk menyeimbangkan jumlah data pada kelas tertentu dengan menambah sampel.

Pandas

Library Python untuk manipulasi dan analisis data berbasis tabel (DataFrame).

Pembagian Dataset

Proses membagi data menjadi data latih dan data uji untuk melatih dan menguji model analisis sentimen.

Machine Learning

Metode kecerdasan buatan yang memungkinkan komputer belajar dari data untuk membuat prediksi atau keputusan.

Persepsi Publik

Pandangan atau opini masyarakat terhadap suatu hal berdasarkan pengalaman dan informasi yang diperoleh.



Pie Chart	Grafik berbentuk lingkaran yang digunakan untuk menunjukkan distribusi kategori dalam suatu dataset.
Pre-processing	Tahap awal dalam pengolahan data teks yang mencakup pembersihan dan standarisasi data.
Precision	Metrik evaluasi yang mengukur seberapa akurat model dalam memprediksi kelas tertentu.
Probabilitas Kondisional	Perhitungan probabilitas suatu peristiwa berdasarkan informasi yang sudah diketahui sebelumnya.
Rating	Penilaian pengguna terhadap aplikasi dalam bentuk bintang (1-5).
Recall	Metrik yang mengukur seberapa baik model dalam menemukan semua data dari kelas tertentu.
Reputasi Aplikasi	Citra atau persepsi publik terhadap aplikasi berdasarkan pengalaman pengguna.
Scatter Plot	Grafik yang menunjukkan hubungan antara dua variabel dalam bentuk titik-titik.
Scikit-learn	Library Python untuk machine learning, termasuk algoritma Naive Bayes.
Sentimen	Kategori opini pengguna yang dapat berupa positif, netral, atau negatif berdasarkan rating.
Sentimen Pengguna	Opini atau perasaan pengguna terhadap suatu produk atau layanan.
Stopwords	Kata-kata umum yang biasanya dihapus dalam pemrosesan teks

TF-IDF

Metode pembobotan kata berdasarkan frekuensi kemunculannya dalam dokumen.

Tokenisasi

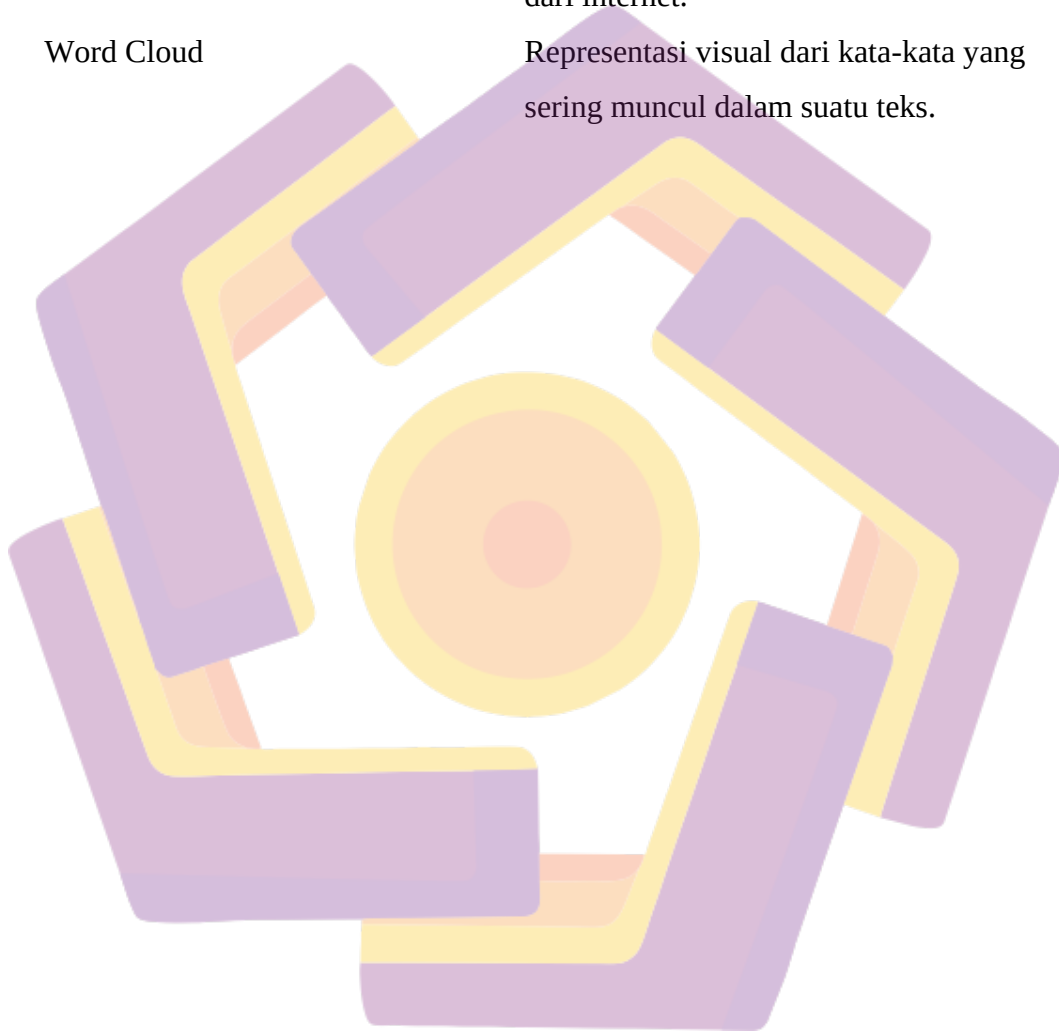
Proses membagi teks menjadi unit yang lebih kecil, seperti kata atau frasa.

Web Scraping

Teknik pengumpulan data secara otomatis dari internet.

Word Cloud

Representasi visual dari kata-kata yang sering muncul dalam suatu teks.



INTISARI

Perkembangan pesat *e-commerce* di Indonesia, khususnya pada aplikasi Shopee, telah menimbulkan banyak ulasan di *Google Play Store* yang dapat mempengaruhi citra dan kepercayaan pengguna. Penelitian ini bertujuan untuk menganalisis sentimen dari *review* pengguna aplikasi Shopee di *Google Play Store* guna memahami persepsi publik terhadap aplikasi tersebut. Masalah yang muncul adalah volume ulasan yang sangat besar dan beragam, sehingga memerlukan metode otomatis untuk melakukan analisis sentimen secara efisien. Sentimen pengguna yang tidak teridentifikasi dengan tepat dapat berdampak negatif pada reputasi aplikasi dan pengembangan lebih lanjut. Untuk menyelesaikan masalah ini, penelitian ini menggunakan kombinasi algoritma *Naive Bayes* dan *BERT* (*Bidirectional Encoder Representations from Transformers*). *Naive Bayes* digunakan karena keandalannya dalam klasifikasi teks, sementara *BERT* dipilih untuk menangkap makna kontekstual dari setiap kata dalam kalimat. Data dikumpulkan dari ulasan pengguna di *Google Play Store*, yang kemudian diproses dengan teknik *pre-processing* seperti *tokenisasi*, *stemming*, dan penghapusan *stopwords*. Model *Naive Bayes* dan *BERT* kemudian diterapkan untuk melakukan analisis sentimen, mengkategorikan ulasan ke dalam sentimen positif, negatif, atau netral.

Kata kunci: Analisis Sentimen, Shopee, *Naive Bayes*, *BERT*, *Google Play Store*.

ABSTRACT

The rapid development of e-commerce in Indonesia, especially the Shopee application, has given rise to many reviews on the Google Play Store which can affect the image and trust of users. This research aims to analyze sentiment from user reviews of the Shopee application on the Google Play Store in order to understand public perceptions of the application. The problem that arises is that the volume of reviews is very large and diverse, requiring automated methods to perform sentiment analysis efficiently. Inappropriately identified user sentiment can have a negative impact on the application's reputation and further development. To solve this problem, this research uses a combination of the Naive Bayes algorithm and BERT (Bidirectional Encoder Representations from Transformers). Naive Bayes was used because of its reliability in text classification, while BERT was chosen to capture the contextual meaning of each word in the sentence. Data is collected from user reviews on the Google Play Store, which are then processed using pre-processing techniques such as tokenization, stemming, and removing stopwords. Naive Bayes and BERT models are then applied to perform sentiment analysis, categorizing reviews into positive, negative, or neutral sentiment.

Keywords: Sentiment Analysis, Shopee, Naive Bayes, BERT, Google Play Store.