

**ANALISIS SENTIMEN ULASAN HOTEL DI YOGYAKARTA
MENGUNAKAN ALGORITMA DISTILBERT
(DISTILLATION VERSION OF BERT)**

SKRIPSI

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



disusun oleh

FAHRUL FUGUH VERDIAN

21.11.4331

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**ANALISIS SENTIMEN ULASAN HOTEL DI YOGYAKARTA
MENGUNAKAN ALGORITMA DISTILBERT
(DISTILLATION VERSION OF BERT)**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

FAHRUL FUGUH VERDIAN

21.11.4331

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN ULASAN HOTEL DI YOGYAKARTA
MENGUNAKAN ALGORITMA DISTILBERT
(DISTILLATION VERSION OF BERT)**

yang disusun dan diajukan oleh

Fahrul Fuguh Verdian

21.11.4331

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 11 Agustus 2025

Dosen Pembimbing,



Theopilus Bayu Sasongko, S.Kom, M.Eng
NIK. 190302375

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS SENTIMEN ULASAN HOTEL DI YOGYAKARTA
MENGUNAKAN ALGORITMA DISTILBERT
(DISTILLATION VERSION OF BERT)

yang disusun dan diajukan oleh

Fahrul Fuguh Verdian

21.11.4331

Telah dipertahankan di depan Dewan Penguji
pada tanggal 11 Agustus 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ike Verawati, S.Kom., M.Kom.
NIK. 190302237

Nuri Cahyono, M.Kom.
NIK. 190302278

Theopilus Bayu Sasongko, S.Kom., M.Eng
NIK. 190302375

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 11 Agustus 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusri, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Fahrul Fuguh Verdian
NIM : 21.11.4331

Menyatakan bahwa Skripsi dengan judul berikut:

**ANALISIS SENTIMEN ULASAN HOTEL DI YOGYAKARTA
MENGUNAKAN ALGORITMA DISTILBERT (DISTILLATION
VERSION OF BERT)**

Dosen Pembimbing : Theopilus Bayu Sasongko, S.Kom, M.Eng

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, 11 Agustus 2025

Yang Menyatakan,


F76ANX010373706
Fahrul Fuguh Verdian

HALAMAN PERSEMBAHAN

Segala puji dan syukur kepada Allah SWT atas limpahan nikmat, rahmat dan karunia-Nya yang telah memberikan kesehatan, kekuatan, kesabaran dan pertolongan-Nya disepanjang penyelesaian skripsi ini. Terima kasih kepada semua pihak yang sudah mendoakan, membimbing, memotivasi dan mendukung selama pembuatan skripsi ini. Tak lupa penulis mengucapkan terima kasih kepada:

1. Diri sendiri yang selalu berjuang menjadi pribadi yang lebih baik dari waktu ke waktu.
2. Kedua orang tua yang selalu mendukung dalam diam dan bersabar dalam mendidik penulis dan percaya kepada penulis dalam menyelesaikan skripsi.
3. Bapak Theopilus Bayu Sasongko, S.Kom, M.Eng., selaku dosen pembimbing yang selalu memberikan masukan, arahan serta kesabaran yang membuat skripsi terselesaikan.
4. Teman Nongkrong “JMK48”, Khalis si “Mas Imut”, Azzam si “Mas Fuad”, Djojo si “Mas Rusdi”, “Onil”, Arkhan si “Mas Amba”.Terima kasih saya sampaikan kepada kalian yang telah menyediakan tawa dan waktu untuk menemani penulis dalam menyelesaikan skripsi.
5. Teman-teman kuliah yang selalu mendukung dan senantiasa ada di waktu saya terpuruk.
6. Tempat cafe yang berada di daerah banyuraden sleman yang selalu buka 24 jam. Tanda terima kasih ini peneliti haturkan karena selalu menyediakan tempat untuk menyelesaikan skripsi.

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah Yang Maha Esa atas limpahan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul " Analisis Sentimen Ulasan Hotel Di Yogyakarta Menggunakan Algoritma Distilbert (Distillation Version Of Bert)" dengan lancar dan tepat waktu. Skripsi ini disusun guna memenuhi salah satu syarat untuk memperoleh gelar S.Kom pada Program Studi S1 Informatika di Universitas Amikom Yogyakarta. Penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Theopilus Bayu Sasongko, S.Kom, M.Eng., selaku dosen pembimbing yang telah memberikan bimbingan, arahan, dan motivasi selama proses penyusunan skripsi ini.
2. Tim Dosen Penguji yang telah memberikan masukan dan saran yang sangat berharga untuk perbaikan skripsi ini.
3. Kedua orang tua yang senantiasa memberikan dukungan moril dan materil serta doa yang tiada henti.
4. Universitas Amikom Yogyakarta yang telah memberikan kesempatan dan dukungan dalam penelitian ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari berbagai pihak untuk perbaikan di masa mendatang. Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi pembaca dan pihak-pihak yang berkepentingan.

Yogyakarta, 01 Agustus 2025

Penulis

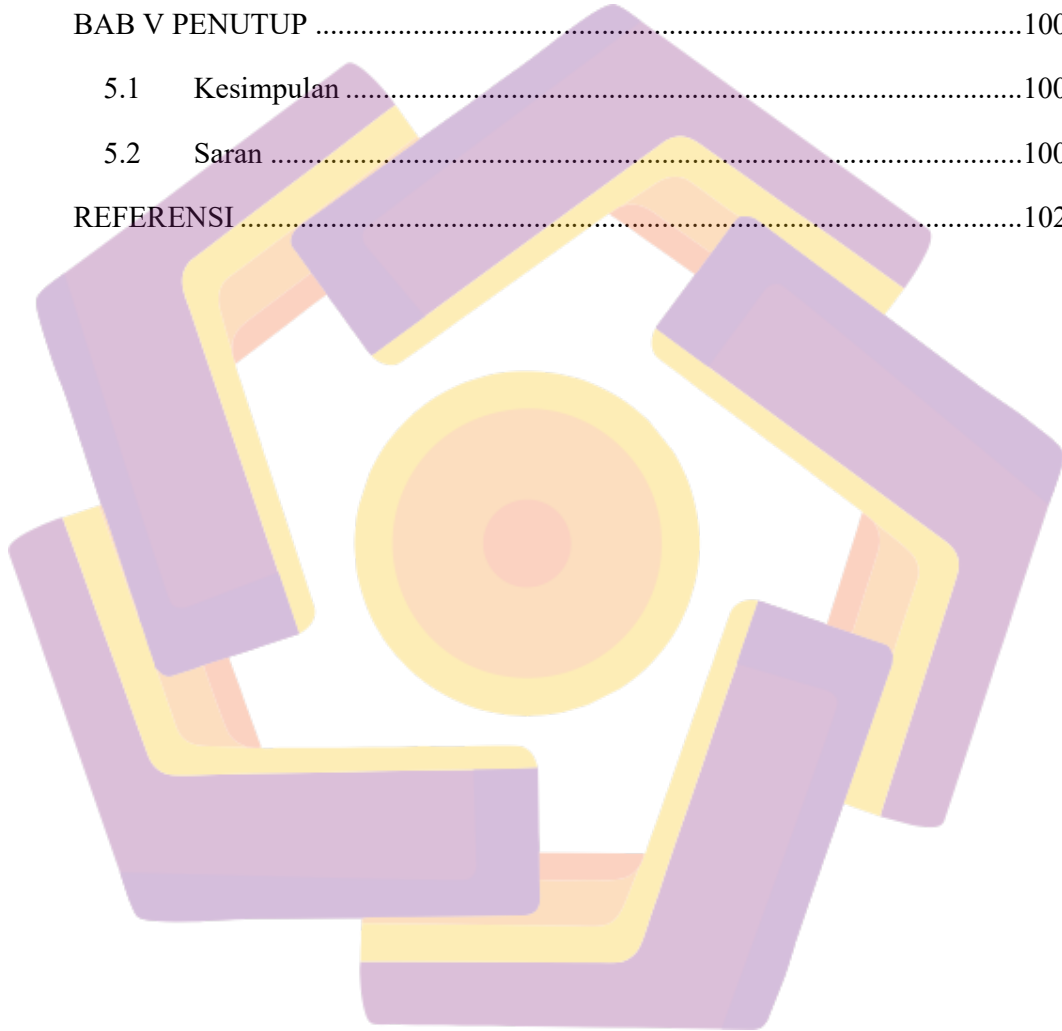
DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMBANG DAN SINGKATAN	xv
DAFTAR ISTILAH	xvi
INTISARI	xviii
ABSTRACT.....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori.....	9

2.2.1	Google Colab	9
2.2.2	NLP	10
2.2.3	Deep Learning.....	10
2.2.4	Analisis Sentimen	11
2.2.5	Python	14
2.2.6	BERT	15
2.2.7	DistilBERT	17
2.2.8	Lexicon Based.....	19
2.2.9	ROS.....	20
2.2.10	Matriks Evaluasi	21
2.2.11	Wordcloud.....	22
BAB III METODE PENELITIAN		23
3.1	Objek Penelitian.....	23
3.2	Alur Penelitian	23
3.2.1	Identifikasi Masalah.....	24
3.2.2	Litelatur Review.....	25
3.2.3	Pengumpulan data	25
3.2.4	EDA	26
3.2.5	Pre-procesing data.....	30
3.2.6	Labeling data.....	37
3.2.7	Balancing data.....	39
3.2.8	Data Preparation.....	39
3.2.9	Modeling	41
3.2.10	Evaluasi.....	42
3.2.11	Kesimpulan Dan Saran	42

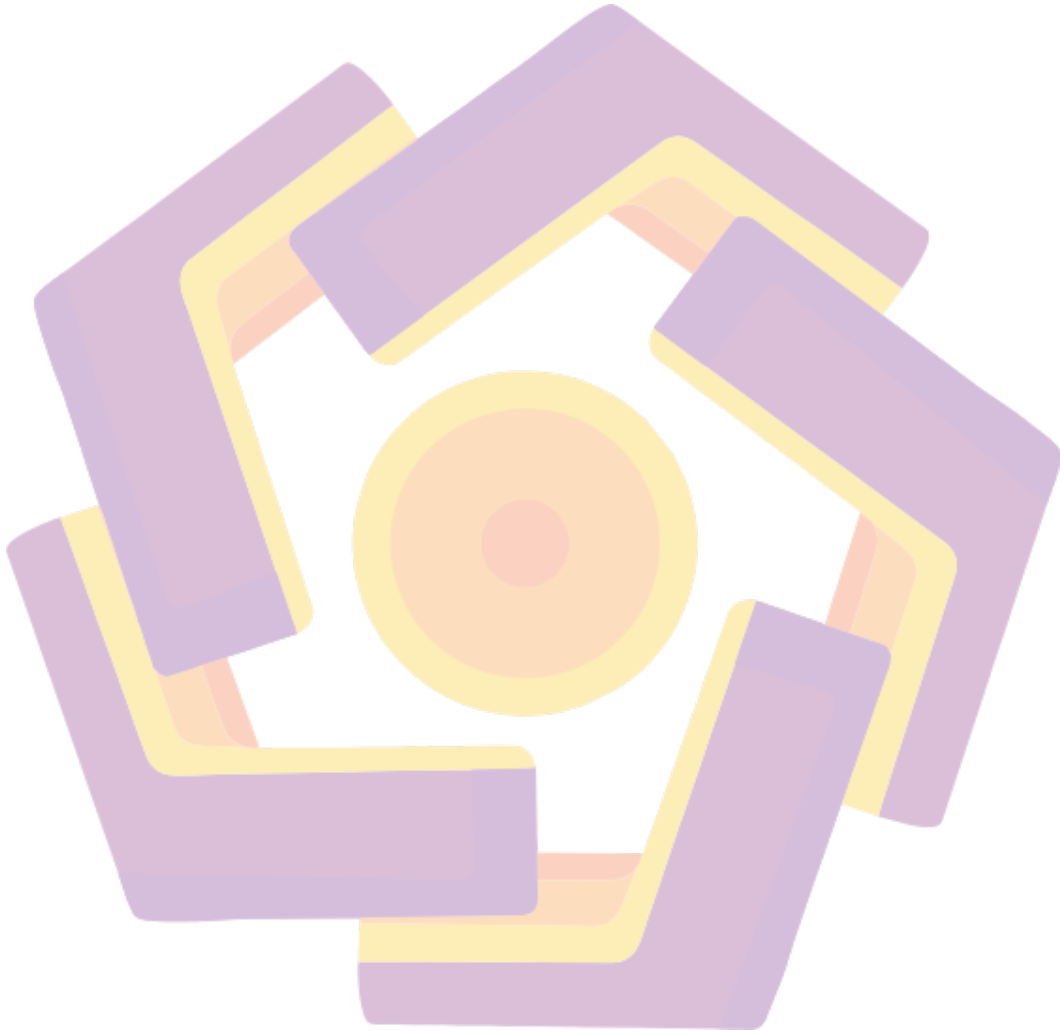
3.3	Alat dan Bahan.....	43
BAB IV HASIL DAN PEMBAHASAN		44
4.1	Hasil pengumpulan data.....	44
4.2	EDA	46
4.2.1	Translate.....	47
4.2.2	Mengecek Nilai Null.....	50
4.2.3	Remove Duplicates	51
4.3	Pre-processing Data	53
4.3.1	Case Folding	54
4.3.2	Remove Multi-Space	54
4.3.3	Filtering.....	55
4.3.4	Clean Hashtag	56
4.3.5	Remove Punctuation	57
4.3.6	Clean Emoji	58
4.3.7	Stopword removal.....	59
4.3.8	Stemming	60
4.4	Labeling Data.....	60
4.5	Balancing Data.....	61
4.6	Data Preparation.....	63
4.6.1	Tokenezing.....	64
4.6.2	Spliting Data	65
4.6.3	Data Loader.....	66
4.7	Modeling	66
4.7.1	Arsitektur Model DistilBERT	67
4.7.2	Setting Optimizer And Hyperparameter	68

4.7.3	Training Model	69
4.8	Evaluasi	70
4.8.1	Evaluasi Model	71
4.8.2	Evaluasi Word Cloud	84
BAB V PENUTUP		100
5.1	Kesimpulan	100
5.2	Saran	100
REFERENSI		102



DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	7
Tabel 4.1 Tabel Perbandingan Dataset Setelah Data Duplicate Dihapus	52
Tabel 4.2 Perbandingan Akurasi Dan Loss Setiap Hotel.....	69



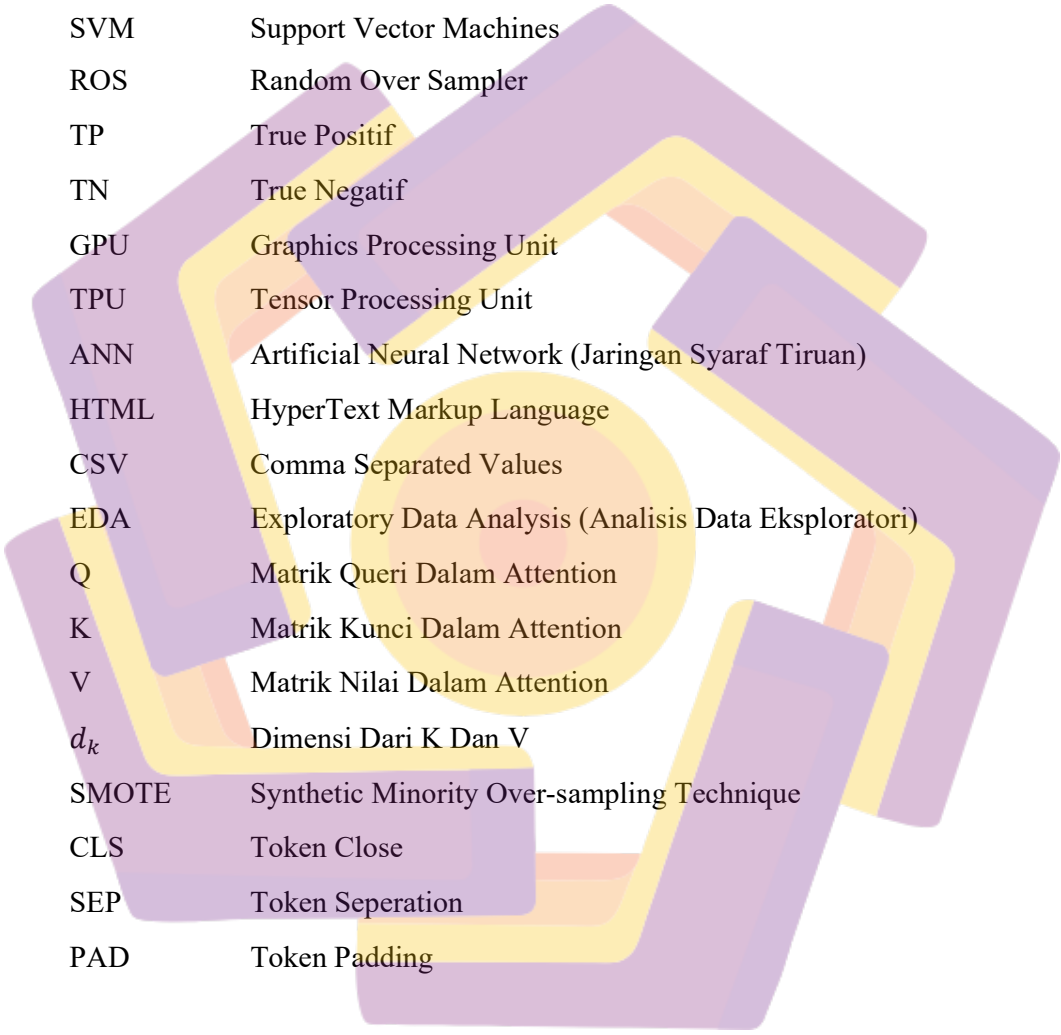
DAFTAR GAMBAR

Gambar 2.1 Arsitektur Deep Learning.....	11
Gambar 2.2 Ilustrasi Proses Sentimen Analisis	12
Gambar 2.3 Arsitektur Tranformer	16
Gambar 2.4 Representasi input pada BERT	17
Gambar 2.5 Perbedaan Arsitektur BERT Dan DistilBERT	18
Gambar 2.6 ROS(Random Oversampling)	20
Gambar 2.7 Matrix Evaluasi 3x3	21
Gambar 3.1 Alur Penelitian	24
Gambar 3.2 Pengumpulan Data	25
Gambar 3.3 Alur Scraping Data.....	26
Gambar 3.4 Alur EDA	27
Gambar 3.5 Alur Penanganan Missing Value.....	28
Gambar 3.6 Gambar Alur Remove Duplicate.....	29
Gambar 3.7 Alur Pre-Processing Data.....	30
Gambar 3.8 Alur Case Folding	31
Gambar 3.9 Alur Remove Multi Space.....	32
Gambar 3.10 Alur Filtering.....	33
Gambar 3.11 Alur Remove Hashtag.....	34
Gambar 3.12 Alur Remove Punctuation.....	35
Gambar 3.13 Alur Clean Emoji	36
Gambar 3.14 Alur Stemming.....	37
Gambar 3.15 Alur Labeling Data.....	38
Gambar 3.16 Alur ROS.....	39
Gambar 3.17 Alur Data Preparation	40
Gambar 4.1 Riview Dari Web Booking.com.....	44
Gambar 4.2 Dataset Riview Dari Scraping.....	45
Gambar 4.3 Jumlah Data Sebelum Translate.....	46
Gambar 4.4 Jumlah data Bahasa Indonesia dan Bukan	47
Gambar 4.5 Contoh Data Yang Tidak Terdeteksi.....	48

Gambar 4.6 Jumlah Data Setelah Data Tidak Terdeteksi dihapus.....	48
Gambar 4.7 Sebelum dan Sesudah Translate.....	49
Gambar 4.8 Jumlah Data Sesudah Translate	49
Gambar 4.9 Data Yang Terdeteksi Bukan Indonesia.....	50
Gambar 4.10 Jumlah Nilai Null Dari Dataset.....	51
Gambar 4.11 Contoh Data Duplikasi.....	52
Gambar 4.12 Perbedaan Dataset Sebelum Dan Sesudah Case Folding.....	54
Gambar 4.13 Perbedaan Dataset Sebelum Dan Sesudah Remove Multi-Space....	55
Gambar 4.14 perbedaan dataset sebelum dan sesudah Filtering.....	56
Gambar 4.15 Perbedaan Dataset Sebelum Dan Sesudah Clean Hashtag.....	57
Gambar 4.16 Perbedaan Dataset Sebelum Dan Sesudah Remove Punctuation...58	
Gambar 4.17 Perbedaan Dataset Sebelum Dan Sesudah Clean Emoji	59
Gambar 4.18 Perbedaan Dataset Sebelum Dan Sesudah Stopword Removal	59
Gambar 4.19 Perbedaan Dataset Sebelum Dan Sesudah Stemming.....	60
Gambar 4.20 Dataset Setelah Ditambahkan Label	61
Gambar 4.21 Perbandingan Jumlah Class	62
Gambar 4.22 Ilustrasi Diagram Batang Untuk Semua Class Dataset	63
Gambar 4.23 Gambar Dataset Setelah Melalui Tokenisasi	64
Gambar 4.24 Gambar Dataset Setelah Melalui Encoding	65
Gambar 4.25 Gambar Dataset Setelah Melalui Tahap Masking.....	65
Gambar 4.26 Gambar Jumlah Data Training Dan Testing	66
Gambar 4.27 Gambar Arsitektur DistilBERT	67
Gambar 4.28 Confusion Matrix Hotel Jambuluwuk Malioboro	72
Gambar 4.29 Classification Report Hotel Jambuluwuk Malioboro.....	73
Gambar 4.30 Confusion Matrix Aveta Hotel Malioboro	74
Gambar 4.31 Classification Report Aveta Hotel Malioboro.....	75
Gambar 4.32 Confusion Matrix Phoenix Hotel.....	77
Gambar 4.33 Classification Report Phoenix Hotel.....	78
Gambar 4.34 Confusion Matrix Melia Purosani	79
Gambar 4.35 Classification Report Melia Purosani	80
Gambar 4.36 Confusion Matrix Gallery Prawirotaman	82

Gambar 4.37 Classification Report Gallery Prawirotaman	83
Gambar 4.38 Visualisasi Wordcloud Untuk Sentiment Positif.....	85
Gambar 4.39 Visualisasi Wordcloud Untuk Sentiment Negatif.....	86
Gambar 4.40 Visualisasi Wordcloud Untuk Sentiment Netral	87
Gambar 4.41 Visualisasi Wordcloud Untuk Sentiment positif.....	88
Gambar 4.42 Visualisasi Wordcloud Untuk Sentiment negatif.....	89
Gambar 4.43 Visualisasi Wordcloud Untuk Sentiment Netral	90
Gambar 4.44 Visualisasi Wordcloud Untuk Sentiment Positif.....	91
Gambar 4.45 Visualisasi Wordcloud Untuk Sentiment Negatif.....	92
Gambar 4.46 Visualisasi Wordcloud Untuk Sentiment Netral	93
Gambar 4.47 Visualisasi Wordcloud Untuk Sentiment Positif.....	94
Gambar 4.48 Visualisasi Wordcloud Untuk Sentiment Negatif.....	95
Gambar 4.49 Visualisasi Wordcloud Untuk Sentiment Netral	96
Gambar 4.50 Visualisasi Wordcloud Untuk Sentiment Positif.....	97
Gambar 4.51 Visualisasi Wordcloud Untuk Sentiment Negatif.....	98
Gambar 4.52 Visualisasi Wordcloud Untuk Sentiment Netral	99

DAFTAR LAMBANG DAN SINGKATAN



DistilBERT	Distillation Version Of Bert
Adam	Adaptive Moment Estimation
AdamW	Adaptive Moment Estimation weight decay
SVM	Support Vector Machines
ROS	Random Over Sampler
TP	True Positif
TN	True Negatif
GPU	Graphics Processing Unit
TPU	Tensor Processing Unit
ANN	Artificial Neural Network (Jaringan Syaraf Tiruan)
HTML	HyperText Markup Language
CSV	Comma Separated Values
EDA	Exploratory Data Analysis (Analisis Data Eksploratori)
Q	Matrik Queri Dalam Attention
K	Matrik Kunci Dalam Attention
V	Matrik Nilai Dalam Attention
d_k	Dimensi Dari K Dan V
SMOTE	Synthetic Minority Over-sampling Technique
CLS	Token Close
SEP	Token Seperation
PAD	Token Padding

DAFTAR ISTILAH

Akurasi	Persentase Prediksi Yang Benar Dibanding Total Semua Prediksi
Web Scrapping	Mengumpulkan Informasi Dari Situs Web Secara Terstruktur
Downsapling	Mengurangi Jumlah Data Untuk Menyeimbangkan Dataset
Upsampling	Menambahkan Jumlah Data Untuk Menyeimbangkan Dataset
Precision	Seberapa Banyak Dari Prediksi Positif Yang Benar-Benar Benar
Recall	Seberapa Banyak Data Positif Yang Berhasil Ditemukan Oleh Model
F1-Score	Mengukur Keseimbangan Antara Precision Dan Recall
Dataset	Kumpulan Data
Case Folding	Mengubah Huruf Kalimat
Stopword	Kata-Kata Yang Dianggap Tidak Penting Dalam Pemrosesan Teks/Tidak Memiliki Makna.
Stemming	Mengubah Kata Ke Bentuk Dasarnya
Tokenizing	Memecah Teks Menjadi Unit-Unit Kecil Token
Ekstraksi Fitur	Mengubah Data Mentah Menjadi Representasi Numerik
Hyperparameter	Parameter Yang Diubah-Ubah
Neuron	Unit Dasar Pemrosesan
Hashtag	Tanda Pagar
Bias	Kesalahan Sistematis Yang Membuat Model Keliru
Interpreted	Kode Langsung Dibaca Dan Dijalankan Tanpa Perlu Dikompilasi
Pandas	Modul Python Untuk Analisis Data
Matplotlib	Library Visualisasi Data Di Python
Pre-Traning	Model Yang Sudah Dilatih

Encoder	Proses Mengubah Data Tesk Menjadi Format Numerik
Self-Attention	Komponen Inti Dalam Arsitektur Transformer
Feed Forward	Memproses Setiap Token Secara Mandiri
Normalisasi	Menstandarkan Skala Nilai Data
Hugging Face	Situs Penyedia Banyak Model
Output	Hasil Yang Dihasilkan Oleh Suatu Proses, Fungsi, Atau Model Setelah Menerima Input.
Transformer	Memproses Data Dalam Bentuk Urutan Mengandalkan Mekanisme Attention
Layers	Sekumpulan Unit (Neuron)
Optimizer	Mengarahkan Model Agar Semakin Baik Dari Epoch Ke Epoch
Epoch	Satu Kali Siklus Penuh Seluruh Data Training Diproses Oleh Model
Batch Size	Jumlah Data (Sampel) Yang Diproses Oleh Model Dalam Satu Kali Iterasi
Loss	Angka Yang Menunjukkan Seberapa Buruk Prediksi Model Dibandingkan Dengan Nilai Yang Sebenarnya
Overfitting	Model Belajar Terlalu Baik Dari Data Training
Attention Mask	Petunjuk Yang Diberikan Ke Model Transformer
Word Embedding	Representasi Kata Dalam Bentuk Vektor Angka
Position Embedding	Cara Untuk Memberikan Informasi Urutan Posisi Kata Ke Dalam Model Transformer
Wordcloud	Kumpulan Kata Dalam Bentuk Gambar
Confusion Matrix	Tabel Evaluasi Kinerja Model Klasifikasi
Classification Report	Ringkasan Metrik Evaluasi Untuk Model Klasifikasi

INTISARI

Perkembangan teknologi yang pesat telah mengubah cara manusia dalam menerima dan menyampaikan informasi. Jika sebelumnya informasi disebarkan secara lisan, kini penyebaran dapat dilakukan hanya melalui sentuhan jari. Salah satu contohnya adalah ulasan pelanggan terhadap layanan hotel, yang kini banyak disampaikan melalui platform digital. Ulasan ini dapat berdampak positif maupun negatif bagi reputasi hotel, tergantung pada pengalaman yang dituliskan oleh pelanggan.

Namun, banyaknya ulasan yang tersedia sering kali menyulitkan calon pelanggan dalam memperoleh informasi yang relevan. Oleh karena itu, analisis sentimen menjadi pendekatan penting untuk mengelompokkan ulasan ke dalam kategori positif, negatif, atau netral. Penelitian ini bertujuan melakukan analisis sentimen terhadap ulasan hotel di Indonesia dengan menggunakan algoritma *DistilBERT* (*Distilled Version of BERT*), guna menyediakan informasi yang valid bagi pelanggan serta memberikan insight bagi pihak pengelola hotel.

Pemilihan *DistilBERT* didasarkan pada kemampuannya yang efisien, dengan ukuran model lebih kecil dari *BERT*, sehingga lebih cepat dan ringan secara komputasi. Penelitian ini juga melakukan tuning parameter untuk meningkatkan performa model, dengan konfigurasi terbaik pada *batch_size* 16, *epoch* 5, dan *learning_rate* 0.0001 menggunakan optimizer *AdamW*. Hasil evaluasi menunjukkan model menghasilkan *akurasi*, *precision*, *recall*, dan *F1-score* masing-masing sebesar 96% pada hotel Gallery Prawirotaman. Hal ini menunjukkan bahwa *DistilBERT* mampu menganalisis sentimen ulasan hotel secara efektif dan akurat.

Kata kunci: DistilBERT, Ulasan Hotel, Sentimen Analisis, Bahasa Indonesia, Hotel.

ABSTRACT

The rapid advancement of technology has significantly transformed the way humans receive and communicate information. Whereas information was previously conveyed orally, it can now be easily disseminated with just a touch of a finger. One example of this transformation is customer reviews of hotel services, which are now predominantly shared through digital platforms. These reviews can have either a positive or negative impact on a hotel's reputation, depending on the experiences described by the customers.

However, the abundance of reviews often makes it difficult for prospective customers to obtain relevant and accurate information. Therefore, sentiment analysis becomes a crucial approach to categorize reviews into positive, negative, or neutral sentiments. This study aims to conduct sentiment analysis on hotel reviews in Indonesia using the DistilBERT algorithm (Distilled Version of BERT), in order to provide more valid information for future customers and valuable insights for hotel management.

DistilBERT was selected for its efficiency, as it is a smaller and faster version of BERT that requires less computational resources. This study also involved tuning several parameters to enhance the model's performance, with the optimal configuration being a batch size of 16, 5 epochs, and a learning rate of 0.0001 using the AdamW optimizer. The evaluation results show that the model achieved 96% accuracy, precision, recall, and F1-score on the Gallery Prawirotaman hotel dataset. These findings demonstrate that DistilBERT is capable of performing sentiment analysis on hotel reviews effectively and accurately.

Keyword: DistilBERT, Hotel Reviews, Sentiment Analysis, Indonesian Language, Hotel.