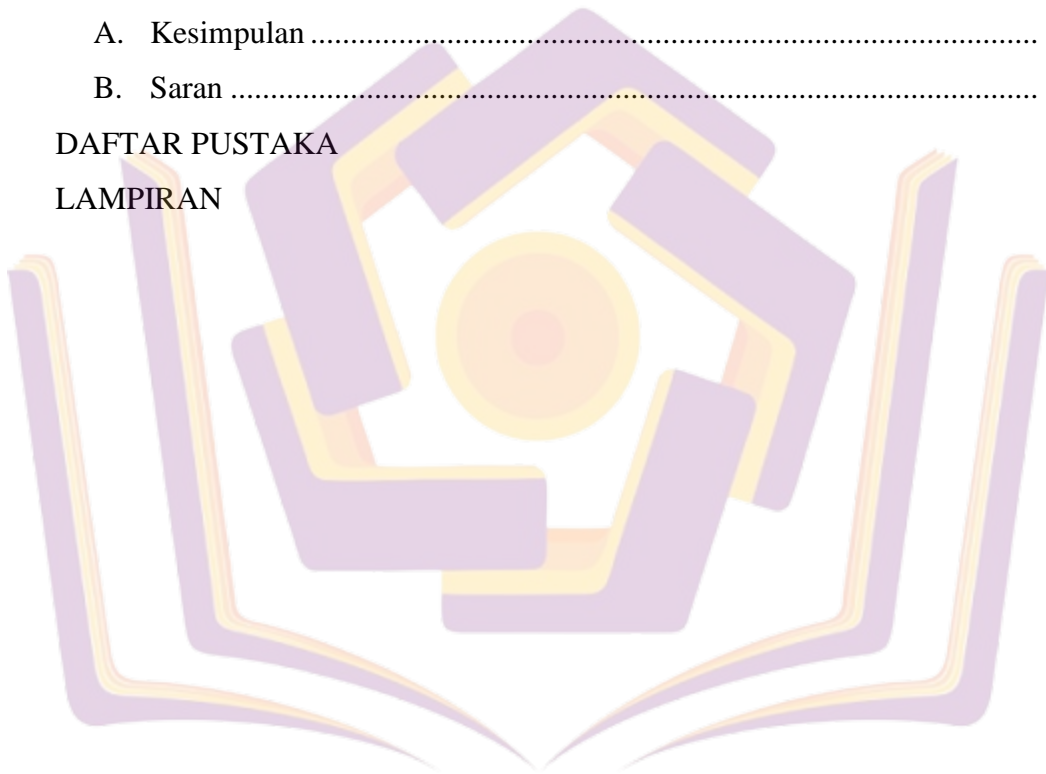


DAFTAR ISI

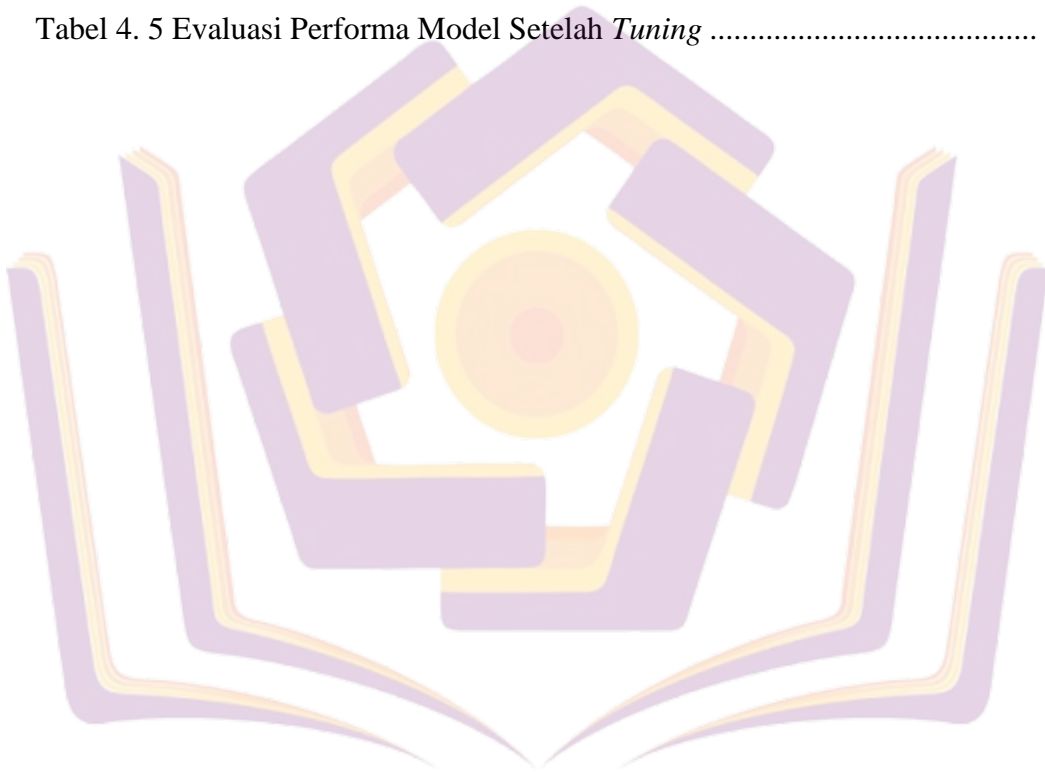
HALAMAN SAMPUL.....	i
HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN.....	v
HALAMAN PERSEMBAHAN	vi
HALAMAN MOTTO.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR ISTILAH.....	xv
DAFTAR LAMPIRAN	xvii
ABSTRAK.....	xviii
<i>ABSTRACT</i>	xix
BAB I PENDAHULUAN.....	1
A. Latar Belakang Masalah	1
B. Rumusan Masalah.....	4
C. Batasan Masalah	4
D. Tujuan Penelitian	6
E. Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA	8
A. Landasan Teori.....	8
B. Penelitian Sebelumnya.....	29
BAB III METODE PENELITIAN	37
A. Waktu Penelitian	37
B. Metode Pengumpulan Data.....	37
C. Alat dan Bahan Penelitian.....	40
D. Konsep Penelitian	40

BAB IV HASIL DAN PEMBAHASAN	55
A. Pengumpulan Data	55
B. Prapemrosesan Data (<i>Preprocessing</i>)	56
C. Distribusi Data	77
D. Pembobotan Kata dan Analisis Karakteristik Data.....	84
E. <i>Modeling</i> dan Analisis	86
F. <i>Deployment</i>	99
BAB V PENUTUP	105
A. Kesimpulan	105
B. Saran	106
DAFTAR PUSTAKA	
LAMPIRAN	



DAFTAR TABEL

Tabel 2. 1 Penelitian Sebelumnya.....	33
Tabel 3. 1 Contoh Sampel Data Mentah Komentar TikTok	43
Tabel 4. 1 Rincian Sumber Data	56
Tabel 4. 2 Hasil <i>Train</i> Sebelum <i>Tuning</i>	87
Tabel 4. 3 Hasil <i>Train</i> Setelah <i>Tuning</i>	88
Tabel 4. 4 Hasil Evaluasi <i>Matrix</i>	94
Tabel 4. 5 Evaluasi Performa Model Setelah <i>Tuning</i>	101



DAFTAR GAMBAR

Gambar 2. 1 Diagram Alur Pra-Pemrosesan Data	17
Gambar 3. 1 Diagram Alur Penelitian.....	41
Gambar 4. 1 Kode Program <i>Import Library</i>	57
Gambar 4. 2 Kode Program <i>Load Data</i>	61
Gambar 4. 3 Kode Program Menyamakan Kolom.....	62
Gambar 4. 4 Kode Program <i>Labeling</i>	64
Gambar 4. 5 Kode Program Menggabungkan <i>Dataset</i>	65
Gambar 4. 6 Kode Program Cek & Hapus Duplikasi Data.....	65
Gambar 4. 7 Kode Program <i>Cleaning Text</i>	67
Gambar 4. 8 Kode Program Tokenisasi	69
Gambar 4. 9 Kode Program <i>Stemming</i>	70
Gambar 4. 10 Kode Program <i>Stopword Removal</i>	73
Gambar 4. 11 Kode Program Cek Data Hasil <i>Preprocessing</i>	75
Gambar 4. 12 Kode Program <i>Data Split</i>	77
Gambar 4. 13 Kode Program <i>WordCloud</i>	78
Gambar 4. 14 <i>Output WordCloud</i>	80
Gambar 4. 15 Kode Program <i>Distribusi Data</i>	81
Gambar 4. 16 Proporsi Pembagian Data.....	83
Gambar 4. 17 Kode Program <i>Fitur Extraction (TF-IDF)</i>	84
Gambar 4. 18 Kode Program <i>Hyperparameter Tuning</i>	85
Gambar 4. 19 Kode Program <i>Train Model</i> Sebelum <i>Tuning</i>	86
Gambar 4. 20 Kode Program <i>Train Model</i> Setelah <i>Tuning</i>	88
Gambar 4. 21 Kode Program <i>Confusion Matrix</i>	89
Gambar 4. 22 <i>Confusion Matrix</i>	91
Gambar 4. 23 Kode Program Tahap Uji dan Evaluasi.....	96
Gambar 4. 24 Kode Program Tahap Uji dan Evaluasi.....	97
Gambar 4. 25 Kode Program Tahap Uji dan Evaluasi.....	98
Gambar 4. 26 Kode Program Pengujian Sistem	99

DAFTAR ISTILAH

Istilah	Definisi
<i>Accuracy</i>	Persentase prediksi model yang benar dari seluruh data.
<i>Algoritma</i>	Langkah sistematis untuk menyelesaikan suatu masalah.
<i>Alpha (Smoothing Parameter)</i>	Parameter pada <i>Naïve Bayes</i> untuk mencegah probabilitas nol.
<i>API</i>	Antarmuka yang memungkinkan aplikasi saling berkomunikasi.
<i>Case Folding</i>	Mengubah seluruh huruf teks menjadi huruf kecil.
<i>Cleansing</i>	Pembersihan data dari simbol atau karakter tidak penting.
<i>Confusion Matrix</i>	Tabel evaluasi perbandingan data aktual dan prediksi.
<i>Data Mining</i>	Proses menemukan pola atau informasi dari data besar.
<i>Dataset</i>	Kumpulan data yang digunakan dalam penelitian.
<i>Deployment</i>	Tahap penerapan model agar dapat digunakan pengguna.
<i>F1-Score</i>	Ukuran performa model dari precision dan recall.
<i>Filtering</i>	Penyaringan data untuk mengambil informasi relevan.
<i>GridSearchCV</i>	Metode pencarian parameter terbaik secara otomatis.
<i>Hate Speech</i>	Konten yang mengandung ujaran kebencian.
<i>Hyperparameter Tuning</i>	Proses optimasi parameter model.
<i>Labeling</i>	Pemberian label kategori pada data.

<i>Machine Learning</i>	Sistem yang belajar dari data untuk membuat prediksi.
<i>Multinomial Naïve Bayes</i>	Algoritma klasifikasi teks berbasis probabilitas kata.
NLP	Teknik pengolahan bahasa alami oleh komputer.
<i>Normalizing</i>	Mengubah kata tidak baku menjadi kata baku.
<i>Overfitting</i>	Model terlalu cocok pada data latih namun buruk di data baru.
<i>Precision</i>	Perbandingan prediksi positif benar terhadap semua prediksi positif.
<i>Preprocessing</i>	Tahap pembersihan dan persiapan data sebelum pemodelan.
<i>Prototipe</i>	Model awal untuk pengujian konsep sistem.
<i>Python</i>	Bahasa pemrograman populer untuk data dan machine learning.
<i>Recall</i>	Perbandingan prediksi positif benar terhadap seluruh data positif aktual.
<i>Root Word</i>	Kata dasar tanpa imbuhan.
<i>Scikit-Learn</i>	Library Python untuk machine learning.
<i>Scraping</i>	Pengambilan data spesifik dari halaman web.
<i>Slang Word</i>	Kata tidak baku yang umum digunakan di media sosial.
<i>Stemming</i>	Proses mengubah kata berimbuhan menjadi kata dasar.
<i>Stopword Removal</i>	Penghapusan kata umum yang tidak bermakna penting.
<i>Testing Data</i>	Data untuk menguji performa model.
<i>Underfitting</i>	Model gagal mempelajari pola data dengan baik.
<i>Vectorization</i>	Konversi teks menjadi bentuk numerik.
<i>Word Cloud</i>	Visualisasi frekuensi kata dalam bentuk grafis.

DAFTAR LAMPIRAN

- Lampiran 1. Kartu Bimbingan Dosen Pembimbing 1
- Lampiran 2. Kartu Bimbingan Dosen Pembimbing 2
- Lampiran 3. Hasil Cek Plagiarisme
- Lampiran 4. *Dataset* Hasil Scraping
- Lampiran 5. *Dataset* Publik
- Lampiran 6. Projek – *Google Colaboratory*
- Lampiran 7. Kode Program

