

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 TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR ISTILAH	xiv
DAFTAR LAMPIRAN.....	xvi
ABSTRAK.....	xvii
<i>ABSTRACT</i>	xviii
BAB I PENDAHULUAN.....	1
A. Latar Belakang Masalah	1
B. Rumusan Masalah.....	4
C. Batasan Masalah	4
D. Tujuan Penelitian	5
E. Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA.....	7
A. Landasan Teori.....	7
B. Penelitian Sebelumnya.....	19
BAB III METODE PENELITIAN.....	23
A. Tempat dan Waktu Penelitian.....	23
B. Metode Pengumpulan Data.....	23
C. Alat dan Bahan Penelitian.....	24
D. Konsep Penelitian	25
BAB IV HASIL DAN PEMBAHASAN	27

A. Pengumpulan Data	27
B. <i>Preprocessing</i> Data.....	31
C. Pelabelan Data Menggunakan <i>Lexicon Based</i>	38
D. <i>Feature Extraction</i> Menggunakan TF-IDF.....	40
E. Pembagian Data	43
F. Klasifikasi Sebelum <i>SMOTE</i>	44
G. Penyeimbangan Data Menggunakan <i>SMOTE</i>	56
H. Klasifikasi Model Setelah <i>SMOTE</i>	57
I. Perbandingan Model Sebelum dan Sesudah <i>SMOTE</i>	68
BAB V PENUTUP.....	70
A. Kesimpulan	70
B. Saran	71
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR TABEL

Tabel 4. 1. Jumlah Dataset Setelah Hapus Data Duplikat.....	29
Tabel 4. 2. Hasil Proses <i>Cleaning</i>	31
Tabel 4. 3. Hasil Proses Case Folding	32
Tabel 4. 4. Hasil Proses Normalisasi	33
Tabel 4. 5. Hasil Proses Tokenisasi	34
Tabel 4. 6. Hasil Proses <i>Stopword Removal</i>	35
Tabel 4. 7. Hasil Proses <i>Stemming</i>	36
Tabel 4. 8. Hasil Proses Pelabelan dan Perhitungan <i>Score</i>	38
Tabel 4. 9. Tabel Distribusi Sentimen.....	39
Tabel 4. 10. Hasil Ekstraksi TF-IDF	41
Tabel 4. 11. Hasil Klasifikasi <i>Naïve Bayes Before Smote</i>	45
Tabel 4. 12. <i>Confusion Matrix Naïve Bayes Before Smote</i>	47
Tabel 4. 13. Hasil Klasifikasi SVM <i>Before Smote</i>	48
Tabel 4. 14. <i>Confusion Matrix SVM Before Smote</i>	50
Tabel 4. 15. Hasil Klasifikasi <i>Random Forest Before Smote</i>	51
Tabel 4. 16. <i>Confusion Matrix Random Forest Before Smote</i>	53

DAFTAR GAMBAR

Gambar 2. 1 Logo Aplikasi Dolan Teluk Penyu.....	8
Gambar 3.1. Proses <i>Scrapping</i> Data di Google Collab.....	23
Gambar 3.2. Alur Penelitian.....	26
Gambar 4.1. Proses <i>Crawling</i> dan <i>Scraping</i> Dataset	28
Gambar 4.2. Visualisasi Data (<i>Wordcloud</i> Sebelum <i>Preprocessing</i>)	29
Gambar 4.3. Visualisasi Data Frekuensi 10 Kata Teratas.....	30
Gambar 4.4. <i>Wordcloud</i> Setelah <i>Preprocessing</i>	37
Gambar 4.5. Frekuensi Kata Setelah <i>Preprocessing</i>	37
Gambar 4.6. Diagram Distribusi <i>Sentiment</i> Sebelum Proses <i>Smote</i>	40
Gambar 4.7. Hasil Proses Pembagian Data <i>Training</i> dan Data <i>Testing</i>	43
Gambar 4. 8. Perbandingan <i>Accuracy</i> Klasifikasi <i>Before Smote</i>	54
Gambar 4.9. Distribusi Data Setelah <i>Smote</i>	56
Gambar 4.10. Perbandingan Akurasi Model <i>After Smote</i>	66
Gambar 4.11. Perbandingan <i>Accuracy</i> Model Sebelum dan Sesudah <i>Smote</i>	69

DAFTAR ISTILAH

1. *Accuracy* : Tingkat ketepatan model dalam mengklasifikasikan data secara benar.
2. *Artificial Intelligence (AI)* : Cabang ilmu komputer yang mempelajari pembuatan sistem cerdas yang mampu meniru kecerdasan manusia.
3. *Case Folding* : Tahap *preprocessing* yang mengubah seluruh huruf menjadi huruf kecil (*lowercase*).
4. *Classification*: Proses pengelompokan data ke dalam kelas tertentu berdasarkan karakteristik yang dimiliki.
5. *Cleaning* : Proses pembersihan data teks dari *URL*, simbol, angka, emoji, dan karakter yang tidak diperlukan.
6. *Confusion Matrix* : Tabel evaluasi yang digunakan untuk mengukur performa model klasifikasi.
7. *Crawling* : Teknik pengambilan data secara otomatis dari internet atau media sosial.
8. *Dataset* : Kumpulan data yang digunakan dalam penelitian atau pelatihan model machine learning.
9. *E-Government* : Pemanfaatan teknologi informasi oleh pemerintah untuk meningkatkan kualitas pelayanan publik.
10. *F1-Score* : Nilai harmonik antara *precision* dan *recall* yang digunakan untuk mengevaluasi model klasifikasi.
11. *Feature Extraction* : Proses mengubah data teks menjadi fitur numerik yang dapat diproses oleh algoritma *machine learning*.
12. *IDF (Inverse Document Frequency)* : Bobot yang menunjukkan tingkat pentingnya suatu kata dalam kumpulan dokumen.
13. *Lexicon Based* : Metode analisis sentimen yang menggunakan kamus kata positif dan negatif untuk menentukan polaritas sentimen.
14. *Machine Learning* : Cabang kecerdasan buatan yang memungkinkan komputer belajar dari data tanpa diprogram secara eksplisit.

15. *Multinomial Naive Bayes* : Algoritma klasifikasi probabilistik yang umum digunakan untuk analisis teks dan dokumen.
16. *Natural Language Processing (NLP)* : Bidang kecerdasan buatan yang mempelajari interaksi antara komputer dan bahasa manusia.
17. *Naive Bayes* : Algoritma klasifikasi berbasis *Teorema Bayes* dengan asumsi independensi antar fitur.
18. *Precision* : Rasio data positif yang diprediksi benar terhadap seluruh data yang diprediksi positif.
19. *Preprocessing* : Tahapan awal pengolahan data teks sebelum dilakukan analisis atau klasifikasi.
20. *Recall* : Kemampuan model menemukan seluruh data positif yang sebenarnya.
21. *Sentiment Analysis*: Teknik untuk mengidentifikasi opini atau sentimen dalam suatu teks.
22. *SMOTE* : *Synthetic Minority Over-sampling Technique*, metode untuk menyeimbangkan jumlah data antar kelas.
23. *Stemming* : Proses mengubah kata berimbuhan menjadi kata dasar.
24. *Stopword Removal* : Proses menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis teks.
25. *TF (Term Frequency)* : Frekuensi kemunculan suatu kata dalam dokumen.
26. *TF-IDF* : Metode pembobotan kata berdasarkan frekuensi kata dalam dokumen dan keseluruhan korpus.
27. *Tokenizing*: Proses memecah kalimat menjadi kata-kata atau token.
28. *WordCloud* : Visualisasi data teks dalam bentuk kumpulan kata berdasarkan frekuensi kemunculannya.

DAFTAR LAMPIRAN

Lampiran 1. Kartu Bimbingan Dosen Pembimbing 1

