

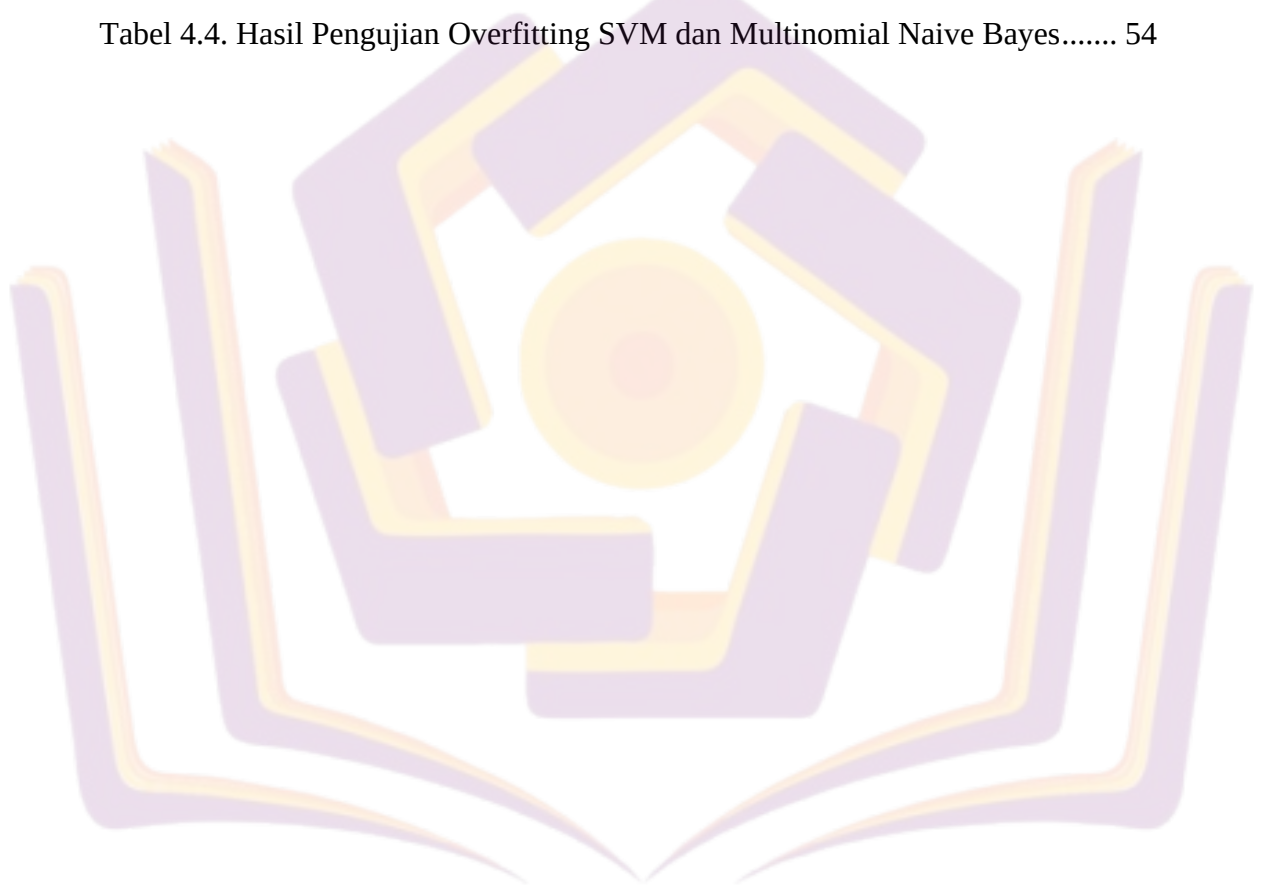
DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PERNYATAAN KEASLIAN	v
HALAMAN PERSEMBAHAN	vi
HALAMAN MOTTO	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR ISTILAH	xiv
INTISARI.....	xvi
ABSTRACT.....	xvii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	6
1.4. Tujuan Penelitian.....	7
1.5. Manfaat Penelitian.....	7
BAB II TINJAUAN PUSTAKA.....	8
2.1. Tinjauan Pustaka	8
2.2. Keaslian Penelitian	11
2.3. Landasan Teori	13
2.4. <i>Accuracy Paradox</i>	20
2.5. <i>Data leakage</i>	20
2.6. <i>Overfitting</i>	21
2.7. Kerangka berpikir	22
BAB III METODE PENELITIAN.....	24

3.1. Jenis, Sifat dan Pendekatan Penelitian	24
3.2. Metode Pengumpulan Data	25
3.3. Metode Analisis Data	29
3.4. Alur Penelitian.....	32
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	38
4.1. Hasil dan Pengumpulan data	38
4.2. Proses Text Preprocessing	39
4.3. Hasil Pelabelan Kelas Sentimen (<i>Labeling</i>)	41
4.4. Ekstraksi Fitur (TF-IDF) dan Implementasi SMOTE	47
4.5. Hasil Pengujian dan Evaluasi Kinerja Algoritma.....	48
4.6. Pengujian Overfitting Model	54
4.7. Pembahasan dan Penegasan Posisi Penelitian	55
4.8. <i>Scientific Findings</i>	56
BAB V PENUTUP.....	58
5.1. Kesimpulan.....	58
5.2. Kontribusi Ilmiah.....	59
5.3. Keterbatasan Penelitian	60
5.4. Saran	60
DAFTAR PUSTAKA	
LAMPIRAN	

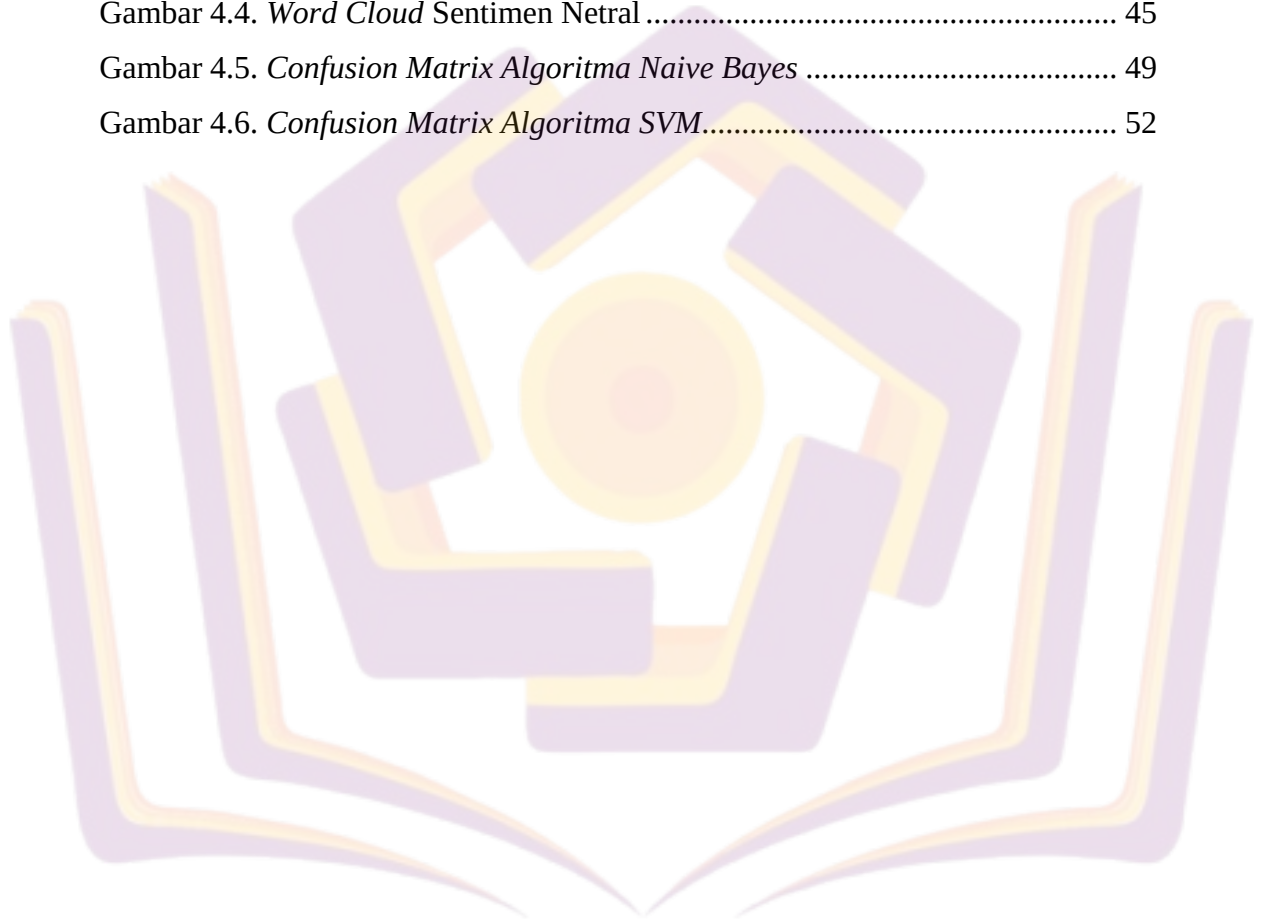
DAFTAR TABEL

Tabel 2.1. Matriks literatur review dan implikasi bagi peneliti	11
Tabel 2.2. Format evaluasi <i>Confusion Matrix</i>	19
Tabel 3.1. Tahap seleksi dan kontrol dataset	27
Tabel 4.1. Hasil Transformasi Data pada Tahap <i>Preprocessing</i>	40
Tabel 4.2. Hasil Evaluasi Kinerja <i>Naive Bayes</i>	50
Tabel 4.3. Hasil Evaluasi Kinerja <i>Support Vector Machine (SVM)</i>	52
Tabel 4.4. Hasil Pengujian Overfitting SVM dan Multinomial Naive Bayes.....	54



DAFTAR GAMBAR

Gambar 2.1. Ilustrasi pemisahan kelas menggunakan hyperplane pada SVM	17
Gambar 3.1 Alur Penelitian.....	33
Gambar 4.1. Distribusi Kelas Sentimen Hasil Pelabelan (Sebelum SMOTE).....	42
Gambar 4.2. <i>Word Cloud</i> Sentimen Positif.....	44
Gambar 4.3. <i>Word Cloud</i> Sentimen Negatif	44
Gambar 4.4. <i>Word Cloud</i> Sentimen Netral	45
Gambar 4.5. <i>Confusion Matrix</i> Algoritma Naive Bayes	49
Gambar 4.6. <i>Confusion Matrix</i> Algoritma SVM.....	52



DAFTAR ISTILAH

1. *Accuracy* : Tingkat ketepatan model secara keseluruhan dalam memprediksi kelas yang benar.
2. *Accuracy Paradox* : Kondisi anomali di mana tingkat akurasi terlihat sangat tinggi, namun sebenarnya gagal mengenali kelas minoritas akibat ketimpangan data.
3. *Case Folding* : Proses penyeragaman seluruh huruf di dalam teks menjadi huruf kecil.
4. *Cleansing* : Pembersihan data teks dari karakter yang tidak relevan (seperti simbol, URL, angka).
5. *Confusion Matrix* : Tabel matriks evaluasi kinerja algoritma yang membandingkan hasil tebakan mesin dengan label data aktual.
6. *Dataset* : Kumpulan atau himpunan data yang digunakan sebagai objek penelitian.
7. *F1-Score* : Metrik evaluasi yang berisi rata-rata harmonik dari nilai Presisi dan *Recall*.
8. *Hyperplane* : Garis batas pemisah antar kelas dalam ruang multi-dimensi pada algoritma SVM.
9. *Imbalanced Data* : Kondisi di mana distribusi atau jumlah sampel antar kelas dalam sebuah himpunan data sangat tidak seimbang.
10. *Lexicon-based* : Pendekatan klasifikasi sentimen menggunakan rujukan kamus (leksikon) yang berisi nilai bobot kata.
11. *Overfitting* : Kondisi di mana model ML terlalu menghafal data latih namun gagal beradaptasi pada data uji yang baru.
12. *Precision* : Rasio tingkat ketepatan prediksi positif yang diberikan oleh model.
13. *Recall* : Sensitivitas model dalam mengenali dan menemukan seluruh data pada suatu kelas (terutama kelas minoritas).
14. *SMOTE* : (*Synthetic Minority Over-sampling Technique*), algoritma untuk menyeimbangkan kelas data minoritas secara sintetis matematis.
15. *Stemming* : Proses pengubahan kata berimbuhan kembali ke bentuk kata dasarnya.

16. *Stopword* : Kata tugas atau kata hubung dalam bahasa Indonesia yang sering muncul namun tidak memiliki nilai/makna sentimen.
17. *SVM* : *Support Vector Machine*, algoritma machine learning berbasis pencarian batas margin maksimal antar kelas vektor.
18. *TF-IDF* : *Term Frequency - Inverse Document Frequency*, yaitu metode statistik yang digunakan untuk menghitung bobot nilai penting suatu kata dalam sekumpulan dokumen.
19. *Word Cloud* : Representasi visual dari frekuensi teks yang menonjolkan kata kunci dominan di dalam sebuah korpus data

