

DAFTAR ISI

ABSTRAK	ii
<i>ABSTRACT</i>	iii
LEMBAR PENGESAHAN.....	iv
LEMBAR PERNYATAAN ORISINALITAS.....	v
Kata Pengantar	vi
Daftar Isi.....	viii
Daftar Gambar	xi
Daftar Tabel.....	xii
Daftar Istilah.....	xiii
Bab I PENDAHULUAN	1
I.1 Latar Belakang.....	1
I.2 Perumusan Masalah.....	4
I.3 Tujuan Tugas Akhir.....	5
I.4 Batasan Tugas Akhir	5
I.5 Manfaat Tugas Akhir.....	5
I.6 Sistematika Penulisan.....	6
Bab II TINJAUAN PUSTAKA.....	8
II.1 <i>E-commerce</i>	8
II.2 Bukalapak	9
II.3 <i>Text Mining</i>	10
II.3.1 <i>Preprocessing</i>	10
II.3.2 Mengatasi Missing Value.....	12
II.4 Pembobotan TF-IDF.....	13
II.5 Metode <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	15
II.6 Analisis Klasifikasi.....	15

II.7 K-Nearest Neighbors	16
II.8 Metode Evaluasi	18
II.8.1 Confusion Matrix	18
II.8.2 K-Fold Cross Validation	20
II.9 <i>Word clouds</i>	20
II.10 <i>State of the Art</i>	21
Bab III SISTEMATIKA PENYELESAIAN MASALAH.....	27
III.1 Kerangka Berpikir	27
III.2 Sistematika Penyelesaian Masalah	28
III.3 Pengumpulan Data.....	29
III.4 Pengolahan Data	29
III.5 Metode Evaluasi	30
III.6 Alasan Pemilihan Metode.....	31
Bab IV Analisis dan Perancangan.....	32
IV.1 Pengumpulan Data.....	32
IV.2 <i>Preprocessing Data</i>	32
IV.2.1 Pembersihan Data.....	33
IV.2.2 <i>Case Folding</i>	33
IV.2.3 <i>Slang Word</i>	34
IV.2.4 <i>Tokenization</i>	34
IV.2.5 Stopwords Removal	35
IV.2.6 <i>Stemming</i>	35
IV.2.7 Mengatasi Nilai Yang Hilang.....	35
IV.3 Pembobotan TF-IDF.....	36
IV.4 Pembagian Data (<i>Splitting Data</i>).....	37
IV.5 Klasifikasi.....	40
IV.6 Penerapan Performasi	41
Bab V Analisis Dan Evaluasi Hasil.....	42

V.1	Klasifikasi dengan K-Nearest Neighbors	42
V.2	Evaluasi Performasi.....	45
V.2.1	Hasil KNN pada <i>Testing</i>	45
V.2.2	Confusion Matrix	46
V.2.3	K-fold Cross Validation	47
V.3	Visualisasi.....	48
Bab VI	KESIMPULAN DAN SARAN	50
VI.1	Kesimpulan.....	50
VI.2	Saran	50
Daftar Pustaka	51
Lampiran	55
1.	Source Code Menentukan K.....	55
2.	Perhitungan Confusion matrix	55
3.	Struktur data	57