

DAFTAR ISI

ABSTRAK	i
<i>ABSTRACT</i>	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN ORISINALITAS	iv
KATA PENGANTAR	v
Daftar Isi.....	vi
Daftar Gambar.....	ix
Daftar Tabel	x
Daftar Lampiran	xi
Daftar Simbol	xii
Daftar Istilah.....	xiii
Bab I Pendahuluan	1
I.1 Latar Belakang	1
I.2 Perumusan Masalah.....	3
I.3 Tujuan Penelitian.....	3
I.4 Batasan Penelitian	3
I.5 Manfaat Penelitian.....	4
I.6 Sistematika Penulisan.....	4
Bab II Tinjauan Pustaka	6
II.1 Taksonomi Bloom Revisi (<i>Revised Bloom's Taxonomy</i>).....	6
II.2 Klasifikasi Pertanyaan	8
II.3 Penambangan Teks (<i>Text Mining</i>).....	8
II.4 <i>Text Preprocessing</i>	8
II.5 <i>K-Nearest Neighbour</i> (KNN)	10
II.6 Evaluasi Performansi.....	12

II.7	<i>Cross-Validation</i>	14
II.8	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	14
II.9	<i>Synthetic Minority-Over Sampling Techniques (SMOTE)</i>	15
II.9.1	Penelitian Terdahulu	15
II.9.2	Alasan Pemilihan Teori, Kerangka Kerja, atau Mekanisme	17
Bab III	Metodologi Penelitian.....	18
III.1.1	Pengembangan Model Konseptual.....	18
III.1.2	Sistematika Penyelesaian Masalah.....	19
III.1.3	Pengumpulan Data	20
III.1.4	Pengolahan Data.....	21
III.1.5	Metode Evaluasi.....	22
III.2	Alasan Pemilihan Metode.....	22
Bab IV	ANALISIS DAN PERANCANGAN	24
IV.1	Analisis Ruang Lingkup Studi Kasus.....	24
IV.2	Pengumpulan Data	24
IV.2.1	Penggabungan <i>Dataset</i>	25
IV.2.2	<i>Labeling</i>	26
IV.3	<i>Data Preprocessing</i>	26
IV.3.1	<i>Case Folding</i>	27
IV.3.2	<i>Tokenizing</i>	27
IV.3.3	<i>Filtering</i>	28
IV.3.4	<i>Stemming</i>	28
IV.4	Pembobotan TF-IDF.....	29
Bab V	IMPLEMENTASI DAN PENGUJIAN.....	30
V.1	Implementasi Metode <i>Oversampling SMOTE</i>	30
V.2	Persebaran <i>Data Train</i> dan <i>Data Test</i>	31

V.3	Implementasi <i>K-Fold Cross Validation</i> dan Algoritma KNN.....	31
V.4	Pengukuran Kinerja Performansi Model.....	32
Bab VI	KESIMPULAN DAN SARAN	35
VI.1	Kesimpulan.....	35
VI.2	Pengembangan Selanjutnya.....	35
	Daftar Pustaka	37
	LAMPIRAN A – Dataset Soal dan Label	43
	LAMPIRAN B – Persebaran Data Train dan Data Test	71
	LAMPIRAN C – Persebaran Data Train dan Data Test	75