

DAFTAR ISI

HALAMAN PENGESAHAN PEMBIMBING	iii
HALAMAN PENGESAHAN PENGUJI.....	iv
SURAT PERNYATAAN KARYA ASLI TUGAS AKHIR	v
PERNYATAAN BEBAS PLAGIASI	vi
ABSTRAK	vii
ABSTRACT	viii
KATA PENGANTAR	ix
DAFTAR ISI.....	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xv
DAFTAR MODUL PROGRAM	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Metodologi Penelitian dan Pengembangan Sistem.....	3
1.7 Sistematika Penulisan	4
BAB II TINJAUAN LITERATUR.....	5
2.1 <i>Email spam</i>	5
2.2 <i>Scraping</i>	5
2.3 <i>Text Preprocessing</i>	6
2.3.1 <i>Cleaning</i>	6
2.3.2 <i>Case Folding</i>	6
2.3.3 <i>Tokenization</i>	6
2.3.4 <i>Stopword Removal</i>	7
2.3.5 <i>Lemmatization</i>	7
2.3.6 <i>Stemming</i>	8
2.4 <i>Word Embedding</i>	8
2.5 <i>Word2Vec</i>	9
2.5.1 <i>Continuous Bags of Words (CBOW)</i>	10
2.5.2 <i>Skip-Gram</i>	12
2.6 <i>Term frequency-Inverce Document Frequency (TF-IDF)</i>	14
2.7 <i>Machine Learning</i>	16
2.8 <i>Support Vector Machine (SVM)</i>	16

2.9	<i>Grid Search</i>	21
2.10	<i>Hyperparameter Tuning</i>	21
2.11	<i>K-Fold Cross Validation</i>	22
2.12	Pengujian	22
2.12.1	<i>Confusion Matrix</i>	22
2.13	Penelitian Terdahulu	23
BAB III METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM.....		27
3.1	Metodologi Penelitian	27
3.1.1	Studi Literatur	28
3.1.2	Pengumpulan Data	28
3.1.3	Pelabelan Data	29
3.1.4	Translasi Teks	30
3.1.5	<i>Preprocessing</i>	30
3.1.6	Representasi Teks <i>Word Embedding</i>	35
3.1.7	Representasi Teks TF-IDF	43
3.1.8	<i>Splitting Data</i>	46
3.1.9	Modeling	46
3.1.10	Pengujian Model	55
3.2	Metodologi Pengembangan Sistem	56
3.2.1	<i>Requirements analysis and definition</i>	56
3.2.2	<i>System and software design</i>	57
3.2.3	<i>Implementation and unit testing</i>	61
3.2.4	<i>Integration and system testing</i>	61
BAB IV HASIL PENGUJIAN DAN PEMBAHASAN		62
4.1	Hasil Penelitian	62
4.2	Implementasi Pengumpulan Data	62
4.3	Implementasi Translasi Teks	62
4.4	Implementasi <i>Preprocessing</i> Data	63
4.4.1	<i>Cleaning</i>	63
4.4.2	<i>Case Folding</i>	63
4.4.3	<i>Tokenizing</i>	64
4.4.4	<i>Lemmatization</i>	64
4.4.5	<i>Stemming</i>	65
4.4.6	<i>Stopword Removal</i>	65

4.4.7 Implementasi <i>Splitting</i> Data	66
4.5 Implementasi Representasi Teks Word2vec.....	66
4.6 Implementasi Representasi Teks TF-IDF	67
4.7 Implementasi Modelling	68
4.8 Implementasi Pengembangan Sistem	71
4.9 Pengujian	73
4.9.1 Pengujian Model.....	74
4.9.2 Pengujian Model Menggunakan Data Primer	84
4.9.3 Pengujian Pengembangan Sistem.....	87
4.10 Pembahasan	87
4.10.1 Analisis <i>Confusion Matriks Baseline</i>	88
4.10.2 Analisis <i>Confusion Matriks</i> Setelah Optimasi <i>Hyperparameter</i>	88
4.10.3 Analisis Evaluasi Matrik <i>Baseline</i>	89
4.10.4 Analisis Evaluasi Matrik Setelah Optimasi <i>Hyperparameter</i>	90
4.10.5 Pembahasan Pengujian Model Menggunakan Data Primer	91
BAB V KESIMPULAN DAN SARAN	93
5.1 Kesimpulan	93
5.2 Saran	93
LAMPIRAN	97