

DAFTAR ISI

HALAMAN PENGESAHAN PEMBIMBING.....	ii
HALAMAN PENGESAHAN PENGUJI.....	iii
HALAMAN PERNYATAAN KARYA ASLI	iv
HALAMAN PENGESAHAN BEBAS PLAGIASI.....	v
ABSTRAK.....	vi
ABSTRACT.....	vii
PRAKATA.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	2
1.6.1 Metodologi penelitian.....	3
1.6.2 Pengembangan sistem.....	3
1.7 Sistematika Penulisan.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 SMS Spam.....	4
2.2 Preprocessing Data.....	4
2.2.1 Case folding.....	4
2.2.2 Remove URL.....	4
2.2.3 Remove emoji.....	5
2.2.4 Remove emoticon.....	5
2.2.5 Remove number.....	5
2.2.6 Remove Punctuation.....	5
2.2.7 Remove Single Character.....	5
2.2.8 Stemming.....	6
2.2.9 Remove stopwords.....	7
2.2.10 Tokenizing.....	7
2.2.11 Algoritma Preprocessing Data.....	7
2.3 Word2Vec.....	8
2.4 Fasttext.....	9

2.5 Support Vector Machine.....	12
2.6 Pengujian dan Validasi.....	15
2.6.1 K-Fold Cross Validation.....	15
2.6.2 Confusion Matrix.....	15
2.6.3 Evaluation Metrics.....	15
2.7 Tinjauan Literatur.....	16
2.8 Research Gap.....	18
BAB III METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM.....	19
3.1 Metodologi Penelitian.....	19
3.1.1 Analisa masalah.....	20
3.1.2 Pengumpulan data.....	20
3.1.3 Preprocessing data.....	20
3.1.4 Ekstraksi fitur menggunakan fastText.....	37
3.1.4.1 fastText.....	37
3.1.4.2 Encoding data.....	45
3.1.5 Pembuatan model SVM.....	46
3.1.6 Pengujian model.....	51
3.1.7 Hasil.....	52
3.1.8 Laporan.....	53
3.2 Pengembangan Sistem.....	53
3.2.1 Communication.....	53
3.2.2 Quick plan & modeling quick design.....	54
3.2.2.1 Alur sistem.....	54
3.2.2.2 Arsitektur Sistem.....	55
3.2.2.3 Desain antarmuka sistem.....	56
3.2.3 Construction of prototype.....	57
3.2.4 Deployment delivery & feedback.....	57
BAB IV HASIL, PENGUJIAN DAN PEMBAHASAN.....	58
4.1 Implementasi.....	58
4.1.1 Implementasi Preprocessing Data.....	58
4.1.1.1 Modul case folding.....	58
4.1.1.2 Modul remove URL.....	58
4.1.1.3 Modul remove emoji.....	59
4.1.1.4 Modul remove emoticon.....	59
4.1.1.5 Modul remove number.....	60
4.1.1.6 Modul remove punctuation.....	60
4.1.1.7 Modul remove single character.....	60
4.1.1.8 Modul stemming.....	61
4.1.1.9 Modul remove stopwords.....	61

4.1.1.10 Modul tokenizing.....	62
4.1.3 Implementasi SVM.....	63
4.1.3.1 Training Model SVM.....	63
4.1.4.1 Testing Model SVM.....	63
4.1.4 Implementasi Antarmuka.....	64
4.2 Hasil.....	66
4.2.1 fastText.....	66
4.2.2 Model SVM.....	66
4.3 Pengujian.....	68
4.4 Pembahasan.....	69
BAB V PENUTUP.....	70
5.1 Kesimpulan.....	70
5.2 Saran.....	70
DAFTAR PUSTAKA.....	71

DAFTAR TABEL

Tabel 2.1. Contoh Case Folding.....	4
Tabel 2.2. Contoh Remove URL.....	4
Tabel 2.3. Contoh Remove Emoji.....	5
Tabel 2.4. Contoh Remove Emoticon.....	5
Tabel 2.5. Contoh Remove Number.....	5
Tabel 2.6. Contoh Remove Punctuation.....	5
Tabel 2.7. Contoh remove single character.....	5
Tabel 2.8. Contoh Stemming.....	6
Tabel 2.9. Contoh Remove Stopwords.....	7
Tabel 2.10. Contoh Tokenizing.....	7
Tabel 2.11. One Hot Encoding.....	9
Tabel 2.12. Window Size fastText.....	10
Tabel 2.13. Training Data.....	10
Tabel 2.14. One Hot Encoding Data fastText.....	10
Tabel 2.15. Training Data setelah One Hot Encoding.....	11
Tabel 2.16. Confusion Matrix.....	15
Tabel 2.17. Evaluation Metrics.....	15
Tabel 2.18. State of The Art.....	17
Tabel 2.18. State of The Art (lanjutan).....	18
Tabel 3.1. Hasil character n-gram.....	37
Tabel 3.2. Training data sebelum One Hot Encoding.....	38
Tabel 3.3. One hot encoding.....	38
Tabel 3.4. Training Data setelah one hot encoding.....	38
Tabel 3.4. Training Data setelah one hot encoding (lanjutan).....	39
Tabel 3.5. Weight 1.....	39
Tabel 3.6. Training Data 1.....	40
Tabel 3.7. Hidden Layer.....	40
Tabel 3.8. Weight 2.....	40
Tabel 3.9. Output Layer.....	41
Tabel 3.10. Softmax.....	41
Tabel 3.11. Error.....	42
Tabel 3.12. Delta W2.....	42
Tabel 3.13. Hasil dot product W2 dengan error.....	43
Tabel 3.14. Delta W1.....	43
Tabel 3.15. Weight 1 Updated.....	44
Tabel 3.16. Weight 2 Updated.....	44
Tabel 3.17. Vektor character n-gram.....	44
Tabel 3.17. Vektor character n-gram (lanjutan).....	45
Tabel 3.18. Vektor seluruh kata.....	45

Tabel 3.19. Hasil encoding.....	46
Tabel 3.20. Contoh Data.....	47
Tabel 3.21. Inisialisasi Variabel.....	47
Tabel 3.22. Matriks Kernel RBF.....	47
Tabel 3.23. Matriks Hessian.....	48
Tabel 3.24. Hasil perhitungan E_i	48
Tabel 3.25. Hasil perhitungan a_i	48
Tabel 3.26. Perhitungan a baru.....	49
Tabel 3.27. Matriks Kernel RBF Data Uji Pertama.....	50
Tabel 3.28. Matriks Kernel RBF Data Uji Kedua.....	51
Tabel 3.29. Rancangan Confusion Matrix.....	52
Tabel 3.30. Rancangan Evaluation Metrics.....	52
Tabel 3.31. Kebutuhan perangkat lunak.....	53
Tabel 3.32. Spesifikasi komputer cloud.....	54
Tabel 3.33. Spesifikasi komputer lokal.....	54
Tabel 3.34. Pengujian Sistem.....	57
Tabel 4.1. Performa Model.....	68
Tabel 4.1. Performa Model (lanjutan).....	69

DAFTAR GAMBAR

Gambar 2.1. Arsitektur Skip-gram (Chia, 2018).....	8
Gambar 2.2. Visualisasi window size (Chia, 2018).....	8
Gambar 3.1. Metodologi Penelitian.....	19
Gambar 3.2. Alur preprocessing data.....	21
Gambar 3.3. Alur case folding.....	22
Gambar 3.4. Alur remove URL.....	23
Gambar 3.5. Alur remove emoji.....	25
Gambar 3.6. Alur remove emoticon.....	26
Gambar 3.6. Alur remove emoticon (lanjutan).....	27
Gambar 3.7. Alur remove number.....	28
Gambar 3.8. Alur remove punctuation.....	29
Gambar 3.8. Alur remove punctuation (lanjutan).....	30
Gambar 3.9. Alur remove single character.....	31
Gambar 3.10. Alur stemming.....	33
Gambar 3.11. Alur remove stopword.....	34
Gambar 3.11. Alur remove stopword (lanjutan).....	35
Gambar 3.12. Alur tokenizing.....	36
Gambar 3.13. Window Size.....	37
Gambar 3.14. Proses encoding.....	46
Gambar 3.15. Visualisasi K-Fold Cross Validation.....	51
Gambar 3.16. Alur Sistem.....	55
Gambar 3.17. Arsitektur Sistem.....	56
Gambar 3.18. Desain Antarmuka Sistem.....	56
Gambar 3.19. Desain Antarmuka Jika Terdeteksi SMS Spam.....	56
Gambar 3.20. Desain antarmuka jika terdeteksi SMS ham/ non spam.....	57
Gambar 4.1. Halaman awal.....	65
Gambar 4.2 Halaman web dengan hasil spam.....	65
Gambar 4.3 Halaman web dengan hasil ham.....	66
Gambar 4.4 Grafik Akurasi Model SVM dengan Data Train.....	67
Gambar 4.5 Grafik Akurasi SVM dengan Data Test.....	67

DAFTAR LAMPIRAN

LAMPIRAN A. Sample Dataset.....	78
LAMPIRAN B. Hasil fastText.....	79
LAMPIRAN C. Hasil Akurasi Model Dengan Data Train dan Data Test.....	84
LAMPIRAN D. Hasil Confusion Matrix.....	85