

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	i
HALAMAN PENGESAHAN	i
HALAMAN PERNYATAAN	ii
HALAMAN PERSEMBAHAN	iii
HALAMAN MOTTO	iv
PRAKATA	v
DAFTAR ISI	vii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xiv
INTISARI	xix
ABSTRACT	xx
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Perumusan Masalah	7
1.3 Batasan Masalah	7
1.4 Tujuan Penelitian	8
1.5 Manfaat Penelitian	8
1.6 Kontribusi Penelitian	9
II TINJAUAN PUSTAKA	10
2.1 Deteksi Konten Spam pada Data Teks dan Media Sosial	10
2.2 Penelitian <i>Sentence Pair Classification</i>	22

III DASAR TEORI	30
3.1 <i>Text Mining</i>	30
3.1.1 <i>Pre-processing</i>	32
3.1.2 <i>Feature generation and selection</i>	34
3.1.3 <i>Vector embedding</i>	36
3.1.4 <i>Classification</i>	39
3.1.5 <i>Evaluation</i>	40
3.2 Konten Spam pada Media Sosial	41
3.3 Model <i>Deep Learning</i>	44
3.3.1 Convolutional Neural Network (CNN)	45
3.3.2 Recurrent Neural Network (RNN)	47
3.3.3 Long Short Term Memory (LSTM)	48
3.3.4 <i>Attention mechanism</i>	51
3.4 <i>Sentence Pair Classification</i>	54
3.5 <i>Browser Extension</i>	61
IV METODOLOGI PENELITIAN	63
4.1 Deskripsi Penelitian	63
4.2 Kerangka Penelitian	67
4.3 Pengumpulan Data dan Pelabelan	68
4.4 <i>Pre-Processing</i> dan <i>Feature Generation</i>	73
4.5 Pengembangan Model	78
4.5.1 Pengembangan model <i>deep learning</i>	80
4.5.2 Pengembangan model <i>machine learning</i>	89
4.6 Evaluasi dan Rancangan Pengujian	91
4.7 Pengembangan Prototipe <i>Browser Extension</i>	93
4.8 Alat Penelitian	96
V DATASET SPAMID-PAIR	97
5.1 Pengumpulan Data	97
5.2 Pelabelan Data	98
5.3 Pembersihan dan Pemrosesan Data	100
5.4 Profil Dataset	100
5.5 Validasi Dataset dan Pengujian Sebagai <i>Benchmark</i> Awal	104

VI ENSEMBLE MACHINE LEARNING PADA DATASET SPAMID-PAIR 108

6.1	Langkah Pengembangan Model	108
6.2	Tahap <i>Pre-processing</i> dan Normalisasi	108
6.3	Skenario yang Digunakan	110
6.4	Spesifikasi Perangkat	113
6.5	<i>Features Generation</i>	114
6.6	Implementasi Model <i>Machine Learning</i>	117
6.6.1	Model berbasis Naive Bayes	119
6.6.2	Model berbasis Support Vector Machine	120
6.6.3	Model berbasis <i>nearest neighbor</i>	121
6.6.4	Model berbasis <i>tree</i>	121
6.6.5	Model berbasis <i>boosting</i>	122
6.6.6	Model berbasis regresi	123
6.6.7	Model berbasis <i>multilayer perceptron</i>	123
6.6.8	Model <i>ensemble machine learning</i>	124
6.7	Evaluasi Model <i>Machine Learning</i>	124
6.7.1	Hasil evaluasi dataset komentar saja dan pasangan posting-komentar tanpa emoji	124
6.7.2	Hasil evaluasi dataset komentar dan pasangan posting-komentar dengan emoji	127
6.7.3	Analisis peningkatan performa pada pasangan posting-komentar (dengan dan tanpa emoji)	131
6.7.4	Perbandingan peningkatan performa pada data pasangan posting-komentar dengan komentar saja (dengan emoji)	134
6.7.5	Perbandingan performa antara emoji teks dan emoji simbol pada data komentar saja dan pasangan posting-komentar	136
6.7.6	Hasil evaluasi pada dataset dengan fitur tambahan, skenario seimbang, posting-komentar <i>pair stacked</i> dan <i>concatenation</i> , serta metode <i>ensemble</i>	137
6.7.7	Analisis dan kesimpulan awal penggunaan metode <i>machine learning</i>	139

VII ARSITEKTUR MODEL LSTM-CNN DENGAN PENDEKATAN DATA PASANGAN POSTING-KOMENTAR UNTUK DE-

TEKSI KOMENTAR SPAM DENGAN FITUR EMOJI	142
7.1 <i>Pre-Processing</i>	143
7.2 <i>Features Generation and Selection</i>	143
7.3 Implementasi <i>Embedding Layer</i>	147
7.4 Implementasi Model <i>Deep Learning</i>	151
7.4.1 Model <i>deep learning</i> dasar komentar saja	152
7.4.2 Model <i>deep learning</i> dasar posting-komentar	165
VIMODEL EMOJI-AWARE INTER ATTENTION PAIR BiLSTM – CNN (EiAP-BC) UNTUK DETEKSI KOMENTAR BERBASIS KONTEKS DENGAN FITUR EMOJI	181
8.1 Model <i>Deep Learning</i> EiAP-BC	181
8.1.1 Model <i>deep learning</i> pembandingan untuk posting-komentar	199
8.1.2 Penggunaan dan pengujian model EiAP-BC untuk studi kasus lain	206
8.1.3 Pengujian lapisan arsitektur dengan studi ablasi pada model EiAP-BC	207
8.1.4 <i>Hyperparameters tuning</i> pada model EiAP-BC	217
8.2 Perbandingan Performa Model <i>Machine Learning</i> dan <i>Deep Learning</i> Terbaik	220
8.3 Prototipe Web Services Berbasis Protokol REST	222
8.4 Implementasi Model <i>Machine Learning</i> dan <i>Deep Learning</i> pada <i>Browser Extension</i> InstaSpamEx menggunakan Web Service REST	230
IX KESIMPULAN DAN SARAN	236
9.1 Kesimpulan	236
9.2 Saran	237
DAFTAR PUSTAKA	239
LAMPIRAN	265
A LUARAN PUBLIKASI	265
B MODEL	267