

DAFTAR ISI

HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTTO	vi
PRAKATA	vii
DAFTAR ISI	ix
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
ABSTRAKSI	xv
ABSTRACT	xvi
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Metodologi Penelitian	5
1.6.1 Studi Literatur	5
1.6.2 Pengumpulan Data	5
1.6.3 Perancangan Sistem	5
1.6.4 Implementasi Sistem	6
1.6.5 Pengujian	6
1.6.6 Penulisan Laporan	6
1.7 Sistematika Penulisan	6
II TINJAUAN PUSTAKA	8

III LANDASAN TEORI	15
3.1 Information Retrieval	15
3.2 Text Mining	15
3.3 Sentiment Analysis	16
3.4 Preprocessing	17
3.4.1 Tokenization	17
3.4.2 Stopwords Removal	17
3.4.3 Stemming	18
3.5 Term Frequency-Inverse Document Frequency	18
3.6 Chi Square	19
3.7 F-test Analysis of Variance	20
3.8 Mutual Information	21
3.9 Support Vector Machine	22
3.10 Decision Tree	23
3.11 Random Forest	23
3.12 Extra Tree Classifier	23
3.13 Gaussian Naive Bayes	24
3.14 Multinomial Naive Bayes	25
3.15 K-Nearest Neighbour	25
3.16 Logistic Regression	26
3.17 Multilayer Perceptron	27
3.18 Gradient Bosting	29
3.19 Adaptive Boosting	29
3.20 Pengujian	31
3.20.1 K-Fold Cross Validation	31
3.21 Perhitungan Performa	32
IV ANALISIS DAN PERANCANGAN	35
4.1 Analisis Sistem	35
4.2 Tahapan Penelitian	35
4.2.1 Studi Literatur	36
4.2.2 Pengumpulan Data	36
4.2.3 Perancangan Sistem	37
4.2.4 Implementasi Sistem	40
4.2.5 Klasifikasi dan Validasi	40

4.2.6	Penulisan Laporan	41
4.3	Rancangan Pengumpulan Data	41
4.3.1	Line-Today	41
4.3.2	Twitter	42
4.4	Rancangan Preprocessing	43
4.4.1	Case Folding	43
4.4.2	URL, Hashtag, Mention Removal	44
4.4.3	Punctuation Removal	44
4.4.4	Whitespace Removal	44
4.4.5	Stopwords Removal	45
4.4.6	Stemming	45
4.5	Pelabelan Data	45
4.6	Ekstraksi Fitur	46
4.6.1	Count Vectorizer	46
4.6.2	Term Frequency Inverse-Document Frequency	46
4.7	Seleksi Fitur	47
4.8	Klasifikasi	48
4.9	Validasi	50
V	IMPLEMENTASI	51
5.1	Spesifikasi Sistem	51
5.2	Implementasi Pengumpulan Data	52
5.2.1	Line-Today	52
5.2.2	Twitter	52
5.3	Preprocessing	53
5.3.1	Case Folding	53
5.3.2	Url, Hashtag, Mention Removal	54
5.3.3	Punctuation Removal	54
5.3.4	Whitespace Removal	55
5.3.5	Stopwords Removal	55
5.3.6	Stemming	57
5.4	Pelabelan Data	57
5.5	Ekstraksi Fitur	58
5.5.1	Count Vectorizer	58
5.5.2	Term Frequency Inverse Document Frequency	58

5.6	Seleksi Fitur	59
5.6.1	Chi Square	59
5.6.2	Analysis of Variance	60
5.6.3	Mutual Information	60
5.7	Klasifikasi	61
5.8	Validasi	63
VI HASIL DAN PEMBAHASAN		66
6.1	Proses Pengumpulan Data	66
6.1.1	Line-Today	66
6.1.2	Twitter	66
6.2	Preprocessing	67
6.3	Pelabelan Data	68
6.4	Perbandingan Metode Ekstraksi Fitur, Seleksi Fitur, dan Pengklasifikasi	70
VII KESIMPULAN DAN SARAN		75
7.1	Kesimpulan	75
7.2	Saran	76
DAFTAR PUSTAKA		77