

DAFTAR ISI

HALAMAN PENGESAHAN TESIS.....	iii
HALAMAN PERNYATAAN.....	iv
PRAKATA	v
DAFTAR ISI	vii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xiii
INTISARI	xv
ABSTRACT	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	4
1.4 Tujuan	4
1.5 Manfaat	4
BAB II TINJAUAN PUSTAKA	5
BAB III LANDASAN TEORI	9
3.1 Analisis Sentimen	9
3.2 Imbalance Data	10
3.3 Text Preprocessing.....	11
3.4 Data Augmentation	12
3.4.1 Word augmentation.....	12
3.4.2 Character augmentation	16
3.4.3 Back Translation	16

3.5	Term Weighting.....	17
3.5.1	Term Frequency (TF).....	18
3.5.2	Inverse Document Frequency (IDF)	18
3.5.3	Inverse Class Frequency (ICF)	19
3.5.4	Inverse Class Space Density Frequency (ICSDF)	20
3.5.5	Relevance Frequency (RF)	21
3.6	Random Forest.....	22
3.7	Support Vector Machine (SVM).....	23
3.7.1	Kernel dan Klasifikasi Non-Linier	26
3.8	Evaluasi Performa.....	27
BAB IV METODOLOGY PENELITIAN		31
4.1	Studi Literatur	31
4.2	Alat	31
4.3	Tahapan Penelitian.....	31
4.4	Data Collection	33
4.4.1	SmSA IndoNLU	33
4.4.2	Indonesian General Sentiment Analysis	34
4.4.3	Dataset COVID-19	35
4.4.4	Dataset PPKM	35
4.5	Preprocessing	36
4.5.1	Case Folding	37
4.5.2	Filtering.....	37
4.5.3	Tokenisasi	37
4.5.4	Stopword.....	38
4.6	Easy Data Augmentation (EDA)	38
4.6.1	Synonym Replacement (SR).....	40
4.6.2	Random Insertion (RI).....	40
4.6.3	Random Swap (RS)	47

4.6.4	Random Deletion (Deletion).....	49
4.7	Term Weighting.....	52
4.7.1	Term Frequency (TF).....	52
4.7.2	Inverse Document Frequency (IDF).....	52
4.7.3	Inverse Class Frequency (ICF).....	53
4.7.4	Inverse Class Space Density Frequency (ICSDF).....	54
4.7.5	Relevancy Frequency (RF).....	55
4.8	Development Model Classification.....	56
4.8.1	Random Forest.....	56
4.8.2	Support Vector Machine.....	59
4.9	Model Evaluation.....	61
BAB V IMPLEMENTASI		63
5.1	Spesifikasi Perangkat Keras.....	63
5.2	Spesifikasi Perangkat Lunak.....	63
5.3	Implementasi Pengumpulan Data	64
5.4	Implementasi Preprocessing	64
5.4.1	Case Folding	64
5.4.2	Special Character Removal.....	64
5.4.3	Number Removal.....	65
5.4.4	Multiple Character Removal.....	66
5.4.5	Remove Emoticon	66
5.4.6	Memeriksa Typo	67
5.4.7	Penanganan Simbol Khusus	67
5.4.8	Stopword Removal	68
5.4.9	Stemming.....	68
5.4.10	Label Encoding.....	69
5.5	Implementasi Easy Data Augmentation	69
5.5.1	Synonym Replacement (SR).....	69

5.5.2	Random Insertion (RI)	71
5.5.3	Random Swap (RS)	72
5.5.4	Random Deletion (RD)	72
5.6	Implementasi Skema Term Weighting	74
5.6.1	TF-IDF	74
5.6.2	TF-RF	75
5.6.3	TF-IDF-ICF	76
5.6.4	TF-IDF-ICSDF	76
5.7	Support Vector Machine (SVM)	77
5.8	Random Forest	78
BAB VI HASIL DAN PEMBAHASAN		79
6.1	Hasil Augmentasi Easy Data Augmentation (EDA)	79
6.1.1	Analisa Augmentasi Synonym Replacement	79
6.1.2	Analisa Augmentasi Random Insertion	82
6.1.3	Analisa Augmentasi Random Swap	85
6.1.4	Analisa Augmentasi Random Deletion	88
6.2	Analisa Kinerja Skema Term Weighting	90
6.2.1	Analisa Kinerja TF-IDF	91
6.2.2	Analisa Kinerja TF-IDF-ICF	94
6.2.3	Analisa Kinerja TF-IDF-ICSDF	97
6.2.4	Analisa Kinerja TF-RF	100
BAB VII KESIMPULAN DAN SARAN		103
7.1	Kesimpulan	103
7.2	Saran	104
DAFTAR PUSTAKA		105