

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENSAHAN	iv
HALAMAN PERNYATAAN KEASLIAN	v
HALAMAN PERSEMBAHAN.....	vi
HALAMAN MOTTO	ix
KATA PENGANTAR	x
DAFTAR ISI.....	xiii
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xv
DAFTAR PERSAMAAN	xvi
DAFTAR LAMPIRAN.....	xvii
DAFTAR ISTILAH	xviii
INTISARI.....	xix
ABSTRACT	xx
BAB I PENDAHULUAN	
A. Latar Belakang Masalah	1
B. Rumusan Masalah.....	6
C. Batasan Masalah	6
D. Tujuan Penelitian	7
E. Manfaat Penelitian	7
BAB II TINJAUAN PUSTAKA	
A. Landasan Teori.....	8
B. Penelitian Sebelumnya.....	27
BAB III METODE PENELITIAN	
A. Tempat dan Waktu Penelitian.....	35
B. Metode Pengumpulan Data.....	34\5

C. Alat dan Bahan Penelitian.....	35
D. Konsep Penelitian	36
BAB IV HASIL DAN PEMBAHASAN	
A. Pengumpulan Data	43
B. Tahap <i>Pre-processing</i>	44
C. <i>Balancing</i>	49
D. Pembobotan TF-IDF	53
E. Pembagian Data	54
F. Algoritma <i>Support Vector Machine (SVM)</i>	55
G. Evaluasi.....	56
BAB V PENUTUP	
A. Kesimpulan	74
B. Saran	75
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR TABEL

Tabel 2. 1 <i>Confusion matrix 2x2</i>	22
Tabel 2. 2 Penelitian sebelumnya.....	33
Tabel 2. 2 Penelitian sebelumnya (Lanjutan).....	34
Tabel 4. 1 Contoh data yang diperoleh dari pengumpulan data.....	43
Tabel 4. 2 Contoh hasil proses <i>delete punctuation</i>	44
Tabel 4. 3 Contoh baris data yang dihapus	45
Tabel 4. 4 Contoh hasil <i>replacing</i>	46
Tabel 4. 5 Contoh hasil <i>case folding</i>	46
Tabel 4. 6 Contoh hasil <i>labeling</i>	47
Tabel 4. 7 Contoh hasil <i>stopwords</i> dan <i>stemming</i>	48
Tabel 4. 8 Contoh hasil <i>tokenizing</i>	49
Tabel 4. 9 Kondisi data setelah <i>balancing</i> menggunakan metode <i>Near Miss</i>	51
Tabel 4. 10 Kondisi data setelah <i>balancing</i> metode ADASYN.....	51
Tabel 4. 11 Kondisi data setelah <i>balancing</i> metode SMOTE.....	52
Tabel 4. 12 Kondisi data setelah <i>balancing</i> metode RCS	53
Tabel 4. 13 Hasil pembobotan TF-IDF	54
Tabel 4. 14 Hasil akurasi training dari setiap metode dan komposisi.....	55
Tabel 4. 15 Nilai akurasi dari setiap metode dan komposisi.....	68
Tabel 4. 16 Hasil <i>F1-Score</i> dari setiap metode dan komposisi.....	70

DAFTAR GAMBAR

Gambar 1. 1 Diagram pengguna sosial media yang paling sering digunakan	3
Gambar 2. 1 Proses kerja dari teknik klasifikasi.....	9
Gambar 2. 2 Kondisi dataset asli	16
Gambar 2. 3 Kondisi <i>imbalanced dataset</i>	16
Gambar 2. 4 Kondisi data yang telah dibalancing dengan SMOTE	17
Gambar 2. 5 Kondisi data yang telah dibalancing dengan ADASYN	18
Gambar 2. 6 Grafik dari ROC-AUC	25
Gambar 4. 1 Visualisasi data sebelum <i>balancing</i>	37
Gambar 4. 2 Visualisasi dari nilai akurasi training setiap metode dan komposisi	55
Gambar 4. 3 <i>Confusion matrix</i> metode <i>Near Miss</i> (70:30).....	57
Gambar 4. 4 <i>Confusion matrix</i> metode ADASYN (90:10).....	57
Gambar 4. 5 Nilai TP pada <i>confusion matrix</i> 4x4 (<i>Near Miss</i> 70:30)	58
Gambar 4. 6 Nilai FP pada <i>confusion matrix</i> 4x4 (<i>Near Miss</i> 70:30).....	58
Gambar 4. 7 Nilai FN pada <i>confusion matrix</i> 4x4 (<i>Near Miss</i> 70:30).....	59
Gambar 4. 8 Nilai TP pada <i>confusion matrix</i> 4x4 (ADASYN 90:10)	63
Gambar 4. 9 Nilai FP pada <i>confusion matrix</i> 4x4 (ADASYN 90:10)	63
Gambar 4. 10 Nilai FN pada <i>confusion matrix</i> 4x4 (ADASYN 90:10).....	64
Gambar 4. 11 Visualisasi dari nilai akurasi setiap metode dan komposisi	69
Gambar 4. 12 Kurva ROC- metode <i>Near Miss</i> (70:30)	71
Gambar 4. 13 Kurva ROC metode ADASYN (90:10)	72

DAFTAR PERSAMAAN

Persamaan 2. 1 Rumus menghitung <i>Imbalance Ratio</i> (IR).....	13
Persamaan 2. 2 Rumus menghitung IDF	14
Persamaan 2. 3 Rumus menghitung TF-IDF.....	15
Persamaan 2. 4 Rumus SVM kernel Linear	20
Persamaan 2. 5 Rumus SVM kernel RBF.....	21
Persamaan 2. 6 Rumus menghitung <i>Error Rate</i>	23
Persamaan 2. 7 Rumus menghitung <i>Accuracy</i>	23
Persamaan 2. 8 Rumus menghitung <i>Recall</i> atau <i>Sensitivity</i>	23
Persamaan 2. 9 Rumus menghitung <i>Precision</i>	24
Persamaan 2. 10 Rumus menghitung <i>F1-Score</i>	24

DAFTAR LAMPIRAN

Lampiran 1. Kartu Bimbingan

Lampiran 2. Sampel Data Komentar

Lampiran 3. Hasil *Confusion Matrix* dan Kurva ROC

Lampiran 4. *Source Code*

DAFTAR ISTILAH

SVM	: <i>Support Vector Machine</i>
RBF	: <i>Radial Basis Function</i>
TF-IDF	: <i>Term Frequency — Inverse Document Frequency</i>
RCS	: <i>Random Combination Sampling</i>
SMOTE	: <i>Synthetic Minority Over-sampling Technique</i>
ADASYN	: <i>Adaptive Synthetic Sampling Approach</i>
ROC-AUC	: <i>Receiver Operating Characteristic - Area Under the Curve</i>