



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	iii
PERNYATAAN BEBAS PLAGIASI	iv
HALAMAN PRAKATA	vi
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xiii
INTISARI	xvi
ABSTRACT	xvii
BAB 1 PENDAHULUAN	1
1. 1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA	4
BAB III DASAR TEORI	10
3.1 Sarkasme.....	10
3.2 Instagram.....	12
3.3 Natural Language Processing (NLP).....	12
3.4 Analisis Sentimen.....	13
3.5 Peningkatan Akurasi Analisis Sentimen dengan Deteksi Sarkasme.....	14
3.6 <i>Preprocessing</i>	14
3.6.1 Pembersihan Data.....	15
3.6.2 <i>Case Folding</i>	15
3.6.3 <i>Tokenization</i>	15
3.6.4 Penghapusan <i>Stopwords</i>	16
3.6.5 Konversi emotikon, angka, <i>slang</i> dan bahasa Inggris.....	16
3.6.6 <i>Stemming</i>	16
3.7 <i>Feature Extraction</i>	17
3.7.1 <i>Feature Extraction</i> pada Analisis Sentimen.....	17



3.7.2 <i>Feature Extraction</i> pada Deteksi Sarkasme.....	18
3.8 Klasifikasi.....	19
3.8.1 <i>Naïve Bayes</i>	19
3.8.2 <i>Decision Trees</i>	22
3.8.3 <i>Random Forest</i>	24
3.9 <i>K-Fold Cross Validation</i>	25
3.10 Evaluasi Model.....	26
BAB IV METODOLOGI.....	27
4.1 Gambaran Umum Penelitian.....	27
4.2 Alat dan Bahan	28
4.3 Tahapan Penelitian.....	29
4.4 Rancangan Sistem.....	31
4.4.1 Data Preprocessing	31
4.4.2 Pembobotan TF-IDF.....	33
4.4.3 <i>Feature Extraction</i>	36
4.4.4 <i>Naïve Bayes Classifier</i>	37
4.4.5 <i>Random Forest Classifier</i>	39
BAB V IMPLEMENTASI.....	39
5.1 Pengumpulan Data dan Implementasi Pelabelan Data	40
5.1.1 Profil Volunteer.....	42
5.1.2 Aturan Pelabelan Data.....	42
5.2 Implementasi <i>Preprocessing</i> Data.....	44
5.2.1 Pengubahan Emoji ke String.....	45
5.2.2 Pembersihan Data.....	46
5.2.3 Tokenisasi.....	46
5.2.4 Pengubahan Kata <i>Slang</i> ke Kata Baku	47
5.2.5 <i>Stopword Removal</i>	47
5.2.6 Stemming.....	48
5.3 Implementasi <i>Features Extraction</i> Analisis Sentimen.....	49
5.4 Implementasi <i>Features Extraction</i> Klasifikasi Sarkasme.....	50
5.4.1 TF-IDF <i>Unigram</i>	50
5.4.2 <i>Sentiment-related Features</i>	50
5.4.3 <i>Punctuation-related Features</i>	53
5.4.4 <i>Lexical Features</i>	54



5.4.5 Normalisasi Fitur.....	55
5.5 Implementasi <i>K-fold Cross Validations</i>	55
5.6 Implementasi Model dan Evaluasi Model.....	56
5.6.1 Implementasi Model dan Evaluasi Model Analisis Sentimen <i>Random Forest</i>	58
5.6.2 Implementasi Model dan Evaluasi Model Analisis Sentimen <i>Multinomial Naive Bayes</i>	59
5.6.3 Implementasi Model dan Evaluasi Model Klasifikasi Sarkasme <i>Random Forest</i>	60
5.7 Implementasi Perubahan Label Sentimen	61
BAB VI HASIL DAN PEMBAHASAN	39
6.1 Hasil Pengumpulan dan Pelabelan Data	39
6.2 Hasil <i>Preprocessing</i> Data.....	40
6.2.1 Hasil Perubahan Emoji ke <i>String</i>	41
6.2.2 Hasil Pembersihan Data.....	41
6.2.3 Hasil Tokenisasi.....	42
6.2.4 Hasil Perubahan Kata <i>Slang</i> ke Kata Baku	42
6.2.5 Hasil <i>Stopword Removal</i>	44
6.2.6 Hasil <i>Stemming</i>	44
6.3 Hasil <i>Features Extraction</i>	45
6.3.1 Hasil TF-IDF.....	45
6.3.2 Hasil <i>Sentiment-related Features</i>	46
6.3.3 Hasil <i>Punctuation-related Features</i>	46
6.3.4 Hasil <i>Lexical Features</i>	47
6.3.5 Hasil Normalisasi Fitur	49
6.5 Hasil Evaluasi Model	49
6.5.1 Evaluasi Model Analisis Sentimen dengan <i>Random Forest</i>	51
6.5.2 Evaluasi Model Analisis Sentimen dengan <i>Multinomial Naive Bayes</i>	52
6.5.3 Evaluasi Model Klasifikasi Sarkasme dengan <i>Random Forest</i>	53
6.6 Hasil Perubahan Label Sentimen.....	55
BAB VII KESIMPULAN DAN SARAN.....	39
7.1 Kesimpulan.....	39
7.2 Saran.....	39
DAFTAR PUSTAKA.....	39
LAMPIRAN.....	39



DAFTAR TABEL

Tabel 2.1 Perbandingan penelitian sebelumnya.....	8
Tabel 3.1 Confusion Matrix	26
Tabel 4.1 Contoh data training	31
Tabel 4.2 Teks setelah dilakukan preprocessing data.....	33
Tabel 4.3 Frekuensi term pada masing-masing dokumen.....	33
Tabel 4.4 Term frequency.....	34
Tabel 4.5 Idf dari masing-masing term.....	35
Tabel 4.6 TF-IDF dari masing-masing term.....	35
Tabel 4.7 Feature extraction untuk model deteksi sarkasme.....	36
Tabel 4.8 Fitur dan label sarkasme pada data training	39
Tabel 6.1 Hasil pelabelan data.....	39
Tabel 6.2 Hasil pengubahan emoji ke string.....	41
Tabel 6.3 Hasil pembersihan data.....	42
Tabel 6.4 Hasil tokenisasi.....	43
Tabel 6.5 Hasil pengubahan kata slang ke kata baku	43
Tabel 6.6 Hasil penghapusan kata stopword.....	44
Tabel 6.7 Hasil stemming	45
Tabel 6.8 Hasil TF-IDF unigram	46
Tabel 6.9 Hasil fitur sentiment-related.....	47
Tabel 6.10 Hasil fitur punctuation-related.....	48
Tabel 6.11 Hasil fitur punctuation-related.....	49
Tabel 6.12 Hasil normalisasi fitur sentiment-related.....	50
Tabel 6.13 Hasil normalisasi fitur punctuation-related dan lexical	50
Tabel 6.14 Nilai akurasi, presisi, recall, f1-score model analisis sentimen Random Forest.....	51
Tabel 6.15 Confusion Matrix model analisis sentimen Random Forest	52



Tabel 6.16 Nilai akurasi, presisi, recall, f1-score model analisis sentimen Multinomial Naive Bayes	53
Tabel 6.17 Confusion Matrix model analisis sentimen Multinomial Naive Bayes	53
Tabel 6.18 Nilai presisi, recall, f1-score model klasifikasi sarkasme Random Forest.....	54
Tabel 6.19 Confusion Matrix model deteksi sarkasme Random Forest.....	55
Tabel 6.20 Perbandingan model analisis sentimen Random Forest dengan deteksi sarkasme dan tanpa deteksi sarkasme	55
Tabel 6.21 Perbandingan model analisis sentimen Multinomial Naive Bayes dengan deteksi sarkasme dan tanpa deteksi sarkasme	56



DAFTAR GAMBAR

Gambar 3.1 Decision Tree untuk pembelian komputer (Han dkk., 2011)	23
Gambar 3.2 Metode <i>ensemble</i> (Han dkk., 2011).....	25
Gambar 3.3 <i>K-Fold Cross Validation</i> (Anon, 2020).....	26
Gambar 4.1 <i>Flowchart</i> analisis sentimen dengan deteksi sarkasme modifikasi dari penelitian Yunitasari, dkk (2019)	28
Gambar 4.2 <i>Flowchart preprocessing</i> data	32
Gambar 4.3 <i>Flowchart features extraction</i>	37
Gambar 4.4 <i>Random Forest trees</i>	40
Gambar 5.1 Potongan kode pengambilan data.....	41
Gambar 5.2 Potongan kode proses preprocessing data.....	45
Gambar 5.3 Potongan kode pengubahan emoji ke string.....	45
Gambar 5.4 Potongan kode pembersihan data.....	46
Gambar 5.5 Potongan kode tokenisasi.....	47
Gambar 5.6 Potongan kode pengubahan kata gaul	48
Gambar 5.7 Potongan kode proses stopword removal	48
Gambar 5.8 Potongan kode proses stemming.....	48
Gambar 5.9 Potongan kode proses pembobotan TF-IDF	49
Gambar 5.10 Potongan kode perhitungan nilai sentimen dan kontras sentimen emoji.....	51
Gambar 5.11 Potongan kode perhitungan nilai sentimen dan kontras sentimen kata	52
Gambar 5.12 Potongan kode perhitungan jumlah kemunculan tanda baca.....	53
Gambar 5.13 Potongan kode perhitungan jumlah kemunculan huruf kapital dalam teks.....	54
Gambar 5.14 Potongan kode perhitungan jumlah pengulangan huruf dalam teks	54
Gambar 5.15 Potongan kode perhitungan jumlah pengulangan tertawa dalam teks	55



Gambar 5.16 Potongan kode normalisasi <i>features</i>	55
Gambar 5.17 Potongan kode k-fold cross validations	56
Gambar 5.18 Potongan kode inisiasi model dan parameter <i>Random Forest</i>	57
Gambar 5.19 Potongan kode implementasi model analisis sentimen Random Forest.....	58
Gambar 5.20 Potongan kode implementasi model analisis sentimen Multinomial Naive Bayes.....	59
Gambar 5.21 Potongan kode implementasi model klasifikasi sarkasme Random Forest.....	61
Gambar 5.22 Potongan kode implementasi perubahan label analisis sentimen .	62
Gambar 6.1 Hasil perbandingan jumlah masing-masing kelas	40



LAMPIRAN

Lampiran 1 Sampel dataset.....	39
--------------------------------	----