DAFTAR ISI

BAB I PENDAHULUAN 1
1.1 Latar Belakang Masalah 1
1.2 Rumusan Masalah 2
1.3 Batasan Masalah 3
1.4 Tujuan Penelitian 3
1.5 Manfaat Penelitian 3
1.6 Tahapan Penelitian 4
1.7 Sistematika Penulisan 5
BAB II TINJAUAN LITERATUR 7
2.1 Klasifikasi Sentimen 7
2.2 Text Mining 7
2.3 Representasi Teks 8
2.3.1 Term Frequency-Inverse Document Frequency (TF-IDF) 9
2.3.2 Count Vectorizer 10
2.4 Machine Learning 11
2.4.1 Supervised Learning 11
2.4.2 Semi-Supervised Learning 12
2.4.3 Pseudo-Labeling 12
2.5 Algoritma Klasifikasi 13
2.5.1 Naive Bayes 13
2.5.2 Random Forest 14
2.6 Evaluasi Model 16
2.6.1 Confusion Matrix 16
2.6.2 Accuracy 17
2.6.3 Precision 17
2.6.4 Recall 18
2.6.5 F1-Score 19
2.7 Penelitian Terdahulu 19
BAB III METODOLOGI PENELITIAN 23
3.1 Jenis dan Pendekatan Penelitian 23
3.2 Objek dan Sumber Data Penelitian 23
3.3 Tahapan Penelitian 24
3.3.1 Pengumpulan Data 26
3.3.2 Preprocessing Data 27
3.3.3 Pelabelan Data 28
3.3.4 Pembagian Data 28
3.3.5 Vektorisasi Teks 29
3.3.6 Base Model Training 32

3.3.7 Semi-Supervised Learning 34
3.3.8 Evaluasi Model 35
BAB IV HASIL PENGUJIAN DAN PEMBAHASAN 38
4.1 Implementasi 38
4.1.1 Lingkungan Implementasi 38
4.1.2 Implementasi Pengumpulan dan Persiapan Data 38
4.1.3 Implementasi Preprocessing Data 39
4.1.4 Implementasi Vektorisasi Teks 41
4.1.5 Implementasi Base Model Training 42
4.1.6 Implementasi Semi-Supervised Learning 44
4.1.7 Implementasi Evaluasi Model 46
4.2 Hasil 48
4.2.1 Hasil Pengumpulan dan Persiapan Data 48
4.2.2 Hasil Preprocessing Data 49
4.2.3 Hasil Vektorisasi Teks 50
4.2.4 Hasil Base Model Training 52
4.2.5 Hasil Semi-Supervised Learning 54
4.2.6 Hasil Evaluasi Movel 60
4.3 Pembahasan 65
BAB V PENUTUP 67
5.1 Kesimpulan 67
5.2 Saran 67
DAFTAR PUSTAKA 68

DAFTAR TABEL

Tabel 2.1 State of The Art 21
Tabel 3.1 Contoh Preprocessing Teks 27
Tabel 3.2 Contoh Teks Ulasan 30
Tabel 3.3 Cara Kerja Count Vectorizer 30
Tabel 3.4 Contoh Perhitungan DF dan IDF 31
Tabel 3.5 Contoh Perhitungan TF pada Dokumen D1 31
Tabel 3.6 Contoh PerhitunganTF-IDF pada Dokumen D1 32
Tabel 3.7 Contoh Ekstraksi Unigram dan Bigram 33
Tabel 3.8 Hyperparameter Naive Bayes dan Random Forest 34
Tabel 3.9 Contoh Prediksi Pohon Keputusan 35
Tabel 3.10 Contoh Perhitungan Pseudo Labeling 36
Tabel 3.11 Contoh Confusion Matrix 36
Tabel 4.1 Lingkungan Implementasi 38
Tabel 4.2 Jumlah Data Setiap Dataset 49
Tabel 4.3 Hasil Teks Preprocessing 49
Tabel 4.4 Jumlah Data Setelah Preprocessing 50
Tabel 4.5 Hasil Vektorisasi Count Vectorizer 51
Tabel 4.6 Hasil Vektorisasi TF-IDF 51
Tabel 4.7 Konfigurasi Hyperparameter Base Model 52
Tabel 4.8 Hasil Evaluasi Base Model IndoNLU 53
Tabel 4.9 Hasil Evaluasi Base Model Ecommerce 53
Tabel 4.10 Hasil Evaluasi Base Model BCA Mobile 54
Tabel 4.11 Konfigurasi Hyperparameter per Iterasi Dataset IndoNLU 54
Tabel 4.12 Konfigurasi Hyperparameter per Iterasi Dataset Ecommerce 55
Tabel 4.13 Konfigurasi Hyperparameter per Iterasi Dataset BCA Mobile 56
Tabel 4.14 Hasil Data Disepakati Tiap Iterasi Pseudo-Labeling 57
Tabel 4.15 Hasil Jumlah Rata-rata Data Tiap Threshold 58
Tabel 4.16 Perbandingan Evaluasi Akurasi Model Count Vectorizer 61
Tabel 4.17 Perbandingan Evaluasi Akurasi Model TF-IDF 62
Tabel 4.18 Perbandingan Evaluasi F1-Score Model Count Vectorizer 63
Tabel 4.19 Perbandingan Evaluasi F1-Score Model TF-IDF 64

DAFTAR GAMBAR

Gambar 2.1 Arsitektur Random Forest..... 15
Gambar 3.1 Tahapan Penelitian..... 25
Gambar 3.2 Tahapan Data Preprocessing..... 27
Gambar 3.3 Tahapan Semi-Supervised Learning..... 34