Daftar Isi

Lembar Pernyataan i
Lembar Pengesahan ii
Abstrak iii
Abstract iv
Lembar Persembahan v
Kata Pengantar vii
Daftar Isi viii
Daftar Gambar x
Daftar Tabel xi
Daftar Istilah xii
1. Pendahuluan 1
   1.1 Latar Belakang 1
   1.2 Perumusan Masalah 1
   1.3 Tujuan 2
   1.4 Metodologi Penyelesaian Masalah 2
2. Landasan Teori 3
   2.1 Text Mining 3
   2.2 Keyword Extraction 3
   2.3 Preprocessing Data 3
      2.3.1 Case Folding dan Cleaning 3
      2.3.2 Tokenization 4
      2.3.3 Stopword Removal 4
      2.3.4 POS Tagging 4
   2.4 Pembobotan Kata 5
   2.5 Machine Learning 6
      2.5.1 Support Vector Machine 6
         2.5.1.1 SVM pada Linearly Separable Data 6
         2.5.1.2 SVM pada Nonlinearly Separable Data 8
   2.6 Parameter Pengukuran 10
      2.6.1 Confusion Matrix 10
      2.6.2 Accuracy 10
3. Deskripsi Sistem 12
   3.1 Tahap Preprocessing 12
      3.1.1 Case Folding 13
      3.1.2 Tokenization 14
      3.1.3 Cleaning dan Stopword Removal 14
      3.1.4 POS Tagging 15
      3.1.5 Penandaan Frase 15
   3.2 Feature Extraction 16
      3.2.1 Fitur TFIDF 16
      3.2.2 Fitur Position 16
   3.3 Learning 17
   3.4 Extraction 18
   3.5 Analisis dan Kebutuhan Sistem 19
      3.5.1 Kebutuhan Fungsionalitas 19
      3.5.2 Spesifikasi Perangkat Keras 19

3.5.3 Spesifikasi Perangkat Lunak 20
3.6 Perancangan Sistem 20
3.6.1 Diagram Konteks 20
3.6.2 Diagram Aliran Data Level 1 20
3.6.3 Diagram Aliran Data Level 2 Proses 1 21
3.6.4 Diagram Aliran Data Level 2 Proses 2 21
3.6.5 Diagram Aliran Data Level 2 Proses 3 22
3.6.6 Diagram Aliran Data Level 2 Proses 4 22
3.6.7 Kamus Data 23
3.6.8 Spesifikasi Proses 24
4. Implementasi dan Analisis Hasil Pengujian 29
4.1 Implementasi 29
4.2 Pengujian Sistem 29
4.2.1 Karakteristik Dataset 29
4.2.2 Tujuan Pengujian 30
4.3 Skenario Pengujian 30
4.3.1 Mengetahui pengaruh komposisi jumlah data *training* dan *testing* 30
4.3.2 Mengetahui pengaruh parameter C, d, dan γ 30
4.3.3 Mengetahui pengaruh keragaman data yang digunakan 31
4.4 Hasil Pengujian 31
4.4.1 Hasil skenario pengaruh komposisi jumlah data training dan testing 31
4.4.2 Hasil skenario pengaruh parameter C, d, dan γ 32
4.4.3 Hasil skenario pengaruh keragaman data yang digunakan 33
4.5 Analisis Pengujian 34
4.5.1 Analisis skenario pengaruh komposisi jumlah data training dan testing 34
4.5.2 Analisis skenario pengaruh parameter C, d, dan γ 36
4.5.3 Analisis skenario pengaruh keragaman data yang digunakan 38
5. Kesimpulan dan Saran 40
5.1 Kesimpulan 40
5.2 Saran 40
Daftar Pustaka 41
Lampiran A: Tagset Used Untuk Bahasa Indonesia 43
Lampiran B: Stopword List 44
Lampiran C: Phrase List 45
Lampiran D: Masukan Untuk libSVM 47
Lampiran E: Model Hasil Training Pada libSVM 49