

DAFTAR ISI

Lembar Persetujuan	i
Lembar Pengesahan	ii
Lembar Orisinalitas.....	iii
Kata Pengantar	iv
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan	4
1.4 Batasan Masalah	4
1.5 Manfaat	5
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	
2.1 Emosi	6
2.2 Text Mining	7
2.2.1 Tahapan Text Mining.....	8
2.3 Tahap <i>Preprocessing</i>	9
2.3.1 <i>Tokenizing</i>	9
2.3.2 <i>Stemming</i>	9
2.3.2.1 Struktur Morfologi Kata Bahasa Indonesia	10
2.4 Proses Stemming Untuk Bahasa Indonesia	13
2.5 Pembobotan (Term Weighting)	15
2.5.1 Pembobotan tf-idf	16
2.6 Data Transformation	17
2.6.1 Distribusi Frekuensi	18
2.7 Decision Tree	19
2.8 Algoritma <i>C5.0</i>	22
2.9 Evaluasi.....	25
2.10 <i>Precision</i>	25
2.11 <i>Recall</i>	26
2.12 <i>Accuracy</i>	27
2.13 Penelitian Yang Berkaitan	27

BAB III METODELOGI DAN PERANCANGAN

3.1 Analisis Data	28
3.1.1 Pengambilan Data	29
3.2 Deskripsi Umum Sistem	30
3.2.1 Batasan Sistem	31
3.3 Rancangan Proses	32
3.3.1 Perancangan <i>Preprocessing</i>	32
3.3.1.1 Perancangan Proses <i>Tokenizing</i>	33
3.3.1.1 Perancangan Proses <i>Stopword</i>	34
3.3.2 Perancangan Proses <i>tf-idf</i>	36
3.3.3 Perancangan Data Transformasi	37
3.3.4 Perancangan Proses Latih	42
3.3.5 Perancangan Proses Testing	46
3.4 Perhitungan Manual	49
3.4.1 Sumber Data	49
3.4.2 Proses Pembobotan	49
3.4.3 Proses Data Transformation	51
3.5 Pembentukan Pohon Keputusan	53
3.5.1 Menghitung <i>entropy</i> atribut dan <i>information gain</i>	54
3.5.2 Perancangan Rule Tree	59
3.6 Perancangan Proses Klasifikasi	59
3.7 Perancangan Interface	60
3.8 Metode Pengujian	61
3.8.1 Perhitungan <i>Accuracy</i>	62

BAB IV IMPLEMENTASI

4.1 Implementasi Sistem	64
4.2 Spesifikasi Perangkat Keras Dan Perangkat Lunak	64
4.3 Implementasi Antarmuka Sistem	64
4.4 Implementasi Preprocessing	65
4.4.1 Stopword	66
4.4.2 Stemming	67
4.4.3 Frekuensi	68
4.4.4 Proses Pembobotan	69
4.4.5 Transformasi Data	71

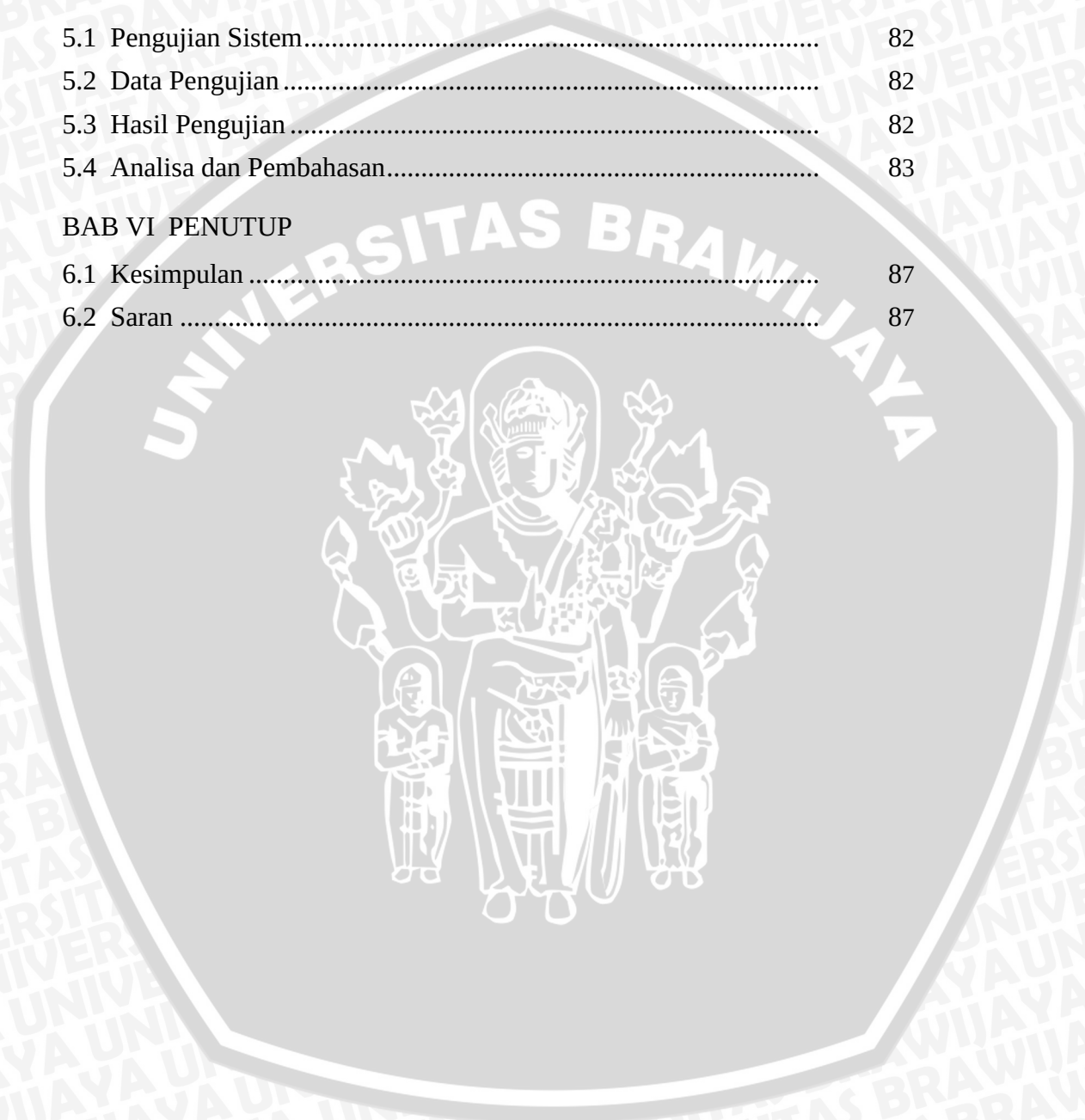
4.5 Implementasi Pembentukan Pohon.....	73
4.6 Implementasi Klasifikasi	77
4.7 Penerapan Aplikasi	79

BAB V PENGUJIAN DAN ANALISIS

5.1 Pengujian Sistem.....	82
5.2 Data Pengujian	82
5.3 Hasil Pengujian	82
5.4 Analisa dan Pembahasan.....	83

BAB VI PENUTUP

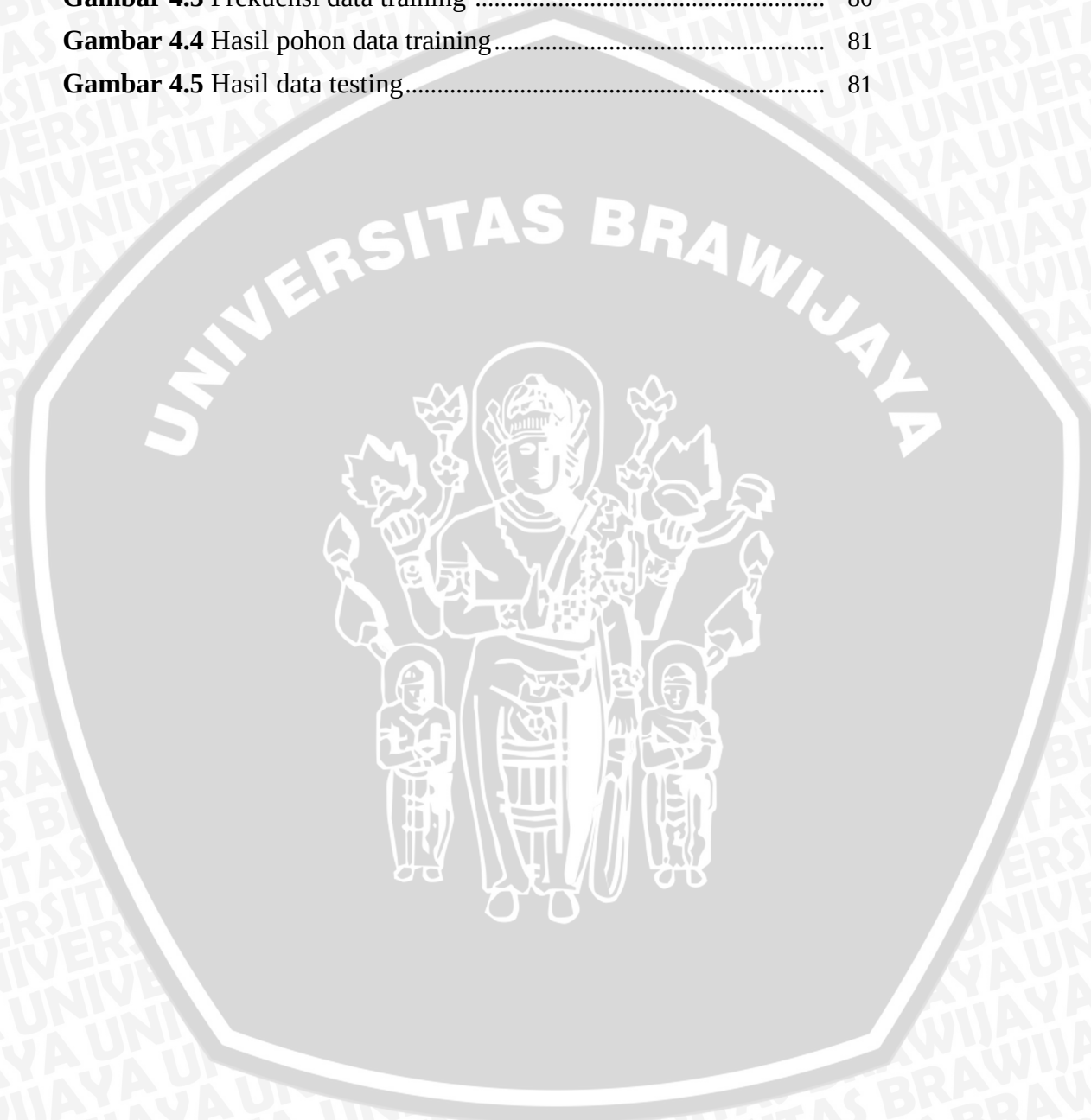
6.1 Kesimpulan	87
6.2 Saran	87



DAFTAR GAMBAR

Gambar 2.1 Tahap dalam teks mining [EVE-02].....	7
Gambar 2.2 Proses tokenizing text [WIB-06]	9
Gambar 2.5 Contoh Decision Tree [SAN-07].....	Error!
Bookmark not defined.0	
Gambar 2.4 Decision tree dalam proses partisi data[JIA-06].	Error!
Bookmark not defined.1	
Gambar 3.1 Alur penelitian.....	29
Gambar 3.2 Flowchart deskripsi umum sistem	Error!
Bookmark not defined.1	
Gambar 3.3 Skema rancangan proses.	Error!
Bookmark not defined.2	
Gambar 3.4 Flowchart perancangan preprocessing	Error!
Bookmark not defined.3	
Gambar 3.5 Flowchart proses tokenizing	Error!
Bookmark not defined.4	
Gambar 3.6 Flowchart proses stopword	Error!
Bookmark not defined.5	
Gambar 3.7 Flowchart proses pembobotan.....	Error!
Bookmark not defined.7	
Gambar 3.8 Flowchart proses data transformasi	38
Gambar 3.9 Flowchart proses distribusi frekuensi.....	39
Gambar 3.10 Flowchart proses transformasi bobot.....	Error!
Bookmark not defined.1	
Gambar 3.11 Flowchart proses latih C5.0	Error!
Bookmark not defined.2	
Gambar 3.12 Flowchart proses pencarian root	43
Gambar 3.13 Flowchart proses pembentukan node	45
Gambar 3.14 Proses klasifikasi.....	Error!
Bookmark not defined.6	
Gambar 3.15 Flowchart proses testing	48
Gambar 3.16 Tree hasil perhitungan root.....	55
Gambar 3.17 Tree hasil cabang Int1 untuk atribut suka.....	56
Gambar 3.18 Tree cabang Int1 untuk atribut senang.....	57
Gambar 3.19 Tree cabang Int2 untuk atribut senang.....	58
Gambar 3.20 Tree cabang Int3 untuk atribut senang.....	58
Gambar 3.21 Tree cabang Int3 untuk atribut suka.....	59

Gambar 3.22 Interface sistem.....	60
Gambar 4.1 Main form	Error!
Bookmark not defined.	
Gambar 4.2 Pengambilan data training	79
Gambar 4.3 Frekuensi data training	80
Gambar 4.4 Hasil pohon data training.....	81
Gambar 4.5 Hasil data testing.....	81



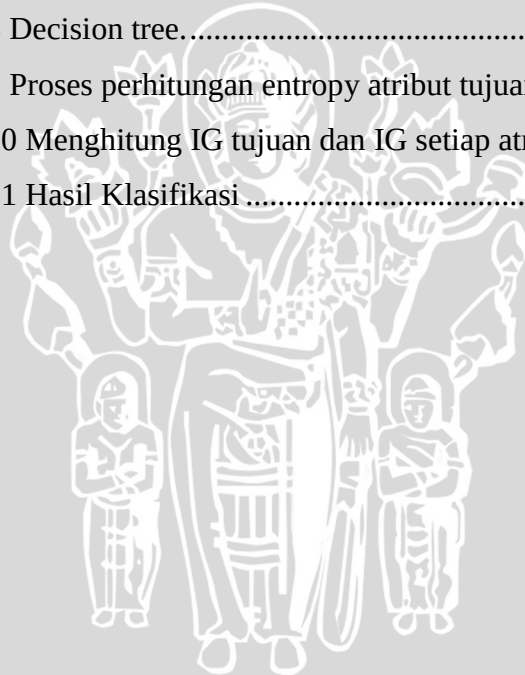
DAFTAR TABEL

No table of contents entries found.



DAFTAR SOURCE CODE

Source Code 4.1 Stopword	66
Source Code 4.2 Stemming	6Error!
Bookmark not defined.	
Source Code 4.3 Frekuensi.....	6Error!
Bookmark not defined.	
Source Code 4.4 Pembobotan data training.....	69
Source Code 4.5 Perhitungan pembobotan.....	71
Source Code 4.6 Distribusi Frekuensi	71
Source Code 4.7 Transformasi data.....	72
Source Code 4.8 Decision tree.....	73
Source Code 4.9 Proses perhitungan entropy atribut tujuan.....	75
Source Code 4.10 Menghitung IG tujuan dan IG setiap atribut. .	76
Source Code 4.11 Hasil Klasifikasi	78



DAFTAR PUSTAKA

- [ASS-91] Assegaf, Dja'far H. 1991. *Jurnalistik Masa Kini*. Jakarta: Ghalia Indonesia.
- [BAS-08] Basuki, Sulisty. 2008. Teknik dan Jasa dokumentasi. Jakarta : Gramedia Pustaka Utama.
- [DEF-10] Defiyanti, Sofi dan Pardede, Crispina D. L. 2010. *Perbandingan Kinerja Algoritma ID3 dan C4.5 Dalam Klasifikasi Spam-Mail*. Depok: Universitas Gunadarma.
- [EVE-02] Even, Yahir dan Zohar. 2002. *Introduction to Teks Mining. Automated Learning Group National Center For Supercomputing Applications*. Chicago: University of Illionis.
- [HAN-06] Han, Jiawei dan Micheline Kamber. 2006. *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers is an imprint of Elsevier.
- [HAR-11] Hardianto, Agung. 2011. *Kasifikasi Berita Berbahasa Indonesia Dengan Pengklasifikasi SVM (support vector machine)* . Malang: Universitas Brawijaya Malang.
- [HEA-99] Hearst, M.A. 1999. *Untangling text data mining. In: Proceedings of ACL'99: the 37 th Annual Meeting of the Association for Computational Linguistics*. Maryland: University of Maryland.

- [HAM-03] Hamilton, H dan Olive, W. 2003. Confusion Matrix.
http://www2.cs.uregina.ca/dbd/cs831/notes/confusion_matrix/confusion_matrix.pdf tanggal akses: 21 Februari 2012
- [MUS-09] Musthafa, Aziz. 2009. *Klasifikasi Otomatis Dokumen Berita Kejadian Berbahasa Indonesia*. Malang: Universitas Islam Negri (UIN) Maulana Malik Ibrahim Malang.
- [PRA-10] Pratiwi, Ifticha D. 2010. *Klasifikasi Dokumen Berita Berbahasa Indonesia Dengan Algoritma Agglomerative Single Linkage Hierarchical Clustering*. Malang : Universitas Brawijaya Malang.
- [RAM-10] Ramlan, M. 1995. *Ilmu Bahasa Indonesia : Morfologi Suatu Tinjauan Deskriptif*. CV. Karyono. Yogyakarta.
- [TAL-03] Tala, Fadillah Z. 2003. *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. Netherlands: Universiteit van Amsterdam The Netherlands.
<http://www.illc.uva.nl/Research/Reports/MoL-2003-02.teks.pdf>
tanggal akses : 18 Juli -11.
- [POE-98]. Poerwadarminta. 1998. *Kamus Umum Bahasa Indonesia*. Jakarta: Balai Pustaka.
- [WIB-05] Wibisono, yudi. Masayu Leylia Khodra. 2005. *Clustering Berita Berbahasa Indonesia*. FPMIPA Universitas Pendidikan Indonesia, Jln Setiabudhi 229 Bandung.
- [YIM-99] Yiming Yang, Jaime G. Carbonell, Rulf D. Brown, Thomas Pierce, Brian T. Achiba Id, Xin Liu. 1999. Learning Approaches for Detecting and Tracking News Events. IEEE Intelligent Systems, Language Technologies Institute, Carbegie Mellon University.

