

TESIS

**DETEKSI JENIS EMOSI DARI TEKS BAHASA  
INDONESIA MENGGUNAKAN *KEYWORD-SPOTTING*  
DAN NAIVE BAYES**



JULIUS VICTOR MANUEL BATA  
No. Mhs.: 145302168/PS/MTF

PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA  
PROGRAM PASCASARJANA  
UNIVERSITAS ATMA JAYA YOGYAKARTA  
2016



PENGESAHAN TESIS

Nama : JULIUS VICTOR MANUEL BATA  
Nomor Mahasiswa : 145302168/PS/MTF  
Konsentrasi : Soft Computing  
Judul Tesis : DETEKSI JENIS EMOSI DARI TEKS BAHASA  
INDONESIA MENGGUNAKAN *KEYWORD-  
SPOTTING* DAN NAIVE BAYES

Nama Penguji	Tanggal	Tanda tangan
Prof. Ir. Suyoto, M. Sc., Ph.D	29-1-2016	
Dr. Pranowo, S.T., M.T.	29-1-2016	
Dr. Ir. Alb. Joko Santoso, M.T.	29-1-2016	

Ketua Program Studi



Prof. Ir. Suyoto, M. Sc., Ph.D

PROGRAM  
PASCASARJANA



UNIVERSITAS ATMA JAYA YOGYAKARTA  
PROGRAM PASCASARJANA  
PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

---

PERNYATAAN

Bersamaan dengan penelitian ini, maka saya yang bertanda tangan di bawah ini :

Nama : JULIUS VICTOR MANUEL BATA

Nomor Mahasiswa : 145302168/PS/MTF

Konsentrasi : Soft Computing

Judul Tesis : Deteksi Jenis Emosi dari Teks Bahasa Indonesia

Menggunakan *Keyword-Spotting* dan Naive Bayes

menyatakan bahwa penelitian ini adalah hasil pemikiran sendiri dan bukan duplikasi dari karya tulis yang telah ada sebelumnya. Karya tulis yang telah ada sebelumnya dijadikan acuan oleh penulis guna melengkapi penelitian ini dan dinyatakan secara tertulis dalam penulisan acuan dan daftar pustaka.

Demikian pernyataan ini dibuat untuk digunakan sebagaimana mestinya.

Yogyakarta, 29 Januari 2016

Julius Victor Manuel Bata

## INTISARI

Sekarang ini emosi dari user memegang peranan penting untuk meningkatkan kinerja sistem komputer. Keadaan emosi user dapat dideteksi dari berbagai media seperti suara, ekspresi wajah dan teks. Deteksi emosi dari teks menjadi penting karena pertumbuhan jumlah data teks pada media online dan media komunikasi berbasis komputer. Sementara itu, penelitian menyangkut deteksi emosi dari teks bahasa Indonesia masih sedikit dilakukan.

Penelitian tesis ini berfokus pada persoalan deteksi emosi dari teks bahasa Indonesia. Untuk menyelesaikan persoalan ini, dua model deteksi dikembangkan. Model yang pertama yaitu berbasis *keyword-spotting* dan yang kedua berbasis *Naive Bayes Classifier*. Model *keyword-spotting* bergantung pada leksikon emosi. Sebagai bagian dari model *keyword-spotting*, tesis ini juga mengembangkan dua leksikon : *baseLex* dan *SoALex*. Model deteksi didasarkan pada struktur hirarki dari leksikon emosi bahasa Indonesia. Bentuk hirarki ini terdiri dari dua kelas *superordinate* dan lima jenis emosi pada tingkat *basic*. Empat percobaan dilakukan untuk mengevaluasi model deteksi.

Berdasarkan hasil percobaan, unjuk kerja *Naive Bayes Classifier* lebih baik untuk deteksi pada tingkat *superordinate*. Pada tingkat *basic*, *Naive Bayes Classifier* lebih baik untuk deteksi pada kelas positif. Sedangkan, *keyword-spotting* lebih baik untuk deteksi kelas negatif.

**Kata-kata kunci : deteksi emosi teks, leksikon emosi, klasifikasi hirarki**

## **ABSTRACT**

Nowadays, user emotions are vital to improve computer system performance. User emotion can be detected from various media such as speech, facial expressions and text. Detection emotion from text is important because the increasing amount of text in on-line media and computer-mediated communication. Meanwhile, there are few research focuses on emotion detection from Indonesian text which is considered as under-resource languages.

This thesis deals with emotion detection from Indonesian text. To solve this problem, two detection models was developed. The first one was based on keyword-spotting and the second was based on Naive Bayes Classifier. As keyword-spotting strongly depend on the availability of emotion-conveying words (emotion lexicon), we also develop two emotion lexicon : baseLex and SoALex. The models based on hierarhical structure of Indonesian emotion lexicon. This structure consist of two superordinate level and five basic emotion. Four type experiment were performed to evaluate the models.

The experimental results using 500 tweet have demonstrated that keyword-spotting and Naive Bayes Classifier is able to detect emotion. The findings of this thesis indicate that Naive Bayes tend to do better for superordinate detection. For basic level, Naive Bayes Classifier is better for detect positive class while keyword-spotting is better for negative class.

**Keywords : text emotion detection, emotion lexicon, hierarhical detection**

## KATA PENGANTAR

Puji syukur kepada BAPA di Surga atas berkat-Nya penulis berhasil menyelesaikan tesis ini. Penulis menyadari bahwa selesainya pembuatan tesis ini tidak bisa terlepas dari bantuan yang diberikan berbagai pihak, untuk itu penulis ingin mengucapkan terima kasih yang sebanyak – banyaknya kepada:

1. Prof. Ir. Suyoto, M.Sc., Ph.D., selaku Ketua Program Studi Magister Teknik Informatika sekaligus menjadi Pembimbing utama tesis ini.
2. Bapak Dr. Pranowo, S.T., M.T., selaku Dosen Pembimbing II yang telah memberikan banyak sekali masukan dan pertanyaan kepada penulis untuk menyelesaikan tesis ini.
3. Bapak Dr. Ir. Alb. Joko Santoso, M.T., selaku Dosen Penguji tesis yang memberikan perbaikan terhadap tesis ini.
4. Seluruh dosen Program Studi Magister Teknik Informatika dan karyawan Program Pascasarjana Universitas Atma Jaya Yogyakarta yang telah membantu penulis selama menempuh studi.
5. Bapa dan ke-empat kakak serta seluruh keluarga yang selalu memberi doa, dukungan serta semangat bagi penulis, hingga akhirnya penulis dapat menyelesaikan tesis ini.
6. Teman – teman MTF September 2014 ta'anters, antgineerd dan 3 idiots serta ci'mon. Selalu lupakan belajar, mulai berpikir dan ciptakan sesuatu.
7. Likas yang selalu menjadi angin serta tujuan bagi penulis.

Demikian penulisan tesis ini dibuat dengan sebaik – baiknya oleh penulis. Penulis menyadari bahwa penulisan tesis ini masih memiliki banyak kekurangan dan jauh dari kata sempurna. Oleh karena itu saran dan kritik yang membangun demi penyempurnaan penulisan tesis ini akan selalu penulis nantikan.

Akhir kata, penulis mengharapkan semoga penulisan tesis ini dapat berguna dan bermanfaat sehingga dapat memberikan inspirasi bagi pembacanya.

Yogyakarta, 29 Januari 2016

Penulis

## DAFTAR ISI

<b>INTISARI .....</b>	<b>iii</b>
<b>ABSTRACT .....</b>	<b>iv</b>
<b>KATA PENGANTAR.....</b>	<b>v</b>
<b>DAFTAR ISI.....</b>	<b>vii</b>
<b>DAFTAR GAMBAR.....</b>	<b>ix</b>
<b>DAFTAR TABEL .....</b>	<b>x</b>
<b>BAB I.....</b>	<b>1</b>
<b>PENDAHULUAN.....</b>	<b>1</b>
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah .....	3
1.3. Batasan Masalah.....	4
1.4. Keaslian Penelitian .....	4
1.5. Manfaat Penelitian.....	5
1.6. Tujuan Penelitian.....	5
1.7. Sistematika Penulisan.....	6
<b>BAB II .....</b>	<b>8</b>
<b>TINJAUAN PUSTAKA .....</b>	<b>8</b>
2.1. Deteksi Emosi dari Teks.....	8
2.2. Deteksi Emosi dari Teks Bahasa Indonesia.....	11
2.3. Rangkuman Penelitian Terkait .....	12
<b>BAB III.....</b>	<b>14</b>
<b>LANDASAN TEORI.....</b>	<b>14</b>
3.1. Model Emosi .....	14
3.1.1. Model <i>categorical</i> .....	15
3.1.2. Model <i>dimensional</i> .....	15
3.2. Deteksi Emosi dari Teks.....	16
3.2.1. Pemilihan sumber data .....	18
3.2.2. Pra-proses.....	18
3.2.3. Pengembangan model deteksi .....	19
3.2.4. Evaluasi.....	21



3.3.	Pendekatan <i>Keyword-Spotting</i> .....	22
3.4.	Pendekatan <i>Learning-Based</i> .....	23
3.4.1.	Tahap pelatihan .....	24
3.4.2.	Tahap klasifikasi .....	24
<b>BAB IV</b>	.....	<b>26</b>
<b>METODOLOGI PENELITIAN</b>	.....	<b>26</b>
4.1.	Data Set .....	26
4.2.	Pra-proses .....	31
4.3.	Pengembangan Model Deteksi.....	34
4.3.1.	Model deteksi hirarki .....	35
4.3.2.	<i>Keyword-spotting</i> .....	37
4.3.3.	<i>Learning based</i> .....	39
4.4.	Evaluasi .....	41
<b>BAB V</b>	.....	<b>43</b>
<b>HASIL PENELITIAN DAN PEMBAHASAN</b>	.....	<b>43</b>
5.1.	Hasil Penelitian.....	43
5.1.1.	Pra-proses .....	44
5.1.2.	Model deteksi <i>keyword-spotting</i> .....	46
5.1.3.	Model deteksi <i>learning based</i> .....	52
5.2.	Pembahasan .....	55
<b>BAB VI</b>	.....	<b>62</b>
<b>KESIMPULAN DAN SARAN</b>	.....	<b>62</b>
6.1.	Kesimpulan.....	62
6.2.	Saran.....	63
<b>DAFTAR PUSTAKA</b>	.....	<b>64</b>

## DAFTAR GAMBAR

<b>Gambar 3.1. Proses Utama Deteksi Emosi dari Teks</b> .....	18
<b>Gambar 3.2. Proses Keyword-Spotting</b> .....	23
<b>Gambar 4.1. . Contoh Tampilan Tweet di Twitter</b> .....	27
<b>Gambar 4.2. Implementasi Tweepy pada Python</b> .....	28
<b>Gambar 4.3. Hashtag Kata Kunci sebagai Filter</b> .....	30
<b>Gambar 4.4. Contoh Data Tweet</b> .....	31
<b>Gambar 4.5. Tahapan Pra-proses</b> .....	32
<b>Gambar 4.6. Struktur Hirarki Emosi</b> .....	36
<b>Gambar 4.7. Langkah Deteksi menggunakan Keyword-Spotting</b> .....	37
<b>Gambar 4.8. Langkah Deteksi menggunakan Naive Bayes Classifier</b> .....	40
<b>Gambar 4.9. Model Evaluasi</b> .....	41
<b>Gambar 5.1. PraProses.java</b> .....	44
<b>Gambar 5.2. Fungsi hapusUser</b> .....	44
<b>Gambar 5.3. Fungsi hapusHTML</b> .....	45
<b>Gambar 5.4. Fungsi hapusURL</b> .....	45
<b>Gambar 5.5. Fungsi cekUnicode</b> .....	46
<b>Gambar 5.6. Hasil Implementasi PraProses</b> .....	46
<b>Gambar 5.7. Contoh Leksikon baseLex</b> .....	47
<b>Gambar 5.8. Kode Program untuk Membuat Leksikon</b> .....	48
<b>Gambar 5.9. Contoh Leksikon SoALex</b> .....	48
<b>Gambar 5.10. Kode Program untuk Tokenisasi</b> .....	50
<b>Gambar 5.11. Cuplikan Kode Program Fungsi doDeteksi</b> .....	50
<b>Gambar 5.12. Kode Program hitungBobotTotal</b> .....	51
<b>Gambar 5.13. Kode Program Penentuan Jenis Emosi</b> .....	51
<b>Gambar 5.14. EkstraksiCiri.java</b> .....	52
<b>Gambar 5.15. Hasil Ekstraksi Ciri</b> .....	53
<b>Gambar 5.16. Potongan Kode Program NaiveBayesClassifier</b> .....	53
<b>Gambar 5.17. Kode Program Fungsi hitungPrior</b> .....	54
<b>Gambar 5.18. Kode Program Fungsi hitungCondProb</b> .....	54
<b>Gambar 5.19. Contoh Isi Variabel modelNB</b> .....	54
<b>Gambar 5.20. Cuplikan Kode Program doDeteksi</b> .....	55
<b>Gambar 5.21. Hasil Proses Deteksi</b> .....	55
<b>Gambar 5.22. Contoh Data Evaluasi</b> .....	56

## DAFTAR TABEL

<b>Tabel 2.1. Rangkuman Penelitian Terkait</b> .....	13
<b>Tabel 3.1. Confusion Matrix</b> .....	21
<b>Tabel 4.1. Daftar Kata Kunci (Shaver, et al., 2001)</b> .....	30
<b>Tabel 4.2. Hasil Pra-proses</b> .....	34
<b>Tabel 5.1. Persebaran Kata dalam Leksikon baseLex</b> .....	48
<b>Tabel 5.2. Perbandingan Bobot baseLex dan SoALex</b> .....	49
<b>Tabel 5.3. Contoh Data Latih NBC</b> .....	52
<b>Tabel 5.4. Hasil Percobaan #KS-SoALex pada Tingkat Superordinate</b> .....	57
<b>Tabel 5.5. Hasil Percobaan #KS-SoALex pada Tingkat Basic Positif</b> .....	57
<b>Tabel 5.6. Hasil Percobaan #KS-SoALex pada Tingkat Basic Negatif</b> .....	58
<b>Tabel 5.7. Confusion Matrix #NBC-bi</b> .....	58
<b>Tabel 5.8. Evaluasi #NBC-bi</b> .....	59
<b>Tabel 5.9. Evaluasi pada Tingkat Superordinate</b> .....	60
<b>Tabel 5.10. Evaluasi pada Tingkat Basic Positif</b> .....	60
<b>Tabel 5.11. Evaluasi pada Tingkat Basic Negatif</b> .....	60