TESIS

KLASIFIKASI TEKS *SHORT MESSAGE SERVICE* (SMS)

HARLIANDI
No. Mhs : 135302026/PS/MTF

PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA
PROGRAM PASCA SARJANA
UNIVERSITAS ATMA JAYA YOGYAKARTA
2016

UNIVERSITAS ATMA JAYA YOGYAKARTA
PROGRAM PASCASARJANA
PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PENGESAHAN DOSEN PEMBIMBING TESIS

Nama	: HARLIANDI
Nomor Mahasiswa	: 135302026 /PS/MTF
Konsentrasi	: Mobile Computing
Judul Tesis	: Klasifikasi Teks *Short Message Service*

Nama Pembimbing	**Tanggal**	**Tanda tangan**
Ir. A. Djoko Budiyanto, M.Eng., Ph.D.	3/8/2016	
Kusworo Anindito, ST., MT.	3/3/2016	

UNIVERSITAS ATMA JAYA YOGYAKARTA
PROGRAM PASCASARJANA
PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PENGESAHAN TIM PENGUJI TESIS

Nama	: HARLIANDI
Nomor Mahasiswa	: 135302026/PS/MTF
Konsentrasi	: Mobile Computing
Judul Tesis	: KLASIFIKASI TEKS *SHORT MESSAGE SERVICE*

Nama Penguji	**Tanggal**	**Tanda tangan**
Ir. A. Djoko Budiyanto, M.Eng., Ph.D. (Ketua)	3/8/2016	
Kusworo Anindito, ST., MT. (Anggota)	3/8/2016	
Prof. Ir. Suyoto, M.Sc., Ph.D. (Anggota)	3/8/2016	

Ketua Program Studi

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

PERNYATAAN KEASLIAN TESIS

Dengan ini penulis menyatakan bahwa tesis ini, yang diberi judul: "Klasifikasi Teks *Short Message Service*" merupakan karya asli penulis. Belum pernah diajukan sebagai persyaratan untuk memperoleh gelar akademik, baik di Universitas Atma Jaya Yogyakarta maupun di Perguruan Tinggi lainnya. Seluruh informasi di dalam tesis ini yang berasal dari penulis lain telah diberikan penghargaan dengan menyebut nama pengarang, judul buku atau tulisan aslinya dan dicantumkan di dalam daftar pustaka.

Yogyakarta, April 2016

Yang Menyatakan,

Harliandi
NPM. 135302026

INTI SARI

Short Message Service (SMS) merupakan salah satu cara untuk melakukan pertukaran pesan melalui perangkat genggam seperti handphone. SMS terdiri dari 160 karakter yang dapat berisi angka, huruf, dan karakter alfanumerik. Klasifikasi SMS merupakan salah satu sub bidang yang sedang banyak dipelajari karena melakukan klasifikasi untuk suatu teks. Dokumen yang berisi teks dengan jumlah sedikit akan berbeda dengan melakukan klasifikasi teks dengan jumlah kata atau kalimat yang banyak dalam sebuah dokumen.

Metode yang digunakan dalam melakukan klasifikasi teks adalah metode Learning Vector Quantization (LVQ) untuk melakukan klasifikasi teks secara otomatis yang termasuk dalam kategori algoritma dari *machine learning*. Selain itu untuk melakukan perbaikan kata pada saat *preprocessing* data akan menggunakan metode Damerau Levensthein.

Klasifikasi SMS menggunakan metode LVQ ini diharapkan memberikan kontribusi terhadap jenis klasifikasi untuk teks dengan jumlah kata yang sedikit seperti SMS dengan akurasi yang tinggi terhadap klasifikasi SMS dengan kategori yang berbeda seperti kejahatan, bencana, dan kecelakaan. Ditambah dengan fitur perbaikan kata pada saat *preprocessing* data diharapkan mampu menambah akurasi pada saat melakukan klasifikasi SMS.

Kata Kunci :*sms, klasifikasi teks, machine learning, lvq, perbaikan feature*

ABSTRACT

Short Message Service (SMS) is one of the way to exchange messages via handheld devices such as mobile phones. SMS consists of 160 characters that can contain numbers, letters, and alphanumeric characters. SMS Classification is one of the sub-fields that are being studied for the classification of a document or text. Documents containing text with a little amount will vary with the classification of text with large number of words or phrases in a document.

The method used in conducting text classification is Learning Vector Quantization (LVQ) to perform automatic text classification are included in the category of machine learning algorithms. In addition to repair word on when preprocessing the data will use Damerau Levensthein methods.

Short message service classification using LVQ is expected for classifying short message text but with the preprocessing of data is aimed to improve the accuracy.

Keyword :*sms, text classification, machine learning, lvq, feature repairing*

KATA PENGANTAR

Puji syukur kepada Tuhan Yesus atas semua berkat dan karunia yang telah diberikan hingga akhirnya Laporan Tesis dengan judul "Klasifikasi Teks *Short Message Service*" ini dapat selesai dengan baik.

Tesis ini disusun untuk memenuhi persyaratan dalam menyelesaikan pendidikan pada Program Studi Magister Teknik Informatika Universitas Atma Jaya Yogyakarta. Dengan penuh rasa syukur karena telah menerima banyak bimbingan, dukungan dan bantuan, penulis mengucapkan banyak terima kasih kepada segenap pihak yang telah berkontribusi dengan penuh ketulusan, keikhlasan, dan kesetiaan hingga tesis ini dapat diselesaikan dengan baik. Penulis ingin mengucapkan terima kasih kepada beberapa pihak yang sangat membantu dalam penyelesaian tesis ini, antara lain :

1. Bapak Prof. Ir. Suyoto, M.Sc., Ph.D selaku Ketua Program Studi Magister Teknik Informatika Program Pascasarjana Universitas Atma Jaya Yogyakarta yang telah memberikan baik ilmu dan motivasi selama menempuh pendidikan.
2. Bapak Ir. A. Djoko Budiyanto, M. Eng., Ph.D selaku pembimbing I dan Bapak Kusworo Anindito, S.T., M.T. selaku pembimbing II serta Bapak Prof. Ir. Suyoto, M.Sc., Ph.D selaku dosen penguji yang telah memberikan waktu, ilmu, bimbingan, arahan, dan saran dari awal penyusunan tesis ini sampai dengan selesai.
3. Segenap Dosen Program Studi Magister Teknik Informatika yang telah memberikan ilmu dan motivasi selama pendidikan, serta Staf Administrasi

dan *student staff* Kantor Administrasi Pasca Sarjana Universitas Atma Jaya Yogyakarta yang telah memberikan bantuan dan pelayanan yang baik selama proses pendidikan.

4. Ayah Juandi, SH dan Ibu Karinah yang selalu memotivasi serta memberikan perhatian, kasih sayang yang terbaik untuk saya.
5. Kedua kakak saya David Perdana Kusuma, SH dan Novia Marcellina, S.Pd yang terus memberikan saya nasehat dan mengingatkan saya untuk terus berjuang.
6. Kristin Gayatri, ST penyemangat dalam setiap langkah yang saya tempuh.
7. Saudaraku Febrian Berthanio, ST, MT atas semua bantuan yang diberikan selama penulis menyelesaikan penelitian ini.
8. Joseph Carlo Kotualubun, ST, MT atas bantuannya dalam menyelesaikan program dan laporan penelitian ini.
9. Semua pihak baik saudara maupun teman yang telah membantu yang tidak bisa disebutkan satu per satu, sekali lagi penulis menyampaikan banyak terima kasih. Tuhan Yesus memberkati kita sekalian.

Penulis menyadari bahwa laporan tesis ini masih banyak memiliki kekurangan dan jauh dari sempurna. Oleh karena itu diharapkan adanya masukan, saran, atau kritik yang membangun agar penuli dapat menjadi lebih baik lagi di masa yang akan datang. Semoga pembaca sekalian memperoleh manfaat yang baik dari laporan tesis ini.

Yogyakarta, Agustus 2016

Penulis

DAFTAR ISI

PENGESAHAN DOSEN PEMBIMBING TESIS ..... **i**

PENGESAHAN TIM PENGUJI TESIS ..... **ii**

PERNYATAAN KEASLIAN TESIS ..... **iii**

INTI SARI ..... **iv**

ABSTRACT ..... **v**

KATA PENGANTAR ..... **vi**

DAFTAR ISI ..... **viii**

DAFTAR TABEL ..... **x**

DAFTAR GAMBAR ..... **xi**

BAB I PENDAHULUAN **1**

A. Latar Belakang 1

B. Perumusan Masalah 4

C. Batasan Masalah 4

D. Keaslian Penelitian 5

E. Manfaat yang diharapkan 5

F. Tujuan Penelitian 5

G. Sistematika Penulisan 6

BAB II TINJAUAN PUSTAKA **9**

BAB III LANDASAN TEORI **15**

A. Machine Learning 15
B. Klasifikasi Teks 15
C. Short Message Service (SMS) 16
D. Linear Vector Quantization (LVQ) 16
E. Damerau Levensthein 18
F. Algoritma Nazief dan Adriani 19
G. Performance Evaluation 23
BAB IV METODOLOGI PENELITIAN 24
A. Alat dan Bahan Penelitian 24
BAB V HASIL PENELITIAN DAN PEMBAHASAN 26
A. Analisa Implementasi Algoritma 26
B. Pengujian 27
C. Pembahasan Hasil Pengujian 27
BAB VI KESIMPULAN DAN SARAN 37
A. Kesimpulan 37
B. Saran 37
DAFTAR PUSTAKA 38

DAFTAR TABEL

Tabel 3. 1 Kombinasi Awalan dan Akhiran yang tidak diijinkan 20

Tabel 3. 2 Aturan awalan me-, be-, pe-, te- 21

Tabel 3.3 Aturan awalan me-, be-, pe-, te- (lanjutan) 22

Tabel 5. 1 Teks SMS kategori Kecelakaan 28

Tabel 5. 2 Teks SMS kategori Kecelakaan (lanjutan) 29

Tabel 5. 3 Teks SMS kategori Bencana 29

Tabel 5. 4 Teks SMS kategori Bencana (lanjutan) 30

Tabel 5. 5 Teks SMS kategori kecelakaan 30

Tabel 5. 6 Teks SMS kategori kecelakaan (lanjutan) 31

Tabel 5. 7 Hasil percobaan terhadap nilai learning rate 34

Tabel 5. 8 Hasil Percobaan Terhadap Nilai Epsilon 35

Tabel 5. 9 Hasil Percobaan Terhadap Nilai MaxEpoch 35

Tabel 5. 10 Hasil pengujian teks SMS untuk 3 kategori 36

Tabel 5. 11 Hasil perhitungan performance evaluation 36

DAFTAR GAMBAR

Gambar 3. 1 Arsitektur Jaringan LVQ 17

Gambar 3. 2 Rumus untuk algoritma edit distance 19

Gambar 5. 1 Flowchart Implementasi Algoritma LVQ 26