

**IMPLEMENTASI *GENERALIZED VECTOR SPACE MODEL*
UNTUK MENGLASIFIKASI JENIS *BULLYING***

Skripsi

**Disusun untuk memenuhi salah satu syarat
memperoleh gelar Sarjana Komputer**



Oleh:

Monica Ratna Andani

3145150806

**PROGRAM STUDI ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI JAKARTA**

2020

ABSTRAK

MONICA RATNA ANDANI. Implementasi *Generalized Vector Space Model* untuk Mengklasifikasi Jenis *Bullying*. Skripsi. Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta. 2020. Di bawah Bimbingan Med Irzal, M. Kom, Ria Arafiyah, M. Si

Bullying yang kerap terjadi, terbagi dalam beberapa tindakan. Beberapa komunitas anti-*bullying* memiliki sebuah fitur berupa pengaduan kasus *bullying*. Namun, saat ini belum memiliki sistem yang dapat mengklasifikasi laporan *bullying*. Oleh karena itu, diperlukan sebuah metode yang dapat mengorganisir dokumen. *Information Retrieval* (IR) merupakan cara yang bisa digunakan untuk mendapatkan informasi yang relevan dengan mengambil data yang tersimpan dalam koleksi dokumen. *Generalized Vector Space Model* (GVSM) merupakan salah satu model sistem IR yang termasuk dalam *Algebraic Model* yang merupakan perluasan dari VSM. Pada penelitian ini akan dibuat suatu program untuk mengklasifikasi jenis *bullying* menggunakan konsep IR dengan metode GVSM. Kumpulan dokumen pengaduan *bullying* disimpan dalam suatu folder. Term indikasi *bullying* dijadikan *query* untuk mengklasifikasi jenis *bullying*. *Query* dan dokumen *bullying* yang dimasukkan akan diproses terlebih dahulu. Pengolahan kata meliputi tokenizing, stemming, dan filtration. Kemudian hasilnya akan diproses menggunakan GVSM untuk dicari nilai kemiripan dengan *query* dan diurutkan dari yang tertinggi. Dari hasil pengujian program IR telah dikembangkan dengan metode GVSM dalam menemukan kembali dokumen *bullying* yang berformat *.txt. Sistem IR dengan metode GVSM ini mampu menemukan kembali dokumen dan mengurutkan sesuai dengan *query*. Hasil dari dokumen yang muncul merupakan representasi dari klasifikasi jenis *bullying* yang ada pada dokumen. Sistem dapat membedakan dan mengklasifikasi data *bullying* yang ada berdasarkan *query*.

Kata kunci : *Bullying, Information Retrieval, Generalized Vector Space Model*

ABSTRACT

MONICA RATNA ANDANI. Implementasi *Generalized Vector Space Model* untuk Mengklasifikasi Jenis *Bullying*. Skripsi. Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta. 2020. Di bawah Bimbingan Med Irzal, M. Kom, Ria Arafiyah, M. Si

*Bullying that often occurs is divided into several types of actions. Some anti-bullying communities have a feature to complaints of bullying cases. However, currently does not have a system that can classify existing bullying reports. Therefore, we need a method that can organize documents. Information Retrieval (IR) is a way that can be used to obtain relevant information by retrieving data stored in a document collection. Generalized Vector Space Model (GVSM) is one of the IR system models included in the Algebraic Model which is an extension of VSM. In this study a program will be made to classify the type of bullying using the concept of IR with the GVSM method. A collection of bullying complaint documents is stored in a folder. Bullying indication is used as a query to classify the type of bullying. The entered query and bullying document will be processed first. Word processing includes tokenizing, stemming, and filtration. Then the results will be processed using GVSM to find the similarity value of the query and sorted from the highest. From the results of testing the IR program has been developed by the GVSM method in rediscovering the bullying document format *.txt. The IR system with the GVSM method is able to find documents back and sort according to the query. The results of the documents that appear are a representation of the classification of the types of bullying. The system can distinguish and classify existing bullying data based on queries.*

Keywords : *Bullying, Information Retrieval, Generalized Vector Space Model*


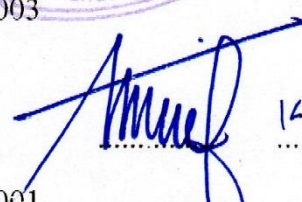


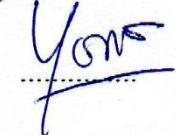
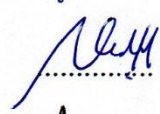

LEMBAR PERSETUJUAN HASIL SIDANG SKRIPSI

Implementasi *Generalized Vector Space Model*

untuk Mengklasifikasi Jenis *Bullying*

Nama : Monica Ratna Andani

No. Registrasi : 3145150806

	Nama	Tanda Tangan	Tanggal
Penanggung Jawab			
Dekan	: Dr. Adisyahputra, M.S. NIP. 19601111 198703 1 003		13-02-2020
Wakil Penanggung Jawab			
Wakil Dekan I	: Dr. Muktiningsih N, M.Si. NIP. 19640511 198903 2 001		14-02-2020
Ketua	: Ir. Fariani Hermin I, M.T. NIP. 19600211 198703 2 001		10-02-2020
Sekretaris	: Ibnu Hadi, M.Si NIP. 19810718 200801 1 017		10-02-2020
Penguji	: Drs. Mulyono, M.Kom. NIP. 19660517 199403 1 003		12-02-2020
Pembimbing I	: Med Irzal, M.Kom. NIP. 19770615 200312 1 001		12-02-2020
Pembimbing II	: Ria Arafiyah, M. Si NIP. 19751121 200501 2 004		12-02-2020

Dinyatakan lulus ujian skripsi tanggal: 07 Februari 2020

LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul **“Implementasi *Generalized Vector Space Model* untuk Mengklasifikasi Jenis *Bullying*”** yang disusun sebagai syarat untuk memperoleh gelar Sarjana komputer dari Program Studi Ilmu Komputer Universitas Negeri Jakarta adalah karya ilmiah saya dengan arahan dari dosen pembimbing.

Sumber informasi yang diperoleh dari penulis lain yang telah dipublikasikan yang disebutkan dalam teks skripsi ini, telah dicantumkan dalam Daftar Pustaka sesuai dengan norma, kaidah dan etika penulisan ilmiah.

Jika dikemudian hari ditemukan sebagian besar skripsi ini bukan hasil karya saya sendiri dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sanding dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku.

Jakarta, 25 Januari 2020



Monica Ratna Andani



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS NEGERI JAKARTA
UPT PERPUSTAKAAN

Jalan Rawamangun Muka Jakarta 13220
Telepon/Faksimili: 021-4894221
Laman: lib.unj.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : MONICA RATNA ANDANI
NIM : 31AS150806
Fakultas/Prodi : FMIPA / ILMU KOMPUTER
Alamat email : moncandani@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

Skripsi Tesis Disertasi Lain-lain (.....)

yang berjudul :

Implementasi Generalized Vector Space Model
Untuk Mengklasifikasi Jenis Bullying

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 19 Februari 2020

Penulis

(MONICA RATNA ANDANI)
nama dan tanda tangan

KATA PENGANTAR

Alhamdulillah rabbil'alamiin

Segala puji bagi Allah Subhanahu wa ta'ala, karena curahan berkah dan rahmat-Nya yang telah memberikan segala bentuk pertolongan pada setiap langkah penyusunan skripsi ini sehingga bisa terselesaikan dengan baik.

Shalawat serta salam semoga selalu tercurah untuk tauladan sepanjang masa, Rasulullah Muhammad shallallahu'alaihiwasallam. Karena perjuangan beliau yang indah Islam dapat dirasakan dan petunjuk Allah SWT dapat tersampaikan yang merupakan karunia bagi seluruh alam semesta.

Pada proses penyusunan skripsi ini terdapat banyak rintangan dan tantangan yang penulis hadapi. Namun karena motivasi dan bimbingan dari kerabat dekat dan orang-orang luar biasa menjadi salah satu faktor penulis bisa menyelesaikan proposal ini sebagai salah satu tugas akhir guna memperoleh gelar Sarjana Komputer dari Program Studi Ilmu Komputer Universitas Negeri Jakarta dengan Judul: **“Implementasi *Generalized Vector Space Model* untuk Mengklasifikasi Jenis *Bullying*”**

Oleh karena itu, pada kesempatan ini dengan kerendahan hati penulis ingin mengucapkan terima kasih kepada:

1. Ibu Ir. Fariani Hermin I, M.T selaku Koordinator Program Studi Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Jakarta.
2. Bapak Med Irzal, M. Kom selaku dosen pembimbing I dan pembimbing akademik penulis yang telah membimbing, mengarahkan, serta memberikan saran yang baik untuk penulis dalam menyelesaikan Tugas Akhir ini.
3. Ibu Ria Arafiyah, M. Si selaku dosen pembimbing II yang telah membimbing, mengarahkan, serta memberikan saran yang baik untuk penulis dalam menyelesaikan Tugas Akhir ini.
4. Seluruh Dosen Prodi Ilmu Komputer FMIPA UNJ yang tidak bisa disebutkan satu per satu, atas ilmu dan bimbingannya selama penulis berkuliah di Ilmu Komputer FMIPA UNJ,

5. Kedua orang tua penulis yang selama ini senantiasa memberikan apapun yang mampu diberikan untuk penulis baik dalam bentuk materi maupun moriil, telah membesarkan penulis dengan sabar membimbing hingga penulis mendapat pendidikan yang layak, serta tak pernah lelah mendoakan penulis.
6. Seluruh teman-teman Program Studi Ilmu Komputer 2015, khususnya Tiwi yang selama ini selalu kebersamai dan setia menjadi teman sebangku penulis, Hida & Rani yang selalu ada saat dibutuhkan, Saulia, Karina, Diana, Mesita, dan Herlinda yang selalu mendukung penulis mengerjakan skripsi, serta Erik yang selalu bersedia membantu penulis selama kuliah.
7. Kakak tingkat Prodi Ilmu Komputer 2013-2014, khususnya Ka Saphira, Ka Ummu, Ka Koko, dan Ka Ardiansyah yang telah menjadi contoh yang baik untuk penulis dan selalu membimbing sejak maba.
8. Sahabat tercinta, Mitha dan Thaniya, yang telah memberikan dorongan serta bantuan yang besar kepada penulis selama menyelesaikan skripsi ini, serta Media yang selalu mendukung dan menemani penulis.
9. Teman-teman yang tak lekang oleh waktu, Hana, Mega, dan Rosita, yang selalu bersedia meluangkan waktunya mendengarkan keluh kesah penulis.
10. Seluruh teman-teman organisasi penulis mulai dari BEM Matematika, BEM FMIPA, dan BEM UNJ khususnya Refani, Mutiara, Ulan dan Assofa yang telah memberikan semangat kepada penulis dalam mengemban amanah.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Masih terdapat banyak kekurangan baik dari segi penyajian materi, bahasa maupun tata cara penulisan. Akhir kata, teriring permintaan maaf apabila terdapat kesalahan maupun kekeliruan dalam penulisan Skripsi ini. Besar harapan Penulis agar skripsi ini dapat berguna sebagai sumber bacaan bagi seluruh kalangan sehingga dapat bermanfaat bagi lingkungan akademik maupun pihak lain yang membutuhkannya. Atas perhatiannya, Penulis mengucapkan terima kasih.

Jakarta, Januari 2020

Penulis

DAFTAR ISI

KATA PENGANTAR	vi
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xii
I. PENDAHULUAN	1
A. Latar Belakang Masalah	1
B. Perumusan Masalah	3
C. Pembatasan Masalah	4
D. Tujuan Penelitian	4
E. Manfaat Penelitian	4
II. LANDASAN TEORI	6
A. Bullying	6
B. Information Retrieval	9
C. Algebraic Model	14
1. Vector Space Model	15
2. Generalized Vector Space Model (GVSM)	17
III. DESAIN MODEL	19
A. Data Penelitian	19
B. Metode Pengumpulan Data	19
C. Alur Kerja Sistem	20
1. Pre-processing	21
2. Metode Generalized Vector Space Model	27
IV. HASIL DAN PEMBAHASAN	28
A. Pemodelan Data	29
B. Simulasi Generalized Vector Space Model	30
C. Implementasi	34
1. Pre-processing	34
2. Implementasi Algoritma Generalized Vector Space Model	35
D. Evaluasi	40

V. KESIMPULAN DAN SARAN	42
A. Kesimpulan	42
B. Saran	43
DAFTAR PUSTAKA	44
LAMPIRAN A	46
LAMPIRAN B	51
LAMPIRAN C	54
LAMPIRAN D	69
DAFTAR RIWAYAT HIDUP	71



DAFTAR GAMBAR

Gambar 2. 1Arsitektur IR.....	11
Gambar 2. 2 Ilustrasi <i>Vector Space Model</i>	16
Gambar 3. 1 Alur Kerja Sistem.....	20
Gambar 3. 2 Alur proses Pre-Processing.....	21
Gambar 3. 3 Alur Proses Tokenizing.....	23
Gambar 3. 4 Alur Proses Stemming.....	24
Gambar 3. 5 Alur Proses Filtration.....	26
Gambar 3. 6 Alur Proses Implementasi Metode GVSM.....	27
Gambar 4. 1 Hasil Pre-Processing.....	35
Gambar 4. 2 Hasil penentuan minterm berdasarkan query.....	36
Gambar 4. 3 Hasil hitung frekuensi kemunculan kata.....	37
Gambar 4. 4 Array Korelasi.....	38
Gambar 4. 5 Vektor Query dan Vektor Dokumen.....	39
Gambar 4. 6 Nilai Similaritas.....	40