

**KLASIFIKASI TEMA MENGGUNAKAN ALGORITMA
GENERALIZED VECTOR SPACE MODEL (GVSM) – IMPROVED
KNN PADA SOAL UJIAN NASIONAL**

SKRIPSI

**Diajukan untuk memenuhi salah satu syarat mendapatkan gelar Strata Satu
Program Studi Informatika**



Disusun oleh :

NURMA AYU WIGATI SLAMET SUBROTO

M0513035

**PROGRAM STUDI INFORMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SEBELAS MARET
SURAKARTA**

2019

PERSETUJUAN

SKRIPSI

KLASIFIKASI TEMA MENGGUNAKAN ALGORITMA GENERALIZED VECTOR SPACE MODEL (GVSM) – IMPROVED KNN PADA SOAL UJIAN NASIONAL

Disusun Oleh :

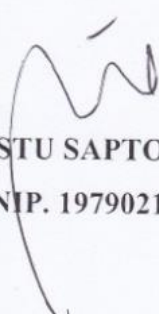
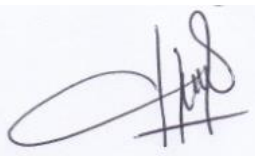
NURMA AYU WIGATI SLAMET SUBROTO

M0513035

**Skripsi ini telah disetujui untuk dipertahankan di hadapan dewan penguji,
pada tanggal 2 April, 2019**

Pembimbing I

Pembimbing II

	
RISTU SAPTONO, S.Si., M.T.	DENIS EKA CAHYANI, S.Kom., M.Kom.
NIP. 197902102002121001	NIP. 1991031020161001

PENGESAHAN
SKRIPSI
KLASIFIKASI TEMA MENGGUNAKAN ALGORITMA
GENERALIZED VECTOR SPACE MODEL (GVSM) – IMPROVED
KNN PADA SOAL UJIAN NASIONAL

Disusun oleh :

NURMA AYU WIGATI SLAMET SUBROTO
M0513035

telah dipertahankan di hadapan dewan penguji
pada tanggal : 2 April 2019

Susunan Dewan Penguji

1. Ristu Saptono, S.Si., M.T.
NIP. 197902102002121001
2. Denis Eka Cahyani, S.Kom, M.Kom.
NIP. 1991031020161001
3. Dr. Tech. Dewi Wisnu Wardani, S.Kom, M.S.
NIP. 197810262005012002
4. Dr. Wisnu Widiarto, S.Si., M.T.
NIP. 197006012008011009

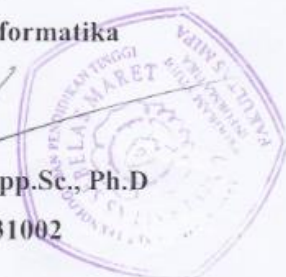
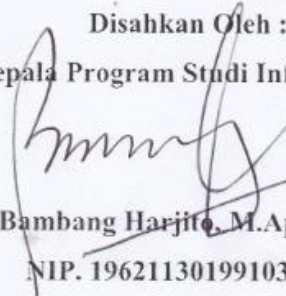


Disahkan Oleh :

Kepala Program Studi Informatika

Drs. Bambang Harjito, M.App.Sc., Ph.D

NIP. 196211301991031002



KATA PENGANTAR

Puji syukur kepada Allah Subhanahu Wa Ta'Ala atas segala limpahan rahmat, nikmat, dan karuniaNya, sehingga penulis dapat menyelesaikan skripsi dengan judul “Klasifikasi Tema Menggunakan Algoritma *Generalized Vector Space Model (GVSM)* – *Improved KNN* pada Soal Ujian Nasional”. Sholawat dan salam senantiasa penulis haturkan kepada Rasulullah Sallahu ‘Alaihi Wassalam sebagai pembimbing seluruh umat manusia.

Laporan skripsi ini disusun untuk memenuhi salah satu syarat dalam memperoleh gelar Sarjana Komputer. Dalam melakukan penelitian dan penyusunan laporan skripsi, penulis mendapatkan dukungan dan bantuan dari banyak pihak, karena itu penulis menyampaikan terimakasih kepada :

1. Bapak dan Ibu, serta keluarga yang selalu memberikan dukungan dan semangat untuk menyelesaikan tugas akhir.
2. Bapak Drs. Bambang Harjito, M.App.Sc., Ph.D., Kepala Program Studi Informatika Universitas Sebelas Maret.
3. Bapak Ristu Saptono, S.Si., M.T., Dosen Pembimbing I dan Ibu Denis Eka Cahyani, S.Kom, M.Kom., Dosen Pembimbing II yang telah dengan sabar memberikan bimbingan dan petunjuk dalam menyelesaikan skripsi ini.
4. Khoirun Nissa Isti, S.Kom., dan Widyaningdyah Hidayati, S.Kom. yang telah membantu menginputkan data soal ujian nasional.

Semoga Allah membalas jerih payah dan pengorbanan yang telah diberikan dengan balasan yang lebih baik. Aamiin. Penulis menyadari bahwa laporan skripsi ini masih jauh dari sempurna, meskipun demikian, penulis berharap semoga karya kecil ini dapat memberikan manfaat dan wawasan tambahan bagi pembaca, khususnya bagi penulis.

Surakarta, April 2019

Penulis

KLASIFIKASI TEMA MENGGUNAKAN ALGORITMA *GENERALIZED VECTOR SPACE MODEL (GVSM)* – *IMPROVED KNN* PADA SOAL UJIAN NASIONAL

Nurma Ayu Wigati Slamet Subroto

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Sebelas Maret

ABSTRAK

Ujian Nasional (UAN atau UNAS atau UN) sebagai alat ukur evaluasi pemerintah untuk menentukan kualitas pendidikan di Indonesia. Kualitas ditunjukkan saat siswa dapat mengerjakan soal ujian nasional berdasarkan materi sesuai dengan Standar Kompetensi Lulusan (SKL). Soal dikelompokkan ke berbagai tema. Klasifikasi tema berguna untuk mengetahui golongan soal yang ada di dalam materi Standar Kompetensi Lulusan (SKL). Penelitian ini bertujuan untuk mengetahui kinerja algoritma *Generalized Vector Space Model (GVSM)* – *improved KNN* dalam melakukan klasifikasi soal berdasarkan tema. Algoritma *GVSM* digunakan untuk mengidentifikasi kemiripan kata yang muncul di dokumen yang satu dengan dokumen yang lain. Algoritma *improved KNN* ini mengklasifikasikan soal ujian nasional berdasarkan tema dalam mata pelajaran dengan menganalisa semua kata yang muncul pada soal ujian nasional. Pengujian dilakukan sebanyak 10 kali dengan metode *k-fold cross validation* menghasilkan nilai akurasi sebesar 0,7939, presisi sebesar 0,7771, dan *recall* sebesar 0,7633.

Kata kunci : Ujian Nasional, Standar Kompetensi Lulus (SKL), Klasifikasi Tema, *Generalized Vector Space Model*, *Improved KNN*.

THEME CLASSIFICATION USING GENERALIZED VECTOR SPACE MODEL (GVSM) – IMPROVED KNN ALGORITHM IN NATIONAL EXAM QUESTIONS

Nurma Ayu Wigati Slamet Subroto

*Department of Informatics, Faculty of Mathematics and Natural Science,
Sebelas Maret University*

ABSTRACT

The National Examination (UAN or UNAS or UN) is used as a measure of government evaluation to determine the quality of education in Indonesia. The quality of education obtained depends on the success of students in working on national exam questions based on the material in the graduate competency standard (Standar Kompetensi Lulusan or SKL). The National exam questions are grouped into various themes. The theme classification is useful for knowing the class of questions that are in the material of graduate competency standard. This research aims to determine the performance of the Generalized Vector Space Model (GVSM) – improved KNN algorithm in classifying national exam questions based on themes. The GVSM algorithm is used to identify the similarity of words that appear in one document with another document. This improved KNN algorithm classifies national exam questions with analyzing all the words that appear on national exam questions based on subject's themes. The classification evaluation with the GVSM - improved KNN algorithm got 0,7939 accuracy, 0,7771 precision, and 0,7633 recall on evaluation using 10-fold cross validation.

Keywords : *The National Examination, Graduate Competency Standard, Theme Classification, Generalized Vector Space Model, Improved KNN.*

DAFTAR ISI

PERSETUJUAN	ii
PENGESAHAN	iii
MOTTO	iv
PERSEMBAHAN	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
BAB 1 PENDAHULUAN	1
1.1.Latar Belakang Masalah	1
1.2.Rumusan Masalah	3
1.3.Batasan Masalah	4
1.4.Tujuan Penelitian	4
1.5.Manfaat Penelitian	4
1.6.Sistematika Penulisan	4
BAB 2 TINJAUAN PUSTAKA	6
2.1.Dasar Teori	6
2.1.1. <i>Text Preprocessing</i>	6
2.1.1.1. <i>Case folding dan tokenizing</i>	6
2.1.1.2. <i>Stopword removal atau filtering</i>	6
2.1.1.3. <i>Stemming</i>	7

2.1.2. <i>Generalized Vector Space Model</i>	14
2.1.3. Pembobotan <i>TF - IDF</i>	15
2.1.4. <i>Text Similarity</i>	16
2.1.5. <i>Improved KNN</i>	18
2.1.6. <i>Cross Validation</i>	19
2.2. Penelitian Terkait	20
BAB 3 METODOLOGI PENELITIAN	25
3.1. Pengumpulan Data	26
3.2. <i>Text Preprocessing</i>	26
3.3. Pemodelan <i>GVSM</i>	26
3.4. Pembobotan <i>TF – IDF</i>	27
3.5. <i>Text Similarity</i>	28
3.6. Perankingan	28
3.7. Klasifikasi <i>Improved KNN</i>	29
3.8. Evaluasi	29
BAB 4 HASIL DAN PEMBAHASAN	32
4.1. Pengumpulan Data	32
4.2. <i>Text Preprocessing</i>	35
4.3. Pemodelan <i>GVSM</i>	38
4.4. Pembobotan <i>TF – IDF</i>	42
4.5. <i>Text Similarity</i>	48
4.6. Perankingan	50
4.7. Klasifikasi <i>Improved KNN</i>	51
4.8. Evaluasi	62
BAB 5 PENUTUP	71
5.1. Kesimpulan	71

5.2. Saran.....	71
DAFTAR PUSTAKA	72
LAMPIRAN.....	76

