

**PENERAPAN *TEXT MINING* UNTUK KLASTERISASI DATA KRITIK  
ALUMNI UNIVERSITAS AMIKOM YOGYAKARTA  
DENGAN *K-MEANS* DAN TF-IDF**

**SKRIPSI**



disusun oleh  
**Nur Hidayah**  
**17.11.1609**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2022**

**PENERAPAN *TEXT MINING* UNTUK KLASTERISASI DATA KRITIK  
ALUMNI UNIVERSITAS AMIKOM YOGYAKARTA  
DENGAN *K-MEANS* DAN TF-IDF**

**SKRIPSI**

Untuk memenuhi sebagian persyaratan mencapai gelar Sarjana  
pada Program Studi Informatika



disusun oleh

**Nur Hidayah**

**17.11.1609**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2022**

# PERSETUJUAN

## SKRIPSI

**PENERAPAN *TEXT MINING* UNTUK KLASTERISASI DATA KRITIK  
ALUMNI UNIVERSITAS AMIKOM YOGYAKARTA  
DENGAN *K-MEANS* DAN *TF-IDF***

yang dipersiapkan dan disusun oleh

**Nur Hidayah**

**17.11.1609**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 22 April 2022

**Dosen Pembimbing,**

**Windha Mega Pradnya D, M.Kom**  
**NIK. 190302185**

**PENGESAHAN**

**SKRIPSI**

**PENERAPAN *TEXT MINING* UNTUK KLASTERISASI DATA KRITIK  
ALUMNI UNIVERSITAS AMIKOM YOGYAKARTA  
DENGAN *K-MEANS* DAN TF-IDF**

yang dipersiapkan dan disusun oleh

**Nur Hidayah**

**17.11.1609**

telah dipertahankan di depan Dewan Penguji  
pada tanggal 22 April 2022

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Anna Baita, M.Kom**  
**NIK. 190302290**

**Arif Akbarul Huda, S.Si, M.Eng**  
**NIK. 190302287**

**Kusnawi, S.Kom, M.Eng**  
**NIK. 190302112**

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 22 April 2022

**DEKAN FAKULTAS ILMU KOMPUTER**

**Hanif Al Fatta, M.Kom**  
**NIK. 190302096**

## PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 15 September 2022



Nur Hidayah

NIM. 17.11.1609

## PERSEMBAHAN

Dengan segala puji syukur kepada Allah SWT, Tuhan yang Maha Esa dan atas dukungan serta doa dari orang tua dan orang-orang tercinta. Alhamdulillah skripsi ini dapat diselesaikan dengan baik dan tepat waktunya. Dengan rasa bahagia dan bangga saya ucapkan rasa syukur dan terima kasih kepada :

1. Allah SWT atas rahmat, anugerah dan karunia-Nya kepada kita semua sehingga atas izin Allah SWT skripsi ini dapat selesai pada waktunya. Puji syukur kepada Allah SWT yang telah meridhoi dan mengabulkan segala doa.
2. Bapak dan Ibu beserta keluarga besar saya yang senantiasa memberikan dukungan kepada saya baik dari materi maupun doa yang tiada henti untuk kesuksesan saya. Tiada doa paling mujarab melainkan doa kedua orang tua, terima kasih untuk Bapak dan Ibu yang selalu mendukung dan mendoakan saya.
3. Ibu Windha Mega Pradnya D, M.Kom selaku dosen pembimbing yang tulus ikhlas membimbing, mengarahkan dan meluangkan waktunya kepada saya agar lebih baik.
4. Bapak dan Ibu dosen Universitas AMIKOM Yogyakarta yang selama ini telah menuntun dan mengarahkan saya, memberikan bimbingan dan pelajaran kepada saya yang tiada ternilai harganya.
5. Seluruh teman-teman Lab Eksplorasi dan juga Forum Asisten yang selalu mengingatkan saya untuk mengerjakan skripsi.
6. Keluarga besar 17 S1 IF 11 Universitas AMIKOM Yogyakarta yang sudah mendukung saya.
7. Teman-teman yang tidak dapat disebutkan satu-persatu terima kasih atas segala dukungan dan bantuannya.

Terima kasih yang sebesar-besarnya telah mendukung saya dalam mengerjakan skripsi ini, semoga dapat memberikan manfaat bagi semua pihak.

## KATA PENGANTAR

Puji syukur kepada Allah SWT atas rahmat dan karunia-Nya yang telah memberikan taufik dan hidayah sehingga dapat menyelesaikan skripsi yang berjudul “Penerapan Text Mining untuk Klasterisasi Data Kritik Alumni Universitas AMIKOM Yogyakarta dengan K-Means dan TF-IDF ” sebagai syarat memperoleh gelar sarjana Strata-1 (S1) jurusan Informatika dengan baik.

Dalam penulisan skripsi ini, saya dapat menyelesaikan dengan maksimal berkat bantuan dan kerjasama dari berbagai pihak. Oleh karena itu saya mengucapkan banyak terima kasih kepada segenap pihak-pihak yang telah ikhlas berkontribusi dan membantu saya dalam menyelesaikan skripsi ini

Pada penulisan skripsi ini, saya sebagai penulis menyadari masih banyak kekurangan dalam penulisan skripsi yang baik dan benar, baik dari tata Bahasa, susunan kalimat maupun isi. Dengan itu saya menerima segala kritik dan saran yang membangun dari pembaca.

Demikian yang bisa saya sampaikan, semoga skripsi ini dapat menambah wawasan ilmu pengetahuan dan memberi manfaat bagi pembacanya.

Yogyakarta, 15 September 2022

Nur Hidayah

## DAFTAR ISI

<b>JUDUL</b> .....	<b>i</b>
<b>PERSETUJUAN</b> .....	<b>ii</b>
<b>PENGESAHAN</b> .....	<b>iii</b>
<b>PERNYATAAN</b> .....	<b>iv</b>
<b>PERSEMBAHAN</b> .....	<b>v</b>
<b>KATA PENGANTAR</b> .....	<b>vi</b>
<b>DAFTAR ISI</b> .....	<b>vii</b>
<b>DAFTAR TABEL</b> .....	<b>x</b>
<b>DAFTAR GAMBAR</b> .....	<b>xi</b>
<b>INTISARI</b> .....	<b>xiii</b>
<b>ABSTRACT</b> .....	<b>xiv</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang Masalah .....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Masalah .....	3
1.4 Maksud dan Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	4
1.6 Metodologi Penelitian .....	4
1.7 Sistematika Penulisan .....	6
<b>BAB II LANDASAN TEORI</b> .....	<b>8</b>
2.1 Tinjauan Pustaka .....	8
2.2 Landasan Teori .....	12
2.2.1 Data Mining .....	12
2.2.2 Text Mining .....	13
2.2.3 Text Clustering .....	14
2.2.4 Text Preprocessing .....	14
2.2.5 Feature Selection .....	16
2.2.6 Pembobotan Kata ( <i>Term Weighting</i> ) .....	18



2.2.7 K-Means .....	20
2.2.8 Python .....	22
2.2.9 System Development Life Cycle (SDLC) .....	22
2.3 Teori Pengujian .....	24
2.3.1 Sum of Square Error (SSE) .....	24
<b>BAB III ANALISIS DAN PERANCANGAN .....</b>	<b>27</b>
3.1 Analisis Sistem .....	27
3.2 Struktur Organisasi .....	28
3.3 Analisis Masalah .....	28
3.4 Analisis Kebutuhan .....	29
3.4.1 Analisis Kebutuhan Fungsional .....	29
3.4.2 Analisis Kebutuhan Non Fungsional .....	29
3.5 Alur Penelitian .....	32
3.6 Pemodelan Sistem .....	32
3.6.1 Penyajian Data .....	32
3.6.2 Tahap Preprocessing .....	33
3.6.3 Tahap Klasterisasi .....	37
3.6.4 Usecase Diagram .....	42
3.6.5 Activity Diagram .....	42
3.6.6 Class Diagram .....	44
3.6.7 Rancangan Database .....	44
3.6.8 Perancangan Interface .....	46
<b>BAB IV IMPLEMENTASI DAN PEMBAHASAN .....</b>	<b>52</b>
4.1 Database Dan Tabel .....	52
4.1.1 Implementasi Database .....	52
4.1.2 Implementasi Tabel .....	53
4.1.3 Koneksi Database .....	54
4.2 Implementasi Antarmuka Pengguna .....	55
4.2.1 Halaman Login .....	55

4.2.2 Halaman Dashboard .....	57
4.2.3 Halaman Input Data .....	58
4.2.4 Halaman Output Data .....	62
4.2.5 Halaman About .....	63
4.2.6 Halaman History .....	64
4.2.7 Halaman Archive .....	65
4.3 Pengujian Sistem .....	66
4.3.1 Pengujian Black Box .....	66
4.3.2 Pengujian Metode Sum of Square Error (SSE) .....	67
4.4 Implementasi Program .....	68
4.4.1 Manual Program .....	68
4.5 Pemeliharaan Sistem .....	69
<b>BAB V PENUTUP .....</b>	<b>71</b>
5.1 Kesimpulan .....	71
5.2 Saran .....	71
<b>DAFTAR PUSTAKA .....</b>	<b>73</b>

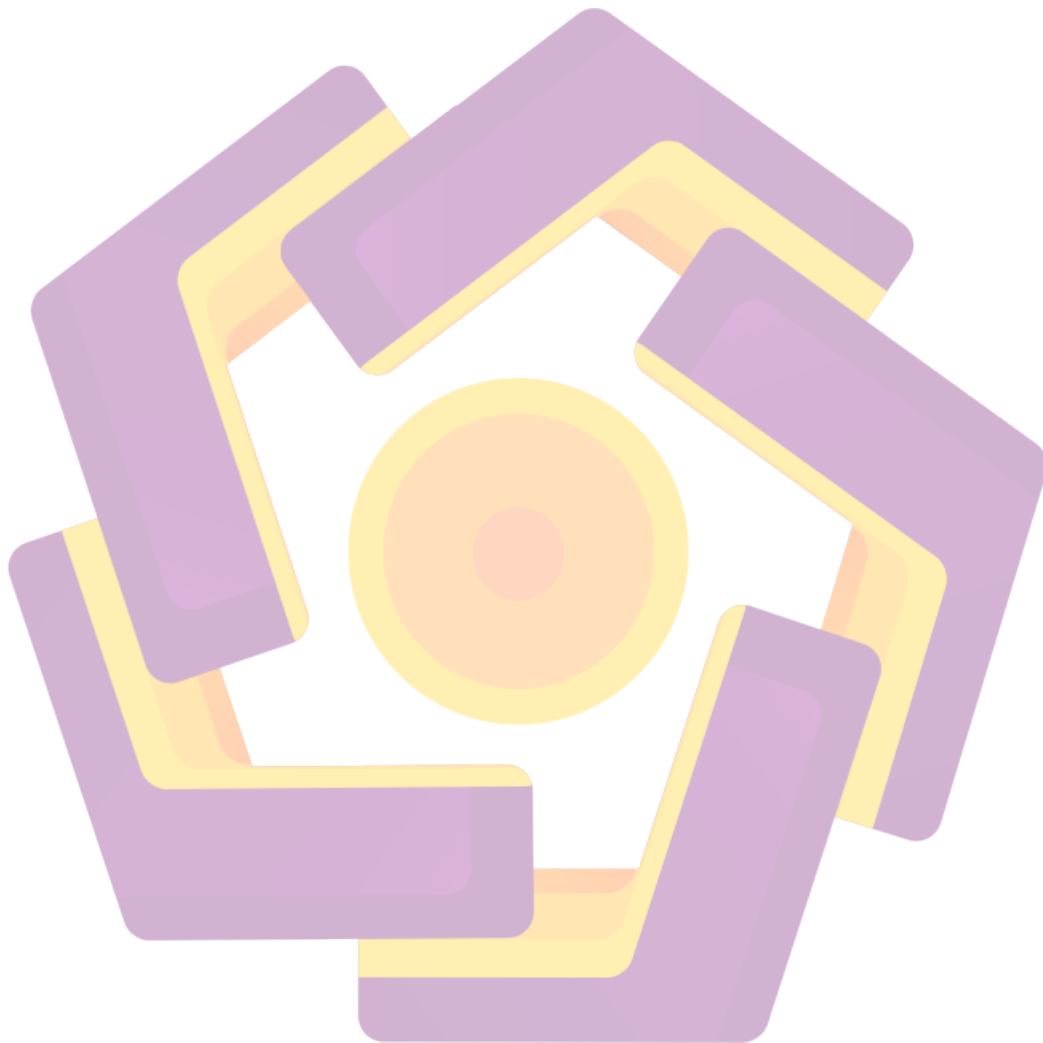
## DAFTAR TABEL

Tabel 2.1 – Studi Literatur .....	10
Tabel 3.1 – Kebutuhan Perangkat Minimum .....	30
Tabel 3.2 – Perangkat Lunak Yang Digunakan .....	30
Tabel 3.3 – Perangkat Keras Yang Digunakan .....	31
Tabel 3.4 – Data Sampel .....	33
Tabel 3.5 – Case Folding .....	33
Tabel 3.6 – Tokenizing .....	34
Tabel 3.7 – Stopword Removal .....	35
Tabel 3.8 – Stemming .....	35
Tabel 3.9 – TF-IDF .....	36
Tabel 3.10 – Hasil Term Weighting .....	37
Tabel 3.11 – Penentuan Centroid .....	38
Tabel 3.12 – Nilai Minimum Dari Titik Pusat Cluster .....	40
Tabel 3.13 – Hasil Iterasi 1 .....	40
Tabel 3.14 – Hasil Centroid Baru .....	41
Tabel 3.15 – Admin .....	45
Tabel 3.16 – Cluster .....	45
Tabel 3.17 – Document .....	45
Tabel 3.18 – Critic .....	46
Tabel 4.1 – Black Box Testing .....	67

## DAFTAR GAMBAR

Gambar 2.1 – Model Waterfall .....	23
Gambar 2.2 – Grafik Metode Elbow .....	25
Gambar 3.1 – Struktur Organisasi.....	28
Gambar 3.2 – Alur Penelitian .....	32
Gambar 3.3 – Usecase Diagram .....	42
Gambar 3.4 – Activity Diagram .....	43
Gambar 3.5 – Class Diagram .....	44
Gambar 3.6 – Rancangan Database .....	44
Gambar 3.7 – Interface Login .....	47
Gambar 3.8 – Interface Dashboard .....	47
Gambar 3.9 – Interface Input Data.....	48
Gambar 3.10 – Interface Input Data Sukses .....	48
Gambar 3.11 – Interface Input Data Gagal .....	49
Gambar 3.12 – Interface Output Data .....	49
Gambar 3.13 – Interface About.....	50
Gambar 3.14 – Interface History .....	50
Gambar 3.15 – Interface Archive .....	51
Gambar 4.1 – Struktur Tabel Database .....	52
Gambar 4.2 – Relasi Tabel Database .....	52
Gambar 4.3 – Tabel Admin .....	53
Gambar 4.4 – Tabel Document .....	53
Gambar 4.5 – Tabel Cluster .....	54
Gambar 4.6 – Tabel Critic .....	54
Gambar 4.7 – Halaman Login .....	56
Gambar 4.8 – Halaman Dashboard .....	58
Gambar 4.9 – Halaman Input Data Sukses .....	61
Gambar 4.10 – Halaman Input Data Gagal .....	62
Gambar 4.11 – Halaman Output Data .....	63
Gambar 4.12 – Halaman About .....	64

Gambar 4.13 – Halaman History ..... 65  
Gambar 4.14 – Halaman Archive ..... 66



## INTISARI

Di Dalam Lembaga Universitas AMIKOM Yogyakarta, setiap alumni diberikan kesempatan untuk memberi kritik kepada pihak lembaga melalui pihak tracer study alumni. Seiring berjalannya waktu, angka alumni dari Universitas AMIKOM Yogyakarta terus bertambah maka data kritik dari alumni pun semakin mengalami peningkatan. Hal itu menyebabkan proses pengevaluasian oleh pihak Lembaga Universitas AMIKOM Yogyakarta relatif semakin sulit.

*Text mining* merupakan metode yang digunakan untuk mencari informasi dari kumpulan koleksi teks dalam jumlah besar, serta mengidentifikasi pola-pola menarik serta keterkaitannya dalam data tekstual. Penggunaan tugas khusus dari *text mining* yaitu *text clustering* yang merupakan metode yang digunakan untuk mengelompokkan data kata ke dalam kelompok-kelompok tertentu yang memiliki kesamaan karakteristik. Berdasarkan pengertian tersebut, dibuat sebuah sistem untuk membantu proses pengelompokkan data berdasarkan kesamaan topik dengan menggunakan metode algoritma *K-Means* dan pembobotan kata yaitu TF-IDF.

Program klasterisasi data pada data kritik adalah solusi dari penanganan masalah tersebut. Melalui program klasterisasi data ini diharapkan mampu membantu Universitas AMIKOM Yogyakarta untuk melakukan proses pengevaluasian kinerja yang didapat melalui data hasil tracer study berdasarkan kritik atau komentar dari alumni mengenai Lembaga Universitas AMIKOM Yogyakarta.

**Kata Kunci :** *Text Mining, Text Clustering, Kritik, K-Means, TF-IDF, Universitas AMIKOM Yogyakarta*

## **ABSTRACT**

*In the Institute of AMIKOM Yogyakarta University, every alumni is allowed to give criticism of the institution through the tracer study alumni. As time goes by, the number of alumni from AMIKOM Yogyakarta University continues to grow, so the critical data from alumni is also increasing. This makes the evaluation process by the Yogyakarta AMIKOM University Institute relatively more difficult.*

*Text mining is a method used to find information from a large collection of text collections, as well as identify interesting patterns and their relationships in textual data. The use of a special task of text mining, namely text clustering is a method used to group word data into certain groups that have similar characteristics. Based on this understanding, a system was created to assist the process of grouping data based on similarity in topics using the K-Means algorithm and word weighting, namely TF-IDF.*

*The data clustering program on critical data is the solution to this problem. Through this data clustering program, it is hoped that it will be able to help AMIKOM University Yogyakarta to carry out the process of evaluating the performance obtained through tracer study data based on criticism or comments from alumni regarding the Yogyakarta AMIKOM University Institute.*

**Keywords :** *Text Mining, Text Clustering, Criticism, K-Means, TF-IDF, AMIKOM University Yogyakarta*