

**PENGELOMPOKAN TOPIK SKRIPSI MENGGUNAKAN
ALGORITMA AGGLOMERATIVE HIERARCHICAL
CLUSTERING DI PROGRAM STUDI SISTEM INFORMASI
UPN VETERAN JAWA TIMUR**

SKRIPSI

**Diajukan untuk memenuhi persyaratan
dalam memperoleh gelar Sarjana Komputer
Program Studi Sistem Informasi**



Disusun Oleh:

MUHAMMAD EVAN HOEDIANSYAH

19082010059

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR
SURABAYA**

2023

SKRIPSI

PENGELOMPOKAN TOPIK SKRIPSI MENGGUNAKAN ALGORITMA
AGGLOMERATIVE HIERARCHICAL CLUSTERING DI PROGRAM
STUDI SISTEM INFORMASI UPN VETERAN JAWA TIMUR

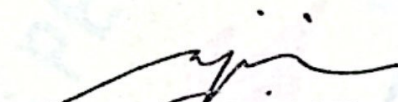
Disusun Oleh:

MUHAMMAD EVAN HOEDIANSYAH
19082010059

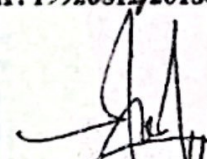
Telah dipertabankan di hadapan dan diterima oleh Tim Penguji Skripsi
Program Studi Sistem Informasi Fakultas Ilmu Komputer
Universitas Pembangunan Nasional "Veteran" Jawa Timur
Pada Tanggal 19 Mei 2023

Pembimbing:

1.


Amalia Anjani Arifryanti, S.Kom., M.Kom.
NIP. 19920812/201803 2 001

2.

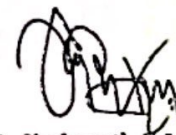

Reisa Perhatasari, S.T., M.Kom.
NIP. 19920514 202203 2 007

Tim Penguji

1.


Prisa Maspa Kusumantara, S.Kom., M.Cs.
NIP. 19821125 202121 1 003

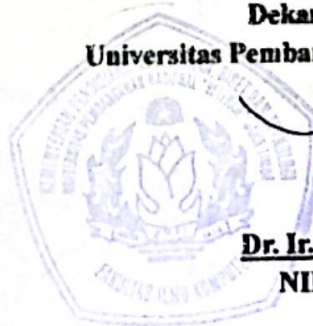
2.



Rizka Hadiwiyanti, S.Kom., M.Kom., MBA
NIP. 19860727 201803 2 001

3.


Abdul Rezha Efrat Najaf, S.Kom., M.Kom.
NIP. 19940929 202203 1 008

Mengetahui,
Dekan Fakultas Ilmu Komputer
Universitas Pembangunan Nasional "Veteran" Jawa Timur




Dr. Ir. Noyirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PENGESAHAN

**PENGELOMPOKAN TOPIK SKRIPSI MENGGUNAKAN ALGORITMA
AGGLOMERATIVE HIERARCHICAL CLUSTERING DI PROGRAM
STUDI SISTEM INFORMASI UPN VETERAN JAWA TIMUR**

Disusun Oleh:


MUHAMMAD EVAN HOEDIANSYAH
19082010059

**Telah disetujui mengikuti Ujian Negara Lisan Gelombang Mei
Periode 2023 pada Tanggal 19 Mei 2023**


Menyetujui,

Dosen Pembimbing 1

Dosen Pembimbing 2


Amalia Anjani Arifiyanti, S.Kom., M.Kom.

NIP. 19920812 201803 2 001


Reisa Permatasari, S.T., M.Kom.

NIP. 19920514 202203 2 007

Mengetahui,

**Ketua Program Studi Sistem Informasi
Fakultas Ilmu Komputer
Universitas Pembangunan Nasional "Veteran" Jawa Timur**


Agung Brastama Putra, S.Kom., M.Kom.

NIP. 19851124 202121 1 003



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
FAKULTAS ILMU KOMPUTER

KETERANGAN REVISI

Kami yang bertanda tangan di bawah ini menyatakan bahwa mahasiswa berikut:

Nama : Muhammad Evan Hoediansyah
NPM : 19082010059
Program Studi : Sistem Informasi

Telah mengerjakan revisi Ujian Negara Lisan Skripsi pada tanggal 19 Mei 2023 dengan judul:

**PENGELOMPOKAN TOPIK SKRIPSI MENGGUNAKAN ALGORITMA
AGGLOMERATIVE HIERARCHICAL CLUSTERING DI PROGRAM
STUDI SISTEM INFORMASI UPN VETERAN JAWA TIMUR**

Oleh karenanya mahasiswa tersebut diatas dinyatakan bebas revisi Ujian Negara Lisan Skripsi dan diijinkan untuk membukukan laporan SKRIPSI dengan judul tersebut.

Surabaya, 25 Mei 2023

Dosen penguji yang memeriksa revisi:

1. Prisa Marga Kusumantara, S.Kom., M.Cs.
NIP. 19821125 202121 1 003
2. Rizka Hadiwiyanti, S.Kom., M.Kom., MBA
NIP. 19860727 201803 2 001
3. Abdul Rezha Efrat Najaf, S.Kom., M.Kom.
NIP. 19940929 202203 1 008

{  }

{  }

{  }

Mengetahui,

Dosen Pembimbing 1

Dosen Pembimbing 2


Amalia Anjan Arifiyanti, S.Kom., M.Kom.
NIP. 19920812 201803 2 001


Reisa Permatasari, S.T., M.Kom.
NIP. 19920514 202203 2 007



**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
FAKULTAS ILMU KOMPUTER**

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Muhammad Evan Hoediansyah
NPM : 19082010059
Program Studi : Sistem Informasi

Menyatakan bahwa Judul Skripsi / Tugas Akhir Saya sebagai berikut:

**PENGELOMPOKAN TOPIK SKRIPSI MENGGUNAKAN ALGORITMA
AGGLOMERATIVE HIERARCHICAL CLUSTERING DI PROGRAM
STUDI SISTEM INFORMASI UPN VETERAN JAWA TIMUR**

Bukan merupakan plagiat dari Skripsi / Tugas Akhir / Penelitian orang lain dan juga bukan Produk / Software / Hasil Karya yang saya beli dari pihak lain.

Saya juga menyatakan bahwa Skripsi / Tugas Akhir ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional "Veteran" Jawa Timur maupun di instansi pendidikan lain.

Jika ternyata dikemudian hari pernyataan ini terbukti tidak benar, maka Saya bertanggung jawab penuh dan siap menerima segala kosekuensinya, termasuk pembatalan ijazah dikemudian hari.

Surabaya, 25 Mei 2023

Hormat Saya,



Muhammad Evan Hoediansyah
19082010059

**Judul : PENGELOMPOKAN TOPIK SKRIPSI
MENGUNAKAN ALGORITMA AGGLOMERATIVE
HIERARCHICAL CLUSTERING DI PROGRAM STUDI
SISTEM INFORMASI UPN VETERAN JAWA TIMUR**

Pembimbing 1 : Amalia Anjani Arifiyanti, S.Kom., M.Kom.

Pembimbing 2 : Reisa Permatasari, S.T., M.Kom.

ABSTRAK

Skripsi menjadi syarat yang wajib diselesaikan oleh seluruh mahasiswa untuk menyelesaikan studi dan mendapatkan gelar sarjana. Program studi Sistem Informasi UPN Veteran Jawa Timur memiliki dua bidang minat skripsi yaitu Manajemen Sistem Informasi dan Solusi Sistem Informasi. Setiap tahunnya, jumlah penelitian skripsi di program studi Sistem Informasi semakin bertambah dan judul penelitian skripsi yang semakin bervariasi. Penelitian ini bertujuan untuk mengelompokkan topik skripsi berdasarkan atribut Judul Skripsi dan Kata Kunci dengan menggunakan algoritma *Agglomerative Hierarchical Clustering*. Data yang didapatkan berasal dari koordinator skripsi dengan rentang tahun antara tahun 2019 hingga tahun 2023. Data tersebut kemudian dilakukan tahap *text preprocessing*, *term weighting* menggunakan TF-IDF, dan *dimensionality reduction* menggunakan UMAP. Hasil pengelompokan topik skripsi tersebut kemudian ditampilkan ke dalam sebuah sistem berbasis *website* dengan menggunakan bahasa pemrograman Python, HTML dan kerangka kerja Flask. Sehingga pengguna seperti mahasiswa maupun dosen dapat melihat data skripsi sekaligus hasil dari pengelompokan skripsi yang ada di program studi Sistem Informasi. Dari penelitian yang telah dilakukan, didapatkan hasil bahwa algoritma *agglomerative hierarchical clustering* mampu mengelompokkan data skripsi dengan cukup baik. Adapun hasil evaluasi *cluster* dengan menggunakan metode *silhouette coefficient* pada *cluster* judul skripsi didapatkan nilai *silhouette coefficient* sebesar 0.439 dengan jumlah *cluster* optimal sebanyak 8 *cluster* dan pada *cluster* kata kunci didapatkan nilai *silhouette coefficient* sebesar 0.487 dengan jumlah *cluster* optimal sebanyak 5 *cluster*.

Kata Kunci: *data mining*, TF-IDF, UMAP, *agglomerative hierarchical clustering*, *silhouette coefficient*, Flask

KATA PENGANTAR

Dengan mengucap puji syukur kepada Tuhan Yang Maha Esa yang telah memberikan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Pengelompokan Topik Skripsi Menggunakan Algoritma Agglomerative Hierarchical Clustering di Program Studi Sistem Informasi UPN Veteran Jawa Timur” sebagai salah satu syarat dalam menyelesaikan studi S1 di program studi Sistem Informasi Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Dalam menyelesaikan skripsi ini tentunya tak lepas dari dorongan dan bantuan dari berbagai pihak yang telah membantu penulis. Oleh karena itu, dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Ayah, Bunda, Mbak Thea, dan Mas Linggar yang senantiasa mendoakan, memberi dukungan, dorongan, serta motivasi penulis dalam menyelesaikan skripsi.
2. Ibu Amalia Anjani Arifiyanti, S.Kom., M.Kom. selaku dosen pembimbing 1 dan Ibu Reisa Permatasari, S.T., M.Kom. selaku dosen pembimbing 2 yang selalu memberikan bimbingan, arahan, dan ilmu selama proses pengerjaan skripsi ini.
3. Bapak Dhian Satria Yudha Kartika, S.Kom., M.Kom. selaku koordinator skripsi program studi Sistem Informasi atas kesediaannya dalam membantu penulis untuk mengumpulkan data penelitian.
4. Seluruh dosen Sistem Informasi yang telah memberikan waktu dan ilmunya selama masa perkuliahan.
5. Teman-teman Dusun Bjiirr, Real, Fajar, Fauzan, Iddo, Yusman, Mail, Fahri, Ali, Dhika, Nazar, Andree, dan Nabil yang telah menemani sejak awal perkuliahan hingga saat ini memberi dukungan dan motivasi kepada penulis dalam menyelesaikan skripsi.
6. Teman-teman PKL, Siti Lailatul Zahrotun Nisak dan Inas Syarifah Abidah yang telah memberi dukungan serta motivasi kepada penulis dalam menyelesaikan skripsi.

7. Yunita Eka Fitrianti, Tito Yosoya, Fenny Gita Cahyani, Raihan Fauzan Rizal, dan M. Amirul Ahzami F. selaku teman-teman penulis yang telah memberikan dukungan, semangat, serta motivasi kepada penulis selama masa pengerjaan skripsi.
8. Mahasiswa Sistem Informasi Universitas Pembangunan Nasional “Veteran” Jawa Timur angkatan 2019 yang tiada hentinya memberikan dukungan dalam penyelesaian pelaksanaan skripsi ini.
9. Kakak dan adik tingkat mahasiswa Sistem Informasi Universitas Pembangunan Nasional “Veteran” Jawa Timur yang selalu memberikan dukungan dan semangat dalam melaksanakan skripsi.

Semoga Tuhan Yang Maha Esa selalu memberikan rahmat dan ridha-Nya kepada seluruh pihak yang memberikan dukungan, bantuan, nasihat, serta bimbingan yang bermanfaat bagi penulis.

Penulis sadar bahwa masih banyak kekurangan dan jauh dari kata sempurna dalam penulisan Laporan Skripsi. Penulis berharap semoga Laporan Skripsi ini dapat menunjang perkembangan ilmu pengetahuan khususnya dalam bidang Sistem Informasi.

Surabaya, Mei 2023

Penulis

DAFTAR ISI

ABSTRAK	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	5
1.4 Tujuan.....	5
1.5 Sistematika Penulisan.....	6
BAB II TINJAUAN PUSTAKA	8
2.1 Dasar Teori	8
2.1.1 Skripsi	8
2.1.2 <i>Data Mining</i>	9
2.1.3 <i>Text Mining</i>	9
2.1.4 Python	10
2.1.5 <i>Text Preprocessing</i>	10
2.1.6 <i>Term Weighting</i>	12
2.1.7 <i>UMAP (Uniform Manifold Approximation and Projection)</i>	14
2.1.8 <i>Clustering</i>	15
2.1.9 <i>Agglomerative Hierarchical Clustering</i>	15
2.1.10 <i>Silhouette Coefficient</i>	18
2.1.11 Flask.....	18
2.1.12 <i>Black Box Testing</i>	19
2.2 Penelitian Terdahulu	19
BAB III METODOLOGI PENELITIAN.....	25
3.1 Studi Literatur.....	25
3.2 Analisis Kebutuhan	26
3.2.1 Kebutuhan Data.....	26
3.2.2 Kebutuhan <i>Software</i> dan <i>Hardware</i>	26
3.3 Pengumpulan Data	27

3.4	Pembuatan Model <i>Data Mining</i>	27
3.4.1	<i>Data Preparation</i>	28
3.4.2	<i>Exploratory Data Analysis</i>	28
3.4.3	<i>Text Preprocessing</i>	28
3.4.3.1	Case Folding	29
3.4.3.2	Cleansing	29
3.4.3.3	Stopword Removal	30
3.4.3.4	Stemming	30
3.4.4	<i>Term Weighting</i>	30
3.4.5	<i>Dimensionality Reduction</i>	31
3.4.6	<i>Clustering</i>	32
3.4.7	Visualisasi <i>Cluster</i>	33
3.5	Evaluasi <i>Cluster</i>	34
3.6	Pembuatan dan Pengujian Sistem	34
BAB IV HASIL DAN PEMBAHASAN		35
4.1	Pengumpulan Data	35
4.2	Proses <i>Clustering</i>	36
4.2.1	<i>Data Preparation</i>	36
4.2.2	<i>Exploratory Data Analysis</i>	38
4.2.3	<i>Text Preprocessing</i>	43
4.2.3.1	Case Folding	43
4.2.3.2	Cleansing	46
4.2.3.3	Stopword Removal	48
4.2.3.4	Stemming	51
4.2.4	<i>Term Weighting</i>	53
4.2.5	<i>Dimensionality Reduction</i>	54
4.2.6	<i>Clustering</i>	56
4.2.7	Visualisasi <i>Cluster</i>	58
4.3	Evaluasi <i>Cluster</i>	61
4.4	Pembuatan dan Pengujian Sistem	65
4.5	Pembahasan	75
BAB V PENUTUP		79
5.1	Kesimpulan	79
5.2	Saran	80
DAFTAR PUSTAKA		81
LAMPIRAN		84

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu yang Terkait.....	20
Tabel 4.1 Contoh Data Skripsi Mahasiswa	35
Tabel 4.2 Karakteristik Bidang Minat Skripsi.....	39
Tabel 4.3 Karakteristik Tahun Angkatan Mahasiswa.....	43
Tabel 4.4 Contoh Hasil Tahap <i>Case Folding</i> pada Atribut JudulSkripsi	44
Tabel 4.5 Contoh Hasil Tahap <i>Case Folding</i> pada Atribut KataKunci	45
Tabel 4.6 Contoh Hasil Tahap <i>Cleansing</i> pada Atribut JudulSkripsi.....	47
Tabel 4.7 Contoh Hasil Tahap <i>Cleansing</i> pada Atribut KataKunci.....	48
Tabel 4.8 Daftar <i>Stopword</i> Tambahan	49
Tabel 4.9 Contoh Hasil Tahap <i>Stopword Removal</i> pada Atribut JudulSkripsi	49
Tabel 4.10 Contoh Hasil Tahap <i>Stopword Removal</i> pada Atribut KataKunci.....	50
Tabel 4.11 Contoh Hasil Tahap <i>Stemming</i> pada Atribut JudulSkripsi	52
Tabel 4.12 Contoh Hasil Tahap <i>Stemming</i> pada Atribut KataKunci	52
Tabel 4.13 Hasil Pengujian Sistem menggunakan <i>Black-box Testing</i>	73

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Dendrogram <i>Agglomerative Hierarchical Clustering</i>	17
Gambar 3.1 Diagram Alir Penelitian.....	25
Gambar 3.2 Diagram Alir Pembuatan Model <i>Data Mining</i>	27
Gambar 3.3 Diagram Alir <i>Text Preprocessing</i>	29
Gambar 3.4 Diagram Alir <i>Term Weighting</i>	31
Gambar 3.5 Diagram Alir <i>Agglomerative Hierarchical Clustering</i>	32
Gambar 4.1 <i>Source Code</i> Deskripsi <i>Dataset</i>	36
Gambar 4.2 Deskripsi <i>Dataset</i>	37
Gambar 4.3 <i>Source Code</i> Penghapusan Data <i>Null</i>	37
Gambar 4.4 Deskripsi <i>Dataset</i> Setelah Penghapusan Nilai <i>Null</i>	38
Gambar 4.5 <i>Source Code</i> Diagram Perbandingan Bidang Minat Penelitian Skripsi	39
Gambar 4.6 Diagram Perbandingan Bidang Minat Penelitian Skripsi	40
Gambar 4.7 <i>Source Code</i> Diagram Perbandingan Tahun Penelitian Skripsi	40
Gambar 4.8 Diagram Perbandingan Tahun Penelitian Skripsi.....	40
Gambar 4.9 <i>Source Code</i> Perbandingan BidangMinat dan TahunPenelitian.....	41
Gambar 4.10 Grafik Batang Perbandingan BidangMinat dan TahunPenelitian ...	41
Gambar 4.11 <i>Source Code</i> Grafik Perbandingan Bidang Minat Skripsi Setiap Tahun.....	42
Gambar 4.12 Tren Perbandingan Bidang Minat Skripsi Setiap Tahun	42
Gambar 4.13 <i>Source Code Case Folding</i> pada Atribut JudulSkripsi.....	44
Gambar 4.14 <i>Source Code Case Folding</i> pada Atribut KataKunci.....	45
Gambar 4.15 <i>Source Code Cleansing</i> pada Atribut JudulSkripsi	46
Gambar 4.16 <i>Source Code Cleansing</i> pada Atribut KataKunci	47
Gambar 4.17 <i>Source Code Stopword Removal</i> pada Atribut JudulSkripsi	49
Gambar 4.18 <i>Source Code Stopword Removal</i> pada Atribut KataKunci	50
Gambar 4.19 <i>Source Code Stemming</i> pada Atribut JudulSkripsi.....	51
Gambar 4.20 <i>Source Code Stemming</i> pada Atribut KataKunci	52
Gambar 4.21 <i>Source Code Term Weighting</i> pada Atribut JudulSkripsi	53
Gambar 4.22 <i>Data Frame</i> Hasil <i>Term Weighting</i> pada Atribut JudulSkripsi.....	54
Gambar 4.23 <i>Source Code Term Weighting</i> pada Atribut KataKunci	54

Gambar 4.24 <i>Data Frame</i> Hasil <i>Term Weighting</i> pada Atribut KataKunci	54
Gambar 4.25 <i>Source Code Dimensionality Reduction</i> (UMAP) JudulSkripsi.....	55
Gambar 4.26 <i>Data Frame</i> setelah dilakukan UMAP pada Atribut JudulSkripsi ..	55
Gambar 4.27 <i>Source Code Dimensionality Reduction</i> (UMAP) KataKunci	56
Gambar 4.28 <i>Data Frame</i> setelah dilakukan UMAP pada Atribut KataKunci.....	56
Gambar 4.29 <i>Source Code Clustering</i> pada Atribut JudulSkripsi.....	57
Gambar 4.30 Dendrogram Hasil <i>Clustering</i> pada Atribut JudulSkripsi	57
Gambar 4.31 <i>Source Code Clustering</i> pada Atribut KataKunci	58
Gambar 4.32 Dendrogram Hasil <i>Clustering</i> pada Atribut KataKunci	58
Gambar 4.33 <i>Source Code Scatter Plot</i> pada Atribut JudulSkripsi.....	59
Gambar 4.34 <i>Scatter Plot</i> Atribut JudulSkripsi	59
Gambar 4.35 <i>Source Code Scatter Plot</i> pada Atribut KataKunci	59
Gambar 4.36 <i>Scatter Plot</i> Atribut KataKunci	60
Gambar 4.37 WordCloud pada Atribut JudulSkripsi	60
Gambar 4.38 WordCloud pada Atribut KataKunci	61
Gambar 4.39 Perhitungan <i>Silhouette Score</i> pada <i>Cluster</i> Judul Skripsi	62
Gambar 4.40 Perhitungan <i>Silhouette Score</i> pada <i>Cluster</i> Kata Kunci.....	62
Gambar 4.41 <i>Source Code</i> Evaluasi <i>Cluster</i> menggunakan <i>Scatter Plot</i>	63
Gambar 4.42 <i>Silhouette Plot</i> dan <i>Scatter Plot</i> pada Atribut JudulSkripsi.....	64
Gambar 4.43 <i>Silhouette Plot</i> dan <i>Scatter Plot</i> pada Atribut KataKunci	64
Gambar 4.44 Potongan <i>Source Code Template Website</i>	66
Gambar 4.45 Potongan <i>Source Code</i> Halaman Utama	67
Gambar 4.46 Tampilan Halaman Utama.....	67
Gambar 4.47 Tampilan Halaman Cluster Judul Skripsi.....	68
Gambar 4.48 Tampilan Halaman Wordcloud Judul Skripsi.....	69
Gambar 4.49 Tampilan Halaman Cluster Kata Kunci.....	70
Gambar 4.50 Tampilan Halaman Wordcloud Kata Kunci	71
Gambar 4.51 Potongan <i>Source Code</i> Halaman Cluster dan Wordcloud.....	71
Gambar 4.52 Potongan <i>Source Code</i> Halaman Grafik	72
Gambar 4.53 Tampilan Halaman Grafik.....	73

DAFTAR LAMPIRAN

Lampiran 1 Data Skripsi Mahasiswa	84
Lampiran 2 Hasil <i>Clustering</i> pada Atribut JudulSkripsi	100
Lampiran 3 Hasil <i>Clustering</i> pada Atribut KataKunci.....	109
Lampiran 4 Surat Permohonan Pengumpulan Data	114