

KLASTERISASI TUGAS AKHIR MAHASISWA FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA MENGGUNAKAN METODE K-MEANS

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh:

Yolanda Marevin
NIM: 09021282025082

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2024

LEMBAR PENGESAHAN SKRIPSI


KLASTERISASI TUGAS AKHIR MAHASISWA FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA MENGGUNAKAN METODE K-MEANS

Oleh:


Yolanda Marevin
NIM: 09021282025082

Indralaya, 14 Mei 2024

Pembimbing I


Rizki Kurniati, M. T.
NIP. 199107122019032016

Pembimbing II,


Anggina Primanita, M.IT, Ph.D.
NIP. 198908062015042002

Mengetahui,
Ketua Jurusan Teknik Informatika

Dr. M. Faehurrozi, S.Si, M.T
NIP. 198005222008121002



TANDA LULUS UJIAN KOMPREHENSIF

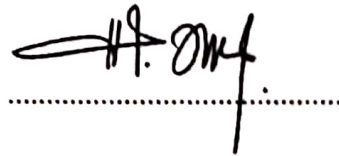
Pada hari Jumat tanggal 03 Mei 2024 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Yolanda Marevin
NIM : 09021282025082
Judul : Klasterisasi Tugas Akhir Mahasiswa Fakultas Ilmu Komputer Universitas Sriwijaya Menggunakan Metode K-Means

Dan dinyatakan LULUS

1. Ketua Penguji

Annisa Darmawahyuni, M.Kom.
NIP. 199006302023212044



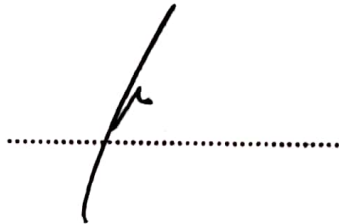
2. Penguji

Kanda Januar Miraswan, M.T.
NIP. 199001092019031012



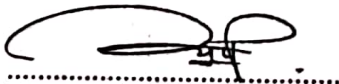
3. Pembimbing I

Rizki Kurniati, M.T.
NIP. 199107122019032016



4. Pembimbing II

Anggina Primanita, M.IT, Ph.D.
NIP. 198908062015042002



Mengetahui,
Ketua Jurusan Teknik Informatika



Dr. M. Fachrurrozi, S.Si, M.T
NIP. 198005222008121002

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Yolanda Marevin
NIM : 09021282025082
Program Studi : Teknik Informatika
Judul Skripsi : Klasterisasi Tugas Akhir Mahasiswa Fakultas Ilmu
Komputer Universitas Sriwijaya Menggunakan Metode K-
Means

Hasil Pengecekan *Software iThenticate/Turnitin*: 19%

Menyatakan bahwa laporan penelitian saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan penelitian ini, maka saya bersedia menerima sanksi akademik Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, 25 Maret 2024



Yolanda Marevin
NIM. 09021282025082

MOTTO DAN PERSEMBAHAN

“Selesaikan apa yang telah kamu mulai, perjuangkan apa yang kamu inginkan”

– Yolanda Marevin

Saya persembahkan karya tulis ini kepada:

- Orangtua,
- Adik-Adik
- Sahabat
- Teman Seperjuangan
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

Abstract

The final project is an important stage in a student's study journey, which often requires exploration and in-depth understanding in a discipline. This research aims to cluster final project topics based on the attributes of student final project titles, especially at the Faculty of Computer Science, Sriwijaya University, using the K-means clustering method. The data analyzed comes from the faculty library in the time span of 2016 to 2023. The data was processed and weighted using TF-IDF method, then reduced using UMAP technique. Topic analysis was also conducted by utilizing the wordcloud generated from the clustering. The clustering results show that the optimal cluster is located in the 7th cluster, with a silhouette score of 0.3515.

Keywords: *Data Mining, Clustering, K-Means Clustering, TF-IDF, Final Project, UMAP*

Abstrak

Tugas akhir merupakan tahap penting dalam perjalanan studi seorang mahasiswa, yang seringkali memerlukan eksplorasi dan pemahaman mendalam dalam suatu disiplin ilmu. Penelitian ini bertujuan untuk mengelompokkan topik tugas akhir berdasarkan atribut judul tugas akhir mahasiswa, khususnya di Fakultas Ilmu Komputer Universitas Sriwijaya, dengan menggunakan metode K-means clustering. Data yang dianalisis berasal dari perpustakaan fakultas dalam rentang waktu 2016 hingga 2023. Data tersebut diproses dan dibobotkan menggunakan metode TF-IDF, kemudian direduksi menggunakan teknik UMAP. Analisis topik juga dilakukan dengan memanfaatkan wordcloud yang dihasilkan dari klasterisasi. Hasil klasterisasi menunjukkan bahwa klaster optimal terletak pada klaster ke-7, dengan nilai silhouette score sebesar 0.3515.

Kata Kunci: Data Mining, Klasterisasi, K-Means Clustering, TF-IDF, Tugas Akhir, UMAP

KATA PENGANTAR

Puji Syukur penulis ucapkan kepada Allah SWT, karena berkat rahmat dan karunianya, penulis dapat menyelesaikan tugas akhir dengan judul “Klasterisasi Tugas Akhir Mahasiswa Fakultas Ilmu Komputer Universitas Sriwijaya Menggunakan Metode K-Means” dengan baik. Penelitian ini disusun sebagai syarat dalam menyelesaikan studi Strata-I program studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Sriwijaya. Pada kesempatan ini penulis mengucapkan banyak terimakasih khususnya kepada:

1. Allah SWT atas berkah, rahmat, dan karunianya.
2. Kedua Orang Tua penulis, papa dan mama atas segala dukungannya.
3. Adik adik penulis yang selalu menghibur dan mendukung penulis,
4. Bapak Dr. M. Fachrurrozi, S.Si, M.T selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Ibu Rizki Kurniati, M.T. selaku Dosen Pembimbing I dan Dosen Pembimbing Akademik yang telah membantu memberikan arahan dan bimbingan serta dukungannya selama masa perkuliahan.
6. Ibu Anggina Primanita, M.IT, P.h.D. selaku Dosen Pembimbing II yang telah membantu memberikan arahan dan bimbingan.
7. Seluruh Dosen Jurusan Teknik Informatika dan Dosen Fakultas Ilmu Komputer yang telah membagikan ilmu dan pengetahuan selama masa perkuliahan.

8. Seluruh Staf Administrasi dan Pegawai yang telah membantu dalam urusan administrasi.
9. Sahabat- sahabat penulis, khususnya Neo dan chan.
10. Seluruh teman-teman penulis di Universitas Sriwijaya.

Penulis menyadari bahwa masih terdapat kekurangan terhadap penyusunan tugas akhir ini. Oleh karena itu, penulis sangat berlapang dada apabila terdapat masukan atau saran yang membantu dalam menyempurnakan tugas akhir ini. Agar bisa bermanfaat bagi kita semua.

Palembang, 02 Mei 2023

Yolanda Marevin
NIM. 09021282025082

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	i
Motto:.....	iv
Abstract	v
Abstrak	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	I-1
1.1. Pendahuluan	I-1
1.2. Latar Belakang Masalah	I-1
1.3. Rumusan Masalah	I-4
1.4. Tujuan.....	I-4
1.5. Manfaat.....	I-5
1.6. Batasan Masalah.....	I-5
1.7. Sistematika Penulisan.....	I-5
1.8. Kesimpulan.....	I-7
BAB II KAJIAN LITERATUR.....	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 Data mining.....	II-1
2.2.2 Unsupervised Learning	II-2
2.2.3 Text Mining.....	II-2
2.2.4 Pre-processing teks.....	II-3
2.2.5 TF-IDF	II-4
2.2.6 Uniform Manifold Approximation and Projection (UMAP)	II-7
2.2.7 Klasterisasi	II-9
2.2.8 K-Means	II-9

2.2.9	Agglomerative Hierarchical Clustering	II-11
2.2.10	Silhouette score	II-12
2.3	Penelitian Relevan	II-12
2.4	Kesimpulan.....	II-14
BAB III METODOLOGI PENELITIAN.....		III-1
3.1	Pendahuluan	III-1
3.2	Pengumpulan Data	III-1
3.2.1	Jenis Data.....	III-1
3.2.2	Sumber data	III-3
3.2.3	Metode Pengumpulan Data.....	III-3
3.3	Tahapan Penelitian.....	III-3
3.3.1	Kerangka Kerja.....	III-3
3.3.1.1	Data Preparation	III-6
3.3.1.2	Text Preprocessing.....	III-6
3.3.1.2.1	Case Folding	III-6
3.3.1.2.2	Cleansing.....	III-7
3.3.1.2.3	Stopword Removal.....	III-8
3.3.1.3	Term Weighting.....	III-9
3.3.1.4	Reduksi Dimensi.....	III-10
3.3.1.5	Klasterisasi.....	III-10
3.3.1.6	Visualisasi cluster	III-11
3.3.2	Kriteria pengujian	III-11
3.3.3	Format Data Pengujian	III-11
3.3.4	Alat Bantu Penelitian	III-12
3.3.5	Pengujian Penelitian	III-13
3.3.6	Analisis Hasil Pengujian dan Membuat Kesimpulan	III-13
3.4	Metode Pengembangan Perangkat Lunak	III-13
3.4.1	Fase Insepsi.....	III-13
3.4.2	Fase Elaborasi.....	III-14
3.4.3	Fase Konstruksi.....	III-14
3.4.4	Fase Transisi	III-15

3.5	Manajemen Proyek Penelitian.....	III-15
BAB IV PENGEMBANGAN PERANGKAT LUNAK.....		IV-1
4.2.1	Pemodelan Bisnis.....	IV-1
4.2.2	Kebutuhan Sistem.....	IV-2
4.2.3	Analisis Desain.....	IV-4
4.2.3.1	Analisis Perangkat Lunak.....	IV-4
4.2.3.2	Analisis Data.....	IV-5
4.2.3.3	Analisis Metode K-Means.....	IV-6
4.2.3.4	Desain Perangkat Lunak.....	IV-16
4.3.1	Pemodelan Bisnis.....	IV-23
4.3.1.1	Perancangan Data.....	IV-24
4.3.1.2	Perancangan Antarmuka.....	IV-24
4.3.2	Kebutuhan Sistem.....	IV-25
4.3.3	Diagram.....	IV-26
4.3.3.1	Diagram Aktivitas.....	IV-26
4.3.3.2	Diagram Alur.....	IV-29
4.4.1	Kebutuhan Sistem.....	IV-32
4.4.2	Kelas Diagram.....	IV-33
4.4.3	Implementasi.....	IV-34
4.4.3.1	Implementasi Kelas.....	IV-34
4.4.3.2	Implementasi Antarmuka.....	IV-35
4.5.1	Pemodelan Bisnis.....	IV-37
4.5.2	Implementasi.....	IV-37
BAB V HASIL DAN ANALISIS PEMBAHASAN.....		V-1
5.2.1	Konfigurasi Percobaan.....	V-1
5.2.2	Data Hasil Konfigurasi.....	V-2
BAB VI KESIMPULAN DAN SARAN.....		VI-1
DAFTAR PUSTAKA.....		xv

DAFTAR TABEL

Tabel III - 1. Data Judul Skripsi Mahasiswa Fakultas Ilmu Komputer	III-2
Tabel III - 2. Data judul Tugas Akhir sebelum.....	III-7
Tabel III - 3. Tabel Judul Tugas Akhir sebelum.....	III-8
Tabel III - 4. Tabel Judul Tugas Akhir sebelum.....	III-8
Tabel III - 5. Rancangan Tabel Nilai Evaluasi	III-12
Tabel IV - 1. Kebutuhan Fungsional	IV-3
Tabel IV - 2. Kebutuhan Non Fungsional	IV-3
Tabel IV - 3. Data Sampel.....	IV-7
Tabel IV - 4. Hasil Preprocessing Judul Tugas Akhir	IV-9
Tabel IV - 5. Reduksi Umap.....	IV-11
Tabel IV - 6. Centroid klaster	IV-12
Tabel IV - 7. Hasil Perhitungan Jarak	IV-13
Tabel IV - 8. Hasil Pengelompokan klaster.....	IV-14
Tabel IV - 9. Definisi Aktor.....	IV-17
Tabel IV - 10. Definisi Use Case.....	IV-17
Tabel IV - 11. Skenario Upload Dataset.....	IV-18
Tabel IV - 12. Skenario Menampilkan Hasil Klasterisasi	IV-20
Tabel IV - 13. Skenario Melakukan filtering klasterisasi.....	IV-21
Tabel IV - 14. Implementasi Kelas.....	IV-34
Tabel IV - 15. Pengujian Use Case Upload Data	IV-38
Tabel IV - 16. Pengujian Use Case Hasil Klasterisasi	IV-39
Tabel IV - 17. Pengujian Use Case Filtering.....	IV-40
Tabel V - 1. Tabel Nilai Evaluasi.....	V-3
Tabel V - 2. Jumlah kata yang sering muncul pada klaster 1	V-5
Tabel V - 3. Jumlah kata yang sering muncul pada klaster 2	V-8
Tabel V - 4. Jumlah kata yang sering muncul pada klaster 3	V-10
Tabel V - 5. Jumlah kata yang sering muncul pada klaster 4	V-13
Tabel V - 6. Jumlah kata yang sering muncul pada klaster 5	V-15
Tabel V - 7. Jumlah kata yang sering muncul pada klaster 6	V-17

Tabel V - 8. Jumlah kata yang sering muncul pada klaster 7V-20

DAFTAR GAMBAR

Gambar III- 1. Tahapan Penelitian.....	III-4
Gambar III- 2. Kerangka Kerja Perangkat Lunak.....	III-5
Gambar III- 3. Tahapan Text Preprocessing.....	III-6
Gambar III- 4. Tahapan Term Weighting	III-9
Gambar IV - 1. Pie Chart Persebaran Jurusan	IV-6
Gambar IV - 2. Use Case Diagram.....	IV-16
Gambar IV - 3. Rancangan Antarmuka Halaman Awal.....	IV-24
Gambar IV - 4. Rancangan Antarmuka Halaman Hasil Klasterisasi.....	IV-25
Gambar IV - 5. Rancangan Antarmuka Hasil Filtering.....	IV-25
Gambar IV - 6. Diagram Aktivitas Upload Dataset	IV-27
Gambar IV - 7. Diagram Aktivitas Menampilkan Hasil Klasterisasi.....	IV-28
Gambar IV - 8. Diagram Aktivitas Filtering Hasil Klasterisasi	IV-29
Gambar IV - 9. Sequence Diagram Upload Dataset.....	IV-30
Gambar IV - 10. Sequence Diagram untuk menampilkan	IV-31
Gambar IV - 11. Sequence Diagram Menampilkan	IV-32
Gambar IV - 12. Diagram Kelas.....	IV-33
Gambar IV - 13. Antarmuka Awal Sistem.....	IV-36
Gambar IV - 14. Antarmuka Hasil klasterisasi.....	IV-36
Gambar IV - 15. Antarmuka Hasil Filtering.....	IV-37
Gambar V - 1. Scatter plot klasterisasi	V-4
Gambar V - 2. Wordcloud pada klaster 1	V-5
Gambar V - 3. Pie Chart Pada Klaster 1	V-6
Gambar V - 4. Wordcloud pada klaster 2	V-7
Gambar V - 5. Pie Chart Pada Klaster 2	V-9
Gambar V - 6. Wordcloud pada klaster 3	V-10
Gambar V - 7. Pie Chart Pada Klaster 3	V-11
Gambar V - 8. Wordcloud pada klaster 4	V-13
Gambar V - 9. Pie Chart Pada Klaster 4	V-14
Gambar V - 10. Wordcloud pada klaster 5	V-15

Gambar V - 11. Pie Chart Pada Klaster 5	V-16
Gambar V - 12. Wordcloud pada klaster 6	V-17
Gambar V - 13. Pie Chart Pada Klaster 6	V-18
Gambar V - 14. Wordcloud pada klaster 7	V-19
Gambar V - 15. Pie Chart Pada Klaster 7	V-21

BAB I

PENDAHULUAN

1.1. Pendahuluan

Bab ini akan membahas secara rinci mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan penelitian.

1.2. Latar Belakang Masalah

Tugas akhir adalah karya tulis yang disusun oleh mahasiswa yang telah menyelesaikan kurang lebih 130 sks dengan dibimbing oleh dosen pembimbing guna mendapatkan gelar pendidikan (Astuti & Satrio, 2019). Signifikansi tugas akhir dalam pendidikan tinggi tak terelakkan seiring pertumbuhan jumlah mahasiswa, yang membawa peningkatan dokumen tugas akhir setiap tahun.

Kenaikan jumlah dokumen tersebut mencakup beragam topik, membuka peluang sekaligus menimbulkan tantangan baru. Mahasiswa sering mengalami kesulitan dalam mencari referensi skripsi di perpustakaan karena jumlahnya yang besar. Untuk itu, perlu ditemukan metode praktis untuk mengetahui topik-topik skripsi yang populer tanpa harus melihat satu per satu.

Kebutuhan mahasiswa akan informasi skripsi menjadi dorongan akan perlunya pengelompokan atau pengkategorian skripsi untuk mempermudah pencarian informasi (Irianto et al., 2022). Sehingga diperlukan metode klusterisasi untuk membantu mengelompokkan tugas akhir mahasiswa untuk memudahkan

proses seleksi penelitian, pemberian bahan rujukan, dan penentuan dosen pembimbing.

Proses klasterisasi judul tugas akhir mahasiswa tiap prodi bisa dilakukan secara manual, namun hal tersebut kurang optimal karena jumlah data yang besar akan membutuhkan waktu yang lama, tenaga yang besar dan memperbesar kemungkinan kesalahan klasterisasi. Dengan demikian, dibutuhkan sistem yang dapat melakukan pengelompokan judul tugas akhir mahasiswa tiap prodi secara otomatis. Sistem tersebut harus dapat mengolah data yang besar dan memberikan informasi dengan tingkat akurasi setinggi mungkin. Ekstraksi informasi dari data teks yang besar ini dapat dilakukan dengan teknik teks mining. Teks mining sendiri termasuk ke dalam variasi data mining (Irianto et al., 2022).

Dalam konteks penggunaan data mining untuk klasterisasi, pemilihan algoritma memiliki peran krusial dalam mengidentifikasi pola dan aturan mekanis dalam data. Terdapat banyak algoritma klasterisasi seperti *K-Means*, *K-Medoids*, *Hierarchical clustering*, *DBSCAN* dan *Fuzzy C-Means*. Salah satu algoritma klasterisasi yang populer dan umum digunakan adalah *K-Means*, sebuah metode tanpa supervisi yang berfokus pada partisi data (Gamadarenda & Waspada, 2020). Algoritma *K-Means* menjadi pilihan yang baik dalam klasterisasi judul tugas akhir mahasiswa karena kemampuannya yang efisien dalam membentuk kelompok berdasarkan kemiripan karakteristik, yang dalam hal ini adalah karakteristik dari judul tugas akhir. Dengan menerapkan *K-Means*, klasterisasi dapat dilakukan dengan lebih cepat, efisien, dan memberikan hasil yang konsisten, mendukung

pengambilan keputusan yang lebih baik dalam memahami pola dari judul tugas akhir mahasiswa.

K-Means juga memiliki keunggulan dalam hal interpretabilitas. Pusat kluster yang dihasilkan oleh algoritma ini dapat dianggap sebagai representasi dari karakteristik rata-rata dari judul tugas akhir dalam kelompok tersebut (Astuti & SatrioWP, 2019). Kemampuan ini memberikan kenyamanan dalam mengidentifikasi tema atau topik utama yang muncul dalam kluster tertentu. Dengan demikian, *K-Means* tidak hanya memberikan efisiensi dalam proses klasterisasi, tetapi juga meningkatkan interpretasi hasil, memfasilitasi pemahaman yang lebih dalam terhadap struktur dan konten dari judul tugas akhir mahasiswa.

Berdasarkan penelitian terkait dapat disimpulkan bahwa penerapan Algoritma *Agglomerative Hierarchical Clustering* berhasil mengelompokkan data menjadi 8 cluster. Penelitian ini juga berhasil mendapatkan nilai *Silhouette score* yang cukup tinggi yaitu 0.439, yang mana menunjukkan bahwa implementasi algoritma tersebut sukses mengelompokkan topik skripsi secara efisien (Hoediansyah, 2023).

Berdasarkan latar belakang yang telah diuraikan, penelitian ini bertujuan untuk melihat hasil dari klasterisasi judul tugas akhir mahasiswa universitas sriwijaya menggunakan algoritma *K-Means clustering*. Penelitian ini akan menghapus metode *stemming* pada bagian *text preprocessing*, hal ini bertujuan untuk memastikan bahwa kata-kata berimbuhan tetap utuh tanpa diubah menjadi kata dasar sehingga dapat diperoleh representasi kata yang lebih tepat dan efisien dalam proses *processing* data. Penelitian ini juga akan menggunakan *dataset* yang

lebih besar dalam pengujian untuk menguji kinerja algoritma pada skala data yang relative besar. Penelitian ini hanya akan berfokus pada klasterisasi judul tugas akhir saja. Selanjutnya penelitian akan melihat hasil *silhouette score* sebagai acuan untuk membuktikan efektivitasnya dalam mengelompokkan judul tugas akhir.

1.3. Rumusan Masalah

Berdasarkan uraian latar belakang di atas maka dapat dihasilkan beberapa rumusan permasalahan pada penelitian ini antara lain.

1. Bagaimana implementasi klasterisasi tugas akhir mahasiswa Universitas Sriwijaya menggunakan metode *K-Means clustering*?
2. Bagaimana kinerja dari algoritma *K-Means* dalam mengelompokkan data judul tugas akhir mahasiswa dan bagaimana evaluasi efektivitasnya?

1.4. Tujuan

Tujuan penelitian ini adalah :

1. Mengetahui cara implementasi klasterisasi tugas akhir mahasiswa Universitas Sriwijaya dengan menggunakan metode *K-Means clustering*.
2. Menentukan jumlah klaster yang optimal untuk mengelompokkan data Judul tugas akhir mahasiswa universitas sriwijaya.
3. Mengetahui hasil kinerja algoritma *K-Means* dalam mengelompokkan data judul tugas akhir mahasiswa.

1.5. Manfaat

Manfaat dari penelitian ini adalah :

1. Memahami dan mempelajari hasil kinerja dari metode *K-means* pada klasterisasi judul tugas akhir mahasiswa
2. Membantu peneliti dan praktisi dalam memilih metode yang optimal untuk mengelompokkan data judul tugas akhir dengan memahami tingkat kinerja algoritma *K-Means*.
3. Hasil penelitian dapat digunakan sebagai dasar untuk penelitian lanjutan dalam bidang pengelompokan tugas akhir dan klasterisasi data.

1.6. Batasan Masalah

Adapun Batasan masalah dalam penelitian adalah :

1. Batasan pada jenis data mencakup data judul tugas akhir mahasiswa di Universitas Sriwijaya.
2. Fokus analisis pada metode *K-Means* dengan implementasi terbatas pada atribut tertentu yaitu atribut judul tugas akhir.

1.7. Sistematika Penulisan

Sistematika penulisan tugas akhir mengikuti standar penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya yaitu sebagai berikut:

BAB I. PENDAHULUAN

Bab ini akan membahas secara rinci mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan penelitian

BAB II. KAJIAN LITERATUR

Pada bab ini akan dijelaskan teori yang melandasi penelitian. Bab ini berisi penjelasan mengenai Data Mining, *Unsupervised Learning*, *text mining*, *k-means*, klusterisasi, *Preprocessing text*, *silhouette score*, dan beberapa penelitian yang relevan, serta kesimpulan.

BAB III. METODOLOGI PENELITIAN

Bab ini membahas mengenai tahapan, metodologi, manajemen proyek dan hal lain yang diperlukan dalam penelitian. Tahap penelitian menjadi acuan pada fase pengembangan yang diharapkan memberi solusi rumusan masalah untuk mencapai tujuan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini akan membahas mengenai proses pengembangan perangkat lunak untuk akurasi hasil klusterisasi pada dataset judul tugas akhir mahasiswa Fakultas Ilmu Komputer Universitas Sriwijaya menggunakan metode *K-Means Clustering* yang akan digunakan pada penelitian ini. Penelitian ini menggunakan metodologi pengembangan perangkat lunak Rational Unified Process (RUP) yang berorientasi objek. Pendekatan RUP meliputi empat fase yaitu inspirasi, elaborasi, pembangunan, dan transisi.

BAB V. HASIL DAN ANALISIS PEMBAHASAN

Pada bab ini akan membahas tentang hasil klusterisasi data Judul tugas akhir mahasiswa fakultas ilmu komputer universitas sriwijaya menggunakan metode *k-means clustering*. Bab ini akan menampilkan hasil yang telah diperoleh dari proses klusterisasi dan hasil pengujian *clustering* yang diterapkan serta mengukur kinerja

dari metode K-Means untuk menentukan jumlah kluster yang optimal pada data yang diberikan menggunakan pengujian *silhouette score*. Selanjutnya akan dilakukan analisis untuk menarik kesimpulan dari penelitian yang telah dilakukan.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini membahas tentang kesimpulan dan saran berdasarkan penelitian judul tugas akhir mahasiswa fakultas ilmu komputer universitas sriwijaya yang telah dilakukan. Hasil dari penelitian ini diharapkan dapat menjadi bahan referensi atau acuan untuk penelitian selanjutnya, serta memperluas pemahaman kita tentang topik yang diteliti.

1.8. Kesimpulan

Berdasarkan latar belakang penelitian mengenai pengelompokan tugas akhir mahasiswa di Universitas Sriwijaya menggunakan metode klusterisasi terutama algoritma *K-Means*, ditemukan bahwa implementasi algoritma tersebut menjadi solusi untuk mengatasi tantangan dalam manajemen informasi akademis. Penelitian ini didorong oleh peningkatan jumlah tugas akhir yang memerlukan kategorisasi, sehingga diperlukan suatu metode efektif untuk mengelompokkan judul-judul tersebut berdasarkan karakteristik tertentu. Diharapkan, hasil analisis dari penelitian ini dapat memberikan manfaat dalam meningkatkan efisiensi pencarian referensi, seleksi penelitian, dan pemilihan dosen pembimbing. Sebagai langkah awal, penelitian ini diarahkan untuk memberikan kontribusi pada optimalisasi manajemen tugas akhir di lingkungan Universitas Sriwijaya.

DAFTAR PUSTAKA

- Abijono, H., Santoso, P., & Anggreini, N. L. (2021). Algoritma Supervised Learning Dan Unsupervised Learning Dalam Pengolahan Data. *Jurnal Teknologi Terapan: G-Tech*, 4(2), 315–318. <https://doi.org/10.33379/gtech.v4i2.635>.
- Allaoui, M., Kherfi, M.L., Cheriet, A. (2020). Considerably Improving Clustering Algorithms Using UMAP Dimensionality Reduction Technique: A Comparative Study. In: El Moataz, A., Mammass, D., Mansouri, A., Nouboud, F. (eds) *Image and Signal Processing. ICISP 2020. Lecture Notes in Computer Science* (), vol 12119. Springer, Cham. https://doi.org/10.1007/978-3-030-51935-3_34
- Astuti, R. W., & SatrioWP, B. (2019). *Prosiding Konferensi Ilmiah Mahasiswa Unissula (Kimu) 2 Rancang Bangun Sistem Klasterisasi Dokumen Menggunakan Metode K-Means Untuk Identifikasi Topik Dokumen Tugas*
- Fuadi, W., Razi, A., & Fariadi, D. (2022). Automasi Penentuan Tren Topik Skripsi Menggunakan Algoritma K-Means Clustering. *Serambi Engineering*, VII(2).
- Gamadarena, I. W., & Waspada, I. (2020). Implementasi Data Mining untuk Deteksi Penyakit Ginjal Kronis (PGK) menggunakan K-Nearest Neighbor (KNN) dengan Backward Elimination. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(2), 417. <https://doi.org/10.25126/jtiik.2020721896>
- Hananto, A. L., Assiroj, P., Priyatna, B., Nurhayati, Fauzi, A., Yuniar Rahman, A., & Shofiah Hilabi, S. (2021b). Analysis of Drug Data Mining with Clustering Technique Using K-Means Algorithm. *Journal of Physics: Conference Series*, 1908(1). <https://doi.org/10.1088/1742-6596/1908/1/012024>
- Hermawan, A., Jowensen, I., Junaedi, J., & Edy. (2023). Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine. *JST (Jurnal Sains Dan Teknologi)*, 12(1), 129–137. <https://doi.org/10.23887/jstundiksha.v12i1.52358>
- Hoediansyah, M. E. (2023). *PENGELOMPOKAN TOPIK SKRIPSI MENGGUNAKAN ALGORITMA AGGLOMERATIVE HIERARCHICAL CLUSTERING DI PROGRAM STUDI SISTEM INFORMASI*. <http://repository.upnjatim.ac.id/id/eprint/14535>
- Hung, P. D., Lien, N. T. T., & Ngoc, N. D. (2019). Customer Segmentation Using Hierarchical Agglomerative Clustering. *ICISS 2019: Proceedings of the 2019 2nd International Conference on Information Science and Systems*, 33–37.

- Irianto, M. R., Maududie, A., & Arifin, F. N. (2022). Implementation of K-Means Clustering Method for Trend Analysis of Thesis Topics (Case Study: Faculty of Computer Science, University of Jember). *BERKALA SAINSTEK*, 10(4), 210. <https://doi.org/10.19184/bst.v10i4.29524>
- Iswara, R. P., Informatika, T., Sains dan Teknologi, F., Syarif Hidayatullah Jakarta, U., & Gotong Royong Jakarta, S. (2019). PENGEMBANGAN ALGORITMA UNSUPERVISED LEARNING TECHNIQUE PADA BIG DATA ANALYSIS DI MEDIA SOSIAL SEBAGAI MEDIA PROMOSI ONLINE BAGI MASYARAKAT. *JURNAL TEKNIK INFORMATIKA*, 12(1).
- Jabri, S., Dahbi, A., Gadi, T., & Bassir, A. (2018). Ranking of Text Documents using TF-IDF Weighting and Association Rules Mining. 2018 4th International Conference on Optimization and Applications (ICOA), 1–6. <https://doi.org/10.1109/ICOA.2018.8370597>
- Kambey, G. E. I., Sengkey, R., & Jacobus, A. (2020). Penerapan Clustering pada Aplikasi Pendeteksi Kemiripan Dokumen Teks Bahasa Indonesia. *Jurnal Teknik Informatika*, 15(2), 75–82.
- McInnes, L., Healy, J., & Melville, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. <http://arxiv.org/abs/1802.03426>.
- Nurdini, S., Nurcahyo, G. W., & Santony, J. (2019). Analisis Perkiraan Jumlah Produksi Tahu Menggunakan Metode Fuzzy Sugeno. *Jurnal Sistim Informasi Dan Teknologi*, 1(3), 19–24. <https://doi.org/10.35134/jsisfotek.v1i3.5>
- Pealat, C., Bouleux, G., & Cheutet, V. (2021). Improved Time-Series Clustering with UMAP dimension reduction method. 2020 25th International Conference on Pattern Recognition (ICPR), 5658-5665. doi: 10.1109/ICPR48806.2021.9412261.
- Permana, R. A., Hapsari, D. P., & Muhima, R. R. (2022). SNESTIK Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika PENERAPAN METODE CLUSTERING K-MEDOIDES UNTUK PENGELOMPOKAN ABSTRAK SKRIPSI BERBASIS WEB. *SNESTIK*, 277–284. <https://doi.org/10.31284/p.snestik.2022.2812>
- Putra, E. D., Muhammad,), Rifqo, H., Dwita Deslianti,), & Krismiyani,); (2022). Analysis of The Theme Clustering Algorithm Using K-Means Method Analisis Algoritma Clustering Tema Skripsi Menggunakan Metode K-Means. *JURNAL KOMITEK*, 2(2), 431–442. <https://doi.org/10.53697/jkomitek.v2i2>
- Ramadhani, S., Azzahra, D., & Z, T. (2022). Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student's Thesis. *Digital Zone: Jurnal Teknologi*

Informasi Dan Komunikasi, 13(1), 24–33.
<https://doi.org/10.31849/digitalzone.v13i1.9292>

- Renaningtias, N., & Apriliani, D. (2021). PENERAPAN METODE PROTOTYPE PADA PENGEMBANGAN SISTEM INFORMASI TUGAS AKHIR MAHASISWA. In *Jurnal Rekursif* (Vol. 9, Issue 1). <http://ejournal.unib.ac.id/index.php/rekursif/92>
- Riyadhi, M. F. (2019). Aplikasi Text Mining Untuk Automasi Penentuan Tren Topik Skripsi Dengan Metode K-Means Clustering (Studi Kasus: Prodi Sistem Komputer). 8(2), 59–64. <https://ojs.unikom.ac.id/index.php/komputika>
- Sang, A. I., Sutoyo, E., & Darmawan, I. (2021). *Analisis Data Mining Untuk Klasifikasi Data Kualitas Udara Dki Jakarta Menggunakan Algoritma Decision Tree Dan Support Vector Machine Data Mining Analysis For Classification Of Air Quality Data Dki Jakarta Using Decision Tree Algorithm And Support Vector Machiner Algorithm.*
- Sept, F., Zentrato, G., Triayudi, A., & Tri, E. (2022). Analisis Clustering Dokumen Tugas Akhir Mahasiswa Sistem Informasi Universitas Nasional menggunakan Metode K-Means Clustering. *Jurnal Teknologi Informasi Dan Komunikasi*, 6(1), 2022. <https://doi.org/10.35870/jti>
- Shahapure, K. R., & Nicholas, C. (2020). Cluster quality analysis using silhouette score. Proceedings - 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020, 747–748. <https://doi.org/10.1109/DSAA49011.2020.00096>
- Sharyanto, & Lestari, dwi. (2022). Penerapan Data Mining Untuk Menentukan Segmentasi Pelanggan Dengan Menggunakan Algoritma K-Means dan Model RFM Pada E-Commerce. *Jurnal Riset Komputer*, 9(4), 867–871. <https://doi.org/10.30865/jurikom.v9i4.4525>
- Wicaksana, D. A., Pandu Adikara, P., & Adinugroho, S. (2018). *Clustering Dokumen Skripsi Dengan Menggunakan Hierarchical Agglomerative Clustering* (Vol. 2, Issue 12). <http://j-ptiik.ub.ac.id>
- Wiyarsih, W. (2021). Analisis Trends Topik Penelitian Mahasiswa Fakultas MIPA UGM Periode 2016-2018. *UNILIB: Jurnal Perpustakaan*, 12(1). <https://doi.org/10.20885/unilib.vol12.iss1.art1>
- Yang, Y., Sun, H., Zhang, Y., Zhang, T., Gong, J., Wei, Y., Duan, Y. G., Shu, M., Yang, Y., Wu, D., & Yu, D. (2021). Dimensionality reduction by UMAP reinforces sample heterogeneity analysis in bulk transcriptomic data. *Cell Reports*, 36(4). <https://doi.org/10.1016/j.celrep.2021.109442>