

**KLASTERISASI ABSTRAK SKRIPSI MAHASISWA FAKULTAS
ILMU KOMPUTER UNIVERSITAS SRIWIJAYA**

SKRIPSI

Sebagai Salah Satu Syarat Untuk Menyelesaikan Studi

Di Program Studi Sistem Informasi



Oleh

Anindya Dewi Maharani

09031281924058

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

2023

HALAMAN PENGESAHAN

SKRIPSI

**KLASTERISASI ABSTRAK SKRIPSI MAHASISWA FAKULTAS
ILMU KOMPUTER UNIVERSITAS SRIWIJAYA**

Sebagai salah satu syarat untuk menyelesaikan studi
di Program Studi Sistem Informasi SI

Oleh :

Anindya Dewi Maharani 09031281924058

Mengetahui,
Pembimbing 1,



Ir. Muhammad Ihsan Jambak, M.Sc., M.M.
NIP. 198706302015041001

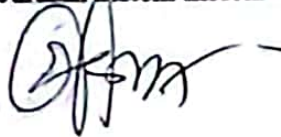
Palembang, 17 Mei 2023
Pembimbing 2,



Ali Bardadi, S.SI., M. Kom
NIP. 198806292019031007



Mengetahui,
Ketua Jurusan Sistem Informasi



Endang Lestari Ruskan, M.T.
NIP. 197811172006042001

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Anindya Dewi Maharani
NIM : 09031281924058
Program Studi : Sistem Informasi Reguler
Judul Skripsi : Klasterisasi Abstrak Skripsi Mahasiswa Fakultas Ilmu
Komputer Universitas Sriwijaya

Hasil Pengecekan *IThenticate/Turnitin* : 7%

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

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



Palembang, 17 Mei 2023



Anindya Dewi Maharani
NIM. 09031281924058

HALAMAN PERSETUJUAN

Teslah diuji dan lulus pada :

Hari : Senin

Tanggal : 17 April 2023

Tim Penguji :

1. Pembimbing 1: Ir. Muhammad Ihsan Jambak, M.Sc., M.M



2. Pembimbing 2: Ali Bardadi, S.SI., M. Kom



3. Ketua Penguji : Yadi Utama, M.Kom

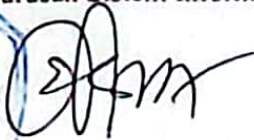


4. Penguji 1 : Fathoni, MMSI

5. Penguji 2 : Rizka Dhini Kurnia, M.Sc



Mengetahui,
Ketua Jurusan Sistem Informasi



Endang Lestari Ruskan, M.T.
NIP. 197811172006042001

HALAMAN PERSEMBAHAN

“ Fa inna ma’al-‘usri yusrā (karena sesungguhnya sesudah kesulitan itu ada kemudahan). Inna ma’al-‘usri yusrā (Sesungguhnya sesudah kesulitan itu ada kemudahan).” (QS. Al-Insyirah:94)

Skripsi ini kupersembahkan untuk:

- Allah SWT
- Diri sendiri, Anindya Dewi Maharani
- Orang tua & Keluarga
- Dosen Pembimbing
- Dosen Penguji
- Seluruh Dosen beserta Staff jurusan Sistem Informasi
- Teman-teman seperjuangan Sistem Informasi 2019
- Semua orang baik yang tidak bisa disebutkan satu-persatu
- Almamater kebanggaan, Universitas Sriwijaya.

KLASTERISASI ABSTRAK SKRIPSI MAHASISWA FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

Oleh

Anindya Dewi Maharani

09031281924058

ABSTRAK

Saat ini, kerap ditemukan penulisan abstrak yang terkesan sembarang. Fakultas Ilmu Komputer Universitas Sriwijaya, sebagai institusi penyelenggara pendidikan strata 1, berkomitmen penuh untuk meningkatkan kualitas mahasiswa dengan menggunakan skripsi sebagai tolok ukur kemampuan mahasiswa dalam bidang studi mereka. Untuk menjaga konsistensi dalam pengembangan program studi, Fakultas Ilmu Komputer Universitas Sriwijaya memerlukan strategi yang tepat dalam penyelenggaraan program studi. Penelitian ini menerapkan pemanfaatan ilmu *data mining* dalam pengolahan data skripsi mahasiswa fakultas ilmu komputer untuk mengelompokan abstrak skripsi berdasarkan prodi (Sistem Informasi, Teknik Informatika, dan Sistem Komputer). Pengolahan data menggunakan aplikasi RapidMiner menghasilkan 3 *cluster* skripsi berdasarkan kemiripan antar abstrak skripsi. Dari analisis *cluster* yang dilakukan, ditemukan bahwa tiap prodi memiliki karakteristik abstrak skripsi yang berbeda-beda berupa kata – kata yang sering muncul dalam abstrak. Temuan penelitian ini dapat menjadi pedoman bagi mahasiswa yang sedang menyusun skripsi dan masukan kepada dosen pembimbing skripsi untuk memberikan arahan dalam menyusun abstrak yang baik.

Kata Kunci: *Data mining*, Klasterisasi, K-Means, RapidMiner, Skripsi, Prodi

Clustering Student Thesis Abstract of Faculty of Computer Science, Sriwijaya University

By

Anindya Dewi Maharani

09031281924058

ABSTRACT

Currently, it is often found that the writing of thesis abstracts seems haphazard. The Faculty of Computer Science, Sriwijaya University, as an institution that organizes undergraduate education, is fully committed to improve the quality of its students by using thesis as a benchmark of students' abilities in their field of study. To maintain its consistency in the development of study programs, the Faculty of Computer Science, Sriwijaya University, requires the right strategy in the implementation of the study programs. This research applies the utilization of data mining science in processing the thesis data of computer science students to group thesis abstracts based on study programs (Information Systems, Informatics Engineering, and Computer Systems). Data processing, using the RapidMiner application, produces 3 thesis clusters based on the similarity between thesis abstracts. From the cluster analysis that had been conducted, it was found that each program has different characteristics of thesis abstracts in the form of words that often appear in abstracts. The findings of this study can serve as a guide for students who are currently preparing their thesis and as feedback for their thesis advisors to provide guidance in creating a good abstract.

Keywords: Data mining, Clustering, K-Means, RapidMiner, Thesis, Study Program

KATA PENGANTAR



Alhamdulillah rabbil'alamini segala puji dan syukur kehadirat Allah SWT atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Penulisan Tugas Akhir ini bertujuan untuk memenuhi persyaratan dalam menyelesaikan studi Strata-1 di Program Studi Sistem Informasi yang diselenggarakan oleh Fakultas Ilmu Komputer di Universitas Sriwijaya. Penulis ingin mengucapkan rasa terima kasih kepada berbagai pihak yang telah memberikan bantuan dan dukungan dalam menyelesaikan Tugas Akhir ini, baik secara langsung maupun tidak langsung, yaitu kepada:

1. Kedua orang tua dan keluarga penulis, yang telah memberikan doa restu dan dukungan yang luar biasa selama penulis menempuh pendidikan di Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Sriwijaya, hingga penulis berhasil menyelesaikan Skripsi ini,
2. Bapak Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Ibu Endang Lestari Ruskan, S.Kom M.T, selaku Ketua Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Ir. Muhammad Ihsan Jambak, M.Sc., M.M dan Bapak Ali Bardadi, S.SI., M.Kom sebagai pembimbing Tugas Akhir yang memberi motivasi dan ilmu dalam proses pengerjaan penelitian ini.
5. Bapak Iman Saladin B. Azhar, S.Kom., M.Msi selaku dosen pembimbing akademik, yang telah membimbing dan memberikan motivasi penulis dalam proses menjalankan studi.
6. Bapak Yadi Utama, M.Kom selaku ketua penguji serta Bapak Fathoni, ST., MMSI dan Ibu Rizka Dhini Kurnia, M.Sc. selaku dosen penguji, yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
7. Seluruh dosen Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Sriwijaya.
8. Seluruh staf jurusan Sistem Informasi yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
9. Para teman seperjuangan Aca, Anabil, Fahlevi, Faiz, Jihan, Julio, Juwinda, Liwi, Pipit dan Rizka yang telah memberikan dukungan, semangat, dan motivasi dalam menyelesaikan tugas akhir ini.
10. Teman-teman dari kelas SI Reguler 2019, kakak tingkat, serta teman-teman lainnya yang telah membantu dalam menyelesaikan tugas akhir

11. Semua orang baik yang tidak bisa disebutkan satu per satu namanya, yang telah membantu dan mendukung saya dalam proses penulisan skripsi ini.

Penulis mengakui bahwa dalam penulisan Tugas Akhir ini masih terdapat banyak kekurangan yang disebabkan oleh keterbatasan pengetahuan dan pengalaman. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun guna meningkatkan kualitas penelitian selanjutnya di masa depan.

Terakhir, penulis berharap Tugas Akhir ini dapat memberikan manfaat dan kontribusi bagi kita semua.

Palembang, 17 Mei 2023



Anindya Dewi Maharani
NIM. 09031281924058

DAFTAR ISI

HALAMAN PENGESAHAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PERNYATAAN	iv
HALAMAN PERSEMBAHAN	v
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan	4
1.4 Manfaat	4
1.5 Batasan Masalah	5
BAB II TINJAUAN PUSTAKA	6
2.1 <i>Data mining</i>	6
2.1.1 Pengertian <i>Data mining</i>	6
2.1.2 Peran <i>Data mining</i>	7
2.2 <i>Text Mining</i>	8
2.3 Klasterisasi	9
2.4 Algoritma <i>K-Means Clustering</i>	11
2.5 <i>Singular Value Decomposition</i>	13
2.6 <i>Davies Bouldin Index (DBI)</i>	14
2.7 Skripsi	15
2.8 <i>Cross Industry Standard Process for Data mining (CRISP-DM)</i>	16
2.9 <i>Text Preprocessing</i>	19
2.10 <i>Confusion Matrix</i>	21
2.11 <i>RapidMiner</i>	23

2.12 Penelitian Terdahulu	24
BAB III METODOLOGI PENELITIAN	26
3.1 Jenis dan Sumber Data	26
3.2 . Tahapan Penelitian	26
3.2.1 Pengumpulan Data dan Pemahaman Bisnis	27
3.2.2 Pemahaman Data.....	28
3.2.3 <i>Data Preparation</i>	29
3.2.4 <i>Exploratory Data Analysis</i>	36
3.2.5 Pemodelan Data dengan RapidMiner.....	37
3.2.5.1 Algoritma K-Means	37
3.2.5.2 Perbandingan Algoritma	38
3.2.5.3 Analisa Hasil <i>Clustering</i>	38
3.2.6 Evaluasi.....	38
3.2.7 <i>Deployment</i>	40
3.3 Pemodelan RapidMiner	40
3.3.1 Pengujian Perbandingan Algoritma	41
3.3.2 Pengujian Parameter Algoritma	43
BAB IV HASIL DAN PEMBAHASAN	45
4.1 Hasil Pengujian	45
4.1.1 Hasil Pengujian Algoritma	45
4.1.2 Hasil Pengujian Parameter Algoritma.....	46
4.2 Hasil <i>Clustering</i>	47
4.3 Analisis Hasil Pengujian	49
4.3.1 Analisis Hasil Pengujian Perbandingan Kombinasi Algoritma K-Means dan SVD & K-Means.....	49
4.3.2 Analisis Hasil Pengujian Parameter Algoritma.....	52
4.3.3 Kesimpulan Hasil Pengujian Algoritma dan Parameter.....	54
4.4 Analisis Hasil <i>Clustering</i>	55
4.5 Evaluasi Abstrak Skripsi dengan Prodi.....	57
4.5.1 <i>Confusion Matrix</i>	57
4.5.2 Analisis Skripsi Tiap Prodi	61
4.5.3 Kesimpulan Analisis Skripsi Tiap Prodi.....	75
4.6 Visualisasi Data	77

BAB V PENUTUP.....	84
5.1 Kesimpulan	84
5.2 Saran	86
DAFTAR PUSTAKA.....	88
LAMPIRAN.....	91

DAFTAR GAMBAR

	Halaman
Gambar 2.1 <i>Framework CRISP-DM</i>	19
Gambar 3. 1 Diagram Batang Komposisi Data Skripsi Fasilkom Unsri.....	26
Gambar 3.2 Tahapan Penelitian	27
Gambar 3.3 Operator <i>Text Processing</i>	30
Gambar 3.4 Parameter <i>Text Processing</i>	30
Gambar 3.5 Operator <i>Replace</i>	31
Gambar 3.6 Parameter <i>Replace</i>	31
Gambar 3.7 Operator <i>Transform Cases</i>	32
Gambar 3.8 Parameter <i>Transform Cases</i>	32
Gambar 3.9 Operator <i>Tokenize</i>	33
Gambar 3.10 Parameter <i>Tokenize</i>	33
Gambar 3.11 Operator <i>Stopword</i>	33
Gambar 3.12 Operator <i>Stemming</i>	34
Gambar 3.13 Tahapan Kerja Algoritma <i>Singular Value Decomposition</i>	34
Gambar 3.14 Parameter Algoritma <i>Singular Value Decomposition</i>	35
Gambar 3. 15 Tahap Data Preparation di RapidMiner	35
Gambar 3.16 <i>Subproses</i> Operator <i>Process Documents From Text</i>	35
Gambar 3.17 Proses Pemodelan <i>Clustering</i>	37
Gambar 3.18 <i>Import</i> Data di RapidMiner.....	40
Gambar 3.19 <i>Input</i> Data Ke <i>Design</i> Proses	41
Gambar 3.20 <i>Design</i> Proses Algoritma K-Means.....	41
Gambar 3.21 Parameter Algoritma K-Means	42
Gambar 3.22 Performa Kombinasi Algoritma dan SVD	42
Gambar 3.23 Performa Algoritma K-Means.....	42
Gambar 3. 24 Pencarian Operation “K-Means” Pada RapidMiner.....	43
Gambar 3. 25 <i>Design</i> Proses Akhir.....	43
Gambar 3. 26 Penyesuaian Parameter Algoritma	44
Gambar 4. 1 Hasil <i>Clustering</i>	47
Gambar 4.2 Contoh Hasil <i>Cluster 0</i>	48
Gambar 4.3 Hasil Tabel <i>Centroid</i>	48

Gambar 4.4 Hasil Visualisasi Plot	49
Gambar 4.5 Hasil <i>Scatter Plot Cluster</i>	55
Gambar 4. 6 <i>Wordcloud</i> Abstrak Skripsi SK Pada <i>Cluster</i> 0(SK).....	66
Gambar 4. 7 <i>Wordcloud</i> Abstrak Skripsi TI Pada <i>Cluster</i> 0 (SK)	67
Gambar 4. 8 <i>Wordcloud</i> Abstrak Skripsi TI Pada <i>Cluster</i> 1 (TI)	68
Gambar 4. 9 <i>Wordcloud</i> Abstrak Skripsi SI Pada Cluste 1 (TI)	70
Gambar 4. 10 <i>Wordcloud</i> Abstrak Skripsi SK Pada <i>Cluster</i> 1 (TI)	71
Gambar 4. 11 <i>Wordcloud</i> Abstrak Skripsi SI Pada <i>Cluster</i> 2 (SI).....	72
Gambar 4. 12 <i>Wordcloud</i> Abstrak Skripsi TI Pada <i>Cluster</i> 2 (SI).....	74
Gambar 4. 13 Abstrak Skripsi SK Pada <i>Cluster</i> 2 (SI)	74
Gambar 4. 14 Diagram Batang Pesebaran Data Skripsi Prodi SK 5 Tahun Terakhir di Tiap <i>Cluster</i>	77
Gambar 4. 15 Diagram Batang Pesebaran Data Skripsi Prodi TI 5 Tahun Terakhir di Tiap <i>Cluster</i>	79
Gambar 4. 16 Diagram Batang Pesebaran Data Skripsi Prodi SI 5 Tahun Terakhir di Tiap <i>Cluster</i>	80
Gambar 4. 17 <i>Pie chart</i> Hasil <i>Clustering</i> Skripsi Fasilkom Unsri.....	81
Gambar 4. 18 <i>Scatter Plot</i> Hasil <i>Clustering</i> Skripsi Fasilkom Universitas Sriwijaya	81
Gambar 4. 19 <i>Scatter Plot</i> Persebaran Skripsi Prodi Teknik Informatika Di Tiap <i>Cluster</i>	82
Gambar 4. 20 <i>Scatter Plot</i> Persebaran Skripsi Prodi Sistem Komputer Di Tiap <i>Cluster</i>	82
Gambar 4. 21 <i>Scatter Plot</i> Persebaran Skripsi Prodi Sistem Informasi Di Tiap <i>Cluster</i>	83

DAFTAR TABEL

	Halaman
Tabel 2.1 Penelitian Terdahulu	24
Tabel 3.1 Jenis Tipe Data.....	28
Tabel 3. 2 <i>Word List</i> Hasil <i>Text Preprocessing</i>	36
Tabel 3.3 <i>confusion matrix</i>	39
Tabel 3. 4 Tabel Perbandingan Parameter Algoritma	44
Tabel 4.1 Hasil Pengujian	45
Tabel 4. 2 Hasil Pengujian Parameter Algoritma.....	46
Tabel 4. 3 <i>Group Statistics</i> Pengujian Algoritma	50
Tabel 4. 4 Hasil Independent Sample T-Test Pengujian Algoritma	51
Tabel 4. 5 Hasil Uji Normalitas	52
Tabel 4. 6 Hasil Uji Kruskal Wallis.....	54
Tabel 4.7 <i>Centroid Table</i> Kombinasi Algoritma K-Means dan SVD.....	55
Tabel 4.8 Hasil Pelabelan <i>Cluster</i>	56
Tabel 4.9 Hasil Analisis <i>Confusion Matrix</i>	57
Tabel 4.10 Hasil Analisis <i>Confusion Matrix II</i>	57
Tabel 4. 11 Ciri-Ciri <i>Cluster</i>	62

DAFTAR LAMPIRAN

	Halaman
Lampiran A. Hasil Pengumpulan Data	A-1
Lampiran B. Data Hasil Pengolahan.....	B-1
Lampiran C. Kurikulum Jurusan Sistem Komputer.....	C-1
Lampiran D. Kurikulum Jurusan Teknik Informatika	D-1
Lampiran E. Kurikulum Jurusan Sistem Informasi.....	E-1
Lampiran F. Kartu Konsultasi.....	F-1
Lampiran G. Surat Rekomendasi Ujian Akhir.....	G-1
Lampiran H. Hasil Pengecekan Similiry	H-1
Lampiran I. Surat Kesiadaan Membimbing	I-1
Lampiran J. Surat Pengumpulan Data.....	J-1
Lampiran K. Berita Acara Ujian Komprehensif	K-1
Lampiran L. Form Perbaikan Ujian Komprehensif	L-1
Lampiran M. Form Daftar Hadir Ujian Komprehensif	M-1
Lampiran N. Surat Keterangan Pengecekan Similiry	N-1

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Skripsi adalah salah satu syarat yang harus dipenuhi mahasiswa S1 untuk mendapatkan gelar. Skripsi dinilai sebagai tolak ukur kemampuan mahasiswa dalam menerima sekaligus mengimplementasikan ilmu yang didapat selama bangku perkuliahan. Pada zaman sekarang sering ditemukan adanya kepenulisan abstrak yang sering luput dari perhatian. Sangat disayangkan abstrak kerap kali dianggap remeh sehingga penyusunannya terkesan sembarang. Abstrak skripsi berisi inti dan tujuan penelitian sehingga mempermudah pembaca dalam memahami penelitian. Pemilihan kata dalam abstrak harus diperhatikan apakah sudah bisa merepresentasikan isi skripsi secara menyeluruh. Abstrak skripsi merupakan bagian yang penting dalam proses penulisan skripsi karena dapat mempengaruhi persepsi pembaca dan calon penguji terhadap kualitas dan relevansi penelitian yang dilakukan. Namun, banyak mahasiswa yang kesulitan menyusun abstrak skripsi yang baik dan benar karena kurangnya pemahaman akan ciri-ciri abstrak skripsi yang baik dan benar.

Fakultas Ilmu Komputer (Fasilkom) Universitas Sriwijaya berdiri sejak tahun 2006 dan sejak saat itu setiap tahunnya Fasilkom mencetak ratusan sarjana. Perpustakaan Fasilkom Universitas Sriwijaya sudah mengarsipkan ratusan skripsi pula. Setiap tahunnya pula mata kuliah menjadi semakin beragam begitu pula dengan bidang penelitian skripsi. Tetapi bertambahnya mata kuliah tidak selamanya menghasilkan

bidang penelitian baru ada kalanya menyebabkan semakin banyak mahasiswa yang melakukan penelitian yang serupa dari segi tema, objek, metodologi dsb. dari penelitian sebelumnya. Pengelompokan skripsi yang sedang dilakukan oleh Perpustakaan Fakultas Ilmu Komputer hanya berdasarkan prodi saja tidak ada proses penemuan karakteristik abstrak tiap prodi. Proses klasterisasi skripsi untuk penemuan karakteristik abstrak tiap prodi bisa saja dilakukan secara manual namun hal tersebut dirasa kurang optimal karena jumlah data yang besar akan membutuhkan waktu yang lama, tenaga yang besar dan memperbesar kemungkinan kesalahan klasterisasi. Oleh karena itu, dibutuhkan sistem yang dapat melakukan penemuan karakteristik abstrak tiap prodi secara otomatis. Sistem tersebut harus dapat mengolah data yang besar dan memberikan informasi dengan tingkat akurasi setinggi mungkin. Ekstraksi informasi dari data *text* yang besar ini dapat dilakukan dengan teknik *text mining*. *Text mining* sendiri termasuk ke dalam variasi *data mining* (Purnajaya et al., 2022).

Seperti yang sudah dijelaskan sebelumnya pengelompokan abstrak skripsi dapat dilakukan dengan memanfaatkan ilmu *data mining* dengan mengadopsi salah satu peran *data mining* yaitu klasterisasi. Klasterisasi dapat dilakukan karena sebelumnya belum ada aturan dari Fakultas Ilmu Komputer Universitas Sriwijaya yang mengatur mengenai karakteristik abstrak tiap prodi. Klasterisasi abstrak skripsi dapat menggunakan berbagai macam algoritma *clustering*. Algoritma *clustering* bekerja untuk data tanpa label dengan membagi objek ke dalam beberapa *cluster* dengan memperhatikan karakteristiknya, objek yang memiliki karakteristik yang sama akan ditempatkan di *cluster* yang sama, begitupun sebaliknya.

Abstrak skripsi ditulis dalam dua bahasa yaitu Bahasa Indonesia dan Bahasa Inggris. Bahasa Inggris memiliki susunan tata bahasa yang kompleks dan beberapa kata memiliki arti ganda yang membuatnya berbeda dari bahasa lain. Karena hal tersebut lah membuat dokumen skripsi menjadi banyak atribut dimana setiap satu atribut mewakili satu dimensi dengan kata lain termasuk ke dalam data berdimensi tinggi. Fenomena ini disebut juga sebagai *curse of dimensionality*, atau fenomena dimana seiring bertambahnya jumlah dimensi, pemrosesan data menjadi lebih lambat dan kurang tepat untuk menangani hal tersebut perlu dilakukan reduksi dimensi agar pemrosesan data lebih efisien (Puspasari et al., 2022). Algoritma *Clustering* biasa seperti K-Means hanya mampu menangani data berdimensi kecil maka hal ini menimbulkan tantangan baru dalam pengelompokan data berdimensi tinggi (Jambak et al., 2019). Solusi yang bisa ditawarkan untuk menangani masalah tersebut adalah dengan melakukan reduksi dimensi. Ada banyak metode dalam mereduksi dimensi salah satunya adalah *Singular Value Decomposition*, SVD akan meningkatkan akurasi hasil klaster karena lebih *precise* menemukan kesamaan antar data.

Oleh karena itu penelitian ini akan berfokus pada mengelompokan abstrak skripsi menggunakan algoritma *clustering* dan *Singular Value Decomposition*. Data yang digunakan berasal dari Perpustakaan Fakultas Ilmu Komputer Universitas Sriwijaya. Dengan tujuan melihat karakteristik abstrak skripsi tiap prodi yang merepresentasikan isi skripsi untuk selanjutnya dijadikan pedoman bagi mahasiswa yang sedang menyusun skripsi dan masukan kepada dosen pembimbing skripsi untuk memberikan arahan dalam menyusun abstrak yang baik.

1.2 Rumusan Masalah

Berdasarkan penjabaran latar belakang masalah diatas, adapun rumusan masalah penelitian ini adalah: Bagaimana tingkat kesesuaian skripsi mahasiswa Fakultas Ilmu Komputer Universitas Sriwijaya. Untuk menjawab rumusan masalah tersebut, maka dapat dipecah ke dalam beberapa pertanyaan penelitian yaitu:

1. Bagaimana keadaan data skripsi yang dimiliki Perpustakaan Fakultas Ilmu Komputer Universitas Sriwijaya?
2. Algoritma apa yang paling tepat untuk mengklasterisasi skripsi berdasarkan abstraknya di Perpustakaan Fakultas Ilmu Komputer Universitas Sriwijaya?
3. Bagaimana karakteristik abstrak skripsi sebagai representasi isi skripsi tiap prodi?

1.3 Tujuan

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui dan menafsirkan keadaan data skripsi yang dimiliki Perpustakaan Fakultas Ilmu Komputer Universitas Sriwijaya.
2. Untuk mengetahui algoritma yang paling tepat untuk mengklasterisasi skripsi berdasarkan abstraknya di Perpustakaan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Untuk mengetahui karakteristik abstrak skripsi sebagai representasi isi skripsi tiap prodi.

1.4 Manfaat

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Dapat mengetahui pengetahuan secara umum mengenai data skripsi Fakultas Ilmu Komputer Universitas Sriwijaya
2. Dapat memilih algoritma *clustering* yang paling tepat untuk diimplementasikan menggunakan data skripsi yang dimiliki oleh pihak Perpustakaan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Dapat mengelompokkan data skripsi sehingga hasil pengelompokan tersebut dapat dijadikan sebagai salah satu referensi untuk mahasiswa Fasilkom Unsri dalam menyusun abstrak skripsi agar lebih mencirikan sesuai dengan prodi yang diambil.

1.5 Batasan Masalah

Adapun batasan masalah penelitian ini sebagai berikut:

1. Data yang digunakan berupa abstrak skripsi berbahasa Inggris milik S1 berasal dari Perpustakaan Fakultas Ilmu Komputer Universitas Sriwijaya dan website resmi repository Unsri.
2. Data preparation menggunakan algoritma *Singular value decomposition* (SVD) untuk mereduksi dimensi guna meningkatkan akurasi *clustering*.
3. Proses kerja mengikuti aturan kerangka berpikir *Cross-Industry Standard Process for Data mining* (CRISP-DM).
4. Hasil *clustering* akan dievaluasi secara eksternal menggunakan *confusion matrix*.
5. *Tools* yang digunakan dalam proses *text mining* adalah RapidMiner Studio.

DAFTAR PUSTAKA

- Aulia, S. (2021). Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means *Clustering* (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja). *Djtechno: Jurnal Teknologi Informasi*, 1(1), 1–5. <https://doi.org/10.46576/djtechno.v1i1.964>
- Aulia, T. M. P., Jamaludin, A., & ... (2021). Extractive Text Summerization Pada Berita Berbahasa Indonesia Menggunakan Algoritma Support Vector Machine. *J-SAKTI (Jurnal Sains ...)*, 5(September), 727–735. <http://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/view/371>
- Dewi, D. A. I. C., & Pramita, D. A. K. (2019). Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma *Clustering* K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Matrix : Jurnal Manajemen Teknologi Dan Informatika*, 9(3), 102–109. <https://doi.org/10.31940/matrix.v9i3.1662>
- Fauziningrum, M. Pd, E., & Encis Indah Suryaningsih, S.T., M. P. (2021). Evaluasi Dan Prediksi Penguasaan Bahasa Inggris Maritim Menggunakan Metode Decision Tree Dan *Confusion Matrix* (Studi Kasus Di Universitas Maritim Amni). *Angewandte Chemie International Edition*, 6(11), 951–952., 5–24.
- Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing*, 5(2), 103–108. <https://doi.org/10.30871/jaic.v5i2.3200>
- Jambak, M. I., Mohammed, F., Hidayati, N., Efendi, R., & Primartha, R. (2019). The impacts of singular value decomposition algorithm toward indonesian language text documents *clustering*. *Advances in Intelligent Systems and Computing*, 843(March), 173–183. https://doi.org/10.1007/978-3-319-99007-1_17
- Kambey, G. E. I., Sengkey, R., & Dkk. (2020). Penerapan *Clustering* pada Aplikasi Pendeteksi Kemiripan Dokumen Teks Bahasa Indonesia. *Penerapan Clustering Pada Aplikasi Pendeteksi Kemiripan Dokumen Teks Bahasa Indonesia*, 15(2), 75–82.
- Kurniana, I. R., Muhima, R. R., Wardana, S., & Hakimah, M. (2021). *Penerapan Algoritma K-Means Untuk Pengelompokan Topik Dokumen Studi Kasus:Dokumen Abstrak Skripsi Jurusan Teknik Informatika ITATS*. 291–296.
- Muliono, R., & Sembiring, Z. (2019). *Data mining Clustering* Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen. *CESS (Journal of Computer Engineering, System and Science)*, 4(2), 2502–2714.
- Nainggolan, R., & Lumbantoruan, G. (2018). Optimasi Performa *Cluster* K-Means Menggunakan Sum of Squared Error (Sse). *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi*, 2(2), 103–108.

<https://ejurnal.methodist.ac.id/index.php/methomika/article/view/138>

- Nurfauzan, R. A., Hidayat, B., & Saidah, S. (2018). Analisis Steganografi Ganda pada Citra Digital Menggunakan Metode Discrete Wavelet Transform dan Singular Value Decomposition dengan Penyisipan Spread Spectrum Image Steganography. *Proceeding of Engineering*, 5(1), 299–304.
- Permana, R. A., Hapsari, D. P., & Muhima, R. R. (2022). Penerapan metode clustering *k-medoids* untuk pengelompokan abstrak skripsi berbasis web. 277–284.
- Purnajaya, A. R., Lieputra, V., Tayanto, V., & Salim, J. G. (2022). Implementasi Text Mining untuk Mengetahui Opini Masyarakat Tentang Climate Change. *Journal of Information System and Technology*, 03(03), 36–44.
- Puspasari, I. N. O., Informatika, P. S., Komunikasi, F., Informatika, D. A. N., & Surakarta, U. M. (2022). Perbandingan Akurasi Algoritma Reduksi Dimensi *Principal Component Analysis (Pca)* Dan *Singular Value Decomposition (SVD)* Dalam Klasifikasi Data.
- Rachman, D. A. C., Goejantoro, R., & Amijaya, F. D. T. (2020). Implementasi Text Mining Pengelompokan Dokumen Skripsi Menggunakan Metode K-Means Clustering. *Jurnal EKSPONENSIAL*, 11(2), 167–174.
- Religia, Y., & Jaya, R. T. B. (2020). Pengelompokan Menggunakan Algoritma K-Medoid Untuk Evaluasi Performa Siswa. *Jurnal Pelita Teknologi*, 15(1), 49–55.
- Rosadi, A., Gustiana, D., Informatika, M., Sti, S. J., No, J. B. R. I., Dalam, R., Baru, K., Selatan, J., Kata, P., Baku, T., & Bernegasi, M. K. (2021). Analisis Sentimen Berdasarkan Opini Pengguna pada Media Twitter Terhadap BPJS Menggunakan Metode Lexicon Based dan Naïve Bayes Classifier. *Jurnal Ilmiah Komputasi*, 20(1), 39–52. <https://doi.org/10.32409/jikstik.20.1.401>
- Saifudin, Kusri, & Fatta, H. Al. (2020). Text Mining Dokumen Tweet Pada Twitter Untuk Klasifikasi Calon Karyawan. *Jurnal Informasi Interaktif*, 5(1), 12–18.
- Sari, H., Ginting, G. L., & Zebua, T. (2021). Penerapan Algoritma Text Mining dan TF-IDF Untuk Pengelompokan Topik Skripsi Pada Aplikasi Repository STMIK Budi Darma. *Terapan Informatika Nusantara*, 2(7), 414–432.
- Setiawan, R. (2021). Klasifikasi Sentimen terhadap Normalisasi Sungai Untuk Pencegahan Banjir dengan Metode Modified K-Nearest Neighbor (MKNN). *Repository.Usd.Ac.Id*, 1–85. https://repository.usd.ac.id/25510/2/084114001_Full%5B1%5D.pdf
- Sikumbang, E. D. (2018). Penerapan *Data mining* Penjualan Sepatu Menggunakan Metode Algoritma Apriori. *Jurnal Teknik Komputer AMIK BSI (JTK)*, Vol 4, No.(September), 1–4.
- Simatupang, M. P., & Utomo, D. P. (2019). Analisa Testimonial Dengan Menggunakan Algoritma Text Mining Dan Term Frequency- Inverse Document Frequence (Tf-

- Idf) Pada Toko Allmeeart. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1), 808–814. <https://doi.org/10.30865/komik.v3i1.1697>
- Siringoringo, R., Jamaluddin, J., & Lumbantoruan, G. (2021). Sistem Rekomendasi Dengan Singular Value Decomposition Dan Teknik Similaritas Pearson Correlation. *METHODIKA: Jurnal Teknik Informatika Dan Sistem Informasi*, 7(1), 19–24. <https://doi.org/10.46880/mtk.v7i1.257>
- Syah, H., & Witanti, A. (2022). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm). *Jurnal Sistem Informasi Dan Informatika (Simika)*, 5(1), 59–67. <https://doi.org/10.47080/simika.v5i1.1411>
- Yainahu, I., Hidayat, B., & Saidah, S. (2019). Klasifikasi Jenis Batik Toraja Berbasis Pengolahan Citra Digital Dengan Metode Singular Value Decomposition Dan Learning Vector Quantization. *EProceedings ...*, 6(1), 834–842. <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/8885>
- Zanuardi, A., & Suprayitno, H. (2018). Analisa Karakteristik Kecelakaan Lalu Lintas di Jalan Ahmad Yani Surabaya melalui Pendekatan Knowledge Discovery in Database. *Jurnal Manajemen Aset Infrastruktur & Fasilitas*, 2(1), 45–55. <https://doi.org/10.12962/j26151847.v2i1.3767>