

Penerapan Metode *K-Means* dan *Euclidean Distance* Untuk Seleksi Metode Judul Tugas Akhir

Leny Tritanto Ningrum^{1*}, Irmayansyah², Lis Utari³

^{1*,2}Sistem Informasi, Fakultas Informatika dan Komputer, Universitas Binaniaga Indonesia, Indonesia

³Teknik Informatika, Fakultas Informatika dan Komputer, Universitas Binaniaga Indonesia, Indonesia

Email: ^{1*}lenytrinie@unbin.ac.id, ²irma@unbin.ac.id, ³lis.utari@gmail.com

Abstrak – Salah satu syarat kelulusan mahasiswa pada sebuah perguruan tinggi yaitu melalui tahap penyelesaian tugas akhir dalam bentuk skripsi sebagai laporan penelitian. Tahap awal penyusunan skripsi yaitu pengajuan judul tugas akhir berdasarkan bidang kajian dan metode yang digunakan. Data skripsi di *repository* perpustakaan semakin bertambah. Bidang kajian dan metode yang dipilih mahasiswa seringkali serupa dengan data judul skripsi yang tersimpan di perpustakaan. Dari data judul skripsi yang tersimpan tersebut dapat digali pengetahuan agar lebih selektif dalam memberikan persetujuan berdasarkan bidang kajian dan metode judul tugas akhir yang diajukan mahasiswa. Objek pada penelitian ini adalah salah satu perguruan tinggi swasta di Bogor. Adapun tujuan penelitian ini adalah untuk mengetahui sebaran bidang kajian dan metode yang pernah digunakan oleh mahasiswa sebagai bahan evaluasi untuk pengajuan judul skripsi berikutnya sehingga dapat lebih memudahkan bagian akademik. Metode yang digunakan pada penelitian ini adalah algoritma *K-Means* dan pengukuran jarak data menggunakan *Euclidean Distance*. Data yang digunakan adalah judul skripsi mahasiswa tahun 2019, 2020, 2021, dan 2022 yang terdiri dari atribut judul, kelas, dan topik penelitian. Dalam proses pengolahannya, dilakukan persiapan data (*data preparation*) dengan mentransformasikan data judul menjadi atribut metode dan bidang kajian. Setelah dilakukan persiapan data dilakukan proses normalisasi yang bertujuan mengubah data kategori menjadi numerik agar dapat diterapkan pada tahap modeling. Penelitian ini menghasilkan 2 klaster berdasarkan bidang kajian dan metode yang terdiri dari metode Data Mining serta *Decision Support System (DSS)*. Hasil uji akurasi klaster menggunakan *Silhouette Coefficient* sebesar 0,62% dan menunjukkan klaster yang dihasilkan baik dan masuk ke dalam kategori *medium structure*.

Kata Kunci – Euclidean, *K-Means*, *Silhouette*, Tugas Akhir

Abstract – One of the requirements for student graduation at a university is to complete the final assignment in the form of a thesis as a research report. The initial stage of preparing a thesis is submitting a final assignment title based on the field of study and method used. Thesis data in the library repository is increasing. The field of study and methods chosen by students are often similar to the thesis title data stored in the library. From the stored thesis title data, knowledge can be

extracted to be more selective in giving approval based on the field of study and method of the final assignment title submitted by the student. The object of this research is a private university in Bogor. The aim of this research is to find out the distribution of fields of study and methods that have been used by students as evaluation material for submitting the next thesis title so that it can make things easier for the academic part. The method used in this research is the K-Means algorithm and measuring data distance using Euclidean Distance. The data used are student thesis titles for 2019, 2020, 2021, and 2022 which consist of the attributes of title, class, and research topic. In the processing process, data preparation is carried out by transforming title data into method attributes and field of study. After data preparation, a normalization process is carried out which aims to convert categorical data into numeric so that it can be applied at the modeling stage. This research produced 2 clusters based on field of study and methods consisting of Data Mining methods and Decision Support Systems (DSS). The results of the cluster accuracy test using the Silhouette Coefficient were 0.62% and showed that the clusters produced were good and fell into the medium structure category.

Keywords – *Euclidean, Final Paper, K-Means, Silhouettes*

I. PENDAHULUAN

Data skripsi di *repository* perpustakaan semakin bertambah. Bidang kajian dan metode yang dipilih mahasiswa seringkali serupa dengan data judul skripsi yang tersimpan di perpustakaan. Dari data judul skripsi yang tersimpan tersebut dapat digali pengetahuan untuk mengetahui sebaran bidang kajian dan metode yang pernah digunakan oleh mahasiswa.

Skripsi merupakan laporan penelitian yang di susun oleh mahasiswa dengan mengikuti kaedah tulisan ilmiah. Laporan ilmiah ini merupakan pengungkapan kecakapan pengetahuan mahasiswa sebagai penanda berakhirnya masa studi mahasiswa bersangkutan di perguruan tinggi, dalam perjalanan penyusunan skripsi mahasiswa akan menjalani beberapa rangkaian kegiatan mulai dari pemilihan judul penelitian, penyusunan Penelitian hingga mempublikasikan hasil penelitian. Pemilihan judul skripsi merupakan kegiatan awal yang harus dilakukan mahasiswa sebelum dilakukan penyusunan dan publikasi hasil penelitian. Untuk dapat melakukan kegiatan tersebut, mahasiswa dapat terlebih dahulu mempelajari permasalahan yang muncul di sekitar atau dapat mempelajari penelitian-penelitian yang sudah dilakukan sebelumnya. Pemilihan judul skripsi merupakan

tahapan penting dalam penyusunan skripsi, dan untuk memastikan judul yang diajukan mahasiswa tidak sama dengan skripsi sebelumnya, perlu dilakukan evaluasi dan kontrol hal tersebut diakibatkan karena semakin banyaknya jumlah skripsi yang telah disusun. Untuk membantu evaluasi dan kontrol terhadap judul skripsi yang diajukan mahasiswa, terdapat beberapa pendekatan yang dapat diterapkan seperti pendekatan *K-means clustering* dengan *Euclidean Distance*, pendekatan tersebut digunakan untuk mengelompokkan metode dan bidang kajian kemudian dilakukan visualisasi yang disesuaikan dengan karakteristik data dari judul yang diajukan mahasiswa.

Pada penelitian sebelumnya K-means terbukti dapat mengelompokkan ide judul skripsi dan jurnal penelitian berdasarkan data skill dan basic yang didominasi pada mata kuliah yang paling banyak diminati [1]. Pada penelitian tersebut k-means digunakan untuk pengelompokan atau klusterisasi data menjadi beberapa kelas berdasarkan kemiripan data. Pada penelitian yang akan dilakukan peneliti akan menggabungkan metode k-means dengan metode *euclidean distance* agar dapat menghasilkan rekomendasi judul skripsi dapat di terima atau ditolak, hal tersebut dilakukan karena metode *K-Means clustering* mampu meningkatkan nilai akurasi moving average pada euclidean distance sebesar 9.92% [2]. Tujuan penelitian ini adalah untuk mengetahui sebaran bidang kajian dan metode yang pernah digunakan oleh mahasiswa sebagai bahan evaluasi untuk pengajuan judul skripsi berikutnya sehingga dapat lebih memudahkan bagian akademik. Hasil akhir dari penelitian ini diharapkan akan memberikan manfaat bagi pihak akademik khususnya sebagai bahan evaluasi pengajuan judul skripsi secara otomatis sehingga dapat lebih memudahkan mahasiswa dan dosen pembimbing.

II. METODE PENELITIAN

Penelitian ini menggunakan konsep data mining. Terdapat beberapa metode data mining yang dapat diterapkan pada penelitian diantaranya CRISP-DM, KDD, SEMMA dan TDSP. Metode penelitian yang diterapkan pada penelitian ini yaitu Cross-Industry Standard Process for Data mining (CRISP-DM). Tahapan penelitian pada CRISP-DM meliputi 6 langkah utama [3], tahapan-tahapan tersebut antara lain :

1. *Business Understanding*, pada tahap ini memerlukan pemahaman terhadap permasalahan yang terjadi pada objek penelitian, pemahaman terhadap teknik memperoleh data, dan mendeskripsikan tujuan yang akan dicapai pada akhir penelitian.
2. *Data understanding*, pada tahap ini dideskripsikan pemahaman terhadap data yang didapatkan pada tahap *Bussiness Understanding* diantaranya sumber data, karakteristik data, atribut yang terdapat pada data, dan jumlah record yang terdapat pada data.
3. *Data Preparation*, pada tahap ini terdapat beberapa hal yang dilakukan antara lain: memilih atribut yang akan digunakan untuk analisa (*Select Data*), melakukan perubahan terhadap struktur data atau atribut tertentu yang dapat disesuaikan untuk tahap modeling (*Transformation*), dan melakukan pembersihan data

apabila terdapat *value error* atau *value null* pada data (*Cleaning*) [4].

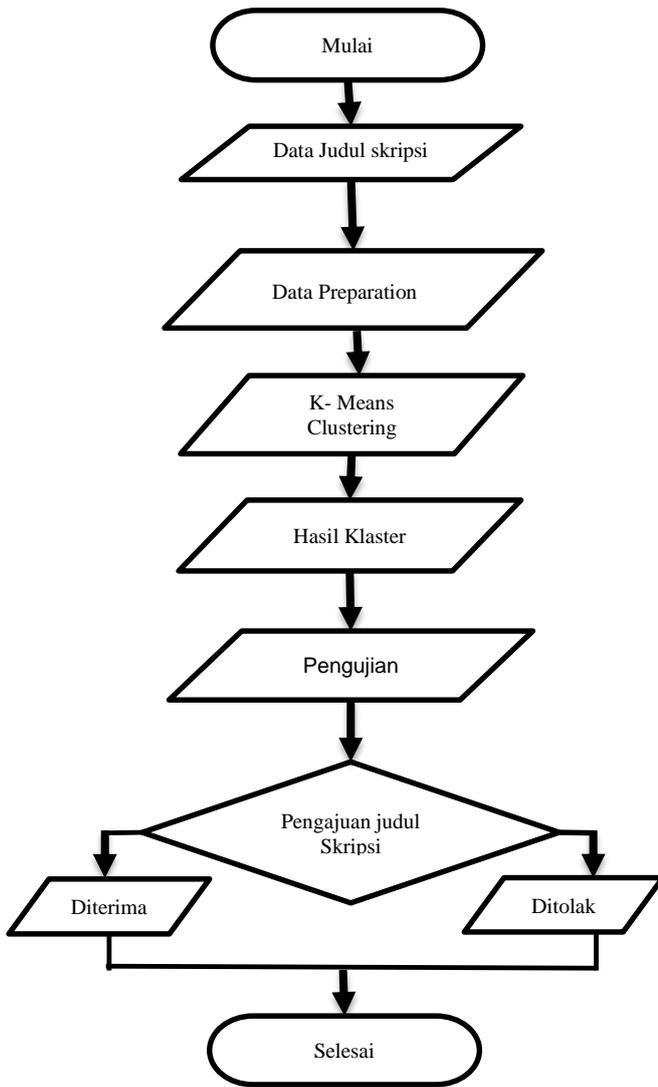
4. *Modeling*, Tahap ini dilakukan penerapan teknik dan algoritma data mining kepada data. Secara garis besar algoritma yang dapat digunakan meliputi model prediktif atau deskriptif.
5. *Evaluation*, Tahap ini dilakukan untuk menguji hasil dari data mining yang telah diselesaikan pada tahap *modeling*. Tahap ini bertujuan untuk mengevaluasi kualitas klaster yang dihasilkan pada tahap modeling dan memastikan sudah sesuai dengan harapan yang akan dicapai.
6. *Deployment*, Merupakan tahapan penggunaan hasil dari proses data mining

langkah-langkah penelitian dan penerapan CRISP-DM yang dilakukan sesuai dengan alur yang terdapat pada gambar 1 .

A. Teknik Pengambilan Data

Untuk penelitian ini, data yang digunakan adalah data judul skripsi mahasiswa program studi sistem informasi di salah satu universitas mulai dari tahun 2019-2022. Proses persiapan data dalam penelitian ini dilakukan dengan mengelompokkan judul skripsi yang diperoleh berdasarkan topik, metode, dan bidang penelitian. Dari proses persiapan data tersebut diperoleh 59 data latih yaitu data judul skripsi mahasiswa tahun 2019 sampai tahun 2022. Alur penelitian dapat dilihat pada gambar 1. Adapun tahapan alur penelitian adalah :

1. *Pengumpulan data judul skripsi*
Pada tahap ini dilakukan pengumpulan data dari perpustakaan yang berisi judul skripsi mahasiswa, juga dilakukan pemahaman terhadap data tersebut.
2. *Data Preparation*
Pada tahap ini terdapat beberapa hal yang dilakukan antara lain: menentukan atribut yang akan digunakan untuk analisa (*select data*) yakni atribut judul dilakukan persiapan data (*data preparation*) dilanjutkan dengan mentransformasikan data judul menjadi atribut metode dan bidang kajian. Setelah itu dilakukan proses normalisasi yang bertujuan mengubah data kategori menjadi numerik agar dapat diterapkan pada tahap modeling.
3. *K-Means Clustering*
Pada tahap ini dilakukan penerapan teknik dan algoritma K-Means dengan atribut judul, metode, dan bidang kajian.
4. *Hasil klaster*
Pada tahap ini telah diperoleh hasil klaster dari penerapan teknik dan algoritma K-Means
5. *Pengujian*
Tahap ini dilakukan untuk menguji kualitas klaster yang telah diperoleh. Setelah diketahui hasil kualitas klaster selanjutnya dapat digunakan sebagai rekomendasi judul skripsi dapat diterima atau ditolak



Gambar 1 Flow Alur Penelitian

B. K-Means Clustering

K-Means merupakan algoritma yang membagi data menjadi kelompok (cluster). Jumlah kelompok didefinisikan terlebih dahulu oleh pengguna sesuai kebutuhan, Tujuannya adalah membentuk kelompok sehingga data dalam satu kelompok memiliki kemiripan yang tinggi. Pada penelitian sebelumnya, k-means dapat melakukan proses pengelompokan [5]. Pada penelitian ini terdapat beberapa judul skripsi yang sama. Oleh karena itu. K-Means clustering akan membantu mengeliminasi judul skripsi yang sebaiknya direkomendasikan. Berikut adalah tahapan dalam Algoritma K-Means

1. Menentukan K sebagai jumlah cluster yang ingin dibentuk. [6]
2. Menentukan record sebagai titik pusat kluster awal (centroid) sejumlah K. [7]
3. Menghitung jarak masing-masing data yang dimasukkan pada centroid dengan *Euclidean Distance* sampai teridentifikasi jarak terdekat masing-masing data dengan centroid. Berikut ini

adalah persamaan *Euclidean Distance* [8]:

$$(x_i, y_j) = \sqrt{(x_i - y_j)^2} \dots (1)$$

Keterangan :

x_i = Data kriteria

y_j = Centroid pada cluster j

4. Menentukan kelompok tiap data berdasarkan jarak yang terkecil dengan titik pusat. [9]
5. Memperbaharui nilai centroid [10], diperoleh dari rata-rata kluster yang berkaitan berdasarkan rumus pada persamaan 2.

$$y_i(t + 1) = \frac{1}{N_{sj}} \sum_{j \in s_j} x_j \dots (2)$$

Keterangan:

$(t+1)$ = Centroid baru pada iterasi ke t + 1

N_{sj} = Banyaknya data pada cluster j

6. Mengulangi langkah 2 sampai langkah 5 hingga anggota tiap kluster tetap [11].
7. Apabila langkah 6 sudah terpenuhi, maka nilai titik pusat kluster pada iterasi (langkah) terakhir akan dijadikan sebagai titik pusat baru pada iterasi selanjutnya.

C. Silhoutte Coefficient

Metode uji kluster yang diterapkan pada penelitian ini yaitu *Silhoutte Coefficient*. Kluster yang telah diperoleh akan diuji kualitasnya dengan membandingkan rerata objek pada satu kluster dan jarak objek pada kluster lainnya [12]. Terdapat 3 langkah yang dilakukan pada *Silhoutte Coefficient*, antara lain:

1. Pada tiap objek ke-i, dihitung rerata jarak objek ke-i dengan semua objek yang terdapat pada satu kluster dan akan diperoleh nilai rerata a_i .
2. Pada tiap objek ke-i, dihitung rerata jarak dari objek ke-i dengan objek yang terdapat pada kluster lain. Berdasarkan jarak rerata tersebut diambil nilai terkecil dan disebut dengan b_i
3. Selanjutnya akan dihitung nilai *Silhoutte Coefficient* dari objek i menggunakan rumus pada persamaan 3.

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \dots (3)$$

Keterangan

a_i : Rerata jarak objek ke-i terhadap objek-objek dalam satu kluster.

b_i : Rerata jarak objek ke-i terhadap objek-objek di luar kluster.

Rentang nilai pada *Silhouette Coefficient* terletak pada nilai -1 sampai 1. Rentang nilai Silhouette Coefficient akan menunjukkan jarak kedekatan atau kemiripan terhadap data yang telah dikelompokkan pada suatu kluster. Apabila rerata *Silhouette Coefficient* mendekati

1, maka kluster tersebut dinilai semakin baik dan sebaliknya. Kriteria kualitas kluster berdasarkan nilai *Silhouette Coefficient* dapat dilihat pada tabel 1. [13]

Tabel 1. Kriteria Sillhouette coefficient

Nilai <i>Silhouette Coefficient</i>	Keterangan
$0,7 < SC \leq 1$	Strong structure
$0,5 < SC \leq 0,7$	Medium structure
$0,25 < SC \leq 0,5$	Weak structure
$SC \leq 0,25$	No structure

III. HASIL DAN PEMBAHASAN

Berdasarkan observasi yang telah dilakukan pada objek penelitian, dapat dijabarkan hasil dan pembahasan yang diperoleh adalah sebagai berikut:

1. *Bussiness Understanding*

Salah satu syarat utama kelulusan mahasiswa adalah dengan menyelesaikan tugas akhir dalam bentuk skripsi. Sebelum proses penyelesaian skripsi terlebih dahulu diperlukan pengajuan judul yang akan diangkat pada penelitian. Semakin bertambahnya judul skripsi yang tersimpan pada *repository* perpustakaan menyulitkan pihak akademik untuk menentukan judul mana yang dapat direkomendasikan dan tidak direkomendasikan. Dengan adanya penelitian ini diharapkan dapat memudahkan dalam pemberian rekomendasi pengajuan judul skripsi mahasiswa.

2. *Data Understanding*

Data yang diperoleh adalah berupa data judul skripsi mahasiswa yang terdiri dari 107 data dan 3 atribut yang terdiri dari : no, judul, kelas, dan topik penelitian. Data judul skripsi dapat dilihat pada tabel 2.

Tabel 2. Data awal judul skripsi

No	Judul	Kelas	Topik
1	Penerapan simple additive weighting (SAW) untuk menentukan prioritas peminjam uang karyawan di yayasan kiranana era prima	D	Information And Managerial Decision Making
2	Penerapan metode K-Means untuk rekomendasi merk sepatu yang diminati pelanggan di toko H.Uci	A	Business Intelligence - Support System
3	Penerapan metode SAW untuk menunjang keputusan dalam menentukan pegawai terbaik di kebun raya	D	Information And Managerial Decision Making
4	Pengembangan sistem informasi keanggotaan menggunakan metode prototype di komunitas chelsea indonesia supporters club (CISC)	D	Computer Supported Cooperative Work
5	Penerapan SAW untuk memilih perangkat komputer yang tersedia di katalog elektronik	D	Information And Managerial

Decision Making

...
106	Penerapan metode simple additive weighting (SAW) untuk rekomendasi kenaikan jabatan pegawai contact center	C	Information And Managerial Decision Making

3. *Data Preparation*

Pada tahap ini telah dilakukan persiapan terhadap data yang digunakan, diantaranya *data selection* dan *data transformation*. Pada tahap *data selection* dipilih data yang sesuai dengan permasalahan pada penelitian yaitu judul skripsi dan topik penelitian. Pada proses *data transformation* dilakukan penambahan 6 atribut yaitu metode, topik, bidang penelitian, *numeric* metode, *numeric* topik, dan *numeric* bidang. Atribut *numeric* metode, *numeric* topik, dan *numeric* bidang merupakan proses mengubah value atribut *categorical* metode, topik, dan bidang penelitian menjadi value numerik. Proses transformasi value *numeric* ini dilakukan dengan cara normalisasi yaitu menghitung frekuensi masing-masing kategori lalu dibagi dengan seluruh jumlah data. Hasil pengolahan data awal setelah melalui *tahap data preparation* dapat dilihat pada tabel 3.

Tabel 3. Data judul skripsi setelah preparation

No	Judul	Topik
1	Penerapan simple additive weighting (SAW) untuk menentukan prioritas peminjam uang karyawan di yayasan kiranana era prima	Information And Managerial Decision Making
2	Penerapan metode K-Means untuk rekomendasi merk sepatu yang diminati pelanggan di toko H.Uci	Business Intelligence - Support System
3	Penerapan metode SAW untuk menunjang keputusan dalam menentukan pegawai terbaik di kebun raya	Information And Managerial Decision Making

Tabel 3. Lanjutan kolom

Metode	Bidang	Numeric Topik	Numeric Metode	Numeric Bidang
SAW	Keuangan	0,651	0,387	0,028
K-Means	Ekonomi	0,245	0,066	0,151
SAW	Kepegawaian	0,651	0,387	0,179

4. *Modelling*

Pada tahap modeling dilakukan klusterisasi menggunakan algoritma K-Means menggunakan atribut *numeric* topik, *numeric* metode, dan *numeric* bidang dimana untuk menghitung jarak data dengan data lainnya menggunakan rumus *similarity measurement* dengan *Euclidean Distance*. Hasil dari tahap modeling ini adalah berupa pengelompokan judul skripsi yang termasuk ke dalam kluster Data

Mining atau DSS. Tahapan proses modeling adalah sebagai berikut:

- 1) Diawali dengan menentukan titik pusat awal data (centroid) sesuai dengan jumlah klaster yang diperlukan dengan cara membagi jumlah seluruh data (N) dengan klaster-n+1 (Centroid = N/klaster n +1) [14]. Pada penelitian ini akan dibentuk menjadi 2 klaster, sehingga ditentukan titik pusat datanya adalah:
 - a. Titik pusat 1 = $(106/1+1) = 53$, artinya data ke 53 akan dijadikan titik pusat untuk C2 (Klaster 2) dengan data pada tabel adalah:

Tabel 4. Menentukan Titik pusat 1

Numeric Topik	Numeric Metode	Numeric Bidang
0,651	0,387	0,057

- b. Titik pusat 2 = $(106/2+1) = 53$, artinya data ke 53 akan dijadikan titik pusat untuk C1 (Klaster 1) dengan data pada tabel adalah:

Tabel 5. Menentukan titik pusat 2

Numeric Topik	Numeric Metode	Numeric Bidang
0,651	0,038	0,179

- 2) Menghitung jarak seluruh data dengan titik pusat menggunakan *Euclidean Distance* dengan rumus pada persamaan 1. Contoh menghitung jarak data ke-1 dengan centroid adalah:
 - a. Jarak data ke-1 dengan C1

$$= \sqrt{(0,651 - 0,651)^2 + (0,387 - 0,038)^2 + (0,028 - 0,179)^2}$$

$$= 0,151$$
 - b. Jarak data ke-1 dengan C2

$$= \sqrt{(0,651 - 0,651)^2 + (0,387 - 0,387)^2 + (0,151 - 0,057)^2}$$

$$= 0,094$$
- 3) Hasil perhitungan jarak data ke-1 dengan C1 dan C2 menunjukkan jarak dengan C2 lebih kecil sehingga data ke-1 dikelompokkan dengan C2.
- 4) Menghitung kembali centroid menggunakan rumus pada persamaan 2.
- 5) Tahap ini dilakukan hingga didapatkan hasil jarak semua data dengan centroid C1 dan C2 lalu dikelompokkan sesuai dengan hasil perhitungan.

Hasil pengelompokkan data dapat dilihat pada tabel 4.

Tabel 4. Hasil pengelompokkan data

Data ke -	C1	C2
1		1
2	1	
3		1
4	1	
5		2
...

Jumlah populasi data pada masing-masing cluster adalah 37 data pada C1 (Klaster 1) dan 69 data pada C2 (Klaster 2).

5. *Evaluation*
 Hasil klaster yang diperoleh dari tahap modeling akan dilakukan evaluasi dengan menggunakan metode evaluasi cluster dengan *Silhouette Coefficient* dengan rumus persamaan 3 dan setelah dihitung didapatkan hasil *Silhouette Score* sebesar 0,62. Berdasarkan tabel 1, *silhouette score* ini termasuk ke dalam *cluster medium structure*.
6. *Deployment*
 Setelah diketahui hasil evaluation tahap selanjutnya yaitu *deployment* dengan menerapkan hasil modeling pada prototype dengan menggunakan bahasa pemrograman python.

NO	Judul	Kelas	Topik	Metode	Objek Penelitian	
0	1	Penerapan simple additive weighting (SAW) untu...	D	Information And Managerial Decision Making	SAW	Peminjaman Uang
1	2	Penerapan metode K-Means untuk rekomendasi mer...	A	Business Intelligence - Support System	K-Means	Rekomendasi Merk Sepatu
2	3	Penerapan metode SAW untuk menunjang keputusan...	D	Information And Managerial Decision Making	SAW	Pegawai Terbaik
3	4	Pengembangan sistem informasi keanggotaan meng...	D	Computer Supported Cooperative Work	Prototype	Komunitas Bola
4	5	Penerapan SAW untuk memilih perangkat komputer...	D	Information And Managerial Decision Making	SAW	Perangkat Komputer
5	6	Penerapan simple additive weighting (SAW) untu...	D	Information And Managerial Decision Making	SAW	Penentuan Pegawai

Normal Metode	Normal Case	Normal Bidang	Normal Topik
0.386792	0.037736	0.028302	0.650943
0.066038	0.198113	0.150943	0.245283
0.386792	0.367925	0.179245	0.650943
0.056604	0.066038	0.028302	0.056604
0.386792	0.113208	0.056604	0.650943
0.386792	0.367925	0.179245	0.650943

Gambar 2. Data awal setelah preprocessing

Pada data awal dilakukan *feature selection* untuk menentukan atribut yang akan digunakan untuk analisa dan dapat dilihat pada gambar 3.

	Normal Metode	Normal Case	Normal Bidang	Normal Topik
0	0.386792	0.037736	0.028302	0.650943
1	0.066038	0.198113	0.150943	0.245283
2	0.386792	0.367925	0.179245	0.650943
3	0.056604	0.066038	0.028302	0.056604
4	0.386792	0.113208	0.056604	0.650943

Gambar 3. Data pemilihan atribut

Gambar 3 menampilkan atribut yang akan dilanjutkan pada perhitungan tahap modeling. Gambar 4 menunjukkan hasil klasterisasi yang telah melalui tahap modeling, ditunjukkan pada kolom cluster_id. Cluster_id 0 untuk kluster 1 dan cluster_id 1 untuk kluster 2.

cluster_id	Metode	Case	Bidang	Topik
0	0	0.251162	0.134400	0.650943
1	1	0.067568	0.174656	0.187914

Gambar 4. Centroid cluster

Gambar 4 menampilkan titik pusat (centroid) kluster 1 dan kluster 2. Selanjutnya pengelompokan hasil kluster setelah melalui tahap analisa dengan Algoritma K-Means menggunakan *Euclidean Distance*.

NMetode	NCase	NBidang	NTopik	cluster_id	
0	0.386792	0.037736	0.028302	0.650943	0
1	0.066038	0.198113	0.150943	0.245283	1
2	0.386792	0.367925	0.179245	0.650943	0
3	0.056604	0.066038	0.028302	0.056604	1
4	0.386792	0.113208	0.056604	0.650943	0

Gambar 5. Pengelompokan kluster 1 dan kluster 2

Gambar 5 menampilkan hasil klasterisasi data. Selanjutnya akan ditampilkan pengelompokan kluster berdasarkan judul skripsi dan metode yang digunakan, dapat dilihat pada gambar 6.

	Judul	Metode	cluster_id
0	Penerapan simple additive weighting (SAW) untu...	SAW	0
1	Penerapan metode K-Means untuk rekomendasi mer...	K-Means	1
2	Penerapan metode SAW untuk menunjang keputusan...	SAW	0
3	Pengembangan sistem informasi keanggotaan meng...	Prototype	1
4	Penerapan SAW untuk memilih perangkat komputer...	SAW	0

Gambar 6. Hasil kluster 1 dan 2 berdasarkan Judul dan Metode Penelitian

Gambar 6 menampilkan hasil klasterisasi yang dibagi menjadi 2 kluster, cluster_id 0 merupakan kluster 1 (C1) dan cluster_id 1 merupakan kluster 2 (C2).

Setelah didapatkan kelompok cluster pada data judul skripsi, langkah selanjutnya dapat divisualisasikan lebih terperinci berdasarkan pilihan atribut yang ingin ditampilkan sebagai acuan penentuan rekomendasi pengajuan judul, misalkan berdasarkan metode penelitian.

Pada Gambar 7 menampilkan metode pada cluster 1, dan Gambar 8 menampilkan metode pada cluster 2. Data ini menginformasikan bahwa metode yang sudah banyak digunakan sebaiknya tidak digunakan pada pengajuan judul skripsi baru dan pengajuan dapat ditolak dan sebaliknya.

SAW	40
AHP	12
Profile Matching	4
Fuzzy Tsukamoto	4
TOPSIS	3
Fuzzy Mamdani	1
Fuzzy Sugeno	1
Monte Carlo	1
Haversine	1
C45	1

Gambar 7. Metode yang digunakan pada cluster 1

Berdasarkan Gambar 7, maka metode yang digunakan pada judul skripsi selanjutnya sebaiknya menggunakan metode lain selain SAW dan AHP karena metode ini paling banyak yang menggunakan.

Naïve Bayes	9
K-Means	7
Prototype	6
C45	6
MPE	3
EOQ	2
Regresi Linear	1
Fuzzy C-Means	1
Apriori	1
SAW	1

Gambar 8. Metode yang digunakan pada cluster 2

Berdasarkan Gambar 8, maka metode yang digunakan pada judul skripsi selanjutnya sebaiknya menggunakan metode lain selain Naive bayes, K-Means, dan C.45 karena metode ini banyak yang menggunakan maka pengajuan dengan metode ini tidak direkomendasikan dan diputuskan untuk ditolak.

Contoh lain penerapan kluster ini dapat dilihat pada Gambar 9 dimana pada gambar ditampilkan pengelompokan kluster berdasarkan bidang penelitian pada kluster.

Kepegawaian	15
Pendidikan	14
Ekonomi	9
Teknologi	5
Sarana dan Prasarana	5
Geografis	5
Umum	4
Keuangan	3
Kesehatan	2
Pemerintahan	2
Name: Bidang Penelitian , dtype: int64	

Gambar 9. Bidang penelitian pada kluster 1

Pendidikan	15
Ekonomi	7
Kepegawaian	4
Sarana dan Prasarana	4
Sosial	2
Lomba	2
Teknologi	1
Pariwisata	1
Geografis	1
Name: Bidang Penelitian , dtype: int64	

Gambar 10. Bidang penelitian pada kluster 2

Berdasarkan Gambar 9 dan Gambar 10 dapat dilihat bahwa bidang penelitian yang paling banyak diajukan pada judul skripsi mahasiswa terkait dengan bidang penelitian yaitu bidang kepegawaian dan pendidikan. Oleh karena itu, untuk ajuan judul skripsi yang bidang penelitiannya termasuk pada bidang kepegawaian dan pendidikan maka pengajuan dengan bidang ini tidak direkomendasikan dan diputuskan untuk ditolak.

Visualisasi data dapat disesuaikan berdasarkan atribut lainnya menyesuaikan dengan judul yang diajukan. Apabila akan melakukan seleksi judul berdasarkan topik penelitian hasil visualisasi datanya dapat terlihat seperti pada Gambar 11 dan Gambar 12.

Information And Managerial Decision Making	69
Name: Topik , dtype: int64	

Gambar 11 Topik penelitian pada kluster 1

Business Intelligence -Support System	26
Computer Supported Cooperative Work	6
Computer Resource Allocation	5
Name: Topik , dtype: int64	

Gambar 12. Topik penelitian pada kluster 2

Dengan demikian, apabila dilihat dari beberapa contoh hasil pola sebaran visualisasi data yang ada mulai Gambar 7 sampai dengan Gambar 12, dapat dilihat bahwa apabila topik penelitian yang diajukan terkait dengan *Information and Managerial Decision Making*, metode yang diajukan SAW, dan bidang kepegawaian dan pendidikan maka akan masuk ke kluster 1 dan diputuskan untuk ditolak, dan diarahkan untuk menentukan topik, metode, dan bidang penelitian lain. Rekomendasi dalam contoh kasus lain, apabila topik penelitian yang diajukan terkait dengan *Computer Resource Allocation* maka ajuan judul ini termasuk ke dalam kluster 2, setelah itu dilihat kembali pada bidang penelitian pada kluster 2, sebaiknya diajukan bidang selain pendidikan dan ekonomi dan metode yang diajukan selain Naive Bayes dan K-Means.

IV. KESIMPULAN DAN SARAN

Berdasarkan permasalahan dan tujuan penelitian dapat disimpulkan bahwa:

1. Hasil akhir dari penerapan metode K-Means ini adalah untuk memberikan rekomendasi pengajuan judul skripsi mahasiswa.
2. Hasil perhitungan dengan algoritma K-Means terbentuk menjadi 2 kluster yaitu:

- a. Berdasarkan metode, kluster 1 menggunakan metode DSS dan kluster 2 Data Mining
- b. Berdasarkan bidang penelitian, kluster 1 *Information and Managerial Decision Making* metode yang dapat digunakan adalah metode DSS, serta kluster 2 populasi data lebih mengarah ke bidang *Bussiness Intelligence – Support System* dan metode yang dapat digunakan yaitu metode Data Mining pada kluster 2 di poin a.
- c. Jumlah data yang terbentuk pada kluster 1 sebanyak 67 data dan kluster 2 sebanyak 37 data.
- d. Kluster yang dihasilkan telah diuji nilai akurasi menggunakan *silhouette coefisient* dan menghasilkan nilai 0,62 yang mengartikan bahwa kualitas kluster ini termasuk ke dalam *medium structure*.
- e. Saran untuk penelitian selanjutnya gunakan metode lain misalkan NLP atau text mining agar dapat menghasilkan klasifikasi rekomendasi pengajuan judul lebih spesifik berdasarkan judul, metode, permasalahan dan topik penelitian.

DAFTAR PUSTAKA

- [1] C. S. D. B. Sembiring, L. Hanum, and S. P. Tamba, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Menentukan Judul Skripsi Dan Jurnal Penelitian (Studi Kasus Ftik Unpri)," *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 80–85, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2393.
- [2] E. Fernando, P. Mudjiraharjo, and M. Aswin, "Implementasi Pendekatan Collaborative Filtering Dan K-Means Clustering Pada Sistem Rekomendasi Mata Kuliah," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 2, pp. 84–91, 2022, doi: 10.33387/jiko.v5i2.4559.
- [3] R. Chapman, P., Clinton, J., Kerber., *CRISP-DM 1.0 Step-by-step data mining guide*, vol. 1. USA: SPSS, 2000.
- [4] L. U. Zogara, A. Sururi, and L. T. Ningrum, "Analysis of COVID-19 Information Based on Social Media Big Data Classification Using the K-Means Data Mining Method," *J. Sisfotek Glob.*, vol. 12, no. 1, p. 18, 2022, doi: 10.38101/sisfotek.v12i1.448.
- [5] R. D. Mustika, A. Zakir, S. Informasi, U. Harapan, and J. Skripsi, "JURNAL MEDIA INFORMATIKA [JUMIN] Implementasi Algoritma K-Means Untuk Clustering Judul Skripsi Universitas Harapan Medan JURNAL MEDIA INFORMATIKA [JUMIN]," vol. 4, pp. 40–47, 2022.
- [6] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru," *Sistemasi*, vol. 7, no. 3, p. 238, 2018, doi: 10.32520/stmsi.v7i3.388.
- [7] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan

Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik),” *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018, doi: 10.18196/st.211211.

- [8] Agus Nur Khormarudin, “Teknik Data Mining: Algoritma K-Means Clustering,” *J. Ilmu Komput.*, pp. 1–12, 2016, [Online]. Available: <https://ilmukomputer.org/category/datamining/>.
- [9] A. Sucipto, “Clasterization of Prospective Students,” *J. Sci. Tech*, vol. 5, no. 2, pp. 50–56, 2019.
- [10] R. Helilintar and I. N. Farida, “Penerapan Algoritma K-Means Clustering Untuk Prediksi Prestasi Nilai Akademik Mahasiswa,” *J. Sains dan Inform.*, vol. 4, no. 2, pp. 80–87, 2018, doi: 10.34128/jsi.v4i2.140.
- [11] L. T. Ningrum and D. Rahmiyati, “Analysis Sentiment of Twitter User on Indonesia’s 2024 Presidential Election Using K-Means Algorithm,” *J. Sisfotek Glob.*, vol. 13, no. 2, p. 102, 2023, doi: 10.38101/sisfotek.v13i2.9609.
- [12] S. ‘Aina Salsabila, T. Widiharah, and S. Sudarno, “METODE K-HARMONIC MEANS CLUSTERING DENGAN VALIDASI SILHOUETTE COEFFICIENT (Studi Kasus : Empat Faktor Utama Penyebab Stunting 34 Provinsi di Indonesia Tahun 2018),” *J. Gaussian*, vol. 11, no. 1, pp. 11–20, 2022, doi: 10.14710/j.gauss.v11i1.34003.
- [13] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *J. Comput. Appl. Math.*, vol. 20, no. C, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.
- [14] F. Nur, M. Zarlis, and B. B. Nasution, “Penerapan Algoritma K-Means Pada Siswa Baru Sekolahmenengah Kejuruan Untuk Clustering Jurusan,” *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 1, no. 2, pp. 100–105, 2017, doi: 10.30743/infotekjar.v1i2.70.