

**PENERAPAN METODE K-MEANS++ DALAM PENENTUAN
TOPIK SKRIPSI MAHASISWA BERDASARKAN
NILAI MATA KULIAH**

(Studi kasus: Mahasiswa Matematika Angkatan 2021 Universitas Tanjungpura)

RIAN PRASETIO

H1011201052

Skripsi

Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Matematika pada Program Studi Matematika

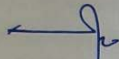


**PROGRAM STUDI MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025**

LEMBAR PENGESAHAN

Judul Tugas Akhir : Penerapan Metode K-Means++ Dalam Penentuan Topik
Skripsi Mahasiswa Berdasarkan Nilai Mata Kuliah
(Studi kasus: Mahasiswa Matematika Angkatan 2021
Universitas Tanjungpura)
Nama Mahasiswa : Rian Prasetio
NIM : H1011201052
Jurusan/Program Studi : Matematika/Matematika
Tanggal Lulus : 24 Januari 2025
SK Pembimbing : No. 700/UN22.8/TD.06/2024 Tanggal 01 Oktober 2024
SK Penguji : No. 64/UN22.8/TD.06/2025 Tanggal 22 Januari 2025

Pembimbing I



Yudhi, M.Si
NIP. 198504072019031004

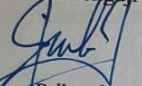
Dosen Pembimbing

Pembimbing II



Dr. Nilamsari Kusumastuti, M.Sc
NIP. 198105102005012003

Ketua Penguji



Dr. Bayu Prihandono, M.Sc
NIP. 197911152005011003

Dosen Penguji

Anggota Penguji



Dr. Yundari, M.Sc
NIP. 198310202008012012

Pimpinan sidang
(Merangkap anggota Penguji)



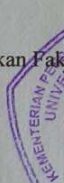
Yudhi, M.Si
NIP. 198504072019031004

Sekretaris sidang
(Merangkap anggota Penguji)



Dr. Nilamsari Kusumastuti, M.Sc
NIP. 198105102005012003

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Tanjungpura



Prof. Dr. Gusrizal, S.Si., M.Si
NIP. 197108022000031001

LEMBAR PERNYATAAN INTEGRITAS AKADEMIK

Saya yang bertanda tangan dibawah ini:

Nama : Rian Prasetyo
NIM : H1011201052
Program Studi/Jurusan : Matematika/Matematika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa dokumen ilmiah Tugas Akhir yang disajikan ini tidak mengandung unsur pelanggaran integritas akademik sesuai Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 39 Tahun 2021. Apabila di kemudian hari dokumen ilmiah Tugas Akhir ini Mengandung unsur pelanggaran integritas akademik sesuai ketentuan perundangan tersebut, maka saya bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Demikian pernyataan ini untuk digunakan sebagaimana mestinya.

Pontianak, 20 Januari 2025



Rian Prasetyo
H1011201052

Penerapan Metode K-Means++ Dalam Penentuan Topik Skripsi Mahasiswa Berdasarkan Nilai Mata Kuliah

Abstrak

Pemilihan topik skripsi merupakan langkah penting bagi mahasiswa dalam menentukan fokus utama penelitian yang akan dilakukan. Proses ini dapat menjadi tantangan karena melibatkan pencocokan antara keahlian, minat, dan potensi individu. Salah satu upaya yang dapat dilakukan untuk mengatasi permasalahan ini adalah membentuk klusterisasi. Penelitian ini bertujuan untuk menentukan pengelompokan topik skripsi berdasarkan nilai mata kuliah yang sudah diambil dan dilakukan uji validasi untuk merekomendasikan topik skripsi berdasarkan hasil clustering. Dalam penelitian ini, digunakan algoritma K-Means++ untuk mengelompokkan mahasiswa berdasarkan nilai akademik, khususnya rumpun ilmu yang menjadi rujukan pemilihan topik skripsi. Data yang diperoleh berasal dari akademik FMIPA Universitas Tanjungpura dengan menggunakan 43 data mahasiswa matematika 2021. Menggunakan metode Elbow untuk menentukan k cluster. Mencari centroid pertama secara acak dan centroid awal lain menggunakan algoritma K-Means++ dan menggeser centroid dengan iterasi hingga konvergen serta uji validasi menggunakan DBI. Dari hasil perhitungan diperoleh 4 cluster, dengan kombinasi dari cluster 1 menunjukkan minat pada statistika komputasi terapan, cluster 2 menunjukkan minat pada analisis komputasi terapan, cluster 3 menunjukkan minat pada komputasi dan matematika terapan dan cluster 4 menunjukkan minat pada analisis dan komputasi terapan. Dari hasil uji validasi, menggunakan uji Elbow dan DBI diperoleh bahwa nilai terbaik berada pada $k=4$ dengan nilai elbow menunjukkan penurunan dan nilai DBI pada $k=4$ adalah 1,15. Hal ini menunjukkan bahwa pengelompokan dengan 4 cluster memberikan kualitas clustering terbaik dibandingkan cluster lain.

Kata kunci: *Clusterisasi, Jarak Euclidean, pusat cluster, elbow, DBI*

Application of K-Means++ Method in Determining Thesis Topics Students Based on Course Grades

Abstract

Selecting a thesis topic is an important step for students in determining the main focus of their research. This process can be challenging because it involves matching individual skills, interests, and potential. One way to overcome this problem is to form clusters. This study aims to determine thesis topic clustering based on previously taken course grades and conduct validation tests to recommend thesis topics based on clustering results. In this study, the K-Means++ algorithm was used to cluster students based on academic performance, particularly the field of study that serves as the reference for thesis topic selection. The data obtained came from the academic department of Faculty of Mathematics and Science, Tanjungpura University, using 43 data points from mathematics students in 2021. The Elbow method was used to determine k clusters. The first centroid was randomly selected, and other initial centroids were determined using the K-Means++ algorithm, with centroids shifted through iterations until convergence. Validation was conducted using DBI. The results yielded 4 clusters: Cluster 1 showed interest in applied computational statistics, Cluster 2 in applied computational analysis, Cluster 3 in computation and mathematic applications, and Cluster 4 in both analysis and computational applications. From the validation test results, using the Elbow test and DBI, it was found that the best value was at $k=4$. with the elbow value showing a decrease and the DBI value at $k=4$ being 1.15. This indicates that grouping with 4 clusters provides the best clustering quality compared to other clusters.

Keyword: *Clustering*, Euclidean Distance, *centroid*, Elbow, Davies Bouldin Index

PRAKATA

Segala puji syukur penulis panjatkan kepada Allah SWT atas segala nikmat dan karunia-Nya, yang memungkinkan penulis menyelesaikan skripsi berjudul “Penerapan Metode *K-Means++* Dalam Penentuan Topik Skripsi Mahasiswa Berdasarkan Nilai Mata Kuliah” Skripsi ini disusun sebagai salah satu syarat untuk meraih gelar Sarjana Matematika (S.Mat) di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura. Penulis mengucapkan terima kasih atas seluruh bantuan, bimbingan dan dukungan kepada:

1. Orang tua penulis yang senantiasa memberikan doa, dukungan dan kasih sayang kepada penulis.
2. Bapak Yudhi, M.Si dan Ibu Dr. Nilamsari Kusumastuti, M.Sc selaku dosen pembimbing yang telah memberikan masukan, arahan dan ilmu-ilmunya untuk penulis dalam menyelesaikan skripsi ini.
3. Bapak Dr. Bayu Prihandono, M.Sc dan Ibu Dr. Yundari, M.Sc selaku Dosen Penguji yang telah memberikan masukan, saran, dan ilmu-ilmunya dalam penyusunan skripsi ini.
4. Tim *OpenAI* selaku pembuat *ChatGPT* dengan hasil ciptaannya telah memberikan pengetahuan kepada penulis dalam pengerjaan skripsi ini.
5. Lady Furina dan Hu Tao selaku *waifu* penulis yang memberikan motivasi penuh kepada penulis sehingga dapat menyelesaikan skripsi ini
6. Seluruh pihak yang telah membantu dan tidak dapat disebutkan satu persatu pada masa perkuliahan hingga akhir penelitian ini.

Semoga segala kebaikan yang telah diberikan mendapatkan balasan dan keberkahan dari Allah SWT. Saya menyadari bahwa skripsi ini belum sempurna karena keterbatasan ilmu yang saya miliki. Oleh karena itu, saya mengharapkan saran dan masukan dari seluruh pihak agar dapat lebih baik lagi.

Pontianak, 16 Januari 2025

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	Error! Bookmark not defined.
LEMBAR PERNYATAAN INTEGRITAS AKADEMIK.....	Error! Bookmark not defined.
Abstrak.....	iv
Abstact	v
PRAKATA.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN.....	xi
DAFTAR SIMBOL	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	3
1.4 Batasan Masalah	3
1.5 Tinjauan Pustaka	4
1.6 Metodologi penelitian	4
BAB II LANDASAN TEORI	8
2.1 Analisis <i>cluster</i>	8
2.2 Algoritma <i>K-means++</i>	8
2.3 <i>Euclidean distance</i>	11
2.4 <i>Elbow</i>	12
2.5 <i>Davies Bouldin Index</i>	12
BAB III PENERAPAN <i>K-MEANS++</i> PADA NILAI MATA KULIAH MAHASISWA	15
3.1 Pelabelan dan Kategorisasi Data.....	15
3.2 Penentuan Jumlah <i>K Cluster</i> Menggunakan Metode Elbow	17
3.3 Perhitungan <i>Cluster</i> dengan metode <i>K-Means++</i>	19
3.3 Perhitungan uji validasi menggunakan <i>Davies Bouldin Index (DBI)</i>	25

3.4 Interpretasi hasil <i>cluster</i> optimal.....	29
BAB IV KESIMPULAN.....	31
DAFTAR PUSTAKA.....	32
LAMPIRAN.....	34

DAFTAR TABEL

Tabel 3. 1 Pelabelan nama mahasiswa	15
Tabel 3. 2 Pelabelan mata kuliah	16
Tabel 3. 3 Kelompok rumpun ilmu pada mata kuliah	17

DAFTAR GAMBAR

Gambar 1. 1 Flowchart metode penelitian.....	7
---	---

DAFTAR LAMPIRAN

Lampiran I Label mahasiswa	34
Lampiran II Label mata kuliah.....	35
Lampiran III Data nilai mahasiswa	36
Lampiran IV Sintaksis python	40

DAFTAR SIMBOL

k	: Banyaknya <i>cluster</i>
K	: <i>Cluster</i> ke- k , untuk $k=1, 2, \dots, K$
i	: Banyaknya objek penelitian
j	: Banyaknya variabel peneltian
x_i	: Baris objek sebanyak i
$x_{i,j}$: Baris objek i pada variabel j
$dC_k x_i$: Jarak antara <i>centroid</i> pada <i>cluster</i> k dengan objek x_i
C_k	: <i>Centroid</i> pada <i>cluster</i> k
$P_k(x_i)$: Probabilitas x_i sebagai <i>centroid</i> baru

BAB I

PENDAHULUAN

1.1 Latar Belakang

Program Studi Matematika merupakan disiplin ilmu yang memiliki penerapan luas di berbagai sektor. Bidang ini mencakup beragam cabang, seperti aljabar, analisis, statistika, dan matematika terapan. Program studi ini dirancang untuk memberikan pemahaman teoritis yang mendalam sekaligus mengasah keterampilan analitis, logis, dan *problem-solving* yang tinggi. Mahasiswa didorong untuk mengintegrasikan konsep-konsep matematika dalam menyelesaikan berbagai tantangan nyata, mulai dari teknologi dan sains hingga pengembangan inovasi yang berdampak luas.

Dalam proses pendidikan di Program Studi Matematika, salah satu tahapan penting yang menjadi akhir pendidikan adalah tugas akhir (TA) atau skripsi. Hal ini sangat krusial karena menjadi kesempatan bagi mahasiswa untuk mengaplikasikan pengetahuan dan keterampilan yang telah diperoleh selama masa studi. Namun, meskipun terlihat sederhana, proses ini sering menjadi tantangan besar bagi mahasiswa. Menurut penelitian yang dilakukan oleh (Kurniawan, 2023), banyak mahasiswa merasa kesulitan menentukan topik karena kurangnya pemahaman mengenai bidang penelitian yang sesuai kompetensi mereka. Dalam konteks program studi matematika, hal ini terjadi karena cakupan bidang penelitian yang sangat luas, mulai dari teori murni seperti aljabar dan analisis hingga aplikasi matematika dalam berbagai disiplin ilmu lainnya. Dengan adanya cakupan bidang penelitian yang beragam, mahasiswa akan kebingungan untuk memilih topik yang tidak hanya sesuai dengan minat mereka, tetapi juga mencakup aspek-aspek yang mereka kuasai secara mendalam. Beberapa mahasiswa merasakan belum cukup memiliki dasar yang kuat di beberapa cabang tertentu atau kesulitan dalam memahami kaitan antara teori dan aplikasinya di dunia nyata (Turmuzi, 2021).

Tantangan ini juga diperburuk oleh tekanan psikologis, seperti rasa takut akan kegagalan atau kekhawatiran bahwa topik yang dipilih tidak sebanding dengan penguasaan materi (Lestari 2021). Hal ini diperkuat oleh temuan (Nugraheni 2022), yang menyebutkan bahwa 64% mahasiswa mengalami kebingungan dalam

menentukan topik karena merasa tidak percaya diri terhadap kemampuan mereka untuk menyelesaikan penelitian di bidang tersebut.

Permasalahan utama yang dihadapi mahasiswa program studi matematika adalah bagaimana cara mengidentifikasi kompetensi mahasiswa secara objektif untuk menentukan topik skripsi yang tepat. Setiap mahasiswa memiliki latar belakang, minat, dan kemampuan yang berbeda. Nilai mata kuliah yang telah ditempuh mahasiswa dapat menjadi indikator potensial dalam menggambarkan minat dan kompetensi tersebut.

Masalah ini telah menjadi perhatian utama bagi lembaga pendidikan, khususnya perguruan tinggi. Untuk mengatasi masalah ini, diperlukan sebuah solusi yang efektif dan efisien. Salah satu metode yang dapat digunakan adalah analisis *clustering*, yang dapat membantu mengelompokkan data dalam kategori-kategori yang berbeda. Dalam analisis *clustering*, *K-means++* merupakan salah satu metode yang dapat digunakan dalam analisis *clustering*. Metode ini memiliki kelebihan dalam menangani dataset yang besar dan memiliki waktu komputasi yang lebih cepat.

Diharapkan dari hasil dari penelitian ini, dapat memberikan kontribusi dalam meningkatkan kualitas pendidikan dan dapat meminimalisir masalah lama waktu tempuh kelulusan mahasiswa. Selain itu, penelitian ini juga dapat menjadi referensi bagi institusi pendidikan dalam memilih metode yang tepat dalam mengatasi masalah serupa di masa depan. Diharapkan melalui penelitian ini, masalah tersebut dapat diatasi dan memberikan dampak positif bagi dunia pendidikan.

1.2 Rumusan Masalah

Berdasarkan latar belakang, rumusan masalah dari penelitian ini adalah bagaimana menerapkan metode *K-Means++* untuk *clustering* mahasiswa berdasarkan nilai mata kuliah mereka sehingga dapat membantu dalam penentuan topik skripsi yang sesuai.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang dikemukakan, maka tujuan yang akan dicapai dari penelitian ini adalah:

1. Menentukan pengelompokkan berdasarkan nilai mata kuliah pada hasil algoritma *K-Means++*.
2. Menentukan nilai uji validasi terbaik untuk menentukan cluster optimal untuk merekomendasikan topik skripsi mahasiswa berdasarkan hasil *clustering* dengan metode *Davies Bouldin Index* (DBI).

1.4 Batasan Masalah

Agar Pembahasan ini lebih terarah dan tidak meluas maka diberikan batasan permasalahan dalam penelitian ini yaitu:

1. Data yang diolah adalah data transkrip nilai mahasiswa matematika angkatan 2021 yang dikumpulkan pada bulan juli 2024. Data transkrip yang digunakan adalah nilai mahasiswa aktif dari semester 1 hingga semester 6 (tidak mengundurkan diri).
2. Pengukuran jarak yang digunakan pada metode *K-Means++* adalah jarak *Euclidean*.
3. Mata kuliah yang digunakan adalah mata kuliah dengan label kode “MPM” dengan kelompok rumpun ilmu analisis dan geometri, aljabar, matematika terapan dan komputasi yang telah ditentukan oleh pedoman akademik FMIPA Universitas Tanjungpura tahun 2020.
4. Mata kuliah yang akan diolah adalah mata kuliah wajib dan mata kuliah pilihan. Untuk mata kuliah wajib mahasiswa harus memperoleh nilai dari semua kelas yang telah ditentukan oleh pedoman akademik FMIPA Universitas Tanjungpura tahun 2020. Untuk mata kuliah pilihan diambil dari mahasiswa yang pernah mengambil mata kuliah pilihan. Mata kuliah pilihan yang diambil akan menggunakan rata-rata yang akan digabung dalam satu variable yang mewakili mata kuliah pilihan rumpun ilmu.
5. Interpretasi hasil *centroid* pada masing-masing *cluster* akan dilakukan rata-rata nilai seluruh mata kuliah setiap rumpun ilmu.

1.5 Tinjauan Pustaka

Penelitian yang dilakukan oleh (Nugroho, 2022) dalam jurnal "Teknika" mengkaji penggunaan algoritma *K-Means* dan *K-Means++* untuk mengelompokkan data persebaran *Covid-19* di Pulau Jawa. Hasil penelitian menunjukkan bahwa *K-Means++* memberikan hasil yang lebih akurat dalam mengelompokkan data kasus *Covid-19* dibandingkan dengan *K-Means*. Hal ini ditunjukkan oleh nilai uji *Silhouette Coefficient*. Hasil penelitian ini adalah *K-Means++* menghasilkan *cluster* yang lebih baik apabila nilai k lebih besar yaitu $K = 5$. Sedangkan, *K-Means* lebih baik jika nilai k yang digunakan lebih kecil, yaitu ketika $K = 4$.

Penelitian *clustering* pada kasus kriminal yang dilakukan oleh (Fastaf 2022) mengenai pemetaan kasus kriminal di kabupaten Bangkalan menggunakan *K-Means* dan *K-Means++* memiliki hasil\ berdasarkan uji *Silhouette Coefficient* diperoleh bahwa *K-Means++* memiliki hasil yang lebih baik pada kasus penganiayaan dan penipuan dibandingkan *K-Means* yang hanya pada kasus perampokan, dengan tujuh kasus lainnya memiliki hasil yang sama.

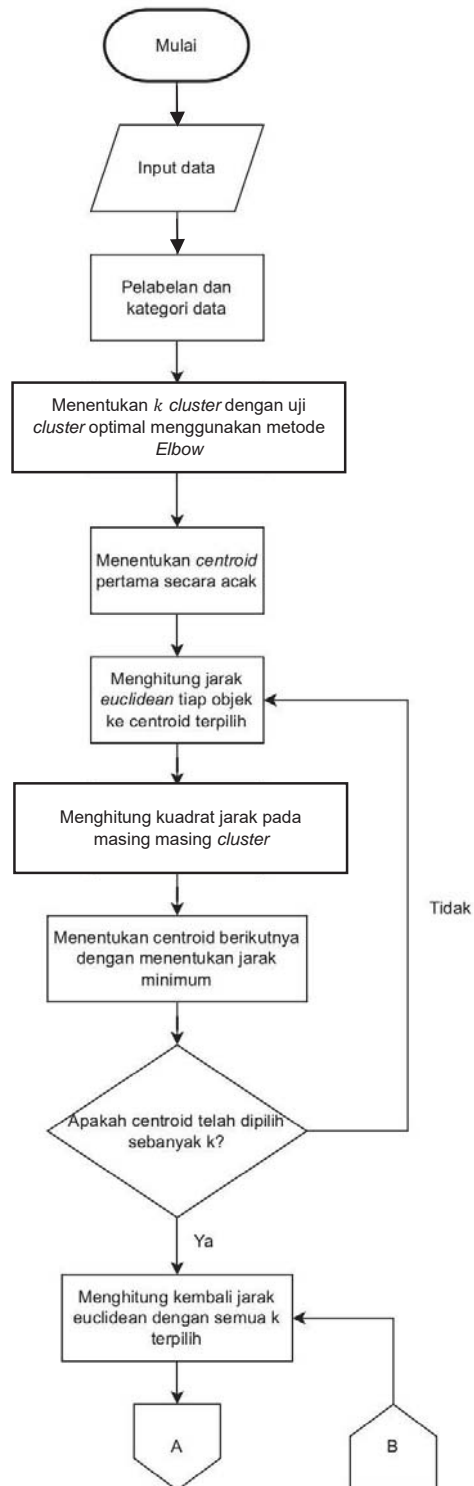
Penelitian *clustering* pada kasus potensi drop out yang dilakukan oleh (Laili 2023) pada mahasiswa IAIN Kediri tahun akademik 2016/2017 dengan *K-Means++*. Hasil penelitian ini diperoleh bahwa *centroid* terbaik adalah 3 yaitu potensi drop out rendah, sedang, dan tinggi. Untuk *centroid* $K=3$ memiliki nilai uji *Silhouette Coefficient* tertinggi dibandingkan *Clustering* lain.

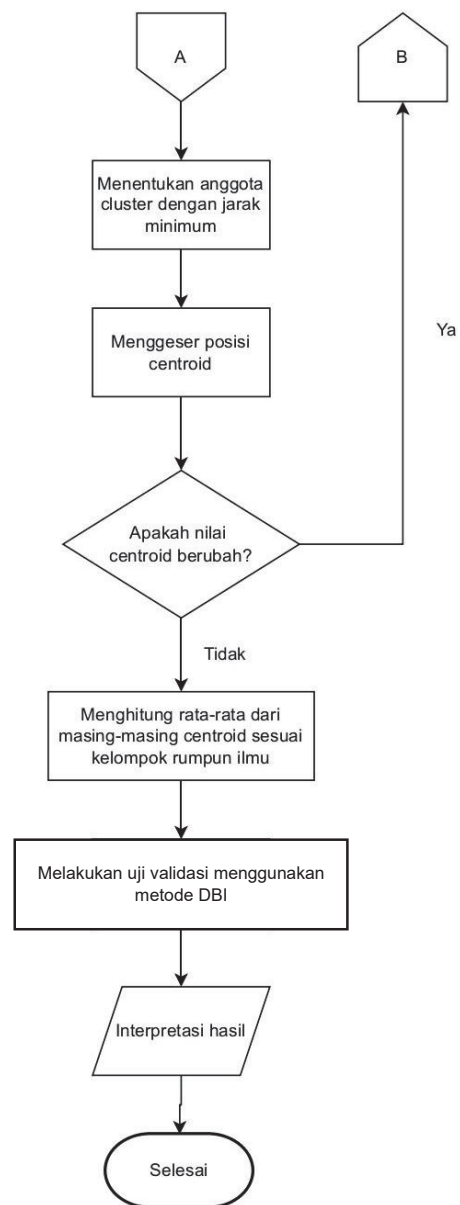
1.6 Metodologi penelitian

Alat bantu analisis yang akan digunakan dalam penelitian ini adalah bahasa pemrograman *python* untuk membantu pengelompokkan peminatan topik skripsi berdasarkan nilai mata kuliah. Berikut ini adalah tahapan analisis data yang digunakan dalam pengelompokan dengan algoritma *K-Means++* dan evaluasi hasil *clustering* dengan *elbow* dan *Davies Bouldin Index*:

1. Analisis pertama yaitu tahap Pengumpulan Data. Mengumpulkan data nilai mata kuliah mata kuliah dari mahasiswa yang akan menjadi penelitian.

2. Menentukan *cluster* awal yang akan dibentuk. Proses ini akan dilakukan dengan menentukan jumlah titik *cluster* awal yang diinginkan sebanyak n *cluster*. Pada algoritma *k-means++* penentuan *centroid* pertama yaitu menentukan satu titik *cluster* secara acak dan hasil perhitungan *centroid* awal berikutnya akan diambil dari jarak maksimum dari setiap objek data dengan *centroid* yang ada untuk menentukan *centroid* berikutnya dan akan terus berulang hingga mencapai *centroid* sebanyak K .
3. Menghitung ulang jarak antara *centroid* awal yang diperoleh sebanyak k dengan menggunakan jarak *Euclidean*. Proses ini akan menggeser posisi titik *cluster* dan anggota *cluster* dengan menggunakan iterasi. Proses iterasi ini akan berhenti jika masing masing *cluster* tidak memiliki perubahan titik pusat *cluster* dan jumlah anggota *cluster*. Proses iterasi dapat dianggap selesai.
4. Mencari *cluster* optimal dan uji akurasi dengan menggunakan metode *elbow* dan *Davies Bouldin Index* (DBI). Langkah pertama dalam penentuan *cluster* optimal yaitu meminimumkan jarak antar setiap titik *cluster* dengan menghitung nilai *Sum of Square Within* (SSW). Langkah selanjutnya menghitung nilai *Sum of Square Between* (SSB) untuk memaksimumkan perbedaan antar setiap *cluster*. Kemudian menghitung nilai rasio dengan menggunakan nilai SSW dan SSB yang sudah diperoleh sebelumnya untuk mendapatkan perbandingan antara satu *cluster* dengan *cluster* lainnya. Setelah mendapatkan nilai rasio, selanjutnya menghitung nilai DBI. Langkah selanjutnya mencari nilai DBI terkecil dari masing-masing jumlah *cluster* yang telah digunakan. Hal tersebut dikarenakan, jumlah *cluster* optimal dilihat dari nilai DBI terkecil.
5. Setelah diperoleh jumlah *cluster* optimal, selanjutnya melakukan interpretasi dari setiap *cluster* dan membuat kesimpulan dari hasil pengelompokan yang diperoleh. Penelitian ini dilakukan dengan menggambarkan alur penyelesaian yang jelas dengan langkah-langkah yang diperlukan melalui diagram alir yang digambarkan sebagai berikut.





Gambar 1. 1 Flowchart *clustering K-Means* dan uji validasi *Elbow* dan DBI dalam menentukan topik skripsi dari nilai mata kuliah