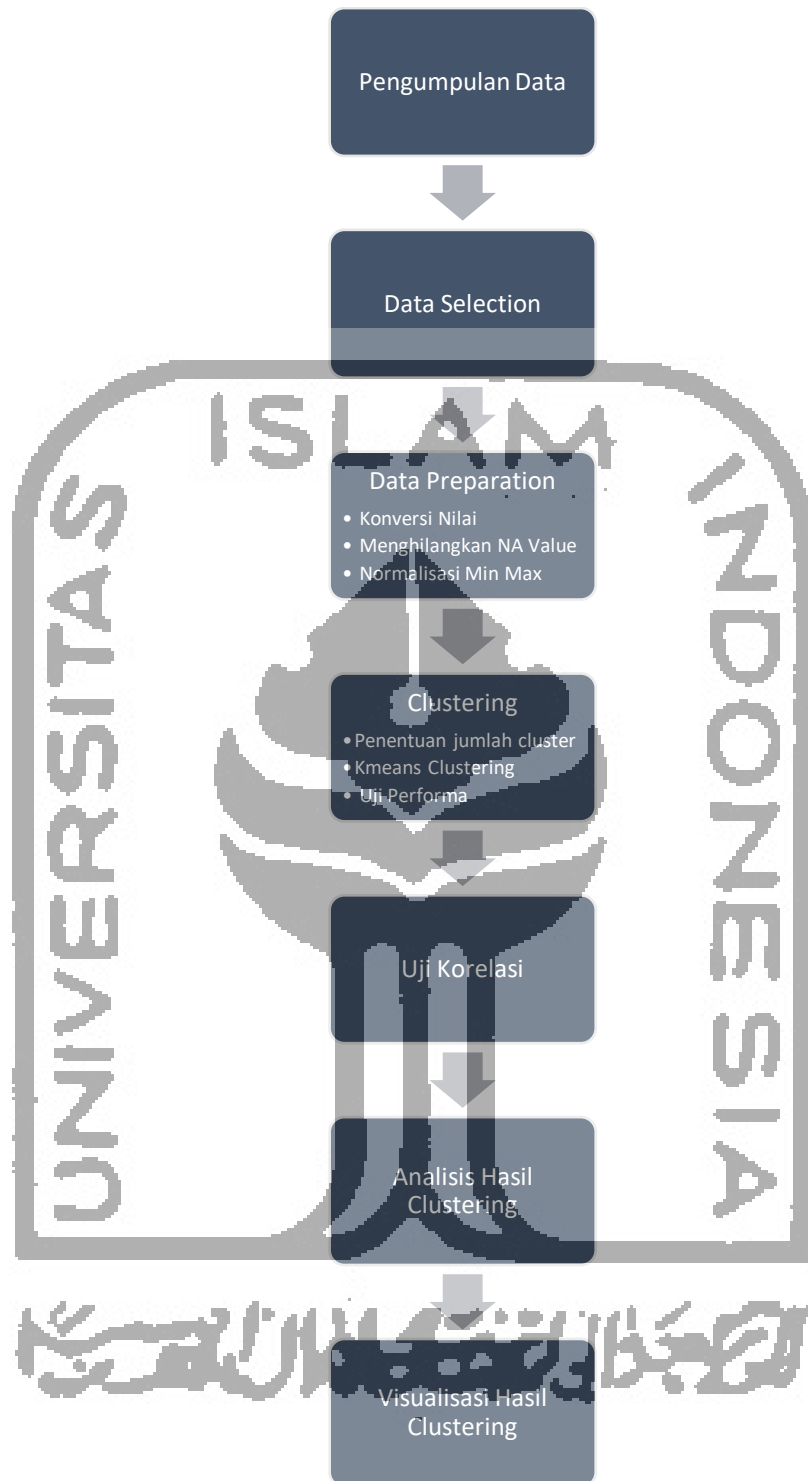


BAB III METODOLOGI PENELITIAN

3.1 Langkah Penyelesaian Tugas Akhir

Pada Gambar 3.1 Langkah Penyelesaian merupakan langkah-langkah yang dilakukan untuk menyelesaikan penelitian.





Gambar 3.1 Langkah Penyelesaian Analisis

3.2 Uraian Metodologi

Berikut di bawah ini merupakan uraian metodologi berdasarkan langkah – langkah penyelesaian yang terdapat pada Gambar 3.1.

3.2.1 Pengumpulan Data

Tahapan pengumpulan data ini dilakukan secara langsung dengan mengambil data langsung melalui Unisys yang merupakan sistem informasi akademik UII dan adapun data yang didapatkan dari BSI (Badan Sistem Informasi) UII. Data yang digunakan adalah data yang disimpan pada *file* excel. Data terbagi menjadi dua yaitu:

- a. Data PMB jurusan Teknik Informatika tahun angkatan 2016, 2017, dan 2018 yang didapatkan dari BSI UII.
- b. Data hasil capaian semester mahasiswa tahun angkatan 2016, 2017, dan 2018 yang didapatkan dari Unisys.

Data dikumpulkan menggunakan metode primer, dengan meminta data dari BSI UII dan juga mengumpulkan data setiap mahasiswa yang ada pada Unisys.. Data PMB merupakan data yang berisi tentang informasi mahasiswa berdasarkan jenis tes yang diikuti saat tes penerimaan mahasiswa baru dan nilai tes yang didapatkan yang berkisar dari 0 sampai dengan 20. Berikut ini merupakan variabel yang terdapat pada data PMB:

1. Nomor Induk Mahasiswa (NIM)
2. Asal Sekolah Menengah Atas (SMA)
3. Kota asal SMA
4. Provinsi asal SMA
5. Jenis SMA
6. Jenis tes PMB
7. Nilai tes agama
8. Nilai tes Bahasa Inggris
9. Nilai tes logika
10. Nilai tes matematika

Sedangkan untuk data hasil capaian mahasiswa juga berisi beberapa variabel, yaitu:

1. Nomor Induk Mahasiswa (NIM)
2. Nilai-nilai mata kuliah yang telah dijalani
3. Indeks Prestasi Semester (IPS)
4. Indeks Prestasi Kumulatif (IPK)

3.2.2 Data Selection

Pada proses ini dilakukan penyeleksian data yang bertujuan untuk mempermudah analisis. Dari data yang telah dikumpulkan akan dipilih variabel yang berkaitan dengan skenario yang akan diterapkan pada analisis yaitu data numerik.

3.2.3 Data Preparation

Setelah seleksi data akan dilakukan proses *data preparation*. Proses ini bertujuan untuk menyiapkan data untuk mempermudah analisis. Pada tahapan ini terdapat menghilangkan NA *value* dan normalisasi min-max.

Konversi Nilai

Konversi nilai ini dilakukan untuk menyamakan nilai tes dan nilai hasil capaian menjadi bentuk yang sama berupa angka. Hal ini dilakukan secara manual pada *file excel*. Konversi ini dilakukan sesuai dengan acuan yang telah ditetapkan oleh Universitas Islam Indonesia, seperti pada Tabel 3.1:

Tabel 3.1 Tabel Konversi Nilai Mata Kuliah

Nilai Huruf	Nilai Angka	Nilai Huruf	Nilai Angka
A	4,00	C+	2,25
A-	3,75	C	2,00
A/B	3,50	C-	1,75
B+	3,25	C/D	1,50
B	3,00	D+	1,25
B-	2,75	D	1,00
B/C	2,50	E dan F	0,00

Menghilangkan NA Value

Proses menghilangkan NA (Not Available) value ini dilakukan karena algoritma *K-Means* tidak dapat melakukan *clustering* terhadap nilai kosong, maka dari itu adanya NA *value* perlu diatasi. Pada kasus ini akan dijalankan sebuah fungsi untuk menghapus nilai kosong yang ada. Penghapusan dilakukan secara menyeluruh untuk satu baris data yang terdapat NA *value* di dalamnya. Karena hal ini maka akan terjadi pengurangan jumlah baris data yang ada.

Normalisasi Min-Max

Normalisasi data dilakukan untuk menyamakan nilai – nilai pada data atau standarisasi data. Pada tugas akhir ini akan digunakan normalisasi min-max dengan rentang nilai 0 sampai 1. Normalisasi pada tugas akhir ini dilakukan bertujuan untuk menyamakan nilai yang ada pada data PMB yang memiliki rentang nilai dari 0 sampai dengan 100 dan nilai hasil capaian mahasiswa yang memiliki rentang nilai dari 0 sampai dengan 4.

Normalisasi pada tugas akhir ini membantu untuk analisis hasil klasterisasi data, karena hasil klasterisasi dari data PMB dan data hasil capaian akademik mahasiswa akan dibandingkan dengan beberapa skenario yang ada. Sehingga akan didapatkan hasil analisis klasterisasi yang menunjukkan pola performa mahasiswa.

3.2.4 Clustering

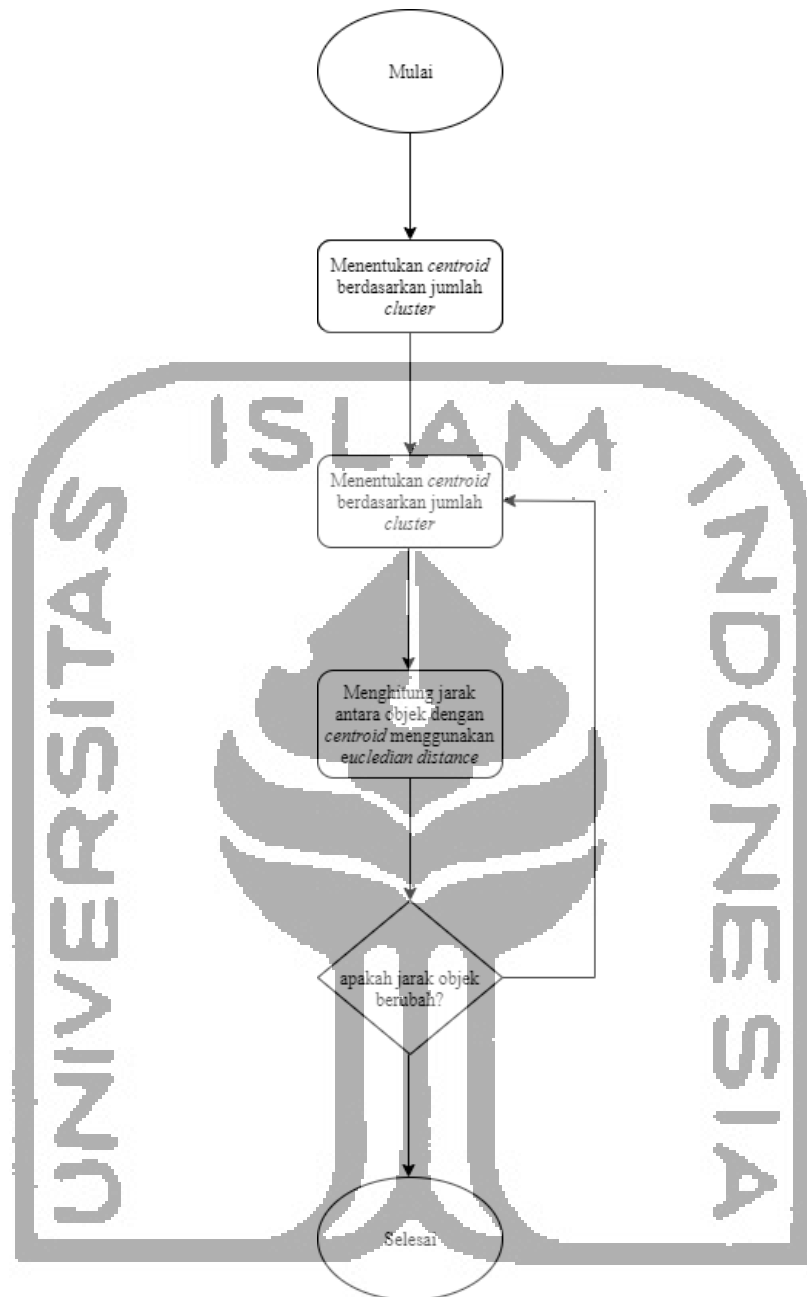
Tahapan yang dilakukan setelah melakukan normalisasi adalah klasterisasi data. Metode klasterisasi yang akan digunakan adalah metode *K-Means*. Untuk menunjukkan persebaran data dari setiap *cluster* juga perlu adanya uji performa *clustering* sebagai validasi apakah *clustering* yang dilakukan sudah baik atau belum.

Penentuan Jumlah Cluster

Penentuan jumlah *cluster* yang akan digunakan untuk proses *clustering* akan ditentukan berdasarkan hasil dari metode *elbow* dan *silhouette*. Diharapkan dengan penentuan jumlah *cluster* dengan metode-metode tersebut akan didapatkan jumlah *cluster* yang tepat.

Kmeans Clustering

Pada penelitian ini akan dilakukan *clustering* dengan menggunakan algoritma *K-Means*. Seperti pada **Error! Reference source not found.** yang menunjukkan proses kerja *K-Means*.



Gambar 3.2 Proses Kerja K-Means

Klatisasi dilakukan terhadap data tes PMB dan data hasil capaian mahasiswa dengan menggunakan beberapa skenario, yaitu:

- Skenario satu: menyandingkan variabel jenis sekolah dan jenis tes PMB yang terhadap hasil capaian mahasiswa pada semester 1.
- Skenario dua: melakukan clustering terhadap data tes PMB dan data hasil capaian mahasiswa semester 1 yang telah digabungkan.

Proses klusterisasi yang dilakukan akan menghasilkan kelompok mahasiswa berdasarkan variabel – variabel yang terdapat pada data. Hasil klusterisasi antara data tes PMB dan hasil capaian mahasiswa akan disimpan menjadi sebuah file untuk kebutuhan tahapan selanjutnya yaitu analisis hasil klusterisasi. *Clustering* akan dilakukan dengan menggunakan fungsi Algoritma *K-Means* yang telah disediakan oleh R tools.

Uji Performa

Uji performa ini dilakukan untuk mengetahui kualitas dari persebaran data yang telah dihasilkan dari proses *clustering*. Akan digunakan sebuah fitur *within sum of square* yang terdapat dalam fungsi *K-Means* yang akan menunjukkan sebuah nilai yang menunjukkan seberapa dekatnya titik objek dengan objek lain yang berada pada satu *cluster* yang sama.

3.2.5 Uji Korelasi

Pada tahapan ini akan dicari hubungan variabel antar variabel yang digunakan untuk proses clustering. Uji korelasi ini akan menunjukkan kedekatan makna atau korelasi dari satu variabel dengan variabel lainnya.

3.2.6 Analisis Hasil Klusterisasi

Tahapan yang dilakukan setelah klusterisasi adalah analisis hasil yang didapatkan. Analisis dilakukan untuk mengetahui bagaimana pola performa mahasiswa berdasarkan hasil klusterisasi. Dengan cara menyandingkan hasil klusterisasi antara data tes PMB dan data hasil capaian mahasiswa berdasarkan dengan skenario yang ada. Hal – hal yang diperhatikan adalah nilai *centroid* yang menjadi acuan titik persebaran objek, bagaimana ciri-ciri dari setiap kluster yang dapat dilihat dari nilai yang terdapat pada tiap *centroid* dengan melihat pada bidang apa yang lebih dominan.

Dengan melakukan analisis tersebut akan ditunjukkan pola performa mahasiswa dan akan diketahui jumlah mahasiswa dengan kelompok yang sama. Ketika analisis dilakukan dengan menyandingkan data tes PMB dan data hasil capaian mahasiswa ini kemungkinan akan ditunjukkan mahasiswa yang stabil, menurun dan juga meningkat pada performanya.

3.2.7 Visualisasi Hasil Klusterisasi

Tahapan akhir dari analisis ini adalah visualisasi hasil dari klusterisasi yang telah dilakukan. Tujuan dari visualisasi adalah untuk mempermudah dalam pemahaman hasil analisis klusterisasi. Visualisasi akan dilakukan dengan menggunakan *web framework open source* dari R yaitu R shiny. Visualisasi akan ditunjukkan pada halaman *dashboard*.

Hasil klasterisasi yang divisualisasikan berupa tabel data hasil *clustering* dan diagram-diagram yang dapat membantu user untuk memahami hasil dari analisis yang telah dilakukan berdasarkan data yang ada yaitu data tes PMB dan data hasil capaian mahasiswa yang sudah diproses dengan tools R sebelumnya.

