

BAB III

METODE PENELITIAN

Bagian dalam penelitian berisi penjelasan tentang langkah-langkah secara detail dalam penyelesaian masalah pada penelitian yang dilakukan.

3.1. Desain Penelitian

Metode yang digunakan dalam penelitian ini adalah metode eksperimen, yaitu menerapkan metode data mining dengan algoritma *K-means Clustering* dan algoritma *K-Medoids Clustering* dengan cara membandingkan akurasi yang dihasilkan sehingga dapat diketahui algoritma yang lebih efektif dalam menangani media promosi program studi teknik informatika Universitas Islam Nahdlatul Ulama Jepara. Tahap-tahap penelitian ini adalah sebagai berikut:

A. Pengumpulan Data.

Pada tahap ini adalah mengumpulkan data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua data ke data set, termasuk variabel yang diperlukan dalam proses kemudian diseleksi dari data yang tidak sesuai.

B. Pengolahan Awal Data

Pada tahap ini dijelaskan tentang tahap awal persiapan data olah. Pengolahan awal data meliputi proses input data ke format yang dibutuhkan, pengelompokkan dan penetapan atribut data.

C. Model yang di usulkan

Model yang diusulkan dalam penelitian ini adalah penerapan Algoritma *K-means Clustering* dan Algoritma *K-Medoids Clustering*.

D. Eksperimen dan Pengujian Data

Untuk eksperimen data penelitian, penulis menggunakan aplikasi *Rapidminer 8.1.1* dan *Microsoft Excel 2016* sebagai pendukung dalam pengolahan data.

E. Evaluasi dan Validasi Hasil

Pada bagian ini dilakukan evaluasi dan validasi hasil penerapan terhadap model yang dilakukan untuk mengetahui tingkat keakurasian model.

3.2. Metode Pengumpulan Data

Pengumpulan data adalah tahapan terpenting pada metode eksperimen. Proses pengumpulan data dilakukan dengan cara merekapitulasi laporan data penerimaan mahasiswa baru teknik informatika UNISNU Jepara yang diambil dari tahun angkatan 2016 sampai tahun angkatan 2019 yang di dapat dari UPT Pusat Data dan Pengembangan IT Universitas Islam Nahdlatul Ulama Jepara. Selanjutnya langkah-langkah yang dilakukan adalah menyusun data-data yang akan diolah termasuk parameter dan alternatif yang akan diranking, merancang dan memproses perhitungan menggunakan algoritma *k-means* dan *k-medoids* untuk mendapatkan nilai pengelompokan mahasiswa. Hasil proses yang telah dinyatakan valid selanjutnya dilakukan analisis yang dikaitkan dengan pengelompokan dari parameter lainnya.

3.3. Tahapan Pengolahan Awal Data

Jumlah data awal yang diperoleh dari pengumpulan data yaitu sebanyak 491 data, tetapi tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan awal (*preparation data*). Untuk mendapat data yang berkualitas, beberapa tahapan yang dilakukan adalah sebagai berikut [21]:

1. Remove duplicate, menghapus nilai-nilai yang bernilai sama.
2. Data validation, untuk mengidentifikasi dan menghapus data yang ganjil(*noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*).
3. Data integration dan transformation, untuk meningkatkan akurasi dan efisiensi algoritma.

Data yang digunakan dalam penelitian ini bernilai kategorikal. Data di transformasikan ke dalam aplikasi Rapidminer.

3.4. Model yang Diusulkan.

Model yang diusulkan pada penelitian ini adalah algoritma *K-means* dan algoritma *K-medoids*.

Tabel 3.1. Sampel Data Calon Mahasiswa Baru Teknik Informatika.

No	NAMA CALON MAHASISWA	JENIS KELAMIN	ASAL KECAMATAN	ASAL SEKOLAH	NILAI UN
1	AKHMAD TOHA	P	BATEALIT	SMK MUHAMMADIYAH 02 JEPARA	77
2	MUHAMMAD UMAR	P	MIJEN	MA YPKM RADEN FATAH	75
3	TAUFIQ HIDAYAT	P	PAKIS AJI	SMK NEGERI 1 PAKIS AJI JEPARA	87.2
...
489	MOHAMMAD ALDI SETIAWAN	P	JEPARA	SMK NEGERI 3 JEPARA	75
490	MUHAMMAD MIFTAKHUR ROZAK ALFARID	P	MLONGGO	SMK NAWA KARTIKA MLONGGO	81
491	M. ANDIS MAULANA AL MUBAROK	P	PECANGAAN	SMK WALISONGO PECANGAAN	50

3.4.1. Algoritma K-Means Clustering.

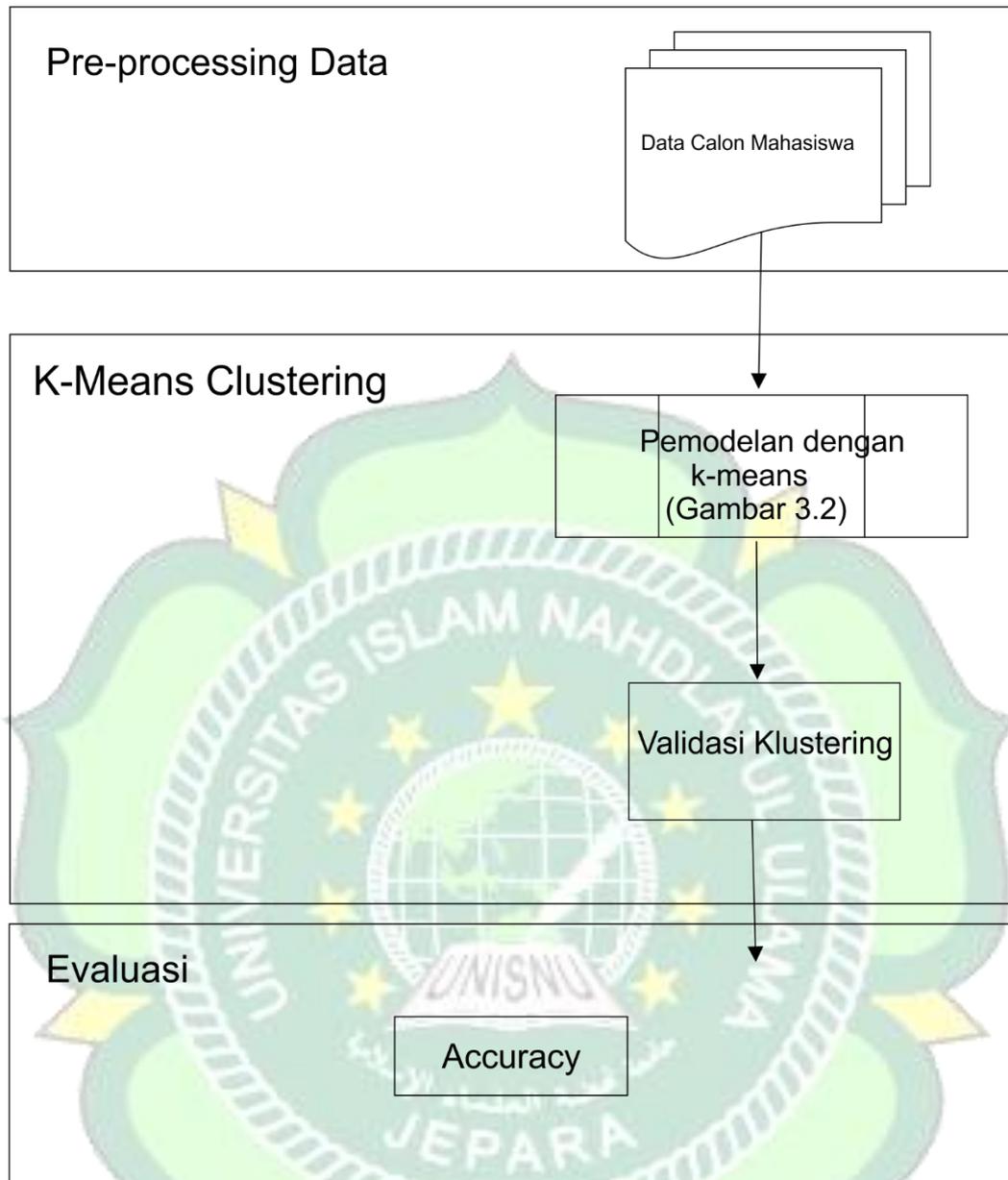
1. Transformasi Data

Transformasi data dilakukan untuk mengubah data supaya bisa diolah menggunakan *Algoritma K-means Clustering*, data berjenis text seperti Jenis Kelamin, Asal Sekolah, dan Asal Kecamatan harus dilakukan proses inialisasi ke dalam bentuk angka atau numerik.

2. Pengolahan Data

Dataset yang dimasukkan pada aplikasi Rapidminer merupakan dataset yang telah di buat dengan aplikasi Microsoft Excel, kemudian file tersebut dipakai untuk penelitian ini. Setelah file dimasukkan ke Rapidminer, tentukan masing-masing indikator atribut yang terdapat didalamnya. Penelitian ini tipe atributnya adalah *polynomial*, karena semua atribut berisi lebih dari dua klasifikasi.

Adapun proses pengolahan data secara detail ditunjukkan pada Gambar 3.1. Bentuk klustering prediksi strategi promosi calon mahasiswa baru dalam penelitian ini dilakukan dengan teknik *smart clustering*, yakni teknik pendekatan data mining dengan mengelompokkan data mahasiswa. *K-means* adalah salah satu metode data *non-hierarchical clustering* yang dapat mengelompokkan data mahasiswa ke dalam beberapa *cluster* berdasarkan kemiripan dari data tersebut. Sehingga data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster* dan yang memiliki karakteristik yang berbeda dikelompokkan dalam *cluster* yang lain [22].



Gambar 3.1. Diagram Proses Pengolahan Data Algoritma *K-means Clustering* [22].

Pengolahan data diawali dengan tahapan *Pre-processing* data, yakni suatu langkah yang dilakukan untuk membuat data mentah menjadi data yang berkualitas, dengan input yang baik untuk data mining tools. Jika data masukan tidak berkualitas, maka hasil data mining juga tidak berkualitas. Dalam tahap ini, akan dilakukan analisis distribusi guna melihat *plot* distribusi data pada *variable input* maupun *output*. Tahapan ini juga merupakan tahap untuk melakukan normalisasi pada data. Data mentah yang didapatkan dikonversi menjadi bentuk data yang dibutuhkan sehingga siap untuk diolah. Tahapan ini meliputi proses seleksi data yang

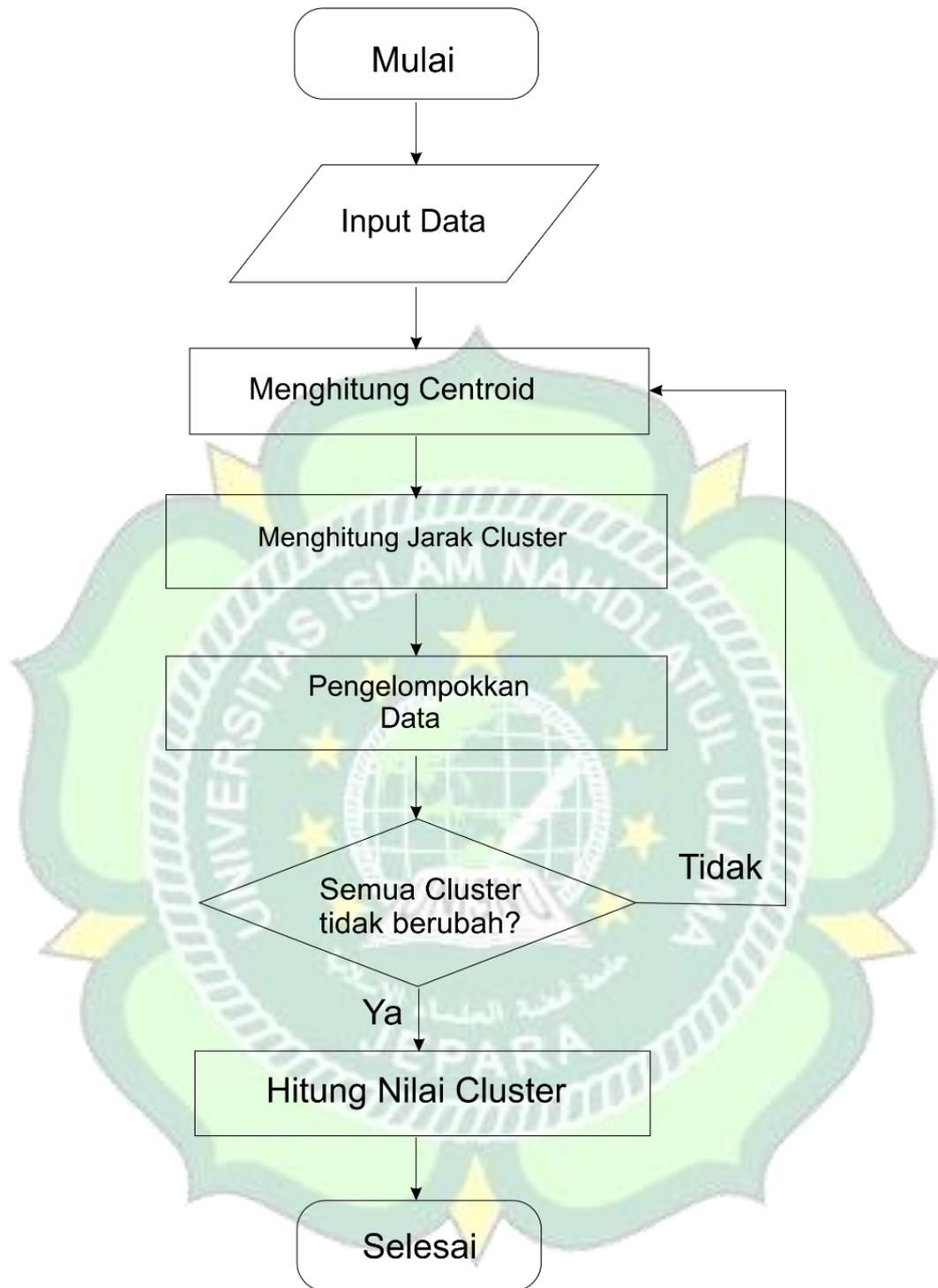
akan digunakan sebagai variabel masukan untuk proses klustering dengan uji kolerasi pada data [22].

Tahapan kedua adalah tahap pemodelan data menggunakan implementasi algoritma *K-means Clustering*. File yang sudah dimasukkan akan mengalami proses pengujian, dimulai dengan dataset Read *Excel*. Kemudian di hubungkan dengan *Set Role* untuk memilih atribut sebagai *Role*. Setelah itu disambungkan dengan algoritma yang di pilih, penelitian ini menggunakan *Algoritma K-means Clustering*. Selanjutnya di hubungkan lagi dengan *Apply Model*, dan di hubungkan lagi dengan *Performance* dan yang terakhir di hubungkan ke *Result* [9].

Setelah proses pengujian selesai maka akan di dapatkan beberapa informasi seperti lama waktu pemrosesan, menunjukkan deskripsi performance yang menghasilkan centroid distance pada setiap cluster, menunjukkan jumlah cluster yang dihasilkan, dapat dilihat juga *Davies Bouldin Index* (DBI) yang ada di performance Vector *K-means*. DBI adalah metode validasi cluster dari hasil clustering. Cluster yang baik adalah semakin rendah nilai DBI yang dihasilkan [9]. Adapun proses detail diagram alur algoritma *K-means clustering* dapat dilihat pada Gambar 3.2.

3. Evaluasi dan Validasi

Tahap ketiga adalah proses evaluasi dengan menggunakan pengukuran evaluasi klustering yaitu pengukuran akurasi, sensitifity, dan juga spesifity dari algoritma *K-means Clustering*. Untuk proses evaluasi penelitian ini menggunakan *metode ROC*. *Metode ROC* merupakan metode untuk mengukur akurasi untuk *clustering*. Selain mengukur akurasi, metode ini juga bisa mengukur nilai *Sensitifity*, dan *Spesififity* Untuk validasi algoritma *Clustering* sudah dijelaskan sebelumnya bahwa metode validasi *clustering* menggunakan *Davies Bouldin Index* (DBI) [9].



Gambar 3.2. Flowchart Algoritma K-means Clustering [22]

3.4.2. Algoritma K-Medoids Clustering

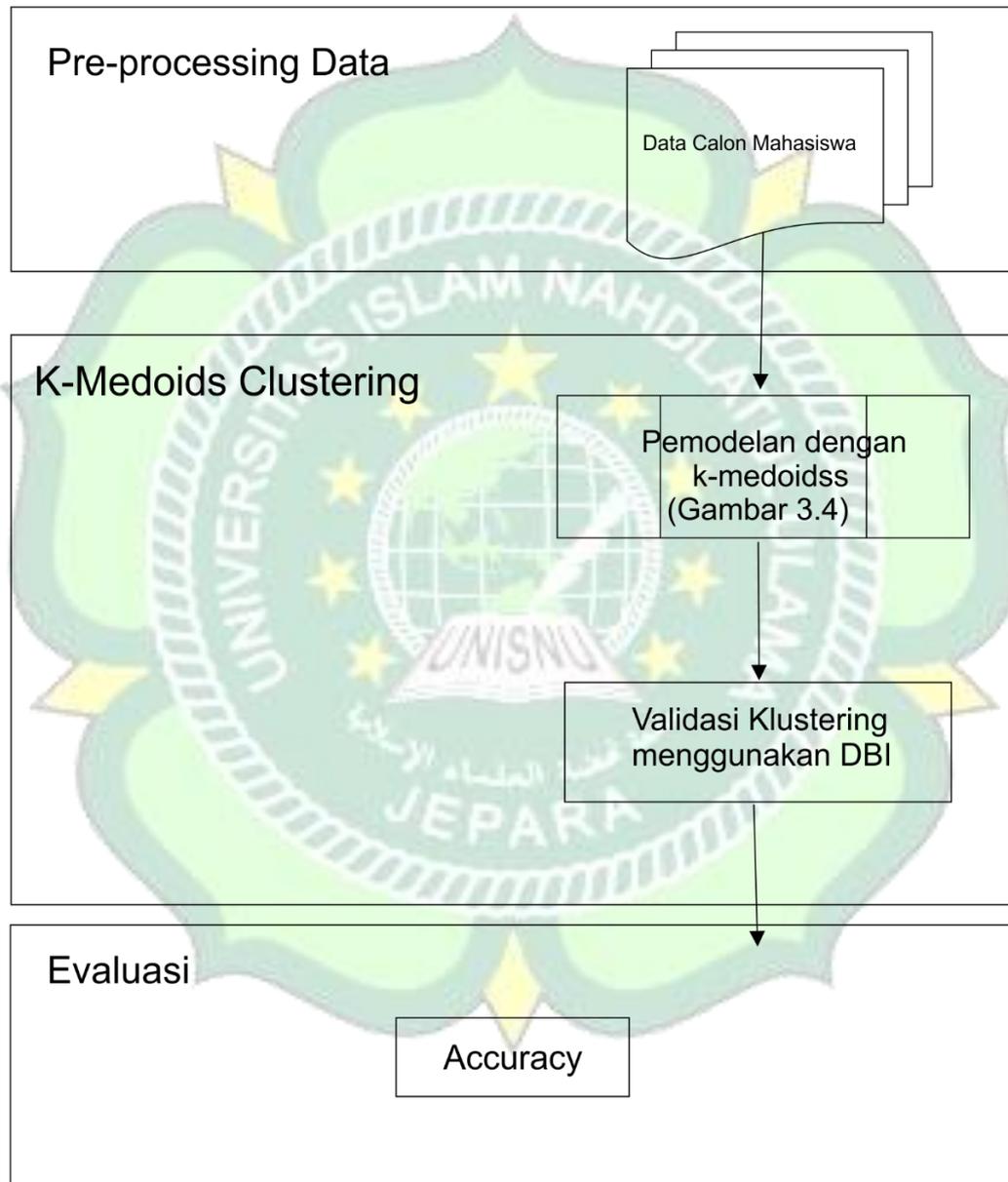
1. Pre-processing Data

Data pada tabel 3.1 yang didapat masih banyak yang *missing value*, oleh sebab itu perlu dilakukan *selection* dan *cleaning data*. *Preprocessing* adalah proses

pembersihan data yang tidak digunakan, data yang *missing value* dan normalisasi dalam perhitungan. Tahapan Preprocessing data terdapat dua proses yaitu *deletion attribute* dan *label transformation*, kemudian setelah tahap tersebut selesai maka akan terbentuk dataset yang baru.

2. Pengolahan Data

Adapun proses pengolahan data secara detail ditunjukkan pada Gambar 3.3.



Gambar 3. 3. Diagram Proses Pengolahan Data Algoritma *K-Medoids* [4].

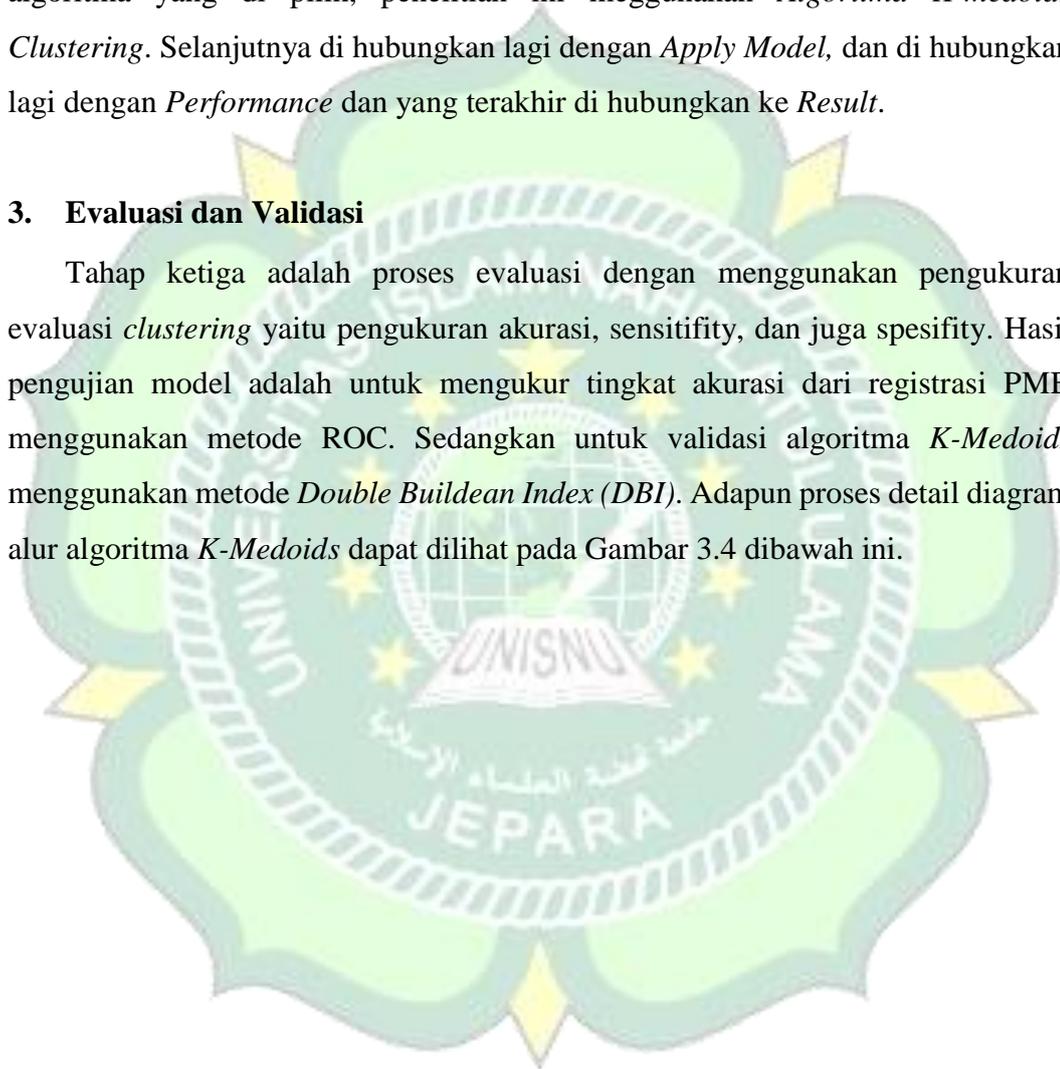
Pada tahap awal pengolahan data (*Preprocessing data*) yang sudah dijelaskan pada proses *Pre-processing data* algoritma *K-means Clustering*. Peneliti melakukan

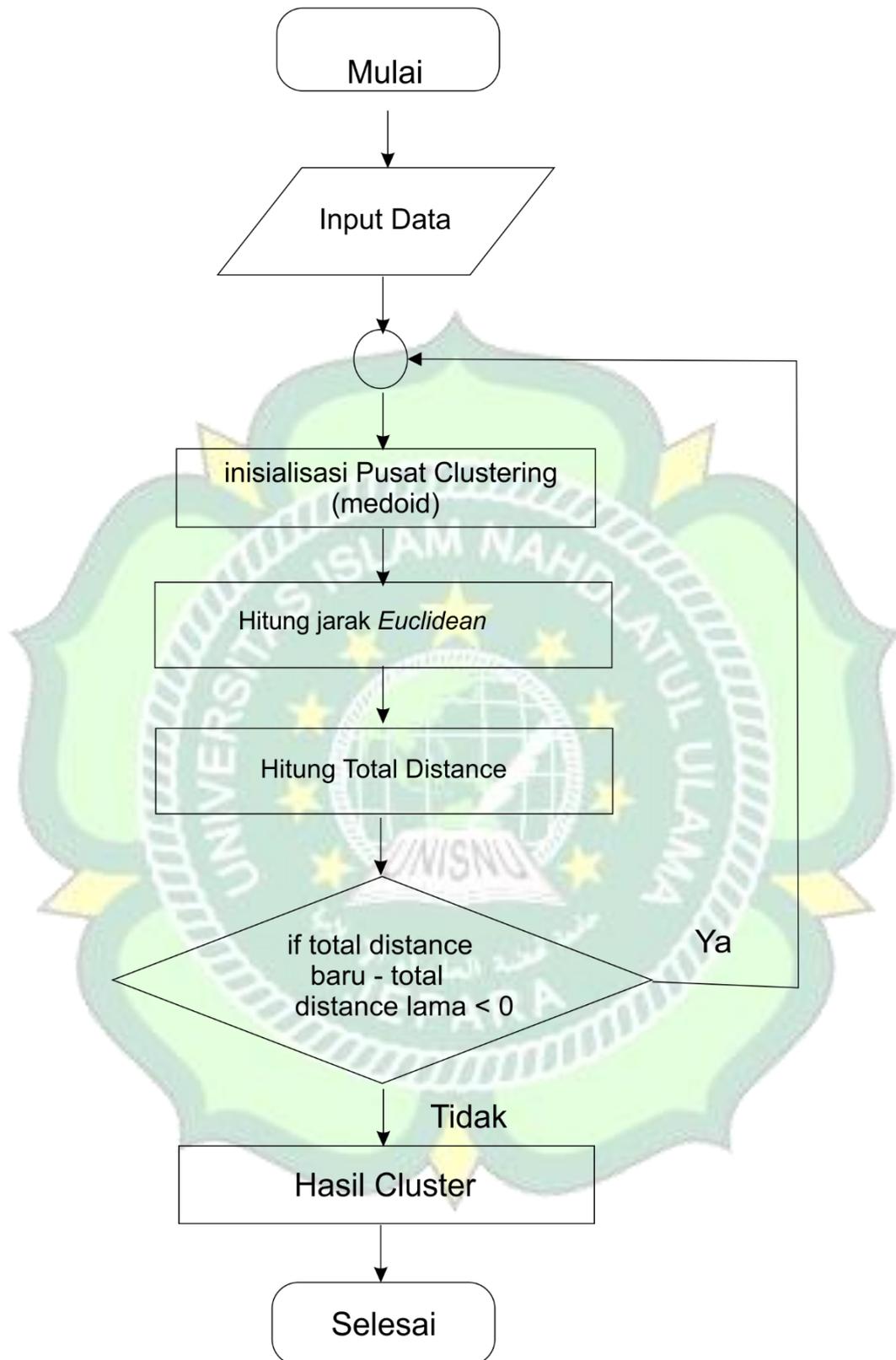
penghapusan atribut-atribut yang tidak diperlukan sekaligus menghapus record ganda sehingga akan tersisa record data yang unik dan berkualitas.

Tahap kedua adalah tahap pemodelan data menggunakan implementasi Algoritma K-medoids. File yang sudah dimasukkan akan mengalami proses pengujian, dimulai dengan dataset Read *Excel*. Kemudian di hubungkan dengan *Set Role* untuk memilih atribut sebagai *Role*. Setelah itu disambungkan dengan algoritma yang di pilih, penelitian ini menggunakan *Algoritma K-medoids Clustering*. Selanjutnya di hubungkan lagi dengan *Apply Model*, dan di hubungkan lagi dengan *Performance* dan yang terakhir di hubungkan ke *Result*.

3. Evaluasi dan Validasi

Tahap ketiga adalah proses evaluasi dengan menggunakan pengukuran evaluasi *clustering* yaitu pengukuran akurasi, sensitifity, dan juga spesifity. Hasil pengujian model adalah untuk mengukur tingkat akurasi dari registrasi PMB menggunakan metode ROC. Sedangkan untuk validasi algoritma *K-Medoids* menggunakan metode *Double Buildean Index (DBI)*. Adapun proses detail diagram alur algoritma *K-Medoids* dapat dilihat pada Gambar 3.4 dibawah ini.





Gambar 3. 4. Flowchart Algoritma *K-Medoids* [4]