

BAB III

METODOLOGI PENELITIAN

3.1 Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan di Universitas Muhammadiyah Yogyakarta di ruang Biro Sistem Informasi, gedung AR. Fachruddin B. Adapun waktu penelitian ini dilaksanakan dari bulan Agustus 2016 - Oktober 2016.

3.2 Alat dan Bahan

Dalam penelitian kali ini penulis menggunakan perangkat keras (*Hardware*) dan juga perangkat lunak (software) dengan spesifikasi sebagai berikut.

3.2.1 Alat

a. Perangkat Keras (*Hardware*)

1. *Personal* laptop,

Demi kelancaran penelitian ini, peneliti menggunakan laptop untuk simulasi dan analisis hasil pengolahan data dengan spesifikasi sebagai berikut:

- 1) RAM 4Gb
- 2) Hardisk 500Gb
- 3) Processor Intel Core i5
- 4) Sistem Operasi Windows 8 64-bit

b. Perangkat Lunak (*Software*)

1. *Visual Studio 2013*

Pada *software Visual Studio 2013* menggunakan *SQL Server Data Tools for Visual Studio 2013* untuk mengolah data dengan menggunakan metode *clustering*.

2. *SQL Server 2014 Management Studio*

Dengan menggunakan *SQL Server 2014 Management Studio* yaitu untuk menghubungkan dengan database server pada Biro Sistem Informasi Universitas Muhammadiyah Yogyakarta.

3. Microsoft Office Excel 2016

Penggunaan Excel untuk *pre-processing* data yaitu untuk membuang duplikasi data dan memeriksa kesalahan pada data. Data yang diproses adalah data yang telah diambil dari basis data.

4. WEKA (*Waikato Environment for Knowledge Analysis*)

Weka sebagai aplikasi untuk menganalisis data *mining* dengan metode dan algoritma yang diinginkan.

3.2.2 Bahan

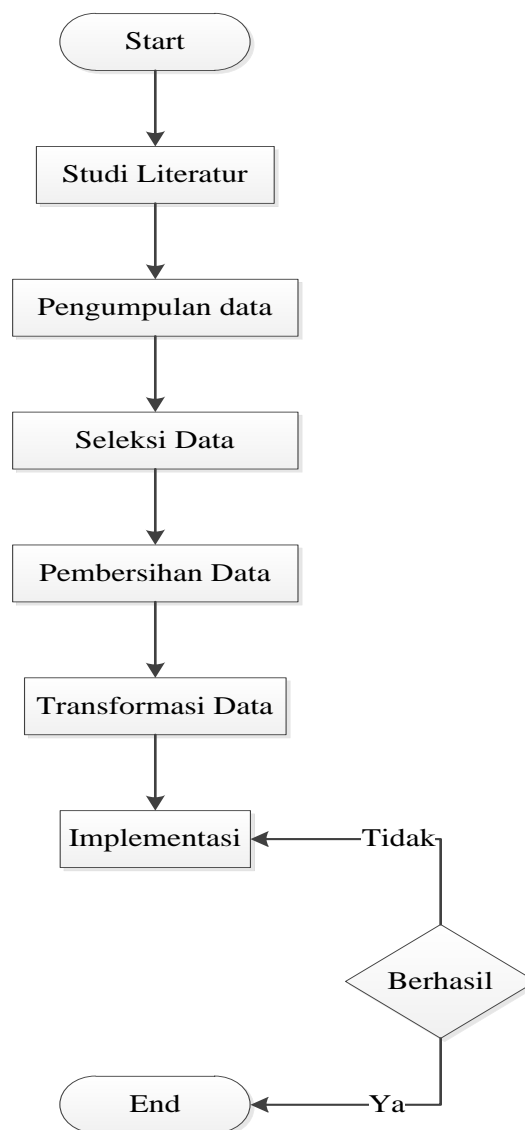
Bahan yang akan digunakan dalam penelitian ini adalah:

- Data yang diperoleh berdasarkan data mahasiswa Fakultas Teknik Universitas Muhammadiyah Yogyakarta

3.3 Prosedur Penelitian

3.3.1 Alur Penelitian

Dalam melakukan penelitian ini, penulis menggunakan model SDLC waterfall dapat dilihat pada **Gambar 3.1**.



Gambar 3.1 *Flowchart* Langkah Penelitian

3.3.2 Studi Literatur

Tahap pertama yang dilakukan dalam penelitian ini adalah Studi literatur. Studi literatur adalah tahap dalam mengumpulkan semua informasi yang diperlukan penulis untuk membangun sistem. Informasi tersebut dapat diperoleh penulis dari berbagai sumber dengan cara membaca literatur yang terdapat pada jurnal, artikel, buku-buku dan skripsi. Dari literatur tersebut dapat diketahui persamaan ataupun perbedaan terhadap penelitian yang dilakukan oleh penulis maupun dengan peneliti lainnya.

3.3.3 Pengumpulan Data

Tahapan berikutnya yaitu pengumpulan data. Dalam penelitian ini tahapan pengumpulan data sangatlah penting untuk penulis karena dari pengumpulan data penulis memperoleh sebuah informasi yang dibutuhkan dalam proses penelitian. Karena data sangatlah penting dalam sebuah penelitian maka dalam tahap proses pengumpulan data harus la dilakukan dengan benar, jika terjadi kesalahan dalam proses pengumpulan data maka akan membuat proses analisis data akan menjadi sulit. Selain itu hasildan kesimpulan akan menjadi rancu jika terjadi kesalahan dalam pengumpulan data.

Analisis data hanya dilakukan di database Universitas Muhammadiyah Yogyakarta, dikarenakan informasi data tentang bagian mahasiswa berada pada database tersebut. Di database Universitas Muhammadiyah Yogyakarta ini lah terdapat informasi yang dibutuhkan dalam pembangunan *data mining*

3.3.4 Seleksi data (*Data Selection*)

Data yang ada di *database* tidak semuanya dipakai untuk penelitian ini, oleh karena hanya data yang sesuai untuk dianalisis yang akan diambil di *database*. Sebagai contoh mengambil ID, tidak perlu mengambil nama pelanggan cukup dengan id pelanggan saja.

3.3.5 Pembersihan data (*Cleaning Data*).

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten. Pada umumnya data yang diperoleh, baik dari *database* Universitas maupun hasil eksperimen, mempunyai isi yang tidak sempurna seperti data yang hilang, data yang tidak valid. Selain itu ada juga atribut-atribut data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performas dari teknik *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

3.3.6 Transformasi data (*Data Transformation*).

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining* karena beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Contoh beberapa metode standar seperti analisis asosiasi dan *clustering* yang hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi – bagi menjadi beberapa interval.

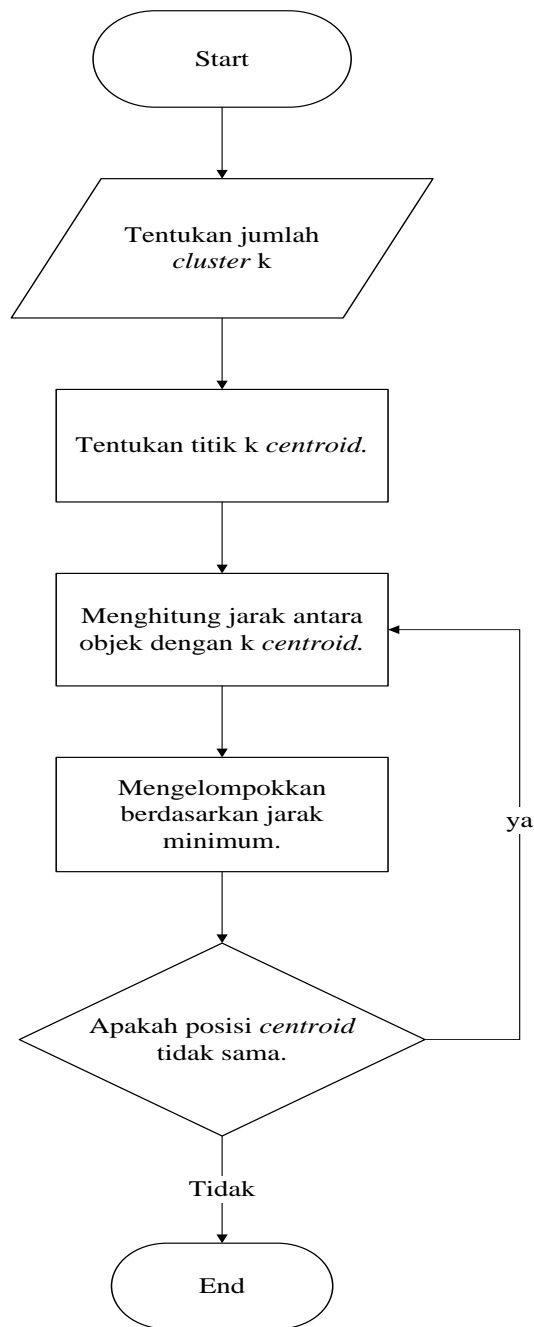
3.3.7 Implementasi

Merupakan suatu proses utama saat metode ini dilakukan untuk menemukan pengetahuan berharga atau sebuah informasi dari data. Teknik yang digunakan pada proses *mining* ini adalah *simple K-Means* dengan menggunakan algoritma C4.5 dan *software* yang digunakan dalam proses *mining* ini yaitu Weka.

3.4 Flowchart K-Means

Algoritma *K-Means* memiliki peroses alur sebagai berikut

1. Menentukan jumlah *cluster* dan menentukan koordinat titik tengah *cluster*. Kelompok *cluster* yang dibuat adalah 4 kelompok berdasarkan jumlah atribut yang digunakan. Total jumlah dari setiap atribut tersebut untuk menentukan kelompok diambil dari frekuensi kurang, frekuensi sedang dan frekuensi padat.
2. Penentuan nilai dari *cluster* untuk dijadikan acuan dalam melakukan perhitungan jarak objek ke *centroid*, perhitungan jarak mengacu pada rumus *Euclidean* yang dilakukan perhitungan di *excel*.
3. Pengelompokkan objek tersebut berdasarkan pada jarak yang dilakukan dengan menggunakan hasil pada perhitungan jarak yaitu pada proses langkah 2. Hasil pada perhitungan jarak tersebut digunakan untuk penentuan kelompok *clustering*.



Gambar 3.2 Flowchart Algoritma K-Means

3.5 Analisis Data Mining

Dalam penelitian ini akan dicari nilai *support* dan *confidence* dari hubungan tingkat mahasiswa *cumlaude* dengan data induk mahasiswa. Tidak semua data induk mahasiswa akan dicari hubungannya dengan data mahasiswa *cumlaude*,

hanya beberapa atribut yang kira-kira berguna dan sebarannya tidak terlalu acak. Karena data yang terlalu acak akan membuat proses *mining* memakan waktu lama dan tingkat hubungannya pun rendah. Data induk mahasiswa yang akan dicari hubungannya meliputi proses masuk, asal sekolah, kota asal sekolah, dan program studi. Adapun yang akan diproses *mining* meliputi:

Tabel 3.1 Atribut yang digunakan untuk klasterisasi

No	Atribut	Keterangan Atribut
1	Student	Mengetahui ID
2	High School Major	Asal Sekolah
3	Gender	Jenis kelamin
4	Province Of Birth	Asal daerah

1. Hubungan mahasiswa *cumlaude* dengan asal sekolah dan proses masuk
 Dari atribut proses masuk dan asal sekolah dicari hubungan mahasiswa *cumlaude* dengan asal sekolah yang melalui proses masuk seleksi penerimaan mahasiswa baru dengan harapan dapat mengetahui tingkat keberhasilan mahasiswa dengan sekolah tertentu.
2. Hubungan mahasiswa *cumlaude* dengan wilayah asal sekolah
 Hubungan mahasiswa *cumlaude* dengan wilayah asal daerah bermanfaat untuk mengetahui daerah-daerah mana yang mempunyai tingkat keberhasilan tinggi ataupun rendah. Diasumsikan bahwa wilayah asal sekolah merupakan kota tempat asal mahasiswa.
 Dalam penelitian ini penulis mencoba memetakan 5 wilayah untuk seluruh Indonesia, yaitu;

- a) Sumatra dan Kepulauan Riau
- b) Jawa – Bali
- c) Kalimantan
- d) Sulawesi
- e) Nusa Tenggara – Maluku – Papua

3. Hubungan mahasiswa *cumlaude* dengan *gender*

Dari atribut proses tersebut harapannya kita dapat mengukur sejauh mana masalah *gender* juga dapat mempengaruhi faktor *cumlaude* mahasiswa Fakultas Teknik

3.6 Prosedur Eksperimen atau Prosedur Validasi

Setelah dilakukan penelitian dengan metode *clustering* menggunakan algoritma *K-Means*, maka dilakukan validasi dari data mahasiswa pada Fakultas Teknik Universitas Muhammadiyah Yogyakarta dalam tingkat *cumlaude* mahasiswa. Adanya hasil penelitian dan data tersebut untuk membuktikan apakah algoritma *K-Means* valid digunakan untuk menganalisa faktor *cumlaude* mahasiswa Fakultas Teknik Universitas Muhammadiyah Yogyakarta.

Apabila dari hasil penelitian ini memiliki validasi dengan data yang sebenarnya, maka algoritma *K-Means* dapat digunakan pada proses analisis faktor mahasiswa *cumlaude* mahasiswa Fakultas Teknik di Universitas Muhammadiyah Yogyakarta (UMY). Jika hasil penelitian ini berbanding jauh dengan data yang sebenarnya, maka algoritma *K-Means* tidak valid digunakan pada data menganalisis faktor mahasiswa *cumlaude*.

3.7 Analisis Data

Setelah dilakukan penelitian dengan metode *clustering* menggunakan algoritma *K-Means*, maka dilakukan validasi dari data mahasiswa *cumlaude* fakultas teknik UMY. Adanya hasil penelitian dan data tersebut untuk membuktikan apakah algoritma *K-Means valid* digunakan untuk menganalisis faktor-faktor yang mempengaruhi mahasiswa *cumlaude* fakultas teknik UMY?

Apabila dari hasil penelitian ini memiliki validasi dengan data yang sebenarnya, maka algoritma *K-Means* dapat digunakan pada proses analisis faktor-faktor yang mempengaruhi mahasiswa *cumlaude* Fakultas Teknik UMY di Universitas Muhammadiyah Yogyakarta (UMY). Jika hasil penelitian ini berbanding jauh dengan data yang sebenarnya, maka *algoritma K-Means* tidak valid digunakan untuk menganalisis faktor-faktor yang mempengaruhi mahasiswa *cumlaude* Fakultas Teknik UMY.