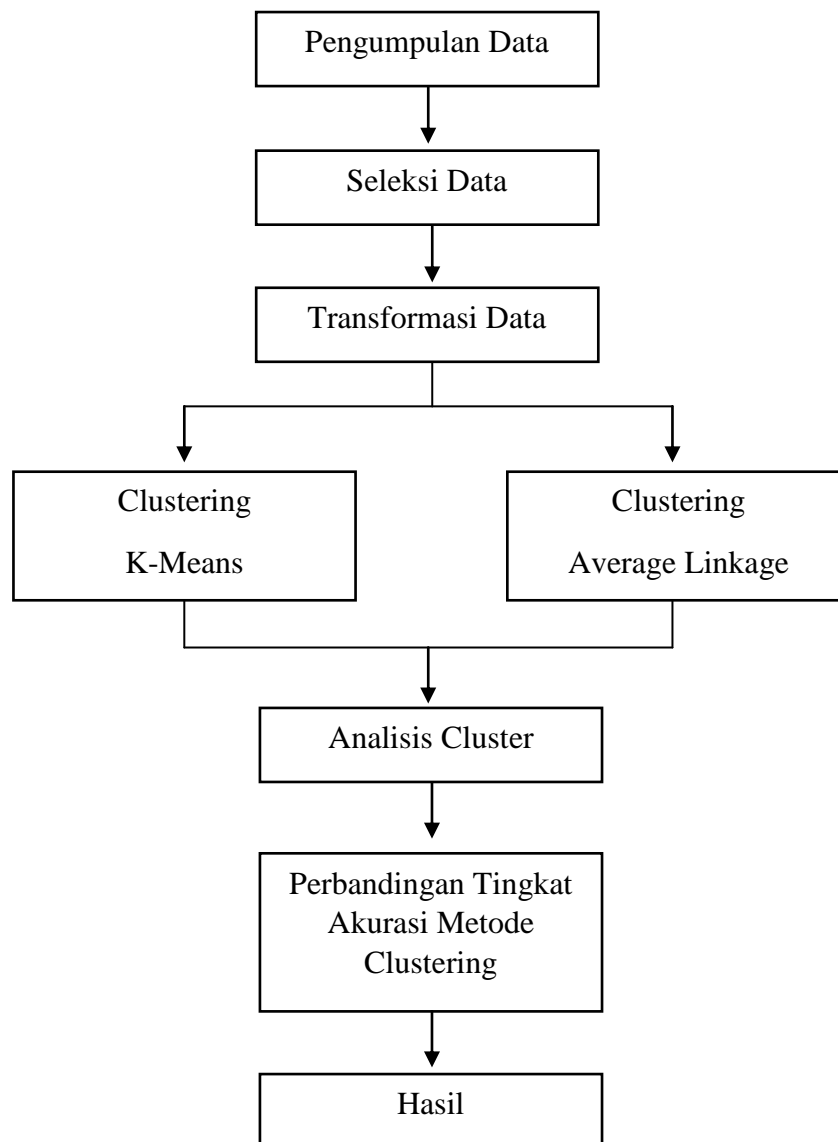


BAB III

METODOLOGI PENELITIAN

Penelitian ini akan dilaksanakan berdasarkan alur penelitian seperti yang ditunjukkan pada Gambar 3.1.



Gambar 3.1 Alur Penelitian

3.1. Tempat dan Waktu Penelitian

Penelitian ini dilakukan di Sekolah SMAN 1 Kotagajah Lampung Tengah, bulan Februari 2017.

3.2. Alat dan Bahan

Adapun alat dan bahan yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Hardware

Kebutuhan perangkat keras (hardware) yang digunakan : Notebook Acer
Processor : Intel (R) Core (TM) CPU @1.80 GHz 1.80 GHz. RAM : 4,00 GB
(3,80 GB usable)

2. Software

Kebutuhan perangkat lunak (software) yang digunakan :

- a. Sistem Operasi Windows 7 Ultimate
- b. Microsoft Excel 2007 untuk melakukan Transformasi (Transformasi) data.
- c. Weka Versi 3.8 untuk melakukan clustering data.

3. Data

Data yang akan digunakan dalam pengklusteran adalah data siswa pemohon KIP SMAN 1 Kotagajah tahun 2017.

3.3. Metode Pengumpulan Data

Dalam penyusunan penelitian ini metode yang digunakan dalam proses pengumpulan data antara lain:

a. Studi Pustaka

Studi pustaka dilakukan dengan cara menghimpun informasi yang relevan dengan topik atau masalah yang akan atau sedang diteliti. Informasi diperoleh dari buku-buku ilmiah, laporan penelitian, karangan-karangan ilmiah, tesis dan disertasi, peraturan-peraturan, ketetapan-ketetapan, ensiklopedia, dan sumber-sumber tertulis baik tercetak maupun elektronik.

b. Dokumentasi

Pengambilan data dilakukan melalui dokumen tertulis maupun elektronik dari lembaga/institusi. Dokumen diperlukan untuk mendukung kelengkapan data yang lain.

3.4. Seleksi Data

Sebelum tahap transformasi data dilakukan, maka perlu dilakukan seleksi data, dari beberapa variabel data siswa yang ada, variabel yang dipilih untuk digunakan dalam tahap clustering antara lain:

1. Jenis Tinggal Siswa
2. Alat Transportasi Ke Sekolah
3. Penghasilan Ayah

4. Penghasilan Ibu
5. Jumlah Tanggungan Orang Tua
6. Tagihan Listrik/Bulan
7. Surat Keterangan Tidak Mampu (SKTM)

3.5. Transformasi Data

Sebelum tahap clustering dilakukan, maka perlu dilakukan transformasi data, untuk merubah tipe data Categorical menjadi data Numeric, karena tipe data yang dapat diolah menggunakan metode clustering adalah tipe data Numeric.

3.5.1. Transformasi Data Jenis Tinggal Siswa

Transformasi data Jenis Tinggal Siswa dilakukan dengan mengubah tipe data categorical menjadi tipe data numeric sebagai berikut:

Tabel 3.1. Tabel Transformasi Data Jenis Tinggal Siswa

Categorical	Numeric
Kost	0.35
Bersama Orangtua	0,70
BersamaWali	1

3.5.2. Transformasi Data Alat Transportasi Ke Sekolah

Transformasi data Alat Transportasi Ke Sekolah dilakukan dengan mengubah tipe data categorical menjadi tipe data numeric sebagai berikut:

Tabel 3.2. Tabel Transformasi Data Alat Transportasi Ke Sekolah

Categorical	Numeric
Sepeda Motor	0.25
Angkutan Umum	0,50
Sepeda	0,75
Jalan Kaki	1

3.5.3. Transformasi Data Penghasilan Orang Tua

Transformasi data Penghasilan Ayah dan Penghasilan Ibu dilakukan dengan mengubah tipe data categorical menjadi tipe data numeric sebagai berikut:

Tabel 3.3. Tabel Transformasi Data Penghasilan Orang Tua

Categorical	Numeric
Lebih dari Rp. 2,000,000	0,20
Rp. 1,000,000 – Rp. 2,000,000	0,40
Rp. 500,000 – Rp. 999,999	0,60
Kurang dari Rp. 500,000	0,80
Tidak Berpenghasilan	1,00

3.5.4. Transformasi Data Jumlah Tanggungan Orang Tua

Transformasi data jumlah tanggungan orang tua dilakukan dengan mengubah tipe data categorical menjadi tipe data numeric sebagai berikut:

Tabel 3.4. Tabel Transformasi Data Jumlah Tanggungan Orang Tua

Categorical	Numeric
1	0.25
2	0,50
3	0,75
≥ 4	1

3.5.5. Transformasi Data Rata-Rata Tagihan Listrik per Bulan

Transformasi data Rata-Rata Tagihan Listrik per Bulan dilakukan dengan mengubah tipe data categorical menjadi tipe data numeric sebagai berikut:

Tabel 3.5. Tabel Transformasi Data Rata-Rata Tagihan Listrik per Bulan

Categorical	Numeric
Lebih Dari Rp. 300.000	0.25
Rp. 200.000 – Rp. 300.000	0,50
Rp. 100.000 – Rp. 199.999	0,75
Kurang Dari Rp. 100.000	1

3.5.6. Transformasi Data Surat Keterangan Tidak Mampu

Transformasi data Surat Keterangan Tidak Mampu dilakukan dengan mengubah tipe data categorical menjadi tipe data numeric sebagai berikut:

Tabel 3.6 Tabel Surat Keterangan Tidak Mampu

Categorical	Numeric
Tidak Ada	0,50
Ada	1

3.6. Proses Clustering

Pada tahap ini akan diterapkan metode K-Means Clustering Dan Average Linkage Clustering untuk mengelompokkan data siswa yang layak menjadi penerima Kartu Indonesia Pintar (KIP). Software yang digunakan adalah Weka 3.8.

Langkah-langkah metode Average Linkage Clustering

1. Dimulai dengan N cluster, dimana masing-masing memuat satu kesatuan. Jika terdapat matriks N x N dengan jarak $D = \{d_{ik}\}$.
2. Mencari matriks jarak untuk pasangan cluster terdekat. Misalkan pasangan cluster paling mirip obyek U dan V maka $D = \{d_{uv}\}$, sehingga U dan V dipilih.
3. Menggabungkan cluster U dan V menjadi cluster baru (UV).
4. Memperbaharui masukan dalam matriks jarak dengan cara:
 - Menghapus baris dan kolom sesuai dengan cluster U dan V
 - Menambahkan baris dan kolom dengan memberikan nilai jarak antara cluster baru (UV) dan semua sisa cluster.

5. Mengulangi langkah (2) dan (3) sebanyak (n-1) kali. (Semua obyek akan berada dalam cluster tunggal pada berakhirnya algoritma).
6. Mencatat identitas dari cluster yang digabungkan dan tingkat (jarak atau similaritas) dimana penggabungan terjadi.

Langkah-langkah metode K-Means Clustering

1. Menentukan k sebagai jumlah cluster yang ingin dibentuk
2. Menentukan centroid (titik pusat)
3. Menghitung jarak setiap data/obyek ke setiap centroid menggunakan rumus euclidean distance

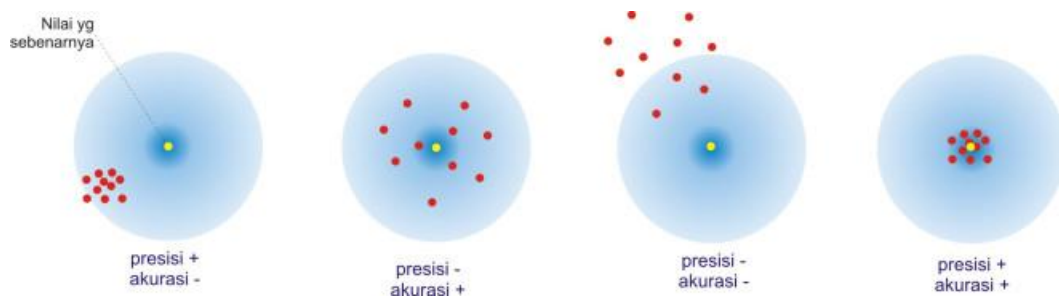
$$d(x_j, c_j) = \sqrt{\sum_{j=1}^n (x_j - c_j)^2}$$

4. Mengelompokkan data/obyek kedalam cluster yang memiliki jarak euclidean terkecil
5. Memperbarui centroid dengan mencari mean (nilai rata-rata) tiap cluster
6. Menghitung jarak setiap data/obyek ke setiap centroid baru
7. Lakukan langkah (3), (4) dan (5) hingga keanggotaan data/obyek terhadap sebuah cluster tidak berubah lagi.

3.7. Perbandingan Akurasi K-Means dan Average Linkage

Dalam pengenalan pola (*pattern recognition*) *precision* dan *recall* adalah dua perhitungan yang banyak digunakan untuk mengukur kinerja dari metode yang digunakan. *Precision* adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Sedangkan *recall* adalah

tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Sedangkan di bidang statistika dikenal juga istilah *accuracy*. Accuracy didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual. Setelah dilakukan Clustering menggunakan dua metode yaitu K-Means Clustering dan Average Linkage Clustering maka dilakukan perbandingan Akurasi untuk menentukan metode terbaik yang hasil cluster nya akan digunakan sebagai dasar pengambilan keputusan penentuan penerima Kartu Indonesia Pintar (KIP). Pada Gambar 3.2 memberikan gambaran perbedaan antara accuracy dan precision.



Gambar 3.2 Accuracy vs Precision