

BAB IV

HASIL DAN PEMBAHASAN

4.1. Data

Proses awal dari Data Mining adalah melakukan proses pengumpulan data dari database. Data yang akan diolah dalam pembahasan ini adalah data siswa yang ada di SMA Negeri 1 Kotagajah Lampung Tengah. Data siswa yang digunakan adalah 1024 data siswa (dapat dilihat pada lampiran 1). Atribut-atribut data yang digunakan adalah :

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)

Data diperoleh dari database siswa yang ada di SMA Negeri Kotagajah Lampung Tengah. Dari 1024 data siswa tersebut akan ditetapkan 300 orang siswa yang berhak menerima KIP.

4.2. Seleksi dan Transformasi Data

Data yang telah diambil dari database selanjutnya diseleksi dan dilakukan transformasi data karena terdapat perbedaan tipe data di antara atribut-atribut yang digunakan. Dalam proses clustering, data yang dapat diolah adalah tipe data numeric, maka dari itu data calon penerima KIP pada lampiran 1 harus di transformasi menjadi data numeric, hasil transformasi data selengkapnya dapat dilihat pada lampiran 3.

4.3. Proses Clustering

Dataset yang digunakan pada clustering tidak menampilkan class / target attribute, sedangkan dataset yang digunakan pada classification mutlak harus menampilkan class / target attribute. Pengetahuan yang dihasilkan oleh metode clustering berupa cluster, sedangkan pengetahuan yang dihasilkan oleh metode classification berupa Decision Tree, Ruleset, BackPropagation, dan lain-lain.

Clustering digunakan ketika tidak diketahuinya bagaimana data harus dikelompokkan. Keluaran pendekatan ini adalah data yang sudah dikelompokkan. Sedangkan classification, terdapat informasi mengenai bagaimana data tersebut dikelompokkan. Selanjutnya pada proses clustering dilakukan training pada sistem dengan data yang sudah diberikan label (ke dalam kelompok manakah data tersebut dikelompokkan), selanjutnya sistem akan mengklasifikasikan data-data yang baru ke dalam kelompok yang ada.

Proses clustering pada penelitian ini menggunakan dua metode, yaitu metode K-Means Clustering dan metode Average Linkage Clustering. Proses Clustering data siswa calon penerima KIP dilakukan menggunakan Software Weka versi 3.8. Cluster yang akan dibentuk adalah dua cluster, yaitu cluster siswa yang berhak menerima KIP dan cluster siswa yang tidak berhak menerima KIP.

4.3.1 Clustering Metode K-Means

Dari 1024 data siswa setelah dilakukan Clustering menggunakan metode K-Means dengan software Weka 3.8, maka didapatkan hasil sebagai berikut:



Gambar 4.1 Persebaran Cluster Siswa Calon Penerima KIP (K-Means)

Number of iterations : 6

Within cluster sum of squared errors : 699.4634696086499

Tabel 4.1 Tabel Nilai Centroid Awal (Random)

Cluster 0	0.75	0.25	0.4	0.6	0.75	0.75	0.5
Cluster 1	0.75	1	0.4	0.8	0.25	0.75	0.5

Tabel 4.2 Tabel Nilai Centroid Akhir

Attribute	Full Data	Cluster 0	Cluster 1
	(1024.0)	(568.0)	(456.0)
tinggal	0.7031	0.7536	0.6402
transport	0.5927	0.2837	0.9776
gajiyah	0.4191	0.4049	0.4368
gajiibu	0.6572	0.6285	0.693
tanggungan	0.6348	0.6395	0.6288
listrik	0.4707	0.4652	0.4775
sktm	0.6377	0.6312	0.6458
Mean		0,5438	0,6428

Clustered Instances :

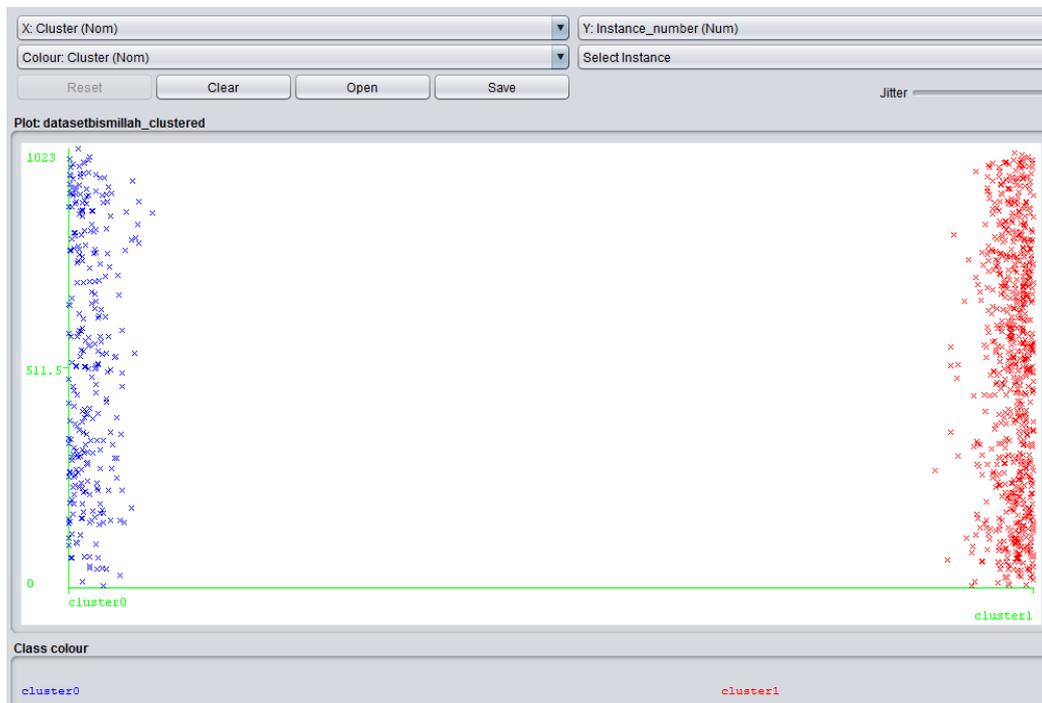
0 568 (55%)

1 456 (45%)

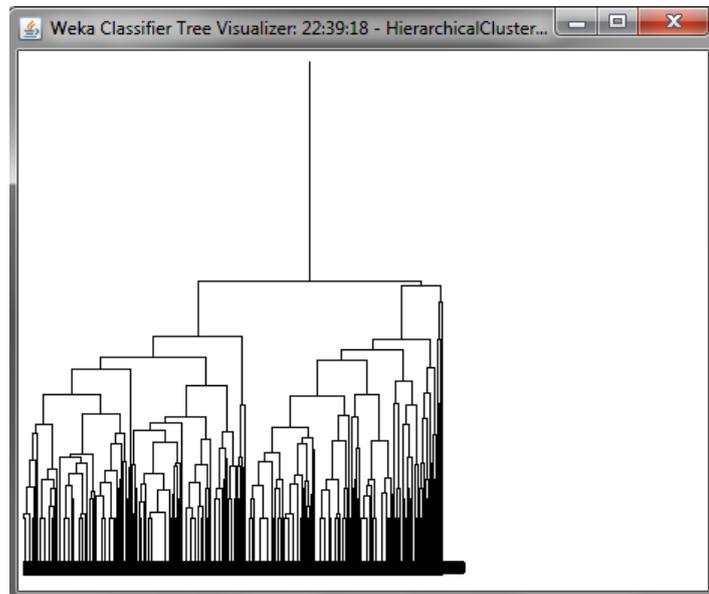
Setelah dilakukan analisis, cluster 0 memiliki rata-rata nilai centroid akhir: 0,5438 dan cluster 1 memiliki rata-rata nilai centroid akhir: 0,6428 maka Cluster 0 ditetapkan sebagai cluster siswa yang tidak berhak menerima KIP karena memiliki nilai rata-rata centroid akhir yang lebih kecil dari Cluster 1 sehingga Cluster 1 ditetapkan sebagai cluster siswa yang berhak menerima KIP. (data hasil cluster menggunakan k-means dengan weka 3.8 selengkapnya dapat dilihat pada Lampiran 3)

4.3.2 Clustering Metode Average Linkage

Dari 1024 data siswa setelah dilakukan Clustering menggunakan metode Average Linkage dengan software Weka 3.8, maka didapatkan hasil sebagai berikut:



Gambar 4.2 Persebaran Cluster Siswa Calon Penerima KIP (Average Linkage)



Gambar 4.3 Dendogram Cluster Siswa Calon Penerima KIP (Average Linkage)

Clustered Instances :

- 0 282 (28%)**
- 1 742 (72%)**

Cluster 0 memiliki nilai awal himpunan cluster (((((((((((((1.0:0,1.0:0)... dan cluster 1 memiliki nilai awal himpunan cluster (((((((((((((0.5:0.25,(0.5:0,0.5:0)... (data hasil cluster menggunakan average linkage dengan weka 3.8 selengkapnya dapat dilihat pada Lampiran 4) dari hasil clustering tersebut dapat ditetapkan Cluster 0 ditetapkan sebagai cluster siswa yang berhak menerima KIP karena memiliki nilai awal himpunan cluster yang lebih besar dari Cluster 1 sehingga Cluster 1 ditetapkan sebagai cluster siswa yang tidak berhak menerima KIP.

4.4. Perbandingan Metode Clustering

Pada penelitian ini untuk mengetahui metode clustering terbaik digunakan perhitungan nilai *precision*, *recall* dan *accuracy*. Secara umum *precision*, *recall* dan *accuracy* dapat dirumuskan sebagai berikut:

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	True Positive (TP)	False Positive (FP)
	FALSE	False Negative (FN)	True Negative (TN)

Precision	$= \frac{TP}{TP+FP}$
Recall	$= \frac{TP}{TP+FN}$
Accuracy	$= \frac{TP+TN}{TP+TN+FP+FN}$

a. Precision, Recall dan Accuracy Metode K-Means Clustering

Dari hasil clustering menggunakan metode K-Means dengan Software Weka 3.8 didapatkan Cluster siswa yang berhak menerima KIP adalah 456 Siswa dari 1024 Siswa. Target penerima KIP adalah 300 Siswa dan Data sebenarnya penerima KIP yaitu 284 Siswa (dapat dilihat pada lampiran 2). Sehingga didapatkan perhitungan sebagai berikut:

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	284 (TP)	172 (FP)
	FALSE	16 (FN)	852 (TN)

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{284}{284+172} = \frac{284}{456} = \mathbf{0,62 (62\%)}$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{284}{284+16} = \frac{284}{300} = \mathbf{0,94 (94\%)}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{284+172}{284+852+172+16} = \frac{456}{1324} = \mathbf{0,344 (34\%)}$$

b. Precision, Recall dan Accuracy Metode Average Linkage Clustering

Dari hasil clustering menggunakan metode Average Linkage dengan Software Weka 3.8 didapatkan Cluster siswa yang berhak menerima KIP adalah 282 Siswa dari 1024 Siswa. Target penerima KIP adalah 300 Siswa dan Data sebenarnya penerima KIP yaitu 284 Siswa (dapat dilihat pada lampiran 2). Sehingga didapatkan perhitungan sebagai berikut:

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	284 (TP)	2 (FP)
	FALSE	16 (FN)	1022 (TN)

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{284}{284+2} = \frac{284}{286} = \mathbf{0,99 (99\%)}$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{284}{284+16} = \frac{284}{300} = \mathbf{0,94 (94\%)}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{284+1022}{284+1022+2+16} = \frac{1306}{1324} = \mathbf{0,986 (98\%)}$$

c. Perbandingan Metode K-Means dan Average Linkage Clustering

Tabel 4.3 Tabel Perbandingan Metode K-Means dan Average Linkage

Metode	Precision	Recall	Accuracy
K-Means	0,62 (62%)	0,94 (94%)	0,344 (34%)
Average Linkage	0,99 (99%)	0,94 (94%)	0,986 (98%)

Teknik pengenalan pola (pattern recognition) dan temu kembali informasi (information retrieval), precision dan recall adalah dua perhitungan yang banyak digunakan untuk mengukur kinerja dari sistem / metode yang digunakan. Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Sedangkan Accuracy didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual.

Dalam menentukan metode clustering terbaik digunakan nilai Precision, Recall dan Accuracy. Clustering menggunakan metode K-Means memiliki nilai Precision : 62 %, Recall : 94% dan Accuracy: 34% sedangkan metode Average Linkage memiliki nilai Precision : 99 %, Recall : 94% dan Accuracy: 98%. Dari kedua hasil tersebut dapat ditarik kesimpulan bahwa Average Linkage Clustering merupakan metode clustering terbaik pada kasus penentuan penerima KIP di SMA Negeri 1 Kotagajah Lampung Tengah karena memiliki nilai Precision, Recall dan Accuracy yang tinggi dibandingkan K-Means Clustering.