

Pengembangan Model Pengambilan Keputusan Penerima Kartu Indonesia Pintar (KIP) Dengan Metode K-Means Dan Average Linkage Clustering (Studi Kasus : SMA Negeri 1 Kotagajah)

Rini Gustini¹, RZ. Abdul Aziz²

^{1,2}Program Pasca Sarjana Prodi Teknik Informatika IIB Darmajaya, Lampung, Indonesia

^{1,2}Jl. ZA. Pagar Alam No.93, Gedong Meneng, Lampung, Indonesia

Corresponding Author E-Mail : rhigoesti@gmail.com

Abstract— Kartu Indonesia Pintar merupakan bagian dari kebijakan dari Presiden Ir. Joko Widodo. Penerima program Kartu Indonesia Pintar (KIP) untuk kalangan siswa-siswi yang bersekolah seringkali tidak tepat sasaran. Di dalam proses pengambilan keputusan siapa yang berhak atas KIP belum jelas aturannya khususnya di SMA Negeri 1 Kotagajah. Proses pengambilan keputusan masih menggunakan input data yang dilakukan oleh operator sekolah melalui aplikasi DAPODIK, sehingga pengambilan keputusan penerima KIP banyak yang tidak tepat sasaran. Untuk itu diperlukan suatu aplikasi sistem pendukung keputusan (SPK) yang dapat memperhitungkan segala kriteria yang mendukung pengambilan keputusan guna membantu, mempercepat dan mempermudah proses pengambilan keputusan dalam penentuan penerima KIP. Salah satu metode yang dapat digunakan untuk meningkatkan kualitas pengambilan keputusan adalah metode clustering. Metode clustering yang digunakan pada penelitian ini adalah metode K-Means Clustering dan Average Linkage Clustering.

Kata Kunci : Average Linkage, Clustering, KIP, K-Means, SMA Negeri 1 Kotagajah.

I. PENDAHULUAN

Pendidikan berperan penting untuk memperoleh kehidupan yang lebih baik, seperti tercantum dalam Undang-Undang Dasar 1945 pasal 31 ayat 1 bahwa setiap warga negara berhak mendapat pendidikan. Pasal 31 ayat 2 berisi tentang setiap warga negara wajib mengikuti pendidikan dasar dan pemerintah wajib membiayainya, ayat 3 menjelaskan bahwa pemerintah mengusahakan dan menyelenggarakan satu sistem pendidikan nasional, yang meningkatkan

keimanan dan ketakwaan serta akhlak mulia dalam rangka mencerdaskan kehidupan bangsa, yang diatur dalam undang-undang.

Kartu Indonesia Pintar merupakan bagian dari kebijakan dari Presiden Ir. Joko Widodo. Kartu ini diresmikan bersamaan dengan Kartu Indonesia Sehat dan Kartu Keluarga Sejahtera pada 3 November 2014. KIP menasar 24 juta siswa kurang mampu yang sebelumnya terdaftar sebagai penerima Bantuan Siswa Miskin. Pada tahap pertama, KIP akan diterapkan di 18 kabupaten/kota dengan sasaran 152.434 siswa di jenjang SD, SMP, SMA/SMK. Berdasarkan data dari Tim Nasional Percepatan Penanggulangan Kemiskinan, KIP direncanakan menasar 20,3 juta siswa kurang mampu. Pada kenyataannya penerima program kartu Indonesia pintar (KIP) untuk kalangan siswa-siswi yang bersekolah tidak tepat sasaran. Didalam proses pengambilan keputusan siapa yang berhak atas KIP belum jelas aturannya khususnya di SMA Negeri 1 Kotagajah. Proses pengambilan keputusan masih menggunakan input data yang dilakukan oleh operator sekolah melalui aplikasi DAPODIK, sehingga pengambilan keputusan penerima KIP banyak yang tidak tepat sasaran. Untuk itu diperlukan suatu aplikasi sistem pendukung keputusan (SPK) yang dapat memperhitungkan segala kriteria yang mendukung pengambilan keputusan guna membantu, mempercepat dan mempermudah proses pengambilan keputusan dalam penentuan penerima KIP. Salah satu metode yang dapat digunakan untuk meningkatkan kualitas pengambilan keputusan adalah metode clustering. Clustering bertujuan menemukan kelompok (cluster) objek yang berguna, dimana gunanya tergantung dari tujuan analisa data (analisis

cluster). Analisis cluster (cluster analysis) adalah salah satu analisis peubah ganda (multivariate analysis) yang digunakan untuk mengelompokkan objek-objek sedemikian rupa sehingga objek dalam satu cluster sangat mirip sedangkan objek di berbagai cluster cukup berbeda (Santosa, 2007).

II. LANDASAN TEORI

A. Pengertian Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. (Turban, dkk. 2005).

Data mining dilakukan dengan *tool* khusus, yang mengeksekusi operasi *data mining* yang telah didefinisikan berdasarkan model analisis. *Data mining* merupakan proses analisis terhadap data dengan penekanan menemukan informasi yang tersembunyi pada sejumlah data besar yang disimpan ketika menjalankan bisnis perusahaan. Kemajuan luar biasa yang terus berlanjut dalam bidang *data mining* didorong oleh beberapa faktor antara lain:

1. Pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam *data warehouse*, sehingga seluruh perusahaan memiliki akses ke dalam *database* yang andal.
3. Adanya peningkatan akses data melalui navigasi web dan internet.
4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Perkembangan teknologi perangkat lunak untuk *data mining* (ketersediaan teknologi).
6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan. (Larose, 2005)

B. Pengelompokan Data Mining

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu:

1. Deskripsi

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpul suara mungkin tidak menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun dengan *record* lengkap menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

5. Pengklusteran

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain.

6. Asosiasi

Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja (*market basket analysis*). (Larose, 2005)

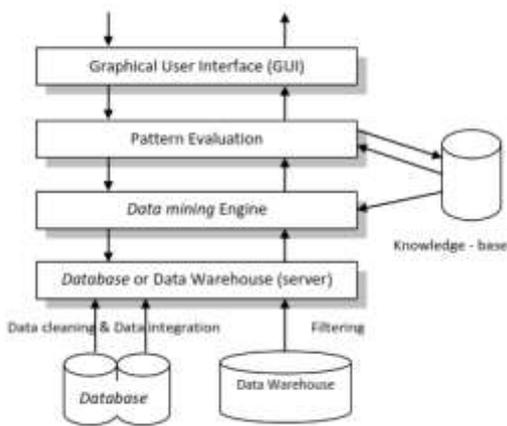
C. Arsitektur Data Mining

Arsitektur utama dari sistem *data mining*, pada umumnya terdiri dari beberapa komponen sebagai berikut:

1. *Database*, *data warehouse*, atau media penyimpanan informasi, terdiri dari satu atau beberapa *database*, *data warehouse*, atau data dalam bentuk lain. Pembersihan data dan integrasi data dilakukan terhadap data tersebut.
2. *Database*, *data warehouse*, bertanggung jawab terhadap pencarian data yang relevan sesuai dengan yang diinginkan pengguna atau *user*.
3. Basis pengetahuan (*Knowledge Base*) merupakan basis pengetahuan yang digunakan sebagai panduan dalam pencarian pola.
4. *Data mining engine*, merupakan bagian penting dari sistem dan idealnya terdiri dari kumpulan modul-modul fungsi yang digunakan dalam proses karakteristik (*characterization*), klasifikasi (*classification*), dan analisis kluster

(cluster analysis). Dan merupakan bagian dari software yang menjalankan program berdasarkan algoritma yang ada.

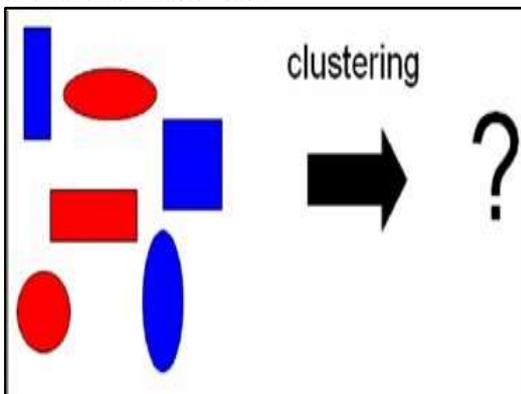
5. Evaluasi pola (*pattern evaluation*), komponen ini pada umumnya berinteraksi dengan modul-modul *data mining*. Dan bagian dari *software* yang berfungsi untuk menemukan *pattern* atau pola-pola yang terdapat dalam *database* yang diolah sehingga nantinya proses *data mining* dapat menemukan *knowledge* yang sesuai.
6. Antar muka (*Graphical user interface*), merupakan modul komunikasi antara pengguna atau user dengan sistem yang memungkinkan pengguna berinteraksi dengan sistem untuk menentukan proses *data mining* itu sendiri.



Gambar 1. Arsitektur *Data mining*

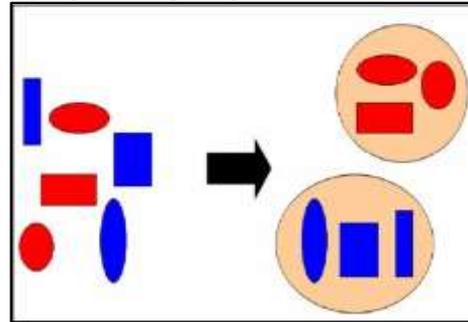
D. Cluster

Cluster adalah sekumpulan objek yang mempunyai “Kesamaan” diantara anggotanya dan memiliki “Ketidaksamaan” dengan objek lain pada *cluster* lainnya, dengan kata lain sebuah *cluster* adalah sekumpulan objek yang digabung bersama karena persamaan atau kedekatannya. *Clustering* adalah proses membuat pengelompokan sehingga semua anggota dari setiap partisi mempunyai persamaan berdasarkan matrik tertentu. Gambar 2.2 berikut menunjukkan contoh data yang akan dilakukan klasterisasi.



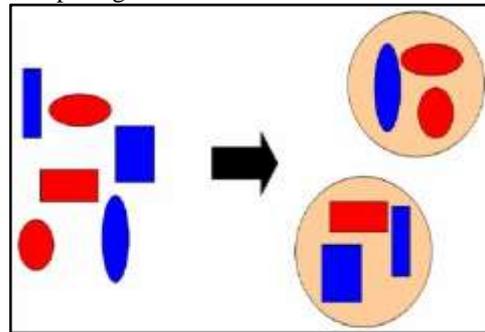
Gambar 2. Data Sebelum di Klasterisasi

Jika data dilakukan *clustering* (pengelompokan) berdasarkan warna, maka pengelompokannya seperti yang terlihat pada gambar 3.



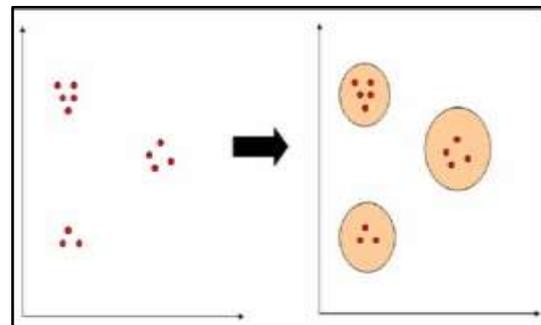
Gambar 3. Klasterisasi Berdasarkan Similaritas (Kesamaan) Warna

Jika data dilakukan *clustering* (pengelompokan) berdasarkan bentuk, maka pengelompokannya dapat dilihat seperti gambar 2.4:



Gambar 4. Klasterisasi Berdasarkan Similaritas (Kesamaan) Bentuk

Selain dengan menggunakan similaritas (kesamaan) berdasarkan bentuk dan warna, *clustering* juga bisa dilakukan dengan menggunakan similaritas berdasarkan jarak, artinya data yang memiliki jarak berdekatan akan membentuk satu *cluster*, contohnya seperti dapat dilihat pada gambar 2.5:



Gambar 5. Klasterisasi Berdasarkan Similaritas (Kesamaan) Jarak

Data *supervised* pada klasifikasi artinya data melalui pembelajaran terbimbing, sedangkan data *unsupervised* pada klasterisasi artinya data

tidak melalui pembelajaran terbimbing. Analisa hasil pada klusterisasi dinyatakan dengan *variance* yang menunjukkan variansi data dalam satu *cluster*, sedangkan klasifikasi analisa hasil diukur menggunakan rasio kesalahan (*error ratio*). Pada dataset yang digunakan oleh klasifikasi terdapat satu *attribut* (label) yang berfungsi sebagai *attribut* target, sedangkan dataset pada klusterisasi tidak terdapat *attribut* (label) sebagai *attribut* target.

E. Metode Pengelompokan

Metode pengelompokan pada dasarnya ada dua, yaitu metode pengelompokan Hirarki (*Hierarchical Clustering Method*) dan metode non Hirarki (*Non Hierarchical Clustering Method*). Metode pengelompokan hirarki digunakan apabila belum ada informasi jumlah kelompok yang akan dipilih. Sedangkan metode pengelompokan Non Hirarki bertujuan untuk mengelompokkan n objek kedalam k kelompok ($k < n$), dimana nilai k telah ditentukan sebelumnya. Salah satu prosedur pengelompokan pada Non Hirarki adalah dengan menggunakan metode *k-means*. Metode ini merupakan metode pengelompokan yang bertujuan untuk mengelompokkan objek sedemikian hingga jarak tiap-tiap objek ke pusat kelompok didalam suatu kelompok adalah minimum.

F. Distance Space

Distance Space berfungsi untuk menghitung jarak antara data dan *centroid*. Ada beberapa macam *distance space* yang sudah diimplementasikan salah satunya adalah : *Euclidean distance space*. *Euclidean* sering digunakan karena penghitungan jarak dalam *distance space* ini merupakan jarak terpendek yang bisa didapatkan antara dua titik yang diperhitungkan. Jarak antara dua titik dapat dihitung dengan cara:

$$d = |x - y| = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$$

Dimana:

- d : jarak
- p : dimensi data
- x : titik data pertama,
- y : titik data kedua,

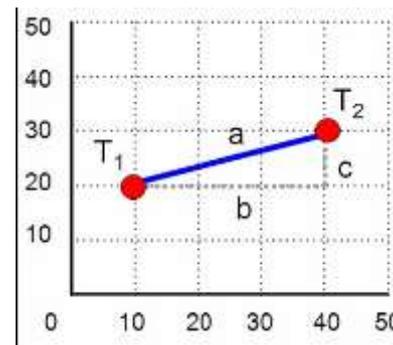
Dalam *Euclidean* perhitungan yang dilakukan merupakan jarak terpendek antara dua titik. Jika ada n titik pengamatan dengan p variabel, maka sebelum dilakukan pengelompokkan data atau objek, terlebih dahulu menentukan ukuran kedekatan sifat antar data.

Ukuran data yang bisa digunakan adalah euclidius (*euclidian distance*), antara dua titik dari p dimensi pengamatan. Jika antar titik X

(x_1, x_2, \dots, x_n) dan titik $Y = (y_1, y_2, \dots, y_n)$ di tentukan dengan rumus:

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Penghitungan jarak dengan *Euclidian Distance* untuk dua titik seperti diilustrasikan pada gambar 2.6:



$$a^2 = b^2 + c^2$$

$$a = \sqrt{b^2 + c^2}$$

$$a = \sqrt{(40 - 10)^2 + (30 - 20)^2}$$

$$a = \sqrt{(30)^2 + (10)^2}$$

$$a = \sqrt{900 + 100}$$

$$a = \sqrt{1000} = 31.628$$

Gambar 6. Penghitungan Jarak Dua Titik

Dari penggambaran diatas dapat diartikan bahwa semakin kecil jarak atau nilai d , maka semakin besar keserupaan antar objek tersebut.

G. Analisa Cluster

Analisis *cluster* adalah suatu analisis statistik yang bertujuan memisahkan obyek kedalam beberapa kelompok yang mempunyai sifat berbeda antar kelompok yang satu dengan yang lain. Dalam analisis ini tiap-tiap kelompok bersifat homogen antar anggota dalam kelompok atau variasi obyek dalam kelompok yang terbentuk sekecil mungkin.

H. K-Means

Algoritma *K-Means* adalah Metode *clustering* non *hierarchical* berbasis jarak yang membagi data kedalam *cluster* dan algoritma ini bekerja pada *attribut* numerik. Algoritma *K-Means* termasuk dalam *partitioning clustering* yang memisahkan data ke k daerah bagian yang terpisah. Algoritma *K-Means* sangat terkenal karena kemudahan dan

kemampuannya untuk mengklaster data besar dan *outlier* dengan sangat cepat.

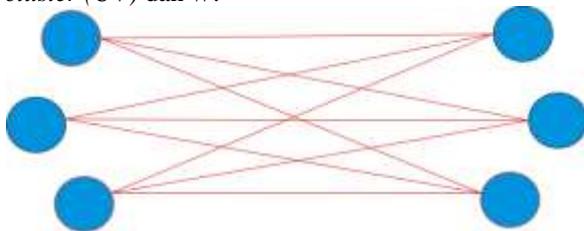
K-Means merupakan metode klasterisasi yang paling terkenal dan banyak digunakan di berbagai bidang karena sederhana, mudah diimplementasikan, memiliki kemampuan untuk mengklaster data yang besar, mampu menangani data *outlier*, dan kompleksitas waktunya linear $O(nKT)$ dengan n adalah jumlah dokumen, K adalah jumlah kluster, dan T adalah jumlah iterasi. Dalam algoritma *K-Means*, setiap data harus termasuk ke *cluster* tertentu pada suatu tahapan proses, pada tahapan proses berikutnya dapat berpindah ke *cluster* yang lain. Pada dasarnya penggunaan algoritma *K-Means* dalam melakukan proses *clustering* tergantung dari data yang ada dan konklusi yang ingin dicapai.

I. Average Linkage Clustering

Average Linkage adalah proses pengelompokan yang didasarkan pada jarak rata-rata antar objeknya. Prosedur ini hampir sama dengan single linkage maupun complete linkage, namun kriteria yang digunakan adalah rata-rata jarak seluruh individu dalam suatu cluster dengan jarak seluruh individu dalam cluster yang lain. (Kusrini 2007). Average Linkage memperlakukan jarak antara dua cluster sebagai jarak rata-rata antara semua pasangan item-item dimana satu anggota dari pasangan tersebut kepunyaan tiap cluster. Mulai dengan mencari matriks jarak $D = \{d_{ik}\}$ untuk memperoleh objek-objek paling dekat (paling mirip) misalnya U dan V . Objek-objek ini digabungkan untuk membentuk cluster (UV). Untuk langkah dari algoritma diatas jarak-jarak antara (UV) dan cluster W yang lain di tentukan oleh:

$$d_{(UV)W} = \frac{\sum_i \sum_k d_{ik}}{N_{(UV)} \cdot N_W}$$

dimana d_{ik} adalah jarak antara objek i dalam cluster (UV) dan objek k dalam cluster W dan N_{UV} dan N_W berturut-turut adalah banyaknya item-item dasar cluster (UV) dan W .



Gambar 7. Ilustrasi Average Linkage Clustering

III. METODOLOGI PENELITIAN

A. 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.

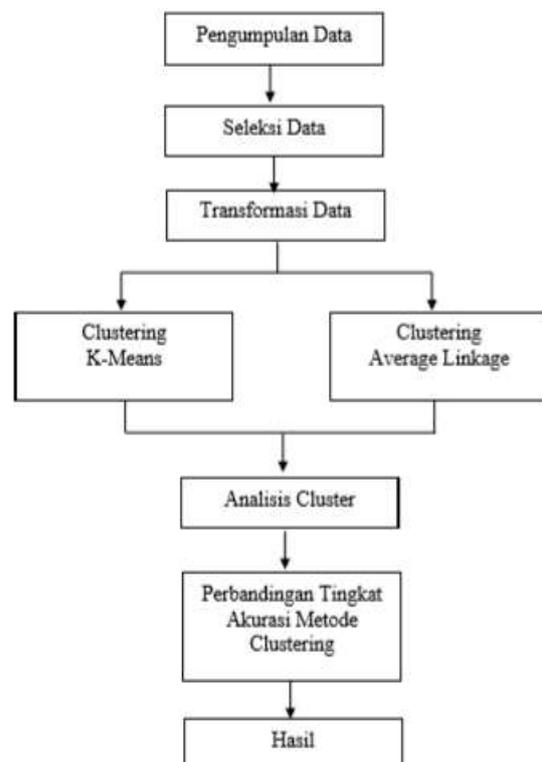
B. Seleksi Data

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)

C. Kerangka Alur Penelitian

Penelitian ini akan dilaksanakan berdasarkan alur penelitian seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

D. Hasil Penelitian

Transformasi Data

Sebelum tahap clustering dilakukan, maka perlu dilakukan transformasi data, untuk merubah tipe data Categorical menjadi data Numeric. Transformasi Data Jenis Tinggal Siswa

Tabel 1. Tabel Transformasi Data Jenis Tinggal Siswa

Categorical	Numeric
Kost	0,35
Bersama Orangtua	0,70
BersamaWali	1

Transformasi Data Alat Transportasi Ke Sekolah

Tabel 2. Tabel Transformasi Data Alat Transportasi Ke Sekolah

Categorical	Numeric
Sepeda Motor	0,25
Angkutan Umum	0,50
Sepeda	0,75
Jalan Kaki	1

Transformasi Data Penghasilan Orang Tua

Tabel 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

Transformasi Data Jumlah Tanggungan Orang Tua

Tabel 4. Tabel Transformasi Data Jumlah Tanggungan Orang Tua

Categorical	Numeric
1	0,25
2	0,50
3	0,75
≥ 4	1

Transformasi Data Rata-Rata Tagihan Listrik per Bulan

Tabel 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

Transformasi Data Surat Keterangan Tidak Mampu

Tabel 6. Tabel Surat Keterangan Tidak Mampu

Categorical	Numeric
Tidak Ada	0,50
Ada	1

IV. HASIL DAN PEMBAHASAN

A. 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. Atribut-atribut data yang digunakan adalah :

8. Jenis Tinggal Siswa
9. Alat Transportasi Ke Sekolah
10. Penghasilan Ayah
11. Penghasilan Ibu
12. Jumlah Tanggungan Orang Tua
13. Tagihan Listrik/Bulan
14. Surat Keterangan Tidak Mampu (SKTM)

B. 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 harus di transformasi menjadi data numeric.

C. Proses Clustering

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.

D. 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 2. Persebaran Cluster Siswa Calon Penerima KIP (K-Means)

Number of iterations : 6
 Within cluster sum of squared errors : 699.4634696086499

Tabel 7. 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 8. Tabel Nilai Centroid Akhir

Attribute	Full Data (1024.0)	Cluster 0 (568.0)	Cluster 1 (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

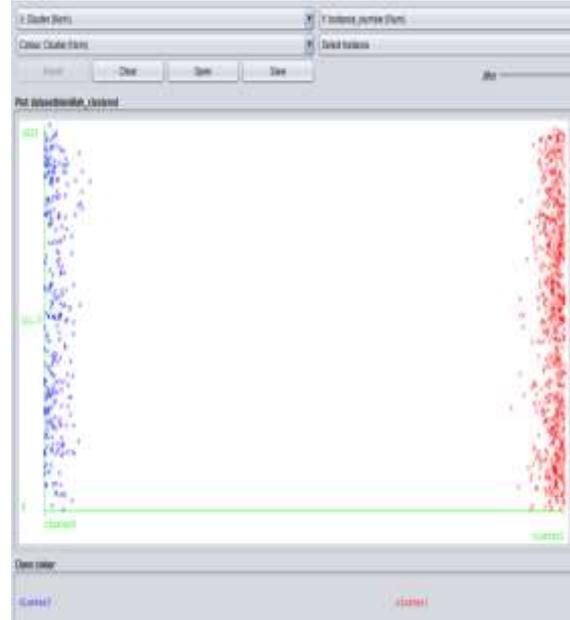
Clustered Instances :

0 568 (55%)
 1 456 (45%)

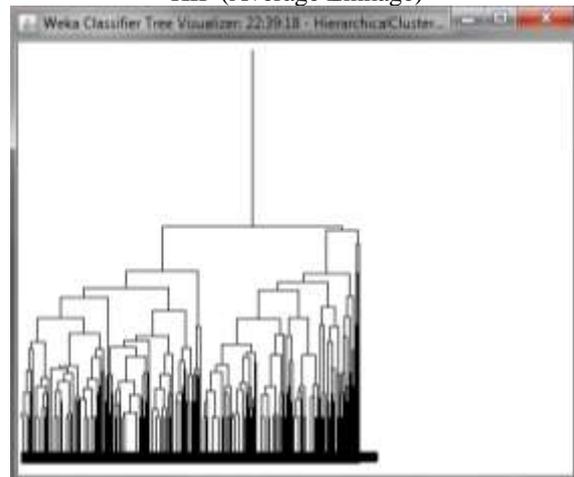
Setelah dilakukan analisis, cluster 0 memiliki rata-rata nilai centroid akhir: 0,5438 dan cluster 1 memiliki rata-rata nilai centroid akhir: 0,6428 maka Cluster 0 ditetapkan sebagai cluster siswa yang tidak berhak menerima KIP karena memiliki nilai rata-rata centroid akhir yang lebih kecil dari Cluster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

E. 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 3. Persebaran Cluster Siswa Calon Penerima KIP (Average Linkage)



Gambar 4. 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.

F. 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:

Tabel 9. Nilai Prediksi Clustering

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

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

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

G. 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. Sehingga didapatkan perhitungan sebagai berikut:

Tabel 10. Hasil clustering menggunakan metode K-Means

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

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

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

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

H. 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:

Tabel 11. Hasil clustering menggunakan metode Average Linkage

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

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

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

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

I. Perbandingan Precision, Recall dan Accuracy Metode K-Means dan Average Linkage Clustering

Tabel 12. Tabel Perbandingan Precision, Recall dan Accuracy Metode K-Means dan Average Linkage Clustering

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%)

Dalam menentukan metode clustering terbaik digunakan nilai Accuracy. Clustering menggunakan metode K-Means memiliki nilai Accuracy: **0,344** atau **34%** sedangkan metode Average Linkage memiliki nilai Accuracy: **0,986** atau **98%**. Sehingga Average Linkage merupakan metode clustering terbaik pada kasus penentuan penerima KIP di SMA Negeri 1 Kota Gajah Lampung Tengah karena memiliki nilai Accuracy yang lebih tinggi dibandingkan dengan metode K-Means.

V. SIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan pembahasan yang telah diuraikan tentang analisis cluster dengan metode K-Means dan Average Linkage Clustering pada data siswa calon penerima KIP di SMA Negeri 1 Kota Gajah, maka dapat disimpulkan beberapa hal sebagai berikut:

1. Menggunakan metode K-Means, data siswa calon penerima KIP terbagi menjadi dua cluster dengan masing-masing jumlah anggota yaitu 568 (55%) dan 456 (45%). Sedangkan metode Average Linkage membagi data siswa calon penerima KIP menjadi dua cluster dengan masing-masing

jumlah anggota yaitu 282 (28%) dan 742 (72%).

2. Berdasarkan nilai accuracy dari metode K-Means dan Average Linkage Clustering, metode Average Linkage Clustering menjadi metode terbaik karena memiliki nilai cluster 0,986 (98%) lebih besar dari metode K-means yang memiliki nilai accuracy 0,344 (34%).

B. Saran

Adapun karena adanya keterbatasan dalam penelitian ini maka saran yang diajukan untuk penelitian selanjutnya adalah sebagai berikut:

1. Penelitian ini dapat dijadikan sebagai salah satu referensi bagi Pemerintah, khususnya Dinas Pendidikan sebagai acuan dalam pengambilan keputusan untuk menentukan kelayakan siswa penerima KIP.
2. Dalam penelitian ini pembahasan hanya terbatas pada pengembangan model pengambilan keputusan menggunakan metode K-Means dan Average Linkage Clustering. Oleh karena itu penulis menyarankan pada peneliti selanjutnya agar dapat mengembangkan metode-metode clustering yang lain serta dapat mengembangkan dan mengimplementasikan pada ilmu dan studi kasus yang lain.

Daftar Pustaka

- [1] Augusta Yudi. Februari 2007. K-Means – Penerapan, Permasalahan dan Metode Terkait. Jurnal Sistem dan Informatika. Volume 3. <http://www.library.gunadarma.ac.id/journal/view/11303/k-means-penerapan-permasalahan-dan-metode-terkait.html/>. Diakses pada tanggal 2 Desember 2015.
- [2] Connolly, T., Begg, C. 2010. Database Systems: A Practical Approach To Design, Implementation, And Management. 5th Edition. America: Pearson Education.
- [3] Dikti. 2014. Pedoman Umum Beasiswa Dan Bantuan Biaya Pendidikan Peningkatan Prestasi Akademik.
- [4] Dista Ariyadi, Bagas. 2013. Sistem Pendukung Keputusan Seleksi Penerima Beasiswa Pada SMA 1 Boja Dengan Menggunakan Metode Analytic Hierarchy Process (AHP). <http://eprints.dinus.ac.id/12514/>. Diakses pada tanggal 30 November 2015.
- [5] Gordon S. Linoff, Michael J. A. Berry. 2011. Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, 3rd Edition. Canada: Wiley.
- [6] Jiawei Han, Micheline Kamber and Jian Pei. 2011. Data Mining: Concepts and Techniques, 3rd edition. Morgan Kaufmann Publishers.
- [7] Junaeni Ina. Juni 2010. Pengelompokan Wilayah Curah Hujan Kalimantan Barat Berbasis Metode Ward Dan Fuzzy Clustering. Jurnal Sains Dirgantara. Volume 7. No.2. http://jurnal.lapan.go.id/index.php/jurnal_sains/article/view/1117. Diakses pada tanggal 5 Desember 2015.
- [8] KBBI. Pengertian Beasiswa. <http://kbbi.web.id/beasiswa> . Diakses pada tanggal 30 November 2015.
- [9] Kusriani, dan Emha Taufik Luthfi. 2009. Algoritma Data Mining. Yogyakarta: Penerbit Andi.
- [10] Kusumadewi Sri, Rismawan Tedy. 2008. Aplikasi K-Means Untuk Pengelompokan Mahasiswa Berdasarkan Nilai Body Mass Index (Bmi) & Ukuran Kerangka. SNATI Yogyakarta. ISSN: 1907-5022.
- [11] <http://journal.uui.ac.id/index.php/Snati/article/view/753>. Diakses pada tanggal 3 Desember 2015.
- [12] Larose Daniel T. 2006. Data Mining Methods and Models. Canada: Wiley.
- [13] Murniasih Erny. 2009. Buku Pintar Beasiswa. Jakarta: Gagas Media.
- [14] Prasetyo Eko. 2012. Data Mining Konsep dan Aplikasi Menggunakan Matlab. Yogyakarta: Andi.
- [15] Rachmatin Dewi. 2014. Aplikasi Metode-Metode Agglomerative Dalam Analisis Klaster Pada Data Tingkat Polusi Udara. Jurnal Infinity. Volume 3. No.2. <http://e-journal.stkipsiliwangi.ac.id/index.php/infinity/article/view/59>. Diakses pada tanggal 5 Desember 2015.
- [16] Ridho Barakbah, Ali dan Arai Kohei. 2004. Determining Constraints of Moving Variance to Find Global Optimum and Make Automatic Clustering. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.469.7736&rep=rep1&type=pdf>. Diakses pada tanggal 30 November 2015.
- [17] Simamora, Bilson. 2004. Riset Pemasaran. Jakarta: Gramedia Utama.
- [18] Suprihatin. Februari 2011. Klastering K-Means Untuk Penentuan Nilai Ujian. JUSI. Volume 1. No. 1. ISSN: 2087-8737. <http://is.uad.ac.id/jusi/files/06-JUSI-Vol-1-No-1-Klastering-K-Means-untuk-Penentuan-Nilai-Ujian.pdf>. Diakses pada tanggal 2 Desember 2015.