

IMPLEMENTASI DATA MINING UNTUK MENENTUKAN KELUARGA YANG LAYAK MENDAPAT KARTU PKH (PROGRAM KELUARGA HARAPAN) DENGAN METODE K-MEANS CLUSTERING

Rika Aprilawati br.Barus, Pilipus Tarigan

Program Studi Teknik Informatika STMIK Budi Darma, Medan, Indonesia
Jl. Sisingamangaraja No. 338 Simpang Limun, Medan

ABSTRAK

PKH (Program Keluarga Harapan) adalah Kartu bantuan untuk anak sekolah yang berasal dari keluarga kurang mampu. Kartu PKH ini sangat dibutuhkan bagi masyarakat yang membutuhkan bantuan anak sekolah. Sebagaimana diketahui bahwa, untuk menentukan keluarga yang layak mendapat kartu PKH (Program Keluarga Harapan) harus membutuhkan data yang akurat untuk mengetahui siapa saja yang layak mendapat kartu PKH (Program Keluarga Harapan). Pada kartu PKH sering hasilnya tidak sesuai, dikarenakan banyak masyarakat yang merasa dirinya pantas untuk mendapatkan kartu PKH tetapi tidak sesuai dengan kriteria yang sudah ditentukan oleh pihak kecamatan. Maka dari itu banyak masyarakat yang kecewa dengan pihak kecamatan. Maka dari itu untuk menerapkan keluarga yang layak mendapat kartu PKH adalah dengan algoritma K-Means Clustering dengan membuat suatu data yang akurat agar data tersebut bisa dibagi sesuai dengan pengelompokan yang ditentukan.

Kata Kunci : Data Mining, PKH, Algoritma K-Means Clustering

I. PENDAHULUAN

Sebagaimana diketahui bahwa, untuk menentukan keluarga yang layak mendapat kartu PKH (Program Keluarga Harapan) harus membutuhkan data yang akurat untuk mengetahui siapa saja yang layak mendapat kartu PKH (Program Keluarga Harapan). Pada kartu PKH sering hasilnya tidak sesuai, dikarenakan banyak masyarakat yang merasa dirinya pantas untuk mendapatkan kartu PKH tetapi tidak sesuai dengan kriteria yang sudah ditentukan oleh pihak kecamatan. Adapun kriteria-kriteria yang dibutuhkan untuk menentukan keluarga yang layak mendapat kartu PKH (Program Keluarga Harapan) adalah nama, pekerjaan, gaji perbulan, jumlah anak dalam satu keluarga, dan kondisi rumah. Hasil data yang didapat dari setiap dusun membuat lambatnya kartu PKH yang didapatkan oleh masyarakat, sehingga untuk mendapatkan kartu PKH membuat prosesnya lama. Maka Untuk menentukan keluarga yang layak mendapat kartu PKH (Program Keluarga Harapan) adalah harus membutuhkan data yang benar-benar sesuai dengan kriteria yang sudah ditentukan oleh pihak kecamatan.

Pemanfaatan data yang ada di dalam sistem informasi untuk menunjang kegiatan pengambilan keputusan, tidak cukup hanya mengandalkan data operasional saja, diperlukan suatu analisis data untuk menggali potensi-potensi informasi yang ada. Para pengambil keputusan berusaha untuk memanfaatkan gudang data yang sudah dimiliki untuk menggali informasi yang berguna membantu mengambil keputusan. Hal ini mendorong munculnya cabang ilmu baru untuk mengatasi masalah penggalian informasi atau pola yang penting atau menarik dari data dalam jumlah besar, yang disebut dengan Data Mining, berkembangnya teknologi dimasa sekarang. Dengan berkaitannyakeluarga yang layak mendapatkartuPKH(ProgramKeluargaHarapan)den

ganmenggunakan metode Clustering maka jurnal pendukung yang digunakan adalah sebagai berikut. [1]

II. TEORITIS

A. Data Mining

Data mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (machine learning) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Menurut Kusriani & Emha Taufiq Luthfi dalam bukunya yang berjudul "Algoritma data mining" adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*.

B. Clustering

Cluster yaitu menemukan kumpulan objek hingga objek-objek dalam satu kelompok sama (atau punya hubungan) dengan yang lain dan berbeda (atau tidak berhubungan) dengan objek-objek dalam kelompok lain. Tujuan dari *cluster* adalah meminimalkan jarak didalam *cluster* dan memaksimalkan jarak antar cluster

C. Algoritma K-means Clustering

K-mean *Clustering* menggunakan pendekatan *partitional clustering*. Tiap *cluster* dihubungkan dengan sebuah centroid (titik pusat). Tiap titik ditempatkan kedalam cluster dengan *centroid* terdekat. Jumlah *cluster*, *K*, harus ditentukan. Algoritma dasarnya sangat sederhana, yaitu:

- Pilih *K* titik sebagai *centroid* awal
- Ulangi
- Bentuk *K cluster* dengan menetapkan semua titik yang terdekat .
- Ulangi perhitungan *centroid* dari tiap *cluster*.

e. Sampai *centroid* tidak berubah.

III. ANALISA

A. Analisa Masalah

Dalam proses menentukan keluarga yang layak mendapat kartu PKH (Program Keluarga Harapan), kecamatan bekerja sama dengan kepala dusun untuk mendata masyarakat pada dusun tersebut untuk melakukan pendataan keluarga yang layak mendapat kartu PKH berdasarkan beberapa kriteria yaitu dengan pekerjaan yang berpenghasilan dibawah Rp.2.500.000/bulan, dengan kondisi rumah semi permanen, maka layak mendapatkan kartu PKH. Data yang digunakan dalam membentuk cluster adalah data keluarga dengan kriteria yang telah ditetapkan pegawai kecamatan dalam menentukan keluarga yang layak mendapat kartu PKH pada kecamatan STM Hilir.

Proses transformasi data :

1. Pekerjaan adalah :
 - Buruh atau Petani = 3
 - Wiraswasta = 2
 - Pegawai = 1
2. Kondisi Rumah :
 - Permanen = 1
 - Semi Permanen = 2

Tabel 1. Transformasi data calon penerima kartu PKH (Program Keluarga Harapan)

N o	Nama	Pekerja an	Gaj i /bl n	Jlh .an ak	Kondi si ruma h
1.	Ari	3	2	1	2
2.	Herma n	1	5	2	1
3.	Anto	3	2	1	2
4.	Ani	2	3	3	1
5.	Santa	3	2	2	2
6.	Mira	1	5	1	1
7.	Sella	3	2	1	1
8.	Santi	2	3	2	2
9.	Ratno	3	2	3	2

B. Penerapan Metode Algoritma K-Means Clustering

Pada penerapan algoritma K-means Clustering dilakukan sebuah penyelesaian untuk data calon penerima kartu PKH. Adapun proses penyelesaian data diatas dengan menggunakan algoritma K-means Clustering.

Proses penyelesaian data calon penerima kartu PKH dengan algoritma K-means adalah sebagai berikut:

1. Tentukan pusat awal cluster "centroid"

Untuk menentukan awal diasumsikan

- Diambil data ke- 2 sebagai pusat Cluster Ke-1: (1,5,2,1)
- Diambil data ke- 12 sebagai pusat Cluster Ke-2: (2,3,2,2).

2. Perhitungan jarak pusat cluster

Untuk mengukur jarak antara data dengan pusat cluster digunakan Euclidian distance, kemudian akan didapatkan matrik jarak sebagai berikut :

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

a. Perhitungan Jarak dari data ke 1 terhadap pusat cluster

$$C1 = \sqrt{(1 - 3)^2 + (5 - 2)^2 + (2 - 1)^2 + (2 - 1)^2} = 3.87$$

$$C2 = \sqrt{(2 - 3)^2 + (3 - 2)^2 + (2 - 1)^2 + (2 - 1)^2} = 2$$

b. Perhitungan Jarak dari data ke 2 terhadap pusat cluster C1

$$C1 = \sqrt{(1 - 1)^2 + (5 - 5)^2 + (2 - 2)^2 + (2 - 2)^2} = 0$$

$$C2 = \sqrt{(2 - 1)^2 + (3 - 5)^2 + (2 - 2)^2 + (2 - 2)^2} = 3$$

c. Perhitungan Jarak dari data ke 3 terhadap pusat cluster

$$C1 = \sqrt{(1 - 3)^2 + (5 - 2)^2 + (2 - 1)^2 + (2 - 2)^2} = 3.74$$

$$C2 = \sqrt{(2 - 3)^2 + (3 - 2)^2 + (2 - 1)^2 + (2 - 2)^2} = 1.73$$

d. Perhitungan Jarak dari data ke 4 terhadap pusat cluster

$$C1 = \sqrt{(2 - 3)^2 + (3 - 2)^2 + (1 - 3)^2 + (2 - 1)^2} = 2.64$$

$$C2 = \sqrt{(2 - 2)^2 + (3 - 3)^2 + (2 - 3)^2 + (2 - 1)^2} = 1.41$$

e. Perhitungan Jarak dari data ke 5 terhadap pusat cluster

$$C1 = \sqrt{(2 - 3)^2 + (3 - 2)^2 + (3 - 2)^2 + (1 - 2)^2} = 2$$

$$C2 = \sqrt{(2-3)^2 + (3-2)^2 + (2-2)^2 + (2-2)^2} = 1.41$$

f. Perhitungan Jarak dari data ke 6 terhadap pusat cluster

$$C1 = \sqrt{(3-1)^2 + (2-5)^2 + (2-1)^2 + (2-1)^2} = 3.87$$

$$C2 = \sqrt{(2-1)^2 + (3-5)^2 + (2-1)^2 + (2-1)^2} = 2.64$$

Tabel 2. Pengelompokan Data

1	2	3	4	5	6	7	8	9	10	11	
0	0	0	1	0	0	0	1	0	1	0	C1
0	1	0	0	0	0	0	0	0	0	1	C2

Karena G1 = G2 dimana anggota yang sama, maka tidak perlu dilakukan iterasi / perulangan lagi. Dan sampai disini hasil Clustering sudah mencapai stabil dan Konvergen

Kesimpulan.

Hasil Clustering adalah

- Data PKH adalah 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20
- Data yang layak mendapat kartu PKH adalah

IV. IMPLEMENTASI

A. Implementasi Sistem

Pada bagian ini dijelaskan secara rinci semua dimasukan dan dikeluarkan dari kebutuhan sistem, yaitu:

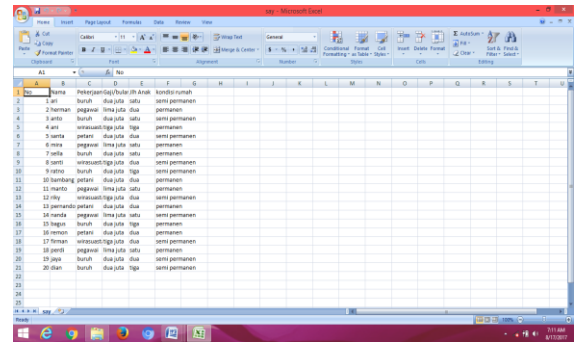
- Sistem Pengguna

Pengguna berinteraksi dengan aplikasi ini dengan menggunakan perangkat sebagai berikut:

 - Keyboard, digunakan oleh pengguna untuk memasukan data maupun perintah kedalam aplikasi ini.
 - Mouse, digunakan untuk melakukan perintah terhadap aplikasi secara modus Grapichal User Interface(GUI).
 - Monitor, digunakan oleh pengguna untuk melihat sistem dan melihat report yang merupakan output dari aplikasi ini

Dibawah ini merupakan langkah-langkah pengujian pada algoritma Algoritma K-Means menggunakan tools Weka, yaitu:

- Langkah pertama pembuatan Format Tabular pada lembar kerja Ms.Excel



Gambar 1. Ms.Excel

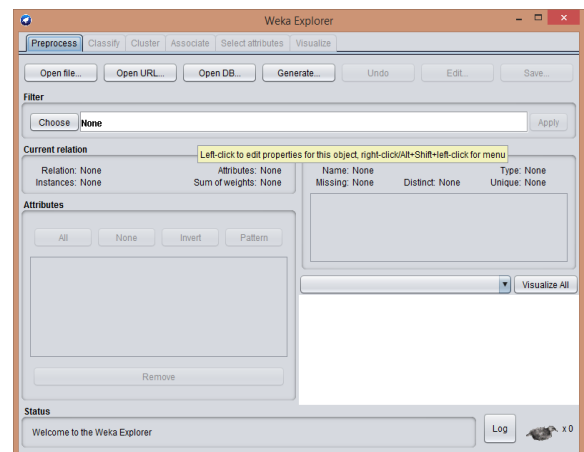
Format tabular tersebut disimpan pada lembar kerja Ms.Excel, Ms.Excel yang menjadi database penyimpanan data tabular, dengan Save as type menjadi Excel 97-2003 Workbook Ms.Excel dengan format ekstensi CSV (*Comma Delimited*) (*.csv), tersebut akan dikoneksikan ke weka.

- Buka Tools Weka 3.81 pada Laptop atau komputer yang telah terinstal weka.
- Berikut adalah tampilan awal weka.



Gambar 2. Tampilan Awal Weka 3.81

- Untuk memulai menggunakan Weka 3.81, pilih *Explorer*

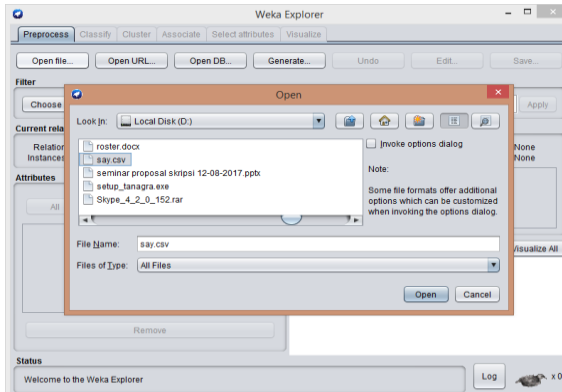


Gambar 3. Isi dari Menu *Explorer*

Dipilih *Open File*—ubah menjadi *all files* - pilih data yang sudah tersimpan pada *folder* – lalau

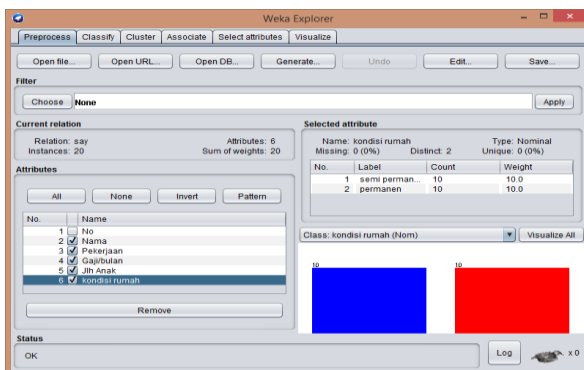
Open karena akan di koneksikan database yang baru dibuat, atau yang belum tersimpan pada Weka 3.81.

3. Kemudian akan tampil tampilan seperti berikut ini.



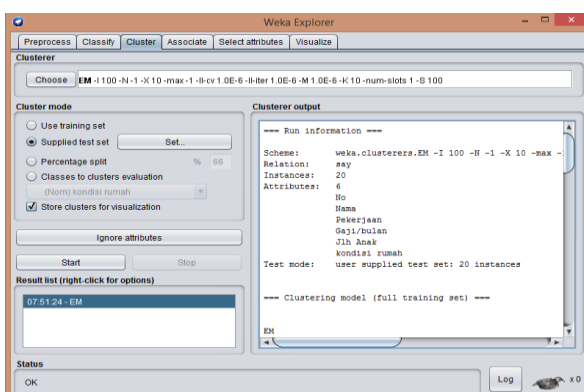
Gambar 4. Tampilan Open File

4. Setelah di klik Menu Open, lalu jangan lupa di ceklis pada sebelah kiri, maka akan muncul tampilan seperti ini.



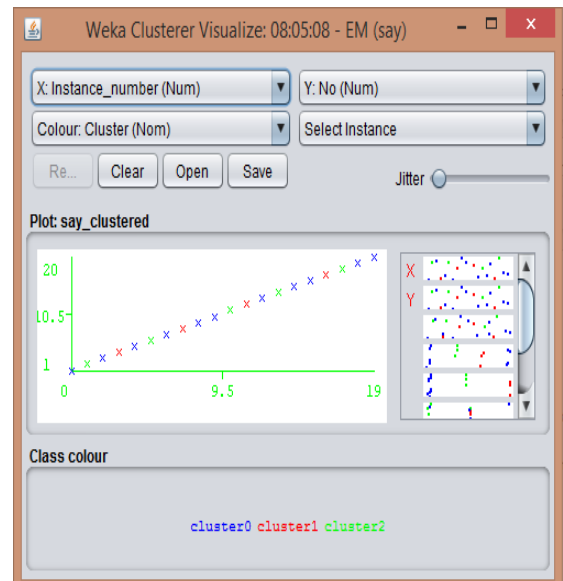
Gambar 5. Pilihan untuk membuka data yang akan di olah

5. Setelah itu pilih menu Cluster pada menu diatas untuk menentukan Cluster, lalu pilih Supplied Test Set, pilih Set, pilih Open File, pilih data yang sudah disimpan, lalu Open, setelah itu klik Start, Maka akan muncul tampilan seperti dibawah ini.



Gambar 6. Tampilan Menu Cluster

6. Setelah itu klik kanan pada 07:51:24 -EM, lalu pilih Visualize Cluster Assignments, maka akan muncul tampilan seperti dibawah ini.



Gambar 7. Weka Cluster

7. Pilih menu Tools pada aplikasi weka tampilan awal, Pilih ArffViewer, Pilih File, lalu Open maka akan tampil hasil akhir seperti gambar di bawah ini.

No.	Instance_number	No	Nama	Pekerjaan	Cajibulan	Jlh Anak	kondisi rumah	Cluster
1	0.0	1.0	arti	buruh	dua juta	satu	semi perman...	cluster0
2	1.0	2.0	herm...	pegawai	lima juta	dua	permanen	cluster2
3	2.0	3.0	anto	buruh	dua juta	satu	semi perman...	cluster0
4	3.0	4.0	ani	wirasuasta	tiga juta	tiga	permanen	cluster1
5	4.0	5.0	santa	pelani	dua juta	dua	semi perman...	cluster0
6	5.0	6.0	mira	pegawai	lima juta	satu	permanen	cluster2
7	6.0	7.0	sella	buruh	dua juta	satu	permanen	cluster2
8	7.0	8.0	santi	wirasuasta	tiga juta	dua	semi perman...	cluster1
9	8.0	9.0	ratho	buruh	dua juta	tiga	semi perman...	cluster0
10	9.0	10.0	bamb...	pelani	dua juta	dua	permanen	cluster0
11	10.0	11.0	manto	pegawai	lima juta	satu	permanen	cluster2
12	11.0	12.0	riky	wirasuasta	tiga juta	dua	semi perman...	cluster1
13	12.0	13.0	perma...	pelani	dua juta	dua	permanen	cluster0
14	13.0	14.0	nanda	pegawai	lima juta	satu	semi perman...	cluster2
15	14.0	15.0	bagus	buruh	dua juta	tiga	permanen	cluster0
16	15.0	16.0	remon	petani	dua juta	dua	permanen	cluster0
17	16.0	17.0	firman	wirasuasta	tiga juta	dua	semi perman...	cluster1
18	17.0	18.0	perdi	pegawai	lima juta	satu	permanen	cluster2
19	18.0	19.0	jaya	buruh	dua juta	dua	semi perman...	cluster0
20	19.0	20.0	dian	buruh	dua juta	tiga	semi perman...	cluster0

Gambar 8. Tampilan hasil Akhir

Pada bab ini adalah hasil pengujian untuk menentukan keluarga yang layak mendapat kartu PKH dengan aplikasi Weka 3.81. Hasil pengujian ini sama hasilnya seperti cara manual, dimana keluarga yang layak mendapat kartu PKH adalah 1,3,5,7,9,10,13,15,16,19,20. Samapai di sini hasil langkah-langkah penyelesaian untuk menentukan

keluarga yang layak mendapat kartu PKH (Program Keluarga Harapan)

V. KESIMPULAN

Sebagai penutup pembahasan dalam penulisan skripsi ini maka diambil kesimpulan-kesimpulan. Dengan adanya kesimpulan ini dapat diambil suatu perbandingan yang akhirnya dapat memberikan perbaikan-perbaikan pada masa yang akan datang. Adapun kesimpulan yang diperoleh adalah :

1. Dengan menggunakan metode *clustering* untuk melakukan perhitungan, ternyata mampu menyelesaikan permasalahan pada Kecamatan STM Hilir dalam menentukan keluarga yang layak mendapat kartu PKH (Program Keluarga Harapan)
2. Dengan memanfaatkan data mining menggunakan metode *clustering* dalam menentukan keluarga yang layak mendapat kartu PKH (Program Keluarga Harapan), informasi yang dihasilkan bersifat *cluster*

Aplikasi weka 3.8.1 digunakan sebagai aplikasi pendukung dalam pengambilan keputusan dan pengujian atas hasil yang didapatkan secara manual, yang menghasilkan sebuah keputusan. Dari keputusan inilah akan menghasilkan sebuah aturan-aturan yang dapat membantu Kecamatan STM Hilir dalam menentukan keluarga yang layak mendapat kartu PKH (Program Keluarga Harapan). Serta mudah dipahami oleh pengguna aplikasi.

REFERENCES

- [1] Rima Dias Ramadhani, "DATA MINING MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN STRATEGI PROMOSI UNIVERSITAS DIAN NUSWANTORO," *Sistem Informasi*, vol. I, no. 8, pp. 1-9, juli 2014.
- [2] Muhammad Faisal Edy Irwansyah, *Advanced Clustering: Teori dan aplikasi*, 2015th ed., Muhammad Faisal Edy Irwansyah, Ed.: DeePublish, 2015.
- [3] Fajar Astuti Hermawati, *Data Mining*, 3rd ed., putri christian, Ed. Andi, Yogyakarta: bowo, 2013.
- [4] Emha taufiq luthfi Kusri, *algoritma data mining*, 2nd ed. andi, 2014.
- [5] Emha taufiq luthfi Kusri, *algoritma data mining*, 11nd ed. andi, 2014.
- [6] Efori Buulolo, "IMPLEMENTASI ALGORITMA APRIORI PADA SISTEM PERSEDIAAN OBAT (STUDI KASUS : APOTIK RUMAH SAKIT ESTOMIHI MEDAN)," *Pelita Informatika Budi Darma*, vol. IV, no. 74, pp. 71-83, agustus 2013.
- [7] Togiartua Nainggolan, *Program keluarga harapan di Indonesia: dampak pada rumah tangga sangat miskin di tujuh provinsi*, Juneman, Ed. P3KS Press, 2012.
- [8] Togiartua Nainggolan, *Program keluarga harapan di Indonesia: dampak pada rumah tangga sangat miskin di tujuh provinsi*, Juneman, Ed. P3KS Press, 2012.
- [9] Fajar Astuti Hermawati, *Data Mining*, 123rd ed., putri christian, Ed. Andi, Yogyakarta: bowo, 2013.
- [10] Fajar Astuti Hermawati, *Data Mining*, 127rd ed., putri christian, Ed. Andi, Yogyakarta: bowo, 2013.
- [11] Muhammad Faisal Edy Irwansyah, *Advanced Clustering: Teori dan aplikasi*, 2015th ed., Muhammad Faisal Edy Irwansyah, Ed.: DeePublish, 2015.
- [12] Fajar Astuti Hermawati, *Data Mining*, 127rd ed., putri christian, Ed. Andi, Yogyakarta: bowo, 2013
- [13] <http://3qepr26caki16dnhd19sv6by6v.wpengine.netdna-cdn.com/wp-content/uploads/2016/06/Weka-GUI-Chooser.png>/diakses tanggal 19 Juni 2017
- [14] <http://3qepr26caki16dnhd19sv6by6v.wpengine.netdna-cdn.com/wp-content/uploads/2016/06/Weka-GUI-Chooser.png>/diakses tanggal 19 Juni 2017