

Penerapan Algoritma C4.5 Dalam Klasifikasi Penerima Kartu Keluarga Sejahtera (KKS)

Arief Prajoko¹, Rahmat Widiya Sembiring², Saifullah³

¹STIKOM Tunas Bangsa Pematangsiantar, Sumatera Utara, Indonesia

^{2,3}STIKOM Tunas Bangsa Pematangsiantar, Sumatera Utara, Indonesia

Jln. Sudirman Blok A No. 1-3 Pematangsiantar, Sumatera Utara

¹riefprajoko13@gmail.com, ³saifullah@amiktunasbangsa.ac.id

Abstract

The purpose of this study is to classify data on families who are entitled to receive a Prosperous Family Card (KKS), because there are still many other poor families who have not had the opportunity to receive this assistance program. Sources of data obtained from the Panghulu Nagori Mekar Rejo office, Bosar Maligas District, Simalungun Regency. The method used in this research is the C4.5 algorithm data mining technique which is implemented with the RapidMiner 5.3 application. The variables used in determining which families are entitled to receive a Prosperous Family Card (KKS) are (1) Occupation, (2) Income of the Head of the Family, (3) Number of Dependents, (4) Highest Education, (5) Condition of the House Building. The results of the classification using the C4.5 algorithm and testing with the Rapidminer software found that the factor that determines the feasibility of providing the Prosperous Family Card (KKS) assistance is C2 (Head of Household Income) with the highest gain value, namely 0.46880137.

Keywords: C4.5 Algorithm, Prosperous Family Card (KKS)

Abstrak

Tujuan penelitian ini untuk mengklasifikasi data keluarga yang berhak menerima Kartu Keluarga Sejahtera (KKS), karena masih banyak keluarga tidak mampu lainnya yang belum berkesempatan menerima program bantuan ini. Sumber data yang diperoleh dari kantor Panghulu Nagori Mekar Rejo Kec.Bosar Maligas Kab.Simalungun. Metode yang digunakan dalam penelitian adalah teknik data mining algoritma C4.5 yang diimplementasikan dengan aplikasi RapidMiner 5.3. Variabel yang digunakan dalam menentukan keluarga yang berhak menerima Kartu Keluarga Sejahtera (KKS) ini, adalah (1) Pekerjaan, (2) Penghasilan Kepala Keluarga, (3) Jumlah Tanggungan, (4) Pendidikan Tertinggi, (5) Kondisi Bangunan Rumah. Hasil dari klasifikasi menggunakan algoritma C4.5 dan pengujian dengan software Rapidminer didapatkan faktor yang menentukan kelayakan dalam pemberian bantuan Kartu Keluarga Sejahtera (KKS) adalah C2(Penghasilan Kepala Keluarga) dengan nilai gain tertinggi yaitu 0.46880137.

Kata kunci: Algoritma C4.5, Kartu Keluarga Sejahtera (KKS)

1. PENDAHULUAN

Kartu Keluarga Sejahtera atau biasa disebut dengan KKS adalah salah satu program pemerintah dalam percepatan penanggulangan kemiskinan, hal ini tercantum pada Peraturan Presiden (Perpres) Nomor 166 Tahun 2014. Pada awalnya program tersebut adalah hasil dari pengembangan program sebelumnya, yaitu dengan nama Kartu Perlindungan Sejahtera (KPS) yang telah dilaksanakan pada era pemerintahan Presiden Susilo Bambang Yudhoyono, dan kini program tersebut diubah secara bertahap menjadi Kartu Keluarga Sejahtera (KKS). Fungsi KKS selain menjadi penanda masyarakat kurang mampu juga berfungsi sebagai kartu identitas untuk mendapatkan Program Simpanan Sejahtera (PSKS). Program bantuan ini dilaksanakan selama pemerintahan Presiden Joko Widodo yang

diberikan sebesar 25% dari total jumlah penduduk yang ada dengan status sosial ekonomi terrendah. Pengambilan data dalam program tersebut berdasarkan hasil pendataan dari program sebelumnya yaitu KPS, pada bulan Juli tahun 2013 . Jumlah tersebut berdasarkan Basis Data Terpadu (BDT) yang diolah oleh Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K), dengan jumlah kisaran bantuan yang diberikan sebesar Rp. 200.000/keluarga/bulan. Pemerintah daerah dalam melakukan pengentasan kemiskinan melalui program Kartu Keluarga Sejahtera tersebut, harus didukung dengan tingkat akurasi data dan juga membutuhkan efisiensi waktu pengolahan data penerima bantuan dan menjumlah total banyaknya jumlah penduduk miskin yang ada.

Nagori Mekar Rejo merupakan salah satu desa di Kecamatan Bosar Maligas Kabupaten Simalungun, dengan tingkat kemiskinan diatas rata-rata. Tetapi dalam melaksanakan program bantuan Kartu Keluarga Sejahtera (KKS) masih ada kesalahan dalam pemberian (KKS) pada masyarakat miskin yang tidak tepat sasaran. Hal ini terjadi karena belum adanya sistem dalam pengolahan data calon penerima bantuan yang cukup besar oleh perangkat desa dikantor Panghulu Nagori Mekar Rejo sehingga terjadi kesalahan dalam pengambilan keputusan.

Dalam masalah tersebut maka penulis akan menyelesaikan masalah diatas dengan menggunakan suatu metode yaitu dengan menggunakan metode teknik *data mining* klasifikasi C4.5. Adapun *variable* yang akan digunakan untuk menyelesaikan masalah diatas diantaranya adalah, Pekerjaan, Penghasilan Kepala Keluarga, Jumlah Tanggungan, Pendidikan Tertinggi, Kondisi Bangunan Rumah. Dengan *variable* tersebut maka penulis akan memberikan masukan kepada perangkat desa Nagori Mekar Rejo dalam memberikan (KKS) kepada masyarakat miskin yang benar-benar membutuhkannya.

Berdasarkan permasalahan diatas, tujuan penelitian ini yaitu untuk mengetahui kelayakan kinerja perangkat desa di Kantor Panghulu Nagori Mekarejo dalam menentukan atau memberikan Kartu Keluarga Sejahtera di Nagori Mekar Rejo dari *variable* yang telah ditentukan. Diharapkan hasil dari penelitian ini dapat memberikan masukan kepada perangkat desa di Kantor Panghulu Nagori Mekar Rejo agar dapat memperbaikinya dalam memberikan Kartu Keluarga Sejahtera (KKS) kepada masyarakat secara layak, sehingga kemudian akan menghasilkan keadilan bagi masyarakat miskin yang benar-benar berhak menerima (KKS) tersebut.

2. METODOLOGI PENELITIAN

2.1. Pengumpulan Data

Pengumpulan data yang dilakukan penulis dengan cara mengambil secara langsung ke Kantor Panghulu Nagori Mekar Rejo. Dalam penelitian ini data yang digunakan akan diolah dari hasil wawancara yang diberikan oleh perangkat Desa Nagori Mekar Rejo dengan jumlah masyarakat sebanyak 86 orang. Penulis melakukan observasi langsung kepada masyarakat dengan jumlah 50 sampel data. Data yang digunakan terdiri dari :

- a) Pekerjaan (C1)
- b) Penghasilan Kepala Keluarga (C2)

- c) Jumlah Tanggungan (C3)
- d) Pendidikan Teritingga (C4)
- e) Kondisi Bangunan Rumah (C5)

Setiap variabel terdiri dari pertanyaan yang diajukan kepada perangkat desa yang diberikan. Data yang digunakan merupakan jenis data *deskriptif*.

2.2. Metode Penelitian

1) Data Mining

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran computer (*machine learning*) untuk menganalisis dan mengekstrasi pengetahuan (*knowledge*) secara otomatis”[1]. Definisi lain *Data Mining* didefinisikan sebagai sebuah proses untuk menemukan hubungan, pola dan tren baru yang bermakna dengan menyaring data yang sangat besar, yang tersimpan dalam penyimpanan, menggunakan teknik pengenalan pola seperti teknik Statistik dan Matematika” [2].

2) Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal” [3]. Kelebihan utama Algoritma C4.5 dapat membuat pohon keputusan (*decision tree*) yang efisien menangani atribut tipe diskrit dan tipe diskrit- numerik, mudah untuk diinterpretasikan dan memiliki tingkat akurasi yang dapat diterima (Luvia et al., 2016).

Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *gain* digunakan rumus seperti yang tertera berikut:

$$Gain(S, A) = Entropy(S) \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

Keterangan:

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|S_i| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

Sebelum mendapatkan nilai *Gain* adalah dengan mencari nilai *Entropi*. *Entropi* digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan sebuah atribut.

Rumus dasar

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

S : Himpunan Kasus

n : Jumlah partisi S

p_i: Proporsi dari S_i terhadap S

3. HASIL DAN PEMBAHASAN

3.1. Perhitungan Menggunakan Metode K-means

Perhitungan algoritma C4.5 dimulai dengan memilih atribut akar terlebih dahulu dengan mencari jumlah kasus keseluruhan, jumlah kasus keputusan layak dan jumlah kasus keputusan tidak layak. Menghitung *entropy* dari semua kasus yang dibagi berdasarkan Pekerjaan (C1), Penghasilan Kepala Keluarga (C2), Jumlah Tanggungan (C3), Pendidikan Tertinggi (C4), dan Kondisi Bangunan Rumah (C5). Setelah itu dilakukan perhitungan *gain* untuk masing – masing atribut. Hasil perhitungan ditunjukkan oleh tabel berikut ini.

Tabel 1. Perhitungan *Node* 1

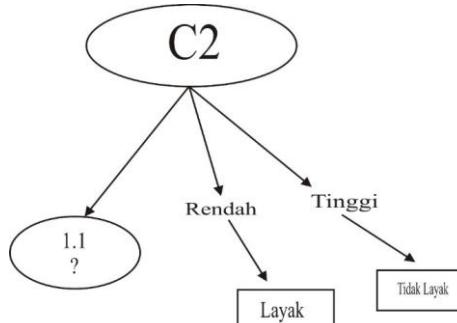
		Jumlah (s)	layak	Tidak layak	Entropy	Gain
Total		86	66	20	0.7824441	
C1						0.225314616
	Wiraswasta	24	13	11	0.9949848	
	Buruh	54	49	5	0.4450649	
	Guru/Honorer	1	1	0	0.0000000	
	Kuli Bangunan	1	1	0	0.0000000	
	Karyawan	5	1	4	0.7219281	
	ART	1	1	0	0.0000000	
C2						0.46880137
	Rendah	53	53	0	0.0000000	
	Cukup	27	13	14	0.9990103	
	Tinggi	6	0	6	0.0000000	
C3						0.121069517
	0	31	30	1	0.2055925	
	1 dan 2	36	22	14	0.9640788	
	>2	19	14	5	0.8314744	
C4						0.031314211
	SD	64	52	12	0.6962123	
	SMP	19	12	7	0.9494520	
	SMA	2	1	1	1.0000000	
	S1	1	1	0	0.0000000	
C5						0.289821966
	Papan	41	41	0	0.0000000	
	Batu	20	14	6	0.8812909	
	Semi Permanen	25	11	14	0.9895875	

Kolom *entropy* pada baris total dihitung dengan persamaan (1) sebagai berikut : $Entropy (total) = - (20/86) * \text{LOG2}(20/86) - (66/86) * \text{LOG2}(66/86)$
 $Entropy (total) = 0.7824441$

Kemudian pada nilai *gain* pada baris C2 dihitung persamaan (2) sebagai berikut : $Gain (total,C2) = (0.46880137) - ((53/86)*0 + (27/86) * 0.9990103 + (6/86)*0)$ $Gain (total,C2) = 0.46880137$.

Dari hasil perhitungan pada tabel 4.1 diperoleh atribut yang menjadi *node* (akar) adalah C2 memiliki gain tertinggi yaitu 0.46880137, dimana terdiri dari 3 sub atribut yaitu Rendah (Berpenghasilan Rendah), Cukup (Berpenghasilan

Cukup), dan Tinggi (Berpenghasilan Tinggi). Berdasarkan nilai *entropy* dari ketiga sub atribut diatas, sub atribut rendah sudah memiliki keputusan dan atribut tinggi sudah memperoleh keputusan, sedangkan sub atribut cukup belum memperoleh keputusan. Maka perlu dilakukan perhitungan lebih lanjut untuk menentukan *node* akar selanjutnya, dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut:



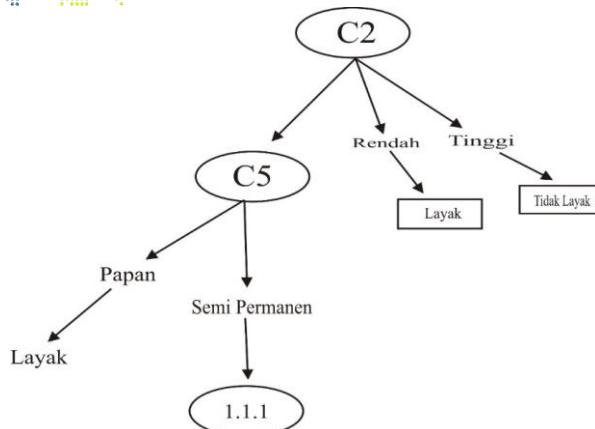
Gambar 1. Pohon Keputusan 1

Untuk hasil perhitungan selanjutnya algoritma C4.5 dapat ditunjukkan pada tabel berikut ini

Tabel 2. Perhitungan *Node* 1.1

	Jumlah (s)	layak	Tidak layak	Entropy	Gain
C2-Cukup	27	13	14	0.9990103	
C1					0.089222904
Wiraswasta	15	6	9	0.9709506	
Buruh	10	5	5	1.0000000	
Guru/Honorer	0	0	0	0.0000000	
Kuli Bangunan	1	1	0	0.0000000	
Karyawan	1	1	0	0.0000000	
ART	0	0	0	0.0000000	
C3					0.1295093
0	6	5	1	0.6500224	
1 dan 2	13	4	9	0.8904916	
>2	8	4	4	1.0000000	
C4					0.041996601
SD	18	8	10	0.9910761	
SMP	8	4	4	1.0000000	
SMA	1	1	0	0.0000000	
S1	0	0	0	0.0000000	
C5					0.4895406
Papan	9	9	0	0.0000000	
Batu	0	0	0	0.0000000	
Semi Permanen	18	4	14	0.7642045	

Dari hasil perhitungan diatas pada tabel 4.2, atribut yang menjadi *Node* cabang dari atribut C2 – Cukup adalah C5. Dimana atribut dari C5 yang terdiri dari Papan, Batu, Semi Permanen. Dimana C5-Papan dan C5-Batu sudah memperoleh keputusan Layak. Sedangkan untuk sub atribut C5-Semi Permanen belum memperoleh keputusan maka akan dilakukan perhitungan kembali. Maka dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut :



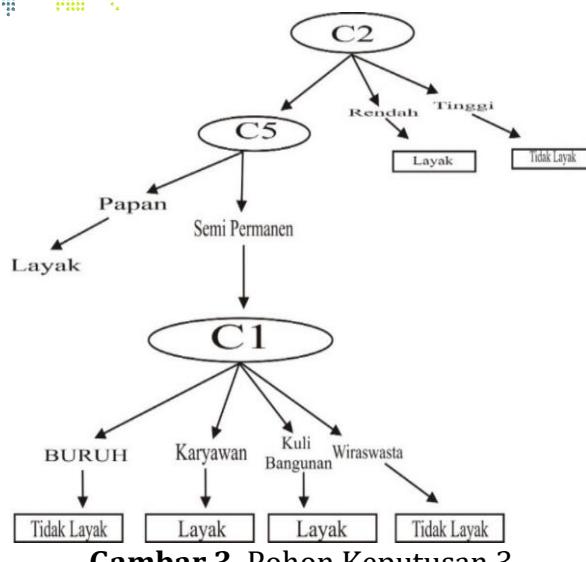
Gambar 2. Pohon Keputusan 2

Dalam mencari hasil perhitungan seanjutnya pada *node* akar C2-Cukup dan C5-Semi Permanen dapat ditunjukkan pada tabel berikut ini.

Tabel 3. Perhitungan *node* 1.1.1

	Jumlah (s)	layak	Tidak layak	Entrophy	Gain
C2-Cukup dan C5-Semi Permanen	18	4	14	0.7642045	
C1					0.346181018
Wiraswasta	11	2	9	0.6840384	
Buruh	5	0	5	0.0000000	
Guru/Honorer	0	0	0	0.0000000	
Kuli Bangunan	1	1	0	0.0000000	
Karyawan	1	1	0	0.0000000	
ART	0	0	0	0.0000000	
C3					0.1148034
0	2	1	1	1.0000000	
1 dan 2	11	1	10	0.4394970	
>2	5	2	3	0.9709506	
C4					0.000805153
SD	13	3	10	0.7793498	
SMP	5	1	4	0.7219281	
SMA	0	0	0	0.0000000	
S1	0	0	0	0.0000000	

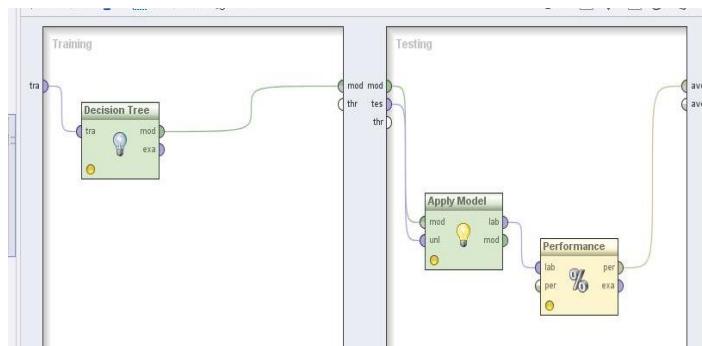
Dari hasil perhitungan pada tabel 4.3, atribut yang menjadi *node* cabang dari C2-Cukup dan C5-Semi Permanen adalah C1. Dimana sub atribut sudah menghasilkan keputusan sehingga perhitungan sudah selesai. Maka dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut :



Gambar 3. Pohon Keputusan 3

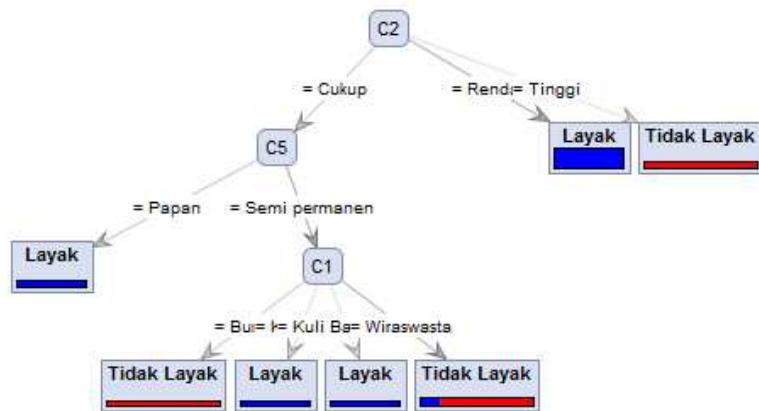
3.2. Pengujian Data Menggunakan Rapidminer

Setelah dilakukan pengolahan data secara manual menggunakan algoritma C4.5, selanjutnya dilakukan pengujian data menggunakan *Tools Rapidminer* untuk mencari rules dan accuracy serta kesamaan dari hasil perhitungan manual dengan menggunakan *tools rapidminer*. Berikut adalah Tampilan dari *port decision tree*, *apply model* dan *performance*.



Gambar 4. Tampilan menghubungkan *port decision tree*, *apply model* dan *performance*

Selanjutnya hubungkan *port-port* dari *operator decision tree*, *operator Apply Model* dan *operator Performance* seperti gambar diatas, kemudian klik *icon Run* pada *toolbar* untuk menampilkan hasil sebagai berikut:



Gambar 5. Hasil *decision tree*

Setelah dilakukan perhitungan dan pengujian data pada masing-masing atribut dengan algoritma C4.5, maka akan didapatkan pola pohon keputusan akhir.



Gambar 6. Deskripsi *decision tree*

Gambar di atas menunjukkan hasil deskripsi secara lengkap dari pohon keputusan (*decision tree*) yang telah terbentuk dengan menggunakan algoritma C4.5. Dari hasil deskripsi juga menunjukkan bahwa penggunaan data mining algoritma C4.5 baik digunakan dalam proses menggali data (*data mining process*) untuk menarik beberapa kesimpulan yang divisualisasikan dengan pohon keputusan (*decision tree*). Berikut ini *rules* yang dihasilkan dari pohon keputusan tersebut:

Tabel 4. *Rule* yang dihasilkan

No	Rules	Keputusan
1.	Jika C2 = Cukup dan C5 = Papan	Layak
2.	Jika C2 = Semi Permanen dan C1 = Buruh	Tidak Layak
3.	Jika C1 = Karyawan	Layak

No	Rules	Keputusan
4.	Jika C1 = Kuli Bangunan	Layak
5.	Jika C1 = Wirausaha	Tidak Layak
6.	Jika C5 = Rendah	Layak
7.	Jika C5 = Tinggi	Layak

accuracy: 100.00%			
	true Layak	true Tidak Layak	class precision
pred. Layak	23	0	100.00%
pred. Tidak Layak	0	3	100.00%
class recall	100.00%	100.00%	

Gambar 7. Nilai Akurasi Algoritma C4.5

Hasil tingkat akurasi dan nilai AUC bisa kita lihat dan ketahui bersama dengan meng-klik tab *PerformanceVector (Performance)*, maka akan terlihat hasil *accuracy* nya seperti gambar diatas. Dimana model yang telah dibentuk diuji tingkat akurasinya dengan memasukan atau uji yang berasal dari data *training* dengan menggunakan *split validation* pada aplikasi *rapidminer studio 5.3* untuk menguji tingkat akurasi. Dengan aplikasi *rapidminer studio 5.3* dihasilkan nilai akurasi, nilai *class recall* dan nilai *class precision* dijelaskan bahwa kategori “Layak” menghasilkan *class recall* sebesar 100.00% dan *class precision* sebesar 100.00 % dan kategori “Berpotensi” menghasilkan *class recall* sebesar 100.00% dan *class precision* sebesar 100 %. Dan hasil akurasi diatas dari perhitungan algoritma C4.5 tersebut adalah sebesar 100.00%.

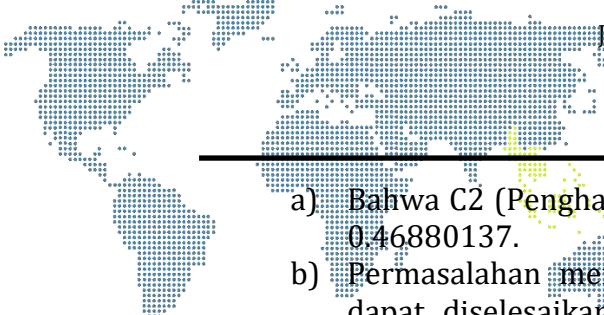
PerformanceVector	
PerformanceVector:	
accuracy:	100.00%
ConfusionMatrix:	
True:	Layak Tidak Layak
Layak:	23 0
Tidak Layak:	0 3

Gambar 8. Nilai Performance Vector Algoritma C4.5

Berdasarkan pengolahan data menggunakan *software RapidMiner* didapat nilai akurasi sistem sebesar 100,00%, artinya bahwa *rule* yang dihasilkan tingkat kebenaran mendekati 100%. Dimana model yang telah dibentuk diuji tingkat akurasinya dengan memasukan atau uji yang berasal dari data *training* dengan menggunakan *split validation* pada aplikasi *rapidminer 5.3* untuk menguji tingkat akurasi.

4. SIMPULAN

Berdasarkan seluruh hasil tahapan penelitian yang telah dilakukan pada Penerapan Algoritma C4.5 dalam mengklasifikasi masyarakat yang berhak menerima bantuan Kartu Keluarga Sejahtera (KKS) secara layak dapat disimpulkan sebagai berikut:

- 
- a) Bahwa C2 (Penghasilan Kepala Keluarga) memiliki gain tertinggi yaitu 0,46880137.
 - b) Permasalahan menentukan masyarakat yang layak dan tidak layak dapat diselesaikan menggunakan teknik *data mining*, yaitu dengan Algoritma C4.5. menghasilkan 8 (Delapan) *rules* dan Tingkat akurasi yang dihasilkan oleh metode tersebut adalah 100,00%.
 - c) Dengan adanya penerapan *data mining* algoritma C4.5 diharapkan mampu memberikan solusi kepada perangkat desa Nagori Mekar Rejo dalam menentukan masyarakat yang berhak dalam menerima bantuan Kartu Keluarga Sejahtera (KKS)

DAFTAR PUSTAKA

- [1] J. Eska, "Penerapan Data Mining Untuk Prekdiksi Penjualan Wallpaper Menggunakan Algoritma C4.5 Stmik Royal Ksiaran," *Jurteksi (Jurnal Teknol. Dan Sist. Informasi)*, Vol. 2, Pp. 9–13, 2016.
- [2] Muhammad, A. P. Windarto, And Suhada, "Penerapan Algoritma C4. 5 Pada Klasifikasi Potensi," Vol. 3, Pp. 1–8, 2019, Doi: 10.30865/Komik.V3i1.1688.
- [3] Rusito And M. T. Firmansyah, "Implementasi Metode Decision Tree Dan Algoritma," 2016.
- [4] Y. S. Luvia, D. Hartama, A. P. Windarto, And Solikhun, "Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di Amik Tunas Bangsa," *Phys. Rev. Mater.*, Vol. 1, No. Issn 2527-5771, P. 7, 2016, Doi: <Https://Doi.Org/10.1103/Physrevmaterials.3.064403>.