Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 6 Nomor 1 Februari, pp 217-225
ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

Analisa Kelayakan Penerima Program Keluarga Harapan (PKH) Menggunakan Algoritma C4.5

Muhammad Amirsyah Tanjung [1], Poningsih[2], Hendry Qurniawan[3]
[1,3]*STIKOM Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia*
[2]*AMIK Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia*
Jln. Sudirman Blok A No. 1-3 Pematangsiantar, Sumatera Utara
[1]amrsyh00@gmail.com, [2] poningsihamiktb@gmail.com

Abstract

In overcoming the problem of poverty, the Government implements the Family Hope Program (PKH) which is a program assistance that provides assistance (subsidies) in cash to poor households as long as they meet the requirements set out in the program. The purpose of this study is to determine whether the family is still eligible or not receive the Family Hope Program (PKH) assistance, where there are many other disadvantaged families who have not had the opportunity to receive this assistance program. Sources of data obtained from the Martoba Village Head Office. The method used in the study is a data mining technique with the C4.5 algorithm which is implemented with the RapidMiner application. The attributes used in determining the family's eligibility are still feasible or not receiving assistance from this assistance program, namely income, number of family dependents, vehicle ownership and residence status. The results of the classification algorithm C4.5 and testing with Rapid Miner software, it is found that the factor that most influences the eligibility of the Family Hope Program (PKH) recipients is the number of stages (C1) with an acquisition value of 0.51827179.

Keywords*: C4.5 Algorithm, Data mining, PKH*

Abstrak

Dalam menanggulangi masalah kemiskinan, Pemerintah melaksanakan Program Keluarga Harapan (PKH) yang merupakan bantuan program yang memberikan bantuan (subsidi) tunai kepada rumah tangga miskin sepanjang mereka memenuhi persyaratan yang ditetapkan dalam program. Tujuan penelitian ini untuk menentukan apakah keluarga tersebut masih layak atau tidak menerima bantuan Program Keluarga Harapan (PKH) ini, dimana masih banyak keluarga tidak mampu lainnya yang belum berkesempatan menerima program bantuan ini. Sumber data yang diperoleh dari Kantor Lurah Martoba. Metode yang digunakan dalam penelitian adalah teknik data mining dengan algoritma C4.5 yang diimpelementasikan dengan aplikasi RapidMiner. Atribut yang digunakan dalam menentukan kelayakan keluarga tersebut masih layak atau tidak menerima bantuan program bantuan ini yaitu penghasilan, jumlah tanggungan keluarga, kepemilikan kendaraan dan status tempat tinggal. Hasil dari klasifikasi menggunakan algoritma C4.5 dan pengujian dengan software RapidMiner maka didapatkan faktor yang paling berpengaruh terhadap kelayakan penerima Program Keluarga Harapan (PKH) adalah jumlah penghasilan (C1) dengan nilai gain sebesar 0,51827179.

Kata kunci*: Algoritma C4.5, Data mining, PKH*

1. PENDAHULUAN

Kemiskinan adalah kondisi ketidakmampuan seseorang dalam memenuhi kebutuhan sehari-hari. “Seseorang disebut miskin dapat dilihat dari pendapatannya, apabila pendapatannya rendah maka seseorang tersebut tidak dapat memenuhi kebutuhannya dan tidak dapat mensejahterakan dirinya” [1]. Dalam menanggulangi kemiskinan tersebut pemerintah melaksanakan Program Keluarga Harapan (PKH) yang merupakan program bantuan yang memberikan bantuan tunai kepada rumah tangga miskin sepanjang mereka memenuhi

persyaratan yang ditetapkan dalam program tersebut. Persyaratan tersebut terdiri dari beberapa komponen yaitu ibu hamil, memiliki anak usia dini, SD, SMP, SMA, disabilitas berat, lanjut usia. Pada saat ini Pemerintah melakukan sistem labelisasi rumah pada setiap rumah yang menerima PKH. Sistem labelisasi rumah tersebut merupakan bentuk sosialisasi yang diupayakan Sumber Daya Manusia (SDM) PKH agar keluarga yang telah mampu mempunyai budaya malu. Namun dilapangan, banyak keluarga yang sudah mampu yang belum mengundurkan diri dari penerimaan bantuan program ini, sehingga tidak memberikan kesempatan kepada keluarga tidak mampu lainnya yang belum berkesempatan untuk mendapatkan bantuan program ini.

Untuk mengatasi permasalahan diatas dapat menggunakan klasifikasi dengan Algoritma C4.5. "Klasifikasi merupakan proses pencarian model yang membedakan kelas data agar model tersebut dapat digunakan untuk memprediksi kelas dari suatu objek yang belum diketahui kelasnya. Salah satu algoritma klasifikasi yang bisa digunakan adalah C4.5" [2]. Dengan menggunakan metode Algoritma C4.5 akan diklasifikasi data penerima PKH untuk memberikan masukan kepada pemerintah dalam menentukan apakah keluarga tersebut masih layak atau tidak menerima bantuan. Kajian penelitian terdahulu yaitu (Irmayansah 2018) yang berjudul "Penerapan Algoritma C4.5 Untuk Klasifikasi Penentuan Penerimaan Bantuan Langsung di Desa Ciomas". Penelitian ini mengklasifikasikan penduduk yang tidak mampu berdasarkan sensus ekonomi penduduk dengan 10 *variabel* yang digunakan yaitu atap rumah, jenis pekerjaan, jenis dinding rumah terluas, jenis lantai rumah terluas, sumber penerangan umum, sumber air minum, bahan bakar utama untuk memasak, fasilitas tempat buang air besar, status rumah, luas rumah. Pada penelitian tersebut perlu ditambahkannya variabel penghasilan terhadap penduduk tersebut dimana variabel ini cukup mempengaruhi penduduk tersebut untuk mendapatkan penerimaan bantuan.

Dengan Tujuan untuk menentukan apakah keluarga tersebut masih layak atau tidak menerima bantuan program ini, dimana masih banyak keluarga tidak mampu lainnya yang belum berkesempatan untuk mendapatkan program bantuan ini. Hasil dari penelitian ini diharapkan dapat memberikan informasi dan masukan kepada Pemerintah dalam menentukan apakah keluarga yang sudah menerima bantuan masih layak atau tidak menerima bantuan Program Keluarga Harapan (PKH) ini, sehingga memberikan kesempatan kepada keluarga yang tidak mampu lainnya untuk mendapatkan bantuan program ini.

2. METODOLOGI PENELITIAN

2.1. Pengumpulan Data

Penelitian ini dilakukan di Kelurahan Martoba yang terletak di Kecamatan Siantar Utara, Kota Pematangsiantar. Pengumpulan data yang penulis lakukan diperoleh melalui Kantor Lurah Martoba akan di olah sebanyak 163 keluarga yang menerima PKH. Penulis melakukan observasi kepada keluarga yang menerima PKH dengan jumlah 100 sampel data.

Atribut data yang digunakan terdiri dari :

a) Jumlah Penghasilan
b) Jumlah Tanggungan Keluarga
c) Kepemilikan Kendaraan
d) Status Tempat Tinggal

Penjelasan atribut dapat dilihat pada tabel dibawah ini :

Tabel 1. Penjelasan Atribut

No.	Atribut	Penjelasan
1.	Jumlah Penghasilan	Rendah (Penghasilan <1jt perbulan), Cukup (Penghasilan ≥1jt - ≤1,5jt perbulan), Tinggi (Penghasilan >1,5jt perbulan)
2.	Jumlah Tanggungan Keluarga	0 (Tidak mempunyai anak), 1 (Jumlah anak 1) 2 (Jumlah anak 2), >2 (Jumlah anak >2)
3.	Kepemilikan Kendaraan	Tidak Punya (Tidak mempunyai kendaraan), Rendah (Sepeda), Cukup (1-2 Sepeda Motor), Tinggi (>2 Sepeda Motor, (Mobil))
4.	Status Tempat Tinggal	Numpang, Kontrak, Milik Sendiri

2.2. Metode Penelitian

a) Data Mining

Data Mining adalah proses mencari pola atau informasi menarik dalam data terpilih menggunakan teknik atau metode tertentu[3]. Data mining dapat diartikan juga sebagai pengesktrakan informasi baru dari data besar. Istilah data mining kadang disebut juga *Knowledge Discovery in Database* (KDD) [4].

b) Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang dicipt9akan oleh J. Rose Quinlan [3]. Algoritma C4.5 merup9akan salah satu metode klasifikasi yang menggunakan represe9ntasi struktur pohon (*tree*) dimana setiap *node* merepre9sentasikan atribut, cabangnya merepresentasikan nilai dari atrib9ut dan daun merepresentasikan kelas [7]. Menurut (Muhamad, Windarto & Suhada 2019) secara umum untuk membangun pohon keputusan Algoritma C4.5 adalah sebagai berikut :

1) Pilih atribut sebagai akar
2) Buat cabang untuk masing-masing nilai
3) Bagi kasus dalam cabang
4) Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

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

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

Keterangan:
S : Himpunan kasus

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 6 Nomor 1 Februari, pp 217-225
ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

A : Atribut
n : Jumlah partisi atribut A
|Si| : Jumlah kasus pada partisi ke i
|S| : Jumlah kasus dalam S

Sebelum mendapatkan nilai *gain* maka dicari terlebih dahulu nilai *entropy*. Untuk mencari nilai *entropy* menggunakan rumus sebagai berikut :

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

S : Himpunan Kasus
n : Jumlah partisi S
pi: Proporsi dari Si 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 *entrophy* dari semua kasus yang dibagi berdasarkan Jumlah Penghasilan (C1), Jumlah Tanggungan (C2), Kepemilikan Kendaraan (C3) dan Status Tempat Tinggal (C4). Setelah itu dilakukan perhitungan *gain* untuk masing – masing atribut. Hasil perhitungan ditunjukkan oleh tabel berikut ini.

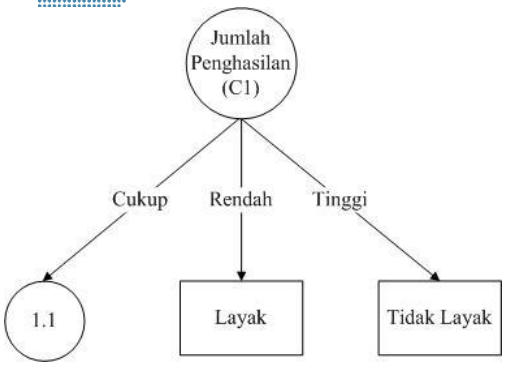
Tabel 2. Perhitungan *Node*

	Jumlah (s)	**Layak**	**Tidak Layak**	***Entrophy***	***Gain***
Total	100	69	31	0,8931735	
Jumlah Penghasilan					0,51827179
	Rendah	40	40	0	0,0000000
	Cukup	42	29	13	0,8926230
	Tinggi	18	0	18	0,0000000
Jumlah Tanggungan					0,40370163
	1	27	13	14	0,9990103
	2	42	31	11	0,8296071
	>2	31	25	6	0,7088357
Kepmilikan Kendaraan					0,12702864
	Rendah	0	0	0	0,0000000
	Cukup	93	69	24	0,8238116
	Tinggi	7	0	7	0,0000000
Status Tempat Tinggal					0,18777421
	Kontrak	90	69	21	0,7837769
	Milik Sendiri	10	0	10	0,0000000

Dari hasil perhitungan pada tabel 3.1 diperoleh atribut yang menjadi *node* (akar) adalah C1 memiliki gain tertinggi yaitu 0,5568321, dimana terdiri dari 3 sub atribut yaitu Rendah (Berpenghasilan Rendah), Cukup (Berpenghasilan Cukup) dan Tinggi (Berpenghasilan Tinggi). Berdasarkan nilai *entrophy* dari ketiga sub atribut diatas, sub atribut rendah sudah memiliki keputusan dan atribut tinggi sudah memperoleh keputusan, sedangkan sub atribut cukup belum memperoleh

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 6 Nomor 1 Februari, pp 217-225
ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

keputusan. Maka perlu dilakukan perhitungan lebih lanjut untuk menetukan *node* akar selanjutnya, dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut:



Gambar 1. Pohon Keputusan 1

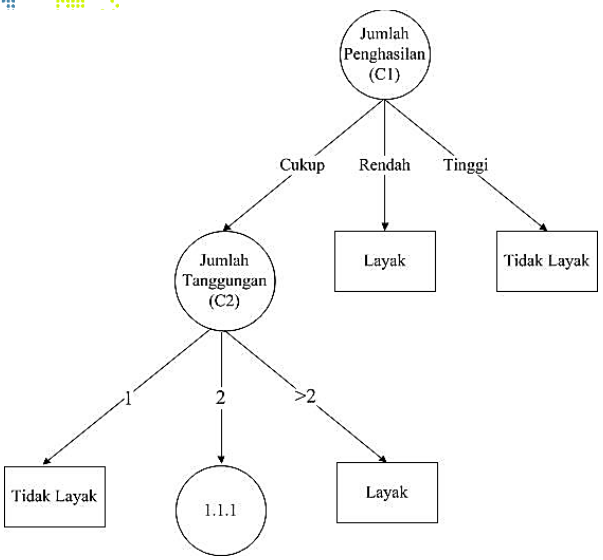
Untuk hasil perhitungan selanjutnya algoritma C4.5 dapat ditunjukkan pada tabel berikut ini.

Tabel 3. Perhitungan *Node* 1.1

penghasilan = cukup		**Jumlah (s)**	**Layak**	**Tidak Layak**	***Entrophy***	***Gain***
		42	29	13	0,8926230	
Jumlah Tanggungan						0,89262301
	1	11	0	11	0,0000000	
	2	13	11	2	0,6193822	
	>2	18	18	0	0,0000000	
Kepmilikan Kendaraan						0,04122698
	Rendah	0	0	0	0,0000000	
	Cukup	41	29	12	0,8721618	
	Tinggi	1	0	1	0,0000000	
Status Tempat Tinggal						0,17808797
	Kontrak	38	29	9	0,7897493	
	Milik Sendiri	4	0	4	0,0000000	

Dari hasil perhitungan diatas pada tabel 3.2, atribut yang menjadi *Node* cabang dari atribut C1 – Cukup adalah Jumlah Tanggungan (C2). Dimana atribut dari Jumlah Tanggungan (C2) yang terdiri dari 1, 2 dan >2. Dimana Jumlah Tanggungan (C2) 1 dan >2 sudah memperoleh keputusan Tidak Layak. Sedangkan untuk sub atribut Jumlah Tanggungan (C2) 2 belum memperoleh keputusan maka akan dilakukan perhitungan kembali. Maka dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut :

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 6 Nomor 1 Februari, pp 217-225
ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik



Gambar 2. Pohon Keputusan 2

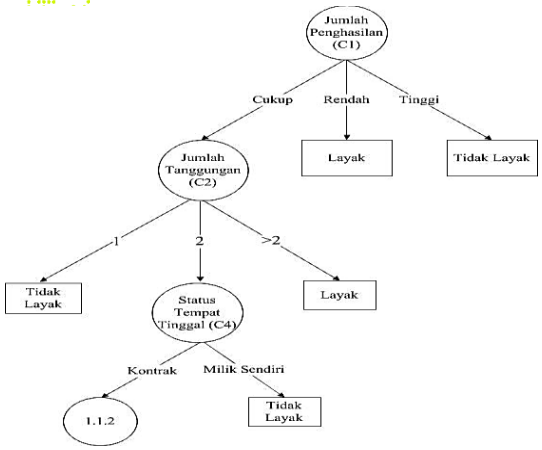
Dalam mencari hasil perhitungan selanjutnya pada *node* akar C1-Cukup dan C2-2 dapat ditunjukkan pada tabel berikut ini.

Tabel 4. Perhitungan *node* 1.1.1

penghasilan = cukup dan tanggungan 2		**Jumlah (s)**	**Layak**	**Tidak Layak**	***Entrophy***	***Gain***
		13	11	2	0,6193822	
Kepmilikan Kendaraan						0
	Rendah	0	0	0	0,0000000	
	Cukup	13	11	2	0,6193822	
	Tinggi	0	0	0	0,0000000	
Status Tempat Tinggal						0,23739741
	Kontrak	12	11	1	0,4138169	
	Milik Sendiri	1	0	1	0,0000000	

Dari hasil perhitungan pada tabel 3.3, atribut yang menjadi *node* cabang dari C1-Cukup dan C2-2 adalah C4. Dimana Status Tempa Tinggal (C4) Milik Sendiri keputusan Tidak Layak. Sedangkan untuk sub atribut Status Tempat Tinggal (C4) Kontrak belum memperoleh keputusan maka akan dilakukan perhitungan kembali. Maka dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut :

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 6 Nomor 1 Februari, pp 217-225
ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

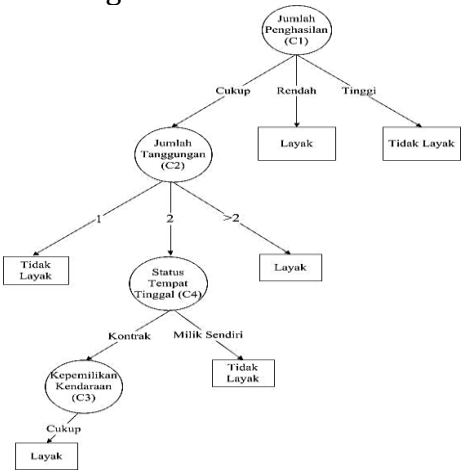


Gambar 3. Pohon Keputusan 3

Tabel 5. Perhitungan *node* 1.1.2

Penghasilan = cukup dan tanggungan 2 dan Kontrak		**Jumlah (s)**	**Layak**	**Tidak Layak**	***Entrophy***	***Gain***
		12	11	1	0,4138169	
Kepmilikan Kendaraan						0
	Rendah	0	0	0	0,0000000	
	Cukup	12	11	1	0,4138169	
	Tinggi	0	0	0	0,0000000	

Dari hasil perhitungan pada tabel 4.4, atribut yang menjadi *node* cabang dari C1-Cukup, C2-2 dan C4-Kontrak adalah C3. Dimana semua atribut sudah memiliki keputusan dan perhitungan sudah selesai. Maka dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut :



Gambar 4. Pohon Keputusan 4

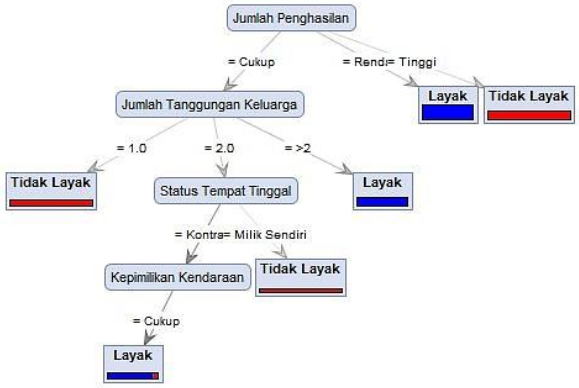
Berikut ini *rules* yang dihasilkan dari pohon keputusan tersebut :

Tabel 6. *Rule* yang dihasilkan

No	Rules	Keputusan
1.	Jika C1 = Cukup dan C2 = 1	Tidak Layak
2.	Jika C1 = Cukup, C2 = 2, C4 = Kontrak dan C3 = Cukup	Layak
3.	Jika C1 = Cukup, C2 = 2 dan C4 = Milik Sendiri	Tidak Layak
4.	Jika C1 = Cukup dan C2 = >2	Layak
5.	Jika C1 = Rendah	Layak
6.	Jika C1 = Tinggi	Tidak Layak

3.2. Pengujian Data Menggunakan Rapidminer

a) Setelah melakukan proses pengolahan data secara manual menggunakan metode algoritma c4.5 maka selanjutnya dilakukan proses pengujian data menggunakan *tools rapidminer*. Berikut adalah hasil pengujian data menggunakan *tools rapidminer.* Tampilan *decision tree*



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.

b) Tingkat Akurasi

PerformanceVector

```
PerformanceVector:
accuracy: 96.67%
ConfusionMatrix:
True:   Layak   Tidak Layak
Layak:  20      1
Tidak Layak:    0       9
```

Gambar 6. Nilai Performance Vector Algoritma C4.5

Berdasarkan pengolahan data menggunakan *software RapidMiner* didapat nilai akurasi sistem sebesar 96,67%, 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 masih berhak atau tidak berhak menerima bantuan Program Keluarga Harapan (PKH) yang merupakan program bantuan yang dilaksanakan oleh Pemerintah dalam menanggulangi masalah kemiskinan dapat disimpulkan sebagai berikut :

a) Permasalahan menentukan masyarakat yang layak dan tidak layak dapat diselesaikan menggunakan teknik *data mining*, yaitu dengan Algoritma C4.5. menghasilkan 6 (Enam) *rules* dan tingkat akurasi yang dihasilkan oleh metode tersebut adalah 96,67%.
b) Dengan adanya penerapan *data mining* algoritma C4.5 diharapkan mampu memberikan solusi kepada Pemerintah Kelurahan Martoba Kecamatan Siantar Utara Kota Pematangsiantar dalam menentukan masyarakat yang berhak dalam menerima bantuan Program Keluarga Harapan (PKH).
c) Dengan adanya penerapan *data mining* algoritma C4.5 diharapkan mampu mempermudah Pemerintah Kelurahan Martoba Kecamatan Siantar Utara Kota Pematangsiantar dalam memberikan bantuan Program Keluarga Harapan (PKH) ini kepada masyarakat yang benar-benar layak dan tidak salah sasaran.

DAFTAR PUSTAKA

[1] V. C. Pamungkas, L. Muflikhah, And R. C. Wihandika, "Klasifikasi Penerimaan Program Keluarga Harapan (Pkh) Menggunakan Metode Learning Vector Quantization (Studi Kasus Desa Kedungjati)," Vol. 3, No. 3, Pp. 2659–2666, 2019.

[2] A. H. Nasrullah, "Penerapan Metode C4.5 Untuk Klasifikasi Mahasiswa Berpotensi Drop Out," Vol. 10, Pp. 244–250, 2018.

[3] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform.*, Vol. 2, No. 2, Pp. 213–219, 2019.

[4] S. Haryati, A. Sudarsono, And E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *J. Media Infotama*, Vol. 11, No. 2, Pp. 130–138, 2015.

[5] E. P. Cynthia And E. Ismanto, "Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji," *Jurasik (Jurnal Ris. Sist. Inf. Dan Tek. Inform.*, Vol. 3, No. July, P. 1, 2018, Doi: 10.30645/Jurasik.V3i0.60.

[6] B. Hermanto And A. Sn, "Klasifikasi Nilai Kelayakan Calon Debitur Baru Menggunakan Decision Tree C4.5," *Ijccs (Indonesian J. Comput. Cybern. Syst.*, Vol. 11, No. 1, P. 43, 2017, Doi: 10.22146/Ijccs.15946.

[7] M. A. Sembiring, "Penerapan Metode Decision Tree Algoritma C45 Untuk Memprediksi Hasil Belajar Mahasiswa Berdasarkan Riwayat Akademik," *J. Teknol. Dan Sist. Inf.*, Vol. 3, No. 1, Pp. 60–65, 2016.

[8] S. Muhamad, Agus Perdana Windarto, "Penerapan Algoritme C4 . 5 Pada Klasifikasi," *Komik (Konferensi Nas. Teknol. Inf. Dan Komputer)*, Vol. 3, Pp. 477–482, 2017, Doi: 10.30865/Komik.V3i1.1688.