

METODE POHON KEPUTUSAN MENGGUNAKAN ALGORITMA C4.5 UNTUK PENGELOMPOKAN DATA PENDUDUK PADA TINGKATAN KESEJAHTERAAN KELUARGA

Lilik Dwi Astuti¹, Hanny Haryanto²

^{1,2}Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
Jalan Iman Bonjol No. 207, Semarang, 50131, (024)3517261
E-mail : 111201106052@mhs.dinus.ac.id¹, 1

Abstrak

Tingkatan keluarga sejahtera merupakan tingkatan dari kesejahteraan keluarga yang berguna sebagai pembinaan dengan kesesuaian dengan indikator dari tiap kesejahteraan keluarganya. Pembinaan keluarga sejahtera bertujuan untuk memperbaiki permasalahan ekonomi maupun non ekonomi yang ada dari tiap keluarga. Dalam pelaksanaannya KBPP dibantu oleh PLKB di kecamatan bertugas untuk melakukan survey data per kepala keluarga. Dalam pengumpulan data PLKB masih menggunakan cara manual yaitu menulis data per kepala keluarga. Setelah itu, data baru diproses satu persatu untuk menghitung tingkatan keluarga sejahtera. Banyaknya data yang dikumpulkan oleh PLKB menyulitkan untuk dilakukan klasifikasi data kedalam tingkatan kesejahteraan keluarga yang dapat memperlambat kinerja PLKB. Untuk menentukan tingkatan keluarga sejahtera dihitung berdasarkan 21 indikator yang dikelompokkan menjadi 5 tahap keluarga sejahtera. Pola-pola dalam kesejahteraan keluarga dapat diklasifikasikan menggunakan metode data mining dengan algoritma C4.5, menghasilkan aturan klasifikasi yang menghasilkan tahapan keluarga sejahtera. Metode klasifikasi algoritma C4.5 berhasil diimplementasikan pada data kesejahteraan keluarga dusun Joho kecamatan Pringapus, kabupaten Semarang. Dari data tersebut menghasilkan nilai akurasi sebesar 95%. Sehingga penelitian dikatakan berhasil karena dapat membantu meringankan tugas PLKB yaitu mengklasifikasikan data kedalam jenis tingkatan kesejahteraan keluarga menggunakan pohon keputusan data mining berbasis algoritma C4.5.

Kata Kunci: pohon keputusan, data mining, algoritma C4.5, klasifikasi kesejahteraan keluarga, menentukan kesejahteraan keluarga.

Abstract

Prosperous family level is the level of family welfare are useful as guidance in conformity with the indicators of each family welfare. Coaching of prosperous family aims to improve the economic and non-economic issues that exist for each family. In the implementation, KBPP assisted by PLKB on duty in the district to conduct a survey of data per household. In the data collection PLKB still use manual way is to write data per household. After that, the new data is processed one by one to calculate the level of family welfare. The amount of data collected by PLKB made it difficult for the classification of data into family welfare levels that can slow down the performance of PLKB. To determine the level of family welfare is calculated based on 21 indicators grouped into 5 stages prosperous family. Patterns in family welfare could be classified using data mining method with C4.5 algorithm, generates rules that generate classifications stages prosperous family. C4.5 classification algorithm method successfully implemented in the data hamlet family welfare Joho Pringapus districts, counties Semarang. From these data, the result of accuracy value is 95%. So the research is successful because it can help ease the task of PLKB which is to classify data into types of family welfare levels using a decision tree data mining based on C4.5 algorithm.

Keywords: decision tree, data mining, C4.5 algorithm, the classification of the welfare of the family, determine the welfare of the family

1. PENDAHULUAN

Pengumpulan data sub bidang KS dibantu PLKB (Petugas Lapangan Keluarga Berencana) kecamatan untuk melakukan sosialisasi secara langsung. Setiap data yang diperoleh dalam kegiatan akan menunjukkan tingkat kesejahteraan keluarga. Tingkatan kesejahteraan keluarga dikelompokkan menjadi lima tahap antara lain : tahapan Keluarga Pra Sejahtera (KPS), tahapan Keluarga Sejahtera I (KS I), tahapan Keluarga Sejahtera II (KS II), Tahapan Keluarga Sejahtera III (KS III), dan tahapan Keluarga Sejahtera III plus (KS III plus) [2]. Dari tahapan keluarga sejahtera terbagi menjadi 21 indikator tahapan keluarga sejahtera. Terdapat 6 tahapan di KS I, 8 indikator KS II, 5 indikator KS III, 2 indikator KS III plus [2]. Banyaknya data yang berkaitan dengan tingkat kesejahteraan keluarga yang dikumpulkan tersebut mengandung berbagai jenis pengetahuan yang berguna untuk memproses pengambilan keputusan, diantaranya adalah yang berkaitan dengan klasifikasi atau pengelompokan data ke tingkatan-tingkatan yang sudah ditentukan di atas, akan tetapi pola data dan pengetahuan tersebut sangat sulit untuk ditemukan dengan cara menganalisis secara manual. Seiring berkembangnya ilmu pengetahuan dan teknologi maka berkembang pula kemampuan dalam mengumpulkan, menganalisa, dan mengolah data pada *database*. Pencarian atau proses penemuan informasi dan pengetahuan yang terkandung di dalam banyaknya data tersebut dapat dilakukan dengan teknologi *data mining*. Dalam kasus ini data mining dapat diterapkan untuk melakukan klasifikasi untuk jenis tingkatan keluarga sejahtera.

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam *database* atau sering disebut *Knowledge*

Discovery in database[3]. *Data mining* digunakan sebagai perhitungan *rule* guna menemukan kelayakan dari data set. Didalam *data mining* terdapat metode klasifikasi. Metode klasifikasi berguna bekerja dengan cara meletakkan data kedalam salah satu dari beberapa kelas yang sudah didefinisikan sebelumnya. Misalnya mengklasifikasikan antara adanya anggota keluarga yang memiliki penghasilan tetap dan keluarga yang memiliki penghasilan tidak tentu. Maka dari data tersebut akan diklasifikasikan kedalam kelas-kelas yang berbeda. Dari data klasifikasi diperoleh kelas-kelas yang dihasilkan dari pohon keputusan (*decision tree*). Kelebihan dari pohon keputusan yaitu dapat menghasilkan aturan yang mudah diimplementasikan dengan cepat. Pohon keputusan dapat dihasilkan dengan berbagai metode, salah satunya adalah algoritma C4.5 Sehingga pengolahan data dilakukan secara komputersasi yaitu dengan menggunakan teknik *data mining*. *Data mining* yaitu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam *database*.

2. METODE

Dalam penelitian di kantor PLKB kecamatan Pringapus melalui persetujuan dari kesbanglinmas kabupaten Semarang yang kemudian dilanjutkan dengan wawancara di KBPP Ungaran dan PLKB Pringapus. Data diambil dari PLKB. Penelitian menggunakan perangkat lunak : Sistem operasi Microsoft Windows 7 32 bit. Dan rapid miner. Dan menggunakan perangkat keras : Processor intel (R) pentium CPU U 5600 @ 1.33 GHz 1.33GHz, RAM 2 GB, dan Printer HP 704. Teknik pengumpulan data yaitu dengan data primer dengan wawancara kepada ibu Srie lestari, S.H selaku ketua subbidang KS di KBPP dan data

diperoleh dari PLKB kecamatan Pringapus. Dan data sekunder diperoleh dari buku, jurnal ilmiah, laporan penelitian yang relevan yang berhubungan dengan kesejahteraan keluarga ataupun klasifikasi C4.5 dan formulir pemutakhiran data keluarga yang didapat dari PLKB sebagai penunjang penyusunan tugas akhir.

Metode analisa digunakan untuk penerapan teknik *data mining* pada keluarga sejahtera yaitu CHRISP-DM yang merupakan standar dari yang digunakan untuk pemecahan permasalahan umum pada permasalahan bisnis ataupun unit penelitian *data mining*. Terdapat enam fase pada CHRISP-DM yaitu : fase pemahaman bisnis, fase pemahaman data, fase pengolahan data, fase pemodelan, fase evaluasi, fase penyebaran.

2.1 Fase Pemahaman Bisnis

Pada fase ini, kesejahteraan keluarga merupakan tercipta suatu keadaan yang harmonis dengan terpenuhi kebutuhan jasmani dan sosial untuk seluruh anggota keluarga, tanpa adanya rintangan yang serius dalam menghadapi permasalahan keluarga, sehingga dapat mewujudkan standar keluarga [18]. Banyak faktor yang mempengaruhi permasalahan keluarga, diantaranya faktor kemiskinan, faktor jumlah keluarga, faktor pekerjaan, maupun faktor pendapatan. Faktor-faktor tersebut sebagai tolak ukur penentu kesejahteraan setiap keluarga.

2.2 Fase Pemahaman Data

Fase ini merupakan fase pengumpulan data. Data penelitian diperoleh dari observasi yang dilakukan oleh PLKB kecamatan Pringapus Kabupaten Semarang tahun 2014. Data yang dikumpulkan yaitu data keluarga di dusun Joho, Kelurahan Wonoyoso, Kecamatan Pringapus, Kabupaten Semarang, Jawa Tengah.

Gambar 1. Sampel data mentah keluarga dusun Joho kelurahan Wonoyoso

Data dengan jumlah atributnya 21 macam. Data disimpan dalam bentuk hardcopy yang selanjutnya ditransformasi ke dalam format Microsoft Excel.

2.3 Fase Pengolahan Data

Data penelitian ini diperoleh dari observasi yang dilakukan oleh PLKB kecamatan Pringapus. Pada tahap ini aktivitas-aktivitas dalam penyusunan dataset akhir dilakukan, seperti pemilihan data, pemilihan atribut, variabel yang dibutuhkan, pembersihan data, dan transformasi data.

Data yang didapat masih berupa data mentah dalam bentuk hardcopy. Data kemudian dibersihkan dengan cara menghapus atribut-atribut yang tidak diperlukan untuk penelitian dan juga menghapus data-data yang *record* fieldnya tidak lengkap. Data yang telah di proses selanjutnya diubah kedalam file dengan format *.xls.

2.4 Fase Pemodelan

Fase selanjutnya adalah pemilihan teknik data mining yang akan dipakai dan selanjutnya membangun model *data mining*. Teknik *data mining* yang dipilih dalam penelitian ini adalah teknik *Decision Tree*. Pemodelan bertujuan menghasilkan pohon keputusan dan aturan-aturan klasifikasi dimana aturan klasifikasi ini selanjutnya dijadikan acuan untuk menentukan keputusan pemberian status keluarga

sejahtera.

Proses awal atau iterasi pertama, yang dilakukan adalah menghitung nilai entropy dari label, menggunakan rumus :

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

Setelah itu ditranslasikan pada masalah yang ada.

yaitu :

$$\text{Entropy}(\text{Hasil}) = \text{Entropi}(kps, ks1, ks2, ks$$

3)

$$= - ((kps/\text{jumlah data}) * \log_2 (kps/\text{jumlah data})) - ((ks1/\text{jumlah data}) * \log_2 (ks1/\text{jumlah data})) - ((ks2/\text{jumlah data}) * \log_2 (ks2/\text{jumlah data})) - ((ks3/\text{jumlah data}) * \log_2 (ks3/\text{jumlah data})). \quad (2)$$

Setelah mendapatkan hasil entropi selanjutnya periksa seluruh atribut untuk mencari atribut yang akan menjadi *root*. Hitung nilai entropi dan gain dari masing-masing atribut (21 atribut). Mulai dari atribut '1'. sampai dengan atribut '21'.

Setelah menghitung entropi di setiap atribut maka langkah selanjutnya adalah menghitung gain. Cara menghitung gain adalah:

$$\text{Gain}(\text{Hasil}, 1) = \text{Entropy}(\text{Hasil}) - \left(\frac{20}{20} * \text{Entropy}(\text{yes}) + \frac{0}{20} * \text{Entropy}(\text{no}) \right) \quad (3)$$

Gain tertinggi akan menjadi *root* diawal. Apabila terdapat beberapa atribut dengan gain tertinggi, maka atribut yang lebih dulu diperiksa dianggap yang memiliki gain tertinggi. Apabila pada data dengan gain tertinggi memiliki sampel tunggal maka otomatis label data tunggal akan menjadi daun dengan label nilai tunggal tersebut. Sedangkan data yang tidak memiliki sampel tunggal akan dilakukan iterasi selanjutnya untuk mencari *node*/simpul dari cabang ini. Atribut yang telah menjadi *node* tidak akan diperiksa lagi, sehingga disetiap prosesnya data akan

berkurang .selanjutnya pemeriksaan data dilanjutkan ke sisa data yang belum dihiung dengan sisa label yang bukan menjadi label tunggal pada perhitungan sebelumnya. Apabila terdapat hasil gain pada iterasi menunjukkan jumlah yang sama, maka dipilih atribut yang memenuhi yaitu yang memiliki jumlah sampel tunggal terbesar.

Iterasi dilanjutkan sampai seluruh atribut menghasilkan output daun atau sampai tidak ada atribut lagi untuk diperiksa.

2.5 Fase Evaluasi

Pada fase ini dilakukan pengujian keakuratan terhadap model yang telah dihasilkan untuk meneliti pemenuhan tujuan yang ada. Pada proses ini pengukuran hasil tingkat akurasi dari aturan klasifikasi yang dihasilkan, dihitung menggunakan tabel *confusion matrix*. Hasil akurasi dapat digunakan untuk menentukan apakah aturan-aturan dari pohon keputusan akan diterapkan atau tidak. Akurasi dari *decision tree* yang dihasilkan dapat dihitung dengan rumus (10) namun karena terdapat 4 label pada kasus ini maka rumus yang digunakan adalah :

$$\text{Accuracy} = \frac{T_{kps} + T_{ks1} + T_{ks2} + T_{ks3}}{T_{kps} + T_{ks1} + T_{ks2} + T_{ks3} + F_{kps} + F_{ks1} + F_{ks2} + F_{ks3}} * 100\% \quad (4)$$

2.5 Fase Penyebaran

Setelah fase Evaluasi selesai pada fase ini model dapat disebar dan digunakan. Hasil dari penelitian ini berupa analisa ke *Decision Support System* (DSS), yang diharapkan dapat digunakan oleh petugas penyuluh tingkat thapan keluarga sejahtera di KBPP kabupaten Semarang untuk menentukan status keluarga sejahtera berdasarkan 21 indikator data dan juga dapat dikembangkan untuk penelitian selanjutnya.

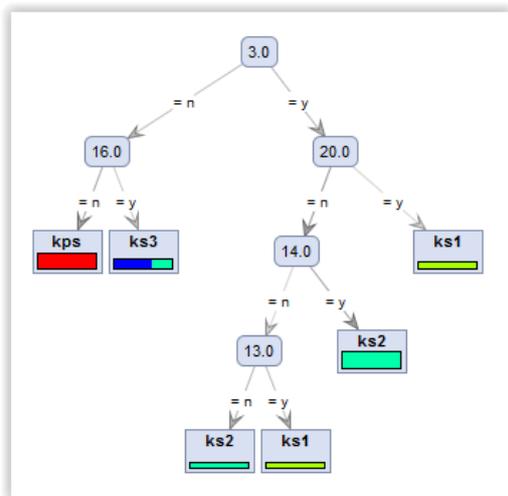
3. HASIL DAN PEMBAHASAN

Data yang digunakan adalah data dari form pemutakhiran data keluarga yang ditransformasi kedalam bentuk *.xls Data terdiri dari 21 atribut dengan status 5 tahapan yaitu : tahapan Keluarga Pra Sejahtera (KPS), Keluarga Sejahtera 1 (KS1), Keluarga Sejahtera 2 (KS2), Keluarga Sejahtera 3 (Ks3), Keluarga Sejahtera 3 Plus (KS 3 Plus). dengan sampel uji sebanyak 20 data

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	HSL
y	y	y	y	y	y	y	y	y	Y	y	y	y	y	Y	Y	y	Y	Y	N	n	ks3
y	y	y	y	y	y	y	y	Y	y	y	y	y	Y	N	n	Y	Y	Y	N	n	ks2
y	y	y	y	y	y	y	y	Y	y	y	Y	Y	Y	Y	n	Y	Y	Y	N	n	ks2
y	y	y	y	y	y	y	y	Y	y	y	Y	n	Y	N	y	Y	Y	Y	N	n	ks1
y	y	y	y	y	y	y	y	Y	y	n	N	n	Y	Y	n	Y	Y	Y	N	n	ks2
y	y	y	y	y	y	y	y	n	n	n	y	Y	y	Y	N	n	Y	Y	Y	n	ks1
y	y	y	y	y	y	Y	y	y	y	y	Y	y	Y	Y	n	N	Y	Y	N	n	ks2
y	y	y	y	y	y	Y	y	y	y	y	N	y	Y	N	n	Y	Y	Y	Y	n	ks1
y	y	y	y	y	y	Y	y	n	y	n	y	Y	y	Y	N	n	N	Y	N	n	ks2
y	y	n	y	y	y	Y	y	y	y	Y	y	Y	Y	Y	n	Y	Y	Y	n	y	kps
y	y	y	y	y	y	Y	y	y	y	y	Y	y	Y	n	Y	Y	N	n	n	ks2	ks2
y	y	y	y	y	y	Y	y	y	y	y	Y	y	Y	Y	Y	N	Y	Y	n	n	ks3
y	n	y	y	y	Y	y	y	y	y	Y	y	Y	N	n	Y	Y	Y	Y	n	n	kps
y	n	y	y	y	Y	y	y	y	y	Y	y	Y	N	n	Y	Y	Y	Y	n	n	kps
y	n	y	y	y	Y	y	n	y	y	N	y	Y	Y	n	N	Y	N	n	n	n	ks2
y	y	y	y	y	Y	y	y	y	y	Y	y	Y	Y	Y	n	Y	Y	Y	n	n	ks2
y	y	y	y	y	Y	y	y	y	y	Y	y	Y	Y	Y	n	Y	y	N	n	y	ks2
y	y	y	y	y	Y	y	y	y	y	Y	y	Y	Y	Y	n	Y	Y	Y	N	y	ks2
y	n	y	n	n	Y	y	y	n	y	N	n	y	N	n	Y	n	N	n	n	n	kps
y	n	y	n	n	Y	y	n	y	Y	n	n	n	N	n	N	n	N	n	n	n	kps
y	n	y	y	y	Y	y	y	y	Y	y	y	Y	n	N	n	Y	y	Y	n	n	kps

Gambar 2. Data uji

Setelah itu hitung entropi hasil, entropi atribut serta gain di tiap atributnya. setiap Gain tertinggi akan menghasilkan *root* dan di teruskan samapai ketemu di setian daunnya. Setelah data dihitung keseluruhan maka menghasilkan pohon.



Gambar 3. Hasil tree

Pada iterasi terakhir menghasilkan pohon seperti pada gambar 3. dengan sampel

tunggal pada nilai 'yes' maupun 'no' sehingga iterasi sudah selesai karena tidak ada atribut lagi yang tersisa. Dari pohon tersebut maka aturan klasifikasi yang terbentuk adalah :

1. Jika '3' = no dan '16' yes maka menghasilkan ks3.
2. Jika '3'= no dan '16' no maka menghasilkan KPS
3. Jika '3' = yes dan '20' yes, maka menghasilkan KS1
4. Jika '3' = yes '20' no, '14' yes maka menghasilkan KS2.
5. jika '3' = yes, '20' no, '14' no dan '13' yes maka menghasilkan KS1
6. Jika '3' = yes, '20' no, '14' no dan '13' yes maka menghasilkan KS2.

Dari data tersebut ddidapatkan tingkat akurasi

accuracy: 95.00%					
	true ks3	true ks2	true ks1	true kps	class precision
pred ks3	2	1	0	0	66.67%
pred ks2	0	8	0	0	100.00%
pred ks1	0	0	3	0	100.00%
pred kps	0	0	0	6	100.00%
class recall	100.00%	88.89%	100.00%	100.00%	

Gambar 4. Akurasi

Cara menghitung akurasi dengan cara :

$$\begin{aligned}
 &= \frac{Tkps + Tks1 + Tks2 + Tks3}{Tkps + Tks1 + Tks2 + Tks3 + Fkps + Fks1 + Fks2 + Fks3} \times 100\% \\
 &= \frac{2 + 8 + 3 + 6}{2 + 8 + 3 + 6 + 1 + 0 + 0 + 0} \times 100\% \\
 &= \frac{19}{20} \times 100\% \\
 &= 95\%
 \end{aligned}$$

akurasi yang didapatkan yaitu sebesar 95%. Sehingga hasil klasifikasi kedalam indikator tingkatan kesejahteraan keluarga sebagai berikut :

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	Hasil	Ymining			
1	T	T	a	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	a	bs3	bs3	
2	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	a	bs2	bs2	
3	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	T	a	bs2	bs2
4	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	T	T	bs1	bs1
5	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	T	T	bs2	bs2
6	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	T	T	bs1	bs1
7	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	T	T	bs2	bs2
8	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	T	T	bs1	bs1
9	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	T	T	bs2	bs2
10	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	T	T	bs2	bs3
11	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	T	T	bs3	bs3
12	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	T	T	bs2	bs2
13	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	T	T	bs2	bs2
14	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	T	T	bs2	bs2
15	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	T	T	bs3	bs3
16	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	T	T	bs3	bs3
17	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	T	T	bs3	bs3
18	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	a	T	T	bs2	bs2

Gambar 5. Hasil klasifikasi kedalam indikator

4. KESIMPULAN DAN SARAN

Kesimpulan : Berdasarkan hasil penelitian pada klasifikasi penduduk kedalam tingkatan keluarga sejahtera diambil kesimpulan bahwa penelitian ini dapat meringankan tugas PLKB dalam mengelompokkan atau klasifikasi data kedalam jenis tingkatan kesejahteraan keluarga menggunakan teknologi *data mining* dengan pohon keputusan berbasis algoritma C4.5.

Saran : Saran yang diajukan agar penelitian ini terus berkembang, maka perlu terdapat saran sebagai pengembang yang diusulkan :

1. Penelitian ini dapat dikembangkan dengan menggunakan metode data mining klasifikasi lain untuk dilakukan perbandingan.
2. Untuk penelitian selanjutnya dapat melengkapi data dengan jumlah yang lebih banyak dan daerah yang lebih luas.

DAFTAR PUSTAKA

[1] (2015, Juni) Portal Informasi Pemerintah Kabupaten Semarang. [Online].
<http://www.semarangkab.go.id/skpd/kbpd/profil/tugas-pokok-dan-fungsi/bidang-ks.html>

[2] (2015, Juli) [Online].
<http://aplikasi.bkkbn.go.id/mdk/BatasanMDK.aspx>

[3] Dewi Kartika Pane, "Implementasi Data Mining Pada Penjualan Produk Elektronik Dengan Algoritma Apriori," *Pelita Informatika Budi Darma*, vol. 4, no. 3, p. 26, Agustus 2013.

[4] Sunjana, "Aplikasi Mining Data Mahasiswa Dengan Metode Klasifikasi Decision Tree," *SNTI*, Juni 2010.

[5] Bundanis Dwi Meilani Achmad and Fauzi Slamet, "Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree," *IPTEK*, vol. 16, no. 1, pp. 17-23, Mei 2012.

[6] Indri Rahmayuni, "Perbandingan Performansi Algoritma C4.5 dan CART Dalam Klasifikasi Data Nilai Mahasiswa Prodi Teknik Komputer Universitas Negeri Padang," *TEKNOIF*, vol. 2, no. 1, pp. 40-46, April 2014.

[7] BKKBN, *Panduan Pembangunan Keluarga Sejahtera Dalam Rangka Penanggulangan Kemiskinan Kantor Menteri Negara Kependudukan/BKKBN*. Jakarta, Indonesia, 1996.

[8] B. Soembodo, ""Kebudayaan dan Politik. Th XIXI No. 4", Aspirasi Sosial Budaya Masyarakat Pedesaan terhadap Kesejahteraan Keluarga," pp. 75-33, 2006.

[9] Ian H Witten, Eibe Frank, and Mark A Hall, *Data Mining : Practical Machine Learning Tools and Techniques Third Edition*. USA:

- Elsevier, 2011.
- [10] Turban, Efraim, R. Kelly Rainer, Jr., and Richard E. Potter, *Introduction to Information Technology, 3th Edition*. New Jersey: John Willey & Sons, Inc, 2005.
- [11] Daniel T Larose, *Discovering Knowledge in Data : An Introduction to Data Mining*. New Jersey: John Wiley & Sons, 2005.
- [12] Ponniah P., *Datawarehouse Fundamentals : A Comprehensive Guide for IT Profesional.*: John Willey & Sons, Inc., 2011.
- [13] Emha Taufiq Luthfi and Kusriani, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009.
- [14] Ethem Alpaydin, "Introduction to Machine Learning," *The MIT Press*, 2004.
- [15] Arif Junanto, "Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa," *Teknologi Informasi Dinamik*, vol. 18, no. 1, pp. 9-16, Januari 2013.
- [16] T. H. Rahmadya and P. P. Herlawati, *Penerapan Data Mining dengan Matlab*. Bandung, Indonesia: Rekayasa Sains, 2013.
- [17] Max Bramer, *Principles of Data Mining*. London: Springer, 2007.
- [18] Achir Y. C. A., "*Prisma*", dalam *dalam Pembangunan Keluarga Sejahtera Sebagai Wahana Pembangunan Bangsa*. Jakarta, Indonesia: LP3ES, 1994.
- [19] Michael J. A Berry and Gordon S Linoff, *Data Mining Techniques For Marketing Sales, Customer Relationship Management.*, 2nd ed. USA: Wiley Publishing, 2004.
- [20] Soetjipto, *Pendidikan Kesejahteraan Keluarga*. Semarang, Indonesia: Satwa Wacana Press, 1992.
- [21] Indah Kuntum Khairina, "Penggunaan Pohon Keputusan Untuk Data Mining".
- [22] Ratih Ariadni and Isye Ariesianti, "Implementasi Metode Pohon Keputusan Untuk Klasifikasi Data Dengan Nilai Fitur yang Tidak Pasti".
- [23] Fatayat and Joko Risanto, "Proses Data Mining dalam Meningkatkan Sistem Pembelajaran pada Pendidikan Sekolah Menengah Pertama," 2013.
- [24] Anik Andriani, "Penerapan algoritma C4.5 Pada program klasifikasi mahasiswa dropout," 2012.
- [25] (2014, October, 13) <http://disdik.patikab.go.id>.
- [26] W., Wu, X., Sun, Y. and Zhang, Q. Li, "Credit Card Customer Segmentation and Target Marketing Based on Data Mining," , 2010, pp. 73-76.
- [27] Sug Hyonta, "A Structural Sampling Technique for Better Decision Trees," 2009.
- [28] Bain.K, Holisatul Munawaroh, and Yeni Kustiyahningsih, "Perbandingan algoritma ID3 dan C5.0 dalam identifikasi penjurusan siswa SMA," Juni 2013.

- [29] Adi Suwondo, Dian Asmarajati, and Heri Surahman, "Algoritma C4.5 Berbasis Adaboost untuk Prediksi Penyakit Jantung Koroner," Juni 2013.
- [30] Zezen Zaenal Abidain Aa, "Implementasi Algoritma C4.5 untuk Menentukan Tingkat Bahaya Tsunami," Juli 2011.
- [31] Swastina Liliana, "Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa," *Gema Aktualita*, Juni 2013.
- [32] Kausar Muchammad, Rika Yunitarini, and Windy Julianto, "Algoritma C4.5 untuk Penilaian Kinerja Karyawan," vol. IX, Juni 2014.