



## IMPLEMENTASI ALGORITMA C4.5 DALAM MENENTUKAN LOKASI PRIORITAS PENYULUHAN PROGRAM KELUARGA BERENCANA DI KECAMATAN DUMAI TIMUR

### *IMPLEMENTATION OF ALGORITHM C4.5 TO DETERMINING LOCATION PRIORITY COUNSELING FAMILY PLANNING PROGRAM IN EAST DUMAI*

Febrina Sari<sup>1</sup>, David Saro<sup>2</sup>

<sup>1,2</sup>Program Studi Informatika Sekolah Tinggi Teknologi  
Jl. Utama Karya Bukit Batrem II, Dumai, Riau, Indonesia  
[febri\\_ghaniya@yahoo.co.id](mailto:febri_ghaniya@yahoo.co.id)

Naskah Diterima: 10 Desember 2017; Direvisi : 15 Maret 2018; Disetujui : 15 Maret 2018

#### Abstrak

Indonesia merupakan salah satu negara dengan penduduk terbanyak di dunia. Ledakan penduduk ini terjadi karena laju pertumbuhan penduduk yang sangat tinggi. Kondisi ini menyebabkan beban negara menjadi semakin besar. Karena berhubungan dengan tinggi rendahnya beban negara untuk memberikan penghidupan yang layak kepada setiap warga negaranya, maka pemerintah memberikan serangkaian usaha untuk menekan laju pertumbuhan penduduk agar tidak terjadi ledakan penduduk yang lebih besar. Salah satu cara yang dilakukan oleh pemerintah adalah dengan menggalakkan Program Keluarga Berencana (KB). Program keluarga berencana yang dicanangkan oleh pemerintah untuk menekan angka kelahiran yang tinggi ini belum sepenuhnya terlaksana dengan baik karena lokasi penyuluhan program KB yang belum tepat sasaran. Oleh karena itu diperlukan suatu sistem yang dapat membantu Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) dalam menentukan lokasi Prioritas Penyuluhan Program Keluarga Berencana agar penyuluhan tepat sasaran. Data Mining adalah kegiatan mengekstraksi atau menambang pengetahuan dari data yang berukuran besar, algoritma C4.5 merupakan algoritma yang digunakan untuk membangun sebuah pohon keputusan (Decision Tree) dari data. Algoritma C4.5 umumnya digunakan untuk melakukan klasifikasi yang nantinya akan digunakan dalam menetapkan lokasi prioritas penyuluhan program Keluarga Berencana.

**Kata kunci:** Data Mining, Algoritma C4.5, Keluarga Berencana

#### Abstract

Indonesia is one of the most populous countries in the world. This population explosion occurs because the rate of population growth is very high. This condition causes the burden of the state becomes larger. Due to the high burden of the state's burden to provide a decent living for every citizen, the government provides a series of efforts to curb the rate of population growth in order to avoid a larger population explosion. One way that the government is doing is by promoting Family Planning Program (KB). The family planning program launched by the government to reduce the high birth rate has not been well implemented because the location of the extension of the family planning program has not been well targeted. Therefore needed a system that can help National Population and Family Planning Agency (BKKBN) in determining location Priority Counseling Family Planning Program so that elucidation on target. Data Mining is the activity of extracting or mining knowledge of large data, C4.5 algorithm is an algorithm used to build a decision tree from the data. The C4.5 algorithm is generally used to classify what will be used in setting priority location of KB program extension.

**Keywords:** Data Mining, Algorithm C4.5, Planning Family

## PENDAHULUAN

Salah satu usaha pemerintah dalam menghadapi masalah kependudukan adalah program keluarga berencana. Program Keluarga Berencana (KB) Nasional Merupakan salah satu program dalam rangka menekan laju pertumbuhan penduduk. Program KB dirumuskan sebagai upaya peningkatan kepedulian dan peran serta masyarakat melalui batas usia perkawinan, pengaturan kelahiran, pembinaan ketahanan keluarga, peningkatan kesejahteraan keluarga, untuk mewujudkan Norma Keluarga Kecil Bahagia dan Sejahtera (NKKBS). Membudayakan norma keluarga kecil bahagia sejahtera dalam rangka meningkatkan mutu sumber daya manusia Indonesia. Cara yang digunakan adalah dengan mengendalikan kelahiran sekaligus menjamin terkendalinya pertumbuhan penduduk.

Program keluarga berencana yang dicanangkan oleh pemerintah untuk membudayakan norma keluarga kecil bahagia dan sejahtera serta menekan angka kelahiran yang tinggi ini belum sepenuhnya terlaksana dengan baik karena penyuluhan program keluarga berencana diadakan di beberapa daerah yang tingkat kelahirannya rendah, jika penetapan lokasi penyuluhan yang selalu tidak tepat sasaran dikawatirkan program ini tidak dapat mencapai tujuan. Oleh karena itu penetapan lokasi prioritas penyuluhan program keluarga berencana yang tepat sangatlah penting untuk dapat menekan angka kelahiran, sehingga diperlukan suatu Algoritma sistem yang dapat membantu Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) dalam menentukan lokasi prioritas Penyuluhan Program Keluarga Berencana agar program tepat sasaran.

Dari pemaparan diatas, rumusan masalah

penelitian ini adalah bagaimana Algoritma C4.5 dapat menentukan lokasi prioritas penyuluhan program keluarga berencana, Karena algoritma C4.5 digunakan untuk melakukan klasifikasi, jadi hasil dari pengolahan *test dataset* berupa pengelompokan data dalam kelas-kelasnya, yang mana kelas dibagi menjadi dua yakni tidak prioritas atau ya prioritas.

Tujuan Penelitian ini adalah menganalisis hasil dari implementasi Algoritma C4.5, dalam melakukan klasifikasi data kelahiran sehingga dapat menentukan lokasi mana yang menjadi prioritas untuk dilakukan penyuluhan program keluarga berencana.

Ada beberapa penelitian terdahulu yang pernah dibuat terkait dengan penggunaan Algoritma C4.5, diantaranya:

1. Penelitian yang dilakukan oleh Liliana Swastina (2013). Tentang Penerapan Algoritma C4.5 untuk penentuan Jurusan Mahasiswa Dengan adanya penerapan Decision Tree C4.5 dapat memberikan solusi bagi mahasiswa dan membantu STMIK Indonesia dalam menentukan jurusan yang sesuai yang akan ditempuh oleh mahasiswa selama studi sehingga peluang untuk sukses dalam studi di perguruan tinggi semakin besar.
2. Penelitian yang dilakukan oleh Mujib Ridwan (2013). pada penelitian ini peneliti menggunakan algoritma C4.5 dalam menentukan prediksi kelulusan berdasarkan atribut jenis kelamin, asal sekolah SMA dan IPK semester satu sampai semester enam.
3. Berikutnya penelitian yang dilakukan oleh Kumara dan Supriyanto (2013), dengan

judul Klasifikasi Data Mining untuk Penerimaan Seleksi Calon Pegawai Negeri Sipil 2014 Menggunakan Algoritma *Decision Tree* C4.5. Tingkat akurasi yang didapatkan menggunakan algoritma C4.5 ini sudah cukup tinggi, oleh karena itu dapat disimpulkan algoritma ini cocok untuk diimplementasikan pada penelitian yang melibatkan proses perekrutan.

4. Penelitian yang dilakukan oleh Hartato (2014). Tentang Penerapan Data mining dengan algoritma C4.5 ini dapat diimplementasikan untuk memprediksi tingkat kelulusan mahasiswa dengan empat kategori yaitu lulus cepat, lulus tepat, lulus terlambat dan *drop out*. Atribut yang paling berpengaruh dalam hasil prediksi adalah IPK semester enam.

Dari beberapa sumber *Literature review* maka peneliti dapat mengetahui bahwa penelitian tentang Implementasi Algoritma C4.5 untuk mengetahui lokasi mana yang menjadi prioritas dalam pelaksanaan penyuluhan program berencana belum pernah dilakukan dan penggunaan algoritma C4.5 dapat digunakan karena hasil akhir berupa *decision tree* yang menggambarkan pengelompokan data berdasarkan kelasnya

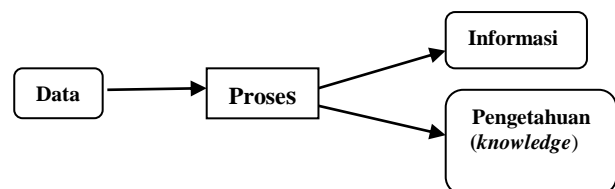
### **Data Mining**

*Data mining* adalah suatu proses menemukan hubungan yang berarti, pola dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. *Data mining* merupakan gabungan dari beberapa disiplin ilmu yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, *database*, dan visualisasi

untuk penanganan permasalahan pengambilan informasi dari *database* yang besar (Larose, 2005).

Secara singkat bahwa *Decision Tree* merupakan salah satu metode klasifikasi pada *text mining*. Klasifikasi adalah proses menemukan kumpulan pola atau fungsi-fungsi yang mendeskripsikan dan memisahkan kelas data satu dengan lainnya, untuk dapat digunakan dalam memprediksi data yang belum memiliki kelas data tertentu (Han, 2006).

Sementara itu (Kusnawi, 2007) menyatakan bahwa *Data mining* merupakan teknologi yang menggabungkan metoda analisis tradisional dengan algoritma yang canggih untuk memproses data dengan volume besar. *Data mining* bermula dari sebuah data yang kemudian diproses sehingga menghasilkan informasi atau menghasilkan pengetahuan (*knowledge*) dan merupakan salah satu tahapan (*step*) dari *Knowledge Discovery in Database* (KDD). Berikut adalah tahapan didalam *Data Mining* yang diilustrasikan pada gambar 1.



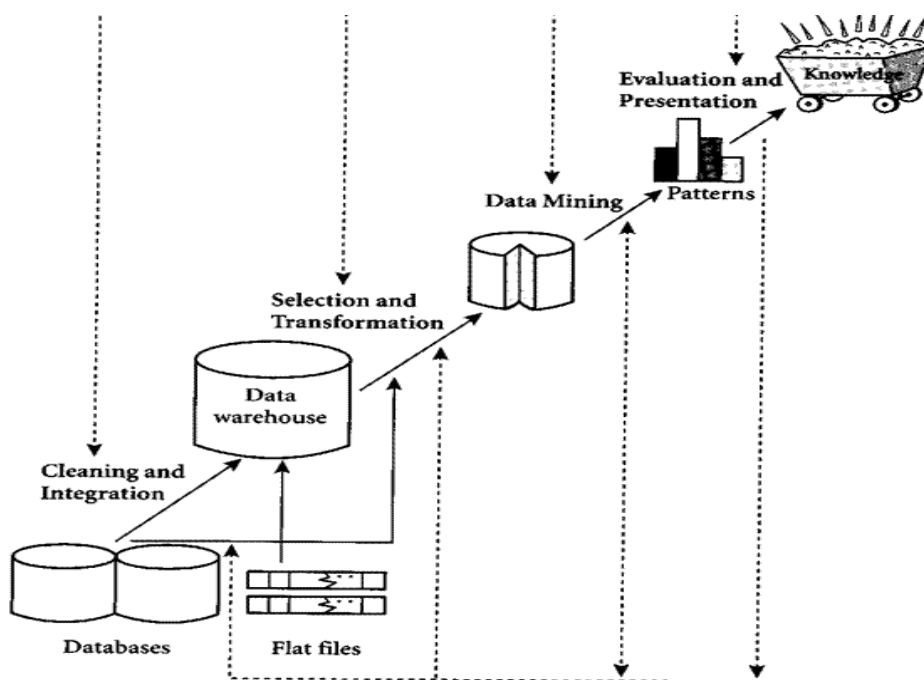
**Gambar 1.** Tahapan *Data Mining*  
Sumber : (Hermawati, 2009)

### **Knowledge Discovery in Database (KDD)**

Pengertian dari KDD adalah penemuan atau pencarian pengetahuan (nilai tambah) di dalam sebuah *database* (Hermawati, 2009), karena *data mining* adalah suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap yaitu :

1. Pembersihan data (untuk membuang data yang tidak konsisten dan *noise*).

2. Integrasi data (penggabungan data dari beberapa sumber).
3. Transformasi data (data diubah menjadi bentuk yang sesuai untuk *Data Mining*).
4. Aplikasi teknik *Data Mining*.
5. Evaluasi pola yang ditemukan (untuk menemukan yang menarik / bernilai).
6. Presentasi pengetahuan (dengan teknik visualisasi).
7. Tahapan dapat diilustrasikan pada Gambar 2.



**Gambar 2.** Tahapan *Knowledge Discovery in Database*  
 Sumber : (Hermawati, 2009)

### Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang sangat populer yang digunakan oleh banyak peneliti di dunia, hal ini dijelaskan oleh Xindong Wu dan Vipin Kumar dalam bukunya yang berjudul *The Top Ten Algorithms in Data Mining* (Wu, 2009), selain itu Algoritma C4.5 merupakan salah satu algoritma *Decision Tree* yang paling efektif untuk melakukan klasifikasi (Chauhan, 2013).

Algoritma C4.5 merupakan salah satu algoritma *machine learning*, dengan algoritma C4.5 mesin (komputer) akan diberikan sekelompok data untuk dipelajari yang disebut *learning dataset* (Hamdan, 2010). Kemudian hasil dari pembelajaran

selanjutnya akan digunakan untuk mengolah data-data yang baru yang disebut *test dataset*. Algoritma C4.5 merupakan algoritma yang digunakan untuk membangun sebuah pohon keputusan (*Decision Tree*) dari data. Algoritma C4.5 merupakan pengembangan dari algoritma *ID3* yang juga merupakan algoritma untuk membangun sebuah pohon keputusan. Algoritma C4.5 secara rekursif mengunjungi tiap simpul keputusan, memilih percabangan optimal, sampai tidak ada cabang lagi yang mungkin dihasilkan ( Rahmayuni, 2014).

Secara umum algoritma C4.5 dalam membangun pohon keputusan langkah-langkahnya adalah sebagai berikut.

- Pilih atribut sebagai akar
- Buat cabang untuk tiap-tiap nilai
- Bagi kasus dalam cabang
- Ulangi proses untuk setiap 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 digunakan rumus seperti tertera dalam persamaan (1) berikut.

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

Keterangan :

- S : himpunan kasus  
 A : atribut  
 n : jumlah partisi atribut A  
 $|S_i|$  : jumlah kasus pada partisi ke-i  
 $|S|$  : jumlah kasus dalam S

Sementara itu perhitungan nilai entropy dapat dilihat pada persamaan (2) berikut.

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

Keterangan :

- S : himpunan kasus  
 A : fitur  
 n : jumlah partisi S  
 $p_i$  : proporsi dari  $S_i$  terhadap S

Proses pengulangan pada metode decision tree ini akan berhenti apabila :

- Semua data telah terbagi rata
- Tidak ada lagi atribut yang bisa dibagi lagi
- Tidak ada data *record* dalam cabang yang kosong

## METODE

Metodologi penelitian merupakan urutan-urutan yang dilakukan dalam sebuah penelitian. Metodologi penelitian ini bertujuan agar penelitian bisa lebih terkonsep dan terstruktur, sehingga setiap tahapan akan dapat dilihat pencapaiannya sesuai dengan tujuan yang diharapkan terhadap penelitian tersebut.

### Desain Penelitian

Penelitian ini menggunakan Metode *Decision Tree*, dengan Algoritma C4.5. yang akan diterapkan ke dalam *software weka* dalam menentukan lokasi prioritas penyuluhan program keluarga berencana, agar penelitian memperoleh hasil yang maksimal, tentunya harus mengikuti kaidah-kaidah (Metode) yang telah ditetapkan.

### Lokasi dan Waktu Penelitian

#### A. Lokasi Penelitian

Penelitian ini dilakukan di BKKBN Kota Dumai. Adapun pertimbangan memilih lokasi ini adalah untuk memperoleh Atribut atau kriteria yang valid, karena sesuai dengan objek penelitian

#### B. Waktu Penelitian

Penelitian ini akan dilaksanakan dalam jangka waktu 10 bulan.

### Sumber dan Sampel Data Penelitian

#### A. Sumber Data Penelitian

Data penelitian ini bersumber dari *database* Kelahiran yang langsung diambil melalui Aplikasi Sistem Informasi Administrasi Kependudukan yang ada di Kantor Dinas Penduduk dan Catatan Sipil Kota Dumai.

#### B. Sampel Penelitian

Sampel dalam penelitian ini adalah data

kelahiran bayi pada setiap Kelurahan yang ada di Kecamatan Dumai Timur Tahun 2016, Karena kesuksesan program KB disuatu daerah di tentukan dari tinggi rendahnya angka kelahiran bayi di daerah tersebut.

### Teknik Pengumpulan dan Analisis Data

A. Teknik pengumpulan data yang akan dilakukan oleh peneliti adalah :

1. Observasi Langsung, peneliti mengamati secara langsung keluhan-keluhan serta masalah-masalah yang dihadapi pihak BKKBN Kota Dumai dalam menetapkan lokasi prioritas untuk diadakannya penyuluhan program KB.
2. Wawancara, peneliti melakukan Tanya jawab dengan pihak terkait yakni BKKBN sebagai pihak yang terkait secara langsung dalam penelitian guna memperoleh data tentang kriteria-kriteria yang diperlukan untuk menentukan lokasi penyuluhan program keluarga berencana dan pihak Kantor Dinas Kependudukan dan Pencatatan Sipil Kota Dumai, selaku pihak yang akan memberikan data penunjang yang diperlukan dalam penelitian.
3. Studi Perpustakaan, peneliti mencari data-data pendukung seperti buku, jurnal-jurnal maupun literatur lain dari akses internet.

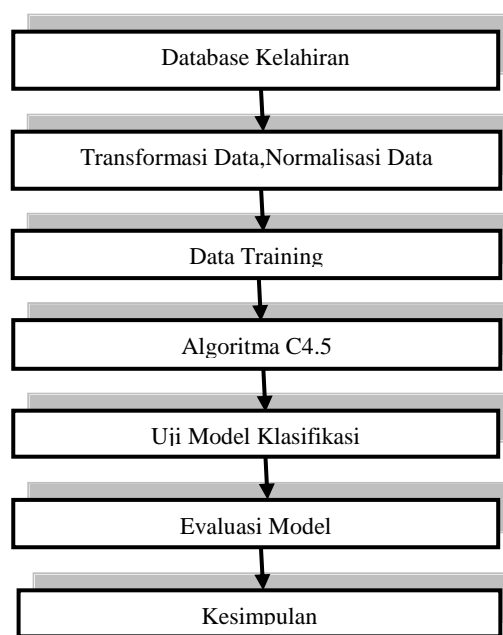
B. Analisis Data

Data Kuantitatif yang telah terkumpul kemudian dianalisa menggunakan

Algoritma C4.5 untuk mengklasifikasikan data kedalam kelas-kelasnya kemudian digunakan untuk membangun sebuah pohon keputusan

### Tahapan Penelitian

Tahapan Penelitian ini merupakan langkah-langkah yang akan dilakukan dalam penyelesaian masalah yang akan dibahas, sedangkan metode yang digunakan dalam penelitian ini bertujuan untuk memperlihatkan bagaimana sebuah model klasifikasi data mining bisa memberikan solusi untuk menentukan lokasi prioritas penyuluhan program keluarga berencana berdasarkan atribut yang ada. Adapun Tahapan Penelitian ini dapat dilihat pada Gambar 3.



**Gambar 3.** Tahapan Penelitian

Berdasarkan Tahapan Penelitian pada gambar 3, maka masing-masing langkahnya dapat diuraikan seperti berikut ini:

1. Transformasi Data

Data yang diperoleh dari database kelahiran yang berasal dari Aplikasi Sistem Informasi

Administrasi Kependudukan Kota Dumai, masih berupa data yang mengandung banyak atribut yang tidak diperlukan sehingga perlu dilakukan transformasi data dengan membuang sebagian atribut yang tidak memiliki kaitan dengan topik penelitian

## 2. Normalisasi Data

Proses normalisasi data yang dimaksud adalah mengubah jenis skala pengukuran yang semula berbentuk numerikal menjadi nominal

## 3. Cleaning Data

Proses Pembersihan data yang tidak relevan termasuk data *missing* dalam atribut

## 4. Training Data

Proses pelatihan data diambil dari sebagian data yang terdapat pada database kelahiran. Besarnya proporsi data yang dilakukan pengujian adalah 70% untuk training, sedangkan sisanya digunakan untuk uji coba model.

## 5. Uji Model Klasifikasi

Proses uji model dilakukan setelah proses training data selesai dilakukan, jumlah data yang dilakukan uji model sebesar 30% dari database kelahiran bayi.

## 6. Evaluasi Model

Evaluasi model dilakukan dengan melihat tingkat akurasi metode melalui *confusion*

matrix dan tabel akurasi serta presisi untuk model yang digunakan.

## Machine Learning WEKA

WEKA adalah sebuah paket *tools machine learning* praktis. WEKA merupakan singkatan dari "Waikato Environment for Knowledge Analysis" yang dibuat di Universitas Waikato New Zealand untuk penelitian, pendidikan dan berbagai aplikasi. WEKA mampu menyelesaikan masalah-masalah data mining di dunia nyata, khususnya klasifikasi yang mendasari pendekatan *machine learning*. Perangkat lunak ini ditulis dalam hirarki *class java* dengan metode berorientasi objek dan dapat berjalan hampir di semua *platform* (Bouckaert, 2008).

## HASIL DAN PEMBAHASAN

Pada bagian ini, akan dijelaskan secara detail proses perhitungan *decision tree* menggunakan algoritma C4.5, untuk menentukan lokasi prioritas penyuluhan program KB. Adapun kriteria yang digunakan meliputi Usia Ibu, Pekerjaan Ayah dan Kelurahan. Masing-masing kriteria memiliki atribut. Salah satu atribut merupakan data solusi per item data yang disebut target atribut, yang menjadi target atribut adalah Keputusan dengan nilai "Ya" atau "Tidak

**Tabel 1.** Data Primer dari Basis Data

No	Atribut	Jumlah Kasus	Ya	Tidak
1	Total	934	849	85
	Usia			
2	Produktif	793	793	0
	Tidak Produktif	141	56	85
	Pekerjaan			
	Belum Bekerja	53	53	0
	Buruh Harian	229	229	0

No	Atribut	Jumlah Kasus	Ya	Tidak	
3	Kelurahan	Wiraswasta	413	359	54
		Karyawan	239	208	31
		Teluk Binjai	242	224	18
		Tanjung Palas	147	137	10
		Jaya Mukti	293	259	34
		Buluh Kasap	104	90	14
		Bukit Batrem	148	146	2

Tahapan berikutnya adalah melakukan proses perhitungan nilai *Entropy* dan nilai *Gain* untuk setiap atribut dengan menggunakan rumus sebagai berikut:

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

$$Entropy(Total) = - (849/934) * \log_2(849/934) - (85/934) * \log_2(85/934)$$

$$= - (0,908 * \log_2 0,908) - (0,090 * \log_2 0,090)$$

$$= - (0,908 * -0,139) - (0,090 * -3,473)$$

$$= 0,126 + 0,312$$

$$= 0,438$$

$$Entropy(Produktif) = 0$$

$$Entropy(Tidak Produktif) = - (56/141) * \log_2(56/141) - (85/141) * \log_2(85/141)$$

$$= - (0,397 * \log_2 0,397) - (0,602 * \log_2 0,602)$$

$$= - (0,397 * -1,332) - (0,602 * -0,732)$$

$$= 0,529 + 0,440$$

$$= 0,969$$

Selanjutnya menghitung nilai *Gain* Masing-masing Atribut dengan rumus sebagai berikut:

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

$$Gain(Usia)$$

$$= 0,438 - (793/934 * 0 + 141/934 * 0,969)$$

$$= 0,438 - (0 + 0,146)$$

$$= 0,438 - 0,146$$

$$= 0,292$$

$$Gain(Pekerjaan)$$

$$= 0,438 - (0 + 0 + 413/934 * 0,558 + 239/934 * 0,555)$$

$$= 0,438 - (0 + 0 + 0,246 + 0,142)$$

$$= 0,438 - 0,388$$

$$= 0,050$$

$$Gain(Kelurahan)$$

$$= 0,438 - (242/934 * 0,382 + 147/934 * 0,359 + 293/934 * 0,519 + 104/934 * 0,569 + 148/934 * 0,101)$$

$$= 0,438 - (0,098 + 0,056 + 0,162 + 0,063 + 0,016)$$

$$= 0,438 - 0,395$$

$$= 0,043$$

Tabel 2. Hasil Perhitungan Nilai *Entropy* dan *Gain* Node 1

Node I	Jumlah Kasus	Ya	Tidak	Entropy	Gain
1 Total	934	849	85	0,438	
Usia					0,292
Produktif	793	793	0	0	

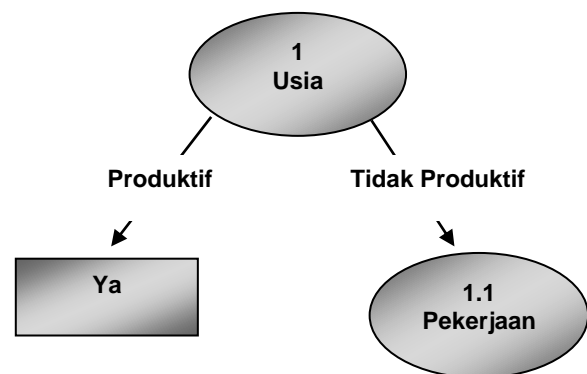


<i>Node 1</i>		<b>Jumlah Kasus</b>	<b>Ya</b>	<b>Tidak</b>	<i>Entropy</i>	<i>Gain</i>
<b>Pekerjaan</b>	Tidak Produktif	141	56	85	0,969	<b>0,050</b>
	Belum Bekerja	53	53	0	0	
	Buruh Harian	229	229	0	0	
	Wiraswasta	413	359	54	0,558	
	Karyawan	239	208	31	0,555	
<b>Kelurahan</b>						<b>0,043</b>
	Teluk Binjai	242	224	18	0,382	
	Tanjung Palas	147	137	10	0,359	
	Jaya Mukti	293	259	34	0,519	
	Buluh Kasap	104	90	14	0,569	
	Bukit Batrem	148	146	2	0,101	

Seperti yang terlihat pada Tabel 2, diperoleh atribut dengan nilai *gain* tertinggi adalah Usia yakni dengan nilai 0,292 maka Usia menjadi node akar. Usia memiliki dua nilai yaitu Produktif dan Tidak Produktif. Usia dengan nilai produktif sudah mengklasifikasikan kasus menjadi 1 yaitu keputusan "Ya" sedangkan untuk atribut Usia tidak produktif masih perlu dilakukan perhitungan lagi karena masih terdapat keputusan Ya dan Tidak.

Dari hasil perhitungan tersebut maka dapat digambarkan pohon keputusan *node 1* sebagai berikut.

Selanjutnya, kembali melakukan langkah-langkah penyelesaian dan proses perhitungan *Entropy* dan *Gain* untuk *node 1.1*. Tabel 3.



**Gambar 4.** Pohon Keputusan *Node.1*

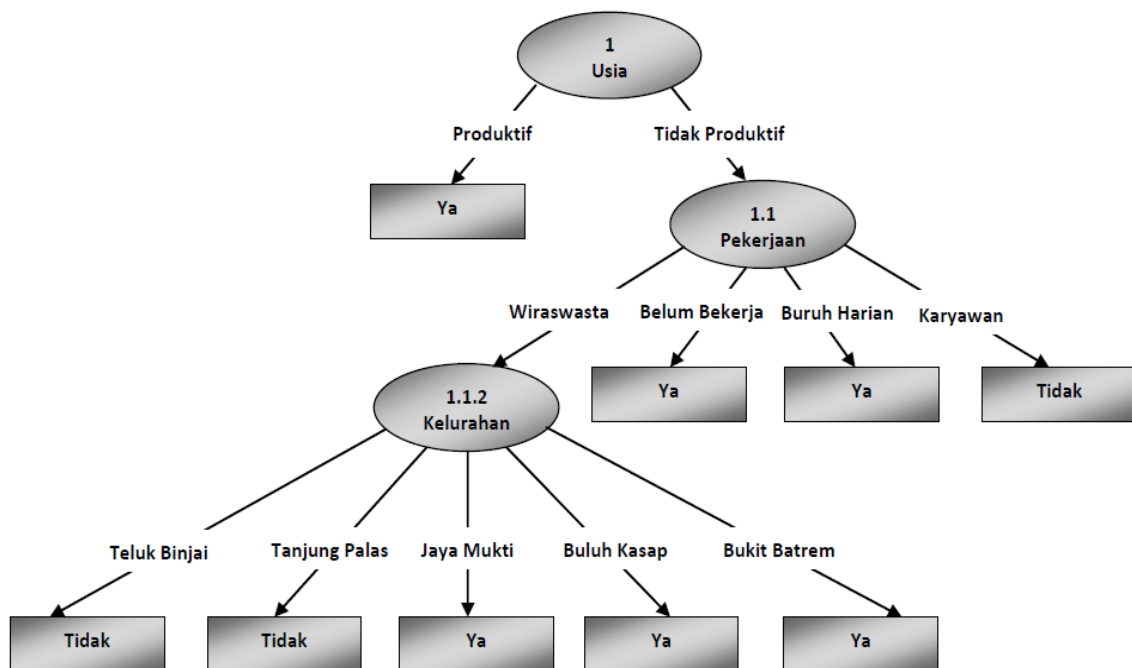
merupakan rekapitulasi hasil perhitungan nilai *Entropy* dan *Gain Node 1.1*

**Tabel 3.** Hasil Perhitungan Nilai *Entropy* dan *Gain Node 1.1*

<b>Node</b>		<b>Jumlah Kasus</b>	<b>Prioritas</b>	<b>Tidak Prioritas</b>	<i>Entropy</i>	<i>Gain</i>
<b>1.1</b>	<b>Usia</b>	Tidak Produktif	141	56	85	0,969
	<b>Pekerjaan</b>					0,509
		Tidak Bekerja	5	5	0	0

Node		Jumlah Kasus	Prioritas	Tidak Prioritas	Entropy	Gain
<b>Kelurahan</b>	Buruh Harian	52	52	0	0	0,411
	Wiraswasta	54	42	12	0,764	
	Karyawan	30	23	7	0,783	
	Teluk Binjai	36	33	3	0,413	
	Tanjung Palas	27	23	4	0,605	
	Jaya Mukti	45	37	8	0,675	
	Buluh Kasap	19	16	3	0,629	
	Bukit Batrem	14	13	1	0,371	

Selanjutnya, kembali melakukan langkah-langkah penyelesaian dan proses perhitungan Entropy dan Gain untuk setiap cabang sampai semua kasus pada cabang memiliki kelas-kelas yang sama. Gambar 5. merupakan pohon keputusan terakhir yang diperoleh dari hasil perhitungan.



Gambar 5. Pohon Keputusan Akhie

Dengan memperhatikan pohon keputusan pada Gambar 5 diketahui bahwa semua kasus sudah masuk dalam kelasnya masing-masing, dari pohon yang terbentuk dihasilkan sejumlah aturan. Adapun aturan yang terbentuk adalah berikut:  
 ”Jika Usia = Produktif Maka Class = Ya

”Jika Usia = Tidak Produktif, dan Pekerjaan = Belum Bekerja Maka *Class* = Ya

”Jika Usia = Tidak Produktif, dan Pekerjaan = Buruh Harian Maka *Class* = Ya

”Jika Usia = Tidak Produktif, dan Pekerjaan = Karyawan Maka *Class* = Tidak

”Jika Usia = Tidak Produktif, dan Pekerjaan = Wiraswasta, dan Kelurahan = Bukit Batrem Maka *Class* = Ya

”Jika Usia = Tidak Produktif, dan Pekerjaan = Wiraswasta, dan Kelurahan = Buluh Kasap Maka *Class* = Ya

”Jika Usia = Tidak Produktif, dan Pekerjaan = Wiraswasta, dan Kelurahan = Jaya Mukti Maka *Class* = Ya

”Jika Usia = Tidak Produktif, dan Pekerjaan = Wiraswasta, dan Kelurahan = Tanjung Palas Maka *Class* = Tidak

”Jika Usia = Tidak Produktif, dan Pekerjaan = Wiraswasta, dan Kelurahan = Teluk Binjai Maka *Class* = Tidak

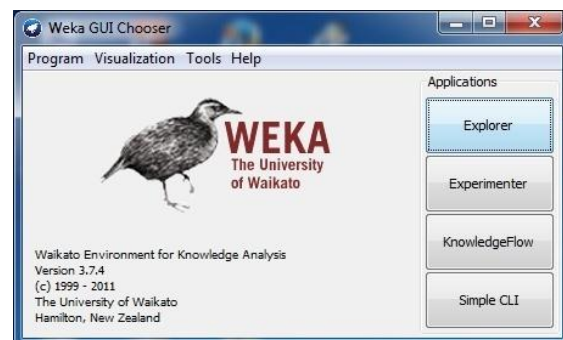
Hasil klasifikasi pada data sampel atribut Usia sebagai *root* pada *decision tree*, sedangkan atribut lainnya sebagai *child node*, dari data sampel dengan 934 *record* dihasilkan jumlah aturan yang terbentuk sebanyak 9 aturan.

### Implementasi *machine learning* WEKA

mengimplementasikan algoritma C4.5 dengan bantuan *tools machine learning* WEKA, Sebelum diproses data dibagi menjadi dua bagian yakni pertama adalah data *training* sebesar 70% dan kedua adalah data *testing* sebesar 30%. Hal ini dilakukan agar terbentuk suatu model dengan menggunakan data *training*, kemudian data yang terbentuk dengan menggunakan data *training* akan

diujikan kembali menggunakan data *testing*. Berikut merupakan langkah-langkah penyelesaian menggunakan *tools WEKA Version 3.7.4*.

1. Sebelum kita mengimplementasikan sistem terhadap data yang ingin diolah, kita harus mempersiapkan terlebih dahulu data yang akan diuji. Data tersebut disimpan dalam bentuk *file* berekstensi *.csv* pada *microsoft excel*.
2. Kemudian jalankan *tools WEKA*. Gambar 6 Berikut ini merupakan tampilan utama aplikasi WEKA.



Gambar 6. WEKA Version 3.7.4.

3. Selanjutnya klik *Explorer*, pilih data yang akan di proses, yang telah tersimpan dengan format *csv*, dengan cara mengklik *Open file*.
4. File yang akan diproses sudah berhasil di import, lalu klik *Classify*, kemudian klik *Choose* untuk memilih metode klasifikasi yang akan diproses, pada penelitian ini peneliti menggunakan metode *Trees* maka pilih J48.
5. Langkah berikutnya klik tombol start, maka akan muncul tampilan seperti pada gambar 7.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      929           99.4647 %
Incorrectly Classified Instances     5            0.5353 %
Kappa statistic                     0.8753
Mean absolute error                  0.0059
Root mean squared error              0.074
Relative absolute error              14.2961 %
Root relative squared error          52.3887 %
Coverage of cases (0.95 level)      99.4647 %
Mean rel. region size (0.95 level)  50.1071 %
Total Number of Instances           934

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
                0.996   0.053   0.999     0.996   0.997     0.971   PRIORITAS
                0.947   0.004   0.818     0.947   0.878     0.971   TIDAK PRIORITAS
Weighted Avg.   0.995   0.052   0.995     0.995   0.995     0.971

=== Confusion Matrix ===

  a  b  <-- classified as
911  4  |  a = PRIORITAS
  1 18 |  b = TIDAK PRIORITAS
    
```

**Gambar 7.** Klasifikasi Data Training

Tabel 4. merupakan Hasil klasifikasi yang akan dihadirkan dalam bentuk *Confusion Matrix*. *Predict Class* dan *Actual Class*. Model *Confusion Matrix* 2x2.

**Tabel 4.** Model *Confusion Matrix*

		Predict Class	
		Class A	Class B
Actual Class	Class A	AA	AB
	Class B	BA	BB

Gambar 8. merupakan hasil tingkat akurasi secara detail yang dihasilkan dari proses Aplikasi WEKA.

```

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
                0.996   0.053   0.999     0.996   0.997     0.971   PRIORITAS
                0.947   0.004   0.818     0.947   0.878     0.971   TIDAK PRIORITAS
Weighted Avg.   0.995   0.052   0.995     0.995   0.995     0.971
    
```

**Gambar 8.** Akurasi Detail *Class Training*

Tingkat positif benar / *True Positive (TP) Rate* adalah proporsi contoh yang diklasifikasikan sebagai kelas x, diantara semua contoh yang benar-benar memiliki kelas x yang bernilai sama dengan *recall*.

Untuk mencari nilai *TP rate* pada *class* Ya adalah

$$True\ Positive = \frac{911}{911 + 4} = 0,996$$

Untuk mencari nilai TN pada *class* Tidak adalah

$$True\ Negative = \frac{18}{18 + 1} = 0,947$$

Tingkat positif salah / *False Positive (FP) Rate* adalah proporsi contoh yang diklasifikasikan sebagai kelas x, tetapi masuk kedalam kelas yang berbeda diantara semua contoh yang bukan dari kelas x.

Untuk mencari nilai FP rate pada class Ya adalah

$$False\ Positive = \frac{1}{1 + 18} = 0,053$$

Untuk mencari nilai FN pada class Tidak adalah

$$False\ Negative = \frac{4}{911 + 4} = 0,004$$

Perhitungan nilai Precision Class Ya dan Tidak adalah sebagai berikut :

$$Precision\ Class\ Ya = \frac{911}{911 + 1} = 0,999$$

$$Precision\ Class\ Tidak = \frac{18}{18 + 4} = 0,818$$

Hasil pengujian ditunjukkan pada Gambar 9. Confusion Matrix Algoritma Klasifikasi C4.5 yang diperoleh merupakan evaluasi dari kinerja model klasifikasi, dan bukti terjadinya hasil proses pada klasifikasi tersebut yang telah tersedia didalamnya.

```

=== Confusion Matrix ===
  a  b  <-- classified as
911  4  |  a = PRIORITAS
  1 18  |  b = TIDAK PRIORITAS
    
```

**Gambar 9.** Confusion Matrix Algoritma C4.5

Selain akurasi dan Confusion Matrix, sebuah model klasifikasi bisa dilihat dari nilai recall dan presisinya. Presisi merupakan probabilitas bahwa sebuah item yang terpilih adalah relevan. Sedangkan recall adalah rasio dari item yang relevan yang dipilih terhadap total jumlah item yang relevan.

Hasil presisi dan recall yang diperoleh dari model klasifikasi diatas adalah 0,995 untuk presisi dan 0,995 untuk recall. Hasil recall dan presisi memiliki nilai antara 0-1. Semakin tinggi nilainya, maka semakin baik

Berdasarkan informasi diatas, kemudian akan

dilakukan proses perhitungan nilai rata-rata persentasi akurasi keberhasilan dengan menggunakan rumus (3) dan error rate pada confusion matrix data training dengan menggunakan rumus (4).

$$Akurasi = \frac{\text{Banyaknya prediksi yang benar}}{\text{Total banyaknya prediksi}} \dots\dots(3)$$

$$Akurasi = \frac{911 + 18}{911 + 4 + 1 + 18} = \frac{929}{934} = 0,995$$

Maka Nilai Persentase Akurasi adalah

$$= 0,995 \times 100\% = 99,5\%$$

$$Error\ Rate = \frac{\text{Banyaknya prediksi yang salah}}{\text{Total banyaknya prediksi}} \dots\dots(4)$$

$$Error\ Rate = \frac{4 + 1}{911 + 4 + 1 + 18} = \frac{5}{934} = 0,005$$

Maka Nilai Persentase Error Rate adalah

$$= 0,005 \times 100\% = 0,5\%$$

Dari nilai akurasi serta error rate data training dengan menggunakan algoritma C4.5 memiliki nilai akurasi lebih dari 90%, yakni dengan nilai 99,5%. hal ini menunjukkan bahwa algoritma C4.5 dapat digunakan pada database kelahiran bayi yang ada di Dinas Penduduk dan Catatan Sipil.

## PENUTUP

Secara rincian jumlah True Positive (TP) 911, False Negative (FN) 1, False Positive (TP) 4, True Negative (FN) 18, nilai presisi dan recall yang hampir mendekati nilai 1 menunjukkan bahwa item atau atribut yang terpilih adalah relevan. Hal ini juga didukung dengan nilai akurasi sebesar 99,5%. Pohon keputusan akhir yang dihasilkan menunjukkan terdapat tiga daerah yang menjadi lokasi prioritas yakni Jaya Mukti, Buluh Kasap, dan Bukit Batrem, Sehingga dapat disimpulkan bahwa Algoritma C4.5 dapat diimplementasikan untuk menentukan lokasi prioritas dalam melakukan penyuluhan program keluarga berencana.

Untuk penelitian berikutnya bisa dilakukan optimasi pada tahap pemilihan atribut sehingga kompleksitas atribut dapat dikurangi, dengan demikian diharapkan nilai akurasi akan meningkat.

## UCAPAN TERIMA KASIH

Terima kasih penulis sampaikan kepada KEMENRISTEK DIKTI P2M yang telah memberikan kepercayaan dan membiayai Penelitian Dosen Pemula ini pada pembiayaan tahun 2017. Terima kasih juga penulis sampaikan kepada Tim Redaksi Jurnal Penelitian Pos dan Informatika yang telah meluangkan waktu untuk kerjasamanya.

## DAFTAR PUSTAKA

- BKKBN. (2010). "KB dan Kontrasepsi".  
Diakses dari  
[http://bkkbn.go.id/difor/program\\_detail.php?prgid=8](http://bkkbn.go.id/difor/program_detail.php?prgid=8). tanggal 3 maret 2016.
- Bouckaert, Remco R. Frank, Eibe. 2008. *WEKA Manual For Version 3-6-0*. New Zealand: University of Waikato.
- Chauhan, H and Chauhan, A. (2013). *Implementation of Decision Tree Algorithm C4.5*. International Journal of Scientific and Research Publication Vol.3 issue 10, October 2013.
- Han, Jiawei., and Kember, Micheline. (2006). *Data Mining Concepts and Techniques*. Second Edition. Morgan Kaufman Publisher.
- Hamdan, A. R., and Othman, Z. A. (2010). *Human Talent Prediction in HRM using Clasiffication Algorithm*. Selangor, Malaysia.
- Hermawati, Fajar Astuti. (2009). *Data Mining*. Surabaya : Andi Offset, Yogyakarta.
- Kusnawi. (2007). *Pengantar Solusi Data Mining*.  
Diakses dari  
<http://p3mamikom.ac.id/p3m/56-pengantas-solusi-dataming>. pdf. tanggal 22 maret 2016.
- Larose, D. T. (2005). *Discovering Knowledge In Data.Mining An Introduction To Data Mining*, Wiley Interscience.
- Rahmayuni, Indri. (2014). *Perbandingan Performasi Algoritma C4.5 dan CART dalam klasifikasi data nilai mahasiswa prodi Teknik Komputer Politeknik Padang*. Jurnal : TEKNOIF. ISSN:2338-2724. Vol2. No.1.
- Wu, X. and Kumar , V. (2009). *The Top Ten Algorithms in Data Mining*. New York : CRC Press.