

**KOMPARASI METODE KLASIFIKASI DATA MINING
ALGORITMA C4.5 DAN NAÏVE BAYES UNTUK PEMILIHAN
ALAT KONTRASEPSI PADA KANTOR DINAS BPKB
KEDUNGREJA CILACAP**

*¹Suprpto, ²Suwarno, ³Maria Yustina
^{1,2,3}Fakultas Teknik dan Informatika Universitas AKI*

Jl. Imam Bonjol No. 15-17, Semarang 50173, Jawa Tengah

suprpto@unaki.ac.id, suwarno@unaki.ac.id, 223180021@student.unaki.ac.id

Abstrak

Jumlah penduduk yang semakin banyak mengharuskan pemerintah untuk mengurangi ledakan penduduk, salah satunya dengan program Keluarga Berencana(KB). Dalam program keluarga Berencana sering terjadi kesalahan dalam memilih alat kontrasepsi. Alat kontrasepsi merupakan alat untuk pencegah kehamilan, kurangnya pengetahuan tentang alat kontrasepsi menyebabkan akseptor menggunakan alat kontrasepsi yang tidak sesuai dengan kondisi tubuhnya atau bahkan tidak menggunakan kontrasepsi. Dengan adanya komparasi algoritma C4.5 dan Naïve Bayes diharapkan dapat mempermudah akseptor dalam memilih alat kontrasepsi. Tujuan dilakukannya komparasi klasifikasi algoritma C4.5 dan Naïve Bayes adalah untuk membantu akseptor memilih alat kontrasepsi dengan menggunakan data akseptor KB pada Kantor Dinas BPKB Kecamatan Kedungreja Cilacap dan akurasi serta kappa yang diperoleh dari hasil klasifikasi yang dilihat dari faktor umur istri, status kerja istri, jenis pelayanan KB, kesertaan KB dan Faskes. Penelitian ini menggunakan metode klasifikasi data mining Algoritma C4.5 dan Naïve Bayes kemudian dilakukan perbandingan kedua metode. Pengolahan dua metode tersebut menggunakan confusion matrix dan kappa. Hasil penelitian ini menghasilkan akurasi algoritma C4.5 sebesar 86.06% dan kappa 0.638 sedangkan tingkat akurasi Naïve Bayes sebesar 92.46% dan kappa 0.880. Sehingga algoritma Naïve Bayes merupakan metode yang lebih baik dalam pengklasifikasian data akseptor KB pada Kantor Dinas BPKB Kecamatan Kedungreja Cilacap tahun 2021 dibandingkan dengan metode algoritma C4.5.

Kata Kunci: *Data Mining, C4.5, Naïve Bayes, Alat kontrasepsi, RapidMiner.*

Abstract

The increasing number of residents requires the government to reduce the population explosion, one of which is the Family Planning(KB) program. In family planning programs, mistakes often occur in choosing contraceptives. Contraceptive device are tools to prevent pregnancy, lack of knowledge about contraceptives causes acceptors to use contraceptives that are not in accordance with their body conditions or even do not use contraception. With the comparison of the C4.5 and Naïve Bayes algorithms, it is hoped that it will make it easier for acceptors to choose contraceptives. The purpose of the comparison of classification algorithms C4.5 and Naïve Bayes is to help acceptors choose contraceptives using family planning acceptor data at the BPKB Office, Kedungreja District, Cilacap and the accuracy and kappa obtained from the classification results seen from the wife's age, wife's work status, types of family planning service, participation in family planning and health facilities. This research uses data mining classification method C4.5 algorithm and Naïve Bayes then a comparison of the two methods is carried out. The processing of these two method in the accuracy of the C4.5 algorithm of 86.06% and kappa 0.638, while the accuracy of Naïve Bayes was 92.46% and kappa 0.880. So the Naïve Bayes algorithm is a better method in classifying family planning acceptor data at the BPKB Office, Kedungreja Cilacap District in 2021 compared to the C4.5 algorithm method.

Key Words: Data Mining, C4.5, Naïve Bayes, Alat kontrasepsi, RapidMiner.

PENDAHULUAN

Kepadatan penduduk terjadi karena lajunya pertumbuhan penduduk sangat tinggi setiap tahunnya. Kondisi ini menyebabkan negara harus mempersiapkan lapangan pekerjaan yang sangat luas untuk mengatasi kemiskinan, pengangguran dan kriminalitas. Pemerintah melakukan berbagai upaya untuk menekan lajunya pertumbuhan penduduk, salah satunya dengan melakukan program Keluarga Berencana(KB). Dengan adanya program KB diharapkan bisa menekan lajunya pertumbuhan penduduk. Banyak dari akseptor atau pengguna KB salah dalam memilih alat kontrasepsi yang akan digunakan dikarenakan belum memahami kegunaan dan fungsi dari alat kontrasepsi dengan benar. Untuk membantu akseptor dalam memilih alat kontrasepsi dilakukan pengolahan data menggunakan metode klasifikasi. Hal ini sesuai dengan penelitian terdahulu yang dilakukan oleh Pertiwi, Adiwisastro, & Supriadi, (2019) dalam penelitiannya yang berjudul *Analisa Komparasi Menggunakan 5 Metode Data Mining dalam Klasifikasi Persentase Wanita Sudah menikah di Usia 15-49 yang Memakai Alat KB (Keluarga Barenca)*. Dari hasil penelitian komparasi kelima algoritma tersebut yang menunjukkan baik dan akurasinya lebih besar adalah algoritma C4.5 dengan nilai *accuracy* 87.50%. Fauzia, (2021) dalam penelitiannya yang berjudul *Komparasi Algoritma Data Mining Naïve Bayes dan C4.5 untuk*

Klasifikasi Penerimaan Peserta Didik Baru di SMPN 35 Semarang. Dapat disimpulkan bahwa algoritma yang lebih baik untuk dataset PPBD SMPN 35 Semarang yaitu *Naïve Bayes*. Septiani, (2017) dalam penelitiannya yang berjudul *Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 dan Naïve Bayes untuk Prediksi Penyakit Hepatitis*. Dalam penelitian ini menurut pengujian berdasarkan *accuracy*, algoritma terbaik adalah *Naïve Bayes*. Sedangkan menurut pengujian berdasarkan ROC Curve (AUC) algoritma yang terbaik adalah C4.5. Untuk itu pada penelitian ini akan dilakukan komparasi klasifikasi menggunakan algoritma C4.5 dan *Naïve Bayes* untuk membandingkan kedua metode dalam penggunaan alat kontrasepsi.

METODE PENELITIAN

Pada penelitian ini data yang digunakan adalah data akseptor KB yang didapat dari Kantor Dinas BPKB Kecamatan Kedungreja Cilacap dengan alamat Kantor Jl.Suhada Desa Ciklapa, Kecamatan Kedungreja, Kabupaten Cilacap. Dalam penelitian ini akan dilakukan beberapa langkah-langkah atau tahapan penelitian seperti di bawah ini:

1. Pengumpulan Data (*Ghatering*)
Dalam tahap ini dilakukan penentuan data yang akan diproses yaitu data Akseptor KB Kantor Dinas BPKB Kecamatan Kedungreja Cilacap. Mengintegrasikan semua data kedalam *dataset*, termasuk membagi variabel yang diperlukan dalam proses.
2. Pengolahan Data Awal (*Data Pre-processing*)
Pada langkah ini dilakukan penyeleksian data, Jumlah data awal yang diperoleh 531 data. Setelah dilakukan pengolahan atau pembersihan data, sebanyak 531 data yang terdapat dalam *dataset* Akseptor KB Kantor Dinas BPKB Kecamatan Kedungreja Cilacap dapat digunakan. Pada tahap pengolahan data awal dilakukan untuk mempersiapkan data yang benar-benar valid sebelum dilanjutkan dengan tahap selanjutnya.
3. Metode yang diusulkan (*Proposed Method*)
Pada langkah ini data akan dianalisis, untuk menemukan variabel mana yang terhubung dengan satu variabel dengan variabel lainnya. Setelah data dianalisis selanjutnya diterapkan model-model yang sesuai dengan jenis data. Pada penelitian ini peneliti menggunakan metode *Confusion Matrix* dan Kappa.
4. Pengujian Model dan Eksperimen (*Model Testing and Experimen*)
Dalam tahap ini model yang diusulkan adalah metode klasifikasi *data mining* algoritma C4.5 dan *Naïve Bayes*. Pengujian model menggunakan *RapidMiner* untuk mengetahui hasil berupa nilai akurasi dan nilai Kappa.
5. Evaluasi dan Validasi Hasil (*Result Evaluation*)
Pada tahap terakhir ini dilakukan evaluasi terhadap model yang ditetapkan untuk mengetahui tingkat akurasi model.

HASIL DAN PEMBAHASAN

Hasil Evaluasi Algoritma C4.5

Pada tahap ini dilakukan eksperimen dan pengujian metode yang digunakan yaitu C4.5. Langkah-langkah yang dilakukan sebagai berikut:

1. Menghitung jumlah kasus, “Suntik”, “Pil”, “Kondom”, “Implan”, dan “IUD” serta nilai Entropy dari semua kasus. Dari data yang ada diketahui jumlah kasus yang “Suntik” sebanyak 448 record, kasus “Pil” sebanyak 65 record, kasus “Kondom” sebanyak 13 record, kasus “Implan” sebanyak 4 record, dan kasus “IUD” sebanyak 1 record total keseluruhan kasus adalah 531 record. Sehingga didapat Entropy keseluruhan:

$$Entropi(S) = - \sum_{j=1}^k P_j * \log_2 P_j$$

$$\begin{aligned} Entropy(S) &= -\left(\frac{448}{531}\right) * \log_2\left(\frac{448}{531}\right) + \left(-\left(\frac{65}{531}\right) * \log_2\left(\frac{65}{531}\right) + \right. \\ &\left. -\left(\frac{13}{531}\right) * \log_2\left(\frac{13}{531}\right) + \left(-\left(\frac{4}{531}\right) * \log_2\left(\frac{4}{531}\right) + \left(-\left(\frac{1}{531}\right) * \right. \right. \\ &\left. \left. \log_2\left(\frac{1}{531}\right) \right) \right) \\ &= 0,7790 \end{aligned}$$

2. Sebelum menghitung *gain* dari masing-masing atribut akan dilakukan perhitungan nilai *entropy* terlebih dahulu. Misalkan menghitung *entropy* atribut umur akseptor.

$$\begin{aligned} Entropy \leq 35 \quad (242,22,10,2,0) &= -\left(\left(\frac{242}{276}\right) * \log_2\left(\frac{242}{276}\right) + \left(\frac{22}{276}\right) * \right. \\ &\left. \log_2\left(\frac{22}{276}\right) + \left(\frac{10}{276}\right) * \log_2\left(\frac{10}{276}\right) + \left(\frac{2}{276}\right) * \log_2\left(\frac{2}{276}\right) + \left(\frac{0}{276}\right) * \right. \\ &\left. \log_2\left(\frac{0}{276}\right) \right) \\ &= 0 \end{aligned}$$

$$\begin{aligned} Entropy > 35 \quad (206,43,3,2,1) &= -\left(\left(\frac{206}{255}\right) * \log_2\left(\frac{206}{255}\right) + \left(\frac{43}{255}\right) * \right. \\ &\left. \log_2\left(\frac{43}{255}\right) + \left(\frac{3}{255}\right) * \log_2\left(\frac{3}{255}\right) + \left(\frac{2}{255}\right) * \log_2\left(\frac{2}{255}\right) + \left(\frac{1}{255}\right) * \right. \\ &\left. \log_2\left(\frac{1}{255}\right) \right) \\ &= 0.8433 \end{aligned}$$

Kemudian menghitung nilai *gain* atribut umur akseptor sebagai berikut:

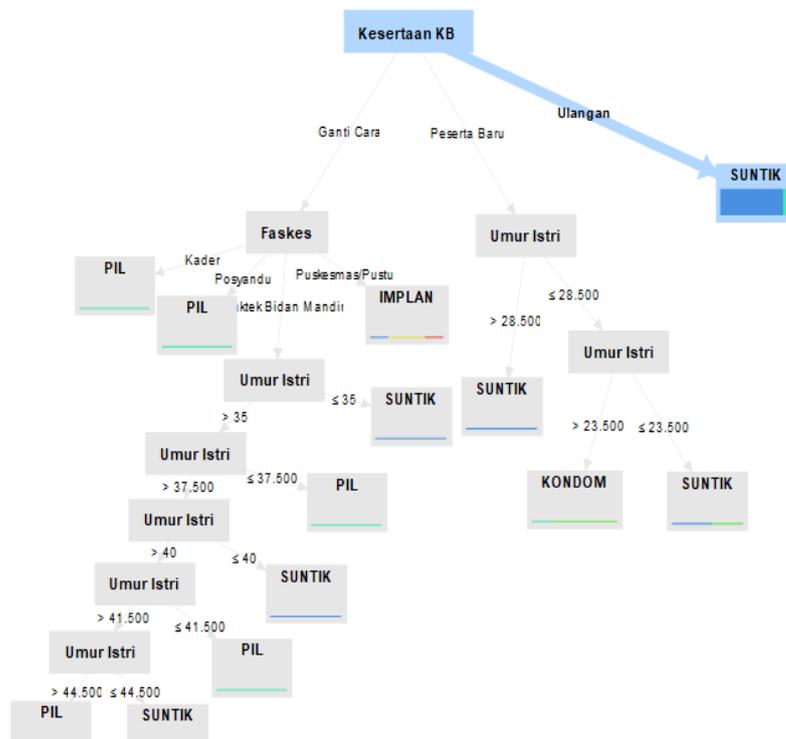
$$Gain(S, A) = Entropi(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropi(S_i)$$

$$Gain(S, A) = (0,7790) - \left(\frac{276}{531} * 0\right) - \left(\frac{255}{531} * 0,8433\right) = 0,374$$

Tabel 5.1 Nilai Entropy dan Gain pada Algoritma C4.5

Simpul		Kasus	Suntik	Pil	Kondom	Implan	IUD	Entropy	Gain
Jumlah Kasus		531	448	65	13	4	1	0,7790	
Umur Istri	≤ 35	276	242	22	10	2	0	0	0,374
	>35	255	206	43	3	2	1	0,8433	
Status Kerja Istri	Kerja	435	373	48	9	4	1	0,7392	0,173
	Tidak Bekerja	96	75	17	4	0	0	0	
Kesertaan KB	Ulangan	478	424	45	7	2	0	0	0,779
	P.Baru	19	12	1	6	0	0	0	
	G.Cara	34	12	19	0	2	1	0	
Faskes	P.Bidan Mandiri	453	432	17	2	2	0	0	0,779
	Posyandu	19	0	12	7	0	0	0	
	Kader	27	0	23	4	0	0	0	
	Puskesmas	22	13	6	0	2	1	0	
	Apotik	2	0	2	0	0	0	0	
	Klinik/KKB	2	2	0	0	0	0	0	
	PPKBD	3	0	3	0	0	0	0	
	Polindes	1	1	0	0	0	0	0	

Setelah melakukan perhitungan semua atribut nilai entropi dan gainnya dapat dilihat nilai gain yang tertinggi. Nilai gain tertinggi akan digunakan sebagai node atau root pada pohon yang akan dibangun. Nilai gain tertinggi pada proses diatas adalah Kesertaan KB dan Faskes, maka kesertaan KB dipilih sebagai node atau root 1. Pohon keputusan yang terbentuk dari *dataset* akseptor KB Kecamatan Kedungreja Cilacap dengan menggunakan perhitungan nilai *gain* dan *RapidMiner* dapat dilihat pada gambar 5.5.



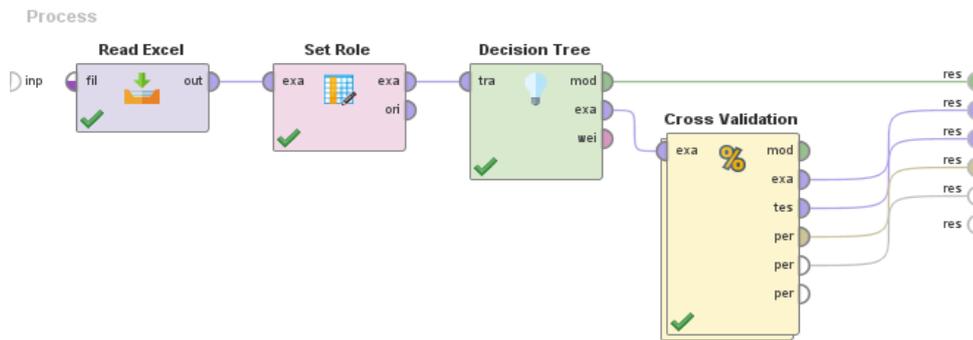
Gambar 5.5 Pohon Keputusan

Dari pohon keputusan didapat *rule* untuk prediksi pemilihan suntik, pil, kondom, implan dan IUD akseptor KB Dinas BPKB Kedungreja Cilacap sebagai berikut:

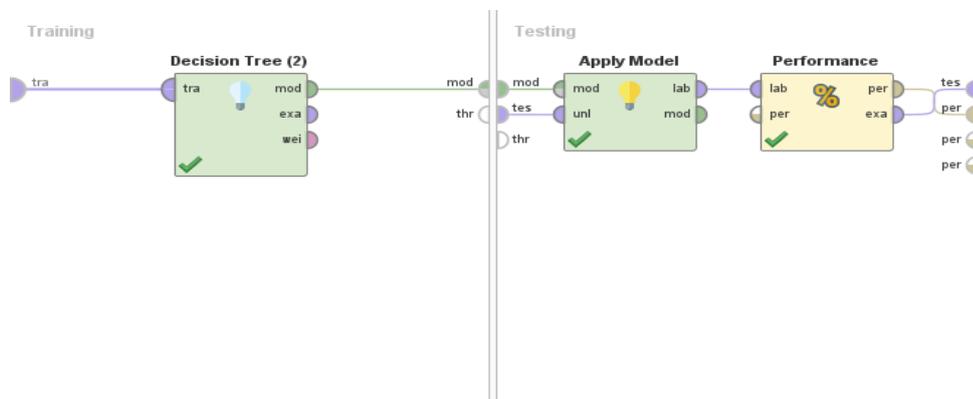
- R1* : JIKA Kesertaan KB ganti cara DAN Faskes kader MAKA *class* = Pil
- R2* : JIKA Kesertaan KB ganti cara DAN Faskes posyandu MAKA *class* = Pil
- R3* : JIKA Kesertaan KB ganti cara DAN Faskes P. Bidan mandiri DAN Umur Istri > 35 DAN Umur Istri > 44.5 MAKA *class* = Pil
- R4* : JIKA Kesertaan KB ganti cara DAN Faskes P. Bidan mandiri DAN Umur Istri >35 DAN Umur Istri ≤ 44.5 MAKA *class* = Suntik
- R5* : JIKA Kesertaan KB ganti cara DAN Faskes P. Bidan mandiri DAN Umur Istri >35 DAN Umur Istri ≤ 41.5 MAKA *class* = PIL
- R6* : JIKA Kesertaan KB ganti cara DAN Faskes P. Bidan mandir DAN Umur Istri >35 DAN Umur Istri ≤ 40 MAKA *class* = Suntik
- R7* : JIKA Kesertaan KB ganti cara DAN Faskes P. Bidan mandiri DAN Umur istri >35 DAN Umur Istri ≤ 37.5 MAKA *Class* = Pil

- R8 : JIKA Kesertaan KB ganti cara DAN Faskes P.Bidan mandiri DAN Umur istri ≤ 35 MAKA Class = Suntik
- R9 : JIKA Kesertaan KB ganti cara DAN Faskes puskesmas/pustu MAKA Class = Implan
- R10 : JIKA Kesertaan KB peserta baru DAN Umur Istri > 28.5 MAKA Class = Suntik
- R11 : JIKA Kesertaan KB peserta baru DAN Umur Istri ≤ 28.5 DAN Umur Istri > 23.5 MAKA Class = Kondom
- R12 : JIKA Kesertaan KB peserta baru DAN Umur Istri ≤ 28.5 DAN Umur Istri ≤ 23.5 MAKA Class = Suntik
- R13 : JIKA Kesertaan KB ulangan MAKA Class = Suntik

Proses *Cross Validation* merupakan teknik validasi yang membagi data menjadi dua bagian secara acak, menjadi data *training* dan *testing*. Berikut ini merupakan proses pengolahan data menggunakan *rapidminer*.



Gambar 5.1 Proses Pengujian Cross Validation Metode C4.5



Gambar 5.2. Proses Cross Validation Metode C4.5

Hasil Evaluasi Algoritma Naïve Bayes

Naïve Bayes merupakan model kedua yang akan dihitung. Langkah-langkah yang akan dilakukan adalah menghitung nilai probabilitas pada rumus berikut ini:

$$P(S/J) = \frac{P(J|S)P(S)}{P(J)}$$

$$P(\text{Jenis Pelayanan KB}|\text{Suntik}) = 448/531 = 0.8436911$$

$$P(\text{Jenis Pelayanan KB}|\text{Pil}) = 65/531 = 0.1224105$$

$$P(\text{Jenis Pelayanan KB}|\text{Kondom}) = 13/531 = 0.0244821$$

$$P(\text{Jenis Pelayanan KB}|\text{Implan}) = 4/531 = 0.0075329$$

$$P(\text{Jenis Pelayanan KB}|\text{Suntik}) = 1/531 = 0.0018832$$

Tabel 5.3 Perhitungan Probabilitas Jenis Pelayanan KB

PROBABILITAS JENIS PELAYANAN KB	
JENIS PELAYANAN KB	NILAI
SUNTIK	0.8436911
PIL	0.1224105
KONDOM	0.0244821
IMPLAN	0.0075329
IUD	0.0018832
Total	1

Perhitungan formula *Naïve Bayes* dapat mereduksi kompleksitas perhitungan $P(x/C_i)$ dengan asumsi naif independensi bersyarat kelas, yaitu nilai atribut saling independen yang artinya antar atribut adalah saling bebas, yang berarti tidak ada ketergantungan sama sekali. Perhitungan *Naïve Bayes classifier* memaksimalkan:

$$P(x/C_i) = \prod_{k=1}^n P(x_k/C_i) * P(x_1/C_i) * \dots * P(x_n/C_i)$$

$$P(C_i/\text{Suntik}) = 0.8436911 * 0.5401785 * 0.832589 * 0.946428 * 0.96428 = 0.346$$

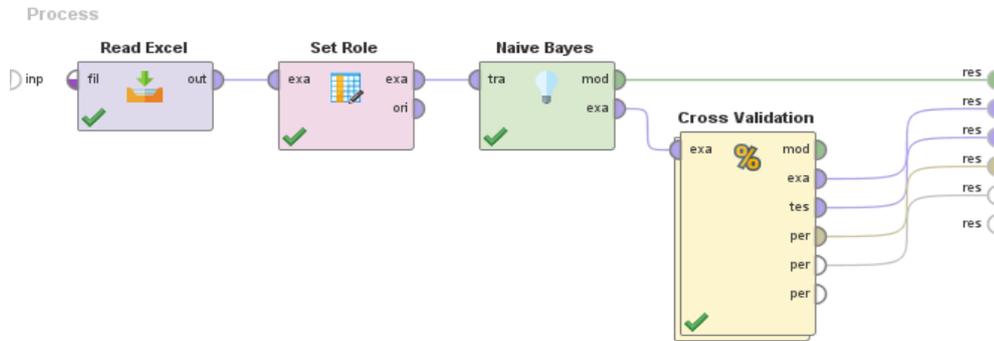
$$P(C_i/\text{Pil}) = 0.1224105 * 0.3385 * 0.7385 * 0.6923 * 0.2615 = 5.539$$

$$P(C_i/\text{Kondom}) = 0.0244821 * 0.769231 * 0.692308 * 0.538462 * 0.153846 = 1.080$$

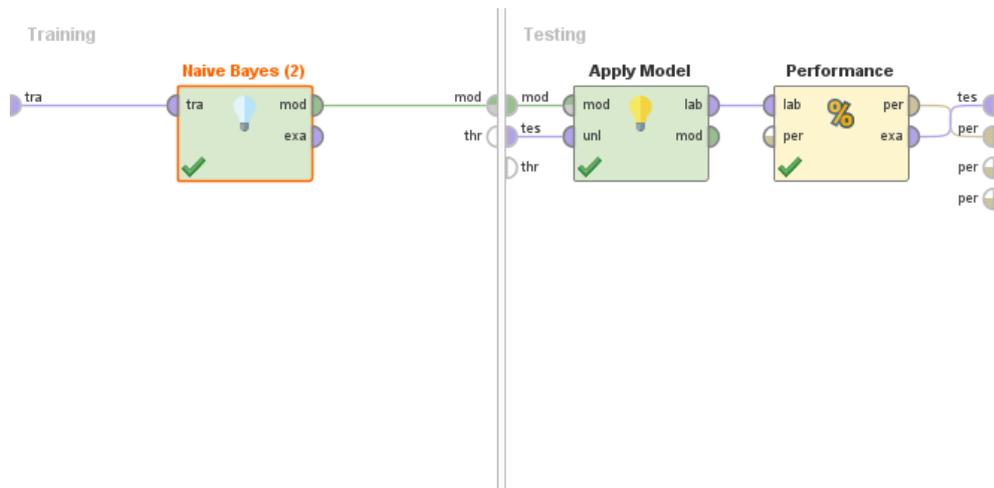
$$P(C_i/\text{Implan}) = 0.0075329 * 0.5 * 1 * 0.5 * 0.5 = 9.1416$$

$$P(C_i/\text{IUD}) = 0.0018832 * 0 * 1 * 0 * 0 = 0$$

Proses *Cross Validation* merupakan teknik validasi yang membagi data menjadi dua bagian secara acak, menjadi data *training* dan *testing*. Berikut ini merupakan proses pengolahan data menggunakan *rapidminer*.



Gambar 5.6. Pengujian Cross Validation Naïve Bayes



Gambar 5.7 Proses Cross Validation Naïve Bayes

EVALUASI DAN VALIDASI HASIL

Hasil proses menggunakan *rapidminer* diperoleh nilai *accuracy* sebesar 86.07% dari 531 sebanyak 1 data diprediksi pengguna implan tetapi pengguna IUD, 4 data diprediksi pengguna suntik tetapi pengguna implan, 3 data diprediksi sesuai pengguna komdom, 10 data diprediksi pengguna suntik tetapi pengguna kondom, 15 data diprediksi sesuai pengguna pil, 50 data diprediksi pengguna suntik tetapi pengguna pil, 1 data diprediksi pengguna implan tetapi pengguna suntik, 8 data diprediksi pengguna pil tetapi pengguna suntik, dan 439 data diprediksi sesuai pengguna suntik. Hasil *accuracy* dapat dilihat pada gambar 5.3 berikut ini:

accuracy: 86.07% +/- 2.01% (micro average: 86.06%)

	true SUNTIK	true PIL	true KONDOM	true IMPLAN	true IUD	class precision
pred. SUNTIK	439	50	10	4	0	87.28%
pred. PIL	8	15	0	0	0	65.22%
pred. KONDOM	0	0	3	0	0	100.00%
pred. IMPLAN	1	0	0	0	1	0.00%
pred. IUD	0	0	0	0	0	0.00%
class recall	97.99%	23.08%	23.08%	0.00%	0.00%	

Gambar 5.3. Hasil accuracy Pada Metode C4.5

Hasil analisis *dataset* evaluasi akseptor KB Kecamatan kedungreja Cilacap dengan menggunakan *rapidminer* pada algoritma *C4.5* diperoleh tingkat akurasi 86.07% yang artinya tingkat akurasi data *good classification*. Klasifikasi dapat dihitung tingkat akurasinya berdasarkan kinerja matriks. Dalam perhitungan akurasi dengan klasifikasi dapat dilakukan dengan perhitungan sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Accuracy = \frac{439+15+3+0+0}{439+50+10+4+0+8+15+0+0+0+0+0+0+3+0+0+1+0+0+0+1+0+0+0+0+0}$$

$$Accuracy = 0.8606 * 100\%$$

$$Accuracy = 86.06\%$$

Untuk menghitung nilai precision adalah sebagai berikut:

$$True\ Positive\ Rate\ (Sensitivity) = \frac{TP}{TP+FN}$$

$$True\ Positive\ Rate\ (Suntik) = \frac{439}{439+50+10+4+0}$$

$$= 0.8727$$

$$= 87.27\%$$

Untuk menghitung nilai *recall* adalah sebagai berikut:

$$PPV = \frac{TN}{TN+FP}$$

$$\begin{aligned} \text{Positive Predictive Value (Suntik)} &= \frac{439}{439+8+0+1+0} \\ &= 0.9799 \\ &= 97.99\% \end{aligned}$$

Perubahan kemungkinan hasil

$$\text{-Suntik} = 92.72\% * 84.36\% = 7.8\%$$

$$\text{-Pil} = 04.33\% * 12.24\% = 52\%$$

$$\text{-Kondom} = 02.44\% * 00.56\% = 1.3\%$$

$$\text{-Implan} = 00.75\% * 00.37\% = 0.2\%$$

$$\text{-IUD} = 0\% * 0.18\% = 0\%$$

$$\text{Total} = 7.8\% + 52\% + 1.3\% + 0.2\% + 0\% = 61.3\% = 0.613$$

$$\begin{aligned} \text{Kappa} &= \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \\ &= \frac{0.8606 - 0.613}{1 - 0.613} \\ &= 0.638 \end{aligned}$$

Hasil proses menggunakan *RapidMiner* diperoleh nilai *accuracy* sebesar 92.46% dari 531 sebanyak 1 data diprediksi pengguna implan tetapi pengguna IUD, 2 data diprediksi pengguna pil tetapi pengguna implan, 2 data diprediksi pengguna suntik tetapi pengguna implan, 5 data diprediksi sesuai pengguna kondom, 6 data diprediksi pengguna pil tetapi pengguna kondom, 2 data diprediksi pengguna suntik tetapi pengguna kondom, 1 data diprediksi pengguna implan tetapi pengguna pil, 43 data diprediksi sesuai pengguna pil, 21 data diprediksi pengguna suntik tetapi pengguna pil, 5 data diprediksi pengguna pil tetapi pengguna suntik, dan 443 data diprediksi sesuai pengguna suntik. Hasil *accuracy* dapat dilihat pada gambar 5.8 dibawah ini.

accuracy: 92.46% +/- 1.56% (micro average: 92.47%)

	true SUNTIK	true PIL	true KONDOM	true IMPLAN	true IUD	class precision
pred. SUNTIK	443	21	2	2	0	94.66%
pred. PIL	5	43	6	2	0	76.79%
pred. KONDOM	0	0	5	0	0	100.00%
pred. IMPLAN	0	1	0	0	1	0.00%
pred. IUD	0	0	0	0	0	0.00%
class recall	98.88%	66.15%	38.46%	0.00%	0.00%	

Gambar 5.8. Hasil accuracy Pada Metode Naïve Bayes

Klasifikasi dapat dihitung tingkat akurasi berdasarkan kinerja matriks. Dalam perhitungan akurasi dengan klasifikasi dapat dilakukan dengan perhitungan sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Accuracy = \frac{443+43+5+0+0}{443+21+2+2+0+5+43+6+2+0+0+0+5+0+0+0+1+0+0+1+0+0+0+0+0}$$

$$Accuracy = 0.9246 * 100\%$$

$$Accuracy = 92.46\%$$

Untuk menghitung nilai precision adalah sebagai berikut:

$$True\ Positive\ Rate\ (Sensitivity) = \frac{TP}{TP+FN}$$

$$\begin{aligned} True\ Positive\ Rate\ (Suntik) &= \frac{443}{443+21+2+2+0} \\ &= 0.9465 \\ &= 94.65\% \\ &= 0\% \end{aligned}$$

Untuk menghitung nilai recall adalah sebagai berikut:

$$PPV = \frac{TN}{TN+FP}$$

$$\begin{aligned} Positive\ Predictive\ Value\ (Suntik) &= \frac{443}{443+5+0+1+0} \\ &= 0.9888 \\ &= 98.88\% \end{aligned}$$

Untuk perhitungan nilai Kappa adalah sebagai berikut:

$$Pr(\alpha) = \frac{443+43+5}{531} = 0.9246$$

Perubahan kemungkinan hasil

$$-Suntik = 88.13\% * 84.36\% = 7.4\%$$

$$-Pil = 10.54\% * 12.24\% = 129\%$$

$$-Kondom = 94\% * 02.44\% = 229\%$$

$$-\text{Implan} = 00.37\% * 00.75\% = 0.2\%$$

$$-\text{IUD} = 0\% * 18\% = 0\%$$

$$\text{Total} = 7.4\% + 129\% + 229\% + 0.2\% + 0\% = 36.5\% = 0.365$$

$$\begin{aligned} \text{Kappa} &= \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)} \\ &= \frac{0.9246 - 0.365}{1 - 0.365} = 0.880 \end{aligned}$$

PEMBAHASAN

Berdasarkan hasil pengujian *dataset* akseptor KB Kecamatan Kedungreja Cilacap tahun 2021 dengan menggunakan algoritma *C4.5* dan *Naïve Bayes*, maka hasil yang diperoleh dirangkum pada Tabel 5.9.

TABEL 5.9

EVALUASI DAN VALIDASI PENGUJIAN DATASET

	<i>C4.5</i>	<i>Naïve Bayes</i>
<i>Accuracy</i>	86.06%	92.46%
Kappa	0.638	0.880
Kategori Kappa	Kuat(Good)	Sangat Kuat(Very good)
Pohon Keputusan	Terbentuk baik	-

Semakin tinggi nilai *accuracy* dan nilai Kappa yang diperoleh, maka semakin baik pula model klasifikasi yang dihasilkan. Dari hasil perbandingan berdasarkan nilai *accuracy* dan Kappa, yang dapat dilihat pada tabel 5.9 algoritma yang lebih baik untuk *dataset* akseptor KB Dinas BPKB Kecamatan Kedungreja Cilacap yaitu *Naïve Bayes*, dengan nilai *accuracy* 92.46% dan Kappa 0.880 yang terbentuk sangat kuat(Very good).

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dengan komparasi menggunakan algoritma *C4.5* dan *Naïve Bayes* untuk pemilihan alat kontrasepsi pada Kantor Dinas BPKB Kedungreja Cilacap dapat disimpulkan bahwa:

1. Pemilihan penggunaan alat kontrasepsi dengan komparasi metode klasifikasi menggunakan algoritma *C4.5* menghasilkan nilai *accuracy* sebesar 86.06% dan nilai Kappa 0.638 yang termasuk kategori kuat (good) dan algoritma *Naïve Bayes* menghasilkan nilai *accuracy* sebesar 92.46% dan nilai Kappa 0.880 yang termasuk kategori sangat kuat (very good).
2. Hasil pengujian komparasi *C4.5* dan *Naïve Bayes* dapat dilihat bahwa algoritma *Naïve Bayes* merupakan algoritma yang lebih baik dengan nilai *accuracy* dan Kappa yang tinggi. Maka algoritma *Naïve Bayes* dapat digunakan sebagai acuan untuk klasifikasi penggunaan suntik, pil, kondom,

implan dan IUD untuk pemilihan penggunaan alat kontrasepsi pada akseptor KB Dinas PLKB Kedungreja Cilacap.

3. Atribut kesertaan KB merupakan atribut yang memiliki pengaruh lebih besar, hal ini dapat dilihat dari hasil perhitungan gain dan pohon keputusan pada *rapidminer* yang memunculkan nilai kesertaan KB sebagai *node* tertinggi.

Penelitian selanjutnya dapat menggunakan data atau menambahkan atribut lebih banyak yang dapat mempengaruhi proses pemilihan alat kontrasepsi.

Untuk keperluan penelitian lebih lanjut mengenai metode klasifikasi data mining dengan menggunakan data akseptor KB, maka disarankan untuk menambahkan beberapa algoritma klasifikasi data mining untuk dikomparasikan seperti *Artificial Neural Network*, *Linear Discriminant Analysis*, *K-Nearest Neighbors*, *Support Vector*, dan lain-lain.

DAFTAR PUSTAKA

- Annisa, R. (2019). Analisa komparasi algoritma klasifikasi data mining untuk prediksi penderita penyakit jantung. *Jurnal Teknik Informatika Kaputama*, 3(1).
- Fauzia, B. (2021). Komparasi algoritma data mining naïve bayes dan C4.5 untuk klasifikasi penerimaan peserta didik baru di SMPN 35 Semarang. Semarang: Unaki.
- Gorunescu, F. (2011). *Data Mining : Concepts, Model and Techniques*. Berlin, Jerman: Springer.
- Han, J. (2007). *Data Mining Concept And Technique*. Champaign: Multiscience Press.
- Han, J., & Kamber, M. (2012). *Data Mining and Techniques Second Edition*. San Fransisco: Morgan Kaufmann Publishers.
- Kurniabudi., Harris, A., & Mintaria, A. E. (2021). Komparasi information gain, gain ratio, cfs-bestfirst dan cfs-pso search terhadap performa deteksi anomali. *Jurnal Media Informatika Budidarma*, 5 (1), 332-343.
- Kurniawan, I. Y. (2018). Perbandingan algoritma naïve bayes dan C4.5 dalam klasifikasi data mining. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 5(4), 455-464.
- Kusrini, & Luthfi, E. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- Pertiwi, M. W., Adiwisastro, M. F., & Supriadi, D. (2019). Analisis komparasi menggunakan 5 metode data mining dalam klasifikasi persentase wanita sudah menikah di usia 15-49 yang memakai alat KB (Keluarga berencana). *Jurnal Khatulistiwa Informatika*.

- Rismia, E. R., Widiharih, T., & Santoso, R. (2021). Klasifikasi regresi logistik multinomial dan fuzzy k-nearest neighbor (FK-NN) dalam pemilihan metode kontrasepsi di kecamatan bulakamba, kabupaten brebes, jawa tengah. *Jurnal Gaussiani*, 10 (4), 476-487.
- Sartika, D., & Sensuse, D. I. (2017). Perbandingan algoritma klasifikasi naïve bayes, nearest neighbour dan decision tree pada studi kasus pengambilan keputusan pemilihan pola pakaian. *Jurnal Jatsi*.
- Septiani, W. D. (2017). Komparasi metode klasifikasi data mining algoritma C4.5 dan naïve bayes untuk prediksi penyakit hepatitis. *Jurnal Pilar Nusa Mandiri*, 13(1).
- Sodik, A. (2015). Dasar metodologi Penelitian. Yogyakarta: Literasi Media Publishing.
- Sugiyono. (2015). *Metodologi Penelitian Kuantitatif, Kualitatif, dan R&Do Title*. Bandung: Alfabete.