



INNOVATIVE: Journal Of Social Science Research

Volume 3 Nomor 1 Tahun 2023 Page 503-513

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naïve Bayes Untuk Pemilihan Penggunaan Jenis KB Di Bidan Swasta H.Enok Hayati

Novia Cahya Utami^{1✉}, Aprilia Hananto², Tukino³, Shofa Shofiah Hilabi⁴

Universitas Buana Perjuangan Karawang

Email: si18.noviautami.mhs@ubpkarawang.ac.id^{1✉}

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. Penggunaan kontrasepsi merupakan hal yang penting, mengingat dapat menurunkan laju pertumbuhan. Penggunaan jenis KB yang tepat untuk para wanita merupakan permasalahan klasifikasi. Dalam penelitian ini diambil data mengenai klasifikasi penggunaan kontrasepsi di bidan swasta H Enok Hayati yang melayani beberapa jenis KB yaitu seperti KB suntik , KB pil , IUD (Intrauterine Device) dan Implan. 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 89.55% dan kappa 0.851 sedangkan tingkat akurasi Naïve Bayes sebesar 34.85% dan kappa 0.016. Sehingga algoritma C4.5 merupakan metode yang lebih baik dalam pengklasifikasian data akseptor KB pada Bidan Swasta HJ Enok Hatiyah dibandingkan dengan metode algoritma Naïve Bayes"

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

Abstract

"The increasing population requires the government to reduce the population explosion, one of which is the family planning programme. In the family planning programme there are often mistakes in choosing contraceptives. The use of contraception is important, considering that it can reduce the growth rate. The use of the right type of birth control for women is a classification problem. In this study, data were taken on the classification of contraceptive use at the private midwife HJ Enok Hayati who serves several types of family planning, such as injectable birth control, birth control pills, IUD (Intrauterine Device) and Implant. This study uses the data mining classification method C4.5 and Naïve Bayes algorithms and then compares the two methods. Processing of the two methods using confusion matrix and kappa. The results of this study resulted in the accuracy of the C4.5 algorithm of 89.55% and kappa 0.851 while the Naïve Bayes accuracy rate was 34.85% and kappa 0.016. So that the C4.5 algorithm is a better method in classifying family planning acceptor data at Private Midwife HJ Enok Hatiyah compared to the Naïve Bayes algorithm method"

Keyword: *Data Mining, C4.5, Naive Bayes, Contraceptives, RapidMiner.*

PENDAHULUAN

Kepadatan penduduk terjadi karena laju pertumbuhan penduduk setiap tahunnya sangat tinggi. Situasi ini menekan negara untuk menciptakan lapangan kerja yang sangat besar untuk mengatasi kemiskinan, pengangguran dan kejahatan. Berbagai upaya telah dilakukan pemerintah untuk menekan pertumbuhan penduduk, diantaranya melalui pelaksanaan Program Keluarga Berencana (KB). Program keluarga berencana harus mampu mengendalikan pertumbuhan penduduk. Banyak orang yang menjalani atau menggunakan KB membuat keputusan yang salah tentang metode kontrasepsi yang akan digunakan karena kesalahpahaman tentang manfaat dan fungsi kontrasepsi. Untuk membantu penerima dalam memilih metode kontrasepsi, data diolah sesuai dengan metode klasifikasi.

"Klasifikasi" memainkan peran penting dalam "teknik penambangan data". Klasifikasi diperkenalkan ke dalam pembelajaran terbimbing karena proses klasifikasi memiliki proses pra-pembelajaran. Proses ini digunakan oleh algoritma yang menganalisis data untuk pola dan kemudian menerapkannya ke grup baru yang tidak diketahui (Bala, 2012). Kajiannya berjudul "Perbandingan Algoritma Data Mining Naive Bayes dan C4.5 Untuk Klasifikasi Penerimaan Mahasiswa Baru SMPN 35 Semarang". Dari sini dapat disimpulkan bahwa "algoritma yang cocok untuk dataset PPBD SMPN 35 Semarang adalah Naive Bayes". Septiani (2017) mempresentasikan penelitian berjudul "Perbandingan Taksonomi C4.5 dan Algoritma Data Mining Naive Bayes untuk Prediksi Hepatitis". Pada penelitian ini, "algoritma terbaik

menurut uji akurasi adalah Naive Bayes. C4.5 adalah algoritma terbaik menurut pengujian berdasarkan kurva ROC (AUC)". Oleh karena itu, dalam penelitian ini kami melakukan "perbandingan kategori menggunakan algoritma C4.5 dan algoritma Naive Bayes untuk membandingkan kedua metode kontrasepsi tersebut"

METODE PENELITIAN

Pada penelitian ini data yang digunakan adalah "data akseptor KB" yang didapat dari Bidan Swasta HJ Enok Hatiyah yang berlokasi di Kampung Cebong Rt 05 Rw 03 Desa Mekarjaya Kec Kedungwaringin Kab Bekasi. Dalam penelitian ini akan dilakukan beberapa langkah-langkah atau tahapan penelitian seperti di bawah ini:

1. Pengumpulan Data (Gathering)

Pada tahap ini ditentukan data apa yang akan diolah yaitu data pasien bidan KB HJ Enok Hatiyah. Gabungkan semua data menjadi satu kumpulan data, termasuk pemisahan variabel yang diperlukan dalam proses".

2. Pengolahan Data Awal (Data Pre-processing)

"Data dipilih dalam langkah ini, Jumlah data akuisisi awal adalah 115 data. Setelah dilakukan pengolahan atau pembersihan data, sebanyak 115 data yang terdapat pada rekam medis dapat digunakan pada Bidan KB HJ Enok Hatiyah. Pengolahan data tahap pertama menyiapkan data yang benar-benar valid sebelum melanjutkan ke tahap berikutnya"

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, model diterapkan tergantung pada tipe data. Pada penelitian ini, peneliti menggunakan metode klasifikasi Naïve Bayes"

4. Pengujian Model dan Eksperimen (Model Testing and Experimen)

Pada tahap ini, model yang diusulkan adalah algoritma C4.5 dan metode klasifikasi data mining Naive Bayes. Pengujian model menggunakan RapidMiner untuk mencari hasil berupa nilai presisi dan 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", "Implan", dan "IUD" serta nilai Entropy dari semua kasus. Dari data yang ada diketahui jumlah kasus yang "Suntik" sebanyak 42 record, kasus "Pil" sebanyak 37 record, kasus "IUD" sebanyak 19 record, dan kasus "Implan" sebanyak 17 record total keseluruhan kasus adalah 115 record. Sehingga didapat Entropy keseluruhan:

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

$$\begin{aligned} \text{Entropy}(S) &= \left(-\left(\frac{42}{115}\right) * \log_2 \left(\frac{42}{115}\right)\right) + \left(-\left(\frac{37}{115}\right) * \log_2 \left(\frac{37}{115}\right)\right) + \left(-\left(\frac{19}{115}\right) * \log_2 \left(\frac{19}{115}\right)\right) + \left(-\left(\frac{17}{115}\right) * \log_2 \left(\frac{17}{115}\right)\right) \\ &= 0,9481 \end{aligned}$$

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

$$\text{Entropy} \leq 30 (32,33,9,13) = \left(-\left(\frac{32}{87}\right) * \log_2 \left(\frac{32}{87}\right)\right) + \left(-\left(\frac{33}{87}\right) * \log_2 \left(\frac{33}{87}\right)\right) + \left(-\left(\frac{9}{87}\right) * \log_2 \left(\frac{9}{87}\right)\right) + \left(-\left(\frac{13}{87}\right) * \log_2 \left(\frac{13}{87}\right)\right) = 1,0368$$

$$\text{Entropy} > 30 (10,4,10,4) = \left(-\left(\frac{10}{28}\right) * \log_2 \left(\frac{10}{28}\right)\right) + \left(-\left(\frac{4}{28}\right) * \log_2 \left(\frac{4}{28}\right)\right) + \left(-\left(\frac{10}{28}\right) * \log_2 \left(\frac{10}{28}\right)\right) + \left(-\left(\frac{4}{28}\right) * \log_2 \left(\frac{4}{28}\right)\right) = 0,9826$$

Kemudian menghitung nilai gain atribut umur akseptor sebagai berikut:

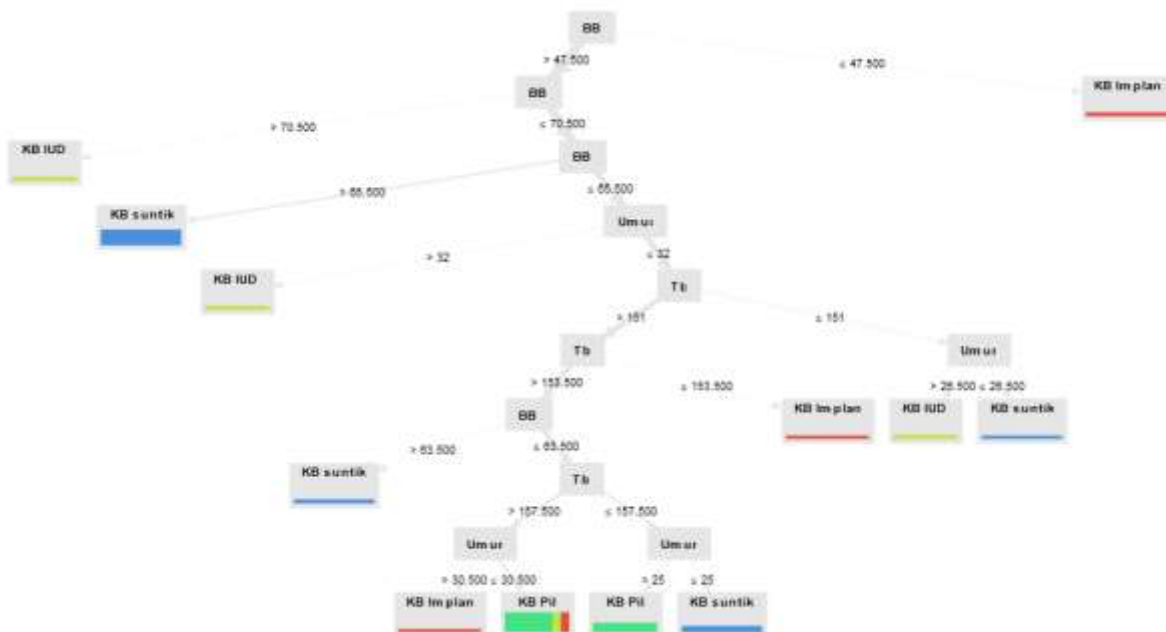
$$\text{Gain}(S, A) = \text{Entropi}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropi}(S_i)$$

$$\text{Gain}(S, A) = (0,9481) - \left(\left(\frac{87}{115}\right) * 1,0368\right) - \left(\left(\frac{28}{115}\right) * 0,9826\right) = 0,0957$$

Tabel 1.1 Nilai Entropy dan Gain pada Algoritma C4.5

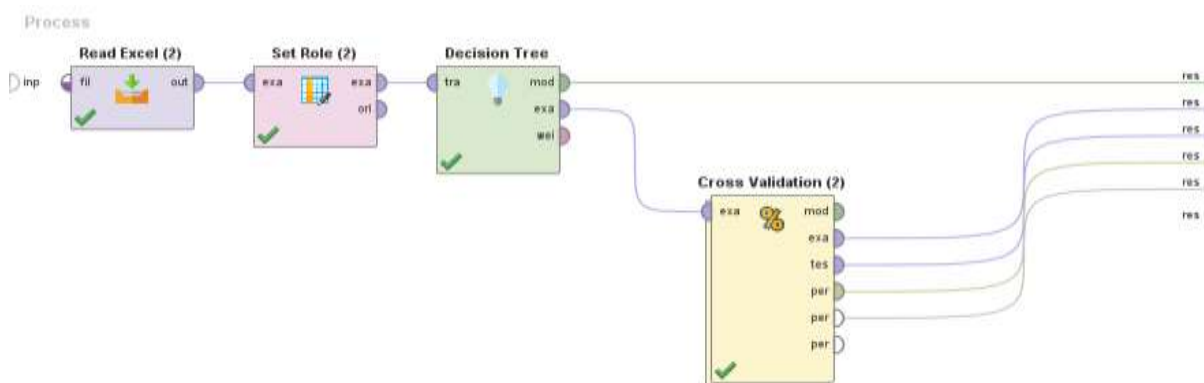
Simpul	Kasus	Suntik	Pil	Impla n	IU D	Entrop y	Gain	
Jumlah Kasus	115	42	37	17	19	0,9481		
Umur Istri	≤30	87	32	33	9	13	1,0368	0,0957
	≥30	28	10	4	10	4	0,9826	
Status Kerja Istri	Kerja	50	21	14	11	4	0,7393	0,0898
	Tidak Bekerja	65	21	23	8	13	0,9986	

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 Umur Istri, maka Umur Istri dipilih sebagai node atau root 1. Pohon keputusan yang terbentuk dari dataset akseptor KB Bidan Swasta HJ Enok Hatiyah dengan menggunakan perhitungan nilai gain dan RapidMiner.

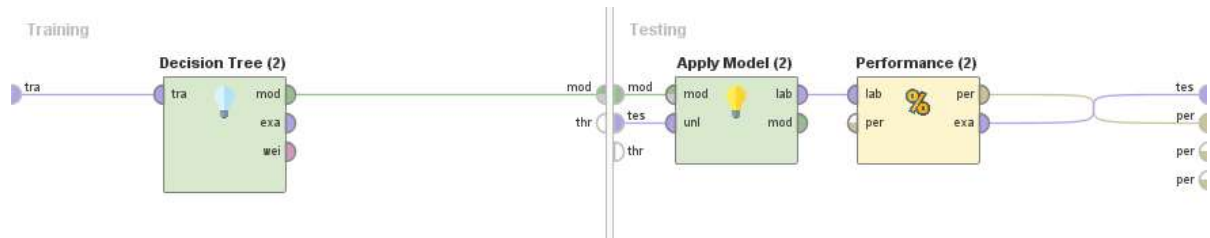


Gambar 1.1 Pohon Keputusan

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 1.2 Proses Pengujian Cross Validation Metode C4.5



Gambar 1.3 Proses Cross Validation Metode C4.5

Hasil Evaluasi Algoritma Naive Bayes

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

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

$$P(\text{Jenis Pelayanan KB}|\text{Suntik}) = 42/115 = 0,3652173$$

$$P(\text{Jenis Pelayanan KB}|\text{Pil}) = 37/115 = 0,3217391$$

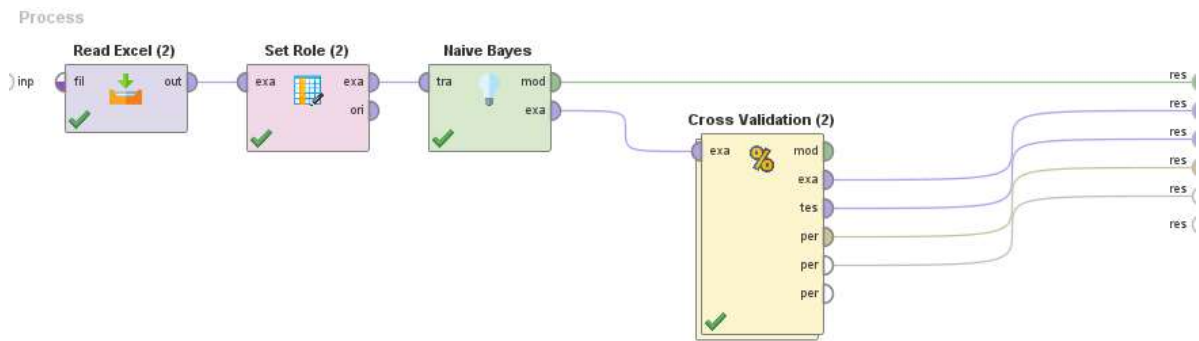
$$P(\text{Jenis Pelayanan KB}|\text{IUD}) = 19/115 = 0,1652173$$

$$P(\text{Jenis Pelayanan KB}|\text{Implan}) = 17/115 = 0,1478260$$

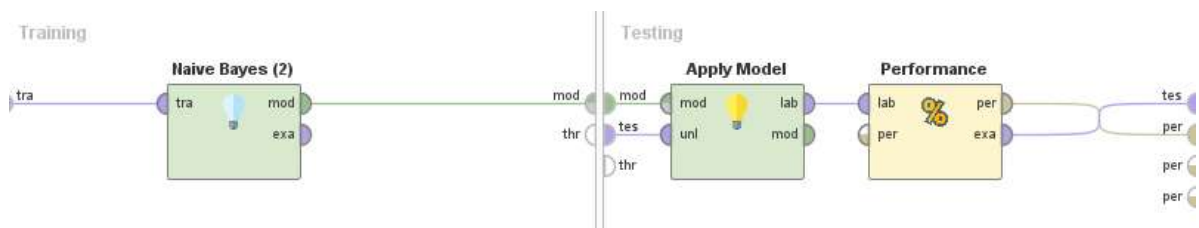
Tabel 1.2 Perhitungan Probabilitas Jenis Pelayanan KB

PROBABILITAS JENIS PELAYANAN KB	
JENIS PELAYANAN KB	NILAI
SUNTIK	0,3652173
PIL	0,3217391
IUD	0,1652173
IMPLAN	0,147826
Total	1

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 1.4 Pengujian Cross Validation Naive Bayes



Gambar 1.5 Proses Cross Validation Naive Bayes

Evaluasi Dan Validasi Hasil

Hasil proses menggunakan rapidminer diperoleh nilai accuracy sebesar 89.55% dari 115 sebanyak 7 data diprediksi pengguna pil tetapi pengguna implan , 4 data diprediksi pengguna pil tetapi pengguna IUD, 10 data diprediksi sesuai pengguna Implan, 1 data diprediksi pengguna pil tetapi pengguna suntik, 15 data diprediksi sesuai pengguna IUD, 37 data diprediksi sesuai pengguna pil dan 41 data diprediksi sesuai pengguna suntik. Hasil accuracy dapat dilihat pada gambar 1.6 berikut ini:

accuracy: 89.55% +/- 9.02% (micro average: 89.57%)

	true KB suntik	true KB Pil	true KB IUD	true KB Implan	class precision
pred. KB suntik	41	0	0	0	100.00%
pred. KB Pil	1	37	4	7	75.51%
pred. KB IUD	0	0	15	0	100.00%
pred. KB Implan	0	0	0	10	100.00%
class recall	97.62%	100.00%	78.95%	58.82%	

Gambar 1.6 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 89.55% 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:

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

$$\text{Accuracy} = \frac{41+37+15+10}{41+0+0+0+1+37+4+7+0+0+15+0+0+0+0+10}$$

$$\text{Accuracy} = 0.8956 * 100 \%$$

$$\text{Accuracy} = 89.56 \%$$

Untuk menghitung nilai precision adalah sebagai berikut:

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

$$\text{True Positive Rate (Suntik)} = \frac{41}{41+0+0+0}$$

$$= 1 \%$$

Untuk menghitung nilai recall adalah sebagai berikut :

$$\text{Positive Predictive Value (Suntik)} = \frac{TN}{TN+FP}$$

$$\text{Positive Predictive Value (Suntik)} = \frac{41}{41+1+0+0}$$

$$= 0.9761$$

$$= 97.61 \%$$

Hasil proses menggunakan rapidminer diperoleh nilai accuracy sebesar 34.85% dari 115 sebanyak 10 data diprediksi pengguna Pil tetapi pengguna Implan , 7 data diprediksi pengguna Suntik tetapi pengguna implan, 3 data diprediksi pengguna implan tetapi pengguna IUD ,3 data diprediksi sesuai pengguna IUD, 1 data diprediksi pengguna pil tetapi pengguna IUD, 12 data diprediksi pengguna Suntik tetapi pengguna IUD, 10 data diprediksi pengguna Implan tetapi pengguna pil, 3 data diprediksi pengguna IUD tetapi pengguna pil, 19 data diprediksi sesuai pengguna Pil, 5 data diprediksi pengguna suntik tetapi pengguna pil, 7 data diprediksi pengguna Implan tetapi pengguna suntik, 7 data diprediksi pengguna IUD tetapi pengguna suntik, 10 data diprediksi pengguna Pil tetapi pengguna suntik dan 18 data diprediksi sesuai pengguna suntik. Hasil accuracy dapat dilihat pada gambar 1.7 berikut ini

accuracy: 34.85% +/- 11.19% (micro average: 34.78%)

	true KB suntik	true KB Pil	true KB IUD	true KB Implan	class precision
pred. KB suntik	18	5	12	7	42.86%
pred. KB Pil	10	19	1	10	47.50%
pred. KB IUD	7	3	3	0	23.08%
pred. KB Implan	7	10	3	0	0.00%
class recall	42.86%	51.35%	15.79%	0.00%	

Gambar 1.7 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:

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

$$\text{Accuracy} = \frac{18+19+3+0}{18+5+12+7+10+19+1+10+7+3+3+0+7+10+3+0}$$

$$\text{Accuracy} = 0.3478 * 100 \%$$

$$\text{Accuracy} = 34.78 \%$$

Untuk menghitung nilai precision adalah sebagai berikut:

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

$$\text{True Positive Rate (Suntik)} = \frac{18}{18+5+12+7}$$

$$= 0.4285 * 100 \%$$

$$= 42.85 \%$$

Untuk menghitung nilai recall adalah sebagai berikut :

$$\text{Positive Predictive Value (Suntik)} = \frac{TN}{TN+FP}$$

$$\text{Positive Predictive Value (Suntik)} = \frac{18}{18+10+7+7}$$

$$= 0.4285$$

$$= 42.85 \%$$

PEMBAHASAN

Berdasarkan hasil pengujian dataset Penggunaan jenis KB pada Bidan Swasta HJ Enok Hatiyah dengan menggunakan algoritma C4.5 dan Naïve Bayes, maka hasil yang diperoleh dirangkum pada Tabel 1.3

Tabel 1.3 EVALUASI DAN VALIDASI PENGUJIAN DATASET

	C4.5	Naïve Bayes
Accuracy	89.55 %	34.85 %
Kappa	0.850	0.016
Kategori Kappa	Sangat Kuat(Very good)	Rendah (poor)
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 pada Bidan Swasta HJ Enok Hatiyah yaitu C4.5, dengan nilai accuracy 89.55% dan Kappa 0.850 yang terbentuk sangat kuat(Very good).

SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dengan komparasi menggunakan algoritma C4.5 dan Naïve Bayes untuk pemilihan alat kontrasepsi pada Bidan Swasta HJ Enok Hatiyah dapat disimpulkan bahwa:

1. Pemilihan penggunaan alat kontrasepsi dengan komparasi metode klasifikasi menggunakan algoritma C4.5 menghasilkan nilai accuracy sebesar 89.55% dan nilai Kappa 0.850 yang termasuk kategori sangat kuat (very good). dan algoritma Naïve Bayes menghasilkan nilai accuracy sebesar 34.85% dan nilai Kappa 0.016 yang termasuk kategori rendah (poor)
2. Hasil pengujian komparasi C4.5 dan Naïve Bayes dapat dilihat bahwa algoritma C4.5 merupakan algoritma yang lebih baik dengan nilai accuracy dan Kappa yang tinggi. Maka algoritma C4.5 dapat digunakan sebagai acuan untuk klasifikasi penggunaan suntik, pil, implan dan IUD untuk pemilihan penggunaan alat kontrasepsi pada akseptor KB Bidan Swasta HJ Enok Hatiyah.

DAFTAR PUSTAKA

- Fauzia, B. (2021). Komparasi algoritma data mining naïve bayes dan C4.5 untuk klasifikasi penerimaan peserta didik baru di SMPN 35 Semarang. Semarang: Unaki.
- Han, J. And Kamber, M, 2006. "Data Mining Concept and Techniques Second Edition". Morgan Kauffman, San Fransisco
- Han, J. (2007). Data Mining Concept And Technique. Champaign: Multiscience Press.
- 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.
- 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.
- 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).
- 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 Jatisi*.
- 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.
- Kusrini, & Luthfi, E. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- Gorunescu, F. (2011). *Data Mining : Concepts, Model and Techniques*. Berlin, Jerman: Springer.
- Sodik, A. (2015). *Dasar metodologi Penelitian*. Yogyakarta: Literasi Media Publishing.