Jurnal Informatika Universitas Pamulang **ISSN: 2541-1004**
Penerbit: Program Studi Teknik Informatika Universitas Pamulang **e-ISSN: 2622-4615**
Vol. 6, No. 3, September 2021 (456-462) **10.32493/informatika.v6i3.9160**

Perbandingan Metode Klasifikasi C4.5 dan Naïve Bayes untuk Mengukur Kepuasan Pelanggan

Devi Yunita[1], Ines Heidiani Ikasari[2]

[1,2]Teknik Informatika, Fakultas Teknik, Universitas Pamulang, Tangerang Selatan, Indonesia
e-mail: [1] dosen00846@unpam.ac.id; [2] dosen01374@unpam.ac.id

Submitted Date: January 17th, 2021 Reviewed Date: June 02nd, 2021
Revised Date: August 24th, 2021 Accepted Date: October 12th, 2021

Abstract

Data mining is a process of collecting meaningful data from some large information contained in information bases, information warehouses or other documentation places. Classification is an educated educational process (supervised learning). The universal classification procedures used include: Decision tree, K-Nearest Neighbor, Naïve Bayes, Neural Network, C4. 5 as well as SVM. Consideration of the classification method is tried to ensure the type of classification that shares the highest accuracy value of an object. The information used in this research is information from a questionnaire about customer satisfaction with the services of PT. Media Semesta Solutions, the factors used in this research are Reliability, Responsiveness, Assurance, Empathy, and Tangibility. Based on the results of the comparative analysis, the information mining classification procedure is C4.5 and Naïve Bayes proves that the C4 method. 5 is more accurate than the Naïve Bayes method, this result is seen from the accuracy value where the C4 procedure. 5 has an accuracy value of 94, 17%, greater than Naïve Bayes with an accuracy value of 85.83%.

Keywords*: Data mining; Classification; Accuracy; C4.5; Naïve Bayes*

Abstrak

Data mining merupakan proses pengumpulan data berarti dari beberapa informasi besar yang terkandung di basis informasi, gudang informasi ataupun tempat dokumentasi lainnnya. Klasifikasi merupakan proses pendidikan secara terdidik (*supervised learning*). Tata cara klasifikasi yang universal digunakan antara lain: *Decision tree, K- Nearest Neighbor, Naïve Bayes, Neural Network*, C4. 5 serta SVM. Pertimbangan metode klasifikasi dicoba guna memastikan tipe klasifikasi yang membagikan nilai akurasi paling tinggi dari suatu objek. Informasi yang digunakan dalam riset ini merupakan informasi hasil angket tentang kepuasan pelanggan terhadap pelayanan PT. Solusi Media Semesta, faktor yang dipakai dalam penelitian ada yaitu *Reliability, Responsiveness, Assurance, Empathy,* dan *Tangibility*. Bersumber pada hasil Analisa perbandingan, tata cara klasifikasi Informasi mining ialah C4.5 serta Naïve Bayes membuktikan kalau tata cara C4. 5 lebih akurat dari pada tata cara Naïve Bayes, hasil ini dilihat dari nilai accuracy dimana tata cara C4. 5 mempunyai nilai accuracy sebesar 94, 17%, lebih besar dibandingkan Naïve Bayes dengan nilai accuracy 85, 83%.

Kata kunci: Data mining; Klasifikasi; Akurasi; C4.5; Naïve Bayes

1. Pendahuluan

Evolusi data tidak lepas dari perkembangan teknologi informasi saat ini yang mampu mengumpulkan sejumlah data yang sangat banyak. Selain kebutuhan *data mining* yang terus meningkat, berbagai algoritma klasifikasi telah muncul agar data dalam jumlah yang sangat banyak itu dapat diproses.

Klasifikasi ialah proses menciptakan suatu model ataupun guna yang menarangkan serta menandai kelas informasi. Klasifikasi ialah wujud analisis informasi model ekstraksi yang mengelompokkan mengelompokkan kelas informasi. Klasifikasi jadi berarti dalam memastikan kelas layanan, sebab kelas layanan pengaruhi mutu layanan. Hasil dari proses klasifikasi pada kelas layanan dijadikan sumber

informasi untuk kebutuhan bisnis serta kebijakan nasional bidang teknologi data serta komunikasi. (Adriyendi & Melia, 2020).

Perbandingan algoritma merupakan perbandingan 2 ataupun lebih algoritma buat mengenali algoritma mana yang terbaik dari algoritma tersebut (Safri et al., 2018). Perbandingan algoritma klasifikasi ini untuk menentukan klasifikasi mana yang menghasilkan nilai akurasi tertinggi berdasarkan data kepuasan pelanggan PT. Solusi Media Semesta. Dipilihnya kepuasan pelanggan sebagai data dalam penelitian ini, dikarenakan kepuasan pelanggan merupakan dasar tujuan dari suatu perusahaan, karena bila tingkat kepuasan pelanggan terhadap layanan semakin tinggi, maka kualitas perusahaan semakin baik pula. Kepuasan konsumen ialah tingkatan dimana asumsi terhadap produk cocok dengan harapan para konsumen. Harapan konsumen biasanya ialah prakiraan ataupun kepercayaan konsumen tentang apa yang hendak diterimanya apabila sudah membeli ataupun komsumsi sesuatu produk (Shiddiq et al., 2018).

Adapun dipilihnya Metode Naïve Bayes dan metode C4.5 sebagai algoritma pengklasifikasi kepuasan pelanggan karena tingkat akurasi dari kedua algoritma ini relatif tinggi. Penelitian sebelumnya tentang penerapan metode klasifikasi dalam data mining telah dilakukan oleh (Sugianto, 2017) Dalam penelitian tersebut metode C4.5 digunakan untuk menentukan hasil seleksi sekolah menengah atas dam hasil penelitian tersebut, penggunaan C4.5 berhasil dengan tepat digunakan dalam pengambilan keputusan untuk pihak sekolah dalam seleksi masuk sekolah menengah atas. Penelitian lain oleh (Rifqo & Wijaya, 2017) yang menggunakan metode naïve bayes untuk penentuan pemberian kredit. Penggunaan metode naïve bayes dapat digunakan dengan baik, dimana hasil implementasi pemberian kredit kepada nasabah dapat mengurangi terjadinya kredit macet.

2. Metodologi Penelitian

Informasi primer yang dipakai didalam studi ini diperoleh dari angket yang disebarkan pada para pelanggan PT. Solusi Media Semesta. Penelitian ini berhubungan dengan masalah yang terkait dengan layanan pelanggan.

Setelah data diperoleh, lalu tentukan variabel-variabelnya yang akan dipakai. Variabel respon diambil dari kepuasan pelanggan atau ketidakpuasan pelanggan dengan pelayanan PT. Solusi Media Semesta. Sedangkan variabel bebasnya, yaitu *Reliability, Responsiveness, Assurance, Empathy*, dan *Tangibility*(Shiddiq et al., 2018)

Tahapan-tahapan penelitian yang dilakukan antara lain:

1. Pra-pengolahan data pelanggan.
2. Hitung manual data tersebut untuk menentukan jumlah pelanggan yang puas dan tidak puas pada.
3. Hasil perhitungannya diolah ke dalam perhitungan dengan metode algoritma Naïve Bayes dan Metode algoritma C4.5 mengupayakan akurasi analisis data dengan data uji.
4. Peneliti menggunakan aplikasi RapidMiner untuk menguji akurasi dari metode Naïve Bayes dan metode C4.5.
5. Hasil perbandingan algoritmanya dapat digunakan untuk mengetahui akurasi terbaik terhadap kepuasan pelanggan atas pelayanan PT. Solusi Media Semesta.

Penelitian ini menerapkan algoritma dengan 2 jenis klasifikasi, yaitu klasifikasi C4.5 dan kalsifikasi Naïve Bayes. Langkah-langkah penerapan analisis metode klasifikasi dalam penelitian ini adalah(Mardi, 2017):

a. Algoritma C4.5
    1. Hitung jumlah respon yang puas, jumlah respon yang tidak puas, serta *entropy* dari keseluruhan.
    2. Hitung *gain* untuk setiap atribut.
    3. Tentukan *node root* untuk mendapatkan nilai *gain* tertinggi.
    4. Lalu hitung jumlah kasus, jumlah respon yang puas, jumlah respon tidak puas, dan *entropy* atas seluruh permasalahan dan permasalahan yang dibagi dengan tanda selain *node root*, lalu gunakan *node root* rendah nilai atributnya, sehingga seluruh atribut dapat menjadi *node* dan membentuk pohon keputusan.

b. Algoritma Naïve Bayes(Wijaya, 2017)
    1. Data latih diinput.
    2. Hitung jumlah data, probabilitas, dan bila datanya numerik:
        - Cari nilai *mean* juga standar deviasi variabelnya.
        - Cari data nilai probabilistik. Caranya hitung data-data yang benar

berdasarkan kelompok yang sejenis, lalu pisahkan dengan total data dalam kelompok tersebut.

- tabulasi *mean*, tabulasi standar deviasi, dan tabulasi probabilitas untuk memperoleh nilainya yang diharapkan.

3. Hasil dan Pembahasan

Informasi primer pada penelitian yang dilakukan merupakan hasil dari angket yang dibagikan kepada pelanggan PT. Solusi Media Semesta. Total data yang didapat yakni berjumlah 1000 responden.

Setelah informasi diperoleh, lalu tentukan variabel-variabelnya yang akan dipakai. Variabel independen meliputi *Reliability*, *Responsiveness, Assurance, Empathy* dan *Tangibility* yang masing-masing memuat indikator-indikator pendukung dari variabel tersebut. Sedangkan variabel responnya adalah kepuasan atau ketidakpuasan terhadap pelayanan PT. Solusi Media Semesta.

Langkah pertama dalam analisis penelitian adalah pra-pengolahan. Hasil kuesioner menunjukkan bahwa setiap variabel menagndung sejumlah sub-faktor, sehingga sub-faktor tersebut ditambahkan menjadi faktor tunggal utama. Selain itu, diterapkan juga pra-pengolahan informasi, Berbagai teknik pra-pengolahan data, antara lain:

a. Data *validation*, bertujuan identifikasi dan hapus data eksternal (*outlier*/ *noise*), tidak stabil, serta nilai yang tidak kompit atau *missing value*.
b. Data *transformation and integration*, bertujuan menaikkan tingkat keefisiensian dan ketepatan metode-metode dalam penelitian ini. Untuk meningkatkan akurasi, data diproses menjadi angka dengan aplikasi RapidMiner.
c. *Dicretization and size reduction*, bertujuan mendapatkan kumpulan data yang lebih sedikit atributnya dan *record*, namun ini informatif. RapidMiner digunakan untuk memilih atribut dan menghapus duplikasi data pada data latih yang digunakan.

Pra-pengolahan data diperoleh hingga 1000 *record*, kemudian dikurangi menjadi 120 *record* untuk data latih dengan sedikit duplikasi.

Tabel 1**.** Contoh Hasil Kuesioner setelah pra-pengolahan

REABILITY			RESPONSIVE			ASSURANCE			EMPATY			TANGIABLE			RESPON
A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3	E1	E2	E3	
5	5	4	5	5	5	4	3	4	4	5	5	5	3	5	PUAS
5	5	4	5	5	4	5	3	3	5	5	4	5	4	4	PUAS
5	5	5	5	4	4	4	4	3	4	4	4	5	4	4	PUAS
5	5	4	3	5	4	5	4	3	4	3	4	5	4	3	PUAS
5	5	3	5	5	4	3	3	5	4	5	4	4	5	3	PUAS
4	5	1	3	4	4	3	4	5	4	4	4	3	5	3	TIDAK PUAS
4	5	4	4	4	5	4	4	3	4	4	4	4	5	5	PUAS
4	4	4	4	5	5	4	4	4	3	4	5	4	4	4	PUAS
4	5	4	3	5	3	3	4	4	3	4	4	5	3	3	PUAS

Setelah pra-pengolahan data, setiap data masing-masing variabel dirata-ratakan untuk menyederhanakan hasil dari setiap variabel. Tabel 2 menunjukkan hasil rata-rata perhitungan.

Tabel 2**.** Hasil perhitungan rata-rata

Reability	Responsiveness	Assurance	Empaty	Tangiabel	Respon
4,67	5,00	3,67	4,67	4,33	PUAS
4,67	4,67	3,67	4,67	4,33	PUAS
5,00	4,33	3,67	4,00	4,33	PUAS
4,67	4,00	4,00	3,67	4,00	PUAS
4,33	4,67	3,67	4,33	4,00	PUAS
3,33	3,67	4,00	4,00	3,67	TIDAK PUAS
4,33	4,33	3,67	4,00	4,67	PUAS
4,00	4,67	4,00	4,00	4,00	PUAS
4,33	3,67	3,67	3,67	3,67	PUAS
4,00	3,67	3,00	3,67	4,33	TIDAK PUAS
4,00	5,00	3,67	3,00	3,67	TIDAK PUAS
4,00	5,00	4,67	3,00	4,00	TIDAK PUAS
4,00	4,33	3,67	4,33	4,00	PUAS

Tahap selanjutnya membuat *coding* untuk data latih. Ini bermaksud memberikan kode pada data dari kuesioner yang memiliki kelompok kategori sama lalu diberi simbol(Takalapeta, 2018). Tiap variabel dibagi menjadi dua kelas, kelas dengan nilai antara 0 hingga 3,5 (Rendah) dan 3,6 hingga 5 (Tinggi).

Tabel 3. Data Latih setelah penerapan *coding*

Reability	Responsiveness	Assurance	Empaty	Tangiabel	Respon
Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	PUAS
Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	PUAS
Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	PUAS
Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	PUAS
Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	PUAS
Rendah	Tinggi	Tinggi	Tinggi	Tinggi	TIDAK PUAS
Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	PUAS
Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	PUAS
Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	PUAS
Tinggi	Tinggi	Rendah	Tinggi	Tinggi	TIDAK PUAS
Tinggi	Tinggi	Tinggi	Rendah	Tinggi	TIDAK PUAS
Tinggi	Tinggi	Tinggi	Rendah	Tinggi	TIDAK PUAS

Algoritma C4.5

Metode C4.5 dapat disebut juga dengan metode Pohon keputusan. Pertama-pertama hitung nilai *entropy*. Diketahui dari data latih terdapat 120 kasus, yang meliputi 76 *record*

kelas puas dan 44 *record* kelas tidak puas, sehingga diperoleh *entropy:*

$$Entropy(S) = \sum_{i=1}^{n} -\mathbf{pi.log2pi} \ldots\ldots\ldots\ldots (1)$$

= (-44/120 * log2 (44/120)) + (-76/120 * log2 (76/120))

= 0,948

Lalu nilai *gain* dihitung untuk setiap atribut, kemudian tentukan *gain value* yang paling maksimal. Data yang memiliki nilai tertinggi atau maksimal ini menjadi digunakan sebagai akar pohon. Misal *gain value* variabel *Reliability*, jumlah kasus dengan nilai rendah sebanyak 6 *record*, terdapat respon tidak puas ada 0 dan respon puas ada 6. Untuk jumlah kasus nilai tinggi sebanyak 114 *records*, respon tidak puas ada 38 dan respon puas ada 76.

Nilai *Entropy* rendah dari *Reliability* adalah:

$$Entropy(S) = \sum_{i=1}^{n} -pi.log2pi \ldots\ldots\ldots\ldots (2)$$

$$= \left(-\frac{0}{6} * log2\ \left(\frac{0}{6}\right)\right) + \left(-\frac{6}{6} * log2\ \left(\frac{6}{6}\right)\right) = 0$$

Nilai *Entropy* tinggi dari *Reliability* adalah:

$$Entropy(S) = \sum_{i=1}^{n} -pi.log2pi \qquad (3)$$

= (-38/114 * log2 (38/114)) + (-76/114* log2 (76/114))

= 0,918

Nilai *gain* dari variabel *Reliability* adalah:

$$Gain\ (S, A)\ =\ Entropy\ (S) - \sum_{i=}^{n} \frac{|Si|}{|S|} * \text{Entropy (Si)} \quad (4)$$

Gain (S, A) = 0,948 - (((6/120) * 0,00) + ((114/120) * 0,918))

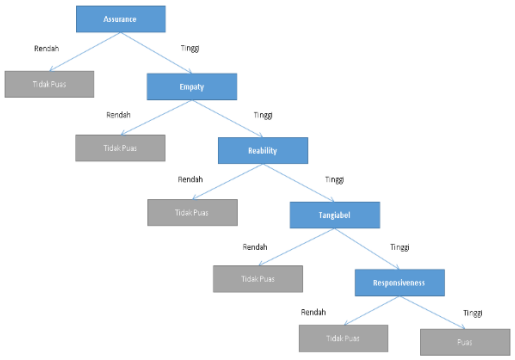
$$Gain\ (S, A) = 0{,}075697$$

Setelah itu, jumlahkan faktor lainnya menggunakan langkah yang sama. Perhitungan *gain* dan *entropy* dilakukan untuk seluruh karakter, sehingga didapatkan *gain value* paling tinggi. Tabel 4 menampilkan hasil perhitungan semua variabel.

Tabel 4. Nilai *Entropy* dan *gain* untuk menentukan *node root*

Node				Jumlah Kasus	Tidak Puas	Puas	Entropy	Gain
1	Total			120	44	76	0,948	
		Reability						0,075697
			Rendah	6	0	6	0,000	
			Tinggi	114	38	76	0,918	
		Responsiveness						0,035099
			Rendah	6	5	1	0,650	
			Tinggi	114	39	75	0,927	
		Assurance						0,235702
			Rendah	17	17	0	0,000	
			Tinggi	103	27	76	0,830	
		Empaty						0,251966
			Rendah	18	18	0	0,000	
			Tinggi	102	26	76	0,819	
		Tangiabel						0,070314
			Rendah	14	11	3	0,750	
			Tinggi	106	33	73	0,895	

Gambar 1 menampilkan pohon keputusan akhir yang dibangun.



Gambar 1 Pohon Keputusan akhir yang dibangun

Algoritma Naïve Bayes

Perhitungan tata cara Naïve Bayes memakai informasi latih. Dari informasi latih dikenal jumlah permasalahan terdapat 120, jumlah informasi puas terdapat 76 serta jumlah informasi tidak puas terdapat 44. Buat menghitung P (X= Puas) serta P (X= Tidak Puas) adalah

P (X = Puas |Ci) = 76 / 120 = 0,6333

P (X = Tidak Puas | Ci) = 44 / 120 = 0,3667

Jumlah puas dipecah total hingga hasilnya merupakan 0, 5067, serta buat nilai P (X= Tidak Puas) merupakan jumlah tidak puas dipecah total serta hasilnya merupakan 0, 2933. Berikutnya merupakan perhitungan nilai

Probabilitas prior, ialah nilai probabilitas kepuasan serta ketidakpuasan tiap- tiap variabel dibanding dengan total puas serta tidak puas dari seluruh informasi.

Contoh perhitungan Variabel *Reliability:*

P (*Reliability* = Rendah | Puas) = 0 / 76 = 0

P (*Reliability* = Rendah | Tidak Puas) = 9 / 44 = 0,2045

P (*Reliability* = Tinggi | Puas) = 79 / 76 = 1,0395

P (*Reliability* = Tinggi | Tidak Puas) = 35 / 44 = 0,7955

Setelah semua variabel perhitungan Probabilitas *Prior*, maka hasilnya akan seperti pada Tabel 5.

Tabel 5. Probabilitas *Prior*

	Variabel		Puas	Tidak Puas	P(X|Ci)	
					Puas	Tidak Puas
Total	120		76	44	0,6333	0,3667
	Reability					
		Rendah	0	9	0,0000	0,2045
		Tinggi	79	35	1,0395	0,7955
	Responsiveness					
		Rendah	1	5	0,0132	0,1136
		Tinggi	78	39	1,0263	0,8864
	Assurance					
		Rendah	0	33	0,0000	0,7500
		Tinggi	76	11	1,0000	0,2500
	Empaty					
		Rendah	0	29	0,0000	0,6591
		Tinggi	76	15	1,0000	0,3409
	Tangiabel					
		Rendah	3	24	0,0395	0,5455
		Tinggi	73	20	0,9605	0,4545

Buat memastikan permasalahan baru, hitung probabilitas posterior dari probabilitas prior yang sudah dihitung (ditampilkan di Tabel 5). Perhitungan probabilitas posterior buat memastikan informasi uji tercantum ke dalam klasifikasi yang mana ada pada Tabel 6. Misalkan informasi uji X semacam yang ditunjukkan pada tabel berikut diambil.

Tabel 6. Data Uji

REABILITY			RESPONSIVENESS			ASSURANCE			EMPATY			TANGIABLE			RESPON
A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3	E1	E2	E3	
4	5	4	3	4	4	4	5	5	4	5	4	5	5	5	PUAS

Nilai masing-masing variabel dijumlahkan, maka hasilnya akan seperti table berikut:

Tabel 7. Penjumlahan Variabel Data Uji

REABILITY	RESPONSIVENESS	ASSURANCE	EMPATY	TANGIABLE	RESPON
13	11	14	13	15	PUAS

Dengan nilai seperti pada Tabel, dilakukan perhitungan Probabilitas *posterior* dengan acuan nilai pada Probabilitas *Prior* sepert pada table berikut:

Tabel 8. Perhitungan untuk Menentukan Klasifikasi Data uji X

Data X			Puas	Tidak Puas
Reability	13	Tinggi	1,0395	0,7955
Responsiveness	11	Rendah	0,0132	0,1136
Assurance	14	Tinggi	1,0000	0,2500
Empaty	13	Tinggi	1,0000	0,3409
Tangiabel	15	Tinggi	0,9605	0,4545

P (X|Ci) = P (X| *remark* = puas) (5)

$$1.0395 * 0.0132 * 1.0000 * 1.0000 * 0.9605 = 0.01313739$$

P (X|Ci) = P (X| remark = tidak puas) (6)

$$0.7955 * 0.1136 * 0.2500 * 0.3409 * 0.4545 = 0.00350178$$

P (X|Ci) P (Ci) = P (X| *remark* = puas) P (*remark* = puas) (7)

$$0.6333 * 0.01313739 = 0.0083203$$

P (X|Ci) P (Ci) = P (X| *remark* = tidak puas) P (*remark* = tidak puas) (8)

0,3667 * 0,00350178= 0,001284

Dari perolehan rekapitulasi tersebut maka didapat nilai P (X|Ci) P (Ci) untuk *remark* puas adalah 0,0083203, ini lebih besar dari *remark* tidak puas dengan nilai 0,001284. Dengan demikian tercapai kesimpulan bahwasanya data uji yang tertera tergolong dari klasifikasi puas. Dari perhitungan ini, hasil data sesuai dengan hasil respon dari data uji.

Confusion matrix.

Dengan metode C4.5, pada Tabel 9 terdapat 120 data latih, 72 diklasifikasikan Puas sesuai prediksi, lalu 3 data diprediksi puas ternyata tidak puas, 41 data tidak puas sesuai prediksi, dan 4 data diprediksi tidak puas ternyata puas.

Tabel 9. Model *Confusion matrix* untuk Metode C4.5

Accuracy: 94.17 % +/- 5,34 (mikro 94,17%)			
	True PUAS	True TIDAK PUAS	*Class precision*
Prediksi PUAS	72	3	96%
Prediksi TIDAK PUAS	4	41	91,11%
Class Recall	94,74%	93,18%	

Sedangakan dengan *Naïve Bayes*, berdasarkan data latih, diketahui pada Tabel 10 bahwa dari 120 data, 73 data tergolong puas menurut prediksi, kemudian diprediksi 14 data puas tetapi ternyata tidak puas, 30 data tidak puas sesuai prediksi, dan diprediksi 3 data tidak puas ternyata puas.

Tabel 10. Model *Confusion matrix* untuk Metode *Naïve Bayes*

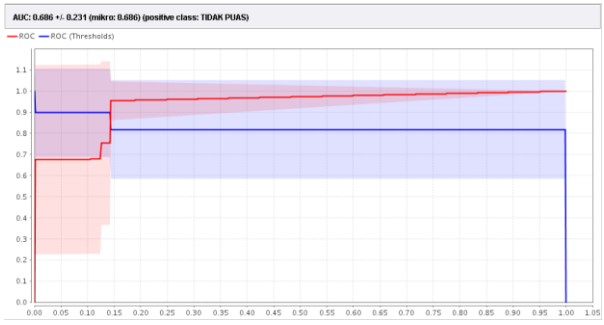
Accuracy: 85,83 % +/- 11,81 (mikro 85,83%)			
	True PUAS	True TIDAK PUAS	*Class precision*
Prediksi PUAS	73	14	83,91%
Prediksi TIDAK PUAS	3	30	90,91%
Class Recall	96,05%	68,18%	

Setelah dua tabel *Confusion matrix* dibuat, hitung *Accuracy, Precision, dan Recall*. Tabel 11 menampilkan perbandingan nilai tersebut untuk metode C4.5 dan Naïve Bayes.

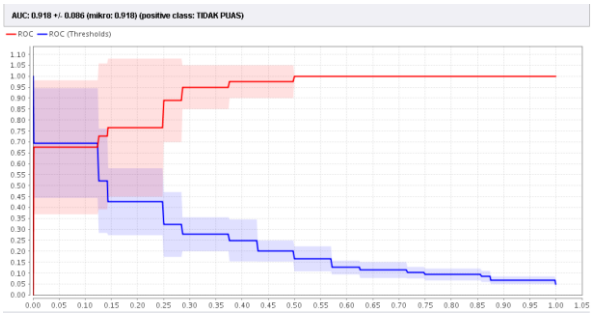
Tabel 11. Perbandingan perhitungan *Accuracy, Precision, dan Recall*

	C4.5	*Naïve Bayes*
Accuracy	94,17%	85,83%
Precision	92,50%	89,67%
Recall	92,50%	67,50%

Kurva ROC



Gambar 2 Kurva ROC dengan algortima C4.5



Gambar 3 Kurva ROC dengan algoritma *Naïve Bayes*

Analisa Hasil Perbandingan

Cross validation digunakan untuk menguji model hasil metode C4.5 dan Naïve Bayes.

Tabel 12. Perbandingan nilai *accuracy* dan AUC

	C4.5	*Naïve Bayes*
Accuracy	94,17%	85,83%
AUC	0.686	0.918

Pembagian kelompok nilai AUC untuk klasifikasi data mining (Suwarno, 2016), yaitu:

a. 0.90 hingga 1.00 untuk Klasifikasi sangat baik
b. 0.80 hingga 0.90 untuk Klasifikasi Baik
c. 0.70 hingga 0.80 untuk Klasifikasi Cukup
d. 0.60 hingga 0.70 untuk Klasifikasi Buruk
e. 0.50 hingga 0.60 untuk Klasifikasi Salah

Dapat dilihat dari perbandingan nilai *accuracy* dan AUC, nilai *accuracy* C4.5 94,17%, namun nilai AUC dengan kriteria klasifikasinya yang buruk. Sedangkan, nilai *accuracy* Naïve Bayes memiliki nilai 85,83%, tapi nilai AUC termasuk ke dalam klasifikasi sangat baik.

4. Kesimpulan

Dari penelitian yang telah dilakukan, perbandingan metode-metode untuk menentukan kepuasan pelanggan, dapat ditarik kesimpulan:

1. Perbandingan metode klasifikasi *Data mining*, C4.5 dan Naïve Bayes, menunjukan C4.5 lebih akurat dari pada metode Naïve Bayes. Hal tersebut dibuktikan dengan nilai *accuracy* dimana metode C4.5 memiliki nilai *accuracy* 94,17%, sedangkan *Naïve Bayes* 85,83%.
2. Berdasarkan metode C4.5, atribut *Assurance* merupakan atribut yang paling mempengaruhi kinerja perusahaan. Dilihat dari atribut *Assurance* sebagai *node root.*
3. Berdasarkan nilai AUC, *Naïve Bayes* termasuk dalam kategori sangat baik, C4.5 termasuk dalam kategori klasifikasi buruk.

Daftar Pustaka

Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, *2*(2), 213–219. https://doi.org/10.22202/ei.2016.v2i2.1465

Rifqo, M. H., & Wijaya, A. (2017). Implementasi Algoritma Naive Bayes Dalam Penentuan Pemberian Kredit. *Jurnal Pseudocode*, *4*(2), 120–128. https://doi.org/10.33369/pseudocode.4.2.120-128

Sipayung, E. M., Maharani, H., & Zefanya, I. (2016). Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Informasi (JSI)*, *8*(1), 958–965. https://ejournal.unsri.ac.id/index.php/jsi/article/view/3250/1907

Shiddiq, A., Niswatin, R. K., & Farida, I. N. (2018). Analisa Kepuasan Konsumen Menggunakan Klasifikasi Decision Tree Di Restoran Dapur Solo. *Generation Journal*, *2*(1), 9.

Mashuri, M., & Mardianis, N. (2020). Pengaruh Jumlah Pelanggan Terhadap Tingkat Profitabilitas Pada Perusahaan Daerah Air Minum Di Kota Bengkalis. *JAS (Jurnal Akuntansi Syariah)*, *4*(1), 83–94. https://doi.org/10.46367/jas.v4i1.220

Sugianto, C. A. (2017). Penerapan Teknik Data Mining Untuk Menentukan Hasil Seleksi Masuk Sman 1 Gibeber Untuk Siswa Baru Menggunakan Decision Tree. *Jurnal Teknologi Rekayasa*, *21*(1), 49–54. https://doi.org/10.31227/osf.io/vedu7

Adriyendi, A., & Melia, Y. (2020). Klasifikasi Menggunakan Naïve Bayes Dan K-Nearest Neighbor Pada Manajemen Layanan Teknologi Informasi. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *2*(2), 99–107. https://doi.org/10.47233/jteksis.v2i2.121

Takalapeta, S. (2018). Penerapan Data Mining Untuk Menganalisis Kepuasan Konsumen Menggunakan Metode Algoritma C4.5. *J I M P - Jurnal Informatika Merdeka Pasuruan*, *3*(3), 34–38. https://doi.org/10.37438/jimp.v3i3.186

Safri, Y. F., Arifudin, R., & Muslim, M. A. (2018). K-Nearest Neighbor and Naive Bayes Classifier Algorithm in Determining The Classification of Healthy Card Indonesia Giving to The Poor. *Scientific Journal of Informatics*, *5*(1), 18. https://doi.org/10.15294/sji.v5i1.12057

Suwarno, A. A. (2016). Penerapan Algoritma Bayesian Regularization Backpropagation Untuk Memprediksi Penyakit Diabetes. *Jurnal MIPA*, *39*(2), 98–106.