



Analysis of Passenger Satisfaction on Airlines Using C4.5 and Naïve Algorithm

Analisis Kepuasan Penumpang pada Maskapai Penerbangan Menggunakan Algoritma C4.5 dan Naïve Bayes

**Yusuf Hendra Kusuma¹, Seno Suprapto², Sariyono³,
Yandra Setiawan⁴, Rahmaddeni^{5*}**

^{1,2,3,4,5}Program Studi Teknik Informatika, STMIK AMIK Riau, Indonesia

E-Mail: ¹1910031802121@sar.ac.id, ²1910031802123@sar.ac.id, ³1910031802143@sar.ac.id,
⁴1810031802161@sar.ac.id, ⁵rahmaddeni@sar.ac.id

Corresponding Author: Rahmaddeni

Abstract

Indonesia is a country that has many islands where people tend to need transportation that makes it easier for people to travel from one area to another. The problem of service quality of user satisfaction often arises because many companies are judged from the company's point of view, not from the user's point of view. Formation of a good model must use a good algorithm for example Naïve Bayes and C4.5 which are algorithms in the decision tree that are used for classification. This research was conducted using data obtained from the website www.kaggle.com. Before getting the variables to be used, it has to go through data processing. Based on data processing using Rapid Miner software, the accuracy value is 81.19% for the C4.5 algorithm and 76.24% for the Naive Bayes algorithm, meaning that the resulting rule has a level of truth close to 100%. In accordance with the provisions of the application of data mining.

Keywords: C4.5, Data Mining, Decision Tree, Naïve Bayes

Abstrak

Indonesia adalah sebuah negara yang memiliki banyak pulau dimana masyarakat cenderung membutuhkan alat transportasi yang memudahkan masyarakat bepergian dari satu daerah ke daerah yang lain. Permasalahan kualitas pelayanan kepuasan pengguna sering timbul karena banyaknya perusahaan yang memberi penilaian dari sudut pandang perusahaan bukan dari sudut pandang pengguna. Pembentukan model yang baik harus menggunakan algoritma yang baik, contohnya Naïve Bayes dan C4.5 yang merupakan algoritma dalam pohon keputusan (*decision tree*) yang digunakan untuk klasifikasi. Penelitian ini dilakukan dengan menggunakan data yang diperoleh dari situs www.kaggle.com. Sebelum mendapatkan variabel yang akan digunakan, proses yang harus dilalui adalah *prosessing* data. Berdasarkan pengolahan data menggunakan *software* Rapid Miner didapat nilai akurasi sebesar 81,19% untuk algoritma C4.5 dan 76,24% untuk algoritma Naïve Bayes, artinya bahwa *rule* yang dihasilkan tingkat kebenarannya mendekati 100%. Sesuai dengan ketentuan penerapan *Data Mining*.

Kata Kunci : C4.5, Data Mining, Decision Tree, Naïve Bayes

1. PENDAHULUAN

Indonesia adalah sebuah negara yang memiliki banyak pulau dimana masyarakat cenderung membutuhkan alat transportasi yang memudahkan masyarakat bepergian dari satu daerah ke daerah yang lain, dari sekian banyak transportasi yang paling mudah melalui jalur udara. Ini adalah potensi yang sangat bagi perusahaan maskapai penerbangan[1].

Permasalahan kualitas pelayanan kepuasan penumpang sering timbul karena banyaknya perusahaan yang dinilai dari sudut pandang perusahaan bukan dari sudut pandang pengguna. Maka dari itu, perusahaan harus membuat strategi dan program pelayanan, perusahaan harus berorientasi pada kepentingan pelanggan

dengan memperhatikan komponen kualitas pelayanan supaya tujuan perusahaan untuk meningkatkan kualitas pelayanan penumpang secara maksimal[2].

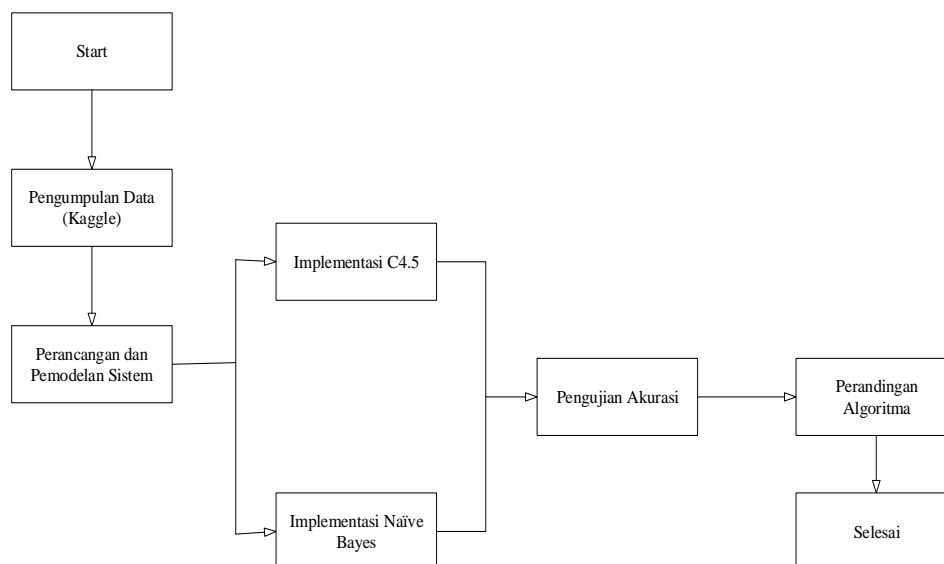
Dari sekian banyak cara yang mudah digunakan untuk mengetahui tingkat kepuasan pelanggan adalah teknik data mining[3]. Data mining merupakan proses pencarian pola atau informasi dalam data dengan menggunakan teknik atau metode tertentu[4]. Website *www.kaggle.com* menyediakan data penerbangan yang terdiri dari 500 data, 7 atribut dan 1 label dimana jenis data ini termasuk dalam kategori data *supervised learning*.

Pembentukan model klasifikasi yang baik membutuhkan algoritma yang baik, salah satunya menggunakan Naïve Bayes. Berdasarkan beberapa penelitian yang telah dilakukan, algoritma Naïve Bayes mampu menunjukkan hasil yang lebih baik[5][6][7]. Algoritma C4.5 adalah salah satu algoritma untuk mencari pohon keputusan (*decision tree*). Algoritma C4.5 menggunakan *Information Gain* dan *Gain Ratio* untuk memilih atribut [8].

Algoritma C4.5 digunakan dalam pembuatan pohon keputusan untuk mengambil keputusan sebab Algoritma C4.5 merupakan algoritma yang paling banyak digunakan dalam melakukan klasifikasi. Keunggulan Algoritma C4.5 dapat menangani atribut kontiniu dan diskrit, dapat menyelesaikan data tanpa nilai dari atribut., serta dapat meminimalkan pohon keputusan sesudah dibuat. Algoritma C4.5 dapat diolah dari model pohon keputusan menjadi *rule* yang sederhana sehingga mudah untuk dipahami[8].

2. METODOLOGI PENELITIAN

Penelitian diawali dengan tahap pengumpulan data, kemudian dilanjutkan dengan pemodelan sistem yang mencakup implementasi algoritma C4.5 dan Naïve Bayes, pada proses selanjutnya dilakukan proses pengujian akurasi dengan membandingkan hasil dari algoritma yang diperoleh. Diagram alur penelitian dijelaskan pada gambar 1.



Gambar 1. Alur Penelitian

2.1 Data Mining

Data mining adalah langkah pengumpulan data untuk diolah dan dijadikan informasi pengetahuan yang bermanfaat. Tujuan data mining yaitu untuk menghasilkan informasi baru. Data mining dibagi menjadi lima tahapan yang dibuat secara berurutan yaitu estimasi, prediksi, klasifikasi, clustering, asosiasi[4].

2.2 Algoritma C4.5

Algoritma C4.5 adalah teknik untuk mencari pohon keputusan yang dipakai untuk memperoleh aturan dan pohon keputusan dengan tujuan mengembangkan akurasi dan prediksi. Selain itu algoritma C4.5 juga salah satu algoritma yang mudah dipahami.

Entropy adalah informasi yang bisa melihat karakter dari input dan homogeneity dari sekumpulan data. Nilai *entropy* akan dihitung untuk mencari nilai *gain* nilai *Information Gain* (IG) tiap atribut. Nilai *entropy* dihitung menggunakan rumus (1).

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

Keterangan:

S	: Himpunan Kasus
n	: Jumlah Partisi S
p_i	: Proporsi dari terhadap S

Setelah mendapatkan nilai entropy untuk sekumpulan data, selanjutnya dapat mengukur keefektifan sebuah atribut untuk pengkolompokan data. Ukuran efektivitas ini disebut perolehan informasi. Hal ini dapat dilihat dalam rumus (2).

$$Gain(S,A) = Entropy - \sum_{i=1}^n \frac{S_i}{S} \times Entropy(S_i) \quad (2)$$

Dengan :

S	: Himpunan Kasus
A	: Atribut
n	: Jumlah Partisi Atribut A
$ S_i $: Jumlah Kasus Pada Partisi ke i
$ S $: Jumlah Kasus Dalam S

Algoritma C4.5 secara berkala mengunjungi setiap simpul keputusan, memilih pembagian yang optimal, sampai tidak dapat dibagi lagi. Proses akan terus berlanjut hingga setiap cabang mempunyai kelas yang sama.

2.3 Algoritma Naïve Bayes

Naïve Bayes adalah suatu metode klasifikasi probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menambahkan frekuensi dan kombinasi nilai pada data yang ada. Seorang ilmuwan Inggris Thomas Bayes memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Naïve Bayes memiliki keuntungan yaitu hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Untuk menyelesaikan metode Naïve Bayes dapat dilakukan dengan persamaan (3) dan (4) [11].

- Baca data training
- Hitung Jumlah dan Probabilitas Prior.

$$P(C_i) = S_i/s \quad (3)$$

- Hitung Probabilitas Training.

$$P(K | D) = (P(D | K) \times P(K)) / (P(D)) \quad (4)$$

Di mana D adalah kelas data yang tidak diketahui. Selanjutnya K adalah hipotesis khusus kelas. $P(K/D)$ adalah Probabilitas hipotesis K berdasarkan kondisi D . $P(K)$ yaitu probabilitas sesuai dengan hipotesis D . $P(K/D)$ merupakan probabilitas D kondisi yang berdasarkan pada hipotesis D . Sedangkan $P(D)$ adalah probabilitas K .

- Hitung nilai probabilitas data testing.

$$P(K | a_1, \dots, a_n) = P(D) P(a_1, \dots, a_n | D) \quad (5)$$

- Menghitung Probabilitas Akhir.

$$P(K | D) = P(D | K) \times P(K | D) \quad (6)$$

- Menghitung Nilai Akhir

$$C = \arg \max P(D | K) \quad (7)$$

2.4 Dataset

Dataset berasal dari website www.kaggle.com[9] yang menyediakan data penerbangan yang terdiri dari 500 data, 7 atribut dan 1 label dimana jenis data ini termasuk dalam kategori data *supervised learning*[5]. Dataset penerbangan dapat dilihat pada tabel 1.

Tabel 1. Dataset Penerbangan

No	Gender	Customer Type	Range of Age	Class	Range of Dapeture Delay	Range Of Arrival Delay	Satisfaction
1	Pria	Pelanggan	1	Ekonomi Plus	2	2	netral atau tidak puas
2	Pria	tidak pelanggan	1	Bisnis	2	2	netral atau tidak puas
3	Wanita	Pelanggan	2	Bisnis	1	1	puas
4	Wanita	Pelanggan	1	Bisnis	2	2	netral atau tidak puas
5	Pria	Pelanggan	3	Bisnis	1	1	puas
6	Wanita	Pelanggan	2	Ekonomi	1	1	netral atau tidak puas
7	Pria	Pelanggan	3	Ekonomi	2	2	netral atau tidak puas
8	Wanita	Pelanggan	3	Bisnis	2	1	puas
9	Wanita	Pelanggan	2	Bisnis	1	1	netral atau tidak puas
10	Pria	tidak pelanggan	1	Ekonomi	1	1	netral atau tidak puas
...
497	Wanita	Pelanggan	2	Ekonomi	1	1	netral atau tidak puas
498	Wanita	Pelanggan	2	Ekonomi Plus	1	1	netral atau tidak puas
499	Pria	Pelanggan	3	Ekonomi	1	1	netral atau tidak puas
500	Wanita	Pelanggan	3	Ekonomi Plus	1	1	netral atau tidak puas

2.5 Split Data

Split data adalah proses pembagian dataset menjadi dua bagian yaitu data latih dan data uji. Dalam penelitian ini, split data yang digunakan adalah 90:10. Tujuan dari split data ini untuk mendapatkan hasil akurasi terbaik[10]. Pembagian data latih dan data uji ditunjukkan pada tabel 2 dan 3.

Tabel 2. Data Latih

No	Gender	Customer Type	Range of Age	Class	Range of Dapeture Delay	Range of Arrival Delay	Satisfaction
1	Pria	Pelanggan	1	Ekonomi Plus	2	2	Netral Atau Tidak Puas
2	Pria	Tidak Pelanggan	1	Bisnis	2	2	Netral Atau Tidak Puas
3	Wanita	Pelanggan	2	Bisnis	1	1	Puas
4	Wanita	Pelanggan	1	Bisnis	2	2	Netral Atau Tidak Puas
5	Pria	Pelanggan	3	Bisnis	1	1	Puas
...
395	Wanita	Pelanggan	1	Ekonomi	1	2	Netral Atau Tidak Puas
396	Wanita	Pelanggan	3	Bisnis	1	1	Netral Atau Tidak Puas
397	Wanita	Pelanggan	1	Ekonomi	1	1	Netral Atau Tidak Puas
398	Wanita	Pelanggan	2	Ekonomi	2	2	Netral Atau Tidak Puas
399	Pria	Pelanggan	2	Bisnis	1	1	Puas
400	Pria	Pelanggan	1	Ekonomi	2	1	Netral Atau Tidak Puas

Tabel 3. Data Uji

No	Gender	Customer Type	Range Of Age	Class	Range of Dapeture Delay	Range of Arrival Delay	Satisfaction
401	Wanita	Pelanggan	2	Bisnis	1	1	puas
402	Pria	Pelanggan	1	Ekonomi	2	1	netral atau tidak puas
403	Pria	Pelanggan	2	Bisnis	2	3	puas
404	Wanita	Pelanggan	3	Ekonomi Plus	1	2	puas
405	Pria	tidak pelanggan	1	Ekonomi	4	4	netral atau tidak puas
...
495	Wanita	Pelanggan	3	Bisnis	1	1	puas
496	Pria	Pelanggan	2	Ekonomi Plus	4	4	netral atau tidak puas
497	Wanita	Pelanggan	2	Ekonomi	1	1	netral atau tidak puas
498	Wanita	Pelanggan	2	Ekonomi Plus	1	1	netral atau tidak puas
499	Pria	Pelanggan	3	Ekonomi	1	1	netral atau tidak puas
500	Wanita	Pelanggan	3	Ekonomi Plus	1	1	netral atau tidak puas

3. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan dengan menggunakan data yang diambil dari situs *www.kaggle.com*. Setelah dilakukan *processing* data maka diperoleh variabel yang akan digunakan.

- Gender
- Costumer
- Range of age
- Range of dapture delay
- Range of dapture arrival delay
- Statisfaction

3.1 Algoritma C4.5

Perhitungan Algoritma C4.5 untuk memperoleh model aturan pohon keputusan dapat diuraikan dalam beberapa tahapan.

Langkah 1: Menghitung jumlah kasus, jumlah kasus untuk keputusan puas, jumlah kasus untuk keputusan netral atau tidak puas.

Langkah 2: Menghitung Entropy dari semua kasus dan kasus yang dibagi berdasarkan kelas atribut dengan persamaan (1). Selanjutnya dilakukan penghitungan *Gain* untuk masing-masing atribut dengan persamaan (2). Perhitungan nilai *Entropy* dan *Gain* ditunjukkan oleh Tabel 4.

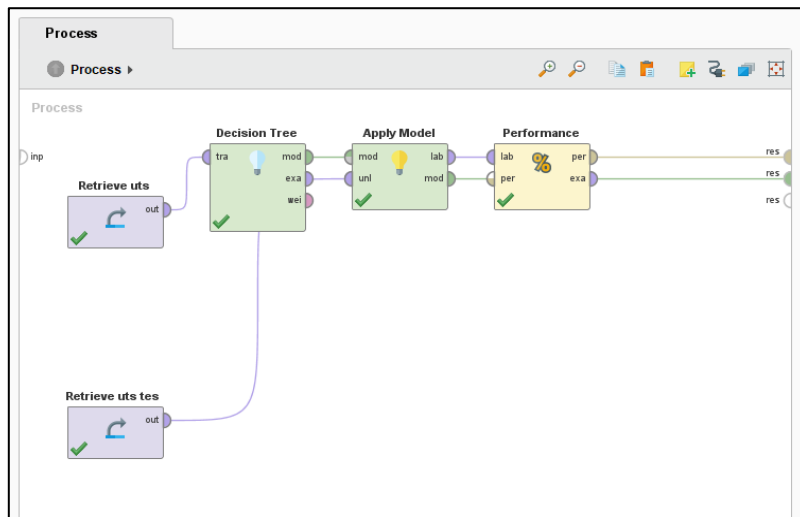
Tabel 4. Tabel Perhitungan Entropy dan Gain

	Jumlah	Puas	Netral atau Tidak Puas	Entropy	Gain
Total	500	222	278	0,990932404	
Gender					0,002912514
Pria	239	114	125	0,998471422	
Wanita	261	108	153	0,978449329	
Customer Type					0,017204447
Pelanggan	422	201	221	0,998379152	
tidak pelanggan	78	21	57	0,840358672	
Age					0,058719878
0-25	128	27	101	0,743270919	
26-45	203	110	93	0,994935238	
>45	169	85	84	0,999974743	
Class					1,107012683
Ekonomi	208	38	170	0,685931523	
Ekonomi Plus	36	11	25	0,88797632	
Bisnis	256	173	83	0,908910327	

	Jumlah	Puas	Netral atau Tidak Puas	Entropy	Gain
Departure Delay in Minutes					0,00179472
0	280	130	150	0,99631652	
1-30 Menit	152	64	88	0,981940787	
31-60 Menit	36	14	22	0,964078765	
>60 menit	32	14	18	0,988699408	
Arrival Delay in Minutes					0,010559611
0	269	134	135	0,999990031	
1-30 Menit	164	62	102	0,956652272	
31-60 Menit	34	12	22	0,936667382	
>60 menit	33	14	19	0,98337619	

Dari hasil perhitungan *Entropy* dan *Gain* diatas, atribut yang menjadi node pertama adalah class kemudian diikuti dengan atribut Customer Type, Range of Age, Range of Dapture Delay, dan Range of Dapture Arrival Delay.

Dilihat dari hasil pengelolaan data yang diimplementasikan pada Rapid Miner memperoleh hasil sebesar 81,19%, yang berarti keaslian aturan yang dihasilkan mendekati 100%. Berdasarkan ketentuan maka dihasilkan sebanyak 12 model aturan untuk tingkat kepuasan dalam pelayanan penerbangan. Hasil visualisasi pengolahan pada Rapid Miner dapat ditunjukkan pada gambar 1, 2 dan 3.



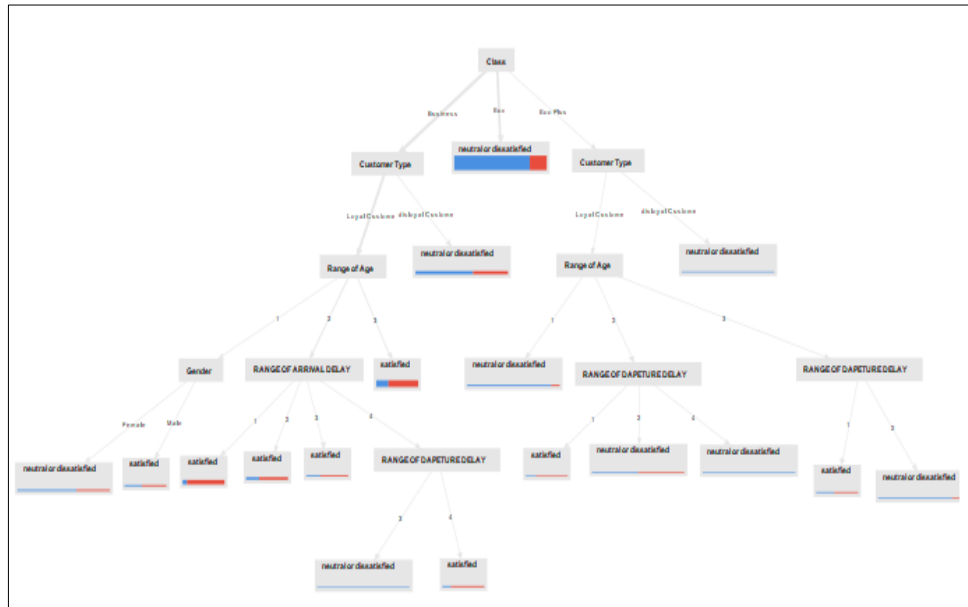
Gambar 1. Design pada Rapid Miner

```

PerformanceVector

PerformanceVector:
accuracy: 81.19%
ConfusionMatrix:
True:   neutral or dissatisfied satisfied
neutral or dissatisfied:      45    11
satisfied:      8      37
    
```

Gambar 2. Performance Vector



Gambar 3. Pohon Keputusan

3.2 Algoritma Naïve bayes

Hasil dari perhitungan menggunakan algoritma Naïve Bayes menghasilkan perhitungan dalam bentuk frekuensi table dan presentasi prediksi kelas yang ditunjukkan pada tabel 5 dan 6.

Tabel 5. Perhitungan Frekuensi Table

P (Puas/netral atau tidak puas)	0,44	0,56
P (Gender)	Puas	netral atau tidak puas
Pria	51%	45%
Wanita	49%	55%
P (Customer Type)	Puas	netral atau tidak puas
Pelanggan	91%	79%
tidak pelanggan	9%	21%
P (Range of Age)	Puas	netral atau tidak puas
1	12%	33%
2	50%	36%
3	38%	30%
P (Type of Travel)	Puas	netral atau tidak puas
Personal Travel	6%	49%
Bisnis travel	94%	51%
P (Class)	Puas	netral atau tidak puas
Ekonomi	17%	61%
Ekonomi Plus	5%	9%
Bisnis	78%	30%
P (Range of Dapeture Delay)	Puas	netral atau tidak puas
0	0%	0%
1	59%	54%
2	29%	32%
3	6%	8%
4	6%	6%
P (Range of Arrival Delay)	Puas	netral atau tidak puas
0	0%	0%
1	60%	49%
2	28%	37%
3	5%	8%
4	6%	7%

Tabel 6. Persentase Puas dan netral atau tidak puas

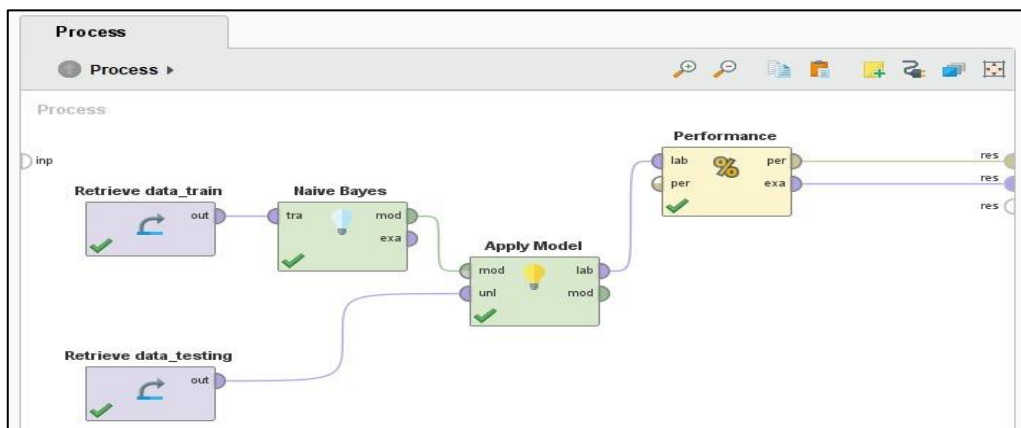
Class Prediction	Puas	Netral atau Tidak Puas
netral atau tidak puas	0,0%	0,2%
Puas	2,5%	0,3%
netral atau tidak puas	0,0%	0,2%
Puas	0,1%	0,0%
Puas	0,1%	0,1%
netral atau tidak puas	0,0%	0,0%
netral atau tidak puas	0,0%	0,0%
Puas	2,5%	0,3%
...
Puas	2,0%	0,2%
Puas	1,9%	0,2%
Puas	0,0%	0,0%
netral atau tidak puas	0,0%	0,6%
netral atau tidak puas	0,0%	0,1%
netral atau tidak puas	0,0%	0,4%
Puas	0,1%	0,1%

Sementara itu, dilakukan pengolahan data secara manual serta menggunakan software Rapid Miner yang memperoleh nilai akurasi sebesar 76.2%, artinya akurasi yang ditunjukkan mendekati 100%. Tabel perolehan akurasi manual ditunjukkan pada tabel 7.

Tabel 7. Akurasi Perhitungan Manual

ConFusion Tabel	CLASS		
	Prediksi Puas	Puas	netral atau tidak puas
netral atau tidak puas		42	18
	netral atau tidak puas	6	35
	AKURASI =	76%	

Hasil perhitungan lainnya menggunakan Rapid Miner dapat divisualisasikan melalui desain algoritma yang digunakan. Secara keseluruhan, hasil yang diperoleh dapat diamati pada performance vector maupun kurva AUC pada Rapid Miner. Desain algoritma dan hasil performa algoritma Naïve Bayes dapat ditunjukkan pada gambar 4, 5, 6 dan 7.



Gambar 4. Desain Algoritma Naïve Bayes

Criterion	Table View	Plot View
accuracy	accuracy: 76.24%	
precision		
recall		
AUC (optimistic)		
AUC		
AUC (pessimistic)		

	true neutral or dissatisfied	true satisfied	class precision
pred. neutral or dissatisfied	40	11	78.43%
pred. satisfied	13	37	74.00%
class recall	75.47%	77.08%	

Gambar 5. Report Accuracy

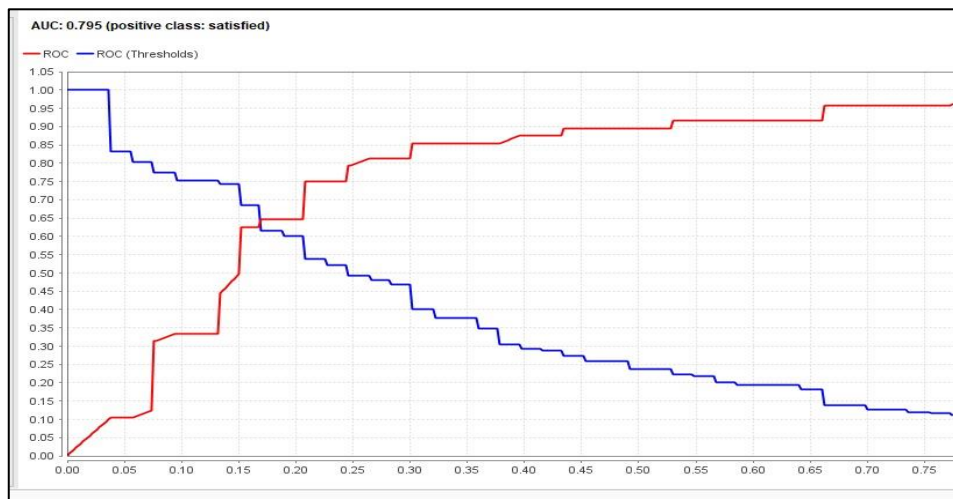
PerformanceVector

```

PerformanceVector:
accuracy: 76.24%
ConfusionMatrix:
True:  neutral or dissatisfied satisfied
neutral or dissatisfied:      40      11
satisfied:      13      37
precision: 74.00% (positive class: satisfied)
ConfusionMatrix:
True:  neutral or dissatisfied satisfied
neutral or dissatisfied:      40      11
satisfied:      13      37
recall: 77.08% (positive class: satisfied)
ConfusionMatrix:
True:  neutral or dissatisfied satisfied
neutral or dissatisfied:      40      11
satisfied:      13      37
AUC (optimistic): 0.799 (positive class: satisfied)
AUC: 0.795 (positive class: satisfied)
AUC (pessimistic): 0.792 (positive class: satisfied)

```

Gambar 6. Performance Vector



Gambar 7. AUC Naïve Bayes

3.3 Analisis Hasil

Berdasarkan hasil implementasi yang telah dilakukan untuk mencari kepuasan penumpang maskapai penerbangan menggunakan algoritma C4.5 serta Naïve Bayes didapat perbandingan hasil pengujian menggunakan software Rapid Miner yang dapat ditunjukkan pada tabel 8.

Tabel 8. Analisis Hasil

No	Algoritma	Accuracy
1	C4.5	81.2%
2	Naïve Bayes	76.2%

4. KESIMPULAN

Penggunaan Data mining klasifikasi menggunakan algoritma C4.5 berkontribusi besar dan dapat menjadi tolak ukur untuk melihat tingkat kepuasan penumpang. Setelah dilakukan perhitungan Algoritma C4.5 dan Naïve Bayes menghasilkan akurasi yang berbeda dimana algoritma C4.5 menghasilkan akurasi yang lebih tinggi. Hasil dari pengujian algoritma C4.5 menggunakan Rapid Miner memperoleh hasil akurasi 81,19%. Pengujian Algoritma Naïve Bayes bisa diuji menggunakan berbagai macam software salah satunya adalah Rapid iner yang menghasilkan akurasi sebesar 76%.

REFERENSI

- [1] M. Darus and K. Mahalli, "Analisis Tingkat Kepuasan Penumpang Terhadap Kualitas Pelayanan Di Bandar Udara Internasional Kualanamu," *J. Ekon. dan Keuang.*, vol. 3, no. 6, p. 14857, 2015.
- [2] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.
- [3] Yoga Religia and A. Amali, "Perbandingan Optimasi Feature Selection pada Naïve Bayes untuk Klasifikasi Kepuasan Airline Passenger," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 527–533, 2021, doi: 10.29207/resti.v5i3.3086.
- [4] R. H. Alawiah, Saifullah, and I. S. Damanik, "Analisis Kepuasan Konsumen Terhadap Pelayanan Bengkel Menggunakan Metode Algoritma C4.5," *Januari*, vol. 2, no. 1, pp. 31–38, 2021.
- [5] I. A. A. Amra and A. Y. A. Maghari, "Students performance prediction using KNN and Naïve Bayesian," *ICIT 2017 - 8th Int. Conf. Inf. Technol. Proc.*, pp. 909–913, 2017, doi: 10.1109/ICITECH.2017.8079967.
- [6] O. F.Y, A. J.E.T, A. O, H. J. O, O. O, and A. J, "Supervised Machine Learning Algorithms: Classification and Comparison," *Int. J. Comput. Trends Technol.*, vol. 48, no. 3, pp. 128–138, 2017, doi: 10.14445/22312803/ijett-v48p126.
- [7] E. N. Azizah, U. Pujianto, E. Nugraha, and Darusalam, "Comparative performance between C4.5 and Naive Bayes classifiers in predicting student academic performance in a Virtual Learning Environment," *2018 4th Int. Conf. Educ. Technol. ICET 2018*, no. 1, pp. 18–22, 2018, doi: 10.1109/ICEAT.2018.8693928.
- [8] A. Angkasa and D. Fitrihanah, "The Implementation of Classification Algorithm C4.5 in Determining the Illness Risk Level for Health Insurance Company in Indonesia," *Int. J. Comput. Appl.*, vol. 177, no. 37, pp. 44–50, 2020, doi: 10.5120/ijca2020919883.
- [9] "Airline Passenger Satisfaction | Kaggle." <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction> (accessed Aug. 22, 2022).
- [10] R. Rahmaddeni, M. K. Anam, Y. Irawan, and ..., "Comparison of Support Vector Machine and XGBSVM in Analyzing Public Opinion on Covid-19 Vaccination," *Ilk. J. ...*, vol. 14, no. 1, pp. 32–38, 2022, [Online]. Available: <http://jurnal.fikom.umi.ac.id/index.php/ILKOM/article/view/1090>.
- [11] Putra, D. S., et al. "Electromyography (EMG) signal classification for wrist movement using naïve bayes classifier." *Journal of Physics: Conference Series*. Vol. 1424. No. 1. IOP Publishing, 2019.