JUSIFOR: Jurnal Sistem Informasi dan Informatika
Vol. 4, No. 1, Juni 2025, hlmn. 75-84
E-ISSN: 2830-3393 | P-ISSN: 2830-2443
DOI: https://doi.org/10.70609/jusifor.v4i1.6491

Performa Algoritma Naïve Bayes Dan C4.5 dalam Perbandingan Evaluasi Paket Trip Terlaris Bromocreative

Puspita Najmulaili[1], Listanto Tri Utomo[2*]

[1,2] Sistem Informasi, Fakultas Teknologi Informasi, Universitas Merdeka Malang, Indonesia

Info Artikel

Riwayat Artikel:
Diterima : **17 Januari 2025**
Direvisi : **05 Juni 2025**
Diterbitkan : **25 Juni 2025**

Kata Kunci:

Naïve Bayes,

C4.5,

Evaluasi Paket Terlaris ,

Klasifikasi

ABSTRAK

Perkembangan yang pesat dalam industry pariwasata telah membuka peluang bagi para agen tour perjalanan untuk menawarkan berbagai paket wisata. Salah satu penyedia layanan wisatanya yaitu Bromocreative, dengan menghadapi tantangan dalam menentukan paket trip yang paling laris sebagai acuan untuk strategi pemasaran. Peneliti akan melakukan pengujian pada dua metode klasifikasi yaitu menggunakan metode algoritma Naïve Bayes dan C4.5. untuk melakukan evaluasi pada paket-paket wisata yang paling diminati atau terlaris. Dalam melakukan analisis ini akan dilakukan menggunakan dataset yang mencakup 1000 data pelanggan pada Bromocreative, yang meliputi jenis paket, jumlah peserta, harga, status, dan kategori (murah, sedang dan mahal. Untuk mengambil keputusan, proses pengambilan keputusan sebaiknya menggunakan aplikasi Orange sebagai basis dari kedua algoritma tersebut, dengan fokus pada akurasi, presisi, dan recall. Hasil evaluasi yang disajikan melalui confusion matrix akan mendapatkan keakuratan pada masing-masing metode. Agar diharapkan peneliti akan membantu Bomocreative dalam melakukan strategi pemasaran yang lebih efektif serta memberikan wawasan yang berharga bagi peneliti lainnya.

***Keywords*:**

Naïve Bayes,

C4.5,

Evaluation of the best-selling packages,

Classification

ABSTRACT

Rapid developments in the tourism industry have opened up opportunities for travel tour agents to offer various tour packages. One of the tourism service providers, Bromocreative, faces the challenge of determining the best-selling trip packages as a reference for marketing strategies. Researchers will test two classification methods, namely using the Naïve Bayes and C4.5 algorithm methods. to evaluate the most popular or best-selling tour packages. In carrying out this analysis, it will be carried out using a dataset that includes 1000 customer data on Bromocreative, which includes package type, number of participants, price, status and category (cheap, medium and expensive. To make a decision you need to use the Orange application as a comparison tool between the two This algorithm method measures the level of accuracy, precision and recall. The evaluation results presented through the confusion matrix will provide accuracy for each method. It is hoped that researchers will help Bomocreative carry out more effective marketing strategies and provide valuable insights for other researchers .

Penulis Korespondensi:
Listanto Tri Utomo,
Program Studi S1 Sistem Informasi,
Universitas Merdeka Malang
Email: listanto.utomo@unmer.ac.id

This is an open access article under the CC BY-SA license

1. PENDAHULUAN

Pariwisata adalah salah satu sektor ekonomi terpenting di Indonesia. Dengan adanya pariwisata ini, maka suatu negara atau lebih khusus lagi pemerintah daerah tempat objek wisata itu berada, akan mendapat pemasukkan atau devisa [2] . Destinasi wisata paling terkenal di Indonesia adalah Taman Nasional Bromo Tengger Semeru (TNBTS). Destinasi ini memikat para wisatawan domestic maupun mancanegara berkat keindahan alamnya yang memukau.

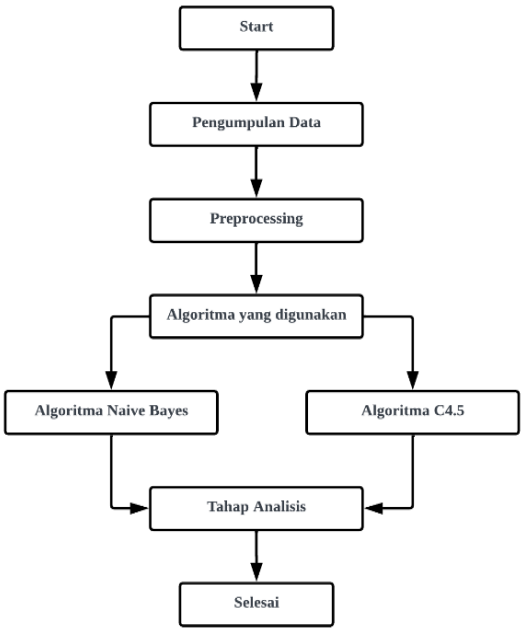
Kemajuan teknologi informasi dan transportasi tersebut juga telah memberikan dampak bagi sektor pariwisata yang membuat beberapa destinasi mampu berinteraksi dan bertukar pengalaman [8]. Bromocreative adalah penyedia wisata yang beroperasi di Kawasan Malang, Batu, dan Bromo. Dengan menawarkan para wisatwan untuk membantu menjelajahi keindahan Kota Malang dengan menyediakan berbagai paket tour yang beragam serta fasilitas yang sesuai dengan masing-masing paket. Dalam menjalankan bisnis, Bromocreative menawarkan beberapa pilihan paket trip yaitu Paket A, B, C, D, E, dan F. stiap masing-masing paket memiliki jumlah peminat yang bervariasi. Sehingga ingin mengidentifikasi paket mana yang paling diminati, baik yang harganya murah, sedang, maupun mahal.

Pada penelitian ini menggunakan klasifikasi data mining untuk melakukan proses menemukannya hubungan-hubungan yang tersembunyi. Hubungan untuk kemudian dapat menjadi landasan dalam pengambilan keputusan. Data Mining adalah proses penting dalam menganalisis jumlah data yang dikumpulkan mengenai paket-paket yang paling laris, yang memungkinkan optimalisasi dalam pnawaran layanan yang lebih tepat sasaran. Metode algoritma yang akan dilakukan adalah metode Naive Bayes dan C4.5, pada pengolahan data tersebut akan dilakukan membandingkan dari kedua metode tersebut dan menentukan hasil perbandingan evaluasi dengan menggunakan visualisasi aplikasi Orange. Melalui evaluasi ini, kami berharap dapat meningkatkan minat klien terhadap paket-paket yang kami tawarkan, sehingga selisih peminat antar paket tidak terlalu jauh.

Dalam konteks ini, klasifikasi melalui data mining memungkinkan pengolahan data yang akurat, sehingga dapat menghasilkan informasi mengenai penawaran paket perjalanan yang menguntungkan. Selain itu, pendekatan ini juga berkontribusi dalam merumuskan strategi yang lebih efektif dan efisien.

2. METODE PENELITIAN

Penelitian ini bertujuan untuk mengidentifikasi potensi permasalahan dan menggunakan metode pengumpulan data. Dalam mengklasifikasi paket trip yang terlaris, mengolah data yang diperoleh dan membandingkan metode yang akan diterapkan. Penelitian ini yang dipilih adalah metode algoritma Naïve Bayes dan C4.5 dengan visualisasi datanya menggunakan aplikasi Orange.



Gambar 1. Kerangka Kerja Penelitian

1) Pengumpulan Data

Pengumpulan data adalah metode penelitian yang menggambarkan cara memperoleh informasi yang diperlukan untuk mencapai tujuan penelitian. Proses ini bertujuan mengumpulkan berbagai informasi secara menyeluruh dan mendetail, sehingga dapat mencapai sasaran yang diinginkan.

2) Preprocessing

Data *Preprocessing* merupakan tahap penting yang bertujuan untuk mengubah data mentah menjadi data yang bersih dan siap digunakan. Data preprocessing bertujuan untuk mengurangi data ukuran, menemukan hubungan antara data, data yang menormalkan, outliers menghapus dan fitur ekstrak data [5]. Proses ini melakukan pemilihan fitur-fitur yang relevan dari data. Dalam penelitian ini, Langkah-langkah data *preprocessing* mencakup tiga tahap utama seperti Cleaning data, Trnformasi data, dan Reduksi data.

3) Algoritma Naïve Bayes

Metode klasifikasi Naive Bayes, yang berakar pada teorema Bayes, didasarkan pada probabilitas sederhana dan dimaksud untuk digunakan dengan asumsi bahwa antar variabel penjelas saling bebas (variabel independen) [4]. Dalam suatu metode klasifikasi yang menggunakan teorema Bayes dan berfokus pada nilai probabilitas, seperti :

$$p(H|X) = \frac{p(X|H).p(H)}{p(X)} \quad (1)$$

Keterangan :

H : hipotesis data *class* spesifik

X : data dengan class yang belum diketahui

p(H|X) : probabilitas hipotesis H diberikan bukti X (*Posteriori probability)*

p(X|H) : probabilitas hipotesis X terjadi apabila hipotesis H benar (*Probabilitas Likelihood)*

p(H) : probabilitas hipotesis H sebelum bukti X diamati

p(X) : probabilitas X terjadi dari hipotesis H

4) Algoritma C4.5

Decision Tree, juga dikenal sebagai "pohon keputusan", adalah sebuah struktur pohon di mana setiap node (simpul) pohon menunjukkan sifat yang telah diuji, setiap cabang menunjukkan pembagian hasil uji, dan setiap node daun menunjukkan kelompok kelas tertentu [1]. Dengan memiliki beberapa data *training* yang telah disiapkan yang memiliki nilai variabel target. Dalam metode ini dievaluasi semua atribut menggunakan ukuran statistik berupa *information gain* dan perhitungan entropi [6]. Rumus entropy seperti berikut :

$$\boldsymbol{Entropy\ (S) = -\sum_{pi}^{n} -pi * log_2 pi} \quad (2)$$

Keterangan :

S : Jumlah dataset

n : Jumlah total kelas

p_i : rasio antara jumlah sampel kelas i dengan jumlah sampel pada himpunan data

Perolehan informasi (*information gain*) adalah efisiensi suatu atribut dalam klasifikasi data, yang dicapai dengan menggunakan aturan tertentu. Nilai gain ini dapat dihitung menggunakan rumus berikut :

$$\boldsymbol{Gain\ (S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy(S_i)} \quad (3)$$

Dimana :

Gain (S, A) : Atribut A terhadap dataset S

N : Jumlah pada atribut A

$|S_i|$: Jumlah sampel dari S sesuai dengan kategori i pada atribut A

|S| : Jumlah sampel pada dataset S

5) Tahap Analisis

Pengumpulan data digunakan untuk menggali fakta mengenai variabel yang diteliti, misalnya perkiraan paket perjalanan yaitu penentuan paket trip yang terlaris. Untuk menguji keakuratan hasil penelitian ini menggunakan matriks akurasi. Berikut adalah table confusion matrix yang terdiri dari 3 kelas.

Tabel 1. Confusion Matrix

	Class	**Predicted**		
		A	B	C
Actual	A	True Positive (TP)	False Negative (FN)	False Negative (FN)
	B	False Positive (FP)	True Negative (TN)	True Negative (TN)
	C	False Positive (FP)	True Negative (TN)	True Negative (TN)

Keterangan :

TP : Jumlah data benar dalam memprediksi data positif

TN : Jumlah data benar namun diprediksi negatif

FN : Jumlah data salah diprediksi data positif sebagai negatif

FP : Jumlah data salah diprediksi data negatif sebagai positif

Confusion Matrix berfungsi sebagai alat untuk mengevaluasi kinerja klasifikasi yang didasarkan pada kemampuan akurasi prediktif suatu model. Akurasi dinyatakan dalam presentase, yang dimana 100% akurasi menunjukkan bahwa semua kasus berhasil dikelompokkan dan di klasifikasikan dengan tepat ke dalam kelas prediksinya. Nilai yang diperoleh dari True Positive (TP), True Negative (TN), False Negative (FN), dan False Positive (FP) dapat memberikan gambaran menyeluruh mengenai performa system dalam mengklasifikasi data. Berikut adalah metrik yang akan digunakan dalam evaluasi :

1. Rumus Akurasi

$$Accuracy = \frac{(TP+TN)}{TP+FP+TN+FN} \quad (4)$$

2. Rumus Recall

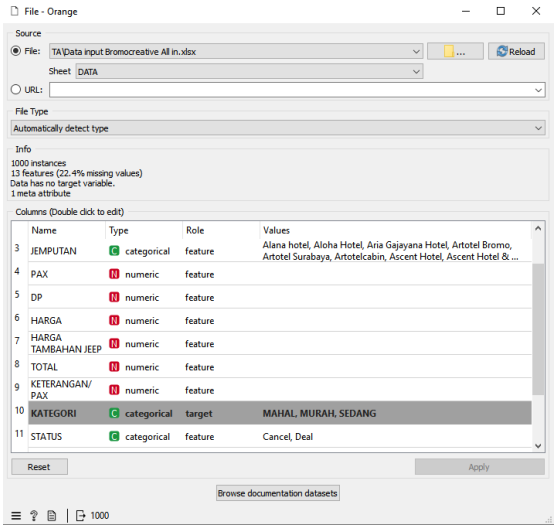
$$Recall = \frac{TP}{TP+FN} \quad (5)$$

3. Rumus Presisi

$$Precision = \frac{TP}{(TP+FP)} \quad (6)$$

3. HASIL DAN ANALISIS

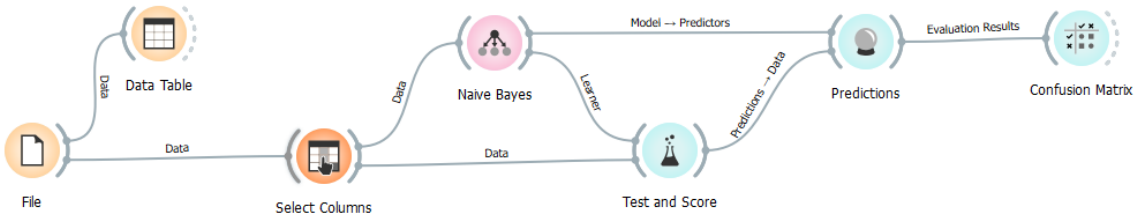
Penelitian menggunakan metode Naïve Bayes dan C4.5 untuk menganalisis paket trip yang terlaris, menggunakan aplikasi data mining Orange untuk menghitung hasil yang berdasarkan pada widget. Orange dikenal juga dengan Orange Data Mining merupakan perangkat lunak open source melakukan proses data mining atau data analytic melalui konsep visual programming [9]. Dataset yang digunakan pada penelitian ini adalah dataset klien Bromocreative dari tahun 2019 hingga 2024, dengan total sebanyak 1000 data. Tampilan awal Aplikasi Orange yang akan digunakan untuk pengolahan data, dengan cara klik dua kali untuk memasukkan dataset maka form columns akan menampilkan meta attribute, type dan target variabel. Berikut seperti gambar 2 dibawah ini.



Gambar 2. Kategori Target Values

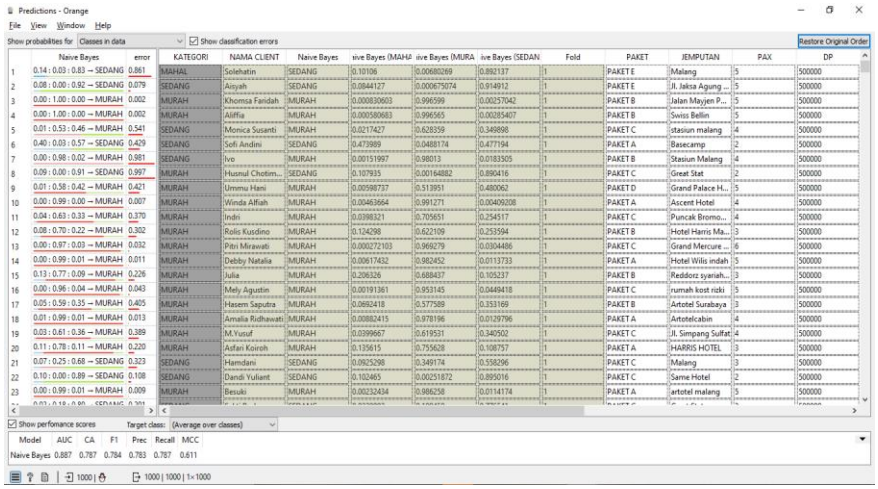
3.1. Algoritma Naïve Bayes

Widget akan digunakan dalam klasifikasi menggunakan metode Naïve Bayes dari dataset yang terhubung ke tabel data. Hal ini memungkinkan untuk melihat data mentah dan menggunakan widget Select Columns untuk menentukan kolom fitur dan target dengan menyertakan data berupa Data, Test Data, dan Leaner. Dengan demikian, dapat menghasilkan predictions serta hasil Confusion Matrix Evaluations Results. Berikut adalah gambaran widget dari Naïve Bayes.



Gambar 3. Widget Naïve Bayes

Hasil prediction ini akurasi klasifikasi yang dihitung menggunakan AUC (*Area Under Curve)*, CA (*Classification Accuracy*), dan Recall. Sehingga diperoleh hasil seperti gambar berikut.



Gambar 4. Predictions Naïve Bayes

Berdasarkan hasil perhitungan yang ditampilkan pada gambar diatas dapat diperoleh nilai Accuracy sebesar 0,887, Precision sebesar 0.783, dan Recall sebesar 0.787. Nilai AUC pada Naïve Bayes terdapat nilai paling tinggi yaitu 0,887. Akurasi digunakan untuk mengukur kinerja diskriminatif dan memperkirakan

probabilitas keluaran berdasarkan data yang dikumpulkan dari populasi positif dan negatif, yang menunjukkan kualitas data. Pada Confusion matrix digunakan untuk mengevaluasi performa dalam masalah klasifikasi machine learning, yang dapat menghasilkan output dalam bentuk 3 kelas, seperti berikut.

		Predicted			
		MAHAL	MURAH	SEDANG	Σ
Actual	MAHAL	51	15	51	117
	MURAH	5	549	34	588
	SEDANG	15	82	198	295
	Σ	71	646	283	1000

Gambar 5. Output Confusion Matrix Naïve Bayes

Grafik menunjukkan nilai True Positive (TP) sebesar 57, nilai True Negative (TN) sebesar 508, 75, 70, dan 176, nilai False Positive (FP) sebesar 5 dan 49, serta nilai False Negative (FN) sebesar 13 dan 47. Dengan melakukan perhitungan multiclass (mempunyai 3 class) maka menggunakan perhitungan seperti :

$Total\ TP = 51 + 549 + 198 = 798$

$Total\ FP = (5 + 15) + (15 + 82) + (51 + 34) = 202$

$Total\ FN = (15 + 51) + (5 + 34) + (15 + 82) = 202$

$Total\ TN = 863 + 315 + 620 = 1798$

Sehingga perhitungan rumusnya adalah :

$$Accuracy = \frac{(798 + 1798)}{(798 + 202 + 1798 + 202)} x100 = \frac{2596}{3000} x100 = 85{,}5\%$$

$$Precision = \frac{(798)}{(798 + 202)} x100 = \frac{798}{1000} x100 = 79{,}8\%$$

$$Recall = \frac{(798)}{(798 + 202)} x100 = \frac{798}{1000} x100 = 79{,}8\%$$

3.2. Algoritma C4.5

Algoritma C4.5 adalah salah satu metode klasifikasi yang digunakan dalam konstruksi pohon keputusan. Dimana setiap cabang kemudian mengarah ke node lain atau keputusan akhir [7]. Pada 1000 data latih ini digunakan sebagai untuk langkah awal dalam menghitung algoritma C4.5 seperti gambar di bawah.

Tabel 2. Data Latih

No	Nama Client	Paket	Fasilitas	Kategori	Status
1	Regina	Paket A	JeepLong	Murah	Deal
2	Upik	Paket A	Standart	Murah	Cancel
3	Rita Eka	Paket D	Jeeplong	Sedang	Deal
4	Afifah Nur	Paket E	Standart	Sedang	Cancel
5	Rizka Nazula	Paket A	Standart	Murah	Cancel
6	Flo	Paket C	Standart	Sedang	Deal
7	Normaila	Paket C	Standart	Mahal	Cancel
8	Solehatin	Paket E	Standart	Mahal	Deal
9	Knang Sadewo	Paket A	JeepLong	Murah	Deal
...	...	...	...	...	...
1000	Tonny	Paket D	JeepLong	Murah	Deal

Data latih ini akan diproses untuk menentukan nilai yang tertinggi di setiap atribut dengan menggunakan rumus *entropy* dan rumus *gain*. Dari hasil hitungan rumus *enttropy* dan rumus *gain*, atribut dengan nilai *gain* yang tertinggi adalah atribut "Fasilitas" dan "Standart". Langkah berikutnya melakukan uji coba cabang seperti gambar 6 dan Gambar 7. Hasil yang diperoleh adalah atribut nilai tertinggi adalah atribut status memiliki nilai 0,793514 pada node 1.1 adalah atribut status. Data tersebut selanjutnya data tersebut diuji menggunakan aplikasi Orange Data Mining. Langkah pertama adalah menghubungkan file dataset dengan data tabel untuk melihat data mentah dan menghubungkan widget Select Columns untuk menentukan kolom fitur dan target.

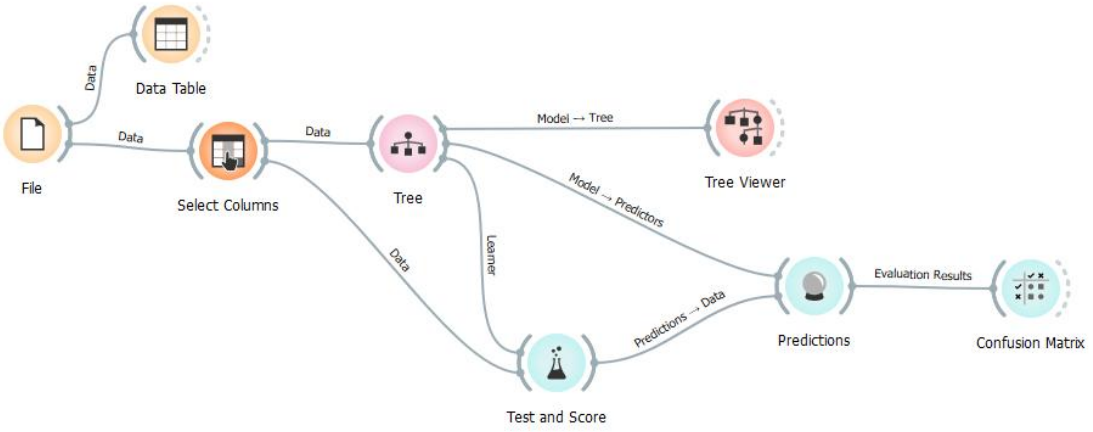
Memasukkan berupa Data, Test Data, Model Tree dan Leaner. Sehingga menghasilkan predictions dan Confusion Matrix Evaluations Results, dapat divisualisasikan seperti Gambar 8.

Node 1		Jumlah Kasus (S)	Mahal (S1)	Sedang (S2)	Murah (S3)	Entropy	Gain
Total		1000	588	295	117	1,332196	
Paket							0,114783
	Paket A	282	33	63	186	1,241263	
	Paket B	285	23	54	208	1,079389	
	Paket C	195	15	65	115	1,26226	
	Paket D	117	9	45	63	1,295738	
	Paket E	70	18	43	9	1,316178	
	Paket F	50	19	25	6	1,39752	
Fasilitas							1,217364
	Standart	868	115	275	478	1,385687	
	JeepLong	132	2	20	110	0,723272	
Status							1,045684
	Deal	814	86	232	496	1,294202	
	Cancel	186	31	63	92	1,462189	

Gambar 6. Node 1

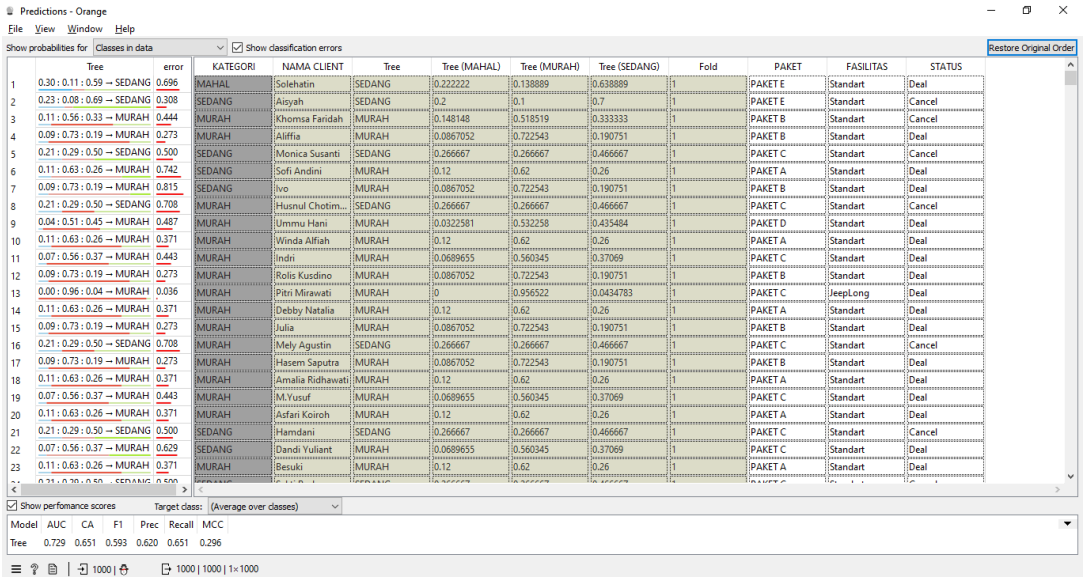
Node 1.1		Jumlah Kasus (S)	Mahal (S1)	Sedang (S2)	Murah (S3)	Entropy	Gain
Total		868	15	275	478	1,100518	
Paket							-0,17529
	Paket A	253	33	65	157	1,314186	
	Paket B	251	23	52	176	1,145549	
	Paket C	164	15	64	85	1,336811	
	Paket D	102	9	44	49	1,340415	
	Paket E	59	17	36	6	1,287491	
	Paket F	39	18	16	5	1,422113	
Status							0,793514
	Deal	707	84	216	407	1,346406	
	Cancel	161	31	59	71	1,509252	

Gambar 7. Node 2

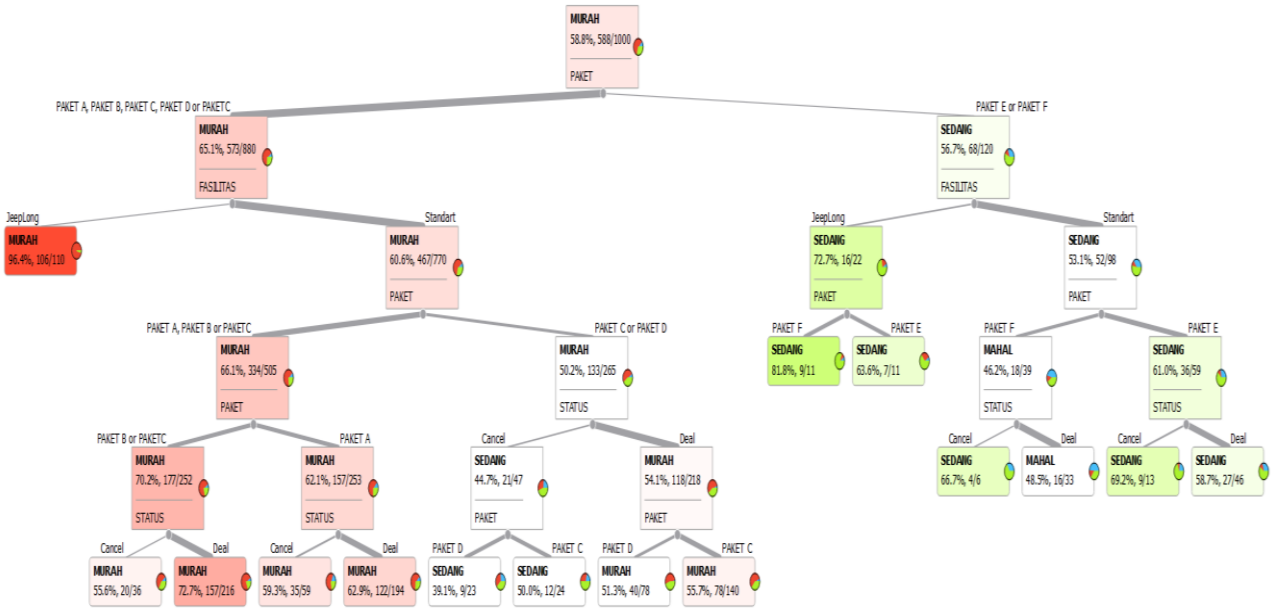


Gambar 8. Widget C4.5

Hasil prediction ini akurasi klasifikasi yang dihitung menggunakan AUC (Area Under Curve), CA (Classification Accuracy), dan Recall. Sehingga diperoleh hasil seperti gambar *9* berikut. Hasil untuk metode C4.5 diperoleh nilai accuracynya: 0.651, Precision : 0.620, dan Recall : 0.651. Dapat diketahui nilai Accuracynya tertinggi dengan nilai 0.651. Sehingga pada Tree Viewer menampilkan pohon keputusan yang dihasilkan seperti gambar dibawah.



Gambar 9. Predictions C4.5



Gambar 10. *Pengujian Pohon Keputusan Menggunakan Aplikasi Orange* C4.5

Pada gambar 4.8 terlihat percabangan dari pohon keputusan dalam visualisasi aplikasi Orange. *Tree* tersebut memiliki ukuran 29 dan jumlah 15 *leaf*, yang artinya terdapat 16 *rule* yang terbentuk diantaranya :

1. *If* Paket = Paket A, Paket B, Paket C, or Paket D, Fasilitas = Jeeplong then kategori = Murah dengan probabilitas 96,4%
2. *If* Paket = Paket A, Paket B, Paket C, or Paket D, Fasilitas = Standart, dengan Paket = Paket A, Paket B or Paket C, Paket = Paket B or Paket C, status = Cancel then kategori = Murah dengan probabilitas 55,6%
3. *If* = Paket A, Paket B, Paket C, or Paket D, Fasilitas = Standart, dengan Paket = Paket A, Paket B or Paket C, Paket = Paket B or Paket C, status = Deal then kategori = Murah dengan probabilitas 72,7%
4. *If* = Paket A, Paket B, Paket C, or Paket D, Fasilitas = Standart, dengan Paket = Paket A, Paket B or Paket C, Paket = Paket A, status = Cancel then kategori = Murah dengan probabilitas 59,3%
5. *If* = Paket A, Paket B, Paket C, or Paket D, Fasilitas = Standart, dengan Paket = Paket A, Paket B or Paket C, Paket = Paket A, status = Deal then kategori = Murah dengan probabilitas 62,9%
6. *If* = Paket A, Paket B, Paket C, or Paket D, Fasilitas = Standart, dengan Paket = Paket C or Paket D, status = Cancel, Paket = Paket D then kategori = Sedang dengan probabilitas 39,1%
7. *If* = Paket A, Paket B, Paket C, or Paket D, Fasilitas = Standart, dengan Paket = Paket C or Paket D, status = Cancel, Paket = Paket C then kategori = Sedang dengan probabilitas 50,0%
8. *If* = Paket A, Paket B, Paket C, or Paket D, Fasilitas = Standart, dengan Paket = Paket C or Paket D, status = Deal, Paket = Paket D then kategori = Murah dengan probabilitas 51,3%
9. *If* = Paket A, Paket B, Paket C, or Paket D, Fasilitas = Standart, dengan Paket = Paket C or Paket D, status = Deal, Paket = Paket C then kategori = Murah dengan probabilitas 55,7%
10. *If* Paket = Paket E or Paket F, Fasilitas = Jeeplong, Paket = F then kategori = Sedang dengan probabilitas 81,8%
11. *If* Paket = Paket E or Paket F, Fasilitas = Jeeplong, Paket = E then kategori = Sedang dengan probabilitas 63,6%
12. *If* Paket = Paket E or Paket F, Fasilitas = Standart, Paket = F, Status = Cancel then kategori = Sedang dengan probabilitas 66,7%
13. *If* Paket = Paket E or Paket F, Fasilitas = Standart, Paket = F, Status = Deal then kategori = Mahal dengan probabilitas 48,5%
14. *If* Paket = Paket E or Paket F, Fasilitas = Standart, Paket = E, Status = Cancel then kategori = Sedang dengan probabilitas 69,2%
15. *If* Paket = Paket E or Paket F, Fasilitas = Standart, Paket = E, Status = Deal then kategori = Sedang dengan probabilitas 58,7%

Confusion matrix dapat digunakan untuk melakukan perhitungan akurasi, presisi dan recall yang dihasilkan oleh aplikasi Orange Data Mining [3]. Sehingga mendapatkan hasil dari perhitungan confusion matrix seperti gambar berikut :

Actual \ Predicted	MAHAL	MURAH	SEDANG	Σ
MAHAL	16	69	32	117
MURAH	5	558	25	588
SEDANG	12	206	77	295
Σ	33	833	134	1000

Gambar 9. Confusion Matric C4.5

Grafik menunjukkan nilai True Positive (TP) sebanyak 16, nilai True Negative (TN) sebanyak 558, 25, 206, dan 77, nilai False Positive (FP) sebanyak 5 dan 12, serta nilai False Negative (FN) sebanyak 69 dan 32. Dengan melakukan perhitungan multiclass (mempunyai 3 kelas) maka menggunakan perhitungan seperti :

$Total\ TP = 16 + 558 + 77 = 651$

$Total\ FP = (69 + 32) + (5 + 25) + (12 + 206) = 349$

$Total\ FN = (5 + 12) + (69 + 206) + (32 + 25) = 349$

$Total\ TN = 866 + 137 + 648 = 1651$

Sehingga perhitungan rumusnya adalah :

$$Accuracy = \frac{(651 + 1651)}{(651 + 349 + 1651 + 349)} x100 = \frac{2302}{3000} x100 = 76{,}7\%$$

$$Precision = \frac{(651)}{(651 + 349)} x100 = \frac{651}{1000} x100 = 65{,}1\%$$

$$Recall = \frac{(651)}{(651 + 349)} x100 = \frac{651}{1000} x100 = 65{,}1\%$$

3.3. Hasil Evaluasi

Penilaian ini sepenuhnya diarahkan untuk menentukan tingkat nilai akurasi dari setiap teknik yang digunakan, sehingga hasilnya dapat diamati dan diteliti untuk setiap klasifikasi. Hasil dari penelitian ini memberikan wawasan baru dalam proses pencarian dari nilai-nilai performa dengan mempertimbangkan atribut-atribut yang digunakan dalam mengklasifikasi. Evaluasi dan validasi menggunakan Confusion Matrix menunjukkan perbedaan akurasi, presisi, dan recall antara metode Naïve Bayes dan C4.5, seperti terlihat pada table3 dibawah ini :

Tabel 3. Perbandingan Hasil

Model Algoritma	Accuracy	Precision	Recall
Naïve Bayes	85,5%	79,8%	79,8%
C4.5	76,7%	65,1%	65,1%

4. KESIMPULAN

Penelitian ini bertujuan untuk mengetahui keakuratan pada paket trip Bromocreative menggunakan algoritma Naïve Bayes dan C4.5, bertujuan untuk mengetahui penerimaan dari kategori paket yang terlaris :

1. Menghasilkan perbandingan dari dua metode dengan menampilkan evaluasi paket trip yang terlaris dari kategori murah, sedang dan mahal.
2. Hasil dari nilai metode Naïve Bayes akurasinya 85,5% dan presisi 79,8%, sedangkan C4.5 hanya memiliki nilai akurasi 76,3% dan presisinya 64,4%. Dari hasil metode Naïve Bayes dan C45 perbandingannya memiliki nilai akurasinya paling tinggi adalah metode Naïve Bayes.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada pihak yang berperan dalam memberikan dukungan, baik secara langsung dan tidak langsung, dan penelitian ini dapat selesai secara baik.

REFERENSI

[1] Ardiansyah, M., Sunyoto, A., & Luthfi, E. T. (2021). Analisis Perbandingan Akurasi Algoritma Naïve Bayes Dan C4. 5 untuk Klasifikasi Diabetes. *Edumatic J. Pendidik. Inform*, *5*(2), 147-156.

[2] Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *Jurnal Media Informatika Budidarma*, *5*(2), 640. https://doi.org/10.30865/mib.v5i2.2937

[3] Azizah, L. (2021). Strategi Pengembangan Pariwisata Dalam Perspektif Islam Menggunakan Metode Analisis SWOT Halal Tourism. *Jurnal Manajemen Dan Inovasi (MANOVA)*, *4*(2), 18–36. https://doi.org/10.15642/manova.v4i2.414

[4] Bahri, S., Midyanti, D. M., & Hidayati, R. (2018, August). Perbandingan Algoritma Naive Bayes dan C4. 5 Untuk Klasifikasi Penyakit Anak. In *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*.

[5] Kohsasih, K. L., Situmorang, Z., & Artikel, I. (2022). Analisis Perbandingan Algoritma C4. 5 dan Naïve Bayes Dalam Memprediksi Penyakit Cerebrovascular. *Jurnal Informatika*, *9*(1), 13-17.

[6] Kusuma, J., Jinan, A., & Situmorang, Z. (2022). Penerapan Decision Tree Algoritma C4.5 Dalam Penentuan Izin Pembongkaran Muatan Kapal. *MEANS (Media Informasi Analisa Dan Sistem)*, *7*(1), 72–76. https://doi.org/10.54367/means.v7i1.1632

[7] Mardiani, E., Rahmansyah, N., Ningsih, S., & ... (2023). Komparasi Metode Knn, Naive Bayes, Decision Tree, Ensemble, Linear Regression Terhadap Analisis Performa Pelajar Sma. *Innovative: Journal Of …*, *3*(2), 13880–13892. http://j-innovative.org/index.php/Innovative/article/view/1949%0Ahttp://j-innovative.org/index.php/Innovative/article/download/1949/1468

[8] Rahmayanti, A., Rusdiana, L., & Suratno, S. (2022). Perbandingan Metode Algoritma C4. 5 Dan Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa. *Walisongo Journal of Information Technology*, *4*(1), 11-22.

[9] Rosandy, T. (2016). Perbandingan Metode Naive Bayes Classifier Dengan Metode Decision Tree (C4. 5) Untuk Menganalisa Kelancaran Pembiayaan (Study Kasus: Kspps/Bmt Al-fadhila. *Jurnal Teknologi Informasi Magister*, *2*(01), 52-62.

[10] Sunanto, N., & Falah, G. (2022). Penerapan Algoritma C4.5 Untuk Membuat Model Prediksi Pasien Yang Mengidap Penyakit Diabetes. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, *7*(2), 208–216. https://doi.org/10.36341/rabit.v7i2.2435

[11] Widaningsih, S. (2020). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, *13*(1), 16–25. https://doi.org/10.36787/jti.v13i1.78

[12] Yulia, D., Pandu Kusuma, A., & Fanny Hebrasianto Permadi, D. (2022). Penerapan Algoritma C4.5 Untuk Prediksi Minat Penjurusan Siswa Di Smkn 1 Kademangan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *6*(2), 893–900. https://doi.org/10.36040/jati.v6i2.5781

[13] Yuliasih, M. (2021). Pemanfaatan Instagram Sebagai Media. *Jurnal Dakwah*, *4*(2), 65–76.

[14] Yunus, Y., & Nurcahyo, G. W. (2023). Perbandingan algoritma c4. 5 dan naive bayes dalam prediksi kelulusan mahasiswa. *Jurnal CoSciTech (Computer Science and Information Technology)*, *4*(1), 193-199.

[15] Zubair, A. (2022). *Data Mining Menggunakan Orange.* Yogyakarta: Teknosain.