



Penerapan algoritma ID3 dan Naïve Bayes untuk memprediksi keputusan konsumen dalam pembelian sports utility vehicle

Implementation of ID3 algorithm and Naïve Bayes to predict consumer decisions in purchasing sports utility vehicles

Julianto Lemantara*

* Universitas Dinamika, Jl. Raya Kedung Baruk No.98, Kedung Baruk, Kec. Rungkut, Surabaya, Jawa Timur 60298

INFORMASI ARTIKEL

Article History:

Submission: 03-12-2024

Revised: 08-12-2024

Accepted: 16-12-2024

Kata Kunci:

Algoritma ID3; Naïve Bayes; Prediksi; Konsumen; SUV.

Keywords:

ID3 Algorithm; Naïve Bayes; Prediction; Consumers; SUV.

*** Korespondensi:**

Julianto Lemantara
julianto@dinamika.ac.id

ABSTRAK

Banyak konsumen yang merasa kebingungan dan membutuhkan pertimbangan saat melakukan pembelian *Sports Utility Vehicle* (SUV). Pertimbangan ini dilakukan agar dapat melakukan pembelian SUV sesuai dengan kebutuhan. Karena itu, penelitian ini membantu konsumen dalam memutuskan untuk membeli mobil SUV atau tidak, didasarkan atas kebutuhan dan beberapa kriteria. Penelitian ini melakukan prediksi untuk konsumen saat membuat keputusan agar yakin dalam melakukan pembelian SUV. Penelitian ini menggunakan metode CRISP-DM yang terdiri atas fase pemahaman bisnis, pemahaman data, pengolahan data, pemodelan dan evaluasi. Sedangkan algoritma yang digunakan dalam penelitian ini yaitu algoritma ID3 dan Naive Bayes dengan menggunakan bantuan tools Weka 3.8.4. Data yang diperoleh berasal dari website kaggle untuk dataset yang diperoleh sejumlah 399 data yang kemudian dibagi menjadi 300 data latih dan 99 data uji. Data tersebut terdiri dari 4 atribut yaitu, *gender*, *age*, *estimated salary* dan *purchased*. Dari pengolahan data kedua algoritma yang digunakan menghasilkan nilai akurasi yang baik. Namun, algoritma ID3 memiliki akurasi yang lebih baik daripada Naive Bayes dengan rincian nilai akurasi sebanyak 80.8% untuk ID3 dan 78.8% untuk Naive Bayes. Karena tingkat akurasinya tergolong tinggi, maka hasil prediksi dari penelitian ini dapat memberikan rekomendasi yang akurat kepada konsumen dalam mengambil keputusan terkait pembelian SUV.

ABSTRACT

Many consumers feel confused and need consideration when making a purchase of a *Sports Utility Vehicle* (SUV). This consideration is done so that they can make an SUV purchase according to their needs. Therefore, this study helps consumers decide whether to buy an SUV or not, based on their needs and several criteria. This study makes predictions for consumers when making decisions to be sure about buying an SUV. This study used the CRISP-DM method, which consists of the phase of business understanding, data understanding, data processing, modeling and evaluation. While the algorithms used in this study were the ID3 and Naive Bayes algorithms using the Weka 3.8.4 tools. The data obtained came from the kaggle website for the dataset obtained as many as 399 data which were then divided into 300 training data and 99 testing data. The data consists of 4 attributes, namely: *gender*, *age*, *estimated salary*, and *purchase*. From data processing, the two algorithms used produced good accuracy values, but the ID3 algorithm has better accuracy than Naive Bayes with details of the accuracy value of 80.8% for ID3 and 78.8% for Naive Bayes. Because the accuracy level is relatively



high, this prediction results can provide accurate recommendations to consumers in making decisions regarding SUV purchases.

1. PENDAHULUAN

SUV adalah singkatan dari *Sports Utility Vehicle* yang memiliki ukuran lebih besar dan lebih tinggi dari mobil biasa dengan atap yang lebih tinggi dan posisi duduk yang memberikan pandangan lebih baik kepada pengemudi. Saat ini SUV tengah digemari banyak orang dari berbagai negara. Hal ini dimulai sejak beberapa tahun yang lalu banyak orang lebih memilih untuk membeli kendaraan jenis SUV khususnya tujuh seater sebagai kendaraan keluarga. Mobil SUV semakin menarik perhatian konsumen dikarenakan desain interior yang luas dan desain eksterior yang tangguh, serta berpadu dengan mesin yang handal sehingga memberikan kenyamanan bagi konsumen [1]. Di Indonesia sendiri, mobil SUV ini sudah menjadi segmen yang memiliki penjualan terbesar di Indonesia pada tahun 2021 [2], [3]. SUV memiliki penjualan terlaris dibandingkan dengan *Low-Cost Green Car* (LCGC) dan *Multi Purpose Vehicle* (MPV). SUV merupakan segmen dengan pertumbuhan terpesat di Indonesia. Tren penjualan SUV juga terus mengalami peningkatan setiap tahun. Peningkatan ini terjadi secara konsisten dari 15 persen pada 2017 menjadi 26 persen pada 2021 [3]. Dari data tersebut menunjukkan bahwa jumlah peminat SUV semakin banyak jika dibandingkan dengan peminat jenis mobil yang lain [4].

Berkaitan dengan minat SUV yang semakin tinggi, maka konsumen mempunyai peranan yang besar dalam memutuskan membeli SUV atau tidak, disesuaikan dengan kebutuhannya. Keputusan saat membeli adalah suatu pemikiran seseorang dengan mengevaluasi bermacam-macam pilihan, lalu membuat keputusan pada suatu pilihan tertentu dari berbagai macam pilihan yang tersedia [5], [6]. Keputusan pembelian merupakan tahapan dari proses pengambilan keputusan dimana konsumen benar-benar melakukan pembelian setelah melakukan berbagai pertimbangan [7]. Keputusan pembelian juga bisa diartikan sebagai penentuan pilihan produk yang dibeli dari dua atau lebih pilihan [8]. Keputusan pembelian merupakan proses konsumen mengenali masalah, lalu mencari informasi dan mengevaluasi tiap-tiap alternatif yang dapat memecahkan masalah, yang kemudian mengarah pada keputusan pembelian [9]. Adapun struktur keputusan pembelian adalah keputusan mengenai penjual, merek, bentuk produk, jenis produk, jumlah produk, waktu pembelian, dan cara pembayaran [10].

Berdasarkan kondisi yang terjadi diperlukan sebuah analisis ilmiah yang menunjukkan perbandingan atas keputusan dari konsumen untuk melakukan pembelian SUV dengan menggunakan metode klasifikasi Naïve Bayes dan *Iterative Dichotomizer 3* (ID3). Dalam melakukan analisis ini peneliti menerapkan dua metode. Yang pertama, metode Naïve Bayes dipilih karena mampu melakukan prediksi peluang untuk masa depan berdasarkan pengalaman pada masa sebelumnya [11]–[13]. Yang kedua adalah metode ID3 dipilih dengan alasan algoritma ini melakukan pencarian secara menyeluruh pada semua kemungkinan pohon keputusan sehingga lebih komprehensif [14].

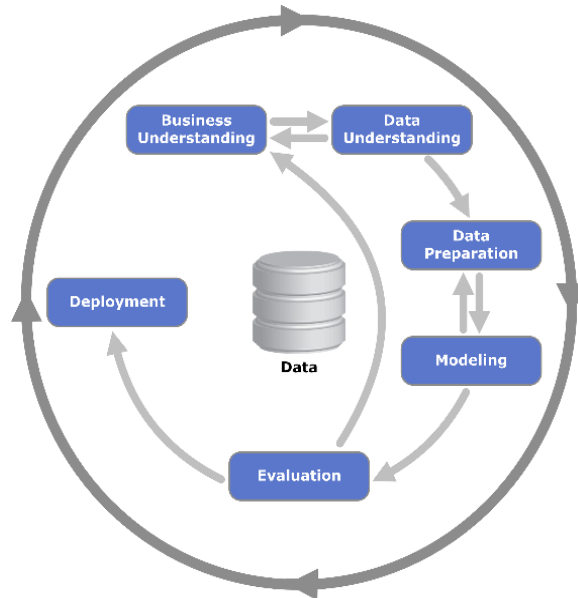
Beberapa penelitian sejenis telah dilakukan menggunakan metode data mining dengan jumlah data dan jumlah atribut yang berbeda-beda. Penelitian pernah dilakukan pada tahun 2016 tentang penentuan minat konsumen membeli mobil menggunakan metode ID3 di Daihatsu kota Jakarta [15]. Penelitian tersebut mengolah 100 data dengan 13 atribut, menggunakan aplikasi *Rapidminer*. Hasil dari penelitian tersebut mengemukakan bahwa tingkat akurasi sebesar 90% dalam menentukan minat konsumen membeli mobil Daihatsu. Beberapa penelitian yang terkait dengan keputusan pembelian mobil hanya menggunakan satu metode saja yaitu ID3. Sedangkan dalam penelitian ini menggunakan dua metode yaitu ID3 dan Naïve Bayes, dengan bantuan aplikasi WEKA.

Penelitian terdahulu yang lain dilakukan pada tahun 2022 terkait prediksi segmentasi pelanggan pada penjualan mobil menggunakan Naïve Bayes dan ID3 [16]. Hasil penelitian tersebut menunjukkan ID3 memiliki tingkat akurasi sebesar 96,43%, sedangkan Naïve Bayes memiliki tingkat akurasi sebesar 56,43%. Fokus penelitian tersebut terletak pada klasifikasi segmentasi pelanggan, bukan pada klasifikasi keputusan konsumen dalam pembelian mobil. Jika

dibandingkan dengan penelitian terdahulu, penelitian ini memiliki kesamaan dalam metode data mining. Namun, penelitian ini memiliki keunggulan yaitu lebih spesifik terkait pembelian mobil dengan jenis/tipe SUV. Hal ini yang merupakan kekhususan dan keterbaharuan yang diangkat dalam penelitian ini, dibandingkan dengan berbagai penelitian-penelitian sebelumnya.

2. METODE

Tahapan yang diterapkan dalam pelaksanaan penelitian ini yakni *Cross Industry Standard Process for Data Mining* (CRISP-DM). Di dalam metode CRISP-DM terdapat 6 fase utama, yang tampak pada [Gambar 1 \[17\]](#).



Gambar 1. Tahapan CRISP-DM

Tahapan yang muncul di [Error! Reference source not found.](#) bisa dijelaskan lebih detail sebagai berikut:

Fase pemahaman bisnis (business understanding phase)

Hasil utama dari tahap ini adalah pemahaman atas tujuan dari bisnis, situasi dan penerjemahan semua hal ke dalam data mining. Hal ini dilakukan dengan memenuhi kebutuhan seperti mengetahui variabel atau faktor-faktor yang mempengaruhi keputusan konsumen dalam pembelian SUV. Hal ini bertujuan agar bisa memberikan rekomendasi kepada konsumen untuk melakukan pembelian mobil SUV atau tidak, sesuai dengan kebutuhan dan kemampuan konsumen.

Fase pemahaman data (data understanding phase)

Dalam tahapan ini terdapat proses pengumpulan data yang diperoleh dari website kaggle, dengan alamat di: <https://www.kaggle.com/datasets/iamaniket/suv-data/data>. Dataset yang diperoleh berupa dokumen excel sejumlah 399 data. Data tersebut terdiri dari 4 atribut yaitu, *gender*, *age*, *estimated salary*, dan *purchased*. Data ini merupakan faktor pertimbangan bagi konsumen dalam membeli kendaraan SUV yang sudah disesuaikan dengan kategori yang telah dibuat.

Fase persiapan data (data preparation phase)

Dalam tahapan ini dilakukan pemilihan data dan memastikan data yang telah diperoleh telah layak untuk dilakukan proses pengolahan. Data-data yang diambil disesuaikan lagi untuk setiap atribut yang ada dan dilakukan pengklasifikasian untuk setiap data yang tercantum sesuai dengan

atribut. Contohnya: atribut *age* diklasifikasikan ke 4 kelompok, yaitu: ≤ 25 , 26-40, 41-55, dan ≥ 56 (satuan: tahun). Untuk atribut *estimated salary* diklasifikasikan ke dalam 4 kelompok, yaitu: 15.000 – 25.000, 26.000-50.000, 51.000-100.000, dan 100.000-150.000. Pengklasifikasian nilai atribut ini dilakukan agar dapat memperkecil ruang lingkup dari nilai atribut yang ada sehingga mempercepat proses komputasi Naïve Bayes dan ID3. Untuk lebih jelasnya, hasil pengklasifikasian atribut dapat dilihat di [Tabel 1](#). Pada data yang tersedia, dilakukan pula penghapusan pada satu data yang tidak bisa terklasifikasikan (*unclassified*) oleh aplikasi Weka.

Tabel 1. Atribut pembelian SUV

Atribut	Nilai
Gender	1. Male 2. Female
Age	1. ≤ 25 2. 26-40 3. 41-55 4. ≥ 56
Estimated Salary	1. 15.000-25.000 2. 26.000-50.000 3. 51.000-100.000 4. 101.000-150.000
Purchase SUV	1. Yes 2. No

Fase pemodelan (*modeling phase*)

Dua model yang digunakan di penelitian adalah Naïve Bayes dan ID3. Kedua metode ini digunakan untuk melakukan prediksi klasifikasi keputusan pelanggan dalam pembelian SUV. Adapun *tools* yang digunakan di dalam studi ini adalah Weka 3.8.4 dan model validasi klasifikasi yang diterapkan adalah *Supplied Test Set*. Model validasi ini digunakan untuk mengukur seberapa bagus hasil prediksi dari data training. Jumlah data yang dipakai pada studi ini sebanyak 399 data, dimana data dipecah menjadi 300 data latih dan 99 data uji. Setelah melakukan klasifikasi data di *Microsoft Excel*, selanjutnya melakukan ekstraksi pada tipe *file* menjadi CSV untuk dilakukan proses training pada data menggunakan Weka.

Langkah-langkah yang dilakukan pada Weka untuk metode ID3 adalah buka aplikasi Weka, klik *open file*, lalu pilih *file*, pada tipe *file* diganti CSV. Hal ini berfungsi untuk melakukan filterisasi pada tampilan tipe *file* yang diperlukan. Karena menggunakan metode ID3 maka pilih tombol *clasify*, kemudian pilih *choose*, pilih *tree*, lalu klik ID3, klik *supplied test set* lalu pada set pilih data yang akan dilakukan pengujian. Selanjutnya klik *more options*, pada *output prediction* diganti dengan *plain text* klik *start*, kemudian muncul hasil sesuai dengan metode yang diinginkan. Selanjutnya, untuk metode *Naïve Bayes*, langkah-langkah yang harus dilakukan adalah pilih tombol *clasify*, kemudian pilih *choose*, pilih *bayes*, lalu klik *Naïve bayes simple*, klik *supplies test set* lalu pada set pilih data yang akan dilakukan pengujian. Selanjutnya pada *more options*, pada *output prediction* diganti *plain text* klik *start*, kemudian muncul hasil sesuai dengan metode yang diinginkan.

Naïve Bayes

Naïve bayes merupakan sebuah metode perhitungan statistik yang mengkalkulasi probabilitas berdasarkan kesamaan antara kasus lama dalam basis data dengan kasus baru. Naïve Bayes dikenal karena tingkat akurasi yang tinggi serta kecepatannya baik saat diterapkan pada basis data berskala besar. Naïve Bayes diakui sebagai salah satu algoritma yang cepat dan sederhana dalam penerapannya [18]. Kelebihan lain dari Naïve Bayes tidak memerlukan jumlah data yang besar dalam melakukan klasifikasi [18]–[20]. Naïve Bayes termasuk *supervised learning*, sehingga membutuhkan data *training* pada proses pembelajaran untuk melakukan pengambilan keputusan [21]. Dalam proses pengklasifikasian, nilai probabilitas dari tiap-tiap label kelas yang

ada dihitung terhadap masukan yang diberikan. Label kelas dengan nilai probabilitas tertinggi akan dijadikan sebagai label kelas untuk data masukan tersebut. Naïve Bayes adalah metode yang sederhana karena dapat memperkecil kompleksitas komputasi menjadi perkalian sederhana dari sebuah probabilitas. Di samping itu, Naïve Bayes juga sanggup mengelola dataset yang mempunyai jumlah atribut yang banyak. Persamaan Naïve Bayes dapat pada Persamaan 1.

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Dimana:

X	= data sampel dengan label/kelas yang belum diketahui
H	= hipotesis bahwa data X adalah suatu label/kelas tertentu
P(H X)	= peluang hipotesis H didasarkan pada kondisi X (posteriori probability)
P(X H)	= peluang X didasarkan pada kondisi pada hipotesis H
P(H)	= peluang hipotesis H (prior probability)
P(X)	= peluang data sampel yang digunakan

Agar nilai probabilitas pada perhitungan Naive Bayes tidak menghasilkan 0 karena tidak ada data pada kategori tertentu dalam suatu kelas, maka teknik estimasi yang dapat digunakan yaitu *Laplace Correction* atau *Laplace Estimator* dengan rumus pada Persamaan 2. Teknik *Laplace Correction* mempunyai hasil yang lebih bagus daripada teknik-teknik lainnya [22].

$$P(H|X) = \frac{P(X|H) \cdot P(H) + 1}{P(X) + V} \quad (2)$$

Dimana:

X	= data sampel dengan label/kelas yang belum diketahui
H	= hipotesis bahwa data X adalah suatu label/kelas tertentu
P(H X)	= peluang hipotesis H didasarkan pada kondisi X (posteriori probability)
P(X H)	= peluang X didasarkan pada kondisi pada hipotesis H
P(H)	= peluang hipotesis H (prior probability)
P(X)	= peluang data sampel yang digunakan
V	= Jumlah kelas/label

Algoritma Iterative Dichotomiser 3 (ID3)

ID3 tergolong dalam algoritma pembelajaran pohon yang paling mendasar [23]. Algoritma ini mencari semua kemungkinan pada pohon keputusan secara menyeluruh [24]. Pohon keputusan yang dibentuk menggunakan algoritma ID3, dilakukan dengan dua nilai penentu, yaitu nilai *entropy* dan nilai *information gain* dari masing-masing variabel. *Entropy* berfungsi mengukur jumlah informasi yang terdapat pada suatu atribut. Sementara itu, *information gain* berfungsi mengukur seberapa baik suatu atribut memisahkan data pelatihan ke dalam suatu kelas target. Pemilihan atribut dilakukan berdasarkan nilai *information gain* [16]. Atribut yang mempunyai *information gain* terbesar akan dipilih dalam algoritma ini. Algoritma ID3 mampu menangani kasus pada pelbagai bidang, diantaranya: kesehatan, pendidikan, pertanian, penjualan retail, dan lain-lain. Dalam rangka mendefinisikan *information gain*, pertama-tama digunakan teori yang disebut *entropy*, dengan persamaan 3.

$$Entropi (S) = \sum_{j=1}^k - P_j \log_2 P_j \quad (3)$$

Dimana:

S	= himpunan dari dataset
k	= banyaknya kelas/label/label
P _j	= peluang yang diperoleh dari Sum (Y _a) dibagi dengan banyaknya kasus

Entropy(S) bernilai 0, bila semua contoh dalam himpunan S ada di kelas yang sama.

Entropy(S) bernilai 1, bila contoh positif dan negatif dalam himpunan S berjumlah sama.

$0 < Entropy(S) < 1$, bila contoh positif dan negatif dalam himpunan S jumlahnya berbeda.

Setelah memperoleh nilai *entropy*, selanjutnya atribut dipilih dengan mengacu pada nilai *information gain* yang terbesar. Untuk memperoleh nilai *information gain* dapat menggunakan Persamaan 4.

$$Gain(A) = Entropi(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropi(S_i) \quad (4)$$

Dimana:

- A = atribut
- S = data sampel yang dipakai untuk proses pelatihan
- [S] = jumlah keseluruhan data sampel
- [Si] = jumlah sampel untuk nilai V
- Entropi (Si) = nilai entropi untuk data sampel yang mempunyai nilai i

Fase evaluasi (*evaluation phase*)

Evaluasi terhadap hasil klasifikasi dengan metode ID3 dan Naïve Bayes dilakukan menggunakan *confusion matrix*. Dengan *confusion matrix* akan dapat terlihat nilai *accuracy*, *recall*, dan *precision* dari kedua metode yang digunakan.

Fase penyebaran (*deployment phase*)

Pada fase *deployment*, model algoritma ID3 dan naïve bayes diterapkan untuk memberikan rekomendasi kepada konsumen dalam memutuskan pembelian SUV atau tidak.

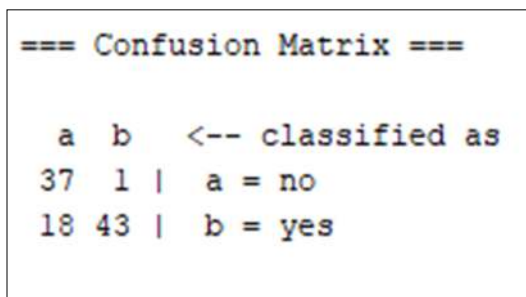
3. HASIL DAN PEMBAHASAN

Pada [Tabel 2](#), penelitian ini memecah data ke dalam 2 jenis, yaitu: a) Data latih, sejumlah 300 data, b) Data uji, sejumlah 99 data.

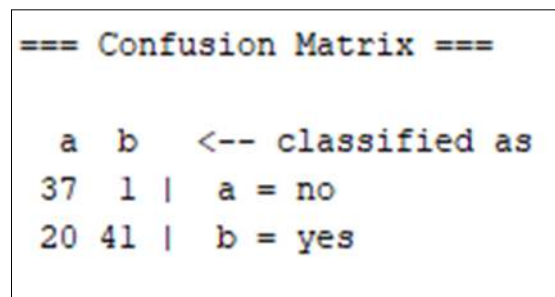
Tabel 2. Hasil pemecahan data training dan testing

Kelas Klasifikasi	Total Data <i>Training</i>	Total Data <i>Testing</i>	Total Seluruh Data
No	219	38	257
Yes	81	61	142

Hasil klasifikasi pada penelitian ini digunakan untuk menentukan keputusan customer dalam pembelian SUV. Hasil klasifikasi akan ditampilkan dengan bentuk *confusion matrix*. Dengan menggunakan tools WEKA, maka diperoleh *confusion matrix* untuk metode ID3 seperti yang tampak pada [Gambar 2](#). Sementara itu, *confusion matrix* untuk metode Naïve Bayes dapat ditunjukkan seperti pada [Gambar 3](#).



Gambar 2. *Confusion Matrix* – ID3



Gambar 3. *Confusion Matrix* – Naïve Bayes

Untuk perbandingan hasil akurasi dari kedua metode di [Tabel 3](#). Nilai akurasi dari metode ID3 ternyata lebih tinggi 2% daripada metode Naïve Bayes. ID3 memiliki nilai akurasi sebesar 80.8%, sedangkan Naïve Bayes memiliki nilai akurasi sebesar 78.8%

Tabel 3. Hasil akurasi metode ID3 dan Naïve Bayes

Metode	Akurasi
ID 3	80.8 %

Naïve Bayes 78.8 %

Selain dari tingkat akurasi, kualitas hasil klasifikasi juga dapat ditinjau dari nilai *recall* dan nilai presisi. Presisi menunjukkan probabilitas sebuah item yang terpilih adalah relevan. Sementara itu, *recall* menunjukkan rasio dari item relevan yang dipilih terhadap total keseluruhan item yang relevan. Pada [Tabel 4](#) menunjukkan bahwa dengan menggunakan metode ID3 nilai *precision* No yaitu 0.673 dan Yes yaitu 0.977 sedangkan nilai *recall* No yaitu 0.974 dan Yes 0.705. Dengan menggunakan algoritma Naïve bayes nilai *precision* No yaitu 0.649 dan Yes 0.976 sedangkan nilai *recall* No yaitu 0.974 dan Yes 0.672.

Tabel 4. Nilai presisi dan recall

Kelas/Label	ID3		Naïve Bayes		
	Klasifikasi	Presisi	Recall	Presisi	Recall
No		0.673	0.974	0.649	0.974
Yes		0.977	0.705	0.976	0.672

Jika ditinjau dari [Tabel 4](#), maka metode ID3 secara umum juga mempunyai nilai presisi dan nilai *recall* yang lebih bagus dibandingkan dengan Naïve Bayes. Untuk hasil klasifikasi dengan Naïve Bayes pada aplikasi Weka di [Gambar 4](#).

```

Class no: P(C) = 0.72847682

Attribute Gender
Male   Female
0.51131222   0.48868778

Attribute Age
<=25   26 - 40  41 - 55  >=56
0.22421525   0.64125561   0.12107623   0.01345291

Attribute Estimated Salary
15000 - 25000   26000 - 50000   51000 - 100000   101000 - 150000
0.13004484   0.22421525   0.59641256   0.04932735

Class yes: P(C) = 0.27152318

Attribute Gender
Male   Female
0.45783133   0.54216867

Attribute Age
<=25   26 - 40  41 - 55  >=56
0.01176471   0.35294118   0.47058824   0.16470588

Attribute Estimated Salary
15000 - 25000   26000 - 50000   51000 - 100000   101000 - 150000
0.08235294   0.15294118   0.25882353   0.50588235

```

Gambar 4. Hasil klasifikasi Naïve Bayes pada Aplikasi Weka

Hasil klasifikasi Naïve Bayes pada aplikasi Weka sudah sesuai dengan perhitungan secara manual.

$P(C_i)$

$P(\text{Purchased} = \text{"no"}) = 219/300 = 0,73$

$P(X|C_i)$

$P(\text{Gender} = \text{"male"} \mid \text{purchased} = \text{"no"}) = 112/219 = 0,5114155$

$P(\text{Gender} = \text{"female"} \mid \text{purchased} = \text{"no"}) = 107/219 = 0,4885845$

Hasil perhitungan manual menggunakan metode ID3 pada [Tabel 5](#) dan [Tabel 6](#).

Tabel 5. Hasil perhitungan entropy total dengan ID3

Total Kasus	Sum (No)	Sum (Yes)	Entropi Total
300	219	81	0,841464636

Tabel 6. Hasil perhitungan information gain dengan ID3

Node	Atribut	Nilai	Sum (Nilai)	Sum (No)	Sum (Yes)	Entropi	Gain
1	Gender	Male	149	112	37	0,808608	0,001699
		Female	151	107	44	0,870511	
	Age	<=25	49	49	0	0	0,228409
		26-40	171	142	29	0,656767	
		41-55	65	26	39	0,970951	
		>=56	15	2	13	0,56651	
		Estimated Salary	15000-25000	34	28	6	
		26000-50000	61	49	12	0,715322	
		51000-100000	153	132	21	0,577004	
		101000-150000	52	10	42	0,706274	

Nilai *information gain* dari atribut *Age* tertinggi, maka atribut *Age* dipilih menjadi "root". Untuk pembuatan pohon keputusan dengan algoritma ID3 yang lebih lengkap, bisa dilihat melalui [Gambar 5](#) dan [Gambar 6](#). Kedua gambar ini merupakan hasil perhitungan lebih lanjut dari ID3 dengan menggunakan *tools* WEKA.

```

Age = <=25: no
Age = 26 - 40
| Estimated Salary = 15000 - 25000: no
| Estimated Salary = 26000 - 50000: no
| Estimated Salary = 51000 - 100000
| | Gender = Male: no
| | Gender = Female: no
| Estimated Salary = 101000 - 150000
| | Gender = Male: yes
| | Gender = Female: yes
Age = 41 - 55
| Estimated Salary = 15000 - 25000
| | Gender = Male: yes
| | Gender = Female: yes
| Estimated Salary = 26000 - 50000
| | Gender = Male: yes
| | Gender = Female: yes
| Estimated Salary = 51000 - 100000
| | Gender = Male: no
| | Gender = Female: no
| Estimated Salary = 101000 - 150000
| | Gender = Male: yes
| | Gender = Female: yes
    
```

Gambar 5. Hasil klasifikasi metode ID3 pada aplikasi weka


```

Age = >=56
| Estimated Salary = 15000 - 25000: null
| Estimated Salary = 26000 - 50000
| | Gender = Male: yes
| | Gender = Female: no
| Estimated Salary = 51000 - 100000
| | Gender = Male: yes
| | Gender = Female: yes
| Estimated Salary = 101000 - 150000: yes

```

Gambar 6. Kelanjutan hasil klasifikasi metode ID3 pada aplikasi weka

4. SIMPULAN

Dengan mengacu pada hasil perbandingan antara algoritma ID3 dan Naïve Bayes untuk memprediksi keputusan konsumen dalam pembelian SUV, dapat disimpulkan ID3 memiliki nilai akurasi 80,8% sedangkan metode Naïve Bayes memiliki nilai akurasi sebesar 78,8%. Hasil akurasi dari kedua algoritma memiliki nilai yang cukup besar tetapi lebih baik menggunakan algoritma ID3 dengan perbedaan 2% dengan Naïve Bayes. Karena tingkat akurasinya tergolong tinggi, maka bisa disimpulkan bahwa hasil penelitian ini dapat memberikan rekomendasi yang akurat kepada konsumen dalam mengambil keputusan terkait pembelian SUV. Penelitian ke depan akan lebih baik jika perbandingan bisa melibatkan lebih dari 2 metode sehingga bisa mendapatkan tingkat akurasi yang lebih baik dan lebih beragam bagi pengguna.

REFERENSI

- [1] Y. Suryanto, "Daftar 15 SUV Tangguh yang Siap Menaklukkan Segala Medan," *Merdeka.Com*, 2024.
- [2] N. Yuniar, "Tren penjualan SUV kian meningkat, ini kelebihanannya," *Antara News*, 2022.
- [3] CNN, "Minat Konsumen Bergeser, SUV Kini Jadi Favorit Tekuk LCGC dan MPV," *CNN Indonesia*, 2022.
- [4] N. R. Thadsyah and R. L. Batu, "Pengaruh Harga dan Kualitas Produk Terhadap Keputusan Pembelian Mobil Toyota Fortuner di Karawang," *EKOMBIS Rev. J. Ilm. Ekon. dan Bisnis*, vol. 10, no. 2, pp. 947–958, 2022, doi: 10.37676/ekombis.v10i2.2284. <https://doi.org/10.37676/ekombis.v10i2.2284>
- [5] I. R. Puspita, A. Fadillah, and Y. Taqyudin, "Tinjauan Atas Keputusan Pembelian Pada Marketplace Shopee," *J. Apl. Bisnis Kesatuan*, vol. 2, no. 1, pp. 67–74, 2022, doi: 10.37641/jabkes.v2i1.1358. <https://doi.org/10.37641/jabkes.v2i1.1358>
- [6] A. C. Devi and U. M. Fadli, "Analisis Keputusan Pembelian Produk E-Commerce Shopee Di Kalangan Mahasiswa Universitas Buana Perjuangan Karawang," *J. Student Res.*, vol. 1, no. 5, pp. 113–123, 2023. <https://doi.org/10.31539/budgeting.v5i1.7372>
- [7] L. P. Panji Ragatirta and E. Tiningrum, "Pengaruh Atmosphere Store, Desain Produk, Dan Citra Merek Terhadap," *Excell. J. Manajemen, Bisnis, dan Pendidik.*, vol. 7, no. 2, pp. 143–152, 2020, [Online]. Available: <https://e-journal.stie-aub.ac.id/index.php/excellent%0Ahttps://doi.org/10.36587/exc.v7i2.793>
- [8] W. Alawiah and A. P. Utama, "Pengaruh Kualitas Produk, Harga, dan Citra Merek pada Keputusan Pembelian Mobil," *Kaji. Brand. Indones.*, vol. 5, no. 1, p. 17, 2023, doi: 10.21632/kbi.5.1.17-34. <https://doi.org/10.21632/kbi.5.1.17-34>
- [9] G. V. M. Tua, Andri, and I. M. Andariyani, "Pengaruh Kualitas Produk Dan Harga Terhadap Keputusan Pembelian Mesin Pompa Air Submersible Dab Decker Di Cv. Citra Nauli Electricsindo Pekanbaru," *J. Ilm. Multidisiplin*, vol. 1, no. 04, pp. 140–154, 2022, doi: 10.56127/jukim.v1i04.266. <https://doi.org/10.56127/jukim.v1i04.266>
- [10] S. P. Sari, "Hubungan Minat Beli Dengan Keputusan Pembelian Pada Konsumen," *Psikoborneo J. Ilm. Psikol.*, vol. 8, no. 1, p. 147, 2020, doi: 10.30872/psikoborneo.v8i1.4870.

- <https://doi.org/10.30872/psikoborneo.v8i1.4870>
- [11] W. A. Dewa, J. Maknulah, and A. D. Putri, "Penerapan Metode Naïve Bayes untuk Menentukan Pengajuan Polis Baru pada PT. 'XYZ,'" *J. Ilm. Komputasi*, vol. 20, no. 1, pp. 83–91, 2021, doi: 10.32409/jikstik.20.1.2696. <https://doi.org/10.32409/jikstik.20.1.2696>
- [12] M. F. Rifai, H. Jatnika, and B. Valentino, "Penerapan Algoritma Naïve Bayes pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS)," *Petir J. Pengkaj. dan Penerapan Tek. Inform.*, vol. 12, no. 2, pp. 131–144, 2019, doi: 10.33322/petir.v12i2.471. <https://doi.org/10.33322/petir.v12i2.471>
- [13] A. Z. Macfud, A. P. Kusuma, and W. D. Puspitasari, "Analisis Algoritma Naive Bayes Classifier (NBC) pada Klasifikasi Tingkat Minat Barang di Toko Violet Cell," *J. Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 87–94, 2023. <https://doi.org/10.36040/jati.v7i1.5692>
- [14] J. T. M. A. Nazanah and M. I. Jambak, "Pemanfaatan Algoritma Decision Tree ID3 Bagi Manajemen Bimbel Untuk Menentukan Faktor Kelulusan Pada Sekolah Kedinasan," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 6, pp. 915–924, 2023, doi: 10.30865/klik.v3i6.791.
- [15] W. W. Arietya, W. Supriyatin, and I. Astuti, "Penentuan Minat Konsumen Dalam Membeli Mobil Menggunakan Algoritma Id3 Studi Kasus Daihatsu," in *Konferensi Nasional Teknologi Informasi dan Aplikasinya*, 2016, pp. 25–30.
- [16] J. Lemantara, "Penerapan Algoritma Naïve Bayes dan ID3 untuk Memprediksi Segmentasi Pelanggan pada Penjualan Mobil," *J. Technol. Informatics*, vol. 4, no. 1, pp. 31–40, 2022, doi: 10.37802/joti.v4i1.265. <https://doi.org/10.37802/joti.v4i1.265>
- [17] E. Anderson and F. F. Adiwijaya, "Penerapan Data Mining Asosiasi Genre Game Untuk Market Research," *J. Penelit. Mhs. Tek. dan Ilmu Komput.*, vol. 2, no. 2, pp. 77–84, 2022, doi: 10.34010/jupiter.v2i2.8676. <https://doi.org/10.34010/jupiter.v2i2.8676>
- [18] D. Sartika and R. Gustriansyah, "Comparison of Naive Bayes and Decision Tree Algorithms to Assess The Performance of Palembang City fire and Disaster Management Employees," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 11, no. 1, pp. 132–138, 2024, doi: 10.37373/tekno.v11i1.843. <https://doi.org/10.37373/tekno.v11i1.843>
- [19] J. Lemantara, "Rancang Bangun Aplikasi Hipertensi.edu sebagai Media Edukasi dan Diagnosis Penyakit Hipertensi Menggunakan metode Naïve Bayes dengan Laplace Correction," *INFOTECH J. Inform. Teknol.*, vol. 5, no. 1, pp. 146–160, 2024, doi: 10.37373/infotech.v5i1.1197. <https://doi.org/10.37373/infotech.v5i1.1197>
- [20] R. Achmad and A. S. Girsang, "Implementation of naive bayes classifier algorithm in classification of civil servants," *J. Phys. Conf. Ser.*, vol. 1485, no. 1, 2020, doi: 10.1088/1742-6596/1485/1/012018. <https://doi.org/10.1088/1742-6596/1485/1/012018>
- [21] G. Firmansyah and A. Hermawan, "Implementasi Algoritma Naive Bayes Untuk Klasifikasi Kesehatan Buah Jeruk," *J. Inform.*, vol. 10, no. 2, pp. 180–184, 2023, doi: 10.31294/inf.v10i2.16115. <https://doi.org/10.31294/inf.v10i2.16115>
- [22] I. Listiowarni, "Implementasi Naïve Bayesian dengan Laplacian Smoothing untuk Peminatan dan Lintas Minat Siswa SMAN 5 Pamekasan," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 8, no. 2, pp. 124–129, 2019, doi: 10.32736/sisfokom.v8i2.652. <https://doi.org/10.32736/sisfokom.v8i2.652>
- [23] N. A. Fatmandini, R. A. Saputra, and R. Yulistria, "Komparasi Criteria Splitting Pada Algoritma Iterative Dichotomizer 3(ID3) Untuk Klasifikasi Kelayakan Kredit," *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 1, pp. 79–84, 2020, doi: 10.31294/p.v22i1.6711. <https://doi.org/10.31294/p.v22i1.6711>
- [24] Nardiono and J. Sanjay, "Penerapan Algoritma Iterative Dichotomiser 3 (Id3) Pada Sistem Cerdas Identifikasi Faktor (Studi Kasus : Pt . Global Jet Cargo)," *J. Jar. Sist. Inf. Robot.*, vol. 7, no. 2, pp. 222–231, 2023.