

KLASIFIKASI AKSEPTABILITAS MOBIL MENGGUNAKAN ALGORITMA NAIVE BAYES

Laurentz Anne Safilla¹, Gusmelia Testiana²

^{1,2}Jurusan Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Raden Fatah Palembang

Jl. Pangeran Ratu 15 Ulu Palembang Sumatera Selatan

[1laurentzanne6@gmail.com](mailto:laurentzanne6@gmail.com), [2gusmeliatestiana_uin@radenfatah.ac.id](mailto:gusmeliatestiana_uin@radenfatah.ac.id)

Abstrak

Dengan industri otomotif yang berkembang pesat, persaingan antar produsen mobil semakin ketat. Untuk memenangkan persaingan ini, pemahaman yang mendalam tentang preferensi konsumen dan mobil sangatlah penting. Oleh karena itu, dapat dilakukan penelitian untuk mengidentifikasi atribut-atribut yang mempengaruhi akseptabilitas mobil. Pengklasifikasian dilakukan dengan menggunakan algoritma Naïve Bayes. Hasil penelitian menunjukkan bahwa algoritma klasifikasi Naïve Bayes efektif dalam mengklasifikasi penerimaan kendaraan. Menggunakan kombinasi atribut yang relevan, model klasifikasi Naïve Bayes dapat mengidentifikasi mobil yang dapat diterima dan tidak dapat diterima dengan akurasi yang wajar.

Kata Kunci: Naïve Bayes, Data Mining, RapidMiner, Klasifikasi

Abstract

With the automotive industry growing rapidly, competition among car manufacturers is intensifying. To win this competition, a deep understanding of consumer preferences and cars is essential. Therefore, research can be conducted to identify attributes that affect car acceptability. Classification using the Naïve Bayes algorithm. The results show that the Naïve Bayes classification algorithm is effective in classifying vehicle acceptance. Using a combination of relevant attributes, the Naïve Bayes classification model can identify acceptable and unacceptable cars with reasonable accuracy.

Keyword: Naïve Bayes, Data Mining, RapidMiner, Classification

1. Pendahuluan

Kehidupan masyarakat modern saat ini berdampak pada perubahan pola perilaku masyarakat dalam melakukan pembelian, terutama untuk mobil (Bimantara et al., 2022). Permintaan mobil terus tumbuh seiring dengan pertumbuhan ekonomi dan perkembangan teknologi. Dalam industri otomotif, penting untuk memahami preferensi konsumen dalam hal fitur dan informasi teknis dari mobil yang mereka cari. Untuk memenangkan persaingan ini, pemahaman yang mendalam tentang preferensi konsumen dan mobil sangatlah penting. Pabrik dan

perancang mobil harus mengidentifikasi faktor-faktor yang mempengaruhi penerimaan mobil untuk menghasilkan produk yang memenuhi kebutuhan dan harapan pelanggan. Klasifikasi penerimaan mobil sulit dilakukan karena faktor-faktor yang mempengaruhinya sangat kompleks dan subyektif. Atribut-atribut seperti harga beli, harga perawatan, jumlah pintu, kapasitas orang dalam mobil, ukuran bagasi, dan keamanan. Oleh karena itu, dapat dilakukan penelitian untuk mengidentifikasi atribut-atribut yang mempengaruhi akseptabilitas mobil

Data mining adalah serangkaian proses untuk menggali

nilai tambah dalam suatu bentuk informasi yang tidak diketahui secara manual dari database (Waluyo, 2017). Informasi tentang itu dibuat dengan merekam dan menggali spesimen penting atau menggunakan informasi yang terdapat dalam database (Fadlan, Ningsih, & Windarto, 2018).

Dalam konteks ini, klasifikasi penerimaan kendaraan menjadi topik penelitian yang menarik. Klasifikasi mobil melibatkan pengelompokan mobil menjadi kategori diterima atau diterima berdasarkan karakteristiknya. Algoritma Naïve Bayes dapat menjadi pendekatan yang efektif untuk klasifikasi ini. Algoritma Naïve Bayes adalah teknik peramalan probabilistik sederhana berdasarkan penerapan teorema Bayes (*rule of Bayes*) dengan asumsi independensi (mandiri) kuat (naif).

Di algoritma Naïve Bayes, independensi fitur yang kuat adalah fitur dalam data yang tidak terkait ada tidaknya karakteristik lain dalam data yang sama (Tarigan, 2023). Data yang digunakan bersumber dari situs penyedia data publik, yaitu Kaggle dengan jumlah 1707 record, 6 atribut dan class atau labelnya.

2. Kajian Pustaka

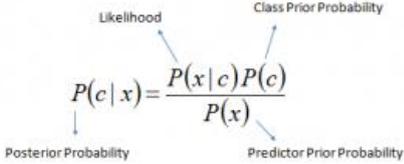
Data Mining

Data mining adalah suatu ilmu yang mempelajari alur kerja pengalian data atribut-atributnya yang saling berhubungan untuk menemukan suatu pola dari sebuah dataset (Ridwan, 2020). Secara sederhana, data mining sering dijelaskan sebagai proses mengekstraksi atau menambang pengetahuan dari sejumlah besar data. Istilah data mining dan *Knowledge Discovery in Database* (KDD) sering digunakan secara bergantian untuk menggambarkan proses mengungkapkan informasi

tersembunyi dalam suatu basis data yang besar (Prasetya et al., 2022).

Algoritma Naïve Bayes

Algoritma Naïve Bayes adalah teknik peramalan probabilistik sederhana berdasarkan penerapan teorema Bayes (rule of Bayes) dengan asumsi independensi (mandiri) kuat (naif) (Suharyanto & Zein, 2022). Di Bayes (terutama Naïve Bayes) Independensi fitur yang kuat adalah fitur dalam data yang tidak terkait ada tidaknya karakteristik lain dalam data yang sama (Laia, Buulolo, & Sirait, 2018). Rumus teorema Naïve Bayes:

$$P(c | x) = \frac{P(x | c)P(c)}{P(x)}$$


$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$

Gambar 1. Rumus Naïve Bayes

RapidMiner

RapidMiner adalah sebuah platform software data ilmu pengetahuan yang dikembangkan oleh perusahaan dengan nama yang sama, yang menyediakan lingkungan terpadu untuk pembelajaran mesin (*machine learning*), pembelajaran mendalam (*deep learning*), penambangan teks (*text mining*), dan analisis prediktif (*predictive analytics*) (Nofitri & Irawati, 2019). Contoh penerapan Rapidminer adalah menerapkan suatu metode pembelajaran ke dataset dan menganalisis hasilnya guna mendapatkan informasi mengenai data atau menerapkan lebih dari satu metode lalu membandingkan performasinya untuk dipilih (Ridwan, 2020).

3. Metode

Pengumpulan Data

Metode yang digunakan dalam penelitian ini melibatkan metode pengumpulan data yang mencakup atribut-atribut mobil dan label akseptabilitas. Data ini diperoleh dari situs Kaggle. Dengan record yang berjumlah 1708 data training dan 21 data training.

Pra-pemrosesan Data

Setelah data dikumpulkan, selanjutnya dilakukan pra-pemrosesan data untuk mempersiapkan data tersebut dengan melakukan pembersihan data guna menghilangkan data yang tidak lengkap, duplikat, atau tidak valid.

Pembagian Dataset

Setelah pra-pemrosesan data dilakukan, dataset dibagi menjadi 2 data, yaitu data training dan data testing. Data training akan digunakan untuk melatih model yang telah dilatih.

Pemodelan

Tahap ini merupakan gambaran dari serangkaian tahapan yang dilakukan untuk *training* dan *testing* terhadap data (Jaman et al., 2020). Pemodelan dilakukan dengan bantuan *software* Rapidminer.

Evaluasi

Pada fase ini, hasil dari fase pemodelan diinterpretasikan secara grafis menggunakan ukuran tingkat ketelitian, yaitu tingkat kedekatan nilai-nilai klasifikasi nilai sebenarnya (Jaman et al., 2020).

4. Hasil Dan Pembahasan

Proses awal yang dilakukan adalah mengumpulkan data dari situs Kaggle yang merupakan situs untuk berbagi

ide dan mendapatkan inspirasi. Adapun variable data yang digunakan disajikan pada tabel berikut.

Tabel 1. Deskripsi Variable Dataset

No	Atribut	Value
1	Buying_price	Vhigh, high, med, low
2	Maintance_price	Vhigh, high, med, low
3	No_of_Doors	2, 3, 4, 5more
4	Person_capacity	2, 4, more
5	Size_of_Luggage	Small, med, big
6	Safety	Low, med, high
7	Class (car_acceptability)	Unacc, acc, good, vgood

Dari tabel 1. Ada 6 atribut dataset akseptabilitas mobil dan 1 variabel class penentu klasifikasi. Dari data set yang akan dianalisis dibagi menjadi data training dan data testing. Setelah praproses, dilakukan perhitungan probabilitas secara manual dan komputerisasi untuk dibandingkan. Berikut total dan probabilitas tiap kelas.

Menggunakan excel:

	jumlah	probabilitas class/label
unacc	1201	0,704
acc	382	0,224
good	63	0,037
vgood	61	0,036
total	1707	

Gambar 2. Perhitungan Probabilitas Secara Manual

Menggunakan Rapidminer:

```
Distribution model for label attribute Car_Acceptability

Class unacc (0.704)
6 distributions

Class acc (0.224)
6 distributions

Class vgood (0.036)
6 distributions

Class good (0.037)
6 distributions
```

Gambar 3. Perhitungan Probabilitas Rapidminer

Setelah pra-pemrosesan, data training dan data testing akan diklasifikasi menggunakan rapidminer. Adapun hasilnya sebagai berikut. Adapun salah satu contoh hasilnya klasifikasi dan akurasi sebagai berikut.

Tabel 2. Hasil Klasifikasi dari Data Testing

N o	Atribut	Testin g	Hasil
1	Buying_price	Vhigh	Vhigh
2	Maintance_price	Vhigh	Vhigh
3	No_of_Doors	2	2
4	Person_capacity	2	2
5	Size_of_Luggag e	Small	Small
6	Safety	Low	Low
7	Class (car_acceptabilit y)	unacc	unacc

	true unacc	true acc	true good	true vgood	cl
pred. unacc	9	0	0	0	10
pred. acc	0	0	3	0	0.0
pred. good	0	2	3	0	60
pred. vgood	0	0	0	4	10
class recall	100.00%	0.00%	50.00%	100.00%	

Gambar 4. Hasil Recall dan Presisi

Selanjutnya dilakukan evaluasi dengan *confusion matrix* Hasil dari evaluasi pengklasifikasian dengan Naïve Bayes menghasilkan:

```
accuracy: 76.19%
ConfusionMatrix:
True:   unacc  acc   good  vgood
unacc:  9     0    0     0
acc:    0     0    3     0
good:   0     2    3     0
vgood:  0     0    0     4
```

Gambar 5. Hasil Akurasi

Dari gambar diatas diperoleh akurasi klasifikasi dataset akseptabilits mobil adalah sebesar 76.19%.

5. Kesimpulan

Ada 6 atribut yang mempengaruhi klasifikasi dataset Car Acceptability yaitu buying price, maintance price, number of doors, person capacity, size of luggage, dan safety dengan label 4 kategori (unacc, good, vgood, acc). Penelitian ini menggunakan Algoritma Naïve Bayes untuk pengklasifikasian dataset Car Acceptability. Dari 1728 data lalu dibagi menjadi 1707 data training dan 21 data testing. Dari hasil evaluasi klasifikasi dengan algoritma Naïve Bayes diperoleh akurasi pengklasifikasian dataset car accepttability yaitu 76.19%. ini berarti akurasi dari klasifikasi permodelan algoritma Naïve Bayes terhadap dataset car acceptability sudah cukup baik dan perlu peningkatan akurasi dengan cara yang lain.

REFERENSI

Bimantara, Y., Novita, D., & Jaelani. (2022). Pengaruh Harga, Desain Produk dan Gaya Hidup Terhadap

- Keputusan Pembelian (Survei Pada Konsumen Mobil Honda Brio di Lampung). *Journal Strategy of Management and Accounting Through Research and Technology*, 1(2), 27–36. <http://jim.teknokrat.ac.id/index.php/smart/issue/archive>
- Fadlan, C., Ningsih, S., & Windarto, A. P. (2018). PENERAPAN METODE NAÏVE BAYES DALAM KLASIFIKASI KELAYAKAN KELUARGA PENERIMA BERAS RASTRA. *JUTIM*, 3(1), 1–8.
- Jaman, J. H., Jaman, J. H., . C., & Sanjaya, A. R. (2020). Klasifikasi jenis mobil paling diminati Di indonesia menggunakan algoritma Naive bayes. *Faktor Exacta*, 13(1), 18. <https://doi.org/10.30998/faktorexacta.v13i1.5573>
- Laia, D., Buulolo, E., & Sirait, M. J. F. (2018). Implementasi Data Mining Untuk Memprediksi Pemesanan Driver Go-Jek Online Dengan Menggunakan Metode Naïve Bayes (Studi Kasus: Pt. Go-Jek Indonesia). *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 2(1),434–439. <https://doi.org/10.30865/komik.v2i1.972>
- Nofitri, R., & Irawati, N. (2019). Analisis Data Hasil Keuntungan Menggunakan Software Rapidminer. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 5(2), 199–204. <https://doi.org/10.33330/jurteksi.v5i2.365>
- Prasetya, T., Yanti, J. E., Purnamasari, A. I., Dikananda, A. R., & Nurdiawan, O. (2022). Analisis Data Transaksi Terhadap Pola Pembelian Konsumen Menggunakan Metode Algoritma Apriori. *INFORMATICS FOR EDUCATORS AND PROFESSIONAL: Journal of Informatics*, 6(1), 43. <https://doi.org/10.51211/itbi.v6i1.1688>
- Ridwan, A. (2020). Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, 4(1), 15–21. <https://doi.org/10.47970/siskom-kb.v4i1.169>
- Suharyanto, E., & Zein, A. (2022). Analisis Data Minat Calon Mahasiswa Universitas Pamulang Dengan Menggunakan Algoritma Naive Bayes Classifier. *Sainstech: Jurnal Penelitian Dan ...*, 70–76. <https://ejournal.istn.ac.id/index.php/sainstech/article/view/1434%0Ahttps://ejournal.istn.ac.id/index.php/sainstech/article/download/1434/943>
- Tarigan, V. (2023). Pembuatan Aplikasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma Naive Bayes. *Informatika*, 11(1), 54–62. <https://doi.org/10.36987/informatika.v11i1.3847>
- Waluyo, S. H. (2017). Klasifikasi Pemanfaat Program Beras Sejahtera (RASTRA) Berdasarkan Tingkat Kemiskinan Dengan Menggunakan Algoritma Decision Tree C4 . 5 Berbasis Particle Swarm Optimization. *JUTIM*, 7(2), 19–24.