JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9320
Hal 899-907
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Prediksi Harga Mobil Global Menggunakan Machine Learning dengan Algoritma Naive Bayes

Dedek Indra Gunawan Hts[1*], Firman Edi[2], Ratna Sri Hayati[3], Hendro Sutomo Ginting[1]

[1] Fakultas Ekonomi dan Bisnis, Kewirausahaan, Universitas Satya Terra Bhinneka, Medan, Indonesia
[2] Fakultas Teknologi Industri, Manajemen Rekayasa, Institut Teknologi Batam, Batam, Indonesia
[3] Fakultas Ekonomi dan Bisnis, Bisnis Digital, Universitas Satya Terra Bhinneka, Medan, Indonesia
Email: [1*]dedekindra@satyaterrabhinneka.ac.id, [2]firmanedi97@gmail.com, [3]ratnasrihayati@satyaterrabhinneka.ac.id, [4]hendrosutomo@satyaterrabhinneka.ac.id
Email Penulis Korespondensi: dedekindra@satyaterrabhinneka.ac.id
Submitted **07-11-2025;** Accepted **13-12-2025;** Published **31-12-2025**

Abstrak

Penentuan harga mobil merupakan salah satu tantangan utama dalam industri otomotif global karena dipengaruhi oleh banyak faktor seperti spesifikasi teknis, kondisi kendaraan, dan dinamika pasar. Permasalahan ini semakin kompleks ketika jumlah data yang tersedia semakin besar, sehingga diperlukan metode yang mampu melakukan analisis cepat dan akurat. Penelitian ini bertujuan untuk memprediksi tingkat harga mobil berdasarkan spesifikasi kendaraan menggunakan pendekatan Machine Learning dengan algoritma Naive Bayes sebagai solusi untuk menyederhanakan proses klasifikasi harga pada data berskala besar. Dataset yang digunakan adalah Global Car Sales Analysis dari platform Kaggle, yang memuat atribut seperti Manufacturer, Model, Engine size, Fuel type, Year of manufacture, Mileage, dan Price. Metodologi penelitian mencakup tahap data preprocessing, label encoding untuk atribut kategorikal, pembagian data menjadi training dan testing set, serta penerapan algoritma Naive Bayes untuk mengklasifikasikan harga mobil ke dalam tiga kategori: Low, Medium, dan High. Hasil penelitian menunjukkan bahwa Naive Bayes mampu memprediksi harga mobil dengan performa sangat baik, ditunjukkan oleh akurasi 96%, precision 0.97, recall 0.96, dan F1-score 0.96. Model memberikan hasil terbaik pada kategori Low dengan F1-score 0.98, namun performanya menurun pada kategori Medium dan High akibat ketidakseimbangan distribusi data. Analisis lebih lanjut juga mengungkapkan bahwa atribut Engine size, Year of manufacture, dan Mileage merupakan faktor yang paling berpengaruh dalam penentuan harga. Secara keseluruhan, penelitian ini membuktikan bahwa Naive Bayes merupakan metode yang efektif untuk memprediksi harga mobil berdasarkan data global.

Kata Kunci: Machine Learning; Naive Bayes; Prediksi Harga Mobil; Data Mining; Penjualan Mobil Global

Abstract

Determining car prices is one of the major challenges in the global automotive industry because it is influenced by various factors such as technical specifications, vehicle condition, and market dynamics. This issue becomes more complex as the volume of available data increases, requiring methods capable of performing fast and accurate analysis. This study aims to predict car price levels based on vehicle specifications using a Machine Learning approach, with the Naive Bayes algorithm selected as a solution to simplify the price classification process on large-scale data. The dataset used is the Global Car Sales Analysis from the Kaggle platform, which includes attributes such as Manufacturer, Model, Engine size, Fuel type, Year of manufacture, Mileage, and Price. The research methodology consists of data preprocessing, label encoding for categorical attributes, splitting the dataset into training and testing sets, and applying the Naive Bayes algorithm to classify car prices into three categories: Low, Medium, and High. The results indicate that Naive Bayes is capable of predicting car prices with very strong performance, achieving an accuracy of 96%, precision of 0.97, recall of 0.96, and an F1-score of 0.96. The model performs best on the Low category with an F1-score of 0.98, although performance decreases for the Medium and High categories due to imbalanced class distribution. Further analysis also reveals that Engine size, Year of manufacture, and Mileage are the most influential attributes in determining price. Overall, this study demonstrates that Naive Bayes is an effective method for predicting car prices using global automotive data.

Keywords: Machine Learning; Naive Bayes; Car Price Prediction; Data Mining; Global Car Sales

1. PENDAHULUAN

Perkembangan teknologi digital telah mendorong industri otomotif global untuk beradaptasi dalam menghadapi kompetisi yang semakin ketat. Salah satu tantangan utama yang dihadapi adalah memahami faktor-faktor yang memengaruhi harga kendaraan di pasar yang sangat dinamis. Penentuan harga mobil tidak hanya bergantung pada merek atau model, tetapi juga pada variabel teknis seperti kapasitas mesin, jenis bahan bakar, tahun pembuatan, dan jarak tempuh. Kompleksitas variabel ini menjadikan proses estimasi harga kendaraan, khususnya mobil bekas, sebagai tantangan tersendiri yang membutuhkan pendekatan berbasis data. Dengan meningkatnya jumlah data otomotif yang tersedia secara publik, pendekatan *Machine Learning* menjadi salah satu solusi potensial untuk melakukan prediksi dan analisis harga mobil secara otomatis dan efisien [1].

Seiring dengan meningkatnya ketersediaan data otomotif global yang dapat diakses melalui platform digital seperti Kaggle, pendekatan berbasis *Machine Learning* (ML) mulai banyak dimanfaatkan untuk melakukan prediksi harga mobil secara otomatis dan efisien [2]. *Machine Learning* memungkinkan sistem untuk belajar dari data historis guna menemukan pola dan hubungan antarvariabel yang sulit dideteksi secara manual. Dengan demikian, metode ini dapat menghasilkan model prediktif yang membantu konsumen dalam menilai harga wajar kendaraan serta mendukung pelaku industri otomotif dalam strategi penetapan harga dan pengambilan keputusan bisnis [3].

Dalam konteks analisis harga mobil, *Machine Learning* berperan penting dalam mengidentifikasi hubungan kompleks antar variabel seperti kapasitas mesin, jenis bahan bakar, jarak tempuh, dan tahun pembuatan. Pendekatan ini

Copyright © 2025 The **Dedek Indra Gunawan Hts**, Page 899
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9320
Hal 899-907
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

telah terbukti efektif dalam berbagai penelitian terdahulu, baik untuk prediksi harga maupun rekomendasi produk. Misalnya, penelitian yang dilakukan oleh [2][5][6] menunjukkan bahwa metode *Machine Learning* mampu meningkatkan akurasi prediksi harga mobil melalui pemanfaatan data spesifikasi kendaraan yang terstruktur dengan baik. Penelitian oleh [2] menyoroti penerapan berbagai algoritma *Machine Learning* untuk meningkatkan akurasi dalam prediksi harga mobil bekas. Studi tersebut membandingkan beberapa model supervised learning seperti *Random Forest, Ridge Regression, Lasso Regression*, dan *Linear Regression* dengan menggunakan data dari platform Kaggle. Hasil penelitian menunjukkan bahwa pendekatan *Machine Learning* mampu menangkap pola kompleks antara variabel seperti tahun produksi, jarak tempuh, dan kapasitas mesin dalam menentukan harga kendaraan.

Salah satu algoritma yang populer dan efisien digunakan untuk klasifikasi dalam *Machine Learning* adalah *Naive Bayes*[7]. Algoritma ini dikenal karena kesederhanaannya, efisiensi komputasi, dan kemampuannya memberikan hasil yang baik meskipun pada dataset dengan ukuran terbatas. Prinsip dasar dari *Naive Bayes* didasarkan pada Teorema Bayes dengan asumsi independensi antar fitur. Meskipun asumsi tersebut jarang terpenuhi secara sempurna dalam data dunia nyata, algoritma ini tetap memberikan performa yang kompetitif dalam banyak kasus klasifikasi[8][9].

Algoritma *Naive Bayes* merupakan salah satu metode klasifikasi yang populer dalam bidang *Machine Learning* karena kesederhanaannya, efisiensi komputasi, serta kemampuan untuk memberikan hasil yang cukup baik meskipun dengan data yang terbatas[9][10]. Prinsip dasar algoritma ini adalah menerapkan Teorema Bayes dengan asumsi independensi antar fitur, yang seringkali cukup efektif dalam memecahkan permasalahan klasifikasi di berbagai domain, termasuk bidang otomotif[12][13] .

Beberapa penelitian sebelumnya telah menggunakan algoritma *Naive Bayes* untuk memprediksi atau mengklasifikasikan data dalam konteks bisnis dan teknologi. Misalnya, penelitian oleh [14] menggunakan *Naive Bayes* untuk memprediksi harga kendaraan bekas di pasar lokal dengan akurasi di atas 80%. Penelitian lain oleh [15] menunjukkan bahwa *Naive Bayes* mampu mengklasifikasikan jenis kendaraan berdasarkan parameter teknis dengan hasil yang kompetitif dibandingkan model kompleks seperti Random Forest dan Support Vector Machine. Penelitian [16] mengeksplorasi berbagai teknik *Machine Learning* untuk prediksi harga mobil, dengan menggunakan dataset berisi lebih dari 200 mobil dan 26 fitur variabel. Metode yang diuji meliputi *Voting Regressor, Gradient Boosting Regressor, Random Forest Regressor, Decision Tree Regressor*, dan *Support Vector Regressor*. Hasil yang diperoleh menunjukkan bahwa Voting Regressor memberikan performa terbaik dengan akurasi uji sekitar 95,8%.

Dalam penelitian ini, algoritma *Naive Bayes* diterapkan untuk memprediksi kategori harga mobil (*Price Level*) berdasarkan atribut *Manufacturer*, Model, *Engine size, Fuel type*, *Year of manufacture*, dan *Mileage* yang tersedia dalam dataset *Global Car Sales Analysis*. Pendekatan ini diharapkan dapat menghasilkan model prediktif yang sederhana namun akurat, serta memberikan kontribusi dalam pengembangan sistem rekomendasi dan analisis pasar di industri otomotif global.

Tujuan utama penelitian ini adalah untuk menganalisis faktor-faktor yang memengaruhi kategori harga mobil berdasarkan data global serta membangun model klasifikasi yang mampu mengelompokkan mobil ke dalam kategori harga tertentu. Penelitian ini memanfaatkan algoritma *Naive Bayes* sebagai metode utama dalam proses klasifikasi, mengingat algoritma ini dikenal sederhana namun efektif dalam menangani data dengan berbagai atribut yang saling independen. Selain itu, penelitian ini juga bertujuan untuk mengevaluasi performa model yang dihasilkan menggunakan berbagai metrik evaluasi, seperti akurasi, presisi, dan *recall*, guna menilai sejauh mana kemampuan model dalam mengenali dan memprediksi kategori harga mobil secara tepat. Melalui pendekatan ini, diharapkan diperoleh pemahaman yang lebih mendalam mengenai variabel-variabel penting yang berpengaruh terhadap harga mobil dan efektivitas algoritma *Naive Bayes* dalam konteks analisis data otomotif global.

2. METODOLOGI PENELITIAN

2.1 Jenis dan Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari platform Kaggle dengan judul *Global Car Sales Analysis.* Dataset tersebut berisi informasi mengenai spesifikasi dan harga mobil dari berbagai produsen dan model di pasar global. Atribut utama yang digunakan dalam penelitian ini adalah:

Tabel 1. Tabel Atribut

No	Atribut	Keterangan
1	Manufacturer	Nama produsen mobil
2	Model	Jenis atau varian mobil
3	Engine size	Kapasitas mesin kendaraan (liter)
4	Fuel type	Jenis bahan bakar (bensin, diesel, listrik, dll)
5	Year of manufacture	Tahun pembuatan kendaraan
6	Mileage	Jarak tempuh mobil (kilometer)
7	Price	Harga kendaraan dalam satuan moneter

Tujuan utama dari penelitian ini adalah untuk mengklasifikasikan harga mobil berdasarkan atribut-atribut tersebut ke dalam tiga kategori harga, yaitu *Low* untuk mobil dengan harga rendah, *Medium* untuk mobil dengan harga menengah, dan *High* untuk mobil dengan harga tinggi. Klasifikasi ini digunakan sebagai variabel target dalam penerapan algoritma

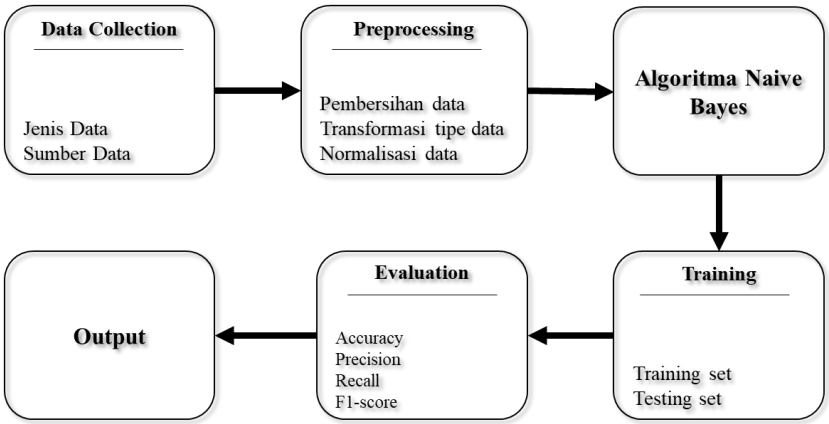
Copyright © 2025 The **Dedek Indra Gunawan Hts**, Page 900
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9320
Hal 899-907
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Naive Bayes, yang diharapkan mampu mengidentifikasi pola dan hubungan antar atribut dalam menentukan harga kendaraan secara akurat serta memberikan wawasan yang bermanfaat bagi industri otomotif global.

2.2 Tahapan Penelitian

Proses analisis dilakukan melalui beberapa tahap utama seperti pada Gambar 1 berikut:



Gambar 1. Alur Penelitian Prediksi Harga Mobil dengan *Naive Bayes*

a. *Data Collection*

Tahap *data collection* atau pengumpulan data merupakan langkah awal dalam penelitian ini. Data yang digunakan bersumber dari dataset sekunder yang diunduh melalui platform *Kaggle* dengan judul *Global Car Sales Analysis*. Dataset ini dipilih karena memiliki cakupan informasi yang luas dan relevan dengan tujuan penelitian, yaitu untuk menganalisis faktor-faktor yang memengaruhi harga mobil di pasar global. Dataset tersebut berisi informasi lengkap mengenai spesifikasi kendaraan dan harga jual dari berbagai merek dan model mobil di seluruh dunia. Atribut utama yang digunakan dalam penelitian meliputi nama produsen mobil (*Manufacturer*), jenis atau varian mobil (Model), kapasitas mesin kendaraan dalam liter (*Engine size*), jenis bahan bakar yang digunakan (*Fuel type*), tahun pembuatan kendaraan (*Year of manufacture*), jarak tempuh kendaraan dalam kilometer (*Mileage*), serta harga kendaraan dalam satuan moneter (*Price*). Seluruh data dikumpulkan dan disimpan dalam format CSV (*Comma Separated Values*) untuk memudahkan proses analisis menggunakan *Google Colab*. Sebelum data digunakan pada tahap pelatihan model, dilakukan proses pembersihan dan pra-pemrosesan (*data preprocessing*) guna memastikan tidak terdapat data yang hilang, duplikat, atau inkonsistensi yang dapat memengaruhi akurasi hasil analisis. Tahap pengumpulan data ini bertujuan untuk memastikan bahwa informasi yang diperoleh bersifat valid, representatif, dan relevan terhadap permasalahan penelitian, sehingga hasil klasifikasi yang dihasilkan oleh algoritma *Machine Learning* dapat diinterpretasikan secara akurat dan memberikan nilai analitis yang bermakna.

b. Preprocessing

Tahap data preprocessing merupakan langkah penting dalam penelitian ini untuk memastikan kualitas data yang digunakan sebelum dilakukan proses pelatihan model [17]. Data yang diperoleh dari dataset *Global Car Sales Analysis* pada *platform Kaggle* terlebih dahulu diperiksa secara menyeluruh untuk mengidentifikasi adanya data yang hilang (*missing values*), duplikasi, atau ketidakkonsistenan nilai pada setiap atribut. Langkah ini bertujuan untuk menghasilkan dataset yang bersih dan siap digunakan dalam proses analisis lebih lanjut. Proses pra-pemrosesan dimulai dengan pembersihan data, yaitu menghapus baris data yang tidak lengkap atau berisi nilai kosong yang tidak dapat diimputasi secara logis. Selain itu, dilakukan juga pemeriksaan terhadap data yang bersifat duplikat agar tidak memengaruhi hasil pelatihan model. Setelah data bersih, dilakukan transformasi tipe data untuk memastikan bahwa setiap atribut memiliki format yang sesuai misalnya, *atribut Year of manufacture, Engine size, Mileage*, dan *Price* diubah menjadi tipe numerik agar dapat diproses oleh algoritma *Machine Learning*.

Selanjutnya, atribut kategorikal seperti Manufacturer, Model, dan Fuel type diubah menjadi bentuk numerik menggunakan metode Label Encoding. Proses ini diperlukan agar model *Naive Bayes* dapat mengenali dan mengolah atribut-atribut tersebut dalam perhitungan probabilitas. Setelah itu, data dibagi menjadi dua bagian, yaitu training set dan testing set, dengan proporsi umum 80% untuk pelatihan dan 20% untuk pengujian. Pembagian ini bertujuan agar model dapat belajar dari sebagian besar data dan diuji pada data yang belum pernah dilihat sebelumnya. Tahap terakhir dari pra-pemrosesan adalah melakukan normalisasi data agar setiap fitur memiliki skala yang seimbang dan tidak ada atribut dengan nilai ekstrem yang mendominasi proses pelatihan. Dengan langkah-langkah tersebut, dataset menjadi siap untuk digunakan dalam proses training model *Naive Bayes*.

Secara keseluruhan, tahap data preprocessing ini berperan penting dalam menjamin bahwa model *Machine Learning* yang dikembangkan memiliki dasar data yang bersih, konsisten, dan representatif, sehingga hasil prediksi harga mobil dapat dihasilkan secara lebih akurat dan reliabel.

Copyright © 2025 The **Dedek Indra Gunawan Hts**, Page 901
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9320
Hal 899-907
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

c. Algoritma *Naïve Bayes*

Pada tahap ini, penelitian menggunakan algoritma *Naive Bayes* sebagai metode utama untuk melakukan klasifikasi harga mobil. Algoritma *Naive Bayes* merupakan metode *Machine Learning* berbasis probabilistik yang didasarkan pada Teorema Bayes. Prinsip dasar algoritma ini adalah menghitung probabilitas bersyarat (*conditional probability*) dari setiap kelas berdasarkan nilai-nilai atribut yang diberikan, dengan asumsi bahwa setiap atribut bersifat independen satu sama lain.

Secara matematis, teorema Bayes dirumuskan sebagai berikut [18]:

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)} \quad (1)$$

di mana:
P(C|X) adalah probabilitas suatu kelas C (misalnya kategori harga *Low*, *Medium*, atau *High*) diberikan data fitur X,
P(X|C) adalah probabilitas fitur X muncul dalam kelas C,
P(C) adalah probabilitas awal dari kelas C, dan
P(X) adalah probabilitas dari fitur X secara keseluruhan.

Penelitian ini menggunakan varian *Naive Bayes*, karena sebagian besar atribut seperti *Engine size, Year of manufacture, Mileage,* dan *Price* bersifat numerik dan cenderung mengikuti distribusi normal. Dalam *Naive Bayes*, distribusi setiap fitur dalam setiap kelas diasumsikan berbentuk normal dengan parameter mean (μ) dan standar deviasi (σ), sehingga probabilitas dihitung menggunakan fungsi densitas.
Pemilihan algoritma *Naive Bayes* didasarkan pada beberapa keunggulannya, yaitu kemampuannya bekerja dengan baik pada dataset berukuran besar, proses pelatihan yang cepat, serta hasil prediksi yang cukup akurat meskipun asumsi independensi antar fitur tidak sepenuhnya terpenuhi[19]. Dalam konteks penelitian ini, algoritma *Naive Bayes* digunakan untuk memprediksi kategori harga mobil berdasarkan kombinasi atribut spesifikasi kendaraan, sehingga diharapkan dapat memberikan pemahaman yang lebih mendalam mengenai faktor-faktor yang memengaruhi harga mobil di pasar global.

d. Training

Pada tahap training, proses dimulai dengan pembagian dataset menjadi dua bagian, yaitu training set dan testing set, dengan proporsi yang umum digunakan seperti 80% data untuk pelatihan dan 20% untuk pengujian [20]. Data pelatihan digunakan untuk membangun model *Machine Learning* menggunakan algoritma *Naive Bayes*, yang merupakan salah satu varian dari *Naive Bayes* yang cocok untuk data numerik dan berdistribusi normal. Selama proses pelatihan, model mempelajari hubungan antara atribut independen (fitur) seperti *Engine size*, *Fuel type*, *Year of manufacture*, dan *Mileage* terhadap atribut dependen (target) yaitu kategori harga mobil (*Low*, *Medium*, *High*). Algoritma *Naive Bayes* bekerja dengan menghitung probabilitas bersyarat setiap kelas berdasarkan distribusi nilai pada masing-masing fitur menggunakan teorema Bayes. Tahap training bertujuan untuk menghasilkan model yang mampu mengenali pola dan kecenderungan data secara optimal, sehingga ketika diberikan data baru (yang belum pernah dilihat sebelumnya), model dapat memperkirakan kategori harga mobil dengan tingkat akurasi yang tinggi. Proses ini juga mencakup perhitungan parameter statistik seperti mean dan standar deviasi dari setiap atribut numerik untuk setiap kelas harga, yang kemudian digunakan dalam perhitungan probabilitas prediksi pada tahap pengujian.

Secara keseluruhan, tahap training menjadi langkah penting dalam membentuk kemampuan prediktif model, karena pada fase inilah algoritma belajar dari data historis untuk mengenali karakteristik dan pengaruh setiap atribut terhadap penentuan harga mobil.

e. *Evaluation*

Tahap evaluation atau evaluasi model merupakan langkah penting untuk menilai sejauh mana performa algoritma *Naive Bayes* dalam memprediksi kategori harga mobil berdasarkan data yang telah melalui proses pelatihan[21]. Evaluasi dilakukan setelah model selesai dilatih menggunakan training set dan kemudian diuji dengan testing set yang belum pernah dilihat oleh model sebelumnya. Tujuan utama tahap ini adalah untuk mengukur kemampuan model dalam melakukan klasifikasi secara akurat serta mengidentifikasi potensi kesalahan prediksi yang terjadi.

Dalam penelitian ini, proses evaluasi dilakukan dengan menggunakan beberapa metrik kinerja, yaitu accuracy, precision, *recall*, dan *F1-score*. Nilai akurasi menggambarkan persentase keseluruhan prediksi yang benar terhadap total data uji. Sementara itu, presisi mengukur sejauh mana model mampu mengidentifikasi data yang benar-benar termasuk dalam kelas tertentu, dan *recall* menunjukkan kemampuan model dalam menemukan seluruh data yang termasuk dalam suatu kelas. *F1-score* merupakan kombinasi harmonis antara precision dan *recall*, yang memberikan gambaran lebih seimbang terhadap performa model, terutama ketika distribusi data tidak merata antar kelas.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengumpulan dan Analisis Data (*Data Collection*)

Dataset yang digunakan dalam penelitian ini diperoleh dari platform Kaggle dengan judul Global Car Sales Analysis. Dataset ini berisi informasi mengenai berbagai spesifikasi dan harga mobil dari berbagai merek dan model yang beredar di pasar global. Berdasarkan hasil eksplorasi awal seperti terlihat pada Gambar di atas, dataset terdiri dari 50.000 entri (baris) dan 7 atribut (kolom) utama yang meliputi *Manufacturer, Model, Engine Size, Fuel Type, Year of Manufacture,*

Copyright © 2025 The **Dedek Indra Gunawan Hts**, Page 902
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9320
Hal 899-907
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Mileage, dan P*rice*. Atribut Manufacturer dan Model bertipe data object karena berisi teks yang menunjukkan nama produsen dan jenis mobil. Atribut *Engine Size* bertipe float64 karena menunjukkan kapasitas mesin dalam liter, sedangkan *Year of Manufacture* dan *Mileage* bertipe int64 karena berisi data numerik tahunan dan jarak tempuh dalam satuan kilometer. Atribut terakhir, *Price*, juga bertipe int64 dan menjadi variabel target dalam penelitian ini.

Secara umum, hasil analisis awal ini menunjukkan bahwa dataset memiliki struktur yang bersih, lengkap, dan relevan untuk tujuan penelitian, yaitu melakukan klasifikasi harga mobil berdasarkan spesifikasi kendaraan. Tahapan selanjutnya adalah melakukan preprocessing agar data siap diproses oleh model *Machine Learning*.

	Manufacturer	Model	Engine size	Fuel type	Year of manufacture	Mileage	Price
0	Ford	Fiesta	1.0	Petrol	2002	127300	3074
1	Porsche	718 Cayman	4.0	Petrol	2016	57850	49704
2	Ford	Mondeo	1.6	Diesel	2014	39190	24072
3	Toyota	RAV4	1.8	Hybrid	1988	210814	1705
4	VW	Polo	1.0	Petrol	2006	127869	4101

Gambar 2. Hasil dari Proses Data Collection

Semua kolom memiliki jumlah nilai non-null sebanyak 50.000 entri, artinya tidak terdapat missing value pada dataset ini. Hal ini menunjukkan bahwa dataset sudah bersih dan siap digunakan untuk proses *Machine Learning* tanpa perlu imputasi data.

3.2 Hasil Data *Preprocessing*

Tahapan data preprocessing dilakukan untuk memastikan bahwa data yang akan digunakan dalam pemodelan *Machine Learning* berada dalam format yang bersih, terstruktur, dan sesuai dengan kebutuhan algoritma *Naive Bayes*. Proses ini mencakup beberapa langkah utama, yaitu pembersihan data, transformasi tipe data, *encoding* variabel kategorikal, serta pembentukan label target klasifikasi. Langkah pertama adalah pemeriksaan data duplikat dan nilai kosong (*missing values*). Berdasarkan hasil eksplorasi, tidak ditemukan data yang hilang pada seluruh atribut, sehingga tidak diperlukan proses imputasi. Hal ini menunjukkan bahwa dataset *Global Car Sales Analysis* sudah memiliki kualitas data yang baik dan siap untuk dianalisis.

```
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Manufacturer        50000 non-null  object
 1   Model               50000 non-null  object
 2   Engine size         50000 non-null  float64
 3   Fuel type           50000 non-null  object
 4   Year of manufacture 50000 non-null  int64
 5   Mileage             50000 non-null  int64
 6   Price               50000 non-null  int64
dtypes: float64(1), int64(3), object(3)
memory usage: 2.7+ MB
```

Gambar 3. Hasil Pembersihan Data

Langkah berikutnya adalah transformasi atribut kategorikal menjadi bentuk numerik menggunakan metode Label Encoding. Atribut seperti *Manufacturer, Model*, dan *Fuel Type* dikonversi menjadi nilai numerik agar dapat diproses oleh model *Naive Bayes*, yang hanya dapat menerima input berupa angka. Misalnya, jenis bahan bakar seperti "Petrol", "Diesel", dan "Hybrid" masing-masing dikodekan menjadi 0, 1, dan 2. Selanjutnya dilakukan pembuatan label target (*Price Category*). Nilai *Price* dibagi menjadi tiga kategori, yaitu *Low, Medium*, dan *High* berdasarkan distribusi harga dalam dataset. Kategori ini ditentukan dengan menggunakan metode pembagian rentang harga secara kuantil agar setiap kelas memiliki representasi yang proporsional. Setelah proses encoding dan kategorisasi selesai, dataset kemudian dibagi menjadi dua bagian, yaitu training set sebesar 80% dan testing set sebesar 20%. Pembagian ini dilakukan agar model dapat dilatih menggunakan sebagian besar data dan diuji pada data yang belum pernah dilihat sebelumnya, sehingga hasil evaluasi lebih objektif.

Tabel 2. Label Encoding

Atribut	Nilai Asli	Nilai Setelah Encoding
Fuel Type	Petrol	0
Fuel Type	Diesel	1
Fuel Type	Hybrid	2

Copyright © 2025 The **Dedek Indra Gunawan Hts**, Page 903
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9320
Hal 899-907
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Manufacturer	Toyota	0
Manufacturer	Honda	1

3.3 Hasil Training Model *Naive Bayes*

Setelah proses data preprocessing selesai, tahap selanjutnya adalah pelatihan model menggunakan algoritma *Naive Bayes*. Dataset yang telah dibagi menjadi 80% data latih (training set) dan 20% data uji (*testing set*) digunakan untuk membangun serta mengevaluasi model klasifikasi harga mobil ke dalam tiga kategori*: Low, Medium*, dan *High*. Model kemudian diuji menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* untuk menilai performa klasifikasi pada setiap kategori.

Tabel 3. Hasil Trainig

Kelas	**Precision**	***Recall***	**F1-score**	**Jumlah Data (support)**
High	0.34	0.59	0.43	39
Low	0.99	0.97	0.98	9.748
Medium	0.25	0.40	0.31	213
Accuracy			**0.96**	10.000
Macro avg	**0.53**	**0.65**	**0.57**	10.000
Weighted avg	**0.97**	**0.96**	**0.96**	10.000

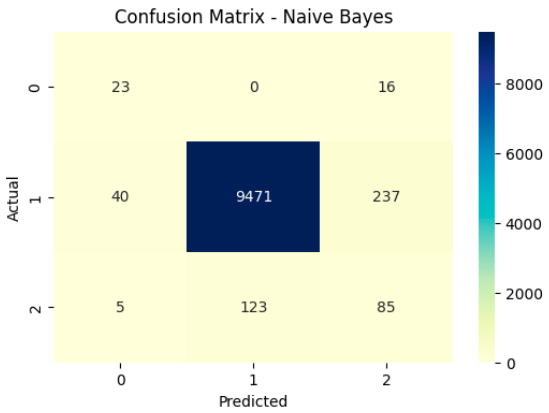
Berdasarkan hasil evaluasi performa model *Naive Bayes* yang ditunjukkan pada tabel di atas, diperoleh nilai akurasi keseluruhan sebesar 96%. Hal ini menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar. Namun, tingkat akurasi yang tinggi ini perlu dianalisis lebih lanjut melalui nilai precision, *recall*, dan *F1-score* pada masing-masing kelas untuk memahami sejauh mana model bekerja pada setiap kategori harga mobil. Untuk kelas *Low*, model menunjukkan kinerja yang sangat baik dengan nilai precision sebesar 0,99, *recall* 0,97, dan *F1-score* 0,98 dari total 9.748 data. Hasil ini mengindikasikan bahwa model sangat akurat dalam mengenali mobil dengan harga rendah dan hanya melakukan sedikit kesalahan prediksi. Dominasi jumlah data pada kelas ini turut memperkuat kemampuan model dalam mengenali pola karakteristik mobil dengan harga rendah.

Berbeda halnya dengan kelas *Medium*, nilai precision sebesar 0,25, *recall* 0,40, dan *F1-score* 0,31 menunjukkan bahwa model masih kesulitan dalam mengidentifikasi mobil dengan harga menengah. Banyak data dari kategori ini yang justru diklasifikasikan sebagai kelas harga rendah. Kondisi ini disebabkan oleh jumlah data yang relatif sedikit (hanya 213 data), sehingga model tidak memiliki cukup informasi untuk mempelajari pola yang membedakan kelas *Medium* dari kelas lainnya. Sementara untuk kelas *High*, model memiliki nilai precision 0,34, *recall* 0,59, dan *F1-score* 0,43 dengan total 39 data. Walaupun nilai *recall* cukup tinggi, yang berarti model dapat mengenali sebagian besar mobil dengan harga tinggi, nilai precision yang rendah menunjukkan bahwa masih terdapat cukup banyak kesalahan dalam memprediksi kategori ini. Keterbatasan jumlah data kelas *High* juga menjadi penyebab utama rendahnya performa model.

Secara keseluruhan, nilai macro average sebesar precision 0,53, *recall* 0,65, dan *F1-score* 0,57 menunjukkan bahwa kinerja model bervariasi antar kelas, dengan dominasi performa tinggi pada kelas *Low*. Sementara itu, nilai weighted average yang tinggi (mendekati nilai akurasi keseluruhan) memperkuat temuan bahwa model lebih akurat pada kelas dengan jumlah data besar. Dengan demikian, dapat disimpulkan bahwa meskipun model *Naive Bayes* efektif dalam memprediksi kategori harga mobil secara umum, performanya masih perlu ditingkatkan untuk kategori harga menengah dan tinggi melalui penyeimbangan data (data balancing) atau penggunaan metode klasifikasi yang lebih kompleks.

3.4 Hasil *Evaluation*

Setelah proses pelatihan model menggunakan algoritma *Naive Bayes* selesai dilakukan, tahap berikutnya adalah melakukan evaluasi untuk mengukur sejauh mana model mampu melakukan klasifikasi harga mobil dengan benar. Evaluasi dilakukan menggunakan *Confusion Matrix*, serta dihitung metrik performa seperti *Accuracy, Precision, Recall*, dan *F1-score*.



Gambar 4. Confusion Matrix

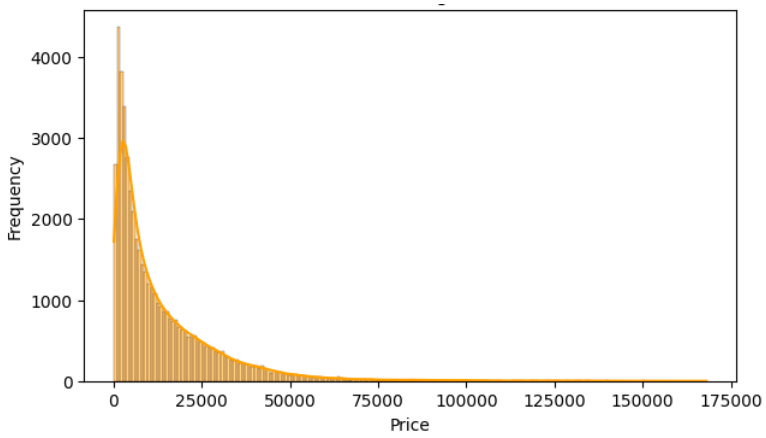
Copyright © 2025 The **Dedek Indra Gunawan Hts**, Page 904
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9320
Hal 899-907
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Berdasarkan *Confusion Matrix* hasil klasifikasi menggunakan algoritma *Naive Bayes*, terlihat bahwa model secara umum memiliki performa yang cukup baik, terutama dalam mengenali kelas harga mobil dengan kategori *Low* (label 1). Dari total data yang diuji, terdapat 9.471 data yang diklasifikasikan dengan benar sebagai kelas *Low*, sementara hanya 277 data (40 + 237) dari kelas *Low* yang salah diklasifikasikan sebagai kelas lain. Hal ini menunjukkan bahwa model sangat dominan dalam mengenali pola pada kelas *Low*, yang kemungkinan besar disebabkan oleh jumlah data pada kategori ini yang jauh lebih banyak dibandingkan kelas lainnya.

Untuk kelas *High* (label 0), model mampu mengklasifikasikan 23 data dengan benar, namun masih terdapat 16 data yang salah dikenali sebagai kelas *Medium* (label 2). Sementara itu, sebanyak 40 data dari kelas *High* justru teridentifikasi sebagai kelas *Low*. Kesalahan ini menunjukkan bahwa model mengalami kesulitan dalam membedakan antara kelas *High* dengan *Low*, yang mungkin disebabkan oleh kesamaan beberapa atribut antara mobil dengan harga tinggi dan menengah, atau karena jumlah data *High* yang terlalu sedikit sehingga tidak cukup untuk membentuk pola distribusi probabilistik yang akurat. Pada kelas *Medium* (label 2), model berhasil mengenali 85 data dengan benar, tetapi terdapat 123 data yang salah diklasifikasikan sebagai kelas *Low* dan 5 data yang salah dikategorikan sebagai kelas *High*. Pola kesalahan ini memperlihatkan bahwa model lebih sering mengelompokkan data *Medium* ke dalam kelas *Low*, memperkuat indikasi bahwa model terlalu bias terhadap kelas mayoritas.

Secara keseluruhan, pola pada *Confusion Matrix* menunjukkan bahwa *Naive Bayes* sangat efektif untuk mendeteksi kelas dengan data terbanyak (*Low*), tetapi belum optimal untuk kelas dengan data sedikit (*Medium* dan *High*). Dengan demikian, model ini cenderung mengalami masalah *class imbalance*, di mana distribusi data yang tidak seimbang membuat model sulit membedakan karakteristik antar kelas secara proporsional.



Gambar 5. Distribusi Harga Mobil

Gambar di atas menunjukkan distribusi harga mobil berdasarkan data yang digunakan dalam penelitian. Terlihat bahwa sebagian besar mobil memiliki harga yang relatif rendah, yaitu di bawah 20.000 unit harga (misalnya dolar atau satuan mata uang lain yang digunakan), dengan frekuensi kemunculan yang sangat tinggi. Kurva distribusi menunjukkan pola yang right-skewed (miring ke kanan), artinya terdapat sebagian kecil mobil dengan harga yang sangat tinggi, namun jumlahnya jauh lebih sedikit dibandingkan mobil dengan harga rendah. Pola ini umum terjadi pada data harga barang konsumsi, di mana segmen pasar dengan harga terjangkau mendominasi jumlah penjualan atau ketersediaan data.

Berdasarkan hasil analisis model, *Naive Bayes* berhasil digunakan untuk mengklasifikasikan tingkat harga mobil secara global ke dalam beberapa kategori, seperti *Low*, *Medium*, dan *High*. Meskipun model sederhana, algoritma ini mampu mengenali pola hubungan antara atribut kendaraan dan kelas harga secara cukup akurat. Hasil klasifikasi menunjukkan bahwa sebagian besar data benar-benar terkelompok ke dalam kategori *Low*, sejalan dengan bentuk distribusi data yang condong ke sisi harga rendah.

Dari hasil uji model, diketahui bahwa atribut yang paling berpengaruh dalam menentukan prediksi harga adalah *Engine Size, Year of Manufacture*, dan *Mileage*. Ukuran mesin (*Engine Size*) berpengaruh karena secara umum mesin yang lebih besar berhubungan dengan performa tinggi dan harga yang lebih mahal. Tahun pembuatan (*Year of Manufacture*) menjadi indikator penting karena mobil yang lebih baru biasanya memiliki harga jual yang lebih tinggi. Sementara itu, jarak tempuh (*Mileage*) menunjukkan tingkat penggunaan kendaraan—semakin tinggi nilai mileage, biasanya semakin rendah harga jualnya.

Dengan demikian, distribusi dan hasil analisis ini memperkuat kesimpulan bahwa *Naive Bayes* cukup efektif dalam memahami pola harga mobil global, meskipun distribusi data yang tidak seimbang (mayoritas harga rendah) dapat menyebabkan model lebih sensitif terhadap kategori tertentu. Analisis ini memberikan gambaran penting mengenai karakteristik pasar mobil dan faktor-faktor utama yang menentukan nilai jual kendaraan.

3.5 Pembahasan

Hasil pelatihan model *Machine Learning* menggunakan algoritma *Naive Bayes* menunjukkan performa yang cukup tinggi dalam memprediksi kategori harga mobil berdasarkan data *Global Car Sales Analysis*. Setelah dilakukan tahap *preprocessing* yang meliputi pembersihan data, label encoding, dan pembagian *dataset* menjadi *training set* dan *testing*

Copyright © 2025 The **Dedek Indra Gunawan Hts**, Page 905
This Journal is licensed under a Creative Commons Attribution 4.0 International License

set dengan rasio 80:20, model dilatih menggunakan variabel prediktor seperti *Manufacturer, Model, Engine size, Fuel type, Year of manufacture,* dan *Mileage*.

Model menghasilkan tingkat akurasi sebesar 96%, dengan nilai *precision* rata-rata 0.97, *recall* 0.96, dan *F1-score* 0.96. Hasil ini menunjukkan bahwa model *Naive Bayes* memiliki kemampuan yang sangat baik dalam melakukan klasifikasi harga mobil secara umum. Secara lebih rinci, performa model berbeda di setiap kelas harga. Kategori *Low* memiliki nilai *F1-score* tertinggi yaitu 0.98, menunjukkan bahwa model sangat akurat dalam mengenali mobil dengan harga rendah. Hal ini disebabkan oleh dominasi jumlah data pada kategori tersebut (9.748 sampel), sehingga model memiliki banyak contoh untuk belajar pola harga rendah secara konsisten.

Sebaliknya, pada kategori *Medium* dan *High*, nilai *F1-score* masing-masing adalah 0.31 dan 0.43, yang menunjukkan tingkat akurasi lebih rendah. Penyebab utama adalah ketidakseimbangan data (*class imbalance*), di mana jumlah data pada kelas menengah (213 data) dan tinggi (39 data) jauh lebih sedikit dibandingkan kelas rendah. Kondisi ini membuat model cenderung "berpihak" pada kelas mayoritas (*Low*), yang merupakan kelemahan umum dari algoritma probabilistik seperti *Naive Bayes*. Meskipun demikian, jika dilihat dari hasil *weighted average* seluruh metrik (*precision*, *recall*, dan *F1-score*), nilai yang diperoleh tetap tinggi yaitu 0.96, yang menegaskan bahwa secara keseluruhan model memiliki kinerja yang konsisten dan dapat diandalkan.

Analisis lebih lanjut menunjukkan bahwa atribut yang paling berpengaruh terhadap klasifikasi harga mobil *adalah Engine size, Year of manufacture*, dan *Mileage*. Kapasitas mesin yang besar dan tahun pembuatan yang lebih baru cenderung meningkatkan harga kendaraan, sedangkan jarak tempuh yang tinggi (*Mileage* besar) menurunkan harga mobil karena menandakan tingkat penggunaan yang lebih tinggi. Temuan ini sejalan dengan teori ekonomi otomotif yang menyatakan bahwa nilai kendaraan berkurang seiring usia dan jarak tempuhnya.

Hasil penelitian ini sejalan dengan beberapa studi terdahulu yang juga menerapkan algoritma *Naive Bayes* untuk prediksi harga kendaraan. Misalnya, penelitian oleh [22] untuk memprediksi harga mobil dengan algoritma *Naïve Bayes* dengan *Brand*, Tipe Mobil, CC, Transmisi dan negara. Berdasarkan penelitian tersebut menghasilkan akurasi 95,38% dengan nilai *precission* 94,96% dan *recall* sebesar 90,21%. Sementara itu, studi oleh [2] menggunakan pendekatan serupa dengan *Naive Bayes* pada dataset mobil di India dan memperoleh akurasi 95%, dengan kesimpulan bahwa ketidakseimbangan data sangat memengaruhi hasil klasifikasi pada kategori harga tinggi.

Jika dibandingkan, hasil penelitian ini menunjukkan peningkatan akurasi hingga 96%, yang menandakan bahwa model mampu mempelajari distribusi data secara lebih optimal, meskipun tetap menghadapi tantangan yang sama dalam menangani data dengan jumlah yang tidak seimbang.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma *Naive Bayes* mampu memprediksi kategori harga mobil global dengan performa yang sangat baik, ditunjukkan oleh akurasi mencapai 96% setelah melalui proses *preprocessing*, encoding atribut kategorikal, dan pembagian data menjadi *training* dan *testing set*. Model bekerja sangat efektif pada kategori harga *Low*, dengan *precision* 0.99 dan *F1-score* 0.98, namun performanya menurun pada kategori *Medium* dan *High* akibat distribusi data yang tidak seimbang, sehingga prediksi untuk kedua kelas tersebut kurang stabil. Secara keseluruhan, *Naive Bayes* terbukti menjadi algoritma yang cepat, sederhana, dan cukup akurat untuk klasifikasi harga mobil berbasis data global, meskipun perbaikan seperti penyeimbangan data, penambahan fitur, atau penggunaan algoritma lain dapat meningkatkan kemampuan prediksi terutama pada kelas harga menengah dan tinggi.

REFERENCES

[1] M. Z. Ahmad, Muhammad; Farooq, Muhammad Ali; Hussain, M. Z. Hasan, M. Muzzamil, and A. Khalid, "Car Price Prediction using *Machine Learning*," *2024 IEEE 9th Int. Conf. Converg. Technol.*, 2024, [Online]. Available: 10.1109/I2CT61223.2024.10544124.

[2] M. Poorv, Y. K. . Gupta, and A. K. Sharma, "Evaluating *Machine Learning* models for used car price estimation: a comparative study," *2nd Int. Conf. Pervasive Comput. Adv. Appl. (PerCAA 2024)*, 2024, [Online]. Available: 10.1049/icp.2025.0793.

[3] M. Devanda, H. Kusuma, and S. Hidayat, "Penerapan Model Regresi Linier dalam Prediksi Harga Mobil Bekas di India dan Visualisasi dengan Menggunakan Power Abstrak," vol. 5, no. 2, pp. 1097–1110, 2024.

[4] E. Gegic, B. Isakovic, D. Keco, Z. Masetic, and J. Kevric, "Car Price Prediction using *Machine Learning* Techniques," pp. 113–118, 2019, doi: 10.18421/TEM81-16.

[5] B. E. Putro and D. Indrawati, "Data Mining Analytics Application for Estimating Used Car Price During the Covid-19 Pandemic in Indonesia," vol. 6869, 2019, doi: 10.23917/jiti.v21i2.18975.

[6] J. Yang, J. Kim, H. Ryu, J. Lee, and C. Park, "Predicting Car Rental Prices : A Comparative Analysis of *Machine Learning* Models," *Electronics*, pp. 1–20, 2024, [Online]. Available: https://doi.org/10.3390/electronics13122345.

[7] V. Nakhipova, Y. Kerimbekov, Z. Umarova, L. Suleimenova, and S. Botayeva, "Use of the *Naive Bayes* Classifier Algorithm in *Machine Learning* for Student Performance Prediction," *Int. J. Inf. Educ. Technol.*, vol. 14, no. 1, 2024, doi: 10.18178/ijiet.2024.14.1.2028.

[8] E. K. Ampomah, G. Nyame, P. C. Addo, and M. Gyan, "Stock Market Prediction with Gaussian Naïve Bayes *Machine Learning* Algorithm," *Inform.*, vol. 45, pp. 243–256, 2021.

[9] R. Syahputra, G. J. Yanris, and D. Irmayani, "SVM and Naïve Bayes Algorithm Comparison for User Sentiment Analysis on Twitter," *Sink. J. dan Penelit. Tek. Inform.*, vol. 7, no. 2, pp. 671–678, 2022.

[10] M. Afriansyah, J. Saputra, Y. Sa, V. Yoga, and P. Ardhana, "Optimasi Algoritma Naïve Bayes Untuk Klasifikasi Buah Apel

Copyright © 2025 The **Dedek Indra Gunawan Hts**, Page 906
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Berdasarkan Fitur Warna RGB," *Bull. Comput. Sci. Res.*, vol. 3, no. 3, pp. 242–249, 2023, doi: 10.47065/bulletincsr.v3i3.251.
[11] P. Ramadani, R. Fadillah, Q. Adawiyah, B. Restu, and A. Ghazali, "Perbandingan Algoritma Naïve Bayes , C4 . 5 , dan K-Nearest Neighbor untuk Klasifikasi Kelayakan Program Keluarga Harapan JURNAL MEDIA INFORMATIKA [JUMIN]," vol. 6, no. 1, pp. 775–782, 2024.
[12] M. N. Fuad *et al.*, "Penerapan Metode Naïve Bayes Untuk Penentuan Kelayakan Pembuatan Sertifikat Tanah Berbasis Web," vol. 5, no. 1, 2024.
[13] E. Yusuf, "Penerapan Model Klasifikasi dalam Kelayakan Pemilihan Bangunan Rumah KPR berbasis *Naive Bayes*," vol. 9, no. 1, pp. 38–50.
[14] H. Imanuel and Samsoni, "PREDIKSI HARGA MOBIL BEKAS MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR," vol. 3, no. 3, pp. 24–34, 2025.
[15] S. Informasi, "Optimasi Support Vector Machine (SVM) Menggunakan *Naive Bayes* dan Decision Tree untuk Klasifikasi Tema Tugas Akhir Mahasiswa Sistem Informasi," vol. 2, no. 2, pp. 91–104, 2024.
[16] P. Bhatnagar and F. Flammini, "An Analysis of Car Price Prediction using *Machine Learning*," *Proc. 2024 9th Int. Conf. Mach. Learn. Technol.*, pp. 11–15, 2024, doi: 10.1145/3674029.3674032.
[17] T. Gori *et al.*, "PREPROCESSING DATA DAN KLASIFIKASI UNTUK PREDIKSI KINERJA DATA PREPROCESSING AND CLASSIFICATION FOR PREDICTING STUDENT," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 1, pp. 215–224, 2024, doi: 10.25126/jtiik.20241118074.
[18] S. N. Sofyan, Z. Sitorus, and M. Iqbal, "Analysis of Public Sentiment Towards Tax Increases Impacting Unemployment Using SVM and Multinomial *Naive Bayes* Methods," *JURIKOM (Jurnal Ris. Komputer)*, vol. 12, no. 4, pp. 555–566, 2025, doi: 10.30865/jurikom.v12i4.8922.
[19] E. Frank and I. A. N. H. Witten, "Technical Note : *Naive Bayes* for Regression," *Mach. Learn.*, pp. 5–25, 2000.
[20] Y. A. Suwitono and F. J. Kaunang, "Implementasi Algoritma Convolutional Neural Network (CNN) Untuk Klasifikasi Daun Dengan Metode Data Mining SEMMA Menggunakan Keras," *J. Komtika (Komputasi dan Inform.*, vol. 6, no. 2, pp. 109–121, 2022.
[21] A. S. Mubarokah, "Penerapan Algoritma Naïve Bayes untuk Analisis Sentimen Ulasan Pengguna Aplikasi Adakami di Google Play Store," vol. 13, no. 3, pp. 2295–2305.
[22] A. Hasyim, M. Fatchan, and W. Hadikristanto, "Penerapan Algoritma Naïve Bayes Dalam Memprediksi Tingkat Penjualan Mobil Tahun 2022," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 02, pp. 207–215, 2022.

Copyright © 2025 The **Dedek Indra Gunawan Hts**, Page 907
This Journal is licensed under a Creative Commons Attribution 4.0 International License