**Implementasi Algoritma Naїve Bayes Untuk Memprediksi**

**Tingkat Penjualan Mobil Tahun 2024**

**Yasser Arafat1, Ade Setiawan2**

1,2Program Studi Teknologi Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika Jl.

Kramat Raya No. 98 Senen, Jakarta Pusat – 10420, Indonesia

Email: yssrarft888444@gmail.com

|  |
| --- |
| Artikel Info : Diterima : 00-00-0000 | Direvisi : 00-00-0000 | Disetujui : 00-00-0000 |

**Abstrak** - Di website ini, GAIKINDO menyediakan berbagai data dan informasi terkait dengan industri otomotif, termasuk data penjualan mobil bulanan dan tahunan, laporan dan analisis industri, serta berita dan informasi terbaru. Pada pembuatan skripsi ini, Penulis mencari permasalahan pada data penjualan mobil kedepannya, penelitian ini bertujuan untuk memprediksi penjualan mobil di Indonesia tahun 2024 menggunakan algoritma Naive Bayes. Algoritma ini dipilih karena kesederhanaannya dan kemampuannya untuk memprediksi dengan akurasi yang wajar. Proses awal melibatkan menganalisa data penjualan mobil dari website GAIKINDO yang berfokus pada periode Januari-April 2024. Setelah itu, untuk mempermudah desain dan pengujian, peneliti menggunakan teknik analisa data mining Cross Industry Standard Process for Data Mining (CRISPDM) dengan langkah-langkah: Pengetahuan Bisnis, Fase Pemahaman Data, Fase Persiapan Data, Fase Pembangunan Model, Fase Evaluasi, Fase Implementasi. Hasil recall klasifikasi naive bayes pada gambar menunjukkan bahwa tingkat recall mencapai 100%. sehingga, dapat disimpulkan bahwa hasil akurasi klasifikasi naive bayes pada ketiga gambar mencapai nilai sebagai berikut: akurasi 94,21%, presisi 95,24% dan recall 100%. hal ini menunjukkan bahwa metode naive bayes merupakan metode yang cukup baik dalam penelitian ini.

Kata Kunci : Algoritma, Naïve Bayes, Klasifikasi, Akurasi, RapidMiner

***Abstracts*** *- On this website, GAIKINDO provides various data and information related to the automotive industry, including monthly and annual car sales data, industry reports and analysis, as well as news and information on the automotive industry. sales data, industry reports and analysis, as well as the latest news and information. news and information. In making this thesis, the author is looking for problems in car sales data. car sales data in the future, this research aims to predict car sales in Indonesia in 2024 using the in Indonesia in 2024 using the Naive Bayes algorithm. This algorithm algorithm was chosen for its simplicity and its ability to predict with reasonable accuracy. reasonable accuracy. The initial process involves analyzing car sales data from the GAIKINDO website, focusing on the January-April 2024 period. After that, to simplify the design and testing, researchers used the Cross Industry Standard Process for Data Mining (CRISPDM) data mining analysis technique with the following steps: 1. Standard Process for Data Mining (CRISPDM) with steps: Business Knowledge, Data Understanding Phase, Data Preparation Phase, Model Building Phase, Evaluation Phase, Implementation Phase. Model, Evaluation Phase, Implementation Phase. The recall results of naive bayes classification in the figure shows that the recall rate reaches 100%. so, it can be concluded that the results of the naive bayes classification accuracy concluded that the results of naive bayes classification accuracy in the three images achieve the following values: 94.21% accuracy, 95.24% precision and 100% recall. 100%. this shows that the naive bayes method is a fairly good method in this research. method in this research.*

*Keywords : Algorithm, Naïve Bayes, Classification, Accuracy, RapidMiner*

**PENDAHULUAN**

Industri otomotif merupakan salah satu sektor penting dalam perekonomian Indonesia. Kontribusi industri ini terhadap PDB nasional cukup signifikan, dan berperan dalam menciptakan lapangan pekerjaan serta mendorong pertumbuhan ekonomi. Selama beberapa tahun terakhir, industri otomotif di Indonesia telah mengalami pertumbuhan yang signifikan. Ini dapat dilihat dari peningkatan jumlah kendaraan bermotor yang terdaftar di Indonesia, yang mencapai angka 107 juta unit pada tahun 2022.

Meskipun industri otomotif mengalami pertumbuhan, namun terdapat fluktuasi dalam tingkat penjualan mobil setiap tahunnya. Fluktuasi ini dipengaruhi oleh berbagai faktor, seperti kondisi ekonomi, kebijakan pemerintah, harga bahan bakar, dan tren pasar. Bagi pelaku bisnis di industri otomotif, memprediksi tingkat penjualan mobil dengan akurat sangatlah penting untuk pengambilan keputusan strategis, seperti perencanaan produksi, penentuan strategi pemasaran, dan pengelolaan inventori.

Website GAIKINDO (https://www.gaikindo.or.id/) merupakan sumber informasi resmi dan terpercaya tentang industri otomotif Indonesia. Di website ini, GAIKINDO menyediakan berbagai data dan informasi terkait dengan industri otomotif, termasuk data penjualan mobil bulanan dan tahunan, laporan dan analisis industri, serta berita dan informasi terbaru.

Penulis melakukan penelitian ini bertujuan untuk memprediksi penjualan mobil di Indonesia tahun 2024 menggunakan algoritma Naive Bayes. Algoritma ini dipilih karena kesederhanaannya dan kemampuannya untuk memprediksi dengan akurasi yang wajar.

Salah satu metode yang dapat digunakan untuk memprediksi tingkat penjualan mobil adalah dengan menggunakan algoritma Machine Learning. Algoritma ini mampu belajar dari data historis dan mengidentifikasi pola yang dapat digunakan untuk memprediksi nilai di masa depan. Naïve Bayes adalah salah satu algoritma Machine Learning yang populer untuk klasifikasi dan prediksi.

Penelitian oleh Hasyim et al. (2022) membahas penggunaan algoritma Naïve Bayes untuk memprediksi tingkat penjualan mobil pada tahun 2022. Tujuan penelitian ini adalah untuk mengklasifikasikan setiap merek mobil berdasarkan kriteria tertentu menggunakan algoritma Naïve Bayes. Hasil penelitian menunjukkan bahwa akurasi klasifikasi dengan algoritma Naïve Bayes mencapai 95,38%, nilai precision sebesar 94,96% dan nilai recall 90,21%. Dalam penelitian ini membuktikan algoritma Naïve Bayes sangat efektif untuk mengklasifikasikan merek mobil berdasarkan data penjualan.

**METODE PENELITIAN**

Rangkaian penelitian di studi ini digambarkan pada Gambar. Proses awal melibatkan menganalisa data penjualan mobil dari *website* GAIKINDO yang berfokus pada periode Januari-April 2024.

Setelah itu, untuk mempermudah desain dan pengujian, peneliti menggunakan teknik analisa data mining *Cross Industry Standard Process for Data Mining* (CRISPDM) dengan langkah-langkah:

1. Pengetahuan Bisnis

Tahap awal, penulis berusaha memahami kendala yang ada didalam penjualan mobil dalam periode awal tahun 2024. Dengan demikian, penulis bisa mengidentifikasi tujuan dan pola yang ingin dicapai melalui mining data.

1. Fase Pemahaman Data

Dalam fase ini, penulis memahami data yang diperlukan untuk mengumpulkan data relevan dan terkait dengan objektif penelitian. Data yang dipakai mencakup output penjualan berbagai kategori dan merek mobil dengan sejumlah atribut, antara lain: Kategori, Merek, Tipe Model, CC, Transmisi, Bahan Bakar, Kapasitas Tangki, Berat Kendaraan Bersih, Perbandingan Gigi, Ukuran Ban, Daya Mesin, Jarak Poros Roda, Dimensi, Jumlah Kursi, Sistem Penggerak, Kecepatan, Jumlah Pintu, Roda, CBU (*Completely Built-Up*), Negara Asal, dan Jumlah penjualan.

1. Fase Persiapan Data

Pada fase ini, penulis memproses data yang diperoleh melalui sejumlah tahapan KDD, seperti pembersihan data, penyatuan data, pemilihan data, dan transformasi data

1. Fase Pembangunan Model

Pada fase ini, penulis menetapkan teknik mining data yang akan digunakan untuk memproses data yang telah dipersiapkan awalnya. Teknik yang digunakan adalah klasifikasi menggunakan algoritma *Naïve Bayes*. Data yang telah diproses setelah itu akan dianalisis menggunakan RapidMiner. Langkah-langkah yang dilakukan pada tahap ini adalah :

1. Penggunaan *RapidMiner* untuk pengujian *Naive Bayes*

Penggunaan *RapidMiner* untuk pengujian *Naive Bayes* bertujuan untuk memudahkan pengolahan data dalam jumlah besar. Pengetesan ini bertujuan menemukan skema algoritma.

1. Uji berdasarkan akurasi

Pengujian ini bertujuan melihat persentase data uji terhadap data latih didalam penjualan mobil pada periode awal tahun 2024.

1. Fase Evaluasi

Melakukan pengujian terhadap berbagai model dengan tujuan menemukan model dengan ketepatan tertinggi. Pada fase evaluasi, akan terlihat apakah hasil dari tahap pemodelan dapat mencapai tujuan yang telah ditentukan pada fase pertama.

1. Fase Implementasi

Pada fase pembuatan model, dilakukan analisis dan perhitungan dari fase sebelumnya. Pada fase ini, model dengan keakuratan tertinggi diterapkan dan kemudian dapat dipakai untuk mengukur data baru.

**HASIL DAN PEMBAHASAN**

Perhitungan naive bayes dilakukan dengan menghitung menggunakan data yang diambil dari website GAIKINDO sebanyak 486 data training dan 121 data testing. Kriteria yang digunakan adalah sebagai berikut :

1. Class : Kurang Laku, Sedang, Laku dan Sangat Laku

2. Category : Sedan Type, 4x2 Type, 4x4 Type, Bus, Pick Up, Truck, Double Cabin 4x2/4x4,

 Affordale Energy Saving Cars

3. Brand : Honda, Toyota, Audi, BMW, Hyundai, Lexus, Mazda, Nissan, Chevrolet, Daihatsu,

Datsun, KIA, Mini, Mitsubishi Motors, Renault, Suzuki, Volkswagen Mitsubishi Fuso, Hino, Scania, DFSK, TATA, FAW, Isuzu, Scania, UD Trucks, Wuling

4. Type Model : Semua brand mobil berdasarkan tipe dan jenisnya

5. CC : <1.500, 1.501 – 2.500, 2.501 – 3.000, 3.001 – 16.000

6. Transmisi : Automatic Transmision (AT), Manual Transmision (MT), Continiuously Variable

Transmision (CVT)

7. Origin Country : Negara yang memproduksi mobil tersebut

8. Output : Jumlah total pembelian konsumen selama empat bulan berturut-turut dari Januari,

Februari, Maret dan April 2024

Berikut yaitu Tabel Data Traning dari data set GAIKINDO Pada bulan Januari - April 2024 sebanyak 486 Data yang terdiri dari 8 atribut (80% dari 608 Data set hanya 10 yang ditampilkan) dan Data Testing sebanyak 121 Data (20% dari 608 Data set hanya 10 yang ditampilkan) :

**Tabel 1** Traning Set yang Telah Diolah



**Tabel 2** Testing Set yang Telah Diolah

Data diklasifikasikan ke dalam empat kategori: Kurang Laku, Sedang, Laku, dan Sangat Laku. Dalam proses ini, Parameter Set Role diterapkan dalam data latih dengan nama atributt: Clαss. Proses tersebut memakai data latih dan uji.

 Menunjukkan pemrosesan data memakai *Naive Bayes* pada *RapidMiner* :

**Gambar 1** Permodelan Klasifikasi Naïve Bayes memakai Rapid Miner

 Hasil *Prediction* data uji dan tingkat *Confidence* dalam setiap perhitungan data dengan teknik Naïve Bayes dengan RapidMiner, ditampilkan pada gambar dibawah. (Ditampilkan hanya beberapa bagian teratas).

**Gambar 2** Hasil Example Set Naïve Bayes

Hasil data latih dengan Naïve Bayes yang ditampilkan bentuk *plot view* pada *Example Set* ditampilkan pada Gambar 4.14 :

**Gambar 3** Hasil Plot View Naïve Bayes menggunakan Rapid Miner

Hasil plot view Naïve Bayes pada gambar menunjukkan bahwa merek mobil dengan kategori Affordable Energy Saving Cars adalah kategori yang paling sering dipesan oleh customer.

****Analisa Akurasi Klasifikasi *Naive Bayes***

**Gambar 4** Menunjukkan Akurasi Klasifikasi Naïve Bayes Menggunakan Rapidminer

 Hasil *Recall* Klasifikasi *Naïve Bayes* pada gambar menampilkan bahwa tingkat *Recall* mencapai 100%. Sehingga, dapat disimpulkan bahwa hasil akurasi klasifikasi *Naïve Bayes* pada ketiga gambar mencapai nilai sebagai berikut : akurasi 94,21%, presisi 95,24%, dan *recall* 100%. Hal ini menunjukkan bahwa metode *Naïve Bayes* merupakan metode yang cukup baik dalam penelitian ini.

**KESIMPULAN**

Berdasarkan pembahasan yang telah diuraikan mengenai penerapan Algoritma *Naïve Bayes* dalam menentukan Merek Mobil Terlaris, dapat, disimpulkan sebagai berikut:

1. Berdasarkan data dari GAIKINDO untuk periode Januari hingga April tahun 2024, klasifikasi menggunakan Metode *Naïve Bayes* menunjukkan bahwa merek mobil terlaris adalah TOYOTA berdasarkan hasil dari *Rapid Miner*. Mobil yang paling banyak disukai customer adalah HONDA Brio Satya E AT.
2. Berdasarkan hasil perhitungan Metode Naïve Bayes menggunakan Rapid Miner, nilai akurasi adalah 94,21%, presisi 95,24%, dan *recall* 100%. Hal ini menunjukkan bahwa Metode *Naïve Bayes* memiliki tingkat akurasi yang rendah saat digunakan denga data yang besar.

**REFERENSI**

Aji, N. M., Atina, V., dan Sudibyo, N. A. (2023). *PEMODELAN PREDIKSI KELULUSAN MAHASISWA DENGAN METODE NAÏVE BAYES DI UNIBA*. https://doi.org/10.36595/misi.v5i2

Alkhairi, P., dan Windarto, A. P. (n.d.). *Seminar Nasional Teknologi Komputer dan Sains (SAINTEKS) Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara*. https://seminar-id.com/semnas-sainteks2019.html

Hasyim, A., Fatchan, M., & Hadikristanto, W. (2022). Penerapan Algoritma Naïve Bayes Dalam Memprediksi Tingkat Penjualan Mobil Tahun 2022. Jurnal Ilmiah Intech : Information Technology Journal of UMUS, 4(02), 207–215.

Hermawan, L., Ismiati, M. B., Bangau, J., 60, N., dan Charitas, M. (2020). Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval. *TRANSFORMATIKA*, *17*(2), 188–199.

Indriyani, F., Yunita, Muthia, D. A., Surniandari, A., dan Sriyadi. (2019). *Analisa Perancangan Sistem Informasi*.

Mas Pintoko, B., dan Muslim, K. (2019). *Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier*.

Monowati, I. T., & Setyadi, R. (2023). Penerapan Algoritma Naïve Bayes Dalam Memprediksi Pengusulan Penghapusan Peralatan dan Mesin Kantor. Journal of Information System Research (JOSH), 4(2), 483–491. https://doi.org/10.47065/josh.v4i2.2674

Muslim, M. A., Prasetiyo, B., Mawarni, E. laily H., Herowati, A. J., Mirqotussa’adah, Rukmana, S. H., dan Nurzahputra, A. (2019). *Data Mining Algoritma C4.5 Disertai contoh kasus dan penerapannya dengan program computer*.

Musu, W., dan Ibrahim, A. (2021). *Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4.5*.

Nasrullah, A. (2023). Data Mining Algoritma Naive Bayes Untuk Memprediksi Jumlah Siswa Baru. *EJECTS : E-Journal Computer, Technology and Informations System*, *2*(2).

Prasetyo, V. R., Lazuardi, H., Mulyono, A. A., & Lauw, C. (2021). Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression. Jurnal Nasional Teknologi Dan Sistem Informasi, 7(1), 8–17. https://doi.org/10.25077/teknosi.v7i1.2021.8-17

Putri, S. E. Y., Saniati, dan Surahman, A. (2021). *PENERAPAN MODEL NAIVE BAYES UNTUK MEMPREDIKSI POTENSI PENDAFTARAN SISWA DI SMK TAMAN SISWA TELUK BETUNG BERBASIS WEB*.

Rachman, R., Handayani, R. N., dan Artikel, I. (2021). Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM. *JURNAL INFORMATIKA*, *8*(2). http://ejournal.bsi.ac.id/ejurnal/index.php/ji

Rifky, M. L., Nugraha, Z., Saputra, M. B., Pratama, D., Raswir, E., dan Pratama, Y. (2022). *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM) Implementasi Data Mining Untuk Penjualan Mobil Menggunakan Metode Naive Bayes*. https://ejournal.unama.ac.id/index.php/jakakom.

Siddik, M., dan Desnelita, Y. (2019). Penerapan Naïve Bayes untuk Memprediksi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademis. *Jurnal Infomedia*, *2*(4).