

Penerapan Algoritma Naïve Bayes Dalam Memprediksi Tingkat Penjualan Mobil Tahun 2022

Application of the Naïve Bayes Algorithm to Predicting the Level of Car Sales in 2022

Abdul Hasyim¹, Muhamad Fatchan*², Wahyu Hadikristanto³

^{1,2,3} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa

*Correspondence author email: *fatchan@pelitabangsa.ac.id*

Abstrak

Persaingan industri otomotif terutama pada sektor roda empat atau mobil setiap bulan semakin kompetitif. Berbagai merk mobil saling berlomba untuk menjadi primadona bagi konsumennya terutama masyarakat Indonesia. Tidak dipungkiri hal ini membuat pabrikan atau perusahaan roda empat saling berlomba untuk meningkatkan penjualan mobilnya di sepanjang tahun 2020, tercatat lebih dari 10 merk mobil yang laku dimasyarakat seperti Toyota, Hino, Honda dan yang lainnya. Hal ini mendorong penelitian dengan konsep data mining menggunakan algoritma naïve Bayes yang bertujuan untuk dapat mengklasifikasikan setiap merk mobil berdasarkan kriteria tertentu. Algoritma ini mencatatkan akurasi sebesar 95,38 % dengan nilai precision 94,96% dan recall sebesar 90,21 %.

Katakunci: penjualan, mobil, klasifikasi, Algoritma Naïve Bayes

Abstrack

Competition in the automotive industry, especially in the four-wheel or car sector is getting more competitive every month. Various car brands are competing with each other to be excellent for their consumers, especially the Indonesian people. It is undeniable that this makes four-wheeled manufacturers or companies compete with each other to increase sales of their cars throughout 2020, with more than 10 car brands that are selling well in the community such as Toyota, Hino, Honda and others. This encourages research with the concept of data mining using a naïve Bayes algorithm which aims to be able to classify each car brand based on certain criteria. This algorithm records an accuracy of 95.38% with a precision value of 94.96% and a recall of 90.21%.

Keyword: sales, cars, classification, Naïve Bayes Algorithm

PENDAHULUAN

Persaingan bisnis dalam era globalisasi dewasa ini sangat ketat, di mana setiap perusahaan dituntut untuk dapat memenuhi kebutuhan konsumen serta berusaha untuk menciptakan suatu produk yang mempunyai keunggulan dan menciptakan produk yang berbeda dengan pesaing. Dengan demikian usaha pengembangan produk yang berbeda, dapat menjadi suatu strategi yang efektif bagi perusahaan dalam memberikan penawaran produk yang inovatif sehingga tercapai suatu kepuasan masing-masing pihak[1], baik dari pembeli karena membeli produk yang sesuai dengan kebutuhannya dan selernya maupun bagi pihak perusahaan yang ingin mendapatkan keuntungan dari penjualan produk produknya dan juga menjaga citra baik perusahaan di mata konsumen.

Minat beli (*purchase intention*) merupakan sesuatu yang berhubungan dengan rencana konsumen untuk membeli produk/jasa tertentu, serta berapa banyak unit produk/ jasa yang dibutuhkan pada periode tertentu[2]. Merk suatu produk akan mempengaruhi minat beli konsumen dimana setiap merek mempunyai sifat khas, dan sifat khas inilah yang membedakan produk yang satu dengan produk yang lainnya[3]. Permintaan akan sebuah produk barang yang semakin berkualitas membuat perusahaan berlomba-lomba meningkatkan kualitas produk dan

Informasi Artikel:

Submitted: Oktober 2022, **Accepted:** November 2022, **Published:** November 2022

ISSN: 2685-4902 (media online), Website: <http://jurnal.umus.ac.id/index.php/intech>

mempertahankan citra merek produk yang mereka miliki. Dengan memperhatikan citra merek dan kualitas produksinya, setiap perusahaan khususnya industri mobil terus bersaing untuk bisa meningkatkan penjualan mobilnya kepada *customer*, sehingga dengan melihat grafik penjualan setiap perusahaan akan mengetahui berdasarkan apa saja minat konsumen membeli mobil tersebut.

Data mining merupakan salah satu cabang ilmu komputer yang saat ini banyak digunakan di era big data saat ini[4], setiap detiknya terdapat jutaan data yang diperoleh sehingga peran data mining sangatlah penting guna mencari informasi yang terpendam dari suatu data. Salah satu teknik data mining yang dapat digunakan dari analisa di atas adalah klasifikasi, yang mampu memprediksi atau memperkirakan suatu kejadian yang belum terjadi hanya dengan seringnya melihat pola dari suatu kejadian secara berulang[5]. Maka dari itu teknik klasifikasi dianggap sesuai untuk memprediksi tingkat penjualan mobil guna mengetahui market pasar untuk bulan dan tahun berikutnya[6]. Algoritma klasifikasi Naïve Bayes merupakan yang paling populer dan banyak digunakan. Algoritma ini memegang asumsi akan hubungan antar fitur atau atributnya yang independen sehingga menjadikannya sederhana dan efisien[7]. Penerapan naïve bayes ini diharapkan dapat menilai akurasi dengan prediksi penjualan mobil pada tahun 2020. Dengan hasil klasifikasi data mining akan dihasilkan merek penjualan mobil paling laku sepanjang tahun 2020 yang dapat digunakan acuan untuk menentukan market pasar penjualan di industri otomotif. Selain itu informasi yang dihasilkan dari data mining dapat digunakan sebagai rekomendasi pada perusahaan untuk meningkatkan produksi. Berdasarkan hal tersebut maka penelitian ini bertujuan untuk melakukan klasifikasi terhadap berbagai kategori penjualan mobil serta mengetahui minat beli konsumen terhadap merk suatu mobil.

Penelitian mengenai penerapan Naïve Bayes pernah dilakukan untuk memprediksi penggunaan listrik dalam rumah tangga dengan menggunakan 50 data uji[8], Naïve Bayes pun pernah digunakan untuk melakukan klasifikasi terhadap pelanggan yang akan diberikan bonus [9]. Apabila dilihat dari nilai akurasi algoritma Naïve Bayes memiliki akurasi sebesar 87,24% saat melakukan prediksi terhadap penerimaan mahasiswa baru Universitas Stikubank Semarang [10]. Sedangkan akurasi yang dihasilkan dalam melakukan klasifikasi tingkat rumah tangga miskin dengan Naïve Bayes menghasilkan nilai 93,21% melalui pengujian *10-Fold Cross Validation*[11].

METODE PENELITIAN

Alur penelitian yang dilakukan dalam penelitian ini tergambar pada Gambar 1. Pada tahap awal akan dilakukan analisis dataset penjualan mobil dari situs GAIKINDO yang terfokus pada bulan januari hingga juli 2020, dengan jumlah data sebanyak 712 data dengan 20 atribut. Selanjutnya untuk memudahkan proses desain dan pengujian peneliti akan menggunakan teknik analisis data data mining yaitu CRISP – DM (*Cross Standart Industries for Data Mining*) dengan tahapan:

a. *Bisnis Understanding*

Pada tahapan pertama penulis mencoba untuk memahami permasalahan yang ada pada penjualan mobil sepanjang pertengahan tahun 2020. Sehingga dapat menentukan tujuan dan pola yang akan didapatkan dengan *data mining*.

b. *Data Understanding Phase*

Pada tahap ini penulis melakukan pemahaman terhadap data yang dibutuhkan, untuk kemudian mengambil data yang relevan dan memiliki keterkaitan dengan tujuan penelitian. Adapun data yang digunakan yaitu data hasil penjualan berbagai kategori dan merk mobil yang memiliki beberapa atribut diantaranya: *Category, Brand, Type Model, Cc, Trans, Fuel, Tank Capt, Gvw, GearRatio, Tyre Size, Ps/Hp, W Heel Base, Dimension, Seater, Drive Sys, Speed, Door, W Heels, Cbu, Origin Country, Amount of sales*.

c. *Data Preparation Phase*

Pada tahap ini penulis mengolah data yang didapat dengan beberapa tahapan pada KDD seperti melakukan pembersihan terhadap data / *data cleaning*, melakukan integrasi data / *data integration*, melakukan pemilihan data / *data selection* dan transformasi pada data / *data transformation*

d. *Modelling Phase*

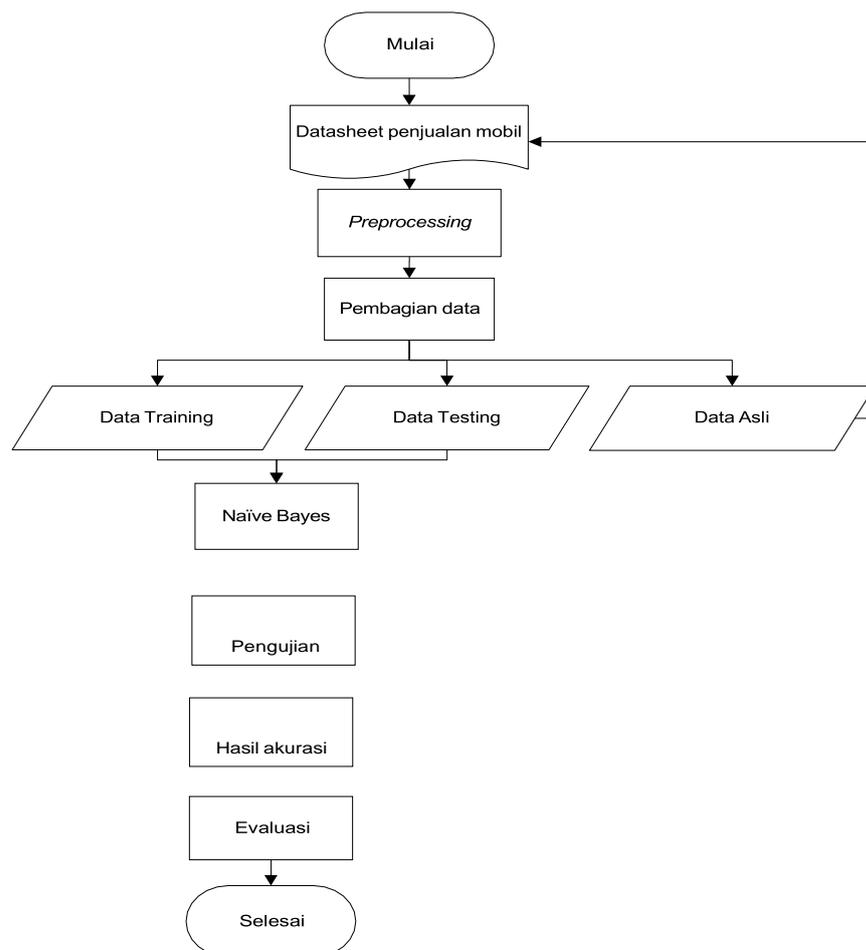
Pada tahap ini penulis menentukan teknik *data mining* yang digunakan untuk mengolah data yang sudah disiapkan sebelumnya. Teknik yang dilakukan yaitu dengan klasifikasi menggunakan algoritma Naïve Bayes. Data yang sudah diolah kemudian akan dilakukan perhitungan dengan menggunakan *tools rapidminer*. Dua langkah yang dilakukan pada tahap ini ialah:

(1) Pengujian Naïve Bayes menggunakan *rapidminer*

Pengujian dengan menggunakan *rapidminer* bertujuan untuk memudahkan dalam pemrosesan data yang berjumlah besar. Pengujian ini bertujuan untuk mendapatkan pola algoritma

(2) Pengujian dengan melihat akurasi

Pengujian ini bertujuan untuk melihat berapa prosentasi data testing terhadap data training pada penjualan mobil sepanjang pertengahan tahun 2020.



Gambar 1. Alur penelitian yang dilakukan

e. *Evaluation Phase*

Melakukan pengujian terhadap model-model yang bertujuan untuk mendapatkan model yang paling akurat. Pada tahap evaluasi, akan diketahui apakah hasil dari tahap pemodelan dapat menjawab tujuan yang telah ditetapkan pada fase pertama.

f. *Deployment Phase*

Pembentukan model selanjutnya melakukan analisa dan pengukuran pada tahap sebelumnya, pada tahap ini diterapkan model atau *rule* yang paling akurat dan selanjutnya dapat digunakan untuk mengevaluasi data baru.

HASIL DAN PEMBAHASAN

Perhitungan *naive bayes* dilakukan dengan menghitung menggunakan data yang diambil dari website GAIKINDO sebanyak 653 data training dan 133 data testing. Kriteria yang digunakan adalah sebagai berikut :

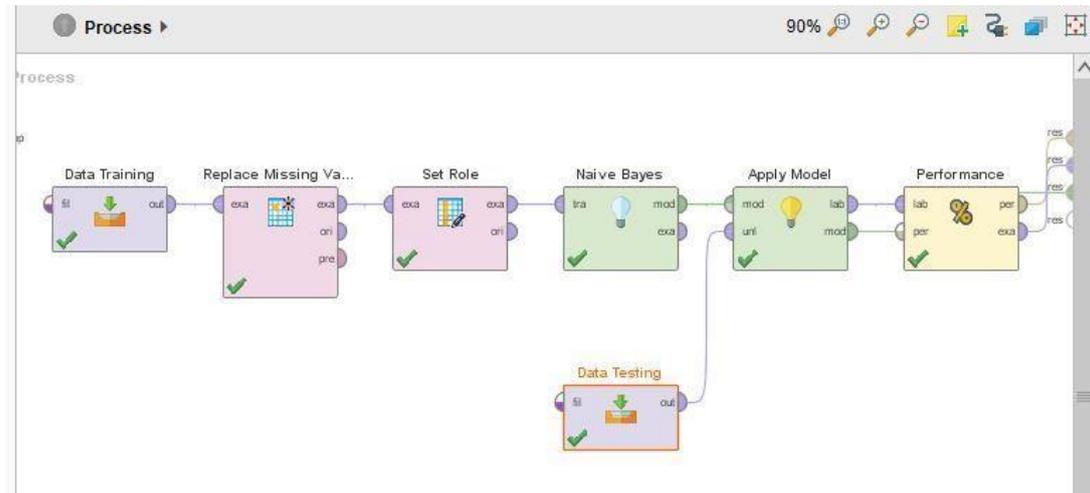
- Class* : Sangat Laku, Laku, Sedang, Kurang laku
- Category* : Sedan type, 4x2 Type, 4x4 Type, Bus, Truck, Pick up, Double Cabin 4x2/4x4, Affordable Energy Saving Cars 4x2
- Brand* : Honda, Toyota, Audi, BMW, Hyundai, Lexus, Mazda, Nissan, Chevrolet, Daihatsu, KIA, Datsun, Mini, Mitsubishi motors, Mitsubishi Fuso, Suzuki, TATA, DFSK, Isuzu, Scania, Ud Truck, Man Truck, Wuling, Renault, FAW, Hino
- Type Model* : Semua brand mobil berdasarkan jenisnya
- CC* : <1500, 15.01 – 2500, 25.01 – 3000, 3.001 – 16000
- Trans*: AT, MT, CVT
- Origin country*: negara tempat produksi
- Amount of sales*: penjualan periode januari – juli 2020

Untuk data training yang akan digunakan adalah 653 data dan 20% diantaranya adalah sebagai data tesing sebanyak 130 data. Rangkuman mengenai dataset training tersebut terdapat pada Tabel 1.

Table 1. Training set

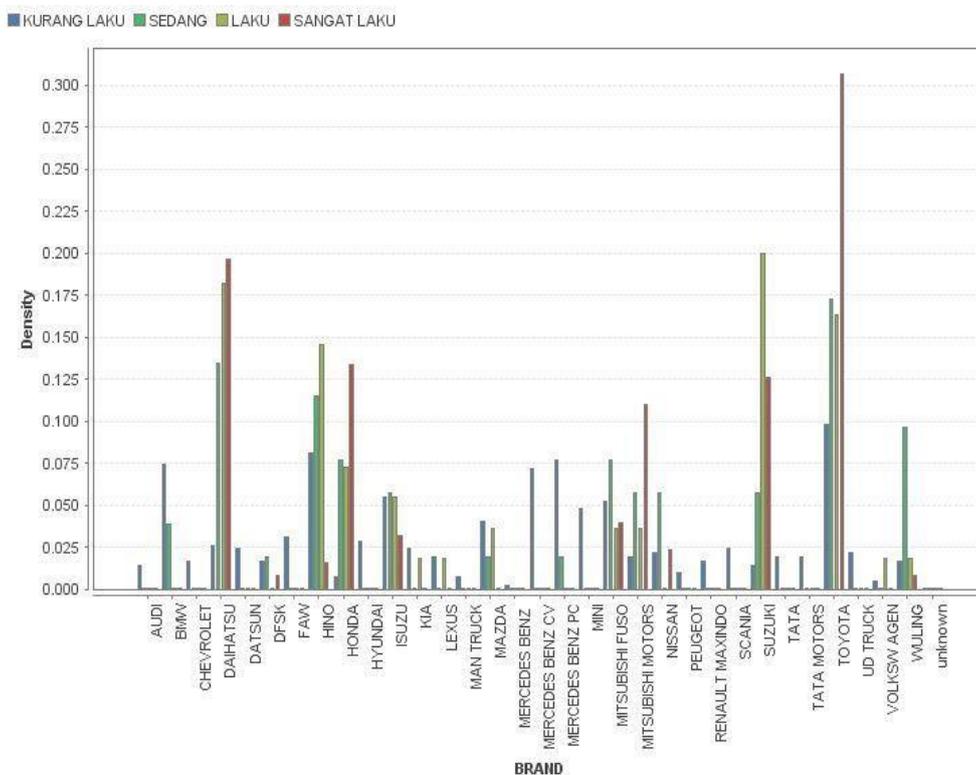
Category	Brand	Type model	CC	Trans	Origin country	Amount of sales	Class
Sedan type	Honda	All new city IVTEC E AT	1497	CVT	Thailand	79	Kurang laku
Sedan type	Honda	All new civic	1500	CVT	Thailand	189	Sedang
Sedan type	Mercedes Benz PC	A 200 sedan (V177)	1332	AT	Ind	46	Kurang laku
Sedan type	Mercedes Benz PC	CLA 200 AMG Line (C118)	1332	AT	Ind	75	Kurang laku
Sedan type	Toyota	Prius	1500	AT	Japan	10	Kurang laku
Sedan type	Toyota	Vios E	1500	MT	Ina	2	Kurang laku
Sedan type	Toyota	Vios G	1500	MT	Ina	34	Kurang laku
Sedan type	Toyota	Vios G AT	1500	AT	Ina	153	Sedang
Sedan type	Audi	A6 2.0TFSI AT	1984	AT	Ina	3	Kurang laku
Sedan type	Audi	A6 3.0 Quattro AT	2995	AT	Germany	1	Kurang laku
Sedan type	BMW	8A36 3201	200	AT	Ina	30	Kurang laku
Sedan type	BMW	8A96 3301	200	AT	Ina	45	Kurang laku
Sedan type	BMW	2U72 M2 Competition	2979	AT	Ina	17	Kurang laku

Selanjutnya data tersebut akan diuji akan diuji dengan rapidminer untuk mengetahui pola dan akurasi yang didapatkan dengan algoritma *Naive Bayes*. Gambar 2 adalah hasil uji coba dengan *rapidminer*.



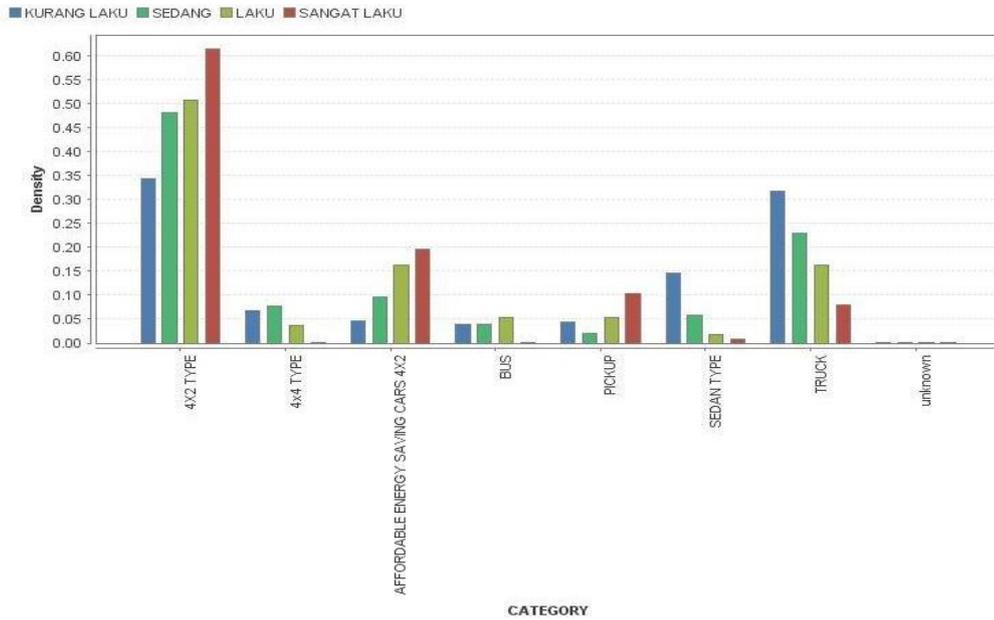
Gambar 2. Klasifikasi Naïve Bayes dengan rapidminer

Berdasarkan gambar 2 data testing akan di *replace missing value* agar tidak ada kekosongan data sehingga data bisa valid untuk digunakan, pada proses tersebut juga digunakan parameter set role dengan menggunakan attribute name: class. Dari hasil pengujian ini didapatkan bahwa implementasi Merk Brand mobil Toyota menempati urutan pertama sebagai Brand mobil paling laku. Gambar 3 menjelaskan hal tersebut.



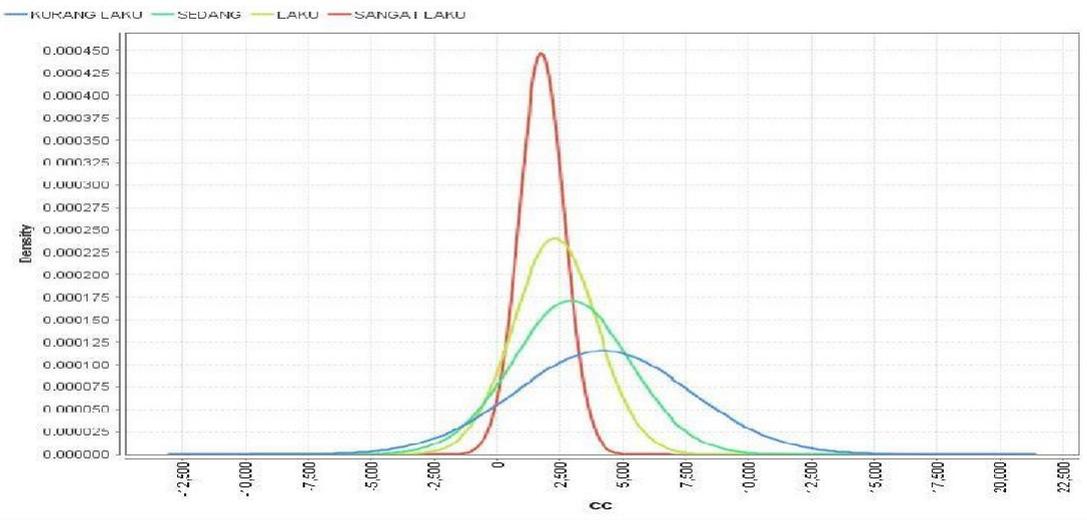
Gambar 2. Penjualan Mobil berdasarkan Brand

Kemudian pada kriteria berdasarkan *Category* penjualan mobil paling laku ada pada *category* 4x2 *Type* disusul *Affordable*, *pick up* dan *Truck*. Gambar 3 Menunjukkan curva penjualan mobil berdasarkan *category*.



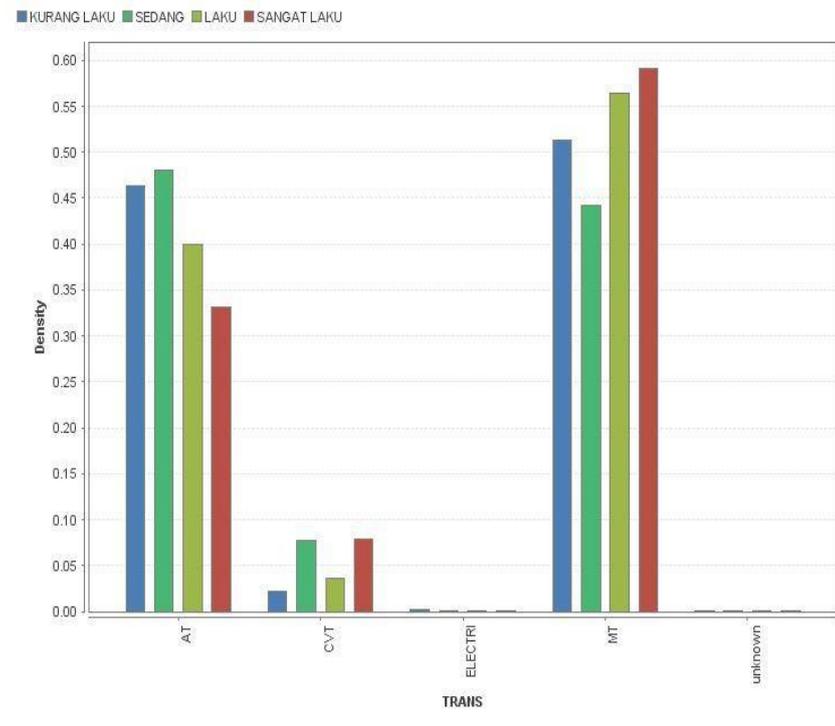
Gambar 3. Penjualan Mobil berdasarkan *category*

Pada kriteria berdasarkan *CC*, penjualan mobil paling laku ada pada nilai $1500 - 5000\text{ cc}$. Gambar 4 menunjukkan curva penjualan mobil berdasarkan *cc*.



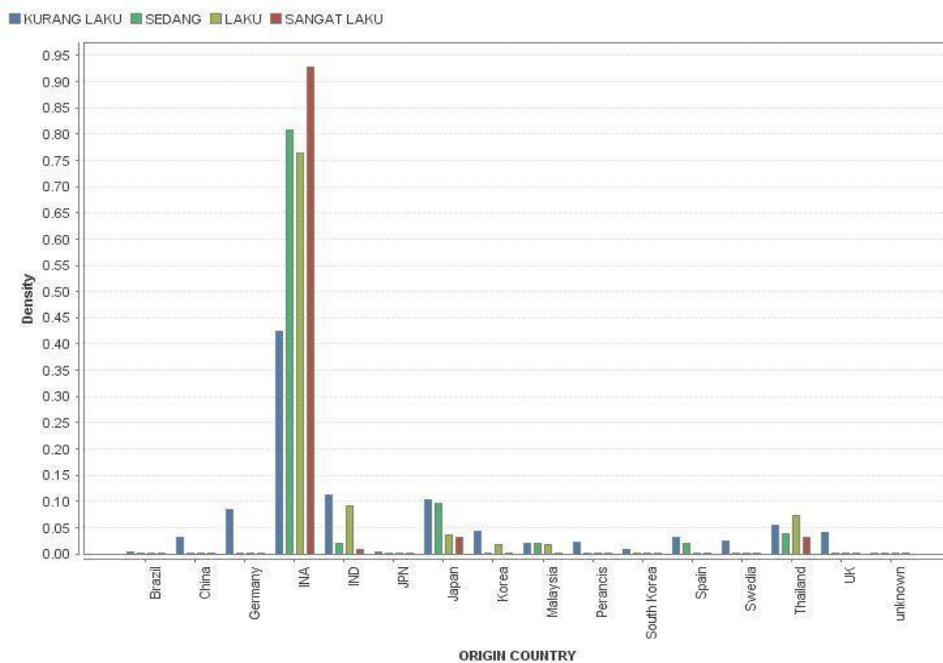
Gambar 4. Penjualan Mobil berdasarkan *cc*

Transmisi *Manual Transmission* (MT) menempati urutan pertama penjualan palinglaku disusul Transmisi *Automatic Transmission* (AT). Gambar 5 menunjukkan kurva penjualan mobil berdasarkan trans.



Gambar 5 Penjualan Mobil berdasarkan Trans

Indonesia menempati urutan pertama sebagai negara tempat produksi mobil palinglaku pada periode Januari hingga Juli 2020. Gambar 6 menunjukkan curva penjualan mobil berdasarkan *Origin Country*



Gambar 6. Penjualan Mobil berdasarkan *Origin*

Analisis akurasi

Berdasarkan hasil uji *performance* terhadap data *training* dan *testing* menggunakan algoritma *naïve bayes* didapatkan hasil *accuracy* 95,38 % dengan nilai *precision* 94,96% dan *recall* sebesar 90,21 %.

accuracy: 95.38%

	true KURANG LAKU	true SEDANG	true LAKU	true SANGAT LAKU	class precision
pred. KURANG LA...	80	4	0	0	95.24%
pred. SEDANG	0	11	2	0	84.62%
pred. LAKU	0	0	14	0	100.00%
pred. SANGAT LAKU	0	0	0	19	100.00%
class recall	100.00%	73.33%	87.50%	100.00%	

Gambar 7. Nilai accuracy

weighted_mean_recall: 90.21%, weights: 1, 1, 1, 1

	true KURANG LAKU	true SEDANG	true LAKU	true SANGAT LAKU	class precision
pred. KURANG LA...	80	4	0	0	95.24%
pred. SEDANG	0	11	2	0	84.62%
pred. LAKU	0	0	14	0	100.00%
pred. SANGAT LAKU	0	0	0	19	100.00%
class recall	100.00%	73.33%	87.50%	100.00%	

Gambar 8. Nilai recall naïve bayes

weighted_mean_precision: 94.96%, weights: 1, 1, 1, 1

	true KURANG LAKU	true SEDANG	true LAKU	true SANGAT LAKU	class precision
pred. KURANG LA...	80	4	0	0	95.24%
pred. SEDANG	0	11	2	0	84.62%
pred. LAKU	0	0	14	0	100.00%
pred. SANGAT LAKU	0	0	0	19	100.00%
class recall	100.00%	73.33%	87.50%	100.00%	

Gambar 9. Nilai Precision Naïve Bayes

KESIMPULAN

Berdasarkan data yang diperoleh, proses Data Mining membantu dalam penerapan metode Naïve Bayes dalam mendapatkan informasi dari hasil prediksi penjualan pada periode Januari hingga Juli Tahun 2020 Brand Toyota menjadi brand paling Laku sepanjang pertengahan tahun ini. Sehingga dengan demikian metode Naïve Bayes ini berhasil memprediksi dengan presentase keakuratan sebesar 95,38 % dengan menggunakan data *Training* sebanyak 653 dan data *Testing* sebanyak 130 data.

DAFTAR PUSTAKA

- [1] R. Astuti, I. Ardila, and R. R. Lubis, "Pengaruh Promosi Dan Kualitas Produk Terhadap Keputusan Pembelian Sepatu Merek Converse," *Akmami Akuntansi, Manajemen, Ekon.*, vol. 2, no. 2, pp. 204–219, 2019.
- [2] M. R. Picaully, "Pengaruh Kepercayaan Pelanggan Terhadap Niat Pembelian Gadget Di Shopee Indonesia," *J. Manaj. Maranatha*, vol. 18, no. 1, pp. 31–40, 2018, doi: 10.28932/jmm.v18i1.1094.
- [3] N. Nurliati and I. Mardian, "Pengaruh Kualitas Produk, Citra Merek Dan Persepsi Harga Terhadap Minat Beli Produk Smartphone Xiaomi Pada Duta Cell Bima," *Manaj. Dewantara*, vol. 5, no. 1, pp. 78–92, 2021, doi: 10.26460/md.v5i1.9651.
- [4] A. Natasuwarna, "Pendekatan Data Mining Memprediksi Profil Sosial Masyarakat Menggunakan Aplikasi RapidMiner," *Semin. Nas. Pengabd. Masy.*, vol. 38, pp. 38–44, 2019.
- [5] A. Riski, I. Pratama, S. A. Latipah, and B. N. Sari, "Optimasi klasifikasi curah hujan menggunakan support vector machine (svm) dan recursive feature elimination (rfe)," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 07, no. 02, pp. 314–324, 2022, doi: 10.29100/jupi.v7i2.2675.g1166.
- [6] H. Budi Jatmiko, N. Tedi Kurniadi, and D. Maulana, "Optimasi Naïve Bayes Dengan Particle Swarm Optimization Untuk Analisis Sentimen Formula E-Jakarta," *JACIS J. Autom. Comput. Inf. Syst.*, vol. 2, no. 1, pp. 22–30, 2022, doi: 10.47134/jacis.v2i1.35.
- [7] A. Army and B. Tujni, "Penerapan Algoritma Naive Bayes Untuk Menentukan Nasabah Potensial Klaim Pada PT Penjaminan Jamkrindo Syariah," *Bina Darma Conf. Comput. Sci.*, vol. 2, no. 1, pp. 207–217, 2020.
- [8] D. Fallo and M. Sogen, "Metode Naïve Bayes Untuk Memprediksi Penggunaan Listrik Rumah Tangga," *J. Pendidik. Teknol. Inf.*, vol. 1, no. 2, pp. 47–51, 2018, doi: 10.37792/jukanti.v1i2.11.
- [9] H. F. Putro, R. T. Vuldari, and W. L. Y. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *J. Teknol. Inf. dan Komun.*, vol. 8, no. 2, 2020, doi: 10.30646/tikomsin.v8i2.500.
- [10] N. Yahya and A. Jananto, "Komparasi Kinerja Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Kegiatan Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Stikubank Semarang)," *Pros. SENDI*, no. 2014, pp. 978–979, 2019, [Online]. Available: <https://www.unisbank.ac.id/ojs/index.php/sendu/article/view/7389>
- [11] H. Harliana and F. N. Putra, "Klasifikasi Tingkat Rumah Tangga Miskin Saat Pandemi Dengan Naïve Bayes Classifier," *J. Sains dan Inform.*, vol. 7, no. 2, pp. 165–173, 2021, doi: 10.34128/jsi.v7i2.339.