

PENERAPAN DATA MINING UNTUK PREDIKSI PENJUALAN CAT MENGGUNAKAN METODE *NAÏVE* *BAYES* (STUDI KASUS: MITRA 10 GADING SERPONG)

Rosa Shinta Putri Fairul¹, Galuh Saputri²

^{1,2} Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia
E-mail: ¹lettca31@gmail.com, ²dosen02693@unpam.ac.id

Abstrak- Perusahaan membutuhkan berbagai teknologi komunikasi yang dapat memprediksi minat beli pelanggan, tujuannya agar perusahaan dapat mempertimbangkan penjualan produk dan menentukan persediaan warna perusahaan. Selama ini Mitra 10 Gading Serpong baru mencatat review penjualan sehingga terkadang banyak produk cat yang kalah bersaing. Algoritma klasifikasi Naive Bayes merupakan salah satu algoritma yang termasuk dalam teknik klasifikasi. Metode ini diharapkan dapat memprediksi pembelian cat sehingga 10 mitra serpong gading lebih banyak menjual produk cat yang kurang kompetitif. Hasil akurasi prediksinya mencapai 82,14% . dari 180 data cat yang diuji, terdapat 148 data pembelian cat yang berhasil diklasifikasi dengan benar.

Kata Kunci: Data Mining, *Naïve bayes*, Penjualan Cat

Abstract- *The company needs several types of communication technology that can predict customer buying interest, the goal is that the company can properly consider product sales and determine the company's paint inventory. So far Mitra 10 gading serpong has only recorded sales recap, so that sometimes many paint products cannot compete. Naïve Bayes classifier algorithm is one of the algorithms included in classification technology. This method is expected to be able to predict paint purchases so that 10 ivory serpong partners sell paint products that are less competitive. The results of the prediction accuracy reached 82.14%. of the 180 paint data tested, there were 148 paint purchase data that were successfully classified correctly.*

Keywords: Data Mining, Naïve bayes, Paint Sales

1. PENDAHULUAN

Persaingan dalam dunia bisnis mendorong para pengusaha untuk terus berinovasi agar operasionalnya dapat mengikuti permintaan. Saat ini sudah banyak perusahaan yang memasok berbagai macam produk cat dengan berbagai jenis, merk dan harga, namun minat konsumen terhadap jenis produk cat tersebut masih belum diketahui.. (Harahap et al., 2021)

Mitra10 sendiri menjual berbagai bahan bangunan salah satunya cat. Permintaan produk cat sendiri cenderung berubah, sehingga pihak manajemen kesulitan dalam mengevaluasi jenis produk cat. Ketidakpastian permintaan konsumen akan produk cat seringkali mengakibatkan produk cat kosong, karena seringkali salah memperkirakan berapa banyak produk cat yang akan dipasok berdasarkan permintaan pasar, sehingga mempengaruhi jumlah pendapatan yang diterima.

Terkadang harga, jenis dan merk cat tidak menjadi prioritas bagi perusahaan, sehingga produk cat tersebut tidak dapat bersaing dengan perusahaan lain. Membeli produk cat yang kurang efisien membuat produk kurang menarik bagi pelanggan. Hal ini tentu saja menyebabkan konsumen beralih ke perusahaan sejenis yang menggunakan produk yang menurut mereka paling mereka butuhkan.

Dalam hal ini, untuk menentukan cat yang dibeli oleh pelanggan, metode Naïve Bayes digunakan untuk klasifikasi data mining untuk memprediksi penjualan cat aplikasi klasifikasi berupa algoritma klasifikasi Naive Bayesian ini diharapkan dapat memprediksi atau mendeteksi minat beli calon pelanggan lebih awal berdasarkan transaksi yang lalu, sehingga perusahaan dapat lebih memahami target minat pelanggannya. Metode klasifikasi adalah proses menemukan konsep fitur atau model atau kelas data yang menjelaskan tujuan mengevaluasi kelas objek dengan pengenalan yang tidak diketahui. (Yuli Mardi, 2019) Pola dari model itu sendiri bisa berupa aturan "jika-maka" atau rumus matematika. Dalam bentuk algoritma klasifikasi Naive Bayesian, aplikasi klasifikasi diharapkan dapat memprediksi atau mengidentifikasi minat beli pelanggan potensial berdasarkan peristiwa masa lalu, sehingga perusahaan dapat lebih memahami minat pelanggannya.

Berdasarkan dari permasalahan yang telah diuraikan maka penulis akan membuat data mining untuk memprediksi penjualan cat, maka penulis bermaksud untuk mengadakan penelitian dengan mengambil judul **“PENERAPAN DATA MINING UNTUK PREDIKSI PENJUALAN CAT MENGGUNAKAN METODE NAÏVE BAYES (STUDI KASUS : MITRA 10 GADING SERPONG)”**.

2. METODE

2.1. Metode Pengumpulan Data

Metode yang digunakan dalam proses pengumpulan data terdiri dari beberapa proses :

1. Studi Pustaka

Studi pustaka merupakan teknik pengumpulan data yang bertujuan untuk mendapatkan lebih banyak informasi tentang studi serupa dengan membaca, meneliti, dan menjelajahi literatur. Untuk memudahkan penulis memahami topik penelitian dan metode yang digunakan. Selain itu, studi literatur diperoleh dari jurnal ilmiah dan karya referensi terkait. (Handayani et al., 2019)

2. Wawancara

Wawancara merupakan suatu cara pengumpulan informasi dengan mengajukan serangkaian pertanyaan kepada informan untuk memperoleh keterangan atau informasi yang relevan dan akurat.

3. Analisa data

Pada tahap analisis data, yang dilakukan setelah pengumpulan data. Informasi ini dihitung dengan menggunakan metode Naive Bayesian yang digunakan untuk memprediksi penjualan cat, sehingga menghasilkan daftar penjualan cat.

2.2. Data Mining

Data mining adalah proses analitis yang bertujuan untuk mengeksplorasi data dalam jumlah besar untuk menemukan informasi tersembunyi yang berharga dan konsisten. Tujuan dari data mining adalah untuk menemukan tren atau pola yang diperlukan dalam database besar untuk membantu membuat keputusan di masa depan. (Harahap et al., 2021)

Karena prosesnya cukup panjang dan kompleks, proses aslinya biasanya menghasilkan sesuatu yang baru yang tidak diketahui sebelumnya. Hal baru ini menambah pengetahuan pengguna atau peneliti dan tentunya sangat bermanfaat karena dapat menjalankan fungsi tertentu. (Noviyanto, 2020)

2.3. Naïve Bayes

Naïve Bayes adalah salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk pembelajaran mesin dan penambangan data. Efisiensi kompetitif Naive Bayes dalam proses klasifikasi, meskipun menggunakan asumsi independensi atribut (tidak ada korelasi antar atribut). Asumsi independensi atribut ini jarang terjadi pada data nyata, tetapi meskipun asumsi independensi atribut ini dilanggar, efisiensi pengklasifikasi Naive Bayes cukup tinggi, sebagaimana dibuktikan oleh beberapa studi empiris. (Syarli & Muin,)

Pada fase learning, sebagian data yang telah diketahui kelas, datanya diumpankan untuk membentuk model perkiraan. Kemudian pada fase testing, model yang sudah terbentuk diuji dengan sebagian data. Dasar dari Naive Bayes yang dipakai dalam pemrograman adalah rumus Bayes:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Keterangan:

$P(A|B)$ = Probabilitas posterior dari A pada kondisi B (posterior probability).

$P(B|A)$ = Probabilitas posterior dari B pada kondisi A (likelihood).

$P(A)$ = Probabilitas prior dari A (class prior probability).

$P(B)$ = Probabilitas prior dari B (predictor prior probability).

Proses perhitungan probabilitas suatu kelas data dimulai dengan menentukan probabilitas berdasarkan himpunan data yang digunakan, dengan menggunakan metode yang sesuai dengan format data yang digunakan. Probabilitas yang dihasilkan dikalikan dengan probabilitas masing-masing kelas. Hasil dari proses ini digunakan sebagai acuan saat mengklasifikasikan data baru. Dalam prakteknya, $P(B)$ sering diabaikan karena nilai $P(B)$ selalu sama. (Amalia, 2020)

3. ANALISA DAN PEMBAHASAN

3.1. Merancang Dan Mendesain Model Sistem

1. Kumpulkan data untuk melatih dan menguji Naive Bayes. Banyaknya informasi yang dikumpulkan menentukan keberhasilan pemecahan masalah
2. Analisis dengan algoritma Naive Bayes. Data yang terkumpul kemudian akan dianalisis, untuk menyesuaikan pengolahan data dengan menggunakan metode Naive Bayes.
3. Pengolahan data. Data yang telah dianalisis akan diolah dengan menggunakan metode Naive Bayes. (Saputro & Sari, 2020)

3.2. Klasifikasi Data

Klasifikasi adalah salah satu tugas data mining untuk mengelompokkan data ke dalam kategori atau kategori yang telah ditentukan sebelumnya untuk memprediksi kelas/kategori yang benar untuk semua data dalam dataset.

a. Mengelompokkan Kualitas Cat

Tabel 1. Klasifikasi kualitas

Klasifikasi	Kualitas
>10 KG	Super

11 L- 25 L	Medium
< 10 L	Standar

Pengelompokkan Kualitas diambil berdasarkan hasil berat cat per produk yang dipasarkan.

b. Mengelompokkan Animo Cat

Tabel 2. Klasifikasi Animo

Animo	Klasifikasi
>Rp. 50.000.000	Tinggi
Rp. 10.100.000 – Rp. 50.000.000	Sedang
Rp. 500.000 – Rp. 10.000.0000	Rendah

Pengelompokkan Animo diambil berdasarkan hasil penjualan per produk cat yang dipasarkan dengan berbagai cara yang dilakukan pihak produsen cat.

c. Mengelompokkan Harga Cat

Tabel 3. Klasifikasi Harga

Harga	Klasifikasi
>Rp. 400.000	Mahal
Rp. 201.000 – Rp. 400.000	Normal
Rp. 401.000 – Rp. 1000.0000	Terjangkau

Pengelompokkan Harga cat diambil berdasarkan harga per produk cat yang dipasarkan pihak produsen cat.

d. Mengelompokkan Kompetisi Cat

Tabel 4. Klasifikasi Kompetisi

Kompetisi	Klasifikasi
>Rp. 100.000.000	Tinggi
Rp. 51.000.000 – Rp. 100.000.000	Sedang
Rp. 10.000.000 – Rp. 50.000.000	Rendah

Pengelompokkan kompetisi diambil berdasarkan hasil penjualan produk cat per supplier yang memasarkan produk cat tersebut

3.3. Data Set

Data yang digunakan dalam penelitian ini diperoleh dari perusahaan yang bergerak di bidang retail dan konstruksi. Data yang digunakan meliputi data historis prakiraan penjualan yang dijual pada perusahaan retail dan konstruksi. Langkah pertama yang harus dilakukan adalah membaca informasi pelatihan. Adapun data latih yang digunakan dapat dilihat pada Tabel berikut:

Tabel 5. Data Latih

NO	Kualitas	Animo	Harga	Kompetisi	Beli
1	Super	Sedang	Normal	Rendah	Beli
2	Standar	Rendah	Terjangkau	Tinggi	Beli
3	Standar	Rendah	Terjangkau	Tinggi	Tidak
4	Standar	Sedang	Terjangkau	Tinggi	Beli

5	Standar	Sedang	Normal	Tinggi	Beli
6	Standar	Sedang	Terjangkau	Tinggi	Beli
7	Standar	Sedang	Normal	Tinggi	Beli
8	Medium	Sedang	Tinggi	Tinggi	Beli
9	Standar	Sedang	Terjangkau	Sedang	Beli
10	Super	Tinggi	Normal	Sedang	Beli
11	Standar	Sedang	Normal	Rendah	Tidak
12	Standar	Sedang	Terjangkau	Tinggi	Beli
13	Standar	Sedang	Normal	Tinggi	Beli
14	Medium	Tinggi	Tinggi	Tinggi	Beli
15	Medium	Tinggi	Tinggi	Tinggi	Beli
16	Super	Sedang	Normal	Rendah	Tidak
17	Medium	Sedang	Normal	Tinggi	Beli
18	Standar	Tinggi	Terjangkau	Tinggi	Beli
19	Standar	Tinggi	Terjangkau	Sedang	Beli
20	Super	Rendah	Tinggi	Sedang	Tidak
21	Super	Sedang	Normal	Rendah	Beli
22	Standar	Rendah	Terjangkau	Sedang	Beli
23	Standar	Rendah	Terjangkau	Sedang	Beli
24	Standar	Sedang	Terjangkau	Sedang	Beli
25	Standar	Sedang	Normal	Sedang	Beli
26	Standar	Sedang	Terjangkau	Sedang	Beli
27	Standar	Sedang	Normal	Sedang	Beli
28	Medium	Rendah	Tinggi	Sedang	Beli
29	Standar	Rendah	Terjangkau	Sedang	Beli
30	Super	Sedang	Normal	Sedang	Tidak
.....
....
180	Super	Tinggi	Tinggi	Sedang	Beli

3.4. Kriteria Kualitas Cat

Tabel 6. Kriteria Kualitas

Probabilitas Kriteria Kualitas				
Kualitas	Jumlah Kejadian Pembelian Cat		Probabilitas	
	Beli	Tidak	Beli	Tidak
Super	27	9	0,194245	0,219512

Medium	31	5	0,223022	0,121951
Standar	81	27	0,582734	0,658537
Total	139	41	1	1

Berdasarkan data target penjualan pada dataset training di atas, diketahui volume dataset training sebanyak 180 dataset. Dari 180 data

1. Terdapat 27 data cat pembelian beli dengan kualitas super dan 9 data cat pembelian tidak dengan kualitas super
2. 31 data cat pembelian beli dengan kualitas medium dan 5 data cat pembelian tidak dengan kualitas medium
3. 81 data cat pembelian beli dengan kualitas standart dan 27 data cat pembelian tidak dengan kualitas standar.

3.5. Kriteria Animo

Tabel 7. Kriteria Animo

Probabilitas Kriteria Animo				
Animo	Jumlah Kejadian Pembelian Cat		Probabilitas	
	Beli	Tidak	Beli	Tidak
Tinggi	51	11	0,366906	0,268293
Sedang	76	24	0,546763	0,585366
Rendah	12	6	0,086331	0,146341
Total	139	41	1	1

Berdasarkan data target penjualan pada dataset training di atas, diketahui volume dataset training sebanyak 180 dataset. Dari 180 data

1. Terdapat 51 data cat pembelian beli dengan Animo Tinggi dan 11 data cat pembelian tidak dengan Animo Tinggi
2. 76 data cat pembelian beli dengan Animo Sedang dan 24 data cat pembelian tidak dengan Animo Sedang
3. 12 data cat pembelian beli dengan Animo Rendah dan 6 data cat pembelian tidak dengan Animo Rendah

3.6. Kriteria Harga Cat

Tabel 8. Kriteria Harga

Probabilitas Kriteria Harga				
Harga	Jumlah Kejadian Pembelian Cat		Probabilitas	
	Beli	Tidak	Beli	Tidak
Mahal	29	7	0,208633	0,170732
Normal	53	19	0,381295	0,463415
Terjangkau	57	15	0,410072	0,365854
Total	139	41	1	1

Berdasarkan data target penjualan pada dataset training di atas, diketahui volume dataset training sebanyak 180 dataset. Dari 180 data

1. Terdapat 29 data cat pembelian beli dengan harga mahal dan 7 data cat pembelian tidak dengan harga mahal
2. 53 data cat pembelian beli dengan harga normal dan 19 data cat pembelian tidak dengan harga normal
3. 57 data cat pembelian beli dengan harga terjangkau dan 15 data cat pembelian tidak dengan harga terjangkau.

3.7. Kriteria Kompetisi Cat

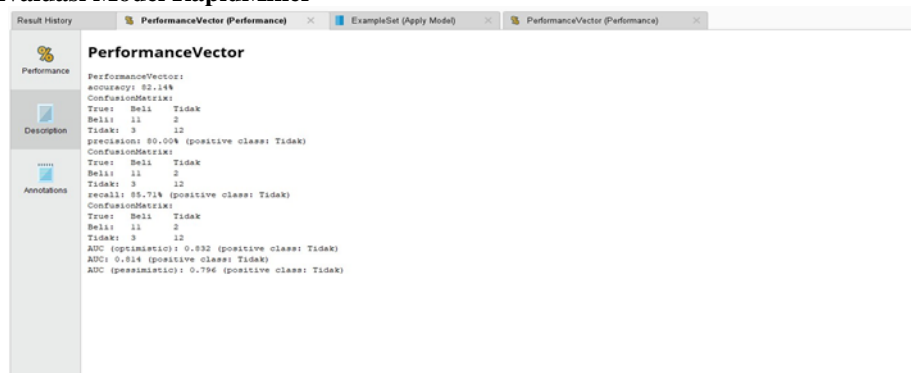
Tabel 9. Probabilitas Kriteria Kompetisi

Probabilitas Kriteria Kompetisi				
Kompetisi	Jumlah Kejadian Pembelian Cat		Probabilitas	
	Beli	Tidak	Beli	Tidak
Tinggi	88	20	0,633094	0,487805
Sedang	38	16	0,273381	0,390244
Rendah	13	5	0,093525	0,121951
Total	139	41	1	1

Berdasarkan data target penjualan pada dataset training di atas, diketahui volume dataset training sebanyak 180 dataset. Dari 180 data

1. Terdapat 88 data cat pembelian beli dengan kompetisi tinggi dan 20 data cat pembelian tidak dengan kompetisi tinggi
2. 38 data cat pembelian beli dengan kompetisi sedang dan 16 data cat pembelian tidak dengan kompetisi sedang
3. 13 data cat pembelian beli dengan kompetisi rendah dan 5 data cat pembelian tidak dengan kompetisi rendah.

3.8. Evaluasi Model RapidMiner



The screenshot shows the PerformanceVector window in RapidMiner. It displays the following performance metrics:

- accuracy: 82.14%
- ConfusionMatrix: True: Beli Tidak; Beli: 11 2; Tidak: 3 12
- precision: 80.00% (positive class: Tidak); ConfusionMatrix: True: Beli Tidak; Beli: 11 2; Tidak: 3 12
- recall: 85.71% (positive class: Tidak); ConfusionMatrix: True: Beli Tidak; Beli: 11 2; Tidak: 3 12
- AUC (optimistic): 0.832 (positive class: Tidak)
- AUC: 0.814 (positive class: Tidak)
- AUC (pessimistic): 0.796 (positive class: Tidak)

Gambar 1. Persentase Hasil

Berdasarkan gambar diatas diketahui bahwa persentase untuk Accuracy adalah sebesar 82,14%. persentase untuk recall adalah 85,71%. Presentase untuk Precision adalah 80.00%. dan presentase untuk AUC adalah 0,814. Di mana dari 180 data pembelian cat, ada sebanyak 148 data pembelian cat yang berhasil diklasifikasikan dengan benar dan sebanyak 32 data pembelian cat yang tidak berhasil diklasifikasikan dengan benar.

4. KESIMPULAN

4.1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka dapat disimpulkan bahwa beberapa kesimpulan dapat ditarik dari studi prediksi pencapaian tujuan, yaitu bahwa metode Naive Bayes menggunakan data pelatihan untuk menghasilkan probabilitas kelas yang berbeda untuk setiap kriteria. Oleh karena itu, nilai probabilitas dari kriteria tersebut dapat dioptimalkan sehingga dapat digunakan untuk memprediksi pencapaian tujuan berdasarkan proses klasifikasi itu sendiri menggunakan metode Naive Bayes. Berdasarkan data penjualan yang digunakan sebagai data latih, metode Naive Bayes berhasil mengklasifikasikan dengan benar 148 dari 180 data sampel. Dengan demikian, metode Naive Bayes berhasil memprediksi gate buying dengan akurasi 82,14%

4.2. Saran

Adapun saran yang ditujukan untuk memberikan masukan yang lebih baik yaitu :

- a. Disarankan untuk mencoba menggunakan metode sistem prediksi yang lain agar dapat menciptakan perbandingan antar metode yang sedang digunakan
- b. Agar dapat menggunakan metode naïve bayes ini dengan baik dibutuhkan lebih banyaknya data agar saat permodelan data yang digunakan seimbang.

REFERENSI

- Amalia, R. (2020). Penerapan Data Mining Untuk Memprediksi Hasil Kelulusan Siswa menggunakan Metode Naïve Bayes. *Jurnal Informatika Dan Sistem Informasi*, 6(1), 33–42.
- Handayani, I., Febriyanto, E., & Yudanto, T. A. (2019). Pemanfaatan Indeksasi Mendeley Sebagai Media Pengenalan. *Jurnal STT Yuppentek. Technomedia Journal*, 3(2), 235–245. <https://doi.org/10.33050/tmj.v3i2.1057>
- Harahap, F., Saragih, N. E., Siregar, E. T., & Sariangisah, H. (2021). Penerapan Data Mining Dengan Algoritma Naive Bayes Classifier Dalam Memprediksi Pembelian Cat. *Jurnal Ilmiah Informatika*, 9(01), 19–23. <https://doi.org/10.33884/jif.v9i01.3702>
- Noviyanto. (2020). Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian. *Jurnal Informatika Dan Komputer*, 22(2), 183–188.
- Saputro, I. W., & Sari, B. W. (2020). Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa. *Creative Information Technology Journal*, 6(1), 1. <https://doi.org/10.24076/citec.2019v6i1.178>
- Syarli, S., & Muin, A. (2016). Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi). *Jurnal Ilmiah Ilmu Komputer*, 2(1), 22–26.
- Yuli Mardi. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . *Jurnal Edik Informatika. Jurnal Edik Informatika*, 2(2), 213–219.