

PENDEKATAN DATA MINING UNTUK MEMILIH PRODUK TERLARIS MENGGUNAKAN ALGORITMA NAIVE BAYES

Hidup Perjuangan Rajagukguk ¹, Rahmat Fauzi²

¹Mahasiswa Program Studi Teknik Informatika, Universitas Putera Batam

²Dosen Program Studi Teknik Informatika, Universitas Putera
email: pb190210074@upbatam.ac.id

ABSTRACT

Technological advances today can be exploited to process data into more useful information. In data collection, information collection is especially useful to maximize profits and develop marketing strategies. One way to increase profits is by using data mining techniques to help business actors in making decisions about stocks, increased profits and more. The Matahari Department Store is the largest retail platform in Indonesia, one of the retail stores located in Batam is the Matahari Department store Nagoya Hill Batam. The transaction data on the store that is still processed does not use a method that causes the processing of product sales data to be less effective and less efficient. Seeing from the number of transactions, a system is needed to predict the sale of the best-selling product as long as it can determine the correct stock for the products sold and can increase the profit, sale and purchase of the product. This research was conducted with the aim of applying data mining methods using the Naive Bayes Classifier algorithm to select the best-selling products in the outlet of the Matahari Nagoya Hill Batam Department Store. By using the collected sales data, the system is expected to increase profits steadily and avoid shortages of product stocks. Through analysis using the Naive Bayes Classifier method, the study achieved an accuracy of 67% and obtained a bag sales result to be the best-selling sale during January 2023 through March 2023 with a sales percentage of 20%.

Keywords: Data Mining, Naive Bayes, Bestselling Sales

PENDAHULUAN

Pada era teknologi saat ini, perkembangan teknologi dapat dimanfaatkan untuk mengolah data menjadi informasi yang lebih bernilai. Proses pengumpulan data dan informasi memiliki peran penting dalam upaya untuk memaksimalkan keuntungan dan mengembangkan strategi pemasaran yang efektif. Salah satu strategi untuk meningkatkan keuntungan adalah dengan menyediakan dan menawarkan

produk yang sesuai dengan permintaan pasar, sehingga menghindari kekurangan stok yang dapat menyebabkan kerugian.

Dan dengan kemajuan waktu, pemanfaatan teknik data mining telah tersebar di berbagai sektor, mulai dari industri bisnis dan perdagangan hingga sektor pendidikan dan telekomunikasi. Terdapat sebuah algoritma *data mining* bernama *Naive Bayes Classifier* yang menjadi sangat bermanfaat bagi para pelaku bisnis dalam mengambil

keputusan mengenai stok barang, peningkatan keuntungan dan lainnya. Aplikasi dari algoritma tersebut dapat diterapkan di berbagai sektor bisnis.

Penggunaan *data mining* dapat digunakan untuk meningkatkan keuntungan, *Naive Bayes Classifier* adalah salah satu metode yang dapat digunakan. Penerapan metode ini pada basis data yang berukuran besar telah terbukti memiliki tingkat akurasi dan kecepatan yang sangat baik. Informasi tersebut bisa menjadi nilai tambah bagi toko, karena dapat membantu toko mengidentifikasi produk yang paling diminati oleh konsumen. Dengan demikian, toko dapat lebih tepat dan akurat dalam menemukan serta menargetkan pasar yang potensial. (Abdullah et al., 2022).

Matahari *Departement Store* adalah salah satu platform ritel terbesar di Indonesia, dengan 155 toko yang tersebar di 81 kota di seluruh negeri, serta memiliki kehadiran daring melalui Matahari.com. Sejak lebih dari enam puluh tahun, Matahari telah menawarkan pakaian, kecantikan, dan alas kaki berkualitas tinggi, modern, dan terjangkau bagi kalangan menengah di Indonesia (matahari, 2023).

Matahari *Departement Store* Nagoya Hill Batam merupakan salah satu outlet / *showroom* Matahari *Departement Store* yang berada di Batam. Matahari *Departement Store* menjual berbagai macam kebutuhan konsumen, sehingga membuat konsumen tertarik melakukan pembelian. Matahari *Departement Store* juga telah menerapkan penjualan secara *online* dengan mempromosikan jualannya. Kategori produk yang di jual berupa sepatu, topi, tas, dan lainnya. Banyaknya transaksi menyebabkan

KAJIAN TEORI

2.1 *Knowledge Discovery in Database* (KDD)

Knowledge Discovery In Database (KDD) adalah suatu pendekatan untuk mendapatkan pengetahuan dari *database* yang telah ada. Di dalam *database* terdapat tabel-tabel yang memiliki hubungan atau relasi satu sama lain. Hasil pengetahuan yang diperoleh dari proses tersebut dapat dijadikan sebagai basis pengetahuan (*knowledge base*) untuk mendukung pengambilan keputusan. (Yuli Mardi, 2019).

Istilah *Knowledge Discovery in Database* (KDD) dan *data mining* seringkali digunakan secara bergantian untuk menggambarkan proses mengeksplorasi informasi tersembunyi dalam *database* besar. Meskipun sebenarnya keduanya memiliki konsep yang berbeda, namun keduanya terkait erat, dan *data mining* merupakan salah satu tahap dalam proses keseluruhan KDD (Yuli Mardi, 2019).

2.2 *Data Mining*

Data mining adalah proses analisa terhadap data (umumnya data berukuran besar) untuk menemukan hubungan yang signifikan serta menyimpulkan informasi yang belum diketahui sebelumnya. Pendekatan yang digunakan adalah menggunakan teknik-teknik terkini untuk memahami dan memberikan manfaat bagi pemilik data tersebut. (Yuli Mardi, 2019).

Salah satu kesulitan untuk mendefinisikan *data mining* adalah kenyataan bahwa *data mining* meneruskan banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan sebelumnya, *data mining* memiliki akar yang berasal dari berbagai bidang

ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik, *database*, dan juga *information retrieval* (Yuli Mardi, 2019).

2.3 Klasifikasi

Klasifikasi adalah proses mengidentifikasi model atau fungsi yang dapat menjelaskan atau membedakan konsep atau data berdasarkan kelasnya, dengan tujuan untuk memprediksi kelas dari suatu obyek yang tidak memiliki label. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan data ke dalam berbagai kelas yang berdasarkan aturan atau fungsi tertentu (Siswandi & Fitriana, 2019).

2.4 Naïve Bayes

Naive Bayes adalah sebuah algoritma klasifikasi yang berbasis pada perhitungan probabilitas dengan menggunakan frekuensi dan kombinasi nilai dalam data tersebut. (Romadhon & Kurniawan, 2021). Metode ini dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes, yaitu metode yang memungkinkan prediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *Teorema Bayes* (Pratama et al., 2022).

Naive Bayes adalah salah satu algoritma klasifikasi yang sederhana namun memiliki tingkat akurasi yang tinggi. Meskipun begitu, *Naive Bayes* memiliki kelemahan yaitu kepekaannya terhadap seleksi fitur, yang dapat mempengaruhi akurasi hasilnya. Namun, dalam penelitian ini, metode klasifikasi berbasis fitur yang dikembangkan berhasil menghasilkan tingkat akurasi yang baik. Keunggulan dari *Naive Bayes* adalah sederhana, cepat dan tingkat akurasi yang tinggi (Putri et al., 2020).

Berikut ini adalah rumus yang menunjukkan persamaan dari *Teorema Bayes*: (Saputra & Herdiansyah, 2022):

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Keterangan:

X: data dengan class yang belum diketahui.

H: hipotesis data menggunakan suatu class spesifik.

P(H|X): probabilitas hipotesis H berdasar kondisi X (parteriori probabilitas).

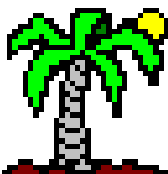
P(X|H): probabilitas X berdasarkan kondisi pada hipotesis H.

P(H): probabilitas hipotesis H (prior probabilitas).

P(X): probabilitas X.

2.5 Tanagra

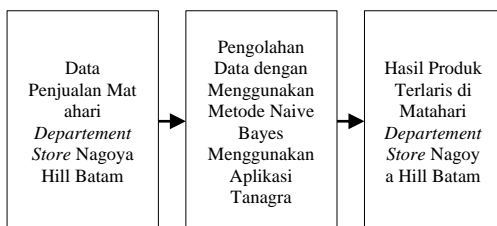
Tanagra adalah suatu perangkat lunak pembelajaran mesin yang gratis, dikembangkan oleh Ricco Rakotomalala di Universitas Lumière Lyon 2, Prancis. Tanagra mendukung berbagai tugas standar dalam penambangan data, seperti visualisasi, statistik deskriptif, pemilihan *instance*, pemilihan fitur, konstruksi fitur, regresi, analisis faktor, pengelompokan, klasifikasi, dan pembelajaran aturan asosiasi (Naik & Samant, 2016). Tanagra membuat kompromi yang baik antara pendekatan statistik (misalnya uji statistik parametrik dan nonparametrik), metode analisis multivariat (misalnya *factor analysis*, *correspondence analysis*, *cluster analysis*, *regression*) dan teknik pembelajaran mesin (misalnya *neural network (NN)*, *support vector machine (SVM)*, *decision trees*, *random forest*) (Naik & Samant, 2016). Berikut aplikasi yang digunakan dalam pengelolaan *data mining* seperti pada Gambar 1 berikut:



Gambar 1. Software Tanagra
(Sumber: Data Penelitian, 2023)

2.6 Kerangka Pemikiran

Kerangka pemikiran pada penelitian ini dapat dilihat pada *Gambar 2*:



Gambar 2. Kerangka Pemikiran
(Sumber: Data Penelitian, 2023)

Gambar 2 memperlihatkan bahwa data yang akan digunakan sebagai data input adalah data penjualan pada Matahari *Departement Store Nagoya Hill Batam*. Kemudian data akan diproses menggunakan metode *Naive Bayes* yang pengimplementasiannya menggunakan aplikasi Tanagra. Setelah melakukan proses tersebut, maka akan mendapatkan hasil prediksi penjualan produk terlaris pada Matahari *Departement Store Nagoya Hill Batam* dan juga hasil akurasi untuk produk terlaris menggunakan *Naive Bayes*.

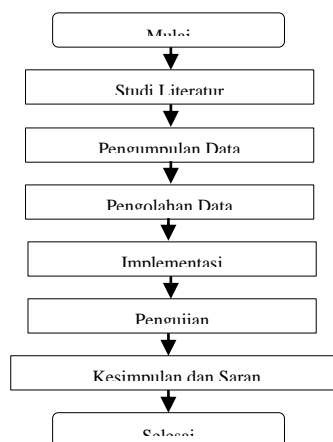
METODE PENELITIAN

3.1 Desain Penelitian

Desain penelitian dalam Pendekatan Data Mining Untuk Memilih Produk

Terlaris Menggunakan Algoritma *Naive Bayes* dapat dilihat pada *Gambar 3*:

Gambar 3 menjelaskan urutan penelitian yang dimana dimulai dari studi literatur yang mendukung dari berbagai sumber. Setelah itu, langkah berikutnya adalah mengumpulkan data, kemudian melakukan pengolahan data dengan metode *Naive Bayes*. Kemudian, dilakukan implementasi dan pengujian metode tersebut. Tahap terakhir dari penelitian ini adalah menarik kesimpulan dan memberikan saran berdasarkan hasil analisis data.



Gambar 3. Desain Penelitian
(Sumber: Data Penelitian, 2023)

Penelitian ini dilaksanakan di Matahari *Departement Store Nagoya Hill Batam*. Data yang akan digunakan dalam penelitian ini yaitu berupa data transaksi penjualan selama periode Bulan Januari–Maret 2023.

Teknik pengumpulan data yang digunakan dalam penelitian ini adalah menggunakan teknik observasi dan wawancara. Observasi dilakukan melakukan pengamanan terhadap data transaksi pembelian dan persediaan

produk, setelah itu peneliti melakukan wawancara secara langsung dengan karyawan Matahari *Departement Store* Nagoya Hill Batam dengan tujuan untuk mengumpulkan data penelitian.

3.2 Operasional Variabel

Operasional variabel dalam penelitian ini diantaranya:

1. Bulan, dalam penelitian ini data dipengaruhi oleh jumlah transaksi untuk setiap produk, yang akan menentukan kombinasi item set mana yang paling banyak terjual. Data untuk penelitian ini dikumpulkan dari periode Januari hingga Maret.
2. Harga, merupakan harga produk yang terjual dengan dikategorikan sebagai murah atau mahal.
3. Kategori, merupakan kategori produk yang menjadi favorit atau

sering dibeli oleh konsumen. Kategori yang digunakan adalah sepatu, sandal, tas, pakaian, kaus kaki, dan topi

4. Label, merupakan hasil penjualan produk yang dikategorikan laris atau tidak laris.

HASIL DAN PEMBAHASAN

4.1 Analisa Data

Data yang akan diolah pada penelitian ini berasal dari data penjualan Matahari *Departement Store* Nagoya Hill Batam selama bulan Januari – Maret 2023. Data terkumpul sebesar 540 data transaksi yang nantinya data akan dibagi menjadi dua yaitu data *training* dan data *testing* dengan perbandingan 90:10. Tabel 1 menunjukkan *dataset* yang akan digunakan dalam penelitian ini.

Tabel 1. Data Transaksi

No	Bulan	Nama Produk	Harga	Kategori	Qty	Label
1	Januari	Connexion Slipper Sandals Double Strap	Rp131.940	Sepatu	49	Tidak Laris
2	Januari	Yongki Komaladi Nayla Heels	Rp149.750	Sepatu	47	Tidak Laris
3	Januari	Peter Keiza Heels	Rp299.500	Sepatu	83	Laris
4	Januari	Yongki Komaladi Olivia Heels	Rp183.770	Sepatu	13	Tidak Laris
5	Januari	Yongki Komaladi Elvina Heels	Rp149.750	Sepatu	10	Tidak Laris
6	Februari	Character Flip Flop Hello Kitty	Rp63.960	Sandal	23	Tidak Laris
7	Februari	Disney Minnie Slipper Sandals	Rp109.950	Sandal	31	Tidak Laris
8	Februari	Disney Frozen Flip Flops	Rp53.162	Sandal	91	Laris
9	Februari	Connexion Crossbody Bag Women Perry1	Rp189.950	Tas	58	Tidak Laris
10	Februari	Nevada Crossbody Bag Women Calla1	Rp174.950	Tas	83	Laris
11	Maret	Details Ki Kemeja Rivet Feb Vapor	Rp168.935	Pakaian	61	Tidak Laris

12	Maret	Pipiniko Socks Dino Dasi	Rp14.320	Kaus Kaki	56	Tidak Laris
13	Maret	Pipiniko Socks Gajah Stripe	Rp14.320	Kaus Kaki	35	Tidak Laris
...
539	Maret	Creart Concept Casual Look Topi Pria	Rp69.000	Topi	17	Tidak Laris
540	Maret	Nevada Caps Bucket Hat Caps	Rp67.960	Topi	5	Tidak Laris

Dari tabel di atas, dilakukan perhitungan menggunakan algoritma *Naive Bayes* untuk menentukan jumlah

produk yang terjual berdasarkan klasifikasi seperti yang terlihat pada Tabel 2 di bawah ini..

Tabel 2. Probabilitas Kelas

Probabilitas Kelas	
Prediksi	Nilai
Laris	38,14%
Tidak Laris	61,86%
Jumlah	100,00%

(Sumber: Data Penelitian, 2023)

4.2 Analisa Proses Agotitma

Pada tahap ini, dilakukan pemodelan data menggunakan algoritma *Naive Bayes*. Data yang telah terkumpul akan diproses untuk mengidentifikasi peluang-peluang yang relevan. Metode ini berguna untuk memproyeksikan kemungkinan kejadian di masa depan berdasarkan pengalaman masa lalu sebagai pembandingnya.

Setelah itu *dataset* yang akan diproses dibagi menjadi dua bagian, yaitu data *training* dan data *testing*, data ini akan dianalisis menggunakan *software* Tanagra. Data *training* terdiri dari 485 data, sementara data *testing* terdiri dari 55 data. Data *training* dapat dilihat pada Tabel 3 dan data *testing* dapat dilihat pada **Error! Reference source not found..**

Tabel 3. Data *Training*

Bulan	Harga	Kategori	Label
Januari	Murah	Sepatu	Tidak Laris
Januari	Murah	Sepatu	Tidak Laris
Januari	Murah	Sepatu	Laris
Januari	Murah	Sepatu	Tidak Laris
Februari	Murah	Sandal	Tidak Laris
Februari	Murah	Sandal	Laris

Februari	Murah	Tas	Tidak Laris
Maret	Murah	Pakaian	Tidak Laris
Maret	Murah	Kaus Kaki	Tidak Laris
...
Maret	Murah	Kaus Kaki	Tidak Laris
Maret	Murah	Topi	Tidak Laris

(Sumber: Data Penelitian, 2023)

Dari data *training* pada Tabel 3 akan dilakukan pencarian probabilitas dengan metode *Naive Bayes* menggunakan rumus *teorema bayes*.

4.3 Probabilitas Kriteria Bulan

Berdasarkan data penjualan produk pada *dataset* yang digunakan, diketahui dari 485 data tersebut 32,62% produk

tidak laris dan 34,42% data laris pada Bulan Januari, pada Bulan Februari sebesar 36,31% produk tidak laris dan 28,84% produk laris, dan pada Maret sebesar 31,08% produk tidak laris dan 36,74% data laris. Tabel 4 menunjukkan probabilitas kriteria Bulan.

Tabel 4. Probabilitas Kriteria Bulan

	Januari	Februari	Maret	Sum
Tidak Laris	0,33	0,36	0,31	1,00
Laris	0,34	0,29	0,37	1,00

(Sumber: Data Penelitian, 2023)

4.4 Hasil Pengujian

Dengan adanya 55 data *testing*, maka hasil yang didapatkan dengan pengujian data menggunakan *software*

Tanagra menghasilkan *class* prediksi dengan nilai akurasi sebesar 67% Tabel 5 menunjukkan *class* prediksi yang didapat dari *software* Tanagra.

Tabel 5. Hasil Pengujian

	Tidak Laris	Laris
Tidak Laris	36	0
Laris	18	1

(Sumber: Data Penelitian, 2023)

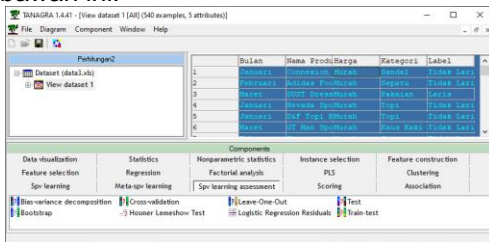
4.5 Implementasi Tanagra

Pada langkah ini, peneliti mengumpulkan data *training* dari Bulan Januari, Februari, dan Maret, dengan jumlah keseluruhan 485 data, serta data *testing* sebanyak 55 data. Data *training*

dan data *testing* yang digunakan dalam penelitian ini diatur dalam format *Excel* agar dapat digunakan untuk memprediksi penjualan produk yang akan menjadi laris atau tidak laris, dan

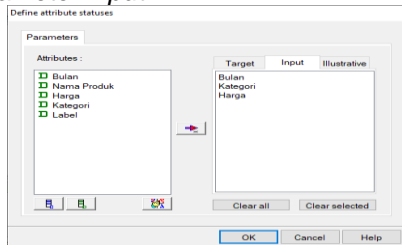
data ini dapat diakses melalui *software* Tanagra.

Setelah itu, langkah selanjutnya adalah melakukan *import dataset* yang digunakan ke *software* Tanagra sesuai dengan Gambar 4 yang ditampilkan di bawah ini:



Gambar 4. Import Dataset
(Sumber: Data Penelitian, 2023)

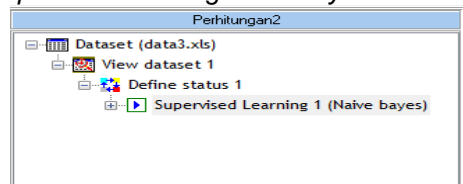
Setelah data di-*import*, tahap selanjutnya adalah memasukkan atribut yang akan dipakai dalam penelitian ini seperti Bulan, Kategori, Harga, dan Label. Atribut yang dimasukkan ke dalam parameter *input* adalah Bulan, Kategori, dan Harga, sedangkan atribut yang dimasukkan kedalam parameter *target* adalah Label. Gambar 5 menunjukkan parameter *input*.



Gambar 5. Parameter Input
(Sumber: Data Penelitian, 2023)

Setelah atribut dimasukkan ke dalam parameter *input* dan *target*, tahap selanjutnya adalah memasukkan *supervised learning* pada *components* dan memilih *Naive Bayes*. Gambar 6

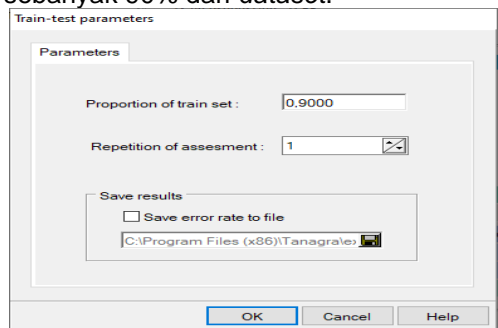
menunjukkan hasil pemasukkan *supervised learning naive bayes*.



Gambar 6. Supervised Learning Naive Bayes

(Sumber: Data Penelitian, 2023)

Setelah memasukkan metode *Naive Bayes*, diberikan komponen *train-test* untuk membagi *dataset* menjadi dua bagian, yaitu data *training* dan data *testing*. Parameter *train-test* yang akan digunakan adalah 0.9 untuk *proportion of train set*. Setelah itu, dilakukan *execute* untuk mengetahui hasil dari proses pengolahan data pada penelitian ini. Gambar 7 menunjukkan parameter yang digunakan dalam *Train-test* yang digunakan. Dan **Error! Reference source not found.** menunjukkan hasil pengolahan data dengan *Naive Bayes* menggunakan data *train* sebanyak 90% dari *dataset*.



Gambar 7. Parameter Train-test
(Sumber: Data Penelitian, 2023)

Pada Gambar 8 terdapat *confusion matrix* yang dapat digunakan untuk mencari akurasi dari hasil perhitungan. Secara umum akurasi dapat dirumuskan sebagai berikut:

Budidarma, 6(4), 2135.
<https://doi.org/10.30865/mib.v6i4.4418>

Putri, D. A., Kristiyanti, D. A., Indrayuni, E., Nurhadi, A., & Hadinata, D. R. (2020). Comparison of Naive Bayes Algorithm and Support Vector Machine using PSO Feature Selection for Sentiment Analysis on E-Wallet Review. *Journal of Physics: Conference Series*, 1641(1), 0–6.
<https://doi.org/10.1088/1742-6596/1641/1/012085>

Romadhon, M. R., & Kurniawan, F. (2021). A Comparison of Naive Bayes Methods, Logistic Regression and KNN for Predicting Healing of Covid-19 Patients in Indonesia. *3rd 2021 East Indonesia Conference on Computer and Information Technology, EIconCIT 2021*, 41–44.
<https://doi.org/10.1109/EIconCIT50028.2021.9431845>

Saputra, M. J., & Herdiansyah, M. I. (2022). PENERAPAN NAIVE BAYES DALAM MEMREDIKSI PENJUALAN TUAN KENTANG PALEMBANG. *Jurnal Mantik*, 6(36), 2502–2507.

Siswandi, A., & Fitriana, M. (2019). *IMPLEMENTASI DATA MINING DENGAN METODE KLASIFIKASI NAIVE BAYES UNTUK MEMREDIKSI STOK BAHAN JADI.pdf* (pp. 1–10). Jurnal Teknologi Pelita Bangsa.

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.

	<p>Biodata Penulis pertama, Hidup Perjuangan Rajagukguk, merupakan mahasiswa Prodi Teknik Informatika Universitas Putera Batam.</p>
	<p>Biodata Penulis kedua, Rahmat Fauzi, S. Kom., M. Kom. merupakan Dosen Prodi Teknik Informatika Universitas Putera Batam. Penulis banyak berkecimpung di bidang Sistem Informasi.</p>