

Analisa Data Shopping Trends Menggunakan Algoritma Klasifikasi Dengan Metode Naive Bayes

by Said Imam Puro

Submission date: 04-Jul-2024 11:14AM (UTC+0700)

Submission ID: 2412351141

File name: REPEATER_-_VOL._2_NO._3_JULI_2024_Hal.119-134..pdf (2.82M)

Word count: 3336

Character count: 20943



Analisa Data Shopping Trends Menggunakan Algoritma Klasifikasi Dengan Metode Naive Bayes

Said Imam Puro ¹, Jodi Hariyan ², Ridho Rafliansyah ³, Muhammad Rival Aziz ⁴, Pebro Vaulina Rajagukguk ⁵

^{1,2,3,4,5} Universitas Bina Sarana Informatika

Abstract In the era of rapid digitalization, understanding consumer behavior through data is becoming increasingly important for retail businesses. Shopping trends, such as those contained in this study, provide in-depth insights into various aspects of consumer behavior, from demographics to purchasing preferences and patterns of discount usage. This data is invaluable in formulating effective marketing strategies, improving customer experience, and optimizing business operations. The data used in this study included a variety of relevant variables, such as age, gender, location, product categories purchased, number of purchases, payment methods, and frequency of purchases. This information allows for a comprehensive analysis of how these factors affect consumer spending decisions. For example, analytics can reveal seasonal trends in purchases, product color and size preferences, and the impact of discounts and promo codes on sales volume. In addition, this dataset also reflects the changes in consumer behavior that have occurred over the past few years. Quantitative methodology is a research approach used to collect and analyze numerical data to understand patterns, relationships, and events in a given population. Data is collected from various sources such as online sales transactions, consumer surveys, Naive Bayesian algorithms are applied to the dataset that has been processed. The data was divided into two sets: training (80%) and testing (20%).

Keywords: Rapidminer, Dataset, Category, Classification, Naivebayes

Abstrak Dalam era digitalisasi yang pesat, memahami perilaku konsumen melalui data menjadi semakin penting bagi bisnis ritel. Tren belanja, seperti yang terdapat dalam penelitian ini, memberikan wawasan mendalam tentang berbagai aspek perilaku konsumen, mulai dari demografi hingga preferensi pembelian dan pola penggunaan diskon. Data ini sangat berharga dalam merumuskan strategi pemasaran yang efektif, meningkatkan pengalaman pelanggan, dan mengoptimalkan operasi bisnis. yang digunakan dalam penelitian ini mencakup berbagai variabel yang relevan, seperti usia, jenis kelamin, lokasi, kategori produk yang dibeli, jumlah pembelian, metode pembayaran, dan frekuensi pembelian. Informasi ini memungkinkan analisis yang komprehensif tentang bagaimana faktor-faktor tersebut mempengaruhi keputusan belanja konsumen. Misalnya, analisis dapat mengungkapkan tren musiman dalam pembelian, preferensi warna dan ukuran produk, serta dampak dari diskon dan kode promo terhadap volume penjualan. Selain itu, dataset ini juga mencerminkan perubahan perilaku konsumen yang terjadi selama beberapa tahun terakhir. Metodologi kuantitatif adalah pendekatan penelitian yang digunakan untuk mengumpulkan dan menganalisis data numerik guna memahami pola, hubungan, dan peristiwa dalam populasi tertentu. Data Dikumpulkan Dari Berbagai Sumber Seperti Transaksi Penjualan Online, Survei Konsumen, Algoritma Naive Bayes Diterapkan Pada Dataset Yang Telah Diproses. Data Dibagi Menjadi Dua Set: Pelatihan (80%) Dan Pengujian (20%).

Kata Kunci: Rapidminer, Dataset, Category, Klasifikasi, Naivebayes

LATAR BELAKANG

Dalam era digitalisasi yang pesat, memahami perilaku konsumen melalui data menjadi semakin penting bagi bisnis ritel. Tren belanja, seperti yang terdapat dalam penelitian ini, memberikan wawasan mendalam tentang berbagai aspek perilaku konsumen, mulai dari demografi hingga preferensi pembelian dan pola penggunaan diskon. Data ini sangat berharga dalam merumuskan strategi pemasaran yang efektif, meningkatkan pengalaman pelanggan, dan mengoptimalkan operasi bisnis. yang digunakan dalam penelitian ini mencakup berbagai variabel yang relevan, seperti usia, jenis kelamin, lokasi, kategori produk yang dibeli, jumlah pembelian, metode pembayaran, dan frekuensi pembelian.

Received: Juni 30,2024 , Accepted: Juli 04,2024 , Published: Juli 31,2024

* Said Imam Puro

Dalam industri retail, memahami pola pembelian konsumen adalah kunci Informasi ini memungkinkan analisis yang komprehensif tentang bagaimana faktor-faktor tersebut mempengaruhi keputusan belanja konsumen. Misalnya, analisis dapat mengungkapkan tren musiman dalam pembelian, preferensi warna dan ukuran produk, serta dampak dari diskon dan kode promo terhadap volume penjualan. Selain itu, dataset ini juga mencerminkan perubahan perilaku konsumen yang terjadi selama beberapa tahun terakhir.

Peningkatan akses internet dan penggunaan perangkat mobile telah mendorong lebih banyak konsumen untuk berbelanja online. Hal ini juga diperkuat oleh pandemi COVID-19 yang mengubah kebiasaan belanja banyak orang, mempercepat adopsi e-commerce, dan meningkatkan permintaan akan pengiriman yang cepat dan aman. Kesadaran yang meningkat terhadap keberlanjutan dan praktik bisnis yang etis juga tercermin dalam tren belanja konsumen. Banyak konsumen sekarang lebih selektif dalam memilih produk yang ramah lingkungan dan berasal dari sumber yang bertanggung jawab.

Analisis terhadap data ini dapat membantu perusahaan memahami preferensi ini dan menyesuaikan penawaran mereka untuk memenuhi harapan konsumen modern.

Penelitian ini bertujuan untuk menganalisis data tren belanja guna mengidentifikasi pola dan tren utama dalam perilaku konsumen. Dengan menggunakan teknik analisis data yang canggih, seperti analisis statistik dan machine learning, penelitian ini akan memberikan wawasan yang dapat digunakan oleh perusahaan ritel untuk meningkatkan strategi pemasaran dan operasi mereka.

Selain itu, hasil analisis ini diharapkan dapat memberikan kontribusi yang signifikan dalam memahami dinamika pasar yang terus berubah dan membantu perusahaan dalam menyesuaikan diri dengan kebutuhan dan preferensi konsumen yang semakin kompleks dan beragam.

Analisis Masalah dan Solusi

Analisis masalah

1. Strategi Pemasaran Di Era Digital Yang Paling Efektif Untuk Meningkatkan Penjualan Di Toko Online, Dengan Banyaknya Saluran Pemasaran Digital Yang Tersedia (Media Sosial, SEO, Iklan Berbayar, Email Marketing), Penting Untuk Mengetahui Mana Yang Paling Efektif Untuk Menarik Dan Mempertahankan Pelanggan Manajemen Inventaris Yang Efisien Adalah Kunci Untuk Memenuhi Permintaan Pelanggan Tanpa Menimbulkan Biaya Penyimpanan Yang Tinggi.

Analisis Solusi

1. Teknologi Seperti AI Dan Analitik Data Bisa Membantu Mengoptimalkan Manajemen Inventaris. Dengan Meningkatnya Ancaman Cyber, Menjaga Keamanan Data Pelanggan Menjadi Prioritas Utama. Menemukan Solusi Untuk Melindungi Data Pelanggan Sangat Penting Untuk Menjaga Kepercayaan Dan Menghindari Kerugian.

LANDASAN TEORI

Data Mining

Pengertian Data Mining

Data Mining Adalah Proses Ekstraksi Pengetahuan Yang Bermanfaat Atau Pola Yang Menarik Dari Sebuah Dataset Besar. Tujuan Utama Dari Data Mining Adalah Untuk Menemukan Pola Yang Tidak Terlihat Sebelumnya, Yang Dapat Memberikan Wawasan Yang Berharga Dan Mendukung Pengambilan Keputusan Yang Lebih Baik.

Data mining suatu proses ekstraksi atau penggalian data dan informasi yang besar, yang belum diketahui sebelumnya, namun dapat dipahamikan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting. Data mining menggambarkan sebuah pengumpulan teknik-teknik dengan tujuan untuk menemukan pola-pola yang tidak diketahui pada data yang telah dikumpulkan. Data mining memungkinkan pemakai menemukan pengetahuan dalam data database yang tidak mungkin diketahui keberadaannya oleh pemakai. Data mining adalah sebuah proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besar.

Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basis data.

Tahapan Data Mining

data mining dapat dibagi menjadi beberapa tahap proses. Tahap-tahap tersebut bersifat interaktif :

- a. **Pembersihan data (Data Cleaning)**
Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan.
- b. **Integrasi data (Data Integration)**
Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru.
- c. **Seleksi data (Data Selection)**
Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database.

d. **Transformasi data (Data Transformation)**

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam Data Mining.

e. **Proses Mining**

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan Data Mining.

f. **Evaluasi pola (Pattern Evaluation)**

Untuk mengidentifikasi pola-pola menarik ke dalam knowledge based yang ditemukan.

g. **Presentasi pengetahuan (Knowledge Presentation)**

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

9 Perancangan dan Implementasi

Pada tahapan ini setelah proses analisis selesai, maka dilakukan tahapan perancangan proses perhitungan menggunakan formula Algoritma Naïve Bayes. Perancangan dilakukan melalui tahapan sebagai berikut:

- a. Pengolahan data secara manual menggunakan Software Microsoft Excel dengan Algoritma Naïve Bayes.
- b. Perhitungan secara manual dengan formula Algoritma Naïve Bayes dengan data yang sama.

Data Warehouse

Pengertian Data Warehouse

Data Warehouse adalah suatu paradigma baru dilingkungan pengambilan keputusan strategik. Data warehouse bukan suatu produk tetapi suatu lingkungan dimana user dapat menemukan informasi strategik. Data Warehouse adalah kumpulan data-data logik yang terpisah dengan database operasional dan merupakan suatu ringkasan (Indarta et al., 2021).

3 Terdapat beberapa definisi seputar data warehouse. Dari sisi praktisi, Barry Devlin, IBM Consultant, menerjemahkan data warehouse sebagai : "Suatu data warehouse sederhananya adalah suatu penyimpanan data tunggal, lengkap dan konsisten, yang diperoleh dari berbagai sumber dan dibuat tersedia bagi end user dalam suatu cara yang bisa mereka pahami dan bisa mereka gunakan dalam suatu konteks bisnis.

Pendapat lain tentang pengertian data *warehouse* yang dikemukakan oleh W. H. Inmon, yang dikenal juga sebagai Bapak data warehousing, adalah : "Suatu data warehouse adalah suatu koleksi data yang bisa digunakan untuk menunjang pengambilan keputusan

manajemen, yang berorientasi subjek (topik), terpadu, time variant, dan tidak mudah berubah". Secara garis besar, data warehouse adalah sebuah database penunjang keputusan yang mengandung data yang biasanya mewakili sejarah bisnis dari suatu perusahaan.

Karakteristik Data Warehouse

a. Subject Oriented

Yaitu mengorganisir data menurut subjek dari suatu perusahaan, misalnya konsumen, produk, dan penjualan. Difokuskan pada pemodelan dan analisis data untuk decision maker, bukan pada operasi harian atau pemrosesan transaksi. Menyediakan wawasan yang sederhana dan ringkas mengenai subjek dengan memisahkan data yg tidak relevan dalam proses pengambilan keputusan.

b. Integrated

Data warehouse harus mengintegrasikan data dari sumber data yang beragam, seperti relational database, flat files, on_line transaction records.

c. Time-variant

Data warehouse tetap menyimpan data-data historis. Setiap struktur key dalam data warehouse mengandung elemen waktu baik eksplisit maupun implisit.

d. Non-volatile

Penyimpanan data transformasi dalam data warehouse selalu terpisah secara fisik dari lingkungan operational. Oleh karena itu, updatedata operasional tidak terjadi pada lingkungan data warehouse, dan data warehouse tidak memerlukan pemrosesan transaksi, recovery, dan concurrency control. Hanya memerlukan dua operasi dalam megakses data, yaitu initial loading of data dan access of data.

Metode Classification

Pengertian Classification

Klasifikasi merupakan kata serapan dari bahasa Belanda, classificatie, yang sendirinya berasal dari bahasa Prancis classification. Istilah ini menunjuk kepada sebuah metode untuk menyusun data secara sistematis atau menurut beberapa aturan atau kaidah yang telah ditetapkan. Karena dalam klasifikasi terdapat target variabel kategori. (Matondang et al., 2021)

Klasifikasi adalah sepengelompokkan yang sistematis pada sejumlah objek, gagasan, buku atau benda-benda lain ke dalam kelas atau golongan tertentu berdasarkan ciri-ciri yang sama.

Algoritma Naïve Bayes

Pengertian Naïve Bayes

10

Naive Bayes adalah metode yang membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Data training digunakan sebagai data rujukan dalam perhitungan setiap algoritma, sedangkan data testing digunakan untuk menilai prediksi maupun penentuan yang dilakukan oleh setiap algoritma sudah tepat atau tidak. (Kumiawan, 2018)

Kegunaan Naive Bayes

1. Mengklasifikasikan dokumen teks seperti teks berita ataupun teks akademis
2. Sebagai metode machine learning yang menggunakan probabilitas
3. Untuk membuat diagnosis medis secara otomatis
4. Mendeteksi atau menyaring spam

Kelebihan Naive Bayes

1. Bisa dipakai untuk data kuantitatif maupun kualitatif
2. Tidak memerlukan jumlah data yang banyak
3. Tidak perlu melakukan data training yang banyak
4. Jika ada nilai yang hilang, maka bisa diabaikan dalam perhitungan
5. Perhitungannya cepat dan efisien
6. Mudah dipahami
7. Mudah dibuat
8. Pengklasifikasian dokumen bisa dipersonalisasi, disesuaikan dengan kebutuhan setiap orang
9. Jika digunakan dalam bahasa pemrograman, code-nya sederhana
10. Bisa digunakan untuk klasifikasi masalah biner ataupun multiclass

Kekurangan Naive Bayes

1. Apabila probabilitas kondisionalnya bernilai nol, maka probabilitas prediksi juga akan bernilai nol
2. Asumsi bahwa masing-masing variabel independen membuat berkurangnya akurasi, karena biasanya ada korelasi antara variabel yang satu dengan variabel yang lain
3. Keakuratannya tidak bisa diukur menggunakan satu probabilitas saja. Butuh bukti-bukti lain untuk membuktikannya.
4. Untuk membuat keputusan, diperlukan pengetahuan awal atau pengetahuan mengenai masa sebelumnya. Keberhasilannya sangat bergantung pada pengetahuan awal tersebut Banyak celah yang bisa mengurangi efektivitasnya dirancang untuk mendeteksi kata-kata saja, tidak bisa berupa gambar.

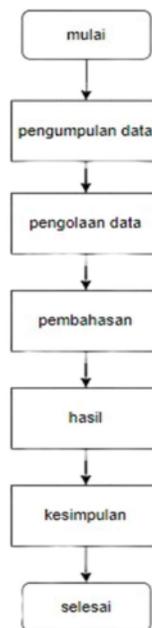
Rapid Miner

Rapidminer Merupakan Solusi Ideal Bagi Pengusaha Dan Calon Pengusaha Yang Tidak Memiliki Latar Belakang Pemrograman Yang Mendalam. Pengusaha Yang Menggunakan Rapidminer Dapat Dengan Mudah Menggali Potensi Data Dan Menerapkan Teknik Data Mining Dalam Bisnis Tanpa Harus Menguasai Bahasa Pemrograman.

Rapid Miner memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi.(Akbar & Rahmanto, 2020) RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik.(Ardiansyah & Walim, 2018)

METODOLOGI PENELITIAN

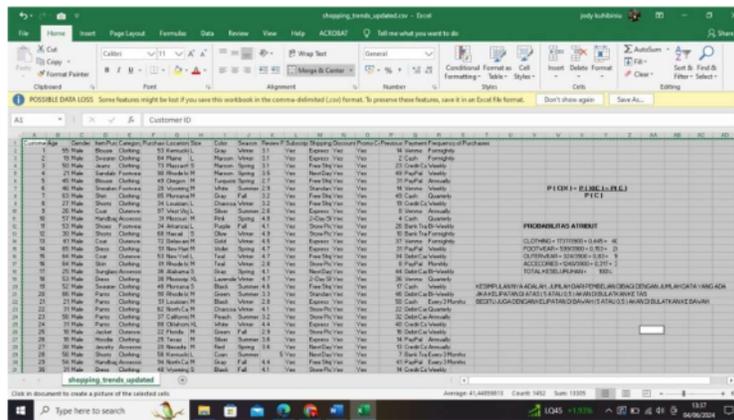
Metodologi kuantitatif adalah pendekatan penelitian yang digunakan untuk mengumpulkan dan menganalisis data numerik guna memahami pola, hubungan, dan peristiwa dalam populasi tertentu. Pendekatan ini bersifat deduktif, dimulai dengan teori atau hipotesis yang kemudian diuji melalui pengumpulan data yang terstruktur.



Gambar 1

Demonstrasi Exel

Analisa Data Shopping Trends Menggunakan Algoritma Klasifikasi Dengan Metode Naive Bayes

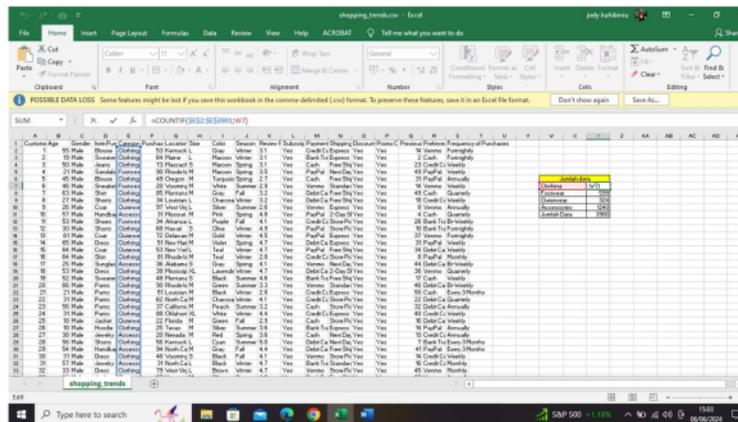


Gambar .2

Jumlah Dataset Yang Digunakan Adalah Berjumlah 3900 Dataset Yang Terbagi Ke Dalam Beberapa Katagori Penjualan Terbanyak.

Jumlah data	
Clothing	1737
Footwear	599
Outerwear	324
Accessories	1240
Jumlah Data	3900

Rumus Mencari Data Clothing Di Exel

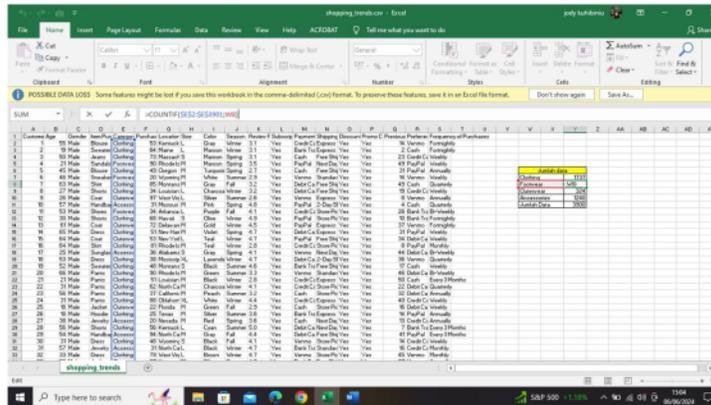


Gambar 3

Masukan rumus =COUNTIF(lalu blok data category lalu klik FN+F4 beri tanda ; lalu klik data clothing lalu beri tutup kurung) dan enter maka akan muncul hasil nya.

=COUNTIF(\$E\$2:\$E\$3901;W7)

Rumus Mencari Data Footwear Di Exel



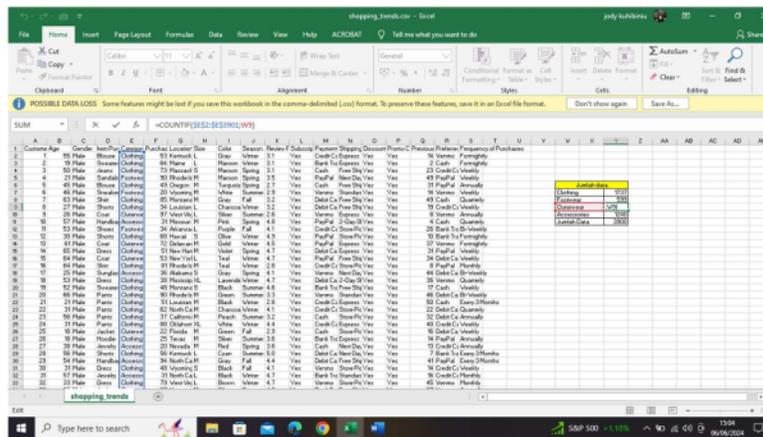
Formula: $=\text{COUNTIF}(\$E\$2:\$E\$3901;W8)$

Category	Jumlah	Produk	Jumlah Data
Footwear	117	300	300

Gambar .4

Masukan rumus $=\text{COUNTIF}$ (lalu blok data category lalu klik FN+F4 beri tanda ; lalu klik data Footwear lalu beri tutup kurung) dan enter maka akan muncul hasil nya.
 $=\text{COUNTIF}(\$E\$2:\$E\$3901;W8)$

Rumus Mencari Data Outerwear Di Exel



Formula: $=\text{COUNTIF}(\$E\$2:\$E\$3901;W9)$

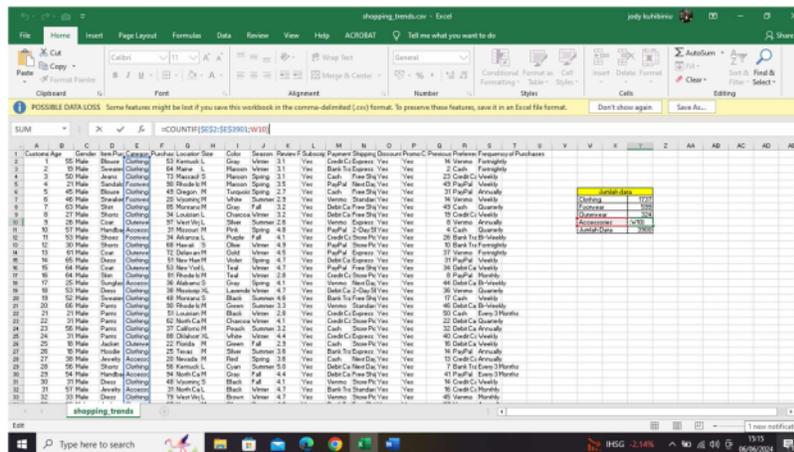
Category	Jumlah	Produk	Jumlah Data
Outerwear	117	300	300

Gambar 5

Masukan rumus $=\text{COUNTIF}$ (lalu blok data category lalu klik FN+F4 beri tanda ; lalu klik data Outerwear lalu beri tutup kurung) dan enter maka akan muncul hasil nya.
 $=\text{COUNTIF}(\$E\$2:\$E\$3901;W9)$

Rumus Mencari Data Accessories Di Exel

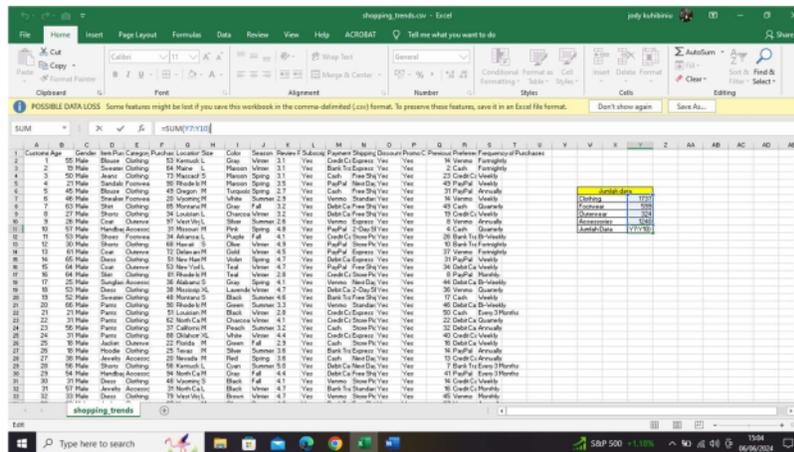
Analisa Data Shopping Trends Menggunakan Algoritma Klasifikasi Dengan Metode Naive Bayes



Gambar 6

Masukan rumus =COUNTIF(lalu blok data category lalu klik FN+F4 beri tanda ; lalu klik data Outerwear lalu beri tutup kurung) dan enter maka akan muncul hasil nya.
 =COUNTIF(\$E\$2:\$E\$3901;W10)

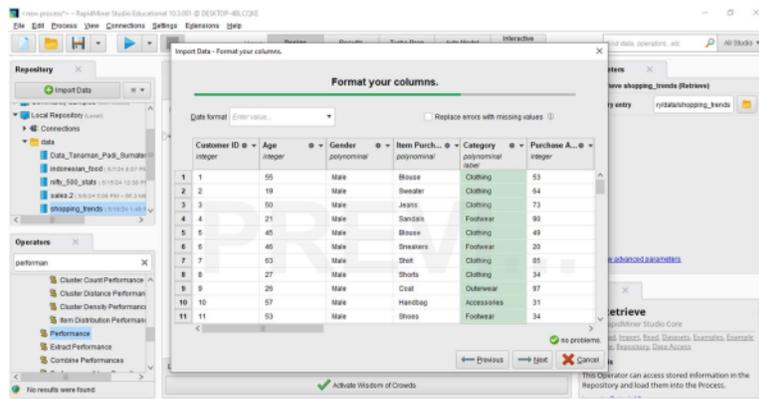
Perhitungan Hasil



Gambar 7

Masukan rumus =SUM (lalu blok data beri tutup kurung) dan enter
 =SUM(Y7:Y10)

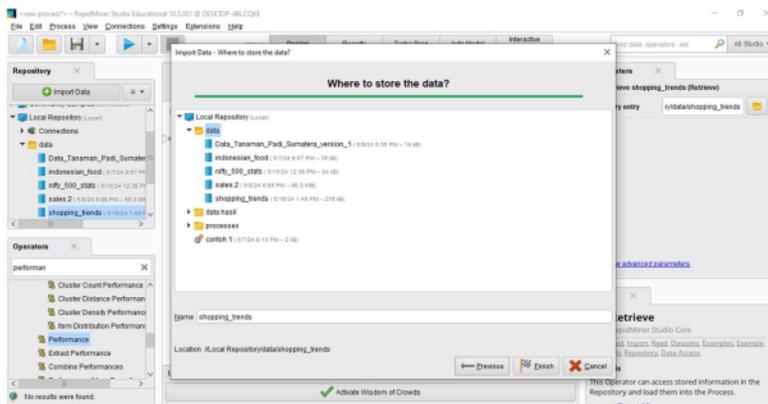
Demonstrasi Rapid Miner



Gambar 8 masukan Dataset di RapidMiner

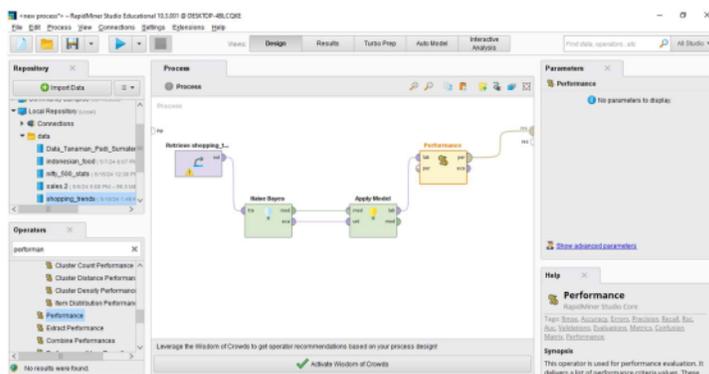
Masukan Dataset Kedalam Rapidminer Lalu Pilih Data Yang Ingin Di Jadikan Data

Label



Gambar 9 Simpan Data di Repository

Simpan Data Yang Telah Kita Beri Label Tadi Ke Repository



Gambar 10 Running

Pilih Operator Untuk Melakukan Running

1. Dataset
2. Naïve bayes
3. Apply model
4. Performance

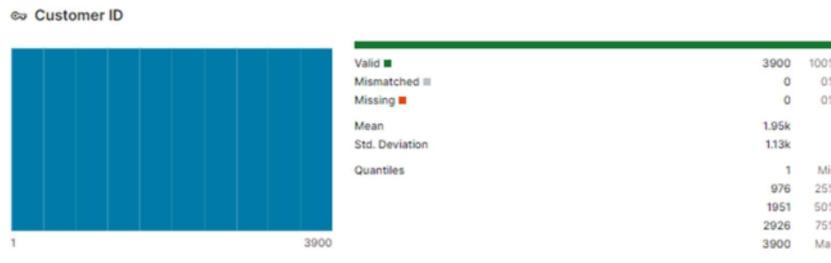
Gambar 11 Hasil Accuracy

accuracy: 100.00%

	true Clothing	true Footwear	true Outerwear	true Accessories	class precision
pred. Clothing	1737	0	0	0	100.00%
pred. Footwear	0	599	0	0	100.00%
pred. Outerwear	0	0	324	0	100.00%
pred. Accessories	0	0	0	1240	100.00%
class recall	100.00%	100.00%	100.00%	100.00%	

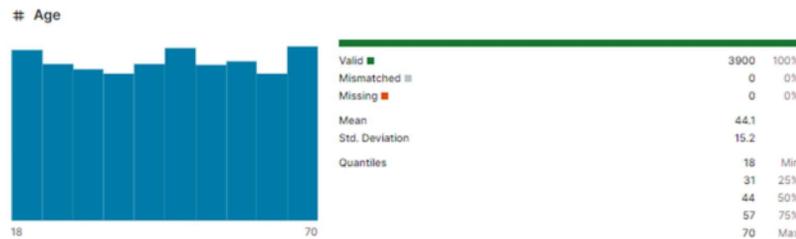
Accuracy di atas adalah jumlah dari penjualan terbanyak,

1. Clothing (pakaian) = 1737
2. Footwear (alas kaki) = 599
3. Outerwear (pakaian luar) = 324
4. Accessories (aksesoris) = 1240
5. Total terjual = 3900



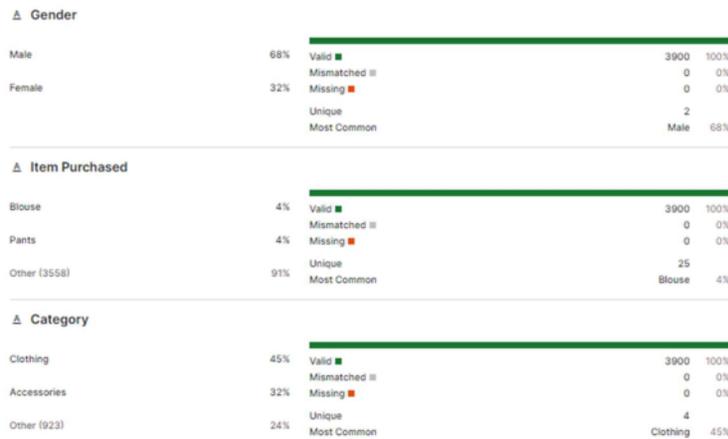
Gambar 12 Customers ID

Dengan Menggunakan Customer ID, Perusahaan Dapat Dengan Mudah Mengelola Data Pelanggan, Seperti Memantau Transaksi, Menanggapi Permintaan Layanan Pelanggan, Dan Menganalisis Pola Pembelian.



Gambar 13 Hasil Age

Customer Mulai Dari Usia 18 Tahun Hingga 70 Tahun



Gambar 14 katalog

- Genre

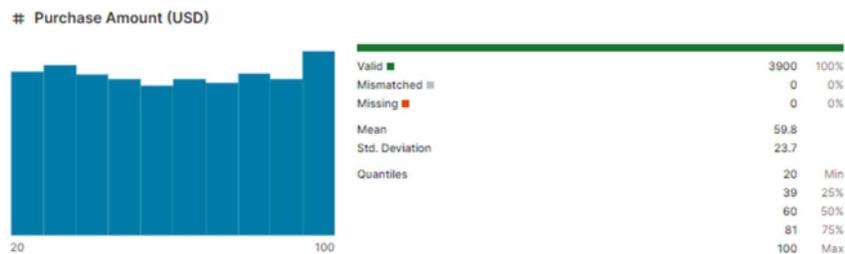
Pembeli Yang Paling Banyak Adalah Dari Kalangan Lelaki Sebanyak 68% Dan Diikuti Pembelian Wanita Sebanyak 32%

- Item Purchase

1. Blouse (baju wanita) 4%,
2. Pants (celana wanita) 4%,
3. Other 91%

- Category

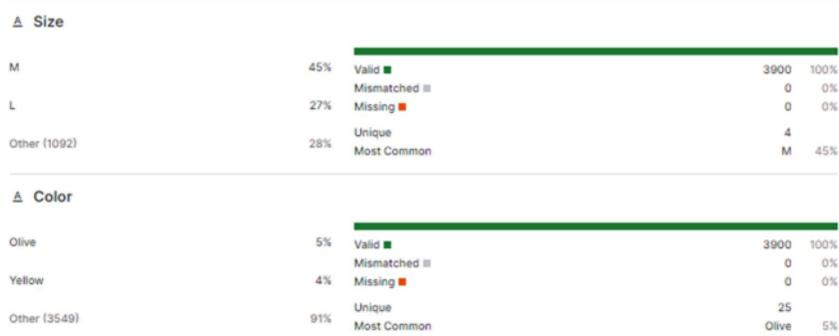
1. Pembelian Terbanyak Adalah
2. Clothing 45%,
3. Accessories 32%,
4. Other 24%



Gambar 15 Purchase

Total Jumlah Penjualan 3900

Analisa Data Shopping Trends Menggunakan Algoritma Klasifikasi Dengan Metode Naive Bayes



Gambar 16 Size

- Size

Ukuran Yang Paling Banyak Diminati Adalah Ukuran M Sebanyak 45%, Dan Ukuran L Sebanyak 27% Dan Other 28%

- Color

1. Olive 5%
2. Yellow 4%
3. Other 91%

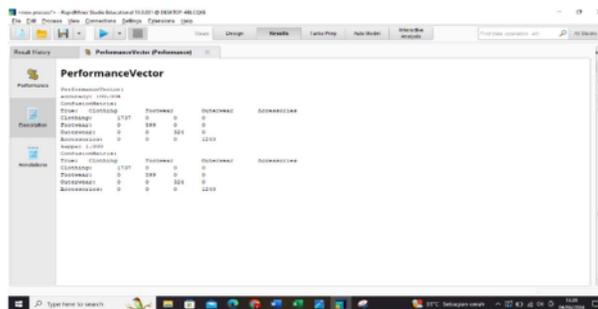
- Season



Gambar 17 Season

- Season

1. Pembelian *Spring* Atau Pembelian Di Musim Semi Sebanyak 26%
2. Dan Pembelian Di *Fall* Atau Pembelian Musim Gugur 25%,
3. Other Sebanyak 49%



Gambar 17 Performance

Description Performance Vector

HASIL HIPOTESIS

Rumus Penghitungan Di Atas Adalah Hasil Dari Pembelian Terbanyak Dibagi

Dengan Jumlah Dataset Yang Ada Disini Saya Menggunakan Rumus

$$P(C|X) = \frac{P(x|c) \cdot P(c)}{P(x)}$$

X sebagai Vektor input

c sebagai sebuah class spesifik

$P(C|X)$ sebagai Probabilitas class berdasar vector input yang diketahui (posteriori propability)

$P(c)$ sebagai Propabilitas class yang dicari (prior propability) dari keseluruhan data

$P(x|c)$ sebagai Propabilitas tiap input berdasarkan kondisi pada class

$P(x)$ sebagai Propabilitas suatu input dari keseluruhan data

- CLOTHING = $1737/3900 = 0,445 = 40\%$
- FOOTWEAR = $599/3900 = 0,153 = 20\%$
- OUTERWEAR = $324/3900 = 0,83 = 10\%$
- ACCECORIES = $1240/3900 = 0,317 = 30\%$
- TOTAL KESELURUHAN = 100%

Jika Kelipatan Di Atas (5 Atau 0,5) Akan Di Bulatkan Ke Tas Begitu Juga Dengan Kelipatan Di Bawah (5 Atau 0,5) Akan Dibulatkan Ke Bawah

PENUTUP

Kesimpulan

Dari penelitian ini, dapat disimpulkan bahwa data tren belanja metode Naive Bayes menunjukkan bahwa teknik ini efektif dalam menemukan pola dan tren belanja konsumen berdasarkan atribut-atribut yang relevan. Dengan menggunakan data historis transaksi, Naive Bayes dapat dengan akurat memprediksi preferensi belanja konsumen, kategori produk favorit, dan perilaku lainnya.

Metode Naive Bayes ideal untuk aplikasi real-time di industri ritel karena memiliki keuntungan dalam hal interpretasi yang mudah dan waktu komputasi yang cepat. Kami dapat menemukan segmentasi pelanggan dan strategi pemasaran yang lebih efektif dengan melihat faktor demografi seperti harga, lokasi belanja, dan jenis produk. Meskipun demikian, beberapa pertimbangan yang harus diperhatikan harus dipertimbangkan, seperti asumsi dasar Naive

Bayes tentang independensi variabel yang kadang-kadang tidak sesuai dengan data yang ada. Oleh karena itu, analisis tren pembelian menggunakan Naive Bayes dapat menjadi alat yang kuat untuk memahami dan memanfaatkan perilaku konsumen secara lebih efisien di pasar global yang kompetitif saat ini, karena kombinasi dengan teknik lain seperti metode kelompok atau deep learning dapat meningkatkan performa model dalam menangani kompleksitas lebih lanjut dari data tren pembelian yang modern dan dinamis.

Saran

Berdasarkan temuan dan keterbatasan penelitian ini, berikut beberapa saran untuk penelitian selanjutnya yang dapat memperdalam pemahaman tentang Analisa data shopping trends menggunakan algoritma klasifikasi dengan metode naïve bayes

1. Penelitian selanjutnya memberi dukungan kepada kolaborasi tim analis data, ahli bisnis, dan tim pemasaran. Secara teratur diskusikan hasil analisis untuk memastikan Anda memahami pola dan tren belanja konsumen dan menemukan cara terbaik untuk meningkatkan pengalaman pelanggan dan penjualan.
2. Selain algoritma Naïve Bayes, penelitian selanjutnya dapat mengeksplorasi penggunaan algoritma lain seperti C.45 atau Apriori yang mungkin lebih efisien dalam menemukan asosiasi itemset di dataset yang sangat besar.

DAFTAR PUSTAKA

- Akbar, M., & Rahmanto, Y. (2020). Desain Data Warehouse Penjualan Menggunakan Nine Step Methodology Untuk Business Intelegency Pada Pt Bangun Mitra Makmur. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 1(2), 137–146. <https://doi.org/10.33365/jatika.v1i2.331>
- Ardiansyah, D., & Walim, W. (2018). Algoritma c4.5 untuk klasifikasi calon peserta lomba cerdas cermat siswa smp dengan menggunakan aplikasi rapid miner. *Jurnal Inkofar*, 1(2), 5–12. <https://politeknikmeta.ac.id/meta/ojs/index.php/inkofar/article/view/29/45>
- Indarta, Y., Irfan, D., Muksir, M., Simatupang, W., & Ranuharja, F. (2021). Analisis dan Perancangan Database Menggunakan Model Konseptual Data Warehouse Sistem Manajemen Transaksi Toko Online Haransaf. *Edukatif: Jurnal Ilmu Pendidikan*, 3(6), 4448–4455. <https://doi.org/10.31004/edukatif.v3i6.1477>
- Kurniawan, Y. I. (2018). Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(4), 455–464. <https://doi.org/10.25126/jtiik.201854803>
- Matondang, M. R., Lubis, M. R., & Tambunan, H. S. (2021). Analisis Data mining dengan Metode C.45 pada Klasifikasi Kenaikan Rata-Rata Volume Perikanan Tangkap. *Brahmana: Jurnal Penerapan Kecerdasan Buatan*, 2(2), 74–81. <https://doi.org/10.30645/brahmana.v2i2.68>

Analisa Data Shopping Trends Menggunakan Algoritma Klasifikasi Dengan Metode Naive Bayes

ORIGINALITY REPORT

24%

SIMILARITY INDEX

22%

INTERNET SOURCES

6%

PUBLICATIONS

7%

STUDENT PAPERS

PRIMARY SOURCES

1	dspace.uui.ac.id Internet Source	4%
2	Admin Teika, Yusran Timur Samuel, Frengky Simbolon. "Perancangan Aplikasi Untuk Memprediksi Seseorang Menderita Penyakit Hipertensi Menggunakan Data Mining", TeIKa, 2017 Publication	3%
3	repo.darmajaya.ac.id Internet Source	3%
4	eprints.unisla.ac.id Internet Source	3%
5	www.dosenpendidikan.co.id Internet Source	2%
6	journal.aptii.or.id Internet Source	2%
7	he-wroteyou.xyz Internet Source	2%

8

Submitted to Universitas Pelita Harapan

Student Paper

2%

9

jsisfotek.org

Internet Source

2%

10

repository.nusamandiri.ac.id

Internet Source

2%

Exclude quotes On

Exclude matches < 2%

Exclude bibliography On