

Implementasi Data Mining Untuk Mengklasifikasikan Data Penjualan Pada Supermarket Menggunakan Algoritma Naïve Bayes

Indriyani¹, Agus Bahtiar²

STMIK IKMI Cirebon, Indonesia

Korespondensi penulis: iyaniindriyani10@gmail.com¹

Abstract

Technology is growing or developing very rapidly. Technological developments also affect the increasing demands of society so that shops or supermarkets compete to provide various types of products or daily necessities. Even though modern market supermarkets are competing with traditional markets, supermarkets will definitely be rivaled by other supermarkets. Every city must have a different trial, where customers will definitely choose which one is really worth visiting for shopping and which one is not. However, in every store and supermarket, it is necessary to know what types of products are in demand by the target customers and what kind of service they want. This supermarket dataset has sales history at 3 (three) different branches for 3 months. Requires sales classification so that the supermarket can increase its development in selling its goods. From the results of research conducted at the 3 supermarket branches, there are 16 attributes and 1000 records. Supermarkets sell general goods, electronics, fashion, food and beverages, health and beauty, home and lifestyle needs, sports and travel. To help shops and supermarkets, the author will carry out an analysis by classifying data mining on supermarket sort using the Naive Bayes algorithm to find out the types of products that customers like in 3 branches in 3 months. Product categories obtained from classifying sales products can produce an accuracy of up to 98.5%. The use of the naive Bayes algorithm and the confusional matrix method has a very important influence in obtaining an accurate and significant classification of supermarket sales.

Keywords: Data Mining, Supermarkets, Naive Bayes

Abstrak

Teknologi semakin berkembang atau berkembang sangat pesat. Perkembangan teknologi juga mempengaruhi tuntutan masyarakat yang semakin meningkat sehingga toko atau supermarket berlomba-lomba menyediakan berbagai jenis produk atau kebutuhan sehari-hari. Walau supermarket pasar moderen yang menyaingi pasar tradisional, akan tetapi supermarket pun pasti akan tersaingi oleh supermarket yang lain. Setiap kota pasti ada trial yang berbeda- beda, yang dimana pelanggan pasti akan memilih mana yang sangat pantas mereka kunjungi untuk berbelanja dan mana yang tidak. Namun, di setiap toko dan supermarket perlu diketahui jenis produk apa saja yang diminati oleh pelanggan yang akan dituju dan pelayanan yang bagaimana yang mereka inginkan. Dataset supermarket ini mempunyai histori penjualan di 3 (tiga) cabang berbeda selama 3 bulan. Membutuhkan klasifikasi penjualan supaya supermarket tersebut bisa meningkatkan perkembangannya dalam menjual barangnya. Dari hasil riset yang dilakukan di supermarket 3 cabang tersebut terdapat 16 artibut dan 1000 record, Supermarket menjual barang umum, elektronik, fashion, makanan dan minuman, kesehatan dan kecantikan, kebutuhan rumah dan gaya hidup, olahraga dan perjalanan. Untuk membantu

toko dan supermarket, penulis akan melakukan analisis dengan mengklasifikasikan data mining pada supermarket sort dengan menggunakan algoritma naive bayes untuk mengetahui jenis produk yang disukai pelanggan di 3 cabang dalam 3 bulan. Kategori produk yang diperoleh dari mengklasifikasikan produk penjualan dapat menghasilkan akurasi mencapai 98.5%. Penggunaan metode algoritma naive bayes dan confusional matrix memberikan pengaruh yang sangat penting dalam mendapatkan klasifikasi penjualan supermarket yang akurasi dan signifikan.

Kata Kunci: Data Mining, Supermarket, Naive Bayes.

I. LATAR BELAKANG

Kedaaan supermaket sendiri sangat mengalahkan pasar tradisional karena lengkapnya produk yang di jual oleh supermaket tersebut, Supermaket sejenis pasar yang dimana mereka menjual berbagai kebutuhan yang sangat mudah di dapat oleh masyarakat. Namun, untuk supermarket sendiri ini menjadi masalah, dengan tingginya minat pelanggan pada produk yang mereka inginkan, pihak supermarket harus menghadapi pengelompokan produk yang terjual. Masalah yang paling sering di alami oleh supermaket adalah terjadinya penumpukan produk yang tidak laku terjual. Masalah lainnya permintaan produk yang banyak namun stok tersedia tidak mencukupi.

Keinginan dan kebutuhan konsumen selalu berubah-ubah, Ketika konsumen melakukan pengambilan produk dengan merek tertentu. Hal itu pasti mengakibatkan penjualan produk ada yang cepat habis dan masih ada. Supermaket sendiri selalu menggunakan pembayaran secara tunai dan non tunai, Untuk mengatasi ini menggunakan pendekatan mengklasifikasikan data memakai data mining, data mining merupakan teknik yang digunakan untuk mengitung dan menggabungkan. Perkembangan data mining saat ini sangat banyak membantu dalam mengambil keputusan untuk menentukan keputusan atas masalah yang di hadapi.

Penelitian tentang klaterisasi produk untuk klasifikasi penjualan di masa covid-19 yang berdampak pada penjualan supermarket yang menjual produk perlengkapan mandi dan mencuci. Tingginya tingkat penjualan dan minat pelanggan penelitian ini menglterisasikan dan mengklasifikasikan prok apa saya yang diminati oleh pelaanggan menggunakan metode data mining algoritma K-Means dan K- Nearest Neighbors. Hasil uji klasifikasi mencapai 85.7143%.Berikut di bawah ini sebuah data penjualan supermarket yang saya dapat dari kaggle.

| Invoice ID | Branch | City | Customer | Gender | Product Line | Unit Price | Quantity | Tax 5% | Total | Date | Time | Payment | Cogs | Gross Margin | Rating |
|------------|-----------|-----------|----------|--------|--------------|------------|----------|----------|------------|------------|-------|-------------|-----------|---------------|----------|
| 1 | Yangon | Yangon | Member | Female | Health arc | 28.89 | 7 | 251.415 | 5.489.715 | 13/10/2019 | 13:08 | Ewallet | 522.81 | 4.761.904.782 | 281.415 |
| 2 | Naypyitaw | Naypyitaw | Normal | Female | Electron c | 15.28 | 5 | 5.181944 | 80.22.00 | 03/08/2019 | 10:29 | Cash | 76.04.00 | 4.761.904.782 | 0.181944 |
| 3 | Yangon | Yangon | Normal | Male | Home arc | 46.33.00 | 7 | 324.335 | 5.485.355 | 03/08/2019 | 13:28 | Credit card | 834.81.00 | 4.761.904.782 | 165.135 |
| 4 | Yangon | Yangon | Member | Male | Health arc | 28.22.00 | 8 | 225.760 | 489.040 | 10/11/2019 | 03:11 | Ewallet | 485.76 | 4.761.904.782 | 21.780 |
| 5 | Yangon | Yangon | Normal | Male | Sports arc | 86.51.00 | 7 | 332.885 | 6.343.785 | 02/08/2019 | 10:37 | Ewallet | 694.17.00 | 4.761.904.782 | 302.885 |
| 6 | Yangon | Yangon | Normal | Male | Electron c | 81.39.00 | 7 | 288.805 | 6.276.105 | 3/25/2019 | 18:30 | Ewallet | 557.73 | 4.761.904.782 | 286.885 |
| 7 | Yangon | Yangon | Member | Female | Electron c | 69.84 | 6 | 305.652 | 413.672 | 2/24/2019 | 14:30 | Ewallet | 413.04.00 | 4.761.904.782 | 26.652 |
| 8 | Naypyitaw | Naypyitaw | Normal | Female | Home arc | 21.36.00 | 10 | 213.600 | 427.260.00 | 2/24/2019 | 11:20 | Ewallet | 720.00.00 | 4.761.904.782 | 0 |
| 9 | Yangon | Yangon | Member | Female | Health arc | 30.20.00 | 2 | 6.020 | 76.140 | 01/10/2019 | 17:15 | Credit card | 72.32.00 | 4.761.904.782 | 6.020 |
| 10 | Mandalay | Mandalay | Member | Female | Food and | 54.84 | 1 | 8.736 | 177.246 | 3/28/2019 | 13:27 | Credit card | 142.57.00 | 4.761.904.782 | 8.736 |
| 11 | Mandalay | Mandalay | Member | Female | Food and | 34.46 | 4 | 2.820 | 80.816 | 02/08/2019 | 18:07 | Ewallet | 51.22 | 4.761.904.782 | 2.820 |
| 12 | Mandalay | Mandalay | Member | Male | Electron c | 25.51.00 | 4 | 5.102 | 197.142 | 09/09/2019 | 17:09 | Cash | 182.04.00 | 4.761.904.782 | 1.102 |
| 13 | Yangon | Yangon | Normal | Female | Electron c | 49.95 | 5 | 249.775 | 2.444.875 | 02/12/2019 | 10:23 | Ewallet | 236.75 | 4.761.904.782 | 117.975 |
| 14 | Yangon | Yangon | Normal | Male | Food and | 31.19.00 | 10 | 311.900 | 631.070 | 10/12/2019 | 18:56 | Ewallet | 611.89.00 | 4.761.904.782 | 21.970 |
| 15 | Yangon | Yangon | Normal | Female | Health arc | 71.38.00 | 10 | 358.49 | 740.49.00 | 3/28/2019 | 19:21 | Cash | 713.08.00 | 4.761.904.782 | 35.89 |
| 16 | Mandalay | Mandalay | Member | Female | Sports arc | 59.72 | 6 | 28.116 | 590.436 | 1/19/2019 | 15:19 | Cash | 542.32.00 | 4.761.904.782 | 26.116 |
| 17 | Yangon | Yangon | Member | Female | Health arc | 18.53 | 7 | 281.255 | 5.084.355 | 10/11/2019 | 11:01 | Credit card | 82.21.00 | 4.761.904.782 | 281.255 |
| 18 | Yangon | Yangon | Normal | Male | Sports arc | 72.81 | 6 | 21.283 | 417.443 | 01/13/2019 | 10:59 | Credit card | 415.85 | 4.761.904.782 | 21.283 |
| 19 | Mandalay | Mandalay | Normal | Male | Food and | 54.87 | 3 | 32.605 | 1.722.105 | 1/21/2019 | 18:00 | Credit card | 144.01.00 | 4.761.904.782 | 32.605 |
| 20 | Mandalay | Mandalay | Normal | Female | Home arc | 40.01.00 | 2 | 80.020 | 160.020 | 10/11/2019 | 15:30 | Ewallet | 80.020 | 4.761.904.782 | 0 |
| 21 | Yangon | Yangon | Member | Male | Electron c | 81.04.00 | 5 | 21.31 | 461.71 | 2/23/2019 | 11:24 | Ewallet | 410.02.00 | 4.761.904.782 | 21.31 |
| 22 | Yangon | Yangon | Normal | Male | Health arc | 47.93 | 3 | 13.197 | 277.419 | 03/05/2019 | 10:40 | Ewallet | 265.94 | 4.761.904.782 | 13.197 |
| 23 | Yangon | Yangon | Normal | Male | Home arc | 37.02.00 | 2 | 63.33 | 69.72 | 3/15/2019 | 12:20 | Credit card | 66.94.00 | 4.761.904.782 | 63.32 |
| 24 | Yangon | Yangon | Normal | Male | Electron c | 34.36.00 | 2 | 3.777.08 | 181.44.00 | 2/11/2019 | 11:12 | Ewallet | 172.08.00 | 4.761.904.782 | 3.777.08 |

Gambar 1. Data penjualan supermarket

Gambar diatas menunjukkan data penjualan supermarket yang dimana berisi invoice ID, City, Customer, Gender, Product Line, Unit Price, Quantity, Tax 5%, Total, Date, Time, Payment, Cogs, Gross Margin Mercentage, gross income, Rating. Pencatatan penjualan supermarket ini berasal dari 3 kota Yangon, Naypyitaw, dan Mandalay. Sistem pembayarannya pun menggunakan Credit Card, Cash, Ewallet. Sehingga proses pencatatannya sesuai barang yang keluar di tiap hari.

Dataset yang di temukan adalah pencatatan penjualan supermarket dari 3 cabang yang berbeda dan di setiap cabang supermarket itu sendiri mempunyai perbedaan dalam pencatatan penjualan supermarket. Karena daya saing supermarket yang semakin meningkat di butuhnya sebuah klasifikasi produk mana saja yang banyak peminatnya, sehingga produk tidak akan terbuang sia-sia.

Dengan begitu maka dilakunnya sebuah klasifikasi data, yang dimana bertujuan mengelompokan data mana saja yang sering dibeli dan kota mana yang banyak menjual suatu peroduk tersebut. Oleh karna itu dibuatlah klasifikasi data mining menggunakan algoritma naive bayes.

II. KAJIAN TEORITIS

Supermarket

Dari beberapa definisi dan klasifikasi tentang ritel di atas, dapat diketahui bahwa supermarket merupakan salah satu bentuk ritel, seperti perusahaan atau industri, dalam hal ini industri jasa. Selain itu, definisi supermarket akan disajikan. Menurut Stanton (1984), supermarket didefinisikan sebagai berikut:

"Supermarket adalah perusahaan ritel dengan bermacam-macam produk sederhana, mencakup bahan makanan dan beberapa produk non-makanan, dan menawarkan layanan pelanggan yang relatif kecil."

Sedangkan menurut McCarthy (1993), Supermarket dapat disimpulkan bahwa supermarket adalah suatu layanan penjualan eceran dengan struktur sektoral yang menawarkan berbagai macam barang (termasuk makanan, minuman dan beberapa lini produk non makanan) yang beroperasi dalam satu sektor. Berikan layanan sesedikit mungkin kepada konsumen berdasarkan swalayan. Dan harga yang ditawarkan menarik.

Klasifikasi

Klasifikasi ialah sesuatu metode dengan memandang pada kelakuan serta atribut dari kelompok yang sudah didefinisikan. Metode ini bisa memberikanklasifikasi pada informasi baru dengan memanipulasi informasi yang terdapat yang sudah diklasifikasi serta dengan memakai hasilnya buat membagikan beberapa ketentuan. Ketentuan ketentuan tersebut digunakan pada data- data baru buat diklasifikasi. Metode ini memakai supervised induction, yang menggunakan kumpulan pengujian dari record yang terklasifikasi buat memastikan kelas- kelas bonus mendefinisikan atribut- atribut, merendahkan dimensi informasi, melenyapkan noise(tercantum outlier), normalisasi, pula memastikan strategi buat menanggulangi informasi yang lenyap.

Data Mining

Data mining adalah istilah yang digunakan untuk membuat pengetahuan tersembunyi di basis data. Data mining adalah proses semi-otomatis yang menggunakan metode statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi dan mengidentifikasi data pengetahuan potensial dan bermanfaat yang disimpan dalam database besar. (Purba et angkatan laut(AL)., 2019)

Naïve Bayes

Algoritma Naïve Bayes ialah tata cara yang menggunakan probabilitas serta statistik buat menyelesaikan permasalahan klasifikasi. Metode menghitung nilai probabilitas $P(x|y)$ dengan mengidentifikasi probabilitas kelas X untuk klasifikasi. Penentuan kelas dalam klasifikasi dilakukan secara probabilistik dengan memilih nilai maksimum $P(x|y)$. Keuntungan dari klasifikasi adalah membutuhkan jumlah informasi pelatihan yang relatif kecil untuk mengestimasi parameter yang dibutuhkan untuk

klasifikasi. Berdasarkan algoritma Naive Bayes, rumus untuk menghitung nilai $P(x|y)$ adalah sebagai berikut:

$$P(X | Y) = \frac{P(Y|X).(X)}{P(Y)} \quad (1)$$

dimana :

$P(X|Y)$ = Posterior|probability yaitu nilai probabilitas X berdasarkan kondisi Y

$P(Y|X)$ = probabilitas Y yang ditentukan X adalah benar

$P(X)$ = Peluang evidence penyakit X

$P(Y)$ = Probabilitas dari nilai Y

III. METODE PENELITIAN

Metode pengumpulan informasi pada riset ini yakni dengan proses pengamatan. Semacam observasi, serta dokumentasi. Observasi dicoba dengan metode pengamatan dari tabel yang berisi tentang penjualan supermarket variabel dari Invoice ID, Branch, City, Customer type, Gender, Product line, Unit price, Quantity, Tax 5%, Total, Date, Time, Payment, cogs, gross margin percentage, gross income, Rating. Berikut ini ialah tata cara pengumpulan informasi serta data yang dicoba merupakan selaku berikut:

1. Observasi

Observasi ialah memperlihatkan suatu dengan memakai mata. Observasi ataupun yang diucap pula dengan pengamatan meliputi aktivitas pemusatan atensi terhadap sesuatu objek dengan memakai segala perlengkapan indra(*Teni& Agus Yudiyanto, 2021*).

2. Dokumentasi

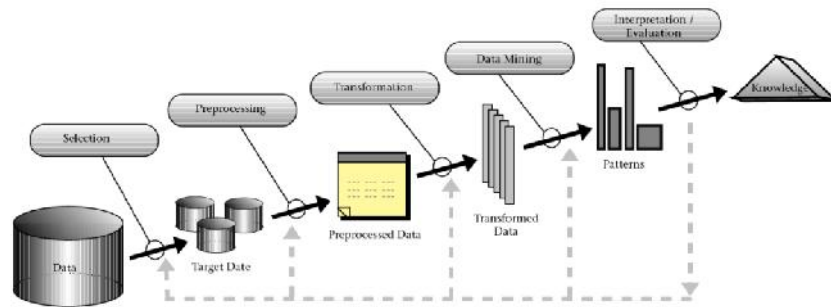
Dokumentasi ialah pengamatan secara langsung buat memperoleh informasi yang diperoleh periset cocok dengan ulasan. Perihal ini dicoba buat mengenali cerminan nyata dilapangan(*Prawiyogi et angkatan laut(AL)., 2021*).

3. Wawancara

Wawancara ialah salah satu metode yang digunakan buat mengumpulkan informasi ataupun data yang berhubungan dengan permasalahan riset di lapangan. Dengan terdapatnya wawancara periset bisa mengenali informasi serta data yang diperlukan(*Evania Dini, 2021*). Penulis mengajukan persoalan kepada narasumber buat mengumpulkan informasi ataupun data yang diperlukan dalam riset ini.

Tahapan Perancangan

Metodologi penelitian untuk digunakan dalam penelitian ini menggunakan Knowledge Discovery in Database (KDD) sebuah proses yang dirancang untuk menggali dan menganalisis data yang sangat besar menjadi informasi yang berguna untuk pengetahuan.(C et al., 2019)



Gambar 2. Alur penelitian menggunakan KDD

Langkah penting dalam proses KDD terdiri dari tahapan-tahapan sebagai berikut:

1) Data clening

Data cleaning ialah proses menyapkan Data duplikat, mengecek data yang tidak tidak berubah- ubah, serta membetulkan kesalahan pada data(semacam kesalahan tulis). Biasanya, data yang diperoleh dari database ataupun riset mempunyai konten yang tidak sempurna, semacam data yang lenyap, data tidak valid, ataupun cuma salah eja. Tidak hanya itu, terdapat atribut data yang tidak relevan dengan anggapan data mining yang Kamu miliki. Pembersihan data pula mempengaruhi hasil informatif dari metode penambangan data, sebab volume serta kompleksitas data yang diproses menurun.

2) Data selection

Data yang relevan dipilih dan dapat dianalisis dari data operasional. Data digabungkan ke dalam tabel, dan data perlu dipilih dari sekumpulan data operasional sebelum tahap penggalian informasi di KDD dimulai. Data yang dipilih untuk proses penambangan data disimpan dalam file, terpisah dari basis data operasional.

3) Data transformation

Proses transformasi informasi kedalam wujud format tertentu sehingga informasi tersebut cocok buat proses informasi mining. Sehabis informasi bersih dari kesalahan, berikutnya dicoba transformasi pada informasi cocok dengan tipe informasi pada tahapan transformasi dimana tipe hendak dikelompokkan jadi informasi yang bertabiat jenis.

4) Data mining

Penambangan data adalah proses penggunaan teknik atau metode tertentu untuk menemukan pola atau informasi menarik dalam data terpilih. Teknik, metode atau algoritma dalam data mining sangat bervariasi. Memilih metode atau algoritma yang tepat sangat bergantung pada tujuan dan keseluruhan proses KDD. Dalam penelitian ini menggunakan metode algoritma Naïve Bayes

5) Pattern evaluation

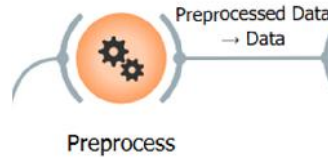
Tahapan ini digunakan untuk mengevaluasi hasil prediksi yang dihasilkan oleh keempat algoritma tersebut dan memilih metode algoritma yang menghasilkan nilai klasifikasi yang mendekati data sebenarnya. Evaluasi dilakukan dengan menggunakan metode matriks konfusi dan kurva ROC (receiver operating Characteristic). Nilai performa yang digunakan adalah akurasi dan error.

6) Knowledge presentation

Skema yang menyajikan informasi yang dihasilkan selama proses klasifikasi, yang menggambarkan hasil, menjadikannya informasi untuk proses akhir dari proses KDD. Memvisualisasikan data yang diproses untuk memudahkan pengguna memahami adalah harapan bahwa tindakan dapat diambil berdasarkan analisis.

IV. HASIL DAN PEMBAHASAN

Tahapan preprocessing data dalam penelitian ini memakai *widget* preprocessing data dengan melakukan *Normalize features* pada tools orange kemudian pilih *Normalize to interval [-1,1]*.



Gambar 3. Widget Proses preprocessing data

Gambar 3 *Widget Proses preprocessing data* digunakan untuk proses preprocessing data mining dalam tools orange. *Widget ini* melakukan Transformasi menghasilkan nilai -1 dan deviasi standar 1 untuk setiap atribut

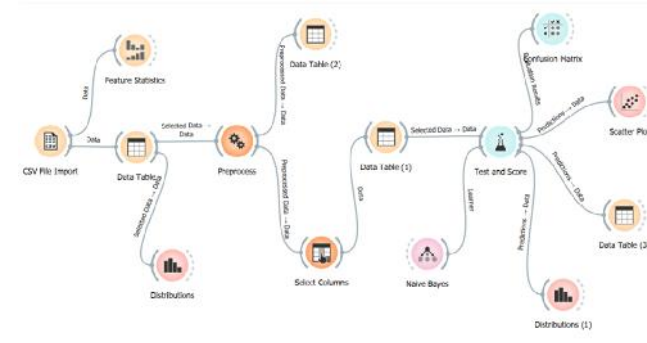
Transformasi

Transformasi data merupakan proses untuk merubah type atribut data diubah menjadi bentuk yang sesuai dengan proses data mining.. Dibawah ini adalah data terdiri dari 16 atribut dan 1000 record yang sudah melalui proses transformasi data seperti yang terlihat dalam gambar data table hasil dari orange.

Gambar 4. Data Tabel Orange

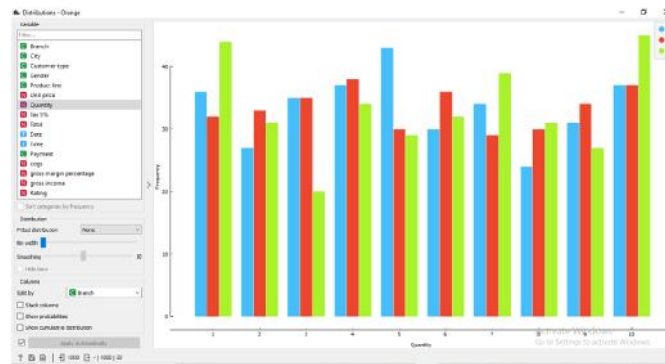
Data Mining

Pada sesi ini dicoba pemilihan metode data mining yang cocok. setelah itu informasi tersebut hendak distandarisasi menajaki proses tahapan data mining supaya informasi tersebut layak serta bisa diolah memakai orange dengan tata cara algoritma Naïve Bayes.



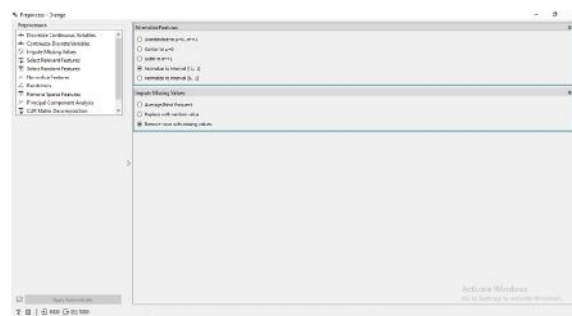
Gambar 5. Proses Algoritma Naïve Bayes Menggunakan Orange

Proses pada gambar 5 proses algoritma Naïve Bayes menggunakan orange yang digunakan untuk melakukan klasifikasi terhadap data supermarket sales menggunakan metode algoritma Naïve Bayes.



Gambar 6. Hasil Distributions-Orange

Pada gambar 6 merupakan hasil dari distribusi dari brand A, B, dan C berdasarkan quantity produk tersebut.



Gambar 7. Hasil preprocessing-Orange

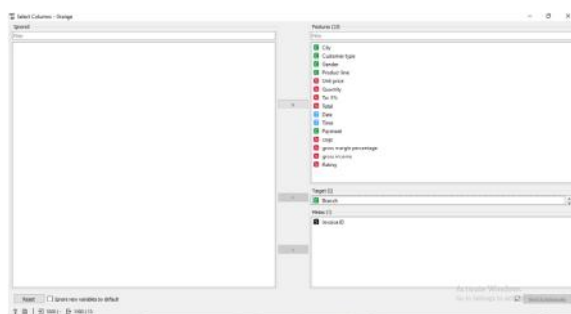
Pada Gambar 7 merupakan hasil dari penerapan widget preprocessing menggunakan tools orange untuk melakukan proses preprocessing terhadap data supermarket sales sebelum dilakukan proses klasifikasi menggunakan algoritma Naïve Bayes.

Implementasi Data Mining Untuk Mengklasifikasikan Data Penjualan Pada Supermarket Menggunakan Algoritma Naïve Bayes

The screenshot shows a data table with the following columns: Invoice ID, Branch, City, Customer type, Gender, Product line, Unit price, Quantity, Tax 5%, Total, Date, Time, Payment, cogs, gross margin percentage, gross income, and Rating. The table contains 20 rows of data representing supermarket sales transactions.

Gambar 8. Hasil dari widget Data Tabel-Orange

Gambar 8 merupakan hasil dari widget Data Tebel-Orange Setelah dilakukan proses preprocessing data menggunakan widget preprocessing terhadap supermarket sales.



Gambar 9. Hasil widget select coloum

Pada gambar 9 tahapan yang digunakan untuk melakukan seleksi atribut dari supermarket sales terdiri dari Invoice ID, Branch, City, Customer type, Gender, Product line, Unit price, Quantity, Tax 5%, Total, Date, Time, Payment, cogs, gross margin percentage, gross income, Rating yang akan dilakukan proses klasifikasi menggunakan algoritma naïve bayes menggunakan tools orange.



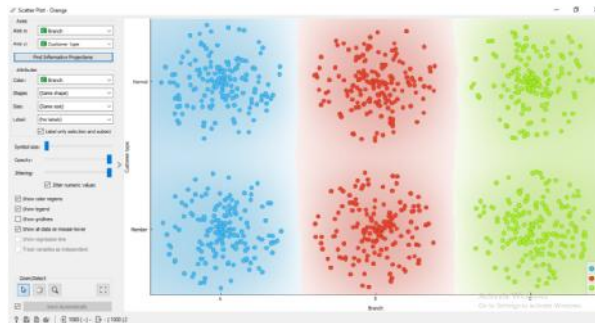
Gambar 10. Hasil Widget Algoritma Naïve Bayes

Berdasarkan gambar 10 merupakan penerapan algoritma Naïve Bayes menggunakan tools orange terhadap data supermarket sales dengan melakukan klasifikasi berdasarkan brand.

| Brand | Brand ID | Product ID | Product Name | Product Type | Product Category | Product Sub-Category | Product Description |
|-------|------------|------------|--------------|--------------|------------------|----------------------|---------------------|
| 1 | 1000000001 | 1000000001 | 1000000001 | 1000000001 | 1000000001 | 1000000001 | 1000000001 |
| 2 | 1000000002 | 1000000002 | 1000000002 | 1000000002 | 1000000002 | 1000000002 | 1000000002 |
| 3 | 1000000003 | 1000000003 | 1000000003 | 1000000003 | 1000000003 | 1000000003 | 1000000003 |
| 4 | 1000000004 | 1000000004 | 1000000004 | 1000000004 | 1000000004 | 1000000004 | 1000000004 |
| 5 | 1000000005 | 1000000005 | 1000000005 | 1000000005 | 1000000005 | 1000000005 | 1000000005 |
| 6 | 1000000006 | 1000000006 | 1000000006 | 1000000006 | 1000000006 | 1000000006 | 1000000006 |
| 7 | 1000000007 | 1000000007 | 1000000007 | 1000000007 | 1000000007 | 1000000007 | 1000000007 |
| 8 | 1000000008 | 1000000008 | 1000000008 | 1000000008 | 1000000008 | 1000000008 | 1000000008 |
| 9 | 1000000009 | 1000000009 | 1000000009 | 1000000009 | 1000000009 | 1000000009 | 1000000009 |
| 10 | 1000000010 | 1000000010 | 1000000010 | 1000000010 | 1000000010 | 1000000010 | 1000000010 |
| 11 | 1000000011 | 1000000011 | 1000000011 | 1000000011 | 1000000011 | 1000000011 | 1000000011 |
| 12 | 1000000012 | 1000000012 | 1000000012 | 1000000012 | 1000000012 | 1000000012 | 1000000012 |
| 13 | 1000000013 | 1000000013 | 1000000013 | 1000000013 | 1000000013 | 1000000013 | 1000000013 |
| 14 | 1000000014 | 1000000014 | 1000000014 | 1000000014 | 1000000014 | 1000000014 | 1000000014 |
| 15 | 1000000015 | 1000000015 | 1000000015 | 1000000015 | 1000000015 | 1000000015 | 1000000015 |
| 16 | 1000000016 | 1000000016 | 1000000016 | 1000000016 | 1000000016 | 1000000016 | 1000000016 |
| 17 | 1000000017 | 1000000017 | 1000000017 | 1000000017 | 1000000017 | 1000000017 | 1000000017 |
| 18 | 1000000018 | 1000000018 | 1000000018 | 1000000018 | 1000000018 | 1000000018 | 1000000018 |
| 19 | 1000000019 | 1000000019 | 1000000019 | 1000000019 | 1000000019 | 1000000019 | 1000000019 |
| 20 | 1000000020 | 1000000020 | 1000000020 | 1000000020 | 1000000020 | 1000000020 | 1000000020 |
| 21 | 1000000021 | 1000000021 | 1000000021 | 1000000021 | 1000000021 | 1000000021 | 1000000021 |
| 22 | 1000000022 | 1000000022 | 1000000022 | 1000000022 | 1000000022 | 1000000022 | 1000000022 |
| 23 | 1000000023 | 1000000023 | 1000000023 | 1000000023 | 1000000023 | 1000000023 | 1000000023 |
| 24 | 1000000024 | 1000000024 | 1000000024 | 1000000024 | 1000000024 | 1000000024 | 1000000024 |
| 25 | 1000000025 | 1000000025 | 1000000025 | 1000000025 | 1000000025 | 1000000025 | 1000000025 |
| 26 | 1000000026 | 1000000026 | 1000000026 | 1000000026 | 1000000026 | 1000000026 | 1000000026 |
| 27 | 1000000027 | 1000000027 | 1000000027 | 1000000027 | 1000000027 | 1000000027 | 1000000027 |
| 28 | 1000000028 | 1000000028 | 1000000028 | 1000000028 | 1000000028 | 1000000028 | 1000000028 |
| 29 | 1000000029 | 1000000029 | 1000000029 | 1000000029 | 1000000029 | 1000000029 | 1000000029 |
| 30 | 1000000030 | 1000000030 | 1000000030 | 1000000030 | 1000000030 | 1000000030 | 1000000030 |

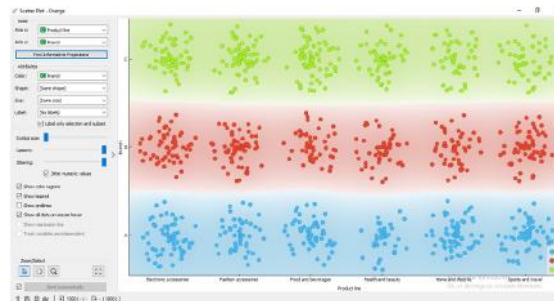
Gambar 11. Hasil dari klasifikasi

Hasil penerapan algoritma Naive Bayes dapat dilihat pada gambar di bawah ini:



Gambar 12. Hasil visualisasi Scatter Plot

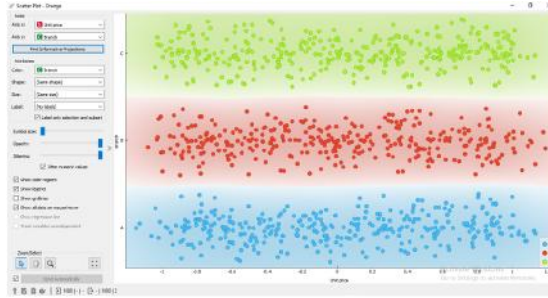
Gambar 12 hasil dari penerapan widget Scatter Plot yang digunakan untuk memvisualisasikan merupakan hasil klasifikasi menggunakan algoritma naïve bayes berdasarkan brand terhadap customer pada tools orange.



Gambar 13. Hasil Visualisasi Scatter Plot Product line

Gambar 13 hasil dari penerapan widget Scatter Plot yang digunakan untuk memvisualisasikan merupakan hasil klasifikasi menggunakan algoritma naïve bayes berdasarkan Product line terhadap branch pada tools orange.

Implementasi Data Mining Untuk Mengklasifikasikan Data Penjualan Pada Supermarket Menggunakan Algoritma Naïve Bayes



Gambar 14. Hasil Visualisasi Scatter Plot Unit Price

Gambar 14 hasil dari penerapan widget Scatter Plot yang digunakan untuk memvisualisasikan merupakan hasil klasifikasi menggunakan algoritma naïve bayes berdasarkan Unit Price terhadap branch pada tools orange.

| | | Predicted | | | Σ |
|----------|---|-----------|-----|-----|----------|
| | | A | B | C | |
| Actual | A | 331 | 2 | 1 | 334 |
| | B | 3 | 329 | 2 | 334 |
| | C | 6 | 1 | 325 | 332 |
| Σ | | 340 | 332 | 328 | 1000 |

Gambar 15. Hasil Visualisasi Confusion Matrik-orange

Melihat hasil gambar 15 jumlah yang diklasifikasi berdasarkan brand A ada 334 brand ternyata hasil preidiksi actual A 331 brand dan actual B 2 brand dan actual C 1, brand B berjumlah 334 hasil prediksi actual B 329 brand, actual A 3 brand dan actual C 2 brand, untuk jumlah brand C 332 hasil prediksi actual C 325 brand, actual A 6 brand dan actual B 1 brand.

Evaluasi

Penilaian kinerja model klasifikasi yang dihasilkan dapat dilihat dari hasil prediksi subjek uji benar dan salah. Model klasifikasi akan ditentukan oleh nilai eksaknya. Akurasi klasifikasi menentukan persentase akurasi untuk catatan data yang diklasifikasikan dengan benar. Hasil perhitungan ditabulasikan dalam bentuk matriks konfusi dengan beberapa nilai true positive (TP), false negative (FN) dan true negative (TN).

Perhitungan akurasi menggunakan rumus : $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100\%$

$$Accuracy = \frac{331 + 329 + 325}{331 + 329 + 325 + 2 + 1 + 2 + 3 + 6 + 1} * 100\%$$

$$Accuracy = \frac{985}{1000} * 100\%$$

$$Accuracy = 98,50\%.$$

V. KESIMPULAN DAN SARAN

Dalam penelitian ini, ditemukan masalah yang terjadi pada Supermarket khususnya pada bagian penjualan. Sehingga digunakan metode data mining dengan menggabungkan proses klasifikasi. klasifikasi digunakan untuk mengklasifikasikan penjualan. Penggunaan metode hybrid antara algoritma Naive Bayes dan Confusion Matrix memberikan pengaruh yang signifikan dalam menentukan klasifikasi.

DAFTAR REFERENSI

- Bertalya. (2009). *Konsep Data Mining, Klasifikasi: Pohon Keputusan*. Universitas Gunadarma.
- C, D. A., Bayes, N., & Svm, D. A. N. (2019). *Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika*. 13(1), 16–25.
- Evania Dini, A. R. S. . (2021). Optimalisasi Pembelajaran Jarak Jauh Dengan Memanfaatkan Sarana Teknologi Informasi Dan Komunikasi. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, 7(02), 348–363. <https://doi.org/10.36989/didaktik.v7i02.222>
- Faid, M. (2019). *Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi*. 8. <https://doi.org/10.34148/teknika.v8i1.95>
- Hadianto, N., Novitasari, H. B., & Rahmawati, A. (2019). Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode Neural Network. *Jurnal Pilar Nusa Mandiri*, 15(2), 163–170. <https://doi.org/10.33480/pilar.v15i2.658>
- Prawiyogi, A. G., Sadiyah, T. L., Purwanugraha, A., & Elisa, P. N. (2021). Penggunaan Media Big Book untuk Menumbuhkan Minat Membaca di Sekolah Dasar. *Jurnal Basicedu*, 5(1), 446–452. <https://doi.org/10.31004/basicedu.v5i1.787>
- Purba, W., Siawin, W., & . H. (2019). Implementasi Data Mining Untuk Pengelompokan Dan Prediksi Karyawan Yang Berpotensi Phk Dengan Algoritma K-Means Clustering. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, 2(2), 85–90. <https://doi.org/10.34012/jusikom.v2i2.429>
- Shafarindu, A. I., Patimah, E., Siahaan, Y. M., & Wardhana, A. W. (2021). *Klasifikasi Data Penjualan pada Supermarket dengan Metode Decision Tree*. April, 660–667.
- Sugianto, C. A., Maulana, F. R., & Mining, D. (2019). *Algoritma Naïve Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai (Studi Kasus Kelurahan Utama)*. 18(4), 321–331.
- Sulistiani, H., Utami, Y. T., Informatika, J., & Indonesia, U. T. (2018). *Penerapan algoritma klasifikasi sebagai pendukung keputusan pemberian beasiswa mahasiswa*. 300–305.