



# Implementasi Naïve Bayes Pada Data Mining Untuk Mengklasifikasikan Penjualan Barang Terlaris Pada Perusahaan Ritel

Ni Wayan Wardani<sup>1\*</sup>, Putu Gede Surya Cipta Nugraha<sup>2</sup>, Gede Surya Mahendra<sup>3</sup> 

<sup>1,2,3</sup> Program Studi Teknik Informatika, STMIK STIKOM Indonesia, Denpasar, Indonesia

## ARTICLE INFO

### Article history:

Received January 07, 2023

Accepted September 20, 2023

Available online October 25, 2023

### Kata Kunci:

Data mining, Naïve Bayes, ritel, klasifikasi.

### Keywords:

Data mining, Naïve Bayes, retail, classification.



This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Copyright © 2023 by Author. Published by Universitas Pendidikan Ganesha.

## ABSTRAK

Perusahaan ritel UD. Mawar Sari memiliki hingga 1,000,541 record data transaksi penjualan pada tahun 2019–2020 yang selama ini belum dipergunakan dengan baik. Data tersebut membutuhkan banyak ruang penyimpanan. Data tersebut juga belum dipergunakan secara efektif untuk tujuan apapun. Berdasarkan permasalahan tersebut, maka penelitian ini bertujuan untuk melakukan klasifikasi terhadap data transaksi penjualan untuk menemukan klasifikasi barang terlaris dengan data mining. Penelitian ini merupakan penelitian studi kasus yang menggunakan pendekatan penelitian kuantitatif. Metode yang digunakan adalah Naïve Bayes. Implementasi penelitian ini menggunakan Rapidminer 9.0. Hasil penelitian dari klasifikasi data transaksi penjualan adalah klasifikasi produk yang laris dan tidak laris. Peneliti melakukan pengujian dengan 6043 data training dan hasil dari split validation terbaik pada split validation 0,7 dengan accuracy sebesar 93,82%. Hasil penelitian ini dapat digunakan dan berimplikasi baik untuk manajemen UD. Mawar Sari dalam pengambilan keputusan untuk menentukan suatu produk setelah mendapatkan hasil perhitungan data mining menggunakan metode Naïve Bayes.

## ABSTRACT

UD retail company. Mawar Sari has up to 1,000,541 records of sales transaction data in 2019–2020 which have not been used properly. Such data takes up a lot of storage space. The data has also not been used effectively for any purpose. Based on these problems, this study aims to classify sales transaction data to find the best-selling classification of goods with data mining. This research is a case study research that uses a quantitative research approach. The method used is Naïve Bayes. The implementation of this research uses Rapidminer 9.0. The research results of the classification of sales transaction data are the classification of best-selling and non-best-selling products. Researchers tested with 6043 training data and the results of the best split validation at 0.7 split validation with an accuracy of 93.82%. The results of this study can be used and have good implications for UD management. Mawar Sari in making decisions to determine a product after obtaining the results of data mining calculations using the Naïve Bayes method.

## 1. PENDAHULUAN

Dewasa ini persaingan di dunia bisnis ritel dan teknologi semakin berkembang pesat dan sangat kompetitif. Hal ini membuat para pelaku bisnis ritel berlomba-lomba untuk mengembangkan bisnis mereka dengan memanfaatkan teknologi yang ada. Perkembangan bisnis ritel yang semakin banyak dan menjamur di suatu wilayah membuat kebutuhan konsumen semakin meningkat. Data transaksi penjualan harian yang terus bertambah dan mengakibatkan penyimpanan data dalam jumlah yang sangat besar. Namun, data tersebut biasanya hanya diproses sebagai catatan tanpa pengolahan lebih lanjut, sehingga tidak ada nilai tambah untuk penggunaan yang tepat (Rahmi & Mikola, 2021; Syahril et al., 2020). Untuk memenuhi keinginan konsumen para pelaku bisnis harus bisa mengklasifikasikan jenis produk yang disukai dan paling sering dibeli oleh konsumen. Hal ini penting dilakukan agar tidak terjadi tumpukan stok yang berisiko terjadinya kerugian. Oleh karena itu diperlukan suatu analisis pada data penjualan pada tahun-tahun sebelumnya dan analisis yang dapat digunakan untuk menemukan informasi dari banyak data. Data mining adalah suatu solusi yang tepat dalam menjawab permasalahan tersebut. Data mining adalah proses menemukan korelasi, pola, dan tren baru dan bermakna dengan menyaring sejumlah besar data yang disimpan dalam repositori, menggunakan teknik penalaran pola, teknik statistik dan matematika (D. Darmansah & Wardani, 2020; Nugraha & Mahendra, 2020). Data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan

\*Corresponding author.

E-mail addresses: [niwayan.wardani@stiki-indonesia.ac.id](mailto:niwayan.wardani@stiki-indonesia.ac.id) (Ni Wayan Wardani)

matematika (Bimantara & Safii, 2021; Rifky et al., 2022). Data mining merupakan bentuk penggalian data yang digunakan untuk menggali pengetahuan dari sejumlah data yang besar sehingga didapat informasi yang dapat digunakan untuk memprediksi penjualan (Apriyani et al., 2023; Rahmat et al., 2023). Bagian yang sangat penting dalam data mining adalah teknik klasifikasi, yaitu bagaimana mempelajari sekumpulan data sehingga dihasilkan aturan yang bisa mengklasifikasi atau mengenali data-data baru yang belum pernah dipelajari. Klasifikasi dapat didefinisikan sebagai proses untuk menyatakan suatu objek data sebagai salah satu kategori yang telah didefinisikan sebelumnya. Salah satu pendekatan yang dapat digunakan untuk mengklasifikasi penjualan barang terlaris pada sebuah ritel yaitu dengan menganalisis data penjualan dan melakukan pengolahan data untuk mengetahui hasil klasifikasi dari data tersebut. Penelitian ini menggunakan model *Naïve Bayes* untuk mengklasifikasikan penjualan barang terlaris yang bekerja berdasarkan sebuah pengklasifikasian probabilistik sederhana dan menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari data set yang diberikan. Algoritma metode *Naïve Bayes* merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan (Ginantra & Wardani, 2019a; Islamiyati & Fikri, 2022).

UD. Mawar Sari adalah salah satu perusahaan ritel yang memanfaatkan data mining dalam mengklasifikasikan penjualan barang terlaris di ritelnya. Berdiri sejak tahun 1976, UD. Mawar Sari berlokasi di Jln. Ahmad Yani No.5, Subagan, Amlapura, Karangasem merupakan perusahaan ritel yang menjual kebutuhan masyarakat sehari-hari. Data yang digunakan adalah data transaksi penjualan selama 2 tahun dari tahun 2019 – 2020. Penelitian ini menggunakan model CRISP-DM karena CRISP-DM adalah de facto standar untuk pengembangan proyek data mining dan knowledge discovery karena paling banyak digunakan dalam pengembangan data mining (Mahendra & Nugraha, 2020; Wantoro, 2020). Data mining sering juga disebut Knowledge Discovery in Database, adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Kant & Ansari, 2016; Vora & Oza, 2013). Data mining adalah proses menggunakan statistik, matematika, kecerdasan buatan dan teknik pembelajaran mesin untuk mengekstrak dan mengidentifikasi informasi yang berguna dan pengetahuan terkait dari database besar (Cholil et al., 2019; D. D. Darmansah & Wardani, 2021; Gunawan et al., 2018; Gupta & Chandra, 2020). Klasifikasi merupakan proses untuk menemukan fungsi dan model yang dapat membedakan atau menjelaskan konsep atau class data dengan tujuan memperkirakan class yang tidak diketahui dari suatu objek (Apriliah et al., 2021; Sisodia & Sisodia, 2018). Dalam proses pengklasifikasian biasa terdapat dua proses yang harus dilakukan yaitu proses training dan proses testing. Algoritma *Naïve Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi (Burhanudin et al., 2018; Wardani & Ariasih, 2019). *Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai teorema Bayes (Indriyani et al., 2020; Wardani, 2020). *Naïve Bayes* memiliki kelebihan antara lain, sederhana, cepat, dan berakurasi tinggi (Fibrianda & Bhawiyuga, 2018; Ginantra & Wardani, 2019b). *Naïve Bayes* adalah suatu klasifikasi yang bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas (Dewanti et al., 2021; Wulandari et al., 2019). K-Nearest Neighbor (KNN) merupakan metode yang biasa digunakan pada klasifikasi data (Faisal & Tutz, 2021; Kück & Freitag, 2021). Algoritma ini digunakan untuk mengklasifikasikan terhadap objek berdasarkan data pembelajaran yang jarak tetangganya paling dekat atau memiliki nilai selisih yang kecil dengan objek tersebut.

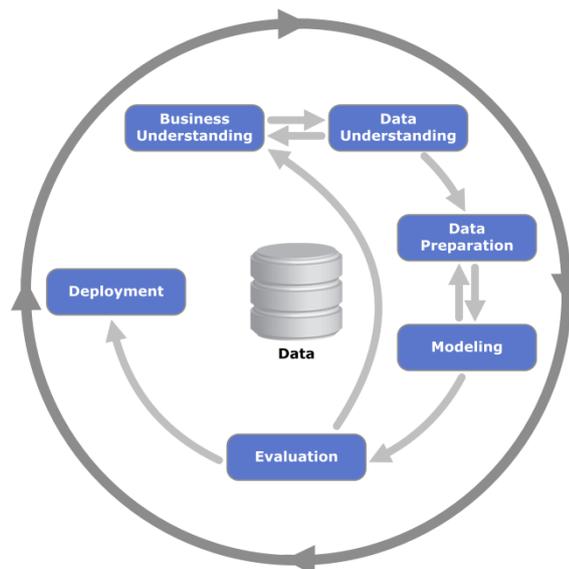
Pohon (*tree*) adalah sebuah struktur data yang terdiri dari simpul (*node*) dan rusuk (*edge*) (Aulia, 2018; Firmansyah & Yulianto, 2021). Simpul pada sebuah pohon dibedakan menjadi tiga, yaitu simpul akar (*root node*), simpul percabangan/internal (*branch/ internal node*) dan simpul daun (*leaf node*). Pohon keputusan merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas berhingga, dimana simpul internal maupun simpul akar ditandai dengan nama atribut, rusuk-rusuknya diberi label nilai atribut yang mungkin dan simpul daun ditandai dengan kelas-kelas yang berbeda (Azhari et al., 2021; Rukmana et al., 2021). Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Confusion matrix berisikan informasi mengenai hasil klasifikasi aktual dan telah di prediksi oleh sistem klasifikasi (Irawan et al., 2019; Mahendra & Aryanto, 2019). Performa dari sistem tersebut biasanya dievaluasi menggunakan data dalam sebuah matrix. *True Negative* (TN) adalah jumlah data negatif yang terdeteksi dengan benar. *False Positive* (FP) adalah data negatif, tetapi terdeteksi sebagai data positif. *True positive* (TP) adalah data yang diuji sebagai benar-benar positif. *False Negative* (FN) adalah kebalikan dari TP, sehingga datanya positif, tetapi terdeteksi sebagai data negatif. Evaluasi dengan confusion matrix menghasilkan nilai accuracy, precision, dan recall (Mahendra & Indrawan, 2020; Saputra & Nugraha, 2021). RapidMiner merupakan perangkat lunak yang bersifat terbuka (*open source*). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data

mining, text mining dan analisis prediksi (Al Zukri et al., 2020; Yuniastari & Wirawan, 2014). RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik (Romadhan, 2018; Sudarsono et al., 2021). RapidMiner memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi. RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi.

Berdasarkan pemaparan yang telah dijelaskan sebelumnya, penelitian ini bertujuan untuk melakukan klasifikasi penjualan barang terlaris pada UD. Mawar Sari menggunakan data mining. Selain untuk dapat mengimplementasikan metode *Naïve Bayes* menggunakan RapidMiner, penelitian ini juga bermanfaat besar bagi UD. Mawar Sari untuk dapat memilih produk yang berpotensi besar untuk terjual, merencanakan stok barang dengan efisien, serta melakukan manajemen ritel dengan lebih baik. Urgensi dari penelitian ini adalah untuk dapat memberikan standar baru pada industri ritel mikro agar dapat menggunakan data mining yang bertujuan untuk meningkatkan laba yang dapat diperolehnya.

## 2. METODE

Penelitian ini merupakan penelitian studi kasus yang menggunakan pendekatan penelitian kuantitatif. Penelitian ini menggunakan model CRISP-DM yang menyediakan standar proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian (Exenberger & Bucko, 2020; Schröer et al., 2021; Wardani et al., 2018). Permasalahan yang berkaitan dengan data seperti data mining dan SPK dapat menggunakan metode CRISP-DM (Feblian & Daihani, 2016; Mahendra et al., 2021; Rivai, 2020). Dalam CRISP-DM yang telah disajikan pada Gambar 1 menunjukkan sebuah proyek data mining memiliki siklus hidup yang terbagi dalam enam fase. Keseluruhan fase berurutan yang ada tersebut bersifat adaptif. Fase berikutnya dalam urutan bergantung kepada keluaran dari fase sebelumnya.



**Gambar 1.** Model CRISP DM  
(Sumber : (Nugraha & Mahendra, 2020))

Pada fase *business understanding*, dilakukan pemahaman mengenai bisnis, penilaian situasi, hingga menerjemahkan tujuan bisnis ke dalam tujuan *data mining*. Beberapa tujuan bisnis pada perusahaan ritel adalah menciptakan tersedianya pilihan terhadap kombinasi produk sesuai dengan yang diinginkan oleh konsumen oleh karena itu diperlukan suatu pengklasifikasian penjualan barang terlaris serta mempermudah bagian pembelian dan penyedia stok barang pada UD. Mawar Sari dalam melakukan perencanaan penyediaan stok barang serta manajemen ritel dapat mengetahui item barang yang paling banyak dibeli oleh konsumen. UD. Mawar Sari merupakan perusahaan ritel yang berdiri sejak tahun 1976 dan berlokasi di Jln. Ahmad Yani No.5, Subagan, Amlapura. UD. Mawar Sari menjual kebutuhan masyarakat sehari-hari secara eceran dan grosir. Di tahun 2019, UD. Mawar Sari memiliki total record sebanyak 563,350 transaksi penjualan dan 9,511 item produk yang berbeda dan rata-rata jumlah transaksi penjualan sebulan sebanyak 46,945 transaksi serta rata-rata pendapatan kotor sebanyak 528 juta rupiah dalam sebulan. Sejak tahun 2009, UD. Mawar Sari telah menggunakan sistem informasi berbasis komputer untuk menangani point of sales (POS) dan inventory. Salah satu tujuan dari *data mining* untuk klasifikasi

penjualan produk terlaris adalah membuat prediksi penjualan barang terlaris di masa depan dengan memanfaatkan data yang sudah ada sebelumnya. Hasil prediksi dapat digunakan oleh pihak manajemen untuk mengetahui item yang paling banyak dibeli oleh konsumen dan mempermudah bagian pembelian serta penyedia stok dalam melakukan perencanaan penyediaan stok barang. Pada fase *data understanding* dilakukan pengumpulan data awal dan mendeskripsikan data. Dataset transaksi penjualan didapatkan dari UD. Mawar Sari pada tahun 2019-2020 (Januari 2019–Agustus 2020) berupa dokumen microsoft excel sejumlah 1,000,541 *record* transaksi penjualan. Dataset transaksi penjualan terdiri dari 25 atribut antara lain yaitu Tanggal, Jam, No\_Ej, Id Barang, Nama Barang, Harga, Qty, Disc1, Disc2, Discrp, Aver, Aver\_Konsi, Bruto, Netto, Bkp, Id, Bo, Tglc, Fee, Cashback, Stn, Stnstd, Isi, Inv, Konsi. Pada fase *data preparation* dilakukan pemilihan data, data *preprocessing* dan transformasi data. Data yang digunakan dalam penentuan pemilihan produk terlaris adalah atribut penjualan caturwulan atau 4 bulan dari setiap data produk selama 2 tahun (Januari 2019 hingga Agustus 2020).

Selanjutnya dilakukan proses pembersihan data dilakukan dengan melakukan metode- metode *query* sederhana untuk menemukan anomali-anomali data yang kemungkinan masih terdapat pada sistem. Dengan mengelompokkan atribut-atribut atau field yang telah terpilih menjadi 1 tabel dengan cara melakukan denormalisasi dari dataset transaksi penjualan pada UD. Mawar Sari. Dalam proses *data preparation* dibangun suatu *data warehouse* untuk mempermudah proses. Data yang dilibatkan dalam penelitian adalah data yang berkaitan dengan transaksi penjualan. Data set yang digunakan dapat disesuaikan selama proses *data mining*, karena dimungkinkan untuk kembali ke fase sebelumnya jika ada informasi dan pengetahuan yang belum terungkap.

Pada proses *modeling*, dilakukan analisa menggunakan metode klasifikasi dengan algoritma *Naïve Bayes*. Data dihitung dengan menggunakan algoritma ini sesuai dengan metodenya kemudian di cari hasil akurasi. Data transaksi penjualan di olah dengan algoritma *Naïve Bayes*, kemudian dilihat tingkat akurasi dari algoritma ini dan keterkaitan antara setiap atribut. Algoritma *Naïve Bayes* mengevaluasi setiap atribut target. *Naïve Bayes* tidak memperhitungkan relasi antar atribut-atribut kontributor prediksi, tidak seperti *Decision Tree*. Bentuk tugas dasar yang dilakukan oleh algoritma *Naïve Bayes* adalah hanyalah klasifikasi. Pada fase *evaluation*, dilakukan pengecekan terhadap setiap nilai atribut dan model yang sudah dibangun. Kemudian melakukan evaluasi terhadap hasil dengan melakukan analisis dari setiap *variable output* atau karakteristik informasi yang dihasilkan oleh model *data mining*. Pada fase ini juga merupakan fase dimana dilakukan perbaikan kembali bila terjadi kekurangan. Pada fase ini bisa saja kembali lagi ke fase yang pertama dan kemudian ke fase berikutnya dengan tujuan perbaikan, sampai sesuai dengan kebutuhan, seterusnya adalah fase akhir dalam pengembangan *data mining*. Pada fase ini digunakan perhitungan akurasi berdasarkan *confusion matrix*. Pada fase *deployment*, dilakukan penyusunan dan presentasi dari pengetahuan yang didapat dari *data mining*. Presentasi dari hasil *data mining* ini, dibuat sedemikian rupa sehingga bisa dimengerti dan dibaca oleh kalangan awam.

### 3. HASIL DAN PEMBAHASAN

#### Hasil

##### Langkah Perhitungan

Pada tahap ini metode yang digunakan dalam perhitungan tingkat akurasi adalah algoritma *Naïve Bayes* dengan melakukan pengujian akurasi data set perhitungan manual. Berikut langkah metode algoritma *Naïve Bayes*. Pada Gambar 2, model algoritma *Naïve Bayes* dimulai dengan identifikasi sampel dari data set baca data. Selanjutnya  $P(X_i|C_i)$ , dimana dilakukan perhitungan jumlah *class* dari klasifikasi yang sudah terbentuk yaitu *class* laris dan tidak laris untuk setiap *class*. Kemudian  $P(X_i|C_i)$ , dilanjutkan dengan menghitung jumlah kasus yang sama dari kelas yang sama X, dalam kasus data set pada penelitian ini terdiri dari 2 *class* yaitu laris yang dinyatakan dengan simbol (+) dan tidak laris yang dinyatakan dengan simbol (-). Kemudian hitung  $P(X_i|C_i)$ , dimana  $i = +/ -$ , untuk setiap *class* atau atribut. Setelah itu dibandingkan, jika  $P(X|C^+) > P(X|C^-)$ , maka kesimpulannya adalah  $C^+$  atau pada penelitian ini berarti data penjualan barang laris. Jika  $P(X|C^+) < P(X|C^-)$ , maka kesimpulannya  $C^-$  atau tidak laris.

##### Seleksi Data

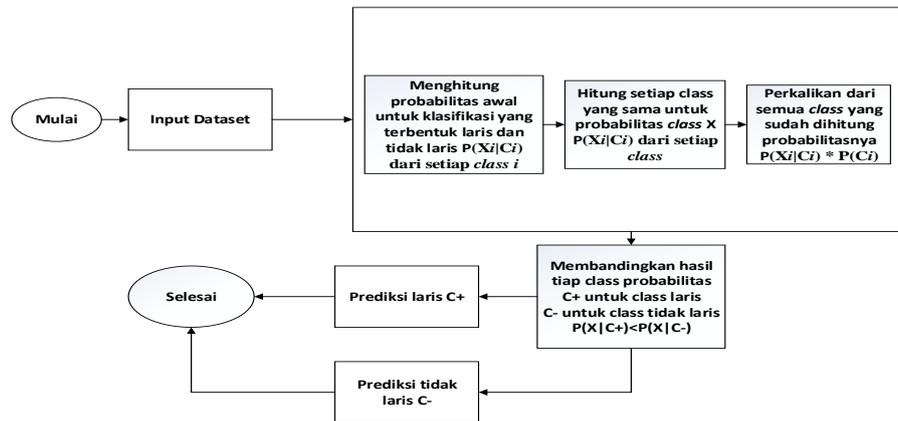
Proses seleksi data yaitu dengan cara mengelompokkan data yang berupa numerik menjadi kategorial, laris, cukup laris dan tidak laris. Contoh tabel klasifikasi data yang telah diimplementasikan pada Microsoft Excel dapat dilihat pada Gambar 3.

##### Model dan Metode yang Diusulkan

Metode yang digunakan dalam penelitian ini adalah algoritma *Naïve Bayes*. Dalam pemodelan ini algoritma *Naïve Bayes* dilakukan perhitungan untuk mendapatkan *performance value (accuracy)*.

### Hasil Pengujian Klasifikasi Klasifikasi Menggunakan Perhitungan Manual

Berikut ini perhitungan dalam penelitian ini menggunakan 6043 data training terdiri dari 6 atribut untuk menentukan sebuah class, yang mana dari 6043 data training tersebut digunakan untuk melakukan perhitungan algoritma *Naive Bayes*. Data training yang telah diimplementasikan pada Microsoft Excel dapat dilihat pada Gambar 4, dan data testing yang telah diimplementasikan pada Microsoft Excel dapat dilihat pada Gambar 5. Sehingga untuk Notasi Nama= “Sabun Lux”, Jenis produk= “Perawatan Diri”, caturwulan ke-1= “Tidak Laris”, caturwulan ke-2= “Cukup Laris”, caturwulan ke-3= “Laris”, caturwulan ke-4= “Laris”, caturwulan ke-5= “Tidak Laris” hasilnya “Laris”



Gambar 2. Model Algoritma *Naive Bayes*

	A	B	D	F	H	J	M
1	TAHUN 2019 - 2020						
2	PRODUK PERAWATAN DIRI	Caturwulan 1 (Jan-apr 19)	Caturwulan 2 (Mei-agt 19)	Caturwulan 3 (Sep-des 19)	Caturwulan 4 (Jan-apr 20)	Caturwulan 5 (Mei-agt)	KLASIFIKASI
4	ASETON HEVINY 35 ML	58	21	32	19	24	LARIS
5	ASETON TOKYO NIGHT 20ML	17	22	20	36	30	LARIS
6	ASETON TOKYO NIGHT 50ML	29	24	28	15	14	LARIS
7	ASETON TOKYO NIGHT PLSTIK 80ML	32	46	36	30	15	LARIS
8	AXE ANARCHY FOR HIM 150ML	33	69	60	50	18	LARIS
9	AXE APOLLO 150ML	4	10	2	3	2	TIDAK LARIS
10	AXE DEO BLACK BODYSPRAY 150ML	9	1	4	4	1	TIDAK LARIS
11	BAR SOAP FRESNATURAL RED 70GR	3	10	10	4	4	TIDAK LARIS
12	BAR SOAP GIV BENGKUANG	30	44	40	54	51	LARIS
13	BAR SOAP GIV BIRU	46	4	94	112	99	LARIS
14	BAR SOAP GIV HIJAU	45	4	55	57	54	LARIS
15	BAR SOAP GIV LEMON	41	15	72	77	60	LARIS

Gambar 3. Klasifikasi Data pada Produk Perawatan Diri pada Microsoft Excel

	A	B	C	D	E	F	G
1	PRODUK DATA UJI	JENIS PRODUK	Caturwulan 1 (Jan-apr 19)	Caturwulan 2 (Mei-agt 19)	Caturwulan 3 (Sep-des 19)	Caturwulan 4 (Jan-apr 20)	Caturwulan 5 (Mei-agt 20)
2	CASABLANCA 200ML VIOLET	Perawatan Diri	TIDAK LARIS				
3	CASABLANCA 65ML BIRU	Perawatan Diri	TIDAK LARIS				
4	CASABLANCA ABU 65ML	Perawatan Diri	TIDAK LARIS				
5	CASABLANCA BIRU 200ML	Perawatan Diri	TIDAK LARIS				
6	CASABLANCA BLACK 65ML	Perawatan Diri	TIDAK LARIS				
7	CASABLANCA BODY MIST 50ML	Perawatan Diri	TIDAK LARIS				
8	CASABLANCA BODY ORANGE 50 ML	Perawatan Diri	TIDAK LARIS				
9	CASABLANCA BODY SPRAY 65ML SI	Perawatan Diri	TIDAK LARIS				
10	CASABLANCA BODY SPRAY 65 ML ME	Perawatan Diri	TIDAK LARIS				
11	CASABLANCA BODY SPRAY HITAM 20	Perawatan Diri	TIDAK LARIS				
12	CASABLANCA BODYMIST 100ML (ABU	Perawatan Diri	TIDAK LARIS				
13	CASABLANCA BODYMIST 100ML HITA	Perawatan Diri	TIDAK LARIS				
14	CASABLANCA COKLAT 200ML	Perawatan Diri	TIDAK LARIS				
15	CASABLANCA COKLAT 65NL	Perawatan Diri	TIDAK LARIS				

Gambar 4. Data Training pada Microsoft Excel

	A	B	C	D	E	F	G
1	PRODUK DATA TESTING	JENIS PRODUK	Caturwulan 1 (Jan-apr 19)	Caturwulan 2 (Mei-agt 19)	Caturwulan 3 (Sep-des 19)	Caturwulan 4 (Jan-apr 20)	Caturwulan 5 (Mei-agt 20)
2	SABUN LUX	Perawatan Diri	Tidak Laris	Cukup Laris	Laris	Laris	Tidak Laris

Gambar 5. Data Testing pada Microsoft Excel

### Susunan Operator *Naïve Bayes*

Tahap pertama yang harus dilakukan adalah *drag and drop operator read excel* kemudian lakukan proses importing tabel Microsoft Excel kedalam tahap proses untuk membaca data training dalam format excel, kemudian ganti atribut produk data training menjadi id dan atribut klasifikasi menjadi label, yang dapat dilihat pada [Gambar 6](#). Selanjutnya *drag and drop operator validation (split validation)* kemudian hubungkan operator read excel dengan *operator validation (split validation)* setelah itu hubungkan pada result, yang dapat dilihat pada [Gambar 7](#), dengan set operator klasifikasi yang digunakan sebagai berikut: (a) Read Excel. Pada aplikasi Rapidminer fitur read excel berfungsi sebagai pembaca file excel, read excel merupakan operator yang paling dasar digunakan sebelum memulai sebuah proses. Operator ini dapat membaca file data dari excel 97, 2000, XP dan 2003. Dalam penelitian ini, file excel yang digunakan adalah .xlsx; (b) Validation (*Split validation*). Validation yang digunakan yaitu *split validation* dengan split ratio 70: 30 dan menggunakan sampling type linear sampling. Merupakan operator yang membagi dua dataset secara terpisah yaitu subproses training dan subproses testing. subproses training digunakan untuk melatih sebuah model algoritma (*Naïve Bayes*). Model yang telah dilatih (diuji atau diproses terlebih dahulu) kemudian diterapkan (*Apply Model*) dalam subproses testing. Kinerja model (*Accuracy*) diukur selama fase testing.

Pada [Gambar 8](#) operator fitur validasi yang digunakan sebagai berikut: (a) Modelling (*Naïve Bayes*). Operator modelling ini memproses set data yang berisi atribut kategorial pada operator read excel yang telah dimasukkan diatas. Modelling klasifikasi yang digunakan pada penelitian ini adalah *Naïve Bayes*; (b) Apply Model. Operator ini menguji set dari operator lain (*Naïve Bayes*) yang merupakan algoritma pembelajaran. Setelah itu, model ini dapat diterapkan pada data set lainnya. Biasanya bertujuan untuk mendapatkan prediksi pada data yang tidak terlihat atau untuk mengubah data dengan menerapkan model preprocessing; (c) Performance. Operator performance digunakan untuk mengevaluasi kinerja model (*Naïve Bayes*) yang memberikan daftar nilai kriteria kinerja secara otomatis sesuai tugas yang diberikan. Setiap operator diatas dihubungkan sesuai dengan kebutuhan. Pertama adalah *read excel* dihubungkan (*out*) dengan operator (*tra*) *validation*. Selanjutnya, *validation suboperator (tra)* dihubungkan dengan operator (*tra*) *Naïve Bayes*. *Naïve Bayes (mod)* dihubungkan dengan suboperator (*mod*) *training*. *Validation suboperator (mod)* terhubung dengan operator (*mod*) *Apply Model* dan *Validation suboperator (tes)* terhubung dengan operator (*unl*) *Apply Model*. *Apply Model (lab)* terhubung dengan operator (*lab*) *performance*. *Performance (per)* dihubungkan dengan *Performance (exa)*. Terakhir, *Validation (mod) (tra) (ave)* dihubungkan dengan (*res*) port akhir untuk mengeluarkan hasil klasifikasi *Naïve Bayes*, performance, dan dataset.

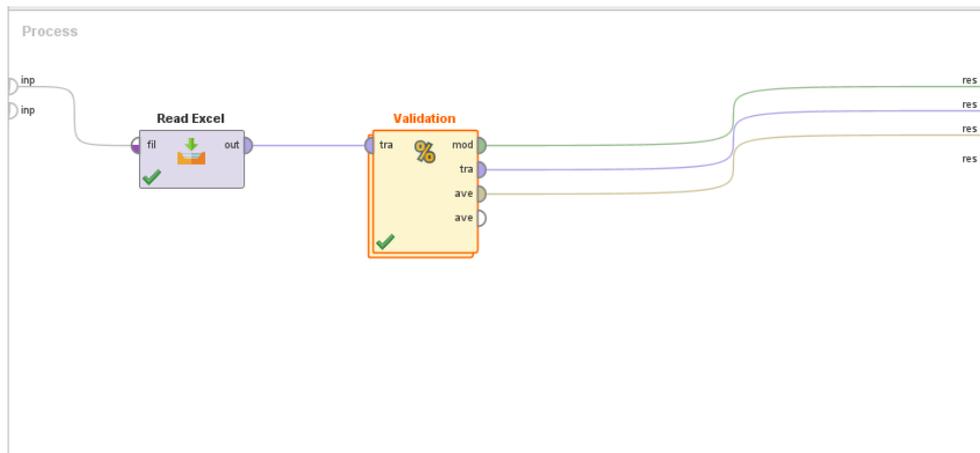
	Caturwulan... polynomial	Caturwulan... polynomial	Caturwulan... polynomial	Caturwulan... polynomial	Caturwulan... polynomial	Klasifikasi polynomial label
1	LARIS	LARIS	LARIS	LARIS	LARIS	LARIS
2	CUKUP LARIS	LARIS	LARIS	LARIS	LARIS	LARIS
3	LARIS	LARIS	LARIS	CUKUP LARIS	CUKUP LARIS	LARIS
4	LARIS	LARIS	LARIS	LARIS	CUKUP LARIS	LARIS
5	LARIS	LARIS	LARIS	LARIS	CUKUP LARIS	LARIS
6	TIDAK LARIS	CUKUP LARIS	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS
7	TIDAK LARIS					
8	TIDAK LARIS	CUKUP LARIS	CUKUP LARIS	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS
9	LARIS	LARIS	LARIS	LARIS	LARIS	LARIS
10	LARIS	TIDAK LARIS	LARIS	LARIS	LARIS	LARIS
11	LARIS	TIDAK LARIS	LARIS	LARIS	LARIS	LARIS
12	LARIS	CUKUP LARIS	LARIS	LARIS	LARIS	LARIS

Gambar 6. Importing Data Training pada Operator Read Excel

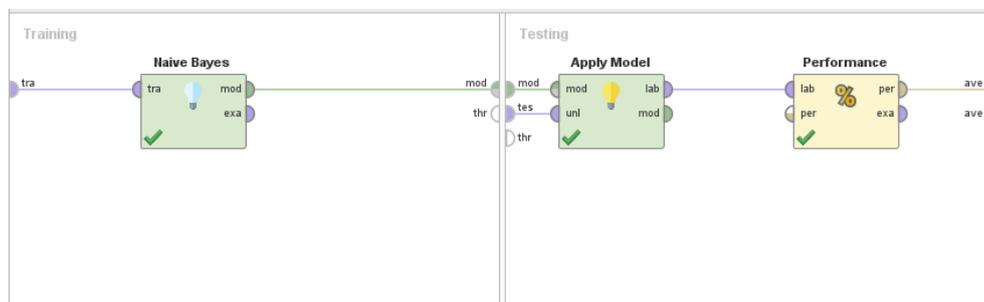
### Hasil Klasifikasi *Naïve Bayes*

Tahap ini adalah tahap terakhir dari proses *data mining* Rapidminer dimana setelah semua operator terhubung kemudian klik icon play tombol F11, maka muncul sebuah tab result, yang isinya sebuah klasifikasi dari pemodelan algoritma *Naïve Bayes* dan hasil accuracy dari performance sebesar 93,82% untuk proses yang dilaksanakan pada model diatas dapat dilihat pada [Gambar 9](#). Model Klasifikasi *Naïve Bayes* pada data transaksi penjualan barang terlaris dapat menghasilkan 2 class utama, yaitu class

LARIS dan *class* TIDAK LARIS. Untuk nilai *class* LARIS yaitu 0.251 dan nilai *class* TIDAK LARIS 0.749. Hal ini dapat dilihat pada Gambar 10.



Gambar 7. Set Operator Klasifikasi



Gambar 8. Validation (Split validation)

### Prediksi Menggunakan Rapid Miner

Uji coba ini dilakukan untuk mengetahui apakah perhitungan yang telah dilakukan diatas sesuai dengan klasifikasi penjualan barang terlaris dengan menggunakan algoritma metode *Naive Bayes* menggunakan Rapid Miner 9.0. Set Operator prediksi seperti yang ditampilkan pada Gambar 11, dapat digunakan sebagai berikut: (a) Read Excel. Pada aplikasi Rapidminer fitur read excel berfungsi sebagai pembaca file excel, read excel merupakan operator yang paling dasar digunakan sebelum memulai sebuah proses. Operator read excel yang pertama menggunakan data suervised (data yang sudah dilatih dan memiliki *class* label) sedangkan operator read excel yang kedua menggunakan data unsuervised sebagai data testing dan belum memiliki *class* label; (b) Validation (*Split validation*). Validation yang digunakan yaitu *split validation* dengan split ratio 70:30 dan menggunakan sampling type linear sampling. Merupakan operator yang membagi dua dataset secara terpisah yaitu subproses training dan subproses testing. subproses training digunakan untuk melatih sebuah model algoritma (*Naive Bayes*). Model yang telah dilatih (diuji atau diproses terlebih dahulu) kemudian diterapkan (*Apply Model*) dalam subproses testing. Kinerja model (*Accuracy*) diukur selama fase testing; (c) Apply Model. Operator ini menguji set dari operator lain (read excel (2)) yang merupakan data testing. Setelah itu, model ini dapat diterapkan pada data set lainnya. Biasanya bertujuan untuk mendapatkan prediksi pada data yang tidak terlihat atau untuk mengubah data dengan menerapkan model preprocessing.

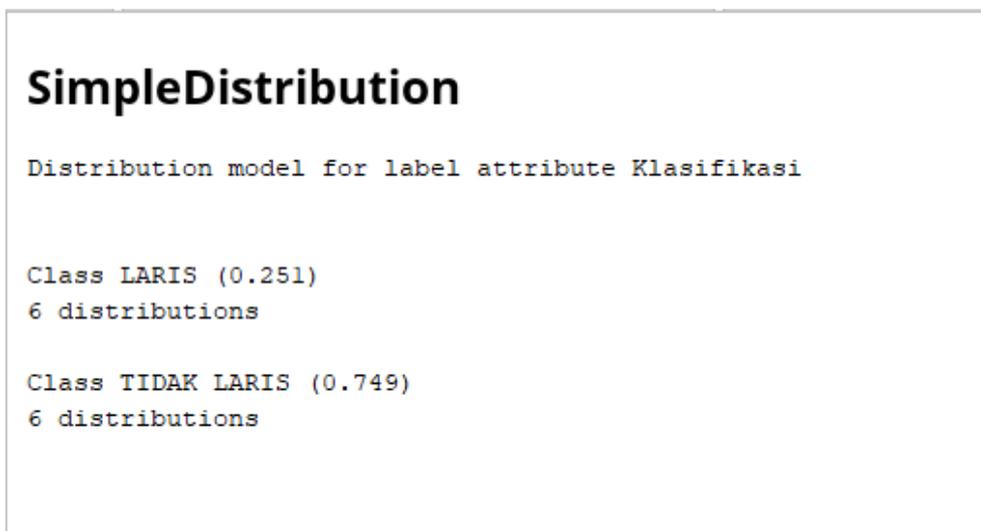
Setiap operator diatas dihubungkan sesuai dengan kebutuhan. Pertama, Read Excel (1) dihubungkan (out) dengan operator (tra) validation. Kemudian Validation suboperator (tra) dihubungkan dengan operator (tra) *Naive Bayes*. *Naive Bayes* (mod) dihubungkan dengan suboperator (mod) training. Validation suboperator (mod) terhubung dengan operator (mod) Apply Model dan Validation suboperator (tes) terhubung dengan operator (uni) Apply Model. Apply Model (lab) terhubung dengan operator (lab) performance. Performance (per) dihubungkan dengan Performance (exa). Read Excel (2) dihubungkan (out) dengan operator (uni) Apply Model (2). Apply Model (2) dihubungkan (mod) dengan operator (mod) validation. Terakhir, Validation (tra) (ave) dan Apply Model (2) (lab) dihubungkan dengan (res) port akhir untuk mengeluarkan hasil prediksi *Naive Bayes*, performance, dan dataset. Tahap selanjutnya adalah tahap terakhir dari proses *data mining* Rapidminer dimana setelah semua operator terhubung kemudian klik icon play tombol F11, maka muncul sebuah tab result, yang isinya sebuah prediksi dari

pemodelan algoritma *Naïve Bayes* untuk proses yang dilaksanakan pada model di atas dapat dilihat pada Gambar 12.

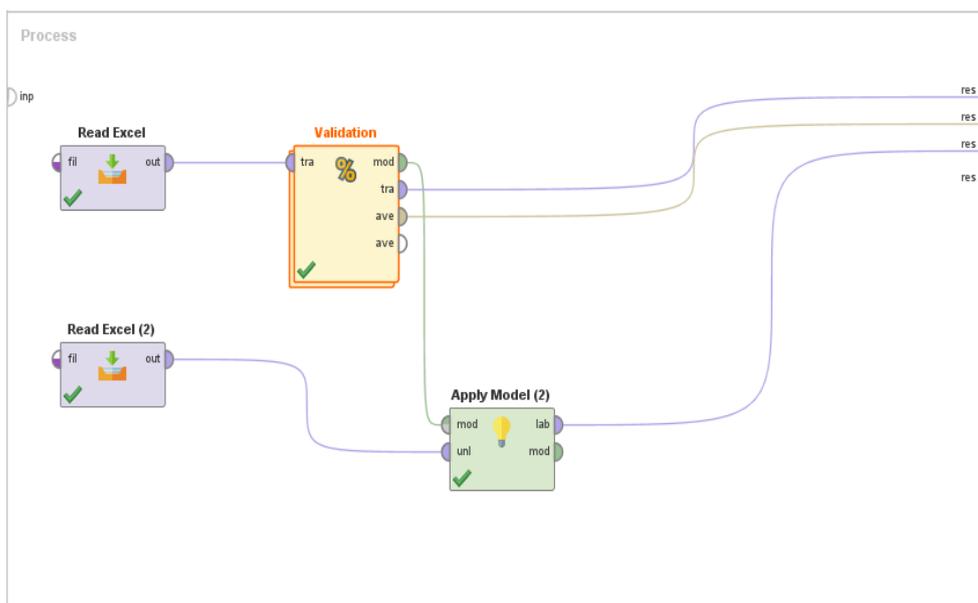
The screenshot shows the PerformanceVector (Performance) window in RapidMiner 9.0. The 'Criterion' is set to 'accuracy', and the 'Table View' is selected. The accuracy is 93.82%. A confusion matrix table is displayed below.

	true LARIS	true TIDAK LARIS	class precision
pred. LARIS	386	90	81.09%
pred. TIDAK LARIS	22	1315	98.35%
class recall	94.61%	93.59%	

Gambar 9. Hasil Perhitungan Nilai Akurasi Menggunakan RapidMiner 9.0



Gambar 10. Model Klasifikasi *Naïve Bayes* Menggunakan RapidMiner 9.0



Gambar 11. Set Operator Prediksi Menggunakan RapidMiner 9.0

PRODUK DA...	prediction(K...	confidence(...	confidence(...	JENIS PROD...	Caturwulan ...	Caturwulan ...	Caturwulan ...	Caturwulan ...	Cat
DOVE SHP D...	TIDAK LARIS	0.001	0.999	Perawatan Diri	CUKUP LARIS	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TID
DOVE SHP D...	TIDAK LARIS	0.252	0.748	Perawatan Diri	TIDAK LARIS	CUKUP LARIS	TIDAK LARIS	CUKUP LARIS	LAF
DOVE SHP D...	TIDAK LARIS	0.000	1.000	Perawatan Diri	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TID
DOVE SHP D...	TIDAK LARIS	0.054	0.946	Perawatan Diri	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	CUKUP LARIS	LAF
DOVE SHP T...	LARIS	0.500	0.500	Perawatan Diri	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	LARIS	LAF
DOVE SHP T...	LARIS	0.854	0.146	Perawatan Diri	TIDAK LARIS	CUKUP LARIS	TIDAK LARIS	LARIS	LAF
SABUN LUX	LARIS	0.996	0.004	Perawatan Diri	Tidak Laris	Cukup Laris	Laris	Laris	Tid
DETTOL SKIN...	TIDAK LARIS	0.000	1.000	Perawatan Diri	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TID
DETTOL B.W...	TIDAK LARIS	0.000	1.000	Perawatan Diri	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TID
DETTOL B.W...	TIDAK LARIS	0.000	1.000	Perawatan Diri	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TID
DETTOL BO...	TIDAK LARIS	0.000	1.000	Perawatan Diri	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TID
DETTOL BO...	TIDAK LARIS	0.000	1.000	Perawatan Diri	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TID
DETTOL BO...	TIDAK LARIS	0.000	1.000	Perawatan Diri	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TID
DETTOL BO...	TIDAK LARIS	0.000	1.000	Perawatan Diri	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TIDAK LARIS	TID

Gambar 12. Hasil Prediksi Naïve Bayes Menggunakan RapidMiner 9.0

Sehingga untuk notasi, Nama = “Sabun Lux”, Jenis produk= “Perawatan Diri”, caturwulan ke-1= “Tidak Laris”, caturwulan ke-2= “Cukup Laris”, caturwulan ke-3= “Laris”, caturwulan ke-4= “Laris”, caturwulan ke-5= “Tidak Laris” yang di lakukan dengan menggunakan RapidMiner menghasilkan prediksi class/label “Laris”. Hal ini menunjukkan hasil yang sama pada perhitungan manual yang telah dilakukan sebelumnya.

**Evaluasi**

Tahapan evaluasi yang dilakukan dalam penelitian ini adalah untuk memberikan penilaian dari hasil penggunaan algoritma Naïve Bayes untuk mengklasifikasi penjualan barang terlaris menggunakan split validation. Bagian yang dievaluasi adalah presentase data, jumlah data training dan nilai akurasi yang di hasilkan.

Performance			
Criterion	accuracy		
accuracy	92.39%		
	true LARIS	true TIDAK LARIS	class precision
pred. LARIS	578	159	78.43%
pred. TIDAK LARIS	25	1655	98.51%
class recall	95.85%	91.23%	

Gambar 13. Split validation 0,6 (60%:40%) Menggunakan RapidMiner 9.0

Performance			
Criterion	accuracy		
accuracy	93.82%		
	true LARIS	true TIDAK LARIS	class precision
pred. LARIS	386	90	81.09%
pred. TIDAK LARIS	22	1315	98.35%
class recall	94.61%	93.59%	

Gambar 14. Split validation 0,7 (70%:30%) Menggunakan RapidMiner 9.0

	true LARIS	true TIDAK LARIS	class precision
pred. LARIS	280	76	78.65%
pred. TIDAK LARIS	8	845	99.06%
class recall	97.22%	91.75%	

Gambar 15. Split validation 0,8 (80%:20%) Menggunakan RapidMiner 9.0

Split validation 0,6 merupakan pembagian data secara sederhana dimana sebanyak 60% data digunakan untuk data training dan 40% data digunakan untuk data testing dan menghasilkan accuracy sebesar 92,39%. Split validation 0,6 dapat dilihat pada Gambar 13. Split validation 0,7 merupakan pembagian data secara sederhana dimana sebanyak 70% data digunakan untuk data training dan 30% data digunakan untuk data testing dan menghasilkan accuracy sebesar 93,82%. Split validation 0,7 dapat dilihat pada Gambar 14. Split validation 0,8 merupakan pembagian data secara sederhana dimana sebanyak 80% data digunakan untuk data training dan 20% data digunakan untuk data testing dan menghasilkan accuracy sebesar 93,05%. Split validation 0,6 dapat dilihat pada Gambar 15.

**Pembahasan**

Penelitian ini telah berhasil mengimplementasikan metode Naive Bayes pada data mining untuk mengklasifikasikan penjualan barang terlaris pada perusahaan ritel dengan menggunakan data dari UD. Mawar Sari. Pendekatan kualitatif dapat digunakan dengan baik pada jenis penelitian studi kasus ini. Model CRISP-DM sangat tepat digunakan pada implementasi data mining dengan studi kasus yang sedang diangkat pada penelitian ini (Bimantara & Safii, 2021; Syahril et al., 2020). Proses pada data preprocessing pada penelitian ini sangat penting dalam memudahkan perhitungan manual ataupun menggunakan RapidMiner (Romadhan, 2018; Sudarsono et al., 2021). Data yang tidak lengkap, noise yang terdapat pada data, duplikasi data, data yang tidak konsisten dapat diperbaiki dalam tahap ini. Kondisi preprocessing yang dilakukan pada tahap data preparation dalam CRISP-DM membantu dalam meningkatkan presisi dan kinerja data mining dan mencegah terjadinya kesalahan pada hasil akhir perhitungan. Penggunaan RapidMiner juga membantu peneliti dalam perhitungan data mining yang lebih cepat dan akurat. Penggunaan RapidMiner juga termasuk mudah dengan tampilan penggunaan yang baik. Penggunaan RapidMiner ketika dibandingkan dengan perhitungan manual juga memiliki hasil yang sama, sehingga perhitungan manual dan RapidMiner memiliki kualitas yang sama.

Tahapan evaluasi yang dilakukan dalam penelitian ini adalah untuk memberikan penilaian dari hasil penggunaan algoritma Naive Bayes untuk mengklasifikasi penjualan barang terlaris menggunakan split validation. Bagian yang akan dievaluasi adalah presentase data, jumlah data training dan nilai akurasi yang di hasilkan. Pada pembagian data secara sederhana dimana sebanyak 60% data digunakan untuk data training dan 40% data digunakan untuk data testing dan menghasilkan accuracy sebesar 92,39%. Pada pembagian data secara sederhana dimana sebanyak 70% data digunakan untuk data training dan 30% data digunakan untuk data testing dan menghasilkan accuracy sebesar 93,82%. Pada pembagian data secara sederhana dimana sebanyak 80% data digunakan untuk data training dan 20% data digunakan untuk data testing dan menghasilkan accuracy sebesar 93,05%. Secara rata-rata hasil accuracy pada penelitian ini mencapai 93,08% dalam kategori yang sangat baik.

Pada penelitian sebelumnya yang dilakukan mengangkat topik tentang klasifikasi komentar spam pada youtube, menunjukkan bahwa Naive Bayes menempati peringkat yang baik dalam kondisi akurasi yang berada diatas 85%, namun masih berada di bawah 90% (Burhanudin et al., 2018; Thoib et al., 2018). Pada penelitian ini akurasi menggunakan Naive Bayes dapat melampaui akurasi tersebut karena akurasinya diatas 90%. Penelitian sebelumnya yang mengangkat topik terkait kualitas pelayanan maskapai penerbangan domestik, menunjukkan pemeringkatan untuk masing-masing alternatif (Indriyani et al., 2020; Mansyur et al., 2021). Namun pada penelitian tersebut, belum memberikan hasil akurasi terhadap hasil perhitungan menggunakan Naive Bayes. Berdasarkan penggunaan metode Naive Bayes dalam penelitian ini menunjukkan bahwa jumlah produk yang

konsisten laris berdasarkan kategori food sebanyak 243 jenis produk yang laris dan non-food sebanyak 215 jenis produk yang laris disetiap caturwulannya. Pihak manajemen UD. Mawar Sari dapat menggunakan hasil dari penelitian ini untuk proses manajemen stok yang dimilikinya. Diharapkan dengan implementasi dari hasil penelitian ini, UD. Mawar Sari akan mendapatkan keuntungan yang lebih baik dari sebelumnya. Pada penelitian selanjutnya, dapat menggunakan data pada UD. Mawar Sari dengan tujuan klasifikasi barang terlaris dengan metode berbeda ataupun dengan software pembantu berbeda untuk dapat memberikan sudut pandang yang baru, sehingga mampu memberikan tambahan saran bagi manajemen UD. Mawar Sari.

#### 4. SIMPULAN

Penelitian ini yang mengangkat topik mengenai Implementasi Naïve Bayes pada Data Mining untuk Mengklasifikasikan Penjualan Barang Terlaris pada Perusahaan Ritel ini, menunjukkan bahwa ini dengan penggunaan model algoritma Naïve Bayes yang dilakukan pada data transaksi penjualan didapatkan beberapa hasil accuracy berdasarkan pembagian data training. Hasil dari data testing (split validation) menunjukkan hasil yang baik dengan akurasi yang tinggi. Hasil penelitian ini dapat digunakan dan berimplikasi baik untuk manajemen UD. Mawar Sari dalam pengambilan keputusan untuk menentukan suatu produk setelah mendapatkan hasil perhitungan data mining menggunakan metode Naïve Bayes.

#### 5. DAFTAR PUSTAKA

- Al Zukri, P., Widyaningrum, S. N., & Aini, Q. (2020). Forecasting permintaan pompa air dangkal shimizu menggunakan metode time series. *Sistemasi: Jurnal Sistem Informasi*, 9(2), 226–234. <https://doi.org/10.32520/stmsi.v9i2.694>.
- Apriliah, W., Kurniawan, I., Baydhowi, M., & Haryati, T. (2021). Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest. *Sistemasi: Jurnal Sistem Informasi*, 10(1), 163–171. <https://doi.org/10.32520/stmsi.v10i1.1129>.
- Apriyani, P., Dikananda, A. R., & Ali, I. (2023). Penerapan Algoritma K-Means dalam Klusterisasi Kasus Stunting Balita Desa Tegalwangi. *Hello World Jurnal Ilmu Komputer*, 2(1), 20–33. <https://doi.org/10.56211/helloworld.v2i1.230>.
- Aulia, W. (2018). Sistem Pakar Diagnosa Penyakit Jantung Koroner Dengan Metode Probabilistic Fuzzy Decision Tree. *Jurnal Sains Dan Informatika*, 4(2), 106–106. <https://doi.org/10.22216/jsi.v4i2.3258>.
- Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 640–651. <https://doi.org/10.30865/mib.v5i2.2937>.
- Bimantara, B., & Safii, M. (2021). Penerapan Data Mining Menentukan Kelayakan Penjualan Kendaraan Bekas Roda Dua Dengan Menggunakan Metode Bayesian Classifier. In *Seminar Nasional Informatika (SENATIKA)* (pp. 67–77).
- Burhanudin, M., Y., & Wihardi, Y. (2018). Klasifikasi Komentar Spam Pada Youtube Menggunakan Metode Naïve Bayes, Support Vector Machine, Dan K-Nearest Neighbors. *Jurnal Informatika Dan Komputer*, 3(2), 54–59. <https://doi.org/10.26798/jiko.v3i2.139>.
- Cholil, W., Dalimunthi, A. R., & Atika, L. (2019). Model Data Mining Dalam Mengidentifikasi Pola Laju Pertumbuhan Antar Sektor Ekonomi di Provinsi Sumatera Selatan dan Bangka Belitung. *Teknika*, 8(2), 103–109. <https://doi.org/10.34148/teknika.v8i2.181>.
- Darmansah, D. D., & Wardani, N. W. (2021). Analisis Pesebaran Penularan Virus Corona di Provinsi Jawa Tengah Menggunakan Metode K-Means Clustering. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(1), 105–117. <https://doi.org/10.35957/jatisi.v8i1.590>.
- Darmansah, D., & Wardani, N. W. (2020). Analisa Penyebab Kerusakan Tanaman Cabai Menggunakan Metode K-Means. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 7(2), 126–134. <https://doi.org/10.35957/jatisi.v7i2.309>.
- Dewanti, M., Muchbarak, A., & Widiyatun, F. (2021). Sistem Pakar Penentuan Menu Diet Bagi Penderita Diabetes Mellitus. *JRKT (Jurnal Rekayasa Komputasi Terapan)*, 1(2). <https://doi.org/10.30998/jrkt.v1i02.4092>.
- Exenberger, E., & Bucko, J. (2020). Analysis of Online Consumer Behavior—Design of CRISP-DM Process Model. *Agris On-Line Papers in Economics and Informatics*, 12(3), 13–22. <https://doi.org/10.7160/aol.2020.120302>.
- Faisal, S., & Tutz, G. (2021). Multiple imputation using nearest neighbor methods. *Information Sciences*,

- 570, 500–516. <https://doi.org/10.1016/j.ins.2021.04.009>.
- Feblian, D., & Daihani, D. U. (2016). Implementasi Model CRISP-DM Untuk Menentukan Sales Pipeline pada PT X. *Jurnal Teknik Industri*, 6(1), 1–12. <https://e-journal.trisakti.ac.id/index.php/tekin/article/view/1526/1316>.
- Fibrianda, M. F., & Bhawiyuga, A. (2018). Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode Naïve Bayes Dan Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(9), 3112–3123. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2559>.
- Firmansyah, F., & Yulianto, A. (2021). Machine Learning Dengan Decision Tree untuk Prediksi Pembayaran Invoice, Case Study: Gramedia Jakarta. *Journal Of Informatics And Telecommunication Engineering*, 5(1), 167–175. <https://doi.org/10.31289/jite.v5i1.5066>.
- Ginantra, N. L. W. S. R., & Wardani, N. W. (2019a). Implementasi Metoda Naïve Bayes dan Vector Space Model dalam Deteksi Kesamaan Artikel Jurnal Berbahasa Indonesia. *Jurnal Infomedia*, 4(2), 94–100. <https://doi.org/10.30811/jim.v4i2.1530>.
- Ginantra, N. L. W. S. R., & Wardani, N. W. (2019b). Measurement of the Similarity of Indonesian Papers on One Journal Topic with the Naive Bayes Algorithm and Vector Space Model. *International Journal of Computer, Network Security and Information System (IJCONSIST)*, 1(1), 20–26. <https://doi.org/10.33005/ijconsist.v1i1.7>.
- Gunawan, R., Nugroho, N. B., & Arbianto, R. (2018). Penerapan Data Mining Untuk Estimasi Laju Pertumbuhan Produk Domestik Regional Bruto (PDRB) Perkapita Atas Dasar Harga Berlaku Menurut Lapangan Usaha Pada Kota Medan Menggunakan Metode Regresi Linier Berganda. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, 1(2), 106–113. <https://doi.org/10.53513/jsk.v1i2.36>.
- Gupta, M. K., & Chandra, P. (2020). A comprehensive survey of data mining. *International Journal of Information Technology*, 1(1), 1–15. <https://doi.org/10.1007/s41870-020-00427-7>.
- Indriyani, N., Ali, E., Rio, U., & Rahmadden, R. (2020). Menentukan Kualitas Pelayanan Maskapai Penerbangan Domestik Dengan Menggunakan Metode Naïve Bayes: SATIN. *Sains Dan Teknologi Informasi*, 6(1), 37–44. <https://doi.org/10.33372/stn.v6i1.605>.
- Irawan, A., Rohaniah, R., Sulistiani, H., & Priandika, A. T. (2019). Sistem Pendukung Keputusan Untuk Pemilihan Tempat Servis Komputer di Kota Bandar Lampung Menggunakan Metode AHP. *Jurnal Tekno Kompak*, 13(1), 30–35. <https://doi.org/10.33365/jtk.v13i1.267>.
- Islamiyati, D. S., & Fikri, A. (2022). Penerapan Algoritma Knuth-Morris-Pratt dalam Mendeteksi Tingkat Kemiripan Judul Skripsi Berbasis Web. *Journal of Information System Research (JOSH)*, 3(2), 58–63. <https://doi.org/10.47065/josh.v3i2.1168>.
- Kant, S., & Ansari, I. A. (2016). An improved K means clustering with Atkinson index to classify liver patient dataset. *International Journal of System Assurance Engineering and Management*, 7, 222–228. <https://doi.org/10.1007/s13198-015-0365-3>.
- Kück, M., & Freitag, M. (2021). Forecasting of customer demands for production planning by local k-nearest neighbor models. *International Journal of Production Economics*, 231, 107837. <https://doi.org/10.1016/j.ijpe.2020.107837>.
- Mahendra, G. S., & Aryanto, K. Y. E. (2019). SPK Penentuan Lokasi ATM Menggunakan Metode AHP dan SAW. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 49–56. <https://doi.org/10.25077/TEKNOSI.v5i1.2019.49-56>.
- Mahendra, G. S., & Indrawan, I. P. Y. (2020). Metode AHP-TOPSIS Pada Sistem Pendukung Keputusan Penentuan Penempatan Atm. *JST (Jurnal Sains Dan Teknologi)*, 9(2), 130–142. <https://doi.org/10.23887/jst-undiksha.v9i2.24592>.
- Mahendra, G. S., Karsana, I. W. W., & Paramitha, A. A. I. I. (2021). DSS for best e-commerce selection using AHP-WASPAS and AHP-MOORA methods. *Matrix: Jurnal Manajemen Teknologi Dan Informatika*, 11(2), 81–94. <https://doi.org/10.31940/matrix.v11i2.2306>.
- Mahendra, G. S., & Nugraha, P. G. S. C. (2020). Komparasi Metode AHP-SAW dan AHP-WP Pada SPK Penentuan E-Commerce Terbaik di Indonesia. *JUSTIN (Jurnal Sistem Dan Teknologi Informasi)*, 8(4), 346–356. <https://doi.org/10.26418/justin.v8i4.42611>.
- Mansyur, T. P., Said, L. B., & Syaifei, I. (2021). Analisis Kualitas Layanan Lanside Terhadap Kepuasan Pengguna Jasa pada Bandar Udara Internasional Sultan Hasanuddin Makassar. *Jurnal Teknik Sipil MACCA*, 6(2), 159–165. <https://doi.org/10.33096/jtms.v6i2.341>.
- Nugraha, P. G. S. C., & Mahendra, G. S. (2020). ). Explorasi Algoritma C4.5 dan Forward Feature Selection untuk Menentukan Debitur Baik dan Debitur Bermasalah pada Produk Kredit Tanpa Agunan (KTA). *JST (Jurnal Sains Dan Teknologi)*, 9(1), 39–46. <https://doi.org/10.23887/jst-undiksha.v9i1.24627>.

- Rahmat, C. A., Permatasari, H., Rasywir, E., & Pratama, Y. (2023). Penerapan K-Means Untuk Clustering Kondisi Gizi Balita Pada Posyandu. *Jurnal Media Informatika Budidarma*, 7(1), 207–213. <https://doi.org/10.30865/mib.v7i1.5142>.
- Rahmi, A. N., & Mikola, Y. A. (2021). Implementasi Algoritma Apriori Untuk Menentukan Pola Pembelian Pada Customer (Studi Kasus: Toko Bakoel Sembako). *Information System Journal*, 4(1), 14–19. <https://doi.org/10.24076/infosjournal.2021v4i1.561>.
- Rifky, L., Nugraha, Z., Pratama, D., & Raswir, E. (2022). Implementasi Data Mining Untuk Penjualan Mobil Menggunakan Metode Naive Bayes. *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, 2(2), 225–230. <https://doi.org/10.33998/jakakom.2022.2.2.109>.
- Rivai, M. A. (2020). Analysis of Corona Virus spread uses the CRISP-DM as a Framework: Predictive Modelling. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(3), 2987–2994. <https://doi.org/10.30534/ijatcse/2020/76932020>.
- Romadhan, M. I. (2018). Personal Branding Jokowi dalam Mempertahankan Brand Image Melalui Video Blog Youtube. *MetaCommunication; Journal of Communication Studies*, 3(2). <https://doi.org/10.20527/mc.v3i2.5446>.
- Rukmana, I., Rasheda, A., Fathulhuda, F., & Cahyadi, M. R. (2021). Analisis Perbandingan Kinerja Algoritma Naive Bayes, Decision Tree- J48 dan Lazy-IBK. *Jurnal Media Informatika Budidarma*, 5(3), 1038–1044. <https://doi.org/10.30865/mib.v5i3.3055>.
- Saputra, M. I. H., & Nugraha, N. (2021). Sistem Pendukung Keputusan Dengan Metode Analytical Hierarchy Process (Ahp)(Studi Kasus: Penentuan Internet Service Provider Di Lingkungan Jaringan Rumah). *Jurnal Ilmiah Teknologi Dan Rekayasa*, 25(3), 199–212. <https://doi.org/10.35760/tr.2020.v25i3.3422>.
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 181, 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>.
- Sisodia, D., & Sisodia, D. S. (2018). Prediction of diabetes using classification algorithms. *Procedia Computer Science*, 132, 1578–1585. <https://doi.org/10.1016/j.procs.2018.05.122>.
- Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. *JBASE - Journal of Business and Audit Information Systems*, 4(1). <https://doi.org/10.30813/jbase.v4i1.2729>.
- Syahril, M., Erwansyah, K., & Yetri, M. (2020). Penerapan Data Mining untuk menentukan pola penjualan peralatan sekolah pada brand wigglo dengan menggunakan algoritma apriori. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, 3(1), 118–136. <https://doi.org/10.53513/jsk.v3i1.202>.
- Thoib, I., Setyanto, A., & Raharjo, S. (2018). Pengaruh Normalisasi Teks Dengan Text Expansion Dalam Deteksi Komentar Spam Pada Youtube. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 2(3), 708–715. <https://doi.org/10.29207/resti.v2i3.602>.
- Vora, P., & Oza, B. (2013). A survey on k-mean clustering and particle swarm optimization. *International Journal of Science and Modern Engineering*, 1(3), 24–26. <https://www.ijisme.org/wp-content/uploads/papers/v1i3/C0150020213.pdf>.
- Wantoro, A. (2020). Kombinasi Metode Analitical Hierarchy Process (Ahp) Dan Simple Addtive Weight (Saw) Untuk Menentukan Website E-Commerce Terbaik. *Sistemasi: Jurnal Sistem Informasi*, 9(1), 131–142. <https://doi.org/10.32520/stmsi.v9i1.608>.
- Wardani, N. W. (2020). *Penerapan Data Mining dalam Analytic CRM*. Yayasan Kita Menulis.
- Wardani, N. W., & Ariasih, N. K. (2019). Analisa Komparasi Algoritma Decision Tree C4.5 dan Naive Bayes untuk Prediksi Churn Berdasarkan Kelas Pelanggan Retail. *International Journal of Natural Sciences and Engineering*, 3(3), 103–112. <https://doi.org/10.23887/ijnse.v3i3.23113>.
- Wardani, N. W., Dantes, G. R., & Indrawan, G. (2018). Prediksi Customer Churn dengan Algoritma Decision Tree C4.5 Berdasarkan Segmentasi Pelanggan pada Perusahaan Retail. *Jurnal Resistor*, 1(1), 16–24. <https://doi.org/10.31598/jurnalresistor.v1i1.219>.
- Wulandari, S., Kridalaksana, A. H., & Khairina, D. M. (2019). Sistem Pakar Penerapan Menu Gizi Pada Penderita Jantung Koroner Menggunakan Metode Teorema Bayes. *Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer*, 14(2), 1–7. <https://doi.org/10.30872/jim.v15i1.2716>.
- Yuniastari, N. L. A. K., & Wirawan, I. W. W. (2014). Peramalan Permintaan Produk Perak Menggunakan Metode Simple Moving Average Dan Exponential Smoothing. *Jurnal Sistem Dan Informatika (JSI)*, 9(1), 97–106. <https://mail.jsi.stikom-bali.ac.id/index.php/jsi/article/view/41/62>.