

# IMPLEMENTASI *DATA MINING* UNTUK PREDIKSI PENJUALAN PESTISIDA PADA CV MITRA ARTHA SEJATI MENGGUNAKAN ALGORITMA *NAIVE BAYES*

Anugrah Angga Ronaldi<sup>1</sup>, Nanang Hunafi<sup>2</sup>

<sup>1</sup>Universitas Adhirajasa Reswana Sanjaya  
Jl. Sekolah International No.1-6 Antapani, Bandung, 022-7100124  
Email : [Anggaronaldi@gmail.com](mailto:Anggaronaldi@gmail.com)

<sup>2</sup>Universitas Adhirajasa Reswana Sanjaya  
Jl. Sekolah International No.1-6 Antapani, Bandung, 022-7100124  
Email : [nanang@ars.ac.id](mailto:nanang@ars.ac.id)

## Abstrak

Salah satu faktor penting untuk terciptanya perusahaan yang sehat dan berkembang adalah dengan adanya penjualan. Prediksi adalah salah satu kunci dari keberhasilan penjualan karena dengan nilai prediksi penjualan bisa dijadikan panduan sebagai acuan menentukan suatu penjualan produk. Salah satu cara untuk membuat prediksi penjualan yang agar lebih baik adalah dengan cara memanfaatkan pengolahan data mining menggunakan algoritma *Naive Bayes*. Algoritma *Naive Bayes* menghitung nilai probabilitas dari masing-masing atribut yang di teliti. Dari data penjualan CV Mitra Artha Sejati pada tahun 2017, 2018, 2019, dan 2020. Hasil perhitungan prediksi menggunakan algoritma *Naive Bayes* menghasilkan tingkat akurasi prediksi mencapai 94,59% dengan *class precision* yaitu "YA" 100.00%, "TIDAK" 94.44%, dan untuk *class recall* yaitu "YA" 33.33%, dan "TIDAK" 100.00%.

**Kata Kunci:** Prediksi, Penjualan, *Data Mining*, Algoritma, *Naive Bayes*.

## Abstract

*One of the important factors for the creation of a healthy and growing company is sales. Prediction is one of the keys to sales success because the predictive value of sales can be used as a guide as a reference for determining a product sale. One way to make better sales predictions is by utilizing data mining processing using the Naive Bayes algorithm. The Naive Bayes algorithm calculates the probability value of each of the attributes studied. From the sales data of CV Mitra Artha Sejati in 2017, 2018, 2019, and 2020. The results of prediction calculations using the Naive Bayes algorithm resulted in a prediction accuracy level of 94.59% with class precision, namely "YES" 100.00%, "NO" 94.44%, and for class recall, namely "YES" 33.33%, and "NO" 100.00%.*

**Keywords:** Prediction, Sales, *Data Mining*, Algorithm, *Naive Bayes*.

## 1. Pendahuluan

CV Mitra Artha Sejati adalah salah satu perusahaan yang bergerak dalam bidang penjualan pestisida yang sudah berdiri dari tahun 2012. Tingginya ketergantungan dan pemakaian pestisida didalam pertanian menjadikan pangsa pasar bagi para pengusaha atau perusahaan untuk berlomba dalam menciptakan dan memasarkan produk pestisida mengingat penggunaan yang sangat tinggi dan wilayah Indonesia sebagai salah satu negara agraris yang memiliki wilayah cukup luas. Namun dengan demikian akhirnya banyak bermunculan perusahaan-perusahaan baru yang mengakibatkan ketatnya persaingan dilapangan. Banyaknya jenis merk dagang yang beredar, ditambah banyaknya produsen juga cuaca yang tidak stabil menjadi masalah baru bagi perusahaan baik mengatur strategi pemasaran, strategi produk, ataupun dalam masalah pengadaan barang.

Namun demikian berdasarkan hasil studi pendahuluan yang dilakukan oleh peneliti ternyata dalam proses penjualan pestisida CV Mitra Artha Sejati masih terdapat berbagai masalah, yaitu yang berkaitan dengan penentuan jenis dan produksi. Hal ini berhubungan dengan peningkatan omset penjualan agar dapat mencapai target. Walaupun tidak terlihat signifikan akan tetapi hasil yang tidak sesuai target penjualan bila diakumulasikan akan menimbulkan kerugian pada perusahaan.

Atas dasar permasalahan diatas, maka penulis tertarik untuk mencoba membuat alternatif jalan keluar dalam sistem yang dapat membantu untuk mengatasi masalah tersebut. Alternatif solusi yang ditawarkan penulis dengan membuat sistem prediksi (*forecast*) penjualan produk CV Mitra Artha Sejati.

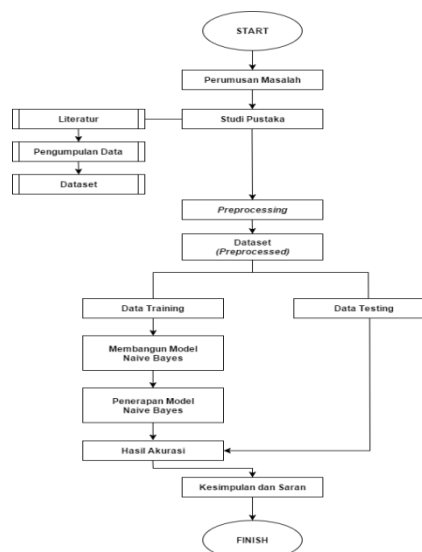
Beberapa peneliti yang melakukan riset mengenai prediksi penjualan diantaranya yaitu terdapat pada jurnal (Tiara et al., 2018) yang berjudul "Prediksi Penjualan Menggunakan Algoritma *Neural Network*" yang memiliki hasil akurasi prediksi mencapai 90,8% yang sebelumnya akurasi prediksi 81,75% sehingga dapat meningkatkan akurasi sebesar 9,05%.

Berdasarkan penelitian sebelumnya maka akan dilakukan penelitian selanjutnya dengan kasus yang sama namun dengan metode yang berbeda. Metode yang

digunakan adalah algoritma *Naïve Bayes*, diharapkan dengan penerapan metode *Naïve Bayes* merupakan metode yang bersifat sederhana dan efektif dalam pengolahan data yang besar serta menghasilkan sistem yang memiliki tingkat akurasi yang tinggi dengan nilai kebenarannya (Chandrasekar & Qian,2016). Metode *Naïve Bayes* adalah sebuah pengklasifikasian *probabilistic* sederhana yang menghitung sekumapulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari *dataset* yang diberikan dan mengansumsikan semua attribut independen atau tidak saling ketergantungan yang diberikan nilai pada variabel kelas (Manalu et al., 2017).

## 2. Metode Penelitian

Untuk melakukan sebuah penelitian, diperlukan adanya tahapan-tahapan yang tersusun dengan baik dan sistematis agar pelaksanaan penelitian tepat mencapai tujuan yang diharapkan. Maka dari itu disusunlah tahapan-tahapan metodologi dalam penelitian ini yang akan dijelaskan pada gambar dibawah ini.



Gambar 2.1 Tahap Metodologi Penelitian

Tahapan metodologi penelitian akan dijelaskan secara rinci pada penulisan dibawah ini :

### 2.1. Perumusan Masalah

Tahap ini merupakan awal dari tahapan metodologi penelitian dengan mencari dan mempelajari tentang permasalahan dalam penelitian di CV Mitra Artha Sejati dengan melakukan riset langsung. Kemudian dilanjutkan dengan

mencari solusi dalam mengatasi permasalahan tersebut. Didalam tahapan ini juga akan ditentukan ruang lingkup dan latar belakang dari topik penelitian.

## 2.2. Studi Pustaka

Tahap ini merupakan lanjutan dari tahapan sebelumnya. Di mana pada tahapan ini, dilakukan pengumpulan literatur dan data yang berhubungan dengan penelitian melalui berbagai media. Studi pustaka dilakukan untuk mendapatkan teori-teori tentang penelitian melalui jurnal, media online, buku atau penelitian lain yang sebelumnya berhubungan dengan penelitian yang akan dilakukan. Pengumpulan data dilakukan langsung ke perusahaan terkait di CV Mitra Artha Sejati yang beralamatkan di Komplek Griya Cikutra Blok E-1 Bojong Koneng Bandung, dan data yang di dapat berupa data mentah penjualan produk pestisida selama 3 tahun kebelakang dari tahun 2017, 2018, 2019 dan 2020 yang harus melalui tahap seleksi dan *preprocessing* sebelum digunakan dalam penelitian.

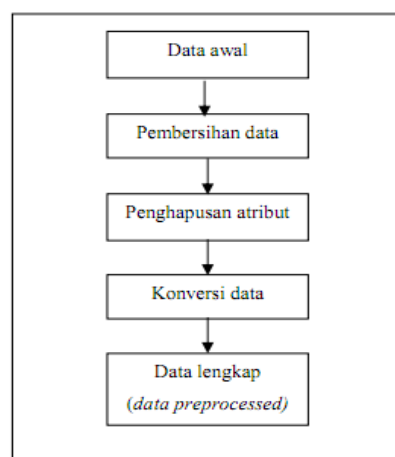
## 2.3. Implementasi Metode *Naive Bayes*

Tahapan ini adalah tahapan yang menerapkan metode dari *data mining* untuk mengolah data yang ada. Metode yang digunakan adalah *Naive Bayes* dengan proses kerja yang telah dijelaskan sebelumnya.

Nantinya akan dilakukan perhitungan berdasarkan tiap-tiap atributnya sesuai dengan kelas yang telah ditentukan. *Output* yang didapat adalah nilai kelas yang tertinggi yang dapat disimpulkan bahwa kelas tersebut merupakan metode yang cocok atau sesuai untuk digunakan berdasarkan data uji yang telah dihitung sebelumnya.

### 2.3.1. *Preprocessing*

Sebelum tahap implementasi diterapkan, pertama-tama tahap *preprocessing* dahulu dilakukan. Jumlah data awal yang dapat diperoleh dari pengumpulan data yaitu 968 data, tetapi tidak semua data digunakan dan tidak semua atribut digunakan karena data tersebut harus melalui tahap pengolahan awal data atau disebut dengan *preparation* data.



Gambar 2.2 Tahap *Preprocessing*

Untuk mendapatkan data yang memiliki kualitas bagus, beberapa tahap harus dilakukan, berikut adalah tahapannya:

#### 2.3.1.1. *Data Cleaning*

Data yang tidak berkualitas akan menghasilkan data mining yang tidak berkualitas. Keputusan yang berkualitas harus didasarkan pada data yang berkualitas (data ganda atau data yang telah hilang dapat menyebabkan ketidakbenaran). Ekstaksi data, pembersihan data, dan transformasi data merupakan tugas utama dalam data warehouse. Tugas-tugas *data cleaning*, yaitu:

1. Mengisi nilai yang telah hilang pada data yang tidak lengkap (*missing value*).
2. Mengidentifikasi atau menghilangkan *outliers* dan memperhalus data *noise*.
3. Memperbaiki ketidak konsistenan data.
4. Memecahkan redundansi yang disebabkan oleh integrasi data.

#### 2.3.1.2. *Data Integration* dan *Transformation*

Dapat meningkatkan akurasi dan efisiensi algoritma. Algoritma *Naive Bayes* mampu melakukan proses data yang bernilai nominal, ordinal, atau kontinu. Sehingga nilai dari tiap-tiap *attribute* yang terdapat pada *dataset* tidak perlu ditransformasikan.

1. Mengintegrasikan berbagai *database* atau *file-file*.
2. Transformasi data (*data transformation*)
3. Normalisasi dan agregasi

### 2.3.1.3. Data Reduction

Proses yang dilakukan dengan mereduksi suatu dataset yaitu dengan mengurangi jumlah atribut dan *record* yang ada sehingga lebih sedikit tetapi masih tetap informatif.

1. Mendapatkan hasil representasi pada volume data yang sebelumnya sudah berkurang namun tetap menghasilkan analisis yang sama.
2. Dekritisasi data adalah bagian dari proses reduksi data, namun pada bagian ini adalah penting untuk data-data yang *numerik*.

### 2.3.2. Dataset (preprocessed)

*Dataset (preprocessed)* adalah *dataset* yang sudah melalui *preprocessing* diantaranya sudah melalui proses *cleaning*, *integration*, dan *transformation*, dan *reduction*, sehingga *dataset* tersebut sudah siap digunakan untuk *data training* dan *data testing*.

### 2.3.3. Data Training dan Data Testing

Pada *data training* biasanya digunakan sebagai pola data pembentukan model *data mining*. Sedangkan *data testing* adalah data yang digunakan untuk pengujian dalam menerapkan model klasifikasi.

### 2.3.4. Membangun Model Naive Bayes

Berdasarkan pada *data training* yang sebelumnya sudah melalui *preprocessing* sehingga dapat digunakan untuk membangun model *Naive Bayes* yaitu menyelesaikan masalah klasifikasi *class* atau atribut dalam data.

### 2.3.5. Penerapan Model Naive Bayes

Penerapan model *Naive Bayes* dilakukan berdasarkan model yang telah dibangun sebelumnya yang digunakan untuk menentukan atribut atau *class* dari suatu data baru yang atribut atau *class* nya belum diketahui sebelumnya.

### 2.3.6. Hasil akurasi

Hasil akurasi merupakan hasil nilai akurasi yang didapat dari penerapan algoritma *Naive Bayes*. Untuk menghitung akurasi dari pola yang didapat dalam metode algoritma yang digunakan.

### 2.3.7. Kesimpulan dan Saran

Pada tahapan ini model yang sudah dibuat dan dievaluasi keakuratannya dan

generalitasnya. Tahap ini mengukur sejauh mana model yang sudah dipilih

## 3. Hasil dan Pembahasan

### 3.1. Pengumpulan Data

Pada pengumpulan data ini yang digunakan adalah data penjualan CV Mitra Artha Sejati Bandung tahun 2017, 2018, 2019 dan 2020. Data yang diambil sebanyak 968 data dan yang dapat digunakan sebanyak 554 data. Kenapa data yang diambil berjumlah sekian karena banyak data yang ada pada variabel tertentu bernilai null sehingga data tersebut tidak dapat digunakan untuk proses data mining.

No. Faktur	Tgl Faktur	Tgl Kirim	Segment	PIC	Cust Code	Nama Toko	Alamat	Wilayah	No Telp	Nama Pelanggan	Tanggal Kirim	Tanggal Jatuh tempo	Umur Faktur	No. PO	No. Surat Jalan	No Faktur	Distribusi	Retur	Sales	Uang Masuk	Sisa Tagihan	Status Lunas	Group	Produk	Kemasan	Isi perbox	Qty	Harga	Diskon	Konsinyasi
000001	2017-01-01	2017-01-01	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Gambar 3.1 Data Penjualan

Data asli yang dihasilkan penulis memiliki 28 atribut diantaranya yaitu : *Segment*, *PIC*, *Cust Code*, Nama Toko, Alamat, Wilayah, No Telp, Nama Pelanggan, Tanggal Kirim, Tanggal Jatuh tempo, Umur Faktur, No. PO, No. Surat Jalan, No Faktur, Distribusi, *Retur*, *Sales*, Uang Masuk, Sisa Tagihan, Status Lunas, *Group*, Produk, Kemasan, Isi perbox, Qty, Harga, Diskon, Konsinyasi.

### 3.2. Seleksi Data

*Data Selection* adalah proses menganalisis data-data yang relevan dari database karena sering ditemukan bahwa tidak semua data dibutuhkan dalam proses *data mining*. Data tersebut dipilih dan diseleksi dari *database* untuk di analisis. Dari semua data yang digunakan hanya tanggal, *salesman*, *segment*, Arizona 55 EC, Rocker 75 EC, Nero 32,5 WP, dan *retur*. Karena informasi yang terkandung didalamnya sudah mewakili informasi yang dibutuhkan untuk dijadikan indikator penelitian.

### 3.3. Pembersihan Data

Setelah tahap pengumpulan data dan filter data maka tahap selanjutnya yaitu *cleaning* data agar tidak ada duplikasi data,

memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data seperti kesalahan cetak, sehingga data tersebut dapat diolah dan dilakukan proses *data mining*.

Setelah semua data yang dibutuhkan telah melalui tahap *cleaning* data maka penulis mendapatkan 554 data yang akan di olah menjadi 2 bagian yaitu *data training* dan *data testing* yang siap dimasukkan pada *software* yang digunakan pada penelitian ini.

### 3.4. Transformasi Data

*Data Transformation* adalah tahap mengubah data menjadi bentuk yang sesuai untuk proses *data mining*. Dalam penelitian ini data yang akan diproses dari *Microsoft Excel* 2016 akan digunakan untuk pengolahan pada *Software Rapidminer*. Dan atribut data juga dirubah menjadi yaitu tanggal menjadi *tgl\_transaksi*, *salesman* menjadi *nm\_salesman*, Arizona 55 EC menjadi *arizona\_55ec*, Rocker 75 EC menjadi *rocker\_75ec*, Nero 32,5 WP menjadi *nero\_32,5wp* dan retur menjadi *retur\_penjualan*

Tabel 3.1 Inisialisasi Retur Penjualan

Retur_Penjualan	Inisialisasi
Retur > 0	YA
Retur = 0	TIDAK

Tabel 3.2 Inisialisasi Arizona 55 EC

Sales Arizona 55 EC	Inisialisasi
Sales 0	Sales = 0
Sales 1 sd. 5.000.000	Sales < 5 juta
Sales 5.000.000 sd. 10.000.000	Sales 5 juta > 10 juta
Sales 11.000.000 sd. 50.000.000	Sales 11 juta > 50 juta
Sales > 50.000.000	Sales > 50 juta

Tabel 3.3 Inisialisasi Rocker 75 EC

Sales Rocker 75 EC	Inisialisasi
--------------------	--------------

Sales 0	Sales = 0
Sales 1 sd. 5.000.000	Sales < 5 juta
Sales 5.000.000 sd. 10.000.000	Sales 5 juta > 10 juta
Sales 11.000.000 sd. 50.000.000	Sales 11 juta > 50 juta
Sales > 50.000.000	Sales > 50 juta

Tabel 3.4 Inisialisasi Nero 32.5 WP

Sales Nero 32.5 WP	Inisialisasi
Sales 0	Sales = 0
Sales 1 sd. 5.000.000	Sales < 5 juta
Sales 5.000.000 sd. 10.000.000	Sales 5 juta > 10 juta
Sales 11.000.000 sd. 50.000.000	Sales 11 juta > 50 juta
Sales > 50.000.000	Sales > 50 juta

### 3.5. Implementasi

Pada tahap ini dilakukan pemodelan data, metode yang dipakai pada penelitian ini adalah probabilitas (prediksi) dengan menggunakan algoritma *Naive Bayes*. Data yang telah dikumpulkan, diseleksi, dan ditransformasikan akan dikelola menggunakan probabilitas. Metode ini dapat digunakan menjadi prediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sebagai perbandingan.

Data yang akan diujikan dibagi menjadi dua bagian yaitu *data training* dan *data testing*. Data penjualan pestisida CV Mitra Artha Sejati memiliki 554 *record*, untuk *data training* memiliki 443 *record* (tahun 2017, 2018, dan 2019) dan *data testing* memiliki 111 *record* (tahun 2019 dan 2020).

tgl_transaksi	nm_saleman	segment	artona_5Jec	rocker_5Jec	nero_5Jec	retur_penjualan
05/01/2017	H.CECEP	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
09/01/2017	H.CECEP	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
09/01/2017	H.CECEP	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
09/01/2017	H.CECEP	RETAIL	SALES 11 JUTA > 50 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
09/01/2017	H.CECEP	RETAIL	SALES 5 JUTA > 10 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
09/01/2017	ALI	RETAIL	SALES 11 JUTA > 50 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
09/01/2017	ALI	RETAIL	SALES 11 JUTA > 50 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
09/01/2017	ALI	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
09/01/2017	ALI	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
09/01/2017	ALI	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
11/01/2017	H.CECEP	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
11/01/2017	H.CECEP	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
11/01/2017	HEDI	SMALL BUSINESS	SALES 11 JUTA > 50 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
12/01/2017	ALI	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
14/01/2017	ALI	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
14/01/2017	ALI	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
14/01/2017	ALI	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
17/01/2017	HEDI	SMALL BUSINESS	SALES < 5 JUTA	SALES = 0	SALES < 5 JUTA	TIDAK
20/01/2017	HEDI	SMALL BUSINESS	SALES < 5 JUTA	SALES = 0	SALES < 5 JUTA	TIDAK
22/01/2017	ALI	RETAIL	SALES < 5 JUTA	SALES = 0	SALES < 5 JUTA	TIDAK
25/01/2017	DULIA	RETAIL	SALES 11 JUTA > 50 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
30/01/2017	ALI	RETAIL	SALES < 5 JUTA	SALES = 0	SALES < 5 JUTA	TIDAK
31/01/2017	DWIKI	SMALL BUSINESS	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK

Gambar 3.2 Data Training

tgl_transaksi	nm_saleman	segment	artona_5Jec	rocker_5Jec	nero_5Jec	retur_penjualan
27/05/2019	HENDRIA	RETAIL	SALES < 5 JUTA	SALES = 0	SALES = 0	TIDAK
27/05/2019	HENDRIA	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
12/06/2019	HENDRIA	RETAIL	SALES < 5 JUTA	SALES = 0	SALES < 5 JUTA	TIDAK
12/06/2019	ARA	RETAIL	SALES < 5 JUTA	SALES = 0	SALES < 5 JUTA	YA
13/06/2019	H.CECEP	COMMERCIAL	SALES < 5 JUTA	SALES = 0	SALES < 5 JUTA	TIDAK
14/06/2019	ARA	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES < 5 JUTA	TIDAK
15/06/2019	HEDI	SMALL BUSINESS	SALES < 5 JUTA	SALES < 5 JUTA	SALES < 5 JUTA	TIDAK
18/06/2019	ARA	RETAIL	SALES < 5 JUTA	SALES = 0	SALES < 5 JUTA	TIDAK
19/06/2019	ARA	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
21/06/2019	HENDRIA	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
26/06/2019	HENDRIA	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
26/06/2019	ARA	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
26/06/2019	BAWON	COMMERCIAL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
26/06/2019	HENDRIA	RETAIL	SALES < 5 JUTA	SALES = 0	SALES < 5 JUTA	TIDAK
26/06/2019	ARA	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES < 5 JUTA	TIDAK
28/06/2019	ARA	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
28/06/2019	ARA	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	YA
28/06/2019	HENDRIA	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
04/07/2019	ARA	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
04/07/2019	ATO	COMMERCIAL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK
05/07/2019	HENDRIA	RETAIL	SALES < 5 JUTA	SALES < 5 JUTA	SALES = 0	TIDAK

Gambar 3.3 Data Testing

3.6. Hasil Pembahasan

Pada tahap ini penulis melakukan perhitungan data testing berdasarkan jumlah data yang dipergunakan, yaitu 55, 138, 222, 277, dan 416 data. Sedangkan data training menggunakan 554 data. Hasil yang didapatkan berbeda-beda, seperti pada tabel dibawah.

Tabel 3.5 Hasil Uji Berdasarkan Jumlah Data

Jumlah Data	Akurasi	Ya (Ya)	Tidak (Tidak)	Tidak (Ya)	Ya (Tidak)
55	96,36%	2	51	2	0
138	93,48%	3	126	9	0
222	92,34%	7	198	17	0
277	92,42%	7	249	21	0
416	92,31%	4	380	32	0

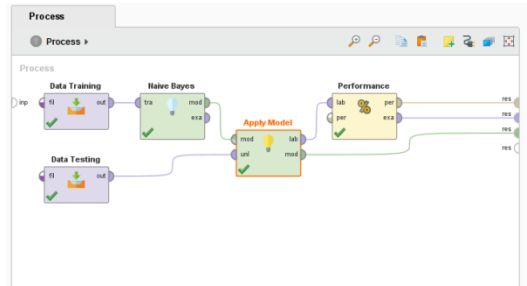
Berdasarkan tabel 5 di atas, data yang di ambil secara acak. Untuk jumlah data 55 memiliki akurasi 96,36% dengan ketepatan prediksi sebanyak 53 data. Untuk jumlah data 138 memiliki tingkat akurasi 93,48% dengan ketepatan prediksi 129 data. Untuk jumlah data 222 memiliki tingkat akurasi 92,42% dengan ketepatan prediksi 205 data. Untuk jumlah data 277 memiliki tingkat akurasi 92,42% dengan ketepatan prediksi 256 data. Untuk jumlah data 416 memiliki tingkat akurasi 92,31%

dengan ketepatan prediksi 384 data.

Dapat dilihat dari hasil perhitungan pada tabel 5 di atas bahwa semakin besar jumlah data yang digunakan maka semakin kecil tingkat akurasi dan sebaliknya. Besar kecilnya tingkat akurasi juga di tentukan oleh isi data dari tiap atribut.

3.7. Pengujian Software Rapidminer

Pada tahap ini penulis menggunakan data training tahun 2017, 2018, dan 2019 sedangkan data testing menggunakan data 2019 dan 2020 sebagai contoh. *Attribute* yang digunakan sebagai label adalah retur\_penjualan, penulis menganalisis 111 data (tahun 2019 dan 2020) yang dipakai sebagai *data testing* dengan format excel untuk memprediksi penjualan yang di retur pada setiap transaksi agar bisa di akses menggunakan *Software Rapidminer*.



Gambar 3.4 Model Naive Bayes

Dari hasil perhitungan menggunakan Rapidminer dengan metode prediksi menampilkan hasil dari data testing dan data training yang telah di uji. Kolom ini memberikan informasi tentang data penjualan yang mengalami retur atau tidak pada tahun 2019 dan 2020. Selanjutnya untuk mengetahui tingkat accuracy algoritma *Naive Bayes*, klik tab *Performance Vector (Performance)* yang

accuracy: 94.59%			
	true YA	true TIDAK	class precision
pred. YA	3	0	100.00%
pred. TIDAK	6	102	94.44%
class recall	33.33%	100.00%	

terletak disebelah kanan, sehingga tabel view seperti gambar dibawah.

Gambar 3.5 Hasil Accuracy

Dapat dilihat pada gambar 3.5. Tingkat accuracy dari PerformanceVector yaitu



94.59%, class precision yaitu “YA” 100.00%, “TIDAK” 94.44%, dan untuk class recall yaitu “YA” 33.33%, sedangkan “TIDAK” 100.00%.

Secara umum precision, recall, dan accuracy dapat di rumuskan sebagai berikut:

□ Untuk kelas “YA”  
 Precision = = 1,00 = 100,00%  
 Recall = = 0,3333333333333333  
 = 33,33%

□ Untuk kelas “TIDAK”  
 Precision = = 0,9444444444444444 = 94,44%

Recall = = 1,00 = 100,00%

□ Untuk tingkat accuracy  
 Accuracy = = 0,9459459459459459  
 = 94,59%

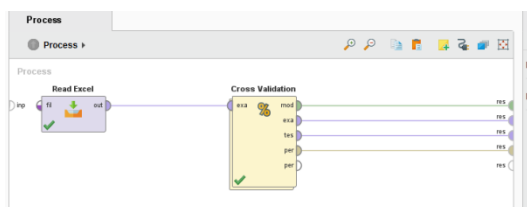
Pengujian kedua menggunakan data penjualan 2017, 2018, 2019, dan 2020 yang “YA” atau “TIDAK” pada retur penjualan sebagai data training dan data testing. Data tersebut kemudian di ujikan dengan menggunakan operator cross

	true TIDAK	true YA	class precision
pred. TIDAK	496	34	93.58%
pred. YA	10	14	58.33%
class recall	98.02%	29.17%	

validation di Rapidminer seperti pada gambar

Gambar 3.6 Model Cross Validation

Setelah semuanya terhubung dan parameter operator cross validation terisi, klik Run pada toolbar. Maka akan muncul hasil perhitungan seperti pada gambar



Gambar 3.7 Hasil Cross Validation

#### 4. Kesimpulan

Dalam penelitian dan pengimplementasian *data mining* untuk prediksi penjualan CV Mitra Artha Sejati

menggunakan algoritma *Naive Bayes*, maka didapatkan beberapa kesimpulan yaitu:

1. Algoritma *Naive Bayes* telah berhasil diterapkan untuk menganalisis prediksi penjualan pestisida pada CV Mitra Artha Sejati dalam hal memprediksi transaksi distribusi produk yang terjadi untuk menjadi acuan penjualan tersebut akan di retur atau tidak.
2. Penerapan *data mining* membantu perusahaan dalam hal pengambilan keputusan yang melakukan prediksi (*forecasting*) terhadap penjualan dan retur penjualan sehingga perusahaan dapat mengambil langkah – langkah penting menghadapinya.
3. Hasil prediksi menggunakan Algoritma *Naive Bayes* mampu memberikan nilai tingkat akurasi sebesar 94,59% dengan *class precision* yaitu “YA” 100.00%, “TIDAK” 94.44%, dan untuk *class recall* yaitu “YA” 33.33%, sedangkan “TIDAK” 100.00% dengan jumlah *data training* 443 data dan 111 untuk *data testing*. Sedangkan perhitungan kedua menggunakan *cross validation* menghasilkan tingkat akurasi sebesar 92.06%.

#### Referensi

- Ahmadi. (2018). *Perancangan Aplikasi Prediksi Penjualan Handphone Dengan Menerapkan Metode Mancova*. 17(April), 150–152.
- Alfarisi, S. (2017). *Sistem Prediksi Penjualan Gamis Toko Qitaz Menggunakan Metode Single Exponential Smoothing*. 4(1), 80–95.
- Bulolo, E. (2020). *Data Mining Untuk Perguruan Tinggi* (pertama). Depublish.
- Fuad, N., & Sulistiono, E. (2016). *Prediksi Penjualan Kerudung Rabbani Di Griya Muslim Store Dukun Gresik Dengan Trend Moment*. 49–53.
- Manalu, E., Sianturi, F. A., & Manalu, M. R. (2017). *Penerapan Algoritma Naive Bayes Untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan Dan Jumlah Pemesanan Pada CV . Papadan Mama Pastries*. 1(2).

Tiara, B., Tinggi, S., Informatika, M., & Pembangunan, K. I. (2018). *Prediksi Penjualan Menggunakan Algoritma Neural Network*. 6(1).