

**PENERAPAN DATA MINING PADA PENJUALAN BARANG  
MENGUNAKAN METODE METODE NAIVE BAYES CLASSIFIER  
UNTUK OPTIMASI STRATEGI PEMASARAN**

Odi Nurdyawan<sup>\*1</sup>, Noval Salim<sup>#2</sup>

*Program Studi Teknologi Informasi, STMIK IKMI Cirebon*  
Jl. Perjuangan No. 10B Majasem – Kota Cirebon, Tlp. (0231) 490480  
E-mail : odynurdyawan@gmail.com<sup>\*1</sup>, novalsalimy@gmail.com<sup>#2</sup>

**Abstrak**

*Perubahan perilaku konsumen terhadap perkembangan zaman, hal ini menyebabkan para pelaku usaha property berinovasi supaya bisnis yang dijalankan tetap bertahan. Dengan adanya kompetensi ini menyebabkan masalah baru yang lebih kompleks yaitu munculnya startup-startup baru yang akhirnya akan menjadi pesaing dalam dunia bisnis. Menyikapi hal tersebut di butuhkan langkah inovasi yang tepat untuk menghadapi persaingan dalam dunia bisnis. Salah satu cara inovasi dengan melakukan analisis penjualan produk untuk menghasilkan pengetahuan baru yang kemudian akan di gunakan untuk optimasi strategi pemasaran produk agar target pasar sesuai dengan yang di harapkan. Hal ini akan berpengaruh besar kepada penghasilan dan omset yang di dapat oleh perusahaan.*

*Pendekatan yang digunakan pada paper ini ialah metode Naive Bayes Classifier. Metode ini bekerja dalam himpunan data kemudian di ekstrak menjadi pengetahuan baru yang akan di gunakan untuk optimasi strategi pemasaran. Algoritma Naive Bayes Classifier juga bekerja dalam tipe data numerik yang dapat memudahkan dalam proses analisa. Proses pada metode ini yaitu proses analisa pola data penjualan yang telah ada sebelumnya (Learning Phase) berdasarkan atribut – atribut yaitu jenis, waktu, ukuran yang di ujikan dan proses dari analisa.*

*Paper penelitian ini menghasilkan pengetahuan baru. Selain hal tersebut dari proses analisa dengan metode Naive Bayes Classifier yaitu menghasilkan pola penjualan berdasarkan atribut-atribut yang telah di tentukan. Hasil dari proses analisa ini akan di gunakan untuk kepentingan perusahaan dalam upaya optimasi strategi pemasaran. Pengetahuan baru ini juga dapat memberikan informasi penting seperti hasil prediksi minat pembeli yang dapat digunakan dalam efektivitas dan efisiensi pemasaran dan peningkatan penjualan.*

**Kata kunci :** *Naive Bayes Classifier, Penjualan, Optimasi*

**1. Pendahuluan**

**A. Latar Belakang.**

Dewasa ini dengan kemajuan teknologi yang ada data tersebut bisa di olah lagi menjadi data yang lebih bermanfaat. Pada data mining Penggalian Informasi ini sangat berguna untuk meningkatkan keuntungan ataupun membantu dalam penyusunan strategi pemasaran. Salah satu penggunaan data mining dengan metode *Naive Bayes Classifier* dalam data penjualan umumnya adalah untuk mengetahui minat dan ketertarikan calon pembeli terhadap produk yang tersedia dengan memprediksi probabilitas keanggotaan suatu class. Informasi ini dapat digunakan untuk mendukung strategi pemasaran agar lebih efektif dan efisien. Dengan informasi ini suatu perusahaan dapat mengetahui tingkat ketertarikan pembeli terhadap suatu produk yang terdapat di perusahaan. Sehingga perusahaan dapat mengetahui serta menentukan target atau sasaran pasar dengan lebih rinci.

Menurut Mohammad Ridwan Effendi dalam penelitiannya yang berjudul Akurasi Data Mining Untuk Menghasilkan Pola Kelulusan Mahasiswa dengan Metode Naive Bayes mengatakan :

“ Data mining adalah kegiatan menemu- kan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam database, data warehouse, atau penyimpanan informasi

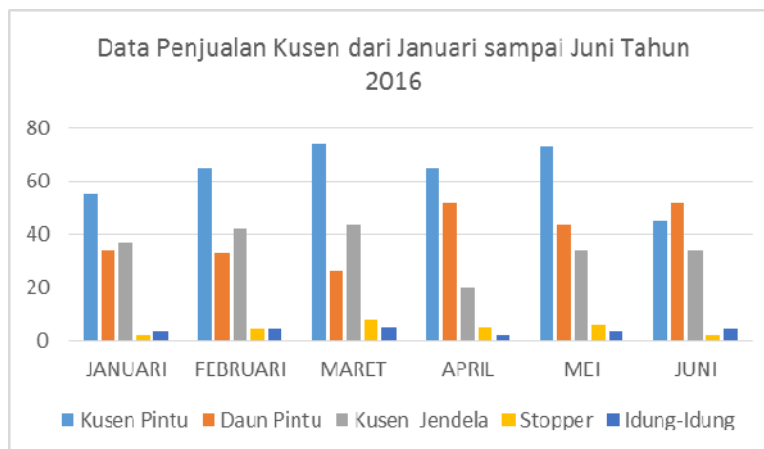
lainnya. Data mining berkaitan dengan bidang ilmu- ilmu lain, seperti database system, data warehousing, statistik, machine learning, information retrieval, dan komputasi ti- ngkat tinggi.”[3]

Penelitian yang dilakukan oleh Selvia Lorena Br, Ginting, dan Reggy Pasya Trinanda pada tahun 2015 dalam penelitiannya yang Berjudul Teknik Data Mining Menggunakan Metode Bayes Classifier Untuk Optimalisasi Pencarian Pada Aplikasi Perpustakaan menjelaskan bahwa Salah satu tugas Data Mining adalah klasifikasi data, yaitu memetakan (mengklasifikasikan) data ke dalam satu atau beberapa kelas yaang sudah didefinisikan sebelumnya. Salah satu metode dalam klasifikasi data adalah Naive Bayes Classifier (NBC). Naive Bayes Classifier merupakan salah satu metoda machine learning yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. [4]. Dari penjelasan terserbut dapat di ambil kesimpulan bahwa metode Naive Bayes Classifier merupakan metode yang di gunakan untuk klasifikasi data dan juga dapat memprediksi probabilitas di masa depan dari hasil analisis data sebelumnya. Hal ini bisa di kaitkan dengan penelitan ini karena data yang di gunakan sama yaitu data penjualan di masa lampau.

Menurut penelitian yang dilakukan oleh Alfa Saleh pada tahun 2015 yang berjudul Implementasi Metode Klasifikasi Naive Bayes dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga mengatakan Penggunaan listrik yang tidak bijak tentu saja akan berdampak pada tingginya penggunaan listrik, hal ini juga mempengaruhi menipisnya persediaan energi listrik dikarenakan kebutuhan akan energi listrik lebih besar dari persediaan akan energi listrik, untuk itu diharapkan setiap masyarakat memahami upaya dalam menggunakan listrik dengan bijak. Selain itu setiap rumah tangga juga harus tahu besarnya penggunaan listrik di rumah tangga masing-masing. Dalam hal mengetahui penggunaan listrik ini maka diterapkan teknik klasifikasi data mining menggunakan metode naïve bayes dalam memprediksi penggunaan listrik rumah tangga.[7]

Dari pembahasan diatas dapat disimpulkan fakta yang ingin dibuat disini ialah manfaat metode Naive Bayes Classifier yaitu dapat memprediksi penggunaan listrik rumah tangga agar kedepan bisa dilakukan penghematan listrik. Oleh karena itu metode ini juga dapat di terapkan dalam memprediksi pola pembelian dari analisis himpunan data penjualan di masa lampau.

Berdasarkan data yang di peroleh dari CV. Maha Karya Kusen terkait data penjualan kusen pintu dan jendela dalam rentan waktu 6 bulan pada tahun 2016, maka dapat di peroleh data sebagai berikut :



Sumber Data : Bag. Administrasi CV. Maha Karya Kusen

Gambar 1 Data Penjualan Kusen dari Januari sampai Juni Tahun 2016

Data penjualan kusen pada bulan januari sampai juni 2016 menjelaskan bahwa pada bulan januari terdapat pembelian 55 unit kusen pintu, 34 unit daun pintu, 37 unit kusen jendela, 2 unit stopper dan 3 unit idung-idung. Kemudian pada bulan februari, maret, april dan mei kusen pintu mengalami kenaikan sebesar 17,24%, 19,63%, 17,24%, dan 19,36%. sedangkan penurunan pada bulan juni sebesar 11,94%. Untuk daun pintu kenaikan hanya ada pada bulan april dan mei sebesar 21,58% dan 18,26%, kusen jendela naik pada bulan februari dan maret sebesar 19,91% dan 20,85%. Stopper naik pada bulan maret dan mei sebesar 29,63% dan 22,22%. Yang terakhir idung-idung mengalami penurunan pada bulan mei sebesar 14,29%.

Data yang telah dipaparkan sebelumnya, masih banyak permasalahan-permasalahan yang terjadi di CV Maha Karya Kusen. Perusahaan ini umumnya memproduksi kusen untuk pembuatan pintu rumah, jendela, dan yang lainnya. Pelanggan perusahaan ini banyak dari luar kota untuk membeli kayu untuk proyek-proyek pembangunan, tetapi dalam strategi pemasaran masih kurang ini terlihat dari jumlah orderan dari perbulannya masih terbilang rendah. Hal itu dikarenakan banyaknya perusahaan lain yang bersaing di bidang bisnis properti ini. Oleh karena itu perlunya optimasi strategi yang baik untuk menghadapi dalam persaingan penjualan produk.

Fokus masalah dalam penelitian ini adalah bagaimana penerapan metode Naive Bayes Classifier pada data penjualan dan transaksi-transaksi penjualan untuk menciptakan pengetahuan baru yang kemudian digunakan untuk optimasi strategi pemasaran produk agar bisa tercapai dengan maksimal. Hal ini dikarenakan butuhnya strategi pemasaran yang baik untuk menghadapi pesaing bisnis lainnya dan juga bisa menambah omset penjualan perusahaan. Permasalahan yang dialami perusahaan sejak dulu dalam pemasaran yaitu masih cenderung menggunakan cara konvensional hanya menggunakan mulut ke mulut saja. masih belum mengetahui bagaimana optimasi yang baik untuk pemasaran, serta kurangnya pengetahuan tentang teknik pengolahan data penjualan yang baik dalam mengatasi solusi terbaik dari permasalahan tersebut.

## **B. Identifikasi Masalah.**

Berdasarkan pemaparan latar belakang diatas yang telah diuraikan maka diidentifikasi masalah yang menjadi fokus masalah pada paper ini sebagai berikut :

- 1) Kurang optimalnya perusahaan dalam mengolah data penjualan.
- 2) Belum adanya analisa data yang dapat membantu dalam mengoptimasi pemasaran pada cv tersebut.
- 3) Penerapan metode yang dapat menganalisa data.

## **C. Tujuan.**

- 1) Mengetahui seberapa besar dampak penerapan metode Naive Bayes Classifier dalam upaya optimasi strategi pemasaran.
- 2) Membuat pengetahuan baru dari hasil analisa menggunakan metode Naive Bayes Classifier.
- 3) Menggunakan hasil analisa untuk optimasi strategi pemasaran.

## **2. Tinjauan Pustaka**

### **A. Strategi Pemasaran**

Menurut Guiltinan dan Paul (1992), definisi strategi pemasaran adalah pernyataan pokok tentang dampak yang diharapkan akan dicapai dalam hal permintaan pada target pasar yang ditentukan [5] meliputi :

- 1) Segmentasi

Menurut Solomon dan Elnora (2003: 221), segmentasi adalah "The process of dividing a larger market into smaller pieces based on one or more meaningful, shared characteristic". Dengan melaksanakan segmentasi pasar, kegiatan pemasaran dapat dilakukan lebih terarah dan sumber daya yang dimiliki organisasi dapat digunakan secara lebih efektif dan efisien dalam rangka memberikan kepuasan bagi konsumen. Selain itu

perusahaan dapat melakukan program-program pemasaran yang terpisah untuk memenuhi kebutuhan khas masing-masing segmen.

2) Demografis

Segmentasi ini dilakukan dengan membagi pasar ke dalam kelompok-kelompok berdasarkan variabel demografis seperti: usia, jenis kelamin, besarnya keluarga, pendapatan, ras, pendidikan, pekerjaan, geografis.

3) Psikografis

Segmentasi ini dilakukan dengan membagi pasar ke dalam kelompok-kelompok yang berlainan menurut kelas sosial, gaya hidup, kepribadian, dan lain-lain. Informasi demografis sangat berguna, tetapi tidak selalu menyediakan informasi yang cukup untuk membagi konsumen ke dalam segmen-segmen, sehingga diperlukan segmen berdasarkan psikografis untuk lebih memahami karakteristik konsumen.

4) Prilaku

Segmentasi ini dilakukan dengan membagi konsumen ke dalam segmen-segmen berdasarkan bagaimana tingkah laku, perasaan, dan cara konsumen menggunakan barang/situasi pemakaian, dan loyalitas merek. Cara untuk membuat segmen ini yaitu dengan membagi pasar ke dalam pengguna dan non-pengguna produk.

## B. Optimasi

Optimasi merupakan penetapan keputusan untuk memperoleh hasil terbaik menurut kondisi. Sasaran akhir yang akan dicapai dari semua keputusan adalah meminimalkan usaha yang dibutuhkan atau memaksimalkan keuntungan. Usaha atau keuntungan dalam situasi yang praktis dapat diekspresikan sebagai fungsi dari harga variabel keputusan.[6]

## C. Naiv Bayes Classifier

Dalam Teorema Bayes probabilitas atau peluang bersyarat dinyatakan sebagai berikut :

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

Gambar 2 Rumus Teorema Bayes

Di mana  $X$  adalah bukti,  $H$  adalah hipotesis,  $P(H|X)$  adalah probabilitas bahwa hipotesis  $H$  benar untuk bukti  $X$  atau dengan kata lain  $P(H|X)$  merupakan probabilitas *posterior*  $H$  dengan syarat  $X$ ,  $P(H|X)$  adalah probabilitas bahwa bukti  $X$  benar untuk hipotesis  $H$  atau probabilitas *posterior*  $X$  dengan syarat  $H$ ,  $P(H)$  adalah probabilitas prior hipotesis  $H$ , dan  $P(X)$  adalah probabilitas prior bukti  $X$ . [2]

## D. Definisi Klasifikasi

Klasifikasi merupakan salah satu teknik yang sering digunakan untuk penambangan data (Baradwaj & Pal, 2011b). Teknik klasifikasi dilakukan dengan memprediksi label kelas yang belum diketahui menggunakan data berbeda yang memiliki label kelas yang sudah diketahui [8]

Klasifikasi dapat didefinisikan sebagai proses untuk menyatakan suatu objek data sebagai salah satu kategori (kelas) yang telah didefinisikan sebelumnya [2]

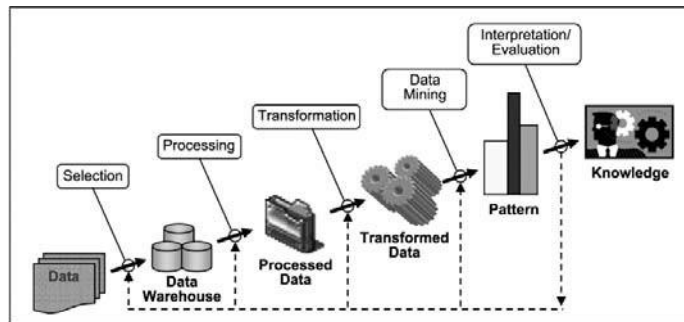
## E. Rapidminer

Rapid Miner adalah sebuah lingkungan machine learning data mining, text mining dan predictive analytics. [9] RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. RapidMiner memiliki kurang lebih

500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi. [1]

### 3. Analisa dan Pembahasan

Metode penelitian yang digunakan adalah metode analisis deskriptif dengan pendekatan kuantitatif artinya penelitian yang dilakukan adalah menekankan analisisnya pada data-data numeric (angka), yang bertujuan untuk mendapatkan gambaran yang jelas mengenai suatu keadaan berdasarkan data yang diperoleh dengan cara menyajikan, mengumpulkan dan menganalisis data tersebut sehingga menjadi informasi baru yang dapat digunakan untuk menganalisa mengenai masalah yang sedang diteliti. Adapun tahapan yang dilakukan berdasarkan kepada Knowledge Discovery Data ( KDD ) adalah sebagai berikut :



Gambar 3 Tahapan KDD

#### A. Data Selection

Menciptakan himpunan data target , pemilihan himpunan data, atau memfokuskan pada subset variabel atau sampel data, dimana penemuan (discovery) akan dilakukan. Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

#### B. Pre-processing/ Cleaning

Pemrosesan pendahuluan dan pembersihan data merupakan operasi dasar seperti penghapusan noise dilakukan. Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Dilakukan proses enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

#### C. Transformation

Pencarian fitur-fitur yang berguna untuk mempresentasikan data bergantung kepada goal yang ingin dicapai. Merupakan proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses ini merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data

#### D. Data mining

Pemilihan tugas data mining; pemilihan goal dari proses KDD misalnya klasifikasi, regresi, clustering, dll. Pemilihan algoritma data mining untuk pencarian (searching), Proses Data mining yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam

data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

#### E. Interpretation/ Evaluation

Penerjemahan pola-pola yang dihasilkan dari data mining. Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

### 4. Hasil

#### A. Deskripsi Data

Dataset penjualan terdiri dari beberapa atribut antara lain No, Bulan, Model Barang, Harga, Terjual, dan Restock. Persiapan data mencakup semua kegiatan untuk membangun dataset penjualan yang akan diterapkan ke dalam alat pemodelan, dari data mentah awal berupa dataset penjualan dan selanjutnya akan melakukan proses data mining.

#### B. Pengolahan data mentah (*Preprocessing Data*)

Pada tahap ini merupakan tahap untuk memastikan data mahasiswa yang dipilih telah layak untuk dilakukan proses pengolahan. Data yang berjenis numerikal seperti harga dan terjual harus dilakukan proses inisialisasi data terlebih dahulu ke dalam bentuk nominal. Untuk melakukan inisialisasi tahun lulus dapat dilakukan dengan:

- 1) Barang yang terjual lebih dari 5 dalam sebulan diberi inisial pada atribut RESTOCK "YES".
- 2) Barang yang terjual kurang atau sama dengan 5 dalam sebulan diberi inisial pada atribut RESTOCK "NO"

E	F
TERJUAL	RESTOCK
7	yes
4	no
5	no
6	yes
7	yes
8	yes
9	yes
9	yes
5	yes

Gambar 4 Membuat Kelas Restok

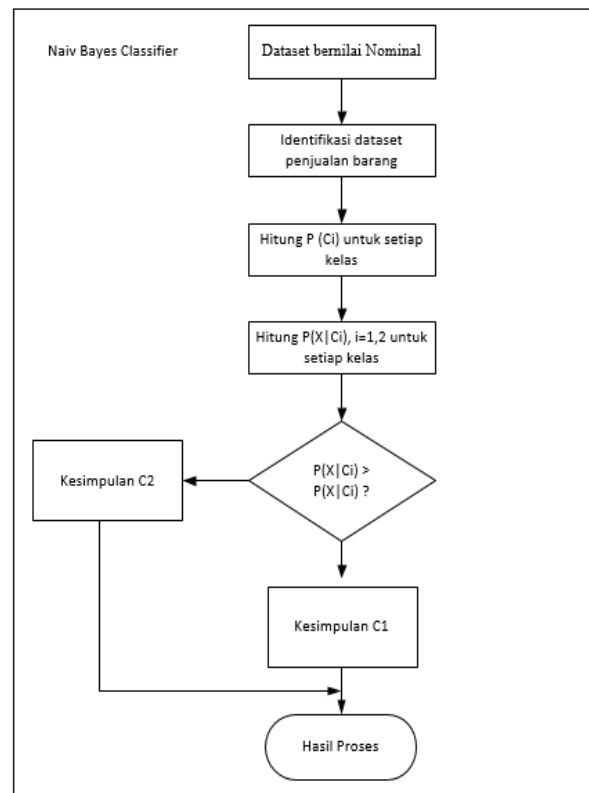
- 3) Berikut adalah dataset penjualan barang yang telah dilakukan inisialisasi yang terdapat pada Gambar

A	B	C	D	E	F
NO	BULAN	MODEL BARANG	HARGA BARANG	TERJUAL	RESTOCK
1	januari	KUSEN MERANTI	367500	7	yes
2	januari	DAUN PINTU MERANTI	245000	4	no
3	Januari	DAUN JENDELA MERANTI	312500	5	no
4	Januari	KUSEN KRUING	105000	6	yes
5	Januari	DAUN PINTU KRUING	550000	7	yes
6	Januari	DAUN JENDELA KRUING	350000	8	yes
7	Januari	KUSEN KAMPER	350000	9	yes
8	Januari	DAUN PNITU KAMPER	175000	9	yes
9	Januari	DAUN JENDELA KAMPER	150000	5	yes
10	Januari	ROSTER	150000	3	no
11	Februari	KUSEN MERANTI	367500	1	no
12	Februari	DAUN PINTU MERANTI	245000	1	no
13	Februari	DAUN JENDELA MERANTI	312500	2	no
14	Februari	KUSEN KRUING	105000	2	no
15	Februari	DAUN PINTU KRUING	550000	1	no
16	Februari	DAUN JENDELA KRUING	350000	1	no
17	Februari	KUSEN KAMPER	350000	1	no
18	Februari	DAUN PNITU KAMPER	175000	2	no
19	Februari	DAUN JENDELA KAMPER	150000	2	no
20	Februari	ROSTER	150000	1	no
21	Maret	KUSEN MERANTI	367500	4	yes
22	Maret	DAUN PINTU MERANTI	245000	1	no
23	Maret	DAUN JENDELA MERANTI	312500	1	no

Gambar 5 Penginisialisasian Data Set

### C. Pemodelan

Pemodelan adalah fase yang secara langsung melibatkan teknik data mining yaitu dengan melakukan pemilihan teknik data mining dan menentukan algoritma yang akan digunakan. Tool yang digunakan adalah Rapid Miner Versi 7.6.001



Gambar 6 Flowchart Naive Bayes

Dari data atribut nominal, kemudian identifikasi dataset Penjualan. Hitung  $P(C_i)$  untuk setiap atribut, dalam kasus dataset pada penelitian ini yaitu atribut restock yang terdiri dari 2 kelas yaitu kelas restock barang dinyatakan “YES” dan tidak direstock dinyatakan “NO”. Kemudian hitung  $P(X|C_i)$ ,  $i=1,2$  untuk setiap kelas atau atribut. Setelah itu bandingkan, jika  $P(X|C_1) > P(X|C_2)$  maka kesimpulannya  $C_1$  adalah kelas di restock = “YES”. Jika  $P(X|C_1) < P(X|C_2)$  maka kesimpulannya  $C_2$  tidak direstock = “NO”.

#### D. Perhitungan Data Mining

Berikut perhitungan manual naiv bayes dengan menggunakan dataset pada tabel berikut jika data training dan data testing di pilih secara acak.

Tabel 1 data set

NO	BULAN	MODEL BARANG	HARGA BARANG	TERJUAL	RESTOCK
1	januari	kusen meranti	367500	7	yes
2	januari	daun pintu meranti	245000	4	no
3	januari	daun jendela meranti	312500	5	no
4	januari	kusen kruing	105000	6	yes
5	januari	daun pintu kruing	550000	7	yes
6	januari	daun jendela kruing	350000	8	yes
7	januari	kusen kamper	350000	9	yes
8	januari	daun pnitu kamper	175000	9	yes
9	januari	daun jendela kamper	150000	5	yes
1...120	januari	roster	150000	3	no

Menghitung jumlah kelas dari tahun lulus berdasarkan klasifikasi yang terbentuk (prior probability) :

- 1)  $C_1$  ( Class Restock = “yes”) = jumlah “YES” pada kolom F Restock =  $57/120 = 0,475$ .
- 2)  $C_2$  (Class Restock = “no”) = jumlah “NO” pada kolom F Restock =  $63/120 = 0,525$

#### E. Menghitung jumlah kasus yang sama pada setiap atribut dari kelas Tahun Lulus ( yes / no ) berdasarkan data testing.

- 1)  $P(\text{MODEL BARANG “DAUN JENDELA KAMPER”} | \text{Class Class Restock = “YES”}) = 5/12 = 0,416666667$ .
- 2)  $P(\text{MODEL BARANG “DAUN JENDELA KAMPER”} | \text{Class Class Restock = “NO”}) = 7/12 = 0,583333333$ .
- 3)  $P(\text{MODEL BARANG “DAUN JENDELA KRUING”} | \text{Class Class Restock = “YES”}) = 4/12 = 0,333333333$ .
- 4)  $P(\text{MODEL BARANG “DAUN JENDELA KRUING”} | \text{Class Class Restock = “NO”}) = 8/12 = 0,666666667$ .
- 5)  $P(\text{MODEL BARANG “DAUN JENDELA MERANTI”} | \text{Class Class Restock = “YES”}) = 5/12 = 0,416666667$ .
- 6)  $P(\text{MODEL BARANG “DAUN JENDELA MERANTI”} | \text{Class Class Restock = “NO”}) = 7/12 = 0,583333333$ .
- 7)  $P(\text{MODEL BARANG “DAUN PINTU KAMPER”} | \text{Class Class Restock = “YES”}) = 8/12 = 0,666666667$ .



- 8)  $P(\text{MODEL BARANG "DAUN PINTU KAMPER"} | \text{Class Class Restock} = \text{"NO"}) = 4/12 = 0,333333333$ .
- 9)  $P(\text{MODEL BARANG "DAUN PINTU KRUING"} | \text{Class Class Restock} = \text{"YES"}) = 7/12 = 0,583333333$ .
- 10)  $P(\text{MODEL BARANG "DAUN PINTU KRUING"} | \text{Class Class Restock} = \text{"NO"}) = 5/12 = 0,416666667$ .
- 11)  $P(\text{MODEL BARANG "DAUN PINTU MERANTI"} | \text{Class Class Restock} = \text{"YES"}) = 4/12 = 0,333333333$ .
- 12)  $P(\text{MODEL BARANG "DAUN PINTU MERANTI"} | \text{Class Class Restock} = \text{"NO"}) = 8/12 = 0,666666667$ .
- 13)  $P(\text{MODEL BARANG "KUSEN KAMPER"} | \text{Class Class Restock} = \text{"YES"}) = 7/12 = 0,583333333$ .
- 14)  $P(\text{MODEL BARANG "KUSEN KAMPER"} | \text{Class Class Restock} = \text{"NO"}) = 5/12 = 0,416666667$ .
- 15)  $P(\text{MODEL BARANG "KUSEN KRUING"} | \text{Class Class Restock} = \text{"YES"}) = 6/12 = 0,5$ .
- 16)  $P(\text{MODEL BARANG "KUSEN KRUING"} | \text{Class Class Restock} = \text{"NO"}) = 6/12 = 0,5$ .
- 17)  $P(\text{MODEL BARANG "KUSEN MERANTI"} | \text{Class Class Restock} = \text{"YES"}) = 7/12 = 0,583333333$ .
- 18)  $P(\text{MODEL BARANG "KUSEN MERANTI"} | \text{Class Class Restock} = \text{"NO"}) = 5/12 = 0,416666667$ .
- 19)  $P(\text{MODEL BARANG "ROSTER"} | \text{Class Class Restock} = \text{"YES"}) = 5/12 = 0,416666667$ .
- 20)  $P(\text{MODEL BARANG "ROSTER"} | \text{Class Class Restock} = \text{"NO"}) = 7/12 = 0,583333333$ .

#### F. Kalikan Semua Hasil Variable

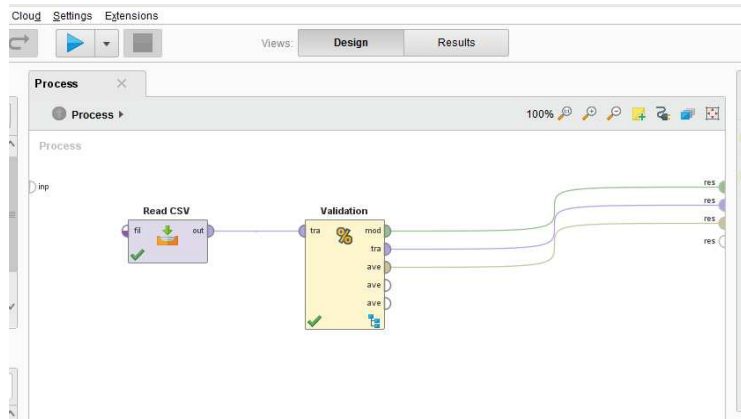
- 1) Untuk semua atribut Class Restock = "yes"  $P(X | \text{Class Tahun Lulus} = \text{"yes"}) = 0,416666667 \times 0,333333333 \times 0,416666667 \times 0,666666667 \times 0,583333333 \times 0,333333333 \times 0,583333333 \times 0,5 \times 0,583333333 \times 0,416666667 = 0,000531806$
- 2) Untuk semua atribut Class Restock = "no"  $P(X | \text{Class Tahun Lulus} = \text{"no"}) = 0,583333333 \times 0,666666667 \times 0,583333333 \times 0,333333333 \times 0,416666667 \times 0,666666667 \times 0,416666667 \times 0,5 \times 0,416666667 \times 0,583333333 = 0,001063611$
- 3) Perkalian prior probability dengan semua atribut Class Restock = "yes",  $P(C_i) | \text{Class Restock} = \text{"yes"} \times P(X | \text{Class Restock} = \text{"yes"}) = 0,475 \times 0,000531806 = 0,000252608$
- 4) Perkalian prior probability dengan semua atribut Class Restock = "no",  $P(C_i) | \text{Class Restock} = \text{"no"} \times P(X | \text{Class Restock} = \text{"no"}) = 0,525 \times 0,001063611 = 0,000558396$

#### G. Bandingkan Hasil Kelas

Class restock = "no" (Perhitungan antara perkalian Class restock "yes" dengan Class restock "no" menunjukkan bahwa nilai Class restock = "no" lebih besar dibandingkan kelas tahun lulus "yes").

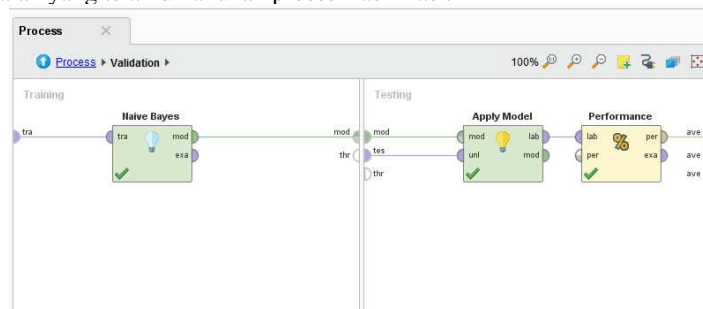
#### H. Implementasi dengan Rapidminer

Berikut adalah pengolahan data dengan menggunakan naïve bayes pada RapidMiner dapat dilihat pada gambar 7 sebagai berikut :



Gambar 7 Pengolahan Data Dengan Naiv Bayes Pada Rapidminer

- 1) Pemodelan adalah tahapan (langkah) dalam membuat model dari suatu sistem nyata (realitas).
- 2) Rapidminer adalah sebuah lingkungan machine learning data mining, text mining dan predictive analytics.
- 3) Jumlah dataset mencapai 120 record maka penulis menggunakan tools Rapidminer untuk membantu proses perhitungan.
- 4) Validation untuk membantu menghasilkan tingkat keakurasian berdasarkan dataset penjualan yang telah di lakukan proses klasifikasi.

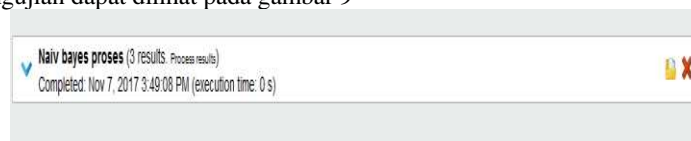


Gambar 8 Proses Klasifikasi

Tujuan utama penelitian ini adalah untuk mengetahui nilai akurasi dari algoritma naive bayes yang digunakan untuk mengklasifikasi kelulusan. Di dalam kolom training terdapat algoritma klasifikasi yang diterapkan yaitu Naive bayes. Sedangkan di dalam kolom testing terdapat Apply Model untuk menjalankan model naive bayes dan Performance untuk mengukur performa dari model Naive bayes tersebut.

#### I. Hasil Pengujian dan Percobaan

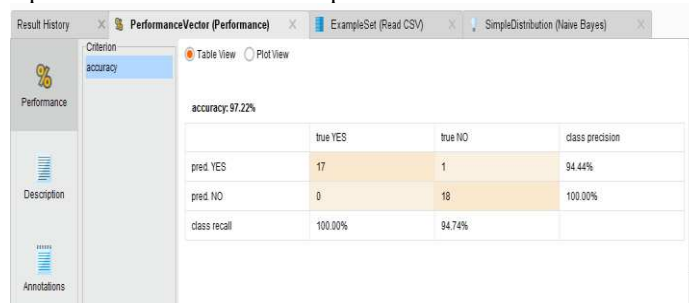
Hasil pengujian dapat dilihat pada gambar 9



Gambar 9 Hasil Pengujian Data Mining

Pada percobaan dengan algoritma naive bayes dengan menggunakan tools Rapidminer diperoleh waktu komputasi adalah 0 second. 0 second disini artinya komputasi menggunakan naive bayes berjalan cukup cepat. Hal ini sesuai dengan kelebihan naive

bayes dibandingkan beberapa algoritma lain seperti neural network yang membutuhkan waktu cukup lama untuk melakukan komputasi data.

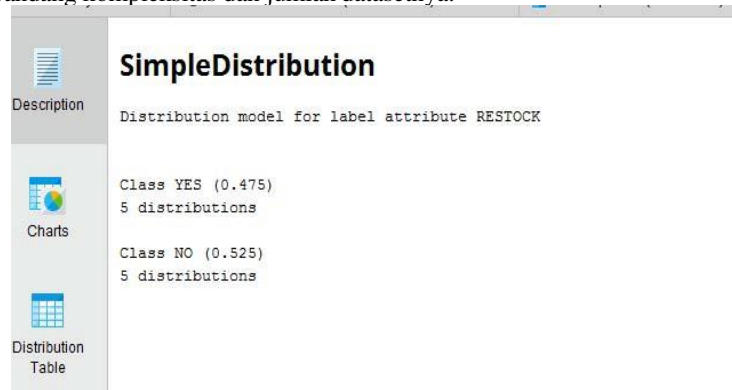


The screenshot shows the 'Performance Vector (Performance)' window in Rapidminer. It displays a table of performance metrics for a Naive Bayes model. The overall accuracy is 97.22%.

	true YES	true NO	class precision
pred. YES	17	1	94.44%
pred. NO	0	18	100.00%
class recall	100.00%	94.74%	

Gambar 10 Hasil Proses Akurasi Data

Hasil akurasi model naive bayes menunjukkan tingkat akurasinya 97.22%, artinya model klasifikasi kelulusan menggunakan naive bayes terbukti baik hal ini dilihat dari tingkat akurasinya yang mencapai 97.22% akan tetapi hal ini perlu di tinjau ulang dari sudut pandang kompleksitas dan jumlah datasetnya.



The screenshot shows the 'SimpleDistribution' window in Rapidminer. It displays the distribution of the 'RESTOCK' attribute for two classes: YES and NO.

Class	Proportion	Distributions
Class YES	0.475	5 distributions
Class NO	0.525	5 distributions

Gambar 11 Proses Simple Distribution

Model distribusi untuk label atribut kelas Restock adalah sebagai berikut:

- 1) Class YES : 5 distributions
- 2) Class YES : 5 distributions

Percobaan pada penelitian ini menggunakan Rapidminer 7.6.001. Algoritma yang digunakan adalah naive bayes. Kemudian divalidasi dan untuk testing menggunakan Apply Model untuk menjalankan algoritma atau model naive bayes serta Performance untuk mengukur performa dari model naive bayes tersebut.

## J. Hasil Analisa

Deployment merupakan tahapan akhir dalam pembuatan laporan hasil kegiatan data mining. Laporan akhir yang berisi mengenai pengetahuan yang diperoleh atau pengenalan pola pada data dalam proses data mining. Berdasarkan penelitian yang dilakukan, telah dihasilkan suatu pola, informasi, dan pengetahuan baru dalam proses data mining untuk klasifikasi restock barang berdasarkan data penjualan cv maha karya kusen tahun 2016. Dari penelitian tersebut dihasilkan suatu pola, informasi, dan pengetahuan baru sesuai dengan tujuan data mining yaitu pola perhitungan data mining yang berisi data training dan data testing serta mencari probabilitas dari setiap atribut berdasarkan data training dan data testing untuk menghasilkan suatu informasi baru, apakah pada data penjualan cv maha karya kusen tahun 2016 barang/produk mana saja yang harus di restock/diperbanyak untuk kedepannya. Kemudian untuk menguji tingkat keakurasiannya maka digunakan Rapidminer sebagai alat bantu dalam proses pengujian

tingkat akurasi dari klasifikasi tersebut. Dari proses perhitungan data mining menggunakan algoritma naive bayes dan tingkat keakurasian, dihasilkan suatu informasi baru yaitu perhitungan data mining berdasarkan data penjualan barang tahun 2016, menunjukkan kelas restock “yes” dengan total perkalian prior probability senilai 0,000531806, sedangkan kelas restock “no” dengan total perkalian prior probability senilai 0,001063611. Untuk tingkat akurasi berdasarkan proses klasifikasi menggunakan algoritma naive bayes, dengan melalui semua tahapan dipastikan tidak ada bagian – bagian penting yang terlewatkan, dihasilkan tingkat akurasi sebesar 97.22 %. Berdasarkan hasil perhitungan data mining dan proses pengujian tingkat akurasi dengan menggunakan Rapidminer, dapat ditarik kesimpulan bahwa tahun 2016 kelas restock “no” lebih besar dari kelas tahun lulus “yes”. Sedangkan analisa yang dilakukan terhadap tingkat akurasi menggunakan algoritma naïve bayes menunjukkan bahwa nilai yang dihasilkan oleh algoritma naïve bayes memiliki tingkat kekuatan yang cukup tinggi. Hal ini di buktikan dengan hasil perhitungan yang mencapai nilai 82.08 %, Nilai 82. 08 % membuktikan bahwa model yang dibangun dapat digunakan untuk melakukan klasifikasi data penjualan. Nilai 82.08 % bisa juga di sebabkan oleh kekurangan kompleksan data yang mengakibatkan model dapat memprediksi dengan akurat.

## 5. Kesimpulan

Dari penelitian yang telah di lakukan dapat ditarik kesimpulan sebagai berikut :

- 1) Penerapan metode naïv bayes classifier pada data set penjualan barang mencapai tingkat optimasi sebesar lebih dari 65%.
- 2) Berdasarkan hasil penelitian didapatkan nilai  $p$  value ( sig ), nilai  $p$  value ( sig ) 0,807 > 0,05 maka  $H_0$  diterima dan  $H_a$  ditolak, nilai t-hitung dan t-tabel, Nilai t-hitung – 0,246 > - t-tabel -2,045 maka  $H_0$  diterima dan  $H_a$  ditolak.

## Daftar Pustaka

- [1] Aprilla Dennis. (2013). Belajar Data Mining dengan RapidMiner. *Innovation and Knowledge Management in Business Globalization: Theory & Practice, Vols 1 and 2*, 5(4), 1–5. <https://doi.org/10.1007/s13398-014-0173-7.2>
- [2] Dr. Suyanto,S.T., M. S. (2017). *DATA MINING UNTUK KLASIFIKASI DAN KLASTERISASI DATA*.
- [3] Effendi, M. R. (2015). Akurasi Data Mining Untuk Menghasilkan Pola Kelulusan Mahasiswa dengan Metode Naive Bayes, 101–106.
- [4] Ginting, S. L. B., & Trinanda, R. P. (2015). Teknik Data Mining Menggunakan Metode Bayes Classifier Untuk Optimalisasi Pencarian Pada Aplikasi Perpustakaan, 1–14.
- [5] Nusantara, B. M., Gunawan, A., & Wijaya, D. A. (2016). Strategi Pemasaran Kampung Wisata Baluwati Sebagai Destinasi Wisata Budaya di Kota Surakarta.
- [6] Rahimi, A., Sugriwan, I., & Manik, T. N. (n.d.). Optimasi Common Mode Rejection Ratio ( Cmr ) Pada Penguat Instrumentasi, 110–118.
- [7] Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Citec Journal*, 2(3), 207–217. Retrieved from [ojs.amikom.ac.id/index.php/citec/article/download/375/355](https://ojs.amikom.ac.id/index.php/citec/article/download/375/355)
- [8] Sari, M. K., Ernawati, & Wisnubhadra, I. (2015). Pembangunan Aplikasi Klasifikasi Mahasiswa Baru untuk Prediksi Hasil Studi Menggunakan Naïve Bayes Classifier. *Buana Informatika*, 7(April), 135–142.
- [9] Sugiyono, P. D. (2015). *Metode Penelitian & Pengembangan Research And Development*. Yuda Septian, N. (2013). Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro. *Jurnal Semantik* 2013, 1–11.