

# **PENERAPAN DATA MINING UNTUK MEMPREDIKSI TINGKAT KETEPATAN JUMLAH PENJUALAN PRODUK AIR MINERAL PADA PT. MARS LESTARI**

**M.Irfan Pratama Putra<sup>1</sup>, Megawaty, M.Kom.<sup>2</sup>, Muhamad Ariandi, M.Kom.<sup>3</sup>**

**Mahasiswa Bina Darma<sup>1</sup>, Dosen Bina Darma<sup>2,3</sup>**

**Jl. Jendral Ahmad Yani No.03 Palembang**

**[irfanpratama9660@gmail.com](mailto:irfanpratama9660@gmail.com)<sup>1</sup>, [megawaty@binadarma.ac.id](mailto:megawaty@binadarma.ac.id)<sup>2</sup>,**

**[muhamad\\_ariandi@binadarma.ac.id](mailto:muhamad_ariandi@binadarma.ac.id)<sup>3</sup>**

## **ABSTRACT**

Sales are a major factor in supporting survival in a company. Therefore companies must be able to determine policies related to sales activities carried out by the company. To deal with sales problems to fit the company's target, a sales forecast is needed. One method that can be used is forecasting. PT Mars Lestari located on Jl. located Sukamoro Talang Kelapa Banyuasin. PT Mars Lestari is a company that produces bottled water. One of the obstacles experienced by PT Mars Lestari is making sales reports according to the calculation of mineral water production. Where it takes 1 (one) to 2 (two) weeks in making a sales report. So that reports that arrive at the party planning and controlling production are not in accordance with the expected time. To solve the problem, a system is needed that can predict the number of sales per month so that data management of product sales is more regular Thus, forecasting (forecast) sales are needed in business management to develop the business, both in the field of services and trade. Therefore, it is expected that the application of datamining using forecasting methods can facilitate recording the amount of sales with available production and to predict the level of accuracy total sales of mineral water products at PT Mars Lestari.

***Keywords:*** Forecasting Method, Product Sales, Sales Forecasting

## **ABSTRAK**

Penjualan merupakan faktor utama dalam menunjang kelangsungan hidup dalam suatu perusahaan. Oleh karena itu perusahaan harus mampu dalam menentukan kebijakan-kebijakan yang berhubungan dengan aktivitas penjualan yang dilakukan oleh perusahaan. Untuk menangani masalah penjualan agar sesuai target perusahaan, maka diperlukan adanya sebuah peramalan penjualan. Salah satu metode yang dapat digunakan adalah forecasting. PT Mars Lestari yang berlokasi di Jl. berlokasi Sukamoro Talang Kelapa Banyuasin. PT Mars Lestari merupakan perusahaan yang memproduksi air minum kemasan. Salah satu kendala yang dialami PT Mars Lestari ialah pembuatan laporan penjualan menurut perhitungan produksi air mineral. Dimana membutuhkan

waktu 1(satu) sampai 2 (dua) minggu dalam pembuatan laporan penjualan. Sehingga laporan yang sampai pada pihak perencanaan dan pengendalian produksi tidak sesuai dengan waktu yang diharapkan. Untuk mengatasi masalah dibutuhkan sistem yang dapat memprediksi jumlah penjualan per-bulanannya agar pengelolaan data penjualan produk lebih teratur. Dengan demikian, *forecatsing* (ramalan) penjualan sangat diperlukan dalam manajemen bisnis guna mengembangkan usaha tersebut, baik bidang jasa maupun dagang.oleh karena itu, diharapkan dengan penerapan datamining menggunakan metode *forecasting* ini dapat mempermudah pencatatan jumlah penjualan dengan produksi yang tersedia dan untuk memprediksi tingkat ketepatan jumlah penjualan produk air mineral pada PT Mars Lestari.

**Kata Kunci :** Metode *Forecasting*, Penjualan produk, Peramalan Penjualan

## 1. PENDAHULUAN

Air merupakan kebutuhan vital dalam kehidupan sehari-hari terutama untuk kebutuhan minum. Kebutuhan masyarakat terhadap air minum tentu harus yang bersih dan sehat. Bisnis air minum kian mengiurkan karena kebutuhan masyarakat akan air minum terus berkembang, dan perusahaan yang menggarap bisnis makin banyak. Masing-masing produsen itu saling berebut dalam menarik perhatian konsumennya untuk meningkatkan penjualan dengan komunikasi pemasaran. Hal ini dapat dilihat dari usaha masing-masing produsen dalam menonjolkan karakteristik produknya.

Produk air minum pada saat ini wajib memenuhi Standar Nasional Indonesia (SNI). Di dalam peraturan itu disebutkan selain dimaksudkan untuk melindungi konsumen, regulasi SNI wajib tersebut juga untuk mendorong peningkatan daya saing, menciptakan persaingan usaha yang sehat, dan melestarikan fungsi lingkungan hidup.

Pemanfaatan sebuah teknologi informasi tentunya akan memberikan perubahan yang signifikan dan dampak positif tersendiri kepada

perusahaan-perusahaan besar yang sedang berkembang. Semakin besar bisnis perusahaan, maka semakin besar pula kebutuhan akan sistem. Sebagai contoh, pada perusahaan yang memiliki proses bisnis produksi barang. Dalam hal pengolahan data dari proses tersebut, sistem memerlukan kesatuan data yang terintegrasi dan tersimpan dalam media penyimpanan komputer yang dikenal dengan basis data. Dengan adanya sistem basis data, perusahaan dapat mengolah data dengan cepat dan akurat, sehingga perusahaan tidak akan mengalami kehilangan data yang penting. Perusahaan juga dapat menghemat waktu dalam melakukan pencarian data.

Perkembangan basis data sendiri di Indonesia saat ini sudah cukup baik. Basis data kini diperlukan oleh sebuah perusahaan skala kecil maupun besar untuk mempermudah pengelolaan data. Menurut Ghencea dan Gieger (2011) dalam judul "*Database Optimizing Services*", basis data dikembangkan dan disesuaikan untuk menjawab tuntutan pelanggan. Dengan basis data, kelalaian manusia akan berkurang.

Tujuan suatu usaha bisnis adalah untuk memperoleh keuntungan baik itu perusahaan dagang maupun perusahaan jasa. Selain itu, setiap perusahaan tentu saja menargetkan penjualan yang ingin dicapai setiap hari, bulan atau tahun. Perusahaan memerlukan *forecasting* (ramalan) penjualan yaitu dapat dicari dengan menggunakan data produksi untuk memperkirakan berapa jumlah penjualan yang kemungkinan terjadi di tahun yang akan datang. Dengan demikian, perusahaan dapat membuat suatu tindakan, kebijakan atau keputusan yang dilakukan secara tepat untuk mencapai target tersebut. Namun, hal itu harus diimbangi dengan upaya-upaya untuk meningkatkan penjualan produk air mineral tersebut seperti iklan yang menarik, dan *event* atau kegiatan-kegiatan yang lainnya untuk menarik minat konsumen. Penjualan meningkat maka perusahaan akan semakin berkembang.

Dengan demikian, *forecastsing* (ramalan) penjualan sangat diperlukan dalam manajemen bisnis guna mengembangkan usaha tersebut, baik bidang jasa maupun dagang. Metode-metode yang digunakan dalam meramal menjadi suatu cara bagaimana manajemen pada akhirnya harus membuat keputusan atau kebijakan yang tepat dalam mengambil suatu tindakan menurut prediksi yang berkaitan dengan data penjualan perusahaan untuk mengambil keputusan berdasarkan klasifikasinya yaitu menggunakan metode *Naive Bayesian*.

Untuk klasifikasi Bayes biasa yang lebih disebut sebagai *Naive Bayesian* dapat diasumsikan bahwa efek dari salah satu nilai atribut sebuah kelas yang diberikan secara

acak dari atribut lain. Dalam *Naive Bayesian* diasumsikan prediksi atribut Yang tidak tergantung pada kelas atau tidak dipengaruhi atribut lain, dikarenakan perhitungan yang tidak menghabiskan waktu dan ruang, kokoh terhadap atribut yang tidak relevan, hanya saja memerlukan sejumlah kecil data untuk mengestimasi parameter.

## 2. METODE PENELITIAN

### 2.1 Database

*Database* atau basis data adalah kumpulan data (elementer) dalam bentuk logic yang berkaitan untuk merepresentasikan fenomena atau fakta secara terstruktur dalam domain tertentu untuk mendukung aplikasi pada sistem tertentu (Hariyanto, 2008). Menurut Yakub (2008), *Database* dapat diartikan sebagai koleksi dari data yang terorganisasi dengan cara sedemikian rupa sehingga data tersebut mudah disimpan dan dimanipulasi.

Menurut Connolly dan Begg (2010), basis data adalah kumpulan data dan keterangannya yang terhubung secara logikal dimana dirancang untuk memenuhi kebutuhan informasi sebuah organisasi.

Ada beberapa definisi dari database yang dikenal diantaranya adalah :

- 1) Basis data terdiri atas 2 kata, yaitu Basis dan data. Basis kurang lebih dapat diartikan sebagai markas atau gudang, tempat bersarang atau berkumpul. Sedangkan Data adalah representasi fakta dunia nyata yang mewakili suatu objek seperti manusia (pegawai, siswa, pembeli,

pelanggan), barang, hewan, peristiwa, konsep, keadaan, dan sebagainya yang terekam dalam bentuk angka, huruf, simbol, teks, gambar, bunyi, atau kombinasi lainnya (Menurut Fathansyah, 2007).

2) Basis data dapat dipahami sebagai suatu kumpulan data terhubung (interrelated data) yang disimpan secara bersama-sama pada suatu media, tanpa mengatap satu sama lain atau tidak perlu suatu kerangkapan data (kalaupun ada maka kerangkapan data tersebut harus seminimal mungkin dan terkontrol (controlled redundancy), data disimpan dengan cara-cara tertentu sehingga mudah untuk digunakan atau ditampilkan kembali, data dapat digunakan satu atau lebih program-program aplikasi secara optimal, data disimpan tanpa mengalami ketergantungan dengan program yang akan menggunakannya, data disimpan dengan sedemikian rupa sehingga proses penambahan, pengembalian, dan modifikasi data dapat dilakukan dengan mudah dan terkontrol.

3) Basis data adalah kumpulan data yang saling berelasi. Data sendiri merupakan fakta mengenai obyek, orang dan lain-lain. Basis data dapat didefinisikan dalam berbagai sudut pandang seperti berikut ini:

- a. Himpunan kelompok data yang saling berhubungan yang diorganisasi sedemikian rupa sehingga kelak dapat dimanfaatkan dengan cepat dan mudah
- b. Kumpulan file/tabel/arsip yang saling berhubungan yang disimpan dalam media penyimpanan elektronik.

Setelah memahami pengertian database, tentunya kita juga harus mengetahui apa fungsi

dari database. Berikut ini adalah beberapa fungsi database:

- a. Mengelompokkan data dan informasi sehingga lebih mudah dimengerti
- b. Mencegah terjadinya duplikat data maupun inkonsistensi data
- c. Mempermudah proses penyimpanan, akses, pembaharuan, dan menghapus data.
- d. Menjaga kualitas data dan informasi yang diakses sesuai dengan yang di-input.
- e. Membantu proses penyimpanan data yang besar
- f. Membantu meningkatkan kinerja aplikasi yang membutuhkan penyimpanan data

## 2.2 Data Mining

*Data Mining*, sering juga disebut *knowledge discovery in database (KDD)*, adalah kegiatan yang meliputi pengumpulan, pemakaian dan historis untuk menemukan keteraturan, pola atau hubungan dalam suatu data berukuran besar (Santosa, 2007). Keluaran dari data mining bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. *Data mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. (Turban, dkk. 2005)

*Data mining* dapat didefinisikan sebagai proses untuk mendapatkan informasi yang berguna dari gudang

basis data yang besar yang membantu dalam pengambilan keputusan. Teknik *data mining* menelusuri data yang ada untuk membangun sebuah model dan menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam data yang tersimpan. Pengelompokan data juga dapat dilakukan menggunakan data mining untuk mengetahui pola keseluruhan data-data yang dapat diambil. (Eko Prasetyo, 2012)

*Data mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Perlu diingat bahwa kata mining sendiri berarti usaha untuk mendapatkan sedikit data berharga dari sejumlah besar data dasar. Karena itu *data mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan basis data.

### 2.3 Klasifikasi

Bagian 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 (kelas) yang telah didefinisikan sebelumnya. Klasifikasi banyak digunakan dalam berbagai aplikasi, diantaranya adalah deteksi kecurangan (*fraud detection*), pengolahan pelanggan, diagnosis medis, prediksi penjualan, dan sebagainya.

Model klasifikasi dapat dibangun berdasarkan pengetahuan seorang pakar (ahli). Namun, mengingat himpunan data yang besar, model klasifikasi lebih sering dibangun menggunakan teknik pembelajaran dalam bidang *machine learning*. Proses pembelajaran secara otomatis terhadap suatu himpunan data mampu menghasilkan model klasifikasi (fungsi target) yang memetakan objek data  $x(\text{input})$  ke salah satu kelas  $y$  yang telah didefinisikan sebelumnya. Jadi, proses pembelajaran memerlukan masukan (*input*) berupa himpunan data latih (*training set*) yang berlabel (memiliki atribut kelas) dan mengeluarkan *output* yang berupa sebuah model klasifikasi.

Terdapat banyak teknik klasifikasi yang telah diusulkan para ahli, yang dapat dikelompokkan ke dalam dua kategori: teknik klasifikasi global (memperhitungkan semua data latih) dan teknik klasifikasi lokal (hanya memperhitungkan sebagian data latih). Masing-masing teknik mempunyai kelebihan dan kekurangan, kekuatan dan kelemahan (Suyanto, 2017).

### 2.4 Knowledge Discovery in Databases

Pengertian *Data Mining* lebih spesifik menyatakan istilah *Data Mining* dan *Knowledge Discovery in Databases* (KDD) secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu kumpulan data yang besar. Akan tetapi kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain dan salah satu tahap dalam proses KDD adalah *data mining* (Han dan Kamber, 2006).

*Data mining* adalah salah satu langkah dalam proses KDD secara

keseluruhan. Secara umum, *data mining* digunakan oleh banyak peneliti sebagai sinonim dari proses KDD. Akhir-akhir ini, *data mining* dan *knowledge discovery* telah diusulkan sebagaimana yang paling memadai untuk keseluruhan proses KDD. *Knowledge Discovery in Databases* berkaitan dengan proses penemuan pengetahuan yang diterapkan pada database. Hal ini juga didefinisikan sebagai proses *non-trivial* untuk identifikasi data yang valid, baru, berpotensi bermanfaat, dan akhirnya memiliki pola yang dapat dimengerti. (Kurgan dan Musilek, 2006).

KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah kumpulan data. *Knowledge discovery in databases* (KDD) adalah keseluruhan proses *non-trivial* untuk mencari dan mengidentifikasi pola (pattern) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti.

## 2.5 Tahapan Data Mining

Menurut Vlandri (2017) menjelaskan bahwa data mining merupakan salah satu dari rangkaian *Knowledge Discovery in Database* (KDD). KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah data. Serangkaian proses tersebut memiliki tahap sebagai berikut (Tan, 2004):

### a. *Data Cleaning*

Ini adalah sebuah fase dimana data-data tidak lengkap, mengandung error dan tidak konsisten dibuang dari koleksi data, sehingga data yang telah bersih dan relevan dapat

digunakan untuk diproses ulang untuk penggalian pengetahuan (*discovery knowledge*).

### b. *Data Integration*

Merupakan data yang diperoleh akan digabungkan menjadi satu tabel data yang besar sebagai fitur pelatihan nantinya untuk menjalankan proses algoritma k-means.

### c. *Data Selection*

Pada langkah ini, data yang relevan terhadap analisis dapat dipilih dan diterima dari koleksi data yang ada. Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

### d. *Data Transformation*

Juga dikenal sebagai data *consolidation*. Pada tahap ini, dimana data-data yang telah terpilih, ditransformasikan ke dalam bentuk-bentuk yang cocok untuk prosedur penggali (*mining procedure*) dengan cara melakukan normalisasi dan agregasi data.

### e. *Data Mining*

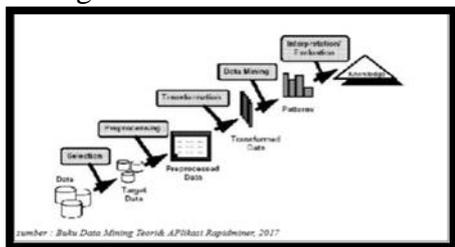
Tahap ini adalah tahap yang paling penting, dengan menggunakan teknik-teknik yang diaplikasikan untuk mengekstrak pola-pola potensial yang berguna. Data mining ialah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

### f. *Pattern Evaluation*

Pada tahap ini, pola-pola menarik dengan jelas merepresentasikan pengetahuan telah diidentifikasi berdasarkan measure yang telah diberikan.

g. *Knowledge Presentation*

Ini merupakan tahap terakhir, dimana pengetahuan yang telah ditemukan secara visual ditampilkan kepada user. Tahap ini menggunakan teknik visualisasi untuk membantu user dalam mengerti dan menginterpretasikan hasil dari data mining.



Gambar 2.1 Tahapan pada data mining

2.6 Forecasting

*Forecasting* atau perkiraan adalah kegiatan yang bertujuan untuk meramalkan atau memprediksi segala hal yang terkait dengan produksi, penawaran, permintaan, dan penggunaan teknologi dalam sebuah industri atau usaha. Perkiraan ini pada akhirnya akan digunakan oleh perusahaan maupun pihak manajemen operasional untuk membuat perencanaan terkait kegiatan usaha dalam beberapa periode tertentu.

Fungsi perkiraan atau *forecasting* terlihat pada saat pengambilan keputusan. Keputusan yang baik adalah keputusan yang didasarkan atas pertimbangan apa yang akan terjadi pada waktu keputusan itu dilaksanakan perkiraan atau *forecasting* memiliki tujuan sebagai berikut:

a. Mengkaji kebijakan perusahaan yang berlaku saat ini dan di masa

lalu, serta melihat sejauh mana pengaruh di masa datang.

b. Perkiraan diperlukan karena adanya *time lag* atau *delay* antara saat suatu kebijakan perusahaan ditetapkan dengan saat implementasi.

c. Perkiraan merupakan dasar penyusunan bisnis pada suatu perusahaan sehingga dapat meningkatkan efektivitas suatu rencana bisnis.

2.7 Naive Bayes

*Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Menurut Olson dan Delen (2008) menjelaskan *Naive Bayesian* untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi obyek. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari "master" tabel keputusan.

*Naive Bayesian* 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. Teorema tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antar atribut independen.

Keuntungan penggunaan algoritma *naive bayes* adalah bahwa metoda ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter yg diperlukan dalam

proses pengklasifikasian. Karena yg diasumsikan sebagai variabel independent, maka hanya varians dari suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians. Tahapan dari proses algoritma Naive Bayes adalah:

1. Menghitung jumlah kelas / label
2. Menghitung Jumlah Kasus Per Kelas
3. Kalikan Semua Variable Kelas
4. Bandingkan Hasil Per Kelas

### 3. ANALISIS DATA MINING

#### 3.1.1 Data Cleaning

Pembersihan data sebenarnya merupakan tahap awal dari proses *KDD*. merupakan proses menghilangkan *noise* yang ditemukan berupa *missing value*. Dikatakan *missing value* jika atribut-atribut dalam *dataset* tidak berisi nilai atau kosong. Pembersihan data juga akan mempengaruhi performansi dari teknik *data mining* karena data yang akan digunakan akan berkurang jumlah dan kompleksitasnya. Pembersihan data terhadap *noise* yang ditemukan berupa *missing value*, *inkonsisten data*, dan *redundant data*.

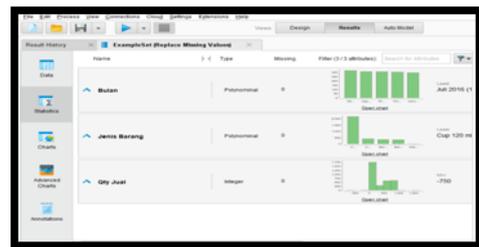
Sebelum melanjutkan ke dalam data mining terlebih dahulu kita harus melakukan langkah-langkah yang berguna untuk meningkatkan kualitas hasil dari algoritma selanjutnya. Proses data mining yang pertama adalah *cleaning* data yaitu proses pembersihan data yang diawali dari prosedur pengambilan data hingga alur kerja dan proses pembersihan data yang dilakukan menggunakan *tools rapidminer*.

Data *Cleaning*, juga dikenal sebagai data *cleansing*. Ini adalah sebuah fase dimana data-data tidak lengkap, mengandung error, dan tidak konsisten dibuang dari koleksi data, sehingga data yang telah bersih dan relevan dapat digunakan untuk diproses ulang untuk penggalian pengetahuan (*Discovery Knowledge*).

No	Nama	Tipe	Missing
1	Batam	Pharmaceutical	0
2	Batam	Pharmaceutical	0
3	Batam	Pharmaceutical	0
4	Batam	Pharmaceutical	0
5	Batam	Pharmaceutical	0
6	Batam	Pharmaceutical	0
7	Batam	Pharmaceutical	0
8	Batam	Pharmaceutical	0
9	Batam	Pharmaceutical	0
10	Batam	Pharmaceutical	0
11	Batam	Pharmaceutical	0
12	Batam	Pharmaceutical	0
13	Batam	Pharmaceutical	0
14	Batam	Pharmaceutical	0
15	Batam	Pharmaceutical	0
16	Batam	Pharmaceutical	0
17	Batam	Pharmaceutical	0
18	Batam	Pharmaceutical	0
19	Batam	Pharmaceutical	0
20	Batam	Pharmaceutical	0

Gambar 3.1 hasil pembersihan data kosong

Pada gambar 3.13 yaitu menunjukkan hasil dari pembersihan data kosong menggunakan *Rapidminer*.



Gambar 3.2 hasil pembersihan data yang tidak konsisten

Pada gambar 3.14 yaitu menunjukkan hasil dari pembersihan data kosong dan data yang tidak konsisten menggunakan *Rapidminer*.

#### 3.1.2 Data Integration

Pada tahap data mining selanjutnya setelah dilakukan pembersihan data yaitu integrasi data dimana data yang berulang, file yang berulang akan digabungkan dan data tersebut akan dihapuskan. Hasil dari penghilangan data yang telah dilakukan dapat dilihat pada gambar 3.15 berikut ini.

Date	Sales	Product	City
2018-01-01	100	Produk A	A
2018-01-02	150	Produk B	B
2018-01-03	200	Produk C	C
2018-01-04	250	Produk D	D
2018-01-05	300	Produk E	E
2018-01-06	350	Produk F	F
2018-01-07	400	Produk G	G
2018-01-08	450	Produk H	H
2018-01-09	500	Produk I	I
2018-01-10	550	Produk J	J
2018-01-11	600	Produk K	K
2018-01-12	650	Produk L	L
2018-01-13	700	Produk M	M
2018-01-14	750	Produk N	N
2018-01-15	800	Produk O	O
2018-01-16	850	Produk P	P
2018-01-17	900	Produk Q	Q
2018-01-18	950	Produk R	R
2018-01-19	1000	Produk S	S
2018-01-20	1050	Produk T	T
2018-01-21	1100	Produk U	U
2018-01-22	1150	Produk V	V
2018-01-23	1200	Produk W	W
2018-01-24	1250	Produk X	X
2018-01-25	1300	Produk Y	Y

Gambar 3.3 hasil penghilangan data duplikat

### 3.1.3 Data Selection

Data Selection merupakan proses seleksi atau memilih atribut yang akan digunakan dalam proses data mining, karena tidak semua atribut pada sumber data dapat digunakan seluruhnya. Data yang digunakan dalam penelitian ini bersumber dari data penjualan harian di PT. Mars Lestari Palembang. Pada struktur tabel data penjualan ini awalnya memiliki 25 atribut yang kemudian belum diseleksi, dan hanya digunakan 3 atribut.

Selection data berfungsi untuk memilih field yang sesuai kebutuhan untuk mendapatkan knowledge, pada penelitian saat ini field data yang sesuai untuk mendapatkan hasil yang diinginkan dan sesuai dengan prosesnya adalah jenis barang dan jenis pengiriman. Pada gambar 3.16 adalah hasil seleksi.

Date	Sales	Product	City
2018-01-01	100	Produk A	A
2018-01-02	150	Produk B	B
2018-01-03	200	Produk C	C
2018-01-04	250	Produk D	D
2018-01-05	300	Produk E	E
2018-01-06	350	Produk F	F
2018-01-07	400	Produk G	G
2018-01-08	450	Produk H	H
2018-01-09	500	Produk I	I
2018-01-10	550	Produk J	J
2018-01-11	600	Produk K	K
2018-01-12	650	Produk L	L
2018-01-13	700	Produk M	M
2018-01-14	750	Produk N	N
2018-01-15	800	Produk O	O
2018-01-16	850	Produk P	P
2018-01-17	900	Produk Q	Q
2018-01-18	950	Produk R	R
2018-01-19	1000	Produk S	S
2018-01-20	1050	Produk T	T
2018-01-21	1100	Produk U	U
2018-01-22	1150	Produk V	V
2018-01-23	1200	Produk W	W
2018-01-24	1250	Produk X	X
2018-01-25	1300	Produk Y	Y

Gambar 3.4 hasil seleksi data

### 3.1.4 Data Transformation

Data Transformation juga dikenal sebagai data consolidation. Pada tahap ini, dimana data-data yang

telah terpilih, ditransformasikan kedalam bentuk-bentuk yang cocok untuk prosedur penggalian (mining procedure) dengan cara melakukan normalisasi dan agregasi data. Hasil dari normalisasi dan agregasi data dapat dilihat pada gambar 3.17

Date	Sales	Product	City
2018-01-01	100	Produk A	A
2018-01-02	150	Produk B	B
2018-01-03	200	Produk C	C
2018-01-04	250	Produk D	D
2018-01-05	300	Produk E	E
2018-01-06	350	Produk F	F
2018-01-07	400	Produk G	G
2018-01-08	450	Produk H	H
2018-01-09	500	Produk I	I
2018-01-10	550	Produk J	J
2018-01-11	600	Produk K	K
2018-01-12	650	Produk L	L
2018-01-13	700	Produk M	M
2018-01-14	750	Produk N	N
2018-01-15	800	Produk O	O
2018-01-16	850	Produk P	P
2018-01-17	900	Produk Q	Q
2018-01-18	950	Produk R	R
2018-01-19	1000	Produk S	S
2018-01-20	1050	Produk T	T
2018-01-21	1100	Produk U	U
2018-01-22	1150	Produk V	V
2018-01-23	1200	Produk W	W
2018-01-24	1250	Produk X	X
2018-01-25	1300	Produk Y	Y

Gambar 3.5 hasil normalisasi agregasi

Pada gambar 3.17 tersebut terdapat outlier atau salah satu teknik agregasi yang terdapat hasil false dan true. Nilai false adalah data yang normal, sedangkan nilai true adalah data unnormal maka harus dihapuskan. Hasil dari transformasi data tersebut dapat dilihat pada gambar 3.18

Date	Sales	Product	City
2018-01-01	100	Produk A	A
2018-01-02	150	Produk B	B
2018-01-03	200	Produk C	C
2018-01-04	250	Produk D	D
2018-01-05	300	Produk E	E
2018-01-06	350	Produk F	F
2018-01-07	400	Produk G	G
2018-01-08	450	Produk H	H
2018-01-09	500	Produk I	I
2018-01-10	550	Produk J	J
2018-01-11	600	Produk K	K
2018-01-12	650	Produk L	L
2018-01-13	700	Produk M	M
2018-01-14	750	Produk N	N
2018-01-15	800	Produk O	O
2018-01-16	850	Produk P	P
2018-01-17	900	Produk Q	Q
2018-01-18	950	Produk R	R
2018-01-19	1000	Produk S	S
2018-01-20	1050	Produk T	T
2018-01-21	1100	Produk U	U
2018-01-22	1150	Produk V	V
2018-01-23	1200	Produk W	W
2018-01-24	1250	Produk X	X
2018-01-25	1300	Produk Y	Y

Gambar 3.6 hasil transformasi data

## 4. HASIL DAN ANALISIS

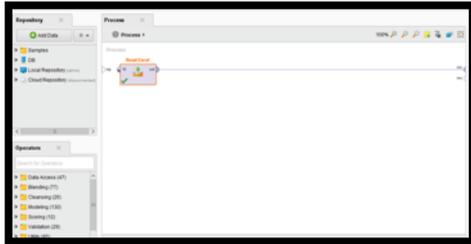
### 4.1 Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam sebuah data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma data mining sangat bervariasi. Pemilihan metode atau algoritma sangat bergantung pada tujuan dan proses Knowledge Discovery in Database (KDD). Pada penelitian ini penerapan data mining

menggunakan teknik *classification* dan algoritma *naïve bayes* untuk mengetahui produk yang banyak terjual di PT.Mars Lestari Palembang.

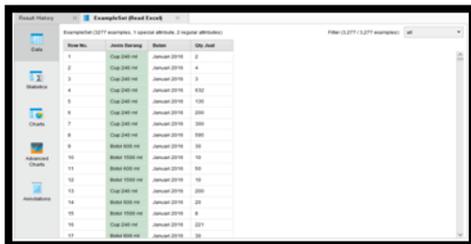
## 4.2 Implementasi Dengan Rapidminer

Berikut ini adalah pengolahan data dengan menggunakan *naïve bayes* pada *Rapidminer*, langkah pertama yaitu memasukkan data yang akan kita olah di *Rapidminer*. Seperti yang ditunjukkan pada gambar 4.1



Gambar 4.1 memasukkan data ke *Rapidminer*

Pilih *ReadExcel* dibagian operator, gunanya untuk membaca data yang penulis gunakan, karena data yang penulis gunakan itu berupa excel.



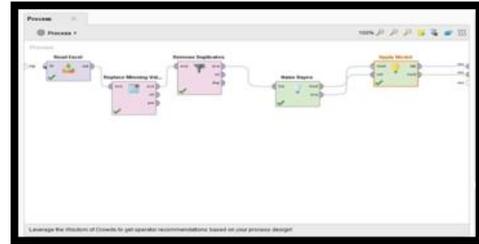
Gambar 4.2 hasil dari data yang telah dikeluarkan oleh *Rapidminer*

Pada gambar 4.2 yaitu data hasil keluaran data oleh *Rapidminer* yang berjumlah 3.277 data.

## 4.3 Pemodelan (Modelling)

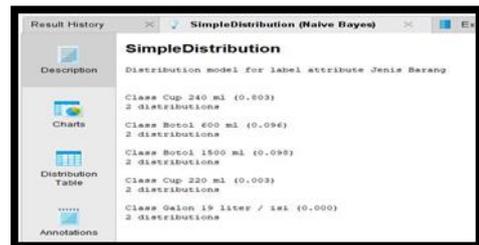
Pemodelan adalah fase yang secara langsung melibatkan teknik *data mining* yaitu dengan melakukan

pemilihan teknik *data mining* dan menentukan algoritma yang digunakan.



Gambar 4.3 Proses *Rapidminer*

Pada gambar 4.3 merupakan suatu proses pada *tools Rapidminer*, dimana terdapat proses memasukkan data, menghapus data yang duplikat atau data yang kosong atau yang *noise*, lalu memasukkan algoritma *naïve bayes* seperti yang sudah di jelaskan diatas.



Gambar 4.4 Simple Distribution

Model distribusi untuk label atribut kelas estimasi adalah sebagai berikut :

Kelas Cup 240 ml	:	2
<i>distribution</i>		
Kelas Botol 600 ml	:	2
<i>distribution</i>		
Kelas Botol 1500 ml	:	2
<i>distribution</i>		
Kelas Cup 220 ml	:	2
<i>distribution</i>		
Kelas Galon 19 liter/isi	:	2
<i>distribution</i>		

Percobaan pada penelitian ini menggunakan *Rapidminer* 8.1.000, algoritma yang digunakan *naïve*

bayes. *Performance(Classification)* untuk mengukur performa dari model *naïve bayes* tersebut dan untuk testing menggunakan *apply model*

	Real Cup 240 ml	Real Botol 600 ml	Real Botol 1500 ml	Real Cup 220 ml	Real Galon 19 liter	class prediction
pred Cup 240 ml	2218	00	00	3	1	84.71%
pred Botol 600 ml	242	207	204	2	1	31.05%
pred Botol 1500 ml	118	00	00	0	1	28.57%
pred Cup 220 ml	0	0	0	0	0	0.00%
pred Galon 19 liter	0	0	0	0	0	0.00%
class total	86.84%	18.81%	22.00%	0.00%	0.00%	

**Gambar 4.5** *Performance Vector(performance)*

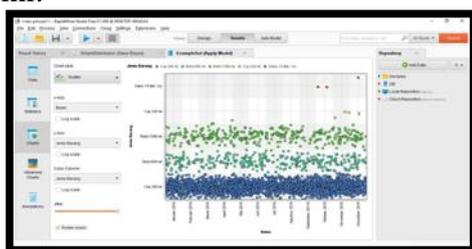
Hasil akurasi model *naïve bayes* menunjukkan tingkat akurasinya 76.44% artinya model klasifikasi penjualan produk menggunakan *naïve bayes* terbukti cukup baik, hal ini dilihat dari tingkat akurasinya yang mencapai 76.44%.

#### 4.4 Evaluasi (*Evaluation*)

Evaluasi adalah fase lanjutan terhadap tujuan *data mining*. Evaluasi dilakukan secara mendalam dengan tujuan agar hasil pada tahap pemodelan sesuai dengan sasaran yang ingin dicapai dalam tahap *business understanding*.

##### 4.4.1 Evaluasi Hasil (*Evaluation Result*)

Tahap ini menilai sejauh mana hasil pemodelan *data mining* memenuhi tujuan yang telah ditentukan pada tahapan *business understanding*. Dalam melihat pola klasifikasi keluaran *rapidminer* dapat dilihat pada gambar dibawah ini:



**Gambar 4.6** Pola *naïve bayes* keluaran *rapidminer*

Dapat dilihat dari gambar 4.6 diatas, bahwa penjualan pada satu tahun terakhir yang paling banyak yaitu di pola berwarna biru atau cup 240ml, sedangkan yang yang paling sedikit penjualannya yaitu galon 19 liter/isi.

#### 4.5 Persentation

Ini merupakan tahapan terakhir, dimana pengetahuan yang telah ditemukan secara visual ditampilkan kepada user. Hasil klasifikasi menggunakan *naïve bayes*,Maka didapat hasil pengetahuan yang bisa dilihat pada gambar dibawah ini:



**Gambar 4.7**Hasil

Dari hasil gambar 4.1 menunjukkan grafik penjualan pada data tahun 2016, dapat dilihat jumlah tingkat penjualan yang paling tinggi itu terdapat pada bulan September 2016 dan bulan Mei 2016, sedangkan jumlah tingkat penjualan yang paling rendah itu terdapat pada bulan Desember 2016.

### 5. PENUTUP

#### 5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan oleh peneliti, maka dapat ditarik kesimpulan sebagai berikut

1. Berdasarkan perhitungan data mining menggunakan algoritma *naïve bayes*, dapat ditarik kesimpulan bahwa jumlah tingkat penjualan air minum mineral pada PT.Mars Lestari pada tahun 2016 lebih banyak

terjual untuk kemasan air mineral cup 240ml, sedangkan untuk jumlah tingkat penjualan air mineral yang paling rendah pada tahun 2016 yaitu galon 19 liter/isi.

2. Dari hasil obeservasi terhadap data backsheet PT.Mars Lestari tahun 2016 melalui proses pengklasifikasian naïve bayes didapatkan jumlah tingkat penjualan air mineral dalam ukuran cup 240ml yang paling banyak atau paling diminati oleh masyarakat.
3. Tools RapidMiner berhasil menentukan jumlah tingkat penjualan air mineral kemasan yang paling banyak diminati atau paling banyak terjual untuk data penjualan tahun 2016 dengan akurasi 76,44%. Dari jumlah seluruhnya 3.277 data.

## 5.2 Saran

Setelah melakukan penelitian mengenai penerapan algoritma naïve bayes untuk mengklasifikasi jumlah tingkat penjualan air mineral kemasan pada PT.Mars Lestari Palembang, penulis memiliki saran sebagai berikut:

1. Dengan adanya penelitian ini diharapkan dapat dimanfaatkan untuk menjadi acuan PT.Mars Lestari Palembang untuk bisa melihat tingkat penjualan air mineral kemasan ukuran berapa saja yang bnyak terjual pada tahun 2016.
2. Penelitian ini dapat digunakan untuk dikembangkan dalam melakukan penelitian-penelitian selanjutnya.
3. Bagi para staff Marketing produk Air Minum Dalam Kemasan (AMDK) PT. Mars

Lestari dalam meningkatkan aktivitas komunikasi pemasaran selalu berusaha untuk tetap mempertahankan dan meningkatkan kualitas pelayanan yang ditawarkan dengan melakukan peningkatan dalam hal pengetahuan produk, serta kreativitas dalam melakukan aktivitas pemasaran.

4. Bagi pihak manajemen produk AMDK, agar dapat menambah jumlah para staff Marketing yang berpengalaman, agar proses penjualan dapat lebih meningkat.
5. Bagi pihak manajemen produk AMDK, agar kiranya menambah peningkatan sistem mempromosikan produk denan menggunakan media teknologi untuk mempercepat strategi komunikasi pemasaran dikalangan masyarakat luas

## DAFTAR PUSTAKA

- Efrain,Turban, Dkk. 2005. *Decision Support System And Intelligent System*. Yogyakarta:ANDI.
- Han, Jiawei Dan Micheline Kamber. 2006. *Data Mining:Concept And Techniques*. Morgan Kaufmann Publishers.
- Hariyanto, Bambang. 2004. *Sistem Manajemen Basis Data*. Bandung:Informatika.
- Nazir, Moh. 2014. *Metode Penelitian*. Bogor:Ghalia Indonesia.
- Olson, David Dan Delen Dursun. 2008. *Advanced Data Mining Techiques*.Springer.
- Rumengan, Jemmy. 2013. *Metodelogi Penelitian Dengan SPSS*. Batam:UNIBA.

Santosa,Budi. 2007. *Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Yogyakarta:Graha Ilmu.

Usman Dan Akbar. 2006. *Pengantar Statistika*. Jakarta:Bumi Aksara.

Yakub. 2008. *Sistem Basis Data Tutorial Konseptual*. Yogyakarta:Graha Ilmu.

Farihin,Muhammad.  
<https://mti.binus.ac.id/2017/11/24/klasifikasi/>

<https://www.maxmanroe.com/vid/teknologi/komputer/pengertian-database.html>

<https://www.kajianpustaka.com/2017/09/data-mining.html>

<https://informatikalogi.com/algorithm-naive-bayes/>

(<http://elib.unikom.ac.id/sriastutip-unikom>,23 April 2018).