

PREDIKSI POTENSI PENJUALAN PRODUK *DELIFRANCE* DENGAN METODE *NAIVE BAYES* DI PT. PANGAN LESTARI

Karsito¹⁾, Winda Monika Sari²⁾

Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa
karsitoawb@pelitabangsa.ac.id

Disetujui, 26 September 2018

Abstraksi

PT. Pangan Lestari adalah perusahaan pemasok, dalam kesehariannya memiliki permasalahan yaitu dalam memprediksi barang yang laku berdasarkan hasil penjualan. Penggunaan data mining, dalam hal ini adalah berupa Metode Naïve Bayes sangat membantu dalam prediksi berapa stok yang disiapkan dan produk apa saja yang tidak laku. Dengan adanya bantuan alat aplikasi ini diharapkan penjualan akan meningkat karena konsumen senang belanja di sini karena barang yang akan di beli selalu tersedia, jenis variannya, dan selalu baru, belum kadaluarsa.

Kata kunci: Penjualan, *Metode Naive Bayes*, *data mining*

Abstract

PT. Pangan Lestari is a supplier company in the form of a franchise, in its daily life it has problems, namely in predicting goods that will sell based on sales results. The use of data mining, in this case is the Naïve Bayes method, is very helpful in predicting how many stocks are prepared and what products are not selling. With the help of this application tool, it is hoped that sales will increase because consumers are happy to shop here because the goods to be purchased are always available, the types of variants, and are always new, have not expired.

Keywords: Sales, *Naive Bayes Method*, *data mining*

1. Pendahuluan

Dewasa ini teknologi informasi menjadi suatu elemen yang penting dalam perkembangan dunia bisnis. Penggalan suatu informasi data dan informasi sangat dibutuhkan suatu perusahaan dan lembaga baik yang berskala besar, sedang, dan kecil. Teknologi informasi terkomputerisasi terbukti mempercepat kinerja, sehingga meningkatkan keuntungan dengan kemajuan teknologi informasi, pengaksesan terhadap data atau informasi yang tersedia dapat berlangsung dengan cepat, efisien serta akurat.

PT Pangan Lestari adalah perusahaan pemasok daging segar ke seluruh pelosok di Indonesia, berlokasi di Cikarang bekasi, memiliki jaringan besar di cabang-cabag di seluruh Indonesia, problema yang dihadapi adalah banyaknya pesanan dari toko-toka agen daging beku dan juga pesanan online. Jumlah data yang begitu besar bisa menjadi masalah bagi perusahaan jika tidak bisa diolah sebaik mungkin. Data mining merupakan proses penggalan data menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data besar. Pada umumnya perusahaan mengumpulkan informasi melalui sistem database yang berguna untuk menampung data transaksi, kemudian nantinya data tersebut diolah sehingga dapat diketahui tingkat dan volume suatu penjualan pada waktu tertentu dan sebagainya.

Permintaan barang oleh customer selanjutnya dilakukan penambahan sesuai keinginan direktur dalam hal prediksi penjualan, hal tersebut menyebabkan kelebihan atau kekurangan barang. Barang yang ditawarkan perusahaan termasuk kedalam jenis *food*, maka penyimpanan barang di gudang tidak dapa terlalu lama. Kelebihan pemesanan barang dari permintaan dapat menyebabkan kerusakan pada barang. Perusahaan juga dapat mengalami kerugian akibat tidak terpenuhinya permintaan dari customer terhadap barang yang diinginkan. Berdasarkan permasalahan tersebut diperlukan sebuah pengelolaan data yang memungkinkan perusahaan untuk melakukan prediksi penjualan terhadap barang baru.

2. Tinjauan Studi

2.1. Tinjauan Pustaka

Pada tinjauan pustaka disini, di tampilkan tiga penelitian pendahuluan yang memiliki konsep penelitian yang sama, yaitu Penelitian pertama, dilakukan oleh Andini Tagshya Izmi (2016) yaitu penelitian potensi pemasaran produk baru dalam menghasilkan area pemasaran penjualan produk baru dengan menggunakan Metode *Naive BayesClassifier*.

Penelitian kedua, penelitian yang dilakukan oleh Saleh, (2015) yang meneliti tentang keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training

Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naïve Bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan.

Penelitian ketiga, penelitian yang dilakukan oleh Manalu (2017) dengan Metode *Naïve Bayes* untuk mengklasifikasikan *probabilistik* sederhana yang menghitung sekumpulan *probabilitas* dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan, dan dapat memprediksi jumlah produksi barang berdasarkan data persediaan dan pemesanan pada CV.Papa dan Mama Pastries sehingga dapat membantu dalam menentukan persediaan stok sesuai dengan kebutuhan.

Dari banyaknya penelitian diatas, menunjukkan penggunaan algoritma *naïve bayes* masih efektif dalam mengklasifikasi dan menentukan akurasi.

2.2. Tinjauan Studi

2.2.1. Data Mining

Menurut Chairul Fadlan, Selfia Ningsih, Agus Perdana Windarto (2015) *data mining* merupakan proses ataupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi-informasi yang nantinya dapat digunakan. Sedangkan menurut Aziz, Hadi, & Moechammad, (2015) *data mining* adalah suatu proses yang menggunakan teknik statistik, matematika, kecerdasan tiruan, dan *machine-learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Istilah *data mining* memiliki hakikat sebagai disiplin ilmu yang tujuannya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki.

Selain itu *Data mining* merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. Ada beberapa teknik yang dimiliki *data mining* berdasarkan tugas yang bisa dilakukan, yaitu: deskripsi, estimasi, prediksi, klasifikasi, *clustering* dan asosiasi. Penelitian ini akan menggunakan fungsi klasifikasi dikarenakan sistem mengelompokkan data didasarkan pada karakteristik alternatif. Dalam klasifikasi variabel, tujuan bersifat kategorik. Istilah *data mining* dan *knowledge discovery in databases* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. KDD dapat dibagi menjadi beberapa tahap proses, yaitu : pembersihan data (*data cleaning*), integrasi data (*data integration*), seleksi data (*data selection*), transformasi data (*data transformation*), proses *data mining*, evaluasi pola (*pattern evaluation*), dan presentasi pengetahuan (*knowledge presentation*).

2.2.2. Metode Algoritma Naive Bayes

Naive Bayes merupakan teknik prediksi berbasis probabilitas sederhana yang berdasarkan pada penerapan teorema Bayes dengan asumsi independensi yang kuat. Dengan kata lain, dalam Naive Bayes menggunakan model fitur independen, maksud independen yang kuat pada fitur adalah bahwa data tidak berkaitan dengan data yang lain dalam kasus yang sama ataupun atribut yang lain.

Persamaan dari Teorema Bayes adalah :

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

Keterangan :

X : Data sampel dengan kelas (label) yang tidak diketahui.

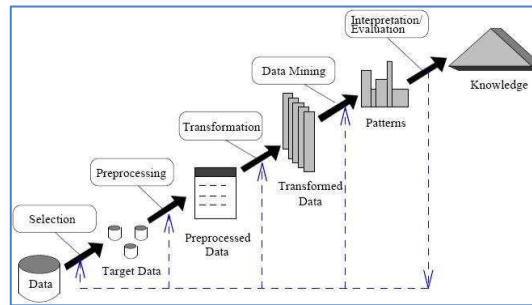
H : Hipotesa bahwa X adalah data dengan kelas (label).

$P(H|X)$: Probabilitas H berdasarkan kondisi X. $P(H)$ = Peluang dari hipotesa H.

$P(X|H)$: Peluang data sampel X, bila diasumsikan bahwa hipotesa benar. $P(X)$ = Peluang data sampel yang diamati.

2.2.3. Knowledge Discovery in Databases (KDD)

Knowledge Discovery in Databases (KDD) merupakan sekumpulan proses untuk menemukan pengetahuan yang bermanfaat dari data. KDD terdiri dari serangkaian langkah perubahan, termasuk data preprocessing dan juga post processing. *Data preprocessing* merupakan langkah untuk mengubah data mentah menjadi format yang sesuai untuk tahap analisis berikutnya. Selain itu *data preprocessing* juga digunakan untuk membantu dalam pengenalan atribut dan data segmen yang relevan dengan tugas *data mining*. Istilah *Data mining* dan *Knowledge Discovery in Databases* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah *Data mining*. Proses *Knowledge Discovery in Databases* (KDD) secara garis besar dapat dijelaskan sebagai berikut.

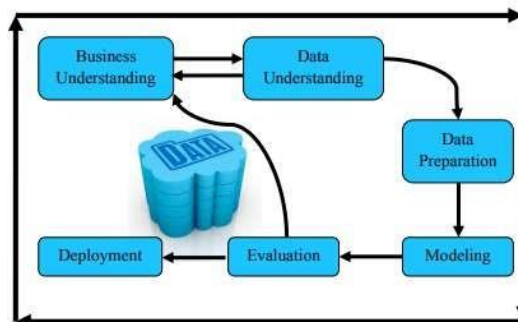


Gambar 1. Proses *Knowledge Discovery in Databases* (KDD)

1. *Data Selection*
 Pemilihan atau seleksi data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *knowledge data discovery* dimulai. Data hasil seleksi yang akan digunakan untuk proses *Data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional. 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). Juga dilakukan proses *enrichment*, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan KDD, seperti data atau informasi eksternal.
2. *Transformation*
Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.
3. *Data mining*
Data mining adalah 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.
4. *Interpretation / Evaluation*
 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 disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.2.4. CRISP-DM

CRISP-DM (*Cross-Industry Standard Process for Data mining*) adalah suatu konsorsium perusahaan yang didirikan oleh Komisi Eropa pada tahun 1996. CRISP-DM telah ditetapkan sebagai proses standar dalam *data mining* yang dapat diaplikasikan di berbagai sektor industri.



Gambar 2. Siklus Pengembangan *Data Mining* menggunakan CRISP_DM

Berikut ini adalah enam tahap siklus hidup pengembangan *datamining* :

1. *Business Understanding* (Fase Pemahaman Bisnis)
 Tahap pertama yaitu memahami tujuan serta kebutuhan dari sudut pandang bisnis, kemudian menterjemahkan pengetahuan ini ke dalam pendefinisian masalah dalam *data mining*. Selanjutnya akan ditentukan rencana dan strategi untuk mencapai tujuan tersebut.
2. *Data Understanding* (Fase Pemahaman Data)
 Tahap ini dimulai dari pengumpulan data kemudian akan dilanjutkan dengan proses untuk mendapatkan pemahaman yang mendalam tentang data, mengidentifikasi masalah kualitas

- data, atau untuk mendeteksi adanya bagian yang menarik dari data yang dapat digunakan untuk hipotesa untuk informasi yang tersembunyi.
3. *DataPreparation* (Fase Pengolahan Data)
Tahap ini mencakup semua kegiatan untuk membangun datasetakhir (data yang akan diproses pada tahap pemodelan/modeling) dari data mentah. Tahap ini dapat diulang beberapa kali. Padatahap ini juga mencakup pemilihan tabel, record, dan atribut- atribut data, termasuk proses pembersihan dan transformasi data untuk kemudian dijadikan masukan dalam tahap pemodelan(modeling).
 4. *Modeling* (Fase Pemodelan)
Pada tahap ini dilakukan pemilihan dan penerapan berbagai teknik pemodelan serta beberapa parameternya akan disesuaikan untuk mendapatkan nilai yang optimal. Secara khusus, ada beberapa teknik berbeda yang dapat diterapkan untuk masalah *datamining* yang sama. Di pihak lain ada teknik pemodelan yang membutuhkan format data khusus. Sehingga pada tahap ini masih memungkinkan kembali ke tahap sebelumnya.
 5. *Evaluation* (Fase Evaluasi)
Pada tahap ini, model sudah terbentuk serta diharapkan memiliki kualitas baik jika dilihat dari sudut pandang analisa data. Pada tahap ini akan dilakukan evaluasi terhadap keefektifan dan kualitas model sebelum digunakan dan menentukan apakah model dapat mencapai tujuan yang ditetapkan pada fase awal (*BusinessUnderstanding*). Kunci dari tahap ini adalah menentukan apakah ada masalah bisnis yang belum dipertimbangkan. Di akhir dari tahap ini harus ditentukan penggunaan hasil proses *data mining*.
 6. *Deployment* (Fase Penyebaran)
Pada tahap ini, pengetahuan atau informasi yang telah diperoleh diatur dan dipresentasikan dalam bentuk khusus sehingga dapat digunakan oleh pengguna. Tahap deployment dapat berupa pembuatan laporan sederhana atau mengimplementasikan proses *data mining* yang berulang dalam perusahaan. Dalam banyak kasus, tahap deployment melibatkan konsumen, di samping analis data, karena sangat penting bagi konsumen untuk memahami tindakan apa yang harus dilakukan untuk menggunakan model yang telah dibuat.

2.2.5. RapidMiner



Gambar 3. Logo RapidMiner

RapidMiner merupakan suatu perangkat lunak yang bertujuan untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi yang bersifat terbuka (*open source*) dan menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. RapidMiner ditulis menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi. RapidMiner memiliki kurang lebih 500 operator *data mining*, termasuk operator untuk *input*, *output*, *data preprocessing* dan visualisasi. RapidMiner merupakan *software* yang berdiri sendiri untuk analisis data sebagai mesin *data mining* yang dapat diintegrasikan pada produknya sendiri.

RapidMiner sebelumnya bernama YALE (*Yet Another Learning Environment*), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh RalfKlinkenberg, Ingo Mierswa, dan SimonFischer di *Artificial Intelligence Unit* dari *University of Dortmund*. RapidMiner didistribusikan di bawah lisensi AGPL (*GNU Affero General Public License*) versi 3. Hingga saat ini telah ribuan aplikasi yang dikembangkan menggunakan RapidMiner di lebih dari 40 negara. RapidMiner sebagai *software open source* untuk *data mining* tidak perlu diragukan lagi karena *software* ini sudah terkemuka didunia.

RapidMiner menempati peringkat pertama sebagai *software data mining* pada polling oleh KDnuggets, sebuah portal *data mining* pada tahun 2010-2011. RapidMiner menyediakan GUI (*Graphic User Interface*) untuk merancang sebuah *pipeline* analitis. GUI ini akan menghasilkan *fileXML* (*Extensible Markup Language*) yang mendefinisikan proses analitis keinginan pengguna untuk diterapkan ke data. File ini kemudian dibaca oleh RapidMiner untuk menjalankan analisis secara otomatis.

RapidMiner memiliki beberapa sifat sebagai berikut.

1. Ditulis dengan bahasa pemrograman java sehingga dapat dijalankan di berbagai sistem operasi.

2. Representasi XML internal untuk memastikan format standar pertukaran data.
3. Bahasa *scripting* memungkinkan untuk eksperimen skalabesar dan otomatisasi eksperimen.
4. Konsep *multi-layer* untuk menjamin tampilan data yang efisien dan menjamin penanganan data.

2.2.6. Penjualan

Penjualan adalah aktivitas maupun kegiatan yang dilakukan manusia yang saling menguntungkan satu sama lain, dimana penjual menawarkan produk yang dimiliki pada konsumen sehingga mampu menarik konsumen untuk mengeluarkan uang guna membeli suatu produk yang telah ditawarkan oleh penjual. Dalam perusahaan penjualan adalah salah satu aktivitas rutin yang dijalani untuk memperoleh laba dan membuat perusahaan berkembang.

Berikut Pengertian penjualan menurut beberapa ahli antara lain :

- a. Menurut Abdullah Thamrin dan Francis Tantri (2016:3) “Penjualan adalah bagian dari promosi dan promosi adalah salah satu bagian dari keseluruhan sistem pemasaran”.
- b. Menurut Dharmmesta Basu Swasta dan T. Hani (2015:8) “Penjualan adalah ilmu dan seni mempengaruhi pribadi yang dilakukan oleh penjual untuk mengajak orang lain agar bersedia membeli barang dan jasa yang ditawarkan”

3. Kerangka Konsep

Tabel 1. Konsep berfikir

Permasalahan	
Belum adanya suatu prediksi untuk melihat potensi penjualan produk <i>frozen food</i>	
Tujuan	
Tujuannya untuk memprediksi penjualan produk pada PT. Pangan Lestari dengan menggunakan algoritma <i>Naïve Bayes</i> .	
Ekperimen	
Inputan	Algoritma
Dari data penjualan pada bulan januari 2017 sampai bulan juli tahun 2018 dengan isi data sebanyak 300 data.	<i>Naïve Bayes</i>
Hasil	
Dapat menentukan pola item mana yang paling laku dijual.	

Kerangka pemikiran pada penelitian kali ini adalah bagaimana menyelesaikan persoalan yang di hadapi oleh PT Pangan Lestari, yaitu memiliki kemampuan memprediksi rencana produksi yang disesuaikan dengan kapasitas dan penjualan. Penggunaan *naïve ayes* yang di aplikasikan kepada data penjualan, penggunaan Rapid miner sebagai alat untuk membantu pengolahan menjadi kunci keberhasilan penelitian ini.

4. Desain Penelitian/ Metodologi

4.1. Pengumpulan data

1. Sumber data

Dataset yang digunakan pada penelitian ini menggunakan data penjualan dari PT.Pangan Lestari sebanyak 3 atribut dan 300 data. atribut tersebut terdiri dari nama customer, wilayah, nama barang. Data dan atribut tersebut digunakan untuk memprediksi material yang sering di order dan mencari seberapa besar nilai akurasi algoritma *naïve bayes*. Data dan atribut tersebut digunakan untuk memprediksi material yang sering di order dan mencari seberapa besar nilai akurasi algoritma *naïve bayes*.

2. Pengumpulan data

Pengumpulan data merupakan bagian paling penting dalam sebuah penelitian. Ketersediaan data akan sangat menentukan dalam proses pengolahan dan analisa selanjutnya. Pada tahap pengumpulan data, penelitian menggunakan data sekunder berupa data penjualan *Delifrance* PT. Pangan Lestari dari data penjualan pada bulan januari 2017 sampai bulan juli tahun 2018 dengan isi data sebanyak 300 data.

4.2. Teknik Analisis data

1. Pengolahan dengan CRISP-DM (Cross Industry Proses Data Mining)
Dalam penelitian ini, peneliti menggunakan CRISP-DM sebagai metode analisis dan pemodelan dengan tahapan fase sebagai berikut.
2. Fase pemahaman Bisnis (*Business Understanding*)

Penelitian ini memiliki tujuan bisnis untuk menentukan aturan asosiasi yang terbentuk dari transaksi penjualan pada PT Pangan Lestari .Sehingga dari hasil yang diperoleh melalui analisis aturan asosiasi dapat bermanfaat untuk Indomaret. Strategi awal yang akan digunakan adalah mencari dataset transaksi penjualan untuk merancang *software data mining*.

3. Fase Pemahaman Data (*Data Understanding*)

Pengumpulan data primer dilakukan dengan cara mengambil data laporan transaksi penjualan pada toko PT Pangan Lestari selama bulan Agustus 2018.

4. Fase Pengolahan Data (*Data Preparation*)

Pada tahap ini penulis mengolah data yang didapat dengan beberapa tahapan pada KDD seperti melakukan pembersihan terhadap data / *data cleaning*, melakukan integrasi data / *data integration*, melakukan pemilihan data / *data selection* dan transformasi pada data / *data transformation*. Pada tahapan pengolahan data, data yang telah diperoleh dari PT Pangan Lestari akan diolah terlebih dahulu sebelum melalui tahap prediksi yaitu melalui beberapa proses antara lain :

Tahap Pertama, penentuan data yang akan diolah pada penelitian ini Berikut pada tabel 3.3 merupakan tabel atribut data penelitian dari dataset.

Tabel 2. Atribut Dataset

No	Atribut	Type	Proses
1	Nama Customer	Char	Digunakan
2	Penjual/wilayah	Char	Digunakan
3	Kode barang	Char	Digunakan
4	Status	Char	Digunakan

Tahap Kedua, melakukan konversi data. Data dengan atribut yang telah dipilih kemudian dikonversikan untuk memudahkan proses *Data Mining* pada sebagian atribut, karena data akan diproses dengan *tools* bantu *Data Mining*.

4.3. Tahap Pemodelan

Pada tahap ini penulis menentukan teknik *Data Mining* yang digunakan untuk mengolah data yang sudah disiapkan sebelumnya. Teknik yang dilakukan yaitu dengan *klasifikasi* menggunakan algoritma *Naïve Bayes*. Data yang sudah melalui proses pengolahan kemudian akan dilakukan perhitungan dengan menggunakan *Rapidminer*.

Dua langkah yang dilakukan pada tahap ini ialah :

Perhitungan *Naïve Bayes* secara manual Data yang akan digunakan dalam perhitungan *Naïve Bayes* secara manual yaitu

1. 300 sampel data. Dimana data yang diambil secara acak oleh peneliti.
2. Penerapan *Naïve Bayes* menggunakan *RapidMiner*. Penerapan dengan menggunakan *RapidMiner* bertujuan untuk memudahkan dalam pemrosesan data yang berjumlah besar. Proses perhitungan *Naïve Bayes* menggunakan *RapidMiner*.

4.4. Tahap Evaluasi

Melakukan pengujian terhadap model-model yang bertujuan untuk mendapatkan model yang paling akurat. Pada tahap evaluasi, akan diketahui apakah hasil dari tahap pemodelan dapat menjawab tujuan yang telah ditetapkan pada fase pertama. Sehingga diharapkan mendapatkan informasi atau pola yang berguna sebagai rekomendasi upaya prediksi potensi penjualan di PT Pangan Lestari.

4.5. Tahap Penyebaran

Pembentukan model selanjutnya melakukan analisa dan pengukuran pada tahap sebelumnya, pada tahap ini diterapkan model atau rule yang paling akurat dan selanjutnya dapat digunakan untuk mengevaluasi data baru.

5. Hasil Penelitian dan Pembahasan

5.1. Perhitungan Naive Bayes

Perhitungan *Naïve Bayes* dilakukan dengan menghitung menggunakan data yang diambil dari data penjualan PT.Pangan Lestari sebanyak 3 atribut dan 300 data. Kriteria yang digunakan adalah sebagai berikut :

5.2. Pengujian Naive Bayes

Pengujian pertama ini akan menggunakan satu algoritma yakni *naïve baye* tanpa menggunakan metode optimasi dalam melakukan klasifikasi data sebanyak 1 *dataset*. Berikut ini model pengujian yang digunakan dapat dilihat pada tabel di bawah ini :

Tabel 3. Data Customer, dan wilayah perdagangan

Nama Customer	GRAN MELIA JAKARTA, HOTEL, THE MARGO, HOTEL, SWISS BELINN KEMAYORAN, ART HOTEL, ASTON HOTEL - ANYER BEACH, ASTON HOTEL - PLUIT – DM, ASTON HOTEL - RASUNA SAID, ASTON IMPERIAL HOTEL – BEKASI, ASTON SENTUL LAKE RESORT & CONFERENCE, BEST SINAR NUSANTARA, PT, BEST WESTERN HOTELKUNINGAN, FAVE CILILITAN, HOTEL, GRAND MERCURE KEMAYORAN, HOTEL, GRANDHIKA, HOTEL, HARISTON HOTEL & SUITES, HORISON CILEDUG, IBIS STYLE GAJAH MADA, KERATON HOTEL, NOVOTEL GOLF RESORT & CONVENTION CENTRE HOTEL, ROYAL KUNINGAN HOTEL, POP HOTEL, OLYMPIC PREMIER, HOTEL, THE ALANA HOTEL & CONFERENCE SENTUL CITY, THE GROVE SUITE HOTEL, THE RITZ CARLTON HOTEL-KUNINGAN		
Penjual/wilayah	JKT FS 1H, JKT FS 2H, JKT FS 3H		
Kode barang	IDT76543, IDT78566, IDT78632, IDT78655, IDT78997, IDT79911, IDT79914,	IDT78126, IDT78192, IDT78622, IDT78628, IDT78644, IDT78651, IDT78662, IDT78688,	IDT78463, IDT78630, IDT78654, IDT78699,
Status	LAKU, SEDANG, TIDAK LAKU		

Tabel 2 merupakan Kriteria data penjualan produk PT. Pangan Lestari yang telah memasuki tahap pengolahan data kemudian akan dilakukan perhitungan dengan menggunakan algoritma *naïve bayes* sebagai berikut :

1. Menghitung probabilitas
Perhitungan ini untuk menentukan keputusan dengan metode *naïve bayes* untuk mencari probabilita darimasing-masing kelas dalam menentukankeputusanakan ditentukan kelas “Laku”, “Sedang”dan “TidakLaku”. Perhitungannya dimana mencari berapa jumlah data “Laku”, “Sdang” dan “Tidak Laku” dari total keseluruhan data training dan membaginya dengan total keseluruhan data.

Tabel 4. Probabilitas Keterangan

(P) Keterangan					
Laku		Sedang		Tidak Laku	
Laku	216	Sedang	37	Tidak Laku	47

2. Menghitung Probabilitas Masing-Masing Atribut
Mencari *probabilitas* suatu atribut dengan membandingkan atribut dari data *testing* dengan atribut dari data *training*. Jumlah atribut dari kelas”Laku” yang ada pada data training, kemudian dibagikan dengan probabilitas kelas “Laku”. Begitu juga mencari *probabilitas* untuk kelas “Sedang” dan “Tidak Laku”.
3. Customer

Tabel 5. Atribut Nama Customer

Nama Customer	Laku	Sedang	Tidak Laku
Tree Park Service Apartment	4	0	0
The Ritz Carlton Hotel-Kuningan	3	0	0
The Margo, Hotel	40	5	8
The Grovesuites Hotel	27	5	6
The Alana Hotel & Conferencsentul City	18	4	5
Swiss Bell Hotel	15	1	1
Swiss Belinn Kemayoran	4	0	2

Swiss Bel Bandara, Hotel	4	1	0
Royal Kuningan Hotel	11	3	3
Pop Hotels	2	1	1
Olympic Premier, Hotel	13	4	4
Novotelgolf Resort & Convention Centre Hotel	1	0	0
Keraton Hotel	5	1	1
Ibis Style Gajah Mada	3	0	1
Horison Ciledug	6	0	2
Hariston Hotel & Suites	2	1	0
Grandhika, Hotel	4	1	1
Grandmercure Kemayoran, Hotel	1	0	0
Gran Melia Jakarta, Hotel	13	1	1
Fave Cililitan, Hotel	3	0	2
Best Western Hotel – Kuningan	3	0	0
Best Sinar Nusantara, Pt	3	3	2
Aston Sentul Lake Resort & Conference	9	2	2
Aston Imperial Hotel – Bekasi	1	0	0
Aston Hotel - Rasuna Said	4	0	2
Aston Hotel - Pluit – Dm	1	1	1
Aston Hotel - Anyer Beach	11	2	1
Art Hotel	5	1	1

4. Penjual/Wilayah

Tabel 6. Atribut Penjual/Wilayah

Penjual/Wilayah	LAKU	SEDANG	TIDAK LAKU
Jkt Fs 1h	64	10	12
Jkt Fs 2h	104	17	25
Jkt Fs 3h	48	10	10

5. Kode barang

Tabel 7. Atribut Kode Barang

Kode Barang	LAKU	SEDANG	TIDAK LAKU
IDT79914	6	1	1
IDT78632	2	1	1
IDT78688	12	3	2
IDT78662	26	3	6
IDT78628	17	4	1
IDT78655	17	1	2
IDT78622	61	9	13
IDT78630	40	11	11
IDT78644	16	2	4
IDT78654	4	0	1
IDT78126	0	0	1

5.3. Menghitung Probabilitas akhir setiap kelas

Menghitung *probabilitas* akhir pada setiap kelas, perlu menggunakan data *training* yang terdapat pada tabel 6 dan mengubahnya menjadi nilai yang sudah ditentukan pada perhitungan probabilitas masing-masing atribut, dari masing-masing atribut dan nilai *probabilitas* kelas dikalikan. dari hasil yang sudah ditentukan pada tiap kelas, bandingkan nilai yang paling tinggi, jika kelas “LAKU” bernilai lebih besar maka hasilnya “LAKU”. Begitu pula dengan “SEDANG” dan “TIDAK LAKU”.

5.4. Kasus Perhitungan Naïve Bayes

Dalam memudahkan pemahaman perhitungan *naïve bayes*, secara manual akan di buat studi kasus sebagai berikut dan *rulanya* berupa data *training* (Lampiran1):

Table 8. Data Testing

Nama Customer	Penjual/wilayah	Kode Barang	Status
ASTON HOTEL - ANYER BEACH	JKT FS 1H	IDT76543	SEDANG
ASTON HOTEL - ANYER BEACH	JKT FS 1H	IDT76543	LAKU
IBIS STYLE GAJAHMADA	JKT FS 1H	IDT76543	SEDANG
THE ALANA HOTEL& CONFERENCE SENTUL CITY	JKT FS 2H	IDT78126	TIDAKI
GRANDMERCURE KEMAYORAN, HOTEL	JKT FS 3H	IDT78192	LAKU
KERATON HOTEL	JKT FS 1H	IDT78463	TIDAK LAKU
KERATON HOTEL	JKT FS 1H	IDT78463	LAKU
KERATON HOTEL	JKT FS 1H	IDT78463	LAKU
KERATON HOTEL	JKT FS 1H	IDT78463	LAKU
ASTON HOTEL - ANYER BEACH	JKT FS 1H	IDT78566	LAKU
ASTON HOTEL - ANYER BEACH	JKT FS 1H	IDT78566	LAKU
SWISS BELL HOTEL	JKT FS 3H	IDT78622	LAKU
GRAN MELIA JAKARTA, HOTEL	JKT FS 2H	IDT78622	LAKU
SWISS BELL HOTEL	JKT FS 3H	IDT78622	LAKU
SWISS BELINN KEMAYORAN	JKT FS 3H	IDT78622	LAKU
THE GROVESUITES HOTEL	JKT FS 2H	IDT78622	LAKU
SWISS BELL HOTEL	JKT FS 3H	IDT78622	TIDAK LAKU
THE ALANA HOTEL& CONFERENCE SENTUL CITY	JKT FS 2H	IDT78622	TIDAK LAKU
THE MARGO, HOTEL	JKT FS 1H	IDT78622	LAKU
GRANDHIKA, HOTEL	JKT FS 2H	IDT78622	LAKU

5.5. Menghitung jumlah kelas atau prediksi :

Tabel 9. Jumlah Kelas atau Prediksi

Keterangan					
P(Ci)					
P (laku)		P (sedang)		P (tidak laku)	
Laku	216/300	Sedang	37/300	Tidak laku	47/300
Laku	0.72	Sedang	0.1233	Tidak laku	0.156667
Atribut Kelas (laku, sedang, tidak laku)					
P (Keterangan) laku, sedang, tidak laku					

Tabel 10. Menghitung Jumlah Kelas atau Prediksi

	no	P	P(Ci)	P(X C1)	P(X C2)	P(X C3)		MAX	STATUS
LAKU	1	216	0,7200	0,0185	0,2546	0,0370	0,0001	0,0001	LAKU
SEDANG		37	0,1233	0,0000	0,1892	0,0000	0,0000		
TIDAK LAKU		47	0,1567	0,0000	0,1277	0,0000	0,0000		
LAKU	2	216	0,7200	0,0648	0,2546	0,2778	0,0033	0,0033	LAKU
SEDANG		37	0,1233	0,0270	0,1892	0,1622	0,0001		
TIDAK LAKU		47	0,1567	0,0426	0,1277	0,3617	0,0003		

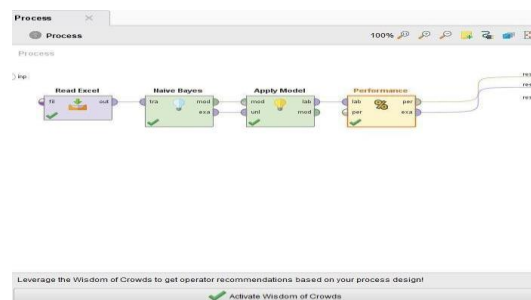
LAKU	3	216	0,7200	0,0509	0,4583	0,2778	0,0047	0,0047	LAKU
SEDANG		37	0,1233	0,0541	0,5405	0,1622	0,0006		
TIDAK LAKU		47	0,1567	0,0426	0,5745	0,3617	0,0014		
LAKU	4	216	0,7200	0,0509	0,4583	0,2176	0,0037	0,0037	LAKU
SEDANG		37	0,1233	0,0541	0,5405	0,2432	0,0009		
TIDAK LAKU		47	0,1567	0,0426	0,5745	0,1277	0,0005		
LAKU	5	216	0,7200	0,0648	0,2546	0,2778	0,0033	0,0033	LAKU
TIDAK LAKU		47	0,1567	0,0426	0,1277	0,3617	0,0003		
LAKU	100	216	0,7200	0,0972	0,4583	0,0000	0,0000	0,0001	TIDAK
SEDANG		37	0,1233	0,0811	0,5405	0,0000	0,0000		LAKU
TIDAK LAKU		47	0,1567	0,0638	0,5745	0,0213	0,0001		
LAKU	101	213	0,7100	0,0657	0,2582	0,2817	0,0034	0,0034	LAKU
SEDANG		213	0,7100	0,0047	0,0329	0,0282	0,0000		
TIDAK LAKU		37	0,1233	0,0541	0,1622	0,4595	0,0005		

Tabel 9 menghitung jumlah kelas dimana :
Perhitungan Jumlah Kelas :

1. Nilai $P(C_i)$ keseluruhan keterangan label “Laku”, “Sedang”, dan “tidak laku”
2. Nilai $P(C_i|x)$ adalah Hasil dimana label “Laku”, “Sedang”, dan “Tidak Laku” dibagi dengan keseluruhan data
3. Nilai $P(X|C_1)$ adalah Nilai Customer yang sudah dihitung dari keseluruhan nilai Laku dibagi keseluruhan data Laku, dan sebaliknya nilai “Sedang” dan “Tidak Laku”
4. Nilai $P(X|C_2)$ adalah Nilai Penjual/Wilayah yang sudah dihitung dari keseluruhan nilai Laku dibagi keseluruhan data Laku, dan sebaliknya nilai “Sedang” dan “Tidak Laku”
5. Nilai $P(X|C_3)$ adalah Nilai Kode Barang yang sudah dihitung dari keseluruhan nilai Laku dibagi keseluruhan data Laku, dan sebaliknya nilai “Sedang” dan “Tidak Laku”
6. Hasil $P(X|C_i)$ adalah perhitungan dari $P(X|C_1)$ sampai $P(X|C_3)$ yang sudah dikalikan

5.2.2. Implementasi Klasifikasi *Naïve Bayes* pada *RapidMiner*

Perhitungan yang telah dilakukan diatas sesuai untuk menentukan kelulusan dengan metode *Naïve Bayes*. Setelah melakukan perhitungan manual terhadap data *training* dan data *testing*, langkah selanjutnya pembuktian dengan menggunakan *RapidMiner*. Pembuktian perhitungan manual *Naïve Bayes* terhadap data *training* dan data *testing*, akan dilakukan perhitungan menggunakan *RapidMiner*. *RapidMiner* yang digunakan adalah versi 7.6



Gambar 4. Proses *RapidMiner* pengujian data *training* terhadap data *testing* dalam menentukan prediksi.

Row No.	status	prediction...	confidence...	confidence...	confidence...	Nama Cesto...	Penjualwa...	Kode Barang
1	LAKU	LAKU	1.000	0.800	0.000	TREE PARK...	_JKT FS 3H	IDT79914
2	LAKU	LAKU	0.890	0.083	0.028	SWISS BELL...	_JKT FS 3H	IDT78622
3	LAKU	LAKU	0.704	0.259	0.088	GRANMELLA...	_JKT FS 2H	IDT78622
4	LAKU	LAKU	0.728	0.097	0.174	GRANMELLA...	_JKT FS 2H	IDT78630
5	LAKU	LAKU	0.890	0.083	0.028	SWISS BELL...	_JKT FS 3H	IDT78622
6	LAKU	LAKU	0.695	0.105	0.200	THE MARGO...	_JKT FS 1H	IDT78628
7	LAKU	LAKU	0.585	0.330	0.085	THE MARGO...	_JKT FS 1H	IDT78655
8	LAKU	LAKU	0.920	0.000	0.080	SWISS BELL...	_JKT FS 3H	IDT78622
9	LAKU	LAKU	0.912	0.000	0.088	SWISS BELL...	_JKT FS 3H	IDT78652
10	LAKU	LAKU	0.603	0.314	0.083	THE GROVE...	_JKT FS 2H	IDT78622
11	LAKU	LAKU	0.453	0.116	0.432	OLYMPIC PR...	_JKT FS 2H	IDT78628
12	TIDAK LAKU	LAKU	0.393	0.348	0.259	OLYMPIC PR...	_JKT FS 2H	IDT78644
13	SEDANG	SEDANG	0.404	0.071	0.525	OLYMPIC PR...	_JKT FS 2H	IDT78688
14	LAKU	LAKU	1.000	0.800	0.000	TREE PARK...	_JKT FS 3H	IDT79914
15	SEDANG	LAKU	0.670	0.228	0.102	THE GROVE...	_JKT FS 2H	IDT78652
16	LAKU	LAKU	0.452	0.392	0.146	THE GROVE...	_JKT FS 2H	IDT78644
17	LAKU	LAKU	0.608	0.157	0.176	THE GROVE...	_JKT FS 2H	IDT78630

Gambar 5. Hasil Prediksi RapidMiner

Gambar 7. Hasil Prediksi dimana data *testing* menggunakan sample acak awal dan akhir data yang digunakan dalam penelitian ini.

Pengujian dengan metode *Naïve Bayes* menggunakan dataset Penjualan. Hasil yang didapatkan pengujian ini mendapatkan hasil akurasi sebesar 72,00% dengan nilai presisi serta *recall* masing-masing kelas dapat dilihat pada gambar 8 berikut ini :

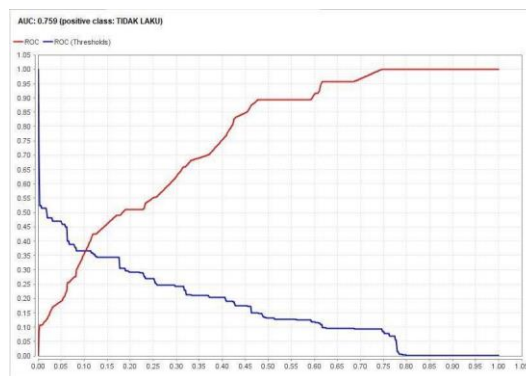
accuracy: 72.00%				
	true LAKU	true TIDAK LAKU	true SEDANG	class precision
pred. LAKU	209	41	32	74.11%
pred. TIDAK LAKU	1	3	1	60.00%
pred. SEDANG	6	3	4	30.77%
class recall	95.76%	6.38%	10.61%	

Gambar 6. Hasil Accuracy RapidMiner

5.3. Pembahasan

Penelitian yang dilakukan menggunakan algoritma *naïve bayes* menentukan *Area Under Curve (AUC)*

5.3.1 Area Under Curve (AUC)



Gambar 7. Kurva ROC

Kurva ROC digunakan untuk mengekspresikan data, garis horizontal mewakili nilai *false positive* dan garis vertical mewakili nilai *true positive*. Dari gambar 4.4 dapat diketahui bahwa nilai *Area Under Curve (AUC)* model algoritma *naïve bayes* adalah 1.000, hal ini menunjukkan bahwa algoritma *naïve bayes* mencapai klasifikasi sempurna.

5.3.2. Pola baru Naïve Bayes

Berdasarkan prediksi menentukan potensi penjualan dalam penelitian ini dapat diperoleh informasi bahwa pada prediksi penjualan berdasarkan label “Laku, Sedang, Tidak Laku”.

5.3.3. Resume :

Berdasarkan pola baru *Naïve Bayes* prediksi potensi penjualan dalam penelitian ini dapat diperoleh prediksi penjualan berdasarkan atribut pada data penjualan untuk status “LAKU” membentuk sebanyak 25 pola, status “SEDANG” dapat diperoleh sebanyak 4 pola, dan untuk status “TIDAK LAKU” diperoleh sebanyak 5 pola.

6. Kesimpulan

Berdasarkan pada hasil penelitian yang telah dilakukan maka dapat disimpulkan :

1. Penerapan Algoritma *Naïve Bayes* untuk memprediksi potensi penjualan berdasarkan data set penjualan *deliverance* pada PT.Pangan Lestari dapat membantu perusahaan dalam memprediksi potensi penjualan.
2. Berdasarkan data yang diperoleh, proses *Data Mining* membantu dalam penerapan metode

Naïve Bayes dalam mendapatkan informasi dari hasil prediksi pada data penjualan *delifrance* di PT.Pangan Lestari dijadikan *Data Mining*. Sehingga dengan demikian metode *Naïve Bayes* ini berhasil memprediksi dengan presentase keakuratan sebesar 72,00 % dengan menggunakan data sebanyak 300.

3. Dari hasil yang diperoleh data penjualan *delifrance* dapat mengetahui tingkat dan volume penjualan barang PT. Pangan Lestari kepada para konsumen.

Daftar Pustaka

- Abdullah, Thamrin dan Francis Tantri. 2016. *Manajemen Pemasaran*. Depok : PT Raja Grafindo Persada
- Andini, Taghsya Izmi (2017). *Prediksi Potensi Pemasaran Produk Baru Dengan Metode Naive Bayes Classifier dan Regresi Linear*.
- As'ad, B. (2016). *Prediksi Keputusan Menggunakan Metode Klasifikasi Naïve, Jurnal Unsada Bayes, One-R dan Decision Tree. Penelitian Komunikasi Dan Opini Publik*, Salemba 4
- Chandra, Andreas. (2017). Penerapan Data Mining Menggunakan Pohon Keputusan Dengan Algoritma C4.5 Dalam Menentukan Kecelakaan Penerbangan, *Jurnal STIMIK AMIKOM*, Yogyakarta.
- Dharmmesta, Basu Swastha dan T. Hani Handoko. 2012. *Manajemen Pemasaran Analisis Perilaku Konsumen*. Edisi Pertama. Yogyakarta: BPPFE.
- Feri sulianta. (2014). Customer Profiling Pada Supermarket Menggunakan Algoritma K-Means Dalam Memilih Produk Berdasarkan Selera Konsumen Dengan Daya Beli Maksimum. *Jurnal Airlangga*
- Gullo, F. (2015). From patterns in data to knowledge discovery: What *data mining* can do. *Physics Procedia Journal*
- Jiawei han, micheline kamber, jian pei. (2012). *Data Mining Concepts and Techniques, Third Edition*.
- 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. *Jurnal Mantik Penusa*.
- Mikael, A. (2017). Penerapan Metode K-Means Clustering untuk pengelompokan Potensi produksi Buah-buahan Di Provinsi Daerah Istimewa Yogyakarta. *Jurnal Teknik Informatika*, Universitas Sanata Darma.
- Muhamad, H., Prasojo, C. A., Sugianto, N. A., Surtiningsih, L., Cholissodin, I., Ilmu, F., ... Optimization, P. S. (2017). Optimasi Naive BayesClassifi dengan Menggunakan Particle Swarm Optimization padaDat Iris. *Jurnal Teknologi Informasi Dan Pendidikan*.
- Prasetyo, E. (2012). Fuzzy K-Nearest Neighbor in Every Class Untuk Klasifikasi Data. *Seminar Nasional Teknik Informatika (SANTIKA 2012)*,(Santika), 57–60.
- R.Danang (2018). Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma FP- GROWTH Pada Data Transaksi Penjualan. *Jurnal Unesa*
- Saleh, A. (2015). Implementasi Metode Klasifikasi *Naive Bayes* dalam Memprediksi besarnya Penggunaan Listrik Rumah Tangga. *Jurnal Teknik informatika, Universitas Potensi Utama*.
- Santosa, B & Umam, A. (2018). *Data mining dan Big data Analytics*. Penerbi t ANDI, Jakarta..
- Wicaksono, A. E. (2016). Implementasi *data mining* dalam pengelompokan data peserta didik di sekolah untuk memprediksi calon penerima beasiswa dengan menggunakan algoritma k- means (studi kasus sman 16 bekasi). *Jurnal Teknik Informatika, UniversitasGunadarma*.