

PENERAPAN ALGORITMA APRIORI UNTUK MENENTUKAN PENJUALAN PRODUK PADA MINIMARKET (STUDI KASUS INDOMARET RUKO VENTURA)

Nurhadi Surojuddin¹⁾, Khoerrudin²⁾

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

Disetujui, 27 Juni 2019

Abstraksi

Toko Indomaret di ruko Ventura merupakan salah satu toko yang menjual berbagai kebutuhan sehari-hari. Setiap hari terjadi transaksi penjualan berbagai item barang. Akan tetapi manajemen Indomaret tidak mengetahui pasangan barang yang sering dibeli oleh konsumen secara bersamaan sehingga sering terjadi kurangnya stok pada item-item yang sering dibeli secara bersamaan oleh konsumen. Teknik *data mining* telah banyak digunakan untuk mengatasi permasalahan yang ada salah satunya dengan penerapan algoritma Apriori guna mendapatkan informasi tentang asosiasi antar produk dari suatu database transaksi. Data transaksi took dapat diolah kembali menggunakan aplikasi *data mining* sehingga menghasilkan aturan asosiasi keterkaitan yang kuat antar item-item penjualan sehingga bias member rekomendasi penyetokan barang di took terlebih barang-barang yang dibeli secara bersamaan oleh konsumen dan meningkatnya barang yang dijual di toko.

Kata Kunci : Toko minimarket, transaksi penjualan, *data mining*, algoritma Apriori.

Abstract

The Indomaret shop at the Ventura shop is one of the shops that sells various daily necessities. Every day there are sales transactions of various items. However, Indomaret management does not know the pairs of items that consumers often buy simultaneously, so there is often a lack of stock on items that consumers often buy simultaneously. Data mining techniques have been widely used to overcome existing problems, one of which is the application of the Apriori algorithm to obtain information about associations between products from a transaction database. Shop transaction data can be reprocessed using a data mining application so as to produce strong association rules between sales items so that members can recommend the stocking of goods in stores, especially goods purchased simultaneously by consumers and an increase in goods sold in stores.

Keywords: *Minimarket stores, sales transactions, data mining, Apriori algorithm.* *Pendahuluan*

1. Pendahuluan

Dunia perdagangan retail saat ini sudah demikian besar, dimana persaingan seru terjadi antara retail tradisional (toko kelontong) dengan minimarket yang bersifat franchise. Biasanya minimarket berbagi sharing keuntungan.

PT Indomarco Prismatama merupakan salah satu perusahaan yang bergerak di bidang retail waralaba di Indonesia yang menyediakan kebutuhan pokok dan kebutuhan sehari-hari. Konsumennya beragam, tetapi umumnya adalah rumah tangga, karyawan dll yang sibuk, sehingga setiap waktu, pola pembelian konsumen di Indomaret akan berbeda. Misalnya pola pembelian konsumen pada awal gajian, maka pelanggan akan berbelanja bahan pokok lebih banyak dibandingkan saat sebelum gajian. Promo produk juga sangat berpengaruh pada minat belanja konsumen. Semakin menarik promo yang ditawarkan, maka akan semakin banyak orang yang tertarik untuk berbelanja. Dengan tujuan menjual produk-produk tersebut sebanyak-banyaknya melalui produk-produk berpasangan yang sering dibeli oleh konsumen. Namun seringkali konsumen hanya akan membeli satu produk tertentu tanpa produk pelengkap. Misalkan jika konsumen sedang memerlukan deterjen baik cair atau bubuk, maka mereka hanya mengambil deterjen tersebut, namun belum tentu membeli pewangi pakaianya sebagai pelengkap. Sehingga ada produk pewangi pakaian tertentu sebagai produk pelengkap yang kurang laku dan tidak memberi keuntungan perusahaan, dikarenakan harga produk yang terlalu mahal ataupun tidak ada promo pada produk tersebut, sehingga produk pelengkap seringkali tidak terjual yang akhirnya dikembalikan lagi ke *Distribution Centre*(DC) dan tidak menjadi keuntungan perusahaan.

Oleh karena itu, diperlukan sebuah aplikasi yang dapat menentukan pola minat pembelian konsumen untuk produk berpasangan atau pelengkap, sehingga dapat diperoleh pola minat beli konsumen untuk

menentukan promo yang akan dibuat untuk waktu berikutnya. Pemanfaatan informasi dan pengetahuan yang terkandung pada banyaknya data tersebut, pada saat ini biasa disebut sebagai *data mining*. *Data mining* dimaksudkan untuk membantu memberikan solusi untuk pengambilan keputusan dalam pengembangan bisnis perusahaan.

Dengan demikian, dari beberapa penelitian yang dilakukan yaitu melalui data transaksi, maka dapat diperoleh pola minat pembelian konsumen dalam membeli suatu produk secara bersamaan di setiap waktu. Pada kasus pembelian produk – produk yang dibeli secara bersamaan disebut asosiasi antar produk.

Untuk mendapatkan informasi tentang asosiasi pembelian antara produk dari sebuah data transaksi, penulis menggunakan algoritma Apriori. Algoritma Apriori merupakan algoritma market basket analisis yang dapat digunakan untuk menghasilkan *association rule* dengan menggunakan pola “*if-then*” atau “jika-maka”. Algoritma ini dapat digunakan untuk menentukan pola konsumen pada saat membeli barang yang bersamaan.

2. Tinjauan Studi

2.1. Tinjauan Studi

Penelitian mengenai itemset data bukanlah pertama kali dilakukan. Sebelumnya sudah pernah dilakukan penelitian mengenai penerapan *data mining* pada penelitian tersebut.

Berikut ini merupakan penelitian yang relevan dengan masalah tersebut yaitu penelitian yang dilakukan oleh Ahmad Fauzi, Iis Joice dan Akim Manao yang berjudul “Sistem Pendukung Pemilihan Pekerjaan Menggunakan Metode A-priori Berdasarkan Korelasi Jurusan Dengan IPK untuk Mengetahui Pekerjaan yang Tepat”. Pada penelitian ini membahas mengenai tidak sedikit dari para lulusan sarjana maupun diploma yang mendapatkan pekerjaan tidak sesuai dengan latar belakang pendidikan yang ditempuhnya saat kuliah, dan bahkan ada juga yang tidak mendapatkan pekerjaan sehingga menganggur berkepanjangan, hal tersebut membuat hati miris bagi para dosen-dosen yang mengajarkan mereka sewaktu kuliah di perguruan tinggi. Indikator keberhasilan alumni selain bisa mengimplementasikan teori yang didapatkan dari perguruan tinggi pada kehidupan sehari-hari adalah mendapatkan pekerjaan sesuai dengan keahlian dan mendapatkan upah sesuai dengan yang mereka harapkan. *Best rule* yang menjadi target keberhasilan dari penelitian ini adalah, mencari *rule* terbaik yang berhubungan dengan Prodi nya masing-masing dan IPK terhadap bidang pekerjaan yang diperoleh oleh alumni, adapun hasil *rule* pada penelitian tersebut diperoleh pada minimum support =0.05 dan confidence=0.7, ditemukan 5 *rule*, yaitu :

- Bidang_Pekerjaan =Pembiayaan IPK=B 97 \implies Prodi=Sistem_Informasi 86 <conf:(0.89)> lift:(1.34) lev:(0.04) conv:(2.73) Artinya : 89% Pekerjaan dibidang Pembiayaan memiliki IPK B (Baik).
- Bidang_Pekerjaan=Pembiayaan 114 \implies Prodi= Sistem_Informasi 101 <conf:(0.89)> lift:(1.34) lev:(0.04) conv:(2.73) Artinya : 89% Pekerjaan dibidang Pembiayaan adalah dari Program Studi Sistem Informasi.
- Prodi=Sistem_Informasi Bidang_Pekerjaan=NULL 103 \implies IPK=B 88 <conf:(0.85)> lift:(1.05) lev:(0.01) conv:(1.2). Artinya 85% Prodi Sistem Informasi bidang pekerjaannya adalah Null (Belum bekerja).
- Prodi=Sistem_Informasi Bidang_Pekerjaan=Pembiayaan 101 \implies IPK=B 86 <conf:(0.85)> lift:(1.05) lev:(0.01) conv:(1.18). Artinya : 85% dari Program Studi Sistem Informasi bekerja dibidang Pembiayaan dengan memiliki IPK B (Baik).
- Bidang_Pekerjaan=Pembiayaan 114 \implies IPK=B 97 <conf:(0.85)> lift:(1.05) lev:(0.01) conv:(1.18). Artinya :85% bidang pekerjaannya Pembiayaan memiliki IPK B (Baik)

Kesimpulan dari penelitian tersebut yaitu alumni STMIK Kaputama Program Studi Sistem Informasi yang memiliki IPK B (Baik) akan bekerja pada bidang Pembiayaan.

Penelitian yang dilakukan oleh Kennedy Tampubolon, Hoga Saragih dan Bobby Reza yang berjudul “Implementasi *Data mining* Algoritma A-priori Pada Sistem Persediaan Alat-Alat Kesehatan”. Pada penelitian ini membahas tentang pentingnya sistem penyediaan obat pada suatu apotek. Untuk mengetahui alat-alat kesehatan apa saja yang dibeli oleh konsumen, maka diperlukan teknik analisis salah satunya teknik analisis keranjang pasar yaitu analisis kebiasaan membeli konsumen. Penerapan algoritma Apriori membantu dalam membentuk kandidat kombinasi item yang mungkin, selanjutnya dilakukan pengujian apakah kombinasi yang terbentuk memenuhi parameter *support* dan *confidence* minimum yang merupakan nilai ambang yang diberikan oleh pengguna. Hasil dari pengujian tersebut menunjukkan bahwa aktifitas penjualan dan pembelian menggunakan komputer yang berbasis jaringan dengan menggunakan *database* terpusat kedalam satu server. Sistem 89) > lift:(1.34) lev:(0.05) informasi persediaan barang di apotek terdiri dari data obat dan alat-alat kesehatan, gudang farmasi, pelayanan resep obat dan penjualan. Tampilan gudang farmasi terdiri atas bagian persediaan, laporan pembelian, laporan penjualan sedangkan pada data alat-alat kesehatan keluar terdiri dari *attribute* nomor, tanggal, nomor slip, nama, jumlah barang, harga, dan total harga dan setiap transaksi penjualan minimal terdiri atas satu jenis alat-alat kesehatan. Untuk mendapat analisa data maka data penjualan di *export* kedalam database Microsoft excel karena database Microsoft excel bersifat *spreadsheet* sehingga sangat mendukung dalam analisa data. Microsoft excel dalam analisis data digunakan sebagai database karena

Microsoft excel sangat mendukung beberapa aplikasi *data mining*, dimana aplikasi (perangkat lunak) *data mining* digunakan sebagai tempat uji coba atau implementasi.

Hasil aturan asosiasi pada penelitian tersebut yaitu pada tabel di halaman berikut :

Tabel 1. Hasil Aturan Asosiasi

No	Kombinasi Item (merek)	Support	Confidence
1.	Jika membeli Stik Asam Urat, maka akan membeli Stuk Gula	20%	100%
2.	Jika membeli Stik Kolesterol, maka akan membeli Stik Gula	16%	100%

Kesimpulan dari asosiasi yang terbentuk yaitu :

1. Jika membeli Stik Asam Urat, maka akan membeli Stik Gula.
2. Jika membeli Stik Colesterol, maka akan membeli Stik Gula.

Penelitian yang dilakukan oleh Moh. Sholik dan Abu Salam yang berjudul “Implementasi Algoritma Apriori untuk Mencari Asosiasi Barang yang Dijual di E- commerce OrderMas” menjelaskan tentang bagaimana mengetahui secara dini tentang ketersediaan barang yang terdapat pada toko / perusahaan tertentu berdasarkan data transaksi penjualan pada sistem agar perusahaan tidak mengalami kekosongan stok untuk kelancaran transaksi pada perusahaan tersebut.

Hasil dari penelitian tersebut yaitu sebagai berikut:

Tabel 2. Hasil Penelitian E-commerce OrderMas

Kode Barang	Jumlah Barang	Proses Support	Support
A1-A2-A3	2	(2/10)*100	20%
A3-A4-A5	1	(1/10)*100	10%
A5-A6-A7	0	(0/10)*100	0%
A6-A7-A-8	2	(2/10)*100	20%
A7-A8-A9	4	(4/10)*100	40%
A8-A9-A10	2	(2/10)*100	20%

Dari tabel diatas dapat disimpulkan bahwa konsumen paling sering membeli produk dengan kode barang A7-A8-A9 yang berisi Pink Fairy Cake, Doormat New englang dan Doormat Fairy Cake karena data dari item tersebut telah memenuhi nilai minimum confidence yang ditentukan.

2.2. Tinjauan Pustaka

2.2.1 Data mining

Definisi *Data mining* Menurut Turban et al (2005), *data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam database besar.

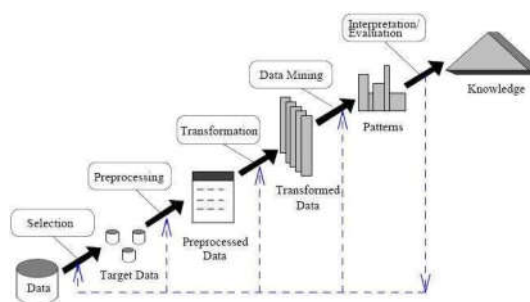
Menurut Larose (2006), *data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika.

Menurut Pramudiono (2006), *data mining* merupakan analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya.

Menurut Tan (2006) mendefinisikan *data mining* sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. *Data mining* juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah *data mining* kadang disebut juga knowledge discovery. Salah satu teknik yang dibuat dalam *data mining* adalah bagaimana menelusuri data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam basis data yang tersimpan. Kebutuhan untuk prediksi juga dapat memanfaatkan teknik ini. Dalam *data mining*, pengelompokan data juga bisa dilakukan. Tujuannya adalah agar kita dapat mengetahui pola universal data-data yang ada. Anomali data transaksi juga perlu dideteksi untuk dapat mengetahui tindak lanjut berikutnya yang dapat diambil. Semua hal tersebut bertujuan mendukung kegiatan operasional perusahaan sehingga tujuan akhir perusahaan diharapkan dapat tercapai.

2.2.2. 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 task *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.

2.2.3. Metode *Data Mining*



Gambar 2. Metode *Data Mining*

Berikut ini merupakan metode yang terdapat pada *data mining* diantaranya:

1. Estimasi (*Estimation*)
Estimasi adalah perkiraan penilaian atau pendapat. Estimasi adalah suatu metode dimana kita dapat memperkirakan nilai dari suatu populasi dengan menggunakan nilai dari sampel. Estimator adalah nilai pendugaan/suatu data statistik, sebagai sampel yang digunakan untuk mengisi suatu parameter. Pada metode estimasi variable target adalah berupa bilangan numerik dan bukan kategorikal. Estimasi nilai dari variable target ditentukan berdasarkan nilai dari variable prediktor (atribut).
2. Prediksi (*Prediction*)
Prediksi adalah hasil dari kegiatan memprediksi atau meramal atau memperkirakan apa yang akan terjadi di masa depan dari data yang sudah ada. Metode prediksi / *forecastings* sama dengan metode estimasi di mana label / target / class bertipe numerik, bedanya adalah data yang digunakan merupakan data rentet waktu (*data time series*). Istilah prediksi tidak hanya untuk prediksi time series, karena sifatnya yang bisa menghasilkan class berdasarkan berbagai atribut yang kita sediakan.
3. Klasifikasi (*Clasification*)
Klasifikasi adalah sebuah metode untuk menyusun data secara sistematis atau menurut beberapa aturan atau kaidah yang telah ditetapkan. Dengan kata lain yaitu pengelompokan data berdasarkan suatu acuan / ketetapan tertentu. Klasifikasi menggunakan data dengan target/class/label berupa nilai kategorikal (nominal). Contoh, apabila target/class/label adalah pendapatan, maka bisa digunakan nilai nominal (kategorikal).
4. Klastering (*Clustering*)
Klastering adalah metode penganalisaan data, yang sering dimasukkan sebagai salah satu metode *data mining*, yang tujuannya adalah untuk mengelompokkan data dengan karakteristik yang sama ke suatu 'wilayah' yang sama dan data dengan karakteristik yang berbeda ke 'wilayah' yang lain.
Perbedaan utama metode klastering dengan klasifikasi adalah klastering tidak memiliki target/class/label, jadi termasuk *unsupervised learning*. Klastering sering digunakan sebagai tahap awal dalam proses *data mining*, dengan hasil klaster yang

2.2.4. Algoritma Apriori

1. Definisi Algoritma Apriori
Algoritma Apriori merupakan algoritma yang digunakan untuk menemukan pola frekuensi tinggi. Pola frekuensi tinggi merupakan pola-pola item di dalam suatu database dimana database tersebut memiliki frekuensi atau support di atas ambang batas tertentu yang biasa disebut dengan istilah minimum support dan confidence. Pola frekuensi tinggi tersebut digunakan untuk menyusun aturan asosiatif.
2. Association Rule Mining
Menurut Kusriani, Analisis asosiasi atau *association rule mining* adalah teknik *data mining* untuk menemukan aturan asosiatif dari kombinasi *itemset*. Algoritma ini sering digunakan pada transaksi penjualan sehingga dikenal juga dengan nama *market basket analysis*. Dari pola atau aturan asosiasi yang telah terbentuk dapat diketahui kemungkinan pembelian barang primer dan barang sekunder lainnya sehingga manajemen dapat mengatur perbaikan susunan pola letak suatu barang serta merancang promosi pemasaran dengan membuat kupon diskon untuk kombinasi barang tertentu.
3. Support dan Confidence
Didalam *association rules data mining* ada dua buah parameter penting yang berfungsi sebagai pembentukan rules yaitu *support* (nilai penunjang) dan *confidence* (nilai kepastian) sehingga dapat dihasilkan aturan asosiasi yang kuat (*strong rules*). *Support* adalah persentase dari kombinasi suatu item di dalam database, sedangkan *confidence* merupakan kuatnya hubungan antar item yang berada di dalam aturan asosiasi. Contoh aturan asosiasi yang terbentuk umumnya dapat dinyatakan dalam bentuk sebagai berikut:
{beras, minyak sayur} => {telur} (support = 40%, confidence = 50%).

Pada aturan asosiasi tersebut memiliki arti 50% dari transaksi pada database yang memuat item beras dan minyak sayur juga memuat item telur. Sedangkan 40% dari seluruh transaksi yang terjadi memuat ketiga item tersebut. Seorang pihak pengelola akan mengambil rules yang telah terbentuk jika aturan asosiasi yang terbentuk kuat atau aturan asosiasi yang terbentuk melebihi batas minimal support dan confidence.

1. Metodologi Dasar
Analisa Pola Frekuensi Tinggi merupakan tahapan untuk pembentukan kombinasi item yang memenuhi syarat minimum dari nilai *support* dalam database. Nilai *support* pada satu itemset

diperoleh dengan rumus:

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi mengandung A}}{\text{Total Transaksi}}$$

Rumus diatas menjelaskan bahwa untuk menentukan suatu nilai *support* pada satu itemset jumlah transaksi yang mengandung item A dibagi dengan jumlah transaksi yang ada pada database.

Misalkan ada data transaksi suatu toko sebagai berikut :

Tabel 3. Data Contoh Transaksi

<u>Id Transaksi (Support)</u>	<u>Items</u>
1	<u>Pena,roti,mentega</u>
2	<u>Roti,mentega,telur,susu</u>
3	<u>Buncis,telur,susu</u>
4	<u>Roti,mentega</u>
5	<u>Roti,mentega,kecap,telur,susu</u>

Maka dengan rumus diatas maka akan terbentuk

1. Support (roti) : 4/5x100 = 80%
2. Support (mentega) : 4/5x100 = 80%
3. Support (telur) : 3/5x100 = 60%
4. Support (buncis) : 1/5x100 = 20%
5. Support (kecap) : 1/5x100 = 20%

Dari hasil diatas, maka didapatkan k-item *support* yang memenuhi minimum *support*, sehingga terpilih k-itemset sebagai pola frekuensi tertinggi yaitu sebagai berikut.

Tabel 4. Contoh Perhitungan *support* k-item

<u>Itemset</u>	<u>Support Count</u>	<u>Support</u>
<u>Roti</u>	80%	4
<u>Mentega</u>	80%	4
<u>Telur</u>	60%	3
<u>Susu</u>	60%	3

Sedangkan untuk dua itemset atau lebih diperoleh dengan rumus:

$$\text{Support (A} \cap \text{B)} = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Total Transaksi}}$$

Pada rumus diatas menjelaskan bahwa untuk menentukan nilai *support* pada dua itemset atau lebih, jumlah transaksi yang mengandung item A dan B dibagi dengan total transaksi yang terjadi pada dataset.

Jika diterapkan pada kasus diatas, maka pertama-tama menentukan 2-itemset sebagai berikut :

Tabel 5. Contoh 2-itemset

<u>Itemset</u>
<u>Roti,mentega</u>
<u>Roti,telur</u>
<u>Roti,susu</u>
<u>Mentega,telur</u>
<u>Mentega,susu</u>
<u>Telur,susu</u>

Dengan menggunakan rumus diatas, maka diperoleh data sebagai berikut:

Tabel 6. Contoh Perhitungan *Support* 2-itemset

<u>Itemset</u>	<u>Support Count</u>	<u>Support</u>
<u>Roti,mentega</u>	4	80%
<u>Roti,telur</u>	2	40%
<u>Roti,susu</u>	2	40%
<u>Mentega,telur</u>	2	40%
<u>Mentega,susu</u>	2	40%
<u>Telur,susu</u>	3	60%

Langkah selanjutnya yaitu menentukan 3-itemset yaitu sebagai berikut.

Tabel 7. Contoh 3-itemset

<u>Itemset</u>
<u>Roti, mentega, telur</u>
<u>Roti, mentega, susu</u>
<u>Mentega, telur, susu</u>

Dengan menggunakan rumus saat menentukan 2-itemset maka data yang dihasilkan yaitu sebagai berikut.

Tabel 8. Contoh Perhitungan *Support* 3-itemset

<u>Itemset</u>	<u>Support Count</u>	<u>Support</u>
<u>Roti, mentega, telur</u>	40%	2
<u>Roti, mentega, susu</u>	40%	2
<u>Mentega, telur, susu</u>	40%	2

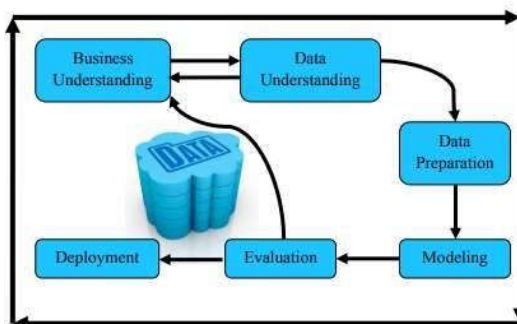
Selanjutnya masih ada satu pola yang ditemukan lagi yaitu pola 4-itemset. Itemset ini dipilih sebagai pola frekuensi tertinggi. 4-itemset tersebut yaitu sebagai berikut.

Tabel 9. Contoh Perhitungan *Support* 4-itemset

<u>Itemset</u>	<u>Support Count</u>	<u>Support</u>
<u>Roti, mentega, telur, susu</u>	40%	2

2.2.1. 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 3. Siklus Pengembangan *Data Mining* menggunakan CRISP_DM

Berikut ini adalah enam tahap siklus hidup pengembangan *data mining* :

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. *Data Preparation* (Fase Pengolahan Data)
Tahap ini mencakup semua kegiatan untuk membangun dataset akhir (data yang akan diproses pada tahap pemodelan/modeling) dari data mentah. Tahap ini dapat diulang beberapa kali. Pada tahap 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 *data mining* 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 (*Business Understanding*). 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 analisis data, karena sangat penting bagi konsumen untuk memahami tindakan apa yang harus dilakukan untuk menggunakan model yang telah dibuat.

2.2.2. RapidMiner



Gambar 4. 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.

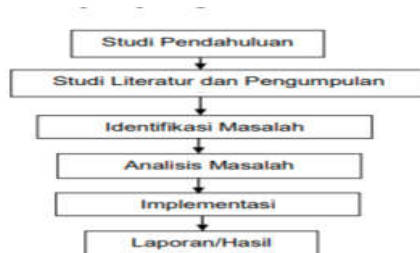
RapidMiner sebelumnya bernama YALE (*Yet Another Learning Environment*), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di *Artificial Intelligence Unit* dari *University of Dortmund*. RapidMiner didistribusikan di bawah lisensi AGPL (*GNU 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 di dunia.

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 *file XML* (*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 skala besar dan otomatisasi eksperimen.
4. Konsep *multi-layer* untuk menjamin tampilan data yang efisien dan menjamin penanganan data.

3. Kerangka Konsep



Gambar 5. Konsep berfikir

Pada penelitian ini adalah penggunaan data mining apriori untuk mengetahui pola belanja pada Mini

Market Indomaret di Ruko Ventura, hal ini ditujukan untuk mengetahui item barang apa saja yang mudah laku. Metode yang dipakai adalah penggunaan Data Mining, Alogaritma Apriori, dengan aplikasi CRISP-DM dan Rapid Miner.

4. Desain Penelitian/ Metodologi

4.1. Pengumpulan data

Pada penelitian ini, peneliti mengambil data transaksi yang diperoleh dari pengumpulan data yaitu sebanyak 973 data transaksi penjualan namun tidak semua data dapat digunakan karena harus melalui beberapa proses pengolahan data sehingga data yang dapat digunakan yaitu sebanyak 250 data transaksi penjualan dan 723 data tidak dapat digunakan. Data yang tidak dapat digunakan tersebut diidentifikasi dari beberapa factor diantaranya:

1. Transaksi berupa satu macam barang.
2. Transaksi berupa rokok.
3. Transaksi berupa pembelian pulsa dan *payment* (virtual)

Untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan oleh penulis yaitu sebagai berikut:

1. Pembentukan kandidat itemset
Kandidat k-itemset dibentuk dari kombinasi (k-1) itemset yang didapat dari iterasi sebelumnya. Satu cara dari algoritma apriori adalah pemangkasan kandidat k-itemset yang subsetnya berisi k-1 item tidak termasuk dalam pola frekuensi tinggi dengan panjang K-1.
2. Perhitungan support dari tiap kandidat k-itemset
Support pada tiap kandidat k-itemset didapat dengan menscan database untuk menghitung jumlah transaksi yang memuat semua item didalam kandidat k-itemset tersebut. Ini adalah juga ciri dari algoritma apriori dimana diperlukan perhitungan dengan cara seluruh database sebanyak k-itemset terpanjang.
3. Tetapkan pola frekuensi tinggi
Pola frekuensi tinggi yang memuat k-itemset ditetapkan dari kandidat k-itemset yang supportnya lebih besar dari minimum support.
4. Bila tidak didapat pola frekuensi tinggi baru maka seluruh proses dihentikan.

4.2. 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.

1. Business Understanding (Fase Pemahaman Bisnis)
Penelitian ini memiliki tujuan bisnis untuk menentukan aturan asosiasi yang terbentuk dari transaksi penjualan pada toko Indomaret ruko Ventura. 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*.
2. Fase Pemahaman Data (Data Understanding Phase)
Pengumpulan data primer dilakukan dengan cara mengambil data laporan transaksi penjualan pada toko Indomaret ruko Ventura selama bulan Agustus 2018.

5. Hasil Penelitian

5.1. Persiapan data

Tahap awal yang dilakukan oleh peneliti pada penelitian ini yaitu persiapan data yang akan digunakan pada penelitian ini. Data yang digunakan yaitu data penjualan pada toko Indomaret ruko Ventura pada bulan Agustus 2018. Data yang didapatkan yaitu 250 *record* data.

5.2. Pengolahan data dengan CRISP –DM

Dalam penelitian ini, peneliti menggunakan CRISP-DM sebagai metode analisis dan pemodelan dengan tahapan fase sebagai berikut.

1. Business Understanding (Fase Pemahaman Bisnis)
Penelitian ini memiliki tujuan bisnis untuk menentukan aturan asosiasi yang terbentuk dari transaksi penjualan pada toko Indomaret ruko Ventura. 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*.
2. Fase Pemahaman Data (Data Understanding Phase)
Pengumpulan data primer dilakukan dengan cara mengambil data laporan transaksi penjualan pada toko Indomaret ruko Ventura selama bulan Agustus 2018.
3. Fase Pengolahan Data (Data Preparation Phase)
Data diambil dari laporan penjualan disimpan dalam bentuk file excel atau *.xlsx atau *.xls. Data transaksi lalu dibersihkan dengan cara menghapus atribut-atribut yang tidak diperlukan pada penelitian yang dilakukan.

5.3. Menjalankan Rapid Miner

Aplikasi RapidMiner merupakan aplikasi yang digunakan untuk proses *data mining*. Di dalam aplikasi ini terdapat banyak algoritma yang terkandung guna kebutuhan data mining. Salah satu algoritma yang akan digunakan pada penelitian ini yaitu Apriori.



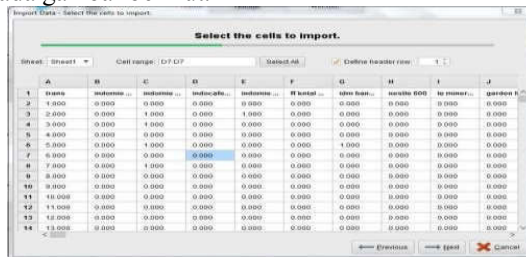
Gambar 6. Halaman Utama Aplikasi

Selanjutnya Tools Add data seperti dibawah ini :



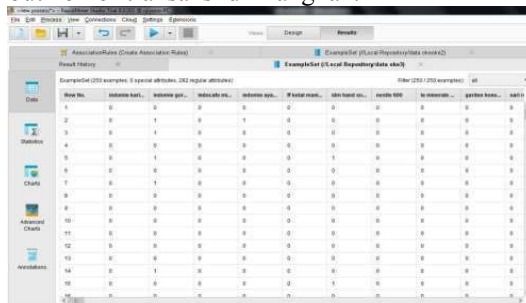
Gambar 7. Tools Add Data

Pada tampilan gambar diatas merupakan *tool* untuk menambahkan data yang telah dipersiapkan untuk dilakukan penelitian melalui pilihan *add data* kemudian pilih letak file yang akan di *input*. Setelah file di pilih, maka tampilan akan seperti pada gambar berikut.



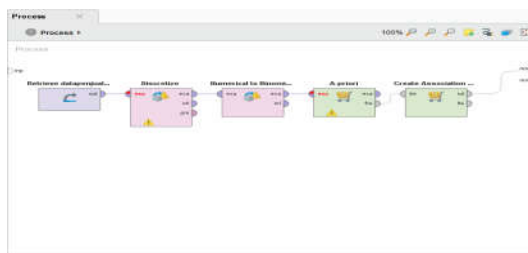
Gambar 8. Lembar Import Data

Pada gambar diatas dapat dijelaskan bahwa semua data telah berhasil di *input* ke aplikasi RapidMiner tanpa ada *error*. Namun untuk atribut nomor transaksi tidak diperlukan, karena akan mempengaruhi hasil dari penelitian, sehingga untuk atribut nomor transaksi dihilangkan.



Gambar 9. Tampilan Data pada RapidMiner

Langkah selanjutnya yaitu proses *data mining* menggunakan *tools* pada aplikasi RapidMiner.



Gambar 10. Proses *Data Mining* Apriori

Pada penelitian ini, ditentukan nilai minimum *support* yang digunakan adalah 6% dan *confidence* 70%. Maksudnya yaitu penulis membatasi jumlah itemset yang mempunyai nilai *support* minimal 6% atau dapat dikatakan 0,06 dari data dan memiliki nilai *confidence* minimal 70% atau 0,7. Setelah dijalankan, maka data yang ditampilkan adalah sebagai berikut :

No.	Premises	Conclusion	Support	Confidence
1.	ekonomi ep 900	idm p lantai aple	0.020	0.714
2.	kopiko 78 caramel	anget sari susu jahe	0.016	0.800
3.	lotte choco pie 165	energen vanilla 5	0.016	0.800
4.	gliko pocky choban	ultra uhtcho 125	0.016	0.800
5.	pop mie baso	pop mie soto	0.016	0.800
6.	sari roti tawar double	inti gandum banana	0.016	0.800
7.	inti gandum banana	sari roti tawar double	0.016	0.800
8.	idm p lantai aple, idm cotton bud bby	ekonomi ep 900	0.016	0.800
9.	idm p lantai aple, ekonomi ep 900	idm cotton bud bby	0.016	0.800

Gambar 11. Hasil Proses *Data Mining*

5.4. Resume Data

1. Pada transaksi penjualan terdapat 71% transaksi mengandung item ekonomi ep 900 dan idm p lantai aple serta 2% di transaksi mengandung kedua item.
2. Pada transaksi penjualan terdapat 80% transaksi mengandung item kopiko 78 caramel dan anget sari susu jahe serta 1.6% di transaksi mengandung kedua item.
3. Pada transaksi penjualan terdapat 80% transaksi mengandung item lotte choco pie 165 dan energen vanilla serta 1.6% di transaksi mengandung kedua item.
4. Pada transaksi penjualan terdapat 80% transaksi mengandung item gliko pocky choban dan ultra uht cho 125 serta 1.6% di transaksi mengandung kedua item.
5. Pada transaksi penjualan terdapat 80% transaksi mengandung item pop mie baso dan pop mie soto serta 1.6% di transaksi mengandung kedua item.
6. Pada transaksi penjualan terdapat 80% transaksi mengandung item sari roti double dan inti gandum banana serta 1.6% di transaksi mengandung kedua item. Pada transaksi penjualan terdapat 80% transaksi mengandung item inti gandum banana dan sari roti double serta 1.6% di transaksi mengandung kedua item.
7. Pada transaksi penjualan terdapat 80% transaksi mengandung item idm p lantai aple, idm cotton bud bby dan ekonomi ep 900 serta 1.6% di transaksi mengandung ketiga item.
8. Pada transaksi penjualan terdapat 80% transaksi mengandung item idm p lantai aple, ekonomi ep 900 dan idm cotton bud bbyserta 1.6% di transaksi mengandung ketiga item.

6. Kesimpulan

Berdasarkan Penelitian diatas didapat beberapa kesimpulan sebagai berikut :

Setelah melalui tahapan pembahasan penelitian penerapan algoritma Apriori untuk menentukan penjualan produk pada minimarket studi kasus Indomaret ruko Ventura, didapatkan kesimpulan bahwa pada penelitian tersebut dapat menemukan kombinasi dari produk yang sering dibeli oleh konsumen secara bersamaan dan diharapkan dapat meningkatkan barang yang terjual pada toko Indomaret ruko Ventura. Hasil dari penelitian ini dapat menjawab permasalahan yang diangkat oleh peneliti.

Daftar Pustaka

Ahmad Fauzi, et.al. 2017. “Sistem Pendukung Pemilihan Pekerjaan Menggunakan Metode A-priori Berdasarkan Korelasi Jurusan Dengan IPK untuk Mengetahui Pekerjaan yang Tepat”.vol. 5, p.57, Sep.2016.

D. Samuel, “Penerapan Struktur Fp-Tree dan Algoritma FP-Growth dalam Optimasi Penentuan Frequent Itemset”, pp.2, Des.2007

Fajar,H. 2013. “Data Mining”.Andi. Yogyakarta.

Fatihatul, A. Setiawan and R. Rosadi, “Asosiasi Data Mining Menggunakan Algoritma FP-Growth Untuk Market Basket Analysis”, pp.2-4, Agu. 2011.

Larose D, T. 2006. *Data Mining Methods and Models*.Jhon Willey and Son, Inc Hoboken New Jersey

- Nuqson,M, "Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa (Studi Kasus di Fakultas MIPA Universitas Diponegoro)," Jurnal Teknik Informatika Jurusan MIPA, p. 90, 2010.
- Pramudiono. 2007. *Pengantar Data Mining : Menambang Permata Pengetahuan di Gunung Data*, Andi, Yogyakarta
- Tan, P. et al, *Introduction to Data Mining.*, Boston: Pearson Education, 2006.
- Turban, et al.2005. *Decision Support System and Intellegent System EdisiBahasa Indonesia Jilid 1*, Andi, Yogyakarta