

Data Mining Menggunakan Algoritma Apriori Untuk Analisis Penjualan

Hasyrif SY, Rismayani, Novita Sambo Layuk

STMIK Dipanegara Makassar

Jl. Perintis Kemerdekaan Km 9 Makassar, telp 0411-587194 fax. 0411-588284

e-mail: hasyrif@gmail.com, maya_setya@ymail.com, fivhy@yahoo.co.id

Abstrak

Penjualan merupakan salah satu bidang bisnis yang banyak digeluti masyarakat, bahkan bisnis penjualan produk akhir akhir ini meningkat pesat dengan terjadinya peningkatan teknologi komunikasi khususnya terkait sosial media, aspek yang penting dalam penjualan salah satunya adalah terkait histori transaksi penjualan, semakin besar skala penjualan semakin banyak pula transaksi yang terjadi. Record transaksi penjualan yang besar selain bisa digunakan untuk menghitung untung rugi, data ini juga bisa digunakan untuk keperluan lainnya. Perancangan aplikasi dibuat dengan menggunakan model waterfall, dalam tahap ini dibagi menjadi enam tahapan yaitu system engineering, design, code, testing dan maintenance. Tulisan ini dibuat dengan tujuan memberikan gambaran tentang penggunaan algoritma apriori untuk analisa transaksi penjualan khususnya terkait dengan kecenderungan dibelinya suatu barang terhadap barang lain. Output penelitian berupa aplikasi data mining menggunakan algoritma apriori dalam penentuan itemset dan aturan asosiasi.

Kata kunci: Penjualan, Algoritma Apriori, Data Mining

Abstract

Sales is one of the business areas that many people cultivate, even business sales of the end of this product increased rapidly with the increase in communication technology, especially related to social media, an important aspect in the sale of one is related to the history of sales transactions, the greater the scale of sales more and more The transaction occurred. Record large sales transactions than can be used to calculate profit and loss, this data can also be used for other purposes. Application design is made by using waterfall model, in this stage is divided into five stages of system engineering, design, code, testing and maintenance. This paper was created with the aim of providing an overview of the use of a priori algorithm for the analysis of sales transactions, especially related to the tendency to buy an item against other barang. Output research in the form of data mining applications using a priori algorithm in determining itemset and association rules.

Keywords: Sales, Apriori Algorithm, Data Mining

1. Pendahuluan

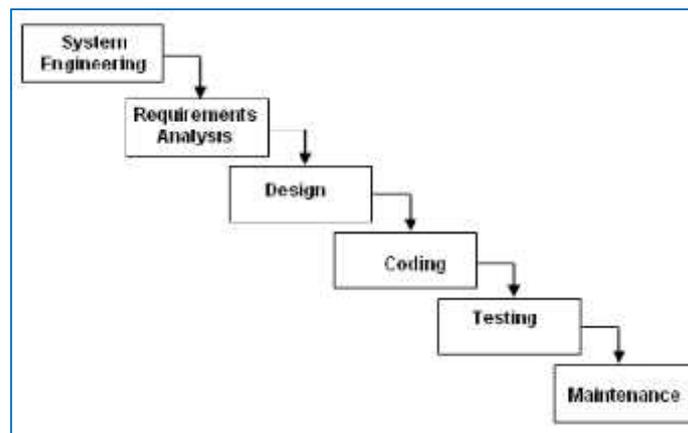
Penjualan merupakan salah satu pekerjaan yang banyak digeluti oleh masyarakat. Terlebih lagi disaat ini kemajuan teknologi komunikasi mempengaruhi pola interaksi masyarakat, yang tadinya interaksinya bersifat fisik berubah pada interaksi secara tidak langsung melalui perantara alat komunikasi, saat ini aktifitas pertemanan, rapat, perkuliahan dan lainnya tidak mengharuskan kita untuk bertemu langsung. Fenomena ini juga berlaku pada pola belanja masyarakat. Kini banyak masyarakat kita melakukan transaksi jual beli online. Trend ini disatu sisi memudahkan masyarakat karena bisa memangkas waktu dan tenaga, namun disisi lain mudahnya akses produk memberi peluang kepada pembeli untuk lebih leluasa dalam memilih produk yang akan dipakainya, hal ini “memaksa” para penjual untuk lebih kreatif dalam menjajakan produk mereka, strategi bisnis dalam penjualan menjadi penentu dilirik atau tidaknya produk oleh masyarakat.

Untuk penjualan sebenarnya sudah banyak aplikasi untuk membantu penjual dalam memudahkan proses jual beli, seperti aplikasi penjualan lengkap dengan analisis perhitungan untung dan rugi. Namun aplikasi ini sebenarnya berfungsi untuk mengukur keadaan dari sisi penjualan, seperti

penentuan status stok barang, keuntungan dll, padahal kebutuhan dari aspek pembeli juga itu sendiri menjadi penting untuk di analisa dalam peningkatan penjualan suatu produk.

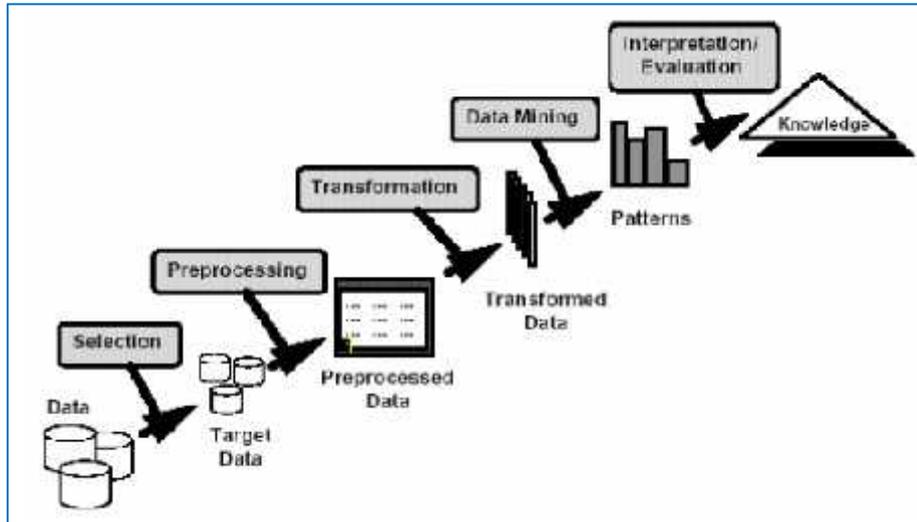
Dalam pembuatan aplikasi ini tentu harus mengikuti salah satu metode yang telah diciptakan untuk memberi patokan dalam pengembangan sebuah software. Dalam pembahasana ini kami memilih menggunakan metode waterfall, metode Waterfall adalah suatu proses pengembangan perangkat lunak berurutan, di mana kemajuan dipandang sebagai terus mengalir ke bawah (seperti air terjun) melewati fase-fase perencanaan, pemodelan, implementasi (konstruksi), dan pengujian. Berikut adalah gambar pengembangan perangkat lunak berurutan/ linear [1]. Sedang menurut Jogiyanto metode warterfall merupakan suatu bentuk pengembangan sistem yang digunakan untuk menggambarkan tahapan utama dan langkah-langkah ditahapan tersebut[2] Tahapan-tahapan pengembangan model waterfall yaitu:

1. Rekayasa perangkat lunak (system engineering),melakukan pengumpulan data danpenetapan kebutuhan semua elemen system.
2. Requirements analysis , melakukan analisis terhadap permasalahan yang dihadapi dan menetapkan kebutuhan perangkat lunak, fungsi performsi dan interfacing
3. Design, menetapkan domain informasi untuk perangkat lunak,fungsi dan interfacing
4. Coding (implementasi), pengkodean yang mengimplementasikan hasil desain kedalam kode atau bahasa yang dimengerti oleh mesin komputer dengan menggunakan bahasa pemrograman tertentu.
5. Testing (pengujian) , kegiatan untuk melakukan pengetesan program yang sudah dibuat apakah udah benar atau belum di uji dengan cara manual.jika testing sudah benar maka program boleh digunakan.
6. Maintenance (perawatan) , menangani perangkat lunak yang sudah selesai supaya dapat berjalan lancar dan terhindar dari gangguan-gangguan yang dapat menyebabkan kerusakan.



Gambar 1. Tahapan Metode Waterfall

Dalam melakukan analisis penjualan idealnya dibutuhkan sebuah koleksi dari ribuan bahkan lebih transaksi dari proses yang sedang dianalisa, record yang sangat banyak ini membutuhkan sebuah ilmu dalam membedah makna-makna yang terkandung didalamnya salah satu konsep terkait yang dibutuhkan dalam analisa record yang sangat banyak ini adalah konsep data mining, data mining adalah proses ekstraksi informasi dari kumpulan data melalui penggunaan algoritma dan teknik yang melibatkan bidang ilmu statistik, mesin pembelajaran, dan system manajemen database [3]. Data Mining digunakan untuk ekstraksi informasi penting yang tersembunyi dari dataset yang besar. Dengan adanya data mining maka akan didapatkan suatu permata berupa pengetahuan di dalam kumpulan data – data yang banyak jumlahnya. Data mining merupakan bagian dari proses Knowledge Discovery in Databases (KDD) – Proses transformasi data mentah menjadi informasi berguna.



Gambar 2. Proses transformasi data mentah menjadi informasi berguna

Tahap-tahap dari proses Knowledge Discovery in Databases (KDD) diperlihatkan dalam point dibawah :

- Selection, penyeleksian atau segmentasi data berdasarkan kriteria ttt.
- Preprocessing, tahap pembersihan dimana info. tidak berguna dibuang. Selain itu data dikonfigurasi ulang utk menjamin format tetap konsisten.
- Transformation, proses transformasi sehingga data dapat digunakan dan ditelusuri. Pemetaan data kompleks
- Data mining, proses ekstraksi pola dari data yg ada.
- Interpretation & evaluation, proses interpretasi pola menjadi pengetahuan yg dapat digunakan utk mendukung pengambilan keputusan (cth prediksi & klasifikasi, ringkasan konten database serta penjelasan fenomena yg diamati).

Algoritma apriori merupakan salah satu algoritma yang digunakan dalam data mining. Algoritma ini merupakan suatu model kerja dalam mencari keterkaitan antar satu atau lebih item dalam suatu dataset. Algoritma ini seringkali digunakan pada rekaman transaksi atau biasa disebut market basket, misalnya sebuah toko online memiliki keranjang belanja online, dengan adanya algoritma apriori, kita dapat mengetahui pola belanja seorang pembeli, jika seorang pembeli memilih item X , Y maka ítem X akan dipilih dengan prosentase 50%. Algoritma apriori adalah algoritma pengambilan data dengan aturan asosiatif (Association rule) untuk menentukan hubungan asosiatif suatu kombinasi item [4]. Association Rule yang dimaksud dilakukan melalui mekanisme penghitungan support dan confidence dari suatu hubungan item.

Berangkat dari permasalahan diatas maka penulis mencoba untuk mengangkat salah satu sisi yang menjadi faktor pendukung proses penjualan, yaitu bagaimana mengetahui keterkaitan antara setiap barang dengan barang lainnya dari sisi pembeli, dengan kata lain bagaimana agar kita bisa meramal keinginan pembeli dengan melihat data barang yang telah dibeli sebelumnya. Tulisan ini akan membahas aplikasi datamining dalam menganalisa penjualan menggunakan algoritma apriori. Algoritma apriori merupakan salah satu algoritma yang digunakan dalam datamining, algoritma ini merupakan jenis aturan asosiasi pada data mining. Algoritma ini digunakan untuk tujuan untuk menentukan frequent item sets. Untuk menerapkan aplikasi ini akan dibuat simulasi kasus penjualan sederhana untuk dibedah.

2. Metode Penelitian

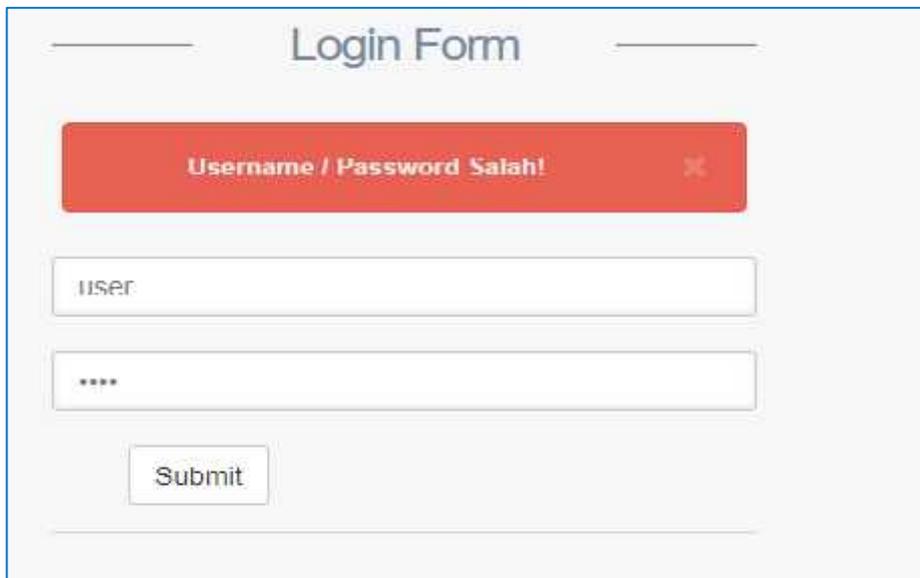
2.1. Metode

Terdapat beberapa metode yang digunakan dalam pengembangan software, dalam tulisan ini pembangunan aplikasi dirancang menggunakan model waterfall, dalam tahap ini langkah-langkah pengembangan dibagi menjadi enam tahapan yaitu system engineering, analisis, design, code, testing dan maintenance.

3. Hasil dan Pembahasan

3.1 Hasil

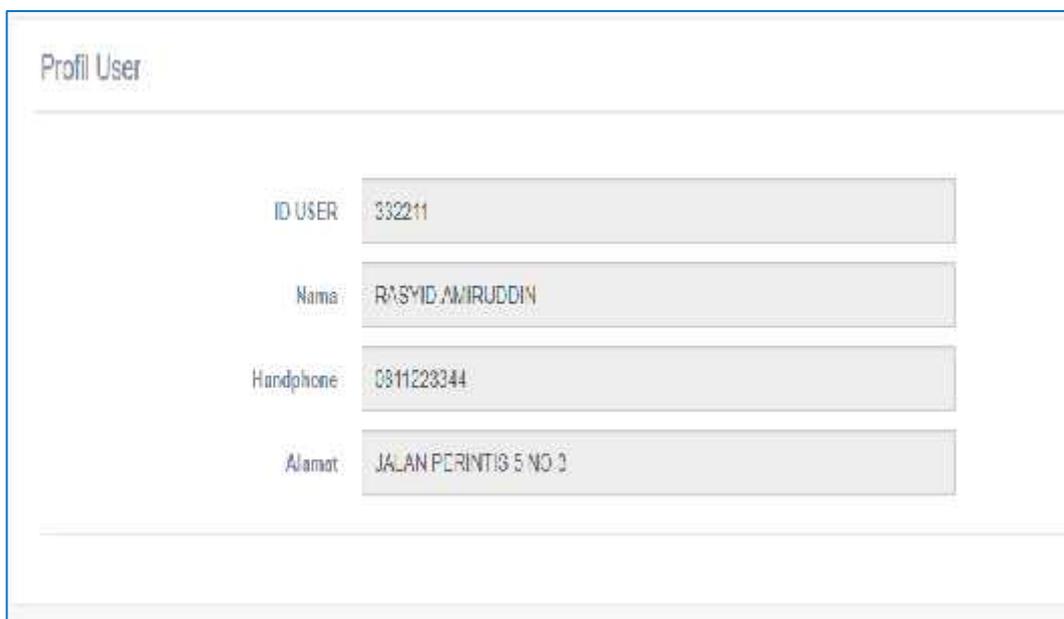
Halaman Login



The image shows a web page titled "Login Form". At the top, there is a red error message box that says "Username / Password Salah!" with a close button (X). Below the error message, there are two input fields: the first contains the text "user" and the second contains four asterisks "****". Below the input fields is a "Submit" button.

Gambar 3. Halaman Login

Profil User



The image shows a web page titled "Profil User". It displays user information in a list format:

ID USER	332211
Nama	RASYID AMIRUDDIN
Handphone	0811223344
Alamat	JALAN PERINTIS 5 NO 2

Gambar 4. Profil User

Histori Harian

HISTORI HARIAN		
#	KODE PENJUALAN	DETAIL DATA
1	Tr 4561	AIR MINERAL, ROTI, MIE
2	Tr 4562	ROTI, MIE TELUR, SUSU
3	Tr 4563	KEJU, TELUR, SUSU
4	Tr 4564	ROTI, MIE
5	Tr 4565	ROTI, MIE TISU, TELUR, SUSU

Gambar 5. Histori Harian

Kombinasi 1 Itemset

C1			
#	ITEMSET	JUMLAH	SUPPORT
1	KACANG	4	80%
2	MIE	4	80%
3	WAJER	3	60%
4	SUSHI	3	60%
5	KEJU	1	20%
6	TISU	1	20%
7	AIR MINERAL	1	20%

Gambar 6. Kombinasi 1 Itemset

Kombinasi 2 Itemset



#	ITEMSET	JUMLAH	SUPPORT
1	KACANG, MIE	4	60%
2	KACANG, WAFER	2	40%
3	KACANG, SUSU	2	40%
4	MIE, WAFER	2	40%
5	MIE, SUSU	2	40%
6	WAFER, SUSU	3	60%

Gambar 7. Kombinasi 2 Itemset

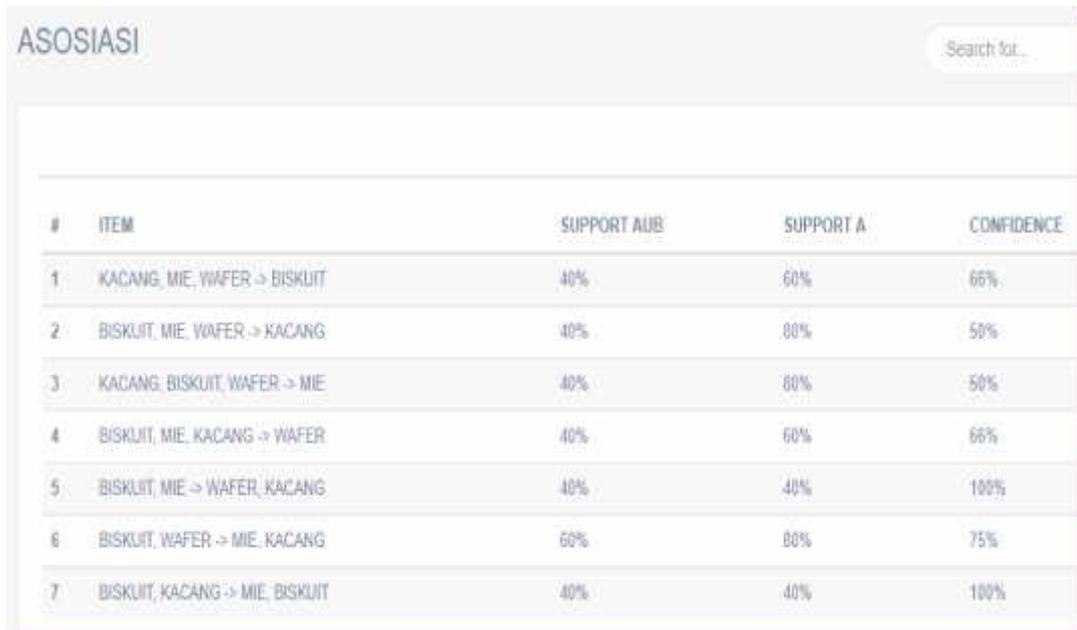
Kombinasi 3 Itemset



#	ITEMSET	JUMLAH	SUPPORT
1	ROTI, MIE, TELUR	2	40%
2	ROTI, MIE, SUSU	2	40%
3	MIE, TELUR, SUSU	2	40%

Gambar 8. Kombinasi 3 Itemset

Aturan Asosiasi



#	ITEM	SUPPORT AUB	SUPPORT A	CONFIDENCE
1	KACANG, MIE, WAFER → BISKUIT	40%	60%	66%
2	BISKUIT, MIE, WAFER → KACANG	40%	80%	50%
3	KACANG, BISKUIT, WAFER → MIE	40%	80%	50%
4	BISKUIT, MIE, KACANG → WAFER	40%	60%	66%
5	BISKUIT, MIE → WAFER, KACANG	40%	40%	100%
6	BISKUIT, WAFER → MIE, KACANG	60%	80%	75%
7	BISKUIT, KACANG → MIE, BISKUIT	40%	40%	100%

Gambar 9. Aturann Asosiasi

3.2. Pembahasan

Data Awal

Aplikasi data mining membutuhkan sumber data berupa histori transaksi penjualan perperiode tertentu. Berikut diperlihatkan contoh dari sebuah transaksi penjualan.

Tabel 1. Contoh transaksi penjualan

No	Kode penjualan	Detail data
1	Tr 4561	Air mineral, roti, mie
2	Tr 4562	Roti,mie,telur,susu
3	Tr 4563	Keju,telur,susu
4	Tr 4564	Roti, mie
5	Tr 4565	Roti, mie, tisu, telur, susu

Penentuan C1

Setelah mengamati tabel transaksi penjualan diatas, langkah berikutnya yaitu menentukan C1 atau sering disebut dengan satu itemset. Proses yang dilakukan adalah melakukan seleksi produk berdasarkan apa yang disebut dengan batas minimum support, misal kita akan menentukan batas minimum support lebih besar dari dua atau 30%, rumus yang dapat digunakan untuk menentukan persentase support adalah :

$$= \frac{\sum \text{Transaksi Mengandung A}}{\sum \text{Transaksi}} * 100\%$$

Tabel 2. Frekuensi kemunculan produk

No	Itemset	Jumlah	Support
1	Roti	4	80%
2	Mie	4	80%
3	Telur	3	60%
4	Susu	3	60%
5	Keju	1	20%
6	Tisu	1	20%
7	Air mineral	1	20%

Di pembahasan sebelumnya kita menentukan bahwa batas minimum support adalah 30% (Minimal) 2, berdasarkan prinsip ini maka akan tersisa 4 produk yang memenuhi syarat diatas, seperti yang diperlihatkan tabel dibawah.

Tabel 3. Frekuensi kemunculan produk dengan batas minimum support 30% (C1)

No	Itemset	Jumlah	Support
1	Roti	4	80%
2	Mie	4	80%
3	Telur	3	60%
4	Susu	3	60%

Penentuan C2

Untuk penentuan C2 (dua itemset), dilakukan dengan melakukan kombinasi dua-dua produk dengan variasi yang unik, item item yang akan di proses merupakan item yang telah diseleksi dari proses sebelumnya atau dengan kata lain yaitu C1, formula yang digunakan yaitu:

$$= \frac{\sum \text{Transaksi Mengandung A dan B}}{\sum \text{Transaksi}} * 100\%$$

Tabel 4. Seleksi C2

No	Itemset	Jumlah	Support
1	Roti, mie	4	80%
2	Roti, telur	2	20%
3	Roti, susu	2	20%
4	Mie, telur	2	20%
5	Mie, susu	2	20%
6	Telur, susu	3	60%

Penentuan C3

Pada prinsipnya penentuan C3 (tiga item set) dilakukan dengan mengikuti prinsip yang sama pada penentuan c2, yang berbeda hanyalah jumlah itemnya, sumber item C3 diambil dari kombinasi dua item yang memenuhi syarat batas minimum support. Rumus yang digunakan yaitu

$$= \frac{\sum \text{Transaksi Mengandung A, B dan C}}{\sum \text{Transaksi}} * 100\%$$

Tabel 5. Seleksi C3

No	Itemset	Jumlah	Support
1	Roti, mie, telur	2	40%
2	Roti, mie, susu	2	40%
3	Mie, telur, susu	2	40%

Penentuan C4

Tabel 6. Seleksi C4

No	Itemset	Jumlah	Support
1	Roti, mie, telur, susu	2	40%

Proses berhenti ketika tidak ada lagi record yang bisa diproses atau itemset tidak lagi memenuhi batas minimum support, untuk kasus diatas karena record tinggal satu (tidak ada lagi itemset yang bisa diproses) maka secara otomatis proses berhenti hingga pada C4.

Pembentukan Aturan Asosiasi

Setelah ditemukan item dengan frekuensi tinggi tahapan terakhir adalah pembentukan aturan asosiasi yang memenuhi syarat minimum dengan melakukan perhitungan confidence association rules A -> B.

Tabel 7. Hasil Confidence

No	Aturan asosiatif	Support aub	Support a	Confidence
1	Kacang, mie, wafer -> biskuit	40%	60%	66%
2	Biskuit, mie, wafer -> kacang	40%	80%	50%
3	Kacang, biskuit, wafer -> mie	40%	80%	50%
4	Biskuit, mie, kacang -> wafer	40%	60%	66%
5	Biskuit, mie -> wafer, kacang	40%	40%	100%
6	Biskuit, wafer -> mie, kacang	60%	80%	75%
7	Biskuit, kacang -> mie, biskuit	40%	40%	100%
8	Mie, wafer -> biskuit, kacang	40%	40%	100%
9	Mie, kacang -> biskuit, wafer	80%	60%	75%
10	Wafer -> mie, kacang	40%	40%	100%
11	Biskuit -> mie, kacang	40%	80%	50%
12	Biskuit -> mie, wafer	60%	80%	75%
13	Mie -> biskuit, wafer	60%	40%	66%
14	Mie -> biskuit, kacang	80%	60%	75%
15	Mie -> biskuit	80%	40%	50%
16	Mie -> kacang	80%	60%	75%
17	Mie -> wafer	80%	80%	100%
18	Biskuit -> kacang	60%	60%	75%
19	Biskuit -> wafer	60%	80%	75%
20	Wafer -> kacang	60%	60%	100%

4. Kesimpulan

Dari hasil perhitungan itemset (C1 hingga C4) ditemukan bahwa itemset dengan pola frequent tinggi yaitu Kacang, Mie, Wafer, Biskut. Aturan asosiasi sebagai tindak lanjut penentuan itemset dengan pola frekuensi tinggi di perlihatkan pada tabel 7.

Referensi

- [1] Pressman, Roger S. 2012. *Rekayasa Perangkat Lunak – Buku Satu, Pendekatan Praktisi (Edisi 7)*. Yogyakarta: Andi.
- [2] Jogiyanto. 2008. [Analisis Dan Desain Sistem Informasi : Pendekatan Terstruktur Teori Dan Praktek Aplikasi Bisnis](#)
- [3] Ranjan, J., 2007, Application of Data Mining Technique in Pharmaceutical Industry, *Journal of Theoretical and Applied Information Technology*, Vol 3, hal 61 – 67.
- [4] Kusriani, 2007, Penerapan Algoritma Apriori pada Data Mining untuk Mengelompokkan Barang Berdasarkan Kecenderungan Kemunculan Bersama dalam Satu Transaksi, http://dosen.amikom.ac.id/.../Publikasi%20Apriori-Kusriani_Feb-13_.pdf, diakses 20 Juli 2011.