

APLIKASI DATA *MINING* MENGGUNAKAN ATURAN ASOSIASI DENGAN ALGORITMA *PINCER SEARCH* UNTUK MENGANALISIS DATA TRANSAKSI PENJUALAN

Studi Kasus: Data Transaksi Penjualan Yomart Margacinta Bandung

Mochamad Ikbal Ashari, Indwiarti, Yuliant Sibaroni

Prodi Ilmu Komputasi Telkom University, Bandung
Universitas Telkom, Bandung

mikbalashari@gmail.com, indwindwi@gmail.com, ysibaroni@gmail.com

Abstrak

Dunia bisnis ritel mengalami pertumbuhan yang sangat pesat, hal ini ditandai meningkatnya jumlah ritel. Umumnya setiap ritel menggunakan sistem komputerisasi dalam penyimpanan data penjualan dan pembelian yang menghasilkan data transaksi. Data transaksi tersebut dapat digunakan sebagai sumber informasi penting dalam mempertahankan keberadaan usaha tersebut. Salah satu upaya yang dilakukan untuk dapat mempertahankan bisnis ritel ini adalah dengan meningkatkan dan mempertahankan tingkat pertumbuhannya.

Penelitian ini bertujuan untuk membangun sebuah aplikasi analisis keranjang belanja (*market basket analysis*) yang berkaitan dengan data transaksi penjualan yang menggunakan aturan asosiasi, mengidentifikasi beberapa tipe dari kaidah asosiasi yang berkaitan dengan data transaksi penjualan yaitu nilai-nilai *support* dan *confidence*, mengetahui kombinasi dari *item-item* yang sering dibeli bersamaan oleh konsumen. Penelitian diolah dengan menggunakan software *Microsoft Office Excel 2010* dan software *matlab 2013*. Data dianalisis menggunakan algoritma *pincer-search* yang menghasilkan aturan asosiasi dengan pola "*if-then*".

Hasil dari Aplikasi Analisis Keranjang Belanja adalah *rule-rule* yang menerangkan kecenderungan pelanggan dalam membeli barang-barang yang dibeli secara bersamaan, *rule-rule* tersebut akan digunakan untuk menentukan penempatan *item-item* agar supermarket dapat memaksimalkan keuntungan yang dapat diperoleh supermarket, dalam studi kasus ini adalah Yomart.

Kata Kunci: *Market Basket Analysis, Algoritma Pincer-search, Support, Confidence*

Abstract

Retail business experienced rapid growth, it is marked increase in the number of retail. Generally every retail uses a computerized system in data storage sales and purchases that generate transaction data. The transaction data can be used as a source of information is important in maintaining the existence of the business. One of the efforts made to maintain the retail business is to increase and maintain the level of sales growth.

This study aims to build a shopping cart application analysis (*market basket analysis*) related to sales transaction data using association rules, identifying several types of association rules relating to sales transaction data that values the support and confidence, knowing combining of items -Item are frequently purchased together by the consumer. Research processed using *Microsoft Office Excel 2010* software and software *matlab 2013*. Data were analyzed using a *pincer-search* algorithm that generates association rules with pattern "*if-then*".

Results of Application Analysis Shopping Cart is some rules that explain the tendency of customers to purchase items purchased at the same time, some rules will be used to determine the placement of items in order to maximize profits supermarket can be obtained supermarket, in this case study is Yomart.

Keywords: *Market Basket Analysis, Pincer-search algorithm, Support, Confidence*

1. Pendahuluan [10 pts/Bold]

Pada dunia yang perkembangan teknologinya berkembang pesat ini, tingkat persaingan ekonomi yang terjadi juga semakin tinggi. Seorang pengambil keputusan suatu perusahaan, dituntut untuk dapat melihat peluang-peluang yang dapat meningkatkan penjualan perusahaan tersebut. Diperlukan kreativitas dan inovasi dari produsen agar dapat memaksimalkan peluang-peluang yang ada untuk

menaikkan tingkat penjualannya. Suatu perusahaan pasti memiliki sekumpulan data yang dapat di olah agar dapat menghasilkan suatu informasi yang bermanfaat untuk perusahaan tersebut. Tetapi jika jumlah data yang dimiliki perusahaan tersebut sangat besar, akan sulit dilakukan analisa secara manual. Untuk itulah diperlukan sebuah kreatifitas dan inovasi yang bisa membantu perusahaan dalam menganalisa data-data tersebut dengan cepat dan

tepat, sehingga dapat dihasilkan suatu knowledge yang berguna bagi peningkatan efisiensi dan efektivitas perusahaan, salah satunya adalah dengan data mining.

Penerapan *association rules* dalam menganalisis data transaksi penjualan dapat meningkatkan penjualan perusahaan tersebut, karena menempatkan *item-item* berdasarkan hasil analisis *association rules* secara tidak langsung mempermudah pelanggan dalam berbelanja, sehingga pelanggan tersebut merasa senang berbelanja di perusahaan tersebut [12]. Hal ini membantu perusahaan tersebut agar tetap bisa bertahan di dunia bisnis ritel.

1.1 Rumusan Masalah

Berdasarkan latar belakang yang dijelaskan diatas, rumusan masalah yang didapatkan adalah sebagai berikut :

1. Bagaimana cara mengimplementasikan Algoritma *Pincer Search* untuk menganalisa data transaksi penjualan ?
2. Bagaimana cara mengetahui performa Algoritma *Pincer Search* untuk menganalisa data transaksi penjualan ?
3. Bagaimana cara menemukan asosiasi antar *item* yang dijual ?

1.2 Tujuan Penelitian

Adapun Tujuan Penelitian yang dilakukan berdasarkan rumusan masalah di atas adalah:

1. Mengimplementasikan Algoritma *Pincer Search* dalam menganalisa data transaksi penjualan.
2. Untuk Mengetahui performa Metode *Pincer Search* dalam menganalisa data transaksi penjualan.
3. Merancang dan membangun aplikasi menggunakan algoritma *Pincer Search* (asosiasi data mining) yang dapat menampilkan asosiasi dari *item-item*.

1.3 Batasan Masalah

Dari permasalahan yang telah diuraikan maka batasan dalam tugas akhir ini sebagai berikut:

1. Data yang digunakan adalah data transaksi penjualan Yomart Margacinta dengan periode waktu selama 1 tahun (Bulan Januari 2009 sampai Desember 2009).
2. Membandingkan hasil dari algoritma *pincer search* dengan algoritma apriori, tidak dengan algoritma lainnya.

2. Landasan Teori

2.1 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database [2]. *Data Mining* adalah proses untuk menemukan knowledge dari berbagai database besar menggunakan teknik

statistik, matematika, kecerdasan buatan dan machine learning dengan cara mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari database tersebut.

2.2 Market Basket Analysis

Market basket analysis merupakan salah satu metode atau teknik yang sering digunakan dan paling bermanfaat untuk lingkungan marketing. *Market basket analysis* adalah suatu metodologi untuk melakukan analisis tentang kebiasaan membeli konsumen dengan menemukan asosiasi antar beberapa *item* yang berbeda, yang diletakkan konsumen dalam *shopping basket* (keranjang belanja) yang dibeli pada suatu transaksi tertentu. Tujuan dari *market basket analysis* adalah untuk menentukan produk – produk apa saja yang paling sering dibeli atau digunakan sekaligus oleh para konsumen. *Market Basket Analysis* merupakan salah satu metode atau teknik yang sering digunakan dan paling bermanfaat untuk lingkungan marketing [2].

2.3 Association Rules

Association rules merupakan sebuah aturan tertentu atau *rules* yang menyatakan sebuah hubungan korelasi antara tingkat kemunculan beberapa atribut dalam sebuah database. *Association Rule* adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi *item*. Karena awalnya berasal dari studi tentang *database* transaksi pelanggan untuk menentukan kebiasaan suatu produk dibeli bersamaan dengan produk apa, maka aturan asosiasi juga sering dinamakan *Market basket analysis*.

2.3.1 Support

Support adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu *item/itemset* dari keseluruhan transaksi. *Support* merupakan persen dari *record-record* yang mengandung satu atau beberapa kombinasi dari *item* dibandingkan dengan jumlah total dari *record-record*. Nilai *support* sebuah *item* diperoleh dengan rumus sebagai berikut : jumlah transaksi yang mengandung *item* didalam *rule* total jumlah transaksi keseluruhan x 100% [2].

Keterangan :

A : *Item* jenis A yang dibeli

2.3 Confidence

Confidence adalah suatu ukuran yang menunjukkan hubungan antar 2 *item* secara *conditional*. *Confidence* merupakan prosentase dari *record-record* yang mengandung kombinasi dari *item* dibandingkan dengan jumlah total dari transaksi yang memilih *item* tersebut (misal, seberapa sering *item* B dibeli jika orang membeli *item* A) [2]. Nilai

confidence juga menentukan kuatnya hubungan antar *item* dalam aturan asosiatif

Nilai *Confidence* dari aturan $A \rightarrow B$ diperoleh dari rumus berikut :

$$|$$

Keterangan :

- A : *Item* jenis A yang dibeli
- B : *Item* jenis B yang dibeli
- P : Peluang

2.3.3 Conviction

Conviction adalah perhitungan untuk menentukan nilai akurasi minimum pada metode *association rules*. Pada proses ini dihitung

performansi yaitu akurasi untuk *rule* yang dihasilkan oleh sistem. Mengukur akurasi dari metode yang digunakan dengan rumus :

$$\text{Conviction}(A \rightarrow B) = \frac{\text{Confidence}(A \rightarrow B) - \text{Support}(A)}{\text{Support}(B)}$$

Dimana :

- A,B : *Item*
- Support (B) : Nilai *Support* dari *Item* B
- Confidence (A→B) : Nilai *Confidence* aturan asosiatif (A→B)

Nilai *range* pada *conviction* ini berada pada 0,5,...,1,...,∞ dengan ketentuan *conviction* dianggap memiliki nilai takhingga (*infinite*) apabila nilai dari *confidence* (A→B) sama dengan 1. Jika *conviction* menghasilkan nilai *rules* yang semakin menjauh dari 1 bahkan sampai tak hingga, maka akan di anggap semakin akurat [4].

2.4 Pincer-Search

Pincer-Search dikembangkan oleh Dao-I Lin dan Zvi M. Kedem pada tahun 1997 [6, 7]. Algoritma ini dinamakan *Two-Way Search* karena memakai 2 cara pendekatan, yaitu *Top Down* serta *Bottom Up*. Dalam prosesnya arah pencarian utama dari *Pincer search* adalah *Bottom Up*, yaitu menggunakan Algoritma *Apriori* untuk menemukan *Frequent Itemset*, bedanya pada *Pincer search* secara *concurrent* dilakukan pencarian secara *Top Down*, yaitu *Maximum Frequent Candidate Set* (MFCS) yang menghasilkan output berupa *Maximum Frequent Set* (MFS).

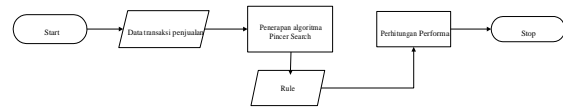
Langkah-langkah perhitungan *pincer-search* [8]:

1. Pisahkan masing-masing item yang dibeli.
2. Bentuk kandidat dari item-item yang dibeli.
3. Bentuk MFCS yg berisi seluruh kandidat item.
4. Hitung *support* dari kandidat.
5. Jika tidak ada kandidat *infrequent*, *generate* kandidat selanjutnya.

6. Jika ditemukan kandidat *infrequent*, kandidat tersebut digunakan untuk mengupdate MFCS hingga tidak ada lagi kandidat yang *infrequent*.
7. Hitung *support* dari MFCS.
8. Update MFS dengan MFCS yang memenuhi nilai minium *support*
9. Pruning kandidat *frequent* yang merupakan *subset* dari MFS.
10. Jika masi ditemukan kandidat yang *frequent* maka akan di *generate* kandidat baru.
11. Langkah no.4 sampai dengan no.9 akan terus di ulang hingga tidak ada lagi kandidat yang bisa di *Join* atau kandidat habis.
12. Output MFS.

3. Analisis dan Perancangan Sistem

3.1 Gambaran Umum Sistem



Berdasarkan gambaran umum di atas, berikut akan dijelaskan mengenai proses pencarian rule beserta akurasinya:

1. Data Transaksi Penjualan
Data yang digunakan diambil dari data transaksi penjualan selama 1 tahun(1 Januari 2009 sampai 31 Desember 2009).
2. Penerapan Algoritma *Pincer-Search*
Pada proses ini data transaksi penjualan tersebut akan diproses dengan algoritma *pincer-search*.
3. Menghitung Akurasi
Setelah proses penerapan algoritma *pincer-search* selesai, rule rule hasil proses tersebut akan dihitung *confidence* dan *convictionnya* pada proses ini.

3.1.1 Pengumpulan Data dan Seleksi Data

a. Pengumpulan Data

Data yang digunakan dalam tugas akhir ini adalah data transaksi sebuah minimarket yomart margacinta (sumber : data tugas akhir kania audrint, 2007). Data tersebut merekam semua data pembelian selama satu tahun (1 Januari – 31 Desember 2009). Berikut beberapa informasi mengenai data transaksi penjualan Super Market :

1. Jumlah record 1 tahun : 10.676 record
2. Jumlah Atribut : 9 Atribut
3. Jumlah record transaksi perhari : rata rata terdapat 296 transaksi

Tabel 0.1 Keterangan Atribut Data

Atribut	Keterangan
atribut 1	Store Code
atribut 2	Date
atribut 3	Order ID
atribut 4	Product ID
atribut 5	Product Description

atribut 6	Quantity
atribut 7	Unit Price
atribut 8	Discount
atribut 9	Discount Price

b. Data Selection

Pada proses *data selection* penulis menggunakan *Feature selection* guna memilih *feature* yang berpengaruh dan mengesampingkan *feature* yang tidak berpengaruh dalam suatu penganalisaan data. Dalam proses Asosiasi atribut yang dibutuhkan adalah *Order ID* dan *Product Description*, oleh karena itu atribut atribut lain dapat dihilangkan karena tidak berpengaruh terhadap proses Asosiasi.

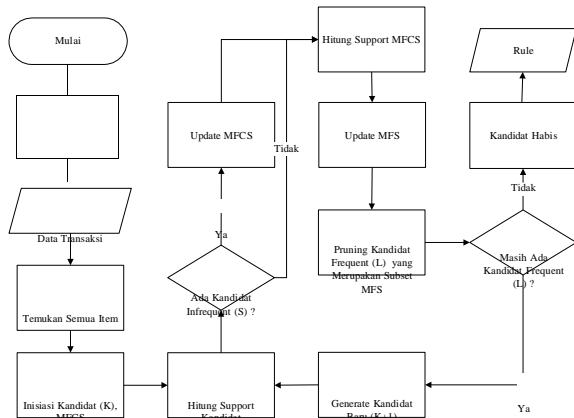
atribut yang tidak berpengaruh terhadap pencarian *association rules* menggunakan algoritma *pincer search* adalah atribut kode transaksi kassa, jumlah *item*, total harga, *discount* dan harga akhir, maka atribut atribut tersebut dapat dihilangkan dari data sehingga data terdiri dari 2 atribut yaitu:

Tabel 0.2 Atribut Setelah Feature Selection

Atribut	Keterangan
atribut 1	ID Transaksi
atribut 2	Jenis Item

3.1.2 Penerapan Algoritma Pincer-Search

Langkah algoritma *pincer-search* adalah sebagai berikut:



Pertama data diinputkan, kemudian akan dicari semua item yang ada di dalam data, setelah semua item ditemukan akan dibentuk kandidat item dan MFCs, selanjutnya kandidat item tersebut akan dihitung *support*nya, jika setelah dihitung *support*nya ditemukan kandidat *infrequent* maka MFCs akan di *update* (update MFCs akan dilakukan sebanyak *infrequent* kandidat didalam loop tersebut), setelah itu *support* dari MFCs akan dihitung untuk mencari *frequent* MFCs untuk dimasukkan ke MFS (*update* MFS). Tetapi, jika

setelah penghitungan *support* kandidat tidak menemukan kandidat yang *infrequent* maka tidak perlu melakukan *update* MFCs, tetapi hanya MFS saja yang di *update*. Setelah ditemukan MFS maka kandidat yang merupakan *subset* dari MFS akan di *pruning*, kemudian dilakukan pengecekan kembali, jika masih ada item yang *frequent* maka akan di *generate* kandidat baru. Proses tersebut akan berulang sampai tidak ada lagi kandidat item yang *frequent*. Setelah tidak ditemukan lagi kandidat item yang *frequent* berarti kandidat telah habis. Selanjutnya akan dilakukan pembangkitan *subset* dari MFS untuk dihitung *support*, *confidence*, dan *conviction*nya untuk mendapatkan *rule* dari data-data tersebut.

3.1.3 Menghitung Akurasi

Pada proses ini dihitung performansi yaitu akurasi untuk *rule* yang dihasilkan oleh sistem. Mengukur akurasi dari metode yang digunakan berdasarkan nilai *confidence* dan *conviction*. Nilai *confidence* adalah nilai akurasi atau biasa disebut nilai kepastian, *rule* tersebut akurat atau tidaknya tergantung dari besar atau kecilnya *confidence*. Jika *confidence* kecil, maka dapat dipastikan *rule* tersebut kurang akurat. Tetapi jika nilai *confidence*nya besar maka dapat dipastikan *rules* tersebut akurat. Nilai *confidence* menentukan kuatnya hubungan antar *item* dalam *pincer search* maupun apriori.

Jika nilai *confidence* sudah ditentukan, maka selanjutnya menentukan nilai *conviction*. Nilai *conviction* dapat digunakan untuk mengetahui tingkat akurasi dan kekuatan *rule* yang terbentuk. Nilai *range* pada *conviction* ini berada pada nilai 0,5, ..., 1, ..., dengan ketentuan *conviction* dianggap memiliki nilai tak terhingga (*infinite*) apabila nilai dari *confidence* (A→B) sama dengan 1. *Conviction* menghasilkan nilai *rule* yang semakin jauh dari 1 bahkan sampai tak hingga (∞), maka akan dianggap semakin akurat.

4. Pengujian dan Analisis

4.1 Data Yang Digunakan

Data yang digunakan dalam tugas akhir ini adalah data transaksi penjualan di Yomart selama 1 (satu) tahun dan tidak ada *missing value*. Jumlah data yang digunakan sebanyak 10.676.

No	Transaksi ID	Jenis Item
1	546742	DIARUM LA LIGHTFLTR,BU,MENU ONON WARNA,INDOONMILK KID CHOU15,CRACKET LOT SOLID
2	546743	234 KRITEX 11 BTG,13MT LIGHT ELECTRIC
3	546745	ROVCO RASA ATAMINGR,RENDO NEW A,NODA,56G,DANGDUT CD LEMON(10)
4	546746	ENERGEN V,AN,30G80CH
5	546747	TELLER,ATM,NGR,ORH
6	546748	MTRABINTEL,ONPLUS 60
7	546749	SAMPORNERA,AMILD 16,DIARUM,COOKLAT 12
8	546750	NYAMC,CHO 30,SOBA,MIE ATAM,BKR,28,BIRDY,KOPI,SUSU 180,INDOONMILK KID CHOU15,INDOONMILK KIDSTRW15,...
9	546751	VICAL,HIJAU,500ML,FRUTELA,FRUTEL135,CHLMIL,DHA,200GBOX
10	546753	ULTRA,CHOCOLATE,125ML,VOUC 1100V,TORIG140
.	.	.
.	.	.
.	.	.
10677	555280	MINUTE,MAID,ORG 550,1000,C 1000,ORG WTR
10676	555281	CLUB,240 ML,ALE,ALE,ORG 200ML,ALE,ALE,STRW 300
10677	555282	WIPOL,CLSPNEBT1400
10676	555283	MITU,TRAVEL,PNK,105,OREO,29,4GR
10677	555284	UPPET,BUBUK,60G,ABC,PLUS 5125 GR
10676	555285	SHEARON,TOAST,KOTAK
10677	555286	SISTER,FANPACK,ISI,SAMPORNERA,AMILD 16,SUNSLK,A,DNDRE,90ML,CALADINE,BDY TLC 60,TANCHO,HAIR,DYE
10676	555287	AQUA,600ML,BEN,BTL,VAN 180,TANGO,COOK&CREAM 20
10677	555288	SOSRO,JTEA,CHEFEL,500,SW,TB ASLI 505
10676	555289	MAOLTO,PWGBLUPECHOU,SUNLIGHT,PPID,NEROR,PORSTEX,WT,500 ML,SCUBER,AT,SS,MUSKUSU,FORM,DVERY COMFORT,...

Tabel 0.3 Tabel Data Transaksi Penjualan

4.2 Pengujian Sistem

a. Tujuan Pengujian

- a. Mengetahui perbandingan akurasi dari setiap *rules*.
- b. Mengetahui Perbandingan waktu dan *rule* dari algoritma *pincer-search* dan algoritma apriori.
- c. Mengetahui perbandingan waktu pada setiap skenario.
- d. Mengetahui *rule* mana yang paling efisien.
- e. Mengetahui asosiasi dari setiap item.

b. Skenario Pengujian

Tabel 0.4 Pembagian Skenario Pengujian.

Nama Skenario	Pembagian Partisi
Skenario 1	Nilai maxRuleLen = 2 dengan minimal confidence 20%,30%, dan 40%
Skenario 2	Nilai maxRuleLen = 3 dengan minimal confidence 20%,30%, dan 40%
Skenario 3	Nilai maxRuleLen = 4 dengan minimal confidence 20%,30%, dan 40%

c. Penentuan Rule

Pada tahap ini akan di seleksi *rules* mana saja yang memiliki tingkat keefisienan paling tinggi, berdasarkan:

1. Memiliki nilai *conviction* tinggi.
Nilai range pada *conviction* ini, berada pada nilai 0.5,...,1,...∞. Apabila *conviction* menghasilkan nilai *rule* yang menjauh dari 1 maka akan dianggap akurat atau *rule* tersebut memiliki tingkat keakuratan yang baik.
2. Menghasilkan *rules* yang tidak terlalu banyak tetapi memiliki keakuratan yang tinggi.
3. Menghasilkan waktu *running* program yang paling cepat.

4. Memiliki nilai *confidence* yang tinggi.

Dalam tugas akhir ini saya akan menganalisis *rules* berdasarkan:

1. Jumlah *rules* yang terbentuk dari nilai minimal maxRuleLen dan nilai minimal *confidence* yang telah ditentukan.

Tabel 0.5 Jumlah Rules Berdasarkan Nilai minimal

maxRuleLen dan Confidence		Confidence		
		20%	30%	40%
maxRuleLen	3	33	27	13
	4	33	27	13
	5	33	27	13

Hasilnya, jumlah *rules* yang paling sedikit ada di nilai minimal maxRuleLen = 2,3 dan 4 dan nilai minimum *confidence* = 40%.

2. Nilai *conviction* berdasarkan nilai minimum *confidence* yang telah ditentukan.

Tabel 0.6 Nilai Conviction Berdasarkan Nilai Minimum Confidence

Confidence	Conviction
20%	1.4562
30%	1.5211
40%	1.6406

Hasilnya, nilai minimum *confidence* 40% menghasilkan rata-rata nilai *conviction* terbesar.

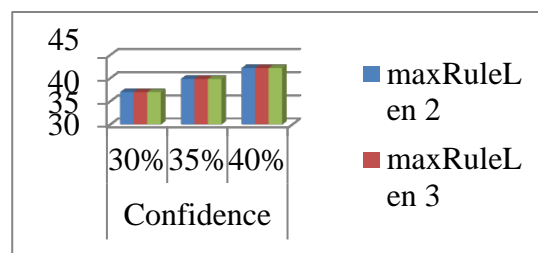
3. Rata-rata Nilai Confidence

Tabel 0.7 Nilai Rata-rata Nilai Confidence

		Confidence		
		20%	30%	40%
maxRuleLen	2	37.0599	39.854	45.4219
	3	37.0599	39.854	45.4219
	4	37.0599	39.854	45.4219

Gambar 0.1 Rata-rata Nilai Confidence

Hasilnya, nilai maxRuleLen= 2,3,4 dan nilai minimum *confidence* = 40% memiliki rata-rata nilai *confidence* terbesar. Jumlah Rule dari maxRuleLen yang dicoba di masing-masing minimum *confidence* tidak berubah, karena *rule* yang dihasilkan dari pengolahan data tersebut tidak ada yang berjumlah lebih ataupun kurang dari 3 itemset, baik dalam minimum *confidence* 30%, 40%, maupun 50%. Dari hasil analisis dan percobaan pada penelitian ini dapat disimpulkan bahwa nilai minimum *confidence* 40% itu memiliki *rule* yang paling efisien, oleh karena itu



rule yang saya ambil berdasarkan minimum *confidence* = 40%, *rules* tersebut merupakan assosiasi antar *item* yang didapat dari data. *Rules* yang dihasilkan adalah sebagai berikut:

Tabel 0.8 Rules Yang Dihasilkan

No.	Rules	Support	Confidence	Conviction
1	FORTUNE POUCH 2LT,IND GRG SPC SAUS ==>IND AYAM BAWANG	0.16%	44.73%	1.7465
2	FORTUNE POUCH 2LT,IND GRG SPC SAUS,YG GULA LOKAL 1KG ==>IND AYAM BAWANG	0.16%	50.00%	1.9174
3	IND AYAM BAWANG,TELUR AYM NGR CRH ==>IND GRG SPC SAUS	0.64%	46.90%	1.8035
4	IND AYAM BAWANG,YG GULA LOKAL 1KG ==>IND GRG SPC SAUS	0.16%	49%	1.8641
5	IND GRG SPC SAUS ==>TELUR AYM NGR CRH	1.76%	43%	1.4943
6	FORTUNE POUCH 2LT,IND AYAM BAWANG ==>TELUR AYM NGR CRH	0.17%	52.94%	1.8217
7	FORTUNE POUCH 2LT,IND GRG SPC SAUS ==>TELUR AYM NGR CRH	0.17%	47.37%	1.6288
8	IND AYAM BAWANG,IND GRG SPC SAUS ==>TELUR AYM NGR CRH	0.64%	48.57%	1.6669
9	IND GRG SPC SAUS,YG GULA LOKAL 1KG ==>TELUR AYM NGR CRH	0.18%	40.43%	1.439
10	SW TB ASLI 25S ==>TELUR AYM NGR CRH	0.33%	40.23%	1.4342
11	SUPERMIE AYAM BWG ==>TELUR AYM NGR CRH	0.48%	48.11%	1.6522
12	SEDAAPMIE MIE GRG90 ==>TELUR AYM NGR CRH	0.67%	40.00%	1.4287
13	EKONOMI CRM LMN 300 ==>TELUR AYM NGR CRH	0.22%	40.00%	1.4287

Dari *rules* diatas kita bisa merekomendasikan *item-item* pada *rules* tersebut untuk diatur penempatannya di dalam rak, agar produsen dapat memaksimalkan keuntungan yang dia dapat, karena

item-item yang sering dibeli oleh pembeli sudah ditempatkan ditempat yang tepat. Rule tersebut diperoleh dari hasil pengujian 10.676 data transaksi, yang di bagi menjadi 5 partisi yaitu 1-2000, 2001-4000, 4001-6000, 6001-8000, dan 8001-10676. Proses tersebut dibagi menjadi 5 partisi dikarenakan perangkat yang tidak mendukung untuk melakukan pengujian dengan data sebanyak 10676 sekaligus. Proses pengujian dengan partisi tidak mempengaruhi hasil rule yang dihasilkan. Hal tersebut telah dibuktikan penulis dengan melakukan pengujian dengan 15 data dummy, membandingkan hasil rule yang didapat dari hasil pengujian 15 data sekaligus dan pengujian secara partisi yaitu per-5 sampel data. Hasil yang diperoleh dari kedua pengujian tersebut sama.

4.3 Perbandingan *Pincer-Search* dengan Apriori

a. Perbandingan Estimasi Waktu *Running*

Tabel 0.9 Perbandingan Estimasi Waktu *Running*

<i>Pincer-Search</i>	Apriori
22885.446 detik	23627.492 detik

Hasilnya *Pincer-Search* terbukti lebih cepat dalam segi estimasi waktu *running*. Untuk menganalisis data yang cukup besar sebaiknya dianalisis menggunakan *pincer-search* karena akan lebih efisien dibandingkan dengan apriori dalam segi estimasi waktu *running*nya.

b. Perbandingan *Rule* Yang Dihasilkan

Tabel 0.10 Perbandingan *Rule* Yang Dihasilkan

		<i>Pincer-Search</i>	Apriori
<i>Confidence</i>	20%	33	24
	30%	27	18
	40%	13	7

Dalam kasus ini, *rule* yang dihasilkan apriori, semuanya terkandung di dalam *rule* yang dihasilkan *pincer-search*, dengan *confidence*, *conviction*, dan *support* yang sama. Tetapi *rule-rule* yang ada di *pincer-search* tidak semuanya tersaring oleh apriori, walaupun *rule* tersebut memiliki *confidence* dan *conviction* yang baik. Hal tersebut disebabkan oleh teknik pencarian *rule* yang berbeda dari kedua algoritma dan juga disebabkan oleh teknik *running* data yang bertahap. Sebagai contoh, kita menetapkan minimum *support* = 10. misalkan pada *running* apriori data 1-2000 terdapat *rule* A→B, tetapi *support*nya hanya 5, otomatis *rule* tersebut tidak tersaring oleh apriori, selanjutnya pada saat melanjutkan di *running* data 2001-4000 *rule* A→B muncul kembali tetapi *support*nya kembali tidak memenuhi minimum *support* yang ditentukan yaitu hanya 6 jadi otomatis *rule* tersebut tidak tersaring kembali oleh algoritma apriori. Tetapi jika langsung melakukan *running* dengan data 1-4000 *rule* A→B

pasti tersaring karena supportnya lebih dari minimum support yang ditentukan. Beda dengan *pincer-search*, algoritma ini mencari MFS, setelah MFS tersebut didapatkan, *rule-rule* akan di *generate* dari MFS tersebut. Bisa saja *rule* $A \rightarrow B$ ini tersaring dikarenakan ada MFS yang mengandung *subset rule* tersebut.

Jadi dalam kasus ini *rule pincer-search* menghasilkan *rule* yang lebih baik dibandingkan apriori berdasarkan rata-rata *confidence* dan *conviction*nya seperti yang terlihat pada tabel berikut:

Tabel 0.11 Perbandingan Rata-rata *Confidence* dan *Conviction*

	<i>Pincer-search</i>	Apriori
<i>Confidence</i>	45.42	43.77
<i>Conviction</i>	1.6406	1.5586

Tabel 4.10 *Rule* yang Dihasilkan Apriori

No.	<i>Rules</i>	<i>Support</i>	<i>Confidence</i>	<i>Conviction</i>
1	IND AYAM BAWANG, TELUR AYM NGR CRH ==>IND GRG SPC SAUS	0.64%	46.90%	1.8035
2	IND GRG SPC SAUS ==>TELUR AYM NGR CRH	1.76%	43%	1.4943
3	SEDAAPMIE MIE GRG90 ==>TELUR AYM NGR CRH	0.67%	40.00%	1.4287
4	SUPERMIE AYAM BWG ==>TELUR AYM NGR CRH	0.48%	48.11%	1.6522
5	IND AYAM BAWANG, IND GRG SPC SAUS ==>TELUR AYM NGR CRH	0.64%	48.57%	1.6669
6	EKONOMI CRM LMN 300 ==>TELUR AYM NGR CRH	0.22%	40.00%	1.4287
7	SW TB ASLI 25S ==>TELUR AYM NGR CRH	0.33%	40.23%	1.4342

5. Kesimpulan dan Saran

a. Kesimpulan

Kesimpulan dari tugas akhir ini adalah:

1. Metode *Pincer-search* bisa di implementasikan dalam data transaksi penjualan Yomart Margacinta Bandung.
2. Dari hasil analisis dan percobaan pada penelitian ini dapat disimpulkan bahwa *rule* yang diperoleh dengan minimum *confidence* 40% memiliki jumlah *rules* yang efisien (diperoleh 13 *rules*), memiliki nilai rata-rata *conviction* terbesar dan memiliki rata-rata *confidence* terbesar, hal

tersebut membuktikan bahwa *rule* tersebut memiliki tingkat keakuratan yang baik.

3. *Pincer-search* terbukti lebih cepat dalam estimasi waktu *running* dibandingkan dengan Apriori.
 4. Dalam kasus ini *pincer-search* menghasilkan *rule-rule* dengan rata-rata akurasi yang lebih baik dibandingkan dengan *rule-rule* yang dihasilkan apriori.
- b. Saran
1. Aplikasi ini masih bisa dikembangkan lebih baik lagi menggunakan metode lainnya seperti FP(*Frequent Pattern*) *Growth*, dan metode *Assosiation* lainnya.
 2. Aplikasi ini belum bisa menjalankan program dengan data keseluruhan, masih dibagi-bagi menjadi beberapa bagian data.
 3. Tampilan dari aplikasi ini masih sederhana, sehingga dapat dibuat lebih menarik lagi.

Daftar Pustaka:

- [1] Andrew McAfee, E. B. (2012). Big Data: The Management Revolution. Harvard Business Review.
- [2] Dunham, M. H. (n.d.). *Data Mining Introductory and Advanced Topics*. Pearson Education.
- [3] Fayyad, U. S. (1996). "From Data Mining to Knowledge Discovery in Databases". *AI Magazine*, 37-54.
- [4] J, H., & M, K. (2006). *Data Mining: Concepts and Techniques*. Sanfrancisco: Morgan Kaufmann Publishers.
- [5] Kamber, J. H. (2000). "Data Mining: Concepts and Techniques : Chapter 6. Mining Association Rules in Large Databases". Simon Fraser University.
- [6] Kusriani. (2007). "Penerapan Algoritma Apriori Pada Data Mining Untuk Mengelompokkan Barang Berdasarkan Kecenderungan Kemunculan Bersama Dalam Satu Transaksi". *E-learning Center STMIK AMIKOM Yogyakarta*.
- [7] Lin, D.-I. (1998). *Fast Algorithm for Discovering the Maximum Frequent Set*. New York University.
- [8] Lin, D.-I., & M.Kedem, Z. (1997). *Pincer-Search: A New Algorithm for Discovering the Maximum Frequent Set*. New York University.
- [9] Lin, D.-I., & M.Kedem, Z. (1999). *Pincer-Search: An Efficient Algorithm for Discovering the Maximum Frequent Set*. Telcordia Technologies, Inc and New York University.
- [10] Lin, D.-I., & M.Kedem, Z. (2002). *Pincer-Search: An Efficient Algorithm for Discovering the Maximum Frequent Set*. *IEEE Transaction of Knowledge and Data Engineering*, Vol.14 No.3, 563-566.

- [11] Luthfi, K. d. (2009). *Algoritma Data Mining*. Yogyakarta: Andi.
- [12] radhiaty, n. H. (2009). Personifikasi Web E-Commerce Menggunakan Basket Algoritma dari Data Mining.
- [13] Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- [14] Sholom M. Weiss, N. I. (n.d.). *Predictive Data Mining a practical guide*. San Fransisco: Morgan Kaufmann Publishers, inc.