

ANALISIS DAN IMPLEMENTASI *FRAMEWORK* CRISP-DM UNTUK MENGETAHUI PERILAKU DATA TRANSAKSI PELANGGAN

Analysis and Implementation CRISP-DM Framework for Customer Behaviour of Transaction Data (Case Study : PT X)

Muhammad Zain Imtiyaz¹, Muhammad Nasrun S.Si, M.T.², Umar Ali Ahmad S.T, M.T.³

^{1,2,3} Prodi S1 Sistem Komputer, Fakultas Teknik, Universitas Telkom

¹muhammadzainimtiyaz@gmail.com, ²nasrun@telkomuniversity.ac.id, ³umar@telkomuniversity.ac.id

ABSTRAK

Bisnis ritel mengalami pertumbuhan yang begitu pesat di Indonesia. Bisnis ini menunjukkan potensi yang cukup tinggi karena menjangkau masyarakat kalangan menengah. Kekuatan utama dalam bisnis ritel ini adalah variasi produk yang ditawarkan. Masalah akan muncul apabila tidak ada strategi khusus dalam pengelolaan variasi produk tersebut. Hal ini berkaitan erat dengan karakteristik atau pola pembelian oleh pelanggan. Pengelola bisnis ritel harus cerdas dalam pengambilan strategi maupun keputusan terutama pemilihan barang yang akan dijual kepada pelanggan.

Tugas akhir ini membahas mengenai perancangan dan implementasi aplikasi *data mining* pada bisnis ritel (studi kasus : PT X) dengan menggunakan data transaksi penjualan dengan tujuan menganalisis perilaku data yang merepresentasikan pola pembelian pelanggan. Pola pembelian pelanggan diperoleh dari kombinasi antar item barang yang dibeli atau biasa disebut Asosiasi. Aplikasi ini menggunakan *framework* CRISP-DM dengan algoritma Apriori sebagai metodenya.

Hasil analisis menunjukkan bahwa kemunculan kombinasi item, *support*, dan *confidence* setiap bulan menunjukkan hasil yang berbeda. Hasil analisis ini menjadi rekomendasi strategi perusahaan dalam memasarkan produk sesuai minat pelanggan.

Kata kunci : bisnis ritel, *data mining*, *framework* CRISP-DM, algoritma Apriori, asosiasi

ABSTRACT

Retail business has grown so rapidly in Indonesia. This business showed a fairly high potential for reaching the middle class. The main strength of the retail business is the variety of products offered. This business would be problematic if there is no specific strategy in the management of the product variations. It is closely related to the characteristics or patterns of purchases by customers. Retail business management must take a smart way in decision-making strategy and especially the selection of goods to be sold to customers.

This final project is to discuss the design and implementation of data mining applications in the retail business (case study: PT X) by using the sales transaction data with the aim of analyzing the data that represent the behavior of the customer buying patterns. Customer buying patterns can be seen from the combination of the items purchased goods or commonly called the Association. This application uses the CRISP-DM framework with Apriori algorithm as the method.

The analysis showed that the appearance of a combination of items, support, and confidence every month showed different results. The results of this analysis will be on the company's strategy in marketing the product according customer interest.

Keywords: retail business, data mining, CRISP-DM framework, Apriori algorithm, association

1. Pendahuluan

Pertumbuhan bisnis ritel, terutama bisnis ritel modern, saat ini semakin berkembang dengan pesat di Indonesia. Diperoleh dari *website* Data Consult (*Business Research Studies Report*), dalam periode lima tahun terakhir (2007-2011) jumlah gerai ritel modern di Indonesia mengalami pertumbuhan hingga 17,57% per tahun. Padahal tahun 2007, jumlah gerai ritel modern hanya 10.365 buah dan pada tahun 2011 jumlah gerai di Indonesia sudah mencapai 18.152 buah yang tersebar di kota-kota besar. Bisnis ritel cukup diperhitungkan dalam memainkan peranan penting di perekonomian sebuah negara. Perekonomian negara ini tertolong dengan adanya bisnis ritel ketika terjadi krisis moneter pada akhir tahun 1997 di Indonesia.

Bisnis ritel memerlukan strategi bisnis yang dapat menyesuaikan kebutuhan pelanggan. Strategi ini diperoleh dari data-data penguat yang bisa dijadikan informasi. Apabila tidak memanfaatkan data yang

diperlukan, strategi bisnis yang diambil bisa tidak sesuai dengan kebutuhan pelanggan sehingga mengakibatkan tidak diminatinya bisnis ritel yang telah dibuka. Untuk itu, pengolahan data yang baik sangat dibutuhkan untuk mendapatkan strategi yang baik pula sehingga pada akhirnya dapat memuaskan pelanggan. Pola pembelian pelanggan dapat dianalisis dengan memanfaatkan data transaksi setiap bulan dan dijadikan bahan pertimbangan dalam pengambilan strategi penjualan.

2. Tinjauan Pustaka

2.1 Data Mining

Data Mining merupakan suatu proses menemukan hubungan yang berarti kecenderungan dengan memeriksa sekumpulan besar data yang tersimpan dalam penyimpanan dengan teknik pengenalan pola seperti teknik statistik dan matematika [1]. *Data mining* dalam konteks implementasi bertujuan untuk meningkatkan bagian pemasaran, penjualan, juga pemahaman yang bisa lebih mendalam mengenai pelanggan yang menjadi fokus pasar [2]. *Data mining* sebagai ilmu yang menemukan informasi tersembunyi untuk memberikan solusi permasalahan dalam bisnis [3]. Penerapan *data mining* antara lain analisa pasar dan manajemen, analisa perusahaan dan manajemen resiko, telekomunikasi, keuangan, asuransi, olahraga, dan astronomi. *Data mining* berisikan berbagai proses untuk dapat menemukan informasi yang tersembunyi [4].

2.2 CRISP-DM

CRISP-DM atau *Cross-Industry Standard Process for Data Mining* merupakan standar yang dikembangkan pada tahun 1996 yang ditujukan untuk proses analisis suatu industry sebagai strategi pemecahan masalah dari bisnis [5]. CRISP-DM tidak menentukan standar atau karakteristik tertentu karena setiap data yang akan dianalisis akan diproses kembali pada fase-fase di dalamnya [5].

2.3 Algoritma Apriori

Algoritma apriori termasuk salah satu algoritma dalam asosiasi pada *data mining*. Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut *affinity analysis* atau *market basket analysis* [8]. *Market basket analysis* merupakan analisis untuk mengetahui isi keranjang belanja di sebuah toko. Analisis asosiasi atau *association rule mining* adalah teknik untuk menemukan aturan kombinasi *item*. Asosiasi dapat diketahui dengan dua tolak ukur atau parameter, yaitu : *support* dan *confidence*. *Support* (nilai penunjang) adalah persentase kombinasi *item* tersebut dalam database, sedangkan *confidence* (nilai kepastian) adalah kuatnya hubungan antar-*item* dalam aturan asosiasi [6].

1. Analisis Pola Frekuensi Tinggi dengan Algoritma Apriori

Nilai *support* sebuah *item* diperoleh dengan menggunakan rumus berikut:

$$\frac{\text{Jumlah item}}{\text{Jumlah data}} \quad (1)[7]$$

Sementara, nilai *support* dari 2 *item* diperoleh dengan menggunakan rumus :

$$(2)[7]$$

$$\frac{\text{Jumlah item}}{\text{Jumlah data}} \quad (3)[7]$$

Frequent itemset menunjukkan *itemset* yang memiliki frekuensi kemunculan lebih dari nilai minimum yang ditentukan (ϕ). [7]

2. Pembentukan Aturan Asosiasi (*Association Rules*)

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiatif $A \rightarrow B$. Nilai *confidence* dari aturan $A \rightarrow B$ diperoleh dengan rumus berikut: [8]

$$\frac{\text{Jumlah item}}{\text{Jumlah data}} \quad (4)[8]$$

Untuk menentukan aturan asosiasi yang akan dipilih maka harus diurutkan berdasarkan *Support* \times *Confidence*. Aturan diambil sebanyak n aturan yang memiliki hasil terbesar.[8]

3. Perancangan dan Implementasi

3.1 Fase pemahaman bisnis (*Business Understanding Phase*)

PT X adalah sebuah mini market yang dibangun di kampus Telkom University. Tujuan PT X ini dibangun untuk dapat memenuhi kebutuhan sehari-hari mahasiswa, baik di kampus maupun di Asrama. Data transaksi

memberikan informasi barang-barang yang terjual. Selain memiliki kegunaan untuk memberikan informasi barang-barang yang terjual, data transaksi digunakan untuk pembuatan rekap bulanan yang berkaitan dengan hasil penjualan perusahaan. Manfaat data transaksi ini dapat memberikan pertimbangan terhadap strategi penjualan pada bulan berikutnya. Dalam *data mining*, terdapat analisis asosiasi yang berguna untuk menganalisis data transaksi. Asosiasi berarti mencari kombinasi antar barang dalam transaksi yang telah terjadi.

3.2 Fase pemahaman data (*Data Understanding Phase*)

Terdapat total 4692 data, sesuai dengan transaksi yang terjadi pada Januari 2014. Data transaksi yang diambil merupakan **Data Penjualan PT X**. Rincian besar data transaksi penjualan per bulan yang terjadi dari PT X bisa dilihat di Tabel dibawah ini.

Tabel 3.1 Rincian Besar Transaksi Per Bulan

Bulan	Jumlah	NoTransaksi Awal	NoTransaksi Akhir
Januari	4692	1	4692
Februari	12886	4693	17578
TOTAL	17578	~	~

3.3 Fase pengolahan data (*Data Preparation Phase*)

Fase persiapan data terdiri dari langkah-langkah berikut ini :

1) *Data Cleaning*

Data cleaning bertujuan untuk membersihkan data-data yang tidak valid. Dalam *database* yang telah didapat, terdapat atribut selain nama barang. Atribut yang dapat dipangkas yang tidak digunakan dalam analisis antara lain

Tabel 3.2 Atribut Data Penjualan

No	Atribut	Value
1	Kode	Numerik
2	Harga	Numerik
3	Jml	Numerik
4	Subtotal	Numerik

2) *Construct Data*

Construct data bertujuan untuk menghasilkan atribut baru sesuai kebutuhan *data mining*. Atribut baru yang bisa dibuat antara lain NoTransaksi dari NoJual dan Tanggal dari tanggal yang tercantum di tiap transaksi.

3) *Data Transformation*

Data Transformation bertujuan untuk mengubah format data sesuai kebutuhan. Format data yang dipakai semula menggunakan .xls. Format data diubah menjadi .csv dengan tujuan mempermudah input ke dalam *database*.

3.4 Fase pemodelan (*Modelling Phase*)

Tujuan utamanya adalah untuk membentuk kombinasi item yang dapat dijadikan sebagai alat rekomendasi strategi penjualan. Pseudocode algoritma Apriori terdapat pada gambar di bawah ini :

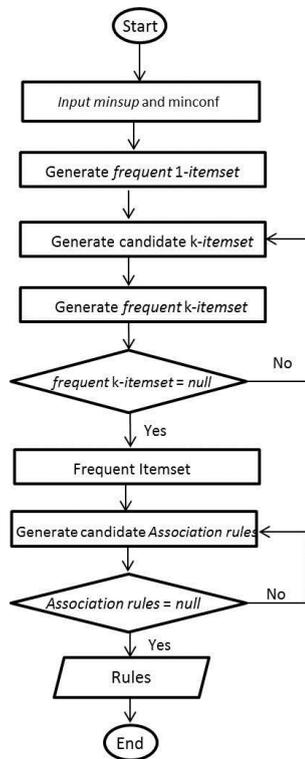
```

1: k = 1
2: Fk = {i | i ∈ I ∧ σ({i}) ≥ N × minsup}. {Find all frequent itemset}
3: repeat
4:   k = k + 1
5:   Ck = apriori-gen(Fk-1). {Generate candidate itemset}
6:   for each transition t ∈ T do
7:     Ct = subset(Ck, t). {Identify all candidates that belong to t}
8:     for each candidate itemset c ∈ Ct do
9:       σ(c) = σ(c) + 1. {Increment support count}
10:    end for
11:  end for
12:  Fk = {c | c ∈ Ck ∧ σ(c) ≥ N × minsup}. {Extract the frequent itemset}
13: until Fk = ∅
14: Result = ∪ Fk

```

Gambar 3.1 Pseudocode Algoritma Apriori[4]

Dari pseudocode tersebut, didapatkan gambaran umum sistem yang dibentuk sebagai berikut :



Gambar 3.2 Gambaran Umum Sistem yang Diusulkan[4]

Sistem dimulai dari input *support* dan *confidence*. Input digunakan untuk pemangkasan dalam proses mendapatkan *frequent itemset*. Setelah input dilakukan, proses selanjutnya adalah pembuatan *frequent 1-itemset*. *Frequent 1-itemset* menentukan pembangkitan *frequent k-itemset*. Pembangkitan *frequent k-itemset* berakhir bila sudah tidak ada lagi *itemset* yang dihasilkan. Bila *Frequent Itemset* sudah didapat proses selanjutnya adalah pembangkitan aturan asosiasi atau *Association Rules*. Pembangkitan kandidat *Association Rules* akan berakhir bila tidak ada lagi kandidat yang dihasilkan.

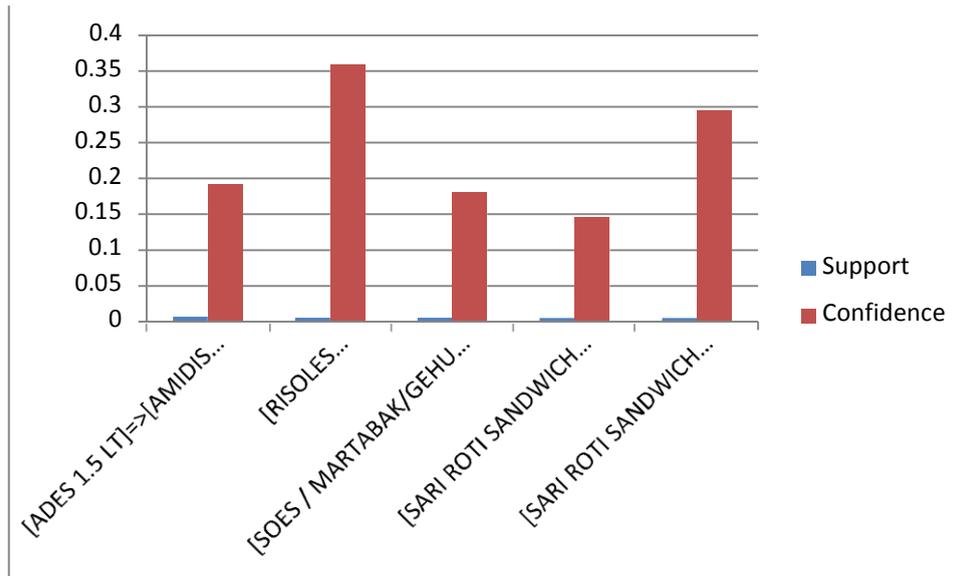
3.5 Fase Evaluasi (Modelling Phase)

Dari pemodelan yang telah dilakukan pada fase sebelumnya, dilakukan implementasi sistem pada Bahasa Pemrograman Java. Hasil berdasarkan *rules* yang terbentuk pada tiap bulan relative berbeda. Berikut hasil yang *rules* yang ada pada Bulan Januari dan Bulan Februari.

Tabel dan grafik di bawah ini dihasilkan dari *minsup* 0.004 dan *minconf* 0.1 :

Tabel 3.3 Tabel Hasil *rules* pada Bulan Januari

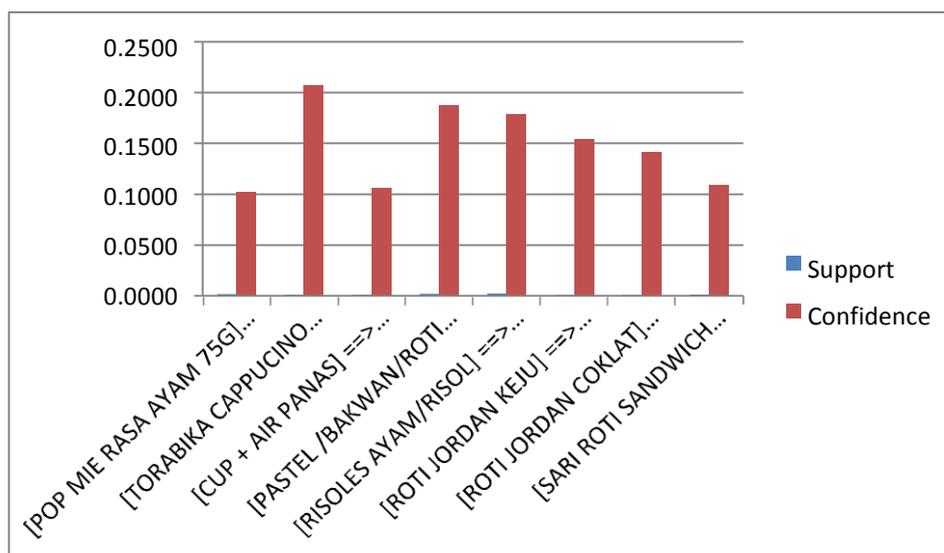
Rules	Support	Confidence
[Ades 1.5 Lt]=>[Amidis Galon 19 Ltr]	0.00683	0.19135802
[Risoles Ayam/ Risol]=>[Soes / Martabak/ Gehu Pedas /Cente /Puding Manis/ Pisang]	0.00529	0.35820896
[Soes / Martabak/ Gehu Pedas/ Cente/ Puding Manis/ Pisang]=>[Risoles Ayam/ Risol]	0.00529	0.18045113
[Sari Roti Sandwich Coklat]=>[Sari Roti Sandwich Keju]	0.00506	0.14465409
[Sari Roti Sandwich Keju]=>[Sari Roti Sandwich Coklat]	0.00506	0.29487179



Gambar 3.3 Grafik rules yang terbentuk Bulan Januari

Tabel 3.4 Tabel Hasil rules pada Bulan Februari

Rules	Support	Confidence
[Pop Mie Rasa Ayam 75g] ==> [Air Panas]	0.0016	0.1020
[Torabika Cappucino 25g] ==> [Cup + Air Panas]	0.0009	0.2069
[Cup + Air Panas] ==> [Torabika Cappucino 25g]	0.0009	0.1062
[Pastel /Bakwan/Roti Goreng/Dadar G] ==> [Soes / Martabak/Gehu Pedas/Cente/Puding Manis/Pisang]	0.0021	0.1875
[Risoles Ayam/Risol] ==> [Soes / Martabak/Gehu Pedas/Cente/Puding Manis/Pisang]	0.0023	0.1786
[Roti Jordan Keju] ==> [Roti Jordan Coklat]	0.0008	0.1538
[Roti Jordan Coklat] ==> [Roti Jordan Keju]	0.0008	0.1408
[Sari Roti Sandwich Kacang] ==> [Sari Roti Sandwich Coklat]	0.0012	0.1087

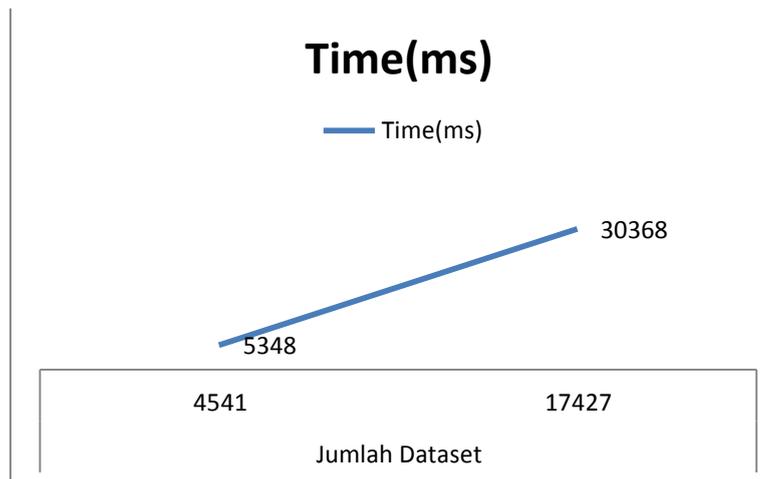


Gambar 3.4 Grafik rules yang terbentuk Bulan Februari

Selain hasil *rules*, waktu komputasi menjadi salah satu parameter yang dilihat pada penelitian ini. Berikut hasil waktu komputasi pada Bulan Januari dan Bulan Februari :

Tabel 3.5 Tabel Waktu Komputasi

	Jumlah Dataset	
	4541	17427
Time(ms)	5348	30368



Gambar 3.5 Waktu Komputasi

4. Kesimpulan

Berdasarkan hasil pengujian dan pembahasan dalam Tugas Akhir dapat disimpulkan bahwa : Berdasarkan hasil pengujian yang didapatkan, tugas akhir ini dapat disimpulkan sebagai berikut :

1. perancangan dengan menggunakan *framework* CRISP-DM dan algoritma Apriori berhasil diimplementasikan,
2. implementasi algoritma Apriori telah tervalidasi,
3. hasil *rules* yang dihasilkan menunjukkan kombinasi item yang berbeda setiap bulannya walaupun menggunakan nilai parameter yang sama,
4. pengaruh jumlah *database* transaksi dengan waktu komputasi menunjukkan bahwa semakin banyak data yang dianalisis, semakin lama waktu yang dibutuhkan.

5. Saran

Pada penelitian berikutnya diberikan beberapa saran yang akan ditujukan untuk memperbaiki sistem sehingga lebih andal :

1. Sistem dapat menggunakan algoritma lainnya. Misal : FP Growth, AprioriTID, dll.
2. Sistem dapat dicoba dengan spesifikasi komputer yang lebih baik sehingga komputasi bisa berjalan lebih cepat. Misal : dijalankan pada PC atau pada processor yang lebih tinggi spesifikasinya

Daftar Pustaka

- [1] Larose, Daneil T. (2005). *Discovering Knowledge in Data : An Introduction to Data Mining*. Canada : John Wiley & Sons, Inc
- [2] Berry, Michael J. A. and Linoff, Gordon S. (2004). *Data Mining Techniques : for Marketing, Sales, and Customer Relationship Management*. Canada : Wiley Publishing, Inc
- [3] Moertini, Veronica S. (2002). *Data Mining Sebagai Solusi Bisnis*. Integral, 7(1), 44-56
- [4] Han, Jiawei and Kamber, Micheline. 2006. *Data Mining: Concepts and Techniques*. America : Morgan Kaufmann Publisher

- [5] Akbar Prajitno, Agora. (2012). Prediksi Kinerja Penjualan Karya Musik Menggunakan Framework CRISP-DM (Studi Kasus: X Music Indonesia). Jurnal Sarjana Institut Teknologi Bandung bidang Teknik Elektro dan Informatika, 1(1), 83-90
- [6] Kusrini, dan Luthfi E,T. (2009). Definisi Algoritma A Priori.
- [7] Kusrini, dan Luthfi E,T. (2009). Algoritma Data Mining. Yogyakarta : Andi Offset.
- [8] Sari, Eka Novita. (2013). Analisa Algoritma Apriori Untuk Menentukan Merek Pakaian Yang Paling Diminati Pada Mode Fashion Group Medan. Pelita Informatika Budi Darma, 4(3), 35-39
- [9] Agrawal, Rakesh and Srikant, Ramakrishnan.1994. Fast Algorithms for Mining Association Rules. Research Report. IBM Almaden Research Center, San Jose, California, June 1994
- [10] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., She et al. (2000). Step-by-step data mining guide. USA
- [11] Junior Simanjuntak, Almon. (2012). Aplikasi Data Mining Untuk Pemodelan Pembelian Barang Dengan Menggunakan Algoritma Apriori.
- [12] Shearer, Colin. (2000). The CRISP-DM Model : The New Blueprint for Data Mining. Journal of Data Warehousing, 5(4), 13-22