

Market Basket Analysis Menggunakan Algoritma FP-Growth (Kasus Data Peminjaman Buku di Perpustakaan Universitas Islam Bandung)

Market Basket Analysis Using FP-Growth Algorithm

¹Febrian Teguh Raharjo, ²Teti Sofia Yanti, ²Abdul Kudus

^{1,2}Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung,
Jl. Tamansari No.1 Bandung 40116

email: ¹febrianteguh34@gmail.com, ²tetisofiyanti@gmail.com

Abstract. Data mining is the mining or discovery of new information by searching for a particular pattern or rule of a very large amount of data, its utilization is already widely applied in various fields. One of the data mining techniques is Market Basket Analysis which is used to find associations between different product sets that customers place in a basket. This thesis will discuss Market Basket Analysis using the Frequent Pattern Growth (FP-Growth) algorithm on the lending data of books at the Library of Islamic University of Bandung. The purpose is to find associations among books that are often borrowed simultaneously measured through the support value which is a measure that shows how much chance of lending transactions containing the itemsets borrowed simultaneously. The confidence value is a measure that shows how much the association between two books borrowed simultaneously from all transactions containing one of the books and found strong association rules of elevated ratio values based on frequent items Obtained. There are 9,804 lending transactions comprising of 250 book codes. The results indicate that the books (codes 657 and 658), (code books 346 and 297) and (1001 and 302 book codes) have support, confidence, and lift ratio values that meet the minimum requirements. Then the three itemsets can be used as a recommendation for the borrowers of books visiting the Library of Islamic University of Bandung and the book can be placed side by side.

Keywords: *Market Basket Analysis, Data Mining, FP-Growth, Support, Confidence.*

Abstrak. *Data mining* adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar, pemanfaatannya pun sudah banyak diterapkan dalam berbagai bidang. Salah satu teknik data pada *data mining* yaitu *Market Basket Analysis* yang digunakan untuk menemukan asosiasi diantara himpunan produk yang berbeda yang diletakkan pelanggan dalam keranjang. Skripsi ini akan membahas *Market Basket Analysis* menggunakan algoritma *Frequent Pattern Growth (FP-Growth)* pada data peminjaman buku di Perpustakaan Universitas Islam Bandung untuk menemukan asosiasi diantara buku-buku yang sering dipinjam secara bersamaan diukur melalui nilai *support* yaitu ukuran yang menunjukkan seberapa besar peluang banyaknya transaksi peminjaman buku yang memuat *itemsets* yang dipinjam secara bersamaan, nilai *confidence* yaitu ukuran yang menunjukkan seberapa besar asosiasi antar 2 buku yang dipinjam secara bersamaan dari seluruh transaksi yang memuat salah satu buku tersebut dan menemukan aturan asosiasi yang kuat dari nilai *lift ratio* berdasarkan *frequent itemsets* yang didapatkan. Terdapat 9,804 transaksi peminjaman buku yang terdiri dari 250 kode buku. Hasilnya menunjukkan bahwa (kode buku 657 dan 658), (kode buku 346 dan 297) serta (kode buku 1001 dan 302) memiliki nilai *support*, *confidence*, dan *lift ratio* yang memenuhi syarat minimal. Maka ketiga *itemsets* tersebut dapat dijadikan rekomendasi bagi peminjam buku yang berkunjung ke Perpustakaan Universitas Islam Bandung dan buku tersebut dapat diletakkan berdampingan.

Kata Kunci: *Market Basket Analysis, Data Mining, FP-Growth, Support, Confidence.*

A. Pendahuluan

Dengan kemajuan teknologi informasi saat ini, kebutuhan akan informasi yang akurat sangat dibutuhkan dalam kehidupan sehari-hari, sehingga informasi akan menjadi suatu elemen penting dalam perkembangan masyarakat saat ini dan masa mendatang. *Data mining* merupakan penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar. Pemanfaatan data yang ada didalam sistem informasi dapat berguna untuk menunjang kegiatan pengambilan keputusan, tidak cukup hanya mengandalkan data operasional saja, diperlukan suatu

analisis data untuk menggali potensi-potensi informasi yang ada.

Salah satu metode dalam *data mining* adalah metode *Market Basket Analysis*. Analisis ini merupakan salah satu metode dalam penambangan data (*data mining*) yang bertujuan untuk menemukan asosiasi antar buku yang sering dipinjam secara bersamaan dari data transaksi peminjaman buku. Tujuan analisis ini adalah untuk merancang strategi penempatan buku yang efektif dengan memanfaatkan data peminjaman buku di perpustakaan. Dalam metode *market basket analysis* ini menggunakan algoritma *Frequent Pattern Growth (FP-Growth)*.

Penggunaan algoritma *FP-Growth* pada data peminjaman buku di perpustakaan dapat memberikan informasi yang akurat terkait asosiasi antar buku berdasarkan nilai *support* yang menunjukkan seberapa besar peluang buku-buku tersebut dipinjam secara bersamaan dari keseluruhan data peminjaman, nilai *confidence* yang menunjukkan seberapa besar asosiasi antar buku-buku yang dipinjam secara bersamaan dari seluruh data peminjaman yang memuat salah satu buku tersebut, dan nilai *lift ratio* untuk melihat kuat tidaknya aturan asosiasi.

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana menemukan aturan asosiasi buku yang sering dipinjam secara bersamaan dari data peminjaman buku di Perpustakaan Universitas Islam Bandung menggunakan Algoritma *FP-Growth*?
2. Berapakah besar nilai *support* dan nilai *confidence* berdasarkan *frequent itemsets* dari data peminjaman buku di Perpustakaan Universitas Islam Bandung?
3. Apakah algoritma *FP-Growth* dapat menemukan aturan asosiasi yang kuat berdasarkan ukuran *lift ratio* dari data peminjaman buku di Perpustakaan Universitas Islam Bandung?

B. Landasan Teori

Data Mining

Dalam *data mining* dibagi kedalam beberapa bagian, diantaranya yaitu pengertian *data mining*, teknik *data mining* dan tahapan *data mining*. Berikut ini merupakan penjelasan dari bagian-bagian *data mining*.

1. Pengertian *Data Mining*
Secara sederhana *data mining* adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Davies, 2004). *Data mining* juga disebut sebagai serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui (Pramudiono, 2007). Pemanfaatan *data mining* ini juga sudah banyak diterapkan dalam banyak bidang salah satu contoh pemanfaatannya yaitu untuk menganalisa pasar sehingga sebuah pelaku bisnis bisa lebih mengetahui pola pembelian para konsumennya.
2. Teknik Data Mining
Beberapa teknik *data mining* antara lain (Larose, 2005) yaitu Klasifikasi (*Classification*), Regresi (*Regression*), Pengelompokan (*Clustering*), Kaidah Asosiasi (*Association Rules*), Analisis Pola Sekuensial (*Sequential Pattern Analysis*).
3. Tahapan Data Mining
Menurut (Han, 2006) tahap-tahap *data mining* tersebut adalah Pembersihan data

(*data cleaning*), Integrasi data (*data integration*), Seleksi data (*data selection*), Transformasi data (*data transformation*), Proses Mining, Evaluasi Pola (*pattern evaluation*), Presentasi Pengetahuan (*knowledge presentation*).

Market Basket Analysis

Fungsi *association rules* seringkali disebut dengan *Market Basket Analysis* yang digunakan untuk menemukan asosiasi diantara himpunan produk-produk yang berbeda yang diletakkan pelanggan dalam keranjang. Megaputer (2007), mengemukakan *market basket analysis*, merupakan salah satu tipe analisis data yang paling sering digunakan dalam dunia pemasaran.

Algoritma FP-Growth

Algoritma *FP-Growth* merupakan pengembangan dari algoritma Apriori. Pada algoritma Apriori diperlukan membangkitkan himpunan kandidat untuk mendapatkan *frequent itemsets*. Akan tetapi pada algoritma *FP-Growth*, membangkitkan himpunan kandidat tidak dilakukan karena algoritma *FP-Growth* menggunakan konsep pohon keputusan dalam pencarian *frequent itemsets*. Hal tersebut yang menyebabkan algoritma *FP-Growth* lebih cepat daripada algoritma Apriori.

Frequent Pattern Growth (FP-Growth) adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul dalam sebuah kumpulan data. Karakteristik algoritma *FP-Growth* adalah struktur data yang digunakan adalah dengan *FP-tree*. Dengan menggunakan *FP-tree*, algoritma *FP-Growth* dapat langsung memperoleh *frequent itemsets*. Pembuatan *FP-tree* dilakukan dengan melakukan *scanning data* dari tabel transaksi.

Algoritma *FP-Growth* dapat dibagi menjadi 3 tahapan utama (Han et al. 2006). Ketiga tahapan ini akan dilakukan berulang-ulang untuk setiap produk di *header table* yang diurutkan berdasarkan frekuensinya:

1. Tahap pembangkitan *conditional pattern base*
2. Tahap pembangkitan *conditional FP-tree*
3. Tahap pencarian *frequent itemsets*

Association Rules

Dalam *association rules* diperlukan ukuran yang ditentukan oleh pengguna untuk mengatur batasan sejauh mana dan sebanyak apa hasil output yang diinginkan. Ukuran tersebut adalah *support*, *confidence* dan *lift ratio*.

Support adalah ukuran yang menunjukkan seberapa besar peluang banyaknya transaksi yang memuat *itemsets* yang dibeli secara bersamaan dari keseluruhan transaksi. Ukuran ini akan menentukan apakah suatu *itemsets* dapat dicari nilai *confidence*. Berikut adalah rumus untuk menghitung *support* produk X dan produk Y.

$$\begin{aligned} \text{Support}(\{X, Y\}) &= P(\{X, Y\}) \\ &= \frac{\text{banyaknya transaksi yang memuat } X \text{ dan } Y}{\text{total banyaknya transaksi}} \end{aligned} \quad (1)$$

Confidence adalah ukuran yang menunjukkan seberapa besar asosiasi antar 2 produk yang dibeli secara bersamaan dari seluruh transaksi yang memuat salah satu produk tersebut. Berikut ini adalah rumus untuk menghitung *confidence* produk X dan produk Y.

$$\begin{aligned} \text{Confidence}(X \Rightarrow Y) &= P(Y|X) \\ &= \frac{P(X \text{ dan } Y)}{P(X)} \end{aligned}$$

$$= \frac{\text{banyaknya transaksi yang memuat } X \text{ dan } Y}{\text{banyaknya transaksi yang memuat } X} \quad (2)$$

Lift Ratio merupakan salah satu cara yang baik untuk melihat kuat tidaknya aturan asosiasi. Cara kerja metode ini adalah membagi *confidence* dengan *expected confidence*. *Confidence* dapat dihitung dengan rumus (2.2). Berikut ini adalah rumus untuk menghitung *expected confidence* produk *X* dan produk *Y*.

$$\begin{aligned} & \text{Expected Confidence } (X \Rightarrow Y) \\ &= \frac{\text{banyaknya transaksi yang memuat } Y}{\text{total banyaknya transaksi}} \end{aligned} \quad (3)$$

Lift Ratio dapat dihitung dengan cara membandingkan antara *confidence* untuk suatu aturan pada persamaan (2.2) dibagi dengan *expected confidence* pada persamaan (2.3). Berikut rumus dari *lift ratio*:

$$\text{Lift Ratio} = \frac{\text{Persamaan (2.2)}}{\text{Persamaan (2.3)}} \quad (4)$$

C. Hasil Penelitian dan Pembahasan

Bab ini berisikan tentang hasil dan pembahasan dari penerapan algoritma *FP-Growth* untuk data peminjaman buku di Perpustakaan Universitas Islam Bandung. Data peminjaman ini diubah kedalam bentuk matriks biner yaitu matriks yang berukuran $n \times p$ dan data yang ada dalam matriks tersebut berbentuk bilangan biner (0 dan 1). Kemudian algoritma *FP-Growth* dilakukan dengan bantuan *software* RapidMiner 5. Algoritma *FP-Growth* digunakan untuk menemukan buku-buku yang sering dipinjam secara bersamaan dengan menghitung nilai *support*, nilai *confidence* dan nilai *lift ratio* berdasarkan *frequent itemsets*.

Hasil Aplikasi Algoritma *FP-Growth*

Dalam bagian ini akan diaplikasikan algoritma *FP-Growth* pada data peminjaman buku di Perpustakaan Universitas Islam Bandung. Dari data peminjaman buku, kemudian disusun kode-kode buku yang dipinjam oleh peminjam perpustakaan. Setelah disusun, kemudian dihitung banyaknya transaksi peminjaman setiap kode buku.

Tabel 1. Frekuensi Kemunculan Tiap Kode Buku

No	Kode Buku	Nama Buku	Banyaknya Dipinjam
1	100	<i>Philosophy & psychology</i>	135
2	101	<i>Theory of philosophy</i>	16
3	102	<i>Miscellany</i>	17
4	107	<i>Education, research & related topics</i>	13
5	109	<i>Historical & collected persons treatment</i>	86
⋮	⋮	⋮	⋮
250	1070	<i>News media, journalism & publishing</i>	331

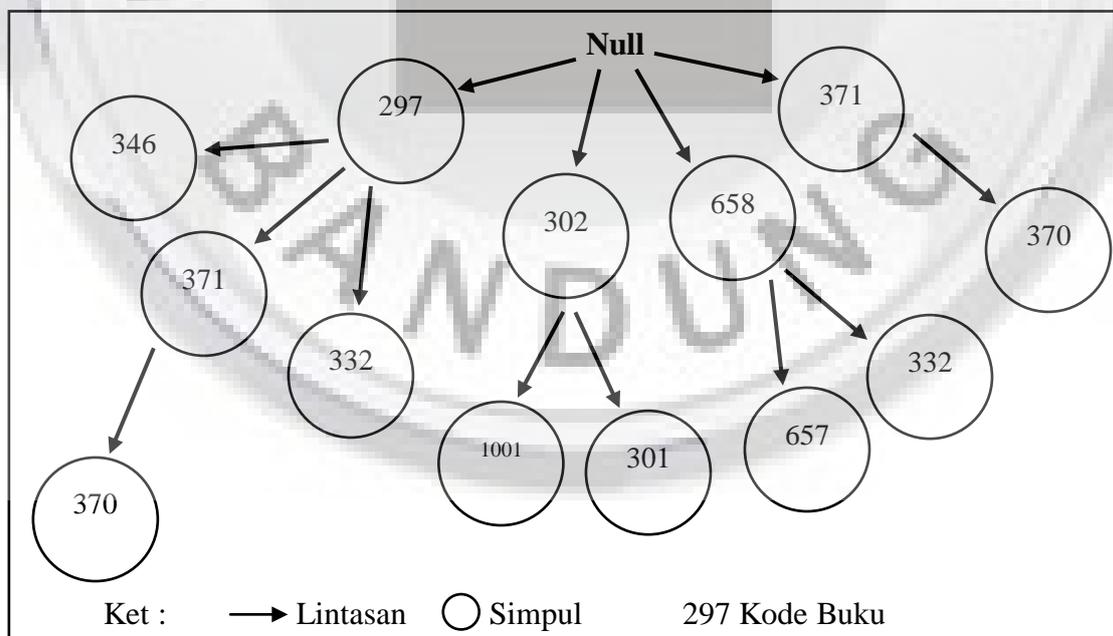
Dalam pembahasan ini akan digunakan batas syarat minimal banyaknya peminjaman buku dengan kode tertentu yaitu sebesar 3% dari total transaksi peminjaman yang terjadi (*minimum support count* = 3%), nilai yang umum digunakan dalam perangkat lunak, buku-buku dan penelitian (Guidici, 2009). Setelah dilakukan perhitungan banyaknya buku yang sering dipinjam (*frequent itemsets*) setiap kode buku pada semua data peminjaman, maka langkah selanjutnya adalah melakukan

scanning untuk menentukan kode buku yang memiliki frekuensi minimal sebesar nilai *minimum support count* yaitu 3% dari 9.804 sebanyak 294 peminjaman.

Tabel 2. Kode Buku yang Memenuhi Syarat *Minimum Support Count*

No	Kode Buku	Nama Buku	Banyaknya Dipinjam
1	297	Islam, Babism & Bahai Faith	1.875
2	302	Social Interaction	1.469
3	658	General management	1.300
4	1001	Knowledge	1.012
5	659	Advertising & public relations	622
6	155	Differential & developmental psychology	508
7	346	Private law	481
8	657	Accounting	471
9	371	Schools & their activities; special education	408
10	370	Education	397
11	332	Financial economics	373
12	338	Production	356
13	1070	News media, journalism & publishing	331
14	301	Sociology & anthropology	309

Informasi dari Tabel 2. akan digunakan dalam pembentukan *FP-tree*. Dengan bantuan *software* RapidMiner 5 diperoleh hasil pembentukan *FP-tree* yang sebagaimana disajikan pada Gambar 1. sebagai berikut:



Gambar 1. Hasil Pembentukan *FP-tree*

Langkah selanjutnya adalah melakukan 3 tahapan utama pada algoritma *FP-Growth* yaitu tahap pembangkitan *conditional pattern base*, tahap pembangkitan *conditional FP-tree* dan tahap pencarian *frequent itemsets*. Hasil dari 3 tahapan utama

tersebut disajikan dalam Tabel 3. sebagai berikut:

Tabel 3. Tiga Tahapan Utama

Kode Buku	Conditional Pattern Base	Conditional FP-tree	Frequent Itemsets
301	{302: 77}	[302: 77]	{302, 301: 77}
332	{297: 108}; {658: 84}	[297: 108, 658: 84]	{297, 332: 108}, {658, 332: 84}
370	[[{297, 371: 63}]; [{371: 43}]]	[297: 63, 371: 63]; [371: 43]	{297, 370: 63}, {371, 370: 106}, {297, 371, 370: 63}
371	{297: 101}	[297: 101]	{297, 371: 101}
657	{658: 299}	[658: 299]	{658, 657: 299}
346	{297: 300}	[297: 300]	{297, 346: 300}
1001	{302: 442}	[302: 442]	{302, 1001: 442}

Hasil Perhitungan *Support*

Pada bagian ini akan diuraikan perhitungan nilai *support*, yaitu ukuran yang menunjukkan seberapa besar peluang banyaknya transaksi peminjaman buku yang memuat *itemsets* yang dipinjam secara bersamaan dari keseluruhan data peminjaman di Perpustakaan Universitas Islam Bandung. Nilai *support* dari *itemsets* yang akan digunakan pada tahapan ini adalah minimal 0.03 (*minimum support* = 0.03) yang merupakan nilai yang umum digunakan. Tabel 4. berisikan nilai *support* untuk seluruh *frequent itemsets*.

Tabel 4. Nilai *Support*

<i>Frequent Itemsets</i>	<i>Support</i>
{302, 1001}	0.05*
{297, 346}	0.03*
{658, 657}	0.03*
{297, 332}	0.01
{371, 370}	0.01
{297, 371}	0.01
{658, 332}	0.01
{302, 301}	0.01
{297, 370}	0.01
{297, 371, 370}	0.01

*: memenuhi syarat *minimum support* 0.03

Hasil Perhitungan *Confidence*

Pada bagian ini akan diuraikan perhitungan nilai *confidence*, yaitu ukuran yang menunjukkan seberapa besar asosiasi antar 2 buku yang dipinjam secara bersamaan dari seluruh transaksi peminjaman yang memuat salah satu buku tersebut. Nilai *confidence* dari *itemsets* yang akan digunakan pada tahapan ini adalah minimal 0.3 (*minimum confidence* = 0.3) yang merupakan nilai yang umum digunakan.

Tabel 5. Nilai *Confidence*

<i>Frequent Itemsets</i>	<i>Confidence</i>
{657 ⇒ 658}	0.63*
{346 ⇒ 297}	0.62*
{1001 ⇒ 302}	0.44*
{302 ⇒ 1001}	0.30*
{658 ⇒ 657}	0.23
{297 ⇒ 346}	0.16

*: memenuhi syarat *minimum confidence* 0.3

Hasil Perhitungan *Lift Ratio*

Pada bagian ini akan diuraikan perhitungan nilai *lift ratio*, yaitu merupakan salah satu cara yang baik untuk melihat kuat tidaknya aturan asosiasi. Tabel 6. berisikan nilai *lift ratio* untuk setiap *itemsets* yang memiliki nilai *minimum confidence* 0.3.

Tabel 6. Nilai *Lift Ratio*

<i>Frequent Itemsets</i>	<i>Lift Ratio</i>
{657 ⇒ 658}	4.74
{346 ⇒ 297}	3.27
{1001 ⇒ 302}	2.92
{302 ⇒ 1001}	2.92

Nilai *lift ratio* diatas menjadi informasi bagi pihak Perpustakaan Universitas Islam Bandung dalam hal kode buku yang dapat direkomendasikan kepada peminjam dan juga penempatan rak buku secara efektif. Maka ketiga *itemsets* tersebut dapat dijadikan rekomendasi bagi peminjam buku yang berkunjung ke Perpustakaan Universitas Islam Bandung dan buku tersebut dapat diletakkan berdampingan.

D. Kesimpulan dan Saran

Kesimpulan

Dalam skripsi ini telah dibahas *market basket analysis* dengan algoritma *FP-Growth* yang diaplikasikan untuk data transaksi peminjaman buku di Perpustakaan Universitas Islam Bandung. Berdasarkan hasil dan pembahasan dapat disimpulkan bahwa:

1. Aplikasi algoritma *FP-Growth* dengan minimum support count 3% dari keseluruhan transaksi didapatkan 14 kode buku yang sering dipinjam dan didapatkan 10 *itemsets* yang sering dipinjam secara bersamaan dari data peminjaman buku di Perpustakaan Universitas Islam Bandung.
- 2.
3. Untuk buku Social Interaction dan buku Knowledge (kode buku 302 dan 1001) memiliki nilai support 0.05 sedangkan untuk nilai confidence {302 ⇒ 1001} sebesar 0.30 dan untuk nilai confidence {1001 ⇒ 302} sebesar 0.44. Untuk buku Islam, Babism & Bahai Faith dan buku Private law (kode buku 297 dan 346) memiliki nilai support 0.03 sedangkan untuk nilai confidence {346 ⇒ 297} sebesar 0.62. Untuk buku General management dan buku Accounting (kode

buku 658 dan 657) memiliki nilai support 0.03 sedangkan untuk nilai confidence {657 \Rightarrow 658} sebesar 0.63.

4. Terdapat 3 itemsets menarik yang dapat direkomendasikan kepada peminjam berdasarkan nilai lift ratio yaitu buku Accounting yang memiliki asosiasi positif terhadap buku General management (kode buku 657 \Rightarrow 658) yaitu dengan nilai 4.74. Buku Private law yang memiliki asosiasi positif terhadap buku Islam, Babism & Bahai Faith (kode buku 346 \Rightarrow 297) yaitu dengan nilai 3.27 dan buku Knowledge memiliki asosiasi positif terhadap buku Social Interaction (kode buku 1001 dan 302) yaitu dengan nilai 2.92.

Saran

Saran yang dapat dikemukakan dalam penulisan skripsi ini adalah:

1. Database yang dimiliki pihak Perpustakaan Universitas Islam Bandung sudah sistematis, tetapi database tersebut belum dapat digunakan secara optimal. Database perlu dikelola dengan baik sehingga dapat dioptimalkan untuk mendapatkan berbagai informasi yang berguna.
2. Disarankan untuk menyimpan buku-buku yang sesuai dengan hasil analisis ini agar membantu pengunjung lebih efektif dalam mencari buku.
3. Disarankan kepada peneliti lain untuk meneliti *Market Basket Analysis* dengan algoritma yang berbeda seperti algoritme *Eclat*, *Generalized Sequential Pattern (GSP)*, dan *GRI algorithm*.

Daftar Pustaka

- Davies., dan Beynon, P. (2004). *Database Systems*. Houndmills, Basingstoke, UK: Palgrave.
- Hajarisman, N. (2009). *Analisis Data Kategorik*. Bandung: Program Studi Statistika, Universitas Islam Bandung.
- Han, J., dan Kamber, M. (2006). *Data Mining: Concept and Techniques*. Waltham: Morgan Kaufmann Publishers.
- Kurniasih, F. (2011). *Analisa dan Perancangan Data Mining dengan Metode Market Basket Analysis untuk Analisa Pola Belanja Konsumen pada Tendencies Store*. Skripsi tidak dipublikasikan. Jakarta: Fakultas Sains dan Teknologi, Universitas Islam Negeri Syarif Hidayatullah.
- Megaputer. (2007). *Market Basket Analysis* (Online). (<http://www.megaputer.com>, diakses pada tanggal 6 Maret 2017).
- Pramudiono, I (2007). *Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data* (Online). (<http://www.ilmukomputer.org/wp-content/uploads/2006/08/iko-datamining.zip>, diakses pada tanggal 6 Maret 2017).
- Venkatachari, K. (2016). Market Basket Analysis using FP-Growth and Apriori Algorithm: A Case Study of Mumbai Retail Store. *Journal of Management Research*, Vol.8 Issue-1, 56-63.