

Penggunaan Algoritma Novel Utility Frequent Itemset Mining dalam Market Basket Analysis

(Kasus Data Transaksional dan Data *Profit* di PT. XYZ Kota Malang)

Use of the Novel Utility Frequent Itemset Mining Algorithm in Market Basket Analysis

¹Sarah Istiqomah, ²Teti Sofia Yanti

^{1,2}*Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung*

Jl. Tamansari No.1 Bandung 40116

email: listiqomahsarah12@gmail.com, 2tetisofiyanti@gmail.com

Abstract. Data mining is mining or discovering new information by looking for certain patterns or rules from a very large amount of data, their use has also been widely applied in various fields. Nowadays, the advancement of information technology and computers has provided data storage facilities in an electronic format so that data storage is no longer a difficult job. One of the data techniques in data mining is Market Basket Analysis, which is used to find associations between different sets of products that customers put in baskets. This thesis will discuss Market Basket Analysis using Novel Utility Frequent Itemset Mining (NUFM) algorithm on customer transactional data and Minimarket profit data at PT. XYZ City of Malang to find associations between products that are often purchased simultaneously and have high utility is measured through the values of support count, transaction weightage, and utility weightage as well as the value of confidence to show how strong the association between the two products purchased simultaneously. There were 159 transactions consisting of 63 product codes. As a result, the overall combination of itemsets is 561 frequent 2 itemsets, two products purchased simultaneously and having a high utility consist of 155 itemsets. Of the 155 itemset that have high utility, there are 31 itemsets that have strong associations based on the value of confidence. The strongest associations are "langsingin" products and "pasta gigi herbal hpai new" with product code {olg1 => dpg4} of 35.00%, so it is advisable to keep the product close together or in a certain rack to make time and energy more effective in looking for products for services in the Minimarket PT. XYZ Malang City.

Keywords: Data Mining, Utility, NUFM, Support

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. Dewasa ini kemajuan teknologi informasi dan komputer telah menyediakan fasilitas penyimpanan data dalam format elektronik sehingga penyimpanan data bukan lagi menjadi satu pekerjaan yang sulit. 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 *Novel Utility Frequent Itemset Mining (NUFM)* pada data transaksional pelanggan dan data *profit* Minimarket PT. XYZ Kota Malang untuk menemukan asosiasi diantara produk-produk yang sering dibeli secara bersamaan dan memiliki utilitas tinggi diukur melalui nilai *support count*, *transaction weightage*, dan *utility weightage* serta nilai *confidence* untuk menunjukkan seberapa besar/kuat asosiasi antara dua produk yang dibeli secara bersamaan. Terdapat 159 transaksi yang terdiri dari 63 kode produk. Hasilnya dari keseluruhan kombinasi *itemsets* yaitu 561 *frequent 2 itemsets*, dua produk yang dibeli secara bersamaan serta memiliki utilitas yang tinggi terdiri dari 155 *itemsets*. Dari 155 *itemset* yang memiliki utilitas tinggi, terdapat 31 *itemsets* yang memiliki asosiasi kuat berdasarkan nilai *confidence*. Asosiasi yang paling kuat yaitu produk "langsingin" dan "pasta gigi herbal hpai new" dengan kode produk {olg1=>dpg4} sebesar 35,00%, maka disarankan untuk menyimpan produk tersebut berdekatan atau dalam suatu rak tertentu agar lebih efektifkan waktu dan tenaga dalam mencari produk untuk pelayanan di Minimarket PT. XYZ Kota Malang.

Kata Kunci: Data Mining, Utility, NUFM, Support

A. Pendahuluan

Data mining merupakan

serangkaian proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Dewasa ini

kemajuan teknologi informasi dan komputer telah menyediakan fasilitas penyimpanan data dalam format elektronik sehingga penyimpanan data bukan lagi menjadi satu pekerjaan yang sulit. Sebagai konsekuensinya jumlah data yang disimpan mengalami peningkatan yang sangat cepat dari segi kuantitas dan kualitas.

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 produk yang sering dibeli secara bersamaan dari data transaksi pelanggan. Terdapat beberapa algoritma dalam metode ini, diantaranya yaitu *Apriori*, *Frequent Pattern Growth (FP-Growth)*, *Novel Utility Frequent Itemset Mining*.

Untuk menanggulangi keterbatasan dari algoritma *Apriori* dan *FP-Growth*, maka diperlukan algoritma baru yaitu algoritma *NUFM*. *NUFM* adalah sebuah algoritma untuk menemukan tingginya utilitas *itemsets*. Algoritma ini diusulkan untuk mengatasi keterbatasan algoritma *apriori*, yang mengabaikan harga dan kuantitas penjualan juga mengatasi keterbatasan algoritma *FP-growth* yang tidak menghitung tingginya utilitas. Selain itu algoritma ini juga untuk merancang strategi penempatan produk yang efektif untuk meningkatkan utilitas bisnis, dengan memanfaatkan komponen lain seperti kuantitas, keuntungan, dan biaya dalam data transaksional pelanggan di salah satu minimarket.

Adapun masalah yang dapat diidentifikasi dalam penelitian ini adalah:

1. Bagaimana menentukan dua produk yang sering dibeli secara bersamaan dan memiliki utilitas tinggi berdasarkan nilai *utility frequent itemset mining* pada

data transaksional di salah satu Minimarket PT. XYZ Kota Malang?

2. Bagaimana menentukan kekuatan asosiasi dua produk dari *utility frequent itemset mining* yang terpilih, berdasarkan nilai *confidence* pada data transaksional di salah satu Minimarket PT. XYZ Kota Malang?

B. Landasan Teori

Data Mining

1. Pengertian *Data Mining*

Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Turban dkk. 2005).

2. Model *Data Mining*

Proses *data mining* menurut Han dan Kamber (2006) dapat dibedakan menjadi dua model yaitu, model prediksi dan model deskripsi. Contoh model prediksi adalah metode klasifikasi, metode regresi dan metode analisis deret waktu. Sedangkan contoh model deksripsi adalah metode *clustering* dan metode *association rules*.

3. Teknik *Data Mining*

Beberapa teknik *data mining* antara lain (Larose, 2005) yaitu Klasifikasi, Regresi, Pengelompokan, Kaidah Asosiasi, dan Analisis Pola Sekuensial.

4. Tahapan *Data Mining*

Menurut Han dan Kamber (2006) tahap-tahap *data mining* tersebut adalah Pembersihan data, Integrasi data, Seleksi data, Transformasi data, Proses *Mining*, Evaluasi Pola, dan Presentasi Pengetahuan.

Association Rule

Mengikuti definisi asli Agrawal dkk. (1993) masalah *association rule mining* didefinisikan sebagai berikut: $I = \{i_1, i_2, \dots, i_n\}$ menjadi satu set atribut biner yang disebut *item* (produk). $D = \{t_1, t_2, \dots, t_m\}$ menjadi sekumpulan transaksi yang disebut *database*. Setiap transaksi dalam D memiliki *ID* transaksi dan berisi sub kumpulan *item* dalam I . Aturan asosiasi didefinisikan sebagai implikasi $X \rightarrow Y$, dijelaskan oleh tiga ukuran menarik yaitu *support*, *confidence*, dan *lift ratio* di mana X dan Y adalah himpunan *item* dan $X \cap Y \neq \emptyset$. $Support(\{X, Y\}) = P(X, Y)$

$$= \frac{\text{banyak transaksi yg memuat } X \text{ dan } Y}{\text{total banyaknya transaksi}} \quad (2.1)$$

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

$$= \frac{\text{banyak transaksi yg memuat } X \text{ dan } Y}{\text{banyaknya transaksi yg memuat } X} \quad (2.2)$$

$$Expected\ Confidence = \frac{\text{banyak transaksi yg memuat } Y}{\text{total banyaknya transaksi}} \quad (2.3)$$

$$Expected\ Confidence = \frac{\text{Persamaan (2.2)}}{\text{Persamaan (2.3)}} \quad (2.4)$$

Market Basket Analysis

Market basket analysis adalah suatu metodologi untuk melakukan analisis *buying habit* konsumen dengan menemukan asosiasi antar beberapa barang yang berbeda, yang diletakkan konsumen dalam *shopping basket* (keranjang belanja) yang dibeli pada suatu transaksi tertentu (Han, 2006).

Algoritma Novel Utility Frequent Itemset Mining (NUFM)

NUFM adalah sebuah algoritma untuk menemukan tingginya utilitas *itemsets*. Algoritma ini diusulkan untuk mengatasi kekurangan pendekatan *apriori*, yang mengabaikan harga dan kuantitas penjualan. Untuk mengidentifikasi *itemset* yang memiliki utilitas cukup tinggi bisa dengan cara mempertimbangkan komponen seperti kuantitas, keuntungan dan biaya.

Tabel 2.1 Prosedur Perhitungan Algoritma *NUFM*

Transaksi ID	I_1	I_2	...	I_m	Transaction Utility
(1)	(2)				(3)
1	x_{11}	x_{12}	...	x_{1m}	$\sum(x_{1j} * p_j)$
2	x_{21}	x_{22}	...	x_{2m}	$\sum(x_{2j} * p_j)$
3	x_{31}	x_{32}	...	x_{3m}	...
...
n	x_{n1}	x_{n2}	...	x_{nm}	$\sum(x_{nj} * p_j)$
Total					$\sum \sum(x_{ij} * p_j)$
Frekuensi	$\sum x_{i1}$	$\sum x_{i2}$...	$\sum x_{im}$	
Profit	p_1	p_2	...	p_m	$\sum p_j$
Significance Weightage	$= \frac{p_1}{\sum p_j}$	$= \frac{p_2}{\sum p_j}$...	$= \frac{p_m}{\sum p_j}$	
Utility Weightage	$= \frac{p_1}{\sum p_j} * x_{i1}$	$= \frac{p_2}{\sum p_j} * x_{i2}$...	$= \frac{p_m}{\sum p_j} * x_{ij}$	

Kolom (1) menjelaskan ID setiap anggota sebanyak i , dimana $i = 1$ sampai n . Kolom (2) menjelaskan produk-produk yang dibeli oleh

konsumen dimana x_{11} adalah transaksi ke-1 membeli produk pertama, x_{12} adalah transaksi ke-1 membeli produk kedua, dan seterusnya sampai x_{1j}

adalah transaksi ke-1 membeli produk ke- j , dimana $j = 1$ sampai m . Kolom (3) menjelaskan tentang perhitungan *transaction utility*.

Item-tree NUFM

Pola pembentukan *item-tree* algoritma *NUFM* menggunakan bilangan prima.

Tabel 2.2 Bilangan Prima yang Ditugaskan untuk Transaksi

Transaksi ID	1	2	3	-	159
Bilangan Prima	2	3	5	-	937

Struktur *item tree* : *item-tree* memiliki tiga simpul yang berbeda.

- Simpul pertama (lingkaran) merupakan atribut.
- Simpul kedua (persegi) merupakan kuantitas produk tertentu yang dibeli.
- Simpul ketiga (oval) merupakan perkalian dari bilangan prima simpul kedua.

C. Hasil Penelitian dan Pembahasan

Hasil Aplikasi Algoritma NUFM

Dalam bagian ini akan diaplikasikan algoritma *NUFM* pada data transaksi penjualan produk PT. XYZ Kota Malang. Dari data transaksi pembelian produk akan dicari satu produk terlebih dahulu yang dibeli oleh konsumen dan dihitung banyaknya transaksi yang membeli satu produk tersebut. Dengan bantuan *software Microsoft Excel* diperoleh himpunan produk serta banyaknya transaksi yang membeli produk tersebut, hasilnya disajikan dalam Tabel 4.1.

Tabel 4.1 Frekuensi untuk Masing-masing Produk

No	Kode Produk	Nama Produk	Frekuensi
1	bdb1	Beauty Care Set	8
2	kdc1	Beauty Day Cream	38
3	knc1	Beauty Night Cream	40
⋮	⋮	⋮	⋮
63	kdb2	Deep Beauty 50 MI	1
	Total		1752

Sumber: PT. XYZ Kota Malang, 2017.

Langkah selanjutnya yaitu melakukan *scanning* untuk menentukan kode produk yang memenuhi syarat *minimum support count* menurut Giudici (2009) yaitu sebesar 2% dari 159 transaksi yang terjadi sehingga didapat *minimum support count* nya sebesar 4 transaksi. Dari transaksi yang terjadi, terdapat 34 jenis produk yang memenuhi syarat tersebut.

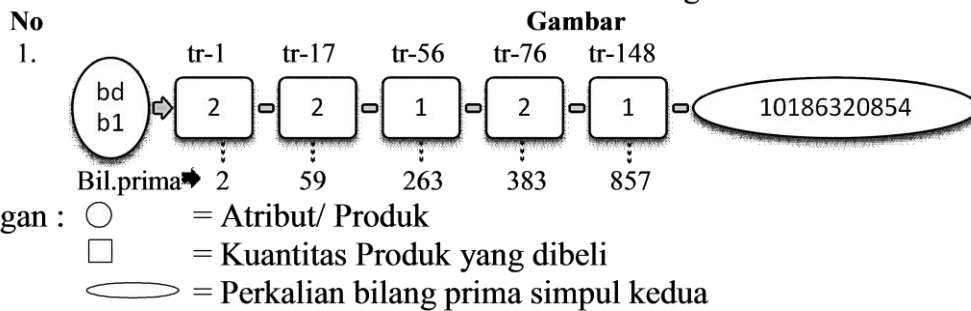
Tabel 4.2 Kode Produk yang Memenuhi Syarat *Minimum Support Count*

No	Kode Produk	Nama Produk	Frekuensi
1	bdb1	Beauty Care Set	8
2	kdc1	Beauty Day Cream	38
⋮	⋮	⋮	⋮
34	mkh1	Hpai Coffee	27

Tabel 4.3 Perhitungan *Utility Frequent Itemset Mining*

No	Transaksi ID	Kode Produk					Transaction Utility (T.U) (Rp)
		bdb1 (1)	kdc1 (2)	knc1 (3)	...	mkh1 (34)	
1	742141	2	11	10	...	0	2.283.000
2	548670	0	0	0	...	0	437.000
3	647792	0	0	0	...	0	90.000
⋮	⋮	⋮	⋮	⋮	...	⋮	⋮
159	862921	0	0	0	...	0	20.000
Total							$\Sigma T.U = 27.764.000$
Frekuensi		8	38	40	...	27	1699
Keuntungan (Profit) (Rp)		500.000	20.000	25.000	...	30.000	1.122.000
Significance Weightage (S.W)		0,44563	0,01783	0,02228	...	0,02674	1
Utility Weightage (U.W)		757,13012	30,28520	37,85651	...	45,42781	

Tabel 4.4 Pembentukan *Item-tree* Algoritma *NUFM*



Selanjutnya menghitung nilai *transaction utility*, *significance weightage*, dan *utility weightage* serta pembentukan *Item-tree* berdasarkan Tabel 4.2

Selanjutnya menghitung syarat validitas *itemsets* berdasarkan Tabel 4.3, syarat validitas *itemset* adalah $support\ count \geq minimum\ support\ count$, $utility\ weightage \geq minimum$

utility weightage, dan *transaction utility* (∂) $\geq minimum\ transaction\ utility$ (∂). Nilai-nilai *minimum support count* = 4, *minimum utility weightage* = 4,54278 \approx 5, dan *minimum transaction utility* (∂) = 50% x 27.764.000 = Rp. 13.882.000.

Tabel 4.5 Kombinasi Dua *Itemsets*

No	Frequent Itemsets	Support Count ($X \cap Y$)	Transaction Weightage	Utility Weightage	Utility Frequent
1	{bdb1, kdc1}	18	8.000.000	6726,30025	Tidak
2	{bdb1, knc1}	16	9.030.000	6036,42330	Tidak
⋮	⋮	⋮	⋮	⋮	⋮
561	{hdg2, mkh1}	21	4.932.000	603,64233	Tidak

Lakukan pembentukan *itemtree* sampai ke-34 produk. Langkah selanjutnya yaitu membuat kombinasi dua *itemsets*. Berdasarkan Tabel 4.4 kombinasi yang terjadi sebanyak $\binom{34}{2} = 561$. Misalnya konsumen membeli produk {bdb1} dan {kdc1} secara bersama-sama. Bilangan prima {bdb1} = {2, 59, 263, 383, 857}, dan bilangan prima {kdc1} = {2, 7, 11, 13, 29, 59, 137, 211, 257, 263, 331, 353, 367, 379, 383, 421, 659}, sehingga bilangan prima yang terbentuk ketika konsumen membeli produk {bdb1} dan {kdc1} adalah {bdb1, kdc1} = {2, 7, 11, 13, 29, 59, 137, 211, 257, 263, 331, 353, 367, 379, 383, 421, 659, 857}. Bilangan prima yang terbentuk tersebut mengindikasikan terjadi 18 transaksi, yaitu transaksi ke-1, 4, 5, 6, 10, 17, 33, 47, 55, 56, 67, 71, 73, 75, 76, 82, 120 dan 148. Banyaknya transaksi tersebut menunjukkan nilai *support count* = 18, selanjutnya menghitung *utility weightage* = $(U.W_{bdb1} + U.W_{kdc1}) \times support\ count = (359,31091 + 14,37244) \times 18 = 6726,30025$, dan *transaction weightage* = $T.U_1 + T.U_4 + \dots + T.U_{148} = 2.283.000 + 147.000 + \dots + 846.000 = Rp\ 8.000.000,-$, dengan cara yang sama akan terbentuk 561 nilai *support count*, *transaction weightage*, dan *utility weightage*.

Dari tabel 4.5 yang termasuk kedalam *utility frequent itemset* ada 155 *itemset*. Dari 155 *itemset* tersebut dihitung kekuatan asosiasi dari kombinasi yang terjadi menggunakan nilai *confidence*.

Hasil Perhitungan Confidence

Frequent Itemset dikatakan kuat apabila memenuhi syarat *minimum confidence* sebesar 2% atau 0,02 sebagaimana yang disarankan oleh Giudici (2009) yaitu *confidence* \geq *minimum confidence*.

Tabel 4.6 Perhitungan Nilai *Confidence*

No	Association Rule ($X \Rightarrow Y$)	Support Count ($X \cap Y$)	Confidence (%) $P(Y X)$	Kekuatan Asosiasi
1	{bdb1 = > msg1}	83	16,60	Ya
2	{bdb1 = > meo1}	65	13,00	Ya
⋮	⋮	⋮	⋮	⋮
155	{ksk1 = > mke1}	46	1,77	Tidak

Tabel 4.7 Pengelompokkan *Association Rule* Berdasarkan Jenis Produk

No	Association Rule ($X \Rightarrow Y$)	Support Count ($X \cap Y$)	Confidence (%) $P(Y X)$	Kekuatan Asosiasi
1	{olg1 = > dpg4}	70	35,00	Ya
2	{olg1 = > omb2}	57	28,50	Ya

∴	∴	∴	∴	∴
3 1	{mhc1= >dpg4}	86	2,10	Ya

D. Kesimpulan

Berdasarkan hasil dan pembahasan dapat disimpulkan bahwa:

1. Dari keseluruhan kombinasi *itemsets* yaitu 561, kemungkinan dua produk yang dibeli secara bersamaan serta memiliki utilitas yang tinggi terdiri dari 155 *itemsets*.
2. Dari 155 *itemset* yang memiliki utilitas tinggi, terdapat 31 *itemsets* yang memiliki asosiasi kuat. Asosiasi yang paling kuat yaitu produk “langsingin” dan “pasta gigi herbal hpai new” dengan kode produk {olg1=>dpg4} sebesar 35,00%. Asosiasi yang kurang kuat yaitu produk “hpai health coffee (wil 1&2)” dan “pasta gigi herbal hpai new” dengan (kode produk mhc1 dan dpg4) sebesar {mhc1=>dpg4} = 2,10%. Sedangkan asosiasi yang paling banyak yaitu produk “beauty night cream” dengan kode produk {knc1}, yaitu {knc1=>msg1}, {knc1=>dpg4}, {knc1=>meo1}, {knc1=>omb2}, {knc1=>mhc1}, {knc1=>ksk1}, dan {knc1=>mke1}.

E. Saran

Saran yang dapat dikemukakan dalam penelitian ini adalah:

1. *Database* yang dimiliki oleh PT. XYZ Kota Malang sudah sistematis, tetapi belum dapat digunakan secara optimal karena perlu input data terlebih dahulu dengan sistem manual.

Database perlu dikelola dengan baik sehingga dapat dioptimalkan untuk mendapatkan berbagai informasi yang berguna.

2. Disarankan untuk menyimpan produk-produk yang sesuai dengan analisis ini secara berdekatan atau dalam suatu rak tertentu agar lebih mengefektifkan waktu dan tenaga dalam mencari produk untuk penjual.
3. Disarankan kepada peneliti lain untuk meneliti *Market Basket Analysis* dengan algoritma yang berbeda seperti *Generalized Sequential Pattern (GSP)*, dan *GRI Algorithm*.

Daftar Pustaka

- Agrawal, R., Imielinski, T., dan Swami, A. (1993). Mining Association Rules between Sets of Items in Large Databases. *International Conference on Management of Data*. Washington DC, USA.
- Cavique, L. (2007). A Scalable Algorithm for Market Basket Analysis. *Journal of Retailing and Consumer Service*, 14; 400-407.
- Fanani, Yunus. (2013). *Teknik dan Metode Data Mining* (online), (<http://yunus-fanani.blogspot.com/2013/01/data-mining-teknik-dan-metodenya.html>, diakses pada 14 Desember 2018).
- Giudici, Paolo., dan Figini, Silvia. (2009). *Applied Data Mining for Business and Industry. Second Edition*. Italy: A John Wiley and Sons, Ltd., Publication.
- Gunawan., dkk. (2015). *Pengembangan Aplikasi Market Bakset Analysis Menggunakan Algoritma Generalized Sequential Pattern pada Supermarket*. Yogyakarta:

- Seminar Nasional Aplikasi Teknologi Informasi (SNATi).
- Han, J., dan Kamber, M. (2006). *Data Mining: Concept and Techniques*. Waltham: Morgan Kaufmann Publishers.
- Jabbar, M.A., Deekshatulu, B.L., dan Chandra, P. (2016). A Novel Algorithm for Utility Frequent Itemset Mining in Market Basket Analysis. *Journal of Advances in Intelligent Systems and Computing* 424, DOI 10.1007/987-3-319-28031-8_29.
- Karmawan, Gusti. M. (2017). *Konsep Utilitas/Daya Guna* (online), (<https://sis.binus.ac.id/2017/01/14/konsep-utilitasdaya-guna-utility/>), diakses pada 08 Mei 2018).
- Kumar, Sunil., dkk. (2012). Improved Apriori Algorithm Based on bottom up approach using Probability and Matrix. *International Journal of Computer Science Issues*, Vol. 9, Issues 2, No. 3, March 2012. ISSN (Online): 1694-0814.
- Kusrini dan E. T. Luthfi. (2009). *Algoritma Data Mining*. Yogyakarta : Andi Offset.
- Larose, D. T., dan Larose, C. D. (2005). *Discovering Knowledge in Data : an Introduction to Data Mining*. Canada: Jhon Wiley & Sons. Inc.
- [Nabarian, T. \(2012\). *Pengenalan Data Mining* \(Online\), \(<http://nabarian.blogspot.com/2012/01/pengenalan-data-mining.html>\), diakses pada 03 Mei 2018\).](http://nabarian.blogspot.com/2012/01/pengenalan-data-mining.html)
- Raharjo, Febrian. T. (2017). *Market Basket Analysis Menggunakan Algoritma FP-Growth*. Skripsi tidak dipublikasikan. Bandung: Fakultas MIPA Universitas Islam Bandung.
- Ramdhan, Faris. L. (2016). *Market Basket Analysis Menggunakan Algoritma Apriori*. Skripsi tidak dipublikasikan. Bandung: Fakultas MIPA Universitas Islam Bandung.
- Samuel, David. (2008). *Penerapan Struktur FP Tree dan Algoritma FP Growth dalam Optimasi Penentuan Frequent Itemset*. Bandung: Institut Teknologi Bandung.
- Turban, dkk. (2005). *Decision Support Systems and Intelligent Systems*. Yogyakarta: ANDI.
- Wideskills. (2015). *Data Mining Tutorial* (Online), (<http://www.wideskills.com/data-mining-tutorial/05-data-mining-tasks>), diakses pada 03 Mei 2018).