

Implementasi Algoritma *Fuzzy c-Covering* untuk Mengetahui Pola Pembelian pada Data Transaksi Swalayan

Kuntari Nur Angraini¹, Indwiarti², Fhira Nhita³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹kuntariangraini@students.telkomuniversity.ac.id

²indwiarti@telkomuniversity.ac.id

³fhiranhita@telkomuniversity.ac.id

Market Basket Analysis merupakan sebuah analisis terhadap kebiasaan pembeli pada saat berbelanja, tujuannya untuk mengetahui item apa saja yang dibeli. Penelitian ini bertujuan untuk mengimplementasikan *Fuzzy c-Covering* pada data transaksi swalayan untuk mengetahui barang yang sering dibeli oleh konsumen dan mengukur performansi yang didapatkan oleh *rule* untuk data transaksi swalayan, peneliti mengumpulkan data-data transaksi swalayan Sastra Mas di Bali dalam waktu satu tahun. Hasil *minimum support* 1% yaitu sebanyak 4 *association rule* dan hasil dari *minimum support* 0.5% yaitu sebanyak 10 *association rule*. Pada setiap penambahan nilai *minimum support* yang diujikan, terlihat *rule* yang terbentuk semakin berkurang.

Kata kunci : *Market Basket, Fuzzy C-Covering,*

Abstract

Market Basket Analysis is an analysis of the habits of buyers when shopping, the goal is to find out what items are purchased. This study aims to implement *Fuzzy c-Covering* on self-service transaction data to find out items that are often purchased by consumers and measure the performance obtained by the rule for self-service transaction data, the researchers collected self-service Sastra Mas transaction data in Bali within one year. The 1% support result is 4 associations rule and the result of the support is 0.5% which is 10 associations rule. At each additional minimum value of support is tested, it appears that the rule that is formed decreases.

Keywords: *Market Basket, Fuzzy C-Covering*

1. Pendahuluan

Latar Belakang

Pemanfaatan teknologi informasi sekarang telah diterapkan hampir di semua aspek kehidupan, seperti yang terjadi pada *superstore*. Dengan memanfaatkan sistem komputer, *superstore* dapat mengolah dan mengumpulkan data transaksi penjualan dengan cepat. Namun pemanfaatan data transaksinya belum maksimal baru sebatas untuk laporan peningkatan dan penurunan penjualan. Sedangkan dalam memproduksi sebuah barang, produsen harus mengetahui berapa banyak penjualan yang dilakukan sebelumnya sehingga ini akan menjadi tolak ukur pada saat produksi selanjutnya. Tolak ukur ini dikenal dengan istilah *market basket*. *Market basket* adalah digunakan untuk melakukan analisis kebiasaan konsumen pada saat berbelanja dengan menemukan asosiasi antar beberapa *item* yang berbeda[1].

Selama ini, data mining yang membahas tentang *market basket* biasanya menggunakan algoritma *Apriori* untuk mendapatkan *association rule*. Data mining sendiri diimplementasikan ke berbagai bidang, diantaranya dalam bidang perdagangan, pendidikan, dan telekomunikasi. Salah satu metode untuk asosiasi adalah metode *Fuzzy C-Covering*. Metode *Fuzzy C-Covering* dapat digunakan untuk mengklasifikasikan elemen-elemen dari suatu himpunan universal yakni keseluruhan produk item pada minimarket menjadi partisi-partisi yang lebih fokus dan detail berdasarkan jenis produk item yang ada. Karena itu *Fuzzy C-Covering* diterapkan agar dapat mengatasi hambatan yang selama ini terjadi pada proses *market basket* di dalam mencari hubungan antar item.

Pada penelitian sebelumnya, telah dilakukan juga penelitian *market basket analysis* dengan metode *Fuzzy C-Covering* untuk menentukan pola pembelian pada toko buku[11]. Selain itu, pernah dilakukan juga penelitian penggunaan metode *Fuzzy c-Covering* untuk analisa *market basket* pada *supermarket*[7]. Dari hasil penelitian yang telah dilakukan menggunakan metode *Fuzzy C-Covering* dapat menghasilkan *association rule* dan tingkat asosiasinya.

Pada penelitian ini akan dilakukan analisa data transaksi swalayan untuk mengetahui pola pembelian dengan menggunakan algoritma *Fuzzy c-Covering* dan hasil yang didapat nantinya akan berupa *rule*, sehingga kita dapat melihat pola pembelian pada data transaksi swalayan yang diamati.

Topik dan Batasannya

Dalam penelitian ini diangkat beberapa permasalahan yaitu, bagaimana implementasi metode *Fuzzy c-Covering* untuk data transaksi swalayan agar mendapatkan *association rule* dan bagaimana mengukur performansi yang didapatkan oleh rule untuk data transaksi swalayan?.

Dari beberapa permasalahan yang telah dijelaskan diatas, penelitian ini memiliki batasan masalah dalam proses pengerjaannya meliputi data yang digunakan adalah data transaksi pembelian yang diperoleh dari transaksi swalayan harian pada tahun 2017

Tujuan

Dari beberapa rumusan masalah yang ada, tujuan yang ingin dicapai dari penelitian ini yaitu mengimplementasikan *Fuzzy c-Covering* pada data transaksi swalayan untuk mengetahui barang yang sering dibeli oleh konsumen dan mengukur performansi yang didapatkan oleh rule untuk data transaksi swalayan.

2. Studi Terkait

2.1. Data Mining

Data mining adalah adalah sebuah proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besar. dan untuk mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Adapun beberapa metode yang terdapat pada data mining ialah *Classification*, *Association rule*, *Clustering*, dan *Anomali*. kemudian beberapa algoritma yang terdapat dalam data mining ialah *Naive bayes*, *Random forest*, *Apriori*, dan sebagainya. Salah satu metode yang sering kali digunakan dalam teknologi data mining adalah metode asosiasi atau *association rule mining*. Di dalam bidang usaha retail metode *association rule mining* ini lebih dikenal dengan istilah analisa keranjang belanja (*Market basket analysis*) [4]. Tugas tersebut dapat diselesaikan dengan melakukan 2 tahap penting yaitu [5] :

- a. Mencari semua frequent itemset.
- b. Mencari aturan asosiasi yang confidence.

2.2. Association Rule

Association rule mining adalah suatu prosedur untuk mencari hubungan antar *item* dalam suatu *data set* yang ditentukan [8]. Dalam menentukan suatu *association rule*, terdapat suatu *interestingness measure* (ukuran kepercayaan) yang didapatkan dari hasil pengolahan data dengan perhitungan tertentu. Umumnya ada dua ukuran, yaitu:

1. *Support*: suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu *item/itemset* layak untuk dicari *confidence*-nya. Nilai *support* sebuah item diperoleh dengan rumus berikut:

$$\text{Support}(A) = \frac{\text{Jumlah Transaksi mengandung } A}{\text{Total Transaksi}}$$

2. *Confidence*: suatu ukuran yang menunjukkan hubungan antar 2 *item* secara *conditional*. Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiatif yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiatif $A \rightarrow B$, diperoleh rumus sebagai berikut:

$$\text{Confidence} = P(B|A) = \frac{\text{Support}(A \cap B)}{P(A)}$$

Kedua ukuran ini nantinya berguna dalam menentukan *interesting association rules*, yaitu untuk dibandingkan dengan batasan (*threshold*) yang ditentukan oleh *user*. Batasan tersebut umumnya terdiri dari *min_support* dan *min_confidence*.

2.3. Apriori Algorithm

Apriori adalah suatu algoritma yang sudah dikenal dalam melakukan pencarian *frequent Itemset* dengan *association rule*. Sesuai dengan namanya, algoritma ini menggunakan *frequent itemset* yang telah diketahui sebelumnya, untuk memproses informasi selanjutnya. Langkah-langkah dari algoritma ini adalah sebagai berikut [9] :

1. Set $k=1$ (menunjuk pada *itemset* ke-1).
2. Hitung semua k *itemset* (*itemset* yang mempunyai k *item*).
3. Hitung *support* dari semua calon *itemset filter itemset* tersebut berdasarkan perhitungan *minimum support*.
4. Gabungkan semua k -sized *itemset* untuk menghasilkan calon *itemset* $k+1$.
5. Set $k=k+1$.
6. Ulangi langkah 3-5 sampai tidak ada *itemset* yang lebih besar yang dapat dibentuk.
7. Buat *final set* dari *itemset* dengan menciptakan suatu *union* dari semua k -*itemset*.

2.4. Fuzzy Logic

Fuzzy Logic dikembangkan untuk menyediakan fungsi, dan aturan-aturan matematis yang memperbolehkan nput berupa bahasa yang alami. Menurut Ter Meulen (2001), bahasa yang digunakan secara umum oleh manusia dalam berkomunikasi secara lisan, ataupun tertulis. Teori Himpunan Fuzzy akan memberikan jawaban terhadap suatu masalah yang mengandung ketidakpastian. Kebanyakandari metode ini berbicara tentang keambiguan, yang mana bisa kita temukan dalam kehidupan kita sehari-hari. Berbeda dengan logika konvensional yang secara alami dapat dihubungkan dengan kondisi Boolean (benar/salah; I/O), *Fuzzy Logic* mencoba untuk menentukan daerah dimana kebenaran yang mutlak tidak dapat dicapai, begitu pula dengan kesalahan yang mutlak, yaitu daerah diantara kebenaran dan kesalahan yang mutlak tersebut. Pada beberapa kasus khusus, seperti nilai keanggotaan yang kemudian akan menjadi 0 atau 1, teori dasar tersebut akan identik dengan teori himpunan biasa dan himpunan Fuzzy akan menjadi himpunan *crisp* biasa.

Logika Fuzzy dikatakan sebagai logika baru yang lama, sebab ilmu tentang logika Fuzzy modern dan metodis baru yang ditemukan beberapa tahun yang lalu, padahal sebenarnya konsep tentang logika fuzzy itu sendiri sudah ada sejak lama. Konsep dari *Fuzzy Logic* diperkenalkan oleh Professor Lotfi A. Zadeh, di Barkley pada Universitas California (University of California) pada 1960an

2.5. Fuzzy c-Covering

Fuzzy c-Covering merupakan salah satu metode yang dipakai untuk mengklasifikasi elemen-elemen dari suatu himpunan universal menjadi partisi-partisi berupa *fuzzy sets*. *Fuzzy c-Covering* sendiri merupakan generalisasi dari metode *fuzzy c-partition* yang telah dikenal sebelumnya [10]. Untuk memperbaiki kelemahan Algoritma Apriori, algoritma baru ini diajukan. Dalam mencari hubungan antar *item*, metode ini berdasarkan pada persepsi bahwa semakin banyak *item* yang dibeli dalam satu *transaksi* itu semakin lemah.

Berikut adalah langkah-langkah dalam algoritma yang dibuat berdasarkan metode baru tersebut:

1. menentukan *max_item_threshold* yang dibutuhkan. *Max_item_threshold* adalah suatu pembatas yang dipakai untuk menyaring transaksi berdasarkan jumlah *item* dalam transaksi tersebut.
2. mencari *record-record* dalam tabel transaksi yang memenuhi *max_item_threshold* dan menyimpannya ke dalam QT, dimana:

$$QT = \{t \mid |t| \leq ith, ith \in \text{positive integer}\}$$

dimana: QT (*Qualified Transaction*): himpunan transaksi yang memenuhi *max_item_threshold*; t : transaksi; $|t|$: jumlah *item* dalam suatu transaksi; ith *max_item_threshold*.

3. set $k=1$ (k sudah variabel untuk menentukan jumlah kombinasi).
4. menentukan *min_support* ke- k .
5. mencari *support* dari setiap kombinasi k -*item* yang memungkinkan yang ada di dalam transaksi tersebut dengan rumus:

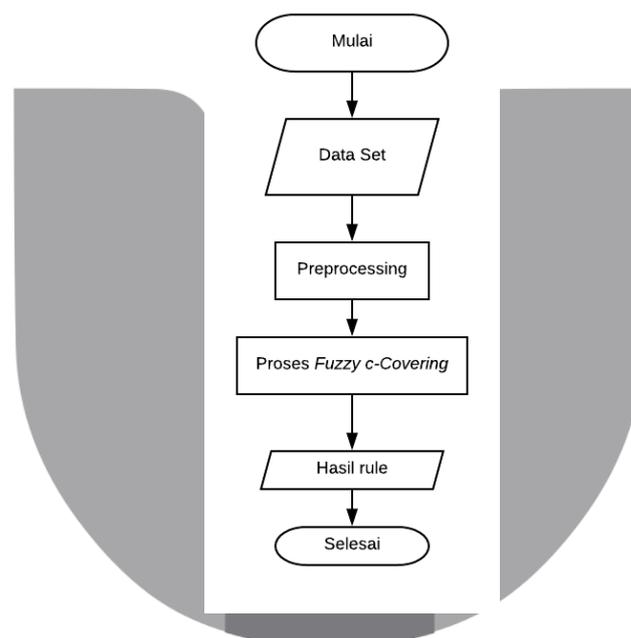
$$Support(A) = \frac{\text{Jumlah Transaksi mengandung } A}{\text{Total Transaksi}}$$

6. melakukan penyaringan terhadap kombinasi *item* yang ada di dalam transaksi tersebut yang tidak memenuhi
7. set $k=k+1$, dimana jika $k > \text{ith}$.
8. mendefinisikan tiap *item* yang telah didapat dari langkah-langkah di atas sebagai *fuzzy set* (disebut *item fuzzy set*) terhadap transaksi QT.
9. mencari *candidate rules* dengan cara menghitung *confidence* dari setiap kombinasi *k-item* yang memenuhi min_support ke- k dari *item fuzzy set* dengan rumus:

$$\text{Confidence} = P(B|A) = \frac{\text{Support}(A \cap B)}{P(A)}$$

3. Sistem yang Dibangun

Dalam penelitian ini, sistem yang di rancang bertujuan untuk menghasilkan association rules dari item-item di dalam transaksi. Sistem yang dibangun dengan mengimplementasikan bantuan *data mining* menggunakan metode *Fuzzy c-Covering*.



Gambar 3.1 Gambaran Umum Flowchart

Pada gambar 3.1 adalah gambaran umum flowchart yang dilakukan dalam penelitian ini. Langkah pertama yang dilakukan adalah dengan mengambil data berupa data transaksi yang didapat dari Swalayan. Kemudian data tersebut dilakukan tahapan preprocessing. Tahapan ini merupakan langkah awal dari proses pengolahan data, kemudian data tersebut dilakukan proses *fuzzy c-Covering*.

3.1. Data Set

Dalam penelitian ini, dataset yang digunakan adalah data transaksi swalayan. Data tersebut diperoleh dari supermarket yang ada di Bali. Pada penelitian ini, data transaksi swalayan tersebut terdapat 81606 record dengan 8 atribut dalam 1 tahun. Penulis disini mengambil disini mengambil 4 variabel yang akan digunakan yaitu Faktur, Tanggal, Kode Barang, dan Nama Barang.

Tabel 3.1 Data transaksi swalayan

No	Faktur	Tgl	Kod Bar	Nama Barang	Jum	Harga	Total
1.	0117/FJ/0001-1	1-Jan-17	9E+12	Mama Lemon	1	7.800	7.800
2.	0117/FJ/0001-1	1-Jan-17	9E+12	Lux White	2	3.200	6.400
3.	0117/FJ/0002-1	1-Jan-17	9017	Aqua dingin	4	4.000	16.000
4.	0117/FJ/0002-1	1-Jan-17	9472	You c-1000	1	6.500	6.500
5.	0117/FJ/0002-1	1-Jan-17	9E+12	Kis Mint	1	4.000	6.500
....	
81606	1217/FJ/2380-0	31-Des-17	9E+12	Dji Sam Soe	1	19.500	19.500

Sumber : Database penjualan Swalayan Sastra Mas

Pada swalayan Sastra Mas banyak sekali jenis produk barang yang tersedia, oleh karena itu disini penulis akan mengelompokkan produk berdasarkan kategori produk tersebut. Berikut adalah kategori produk yang dipilih :

- | | |
|----------------|----------------|
| 1. Coklat | 15. Potato |
| 2. Susu | 16. Larutan |
| 3. Teh pucuk | 17. Mie |
| 4. Kopi | 18. Ale-ale |
| 5. Air mineral | 19. Garuda |
| 6. Fanta | 20. Buavita |
| 7. Jelly | 21. Freastea |
| 8. Roti | 22. You c 1000 |
| 9. Sprite | 23. Yakult |
| 10. Coca cola | 24. Piattos |
| 11. Chitato | 25. Oreo |
| 12. Wafer | 26. Oishi |
| 13. Lays | 27. Permen |
| 14. Pocari | |

Dari pemilihan kategori produk di atas, dilihat berdasarkan data dari kategori produk-produk yang sering dibeli konsumen setiap hari. Setelah melakukan pengkategorian, penulis melakukan *preprocessing* dengan cara diskritisasi. Berikut merupakan data yang merupakan hasil dari diskritisasi :

No	Kategori	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	27
1	8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	...	0
2	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...	1
3	4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	...	0
...
9648	12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	...	0

Dari hasil diskritisasi diatas, maka dapat dilihat jumlah record pada tabel diatas mengalami pengurangan, sebelumnya dari data asli terdapat 81606 record lalu diolah menjadi 9648 record dan 28 atribut.

3.2. Fuzzy c-Covering



Gambar 3.2 Alur tahapan *Fuzzy c-Cocering*

Pada gambar 3.2 akan dijelaskan mengenai alur tahapan dalam *Fuzzy c-Covering*. Langkah pertama yang perlu dilakukan adalah menentukan Batasan atau *max_item_threshold* (ith). Setelah menentukan ith maka selanjutnya mengecek record yang memenuhi ith dan menyimpannya ke dalam QT. Lakukan set k=1 yakni k berupa jumlah kombinasi. Kemudian langkah selanjutnya yaitu menentukan minimum support sebagai batasan, mencari support dari setiap kombinasi dan akan dilakukan penyaringan pada setiap item untuk memilih yang memenuhi minimum support agar bisa melanjutkan ke tahap berikutnya. Hasil dari yang memenuhi *min_sup* didefinisikan tiap itemnya yang akan dijadikan fuzzy set. Setelah itu mencari rule dengan menghitung nilai confidence.

4. Evaluasi

Pada tugasakhir ini data yang digunakan adalah transaksi pada swalayan. Data tersebut diperoleh dari supermarket yang ada di Bali. Pada penelitian ini, data transaksi swalayan yang telah diolah sehingga terdapat 31421 record dengan 8 atribut dalam 1 tahun.

4.1 Hasil Pengujian

Pada pengujian dilakukan analisis dari hasil yang telah diperoleh menggunakan algoritma *fuzzy c-Covering*, adapun permasalahan akan diselesaikan dalam 2 skenario. Pada skenario ini menggunakan data satu tahun dan *max item threshold* yang berbeda. Diasumsikan *min_support* untuk tiap kombinasi bernilai sama.

Tabel 4-1. Dengan *max item threshold* 8, didapatkan:

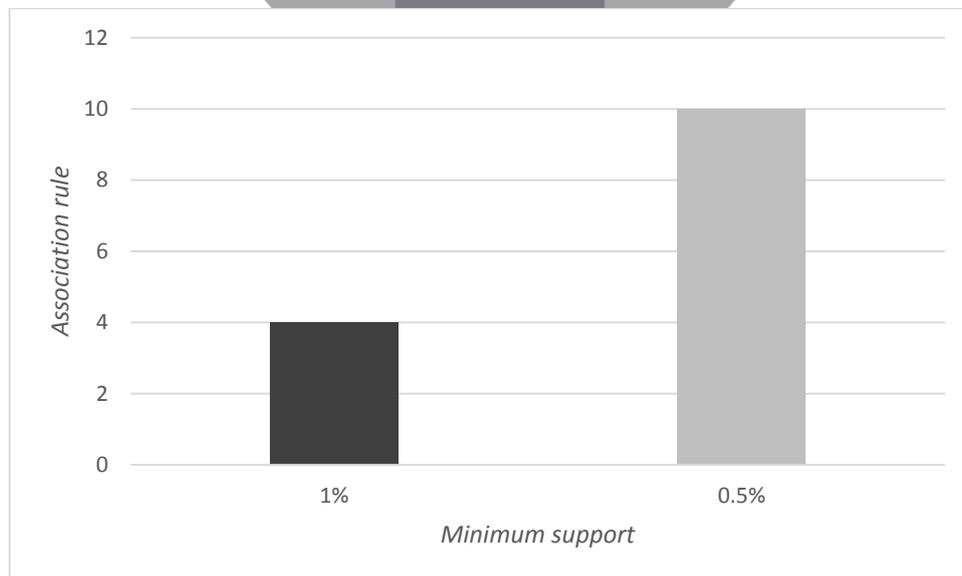
Association rule (AR) yang terbentuk jika minimum confidence	1%	4 AR	support	Waktu	Confidence	Waktu
		Air mineral → roti	0.0195	2.7343	0.0862	9.154
Roti → air mineral	0.0195	0.0659				
Susu → roti	0.0125	0.0713				
Roti → susu	0.0125	0.0424				
Association rule (AR) yang terbentuk jika minimum confidence	0.5%	10 AR	support	Waktu	Confidence	Waktu
		Coklat → roti	0.00814	3.5541	0.0731	15.40
		Roti → coklat	0.00814		0.0267	
		Mie → roti	0.00588		0.0371	
		Roti → mie	0.00588		0.0207	
		Teh pucuk → roti	0.0078		0.0577	
		Roti → teh pucuk	0.0078		0.0263	
		Air mineral → roti	0.0195		0.0862	
		Roti → air mineral	0.0195		0.0659	
		Susu → roti	0.0125		0.0713	

		Roti → susu	0.0125		0.0424	
--	--	-------------	--------	--	--------	--

Tabel 4-2. Dengan *max item threshold* 15, didapatkan:

Association rule (AR) yang terbentuk jika minimum confidence	Confidence	support	Waktu	Confidence	Waktu	
						Waktu
Association rule (AR) yang terbentuk jika minimum confidence	1%	4 AR	support	Waktu	Confidence	9.35
		Air mineral → roti	0.01958	3.747	0.086293	
		roti → air mineral	0.01958		0.065888	
		susu → roti	0.01258		0.071354	
		roti → susu	0.01258		0.042435	
	0.5%	10 AR	support	waktu	Confidence	waktu
		Coklat → roti	0.008106	3.209	0.07335	18.587
		Roti → coklat	0.008106		0.0268	
		Mie → roti	0.00589		0.0375	
		Roti → mie	0.00589		0.021	
		Teh pucuk → roti	0.007803		0.0579	
		Roti → teh pucuk	0.007803		0.0264	
		Air mineral → roti	0.0195		0.0862	
		Roti → air mineral	0.0195		0.0658	
Susu → roti	0.0125	0.07135				
Roti → susu	0.0125	0.0424				

Dapat dilihat pada Tabel di atas merupakan *assosiation rule* yang muncul pada *minimum support* yang telah ditentukan.



Gambar 4-3 Grafik *minumum support* terhadap *assosiation rule*.

Berdasarkan grafis pada gambar di atas, dapat dilihat hasil *minumum support* 1% yaitu sebanyak 4 *assosiation rule* dan hasil dari *minumum support* 0.5% yaitu sebanyak 10 *assosiation rule*. Pada setiap penambahan nilai *minimum support* yang diujikan, terlihat *rule* yang terbentuk semakin berkurang. Pada pengujian ini dapat disimpulkan semakin tinggi nilai *minimum support* yang ditentukan maka *rule* yang dihasilkan akan semakin sedikit dan pola pembelian barang yang sering dilakukan oleh konsumen pada swalayan tersebut dapat semakin terlihat.

4.2 Analisis Hasil Pengujian

Hasil analisis pola diatas menunjukkan semakin tinggi *max item threshold* yang digunakan maka semakin tinggi juga nilai *support* yang dihasilkan, tetapi tidak mempengaruhi nilai *confidence* yang dihasilkan, dan waktu yang dibutuhkan semakin cepat.

5. Kesimpulan

Dalam penelitian yang telah dilakukan, metode *Fuzzy c-Covering* dapat menghasilkan *association rule*. semakin tinggi *max item threshold* yang digunakan maka semakin tinggi juga nilai *support* yang dihasilkan, tetapi tidak mempengaruhi nilai *confidence* yang dihasilkan, dan waktu yang dibutuhkan semakin cepat. Dari beberapa percobaan, nilai *minimum support* 1% menghasilkan *support* dan *confidence* yang paling besar.

Daftar Pustaka

- [1]. Ibnu Sani Wijaya, *Aplikasi Data Mining dengan Konsep Fuzzy c-Covering untuk Analisa Market Basket pada pasar swalayan*.
- [2]. Finn Lee S & Juan Santana. 2010. *Data Mining : Meramalkan Bisnis Perusahaan*. Jakarta : Penerbit PT. Elex Media Komputindo..
- [3]. S, L. W. (2014). pembuatan perangkat lunak Data mining untuk penggalan kaidah Asosiasi menggunakan metode Apriori
- [4]. Kusriani, & Luthfi, e. t. (2013). *Algoritma Data mining*. Bogor.
- [5]. Ramalia, M. (2015). *Agricultural Productivity In South Africa: Literature Review*. Report on agricultural productivity in South Africa. Africa: South Africa.
- [6]. Albino Research Ltd. (2007). *Market Basket Analysis*, <http://www.albinoresearch.com/data-mining/market-basket.php> .
- [7]. Budhi, G. S., Resmana L., dan Osmand P. 2005. *Penggunaan Metode Fuzzy c-Covering untuk Analisa Market Basket pada Supermarket*.
- [8]. Han, Jiawei; Kamber, Micheline, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2001..
- [9]. Ulmer, David; "Mining an Online Auctions Data Warehouse." *The Mid-Atlantic Student Workshop on Programming Languages and Systems*. 19 April 2002. Pace University. <<http://csis.pace.edu/csis/masplas/p8.pdf>>.
- [10]. Klir, George J.; Yuan, Bo, *Fuzzy Sets and Fuzzy Logic – Theory and Applications*. India: Prentice Hall, 2001.
- [11]. Beeh Y.S, Nataliani Yessica. *Market Basket Analysis dengan Metode Fuzzy C-Covering untuk Menentukan Pola Pembelian pada Toko Buku*.