

PENERAPAN ASSOCIATION RULE DENGAN ALGORITMA APRIORI DALAM ANALISIS DATA MINING PENJUALAN BAHAN KIMIA PADA PT GRESIK CIPTA SEJAHTERA

Ariadi Retno Tri Hayati Ririd¹, Yopy Yunhasnawa²,
Dyah Kartika Dwi Andarini³

^{1,2,3} Program Studi Teknik Informatika, Teknologi Informasi, Politeknik Negeri Malang
¹ faniri4education@gmail.com, ² yunhasnawa@polinema.ac.id, ³ dyahkartikaa01@gmail.com

Abstrak

PT Gresik Cipta Sejahtera merupakan perusahaan yang bergerak dalam 7 bidang usaha utama untuk mendukung dan memasok kebutuhan pelanggan dari berbagai sektor, salah satunya dalam sektor pendistribusian bahan kimia. Untuk meningkatkan keuntungan yang diperoleh melalui penjualan bahan kimia diperlukan sebuah strategi bisnis yang mampu memperkirakan *stock* bahan kimia yang harus disiapkan dan dapat membantu dalam perencanaan RKAB untuk tahun selanjutnya.

Data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar. Data yang digunakan dalam proses *data mining* diperoleh dari data penjualan bahan kimia pada tahun sebelumnya. Informasi baru yang diperoleh dari proses *data mining* kemudian digunakan untuk merancang strategi bisnis. *Association rules* merupakan salah satu teknik *data mining* untuk menemukan aturan asosiatif antara suatu kombinasi *item*. Algoritma *apriori* dapat digunakan untuk melihat aturan pembelian antar *item* pada bahan kimia yang sering terjual secara bersamaan. Algoritma *apriori* menghasilkan *association rules* antar *item* pada bahan kimia, sehingga perusahaan dapat menentukan strategi bisnis serta menentukan penambahan *stock* pada *item-item* yang pola kombinasinya besar.

Dari hasil pengujian, dapat disimpulkan bahwa semakin besar *minimum support* maka semakin kecil kemungkinan *item* lolos sehingga semakin sedikit *item* yang dapat dikombinasikan. Tingkat akurasi yang dihasilkan sebesar 16% dengan tren turun.

Kata kunci : *Data Mining, Association Rules, Algoritma Apriori*

1. Pendahuluan

Saat ini, perkembangan teknologi telah memberikan pengaruh yang sangat besar didalam kehidupan manusia. Salah satu pengaruh tersebut dibidang informasi yaitu dalam aplikasi *database*. Pada aplikasi *database*, informasi memegang peranan yang sangat penting dan dibutuhkan dalam berbagai aspek kehidupan, baik dalam dunia pendidikan, bisnis, perbankan, dan lain-lain. Dengan bermanfaatnya informasi tersebut, banyak perusahaan berusaha untuk mengumpulkan informasi sebanyak-banyaknya untuk mendapatkan keuntungan yang maksimal.

PT Gresik Cipta Sejahtera merupakan salah satu perusahaan yang berada di lingkungan Petrokimia Gresik *Group*. PT Gresik Cipta Sejahtera menjalankan 7 bidang usaha utama untuk mendukung dan memasok kebutuhan pelanggan dari berbagai sektor industri, perkebunan, pertanian yang tersebar di seluruh Indonesia. Ketujuh bidang usaha tersebut yaitu distributor pupuk, distributor bahan kimia, distributor pestisida, produsen pupuk dan benih padi, pemasok kebutuhan industri, jasa angkutan barang, dan jasa pergudangan.

Mencari keuntungan pada penjualan bahan kimia merupakan salah satu tujuan yang harus tercapai di PT Gresik Cipta Sejahtera, maka untuk memenuhi tujuan itu diperlukan suatu strategi yang dapat meningkatkan penjualan sekaligus untuk memperhitungkan *stock* bahan kimia yang harus disiapkan serta dapat membantu dalam perancangan RKAB untuk tahun selanjutnya.

Penggunaan *Association Rules* dengan Algoritma *apriori* dapat dimanfaatkan dalam proses penjualan, dengan memberikan hubungan antar data penjualan, dalam hal ini adalah bahan kimia yang dipesan sehingga akan didapat pola pembelian konsumen. Pihak PT Gresik Cipta Sejahtera dapat memanfaatkan informasi tersebut untuk mengambil tindakan bisnis yang sesuai, dalam hal ini informasi dapat menjadi bahan pertimbangan untuk menentukan strategi penjualan selanjutnya.

2. Tinjauan Pustaka

2.1 Data Mining

Data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar.

Data mining, sering juga disebut sebagai *Knowledge Discovery In Database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. *Data mining* didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar Santoso, *et al* (2016).

2.2 Association Rules

Analisis asosiasi atau *association rule mining* adalah teknik *data mining* untuk menemukan aturan asosiatif antara suatu kombinasi *item*. Analisis asosiasi dikenal juga sebagai salah satu teknik data mining yang menjadi dasar dari salah satu teknik data mining lainnya. Secara khusus, salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien, yaitu analisis pola frekuensi tinggi (*frequent pattern mining*) Fauzy, *et al* (2016).

Metode *association rules* ini dapat membantu mengenali pola-pola tertentu di dalam kumpulan data yang besar. Pencarian *association rules* dilakukan melalui dua tahap yaitu pencarian *frequent itemset* (kombinasi yang paling sering terjadi) dan penyusunan *rules*. *Itemset* merupakan suatu kelompok *item*, *support* dari *itemset* X adalah persentase transaksi di P yang mengandung X, biasa ditulis dengan $supp(X)$. Setelah menemukan *frequent itemset* barulah mencari aturan asosiatif atau aturan keterkaitan yang memenuhi syarat yang ditentukan. *Association rule* memerlukan suatu variabel yang ditentukan oleh *user* untuk menentukan banyaknya *output* yang diinginkan *user*.

Sebuah *association rule* dengan *confidence* sama atau lebih besar dari *minimum confidence* dapat dikatakan sebagai *valid association rule* Norsyannah (2016).

2.3 Algoritma Apriori

Dalam bukunya, Kusri dan Luthfi mengemukakan bahwa algoritma Apriori adalah salah satu algoritma yang dapat digunakan pada penerapan *market basket analysis* untuk mencari aturan-aturan asosiasi yang memenuhi batas *support* dan *confidence*. Selama proses tahap pertama, algoritma menghasilkan penggalan secara sistematis tanpa menjelajahi semua kandidat, sedangkan pada tahap kedua dilakukan ekstraksi terhadap aturan yang kuat. *Frequent itemset* biasanya mengacu pada kumpulan *item* yang sering muncul bersamaan dalam sebuah data transaksional. Contohnya jika *item* A dan B sering dibeli bersamaan dalam suatu toko. Setelah menemukan *frequent itemset*, algoritma ini kemudian meneliti *knowledge* dari *frequent item* sebelumnya

untuk menggali informasi selanjutnya. Apriori menggunakan pendekatan iteratif dengan *level-wise search* dimana *k-itemset* dipakai untuk mencari $(k+1)$ -*itemset*.

Untuk mendapatkan nilai *support* dari suatu item X maka dapat dicari dengan menggunakan rumus berikut:

$$Support\ X = \frac{Jumlah\ transaksi\ yang\ mengandung\ item\ X}{Total\ Transaksi} \quad (1)$$

Kemudian untuk mencari nilai *support* dari dua item dapat menggunakan rumus berikut:

$$Support\ X, Y = \frac{P\ X \cap Y}{Total\ Transaksi} \quad (2)$$

Setelah semua *frequent item* didapat, maka ditentukan syarat minimum *Confidence*. *Confidence* adalah nilai ukuran seberapa besar valid tidaknya suatu *association rules*. *Confidence* suatu rule $R (X \rightarrow Y)$ adalah proporsi dari semua transaksi yang mengandung baik X maupun Y dengan yang mengandung X.

$$Confidence\ X \rightarrow Y = \frac{P\ X|Y}{Jumlah\ Transaksi\ yang\ mengandung\ item\ X\ dan\ Y} \quad (3)$$

Prinsip kerja dasar dari algoritma ini yaitu dengan mengembangkan *frequent itemset*. Mulai dari satu *item* dan secara rekursif mengembangkan *frequent itemset* menjadi dua *item*, tiga *item*, dan seterusnya hingga *frequent itemset* tidak dapat dikembangkan lagi.

Untuk mengembangkan *frequent itemset* dengan dua *item*, dapat menggunakan satu *item*, dengan alasan bila set satu *item* tidak dapat mencapai minimum *support*, maka setiap *itemset* dengan ukuran yang lebih besar juga tidak akan melebihi minimum *support*.

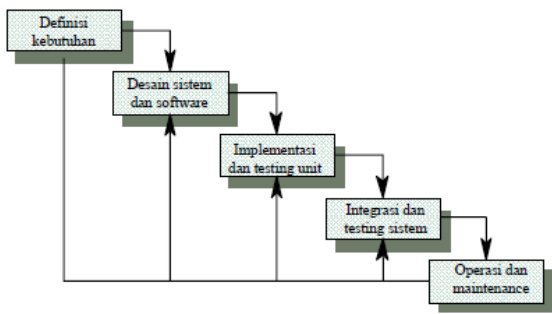
Terdapat dua proses utama dalam algoritma apriori yaitu :

- a. *Join* (Penggabungan)
Dalam proses ini, setiap item dikombinasikan dengan item lain sampai tidak dapat terbentuk kombinasi lagi.
- b. *Pruning* (Pemangkasan)
Pada proses ini, hasil kombinasi item akan dipangkas berdasarkan minimum *support* yang telah ditentukan.

3. Metodologi

3.1 Metodologi Penelitian

Dalam metode penelitian ini akan menjelaskan langkah yang dilakukan untuk merancang sistem yang diajukan menggunakan pemodelan *waterfall* sebagai berikut :



Gambar 1. SDLC Waterfall

3.2 Metode Pengumpulan Data

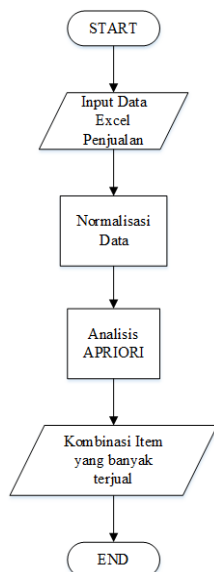
Pengumpulan data dilakukan sesuai dengan sumber dan jenis data yang diperlukan. Pengumpulan data kualitatif dalam penelitian ini dilakukan dengan melakukan wawancara dan observasi kepada pihak PT Gresik Cipta Sejahtera. Data yang diberikan adalah data penjualan bahan kimia dari tahun 2015 sampai 2017 yang berupa Excel.

Gambar 2. Data Excel Penjualan Bahan Kimia

1. Input data berupa data excel yang berisi data penjualan bahan kimia yang masih belum dinormalisasi.
2. Pada normalisasi data yang diterapkan yaitu data cleansing dan data transformation. Pada tahap cleansing data, dilakukan pemilihan data apa yang digunakan untuk tahap selanjutnya dan membuang data-data yang tidak digunakan. Pada tahap ini, dipilih data penjualan pada bagian bahan kimia saja dan membuang transaksi selain data bahan kimia. Setelah dilakukan cleansing data, dilakukan transformasi data. Pada transformasi data ini dilakukan normalisasi data sesuai dengan bentuk data yang diperlukan untuk melakukan perhitungan selanjutnya dan dilakukan pengubahan format excel menjadi csv agar nantinya data tersebut dapat dijalankan pada sistem.
3. Pada perhitungan manual ini, dilakukan perhitungan dengan menentukan minimum support dan minimum confidence. Semua item yang memiliki support di bawah minimum support dipangkas (pruned) sehingga item-item pada yang tersisa setelah dilakukan pemangkasan. Selanjutnya item-item yang tersisa tersebut dikombinasikan sebanyak dua item menjadi 2-itemset, dan kembali dihitung support dari item-item tersebut. Hal ini dilakukan sampai tidak ada kombinasi yang tersisa. Selanjutnya dapat dilakukan perhitungan confidence untuk mengetahui barang mana saja yang menjadi interesting rule.
4. Output dari sistem ini berupa hasil analisa yang menampilkan barang apa saja yang menjadi interesting rule atau bahan kimia apa saja yang sering dibeli secara bersamaan.

4. Implementasi dan Hasil

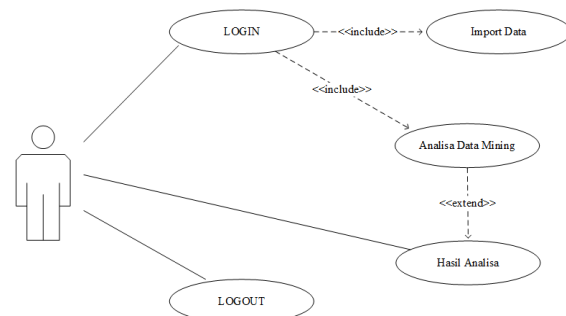
4.1 Deskripsi Sistem



Gambar 3. Flowchart Sistem

4.2 Perancangan Sistem

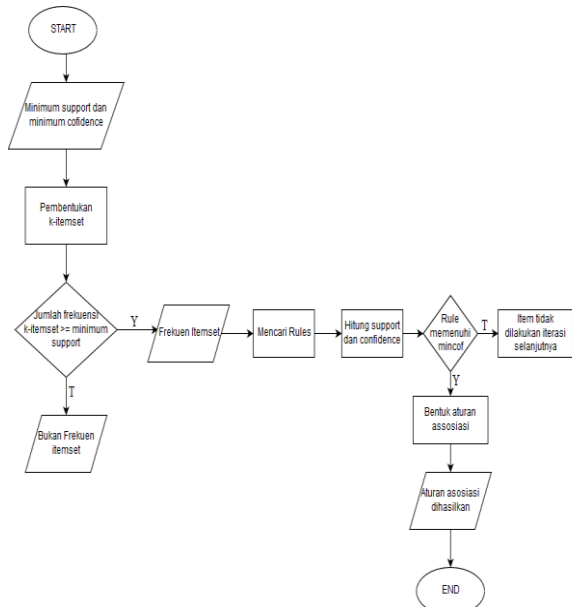
Untuk menangkap kebutuhan fungsional dari sistem yang bersangkutan, Penerapan *Association Rule* Dengan Algoritma Apriori Dalam Analisis *Data Mining* Penjualan Bahan Kimia Pada PT Gresik Cipta Sejahtera adalah sebagai berikut :



Gambar 4. Use Case Diagram

4.3 Analisa dan Hasil

Analisa dalam penyelesaian masalah untuk menganalisa pola data penjualan dengan menggunakan Algoritma Apriori.



Gambar 5. Flowchart Algoritma Apriori

Langkah-langkah dari proses algoritma apriori adalah :

1. Melakukan scan database untuk mendapat kandidat 1-itemset, yaitu C1 (himpunan item yang terdiri dari 1 item) dan menghitung nilai support-nya. Bandingkan nilai support dengan minimum support yang sudah ditentukan, jika nilainya lebih besar atau sama dengan minimum support, maka itemset tersebut termasuk dalam large itemset yaitu L1 (large itemset dengan 1 item).

C1 (Kandidat 1-itemset)				
No	Kategori1	Qty	Support	
1	AMONIA	55	94.50%	
2	DRY ICE	45	77.50%	
3	ASAM SULFAT	51	85.44%	
4	Oksigen	44	74.50%	
5	ASAM CHLORIDA	29	47.88%	
6	CARBONDIOKSIDA	31	52.54%	
7	CHLORINE	3	5.00%	
8	ACETYLENE	6	10.17%	
9	Nitrogen	1	1.69%	

Gambar 6. Kandidat 1 item

L1 (Large 1-itemset)				
No	Kategori1	Qty	Support	
1	AMONIA	55	94.50%	
2	DRY ICE	45	77.50%	
3	ASAM SULFAT	51	85.44%	
4	Oksigen	44	74.50%	
5	CARBONDIOKSIDA	31	52.54%	

Gambar 7. Large Itemset 1 item

2. Itemset yang tidak termasuk dalam large itemset tidak disertakan dalam iterasi selanjutnya (dilakukan pruning).

3. Himpunan L1 hasil iterasi pertama akan digunakan untuk iterasi selanjutnya. Pada L1 dilakukan proses join terhadap dirinya sendiri untuk membentuk kandidat 2-itemset (C2). Bandingkan lagi support dari item-item C2 dengan minimum support, bila tidak kurang dari minimum support, maka itemset tersebut masuk dalam large itemset L2. Pada iterasi selanjutnya, hasil large itemset pada iterasi sebelumnya (Lk-1) akan dilakukan proses join terhadap dirinya sendiri untuk membentuk kandidat baru (Ck), dan large itemset baru (Lk). Setelahnya dilakukan proses pruning pada itemset yang tidak termasuk dalam Lk.

C2 (Kandidat 2-itemset)				
No	Kategori1	Kategori2	Qty	Support
1	AMONIA	DRY ICE	43	72.88%
2	AMONIA	ASAM SULFAT	49	83.95%
3	AMONIA	Oksigen	43	72.88%
4	AMONIA	CARBONDIOKSIDA	29	49.15%
5	DRY ICE	ASAM SULFAT	39	66.1%
6	DRY ICE	Oksigen	33	55.93%
7	DRY ICE	CARBONDIOKSIDA	29	49.15%
8	ASAM SULFAT	Oksigen	41	69.49%
9	ASAM SULFAT	CARBONDIOKSIDA	31	52.54%
10	Oksigen	CARBONDIOKSIDA	25	42.37%

Gambar 8. Kombinasi 2 item

L2 (Large 2-itemset)				
No	Kategori1	Kategori2	Qty	Support
1	AMONIA	DRY ICE	43	72.88%
2	AMONIA	ASAM SULFAT	49	83.95%
3	AMONIA	Oksigen	43	72.88%
4	DRY ICE	ASAM SULFAT	39	66.1%
5	DRY ICE	Oksigen	33	55.93%
6	ASAM SULFAT	Oksigen	41	69.49%
7	ASAM SULFAT	CARBONDIOKSIDA	31	52.54%

Gambar 9. Large Itemset 2 item

4. Tahap pembentukan kandidat (joining) dan pembentukan large itemset (pruning) terus dilakukan hingga terdapat himpunan kosong atau sudah tidak ada lagi kandidat yang bisa dibentuk.

C4 (Kandidat 4-itemset)						
No	Kategori1	Kategori2	Kategori3	Kategori4	Qty	Support
1	AMONIA	DRY ICE	ASAM SULFAT	Oksigen	31	52.54%

Gambar 10. Kandidat 4 item

L4 (Large 4-itemset)						
No	Kategori1	Kategori2	Kategori3	Kategori4	Qty	Support
1	AMONIA	DRY ICE	ASAM SULFAT	Oksigen	31	52.54%

Gambar 11. Large Itemset 4 item

5. Dari seluruh large itemset yang memenuhi minimum support (frequent itemset) dibentuk association rule dan dicari nilai confidence-nya. Aturan-aturan yang nilai confidence nya lebih kecil dari minimum confidence, tidak termasuk dalam association rule yang dipakai.

Aturan Asosiasi 2-item				
No	Rule	Support	Confidence	
1	Jika DRY ICE maka AMONIA	72,88	4346	93,48%
2	Jika AMONIA maka DRY ICE	72,88	4356	76,79%
3	Jika ASAM SULFAT maka AMONIA	83,05	4951	96,98%
4	Jika AMONIA maka ASAM SULFAT	83,05	4956	87,5%
5	Jika Okalgen maka AMONIA	72,88	4344	97,73%
6	Jika AMONIA maka Okalgen	72,88	4356	76,79%
7	Jika ASAM SULFAT maka DRY ICE	66,1	3951	76,47%
8	Jika DRY ICE maka ASAM SULFAT	66,1	3946	84,78%
9	Jika Okalgen maka DRY ICE	55,53	3344	79%
10	Jika DRY ICE maka Okalgen	55,53	3346	71,74%
11	Jika Okalgen maka ASAM SULFAT	69,49	4144	93,16%
12	Jika ASAM SULFAT maka Okalgen	69,49	4151	89,39%
13	Jika CARBONDIKSIDA maka ASAM SULFAT	52,54	3131	100%
14	Jika ASAM SULFAT maka CARBONDIKSIDA	52,54	3151	89,78%

Gambar 12. Aturan Asosiasi 2 item

Aturan Asosiasi 3-item				
No	Rule	Support	Confidence	
1	Jika DRY ICE, ASAM SULFAT maka AMONIA	62,71	3739	94,87%
2	Jika AMONIA, ASAM SULFAT maka DRY ICE	62,71	3749	75,51%
3	Jika AMONIA, DRY ICE maka ASAM SULFAT	62,71	3743	86,95%
4	Jika DRY ICE, Okalgen maka AMONIA	54,24	3233	96,97%
5	Jika AMONIA, Okalgen maka DRY ICE	54,24	3243	74,42%
6	Jika AMONIA, DRY ICE maka Okalgen	54,24	3243	74,42%
7	Jika ASAM SULFAT, Okalgen maka AMONIA	69,49	4141	100%
8	Jika AMONIA, Okalgen maka ASAM SULFAT	69,49	4143	95,35%
9	Jika AMONIA, ASAM SULFAT maka Okalgen	69,49	4149	83,67%
10	Jika ASAM SULFAT, Okalgen maka DRY ICE	52,54	3141	75,61%
11	Jika DRY ICE, Okalgen maka ASAM SULFAT	52,54	3133	93,94%
12	Jika DRY ICE, ASAM SULFAT maka Okalgen	52,54	3139	79,49%

Gambar 13. Aturan Asosiasi 3 item

Aturan Asosiasi 4-item				
No	Rule	Support	Confidence	
1	Jika DRY ICE, ASAM SULFAT, Okalgen maka AMONIA	52,54	3131	100%
2	Jika AMONIA, ASAM SULFAT, Okalgen maka DRY ICE	52,54	3141	75,61%
3	Jika AMONIA, DRY ICE, Okalgen maka ASAM SULFAT	52,54	3152	96,88%
4	Jika AMONIA, DRY ICE, ASAM SULFAT maka Okalgen	52,54	3157	83,70%

Gambar 14. Aturan Asosiasi 4 item

5. Kesimpulan

Berdasarkan uraian yang telah penulis buat tentang implementasi algoritma apriori pada aplikasi data mining, maka penulis dapat menarik beberapa kesimpulan yang diperoleh sebagai berikut:

1. Sistem ini telah berhasil menerapkan Association Rules dengan Algoritma Apriori untuk melakukan analisis data penjualan bahan kimia pada PT Gresik Cipta Sejahtera.
2. Tingkat akurasi sistem yang dihasilkan sebesar 16% dengan tren turun. Hal ini disebabkan oleh banyaknya item yang memenuhi nilai minimum support pada tahun 2015 lebih banyak dari tahun 2016 dan 2017 sehingga aturan yang muncul dalam tahun 2016 dan 2017 sedikit.

3. Dengan menggunakan software ini, manajer dapat mengetahui item-item apa saja yang saling berasosiasi pada sebuah pasar dan berapa besar tingkat asosiasi tersebut.
4. Penentuan minimum support dan minimum confidence sangat mempengaruhi proses analisis, semakin kecil minimum support yang ditentukan maka akan semakin banyak kombinasi yang terbentuk, dan sebaliknya. Sedangkan semakin besar minimum confidence yang ditentukan maka akan semakin sedikit aturan yang diterima begitu pula sebaliknya.

Daftar Pustaka:

D. A. Nurdin, “Penerapan Data Mining Untuk Menganalisis Penjualan Barang Dengan Menggunakan Metode Apriori Pada Supermarket Sejahtera Lhokseumawe,” *Techsi*, pp. 133-155, 2015.

E. Norsyanah, “Penerapan Algoritma Eclat Dalam Menentukan Metode Kontrasepsi Yang Dipilih,” *UIN SUSKA RIAU, Riau*, 2016.

H. Santoso, I. P. Hariyadi and P. , “Data Mining Analisa Pola Pembelian Produk Dengan Menggunakan Metode Algoritma Apriori,” in *Seminar Nasional Teknologi Informasi dan Multimedia*, Yogyakarta, 2016.

Kusrini, *KONSEP DAN APLIKASI SISTEM PENDUKUNG KEPUTUSAN*, Andi Yogyakarta, 2007.

M. Fauzy, K. R. S. W and I. Asror, “Penerapan Metode Association Rule Menggunakan Algoritma Apriori Pada Simulasi Prediksi Hujan Wilayah Kota Bandung,” *JITTER*, pp. 221-227, 2016.

S. Kuswayati, “IMPLEMENTASI MARKET BASKET ANALYSIS MENGGUNAKAN ALGORITMA APRIORI PADA TRANSAKSI RITEL,” in *Seminar Nasional Teknologi Informasi, Bisnis, dan Desain 2017*, Bandung, 2017.

W. Fitri and R. Setiawan, “Clustering Karyawan Berdasarkan Kinerja dengan Menggunakan Logika Fuzzy C-Mean,” *Jurnal Sains, Teknologi & Industri UIN SUSKA RIAU*, vol. 10, no. 2, pp. 1-7, 2013.