

Data Mining Dengan Algoritma Apriori untuk Penentuan Aturan Asosiasi Pola Pembelian Pupuk

Amrin

Program Studi Teknik Komputer

AMIK Bina Sarana Informatika Jakarta

Jl. R.S Fatmawati no. 24 Pondok Labu, Jakarta Selatan, Indonesia

amrin.ain@bsi.ac.id

Abstract—In order to find out what fertilizer purchased by consumers, can be done with analytical techniques that is the analysis of consumer buying habits. Detection of fertilizers often purchased simultaneously is done using association rules. In this research will be used a priori algorithm for determining the rules of association of fertilizer purchases. From the results of the discussion and analysis of data can be concluded that with the application of a priori algorithm in determining the combination between itemsets with minimum support of 20% and minimum confidence 75% found 6 association rules, which has the highest value of support and confidence is if the consumer made a purchase transaction of fertilizer Organic and urea fertilizers simultaneously with the value of 60% support and 86% confidence value. Thus, if there are consumers buying organic fertilizers, then the possibility of consumers buying urea fertilizer is 86%.

Intisari—Agar dapat mengetahui pupuk apa saja yang dibeli oleh para konsumen, dapat dilakukan dengan teknik analisis yaitu analisis dari kebiasaan membeli konsumen. Pendeteksian mengenai pupuk yang sering dibeli secara bersamaan dilakukan dengan menggunakan aturan asosiasi. Pada penelitian ini akan digunakan algoritma apriori untuk penentuan aturan asosiasi pembelian pupuk. Dari hasil pembahasan dan analisis data yang dilakukan dapat disimpulkan bahwa dengan penerapan algoritma apriori dalam menentukan kombinasi antar itemset dengan *minimum support* 20% dan *minimum confidence* 75% ditemukan 6 aturan asosiasi, dimana yang memiliki nilai *support* dan *confidence* tertinggi adalah jika konsumen melakukan transaksi pembelian pupuk organik dan pupuk urea secara bersamaan dengan nilai *support* 60% dan nilai *confidence* 86%. Dengan demikian, jika terdapat konsumen membeli pupuk organik, maka kemungkinan konsumen tersebut membeli pupuk urea adalah 86%.

Kata kunci: data mining, algoritma apriori, aturan asosiasi, *support*, *confidence*.

I. PENDAHULUAN

Saat ini teknologi informasi berkembang begitu cepat sehingga kebutuhan terhadap informasi semakin meningkat. Informasi tidak akan bernilai apabila tidak dikelola dengan baik. Akan tetapi jika data yang tersedia tersebut berjumlah besar maka cara konvensional tidak lagi mampu untuk menganalisa data yang ada. Maka dari itu dibutuhkan metode yang dapat menganalisis, meringkas dan mengekstrak data untuk menjadi sebuah informasi yang berguna. Tidak hanya mengandalkan

data yang ada saja, perlu diadakannya analisis data untuk menggali potensi-potensi yang ada.

Ketersediaan informasi transaksi pelanggan mendorong pengembangan teknik yang secara otomatis mencari hubungan antara item data pada database, seperti halnya dalam transaksi penjualan pupuk. Database penjualan pupuk menyimpan jumlah record transaksi penjualan setiap hari yang kalau dikumpulkan jumlahnya sangatlah besar. Setiap record memberikan daftar item barang yang dibeli oleh pelanggan dalam satu transaksi. Bayangkan jumlah transaksi yang sebegitu besarnya kalau tidak dimanfaatkan untuk menggali informasi yang tersembunyi, hanyalah akan menjadi tumpukan sampah semata. Padahal pemilik atau manajemen akan sangat tertarik untuk mengetahui jika beberapa kelompok item barang secara konsisten dibeli secara bersamaan. Manajemen dapat menggunakan data tersebut dalam pengaturan layout untuk meletakkan item barang secara optimal dengan keterkaitan satu dengan lainnya, dapat pula digunakan dalam promosi, atau dalam design katalog dan untuk mengidentifikasi segmen pelanggan berdasar pola pembelian.

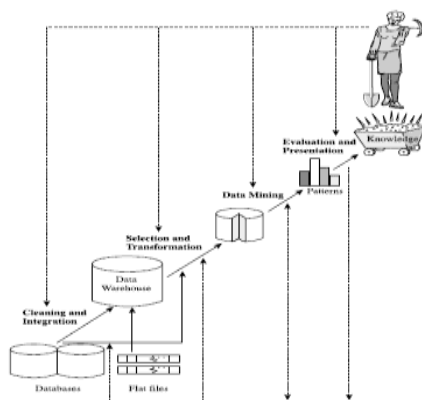
Banyaknya persaingan di dunia bisnis, khususnya dalam industri penjualan pupuk, menuntut manajemen untuk menemukan suatu strategi yang dapat meningkatkan penjualan. Agar dapat mengetahui pupuk apa saja yang dibeli oleh para konsumen, dapat dilakukan dengan teknik analisis yaitu analisis dari kebiasaan membeli konsumen. Pendeteksian mengenai pupuk yang sering dibeli secara bersamaan dilakukan dengan menggunakan association rule (aturan asosiasi), yang mana proses pencarian asosiasi atau hubungan antar item data ini diambil dari data transaksi penjualan pupuk harian di CV. Tani Monta Baru. Pada penelitian ini penulis menggunakan algoritma apriori, yang berfungsi untuk membentuk kandidat kombinasi item yang mungkin, lalu diuji apakah kombinasi tersebut memenuhi parameter minimum support dan minimum confidence yang merupakan nilai ambang yang diberikan oleh pengguna.

II. KAJIAN LITERATUR

A. Data Mining

Data mining adalah rangkaian proses untuk menggali nilai tambah berupa informasi yang belum terekplorasi dari sebuah basis data, melakukan eksplorasi dengan cara-cara tertentu untuk memanipulasi data menjadi informasi yang lebih berharga dengan cara mengekstraksi dan mengenali pola penting dari basis data (Han & Kamber). Menurut Daryl Pregibons dalam (Gorunescu) "*Data mining* adalah perpaduan dari ilmu statistik, kecerdasan buatan, dan penelitian bidang *database*". Nama *data mining* berasal dari kemiripan antara pencarian informasi yang bernilai dari *database* yang besar dengan menambang sebuah gunung untuk sesuatu yang bernilai (Sumathi). Keduanya memerlukan penyangiran melalui sejumlah besar material, atau menyelidiki dengan cerdas untuk mencari keberadaan sesuatu yang disebut bernilai tadi.

Data Mining merupakan teknologi baru yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka. Beberapa aplikasi data mining fokus pada prediksi, mereka meramalkan apa yang akan terjadi dalam situasi baru dari data yang menggambarkan apa yang terjadi di masa lalu (Witten, Frank, & Hall). Data mining sering disebut juga Knowledge Discovery in Database atau disingkat menjadi KDD, adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santoso). Gambar tahapan pembuatan aplikasi data mining ditunjukkan pada gambar 1 berikut ini:



Sumber: Han & Kamber

Gambar 1. Tahapan Proses KDD

Gambar 1 menunjukkan langkah dalam proses *data mining*. Proses dalam tahap *data mining* terdiri dari tiga langkah utama, yaitu (Sogala):

1. Data Preparation

Pada langkah ini, data dipilih, dibersihkan, dan dilakukan *preprocessed* mengikuti pedoman dan *knowledge* dari ahli domain yang menangkap dan

mengintegrasikan data internal dan eksternal ke dalam tinjauan organisasi secara menyeluruh.

2. Algoritma *data mining*

Penggunaan algoritma *data mining* dilakukan pada langkah ini untuk menggali data yang terintegrasi untuk memudahkan identifikasi informasi bernilai.

3. Fase analisa data

Keluaran dari data mining dievaluasi untuk melihat apakah *knowledge* domain ditemukan dalam bentuk *rule* yang telah diekstrak dari jaringan.

B. Association Rules Mining

Aturan asosiasi (*association rule*) adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi data mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Aturan asosiasi (*association rules*) atau analisis afinitas (*affinity analysis*) berkenaan dengan studi tentang 'apa bersama apa'. Ini bisa berupa studi transaksi di supermarket, misalnya seseorang yang membeli susu bayi juga membeli sabun mandi. Di sini berarti susu bayi bersama dengan sabun mandi. Karena awalnya berasal dari studi tentang database transaksi konsumen untuk menentukan kebiasaan suatu produk dibeli bersama produk apa, maka aturan asosiasi juga sering dinamakan market basket analysis (Kusrini, & Emha Taufiq Luthfi).

Dalam menentukan suatu aturan asosiasi, terdapat suatu ukuran ketertarikan (*interestingness measure*) yang didapatkan dari hasil pengolahan data dengan data perhitungan tertentu. Pada umumnya terdapat dua ukuran ketertarikan dalam aturan asosiasi, yaitu :

- Support*, adalah probabilitas konsumen membeli beberapa produk secara bersamaan dari jumlah seluruh transaksi (Yulita dan Moertini). Ukuran ini menentukan apakah suatu item/itemset layak untuk dicari nilai *confidence*-nya (misal dari keseluruhan transaksi yang ada, seberapa besar tingkat dominasi yang menunjukkan bahwa item X dan Y dibeli secara bersamaan).
- Confidence* atau tingkat kepercayaan merupakan probabilitas kejadian beberapa produk yang dibeli bersamaan dimana salah satu produk sudah pasti dibeli (misal, seberapa sering item Y dibeli apabila konsumen membeli item X) (Yulita dan Moertini).

Kedua ukuran (*support* dan *confidence*) berguna dalam menentukan aturan asosiasi, yaitu untuk dibandingkan dengan batasan (*threshold*) yang ditentukan oleh pengguna. Batasan tersebut umumnya terdiri atas minimum *support* dan minimum *confidence*. Langkah-langkah dalam pembentukan aturan asosiasi meliputi dua tahap, yaitu:

a. Analisis Pola Frekuensi Tinggi

Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai *support* dalam basis data. *Support* untuk aturan " $X \Rightarrow Y$ " adalah probabilitas atribut atau kumpulan atribut X dan Y yang terjadi secara bersamaan dalam suatu transaksi (Yulita dan

Moertini). Bentuk persamaan matematis dari nilai *support* (Han & Kamber) adalah:

$$\text{Support}(X \Rightarrow Y) = P(X \cap Y)$$

Dengan keterangan:

$X \Rightarrow Y$ = item yang muncul bersamaan

$P(X \cap Y)$ = probabilitas transaksi yang mengandung X dan Y dibagi dengan jumlah transaksi seluruhnya.

b. Pembentukan aturan asosiasi

Setelah semua pola frekuensi tinggi ditemukan, kemudian dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* dari aturan *if X then Y*. Bentuk rumus matematika dari *confidence* (Han & Kamber) adalah:

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

Dengan keterangan:

$P(X|Y)$ = item yang muncul bersamaan

$X \cap Y$ = probabilitas transaksi yang mengandung X dan Y dibagi dengan jumlah transaksi yang mengandung X.

C. Algoritma Apriori

Algoritma apriori adalah jenis aturan asosiasi pada data mining. Algoritma ini ditujukan untuk mencari kombinasi itemset yang mempunyai suatu nilai keseringan tertentu sesuai kriteria atau filter yang diinginkan. Algoritma ini diajukan oleh R. Agrawal dan R. Srikant tahun 1994 (Kusrini, & Emha Taufiq Luthfi).

Hasil dari algoritma ini dapat digunakan untuk membantu dalam pengambilan keputusan pihak manajemen. Algoritma apriori melakukan pendekatan iteratif yang dikenal dengan pencarian level-wise, dimana k-itemset digunakan untuk mengeksplorasi atau menemukan (k+1)-itemset. Oleh karena itu, algoritma apriori dibagi menjadi beberapa tahap yang disebut iterasi. Tiap iterasi menghasilkan pola frekuensi tinggi dengan panjang yang sama dimulai dari iterasi pertama yang menghasilkan pola frekuensi tinggi dengan panjang satu.

Di iterasi pertama ini, *support* dari setiap item dihitung dengan men-scandatabase. Setelah *support* dari setiap item didapat, item yang memiliki *support* diatas *minimum support* dipilih sebagai pola frekuensi tinggi dengan panjang 1 atau sering disebut Large 1-itemset atau disingkat L1. Iterasi kedua menghasilkan 2-itemset yang tiap set-nya memiliki dua item. Pertama dibuat kandidat 2-itemset atau disingkat C2 dari kombinasi semua 1-itemset. Lalu untuk tiap kandidat 2-itemset ini dihitung *support*-nya dengan men-scandatabase. *Support* disini artinya jumlah transaksi dalam database yang mengandung kedua item dalam C2. Setelah *support* dari semua C2 didapatkan, C2 yang memenuhi syarat *minimum support* dapat ditetapkan sebagai 2-itemset yang juga merupakan pola frekuensi tinggi dengan panjang 2 atau Large 2-itemset (L2).

Untuk selanjutnya pada iterasi ke-k dapat dibagi lagi menjadi beberapa bagian :

1. Pembentukan kandidat itemset, Kandidat k-itemset (C_k) dibentuk dari kombinasi (k-1)-itemset yang didapat dari iterasi sebelumnya. Hal ini disebut juga dengan proses join. Setelah proses *join* dilakukan, selanjutnya proses *prune* yang bertujuan untuk menghasilkan Lk. Proses *prune* merupakan proses pemangkasan kandidat k-itemset yang subset-nya yang berisi (k-1)-item yang tidak termasuk dalam pola frekuensi tinggi dengan panjang k-1.
2. Tetapkan pola frekuensi tinggi. Pola frekuensi tinggi yang memuat k-item atau k-itemset yang ditetapkan dari kandidat k-itemset yang *support*-nya lebih besar dari *minimum support*. Bila tidak didapat pola frekuensi tinggi baru maka seluruh proses dihentikan. Bila tidak, maka k ditambah satu dan kembali ke bagian 1.

III. METODE PENELITIAN

1. Obyek Penelitian

Obyek penelitian adalah transaksi penjualan pupuk pada CV. Tani Monta Baru yang berlokasi di kelurahan Monta Baru, kecamatan Woja, Kabupaten Dompu.

2. Sumber Data

a. Data Primer

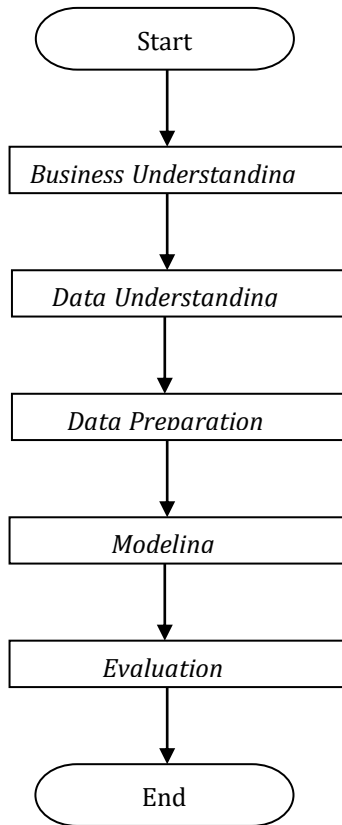
Data primer adalah data yang diperoleh dari sumber-sumber asli. Dalam hal ini data primer diperoleh dari bagian penjualan CV. Tani Monta Baru.

b. Data Sekunder

Data sekunder adalah data yang diperoleh dari berbagai sumber selain CV. Tani Monta Baru, seperti buku dipergustakaan, publikasi-publikasi ilmiah, peraturan-peraturan pemerintah, majalah dan lain sebagainya yang berhubungan dengan masalah yang sedang diteliti.

3. Tahap Penelitian

Terdapat beberapa tahap dalam pengolahan data eksperimen, pada penelitian ini menggunakan model Cross-Standard Industry for Data Mining (CRISP-DM), yaitu seperti terlihat pada diagram di bawah ini: (Sumathi)



Gambar 2. Diagram Tahap Penelitian
Sumber: Sumathi

IV. HASIL DAN PEMBAHASAN

Data transaksi penjualan pupuk diperlihatkan oleh tabel di bawah ini:

Tabel 1. Transaksi Pupuk

Transaksi	Item Produk
1	12,09,01,04,02
2	01,03,04,05,10
3	03,06,02,11,01
4	01,04,09,08,06,12
5	03,06,12,04,11
6	03,07,08,12,09
7	03,06,04,11,05
8	01,02,11,06,04
9	04,01,10,11,05,03
10	01,04,06,10,07,08

Sumber : CV. Tani Monta Baru

Pada penelitian ini, penulis menggunakan algoritma apriori untuk menemukan aturan asosiasi pola pembelian pupuk dengan *minimum support* sebesar 40% dan *minimum confidence* sebesar 75%. Adapun urutan langkah-langkahnya sebagai berikut:

Iterasi 1:

Menghitung support dan menentukan itemset yang memenuhi minimum support untuk 1 itemset (k=1)

Langkah 1: Menghitung support untuk 1 itemset. Hasilnya seperti diperlihatkan oleh table 2 di bawah ini:

Tabel 2. C1

Itemset	Support
P01	70%
P02	30%
P03	60%
P04	80%
P05	30%
P06	60%
P07	20%
P08	30%
P09	30%
P10	20%
P11	50%
P12	40%

Sumber : Hasil Penelitian (2017)

Langkah 2: Menentukan itemset yang memenuhi *minimum support* 40%, maka item yang tidak memenuhi dieliminasi atau dipangkas, sehingga menghasilkan tabel 3 di bawah ini:

Tabel 3. L1

Itemset	Support
P01	70%
P03	60%
P04	80%
P06	60%
P11	50%
P12	40%

Sumber : Hasil Penelitian (2017)

Iterasi 2:

Menghitung support dan menentukan itemset yang memenuhi minimum support untuk 2 itemset (k=2)

Langkah 1: Menghitung support untuk 2 itemset. Hasilnya seperti diperlihatkan oleh tabel 4 di bawah ini:

Tabel 4. C2

Itemset	Support
P01,P03	30%
P01,P04	60%
P01,P06	40%
P01,P11	30%
P01,P12	20%
P03,P04	40%
P03,P06	30%
P03,P11	40%
P03,P12	20%
P04,P06	50%
P04,P11	40%
P04,P12	20%
P06,P11	40%
P06,P12	20%
P11,P12	10%

Sumber : Hasil Penelitian (2017)

Langkah 2: Menentukan itemset yang memenuhi *minimum support* 40%, maka item yang tidak memenuhi dieliminasi atau dipangkas, sehingga menghasilkan tabel 5 di bawah ini:

Tabel 5. L2

Itemset	Support
P01,P04	60%
P01,P06	40%
P03,P04	40%
P03,P11	40%
P04,P06	50%
P04,P11	40%
P06,P11	40%

Sumber : Hasil Penelitian (2017)

Iterasi 3:

Menghitung support dan menentukan itemset yang memenuhi *minimum support* untuk 3 itemset ($k=3$)

Langkah 1: Menghitung support untuk 3 itemset. Hasilnya seperti diperlihatkan oleh tabel 6 di bawah ini:

Tabel 6. C3

Itemset	Support
P01,P04,P06	30%
P01,P04,P03	20%
P01,P04,P11	20%
P01,P04,P12	20%
P01,P06,P03	10%

P01,P06,P11	10%
P01,P06,P12	10%
P01,P03,P11	10%
P01,P03,P12	0%
P03,P11,P04	30%
P03,P11,P06	30%
P04,P06,P11	30%

Sumber : Hasil Penelitian (2017)

Langkah 2: Menentukan itemset yang memenuhi *minimum support* 40%. Berdasarkan tabel 5 di atas tidak ada itemset yang memenuhi *minimum support*. Sehingga nilai $L3 = \{ \}$

Iterasi 3: Stop

Untuk mencari aturan asosiasi hanya menggunakan $L2$ dengan menetapkan *minimum confidence* sebesar 75%. Diperoleh kombinasi aturan asosiasi pada tabel 7 di bawah ini:

Tabel 7. Kombinasi Aturan Asosiasi

(X \rightarrow Y)	Support	Confidence
P01 \rightarrow P03	30%	43%
P03 \rightarrow P01	30%	50%
P01 \rightarrow P04	60%	86%
P04 \rightarrow P01	60%	75%
P01 \rightarrow P06	40%	57%
P06 \rightarrow P01	40%	67%
P01 \rightarrow P11	30%	43%
P11 \rightarrow P01	30%	60%
P01 \rightarrow P12	20%	29%
P12 \rightarrow P01	20%	50%
P03 \rightarrow P04	40%	67%
P04 \rightarrow P03	40%	50%
P03 \rightarrow P06	30%	50%
P06 \rightarrow P03	30%	50%
P03 \rightarrow P11	40%	67%
P11 \rightarrow P03	40%	80%
P03 \rightarrow P12	20%	33%
P12 \rightarrow P03	20%	50%
P04 \rightarrow P06	50%	63%
P06 \rightarrow P04	50%	83%
P04 \rightarrow P11	40%	50%
P11 \rightarrow P04	40%	80%
P04 \rightarrow P12	20%	25%
P12 \rightarrow P04	20%	50%
P06 \rightarrow P11	40%	67%
P11 \rightarrow P06	40%	80%
P11 \rightarrow P12	10%	20%
P12 \rightarrow P11	10%	25%

Sumber : Hasil Penelitian (2017)

Dari tabel 7 di atas yang memenuhi *minimum confidence* 75%, diperlihatkan pada tabel 8 di bawah ini:

Tabel 8. Hasil Aturan Asosiasi

$(X \rightarrow Y)$	Support	Confidence
P01→P04	60%	86%
P04→P01	60%	75%
P11→P03	40%	80%
P06→P04	50%	83%
P11→P04	40%	80%
P11→P06	40%	80%

Sumber : Hasil Penelitian (2017)

KESIMPULAN

Kesimpulan yang dapat diambil berdasarkan penelitian ini adalah:

1. Nilai *support* dan *confidence* yang ditentukan akan mempengaruhi akurasi dalam pembentukan rule, semakin tinggi nilai *support* dan *confidence* maka rule akan semakin akurat.
2. Dengan penerapan algoritma apriori dalam menentukan kombinasi antar itemset dengan *minimum support* 40% dan *miimum confidence* 75% ditemukan 6 aturan asosiasi.
3. Berdasarkan hasil aturan asosiasi yang diperoleh, maka yang memiliki nilai *support* dan *confidence* tertinggi adalah P01 (Pupuk Organik) → P04 (Pupuk Urea) dengan nilai *support* 60% dan nilai *confidence* 86%.

BIODATA PENULIS



Amrin, S.Si, M.Kom. Dompu 10 Agustus 1980. Tahun 2003 lulus dari Program Strata Satu (S1) Jurusan Matematika Universitas Diponegoro Semarang. Tahun 2014 lulus dari Program Strata Dua (S2) Jurusan Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta. Pekerjaan saat ini sebagai Dosen AMIK BSI Jakarta sejak tahun 2007. Telah menulis beberapa paper di beberapa jurnal diantaranya Jurnal TECHNO STMIK Nusa Mandiri, Jurnal Teknik Komputer AMIK BSI Jakarta dan Jurnal PARADIGMA AMIK BSI Jakarta.

REFERENSI

- Gorunescu, Florin. Data Mining: Concepts, Models, and Techniques. Verlag Berlin Heidelberg: Springer, 2011.
- Han, J., & Kamber, M. Data Mining Concept and Tehniques. San Fransisco: Morgan Kauffman, 2006.
- Kusrini, & Luthfi, E. T. Algoritma Data Mining. Yogyakarta: Andi Publishing, 2009.
- Santosa, Budi. Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu, 2007.
- Sogala, Satchidananda S. Comparing the Efficacy of the Decision Trees with Logistic Regression for Credit Risk Analysis. India, 2006.
- Sumathi, & S., Sivanandam, S.N. Introduction to Data Mining and its Applications, 2006.
- Witten, I. H., Frank, E., & Hall, M. A. Data Mining: Practical Machine Learning and Tools. Burlington: Morgan Kaufmann Publisher, 2011.
- Yulita, Marsela dan Veronica S. Moertini. Analisis Keranjang Pasar dengan Algotitma Hash-Based pada transaksi penjualan di Apotik. Bandung: Jurnal Integral Majalah Ilmiah Matematika dan Ilmu Pengetahuan Alam, Vol 9, No 3, Jurusan Ilmu Komputer Universitas Katolik Parahyangan, 2004.