
PENGUJIAN ALGORITMA APRIORI DENGAN APLIKASI WEKA DALAM PEMBENTUKAN ASOSIATION RULE

Paska Marto Hasugian

Teknik Informatika
STMIK Pelita Nusantara Jl Iskandar Muda no 1 Medan, Sumatera Utara, Indonesia 20154

paskamarto86@gmail.com

ABSTRAK

Market Basket Analysis adalah Analisis dari kebiasaan membeli customer dengan mencari asosiasi dan korelasi antara item-item berbeda yang diletakkan customer dalam keranjang belanjanya. Fungsi ini paling banyak digunakan untuk menganalisa data dalam rangka keperluan strategi pemasaran, desain katalog, dan proses pembuatan keputusan bisnis. Tipe association rule bisa dinyatakan dengan persentasi terhadap banyaknya aktifitas yang terjadi dalam kurun waktu tertentu, dalam penelitian dilakukan pengujian terhadap algoritma dalam pembentukan rule dengan menggunakan aplikasi weka, data yang digunakan dengan inisial.

Kata Kunci : DM, Apriori, Rule, Weka

I. PENDAHULUAN

Data mining merupakan serangkaian proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Dalam data mining terdapat banyak teknik dalam pengerjaannya, diantaranya yaitu algoritma naïve bayes, decision tree, jaringan saraf tiruan dan masih banyak lainnya.

Analisis asosiasi (association analysis) berguna untuk mengungkap hubungan yang menarik yang tersembunyi dalam dataset besar. Hubungan yang terungkap tersebut dapat direpresentasikan dalam bentuk association rules) atau himpunan item yang sering muncul (sets of frequent items). Association Rules seringkali disebut dengan "market basket analysis", yang digunakan untuk menemukan relasi atau korelasi diantara himpunan item.

Dalam perancangan data mining informasi yang dihasilkan akan digunakan dalam membantu aktifitas sehari – hari sehingga data- data lama dapat dimanfaatkan dengan informasi yang dihasilkan. Proses ekstraksi secara manual sangatlah menghabiskan waktu dikarenakan data yang cukup banyak sehingga para peneliti dan pengguna prinsip mining tidak dapat menyelesaikan dengan cepat dan akurat.

Untuk membantu proses mining maka dilakukan suatu pengujian dengan aplikasi weka dengan tata cara penguraian data dalam bentuk excel dan memilih proses upload untuk input data awal. Setelah

proses selesai maka dilakukan proses pembentukan mining dan terbentuk rule terbaik berdasarkan data hasil uji coba.

II. TEORI

Analisis asosiasi (association analysis) berguna untuk mengungkap hubungan yang menarik yang tersembunyi dalam dataset besar. Hubungan yang terungkap tersebut dapat direpresentasikan dalam bentuk association rules) atau himpunan item yang sering muncul (sets of frequent items). Association Rules seringkali disebut dengan "market basket analysis", yang digunakan untuk menemukan relasi atau korelasi diantara himpunan item. Market Basket Analysis adalah Analisis dari kebiasaan membeli customer dengan mencari asosiasi dan korelasi antara item-item berbeda yang diletakkan customer dalam keranjang belanjanya. Fungsi ini paling banyak digunakan untuk menganalisa data dalam rangka keperluan strategi pemasaran, desain katalog, dan proses pembuatan keputusan bisnis. Tipe association rule bisa dinyatakan dengan persentasi terhadap banyaknya aktifitas yang terjadi dalam kurun waktu tertentu misalnya "70% dari pengunjung yang membeli mie, juice dan saus akan membeli juga roti tawar". Aturan asosiasi mengcapture item atau kejadian dalam data berukuran besar yang berisi data transaksi. Dengan kemajuan teknologi, data penjualan dapat disimpan dalam jumlah besar yang disebut dengan "basket data." Aturan asosiasi yang didefinisikan pada basket data, digunakan untuk keperluan promosi, desain katalog, segmentasi

customer dan target pemasaran. Secara tradisional, aturan asosiasi digunakan untuk menemukan trend bisnis dengan menganalisa transaksi customer,

Algoritma apriori adalah algoritma paling terkenal untuk menemukan pola frekuensi tinggi. Pola frekuensi tinggi adalah pola-pola item di dalam suatu *database* yang memiliki frekuensi atau *support* di atas ambang batas tertentu yang disebut dengan istilah minimum *support*. Algoritma apriori dibagi menjadi beberapa tahap yang disebut iterasi atau pass yaitu:

1. Pembentukan kandidat *itemset*, kandidat *k-itemset* dibentuk dari kombinasi (k-1)-*itemset* yang didapat dari iterasi sebelumnya. Satu cara dari algoritma apriori adalah adanya pemangkasan kandidat *k-itemset* yang subset-nya yang berisi k-1 item tidak termasuk dalam pola frekuensi tinggi dengan panjang k-1.
2. Penghitungan *support* dari tiap kandidat *k-itemset*. *Support* dari tiap kandidat *k-itemset* didapat dengan *menscan* database untuk menghitung jumlah transaksi yang memuat semua item di dalam kandidat *k-itemset* tersebut. Ini adalah juga ciri dari algoritma apriori dimana diperlukan penghitungan dengan *scan* seluruh *database* sebanyak *k-itemset* terpanjang.
3. Tetapkan pola frekuensi tinggi. Pola frekuensi tinggi yang memuat k item atau *k-itemset* ditetapkan dari kandidat *k-itemset* yang *support*-nya lebih besar dari minimum *support*.

4. Bila tidak didapat pola frekuensi tinggi baru maka seluruh proses dihentikan. Bila tidak, maka k ditambah satu dan kemabali ke bagian 1.

A. Analisis Pola Frekuensi Tinggi dengan Algoritma Apriori

Tahap inimencari kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam database. Nilai *support* sebuah *item* diperoleh dengan menggunakan rumus berikut:

$$Support(A) = \frac{\text{Jumlah transaksi mengandung A} \cdot (1)}{\text{Transaksi}}$$

nilai *support* dari 2 *item* diperoleh dengan menggunakan rumus:

$$Support(A, B) = P(A \cap B)$$

$$Support(A, B) = \frac{\sum \text{Transaksi mengandung A dan B}}{\sum \text{Transaksi}}$$

Frequent itemset menunjukkan *itemset* yang memiliki frekuensi kemunculan lebih dari nilai minimum yang ditentukan (\emptyset). Misalkan $\emptyset = 2$, maka semua *itemsets* yang frekuensinya kemunculannya lebih dari atau sama dengan 2 kali disebut *frequent*. Himpunan dari *frequent k-itemset* dilambangkan dengan F_k .

III. PEMBAHASAN DAN PENGUJIAN

3.1. Pembahasan

Sebelum melakukan pengujian maka data yang digunakan dalam pengujian disajikan pada tabel 1 berikut ini.

TABEL 1 Data Awal

No	Nama	1	2	3	4	5	6	7	8	9	10	11	12
1	A	147	128	86	146	142	148	134	165	76	148	165	78
2	B	84	92	127	83	127	82	90	140	82	120	143	124
3	C	123	124	74	124	92	70	127	82	125	90	125	84
4	D	99	56	90	50	41	95	63	54	93	123	82	89
5	E	92	136	148	120	140	137	124	160	80	97	154	80
6	F	54	40	43	46	34	44	69	49	37	38	52	32
7	G	67	65	52	67	62	65	71	65	59	57	87	50
8	H	86	82	76	89	78	76	84	76	74	74	120	73
9	I	42	42	36	38	30	35	52	43	40	40	46	42
10	J	74	74	65	74	74	72	75	70	70	68	92	68

A. Membentuk Pola Transaksi

Berdasarkan transaksi penjualan yang telah disajikan pada tabel diatas, transaksi tersebut dapat diakumulasikan. Akumulasi transaksi penjualan suku

cadang penjualan bulanan yang diambil dari 3 teratas laporan bulanan, dapat dilihat dalam tabel dibawah ini untuk lebih memahami proses penentuan pola dengan cara melakukan kombinasi untuk setiap

transaksi yang ada dari setiap aktifitas setiap bulannya. Misalnya pola pertama dipilih A,C, D dikarenakan penjualan tertinggi pada bulan pertama adalah A dengan banyak 147, C dengan penjualan 123 dan D dengan penjualan 99, dengan cara yang sama maka dihasilkan tabel pola berikut ini.

TABEL II. Pembentukan Pola

No	1	2	3	4	5	6	7	8	9	10	11	12
Item	A,C,D	E,A,C	E,B,D	A,C,E	A,E,B	A,E,D	A,C,E	A,E,B	C,D,B	A,D,B	A,E,B	B,D,C

B. Penerapan Apriori

Setelah melengkapai proses penentuan data dan membentuk pola transaksi penjualan maka dilanjutkan dengan langkah – langkah mining secara apriori.

1. Representasi Data Transaksi

Reperentasi data transaksi dibuat berdsarkan data transaksi yang terdapat pada tabel, dan p roses terbentuk berdasarkan yang telah dihasilkan dari pola transaksi penjualan. Didalam pola transaksi no item 1 dengan pemjualan A,C dan D dan seterusnya sehingga terbentuk tabel dibawah ini.

TABEL III. Representasi Data Transaksi

Transaksi	Item
1	A
1	C
1	D
2	E
2	A
2	C
3	E
3	B
3	D
4	A
4	C
4	E
5	A
5	E
5	B
6	A
6	E
6	D
7	A
7	C
7	E
8	A
8	E
8	B
9	C
9	D
9	B
10	A

Transaksi	Item
10	D
10	B
11	A
11	E
11	B
12	B
12	D
12	C

2. Pembuatan Format Tabular

Langkah kedua setelah terbentuk representasi data transaksi maka akan dibentuk pembuatan format tabular sesuai teori yang sudah dijelaskan diatas.

TABEL IV Format Tabular

Transaksi	A	B	C	D	E
1	1	1	1	0	0
2	1	1	0	0	1
3	0	0	1	1	1
4	1	1	0	0	1
5	1	0	0	1	1
6	1	0	1	0	1
7	1	1	0	0	1
8	1	0	0	1	1
9	0	1	1	1	0
10	1	0	1	1	0
11	1	0	0	1	1
12	0	1	1	1	0

3. Analisa Pola Frekuensi Tinggi

Tahap ini mencari kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam database.

a. Pembentukan Itemset

Berikut ini adalah penyelesaian dengan kasus berdasarkan data yang sudah disediakan pada tabel V Proses pembentukan C1 atau disebut dengan 1 *itemset* dengan jumlah minimum *support* = 30%, dengan rumus sebagai berikut:

$$Support(A) = \frac{\sum \text{transaksimengandungA}}{\sum \text{transaksi}} * 100\%$$

Berikut merupakan perhitungan pembentukan 1 itemset:

$$S(A) = \frac{\sum \text{transaksi yang mengandung A}}{\sum 12} = \frac{9}{12} *$$

$$100\% = 75\%$$

$$S(C) = \frac{\sum \text{transaksi yang mengandung C}}{\sum 12} = \frac{6}{12} *$$

$$100\% = 50\%$$

$$S(D) = \frac{\sum \text{transaksi yang mengandung D}}{\sum 12} = \frac{6}{12} *$$

$$100\% = 50\%$$

$$S(B) = \frac{\sum \text{transaksi yang mengandung B}}{\sum 12} = \frac{7}{12} * 100\% = 58,33\%$$

$$S(E) = \frac{\sum \text{transaksi yang mengandung E}}{\sum 12} = \frac{8}{12} * 100\% = 66,67\%$$

Dengan cara yang sama maka menghasilkan tabel V berikut ini

TABEL V. *Support* dari tiap item

Itemset	Support
A	75%
C	50%
D	50%
B	58,33%
E	66,67%

b. Kombinasi 2 Itemset

Proses Pembentukan C2 atau disebut dengan 2 *itemset* dengan jumlah minimum *support* = 30%, dapat diselesaikan dengan perhitungan berikut perhitungan pembentukan C2 atau 2 itemset:

$$S(A, C) = \frac{\sum \text{transaksi yang mengandung A dan C}}{\sum 12} * 100\%$$

$$= \frac{4}{12} * 100\% = 33,33\%$$

$$S(A, D) = \frac{\sum \text{transaksi yang mengandung A dan D}}{\sum 12} * 100\%$$

$$= \frac{3}{12} * 100\% = 25\%$$

$$S(A, B) = \frac{\sum \text{transaksi yang mengandung A dan B}}{\sum 12} * 100\%$$

$$= \frac{4}{12} * 100\% = 33,33\%$$

Dengan cara yang sama akan menghasilkan tabel berikut ini

TABEL VI. Calon 2 *itemset*

Itemset	Jumlah	Support
A, C	4	33,33%
A, D	3	25%
A, B	4	33,33%
A, E	7	58,33%
C, E	3	25%
C, B	2	16,67%
C, D	3	25%
D, B	4	33,33%
D, E	2	16,67%
E, B	3	25%

Minimal *support* yang ditentukan adalah 30% jadi kombinasi 2 *itemset* yang tidak memenuhi minimal *support* akan dihilangkan, terlihat seperti tabel dibawah ini:

TABEL VII. Minimal *Support* 2 *itemset* 30%

Itemset	Jumlah	Support
A, C	4	33,33%
A, B	4	33,33%

A, E	7	58,33%
D, B	4	33,33%

c. Kombinasi 3 Itemset

Proses pembentukan C3 atau disebut dengan 3 *itemset* dengan jumlah minimum *support* = 30%, dapat diselesaikan dengan rumus berikut:

$$Support(A, B) = P(A \cap B \cap C)$$

$$Support(A, B \text{ dan } C) = \frac{\sum \text{transaksi yang mengandung A, B dan C}}{\sum \text{transaksi}} * 100\%$$

$$C) = \frac{\sum \text{transaksi yang mengandung A, E dan D}}{\sum 12} * 100\%$$

$$S(A, E, D) = \frac{\sum \text{transaksi yang mengandung A, E dan D}}{\sum 12} * 100\%$$

Dengan cara yang sama maka akan menghasilkan tabel berikut ini

TABEL VIII. Kombinasi 3 *itemset*

Itemset	Jumlah	Support
A, E, D	1	8,33%
A, E, C	3	25%
A, B, E	3	25%
C, A, D	1	8,33%
C, D, B	2	16,67%
C, D, E	0	0%
D, B, E	1	8,33%

Karena kombinasi 3 *itemset* tidak ada yang memenuhi minimal *support*, maka 2 kombinasi yang memenuhi untuk pembentukan asosiasi.

4. Pembentukan Aturan Asosiasi

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* menghitung *confidence* aturan asosiatif A, B.

Minimal \rightarrow *confidence* = 70% Dari kombinasi 2 *itemset* yang telah ditemukan, dapat dilihat besarnya nilai *support* dan *confidence* dari calon aturan asosiasi seperti pada tabel dibawah ini:

TABEL IX. Tabel Aturan Asosiasi

Aturan	Confidence
Jika membeli A, maka akan membeli C	4/9 44,44%
Jika membeli C, maka akan membeli A	4/6 66,67%
Jika membeli A, maka akan membeli B	4/9 44,44%
Jika membeli B, maka akan membeli A	4/7 57,14%
Jika membeli A, maka akan membeli E	7/9 77,77%
Jika membeli E, maka akan membeli A	7/8 87,5%
Jika membeli D, maka akan membeli B	4/6 66,67%
Jika membeli B, maka akan membeli D	4/7 57,14%

5. Aturan Asosiasi Final

Aturan asosiasi final terurut berdasarkan minimal *support* dan minimal *confidence* yang telah ditentukan, dapat dilihat pada tabel dibawah ini:

TABEL X. Aturan Asosiasi Final

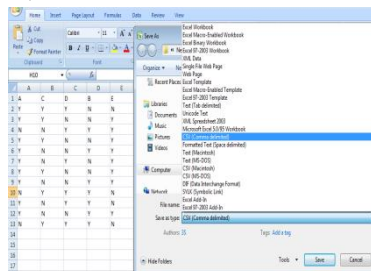
Aturan	Support	Confidence
Jika membeli A, maka akan membeli E	58,33%	77,77%
Jika membeli E, maka akan membeli A	58,33%	87,5%

Jika yang lebih banyak terjual adalah E dan A dengan diketahuinya penjualan yang paling banyak terjual tersebut, sehingga perusahaan dapat menyusun strategi pemasaran dan dapat menambah persediaan E dan A.

3.2. Pengujian Dengan Aplikasi Weka

A. Proses Penentuan Data

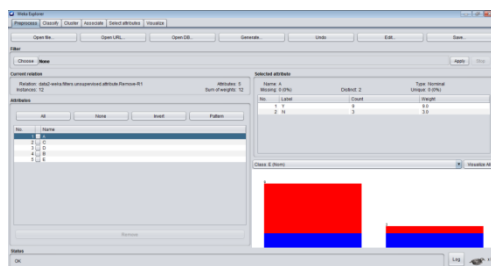
Pada saat dilakukan pengujian dengan weka maka langkah awal yang harus di persiapkan adalah data, untuk uraian data maka dimanfaatkan aplikasi Microsoft excel dengan penyimpanan dengan extension CSV (Comma Delimited), uraian gambar dibawah ini.



Gambar 1. Persiapan Data Pengujian

B. Melakukan Pengujian Dengan Weka

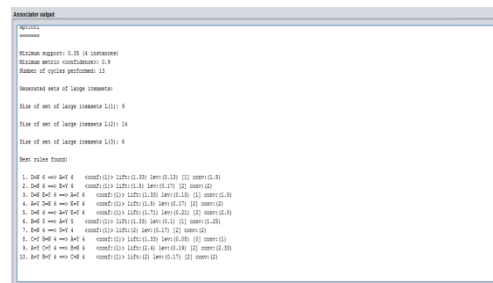
Setelah persiapan data sesuai maka aplikasi Weka diaktifkan dan pilih open file, silahkan cari data yang telah disimpan pada langkah 1, tampilan sesuai dengan gambar berikut ini.



Gambar 2. Melakukan Pengujian Weka

Pada saat gambar pemilihan data selesai maka secara otomatis fitur yang disediakan oleh weka sudah aktif yaitu classify, cluster, Associate dan yang lainnya. Untuk tahapan ini karena yang akan diselesaikan

adalah apriori maka pilihan aktif adalah Associate. Setelah proses ini selesai maka pilih Choose dengan memilih apriori dan Tekan Tombol Start maka otomatis muncul analisa dari apriori seperti gambar berikut ini.



Gambar 3. Hasil Analisa Apriori

Rule yang dihasilkan dari data yang tersedia adalah sebanyak 10 rule dengan tingkat confidence yang berbeda. Salah satu contoh rule 1, pada saat tidak ada D maka Aktifitas A ada. Disana ada Y dan N, dimana dari data awal diposisikan bahwa Y menyatakan ada sedangkan N menyatakan tidak. Untuk melatih pemahaman maka perlu menambahkan data sebanyak mungkin agar proses analisa semakin akurat.

IV. PENUTUP

Berdasarkan pembahasan dengan penerapan apriori menghasilkan beberapa rule yang dapat digunakan oleh para pemangku kepentingan dalam menggali informasi yang terkandung didalam tumpukan data ataupun big data dan dengan pengujian algoritma menggunakan aplikasi weka maka proses semakin efektif dan menghasilkan suatu rule dengan waktu yang singkat.

V. REFERENSI

- [1] Kusriani & Emha Taufiq Luthfi, "Algoritma Data Mining", Penerbit Andi, Yogyakarta, 2010
- [2] Kusriani, "Konsep dan Aplikasi Sistem Pendukung Keputusan", Penerbit Andi, Yogyakarta, 2007
- [3] Larose, Daniel T. 2005. Discovering Knowledge in Data: an Introduction to Data Mining. John Wiley & Sons, Inc.
- [4] MacLennan, J. dan Tang, Z.H. 2005. Data Mining with SQLServer 2005. United States of America : Wiley Publishing.

-
- | | |
|--|--|
| <p>[5] Turban, E., dkk. 2005. Decision Support System and Intelligent Systems. Yogyakarta: Andi Offset.</p> <p>[6] Kusrini, Emha Taufiq Luthfi, Algoritma Data Mining</p> <p>[7] Dana Sulistyo Kusumo, Moch. Arief Bijaksana, Dhinta Darmantoro, Data Mining</p> | <p>Dengan Algoritma Apriori Pada RDBMS Oracle, 2003</p> <p>[8] Nugroho Adi, “Rekayasa Perangkat Lunak Menggunakan UML dan Java”, Andi, Yogyakarta, 2009.</p> |
|--|--|