

PENGGUNAAN ALGORITMA APRIORI DATA MINING UNTUK MENGETAHUI TINGKAT KESETIAAN KONSUMEN (*BRAND LOYALTY*) TERHADAP MEREK KENDERAAN BERMOTOR (STUDI KASUS DEALER HONDA RUMBAL)

Wirdah choiriah

(Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Lancang Kuning
(Jl. Yossudarso KM.8 Rumbal, telp/fak. (0761) 7745164
(e-mail: wirdah@unilak.ac.id)

Abstrak

Algoritma yang umum digunakan dalam proses pencarian frequent itemset (data yang paling sering muncul) adalah Apriori. Tetapi Algoritma Apriori mempunyai memiliki kekurangan yaitu membutuhkan waktu yang lama dalam proses pencarian frequent itemset. Dengan memanfaatkan data Transaksi konsumen yang dihubungkan dengan pola kesetiaan konsumen terhadap merek kendaraan bermotor Honda maka pola hubungan keduanya melalui teknik data mining, association rule. Kategori profesi, jenis kelamin konsumen dan merek kendaraan bermotor di ukur dengan parameter pada tingkat ketertarikan konsumen terhadap merek kendaraan yang di sajikan. Algoritma yang digunakan adalah algoritma apriori, informasi yang ditampilkan berupa nilai support dan confidence dari masing-masing kategori.

Kata kunci: data mining, association rule, data transaksi, algoritma apriori, support, confidence.

Abstract

The algorithm is commonly used in the process of finding frequent itemset (data that most often comes up) is Apriori. But the Apriori algorithm has a disadvantage that has take a long time in the process of finding frequent itemset. By utilizing the data consumer transactions associated with patterns of consumer loyalty to the brand Yamaha motor vehicles then their relationship patterns through data mining techniques, association rule. Professional category, gender consumers and brand of motor vehicles on the parameters measured by the level of consumer interest in the brand vehicles are at present. The algorithm used is a priori algorithm, the information displayed in the form of support and confidence values of each category.

Keywords: data mining, association rule, transaction data, apriori algorithm, support, confidence.

1. Pendahuluan

Istilah *market basket* digunakan untuk menggambarkan kelompok *item* (terdiri dua atau lebih *item*) yang cenderung dibeli oleh satu konsumen sewaktu berbelanja di perusahaan/swalayan dalam satu transaksi pembelian. Untuk saat ini, sarana transportasi yang sangat bisa diandalkan untuk kalangan kelas menengah yaitu Kendaraan bermotor. Dan kendaraan bermotor yang menjadi andalan untuk saat ini adalah sepeda motor Honda. Keunggulan sepeda motor Honda sudah menjadi satu image tersendiri di kalangan masyarakat karena performa mesin yang canggih dan awet. Data kendaraan bermotor dan pemanfaatan dapat diolah menjadi informasi bagi Dealer penjual kendaraan bermotor.

Salah satu bentuk informasi penting yang dibutuhkan oleh dealer kendaraan bermotor adalah informasi tentang seberapa besar pola kesetiaan konsumen terhadap kendaraan bermotor yang dijual, serta bagaimana memprediksi merek kendaraan bermotor yang akan dibeli oleh konsumen. Loyalitas terhadap suatu merek (*Brand Loyalty*) merupakan ukuran keterikatan seorang konsumen pada suatu merek, dan informasi yang penting bagi dealer merupakan inti dari kekuatan suatu merek (*Brand Quality*). Penelitian ini dilakukan agar data penjualan kendaraan bermotor dapat diolah menjadi informasi yang penting bagi dealer penjual kendaraan

bermotor. Bentuk informasi yang penting yang diteliti adalah pola kesetiaan konsumen terhadap merek kendaraan bermotor.

Peneliti menerapkan metode *data mining rule based classification* untuk membentuk rule-rule penjualan kendaraan. Hasil penelitian dapat membantu perusahaan atau dealer kendaraan bermotor untuk mendapatkan informasi *brand loyalty* konsumen terhadap merek kendaraan yang dijualnya dan membantu prediksi merek kendaraan berikutnya yang akan dibeli seorang konsumen

2. Metode Penelitian

2.1 Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Di dalam jurnal yang berjudul “Pembentukan Cluster dalam Knowledge in Database dengan Algoritma K-Means”. *Knowledge Discovery in Database* (KDD) didefinisikan sebagai ekstraksi informasi potensial, implisit dan tidak dikenal dari sekumpulan data.

Istilah *data mining* dan *knowledge discovery in database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*. Proses KDD secara garis besar dapat dijelaskan sebagai berikut: gambar 2.1 Tahapan *data mining* Dalam proses KDD (Fayyad dalam Kusriani dan Luthfi, 2007),

2.2 Data mining

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual (Pramudiono dalam Kusriani dan Luthfi, 2007).

Selain itu juga ada beberapa definisi tentang *data mining*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengestraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Turban, dalam Kusriani dan Luthfi, 2007)

2.3 Algoritma apriori

Algoritma apriori adalah *algoritma* yang paling terkenal untuk menemukan pola frekuensi tinggi. *algoritma apriori* dibagi menjadi beberapa tahap yang disebut narasi atau pass (Devi dinda setiawan, 2009).

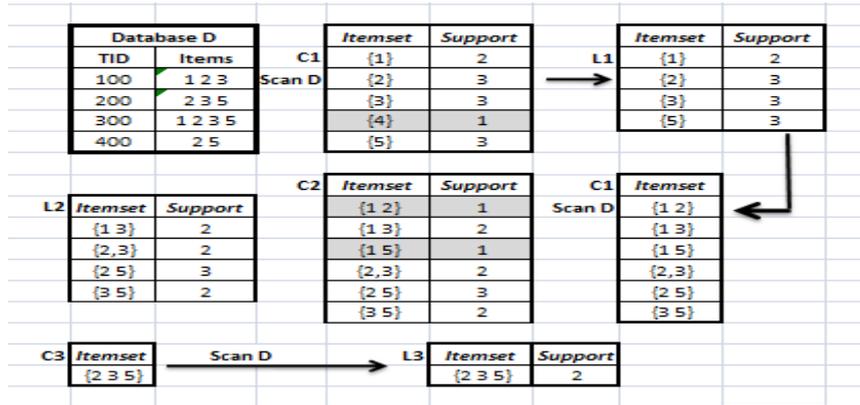
1. Pembentukan kandidat *itemset*, kandidat *k-itemset* dibentuk dari kombinasi $(k-1)$ -*itemset* yang didapat dari iterasi sebelumnya. Satu cara dari *algoritma apriori* adalah adanya pemangkasan kandidat *k-itemset* yang *subset*-nya yang berisi $k-1$ item tidak termasuk dalam pola frekuensi tinggi 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* didalam kandidat *k-itemset* tersebut. Ini adalah juga ciri dari *algoritma apriori* dimana diperlukan penghitungan dengan cara 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 kembali bagian 1.

Contoh dari penerapan *algoritma apriori* adalah diilustrasikan digambar berikut:

Keterangan Code

Ck : Candidate Itemset of Size k

Lk : Frequent Itemset of size k



Gambar 2.1 Gambar Ilustrasi Algoritma Apriori

2.4 Association Rule

Algoritma aturan asosiasi akan menggunakan data latihan, sesuai dengan pengertian *data mining*, untuk menghasilkan pengetahuan. Pengetahuan apakah yang hendak dihasilkan dalam aturan asosiasi? Pengetahuan untuk mengetahui *item-item* belanja yang sering dibeli secara bersamaan dalam suatu waktu. Aturan asosiasi yang berbentuk “if...then...” atau “jika...maka...” merupakan pengetahuan yang dihasilkan dari fungsi aturan asosiasi (Beta Noranita dan Nurdin Bahtiar)..

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap :

1. Analisa pola frekuensi tinggi

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

$$Support(A) = \frac{JumlahTransaksiMengandungA}{TotalTransaksi}$$

Sedangkan nilai dari *support* 2 item diperoleh dari rumus berikut :

$$Support(A,B) = P(A \cap B) = \frac{JumlahTransaksiMengandungAdanB}{TotalTransaksi}$$

2. Pembentukan aturan assosiatif

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan assosiatif yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan assosiatif A »B Nilai *confidence* dari aturan A »B diperoleh dari rumus berikut :

$$Confidence = P(B|A) = \frac{JumlahTransaksiMengandungAdanB}{JumlahTransaksiMengandungA}$$

3. Hasil dan Pembahasan

Tahap ini melakukan analisa terhadap nilai-nilai objek yang akan diangkat, kemudian *value* atau nilai data tersebut akan diolah untuk mendapatkan informasi yang nantinya akan digunakan dalam tahap implementasi. Dan berikut adalah sampel Data Transaksi Konsumen pada bulan januari. Adapun Data Transaksi Penjualan Dapat dilihat pada Tabel 3.1 Data Transaksi Penjualan Bulan Januari di bawah ini :

Tabel 3.1 Data Transaksi Penjualan Bulan Januari 2011

No	Nama Konsumen	Jenis Kelamin	Pekerjaan	Alamat	Merek Kenderaan	Tgl Pembelian
Jan001	Boby Harvan Hidayat	Laki-laki	Mahasiswa	Jl. Sudirman Pekanbaru,	Vario	20/01/2011

Jan00 2	Despa Sari	Perempua n	Mahasiswa	Jl. Pahlawan Kerja No.19 A Raya Pekanbaru.	Supra 125	04/01/201 1
Jan00 3	Maryono	Laki-laki	Karyawan	Jl. Garuda Sakti KM 5 Pekanbaru	Tiger	21/01/201 1

3.1 Analisa Data

Agar data yang akan diolah benar-benar relevan dengan yang dibutuhkan maka dilakukan Penggabungan data (*Integrasi Data*) kedalam suatu penyimpanan baru kemudian ditemukan atribut – atribut yang digunakan dalam data konsumen. Adapun atribut-atribut di pakai adalah:

1. ID adalah Nomor Registrasi transaksi pada data Konsumen.
2. Konsumen adalah data konsumen yang terdiri dari Jenis kelamin dan profesi konsumen.
3. Item adalah Kategori untuk Merek kendaraan bermotor.

Tabel 3.2 Atribut Terpilih

Atribut	Keterangan
ID	ID adalah Nomor registrasi transaksi
Konsumen	Konsumen= profesi dan jenis kelamin Konsumen: 1. MK = Mahasiswa Laki-laki 2. MP = Mahasiswa Perempuan 3. KK = Karyawan Laki-laki 4. KP = Karyawan Perempuan
Item	Item = Kategori untuk Merek kendaraan bermotor disimbolkan dengan : 1. q1 adalah merek kendaraan bermotor Vario 2. q2 adalah merek Kendaraan Supra 125 3. q3 adalah merek kendaraan Tiger 4. q4 adalah merek kendaraan Revo 5. q5 adalah merek kendaraan CBR

Untuk Memudahkan dalam pengolahan data maka dilakukan Analisa data kemudian dari Atribut-atribut yang terpilih maka terbentuklah Data Penjualan pada bulan januari 2011 yang sudah disimbolkan berdasarkan atribut terpilih pada Tabel 4.3 Atribut Terpilih di atas. Maka Terbentuklah Tabel 5.3 Data Penjualan yang dapat dilihat seperti di bawah berikut :

Tabel 3.3 Data Penjualan

Id	Konsumen		Merek Kendaraan (Item)	Jumlah
1	Mahasiswa	Laki-laki	Vario (q1)	1 unit
2	Mahasiswa	Laki-laki	Supra 125(q2)	1 unit
1	Mahasiswa	Perempuan	Vario (q1)	1 unit
2	Mahasiswa	Perempuan	Supra 125(q2)	2 unit
1	Karyawan	Perempuan	Vario (q1)	4 Unit
1	Karyawan	Laki-Laki	VARIO (q1)	2 unit

2	Karyawan	Laki-laki	Tiger (q3)	2 unit
3	Karyawan	Laki-Laki	Supra 125(q2)	1 unit
4	Karyawan	Laki-laki	CBR (q5)	1 unit

Berdasarkan Data Rekap Transaksi selama enam bulan tersebut maka dilakukan *proses Cleaning* pada Data Transaksi Penjualan yang terdapat pada Tabel 5.1 di atas tersebut, sehingga membentuk Tabel Data Awal Transaksi guna untuk memudahkan dalam proses Penentuan atribut yang dipakai. Adapun sampel data yang telah melewati proses validasi adalah terlihat pada Tabel 3.4 Data Awal Transaksi yang disimbolkan pada bulan Januari 2011 adalah sebagai berikut :

Tabel 3.4 Data Awal Transaksi

Id	Konsumen	Item
1	Mk	q1
2	Mk	q2
3	Mp	q1
4	Mp	q2
5	Kp	q1
6	Kk	q1
7	Kk	q2
8	Kk	q3
9	Kk	q5

Di dalam *algoritma apriori* langkah selanjutnya adalah menemukan *frequent itemset*. *Frequent itemset* adalah *set of item* yang *supportnya* besar sama dengan batas *minimum support* yang telah ditentukan sebelumnya (10%). Yang dimulai dengan menemukan *frequent itemset* dengan 1-item. Pada tabel 3.5 di bawah ini adalah *frequent itemset* dengan 1 item. di bawah ini adalah *frequent itemset* dengan 1 item.

$$\text{support } A = \frac{\text{Transaksi } A}{\text{Transaksi}}$$

Tabel 3.5 Frekuensi 1 itemset (F1)

1-item	count	Support
Kk	21	46,67%
kp	11	24,44%
q3	8	17,78%
q1	16	35,56%
q2	11	24,44%
mp	7	15,56%
mk	6	13,33%
q4	6	13,33%
q2	11	24,44%

Pada tabel 3.5 Frekuensi 1-itemset dijelaskan kegiatan transaksi diatas adalah jumlah Konsumen yang berprofesi sebagai karyawan laki-laki adalah sebanyak 21 orang, karyawan perempuan sebanyak 11 orang, mahasiswa laki-laki sebanyak 6 orang, mahasiswa perempuan sebanyak 7 orang. Dan konsumen yang membeli kendaraan merek Vario sebanyak 16 orang, Supra 125sebanyak 11 orang, Tiger sebanyak 8 orang, Vega sebanyak 6 orang.

Selanjutnya setelah ditentukan *support itemset* seperti tabel di atas maka dapat ditentukan *support item 2- itemset* berdasarkan bulan dengan mengkombinasikan masing-masing produk untuk 2 *itemset* yang *frequent*. Dapat dilihat pada Tabel 3.6 Frekuensi 2 *Itemset* (F2) di bawah ini berdasarkan rumus *Support* berikut.

$$support (A, B) = \frac{\sum \text{transaksi mengandung } A \text{ dan } B}{\sum \text{ Taransaksi}}$$

Tabel 3.6 Frekuensi 2 *Itemset* (F2)

Item	Quantity	Support
kk-q3	5	11,11%
kp-q1	5	11,11%
kk-q2	6	13,33%
q3-kk	5	11,11%

Pada tabel 3.6 Frekuensi 2-*itemset* di jelaskan kegiatan transaksi diatas adalah jumlah data Konsumen yang jika karyawan laki-laki dan membeli Tiger adalah sebanyak 11,11% dari jumlah data transaksi, jika karyawan perempuan dan membeli Vario banyak 11,11% dari jumlah data transaksi, jika karyawan laki-laki dan membeli Supra 125 adalah sebanyak 13,33% dari jumlah data transaksi, jika merek kendaraan adalah Tiger maka yang membeli adalah Karyawan laki-laki, jika merek Vario maka yang membeli adalah karyawan perempuan, dan jika Supra 125 maka yang membeli adalah karyawan laki-laki sebanyak 13,33% dari jumlah transaksi.

Untuk mendapatkan nilai *support* pada tabel 4.8 adalah dengan cara menjumlahkan semua nasabah yang mengambil 3 atribut sekaligus dibagi dengan total semua data konsumen perbulan dan dikali dengan 100%. Pada bulan Januari Nilai *support* yang memenuhi *support* minimum $\geq 10\%$. Setelah semua *support* ditentukan dilanjutkan dengan membuat *rule asosiasi* mulai dari 2 *itemset* seperti pada Tabel 3.7 *Association Rule* di bawah ini.

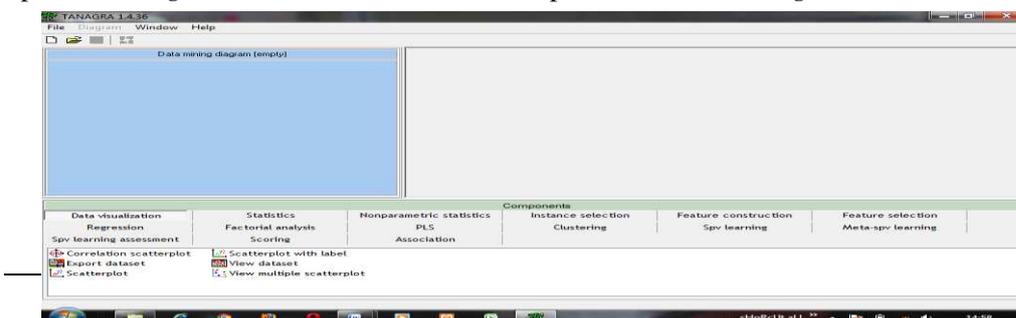
Tabel 3.7 *Association Rule*

<i>Association Rule</i>	<i>Support minimal</i>	<i>Confidence</i>
Jika Merek kendaraan adalah Tiger maka yang membeli adalah Karyawan Laki-laki	10%	62,5%
Jika Merek kendaraan adalah Supra 125 maka yang membeli adalah Karyawan Laki-laki	10%	54,54%

Berdasarkan nilai *confidence* yang telah ditemukan untuk masing-masing *rule asosiasi* seperti pada tabel 4.9 maka dapat diketahui apakah masing-masing *rule asosiasi interesting* adalah 2 rule yang *interesting*. Suatu rule asosiasi dikatakan *interesting* jika *confidence* dan *support* dari masing-masing *rule asosiasi* dari batas minimal *support* $\geq 10\%$ dan *confidence* $\geq 50\%$ yang telah ditentukan. Maka rule-rule yang *interesting* adalah sebagai berikut :

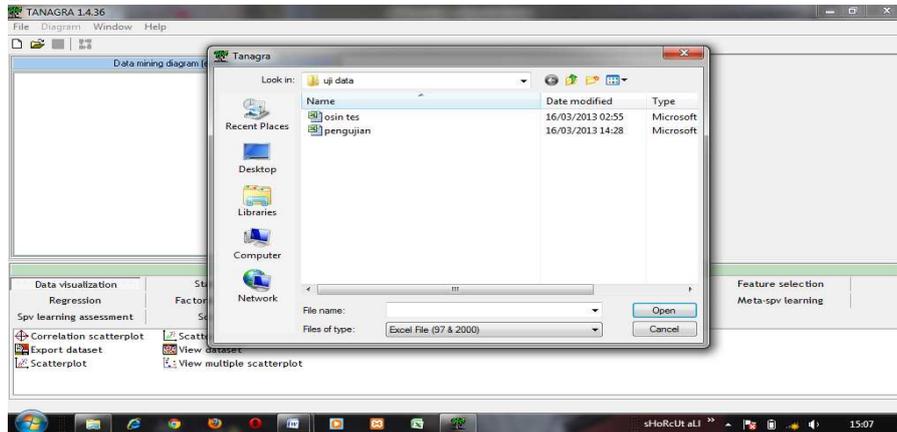
1. Untuk nilai *confidence* rule asosiasi Jika Merek kendaraan adalah Tiger maka yang membeli adalah Karyawan Laki-laki sebanyak 62,66%.
2. Nilai *confidence* dari rule Jika Merek kendaraan adalah Supra 125 maka yang membeli adalah Karyawan Laki-laki sebanyak 54,54%

Untuk melakukan pengujian pada tahap ini yang pertama adalah dengan membuka aplikasi *Tanagra 1.4*. Gambar 3.1 adalah tampilan awal dari *Tanagra 1.4*



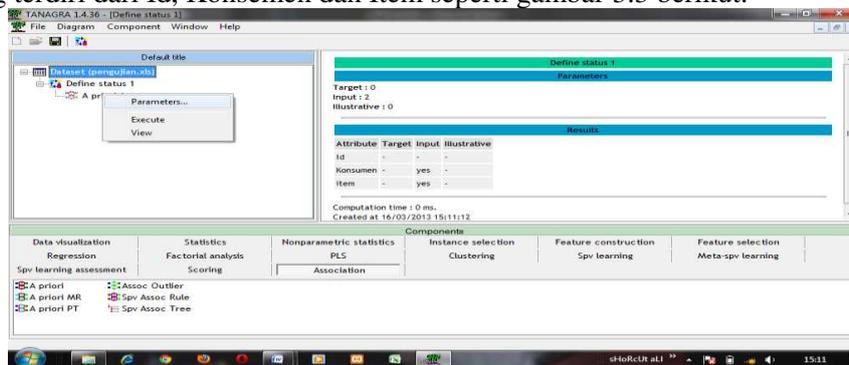
Gambar 3.1 Tampilan Awal Tanagra 1.4

Tampilan selanjutnya seperti gambar 3.1. Kemudian input data file dengan mengubah *type data* yaitu *type file excel file (97 & 2000)*. Di mana pada tahap ini sistem meminta menginputkan nama file. Pada pengujian ini nama filenya adalah hasil akhir seperti pada gambar 3.2



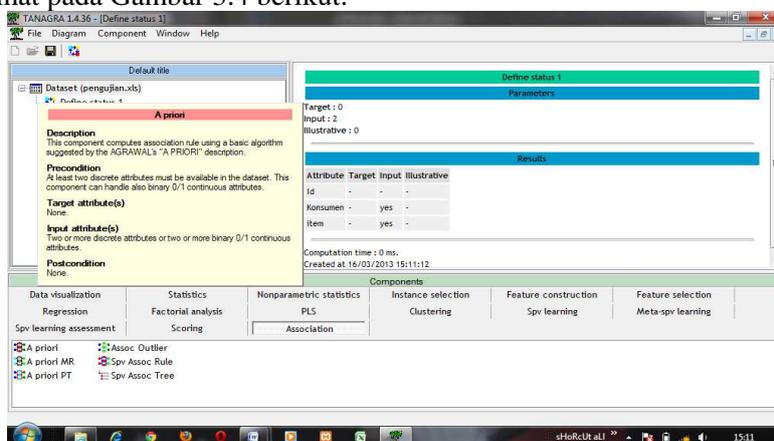
Gambar 3.2 Input data pada File

Setelah melakukan penginputan data, maka akan terlihat pada kolom *default titel dataset*. Setelah melakukan penginputan Atribut, maka akan terlihat pada kolom *define status 1* atribut yang terdiri dari Id, Konsemen dan Item seperti gambar 3.3 berikut.



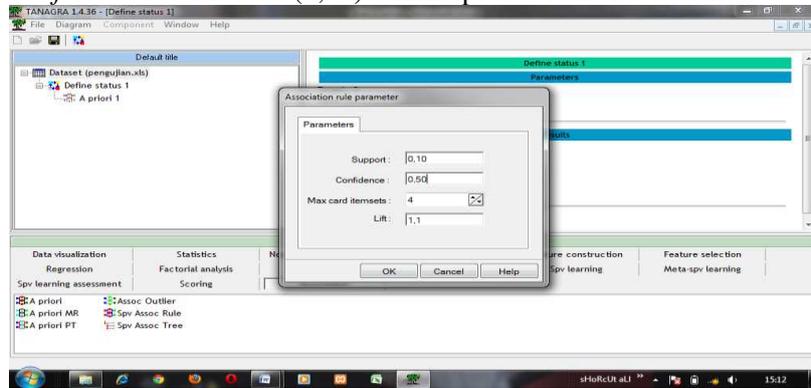
Gambar 3.3 Kolom Atribut

Langkah selanjutnya adalah melakukan penampilan *Algoritma Aprior* pada *define status 1* dengan cara memilih *Association* pada *components* di bagian bawah tampilan tanagra 1.4 seperti terlihat pada Gambar 3.4 berikut.



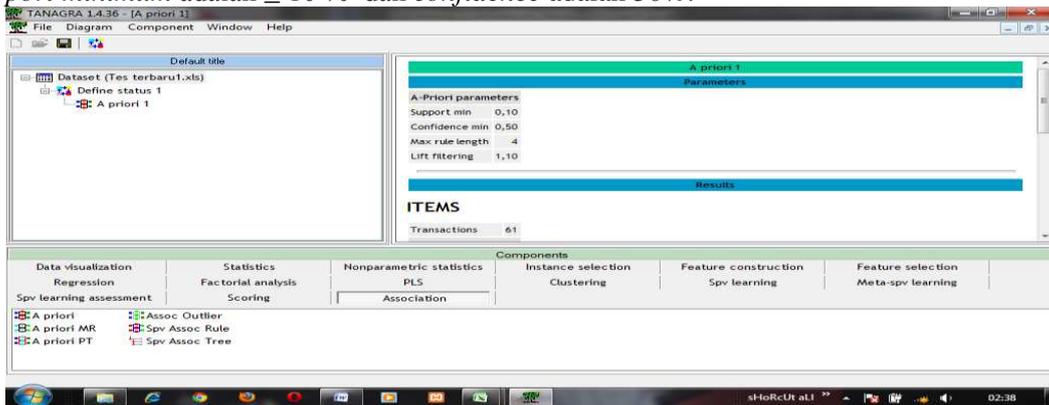
Gambar 3.4 Components Association

Associationrule parameter, untuk menentukan nilai *support* dan *confidence*, maka terlihat *Algoritma Apriori Parameters*, selanjutnya masukkan nilai *support* minimum adalah 10 % (0,10) dan *confidence* adalah 50% (0,50) terlihat pada Gambar 3.5 berikut



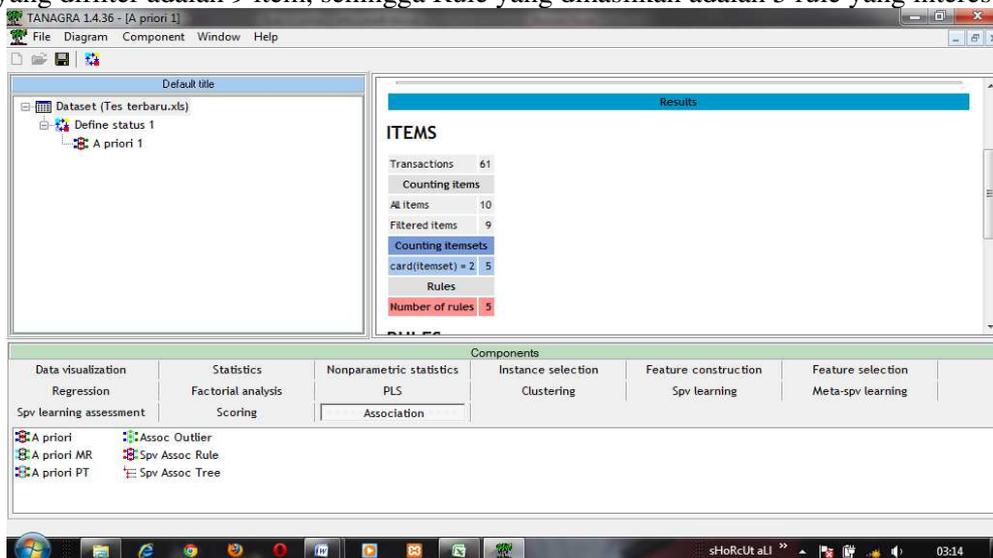
Gambar 3.5 Association parameter

Apriori Parameter yang dihasilkan adalah seperti pada gambar berikut, dimana terlihat *Support minimum* adalah $\geq 10\%$ dan *confidence* adalah 50%.



Gambar 3.6 Apriori Parameter

Kemudian berikut adalah gambar tampilan *Results* dimana item-itemnya adalah Jumlah transaksi adalah 61, kemudian *Counting Items* dimana keseluruhan items adalah 10 item, dan item yang difilter adalah 9 item, sehingga Rule yang dihasilkan adalah 5 rule yang interesting.



Gambar 3.7 Results Item

Berdasar kan data transaksi tersebut, maka ada 5 rule interesting yang di dihasilkan, berikut adalah Gambar 3.8 Tampilan Rule.

Nº	Antecedent	Consequent	Lift	Support	Confidence
1	"Konsumen=mp"	"Item=q1"	1,848	0,131	0,727
2	"Konsumen=kp"	"Item=q1"	1,452	0,131	0,571
3	"Item=q4"	"Konsumen=kj"	1,402	0,098	0,667
4	"Item=q3"	"Konsumen=kj"	1,402	0,098	0,667
5	"Item=q2"	"Konsumen=kj"	1,294	0,131	0,615

Gambar 3. 8 Tampilan Rule

4. Kesimpulan

1. Aplikasi *data mining* ini dapat digunakan untuk mengetahui hubungan tingkat kesetiaan konsumen dengan data transaksi, hubungan tersebut diukur oleh nilai support dan *confidence* antar item.
2. Hasil dari proses *data mining* ini dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih lanjut dengan faktor yang mempengaruhi tingkat kesetiaan konsumen terhadap merek kendaraan bermotor. Serta menggali informasi yang memiliki potensi besar dalam pengambilan keputusan suatu perusahaan dalam menganalisa perilaku konsumen dan menerapkan strategi pemasaran yang tepat sehingga mendatangkan keuntungan bagi pihak perusahaan.

Daftar Pustaka

- [1] Angeline, Magdelene Delighta. *et al* (2012). "Association Rule Generation Using Apriori Mend Algorithm for Student's Placemnt". Department of Computer Science and Engineering. ISSN: 2222-4254.
- [2] Dana Sulistyio Kusumo, *et al* (2003). "Data Mining Dengan Algoritma Apriori Pada RDBMS Oracle". Jurnal Penelitian dan Pengembangan Telekomunikasi. Vol. 8. No. 1.
- [3] Eko Wahyu Tyas (2008). "Penerapan Metode Association Rule Menggunakan Algoritma Apriori untuk Analisa Pola Data Hasil Tangkapan Ikan". Konferensi dan Temu Nasional Teknologi Informasi dan Informasi dan Komunikasi untuk Indonesia. Jakarta.
- [4] Han, Jiawei dan Kamber, Micheline (2001). "Data Mining: Concepts and Techniques". London. Morgan Kaufmann Publishers.
- [5] Xavier Pi-Sunyer, F., Becker, C., Bouchard, R.A., Carleton, G. A., Colditz, W., Dietz, J., Foreyt, R. Garrison, S., Grundy, B. C., 1998, Clinical Guidelines on the identification, evaluation, and treatment of overweight and obesity in adults, *Journal of National Institutes of Health*, No.3, Vol.4, 123-130, :http://journals.lww.com/acsm-mssse/Abstract/1998/11001/paper_treatment_of_obesity.pdf.