

IMPLEMENTASI DATA MINING UNTUK MENEMUKAN HUBUNGAN ANTARA KOTA KELAHIRAN MAHASISWA DENGAN TINGKAT KELULUSAN MAHASISWA PADA FAKULTAS TEKNOLOGI INFORMASI UNISBANK

Pipin Retnosari, Arief Jananto

Abstrak

Basis data mahasiswa UNISBANK berisi data mahasiswa dalam jumlah yang besar dan bervariasi. Data-data tersebut harus diolah agar menjadi informasi yang bernilai lebih. Penelitian ini bertujuan untuk mengetahui hubungan antara kota kelahiran mahasiswa dengan tingkat kelulusan mahasiswa dengan menggunakan teknik data mining dan dihitung melalui 3 cara, yaitu perhitungan manual, dengan software Tanagra dan RapidMiner. Sebelum proses mining dilakukan, data terlebih dahulu melewati proses pembersihan, seleksi dan transformasi. Metode yang digunakan pada penelitian ini adalah metode *association rules* dengan menggunakan algoritma apriori, yaitu algoritma untuk menentukan pola frekuensi tertinggi.

Dari hasil analisa, aturan asosiasi yang dihasilkan melalui ketiga cara tersebut secara umum relatif sama, hanya saja RapidMiner dapat mengolah bilangan minimum *support* kurang dari 1,0 sampai tiga angka dibelakang koma, sedangkan Tanagra hanya dapat mengolah sampai dua angka dibelakang koma.

Dari hasil perhitungan, dapat terlihat daerah mana saja yang mempunyai tingkat keberhasilan tinggi, sehingga dapat menjadi bahan pertimbangan ketika UNISBANK akan melakukan promosi.

Kata Kunci : Data Mining, *Association Rules*, *Algoritma Apriori*, *Tanagra*, *RapidMiner*

1. PENDAHULUAN

1. 1. Latar Belakang Penelitian

Banyaknya kegiatan yang dilakukan secara komputerisasi sekarang ini mempermudah dalam ketersediaan data. Namun data yang tersedia seringkali hanya diperlakukan sebagai rekaman tanpa pengolahan lebih lanjut sehingga tidak bernilai lebih untuk keperluan masa mendatang.

Bisnis dalam bidang apapun saat ini dituntut memiliki keunggulan bersaing dengan memanfaatkan semua sumber daya yang dimiliki. Pelaku bisnis harus mampu melakukan proses evaluasi, perencanaan dan pengelolaan secara baik untuk dapat memenangkan persaingan di era globalisasi ini.

Teknologi informasi dapat digunakan untuk mendapatkan, mengolah dan menyebarkan informasi untuk menunjang kegiatan operasional sehari-hari sekaligus menunjang kegiatan pengambilan

keputusan strategis yang pada akhirnya dapat meningkatkan keunggulan bersaing.

Teknik yang dikenal dengan data mining dapat menghasilkan informasi yang bermanfaat tersebut. Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual (Pramudiono, 2006). Metode data mining yang akan digunakan pada penelitian ini adalah metode *association rules* dengan menggunakan algoritma apriori, yaitu algoritma untuk menemukan pola frekuensi tinggi.

Penelitian ini akan membandingkan hasil perhitungan manual berdasar ilmu data mining dengan hasil perhitungan menggunakan software data mining, yaitu Tanagra dan RapidMiner.

1. 2. Perumusan Masalah

Masalah yang akan diteliti adalah

bagaimana menggali dan menemukan tingkat pengaruh antara kota kelahiran mahasiswa dengan tingkat kelulusan mahasiswa.

Data yang digunakan dalam penelitian ini dibatasi hanya pada data kelulusan mahasiswa tahun 2009 sampai dengan 2011.

1. 3. Tujuan dan Manfaat Penelitian

Penelitian ini bertujuan untuk mengetahui sejauh mana pengaruh kota kelahiran mahasiswa terhadap tingkat kelulusan mahasiswa, dimana perhitungan dilakukan dengan 3 cara, yaitu manual dengan MS. Excel, dengan software Tanagra dan dengan software RapidMiner. Dari hasil penelitian ini akan di peroleh informasi yang dapat digunakan sebagai dasar analisis dalam pengambilan keputusan, sekaligus mengetahui tingkat ketelitian perhitungan dari software-software tersebut.

2. TINJAUAN PUSTAKA

2.1. Pustaka yang Terkait dengan Penelitian

Mengenai implementasi data mining ini telah dilakukan penelitian sebelumnya, yaitu implementasi data mining untuk menemukan pola hubungan tingkat kelulusan mahasiswa dengan data induk mahasiswa yang dilakukan di Fakultas MIPA Universitas Diponegoro, dimana digunakan teknik data mining dengan algoritma apriori.

Tabel yang digunakan yaitu tabel data induk mahasiswa dan tabel data kelulusan. Tabel data induk mahasiswa berisi atribut antara lain NIM, jenis kelamin, nama mahasiswa, kota lahir, tanggal lahir, agama, proses masuk, alamat mahasiswa, nama wali, alamat wali, pendidikan wali, nama asal sekolah, kota asal sekolah dan tahun lulus asal sekolah. Sedangkan tabel data kelulusan berisi atribut antara lain NIM, nama mahasiswa, tempat tanggal lahir, program studi, tanggal lulus, IPK, lama studi, judul skripsi dan periode wisuda.

Atribut NIM digunakan sebagai *primary key* untuk merelasikan kedua tabel tersebut.

Data induk mahasiswa yang diambil adalah data mahasiswa angkatan tahun 2000 sampai 2003 dan data kelulusan yang dipakai adalah data kelulusan tahun 2004 sampai 2008.

Dari penelitian tersebut disimpulkan bahwa aplikasi data mining dapat digunakan untuk mengetahui hubungan tingkat kelulusan dengan data induk mahasiswa yang diukur dari nilai *support* dan *confidence* antar item (Noranita, B. & Bahtiar, N., 2010).

Juga dari penelitian yang lain, yaitu implementasi data mining algoritma apriori pada sistem penjualan yang dilakukan oleh mahasiswa Sekolah Tinggi Manajemen Informatika dan Komputer Yogyakarta.

Perancangan sistem penjualan ini menggunakan UML (*Unified Modeling Language*) dan software yang digunakan untuk mengimplementasikan adalah Netbeans 6.8 dan PHP My Admin.

Terdapat 4 tabel yang digunakan, yaitu tabel user untuk menyimpan data user atau admin, tabel transaksi untuk menyimpan data transaksi, tabel barang untuk menyimpan data barang, dan tabel aturan untuk menyimpan data analisis apriori.

Tabel user berisi atribut id user, username, password dan status. Tabel transaksi berisi atribut id transaksi, no transaksi, id barang, nama barang dan jumlah. Tabel barang berisi atribut id barang, nama barang, harga satuan dan jumlah. Sedangkan tabel aturan berisi atribut id aturan, aturan, support, confidence dan hasil.

Pada proses pengujian, dari data yang telah tersedia ditentukan itemset1, itemset2 dan itemset3. Setelah itu dipilih kombinasi yang jumlahnya lebih dari 1 serta dilakukan perhitungan nilai support dan confidencenya.

Dari penelitian ini disimpulkan bahwa dengan teknik data mining yang diimplementasikan pada sistem penjualan dapat dihasilkan sebuah metode yang bisa meningkatkan penjualan dengan cara memberikan saran kepada konsumen, dimana keterkaitan suatu barang yang

dibeli oleh konsumen bisa dihitung dengan teknik algoritma apriori (Syaifullah, M. A., 2010).

2.2 Perbedaan Penelitian yang Dilakukan dengan Penelitian Terdahulu

Dalam penelitian ini, penulis akan menggunakan metode *association rules* dengan algoritma apriori untuk mencari hubungan antara kota kelahiran mahasiswa dengan tingkat kelulusan mahasiswa pada Fakultas Teknologi Informasi UNISBANK.

Data yang digunakan dalam penelitian ini diperoleh dari Fakultas Teknologi Informasi UNISBANK, dimana data kelulusan yang dipakai adalah data kelulusan mahasiswa tahun 2009 sampai dengan 2011.

Jika pada penelitian sebelumnya yang telah dilakukan di Universitas Diponegoro adalah perancangan aplikasi data mining, maka penelitian kali ini akan membandingkan antara hasil perhitungan manual berdasar ilmu data mining dengan hasil perhitungan software-software data mining.

Dari penelitian ini diharapkan dapat diketahui seberapa besar pengaruh dari kota kelahiran terhadap tingkat kelulusan mahasiswa tersebut, serta mengetahui tingkat kevalidan dari hasil perhitungan software-software data mining, yang nantinya dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih lanjut.

3. LANDASAN TEORI DAN RANCANGAN PENELITIAN

3.1. Pengertian Data Mining

Secara sederhana data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Davies, 2004).

Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat

dipahami dan bermanfaat bagi pemilik data (Larose, 2005).

Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar (Larose, 2005).

Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual (Pramudiono, 2006).

Data mining sering juga disebut *knowledge discovery in database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran data mining bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan.

Hal penting yang terkait dengan data mining adalah :

- 1) Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
- 2) Data yang akan diproses berupa data yang sangat besar.
- 3) Tujuan data mining adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat. (Kusrini & Luthfi, E. T., 2009).

3.2. Rancangan Penelitian

Dalam penelitian ini akan dicari aturan asosiasi melalui perhitungan nilai *confidence* dari hubungan antara kota kelahiran mahasiswa dengan tingkat kelulusan mahasiswa dengan membandingkan 3 (tiga) cara, yaitu dengan perhitungan manual, menggunakan software Tanagra dan menggunakan software RapidMiner. Hubungan antara kota kelahiran mahasiswa dengan tingkat kelulusan mahasiswa bermanfaat untuk mengetahui daerah mana saja yang

mempunyai tingkat keberhasilan tinggi. Perbandingan ini bertujuan untuk mencari tahu apakah akan didapatkan hasil yang sama melalui 3 (tiga) cara yang berbeda tersebut.

Data kelulusan yang digunakan dalam penelitian ini adalah data kelulusan mahasiswa Strata 1 (S1) reguler dari tahun 2009 sampai dengan 2011 yang diperoleh dari Biro Administrasi Akademik Fakultas Teknologi Informasi UNISBANK.

4. METODE PENELITIAN

4.1. Persiapan Data

Dalam penelitian ini akan dicari aturan asosiasi melalui perhitungan nilai *confidence* dari hubungan kota kelahiran mahasiswa dengan tingkat kelulusan mahasiswa dengan membandingkan 3 (tiga) cara, yaitu dengan perhitungan manual, menggunakan software Tanagra dan menggunakan software RapidMiner. Perbandingan ini bertujuan untuk mencari tahu apakah akan didapatkan hasil yang sama melalui 3 (tiga) cara yang berbeda tersebut.

Data kelulusan yang digunakan dalam penelitian ini berupa data matang yang siap diolah. Sebelum diolah menjadi data matang, atribut data kelulusan dapat dilihat pada Tabel 4.1.

Tabel 4.1. Data Kelulusan

Atribut	Keterangan
NIM	Merupakan NIM (Nomor Induk Mahasiswa) sebagai nomor unik identitas di perguruan tinggi. Terdiri dari 10 digit yang merepresentasikan angkatan masuk, fakultas, program studi dan nomor urut saat mendaftar
Nama Mahasiswa	Merupakan nama lengkap mahasiswa
Periode Wisuda	Merupakan tahun saat mahasiswa diwisuda
Program Studi	Merupakan program studi dari mahasiswa
Tempat Lahir	Merupakan Kota tempat mahasiswa dilahirkan
Asal sekolah	Merupakan nama SMA atau SMK mana asal dari mahasiswa
Total SKS	Merupakan jumlah total SKS yang telah ditempuh oleh mahasiswa

IPK	Merupakan IPK (Indeks Prestasi Kumulatif) mahasiswa yang mengukur kemampuan mahasiswa sampai pada waktu tertentu yang dapat dihitung berdasarkan jumlah SKS (Satuan Kredit Semester) mata kuliah yang diambil sampai pada periode tertentu dikalikan dengan nilai bobot masing-masing mata kuliah dibagi dengan jumlah seluruh SKS mata kuliah
-----	--

Perlu dilakukan pembersihan terhadap data-data tersebut, atribut nama sekolah dari data yang ada hanya sedikit yang terisi, maka atribut nama sekolah dihapus. Begitu juga dengan *record* yang isinya tidak lengkap juga dihapus. Data yang mengalami kesalahan pengetikan saat input ke sistem hanya sedikit, sehingga masih tetap bisa digunakan.

Atribut total SKS juga dapat dihapus, karena mahasiswa yang diwisuda mempunyai jumlah total SKS yang hampir sama semua, yaitu 146 SKS sampai dengan 148 SKS. Begitupun dengan atribut program studi yang hanya ada 2 pilihan yaitu SI dan TI, sehingga dapat dihapus.

Dari atribut NIM, dapat diketahui mahasiswa tersebut angkatan tahun berapa, sehingga dapat ditambahkan satu atribut yaitu angkatan. Dari atribut angkatan dan periode wisuda dapat ditransformasikan menjadi masa kuliah. Sehingga didapat satu atribut baru yaitu masa kuliah. Data yang telah diolah tersebut menyisakan atribut pada Tabel 4.2.

Tabel 4.2. Data Olahan

Atribut	Keterangan
NIM	Merupakan NIM mahasiswa
Nama Mahasiswa	Merupakan nama lengkap mahasiswa
Angkatan	Merupakan tahun masuk mahasiswa
Masa kuliah	Merupakan lama mahasiswa menempuh kuliahnya
Tempat Lahir	Merupakan Kota tempat mahasiswa dilahirkan
IPK	Merupakan IPK mahasiswa

Atribut IPK dan masa kuliah ditransformasikan menjadi 6 kategori, yaitu dapat dilihat pada Tabel 4.3.

Tabel 4.3. Kategori berdasar Masa Kuliah dan IPK

Kategori	Keterangan
X1	Lama studi 4 tahun atau kurang dari 4 tahun dan IPK 3,51 – 4,00
X2	Lama studi 4 tahun atau kurang dari 4 tahun dan IPK 2,76 – 3,50
X3	Lama studi 4 tahun atau kurang dari 4 tahun dan IPK 2,00 – 2,75
Y1	Lama studi lebih dari 4 tahun dan IPK 3,51 – 4,00
Y2	Lama studi lebih dari 4 tahun dan IPK 2,76 – 3,50
Y3	Lama studi lebih dari 4 tahun dan IPK 2,00 – 2,75

Atribut NIM, Nama Mahasiswa dan angkatan dapat dihapus karena tidak diperlukan lagi untuk memudahkan pengolahan data. Hasil olahan data telah menghasilkan data matang yang terdiri dari 2 atribut, yaitu terlihat pada Tabel 4.4.

Tabel 4.4. Data Matang

Atribut	Keterangan
Tempat Lahir	Merupakan Kota tempat mahasiswa dilahirkan
Kategori	Kategori X1/ X2 / X3 / Y1 / Y2 dan atau Y3

4.2. Perhitungan Manual

Algoritma apriori adalah algoritma paling terkenal untuk menemukan pola frekuensi tinggi, yaitu pola-pola item di dalam suatu *database* yang memiliki frekuensi atau *support* di atas ambang batas tertentu yang disebut dengan istilah minimum *support* atau *threshold*. *Threshold* adalah batas minimum transaksi. Jika jumlah transaksi kurang dari *threshold* maka item atau kombinasi item tidak akan diikutkan perhitungan selanjutnya. Perhitungan manual yang dimaksud disini adalah perhitungan dengan menggunakan MS Excel. Dari seluruh data yang tersedia yaitu 760 *record*, didapat hasil C1 seperti pada Tabel 4.5.

Tabel 4.5. Kandidat Pertama (C1)

Itemset	Support
X1	55
X2	560
X3	69

Y1	2
Y2	38
Y3	36
AMBARAWA	2
AMBON	1
ATAMBUA	1
BANDUNG	2
BANJARNEGARA	3
BANYUMAS	1
BATANG	8
BAUCAU	1
BEKASI	1
BIAK	1
BLORA	11
BOBONARO	1
BOGOR	2
BOJONEGORO	3
BOYOLALI	7
BREBES	15
BUKIT TINGGI	1
CILACAP	5
DEMAK	34
DILI	6
ENDE	1
GROBOGAN	22
GUNUNG KIDUL	1
JAKARTA	11
JANGGA	1
JEPARA	21
KAB SEMARANG	8
KARANGANYAR	1
KENDAL	46
KENDARI	1
KLATEN	4
KONDO ROS	1
KUDUS	21
KUMAI	1
KUPANG	1
LAHAT	1
LALEIA	1
LUMBIR	1
MAGELANG	5
MAGETAN	1
MAJALENGKA	1
MAKARTI MULIA	1
MALANG	2
MANATUTO	1

MARGA MULYA	1
MATARAM	1
MEULABOH	1
MUARA BUNGO	1
NANGA BULIK	1
NGANJUK	2
NGAWI	1
PADANG SIDEMPUAN	1
PALANGKARAYA	1
PALEMBANG	1
PALOPO	1
PANGKALAN BUN	2
PARDASUKA	1
PATI	23
PAYAKUMBUH	1
PEKALONGAN	13
PEKANBARU	1
PEMALANG	11
PONTIANAK	1
PULOKULON	3
PURBALINGGA	3
PURWAKARTA	2
PURWODADI	3
PURWOKERTO	1
PURWOREJO	3
RAHA	1
REMBANG	19
RIMBO BUJANG	1
SABANG	1
SALATIGA	3
SAMARINDA	1
SAMPIT	1
SANGGAU	1
SARKO	1
SEMARANG	340
SENTEBANG	1
SIMALUNGUN	1
SLEMAN	2
SORONG	1
SRAGEN	4
SUKARAJA KIDUL	1
SUKOHARJO	2
SUNGAILIAT	1
SURABAYA	1
SURAKARTA	2
TANGERANG	1
TAWANGHARJO	2

TEBING TINGGI	1
TEGAL	17
TEMANGGUNG	4
TUBAN	1
TULUNGAGUNG	2
UJUNG PANDANG	1
WAMENA	1
WONOGIRI	2
WONOSOBO	1
YOGYAKARTA	1

Setelah itu ditetapkan *threshold* atau minimum *support* 30, maka *itemset* yang mempunyai *support* kurang dari 30 akan dihapus, seperti yang terlihat pada Tabel 4.6.

Tabel 4.6. Hasil Pemangkasan Pertama (L1)

<i>Itemset</i>	<i>Support</i>
X1	55
X2	560
X3	69
Y2	38
Y3	36
SEMARANG	340
KENDAL	46
DEMAK	34

Dari Tabel 4.6, didapat kandidat kedua seperti pada Tabel 4.7.

Tabel 4.7. Kandidat Kedua (C2)

<i>Itemset</i>	<i>Support</i>
SEMARANG, X1	20
SEMARANG, X2	261
SEMARANG, X3	28
SEMARANG, Y2	15
SEMARANG, Y3	14
KENDAL, X1	2
KENDAL, X2	37
KENDAL, X3	3
KENDAL, Y2	2
KENDAL, Y3	2
DEMAK, X1	2
DEMAK, X2	22
DEMAK, X3	2

DEMAK, Y2	4
DEMAK, Y3	4

Selanjutnya itemset yang mempunyai jumlah *support* kurang dari 30 akan dihapus dan hasilnya seperti pada Tabel 4.8

Tabel 4.8 Hasil Pemangkasan Kedua (L2)

Itemset	Support
SEMARANG, X2	261
KENDAL, X2	37

Dari tabel 4.8 didapatkan $d = 2$ untuk (SEMARANG, X2) dan $d = 2$ pula untuk (KENDAL, X2), maka jumlah aturan asosiasi atau $R = 3^2 - 2^3 + 1 = 2$, sehingga ada 2 aturan asosiasi untuk masing-masing *itemset* tersebut.

```

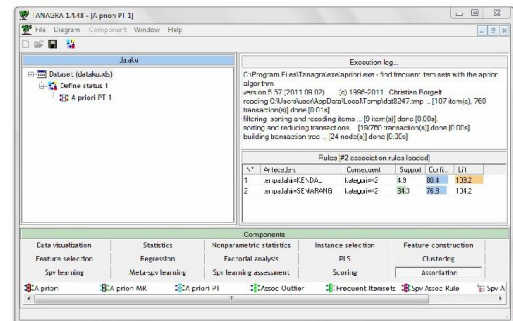
support(A,B) = support(A,B)/support A
support(SEMARANG, X2) = support(SEMARANG, X2)/support SEMARANG
                    = 261/340
                    = 0,768
                    = 76,76%
support(X2, SEMARANG) = support(X2, SEMARANG)/support X2
                    = 261/340
                    = 0,466
                    = 46,61%
support(KENDAL, X2) = support(KENDAL, X2)/support KENDAL
                    = 37/46
                    = 0,804
                    = 80,43%
support(X2, KENDAL) = support(X2, KENDAL)/support X2
                    = 37/46
                    = 0,066
                    = 6,61%
    
```

Misalkan ditetapkan nilai minimum *confidence* adalah 75%, maka aturan asosiasi yang memenuhi hanya 2, yaitu :

- (SEMARANG, X2) dengan nilai *confidence* sebesar 0,768 atau 76,76%
- (KENDAL, X2) dengan nilai *confidence* sebesar 0,804 atau 80,43%

4.3. Perhitungan Dengan Tanagra

Proses perhitungan dengan menggunakan software Tanagra dapat dilihat pada Gambar 4.1 sampai dengan Gambar 4.14.



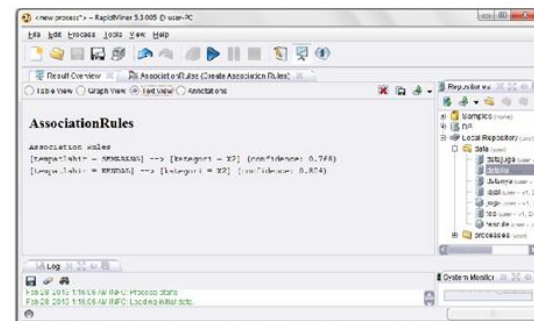
Gambar 4.1 Hasil Aturan Asosiasi

Pada Gambar 4.1 terlihat ada 2 aturan asosiasi yang dihasilkan dari isian minimum *support* 0,039 dan minimum *confidence* 0,75 yaitu :

- (KENDAL, X2) dengan nilai *confidence* sebesar 80,4%
- (SEMARANG, X2) dengan nilai *confidence* sebesar 76,8%

4.4. Perhitungan Dengan RapidMiner

Proses perhitungan dengan menggunakan software RapidMiner seperti pada Gambar 4.2.



Gambar 4.2 Hasil Association Rules

Pada Gambar 4.2 terlihat aturan asosiasi yang dihasilkan, yaitu :

- (SEMARANG, X2), dengan nilai *confidence* sebesar 0,768
- (KENDAL, X2), dengan nilai *confidence* sebesar 0,804

5. HASIL PENELITIAN DAN PEMBAHASAN

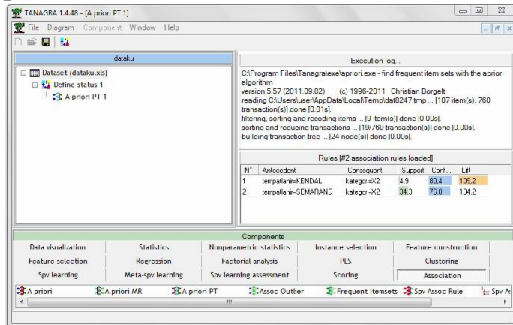
5.1. Hasil Perhitungan Manual

Dari hasil perhitungan manual dengan menggunakan MS Excel, didapatkan 2 aturan asosiasi yaitu :

- (SEMARANG, X2) dengan nilai confidence sebesar 0,768 atau 76,76%
- (KENDAL, X2) dengan nilai confidence sebesar 0,804 atau 80,43%

5.2. Hasil Perhitungan Dengan Tanagra

Dari hasil perhitungan dengan menggunakan software Tanagra, didapatkan aturan asosiasi sebagaimana pada Gambar 5.1.

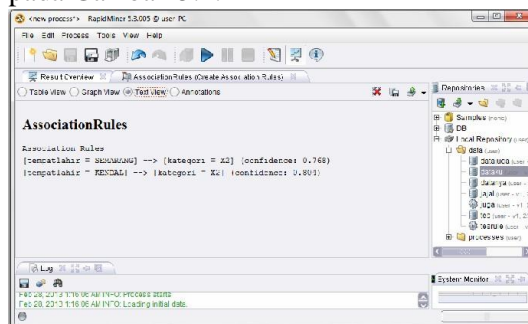


Gambar 5.1. Hasil Aturan Asosiasi dengan Tanagra

- (KENDAL, X2), dengan nilai confidence sebesar 80,4%
- (SEMARANG, X2), dengan nilai confidence sebesar 76,8%

5.3. Hasil Perhitungan Dengan RapidMiner

Dari hasil perhitungan dengan menggunakan software RapidMiner, didapatkan aturan asosiasi sebagaimana pada Gambar 5.2.



Gambar 5.2 Hasil Aturan Asosiasi dengan RapidMiner

Pada Gambar 5.2 terlihat aturan asosiasi yang dihasilkan, yaitu :

- (SEMARANG, X2) dengan nilai confidence sebesar 0,768
- (KENDAL, X2) dengan nilai confidence sebesar 0,804

5.4. Pembahasan

Dari hasil aturan asosiasi yang diperoleh, dapat disimpulkan bahwa hasil perhitungan manual dengan MS Excel sama dengan hasil perhitungan dengan menggunakan software Tanagra dan RapidMiner, seperti pada Tabel 5.1.

Itemset	Hasil Perhitungan		
	Manual	Tanagra	RapidMiner
{SEMARANG, X2}	0,768 = 76,76%	76,8%	0,768
{KENDAL, X2}	0,804 = 80,43%	80,4%	0,804

Tabel 5.1. Perbandingan Hasil Perhitungan

Dari Tabel 5.1 diketahui bahwa perbedaan yang pertama adalah jika dihitung manual maka hasilnya dapat dibuat dalam bentuk desimal ataupun prosentase, dan jika dihitung dengan Tanagra maka hasilnya dalam bentuk prosentase, sedangkan jika dihitung dengan RapidMiner hasilnya dalam bentuk desimal.

Melalui serangkaian proses perhitungan pada bab sebelumnya, dapat diketahui perbedaan lain, yaitu RapidMiner mempunyai nilai maksimum support 1,0, sehingga tidak bisa memasukkan nilai minimum support 30, maka harus diubah dulu menjadi 30 dibagi 760 atau sebesar 0,039.

Sedangkan pada Tanagra, bisa dimasukkan nilai minimum support 30, tetapi untuk mendapatkan hasil yang sama dengan perhitungan manual dan RapidMiner, maka nilai minimum supportnya juga harus diubah menjadi 0,039.

Dari hasil aturan asosiasi yang diperoleh dengan menetapkan nilai confidence 75%, dapat diartikan :

- (SEMARANG, X2) : Jika mahasiswa tersebut lahir di Semarang, maka lama studinya adalah 4 tahun atau kurang dari 4 tahun dan IPKnya 2,76 - 3,50.
- (KENDAL, X2) :

Jika mahasiswa tersebut lahir di

Kendal, maka lama studinya adalah 4 tahun atau kurang dari 4 tahun dan IPKnya 2,76 - 3,50.

Disamping melakukan perhitungan dengan nilai minimum *support* 30 atau 0,039 dan nilai minimum *confidence* 75% atau 0,75, juga dilakukan beberapa pengujian dengan memasukkan nilai minimum *support* yang berbeda, yang hasilnya dapat dilihat pada Tabel 5.2.

Tabel 5.2. Hasil Perhitungan dengan Beberapa Minimum Support

Minimum Support	Itemset	Hasil Perhitungan		
		Tanagra	RapidMiner	Manual
10 atau 0,013	(PEMALANG, X2)	90,9%	0,909	90,91%
	(JAKARTA, X2)	90,9%	0,909	90,91%
	(JEPARA, X2)	85,7%	0,857	85,71%
	(PEKALONGAN, X2)	84,6%	0,846	84,62%
	(KUDUS, X2)	81,0%	0,81	80,95%
	(KENDAL, X2)	80,4%	0,804	80,43%
	(SEMARANG, X2)	76,8%	0,768	76,76%
15 atau 0,020	(JEPARA, X2)	85,7%	0,857	85,71%
	(KUDUS, X2)	81,0%	0,81	80,95%
	(KENDAL, X2)	80,4%	0,804	80,43%
	(SEMARANG, X2)	76,8%	0,768	76,76%
20 atau 0,026	(JEPARA, X2)	85,7%	-	-
	(KUDUS, X2)	81,0%	-	-
	(KENDAL, X2)	80,4%	0,804	80,43%
	(SEMARANG, X2)	76,8%	0,768	76,76%
25 atau 0,033	(KENDAL, X2)	80,4%	0,804	80,43%
	(SEMARANG, X2)	76,8%	0,768	76,76%
30 atau 0,039	(KENDAL, X2)	80,4%	0,804	80,43%
	(SEMARANG, X2)	76,8%	0,768	76,76%
35 atau 0,046	(KENDAL, X2)	80,4%	0,804	80,43%
	(SEMARANG, X2)	76,8%	0,768	76,76%
40 atau 0,053	(SEMARANG, X2)	76,8%	0,768	76,76%
45 atau 0,059	(SEMARANG, X2)	76,8%	0,768	76,76%
50 atau 0,066	(SEMARANG, X2)	76,8%	0,768	76,76%
55 atau 0,072	(SEMARANG, X2)	76,8%	0,768	76,76%

Dari pengujian yang dilakukan dengan menggunakan beberapa minimum *support* seperti yang terlihat pada Tabel 5.2 memperlihatkan bahwa terdapat perbedaan hasil nilai *confidence* saat ditetapkan minimum *support* 20 atau 0,026. Dari hasil analisa yang dilakukan, disimpulkan bahwa RapidMiner dapat membaca inputan

sampai tiga angka di belakang koma, sebagai contoh yaitu 0,026 akan dibaca 0,026 dan 0,020 akan dibaca 0,020, tetapi Tanagra hanya membaca dua angka dibelakang koma, sebagai contoh yaitu 0,026 akan dibaca 0,02 dan 0,020 juga akan dibaca 0,02, sehingga hasil yang diperoleh dari minimum *support* 0,026 dan 0,020 akan sama. Perbedaan ini memperlihatkan bahwa RapidMiner lebih teliti daripada Tanagra, karena hasil perhitungan RapidMiner sama dengan hasil perhitungan manual yaitu menghasilkan 2 aturan asosiasi, sedangkan Tanagra menghasilkan 4 aturan asosiasi.

Dari Tabel 5.2 dapat terlihat daerah mana saja yang mempunyai tingkat keberhasilan tinggi dalam menghasilkan mahasiswa dengan masa kuliah 4 tahun atau kurang dari 4 tahun dengan IPK 2,76 sampai dengan 3,50, serta dapat menjadi pertimbangan dalam memutuskan di daerah mana saja sebaiknya UNISBANK melakukan promosi.

6. KESIMPULAN DAN SARAN

Setelah dilakukan pembahasan pada bab-bab sebelumnya, maka pada bab ini penulis akan mengemukakan kesimpulan serta saran yang dapat diambil dari uraian-uraian pada bab-bab sebelumnya.

6.1. Kesimpulan

1. Dari hasil analisa, didapatkan bahwa aturan asosiasi yang dihasilkan dari ketiga cara tersebut secara umum relatif sama, hanya saja RapidMiner dapat mengolah bilangan minimum *support* kurang dari 1,0 sampai tiga angka di belakang koma, sedangkan Tanagra hanya dapat mengolah sampai dua angka dibelakang koma saja.
2. Dari hasil perhitungan, dapat terlihat daerah mana saja yang mempunyai tingkat keberhasilan tinggi, sehingga dapat menjadi bahan pertimbangan ketika UNISBANK akan melakukan promosi.

6.2. Saran

Peneliti memiliki beberapa saran berkaitan dengan penelitian, yaitu :

1. Dapat dilakukan penambahan atribut lain, sehingga didapatkan hasil aturan asosiasi antar atribut yang lain.
2. Sebelum data diolah, sebaiknya dilakukan uji validasi data terlebih dahulu.

Susanto, S. & Suryadi, D. (2010) *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*, Andi, Yogyakarta.

Syaifulah, M. A., 2010, *Implementasi Data Mining Algoritma Apriori pada Sistem Penjualan*, Amikom, Yogyakarta.

DAFTAR PUSTAKA

Bow, Algoritma Apriori (Apriori Algorithm), 2010, Website : <http://bow-masbow.blogspot.com/2010/10/algoritma-apriori-apriori-algorithm.html>, diakses 1 Nopember 2012.

Kusrini & Luthfi, E. T. (2009) *Algoritma Data Mining*, Andi, Yogyakarta.
 Noranita, B. & Bahtiar, N. (2010) *Implementasi Data Mining untuk Menemukan Pola Hubungan Tingkat Kelulusan Mahasiswa Dengan Data Induk Mahasiswa*, Universitas Diponegoro, Semarang.

Pramudiono, Iko., 2006, *Pengantar Data Mining : Menambang Permata Pengetahuan di Gunung Data*, Website : www.ilmukomputer.org/wp-content/uploads/2006/08/iko-datamining.zip, diakses 1 Nopember 2012.

Rakotomalala, Ricco., 2004, Tanagra, Website : <http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html>, diakses 13 Februari 2013.

RapidMinerTutorial, 2011, RapidMiner Tutorial (part 9/9) Association Rules, Website : <http://www.youtube.com/watch?v=HjqNUfXPuIM>, diakses 13 Februari 2013.

Rapid-GmbH, RapidMiner User Manual, Website : http://docs.rapid-i.com/files/rapidminer/rapidminer-5.0-manual-english_v1.0.pdf, diakses 13 Februari 2013.

Santoso, Budi. (2007) *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Graha Ilmu, Yogyakarta.