

PENERAPAN DATA MINING UNTUK MENGETAHUI FAKTOR-FAKTOR YANG MEMPENGARUHI KELAHIRAN BAYI MENGGUNAKAN ASSOCIATION RULES (Studi Kasus: RSKIA BHAKTI IBU YOGYAKARTA)

¹Hasni Griya Anatasia(09018200), ²Tedy Setiadi (0407016801)

^{1,2}Program Studi Teknik Informatika
Universitas Ahmad Dahlan
Prof. Dr. Soepomo, S.H., Janturan, Umbulharjo, Yogyakarta 55164
²Email: tedy.setiadi@tif.uad.ac.id

ABSTRAK

Tingginya laju pertumbuhan penduduk disebabkan masih lebih tingginya tingkat kelahiran bayi dibanding tingkat kematian. Pertumbuhan penduduk yang tinggi menyebabkan hasil pembangunan kurang bisa dirasakan masyarakat. Pernikahan dini banyak terjadi di kalangan masyarakat miskin. Tinggi rendahnya kelahiran bayi dalam suatu kelompok penduduk tergantung dari faktor-faktor yang mempengaruhi kelahiran bayi. Faktor-faktor yang mempengaruhi kelahiran bayi yang digunakan dalam penelitian ini yaitu usia ibu, tingkat pendidikan ibu, status pekerjaan ibu, dan urutan anak lahir.

Subjek penelitian yang akan dibahas pada penelitian ini adalah di RSKIA BHAKTI IBU Yogyakarta yang membahas mengenai penerapan data mining untuk mengetahui faktor-faktor yang mempengaruhi kelahiran bayi menggunakan Association Rules dengan algoritma apriori. Langkah awal yang dilakukan dalam penelitian ini adalah melakukan pengumpulan data menggunakan metode kepustakaan, observasi, wawancara dan dokumentasi. Kemudian dilakukan analisis kebutuhan sistem, perancangan proses, perancangan antar muka, pengembangan sistem data mining, pembuatan sistem, pengujian menggunakan black box test dan alfa test, evaluasi pola yang ditemukan dan presentasi pengetahuan.

Penelitian ini akan menghasilkan sebuah aplikasi untuk menampilkan informasi dari proses data mining yaitu faktor yang mempengaruhi kelahiran bayi khususnya faktor dalam data rekam medis kelahiran bayi. Informasi yang akan ditampilkan berupa nilai support dan confidence. Semakin tinggi nilai support dan confidence maka semakin kuat nilai hubungan antar item. Hasil dari proses data mining ini dapat digunakan sebagai rekomendasi kebijakan baru bagi BKKBN untuk menekan angka kelahiran bayi yang tinggi.

Kata Kunci : *Faktor Kelahiran Bayi, Data Mining, Association Rules, algoritma apriori, Support dan Confidence.*

A. PENDAHULUAN

Masih cukup tingginya laju pertumbuhan penduduk disebabkan masih lebih tingginya tingkat kelahiran bayi dibanding tingkat kematian. Pertumbuhan penduduk yang tinggi menyebabkan hasil pembangunan kurang bisa dirasakan masyarakat. Oleh karena itu faktor-faktor penyebab tingginya kelahiran bayi masih perlu diperhatikan. Tinggi rendahnya kelahiran bayi dalam suatu kelompok penduduk tergantung dari faktor-faktor kelahiran bayi. Faktor-faktor kelahiran bayi yaitu jenis kelamin bayi, tingkat pendidikan, status pekerjaan wanita, tempat tinggal, agama, tanggal atau usia nikah, jumlah anak dan usia ibu. [1]

Salah satu Rumah Sakit Bersalin yang ada di kota Yogyakarta adalah RSKIA BHAKTI IBU. Setiap satu bulannya ada sekitar 40 sampai 50 bayi yang dilahirkan ke RSKIA BHAKTI IBU.

Untuk mengontrol laju pertumbuhan masyarakat, pemerintah Negara Indonesia membentuk sebuah lembaga bernama BKKBN (Badan Kependudukan dan Keluarga Berencana Nasional). Salah satu program dari pemerintah yang bertujuan menekan angka kelahiran dan meningkatkan kesejahteraan keluarga adalah Keluarga Berencana. Namun, masih banyak warga yang tidak mengikuti program KB. Hal tersebut mengakibatkan jumlah kelahiran bayi tidak terkendali.

Dari permasalahan diatas maka diperlukan adanya suatu pola untuk menentukan faktor yang mempengaruhi kelahiran bayi. Dengan menggunakan metode yang ada di dalam data mining yang digunakan adalah aturan asosiasi (association rules) yang merupakan suatu bentuk algoritma dalam data mining yang memberikan informasi hubungan antar item data di database. Metode *association rules* akan diimplementasikan dalam suatu sistem. Sistem yang akan dibuat mampu menampilkan informasi faktor-faktor yang mempengaruhi kelahiran bayi khususnya faktor dalam data rekam medis kelahiran bayi. Sistem yang akan dibuat dapat membantu persediaan obat KB di Rumah Sakit. Selain itu dari BKKBN akan memunculkan adanya rekomendasi baru untuk menekan angka kelahiran dan menciptakan keluarga sejahtera.

B. KAJIAN PUSTAKA

Data mining adalah teknologi baru yang sangat berguna untuk membantu perusahaan maupun instansi menemukan informasi yang sangat penting dari gudang data mereka.[2]

Data mining, sering juga disebut knowledge discovery in database (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keterkaitan, pola atau hubungan dalam set data berukuran besar. [3]

Suatu bentuk algoritma dalam data mining yang memberikan informasi hubungan antara item data di dalam data base adalah Association rules. Association rules (aturan asosiasi) berkenaan dengan studi tentang “apa bersa ma apa”. Aturan asosiasi ingin memberikan informasi dalam bentuk hubungan “if-then ” atau “jika-maka”.

Penting tidaknya suatu aturan assosiatif dapat diketahui dengan dua parameter, support (nilai penunjang) dan confidence (nilai kepastian). Support adalah rasio antara jumlah transaksi yang memuat antecedent dan consequent dengan jumlah transaksi. Confidence

adalah rasio antara jumlah transaksi yang meliputi semua item dalam antecedent dan consequent dengan jumlah transaksi yang meliputi semua item dalam antecedent.

$$S = \frac{\sum(\text{antecedent} + \text{consequent})}{\sum(\text{antecedent})} \quad (1)$$

Keterangan :

S = Support

$(\text{antecedent} + \text{consequent})$ = Jumlah transaksi yang mengandung *antecedent* dan *consequent*

$\sum(\text{antecedent})$ = Jumlah transaksi

$$C = \frac{\sum(\text{antecedent} + \text{consequent})}{\sum(\text{antecedent})} \quad (2)$$

Keterangan :

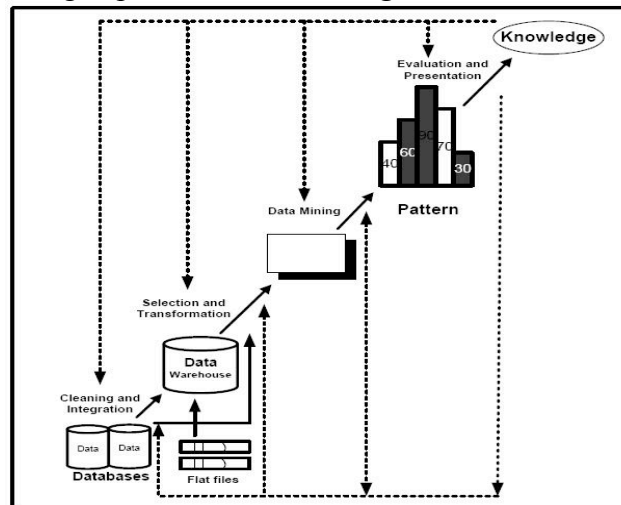
C = Confidence

$(\text{antecedent} + \text{consequent})$ = Jumlah transaksi yang mengandung *antecedent* dan *consequent*

$\sum(\text{antecedent})$ = Jumlah transaksi yang mengandung *antecedent*

1. Tahapan Data Mining

Sebagai suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap yang diilustrasikan di Gambar 2.2. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantara *knowledge base*.



Gambar 2.2: Tahap-tahap *data mining*

[4] Tahap-tahap *data mining* ada enam yaitu :

a. Pembersihan data (*data cleaning*)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Data-data yang tidak relevan itu juga lebih baik dibuang.

b. Integrasi data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya.

c. Seleksi Data (*Data Selection*)

Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

d. Transformasi data (*Data Transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. [4]

e. Proses *mining*

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

f. Evaluasi pola (*pattern evaluation*)

Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan, dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. [4]

Pada penelitian ini evaluasi dilakukan dengan mencari kuatnya hubungan antara *antecedent* dengan *consequent* yang disebut dengan *lift*. *Antecedent* adalah kalimat yang menerangkan jika dan *consequent* adalah kalimat yang menerangkan maka. Tidak semua aturan asosiasi yang ditemukan dalam penelitian ini diinterpretasi. Yang diinterpretasi adalah aturan-aturan yang memiliki nilai *lift* yang tinggi (alasan obyektif) dan aturan yang memiliki relevansi dengan kebutuhan (alasan subyektif).

Lift merupakan sebuah angka ratio yang menunjukkan berapa banyak kemungkinan menemukan sebuah item misal jumlah anak muncul bersama dengan item lainnya misal usia ibu, pendidikan ibu, dan pendidikan ibu dibandingkan dengan seluruh kejadian adanya item yang terpenuhi. *Lift* menunjukkan adanya tingkat kekuatan aturan atas kejadian acak dari *antecedent* dan *consequent* berdasarkan pada *supportnya* masing-masing. Hal ini akan memberikan informasi tentang perbaikan dan peningkatan probabilitas dari *consequent* berdasarkan *antecedent*. *Lift* didefinisikan sebagai berikut:

$$Lift = Confidence / Expected Confidence$$

Dimana

Expected Confidence = (Jumlah Transaksi memiliki transaksi) item consequent) / (Total jumlah

Atau dengan cara:

$$Pr(A|C)$$

Lift =

$$\frac{Pr(A|C)}{Pr(EC)} \dots\dots\dots (3)$$

Keterangan :

Lift = kuat lemah hubungan antara antecedent dan consequent

$Pr(A|C)$ = Probabilitas antecedent maka consequent (probabilitas confident)

$Pr(EC)$ = Expected Confident

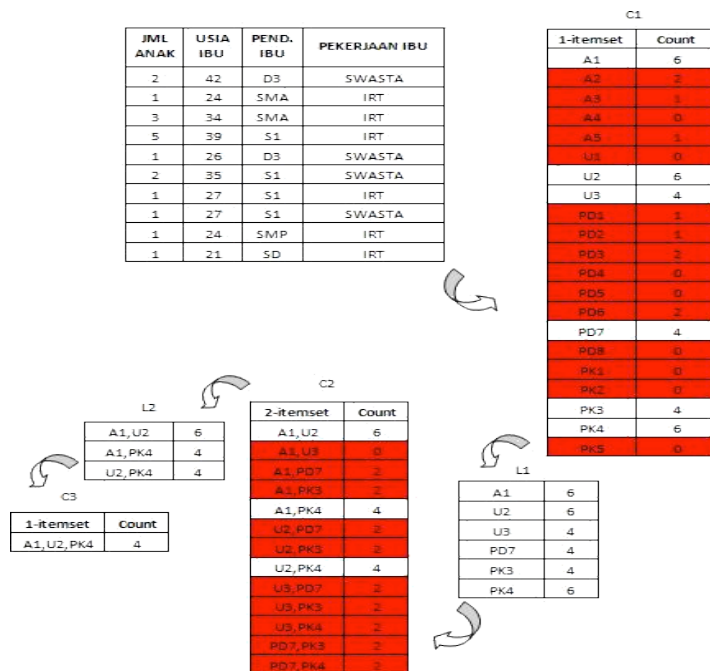
Ketika *lift* sama dengan 1 maka A dan C adalah independen karena $Pr(C|A)=Pr(EC)$. Ketetapan *lift* ratio adalah apabila hasil perhitungan berada di bawah 1 maka item-item tersebut tidak menunjukkan adanya saling keterkaitan antara antecedent dengan consequent. [5]

g. Presentasi pengetahuan (Knowledge Presentation)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. [4]

2. Bentuk Algoritma Apriori

Berikut ini merupakan bentuk algoritma apriori dari kasus yang ada



Gambar 2.3: Bentuk algoritma apriori

Berikut tabel pencarian support dan confidence setelah aturan asosiasi didapatkan. Tabel 2.9: Association Rules, Nilai Support dan Confidence

Item	Keterangan	Support	Confidence
	Jika jumlah anak 1 maka Usia Ibu diantara	$(6/10)*100$ %	$(6/6)*100$ %

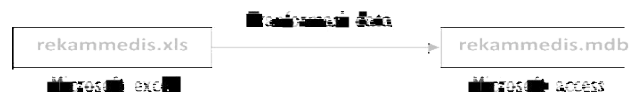
A1,U2	20 sampai 30	= 60%	= 100 %
	Jika usia ibu diantara 20 sampai 30 maka jumlah anak 1		$(6/6)*100$ % = 100 %
A1,PK4	Jika jumlah anak 1 maka pekerjaan ibu IRT	$(4/10)*100$ = 40 %	% $(4/6)*100$ = 66,67 %
	Jika pekerjaan ibu IRT maka jumlah anak 1		$(4/6)*100$ = 66,67 %
U2,PK4	Jika usia ibu diantara 20 sampai 30 maka pekerjaan ibu IRT	$(4/10)*100$ = 40 %	% $(4/6)*100$ = 66.67 %
	Jika pekerjaan ibu IRT maka usia ibu diantara 20 sampai 30		$(4/6)*100$ = 66.67 %
A1,U2,PK4	Jika jumlah anak 1 dan usia ibu antara 20 sampai 30 maka pekerjaan ibu IRT	$(4/10)*100$ = 40 %	% $(4/6)*100$ = 66,67 %
	Jika jumlah anak 1 dan pekerjaan ibu IRT maka usia ibu antara 20 sampai 30		$(4/4)*100$ % = 100 %
	Jika usia ibu antara 20 sampai 30 dan pekerjaan ibu IRT maka jumlah anak 1		$(4/4)*100$ % = 100 %

C. HASIL DAN PEMBAHASAN

Pengembangan Sistem

1. Transformasi Data

Merupakan tahap untuk mengubah data ke dalam format yang sesuai agar dapat diminingkan. Format data yang awalnya microsoft excel dengan format (.xls) ditransformasikan terlebih dahulu menjadi bentuk microsoft access dengan format (.mdb).



Gambar 4.4: Transformasi Data

2. Seleksi Data

Merupakan tahap untuk memutuskan data mana yang akan digunakan pada tahap data mining. Atribut yang dibutuhkan sebagai input untuk algoritma yang digunakan pada tahap data mining yaitu, jml_anak, usia_ibu, pend_ibu dan pek_ibu.

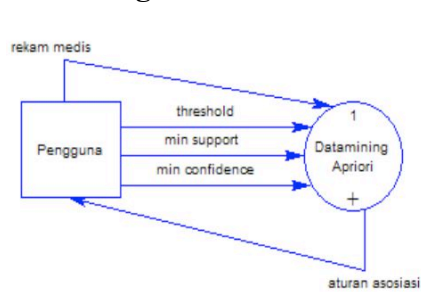
3. Cleaning Data

Merupakan tahap untuk membersihkan data yang mengandung noise, data yang mempunyai nilai kosong, dan data yang tidak konsisten untuk dihilangkan. Data tersebut dihilangkan dan tidak diikuti pada tahap selanjutnya. Dari tahap ini akan dilakukan proses mining. Data yang mengandung noise dari 851 data ada 47 noise dan data rekam medis yang digunakan untuk proses mining ada 804 data.

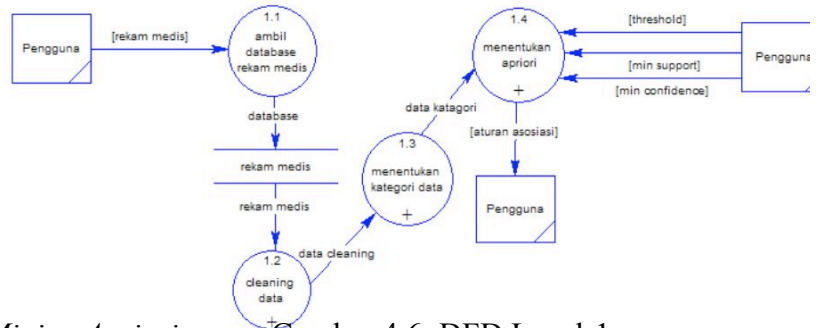
4. Aplikasi Teknik Data Mining

Kegiatan awal yang dilakukan pada tahap ini adalah perancangan proses dan merancang antarmuka (*user interface*).

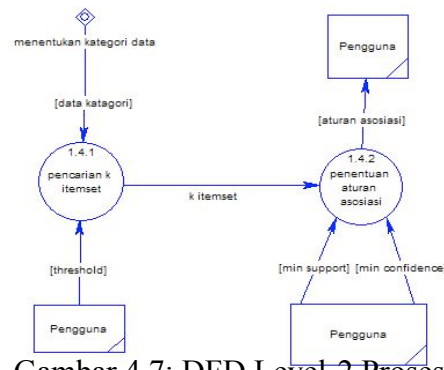
a) Perancangan Proses



Gambar 4.5: DFD Level-0 Sistem *Data Mining Apriori*



Gambar 4.6: DFD Level-1

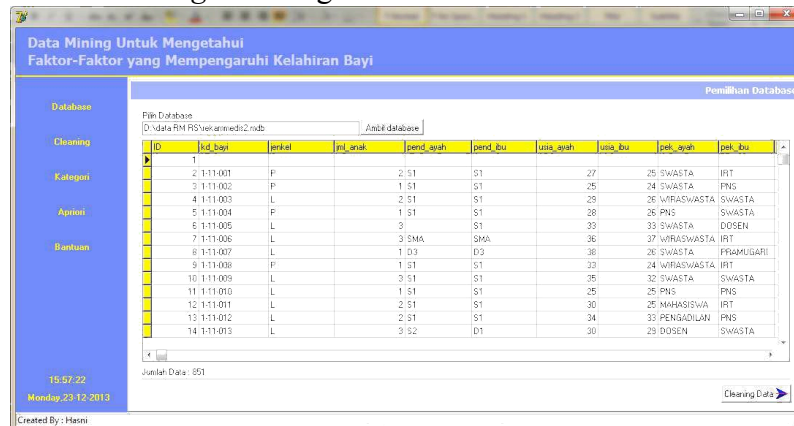


Gambar 4.7: DFD Level-2 Proses 1.4

b) Pembuatan Sistem

1) **Form Ambil Database (Pemilihan Data Base)**

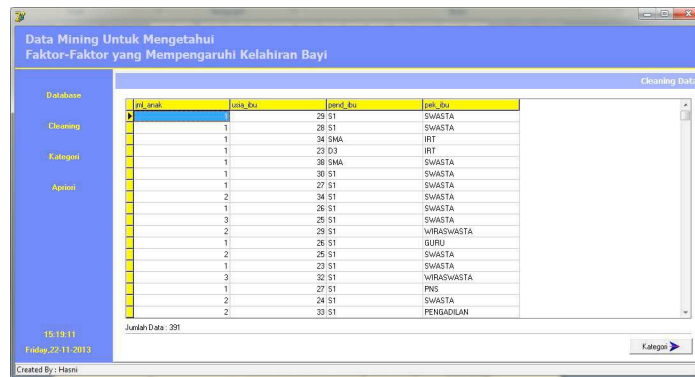
Form ambil database digunakan oleh pengguna untuk mengam bil data yang akan *diminingkan* dengan aturan asosiasi.



Gambar 4.13: Tampilan *Form* Ambil Database

2) **Form *Cleaning Data***

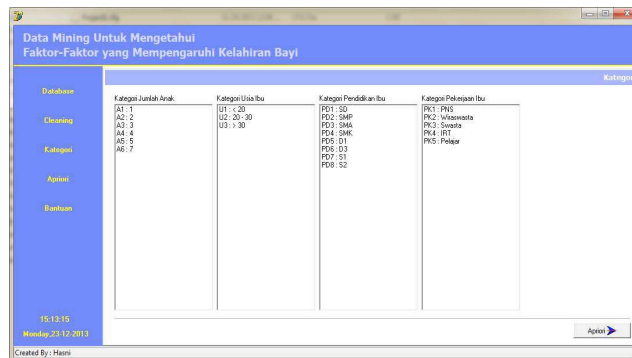
Form cleaning data digunakan oleh sistem untuk membersihkan data base yang tidak digunakan lagi.



Gambar 4.14: Tampilan Form Cleaning Data

3) Form Menentukan Kategori Data

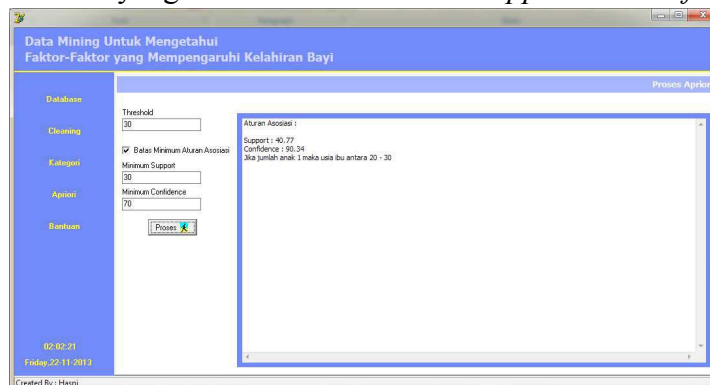
Form ini digunakan oleh sistem untuk menampilkan data yang dikodekan dan dikategorikan sesuai dengan data base rekam medis ada.



Gambar 4.15: Tampilan Form Kategori Data

4) Form Proses Apriori

Form ini digunakan untuk menginputkan nilai *threshold*, *minimum support* dan *minimum confidence*. Hasil output berupa aturan asosiasi yang terbentuk beserta nilai *support* dan *confidencinya*.



Gambar 4.16: Tampilan Form Proses Apriori

5. Evaluasi Pola yang Ditemukan

Pada penelitian ini evaluasi dilakukan dengan mencari kuatnya hubungan antara antecedent dengan consequent yang disebut dengan Lift. Berikut tabel percobaan:

Tabel 4.16: Data Percobaan

Percobaan	Threshold	Minimum Support	Minimum Confidence
1.	10	20%	70%
2.	20	30%	80%
3.	30	30%	90%
4.	40	40%	90%
5.	50	50%	100%

Tabel 4.17: Hasil Percobaan

Percobaan	Threshold	Minimum Support	Minimum Confidence	Aturan Asosiasi
1.	10	20%	70%	Support : 46,14% Confidence: 94,4% Jika jumlah anak 1 maka usia ibu antara 20 – 30
				Support : 27,99% Confidence: 80,65% Jika pekerjaan ibu IRT maka usia ibu antara 20 – 30
				Support : 27,86% Confidence: 81,45% Jika pekerjaan ibu swasta maka usia ibu antara 20 – 30
2.	20	30%	80%	Support : 46,14% Confidence: 94,4% Jika jumlah anak 1 maka usia ibu antara 20 – 30
3.	30	30%	90%	Support : 46,14% Confidence: 94,4% Jika jumlah anak 1 maka usia ibu antara 20 – 30
4.	40	40%	90%	Support : 46,14% Confidence: 94,4%
5.	50	50%	100%	Jika jumlah anak 1 maka usia ibu antara 20 – 30 -

Hasil dari aturan asosiasi pada percobaan 1 menghasilkan aturan asosiasi berjumlah 3 yaitu:

Aturan pertama adalah nilai support 46,14% dan nilai confidence 94,4% aturan asosiasi yang dihasilkan jika jumlah anak 1 maka usia ibu antara 20 - 30. Sehingga nilai liftnya adalah: $Lift = 0,944 / (680/804) = 1,20$

Aturan kedua adalah nilai support 27,99% dan nilai confidence 80,65% aturan asosiasi yang dihasilkan jika pekerjaan ibu IRT maka usia ibu antara 20 - 30. Sehingga nilai liftnya adalah: $Lift = 0,8065 / (680/804) = 1,03$

Aturan ketiga adalah nilai support 27,86% dan nilai confidence 81,45% aturan asosiasi yang dihasilkan jika pekerjaan ibu swasta maka usia ibu antara 20 - 30. Sehingga nilai liftnya adalah: $Lift = 0,8145 / (680/804) = 1,04$

Hasil dari aturan asosiasi pada percobaan 2, 3 dan 4 menghasilkan aturan asosiasi berjumlah 1 yaitu:

Aturan asosiasi yang terbentuk adalah nilai support 46,14% dan nilai confidence 94,4% aturan asosiasi yang dihasilkan jika jumlah anak 1 maka usia ibu antara 20 - 30. Sehingga nilai liftnya adalah: $Lift = 0,944 / (680/804) = 1,20$

Hasil dari perhitungan nilai lift untuk semua aturan asosiasi menunjukkan bahwa semua nilai lift diatas 1 yang menunjukkan bahwa antara antecedence yaitu jumlah anak 1 dengan consequent yaitu usia ibu antara 20 - 30 ter bukti memiliki hubungan atau keterkaitan antar item.

Evaluasi pola dengan melihat nilai batas minimum aturan asosiasi yang diberikan jika semakin tinggi maka aturan asosiasi yang dihasilkan untuk data rekam medis dengan jumlah 851 ternyata menghasilkan aturan asosiasi yang lebih akurat. Terlihat ketika nilai minimum support dan minimum confidence diberikan dengan nilai rendah yaitu minimum support = 20% dan minimum confidence = 70% aturan asosiasi yang terbentuk ada 3 aturan, sedangkan ketika dinaikkan menjadi minimum support = 40% dan minimum confidence = 90% aturan asosiasi yang terbentuk ada 1 aturan. Hal ini menunjukkan bahwa semakin tinggi batas minimum aturan asosiasi yang diberikan maka hasil aturan asosiasi lebih akurat. Nilai support dan nilai confidence juga akan semakin akurat.

6. Presentasi Pengetahuan

mengikuti pelayanan klinik KB.

D. KESIMPULAN

Berdasarkan penelitian dan pembahasan, maka dapat disimpulkan hal-hal sebagai berikut: Aturan asosiasi dengan nilai support dan confidence terbaik yaitu:

Nilai support 46,14% dan nilai confidence 94,4% aturan asosiasi yang dihasilkan jika jumlah anak 1 maka usia ibu antara 20 - 30 yang artinya faktor yang mempengaruhi kelahiran bayi yaitu jumlah anak 1 mempunyai kemungkinan 94,4% juga dapat berpengaruh untuk faktor usia ibu antara 20 - 30 terhadap kelahiran bayi. Aturan ini mewakili 46,14% dari database.

Maka analisis yang didapat adalah faktor kelahiran bayi jumlah anak 1 seringkali dilahirkan oleh seorang ibu ketika usianya antara 20 sampai 30 tahun. Seorang ibu dengan pekerjaan IRT dan swasta juga memiliki peluang untuk melahirkan anak lebih banyak pada usia ibu antara 20 sampai 30 tahun. Dengan aturan asosiasi ini dapat dijadikan sebagai dasar bahwa ibu berusia antara 20 - 30 tahun sudah melahirkan anak dengan jumlah 1. Melihat aturan tersebut berusia 20 - 30 tahun kemungkinan ibu untuk melahirkan anak lebih dari 1, peluang sangat besar. Maka untuk mengurangi tingkat kelahiran berusia 20 - 30 tahun, setiap ibu yang telah melahirkan anak ke-1 harus

1. Telah dibuat aplikasi *data mining* yang mampu digunakan untuk menampilkan informasi hubungan antar item yaitu hubungan faktor-faktor yang mempengaruhi kelahiran bayi berdasarkan database rekam medis.
2. Dari hasil uji coba menghasilkan aturan asosiasi yang dapat dianalisis yaitu faktor kelahiran bayi jumlah anak 1 sering kali dilahirkan oleh seorang ibu ketika usianya antara 20 sampai 30 tahun, seorang ibu dengan pekerjaan IRT dan swasta juga memiliki peluang untuk melahirkan anak lebih banyak pada usia ibu antara 20 sampai 30 tahun.
3. Hasil dari perhitungan nilai *lift* untuk semua aturan asosiasi menunjukkan bahwa semua nilai *lift* diatas 1 yang menunjukkan bahwa antara *antecedence* dengan *consequent* terbukti memiliki hubungan atau keterkaitan antar item, sehingga hasil aturan asosiasi yang telah memenuhi batas parameter dapat diinterpretasi.

DAFTAR PUSTAKA

- [1] Mantra, Ida Bagoes. 2000. *Demografi Umum*. Penerbit Pustaka Pelajar. Yogyakarta
- [2] Kusrini dan Emha Taufiq Luthfi. 2009. *Algoritma Data Mining*. Penerbit Andi. Yogyakarta.
- [3] Santosa, Budi. 2007. *Data mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu. Yogyakarta.
- [4] Han, Jiawei; Kamber, Micheline. 2006. *Data Mining: Concepts and Techniques*. San Fransisco: Morgan Kaufmann.
- [5] Amiruddin. 2007. *Penerapan Association Rules Mining Pada Data Nomor Unik Pendidik dan Tenaga Kependidikan untuk Menenmpuh Pola Sertifikasi Guru*. Teknik Elektro, FTI ITS, Surabaya.
- [6] Badaruddin, Lalu Saptiadi. 2012. *Penerapan Data Mining dengan Algoritma Asosiasi untuk Mencari Hubungan Antar Produk dari Pola Transaksi Penjualan pada Swalayan Pamela*. Teknik Informatika, FTI Universitas Ahmad Dahlan, Yogyakarta.
- [7] Bramer, Max, 2007, “ *Principles of Data Mining* ” , Springer, London.
- [8] Huda, Nugon Masykur. 2010. *Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa (Studi Kasus Di Fakultas MIPA Universitas Diponegoro)*. Teknik Informatika, FMIPA Universitas Diponegoro, Semarang.
- [9] Pramudiono, I., 2007, *Algoritma Apriori*, <http://datamining.japati.net/cgibin/indodm.cgi?bacaarsip&1172210143>
Diakses pada tanggal 1 November 2012 jam 20.00
- [10] Pritahastari, Miranthi. 2012. *Penerapan Teknik Aturan Asosiasi dalam data Mining untuk Pencarian Pola Peminjaman Buku Perpustakaan. (Studi Kasus: Perpustakaan FMIPA UGM)*. Ilmu Komputer, FMIPA Universitas Gadjah Mada, Yogyakarta.