

**PENGUNAAN DATA MINING UNTUK MEMPREDIKSI
CHURN PASIEN DI RUMAH SAKIT
(STUDI KASUS: RUMAH SAKIT X)**

Yaten Purwono

Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI

Jl. Ir. H. Juanda 96 Bandung 40132

ABSTRAK

Rumah sakit “X” merupakan salah satu rumah sakit yang berada di kota Bandung. Di tengah persaingan yang ada, rumah sakit ini mengalami penurunan jumlah persalinan karena pindahnya pasien dari layanan rumah sakit ke kompetitor lainnya (*Churn*). Penggunaan data mining dengan metodologi CRISP-DM (*Cross Industry Standard Process for Data Mining*) merupakan salah satu solusi untuk mendapatkan pola dari pasien yang tidak melahirkan di rumah sakit ini.

Model prediksi yang dihasilkan oleh algoritma CART (*Classification and Regression Trees*) ini mempunyai tingkat akurasi model sebesar 73,40 % dengan akurasi memprediksi pasien tidak melahirkan di rumah sakit sebesar 81,33%. *Knowledge* yang dihasilkan oleh pemodelan ini dapat dijadikan sebagai *rule* untuk memprediksi pasien yang telah melakukan kontrol kehamilan.

Kata-kata Kunci: *Data mining, CRISP-DM, Decision Tree, Pasien Churn*

1. PENDAHULUAN

Bermunculannya rumah sakit baru yang memfokuskan terhadap pelayanan ibu dan anak dengan memberikan tarif yang kompetitif serta sarana yang baik telah memberikan alternatif terhadap pasien untuk menentukan pilihannya. Seiring dengan bertambahnya jumlah rumah sakit, dapat menimbulkan persaingan antar rumah sakit di dalam mempertahankan jumlah pasiennya. Persaingan ini berdampak terhadap jumlah persalinan yang dilakukan di RS. “X” yang mengalami penurunan dari tahun ke tahun.

Salah satu cara untuk meningkatkan jumlah persalinan adalah dengan mempertahankan pasien untuk tidak berpindah ke rumah sakit lain atau *Churn*.

Mempertahankan agar pasien tidak *Churn* sangat penting seperti yang dikemukakan oleh (Govindaraju, Simatupang dan Ari Samadhi, 2008:2) bahwa dibutuhkan biaya lima kali lebih besar untuk mendapatkan pelanggan baru dibanding biaya untuk mempertahankan pelanggan lama.

Oleh karena itu, dilakukan penelitian untuk memprediksi pasien yang tidak melakukan persalinan di rumah sakit “X” tersebut. Salah satu model yang digunakan adalah dengan *data mining* sehingga dapat diketahui pola pasien yang tidak melakukan proses persalinannya di rumah sakit.

Permasalahan yang menjadi fokus penelitian adalah:

1. Perlu adanya suatu model untuk dapat menganalisa pasien mana yang paling besar kemungkinannya yang tidak akan melahirkan di rumah sakit.
2. Masih lambatnya penanganan terhadap pasien yang kemungkinan tidak akan melahirkan di rumah sakit dikarenakan belum ada informasi pasien-pasien tersebut ke bagian *marketing*.
3. Belum adanya segmentasi pasien yang akan melahirkan di rumah sakit dan pasien yang tidak melahirkan di rumah sakit sehingga *treatment* terhadap pasien belum terarah.

Tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Merancang suatu model untuk dapat memprediksi pasien-pasien yang tidak melakukan persalinan di rumah sakit.
2. Memberikan informasi mengenai daftar pasien yang tidak akan melakukan persalinan di rumah sakit.
3. Mengoptimalkan pemanfaatan historis data yang ada di rumah sakit.

Untuk penelitian ini penulis membatasi ruang lingkup sebagai berikut:

1. Pasien yang melakukan kontrol dan berobat kehamilan di rawat jalan khususnya di klinik Obgin dan Klinik Kesehatan Keluarga.
2. Jenis pasien yang akan diteliti adalah pasien asuransi dan pasien umum.
3. Penelitian ini dilakukan dengan menggunakan basis data (*database*) dari rumah sakit ini dengan periode 5 tahun, yaitu tahun 2005 sampai dengan tahun 2009.

Metodologi penelitian yang digunakan adalah CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Untuk algoritma *data mining* yang diambil adalah algoritma klasifikasi dengan menggunakan metode CART (*Classification and Regression Trees*).

2. TINJAUAN PUSTAKA

2.1. DATA MINING

Banyak orang menggunakan istilah *data mining* dan *knowledge discovery in databases* (KDD) untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu kumpulan data yang besar. Tetapi kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Salah satu tahapan dalam proses KDD adalah *data mining*. Sedangkan pendapat lain mengungkapkan:

Data mining is a term that describes different techniques used in a domain of machine learning, statistical analysis, modeling techniques and data base technologies that can be used in different industries. With a combination of these techniques, it is possible to find different kinds of structures and relations in the data, as well as to derive rules and models that enable prediction and decision making in new situations. (Pejic Bach dan Cosic:2008:1)

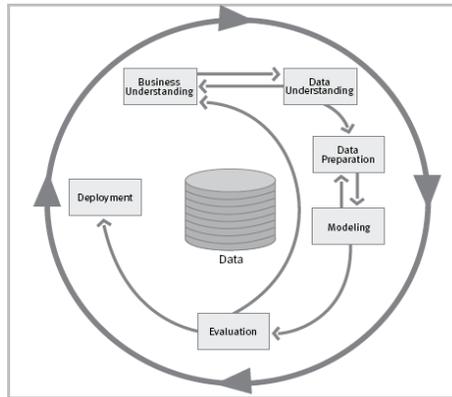
Dengan *data mining* maka akan didapat suatu pola yang nantinya dapat menjadi *knowledge* yang bermanfaat. *Data mining* sekarang sudah dipergunakan untuk mengambil *knowledge* dari basis data (*database*) organisasi.

2.2. CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM)

CRISP-DM merupakan suatu metodologi yang merupakan suatu standar *non-proprietary* dan *free* yang dapat digunakan secara umum untuk *problem solving* strategi bisnis dan unit penelitian. Metodologi ini mempunyai 6 fase yang meliputi (Chapman, Clinton, Kerber, Khabaza, Reinartz, Shearer dan Wirth:2000,10) dan dapat dilihat pada gambar 1.

1. *Bussines Understanding*, memfokuskan terhadap pemahaman dari tujuan organisasi, memberikan solusi terhadap permasalahan yang telah didefinisikan dan mendesain rencana untuk mencapai tujuan.
2. *Data Understanding*, pengambilan awal data dan memprosesnya sehingga dapat mengetahui makna dari suatu data.
3. *Data Preparation*, merupakan aktivitas yang diperlukan untuk persiapan data menjadi *dataset* yang digunakan sebagai dasar untuk pembentukan model.
4. *Modeling*, melakukan pemodelan data dengan model yang berbeda-beda seperti CART, CHAID, *Neural Network* atau model yang lainnya. Terkadang suatu model memerlukan kebutuhan spesifik dalam suatu data, apabila hal tersebut terjadi maka fase ini akan balik ke fase sebelumnya yaitu fase *Data Preparation*.

5. *Evaluation*, melakukan evaluasi dari rangkaian proses untuk menjamin bahwa model yang dibentuk dapat mencapai tujuan organisasi.
6. *Deployment*, pengimplementasian dari proses ini mulai dari *bussines understanding* sampai *evaluation*.



Gambar 1. Fase dari CRISP-DM
(Chapman, Clinton, Kerber, Khabaza, Reinartz, Shearer dan Wirth: 2000:10)

2.3. *Customer Churn*

Seperti dalam industri jasa, di rumah sakit pun sering terjadi pasien yang berpindah ke rumah sakit lain, dikatakan sebagai *Churn*. Hadden, Tiwaria, Roy dan Ruta mengemukakan sebagai berikut:

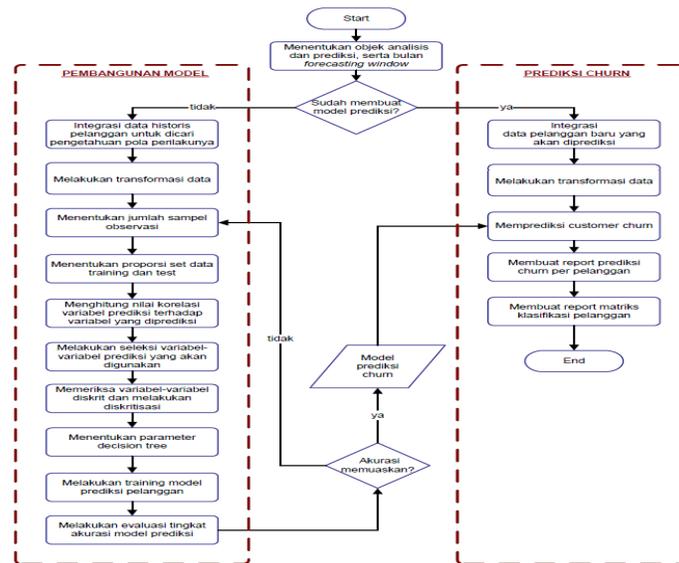
Churn management is the term that has been adopted to define customer turnover. More specifically, churn management is the concept of identifying those customers who are intending to move their custom to a competing service provider. Once identified, these customers can be targeted with proactive marketing campaigns for retention efforts. Customer retention becomes an important aspect of everyday business strategy. (Hadden, Tiwaria, Roy dan Ruta , 2005:2).

2.4. *Framework Customer Churn*

Framework ini digunakan untuk memprediksi *churn* pada industri jasa rumah sakit. Model ini membahas secara lengkap setiap tahapan dari proses pembentukan model *churn* (Govindaraju, Simatupang dan Ari Samadhi, 2008:5). *Framework* ini membagi menjadi dua proses besar, yaitu:

- a. Pembangunan model, digunakan untuk memproses data–data yang nantinya akan digunakan untuk pembangunan pemodelan prediksi *churn*.
- b. Prediksi *Churn* merupakan proses untuk menghasilkan *output* yang akan digunakan untuk memprediksi pelanggan tersebut *churn* atau *non churn*. Ada dua

proses besar dari pembangunan model dan prediksi *churn*, seperti pada gambar berikut ini:



Gambar 2. Proses analisis dan prediksi *churn* (Govindaraju, Simatupang dan Ari Samadhi,2008:5).

2.5. Klasifikasi

Algoritma Klasifikasi merupakan algoritma yang digunakan untuk mengklasifikasikan sesuatu berdasarkan variabel-variabel yang sudah ditentukan sebelumnya. Diperlukan dua jenis variabel yaitu variabel prediktor untuk menentukan variabel tujuan dan variabel tujuan yang ditentukan berdasarkan variabel prediktor (Susanto dan Suryadi, 2010:53).

Variabel yang digunakan dalam penelitian ini adalah jumlah kontrol persalinan, umur pasien, jarak rumah ke rumah sakit, pekerjaan suami, penanggung persalinan dan variabel-variabel lainnya yang akan dijelaskan pada bab berikutnya. Sedangkan variabel-variabel lainnya yang mempengaruhi terhadap persalinan ini adalah umur pasien, kebiasaan merokok dan penggunaan alkohol selama kehamilan, pendapatan keuangan, jumlah kehamilan, jumlah kontrol kehamilan dan historis penggunaan ultrasonography sebelum 20 minggu (Janssen , Lee, Page, Klein, Liston, Lee Shoo, 2009:379).

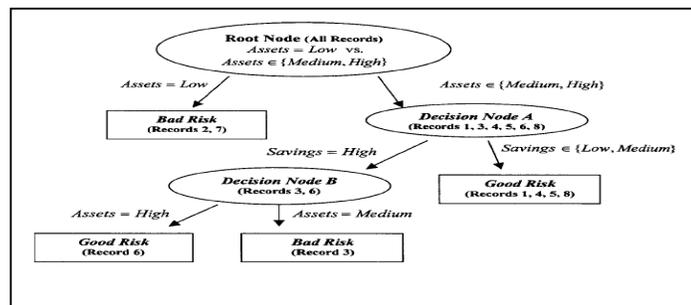
Salah satu model yang ada dalam klasifikasi ini adalah *decision trees*. Pejic Bach dan Cosic mengemukakan tentang *decision trees* ini sebagai berikut:

Decision trees are powerful and popular tools for classification and prediction. They are appealing because they are easily understood, as they can be graphically

presented as trees as well as in the form of rules (in English or in SQL). Most popular algorithms are CHAID (Chisquared automatic induction), CART and C4.5 (later version of ID3 algorithm) (Pejic Bach dan Cosic,2008:59)

2.6. CART (CLASSIFICATION AND REGRESSION TREE)

Algoritma CART merupakan salah satu algoritma yang masuk dalam *The Top Ten Algorithms in Data Mining* (Susanto dan Suryadi, 2010:57). CART mempergunakan model seperti struktur pohon, di mana setiap titik (*node*) akan memiliki dua cabang keputusan. Dalam algoritma ini ada dua noktah yaitu noktah keputusan (*elips*) dan noktah terminasi (persegi panjang) yang digunakan apabila suatu kondisi tidak bercabang lagi. Untuk lebih jelasnya dapat dilihat pada Gambar 3.



Gambar 3. Algoritma CART

Algoritma CART tersusun berdasarkan 4 (empat) langkah (Susanto dan Suryadi, 2010:58)

1. Tentukan dahulu data yang akan diproses seperti pada Gambar 4.

Customer	Savings	Assets	Income (\$1000s)	Credit Risk
1	Medium	High	75	Good
2	Low	Low	50	Bad
3	High	Medium	25	Bad
4	Medium	Medium	50	Good
5	Low	Medium	100	Good
6	High	High	25	Good
7	Low	Low	25	Bad
8	Medium	Medium	75	Good

Gambar 4. Data set yang akan diproses (Larose,2005:110)

2. Menilai kinerja keseluruhan calon cabang yang ada pada daftar calon cabang mutakhir, dengan mempergunakan persamaan:

$$\Phi(s|t) = 2P_L P_R \sum_{j=1}^{\# \text{ classes}} |P(j|t_L) - P(j|t_R)|$$

Dimana :

t_L = cabang calon kiri dari noktah keputusan t

t_R = cabang calon kanan dari noktah keputusan t

P_L = $\frac{\text{jumlah catatan pada calon cabang kiri } t_L}{\text{jumlah catatan pada latihan}}$

P_R = $\frac{\text{jumlah catatan pada calon cabang kanan } t_R}{\text{jumlah catatan pada latihan}}$

$P(j|t_L)$ = $\frac{\text{jumlah catatan berkategori } j \text{ pada calon cabang kiri } t_L}{\text{jumlah catatan pada noktah keputusan } t}$

$P(j|t_R)$ = $\frac{\text{jumlah catatan berkategori } j \text{ pada calon cabang kanan } t_R}{\text{jumlah catatan pada noktah keputusan } t}$

3. Menentukan calon cabang manakah yang akan dijadikan cabang dengan memilih calon cabang yang memiliki nilai kesesuaian terbesar.
4. Setelah itu gambarkan percabangan, jika tidak ada percabangan maka proses CART akan dihentikan. Namun jika ada percabangan maka akan dilanjutkan ke langkah nomor 2 setelah sebelumnya menghapus dahulu daftar calon cabang yang berhasil menjadi cabang.

3. PENGAMBILAN DATA DAN PENGOLAHAN DATA

Tahapan yang dilakukan merujuk ke metode CRISP-DM (*Cross Industry Standard Process for Data Mining*), yang meliputi enam fase seperti pada Gambar 5 di bawah ini:



Gambar 5. Urutan dari proses pengerjaan *data mining* (Chapmen:2000,13)

1. Fase *Bussines Understanding*

Membahas tujuan bisnis dari organisasi dan tujuan *data mining* itu sendiri yaitu meningkatkan jumlah persalinan yang dilakukan di rumah sakit. Kriteria-kriteria untuk penentuan keberhasilan, yaitu:

- a. Dapat memprediksi pasien-pasien yang tidak melakukan persalinan di RS. “X”.
- b. Dapat memberikan informasi mengenai daftar pasien yang tidak melakukan persalinan di rumah sakit.
- c. Dapat mengoptimalkan pemanfaatan data historis rumah sakit.

Asses Situation

Untuk menjalankan proses *data mining* ini diperlukan sumber daya yang akan dibutuhkan selama proses *data mining* berlangsung, terbagi menjadi empat komponen yaitu:

- a. Personel: sistem analis, *database administrator*, *programmer*, dan teknisi komputer.
- b. Data: pengaksesan terhadap *database*.
- c. *Platform*: Sistem *Operating Server* nya Window Server dan *Database*-nya SQL Server 2008.
- d. *Software* yang digunakan untuk proses *data mining* terdiri dari:
 - 1). Clementine 11.1 untuk melakukan proses pemodelan *data mining*
 - 2). Java, [VB.Net](#) dan Visual Foxpro 9 untuk melakukan pengaksesan data ke data operasional

Adapun tujuan *data mining* adalah:

1. Memprediksi seberapa besar jumlah pasien yang tidak melahirkan di rumah sakit.
2. Memberikan pengetahuan tentang *rule* kepada manajemen rumah sakit dari hasil proses *data mining*.
3. Memberikan informasi matrik klasifikasi pasien di mana pasien akan dipetakan berdasarkan *group* pasien dan *group* penanggung persalinan.

2. Fase Data Understanding**a. Pengumpulan data awal**

Beberapa *database* yang digunakan untuk melakukan proses *data mining* ini adalah:

- 1). *Database* Rekam Medis, digunakan untuk mengambil identitas pasien, kunjungan pasien, diagnosa utama pasien, informasi mengenai pasien tersebut dirawat serta ICD X.
- 2). *Database Billing*, digunakan untuk mengambil informasi penanggung persalinan.

b. Gambaran data yang dibutuhkan

Langkah berikutnya adalah menggambarkan secara detail mengenai informasi tabel yang digunakan dalam pemrosesan *data mining* ini, yakni:

- 1). Tabel pasien, untuk mengambil informasi yang berkaitan dengan identitas pasien.

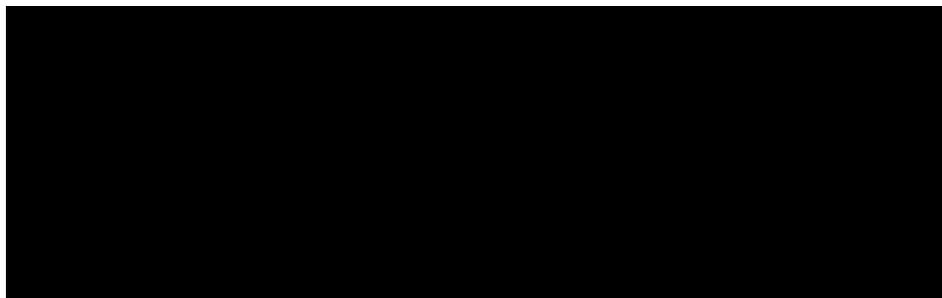
Penggunaan Data Mining Untuk Memprediksi Churn Pasien di Rumah Sakit

- 2). Tabel daftar untuk mengambil informasi yang berkaitan dengan kontrol pasien ke rumah sakit untuk pemeriksaan kehamilannya.
- 3). Tabel Kod_utama berisikan diagnosa-diagnosa utama pasien yang melakukan pemeriksaan di rumah sakit.
- 4). Tabel ICD10, tabel ini berisikan master standar WHO mengenai pengkodean penyakit-penyakit.
- 5). Tabel Kontraktor, digunakan sebagai referensi master penanggung kesehatan dari pasien yang ada di rumah sakit.
- 6). Tabel Idts2, digunakan sebagai tempat menyimpan data pasien yang dirawat di rumah sakit.

c. Explore data

Tahap ini menjelaskan langkah-langkah untuk mendapatkan data yang nantinya akan diproses *data mining* adalah sebagai berikut:

- 1). Pengambilan data diambil selama periode 5 tahun (2005-2009) di *database* rumah sakit.
- 2). Pengambilan data diagnosa utama pasien yaitu kode ICD10 Z34 (Pengawasan Kehamilan Normal) dan Z35 (Pengawasan Kehamilan Dengan Resiko Tinggi) dengan status diagnosa Baru. Total data dalam tahap ini adalah 181.526 *record*.
- 3). Data pasien-pasien dari karyawan atau personel rumah sakit dihilangkan karena ruang lingkup yang akan diteliti adalah pasien umum dan pasien yang ditanggung oleh perusahaan di mana pasien tersebut bekerja. Jumlah *record* untuk tahap ini adalah 140.056 *record*.
- 4). Melakukan proses *cross-tab*, yaitu data kunjungan perpasien yang sebelumnya berurutan ke bawah menjadi berurutan ke samping kanan. Total untuk *record* yang ada pada tahap proses ini adalah 10.929 *record*. Bentuk dari data ini seperti pada Gambar 6.



Gambar 6. Data yang sudah dibuat untuk pemodelan *data mining*

- 5). Langkah terakhir adalah melakukan pengecekan apakah pasien tersebut telah melakukan persalinan di rumah sakit ini atau tidak.

d. Verifikasi kualitas data

Untuk memastikan data yang digunakan terbebas dari kesalahan, maka yang perlu dilakukan adalah:

- 1). Memastikan bahwa tidak ada duplikasi-duplikasi selama proses pengambilan data berlangsung.
- 2). Pengambilan sumber data yang valid dari data rekam medis yang sudah divalidasi oleh bagian rekam medis.

3. Fase *Data Preparation*

Merupakan aktivitas-aktivitas yang diperlukan untuk persiapan data menjadi *dataset*.

a. *Select Data*

Memutuskan data yang akan digunakan untuk dianalisis serta mengkaitkannya dengan metode yang dipilih, yaitu metode klasifikasi. Menentukan variabel prediktor yang digunakan untuk menentukan variabel tujuan.

b. *Clean Data*

Tujuannya membersihkan data-data yang dianggap *error*. *Cleaning data* ini dikategorikan menjadi standarisasi penamaan, perubahan bentuk dan penghilangan nilai.

4. Fase *Modeling*

Pemilihan model klasifikasi beserta *software*nya. Penulis mempergunakan algoritma klasifikasi, yaitu CART (*Classification and Regression Trees*) dan *software* yang digunakan adalah Clementine 11.1. Total *record* yang siap untuk diproses model prediksi adalah 10.929 *record*, di mana 8.929 *record* untuk *training data* yang disebut *dataset* sedangkan 2.000 untuk *test data*.

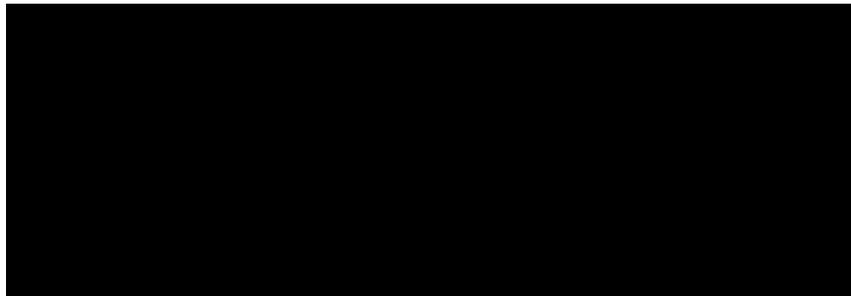
Pemodelan CART (*Classification dan Regression Trees*)

Melakukan *export data* ke dalam format excel kemudian dipergunakan *software* Clementine model untuk CART (*Classification dan Regression Trees*) untuk memproses pemodelan prediksi. Informasi mengenai variabel prediktor yang dianggap sangat penting tersebut dapat dilihat pada Gambar berikut:



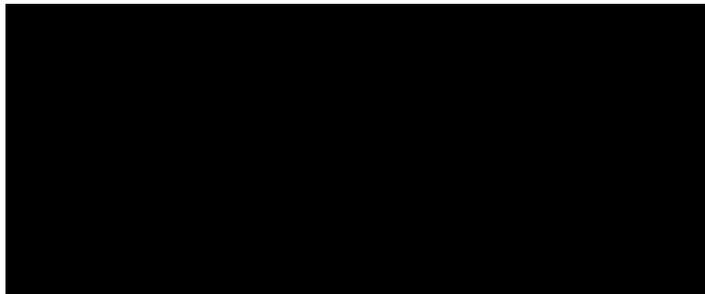
Gambar 7. Pemilihan variabel prediktor yang berpengaruh

Sejauh mana kontribusi tiap-tiap variabel prediktor mempengaruhi terhadap persalinan di rumah sakit, dapat dilihat pada gambar di bawah ini:



Gambar 8. Data audit variabel prediktor

Setelah proses pemilihan variabel prediktor maka dilakukan proses pemilihan model CART. Adapun skema dari proses pemodelan tersebut adalah:



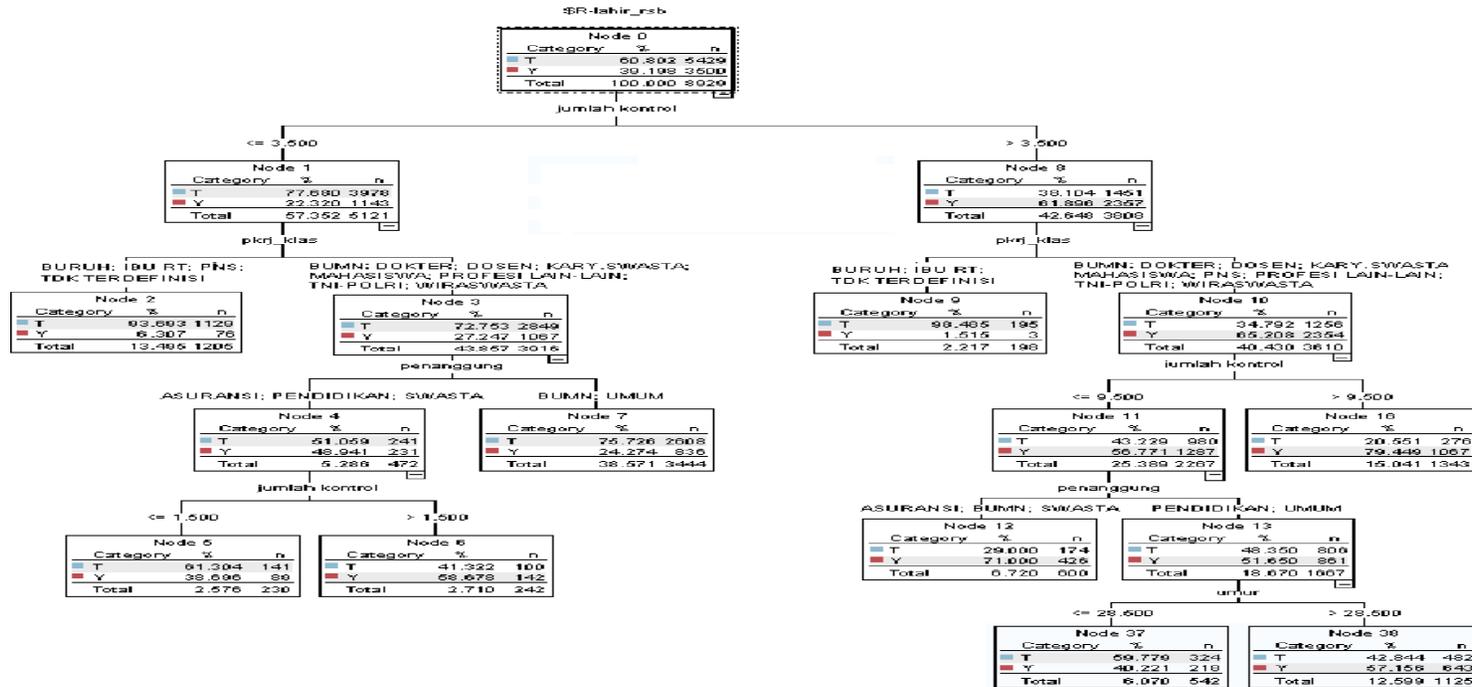
Gambar 9. Proses akhir dalam pemrosesan model

Proses akhir data pemodelan ini adalah terbentuknya model CART (*Classification And Regression Tree*), hasilnya tampak pada Gambar 10.

Kesimpulannya: dari 8.929 kehamilan pasien yang melakukan persalinan di RS. "X" adalah 39,198% (3.500 kehamilan pasien) sedangkan selebihnya 60,802% (5.429 kehamilan pasien) tidak melakukan persalinan di rumah sakit ini. Penulis

menggunakan metode algoritme CART (*Classification dan Regression Trees*) dan didapat suatu penulisan pemodelan dengan bentuk IF kondisi THEN status pasien melahirkan di rumah sakit. Setelah dianalisis dari pemodelan tersebut terdapat 4 (empat) variabel prediktor yang sangat berpengaruh terhadap keputusan pasien melakukan persalinan, yaitu:

- 1). Jumlah kontrol kehamilan ke rumah sakit, didasarkan atas node 1 dan node 8.
- 2). Pekerjaan suami pasien, didasarkan atas node 2,3,9 dan node 10.
- 3). Penanggung persalinan, didasarkan atas node 4,7,12 dan 13.
- 4). Umur pasien pada saat melahirkan, didasarkan atas node 37 dan node 38.



Gambar 10. Model prediksi Churn pasien persalinan

5. Fase *Evaluation*

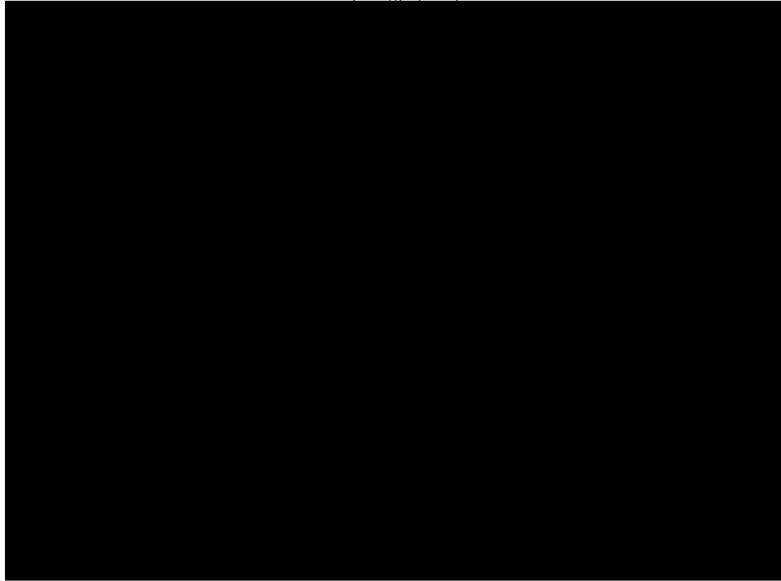
Setelah pembuatan model, didapat bahwa variabel prediktor untuk pasien yang kehamilannya bayi kembar, adanya gangguan selama kehamilannya dan jumlah kehamilan ternyata tidak berpengaruh terhadap model prediksi. Untuk menilai tingkat akurasi, maka dilakukan pengetesan model dengan cara mengambil data yang sudah dipersiapkan di mana dari data tersebut sudah ada hasil mengenai pasien yang akan diprediksi. Dari proses pengetesan tersebut didapat tingkat akurasi dari model ini adalah 73,40 % dengan distribusi akurasi seperti pada Tabel 1. Dengan tingkat akurasi untuk memprediksi kehamilan pasien tidak melahirkan di RS. “X” sebesar 81,33% dan memprediksi kehamilan pasien yang melahirkan di RS. “X” sebesar 61,77% membuktikan bahwa model ini cukup baik. Dengan akurasi tersebut maka *rule* ini mampu memprediksi kehamilan pasien yang tidak melahirkan di RS. “X” dengan tingkat akurasi 81,33%.

Tabel 1. Hasil Akurasi Model

	Prediksi Tidak Melahirkan	Prediksi Melahirkan	% akurasi
Aktual Tidak Melahirkan	967	222	81,33%
Aktual Melahirkan	310	501	61,77%

6. Fase *Deployment*

Diperlukan seorang sistem analis untuk mendesain dan membangun sistem yang dapat terintegrasi yang prosesnya diperlihatkan pada Gambar 11. Keterlibatan staff IT pada pengembangan model ini sangat perlu karena pembuatan pemodelan prediksi tidak statis tetapi dapat berubah-ubah tergantung data yang dijadikan pemodelan. Tidak hanya itu saja keterlibatan IT digunakan untuk *cleaning* dan transformasi data.



Gambar 11. Deployment model prediksi

4. PEMBAHASAN HASIL PEMODELAN

Berdasarkan 9 (sembilan) *rule* yang telah diperoleh dengan menggunakan algoritma CART, maka penulis akan menjelaskan pemanfaatan *rule* yang dapat dilakukan oleh rumah sakit. Adapun pemanfaatan tersebut adalah sebagai berikut:

1. Menawarkan tarif kontrol kehamilan dengan biaya yang terjangkau, mengadakan kegiatan yang berhubungan dengan kehamilan untuk menarik minat pasien melakukan kontrol kehamilan di rumah sakit. Penawaran ini dapat dilakukan sebagai *treatment* terhadap pasien-pasien yang terklasifikasi oleh *rule* 1 dan 5 seperti pada Tabel 2.
2. Membuat paket-paket persalinan yang kompetitif sehingga pasien akan segera mengetahui biaya yang harus disiapkan. Pemanfaatan pembuatan paket persalinan ini mengacu berdasarkan *rule* pada Tabel 2.

Tabel 2. Rule no 1 dan 5

No Rule	Kondisi	Melahirkan di rumah sakit
1	Jika Jumlah Kontrol ≤ 3 dan Pekerjaan=BURUH atau Pekerjaan=PNS atau Pekerjaan=IBU RT atau Pekerjaan=TDK TERDEFINISI Maka	TIDAK
5	Jika Jumlah Kontrol > 3 dan Pekerjaan=BURUH atau Pekerjaan=IBU RT atau Pekerjaan=TDK TERDEFINISI Maka	TIDAK

3. Berdasarkan *rule* yang terdapat pada Tabel 3, RS. “X” berupaya untuk dapat meningkatkan kerja sama dengan penanggung pasien ASURANSI, BUMN, PENDIDIKAN, SWASTA dan UMUM.
4. Meningkatkan program-program yang menunjang terhadap proses pra-persalinan supaya pasien datang ke rumah sakit. Contoh *hypnobirthing* dan *infertilitas*.
5. Melakukan pemantauan pasien-pasien yang sudah melakukan kontrol persalinan di rumah sakit untuk memberikan konfirmasi kontrol kehamilan ulang atau memberikan informasi mengenai layanan kesehatan di rumah sakit. Pemanfaatan point 3, 4 dan 5 di atas mengacu *rule* yang ada pada Tabel 3.

Tabel 3. Rule no 2,3, dan 4

No Rule	Kondisi	Melahirkan di rumah sakit
2	Jika Jumlah kontrol ≤ 3 dan Pekerjaan=(BUMN,DOKTER,DOSEN,KARY.SWASTA,MAHASI SWA,PROFESI LAINNYA,TNI-POLRI,WIRASWASTA) dan penanggung=(BUMN,UMUM) Maka	TIDAK
3	Jika Jumlah kontrol ≤ 3 dan Pekerjaan=(BUMN,DOKTER,DOSEN,KARY.SWASTA,MAHASISWA,PROFESI LAINNYA,TNI-POLRI,WIRASWASTA) dan penanggung=(ASURANSI,PENDIDIKAN,SWASTA) dan jumlah kontrol ≤ 1 Maka	TIDAK
4	Jika Jumlah kontrol > 1 dan ≤ 3 dan Pekerjaan=(BUMN,DOKTER,DOSEN,KARY.SWASTA,MAHASISWA,PROFESI LAINNYA, TNI-POLRI,WIRASWASTA) dan penanggung=(ASURANSI,PENDIDIKAN,SWASTA) Maka	YA

6. Pasien yang masuk ke dalam klasifikasi seperti pada Tabel 4, merupakan pasien yang sebagian besar melakukan persalinan di rumah sakit.
7. Mengembangkan suatu paket persalinan untuk pasien dengan pangsa pasar menengah ke atas, yaitu penggabungan antara persalinan dengan klinik kecantikan. Pemanfaatan point 6 dan 7 di atas mengacu pada Tabel 4.

Tabel 4. Rule no 6,7,8, dan 9

No Rule	Kondisi	Melahirkan di rumah sakit
6	Jika Jumlah kontrol > 3 dan Pekerjaan=(BUMN,DOKTER,DOSEN,KARY.SWASTA,MAHASI SWA,PROFESI LAINNYA,TNI-POLRI,WIRASWASTA) dan jumlah kontrol > 9 Maka	YA
7	Jika Jumlah kontrol > 3 dan	YA

No Rule	Kondisi	Melahirkan di rumah sakit
	Pekerjaan=(BUMN,DOKTER,DOSEN,KARY. SWASTA,MAHASISWA,PROFESI LAINNYA,TNI-POLRI,WIRASWASTA) dan jumlah kontrol ≤ 9 dan penanggung=(ASURANSI,BUMN,SWASTA) Maka	
8	Jika Jumlah kontrol > 3 dan Pekerjaan=(BUMN,DOKTER,DOSEN,KARY. SWASTA,MAHASISWA,PROFESI LAINNYA, TNI-POLRI,WIRASWASTA) dan penanggung=(PENDIDIKAN,UMUM) dan umur ≤ 28 Maka	TIDAK
9	Jika Jumlah kontrol > 3 dan Pekerjaan=(BUMN,DOKTER,DOSEN,KARY. SWASTA,MAHASISWA,PROFESI LAINNYA, TNI-POLRI,WIRASWASTA) dan penanggung=(PENDIDIKAN,UMUM) dan umur > 28 Maka	YA

Ada 3 (tiga) masalah yang harus terjawab dengan adanya penelitian ini.

1. Perlunya suatu model untuk dapat menganalisa pasien mana yang paling besar kemungkinannya tidak akan melahirkan di rumah sakit. Dari 8.929 kehamilan pasien yang melakukan pengontrolan di RS. "X", sebesar 60,802% (5.429 kehamilan pasien) tidak melahirkan di rumah sakit ini dan sisanya 39,198% (3.500 kehamilan pasien) melahirkan di rumah sakit ini. Beberapa faktor yang sangat mempengaruhi persalinan tidak dilakukan di rumah sakit ini adalah:

a. Faktor jumlah kontrol

Pasien yang melakukan kontrol ≤ 3 (tiga) kali maka pasien tersebut tidak akan melahirkan di RS. "X" dengan kontribusi sebesar 77,680% (3.978 kehamilan pasien). Alasannya disebabkan karena faktor biaya persalinan, anggapan apabila kondisi ibu dan bayi sudah baik maka kontrol ke rumah sakit tidak perlu dilakukan kembali. Apabila pasien yang melakukan kontrol pasien ≥ 4 (empat) kali maka pasien tersebut akan melahirkan di RS. "X" dengan kontribusi sebanyak 61,896% (2.357 kehamilan pasien).

b. Faktor jarak

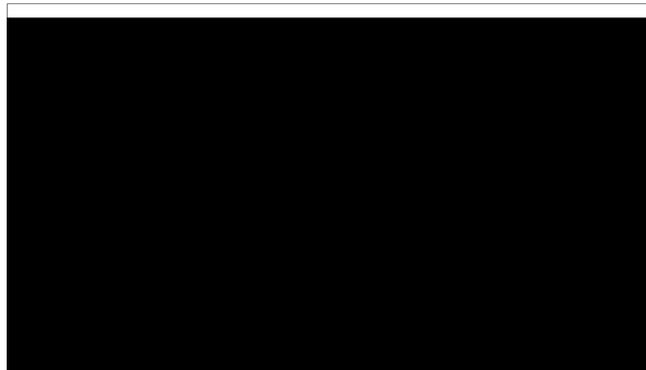
Pasien yang mempunyai jarak tempuh dari rumah ke RS. "X" berkategori JAUH hampir 37% (2.015 kehamilan pasien) dapat dipastikan bahwa pasien tersebut tidak melakukan persalinan di rumah sakit. Detail dari prosentase tersebut seperti terlihat pada Gambar 12.



Gambar 12. Pasien yang tidak melahirkan berdasarkan jarak

c. Faktor pekerjaan suami

Hampir 93,391 % (1.074 kehamilan pasien) untuk suami yang bekerja sebagai BURUH dan PNS tidak melahirkan di rumah sakit, secara lengkap dapat dilihat pada Gambar 13.



Gambar 13. Distribusi jumlah pasien per pekerjaan

d. Faktor penanggung persalinan

Dari total pasien yang hamil yang tidak melahirkan di RS. "X" adalah 5.429 kehamilan pasien, yang merupakan pasien UMUM adalah 87% (4.723 kehamilan pasien).



Gambar14. Pasien yang tidak melahirkan berdasarkan Penanggung

2. Masalah kedua, masih lambatnya penanganan terhadap pasien yang kemungkinan tidak akan melahirkan di rumah sakit dikarenakan belum ada informasi pasien-pasien tersebut ke pihak marketing.
3. Masalah ketiga, belum adanya segmentasi antara pasien dalam pengkategorian pasien yang akan melahirkan di rumah sakit dan pasien yang tidak melahirkan di rumah sakit sehingga *treatment* terhadap pasien belum terarah.

5. KESIMPULAN

Kesimpulan yang didapat dari pengetahuan model prediksi tersebut adalah:

1. Rumah sakit ini perlu memfokuskan terhadap penanganan pasien yang tidak melahirkan di rumah sakit terutama kategori pasien UMUM karena berdasarkan historis data yang ada cukup besar yaitu 87% (4.723 pasien). Jika penanganan ini tidak dilakukan dengan segera maka angka pasien yang tidak melakukan persalinan di rumah sakit ini (*Churn*) akan bertambah. Adapun hal yang dapat dilakukan untuk point ini adalah:
 - a. Untuk target pasien dengan status ekonomi menengah ke bawah maka dapat dibuatkan paket-paket persalinan yang kompetitif untuk ekonomi menengah dan ibu muda yang berumur ≤ 30 tahun.
 - b. Mengembangkan suatu paket persalinan untuk pasien dengan pangsa pasar menengah ke atas, yaitu penggabungan antara persalinan dengan klinik kecantikan.

2. Prosentase kehamilan normal yang tidak mengalami gangguan selama proses persalinan adalah 92,56%(5.025 kehamilan pasien) namun pasien tersebut tidak melahirkan di rumah sakit. Artinya pasien yang melahirkan di rumah sakit sebagian besar adalah dengan resiko persalinan yang tinggi.
3. Sedangkan dengan berdasarkan kriteria umur maka pasien yang berumur < 27 tahun sebesar 42,75%(2.148 kehamilan pasien) dan yang berumur >27 tahun sebesar 57,25% (2.877 kehamilan pasien). Sama seperti point 1 di atas bahwa perlu dibuatkan paket-paket persalinan yang kompetitif yang dapat menjadi daya tarik pasien.
4. Faktor jarak sangat berpengaruh terhadap pasien untuk tidak melakukan persalinan di rumah sakit. Terbukti sekitar 37% (2.015 kehamilan pasien) faktor jarak yang JAUH sangat berpengaruh terhadap jumlah pasien persalinan. Rumah sakit ini tidak menjadi minat pasien untuk melakukannya untuk pasien yang berada dekat dengan rumah sakit, di lihat dari faktor jarak DEKAT sekitar 25% (1.347 kehamilan pasien) dari 2.904 kehamilan pasien, seharusnya jika rumah sakit ini menjadi tempat untuk persalinan maka angka faktor jarak DEKAT ini harusnya kecil.
5. Perlu ditingkatkan kerja sama dengan penanggung persalinan, mengingat angka penanggung persalinan ini seperti ASURANSI, BUMN, SWASTA, PENDIDIKAN jika di total sebesar 35,48% (836 kehamilan pasien) sangat berpengaruh mempengaruhi pasien melahirkan di rumah sakit.

6. DAFTAR PUSTAKA

- [1]. Chapman,Pete., Clinton,Julian., Kerber, Randy., Khabaza,Thomas., Reinartz,Thomas., Shearer, Colin dan Wirth, Rudiger. CRISP-DM 1.0 Step-by-step data mining guide CRISP-DM Consorcium, 2000
- [2]. Hadden,John., Tiwaria,Ashutosh., Roy, Rajkumar., Ruta, Dymitr. Computer assisted customer churn management:State-of-the-art and future trends . Enterprise Integration, Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK. www.elsevier.com/locate/cor, 2005
- [3]. Hadden,John., Tiwaria,Ashutosh., Roy, Rajkumar., Ruta, Dymitr. Churn Prediction using Complaints Data. World Academy of Science, Engineering and Technology 19 2006
- [4]. Larose , Daniel T. Discovering Knowledge in Data - An Introduction to Data Mining.John Wiley & Sons, Inc., Hoboken, New Jersey. 2005

- [5]. Pejic Bach, Mirjana dan Cosic, Dijana . Data mining usage in health care management: literature survey and decision tree application, Medicinski Glasnik, Volume 5, Number 1, January 2008
- [6]. Muata,Kweku dan Bryson,Osei. Evaluation of decision trees: a multi-criteria approach, Department of Information Systems and The Information Systems Research Institute, School of Business, Virginia Commonwealth University Richmond, VA 23284, USA, www.elsevier.com/locate/dsw 2004
- [7]. Janssen , Patricia A., Lee, Saxell MA., Page, Lesley A., Klein , Michael C., Liston , Robert M., Lee, Shoo K . Outcomes of planned home birth with registered midwife versus planned hospital birth with midwife or physician www.cmaj.ca on August 31, 2009
- [8]. Govindaraju, Rajesri., Simatupang,Tota. dan Ari Samadhi, TMA. Perancangan Sistem Prediksi Churn Pelanggan PT. Telekomunikasi Seluler dengan Memanfaatkan Proses Data Mining, 2008
- [9]. Susanto, Sani dan Suryadi, D. Pengantar Data Mining menggali pengetahuan dari bongkahan data, CV. ANDI OFFSET , 2010
- [10]. Stolba , Nevena – Tesis : Towards a Sustainable Data Warehouse Approach for Evidence-Based Healthcare , Vienna University of Technology Facult of Informatics, Vienna, September 27 th 2007