

Tahapan yang dilakukan mengacu pada jurnal “Penggunaan Data Mining untuk Memprediksi Churn Pasien (Kasus Studi Rumah Sakit ‘x’)”

1. ***Bussiness Understanding***

Merupakan tahapan dari CRISP-DM yang memfokuskan terhadap pemahaman dari tujuan organisasi, memberikan solusi terhadap permasalahan yang telah didefinisikan dan mendesain rencana untuk mencapai tujuan. Pada fase ini memiliki tujuan bisnis dan *data mining* untuk meningkatkan jumlah proses persalinan yang ada di rumah sakit. Parameter keberhasilannya meliputi prediksi pasien yang tidak melakukan persalinan di RS. “X”, dapat memberikan informasi daftar pasien yang tidak melakukan proses persalinan di rumah sakit, dan mengoptimalkan data histori di rumah sakit berdasarkan pemanfaatannya.

Komponen	Data Mining
Personel	<i>Database, Administrator, Programmer</i>
Data	<i>Database</i>
Platform	<i>Windows Server, SQL 2008</i>
Software	<i>Clementine, Java</i>

Tabel 1. *Komponen Data Mining*

2. ***Data Understanding***

a. Data Awal

Beberapa *database* yang digunakan untuk melakukan proses *data mining* diantaranya adalah *database* rekam medis (identitas, kunjungan, diagnose, informasi) dan *database billing* (informasi penanggung persalinan).

b. Data yang dibutuhkan

Informasi Tabel	Penjelasan
Tabel Pasien	informasi yang berkaitan dengan identitas pasien
Tabel Daftar	informasi yang berkaitan dengan kontrol pasien ke rumah sakit untuk pemeriksaan kehamilannya
Tabel Kod_Utama	diagnosa-diagnosa utama pasien yang melakukan pemeriksaan di rumah sakit
Tabel ICD10	berisikan master standar WHO mengenai pengkodean penyakit-penyakit
Tabel Kontraktor	referensi master penanggung kesehatan dari pasien yang ada di rumah sakit
Tabel Idts2	tempat menyimpan data pasien yang dirawat di rumah sakit

c. *Explore Data*

Ada beberapa tahapan yaitu (1) pengambilan data berdasarkan *database* rumah sakit selama 5 tahun terakhir. (2) Diagnosa utama pasien berdasarkan pengawasan kehamilan normal dan resiko tinggi. (3) *cross-tab*, data kunjungan pasien. (4) Cek histori apakah telah melakukan persalinan di rumah sakit.

d. Verifikasi kualitas data

Memastikan agar semua data tidak terdapat duplikasi dan data tidak valid selama pengambilan data diambil.

3. **Data Preparation**

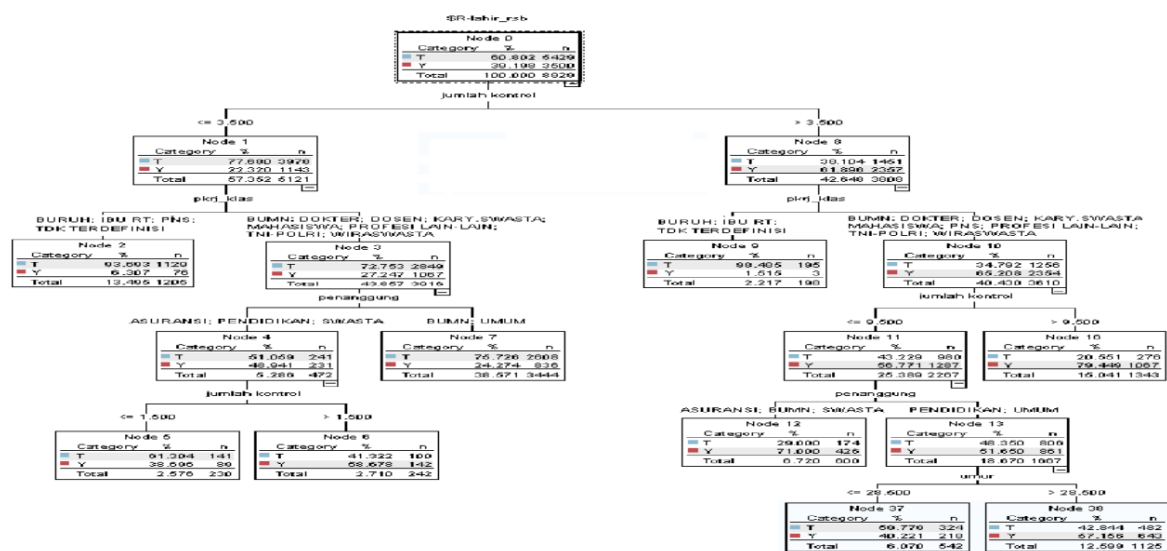
Merupakan aktivitas yang diperlukan untuk persiapan data menjadi dataset yang digunakan sebagai dasar untuk pembentukan model. Pada fase ini terdapat dua tahapan yaitu *Select Data* (memutuskan data yang akan digunakan untuk dianalisis) dan *Clean Data* (membersihkan data yang dianggap *error*).

4. **Modelling**

Pemilihan model klasifikasi beserta softwarenya. 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 and Regression Trees)

Melakukan export data ke dalam format excel kemudian dipergunakan software Clementine model untuk CART (Classification and Regression Trees) untuk memproses pemodelan prediksi.



Gambar 1. Model Prediksi menggunakan CART

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 and Regression Trees) dan didapat suatu penulisan pemodelan dengan bentuk IF kondisi THEN status pasien melahirkan di rumah sakit.

5. **Evaluation**

Melakukan evaluasi dari rangkaian proses untuk menjamin bahwa model yang dibentuk dapat mencapai tujuan organisasi. 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 pengujian model dengan cara mengambil data yang sudah dipersiapkan di mana dari data tersebut sudah ada hasil mengenai pasien yang akan diprediksi. Dari proses pengujian tersebut didapat tingkat akurasi dari model ini adalah 73,40 %. Dari proses pengujian tersebut didapat tingkat akurasi dari model ini

adalah 73,40 % dengan distribusi akurasi. 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%.

6. *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.

DAFTAR PUSTAKA

- Purwono, Y. (2012). PENGGUNAAN DATA MINING UNTUK MEMPREDIKSI CHURN PASIEN DI RUMAH SAKIT (STUDI KASUS: RUMAH SAKIT X) . *Media Informatika Vol. 11. No. 3*, 48-67.