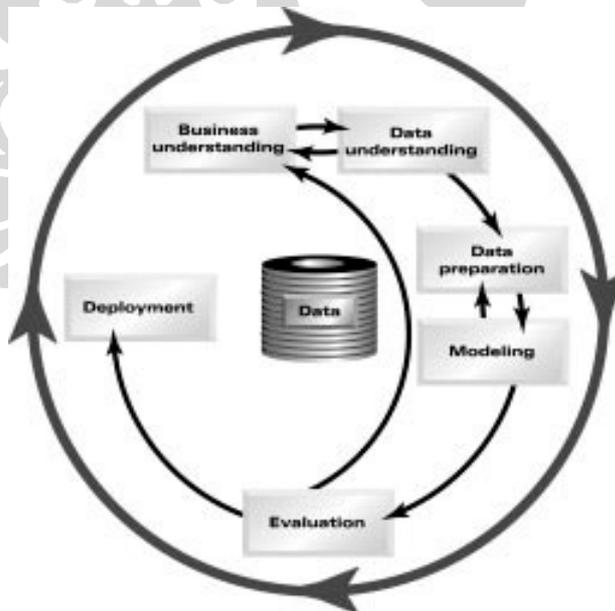


BAB III METODOLOGI PENELITIAN

Bab ini akan menjelaskan mengenai pemilihan metodologi yang akan dipakai dalam penelitian nanti, serta objek dari penelitian.

3.1 Metode Data mining

Proses *data mining* digunakan dengan tujuan untuk dapat memilih teknik *data mining* apa yang cocok untuk diterapkan, dan digunakan juga sebagai acuan untuk keseluruhan proses penambangan data dari obyek bisnis. Saat ini terdapat beberapa metode yang bisa dipakai untuk melakukan *data mining*, diantaranya model CRISP, *generic model*, dan lain sebagainya. Proses *data mining* yang akan digunakan pada tesis yang akan dibuat adalah model referensi CRISP-DM, seperti terlihat pada gambar yang ada dibawah ini[6].



Gambar 6. Model data mining CRISP-DM

Model CRISP-DM berisikan daur hidup kegiatan *data mining* yang terdiri atas enam fase. Gambar 3-1 diatas menunjukkan keseluruhan tahapan pada model CRISP-DM. Salah satu keuntungan dari model ini adalah, tahapan fase dari model bukanlah sebuah tahapan yang kaku. Perpindahan maju dan mundur antara tiap fase yang berbeda bisa selalu dilakukan. Hal ini sesuai dengan sifat alami dari *data mining* itu sendiri, dimana proses *data mining* tidak selesai saat sebuah hasil di temukan, sebab proses *data mining* merupakan sebuah proses pembelajaran terus menerus. Subbab-subbab berikut ini akan dijelaskan secara singkat apa saja yang dilakukan pada tiap fase dari model ini.

3.1.1 Fase business understanding

Fase ini merupakan fase awal dari model CRISP-DM. Pada fase ini dibutuhkan pengertian akan obyektif dari kegiatan *data mining* yang akan dilakukan, dan kebutuhan dari perspektif bisnis, kemudian mengubah informasi tersebut ke dalam sebuah definisi problem yang akan dijawab oleh *data mining* dan sebuah rencana awal yang didesai untuk mencapai obyektif tersebut.

3.1.2 Fase data understanding

Fase ini merupakan fase pengumpulan koleksi data awal, dan melakukan proses pengenalan terhadap data tersebut dengan tujuan untuk lebih mengenal *nature* dari data yang akan dipakai. Selain itu fase ini juga mencoba melakukan identifikasi awal mengenai masalah-masalah yang berkaitan dengan kualitas data, dan mencoba mendeteksi subset-subset yang menarik dari data untuk membentuk hipotesa awal dari informasi yang tersembunyi.

3.1.3 Fase data preparation

Pada fase ini dilakukan aktivitas-aktifitas untuk menyusun dataset *final*, yaitu *dataset* yang akan dijadikan input kedalam perangkat lunak *data mining*. Aktifitas-aktifitas yang dilakukan pada fase ini meliputi pemilihan data (*table*, *record* dan serta atribut yang akan dipakai), transformasi data dan pembersihan data (*Data cleansing*).

3.1.4 Fase modelling

Fase ini mengerjakan aktifitas pemilihan teknik *data mining* yang akan digunakan, serta mengkalibrasi parameter-parameter teknik *data mining* tersebut dengan nilai yang optimal. Teknik *data mining* yang dipakai seringkali memerlukan beberapa kebutuhan yang spesifik, sehingga pada fase ini dibutuhkan pengulangan kembali fase *data preparation*.

3.1.5 Fase evaluasi

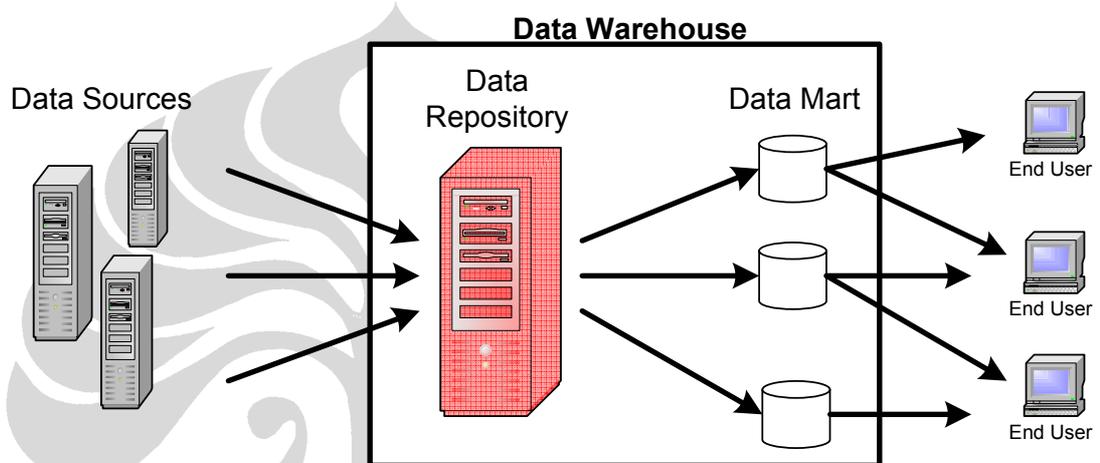
Pada tahap ini akan dimulai pembuatan model *data mining* berkualitas tinggi dari sudut pandang data analysis. Evaluasi yang dilakukan pada tahap dilakukan secara mendalam, dengan tujuan untuk menyesuaikan model yang didapat agar sesuai dengan obyektif bisnis yang dilakukan. Pada akhir tahap ini juga ditentukan keputusan apakah hasil *data mining* akan digunakan atau tidak.

3.1.6 Fase deployment

Pada fase ini dilakukan penyusunan dan presentasi dari pengetahuan yang didapat dari *data mining*. Presentasi dari hasil *data mining* ini, dibuat sedemikian rupa sehingga bisa dimengerti dan dibaca oleh kalangan awam.

3.2 Obyek penelitian

Institusi yang menjadi obyek penelitian adalah sebuah institusi perbankan nasional yang cukup besar. Institusi tersebut saat ini sudah memiliki *data warehouse* yang cukup memadai. Selain *data warehouse*, Institusi tersebut juga memiliki beberapa *data mart* yang tersebar di beberapa divisi. Gambar 3-2 dibawah menggambarkan arsitektur dari data *data warehouse* yang digunakan



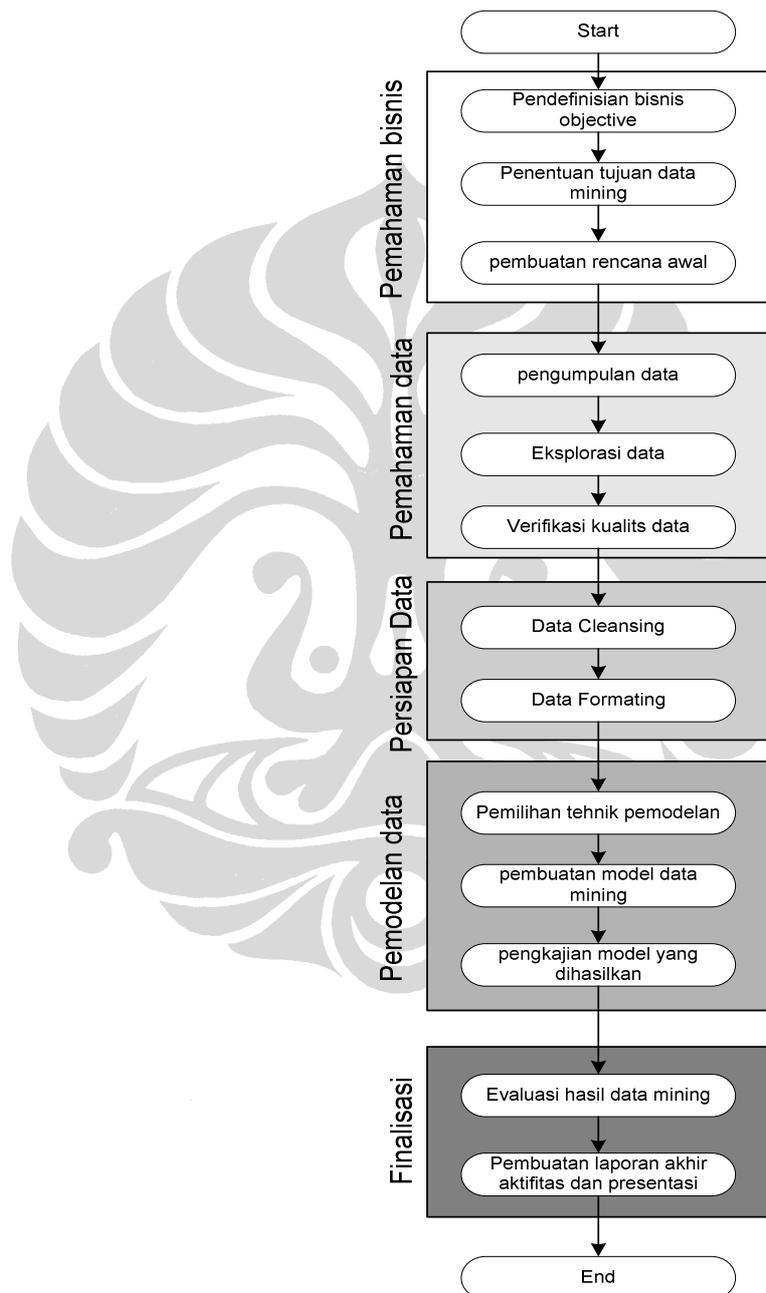
Gambar 7. Arsitektur Datawarehouse

3.3 Dataset

Dataset yang akan digunakan diambil dari sebuah *datamart* yang digunakan pada divisi *sales* nasional sebuah bank ternama. Data yang terdapat pada data mart tersebut merupakan data harian yang diturunkan dari data repositori pusat. Proses importing data dilakukan dalam periode harian secara otomatis. *Dataset* tersebut merupakan kumpulan data fluktuasi saldo harian dari seluruh nasabah yang dimiliki institusi tersebut.

3.4 Skema Pelaksanaan penelitian

Proses pelaksanaan *data mining* meliputi beberapa urutan langkah dasar. Gambar 3-3 dibawah menggambarkan proses pelaksanaan *data mining* yang akan diterapkan pada tesis yang akan dibuat.



Gambar 8. Tahapan pelaksanaan data mining