1. Pendahuluan

1.1 Latar Belakang

Perkembangan sistem komunikasi pada jaringan bergerak telah berkembang semakin pesat seiring dengan kemajuan tekonologi komunikasi itu sendiri. Terdorong oleh pertumbuhan kebutuhan para pelanggannya, membuat para *provider* layanan telekomunikasi melakukan berbagai macam terobosan dengan fitur-fitur dan fasilitas lainnya yang memanjakan para pelanggannya. Kesuksesan peluncuran sistem komunikasi 2G telah mendorong penyebaran dan penyempurnaan kapabilitas untuk memenuhi harapan para pelanggan untuk mendapatkan layananan informasi yang lebih canggih. Kecanggihan akan fitur-fitur yang telah dikembangkan pada jaringan 2G terus dikembangkan hingga saat ini telah memasuki era 3G (*Third Generation*). Jaringan 3G akan membawa perubahan secara revolusioner pada para pengguna *mobile phone* yang sebelumnya berbasiskan layanan suara menjadi layanan yang berbasis multimedia dan akan digunakan secara luas di dunia telekomunikasi.

Munculnya teknologi komunikasi generasi ke-tiga ini diharapkan dapat meningkatkan skala bisnis bagi para *provider* layanan komunikasi dengan tujuan untuk meningkatkan profit perusahaan. Beberapa langkah yang bisa diterapkan dalam mencapai tujuan tersebut adalah : (1) Perumusan target, yaitu memilih pelanggan-pelanggan potensial untuk beralih ke jaringan 3G yang menjadi target pemasaran. Peningkatan profit perusahaan diharapkan semakin besar apabila jumlah pelanggan jaringan 3G semakin meningkat. (2) Estimasi dan prediksi. Estimasi adalah menerka sebuah nilai yang belum diketahui berdasarkan informasi yang telah tersedia dan prediksi adalah memperkirakan nilai estimasi tersebut untuk masa yang akan datang.

Untuk mendukung solusi-solusi tersebut maka digunakan teknologi *data mining*. Dengan menggunakan teknologi ini perusahaan akan lebih intensif untuk menembak target pasar yang potensial. Dengan demikian diharapkan semakin besar peluang para pelanggan 2G yang akan beralih ke layanan 3G semakin besar pula profit yang akan diperoleh perusahaan. Oleh karena itu, dalam tugas akhir ini akan dibuat suatu implementasi perangkat lunak yang dapat memperkirakan para pelanggan 2G yang potensial beralih ke layanan jaringan 3G pada studi kasus PAKDD *(Pasific-Asia Conference on Knowledge Discovery and Data Mining)* 2006 Kompetisi Data Mining.

1.2 Perumusan Masalah

Tujuan utama dalam pengerjaan tugas akhir ini adalah memprediksi pelanggan 2G yang potensial untuk beralih ke layanan 3G. Rumusan masalah untuk tugas akhir ini akan dilakukan, yaitu (1) menentukan profil pelanggan berdasarkan kelas 2G atau 3G dan (2) memprediksi pelanggan 2G yang potensial berdasarkan pemodelan *classifier* yang telah diperoleh pada saat pelatihan sistem.

Namun, masalah yang dihadapi untuk fase ini adalah dimensi *dataset* pelanggan cukup tinggi dan tidak efisien jika seluruh atribut digunakan untuk memprediksi. Untuk mengatasi hal ini, yang pertama dilakukan adalah melakukan pencarian atribut yang relevan dalam menentukan kelas pelanggan dengan teknik *feature selection*. Kemudian untuk melakukan klasifikasi dan prediksi pelanggan akan dilakukan dengan metode *Support Vector Machine* (SVM). Hasil klasifikasi berupa *false positive* akan dianggap sebagai pelanggan 2G yang potensial. Dengan mengacu pada latar belakang masalah di atas, maka permasalahan yang akan dibahas dan diteliti adalah :

1. Bagaimana melakukan *feature selection* yang relevan terhadap label kelas dengan *dataset* yang berdimensi tinggi
2. Bagaimana menerapkan metode *Support Vector Machine* sebagai *classifier* untuk melakukan klasifikasi terhadap *dataset* pelanggan sehingga menghasilkan kelas pelanggan 2G dan 3G dengan akurat.
3. Bagaimana menentukan parameter-parameter masukan SVM agar hasil klasifikasi yang diperoleh hasil yang optimal.

1.3 Tujuan

Berdasarkan rumusan masalah di atas, maka tujuan dari tugas akhir ini adalah:

1. Menganalisis data pelanggan yang terdiri dari data yang telah memiliki kelas (data *training*) dan data yang belum memiliki kelas (data *testing*).
2. Menerapkan metode *feature selection* untuk tahap *pre-process* untuk menangani dimensi data yang tinggi.
3. Menerapkan metode SVM sebagai *classifier* untuk klasifikasi serta prediksi.
4. Melakukan perbandingan akurasi *classifier* terhadap penggunaan *kernel linear, polynomial*, dan RBF.
5. Menghasilkan *list* prediksi pelanggan mana saja yang termasuk ke dalam kelas positif (3G) dan kelas negatif(2G).

1.4 Batasan Masalah

Dalam Tugas Akhir ini, yang akan dibahas adalah suatu implementasi untuk menemukan pola yang dapat memprediksi para pelanggan 2G yang berpotensi menjadi pelanggan 3G dengan batasan masalahnya sebagai berikut :

1. Sistem yang akan dirancang hanya untuk studi kasus PAKDD 2006 Kompetisi *Data Mining*.
2. Untuk kasus yang lain, sistem bisa digunakan jika memiliki bentuk *dataset* yang sama dengan *dataset* PAKDD 2006 Kompetisi *Data Mining*.
3. *Kernel* yang digunakan pada SVM meliputi *kernel linear*, *polynomial*, dan RBF.
4. Melakukan perbandingan hasil klasifikasi dengan pemenang PAKDD yang menggunakan metode SVM berdasarkan *balanced accuracy* (secara kuantitatif).

5. Jumlah *record* data yang digunakan sebesar 4000 record yang terdiri dari 3000 *record* data *training* dan 1000 *record* data *testing*.

1.5 Metodologi Penyelesaian Masalah

Metode yang digunakan dalam penyelesaian tugas akhir ini adalah menggunakan metode studi pustaka atau studi literatur dan analisis dengan langkah kerja sebagai berikut:

1. Studi Literatur:
   a) Pencarian referensi
      Mencari referensi dan sumber-sumber lain yang layak yang berhubungan dengan *data mining*, *feature selection*, klasifikasi dengan metode *Support Vector Machine*, dan pemahaman akan tipe pelanggan 2G dan 3G itu sendiri.
   b) Pendalaman materi
      Mempelajari dan memahami materi yang berhubungan dengan tugas akhir ini.
2. Pencarian dan pengumpulan data. Mengumpulkan data yang berupa data profil pelanggan dengan label kelas 2G dan 3G untuk *training* serta data yang belum memiliki label kelas yang akan diklasifikasi dan diprediksi.
3. Mempelajari konsep dari *feature selection* dan *Support Vector Machine* yang akan digunakan dalam implementasi perangkat lunak.
4. Melakukan analisa terhadap *feature selection*, dan *Support Vector Machine* dalam perancangan perangkat lunak.
5. Melakukan implementasi perancangan perangkat lunak dengan menggunakan Matlab 7 dan Microsoft SQL Server 2000 untuk aplikasi *database*.
6. Melakukan proses pengujian perangkat lunak dengan skenario sebagai berikut, yaitu :
   - Melakukan pelatihan agar model sistem yang telah dirancang mampu memberikan hasil yang optimal dan stabil.
   - Tahap pelatihan dilakukan dengan cara membagi data yang telah memiliki label menjadi dua bagian, yaitu sebagai data untuk *training* dan *testing*.
   - Melakukan proses klasifikasi untuk data pelanggan yang belum memiliki kelas.
   - Melakukan prediksi terhadap data pelanggan yang potensial untuk beralih ke jaringan 3G.
7. Mencatat hasil keluaran program.
   Pengambilan kesimpulan dan penyusunan laporan tugas akhir.