

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada saat sekarang ini perkembangan teknologi sangatlah cepat, terlebih lagi dibidang industri telekomunikasi. Rata-rata setiap orang kini telah memiliki alat komunikasi (*gadget*), mulai dari anak-anak hingga orang dewasa memilikinya. Hal ini menyebabkan persaingan dalam dunia industri telekomunikasi semakin ketat, serta mendorong setiap industri telekomunikasi untuk mendapatkan pelanggan baru sebanyak-banyaknya dan juga mempertahankan pelanggan lama yang telah menggunakan jasa industri telekomunikasi tersebut.

Churn merupakan suatu fenomena berpindahnya pelanggan dari suatu operator ke operator lain, dengan adanya fenomena *churn* ini, nantinya akan berdampak kepada aliran kas suatu perusahaan apabila tidak diatasi dengan solusi yang cepat dan tepat. Perusahaan besar mengimplementasikan model *churn prediction* yang dapat mendeteksi kemungkinan *churn* pelanggan sebelum pelanggan tersebut meninggalkan perusahaan [9]. Namun pada faktanya sendiri terjadinya *churn* pada industri telekomunikasi sangatlah kecil sekitar 1,8% dari jumlah pelanggan [8]. Hal ini menimbulkan masalah yang disebut dengan *imbalance data*, dimana terjadi perbedaan yang sangat besar diantara jumlah data pada satu kelas (kelas major) dengan kelas lainnya (kelas minor) [10]. Masalah lain dari *churn prediction* adalah adanya *noise*, dan meranking pelanggan yang tergolong ke dalam fenomena *churn* atau tidak [1].

Teknik *data mining* merupakan salah satu studi yang dapat menangani permasalahan *imbalance data*, dan diharapkan nantinya akan dapat menghasilkan model *churn prediction* yang handal. Beberapa pendekatan sudah diusulkan untuk menghadapi permasalahan diatas, seperti *Decision Tree Based*, *Neural Network Algorithm*, *Genetic Algoritm*, *Support Vector Machine*, tetapi masih memiliki keterbatasan [1]. *Decision Tree Based* masih rentan terhadap adanya *noise*, *neural*

network masih sangat sulit untuk menterjemahkan *pattern* yang bersifat symbol, *genetic algorithm* yang tidak dapat menentukan kemungkinan terkait dengan prediksinya sendiri [1].

Untuk itu tugas akhir ini menggunakan teknik *Support Vector Machine* sebagai klasifikasinya dan teknik *combine sampling* (*Synthetic Minority Over-sampling Technique & Neighbourhood Cleaning Rule*) untuk mengatasi permasalahan *imbalance data*, teknik *combine sampling* digunakan untuk meningkatkan akurasi dari *support vector machine* pada kelas major dan kelas minor [3]. *Support Vector Machine* dipilih karena nantinya akan dapat meminimalkan error pada *data training*, dan juga secara teori *Support Vector Machine* dibangun untuk permasalahan klasifikasi dengan dua kelas. *Combine Sampling* dipilih karena dengan menggunakan *Neighbourhood Cleaning Rule* dapat mengidentifikasi data yang termasuk *noise*, dan *Synthetic Minority Over-sampling Technique* digunakan untuk menduplikasi data yang sudah bersih dari *noise*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan diatas, maka rumusan masalah yang akan dibahas dan diteliti adalah:

1. Membangun model klasifikasi yang dapat menangani permasalahan *imbalance data* pada *churn prediction*.
2. Bagaimana performansi yang dihasilkan oleh model klasifikasi yang telah dibangun.

Batasan masalah yang ada pada tugas akhir ini adalah:

1. Data yang digunakan adalah data yang diperoleh dari perusahaan telekomunikasi WITEL Area Regional 7.
2. Evaluasi model *churn prediction* yang telah dibangun menggunakan *f1-measure*.

1.3 Tujuan

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

1. Membangun model *churn prediction* dengan menggunakan metode klasifikasi *Support Vector Machine* dan *Combine Sampling*.
2. Menganalisis performansi model *churn prediction* yang telah dibangun menggunakan metode klasifikasi *Support Vector Machine* dan *Combine Sampling* dengan menggunakan *f1-measure*.

1.4 Metodologi Penyelesaian Masalah

Metode atau langkah-langkah yang digunakan dalam penyelesaian tugas akhir ini adalah sebagai berikut:

1. Studi literatur

a) Melakukan pencarian referensi dan sumber-sumber lain yang layak yang berhubungan dengan tugas akhir, antara lain *handling imbalance data*, *churn prediction*, *support vector machine*, *neighbourhood cleaning rule*, *combine sampling*, *f1-measure*, *oversampling*, *undersampling*, dll.

b. Melakukan pendalaman terhadap materi-materi yang telah disebutkan diatas.

2. *Feature selection*, yaitu melakukan pengurangan atribut terhadap data set yang digunakan, pengurangan atribut dilihat berdasarkan keterhubungan setiap atribut prediksi dengan target kelas, sehingga prediksi *churn* dapat diketahui lebih akurat.

3. Membangun model klasifikasi dan melihat performansi dari data yang tidak dilakukan proses *handling imbalance data*.

4. Melakukan penanganan *handling imbalance data* terhadap data set, *Synthetic Minority Oversampling Technique* dan *Neighbourhood Cleaning Rule*.

5. Melihat performansi dari data set yang dilakukan proses *handling imbalance data* berdasarkan *f1-measure* yang dihasilkan.
6. Mengambil kesimpulan dari hasil pengujian yang telah didapatkan.
7. Melakukan penyusunan laporan buku tugas akhir.

1.5 Sistematika Penulisan

Langkah-langkah penulisan yang terdapat dalam pelaksanaan tugas akhir ini adalah sebagai berikut:

BAB I PENDAHULUAN

Pada bab pendahuluan berisi tentang latar belakang masalah, rumusan masalah beserta batasan masalah, tujuan, metodologi penyelesaian masalah, dan sistematika penulisan.

BAB II KAJIAN PUSTAKA

Berisikan tentang hasil ringkasan/ *resume* dari sumber-sumber terkait dengan hal-hal yang berhubungan dengan tugas akhir. Sumber-sumber yang ada dapat berasal dari *paper*, jurnal, textbook, tesis, dan tugas akhir lain yang berhubungan dengan memperhatikan penggunaan teknik acuan sebagai tindakan untuk menghindari plagiat.

BAB III PERANCANGAN SISTEM

Bagian alur pemodelan berisi rancangan dari model klasifikasi yang dibangun dalam bentuk flowchart. Alur pemodelan dapat berisi kebutuhan data (pengumpulan dan preprocessing data), serta skenario pengujian yang dilakukan untuk mencapai tujuan penelitian yang ditulis pada Bab Pendahuluan. Ilustrasi proses pengolahan data juga dijelaskan dalam bagian ini untuk memperjelas kegiatan yang dilakukan dalam pengerjaan tugas akhir.

BAB IV PENGUJIAN DAN ANALISIS

Bagian pengujian dan analisis berisi pembahasan hasil pengujian berdasarkan skenario pengujian yang dituliskan pada bab Alur Pemodelan. Hasil dari kegiatan pengujian ini menjadi dasar pengambilan kesimpulan.

BAB V PENUTUP

Bagian penutup berisi kesimpulan dari keseluruhan hasil pengerjaan tugas akhir yang mengacu pada tujuan penelitian, skenario pengujian dan analisis hasil pengujian pada bab-bab sebelumnya.

DAFTAR PUSTAKA

Bagian daftar pustaka berisikan daftar referensi yang digunakan dalam pembuatan/ pembangunan tugas akhir ini.

LAMPIRAN

Bagian lampiran berisikan deskripsi *dataset*, dan screenshot hasil program yang dihasilkan.