

BAB I

PENDAHULUAN

1.1 Latar belakang

Perkembangan dunia teknologi yang semakin hari semakin berkembang pesat khususnya dunia telekomunikasi, menyebabkan banyaknya *provider-provider* baru yang muncul menawarkan produk yang lebih menguntungkan bagi konsumen. Hal tersebut dapat mengakibatkan perpindahan pelanggan dari *provider* satu ke *provider* yang baru, atau dalam *data mining* disebut *churn*. Oleh karena itu, diperlukan prediksi awal untuk pengklasifikasian pelanggan yang diprediksi *churn* atau tidak, serta menganalisis atribut utama sebagai penentu prediksi awal terjadinya *churn* dari data historis pelanggan *churn* salah satu perusahaan telekomunikasi.

Salah satu cara untuk mendapatkan informasi atau pola dari kumpulan data yang besar adalah dengan menggunakan teknik-teknik dalam data *mining*. Ada banyak teknik dalam data *mining* untuk menghasilkan informasi atau pola dari sekumpulan data. Salah satu teknik tersebut adalah klasifikasi pohon keputusan. Namun, kasus data yang digunakan penelitian ini merupakan data yang *imbalanced class*. *Imbalanced class* merupakan suatu masalah atau tantangan karena biasanya mesin *learning* akan menghasilkan suatu akurasi prediksi yang baik terhadap kelas data latih yang banyak (kelas mayor), sedangkan untuk kelas data latih yang sedikit (kelas minor) akan dihasilkan akurasi prediksi yang buruk. Hal ini terjadi untuk kelas minor bahkan mungkin algoritma tersebut akan menganggapnya sebagai pencilan sehingga *rule* yang terkandung dalam kelas minor tidak dapat dihasilkan dengan baik. Salah satu pendekatan untuk permasalahan kasus data ini adalah dengan menggunakan metode *balanced random forest*.

Balanced Random Forest merupakan salah satu algoritma pohon keputusan dari kelas klasifikasi dengan tingkat akurasi yang baik [1]. *Balanced Random Forest* merupakan sebuah metode *ensemble* yang terdiri dari beberapa pohon keputusan sebagai *classifier*. Kelas yang dihasilkan dari proses klasifikasi ini diambil dari kelas terbanyak yang dihasilkan oleh pohon-pohon keputusan yang ada pada *random forest*. Dengan melakukan *voting* pada pohon-pohon keputusan yang tersedia membuat akurasi dari *balanced random forest* meningkat. Algoritma ini mengatasi permasalahan pada jumlah data kelasnya yang tidak seimbang atau dengan kata lain jumlah kelas *churn* lebih sedikit dari jumlah kelas tidak *churn*. Sehingga algoritma ini sesuai dengan karakteristik data yang digunakan pada tugas akhir.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah yang diuraikan adalah penentuan prediksi atribut utama yang mempengaruhi terjadinya *churn*, karena karakteristik data *churn* yang *imbalanced class* maka pada tugas akhir ini

mengimplementasikan metode *balanced random forest* yang mengatasi masalah data *imbalanced class* dengan akurasi yang baik [1].

1.3 Tujuan

Tujuan dari penelitian Tugas Akhir ini adalah :

- Membuat perangkat lunak yang mengimplementasikan metode *balanced random forest* untuk *churn prediction*.
- Mengidentifikasi atribut utama yang mempengaruhi *churn prediction*.

1.4 Hipotesa

Hipotesa yang dikemukakan adalah bahwa proses klasifikasi khususnya dengan penerapan metode *balanced random forest* terhadap data pelanggan pada kasus *churn*, dapat mengidentifikasi atribut utama yang mempengaruhi terjadinya *churn* serta dapat mengatasi data yang *imbalance* dengan akurasi yang baik.

1.5 Batasan Masalah

Batasan masalah dari tugas akhir ini adalah :

1. Metode *data mining* yang digunakan adalah klasifikasi *decision tree* dengan menggunakan algoritma *balanced random forest*.
2. Data primer yang digunakan adalah data pelanggan untuk kasus *churn* pada sebuah perusahaan telekomunikasi di Indonesia.
3. Asumsi data *churn* yang digunakan sudah mengalami *preprocessing* terlebih dahulu.
4. Data yang digunakan merupakan *imbalance class*.

1.6 Metodologi penyelesaian masalah

Metodologi yang digunakan dalam Tugas Akhir ini antara lain :

1. Studi Literatur, yang dilakukan dengan mempelajari beberapa literatur (makalah, buku ataupun jurnal) yang berkaitan dengan data *mining*, klasifikasi terutama yang berkaitan dengan *balanced random forest* dan metode untuk menganalisa performansi algoritma klasifikasi tersebut.
2. Pengumpulan dan analisis data yang digunakan untuk mendukung implementasi dan analisis algoritma *balanced random forest*.
3. Analisis kebutuhan dan perancangan perangkat lunak yang akan dibangun. Dalam hal ini digunakan metode berorientasi objek.
4. Implementasi dari rancangan perangkat lunak yang telah dibuat, sehingga dihasilkan perangkat lunak yang dapat digunakan untuk mengklasifikasikan data dengan menerapkan algoritma *balanced random forest*. Dalam pembuatan perangkat lunak ini digunakan bahasa pemrograman matlab.
5. Pengujian dan analisa hasil, melakukan uji kebenaran klasifikasi dan mengukur hasil implementasi berdasarkan parameter akurasi yang dihasilkan serta menganalisis atribut-atribut yang teridentifikasi dari nilai akurasi gini yang terendah.

6. Pengambilan kesimpulan dan penyusunan laporan Tugas Akhir. Pengambilan kesimpulan terhadap analisis hasil penelitian yang dilakukan dan pembuatan laporan tugas akhir untuk mendokumentasikan tahap-tahap pelaksanaan.

1.7 Sistematika Penulisan

Tugas akhir ini akan dilaksanakan dengan memiliki sistematika penulisan sebagai berikut:

BAB I Pendahuluan

Bab ini berisi mengenai penjelasan secara umum dalam tugas akhir, yang meliputi latar belakang masalah, perumusan masalah, tujuan, hipotesa, batasan masalah, dan metodologi penyelesaian masalah.

BAB II Tinjauan Pustaka

Bab ini akan menjelaskan landasan teori yang berkaitan dengan pembangunan perangkat lunak.

BAB III Perancangan dan Implementasi

Bab ini berisi rancangan sistem sesuai dengan kasus yang dipilih, dan dilanjutkan dengan pengimplementasiannya.

BAB IV Pengujian dan Analisis

Bab ini akan menjelaskan tentang pengujian terhadap perangkat lunak dan melakukan analisis terhadap *output* pengujian sebelumnya.

BAB V Kesimpulan dan Saran

Bab ini berisi kesimpulan dari tugas akhir mengenai perangkat lunak yang dibangun dan berisi juga saran bagi pengguna yang memakai atau yang ingin mengembangkan penelitian lebih lanjut.