

1. PENDAHULUAN

1.1 Latar Belakang

Aset perusahaan yang paling bernilai adalah pelanggan setia. Oleh sebab itu, perusahaan akan mencari cara untuk tetap mempertahankan pelanggannya agar tidak berpindah langganan ke perusahaan lain, khususnya pada perusahaan telekomunikasi. *Churn* merupakan permasalahan umum yang sering ditemui di industri telekomunikasi karena memberikan pengaruh pendapatan pada perusahaan [1]. *Churn* adalah keadaan dimana pelanggan tidak berlangganan lagi dengan sebuah perusahaan dalam jangka waktu tertentu.

Dari keseluruhan data pelanggan, pelanggan yang mengalami *churn* menjadi kelas minor yang bersifat *imbalance* [2]. Dalam memprediksi *churn*, masalah yang menantang dan sangat penting adalah ketidakseimbangan dalam distribusi data antara kelas *not churn* dan kelas *churn*. *Imbalance* merupakan keadaan dimana porsi data untuk keperluan *learning* di sebuah kelas (mayor) lebih banyak daripada porsi data untuk *learning* di kelas lain (minor). Kelas minor yang dimaksud disini adalah kelas yang dianggap *churn*. Hanya terdapat sedikit sekali *churn* dari keseluruhan data yang digunakan. Rata-rata pelanggan yang mengalami *churn* hanya sekitar 0,7% dari keseluruhan *dataset* [2]. Untuk menangani permasalahan *imbalance data* dan *churn*, beberapa perusahaan menerapkan metode *data mining* [2] untuk memprediksi perilaku pelanggan di masa depan dengan mengetahui data perilaku pelanggan.

Pada permasalahan *imbalance data* dan *churn* diatas, penulis menerapkan teknik *undersampling* dengan menerapkan algoritma *Neighborhood Cleaning Rule* (NCL) dan RUS [3, 4]. NCL dapat meningkatkan pembersihan data dari metode ENN yang ada didalamnya. Keunggulan dari ENN adalah tetap mempertahankan sebagian data dan akurasi klasifikasi tidak terganggu [4]. Sedangkan penerapan RUS dapat membuat lebih efektif dan cepat dalam proses pelatihan prediksi kelas *imbalance* [3]. Selanjutnya dibuat model klasifikasi menggunakan metode *Weighted Random Forest* (WRF) untuk mengklasifikasikan kelas *churn* dan *not churn*. Setelah model klasifikasi dibangun, maka akan dilakukan analisis pengaruh *undersampling* dan *weight* terhadap nilai *precision*, *recall*, *F1-score*, dan *overall accuracy*. Metode WRF ini diklaim dapat menghasilkan model dengan kinerja yang baik dengan membangun banyak *tree*

(pohon) dan memberikan *weight* pada setiap kelas sehingga dapat membuat performansi model klasifikasi menjadi lebih stabil [2].

1.2 Perumusan Masalah

Dengan mengacu pada latar belakang masalah diatas, maka permasalahan yang akan dibahas dan diteliti adalah sebagai berikut.

1. Bagaimana menganalisis pengaruh hasil *undersampling* pada akurasi model klasifikasi *churn*.
2. Bagaimana menganalisis pengaruh *weight* pada WRF terhadap akurasi model klasifikasi *churn*.

Sedangkan batasan masalah dalam penelitian tugas akhir ini adalah :

1. *Preprocessing* dan pembangunan model klasifikasi dilakukan dengan menggunakan tools *Rstudio*.
2. Penelitian tugas akhir ini hanya membangun model klasifikasi yang dapat dianalisis untuk memprediksi *churn* dan mengukur performansi yang dihasilkan dari model klasifikasi yang telah dibangun.

1.3 Tujuan

Berdasarkan rumusan masalah diatas, maka tujuan dari penelitian tugas akhir ini adalah sebagai berikut.

1. Mengetahui pengaruh hasil *undersampling* terhadap akurasi model klasifikasi *churn*.
2. Mengetahui pengaruh *weight* pada WRF terhadap akurasi model klasifikasi *churn*.

1.4 Metodologi Penyelesaian Masalah

Metode yang digunakan dalam penyelesaian tugas akhir ini dengan langkah kerja adalah sebagai berikut.

1. Studi literatur dilakukan dengan dua tahap, yaitu sebagai berikut :
 - a) Tahap awal dilakukan pencarian referensi dan sumber-sumber lain yang layak yang berhubungan dengan tugas akhir, antara lain *handling imbalance data*,

data mining, klasifikasi, *random forest*, *weighted random forest*, *Rstudio*, dan pengukuran performansi untuk model klasifikasi.

- b) Kemudian dilakukan pendalaman materi mengenai hal-hal yang telah disebutkan diatas.
2. Melakukan *feature selection*, yaitu pengurangan atribut dengan memilih atribut mana saja yang mempengaruhi terjadinya *churn*. Sehingga prediksi *churn* dapat diketahui secara akurat.
3. Mempelajari konsep *undersampling* dan algoritma yang cocok untuk diterapkan pada *undersampling* untuk penanganan *imbalance data*.
4. Menggabungkan dua *dataset* menjadi satu, yaitu *training clean* dan *testing clean*. Kemudian setelah itu, di-*split* kembali menjadi *data training* yang berguna untuk membangun model klasifikasi dan *data testing* yang berguna untuk mengukur performansi model klasifikasi.
5. Melakukan penanganan *imbalance data* menggunakan algoritma *Neighborhood Cleaning Rule (NCL)* dan *Random Undersampling (RUS)*.
6. Membangun model klasifikasi menggunakan metode *WRF*.
7. Menganalisis model klasifikasi menggunakan ukuran keberhasilan yang dinyatakan dalam *precision*, *recall*, dan *F1-score* dan *overall accuracy*.
8. Mengambil kesimpulan dari hasil penelitian (eksperimen) dan menyusun laporan tugas akhir.

1.5 Sistematika Penulisan

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

BAB I PENDAHULUAN

Bagian pendahuluan berisi latar belakang masalah, permasalahan dalam penelitian, tujuan, pembatasan masalah, metodologi penyelesaian masalah, dan sistematika penulisan.

BAB II KAJIAN PUSTAKA

Bagian dasar teori diawali dengan ringkasan hasil tinjauan pustaka (*paper* atau *jurnal*) terkait *research question* atau *problem statement* yang diajukan pada Bab

Pendahuluan. Kemudian ditambahkan dengan teori pendukung yang digunakan dalam penyelesaian masalah dalam tugas akhir. Teori pendukung dapat berasal dari *textbook*, *paper/jurnal*, tesis, dan tugas akhir lain dengan memperhatikan penggunaan teknik acuan sebagai bentuk kejujuran akademik.

BAB III ALUR PEMODELAN

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. Selain itu, pada bab ini juga dijelaskan analisis terhadap hasil pengujian tersebut. Hasil dari kegiatan analisis 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 berisi daftar referensi yang digunakan dalam pembuatan buku tugas akhir ini, dimana minimal terdapat 10 referensi yang digunakan dan seluruh referensi yang ada tercatat diacu dalam buku tugas akhir ini.

LAMPIRAN

Bagian lampiran berisi deskripsi *dataset*, contoh *dataset*, *form* yang berkaitan dengan Tugas Akhir, dan *screenshot* hasil program.