

1. Pendahuluan

1.1 Latar Belakang

Perkembangan teknologi sekarang ini semakin pesat. Kebutuhan akan informasi dan komunikasi bertambah. Telah ditemukan berbagai perangkat teknologi yang memudahkan manusia dalam menyelesaikan masalah informasi dan komunikasi. Perusahaan telekomunikasi di Indonesia sekarang sedang melakukan banyak inovasi untuk melakukan persaingan yang semakin ketat karena *customer* berhak memilih dari banyaknya penyedia layanan yang ada. Persaingan ini penting bagi kelangsungan perusahaan karena berpengaruh kepada pendapatan perusahaan.

Persaingan tersebut mengakibatkan beralihnya *customer* ke perusahaan telekomunikasi lain atau bisa disebut *Churn*. *Churn* adalah pemutusan jasa suatu perusahaan oleh pelanggan karena pelanggan tersebut lebih memilih menggunakan layanan jasa perusahaan kompetitor. Pada persaingan pasar ini dilihat dari pengalaman bahwa setiap tahunnya sekitar 30-35% laju *churn* dan membutuhkan 5-10 kali usaha dan biaya untuk menambah customer baru daripada mempertahankan yang sudah ada. Untuk itu industri telekomunikasi lebih memilih untuk mempertahankan *customer*. [1]

Dalam mempertahankan *customer*, perusahaan telekomunikasi membutuhkan cara untuk memprediksi dalam mengetahui resiko *customer* kapan akan menjadi churn. Ketika memprediksi *churn*, terdapat berbagai teknik *data mining* dapat diterapkan. Salah satunya adalah model prediksi *churn*. Perusahaan besar harus mengimplementasikan model *churn prediction* tersebut untuk dapat mendeteksi atau mengetahui kemungkinan adanya *churn* sebelum mereka secara efektif meninggalkan perusahaan tersebut yang bisa berpengaruh terhadap pendapatan perusahaan. Namun, pada dataset pelanggan yang didapat dari perusahaan tersebut seringkali masih tidak seimbang. Data tidak seimbang merupakan jika kelas mayoritas (*not churn*) lebih banyak dari kelas minoritas (*churn*). Data yang tidak seimbang akan mengakibatkan proses klasifikasi yang cenderung keliru

pada kelas minoritas (*churn*). Oleh karena itu, untuk keberhasilan pada proses klasifikasi tersebut diperlukan adanya penanganan masalah kelas tidak seimbang dahulu sebelum membuat model prediksi.

Pada permasalahan *imbalance data* dan *churn* diatas, penulis menerapkan teknik *underbagging* dan *logistic regression*. Metode *underbagging* merupakan metode penggabungan (*ensemble*) antara *undersampling* dan *bagging*. Metode *Logistic Regression* termasuk dalam *predictive modeling*, karena dalam statistika digunakan untuk prediksi probabilitas kejadian suatu peristiwa dengan mencocokkan data pada fungsi logit kurva logistik dalam suatu kelas. [2] Metode ini bisa digunakan untuk penanganan kasus prediksi *churn* karena pada kasus data pelanggan PT. Telekomunikasi dapat diketahui bahwa tipe data variabel respon (Y) adalah nominal. Terdapat dua kemungkinan, yaitu *churn* dan *not churn* yang menghasilkan *binary values* seperti angka 0 adalah *non churn* dan angka 1 mempresentasikan *churn*. [2]

Dengan penelitian menggunakan metode penanganan kelas tidak seimbang *underbagging* dan teknik klasifikasi *logistic regression*, diharapkan dapat melihat pengaruh metode tersebut pada akurasi dan *F1-measure* agar perusahaan Telekomunikasi dapat mencegah terjadinya *churn* pelanggan.

1.2 Perumusan Masalah

Dilihat dari latar belakang masalah, maka perumusan masalah yang akan dibahas dan diteliti pada Tugas Akhir ini adalah :

1. Bagaimana menganalisis pengaruh penanganan *imbalance data* pada performansi model prediksi *churn*.
2. Bagaimana menganalisis pengaruh model prediksi *churn* dari data hasil penanganan *imbalance data*.

Batasan masalah pada penelitian Tugas Akhir ini adalah:

1. Data yang digunakan adalah salah satu dataset pelanggan WITEL PT. Telekomunikasi Regional 7.

2. Menggunakan dataset berjumlah 200.381 pelanggan dengan 53 atribut.
3. Data yang digunakan asumsi semua data numerik.
4. Data *training* dan data *testing* dilakukan penanganan *imbalance data*.
5. Setelah model klasifikasi dibangun, maka akan dilakukan analisis pengaruh *underbagging* dan *logistic regression* terhadap nilai *precision*, *recall*, *F1-measure*, dan *overall accuracy*.

1.3 Tujuan

Berdasarkan rumusan masalah yang ada, maka tujuan penelitian ini adalah sebagai berikut:

1. Menganalisis pengaruh hasil metode *ensemble*, yaitu *undersampling* dan *bagging (underbagging)* terhadap performansi klasifikasi *churn*.
2. Menganalisis pengaruh hasil klasifikasi *Logistic Regression* terhadap performansi model klasifikasi *churn*.
3. Menganalisis performansi model *churn prediction* yang didapat dari penelitian dalam bentuk *precision*, *recall*, *F1-measure*, dan *overall accuracy*.

1.4 Metodologi Penyelesaian Masalah

Tahap metodologi yang digunakan untuk menyelesaikan masalah dalam Tugas Akhir ini adalah :

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, *logistic regression*, *matlab*, 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 *underbagging* dan algoritma yang cocok untuk diterapkan pada *underbagging* 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 *ensemble Random Undersampling (RUS)* dan *Bagging*.
6. Membangun model klasifikasi menggunakan metode *logistic regression*.
7. Menganalisis model klasifikasi menggunakan ukuran keberhasilan yang dinyatakan dalam *precision*, *recall*, dan *F1-measure* 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.