

1. Pendahuluan

1.1 Latar Belakang Masalah

Industri penyedia jasa telekomunikasi merupakan industri yang terus berkembang dan selalu dibutuhkan masyarakat. Dengan semakin banyaknya jumlah perusahaan telekomunikasi baik penyedia layanan GSM (*Global System Mobile*) maupun CDMA (*Code Division Multiple Access*), masing-masing akan saling menerapkan strategi untuk memperebutkan perhatian pelanggan dan menguasai pasar. Berbagai cara dilakukan dalam mendukung strategi tersebut, seperti: penerapan tarif murah, penyediaan layanan/ fitur khusus kepada pelanggan, undian berhadiah, bonus pulsa, jaminan minimalisasi *call drop*, ataupun lainnya. Semua hal tersebut bertujuan untuk mempertahankan atau menambah *revenue* yang didapat oleh perusahaan, serta landasan bahwa biaya untuk mempertahankan pelanggan akan lebih murah dibandingkan biaya untuk menarik pelanggan baru [18].

Churn lahir dari fenomena yang terjadi di atas. Pengertian *churn* adalah pemutusan jasa suatu perusahaan oleh pelanggan karena pelanggan tersebut lebih memilih menggunakan layanan jasa perusahaan kompetitor [5,6]. *Churn* sangat berpotensi terjadi pada operator telekomunikasi selular karena kemudahan untuk berganti layanan dari satu operator ke operator yang lainnya [5]. *Churn* harus diwaspadai oleh perusahaan karena dengan bertambahnya jumlah *churn* akan semakin mengakibatkan penurunan *revenue* perusahaan.

Prediksi *churn* (*Churn Prediction*) merupakan salah satu jenis *task* pada *data mining*, yaitu klasifikasi. *Data mining* sendiri merupakan ilmu yang berkembang akibat semakin menumpuknya kuantitas data dan diperlukan suatu teknik untuk mendapatkan informasi dari tumpukan data tersebut. *Churn* harus bisa diprediksi oleh perusahaan dalam rangka penerapan strategi untuk mempertahankan pelanggan yang potensial *churn* dan agar pemberian retensi atau wujud layanan tambahan lainnya dapat tepat pada sasaran. Kebutuhan aplikasi prediksi *churn* bagi perusahaan telekomunikasi yaitu mampu menghasilkan daftar nama-nama pelanggan potensial *churn* diurutkan secara *descending* berdasarkan nilai *confidence*-nya (bobot *churn*). Adapun parameter *predictors* yang digunakan yaitu data demografis, data *traffic* (data *Call Detail Record*), dan data *payment* (pembayaran) [7].

Permasalahan prediksi *churn* yang akan dipaparkan adalah *imbalance/unbalance class*. *Imbalance class* merupakan salah satu masalah yang vital dalam klasifikasi dimana pada *training set* terdapat ketidakseimbangan yang tajam tentang jumlah *record* suatu kelas dengan *record* pada kelas lainnya. Kelas dengan jumlah *record* lebih besar disebut dengan kelas mayor, sedangkan kelas dengan jumlah *record* lebih kecil disebut dengan kelas minor. *Imbalance* menjadi sebuah masalah karena prediksi untuk kelas minor lebih menarik daripada prediksi untuk kelas mayor. Dari banyak penelitian didapat kesimpulan bahwa pada kondisi data *imbalance*, *classifier* cenderung membias *record* yang seharusnya

merupakan kelas minor menjadi kelas mayor dan kemungkinan menganggap kelas minor hanya sebagai *outlier* (untuk data minor dengan jumlah sangat kecil).

Metode yang akan digunakan adalah *K-Nearest Neighbor* untuk *imbalance class* yaitu modifikasi metode *K-Nearest Neighbor* yang berusaha lebih ‘memihak’ pada kelas minor. *K-Nearest Neighbor* dianggap sebagai salah satu algoritma yang ampuh dalam mengatasi permasalahan klasifikasi. Hal ini dikarenakan objek diklasifikasikan sesuai label kelas terbanyak yang terdapat pada area k buah tetangga terdekat (k buah *data learning*), sehingga bersifat *local approximation*. Penentuan tetangga terdekat dihitung dengan rumus perhitungan jarak antar objek tertentu seperti *Euclidean Distance*. Dengan modifikasi KNN yang dilakukan, diharapkan akan memperbaiki akurasi dari prediksi pada data minor terutama jika dibandingkan dengan penggunaan *classifier lazy learner* IBk (nama lain KNN) pada *tool* Weka 3.5.6 serta *classifier* populer lain dari *tool* Clementine 10.1.

1.2 Perumusan Masalah

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

1. Bagaimana menganalisa dan mengimplementasikan metode *K-Nearest Neighbor* untuk *imbalance class* dalam kasus *churn prediction*.
2. Bagaimana akurasi yang didapat dari *K-Nearest Neighbor* untuk *imbalance class* jika dibandingkan dengan *classifier lazy learner* IBk pada *tool* Weka 3.5.6 serta dengan *classifier* populer lain dari *tool* Clementine 10.1 dalam bentuk *top decile lift*, *lift curve* dan *gini coefficient*.

Batasan masalah yang akan dibahas dalam penelitian tugas akhir ini adalah :

1. *Preprocessing* berupa *filter nominal*(kategoris) *to binary* dan normalisasi ditangani oleh sistem, sedangkan *preprocessing* lainnya menggunakan Weka 3.5.6, Clementine 10.1 dan Microsoft Excel.
2. *Postprocessing* menggunakan Weka 3.5.6 dan Microsoft Excel.
3. Tidak membahas penggunaan *tool* Clementine 10.1 dan *Weka* 3.5.6 dalam pengujian sebagai pembanding terhadap akurasi dari perangkat lunak yang dihasilkan.

1.3 Tujuan

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

1. Mengimplementasikan metode *K-Nearest Neighbor* untuk *imbalance class* dalam kasus *churn prediction*.
2. Menganalisis akurasi hasil klasifikasi yang dihasilkan oleh perangkat lunak *Churn Prediction* menggunakan algoritma *K-Nearest Neighbor* untuk *imbalance class*.
3. Membandingkan akurasi hasil klasifikasi dari algoritma *K-Nearest Neighbor* untuk *imbalance class* dengan hasil klasifikasi oleh *classifier*

lazy learner IBk pada *tool* Weka 3.5.6 serta dengan *classifier* populer lain dari *tool* Clementine 10.1 dalam bentuk *top decile lift*, *lift curve* dan *gini coefficient*.

1.4 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 yang layak dan berhubungan dengan *imbalance class problem*, *churn prediction*, dan *K-Nearest Neighbor*.
 - b. Pendalaman materi, mempelajari dan memahami materi yang berhubungan dengan tugas akhir.
 - c. Pencarian manual dan referensi tentang penggunaan *tool* Clementine 10.1 dan Weka 3.5.6.
2. Analisis Permasalahan :
 - a. Mencari data pelanggan dan memahaminya.
 - b. Mempelajari konsep dari *K-Nearest Neighbor* untuk *imbalance class* yang akan digunakan dalam implementasi perangkat lunak.
 - c. Menganalisis algoritma *K-Nearest Neighbor* untuk *imbalance class* dalam perancangan perangkat lunak
 - d. Simulasi dan analisa data dengan *classifier lazy learner* IBk pada *tool* Weka 3.5.6 serta dengan *classifier* populer lain dari *tool* Clementine 10.1.
3. Mengumpulkan *requirement* terhadap perangkat lunak yang akan dibangun.
4. Melakukan disain/ perancangan perangkat lunak dengan teknik *object oriented*.
5. Melakukan implementasi perancangan perangkat lunak.
6. Melakukan pengujian perangkat lunak dengan memasukkan data yang sudah di-*preprocessing* serta menganalisis hasil keluaran program.
7. Menganalisis hasil prediksi dari perangkat lunak dengan hasil prediksi oleh *classifier lazy learner* IBk pada *tool* Weka 3.5.6 serta dengan *classifier* populer lain dari *tool* Clementine 10.1.
8. Pengambilan kesimpulan dan penyusunan laporan tugas akhir.