

1. Pendahuluan

Pelanggan merupakan salah satu aset perusahaan yang sangat berharga. Karena kepuasan dan jumlah pelanggan dapat menentukan keberhasilan dari perusahaan tersebut, tidak terkecuali perusahaan manapun. Pada contohnya di perusahaan telekomunikasi, jika pelayanan yang diberikan oleh perusahaan baik maka pelanggan akan terus memakai jasa yang diberikan atau bahkan dapat menambah jumlah pelanggan. Sebaliknya jika pelayanan dirasa kurang memuaskan, maka dapat dikhawatirkan pelanggan tersebut akan beralih ke provider lain. Agar dapat menghindari hal tersebut, maka diperlukan adanya suatu prediksi yang berguna untuk melihat pelanggan mana yang memiliki potensi untuk melakukan *churn*.

Topik dan Batasannya

Pada tugas akhir ini, digunakan metode *Random Undersampling* (RUSBoost), yang akan mengurangi jumlah dari kelas mayoritas sampai setara dengan jumlah kelas minoritas secara acak. Kemudian digunakan metode SVM sebagai *classifier* dan yang terakhir metode *boosting* untuk meningkatkan akurasi dari klasifikasi.

Ketidakseimbangan data merupakan hal yang sering dijumpai pada proses klasifikasi. Data akan disebut tidak seimbang apabila distribusi antar kelas mayoritas dan minoritas yang tidak seimbang. Masalah ini dapat memperburuk kinerja dari klasifikasi, yang dimana menganggap distribusi kelas sudah setara dan tidak menghiraukan kelas minoritas di *dataset* yang tidak seimbang [1, 2]. Metode yang dapat digunakan untuk menghilangkan masalah ini adalah *sampling*. *Sampling* sendiri terdapat 2 metode, yaitu dengan menambahkan fitur dari kelas minoritas (*oversampling*) dan mengurangi fitur dari kelas mayoritas (*undersampling*). Kedua metode ini sama-sama memiliki kekurangan, jika menggunakan *undersampling* informasi-informasi penting dapat hilang karena menghapus fitur-fitur dari data yang ada, tetapi metode ini dapat mempercepat waktu karena pengurangan jumlah data. Disisi lain, penggunaan *oversampling* dapat mengakibatkan *overfitting* dan memakan waktu yang lebih lama [3].

Setelah *dataset* sudah seimbang kemudian dapat melanjutkan proses klasifikasi. Metode yang digunakan tugas akhir ini adalah *Support Vector Machine* (SVM). Hasil klasifikasi dari SVM kemudian akan melalui proses *boosting*. *Boosting* sendiri merupakan suatu metode yang dapat menaikkan akurasi dari prediksi dengan cara, memanggil fungsi secara iterasi dan mengkombinasikan output dari tiap fungsi dengan adanya bobot, sehingga total *error* dari prediksi dapat diminimalisir [4].

Batasan dari tugas akhir ini adalah *dataset training* yang berjumlah 99809 data, dengan rincian data mayoritas sebanyak 96003 data dan data minoritas sebanyak 3761 data.

Tujuan

Tujuan dari tugas akhir ini adalah untuk membangun sistem klasifikasi terhadap pelanggan yang memiliki kecenderungan untuk melakukan aksi *churn*, dengan menggunakan metode *random undersampling* untuk menyeimbangkan data mayoritas dan data minoritas, kemudian menggunakan metode SVM untuk melakukan klasifikasi.