

1. Pendahuluan

1.1 Latar Belakang

Perkembangan ilmu pengetahuan dan teknologi telah mengakibatkan perkembangan data di dunia semakin kompleks. Untuk menangani data-data tersebut diperlukan sistem pengolahan data yang baik. Masalah yang timbul dalam pengolahan data adalah *imbalance class*. *Imbalance class* merupakan ketidakseimbangan dalam jumlah data *training* antara dua kelas yang berbeda, salah satu kelasnya merepresentasikan jumlah data yang sangat besar (*majority class*) sedangkan kelas yang lainnya merepresentasikan jumlah data yang sangat kecil (*minority class*). *Imbalance class* biasanya terjadi pada kasus yang anomali. Contoh-contoh kasus untuk *imbalance class* adalah *detecting fraudulent transaction*, *network intrusion detection*, *Web Mining*, dan *direct marketing*.

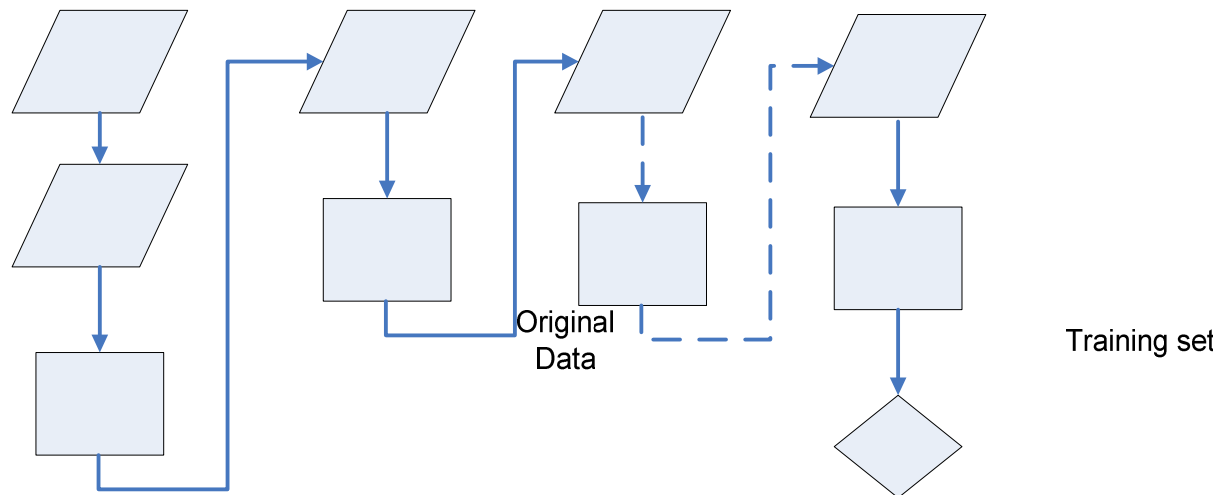
Salah satu cara untuk memprediksi data adalah dengan metode klasifikasi. Klasifikasi adalah suatu metode untuk memprediksi kelas dari suatu data yang belum diketahui sebelumnya dengan mengacu pada model yang telah dibuat dari sekumpulan data *training*. Metode klasifikasi kurang tepat untuk digunakan dalam kasus *imbalance class* karena hanya sebatas untuk kasus yang distribusi kelas labelnya seimbang. Untuk itu perlu dikembangkan suatu pendekatan untuk meningkatkan hasil performansi dari metode klasifikasi tersebut.

Salah satu cara untuk menangani kasus *imbalance class* adalah dengan menggunakan algoritma RareBoost-1. Dibandingkan algoritma lainnya seperti AdaBoost, algoritma ini dapat meningkatkan performansi yaitu keakuratan prediksi *minority class* dalam kasus *imbalance class*[6]. Dalam proses pelatihannya, kelas yang jumlahnya lebih sedikit akan mendapatkan perhatian lebih sehingga diharapkan keseluruhan kelas yang jumlahnya sedikit dapat diprediksikan secara maksimal. Analisa lebih dalam mengenai algoritma RareBoost-1 akan dibahas dalam Tugas Akhir ini.

1.2 Perumusan Masalah

Boosting merupakan salah satu *ensemble-method* yang bisa digunakan untuk meningkatkan performansi klasifikasi dari *classifier*. Salah satu penanganan kasus *imbalance class* adalah dengan menggunakan algoritma RareBoost-1. RareBoost-1 merupakan salah satu algoritma *Boosting* untuk menangani *minority class* sehingga keakuratan hasil prediksi mendekati akurat.

Gambaran umum mengenai Boosting dapat dilihat pada Gambar 1.1 :



Gambar 1-1: Boosting

Penjelasan dari Gambar 1.1 adalah sebagai berikut :

1. *Original data* dibagi menjadi *training data* dan *test data*.
2. Setelah itu metode klasifikasi digunakan untuk memproses *training data* agar terbentuk sebuah model yang nantinya akan digunakan untuk memprediksi *test data*.
3. Untuk memperoleh hasil prediksi yang akurat diperlukan beberapa iterasi dan di dalam setiap iterasi digunakan algoritma RareBoost-1.
4. Setelah itu diadakan voting untuk menentukan hasil prediksi dan hasil dari voting tersebut akan digunakan untuk menghitung performansi dilihat dari parameter *Recall*, *Precision*, dan *F-Measure*.

Berdasarkan penjelasan tersebut maka dapat dirumuskan beberapa permasalahan diantaranya :

1. Bagaimana mengimplementasikan algoritma RareBoost-1 untuk menangani kasus *imbalance class*.
2. Bagaimana performansi algoritma RareBoost-1 untuk menangani kasus *imbalance class*.

Sedangkan batasan masalah dari pembahasan Tugas Akhir ini adalah :

1. Data yang digunakan untuk analisa adalah data yang *supervised* (memiliki *class label*)
2. Tidak menangani tahap *preprocessing*, data latih dan data uji telah bersih dari *noise*.
3. Data yang ditangani hanya 2 (dua) *class problem*.
4. Data yang dianalisis yaitu data riil yaitu data Churn, data UCI, dan data PAKDD 2006.
5. Algoritma yang digunakan adalah RareBoost-1.

1.3 Tujuan

Tujuan pembuatan Tugas Akhir ini adalah :

1. Membuat sebuah perangkat lunak dengan mengimplementasikan algoritma RareBoost-1 untuk mendeteksi *minority class* pada kasus *imbalance class*.
2. Menganalisis performansi algoritma RareBoost-1 pada kasus *imbalance class* dengan menggunakan parameter *Recall*, *Precision* dan *F-Measure*.

1.4 Metodologi Penyelesaian Masalah

Metodeologi yang akan digunakan untuk menyelesaikan tugas akhir ini adalah:

1. Studi Literatur.
Studi Literatur dengan mempelajari literatur-literatur yang relevan dengan permasalahan yang meliputi : melakukan studi pustaka dan referensi mengenai data mining, metode klasifikasi, *imbalance class*, algoritma RareBoost-1, dan Boosting secara umum.
2. Pengumpulan Data.
Mencari data yang akan digunakan sebagai studi kasus.
3. Analisis dan Perancangan Perangkat Lunak.
Menganalisis permasalahan yang akan ditangani, menganalisis metode yang akan digunakan untuk menyelesaikan permasalahan.
4. Implementasi Sistem.
Melakukan *coding* dengan membangun perangkat lunak dengan menggunakan algoritma RareBoost-1.
5. Pengujian Sistem dan Analisis Hasil.
Melakukan proses pengujian terhadap data-data yang digunakan sebagai studi kasus sehingga keakuratan hasil dan performansi algoritma RareBoost-1 dapat terlihat, serta melakukan analisis terhadap kelebihan dan keterbatasan algoritma RareBoost-1 pada permasalahan *imbalance class*.
6. Penyusunan laporan tugas akhir dan kesimpulan akhir.