

1. Pendahuluan

1.1 Latar belakang

Imbalance class merupakan ketidakseimbangan distribusi *class label* pada suatu data latih. Karakteristik dari *imbalance class* adalah pada salah satu classnya direpresentasikan dengan jumlah data yang sangat besar (*majority class*) sedangkan class yang lainnya direpresentasikan dengan jumlah yang sangat kecil (*minority class*). *Imbalance class* biasanya ditemukan pada kasus-kasus anomali misalnya *fraud detection*, *network intrusion detection*, *curb prediction*, dan *thyroid disease*. Kasus-kasus seperti itu kalau tidak segera diatasi bisa menyebabkan kerugian finansial, seperti pada industri telekomunikasi Amerika Serikat kerugian yang diakibatkan fraud antara 4%-6% dari pendapatan [2].

Imbalance class dalam konteks data mining perlu dipelajari karena *minority class* lebih sulit untuk diprediksi daripada *majority class*. Padahal, terkadang *class* yang minoritas inilah yang mempunyai informasi yang sangat berharga. Oleh karena itu, bagaimana memprediksi *class label* yang tepat sangat diperlukan agar kita memperoleh hasil dengan nilai *true positif* yang tinggi.

Pada tugas akhir ini, permasalahan *imbalance class* akan coba ditangani dengan algoritma SMOTEBoost (*Synthetic Minority Over-sampling TEchnique* yang dikombinasikan dengan teknik boosting) oleh karena algoritma SMOTE dapat meningkatkan akurasi dari *minority class* dan penggunaan boosting tidak mengorbankan akurasi dari data set secara keseluruhan [2].

1.2 Perumusan masalah

Dari uraian latar belakang diatas maka dapat dirumuskan beberapa permasalahan yang harus diselesaikan yaitu :

1. Bagaimana memprediksi *minority class* pada kasus *imbalance class*.
2. Bagaimana menentukan prediksi *class label* yang tepat sehingga didapatkan nilai *True Positif* yang tinggi dan *False Positif* yang kecil.
3. Bagaimanakah performansi algoritma SMOTEBoost pada kasus *imbalance class*.

Sedangkan batasan masalah dari pembahasan Tugas Akhir ini diantaranya :

1. Data yang dianalisis yaitu data *Thyroid Disease (sick)*, data PAKDD 2006, data operator ilegal Telkom pada periode tertentu, dan data prediksi *churn* PT. Telkom.
2. Tidak membahas tahap *preprocessing* secara detail.

1.3 Tujuan

Berdasarkan pada rumusan masalah yang telah didefinisikan, maka tujuan tugas akhir ini adalah :

1. Membangun sebuah perangkat lunak dengan mengimplementasikan algoritma SMOTEBoost untuk memprediksi *minority class* pada kasus *imbalanced class*.
2. Menerapkan algoritma SMOTE pada pendekatan Boosting, untuk mendapatkan hasil *True Positif (recall)* yang tinggi dan nilai *False Positif* yang kecil.
3. Menganalisis performansi algoritma SMOTEBoost 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 : studi pustaka dan referensi mengenai data mining, klasifikasi, *imbalanced class*, SMOTE, dan Boosting.
2. Pengumpulan Data.
Mencari data yang akan digunakan sebagai studi kasus.
3. Analisis dan Perancangan Perangkat Lunak.
Menganalisis dan merancang perangkat lunak yang akan digunakan.
4. Implementasi system.
Mengimplementasikan perangkat lunak yang telah didesain.
5. Pengujian sistem dan analisis hasil.
Melakukan proses pengujian terhadap keakuratan hasil dan performansi algoritma SMOTEBoost pada permasalahan *imbalanced class* berdasarkan parameter *recall*, *precision*, dan *F-Measure*.
6. Penyusunan laporan tugas akhir dan kesimpulan akhir.