

BAB I

PENDAHULUAN

1.1 Latar Belakang

Akibat perkembangan teknologi yang demikian cepat, kebutuhan akan data yang ada saat ini sudah berkembang menjadi bagaimana mencari pola tersembunyi yang terdapat di dalam data untuk memperoleh peluang-peluang baru. Pengenalan pola dalam data dipelajari dalam disiplin ilmu *data mining*.

Dalam *data mining* terdapat berbagai metode untuk menemukan pola tertentu dalam data. Salah satu metode yang digunakan ialah klasifikasi. Dalam klasifikasi data dianggap sebagai kumpulan data tunggal, dimana setiap data tunggal terdiri dari beberapa atribut dan satu atribut kelas. Selanjutnya dari kumpulan data tersebut akan dicari sebuah model untuk atribut kelas sebagai fungsi dari nilai atribut lainnya. Tujuan akhirnya ialah agar data yang sebelumnya *invisible* dapat diberikan kelas yang sesuai dengan seakurat mungkin. Permasalahan mulai timbul ketika data pada dunia nyata berkembang menjadi bersifat *imbalance*, yaitu suatu kondisi dimana terdapat ketidakseimbangan atau terjadi ketimpangan dalam jumlah data dimana salah satu kelas dalam data tersebut merepresentasikan kasus yang bersifat jarang terjadi sehingga jumlah datanya relatif sedikit dibandingkan dengan jumlah data yang merepresentasikan kasus normal. Beberapa metode klasifikasi yang ada sekarang ini lebih ditujukan untuk kasus yang jumlah data tiap kelasnya seimbang, dengan tujuan akhir memaksimalkan akurasi keseluruhan, sehingga metode tersebut tidak dapat memprediksi data *imbalance* secara maksimal. Salah satu pendekatan yang dapat digunakan untuk memprediksi data *imbalance* ialah menggunakan *cost sensitive learning*.

Cost sensitive learning merupakan pendekatan pembelajaran dalam *data mining*. Ide dasar dari *cost sensitive learning* berasal dari sifat dasar manusia, dimana dalam mengambil suatu keputusan, keputusan yang diambil yang mengandung resiko seminimal mungkin. Dalam *cost sensitive learning* yang menjadi perhatian bukanlah mencapai klasifikasi yang seakurat mungkin atau tingkat kesalahan seminimal mungkin dalam mengelompokkan data, melainkan bagaimana meminimalkan *misclassification cost*.

Dalam Tugas akhir ini, masalah yang diambil adalah bagaimana penerapan *AdaCost* sebagai metode *cost sensitive learning* dan Naive Bayes sebagai metode klasifikasi untuk memprediksi data *imbalance*. Dalam *AdaCost*, kelas yang jumlahnya lebih kecil akan diberi *misclassification cost* yang lebih besar dibanding data yang jumlahnya lebih besar sehingga kelas tersebut mendapatkan perhatian lebih walaupun jumlahnya sangat sedikit sehingga diharapkan keseluruhan kelas yang jumlahnya kecil dapat diprediksikan secara maksimal.

2.1 Perumusan Masalah

Berdasarkan latar belakang yang dikemukakan di atas, maka masalah yang akan dibahas dalam tugas akhir ini adalah bagaimana mengimplementasikan *cost sensitive learning* dalam mengklasifikasikan data *imbalance* menggunakan

metode pembelajaran *AdaCost* yang diharapkan dapat meningkatkan prediksi klasifikasi oleh metode klasifikasi *Naive Bayes*.

3.1 Tujuan

Tujuan dari penulisan Tugas Akhir dengan judul ” **Penggunaan Cost Sensitive Learning Untuk Klasifikasi Data Dengan Metode Naive Bayes**” yaitu :

1. Mengimplementasikan pendekatan *cost sensitive learning* dalam pembuatan klasifikasi menggunakan metode *AdaCost*.
2. Menganalisa hasil penerapan metode *AdaCost* pada prediksi klasifikasi data *imbalance* dengan metode klasifikasi *Naive Bayes* menggunakan parameter *recall*, *presicion*, *F-Measure* dan *total misclassification cost*.

4.1 Batasan Masalah

Dalam penyusunan tugas akhir ini permasalahan dibatasi dalam beberapa hal yaitu:

1. Data yang digunakan telah melalui tahap *preprocessing* data sehingga bebas dari noise dan data yang tidak konsisten.
2. Sistem menerima masukan berupa sekumpulan data yang kelasnya bertipe nominal.
3. Data yang digunakan terutama yang bersifat *imbalance* dan sudah tersedia dalam bentuk tabel dengan atribut kelas berada pada urutan paling akhir.
4. *Misclassification cost* untuk kelas *imbalance* selalu lebih besar daripada *misclassification cost* untuk kelas bukan *imbalance*.

5.1 Metodologi Penyelesaian Masalah

Metode yang akan digunakan untuk menyelesaikan tugas akhir ini adalah :

- a. Studi Literatur
Mempelajari dasar teori dan literatur-literatur tentang data mining, *imbalance class*, *Cost Sensitive Learning*, klasifikasi *Naive Bayes* , dan *AdaCost*.
- b. Studi Pengembangan Aplikasi
Bertujuan untuk menentukan metodologi pengembangan perangkat lunak yang digunakan dengan pendekatan terstruktur dan melakukan analisa perancangan
- c. Desain dan Implementasi Perangkat Lunak
Bertujuan untuk melakukan implementasi metode pada perangkat lunak sesuai dengan analisa perancangan yang telah dilakukan.
- d. Pengujian dan Analisa Perangkat Lunak

Dalam tahap ini akan diuji dan dianalisis hasil penerapan metode *AdaCost* terhadap prediksi klasifikasi *Naive Bayes* yang dibuat untuk berbagai kasus yang disediakan dan mencari kesalahan-kesalahan yang masih muncul dalam pengimplementasian.

- e. Pengambilan kesimpulan dan penyusunan laporan

6.1 Sistematika Penulisan

Sistematika penulisan tugas akhir ini adalah sebagai berikut :

BAB I PENDAHULUAN

Berisi latar belakang pembuatan tugas akhir, perumusan masalah, tujuan, batasan masalah, metodologi penyelesaian masalah dan sistematika penulisan

BAB II LANDASAN TEORI

Pada bab ini dikemukakan berbagai teori yang mendukung penyusunan tugas akhir, antara lain meliputi teori tentang data mining, *imbalance class*, *Cost Sensitive Learning*, klasifikasi *Naive Bayes*, dan *boosting* (meliputi *AdaBoost* dan *AdaCost*).

BAB III ANALISIS KEBUTUHAN DAN PERANCANGAN SISTEM

Menjelaskan tentang proses analisa masalah dan kebutuhan perangkat lunak. Berisi rancangan menggunakan notasi-notasi Diagram Aliran Data (DAD).

BAB IV IMPLEMENTASI DAN ANALISA HASIL PENGUJIAN

Implementasi yang detail mengenai sistem yang dibangun berdasarkan rancangan pada Bab III serta hasil analisis pengujian terhadap beberapa skenario tes.

BAB V KESIMPULAN DAN SARAN

Berisi kesimpulan akhir dan saran pengembangan dari tugas akhir.