

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi yang seperti sekarang ini menjadikan penumpukan data yang sangat banyak. Namun data tersebut seakan-akan hanya seperti tumpukan sampah yang tidak berguna. Pada dasarnya, data-data tersebut dapat dimanfaatkan seperti halnya dalam data mining. Data mining adalah proses yang memperkerjakan satu atau lebih teknik *machine learning* untuk menganalisis dan mengekstraksi *knowledge* secara otomatis[1]. Salah satu teknik data mining yang sering kita jumpai adalah klasifikasi.

Secara singkat, klasifikasi merupakan pengelompokan obyek berdasarkan kelompoknya atau yang biasanya disebut dengan kelas (*class*)[2]. Dalam teknik klasifikasi terdapat beberapa algoritma yang dapat digunakan, diantaranya yaitu *naïve bayes* dan *support vector machine*. Namun algoritma-algoritma tersebut masih banyak kekurangan salah satunya dalam mengatasi *imbalanced dataset*.

Klasifikasi dengan *imbalanced dataset* adalah topik penelitian populer dalam *machine learning* dan data mining, karena terdapat banyaknya dataset yang tidak seimbang[3]. Pada kenyataannya, hampir semua dataset memiliki distribusi kelas yang miring atau biasa disebut *imbalanced data*[4]. Dalam menangani *imbalanced data* terdapat 3 pendekatan, yaitu pendekatan level data, pendekatan level algoritma dan *Cost-sensitive*[5]. Fokus penelitian ini yaitu pada pendekatan level algoritma dengan menggunakan Teknik ensemble.

Teknik ensemble merupakan kombinasi dari *individual classifiers* dengan

menggabungkan dan mengevaluasinya melalui berbagai cara. Terdapat beberapa Teknik ensemble diantaranya *bagging*, *boosting* dan *stacking*[6]. Namun saat ini masih sedikit penelitian tentang teknik ensemble. Sehingga diperlukan penelitian lebih lanjut untuk menentukan metode terbaik dalam mengatasi *imbalanced dataset*, terutama dalam menggunakan Teknik ensemble.

Berdasarkan uraian diatas, maka diadakan penelitian ini dengan tujuan menganalisis pengaruh teknik ensemble terhadap akurasi dari algoritma klasifikasi yang sudah ditentukan.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka dapat dirumuskan suatu permasalahan yaitu apakah ada pengaruh teknik ensemble terhadap akurasi pada algoritma klasifikasi?

1.3 Hipotesis Penelitian

Adapun hipotesis penelitian yang diajukan:

H_0 : Tidak terdapat pengaruh pada penerapan teknik ensemble terhadap akurasi algoritma klasifikasi.

H_1 : Terdapat pengaruh pada penerapan teknik ensemble terhadap akurasi algoritma klasifikasi.

1.4 Batasan Masalah

Adapun Batasan masalah yang diterapkan pada penelitian ini antara lain adalah:

1. Penelitian ini terfokus pada algoritma klasifikasi yaitu *naïve bayes* dan *support vector machine*.

2. Data set yang dipakai berupa 10 imbalanced dataset dengan *binary class* yang diambil dari KEEL, UCI, dan datahub.io dengan jumlah atribut 7 sampai dengan 10.
3. Teknik ensemble yang digunakan pada penelitian ini merupakan *Adaptive boosting* dan *bagging*.

1.5 Maksud dan Tujuan Penelitian

Berdasarkan latar belakang yang diteliti, maka maksud dan tujuan dari penelitian ini adalah “Menganalisis pengaruh Teknik ensemble terhadap akurasi pada algoritma klasifikasi yang sudah ditentukan”.

1.6 Manfaat Penelitian

Dengan penelitian ini diharapkan dapat memberikan manfaat bagi peneliti untuk mengetahui pengaruh Teknik ensemble terhadap akurasi pada algoritma klasifikasi serta dapat menjadikan referensi untuk penelitian selanjutnya.

1.7 Metode Penelitian

1.7.1 Metode Pengumpulan Dataset

Dataset yang digunakan pada penelitian berupa 10 imbalanced dataset dengan *binary class* yang diambil dari (KEEL), UCI Machine Learning, dan datahub.io dengan jumlah atribut 7 sampai dengan 10.

1.7.2 Pre-processing

Sebelum dilakukan pengklasifikasian, dilakukan data *preprocessing*. Pada tahap ini dilakukan pembersihan data salah satunya dengan mengisi nilai-nilai yang hilang atau kosong.

1.7.3 Implementasi Algoritma Klasifikasi

Pada tahap ini dilakukan pengklasifikasian terhadap dataset menggunakan algoritma klasifikasi yang sudah ditentukan serta diterapkan juga Teknik ensemble yaitu *bagging* dan *boosting*.

1.7.4 Evaluasi Hasil Klasifikasi

Data yang sudah diklasifikasikan menggunakan algoritma yang sudah ditentukan dan telah diterapkan Teknik ensemble selanjutnya dilakukan evaluasi dengan menggunakan *Confusion matrix*.

1.7.5 Uji Statistik

Pada tahap ini dilakukan uji statistik dengan melakukan komparasi dari hasil akurasi yang didapat dari hasil klasifikasi yang menerapkan Teknik ensemble maupun tidak untuk mendapat signifikan atau tidaknya perbedaan akurasi.

1.8 Sistematika Penulisan

BAB I PENDAHULUAN

Bagian ini terdiri atas latar belakang masalah, rumusan masalah, Batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, hipotesa penelitian, metode penelitian dan sistematika penelitian.

BAB II LANDASAN TEORI

Bagian ini menguraikan tinjauan pustaka dan dasar teori yang berkaitan dengan penelitian ini.

BAB III METODE PENELITIAN

Bab ini menjelaskan alur dari penelitian, mulai dari pengumpulan data hingga strategi penanganan masalah yang telah dirumuskan.

BAB IV HASIL DAN PEMBAHASAN

Membahas tentang proses penelitian dan juga menampilkan hasil dari penelitian tersebut.

BAB V PENUTUP

Bab ini menampilkan kesimpulan dari hasil penelitian dan juga saran untuk penelitian selanjutnya.

DAFTAR PUSTAKA

Berisi informasi mengenai sumber-sumber dan referensi yang digunakan penulis sebagai pendukung penulisan skripsi ini.

