

BAB I PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital dan internet saat ini kian meningkat, salah satu efek dari peningkatan tersebut adalah dengan padatnya arus data sehingga berdampak pada kebutuhan akan pengolahan data-data tersebut. Salah satu cara dalam mengolah data yakni dengan teknik data mining *Data mining* merupakan proses dalam membuat sebuah model *machine learning* atau kecerdasan buatan yang dapat mengekstraksi dan mendapatkan informasi dari sekumpulan data sehingga dapat digunakan untuk penentuan keputusan di masa depan [1].

Klasifikasi merupakan salah satu tujuan dalam *data mining* yang bisa diselesaikan menggunakan metode *supervised learning*. Metode ini sangat bergantung pada data latih sehingga kualitasnya akan sangat dipengaruhi oleh data yang akan digunakan. Permasalahan yang biasa terjadi pada dataset yaitu adanya *imbalance class*. *Imbalance class* sendiri adalah ketika jumlah kelas pada sebuah dataset tidak seimbang sehingga membuat model yang dibangun dapat menyebabkan *misleading* pada hasil klasifikasi [2].

Dalam menangani masalah *imbalance class* pada dataset terdapat dua pendekatan yaitu pada tingkat algoritma dan tingkat data [1]. Penelitian ini menggunakan pendekatan pada tingkat data yaitu dengan melibatkan teknik *resampling*. Terdapat dua cara dalam melakukan *resampling*, pertama dengan cara menduplikat data kelas minoritas secara *random* sebanyak kelas mayoritas atau ROS (*random oversampling*) dan cara kedua mengurangi data kelas mayoritas sebanyak kelas minoritas RUS (*random undersampling*) [2].

Penelitian ini menggunakan metode SMOTE dalam menangani *imbalance class* pada dataset *stroke prediction* dan dilakukan perbandingan menggunakan tiga model algoritma klasifikasi yaitu NBC (*Naïve Bayes Classifier*), KNN (*K-Nearest Neighbour*) dan *Decision Tree C4.5*. Tujuan perbandingan ini dilakukan agar dapat mengetahui pengaruh SMOTE terhadap algoritma klasifikasi dan perbedaan performa dari model algoritma klasifikasi ketika menggunakan dataset

yang telah di *oversampling* menggunakan SMOTE dan tanpa *oversampling*.

1.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengetahui perbedaan kinerja pada tiap model algoritma klasifikasi saat menggunakan dataset telah diimplementasikan SMOTE *oversampling* dan tanpa dilakukan *oversampling*.

1.3 Rumusan Masalah

Berdasarkan penyampaian dari latar belakang di atas dapat disimpulkan rumusan masalah pada penelitian ini yaitu :

1. Bagaimana perbedaan kinerja tiap model algoritma klasifikasi ketika menggunakan dataset yang telah diimplementasikan SMOTE *oversampling* dan tanpa dilakukan *oversampling*?
2. Bagaimana pengaruh SMOTE terhadap model algoritma klasifikasi?

1.4 Batasan Masalah

Batasan masalah pada penelitian ini mencakup beberapa hal yaitu :

1. Menggunakan teknik *oversampling* SMOTE dalam menangani *imbalance class*.
2. Menggunakan 3 algoritma klasifikasi yaitu algoritma NBC (*Naïve Bayes Classifier*), KNN (*K-Nearest Neighbour*) dan *Decision Tree C4.5*.
3. Membuat *model machine learning* menggunakan bahasa pemrograman *python* dan *software google collaboratory*.
4. Hanya melakukan perbandingan akurasi antara model yang menggunakan dataset tanpa *oversampling* dan model yang

menggunakan SMOTE *oversampling* pada dataset.

5. Menggunakan dataset publik *stroke prediction* yang diunduh dari situs *kaggle repository*.

1.5 Manfaat Penelitian

Dalam melakukan penelitian ini terdapat beberapa manfaat yang diketahui oleh peneliti seperti menangani *imbalance class* pada dataset, mendapat informasi terkait perbedaan kinerja model algoritma klasifikasi saat menggunakan dataset dengan implementasi SMOTE *oversampling* dan tanpa *oversampling* juga dapat mengetahui pengaruh SMOTE pada tiap algoritma klasifikasi.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN, berisi Latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, sistematika penulisan.

BAB II TINJAUAN PUSTAKA, berisi studi literatur dan dasar-dasar teori yang digunakan pada penelitian.

BAB III METODE PENELITIAN, didalamnya terdapat alur penelitian, alat dan bahan yang digunakan dan perhitungan manual algoritma klasifikasi.

BAB IV HASIL DAN PEMBAHASAN, bab ini merupakan tahapan yang penulis lakukan dalam membangun model klasifikasi sesuai dengan alur penelitian. Berisi urutan proses pada *data mining* seperti akuisisi data, *pre-processing*, split data, implementasi SMOTE, implementasi algoritma klasifikasi, evaluasi confusion matrix dan perbandingan hasil.

BAB V PENUTUP, berisi kesimpulan dan saran yang dapat peneliti rangkum selama proses penelitian.