

1. Pendahuluan

1.1 Latar belakang

Imbalance class merupakan kondisi adanya 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* ini biasanya ditemukan pada kasus-kasus anomali, misalnya: *fraud detection*, *churn prediction*, *network intrusion detection*, dan *thyroid disease* [7].

Imbalance class dalam data mining perlu dipelajari karena *minority class* lebih sulit untuk diprediksi daripada *majority class*. Padahal, terkadang kelas minoritas inilah yang lebih memiliki informasi yang berharga. Oleh karena itu, kita memerlukan algoritma yang dapat memprediksi label kelas yang tepat agar memperoleh hasil dengan nilai akurasi yang tinggi.

Salah satu task yang dapat dilakukan dalam data mining yaitu klasifikasi. Metode klasifikasi dapat digunakan untuk memprediksi kasus-kasus *imbalance class*. Dalam proses klasifikasi yang dilakukan adalah mencari nilai dari salah satu atribut dari data yang tersedia yang disebut atribut *class*. Nilai dari atribut *class* yang akan kita cari dalam proses klasifikasi ini dapat diperoleh dengan mengacu pada nilai-nilai atribut *class* dari data yang telah diketahui nilai-nilai atribut *class*-nya, data ini disebut *training set*.

Ada beberapa metode yang digunakan dalam proses klasifikasi, salah satu metode dalam klasifikasi yaitu K-Nearest Neighbor atau yang disingkat K-NN yang merupakan metode pengklasifikasian data yang bekerja relatif dengan cara yang lebih sederhana dibandingkan dengan metode pengklasifikasian data lainnya. Algoritma ini berusaha mengklasifikasikan data baru yang belum diketahui class-nya dengan memilih data sejumlah K yang letaknya terdekat dari data baru tersebut. Class terbanyak dari data terdekat sejumlah K tersebut dipilih sebagai class yang diprediksikan untuk data yang baru [6].

Salah satu varian dari algoritma k-nearest neighbor yaitu Fuzzy K-Nearest Neighbor atau yang disebut Fuzzy K-NN yang merupakan metode klasifikasi yang menggabungkan teknik fuzzy dan K-Nearest Neighbor classifier. Fuzzy K-Nearest Neighbor memiliki dua keunggulan utama daripada algoritma K-NN [1]. Pertama, algoritma ini mampu mempertimbangkan sifat ambigu dari tetangga jika ada. Algoritma ini telah dirancang sedemikian rupa agar tetangga yang ambigu tidak memainkan peranan penting dalam klasifikasi saat ini. Keunggulan kedua yaitu sebuah *instance* akan memiliki derajat nilai keanggotaan pada setiap kelas sehingga akan lebih memberikan kekuatan atau kepercayaan suatu *instance* berada pada suatu kelas.

Dari penelitian yang telah dilakukan sebelumnya, telah terbukti bahwa metode K-NN untuk *imbalance class* baik dalam memprediksi banyaknya kelas minor pada permasalahan *imbalance class*, tetapi melakukan misclassification (pembiasan) kelas

mayor ke kelas minor yang menyebabkan nilai akurasi kelas mayor menurun [7]. Sehingga diharapkan dengan penggunaan metode Fuzzy K-NN dengan beberapa keunggulannya ini mampu mengatasi kekurangan yang dihadapi metode K-NN pada kasus *imbalance class*.

1.2 Perumusan masalah

1. Bagaimana menerapkan Fuzzy K-Nearest Neighbor pada *imbalance class*?
2. Seberapa besar peningkatan akurasi dalam klasifikasi menggunakan teknik Fuzzy pada K-Nearest Neighbor classifier?
3. Bagaimana melakukan pengujian dan analisis dari implementasi teknik Fuzzy K-Nearest Neighbor dalam klasifikasi?

Batasan Masalah :

1. Data training sudah mengalami data preprocessing
2. Data yang digunakan untuk analisis merupakan data yang supervised (memiliki class label)
3. Dataset yang digunakan merupakan data *imbalance class*
4. Evaluasi diperoleh berdasarkan nilai *recall*, *precision*, dan *f-measure*
5. Data yang digunakan yaitu data churn perusahaan telekomunikasi

1.3 Tujuan

1. Mengimplementasikan metode Fuzzy K-NN pada kasus *imbalance class*
2. Mengetahui besarnya peningkatan akurasi dalam klasifikasi menggunakan metode Fuzzy K-NN
3. Melakukan pengujian dan analisis dari implementasi teknik Fuzzy K-NN

1.4 Metodologi penyelesaian masalah

1. Studi literatur
Pencarian referensi, mencari referensi dan sumber-sumber lain yang layak yang berhubungan dengan data mining, *imbalance class*, dan Fuzzy K-NN
2. Pencarian dan pemahaman data
Mencari data. Setelah data didapatkan maka langkah selanjutnya adalah data understanding yaitu memahami data yang diperoleh
3. Analisis dan perancangan perangkat lunak
Menganalisis permasalahan yang akan diselesaikan dan menganalisis tahapan-tahapan yang diinginkan untuk menyelesaikan permasalahan dengan metode prosedural
4. Implementasi sistem
Melakukan implementasi sistem dengan membangun perangkat lunak sesuai dengan perancangan yang telah dilakukan
5. Pengujian sistem dan analisa hasil

Melakukan proses pengujian terhadap keakuratan hasil dan performansi algoritma Fuzzy K-NN, serta melakukan analisis terhadap kelebihan dan keterbatasan algoritma Fuzzy K-NN pada kasus *imbalance class*

6. Pengambilan keputusan dan penyusunan tugas akhir
Pengambilan kesimpulan berdasarkan pengujian dan analisis disertai dengan penyelesaian laporan tugas akhir.