

BAB I

PENDAHULUAN

I.1. Latar Belakang Masalah

Data mining merupakan sebuah proses ekstraksi untuk mendapatkan suatu informasi yang sebelumnya tidak diketahui dari sebuah data (Witten et al., 2011). Mempersiapkan data adalah tahapan *preprocessing* yang sangat penting pada data mining, alasan utamanya adalah karena kualitas dari *input* data sangat mempengaruhi kualitas *ouput* analisis yang dihasilkan (Junaedi et al., 2011). Pembersihan data dilakukan untuk menghilangkan *noise* didalam data (Karthick & Malathi, 2015) sehingga kualitas data menjadi meningkat. Ada beberapa algoritma untuk *preprocessing data* salah satunya adalah *Noisy Data Filtering*. Pada penelitian ini digunakan metode *Iterative Partitioning Filter*, metode ini merupakan salah satu algoritma *Noisy Data Filtering*. Pada penelitian sebelumnya yang dilakukan oleh Khoshgoftaar & Rebour (2007) tentang *Improving Software Quality Prediction by Noise Filtering Techniques*, metode ini memiliki dua buah skema penyaringan data yaitu *majority* dan *consensus*. Filter ini merupakan upaya untuk meningkatkan kualitas input data dengan menghapus potensi *noisy instance*. *Instance* yang salah diklasifikasikan dan diidentifikasi sebagai *noisy* dihapus dari dataset pelatihan. Data yang digunakan dalam penelitian ini adalah *wine*, *iris*, *wisconsin*, *pima* dan *haberman*. Data ini diambil dari

University of California Irvine (UCI) ,Machine Learning Repository. Dataset ini memiliki karakteristik kompleksitas masing-masing. Dengan adanya kompleksitas data ini maka akan mempengaruhi performansi hasil klasifikasi.

Salah satu tujuan preprocessing data adalah untuk meningkatkan hasil akurasi klasifikasi. Klasifikasi digunakan untuk memprediksi kelas dari objek yang kelasnya belum diketahui (Raviya & Gajjar, 2013). Penelitian ini menggunakan metode jaringan syaraf tiruan dengan algoritma Backpropagation, karena kapasitas untuk memodelkan masalah yang sangat kompleks di area Machine Learning, Data Mining dan Pengenalan Pola (Nawi et al., 2013). Klasifikasi menggunakan Jaringan Syaraf Tiruan dinilai baik karena JST cukup handal dalam memprediksi (Maharani, 2009). Tanpa *preprocessing* ini, pelatihan jaringan saraf akan sangat lambat (Anand et al., 2012). Pada tahap klasifikasi, data yang telah di *preprocessing* sebelumnya diolah menggunakan software weka.

Hasil penelitian disajikan dalam bentuk tabel, ukuran kinerja data *preprocessing* yang digunakan untuk perbandingan adalah waktu pelatihan, akurasi, *kappa statistic*, *Mean Absolute Error (MAE)*, *Root mean squared* dan *confusion matrix*. Hasil penelitian juga disajikan dalam bentuk grafik perbandingan hasil akurasi klasifikasi dengan variasi parameter dan skema penyaringan iterative partitioning filter.

I.2. Rumusan Masalah

Berdasarkan latar belakang permasalahan di atas, maka didapat sebuah rumusan masalah sebagai berikut :

1. Bagaimana melakukan *preprocessing* pada algoritma *Noisy Data Filtering* dengan metode *Iterative Partitioning Filter* pada data Wine, Iris, Wisconsin, Pima dan Haberman dari data Machine Learning Repository dataset di University of California Irvine (UCI) ?
2. Bagaimana hasil dari kinerja *Iterative Partitioning Filter* dalam *preprocessing* data Wine, Iris, Wisconsin, Pima dan Haberman dari data Machine Learning Repository dataset setelah diterapkan dengan metode Backpropagation ?

I.3. Batasan Masalah

Adapun batasan-batasan masalah untuk penelitian ini antara lain :

1. Dataset yang digunakan untuk *preprocessing* adalah dataset yang didapat dari Machine Learning Repository dataset di University of California Irvine (UCI).
2. Algoritma yang digunakan untuk *preprocessing* adalah *Noisy Data Filtering* dengan metode *Iterative Partitioning Filter*, software yang digunakan pada *preprocessing* adalah Keel.
3. Untuk klasifikasi pada *postprocessing* menggunakan algoritma *Backpropagation* pada software Weka.

I.4. Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah :

1. Melakukan proses *preprocessing* dengan algoritma *Noise Data Filtering* dengan metode *Iterative Partitioning Filter* pada data Wine, Iris, Wisconsin, Pima dan Haberman yang diambil dari Machine Learning Repository dataset di University of California Irvine (UCI).
2. Mengetahui hasil dari kinerja *iterative partitioning filter* dalam *preprocessing* data Wine, Iris, Wisconsin, Pima dan Haberman dari data Machine Learning Repository dataset setelah diterapkan dengan metode *Backpropagation*.

I.5. Sistematika Penulisan

Sistematika penulisan dalam penelitian ini adalah:

BAB I Pendahuluan

Bab ini berisi tentang latar belakang, rumusan masalah, maksud dan tujuan, batasan masalah dan sistematika penulisan dalam pembuatan laporan penelitian.

BAB II Tinjauan Pustaka

Bab ini berisi hasil penelitian terdahulu yang berkaitan dengan penelitian yang dilakukan saat ini.

BAB III Landasan Teori

Bab ini berisi landasan teori yang digunakan sebagai acuan dalam pembahasan masalah yang berkaitan dengan *preprocessing* data, klasifikasi data dan software yang digunakan.

BAB IV Metodologi Penelitian

Bab ini berisi penjelasan bahan atau data yang digunakan dalam penelitian serta langkah-langkah dalam melakukan penelitian.

BAB V Pembahasan

Bab ini berisi hasil penelitian dan pembahasan. Pembahasan berisi tentang analisa data sebelum data diolah, ketika diolah dan hasil setelah pengolahan selesai dilakukan.

BAB VI Kesimpulan dan Saran

Bab ini berisi kesimpulan akhir dari pembahasan penelitian secara keseluruhan dan saran-saran untuk penelitian yang akan dilakukan selanjutnya.

Daftar Pustaka

Bagian ini berisi tentang daftar pustaka yang digunakan pada pembahasan tugas akhir ini.