

BAB I

PENDAHULUAN

1.1 Latar Belakang

Seiring kemajuan jaman, data yang perlu diolah menjadi semakin banyak. Tentu data yang banyak ini memang diperlukan untuk mengambil keputusan yang akurat. Dalam satu buah data, bisa saja atribut yang terdapat di dalamnya cukup banyak, tetapi tidak semua atribut itu terpakai dalam proses pengambilan keputusan tertentu.

Menghadapi masalah ini, maka beberapa atribut perlu dikurangi atau dihilangkan memakai proses *feature selection*. Feature selection sendiri adalah proses reduksi atribut guna mendapatkan atribut yang sesuai atau relevan dengan kriteria tertentu [8]. Dengan feature selection, diharapkan atribut-atribut yang terpilih bisa lebih relevan dan memudahkan dalam proses pengolahan data selanjutnya. Untuk klasifikasi misalnya. Dengan dataset yang memiliki lebih sedikit atribut dan relevan, maka hasil proses klasifikasi bisa lebih akurat.

Salah satu algoritma yg dipakai adalah Relief yang dikembangkan tahun 1992. Algoritma ini adalah akar dari algoritma – algoritma selanjutnya yang banyak dijumpai dalam proses feature selection. Algoritma ini hanya bisa menangani dataset dengan 2 kelas saja [4]. Setelah Relief ditemukan, makin banyak algoritma lain muncul sebagai pengembangannya seperti A-Relief dan LAS Relief.

Yang akan dipakai dalam tugas akhir ini adalah C-LAS Relief. Algoritma ini masih cukup baru namun lebih baik dari pendahulunya karena bisa menangani dataset multiclass. Pada dasarnya proses kerja C-LAS Relief tidak terlalu berbeda dengan algoritma Relief. Ide dasarnya adalah menghitung nilai perbedaan jarak antar instance. Perbedaan jarak ini nantinya akan digunakan untuk melakukan penghitungan bobot atribut. C-LAS Relief memakai Chebyshev distance untuk menghitung *nearest hit* (instance terdekat yang satu kelas) dan *nearest miss* (instance terdekat yang berbeda kelas) [3].

Dalam tugas akhir ini, penulis mencoba mengaplikasikan algoritma tersebut untuk melakukan proses feature selection pada dataset yang telah dipersiapkan.

1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, maka masalah yang akan dibahas dalam tugas akhir ini adalah sebagai berikut :

1. Bagaimana melakukan implementasi dan melihat kehandalan algoritma C-LAS Relief dalam proses feature selection?
2. Bagaimana pengaruh pemilihan jumlah sampel dan threshold terhadap kinerja algoritma C-LAS Relief dalam feature selection?

3. Bagaimana mengevaluasi keberhasilan proses feature selection berdasarkan nilai Symetrical Uncertainty, precision, recall dan F-measure?

Dalam tugas akhir juga ada beberapa batasan yaitu :

1. Dataset yang dipakai merupakan dataset multiclass
2. Dataset diambil dari UCI Machine Learning Repository bertipe data csv (numerik, nominal, campuran)
3. Proses klasifikasi dilakukan dengan memakai Weka

1.3 Tujuan

Berdasarkan rumusan masalah di atas, maka tujuan yang ingin dicapai dalam penelitian ini adalah :

1. Mengimplementasikan algoritma C-LAS Relief dalam proses feature selection
2. Melakukan analisis terhadap Pengaruh pemilihan jumlah sampel dan threshold terhadap kinerja algoritma C-LAS Relief
3. Melakukan analisis kinerja feature selection dan classifier yang meliputi Symetrical Uncertainty, precision, recall dan F-measure pada dataset yang belum dan sudah dilakukan proses feature selection

1.4 Hipotesa

Hipotesa yang bisa didapat antara lain :

1. Dataset baru akan memiliki atribut yang relevan berdasarkan perhitungan *Symetrical Uncertainty*. Atribut tersebut akan bernilai mendekati 1

1.5 Metodologi

Metodologi dalam menyelesaikan tugas akhir ini adalah sebagai berikut

1. Studi Literatur

Tahapan dalam mempelajari konsep dan teori yang berkaitan dalam tugas akhir ini seperti konsep tentang feature selection, prinsip kerja algoritma C-LAS Relief, perhitungan akurasi data setelah klasifikasi

2. Mengumpulkan Data

Data yang dikumpulkan, diambil dari UCI Machine Learning Repository

3. Perancangan Sistem

Pada tahap ini dilakukan perancangan terhadap implementasi Algoritma C-LAS Relief untuk melakukan proses feature selection

4. Implementasi dan Pengujian

Pengimplementasian Algoritma C-LAS Relief pada proses feature selection serta melakukan pengujian terhadap dataset yang tersedia

5. Analisis Hasil

Melakukan analisa terhadap hasil pengujian dataset yang sudah dilakukan proses feature selection. Analisis terbagi menjadi 2 yaitu secara langsung dan tidak langsung.

Analisis secara langsung memakai perhitungan *Symetrical Uncertainty*, sementara yang tidak langsung memakai perhitungan akurasi, recall dan F-Measure.

6. Pembuatan Buku TA

Penyusunan laporan dan dokumentasi dari hasil penelitian yang telah dilakukan, serta menyimpulkan hasil analisis dengan mengikuti ketentuan yang telah ditetapkan oleh institusi