

ABSTRAKSI

Algoritma *machine learning* secara garis besar melakukan ekstraksi *knowledge* dari suatu database. Sebagian besar algoritma tersebut biasanya hanya bisa diaplikasikan pada data numerik ataupun nominal. Lain halnya untuk atribut *continuous*, dibutuhkan proses diskretisasi dahulu untuk merubah nilai atribut *continuous* menjadi interval.

Diskretisasi adalah proses mentransformasi nilai atribut *continuous* menjadi sejumlah interval terbatas yang berhubungan dengan nilai diskret, yaitu nilai numerik. Pendekatan yang biasa dilakukan dalam proses *learning* menggunakan *mixed-mode data* (campuran antara data numerik dan *continuous*) adalah melakukan diskretisasi terlebih dahulu sebelum proses *learning* (*pre-processing*).

CAIM (Class-Attribute Interdependence Maximization) adalah salah satu algoritma diskretisasi yang dirancang untuk *supervised learning*. Algoritma ini memaksimalkan saling ketergantungan (*interdependency*) antara kelas dan atribut, dan pada saat bersamaan menghasilkan jumlah interval diskret seminimal mungkin. Algoritma ini bekerja tanpa user harus mendefinisikan dahulu jumlah intervalnya.

Pada tugas akhir ini penulis mengimplementasikan metode diskretisasi CAIM untuk *supervised learning* pada sejumlah dataset. Lalu hasil diskretisasinya diujikan pada algoritma C5.0 untuk menghasilkan rule klasifikasi. Tingkat akurasi dan jumlah rule yang dihasilkan CAIM lalu dibandingkan dengan akurasi dan jumlah rule yang dihasilkan enam metode diskretisasi lain. Hasil perbandingan menunjukkan secara umum CAIM mencapai hasil terbaik – akurasi tinggi dan jumlah rule kecil – dibanding enam metode lain yang diujikan.

Kata kunci : CAIM, Class-Attribute Interdependence Maximization, diskretisasi, supervised learning.