

Abstrak

Data mining adalah proses pencarian pola dari kumpulan data. Dalam *data mining* ada 3 *task* utama yaitu klasifikasi, asosiasi dan klasterisasi. Klasifikasi merupakan *task data mining* untuk menemukan pola dari kumpulan data. Tujuan dari pencarian pola tersebut adalah untuk menjawab nilai dari suatu data yang belum diketahui nilainya. Tahapan klasifikasi pada *data mining* dimulai dengan pembersihan data (*preprocessing*), pembentukan model dan evaluasi (*testing*) model.

Tahap *preprocessing* merupakan salah satu tahap yang sangat penting dalam klasifikasi *data mining*, karena tahap ini merupakan tahap penyediaan data yang digunakan untuk pembentukan pola nantinya. Salah satu bentuk tahap *preprocessing* dalam *data mining* adalah pemilihan variabel atau yang sering juga disebut dengan *variable selection*, *variable subset selection*, *feature selection*, *feature reduction*, atau *attribute selection*.

Tujuan dari pemilihan variabel adalah untuk mendapatkan data yang optimal, sehingga hasil akhir model yang didapat lebih optimal, selain itu juga untuk mendapatkan model yang lebih ringkas dan mempercepat proses *learning*. Dalam tugas akhir ini diimplementasikan pemilihan variabel menggunakan *Rough Set*. *Rough Set* digunakan untuk mendapatkan perkiraan rule yang singkat dari suatu data, dalam hal ini pengurangan variabel atau kolom.

Kata kunci : *Data mining*, *Rough Set*, Klasifikasi, *Preprocessing*,
Variable selection, *Feature selection*