Abstrak

Data mining adalah proses mengekstraksi atau menggali informasi dari sekumpulan data. Dalam data mining ada tiga *task* utama yaitu klasifikasi, asosiasi dan klasterisasi. *Klasifikasi* dalam data mining merupakan proses mencari pola dari sekumpulan data. Tujuan dari pencarian pola tersebut adalah untuk menjawab nilai dari suatu data yang belum diketahui nilainya. Namun kerap kali dibutuhkan resource yang cukup besar untuk melakukan klasifikasi. Hal ini dapat disebabkan karena data yang akan dilakukan klasifikasi memiliki dimensionalitas yang tinggi, banyak variabel yang irrelevant, ataupun variabel yang redundant.

Untuk itu diperlukan tahap *pre-prosessing* sebelum dilakukan proses klasifikasi. Ada banyak jenis tahap preprosesing, salah satunya adalah *Variabel Selection*. *Variabel Selection* adalah suatu proses pengidentifikasian dan penghapusan variabel yang bersifat irrelevant atau mengalami redundant.

Tugas akhir ini secara khusus membahas tentang seleksi variabel dengan menggunakan metode *Correlation-based*. Dalam melakukan seleksi variabel, correlation-based melakukan penghitungan dan pembandingan tingkat korelasi antara masing-masing variabel dengan variabel Classnya dan antara variabel dengan variabel lainnya. Kemudian dilakukan pengujian dengan membandingan nilai akurasi, precision dan recall antara dataset asli dengan dataset hasil seleksi variabel dengan metode correlation-based. Dari hasil pengujian diketahui bahwa seleksi variabel dengan metode correlation-based mampu mengurangi dimensionalitas data dengan nilai akurasi, precison dan recall yang mampu menyamai dataset aslinya.

Kata kunci : *Klasifikasi , Pre-prosessing , Variabel Selection*, *Correlation-based*