

ABSTRAK

Ada beberapa tantangan dalam data mining salah satunya fenomena *curse of dimensionality*. Fenomena *curse of dimensionality* merupakan fenomena yang terjadi pada *data mining* dengan adanya peningkatan jumlah dimensi data sehingga dapat menyebabkan turunnya *performance*. Oleh karena itu harus dilakukan *preprocessing data* yaitu dengan *data reduction*. Penelitian ini bertujuan untuk membandingkan beberapa teknik *data reduction* yang dapat memberikan *performance* yang lebih baik dilihat dari tingkat akurasi dan berapa lama waktu proses yang dibutuhkan dimulai dari *preprocessing* sampai dengan hasil akhir diperoleh. Kemudian menguji apakah ada perbedaan akurasi yang signifikan antara beberapa metode *data reduction* menggunakan pengujian *t-test*. Pada penelitian ini metode klasifikasi *decision tree* digunakan untuk melakukan penilaian *performance data reduction* pada *glass identification dataset* dari *UCI machine learning repository*. Pada pengujian *performance* untuk tingkat akurasi, tahapan *data reduction* menggunakan teknik *forward selection* dan klasifikasi *decision tree* (FS+DT) menghasilkan tingkat akurasi yang paling tinggi sebesar 72,88% +/- 7,89% dengan *performance* untuk rata-rata lama waktu proses sebesar 210,15 ms. Pada pengujian *performance* untuk rata-rata lama waktu proses, tahapan *data reduction* menggunakan teknik *singular value decomposition* dan klasifikasi *decision tree* (SVD+DT) menghasilkan rata-rata lama waktu proses yang paling cepat sebesar 9,70 ms dan *performance* untuk tingkat akurasi sebesar 63,07% +/- 8,04%. Pada pengujian *t-test* secara statistik tidak ada perbedaan yang signifikan antara tahapan *data reduction* menggunakan teknik *forward selection* dan klasifikasi *decision tree* (FS+DT) dengan tahapan *data reduction* menggunakan teknik *backward elimination* dan klasifikasi *decision tree* (BE+DT) dan tahapan *data reduction* menggunakan teknik *principle component analysis* dan klasifikasi *decision tree* (PCA+DT).

Kata kunci: *data reduction, decision tree, akurasi, waktu proses, t-test*