

ABSTRAK

Pertumbuhan yang pesat dari akumulasi data telah menciptakan kondisi yang sering disebut sebagai *rich in data but poor in information*. *Data preprocessing* merupakan salah satu cara yang dapat dilakukan untuk mengatasi masalah tersebut. Salah satu bagian dalam *data preprocessing* adalah *data reduction*, yang merupakan suatu teknik yang dapat diterapkan untuk memperoleh hasil reduksi dari sejumlah dataset dengan jumlah volume data yang lebih kecil. Metode yang digunakan untuk menerapkan reduksi data tersebut adalah Metode *Density-Based Multiscale Data Condensation* yang bertujuan untuk untuk mereduksi data pada skala yang berbeda. Dengan skala yang berbeda ini, diharapkan didapatkan informasi yang dapat mewakili keseluruhan data yang ada. Penggunaan skala yang berbeda ini diterapkan dengan menentukan radius-radius pereduksian data yang bergantung pada jarak kedekatan antar data. Hasil performansi dari data hasil *preprocessing*, yaitu nilai *precision* dan *recall* mengalami peningkatan dan waktu pembentukan model menjadi lebih cepat.

Kata kunci : *data preprocessing, data reduction, Density-Based Multiscale Data Condensation, skala, precision, recall, waktu.*