

## Abstrak

Deteksi outlier merupakan salah satu fungsionalitas dalam data mining yang bertujuan untuk mencari data yang berbeda dengan mayoritas data lainnya. Walaupun memiliki perilaku yang berbeda dengan mayoritas data lainnya, outlier sering mengandung informasi yang sangat berguna. Ada banyak metode untuk mendeteksi outlier, namun kebanyakan didesain untuk data numerik dan tidak cocok diterapkan dalam data kategorikal. Selain itu, banyak algoritma yang membutuhkan waktu proses yang lama seiring bertambahnya jumlah data. CBLOF (*Cluster Based Local Outlier Factor*) merupakan suatu metode untuk mendeteksi outlier pada data kategorikal berbasis kluster. Nilai CBLOF untuk tiap data akan dihitung, berdasarkan kondisi data tersebut termasuk dalam *large cluster* atau *small cluster*, untuk menentukan data tersebut outlier atau tidak. Pengujian dilakukan dengan beberapa skenario untuk mengetahui akurasi berdasarkan *detection rate*, *false positive rate* serta *false negative rate*, pengaruh persentase *rare class* terhadap akurasi dan pengaruh jumlah data terhadap waktu proses. CBLOF dapat mendeteksi outlier dengan tingkat akurasi relatif baik dilihat berdasarkan *detection rate*, *false positive rate* dan *false negative rate*. Selain itu, prosesnya juga cepat karena CBLOF hanya perlu membaca dataset satu kali hingga didapatkan data yang dianggap sebagai outlier atau tidak.

Kata Kunci : outlier, kluster, kategorikal, CBLOF