

## ABSTRAK

Pertumbuhan data yang semakin pesat menyebabkan penumpukan data yang sangat besar. Untuk itu diperlukan penambangan data untuk mengubah data yang sangat banyak dan tidak informatif menjadi data yang memberi suatu informasi. *Outlier* merupakan salah satu bidang penelitian di dalam penambangan data. *Outliers* adalah data yang menyimpang terlalu jauh dari data yang lainnya dalam suatu rangkaian data. Salah satu algoritma deteksi *outlier* adalah algoritma *Enhanced Class Outlier Distance Based* (ECODB) yang merupakan peningkatan algoritma *Class Outlier Distance Based* (CODB). Algoritma ECODB mengurangi parameter dalam CODB dengan melakukan normalisasi. Algoritma ECODB dapat mengidentifikasi *outlier* pada data yang memiliki *class label*. Pada penelitian ini algoritma ECODB dipergunakan untuk mengidentifikasi *outlier* pada data hasil UN, indeks integritas dan akreditasi Sekolah Menengah Atas (SMA) pada Provinsi Daerah Istimewa Yogyakarta tahun 2015. Berdasarkan dari hasil pengujian dapat disimpulkan bahwa algoritma ECODB dapat mengidentifikasi *outlier* pada hasil UN, indeks integritas dan akreditasi Sekolah Menengah Atas (SMA) pada Provinsi Daerah Istimewa Yogyakarta tahun 2015. Nilai jumlah tetangga terdekat (K) dan jumlah *outlier* yang diharapkan (N) dapat berpengaruh terhadap hasil identifikasi *outlier*. Memvariasi nilai K dapat mempengaruhi nilai *Probability Class Label* (PCL) setiap *instance*. Memvariasi nilai N dapat mempengaruhi nilai *Class Outlier Factor* (COF) setiap *instance*.

Kata Kunci : Penambangan Data, Identifikasi *Outlier*, Algoritma *Enhanced Class Outlier Distance Based*

## ABSTRACT

Rapid data growth leads to huge data stacks. Therefore, data mining is needed to change the large and uninformative data into more informative. Outlier detection is one field of research in data mining. Outliers are data that deviate too far from other data in a dataset. One of the outlier detection algorithms is the Enhanced Class Outlier distance based (ECODB) algorithm. ECODB algorithm is enhancement from Class Outlier Distance Based (CODB) algorithm. ECODB algorithm reduces CODB parameter using normalization technique. ECODB algorithm can identify outliers in data that have class label. In this study, the ECODB algorithm is used to identify outliers on national exam results, integrity index and accreditation of senior high schools in the Yogyakarta province year 2015. From the experimental results, it can be concluded that ECODB algorithm can identify outliers on national exam results, integrity index and accreditation of senior high school in the Yogyakarta province year 2015. The value of the nearest neighbor (K) and the expected number of outliers (N) effect the result of outlier identification. Varying the K value can affect the Probability Class Label (PCL) of each instance. Varying the N value can affect the Class Outlier Factor (COF) of each instance.

Keywords: Data Mining, Outlier Identification, Enhanced Class Outlier Distance Based Algorithm