

Abstract

Outlier detection is one of the importance task in data mining to find data that has different characteristic with another data. Many existing methods to find outlier not looking closely to the class label. This thing causes the data that should be detected as outlier in class label is considered as the normal one. In class outlier detection with ECODB algorithm, data searching that deviates from its class can be found. From the existing outlier detection methods, all of them haven't been able to handle dataset that has more than two class labels dan numeric dataset or mixed dataset, while ECODB algorithm can handle it. This ECODB algorithm is based on distance approach, by calculating COF that represents the degree of the data that categorized as class outlier or not. The testing is done with some scenarios to know the system's accuracy, error rate, and its FPR and also how the addition of dataset number and top n class outlier number influence the process time. ECODB can detect class outlier with better accuracy, and better result in numeric dataset. The ECODB's process time keeps increasing exponentially if we keep adding the instances number, so that it perfoms worse to handle high dimation dataset.

Keywords: ECODB, class outlier, distance based approach, outlier detection.