

Abstrak

Deteksi *outlier* merupakan salah satu task penting dalam *data mining*, bertujuan untuk mencari data yang memiliki karakteristik berbeda dengan kebanyakan data lainnya. Dari banyak metode yang sudah ada, pencarian *outlier* tidak memerhatikan secara khusus pada *class label* dalam kasusnya. Hal tersebut mengakibatkan data yang seharusnya terdeteksi sebagai *outlier* dalam suatu *label class* dianggap sebagai data yang normal. Pada deteksi *class outlier* dengan algoritma ECODB, pencarian data yang menyimpang dari *class*-nya dapat ditemukan. Dari beberapa metode pendeteksi *class outlier* yang ada, seluruhnya belum mampu menangani dataset yang memiliki lebih dari dua *label class* dan dataset yang bertipe numerik atau *mixed* dataset, sedangkan algoritma ECODB sudah dapat menanganinya. Algoritma ECODB ini berbasiskan pendekatan jarak, dengan memperhitungkan nilai COF yang merepresentasikan derajat suatu data termasuk dalam *class outlier* atau bukan. Pengujian dilakukan dengan beberapa skenario untuk mengetahui nilai akurasi, *error rate*, dan FPR, pengaruh penambahan jumlah data dan penambahan jumlah *top n class outlier* terhadap waktu proses. ECODB dapat mendeteksi *class outlier* dengan tingkat akurasi yang baik, dengan hasil yang lebih baik pada dataset bertipe numerik. Waktu proses yang dibutuhkan oleh ECODB terus meningkat secara eksponensial seiring dengan penambahan jumlah *instances*, sehingga kurang baik dalam menangani dataset berdimensi besar.

Kata kunci: ECODB, *class outlier*, *distance based approach*, deteksi *outlier*.