

## INTISARI

# METODE DETEKSI OUTLIER DATA KATEGORIK MENGGUNAKAN FUNGSI PEMBOBOTAN

Oleh

**Nur Rokhman**  
**09/294337/SPA/0264**

Outlier adalah data yang menyimpang dari data pada umumnya. Dewasa ini persoalan outlier merupakan hal yang sangat penting. Keberadaan outlier dapat menjadi petunjuk adanya gangguan suatu sistem, penyalahgunaan sistem, dan petunjuk tentang fakta-fakta baru.

Sejumlah metode deteksi outlier telah dikembangkan. Akan tetapi sebagian besar metode deteksi outlier bekerja pada data numerik, hanya sedikit yang bekerja pada data kategorik. Sementara cukup banyak data dalam kehidupan nyata yang merupakan data kategorik.

Pada berbagai pengolahan data, seperti juga deteksi outlier, fungsi pembobotan (*weighting function*) digunakan untuk meningkatkan hasil. Namun demikian, kebanyakan metode deteksi outlier data kategorik belum menggunakan fungsi pembobotan untuk meningkatkan hasil deteksinya.

Pada disertasi ini dikembangkan metode deteksi outlier data kategorik dengan menggunakan fungsi pembobotan yang terdiri atas pembobotan range, variansi, standar deviasi, dan fungsi kuadrat. Selain itu dikenalkan matrik frekuensi nilai atribut untuk mereduksi kompleksitas algoritma deteksi outlier. Penerapan pembobotan ke metode AVF (*Attribute Value Frequency*), WDOD (*Weighted Density Outlier Detection*), dan AEVF (*Automatic Entropy Value Frequency*) menghasilkan metode WAVF (*Weighted Attribute Value Frequency*), WADOD (*Weighted Attribute Density Outlier Detection*), dan WMEVF (*Weighted Matrix Entropy Value Frequency*). Berdasarkan semua pembobotan yang dicobakan, pembobotan fungsi kuadrat memberikan hasil yang paling baik. Penggunaan matrik frekuensi nilai atribut pada metode WMEVF dapat menurunkan kompleksitas waktu dari  $O(m*n^2)$  menjadi  $O(m*n)$ .

Kata kunci : *Outlier detection, Categorical Data, Weighting Function*

## ABSTRACT

# OUTLIER DETECTION METHODS ON CATEGORICAL DATA BY USING WEIGHTING FUNCTION

By

**Nur Rokhman**

**09/294337/SPA/0264**

An Outlier is a data that deviates from the data in general. Today, the problem of outliers is very important. The outliers may indicate a system interference, misuse of the system, or finding new facts.

A number of outlier detection methods have been developed. But most of them works on numerical data, only a few works on categorical data. On the other hand, a lot of data in real life is categorical data.

In many data processing, as well as outlier detection, the weighting functions are used to increase the result. However, most methods of outlier detection method for categorical data are not yet using the weighting function.

In this dissertation three outlier detection method on categorical data using weighting functions have been developed. The weighting functions are the range, variance, standard deviation, and quadratic functions. Beside of that, a matrix of attribute values frequency has been introduced in order to reduce the complexity of outlier detection algorithms. The implementation of the weighting function to the AVF (Attribute Value Frequency), WDOD (Weighted Density Outlier Detection), and AEVF (Automatic Entropy Value Frequency) methods produced WAVF (Weighted Attribute Value Frequency), WADOD (Weighted Attribute Density Outliers Detection), and WMEVF (Weighted Entropy Value Matrix Frequency) methods. Based on the experiments, quadratic function was the best weighting function. The use of attribute values frequency matrix on WMEVF method reduced the time complexity of from  $O(m*n^2)$  to  $O(m*n)$ .

*Key words: Outlier detection, Categorical Data, Weighting Function*