

## ABSTRAK

Data mining (penambangan data) adalah suatu proses untuk menemukan suatu pengetahuan atau informasi yang berguna dari data berskala besar. Sering juga disebut sebagai bagian proses KDD (Knowledge Discovery in Databases) (Santosa, 2007).

Deteksi *outlier* merupakan salah satu bidang penelitian yang penting dalam mendeteksi perilaku yang tidak normal seperti deteksi mahasiswa yang nilai tes masuk universitasnya bagus tetapi pada saat kuliah mahasiswa tersebut mendapat nilai rendah.

Penelitian ini menggunakan algoritma *Local Correlation Integral* dengan teknik *density-based*. Data diambil dari data akademik Mahasiswa Program Studi Teknik Informatika Universitas Sanata Dharma angkatan 2007 dan 2008 yang meliputi nilai tes masuk dan nilai IPS semester 1-4.

Algoritma *Local Correlation Integral (LOCI)* merupakan salah satu algoritma yang memiliki kemampuan untuk mendeteksi *outlier* dalam sekumpulan data. Mendeteksi *outlier* dilakukan untuk menemukan data yang tidak konsisten dengan data lainnya. Data dianggap tidak konsisten (*outlier*) apabila data tersebut tidak memiliki tingkat kemiripan dengan data lainnya (Han & Kamber, 2006). Algoritma *Local Correlation Integral (LOCI)* dapat diimplementasikan pada sekumpulan data numerik untuk mendeteksi adanya *outlier* dengan pendekatan *density-based*.

Hasil dari penelitian ini adalah sebuah perangkat lunak yang dapat digunakan untuk mendeteksi *outlier*. Pengujian terhadap sistem ini meliputi tiga metode pengujian, yaitu pengujian *Blackbox*, pengujian efek perubahan nilai atribut terhadap hasil deteksi *outlier*, dan pengujian *reviewer* dan validitas pengguna.

Hasil pengujian dari ketiga jenis data tersebut dapat disimpulkan bahwa sistem pendeteksi *outlier* ini dapat menghasilkan output yang sesuai dengan yang diharapkan oleh pengguna.

Kata kunci: Penambangan Data, *Local Correlation Integral*, deteksi *outlier*.

## ABSTRACT

Mining data is a process to discover a useful knowledge or information from grand data. It is often said as part of the KDD (Knowledge Discovery in Databases) process. (Santosa, 2007)

The outlier detection is one of the important researches in detecting the abnormal behavior namely detecting the students who have good marks in the university entrance test but after they join in the lecturing they get low marks.

This research uses the Local Correlation Integral algorithm in density-based technique. The data is taken from the academic data of the students of Sanata Dharma University, in Informatics Engineering Study Program, year 2007 and 2008 consisting the marks of the entrance test and the marks of Social Study semester 1 – 4.

The Local Correlation Integral Algorithm (LOCI) is one the algorithms that has ability to detect outlier in a group of data. Detecting outlier is done to discover the inconsistent data with the other data. Data is considered inconsistent (outlier) if the data do not have the grade of similarity to the other data (Han & Kamber, 2006). The Local Correlation Integral (LOCI) algorithm can be implemented in a group of numeric data to detect the existence of the outlier with density-based approach.

The result of this research is a set of software which can be used to detect outlier. The test of this system comprises 3 testing methods, namely The Blackbox testing, the testing of the change of attribute value toward the result of outlier detecting, and the testing of reviewer and the validity of the user.

The test results of the three types of data we can conclude that this outlier detection system can generate output as expected by the user.

Keywords: data mining, Local Correlation Integral, outlier detection.