

ABSTRAK

Penambangan data merupakan ekstraksi pola terhadap data yang menarik dalam jumlah yang besar. Pola tersebut dikatakan menarik apabila tidak diketahui sebelumnya dan berguna bagi perkembangan ilmu pengetahuan. Data tersebut dapat diolah dengan berbagai teknik penambangan data seperti asosiasi, klasifikasi, *clustering* dan deteksi *outlier*.

Deteksi *outlier* merupakan salah satu bidang penelitian yang penting dalam penambangan data. Penelitian tersebut bermanfaat untuk menemukan *outlier* yang mungkin berguna bagi pengguna. *Outlier* merupakan sebuah data yang berbeda dibandingkan dengan sifat umum yang dimiliki data lain pada suatu kumpulan data.

Pada tugas akhir ini, pendekripsi *outlier* dilakukan menggunakan algoritma *Naïve Nested Loop*. Data yang digunakan adalah data akademik mahasiswa program studi Teknik Informatika Universitas Sanata Dharma, Yogyakarta tahun angkatan 2007 dan 2008. Data tersebut terdiri dari data numerik nilai hasil seleksi masuk mahasiswa yang diterima melalui jalur tes tertulis maupun jalur prestasi dan nilai indeks prestasi dari semester satu sampai empat.

Hasil dari penelitian ini adalah sebuah perangkat lunak yang dapat digunakan sebagai alat bantu untuk mendekripsi *outlier* menggunakan algoritma *Naïve Nested Loop*. Pengujian terhadap sistem ini meliputi tiga metode pengujian, yaitu pengujian *blackbox*, pengujian efek perubahan nilai atribut penambangan data, dan pengujian *review* dan validitas oleh pengguna.

Berdasarkan pengujian *blackbox* yang telah dilakukan dapat disimpulkan bahwa sistem pendekripsi *outlier* ini secara fungsional dapat berjalan dengan baik dan menghasilkan keluaran yang sesuai dengan yang diharapkan. Dari hasil pengujian efek perubahan nilai atribut penambangan data disimpulkan bahwa penentuan nilai parameter pada algoritma *Naïve Nested Loop* yaitu nilai M dan d_{min} berpengaruh terhadap jumlah *outlier* yang dihasilkan. Berdasarkan hasil pengujian *review* dan validitas oleh pengguna dapat disimpulkan bahwa sistem dapat menghasilkan data yang dinyatakan sebagai *outlier*.

Kata kunci : penambangan data, deteksi *outlier*, *naïve nested loop*

ABSTRACT

Data mining is an extraction of interesting pattern in large numbers of data. The pattern is said to be interesting if the data is previously unknown and it is useful for the development of knowledge in data mining. The data can be processed by a variety of data mining techniques such as association, classification, clustering and outlier detection.

Outlier detection is one of important researches in data mining. This research is useful to discover outliers which might be useful for users. Outlier is a data which is different from other data in a dataset.

In this thesis, Naïve Nested Loop algorithm was used to perform outlier detection. The data used in this thesis are academic data of students batch 2007 and 2008 of Informatics Engineering Department of Sanata Dharma University. The dataset consists of student admission data from regular admission track as well as students from outstanding track, and student academic data (Grade Point Average) of those students from first semester until fourth semester.

The results of this research is a software that can be used as a tool to determine outliers using Naïve Nested Loop algorithm. The testing of this system includes three testing methods, namely blackbox testing, the effects of attribute changes of M (the maximum number of objects within the d_{min} neighbourhood of an outlier) and d_{min} (the maximum distance between any pair of objects that define as a neighbour), and validation testing by users.

Based on blackbox testing, it can be concluded that the outlier detection's system could perform properly and produce output as expected. Based on the second testing, it can be concluded that the value of M and d_{min} influence the number of generated outliers. Based on the user's validation, it can be concluded that the results of the system are confirmed as outliers.

Keywords: data mining, outlier detection, naïve nested loop