

ABSTRAK

Knowledge Discovery in Database merupakan suatu proses yang memanfaatkan metode *Data Mining* dalam menemukan pola serta informasi yang bermanfaat dari sekumpulan data. *Domain Name System* (DNS) menyediakan data yang kaya dan menarik, serta dapat diekstraksi untuk mengungkap informasi yang bisa dianalisis bagi berbagai keperluan seperti tindakan keamanan, mengukur tingkat penggunaan trafik, *user profiling* hingga kebijakan lain yang diterapkan dalam suatu jaringan. Penelitian ini bertujuan melakukan klasterisasi penggunaan trafik internet pada jaringan sebagai proses untuk mendapatkan klaster *profile* pengguna dalam melakukan aktivitas *online* selama rentang waktu jam kerja, yang terdiri 3 klaster tinggi, sedang, rendah. Klasterisasi ini perlu dilakukan guna mengawasi pemanfaatan *bandwidth* yang telah disediakan serta memantau efisiensi dan produktifitas pegawai dalam lingkungan kerja. Proses klasterisasi dilakukan menggunakan teknik *Data Mining* dengan metode *K-Means Clustering*. *Sample data* yang digunakan terdiri dari 59 *records* yang diperoleh dari ekstraksi *DNS log file* selama 5 hari kerja yang terdiri dari sesi pagi (08:00 s/d 12:00) dan sesi siang (13:00 s/d 17:00). Setelah melalui proses analisis dan pengujian menggunakan metode yang disebutkan, berdasarkan data *sampling* yang digunakan diperoleh hasil 3 kelompok berupa klaster tinggi, sedang dan rendah. Klaster tinggi sebesar 15%, sedang 12% dan rendah 73%. Sedangkan hasil pengujian menggunakan data *real* (sebanyak 213 *records*) diperoleh hasil klaster tinggi 11%, klaster sedang 14% dan klaster rendah 75%. Berdasarkan pengujian data *real* tersebut, diperoleh rata-rata akurasi yang sangat baik yaitu sebesar 3%, sehingga penelitian ini layak untuk direkomendasikan sebagai salah satu metode dalam melakukan klasterisasi penggunaan trafik internet.

Kata kunci: *data mining*, klasterisasi, *k-means*, *DNS log*, *dnsmasq*, trafik internet

ABSTRACT

Knowledge Discovery in Database is a process that utilizes the Data Mining method in finding patterns and useful information from a set of data. The Domain Name System (DNS) provides rich and interesting data, and can be extracted to reveal information that can be analyzed for various purposes such as security audit, measuring traffic usage levels, user profiling to other policies implemented in an internet network. This study aims to cluster the use of internet traffic on the network to obtain a cluster profile of users in carrying out online activities during the working hours, which consists of 3 high, medium, and low clusters. This clustering needs to be done to monitor the utilization of the available bandwidth and monitor the efficiency and productivity of employees in the work environment. The clustering process is carried out using the Data Mining technique with the K-Means Clustering algorithm. The sample data used consisted of 59 records obtained from the extraction of DNS log files for 5 working days consisting of morning sessions (08:00 to 12:00) and afternoon sessions (13:00 to 17:00). After going through the process of analysis and testing using the methods mentioned, based on the sampling data used, the results obtained were 3 groups in the form of high, medium and low clusters. The high cluster is 15%, the medium is 12% and the low is 73%. While the test results using real data (as many as 213 records) obtained the results of a high cluster of 11%, a medium cluster of 14% and a low cluster of 75%. Based on the real data test, the average accuracy is very good, which is 3%, so this research deserves to be recommended as a method for clustering the use of internet traffic.

Keywords: data mining, clustering, k-means, DNS log, dnsmasq, internet traffic