Abstract

The Internet has become a necessity in today’s society; any information is accessible on the internet via web browser. However, these activities could have an impact on users, one of which changes in behavior. This study focuses on the activities of Internet users based on the log data network at an educational institution. The data used in this study resulted from one-week observation from one of the universities in Yogyakarta. Data log network activity is one type of big data, so it is needed to use data mining with K-Means algorithm as a solution to determine the behavior of Internet users. The K-Means algorithm used for clustering based on the number of visitors. Cluster number of visitors divided into three, namely low with 1479 amount of data, medium with 126 amount of data, and high with 33 amount of data. Categorization also performed by the access time and is based on website content that exists in the data. It is to compare the results by the K-Means clustering algorithm. The results of the educational institution show that each of these clusters produces websites that are frequented by the sequence: website search, social media, news, and information. This study also revealed that the cyber-profiling had been done strongly influenced by environmental factors and daily
Log Classification using K-Means Clustering for Identify Internet User Behaviors

activities.

References

4. J. He, A. Wei, Y. Yang, and W. Dong, “Research on Degree of Video Completion of Internet Videos with Clustering Algorithms,” pp. 89–95, 2015.

**Index Terms**

Computer Science  
Algorithms

**Keywords**

Clustering, K-Means, Network, Log, Cyber-profiling