Clustering of internet network usage using the K-Medoid method

Martanto¹, Saeful Anwar², Cep Lukman Rohmat³, Fadhil Muhammad Basysyar⁴, Yudhistira Arie Wijaya⁵

¹Informatics Management [STMIK IKMI Cirebon]  
²Informatics Technology Department, [STMIK IKMI Cirebon]  
³Software Engineering Department, [STMIK IKMI Cirebon]  
⁴Information System [STMIK IKMI Cirebon]  
⁵*martantomusijo@gmail.com

Abstract – The use of the internet network at Wahidin Vocational High School (SMK) Cirebon City is a very significant need in supporting the progress of learning. The ease of using the internet has made the academic community free to use the internet network facilities provided by the school and carry out the activities they want without any control. Causing the academic community to freely access the web that is not related to ongoing education, and certainly this matter will be very disturbing in the teaching and learning process. This matter is in order to identify accurate analysis also using Information Mining as information processing. After that, using the clustering algorithm procedure is K-Medoid. The K-Medoid procedure is able to carry out information grouping on the use of social media access which is often accessed by the academic community throughout the school. Grouping using the K-medoids algorithm can be known that the use of internet network access is on average used to access social media compared to accessing E-learning. It is suitable from the results of research that was tried if the use of WhatsApp reached an average of 126 users, after that accompanied by Instagram, its usage reached an average of 108 users, then accompanied by E-Learning, the average usage reached 90 users, after that it was accompanied by Facebook and YouTube which an average of 71 users. Thus, from the research that has been tried, it is necessary to implement layer 7 protocol firewall security to block social media access so that internet network access at Wahidin Vocational High School (SMK), Cirebon City. Used as a suitable early idea is to facilitate the academic community to maximize E-learning education.

1. Introduction

School is an institution that is intended for teaching and education of students / pupils under the supervision of the teacher[1]. The formal education system in each country has a different system but has the same goal, one of which is to develop the potential of each student so that they are superior, educated according to the system applied in their respective countries[2]. Education in Indonesia is currently implementing smart learning, one of which is through E-learning. E-learning that is applied in Indonesia is still more focused on being used in high schools or vocational high schools. One of the private schools in Indonesia, especially at SMK Wahidin, located in Cirebon, is developing an E-learning system. The online learning system is closely related to the internet network. With the progress of the era that is increasingly rapid, many new social media have sprung up with the many existing social media, it needs school maturity, especially teachers and students to respond to good internet use in the school.

The data that researchers get from Mtyslog, researchers can get information about the use of access that is often used by the academic community at SMK Wahidin, Cirebon City. With the data that the researcher obtained, the researcher can classify what accesses are frequently visited by the academic community of SMK Wahidin, Cirebon City by using the K-Medoid algorithm. The K-Medoid algorithm is an algorithm intended
for grouping or clustering[3]. In order for the data that is obtained to be run in rapid miner, the data must first go through KDD (knowledge discovery in database) so that the data is not noise or missing value.

Based on research conducted by Solikhun et al, in 2019 entitled "K-Medoid Algorithm for Classifying Villages with School Facilities in Indonesia" explained that there are many algorithms contained in classification data mining, one of which is K-Medoids. The K-Medoids method is part of partitioning clustering. The K-Medoids method is quite efficient in small datasets. The initial step for K Medoids is to find the most representative points (medoids) in the dataset by calculating the distance from groups in all possible combinations of medoids so that the distance between points in a cluster is small while the distance between the points between clusters is large. So that the calculation results obtained by testing using a system where low clusters can be input to the government to pay more attention to provinces with low number of school facilities[4].

Another research related to the K-Medoid algorithm was also carried out by Dedi Hartama et al., In 2019 entitled "Application of the K-Medoid Method to Household Grouping in the Treatment of Sorting Waste According to Province" explained that the behavior of sorting waste using a datamining algorithm with K-Medoid. The data were processed into 2 clusters, namely the low level waste sorting cluster (C1) and the high waste sorting level cluster (C2). Where the results of this study concluded from 33 provinces in Indonesia that the cluster of low waste sorting behavior (C1) was obtained by 22 provinces, namely Aceh, North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, Bangka Belitung Islands, Kep. Riau, DKI Jakarta, East Java, Banten, NTB, West Kalimantan, South Kalimantan, Gorontalo, West Sulawesi, Maluku, North Maluku, Papua, and 12 other provinces are included in the high waste sorting rate cluster (C2)[5].

| No | Device Name | Type | Description |
|----|-------------|------|-------------|
| 1  | Router board | RB2011 | Main router |
| 2  | Access Point | TP-Link | As an internet network transmission medium |
| 3  | Access internet | - | School internet network access speed 100Mbps |
| 4  | Smartphone | - | Belongs to the academic community of SMK Wahidin |
| 5  | Laptop | - | Belongs to the academic community of SMK Wahidin |

Based on the table above, the Mikrotik Router board can control the internet access network and the speed of internet access for each user. The access point is available as a means of transmitting internet access from the user's cellphone. Internet access available in schools has a speed of 100 Mbps for all academic communities of SMK Wahidin Cirebon City in carrying out activities related to school. Smartphone is a tool that all academicians of SMK Wahidin Cirebon City use to carry out the learning process of the E-learning system, especially in the school environment. Laptop is a tool that is used by all academic community of SMK Wahidin Cirebon City to carry out the learning process of the E-learning system, especially in the school environment.

Based on the data obtained, the root of the problem is the absence of a grouping of school internet network access which is often used by the academic community of SMP Wahidin, Cirebon, which results in the use of the school internet not only for the benefit of learners due to the absence of network security that is implemented for access restrictions and the absence of tools, which is used to classify internet access that is frequently visited by the academic community of SMK Wahidin, Cirebon City.

This is because there are still many academicians of SMK Wahidin Cirebon City who use internet network access facilities that are not in accordance with school rules because internet network security has not been implemented regarding blocking websites that should not be used in the school environment. In addition, when online learning was carried out, there were still many teachers and students who used the internet not only for online learning but, it was used to access social media which resulted in an increase in internet traffic which should have been intended for online learning implemented in schools, thus impacting access to use online learning media becomes disrupted. With this, schools need to implement internet network security to increase the effectiveness of online learning by using the screen firewall 7 protocol features that are on the Mikrotik router.
2. Methodology
The stages of this research are divided into several stages to be carried out in this research namely:

![Figure 1. Methodology](image)

Data mining is defined as the process of finding patterns in data. This process is automatic or often semi-automatic. The pattern that is found must be meaningful and the pattern provides benefits, usually economic benefits[6]. The process in the data mining stage consists of three steps, namely:

a. Data Preparation in this step, data is selected, cleaned, and preprocessed following the guidelines and knowledge of domain experts who capture and integrate internal and external data into the overall organizational review.

b. Data mining algorithms The use of data mining algorithms is carried out in this step to explore integrated data to facilitate identification of valuable information. The algorithm used is K-Medoids. K-Medoids is one of the partitioning methods, because it uses the most centered objects (medoids) in the cluster to become the cluster center of the mean value of objects in a cluster. The K-Medoids method is more suitable for grouping data than the K-Means method[7]. K-Medoids Clustering, also known as Partitioning Around Medoids (PAM), is a variant of the K-Means method. This is based on the use of medoids instead of observing the mean that each cluster has, with the aim of reducing the sensitivity of the partition with respect to the extreme values that are in the dataset[8]. The K-Medoids algorithm has the advantage of overcoming weaknesses in the K-Means algorithm, sensitive to noise and outliers, which are objects with large values that may deviate from the data distribution. Another advantage is that the results of the clustering process do not depend on the order entered in the dataset.

c. Data analysis phase The output of data mining is evaluated to see if domain knowledge is found in the form of rules that have been extracted from the network.

3. Result and Discussion
In conducting this research for 8 months with data obtained from monitoring the use of network access SMK Wahidin Cirebon City. These are the monitoring data obtained.

![Figure 2. Determining Data](image)

After determining a data, the next step is to determine the dataset so that the data can be processed properly, the following data sets will be used:
The data used as a sample of 165 data. So that there is no more data that contains noise or missing value and the data is declared normal, the data must first go through the KDD process. The KDD process itself is a process in which to look for and identify patterns before this data can be processed further, so that we can use the data as a dataset. a process model for clustering social media access data using the Rapid Miner tools.

**Figure 3. Clustering Social Media Access Data Using the Rapid Miner Tools**

The next step is to execute the K-Medoids clustering process model and the results will be displayed here.

**Cluster Model**

Cluster 0: 66 items  
Cluster 1: 55 items  
Cluster 2: 5 items  
Cluster 3: 39 items  
Total number of items: 165

The picture above illustrates that a data set of 165 records results in the following groupings:

| Cluster | Number of items |
|---------|-----------------|
| 0       | 66              |
| 1       | 55              |
| 2       | 5               |
| 3       | 39              |

Number of Clusters= 4  Number of datasets= 265

**Figure 4. Clustering Social Media Access Data Using the Rapid Miner Tools**
The conclusion from the Figure above is that cluster 0 is the cluster with the most members followed by cluster 1, cluster 3, and cluster 2. To know the performance of the K-Medoids algorithm, it can be seen from the Performance Vector. Where the Davies Baouldin Index approaches the number 0 value when executed, it means that the performance of the K-Medoids algorithm is getting better.

| K  | Dbi K-Medoids |
|----|----------------|
| K2 | -1.042         |
| K3 | -1.194         |
| K4 | -0.974         |
| K5 | -1.312         |
| K6 | -1.165         |
| K7 | -1.050         |
| K8 | -1.622         |
| K9 | -1.156         |

Experiments that have been carried out get the results of k which has the best performance, namely k which is close to the number 0. K which is close to the number 0 is k 4, which means that k 4 is the best out of 9 experiments. This means that the focus of the cluster focuses on cluster 4. As for the centroid value, the following values are obtained:

| Attribute     | cluster_0 | cluster_1 | cluster_2 | cluster_3 |
|---------------|-----------|-----------|-----------|-----------|
| Instagram     | 124       | 89        | 132       | 84        |
| Facebook      | 43        | 83        | 46        | 72        |
| E-Learning    | 96        | 83        | 158       | 93        |
| YouTube       | 87        | 53        | 52        | 50        |
| WhatsApp      | 111       | 173       | 46        | 89        |

Concluded that the grouping of internet network access usage, the highest frequency of use is WhatsApp, followed by Instagram, and E-Learning, Facebook then the last is YouTube. Grouping using the K-Medoids algorithm and from the results of 9 times the experiment, k-4 is the best k with Dbi -0.974 which is the value of k which is closer to the value of 0 compared to the other k. The result of k-4 is k-0 with 66 members, k-1 55 members, k-2 with 5 members, and k-3 with 39 members.

![Figure 5. The Results of Grouping Using the K-Medoids](image_url)
an average of 126 users, then followed by Instagram whose average usage reaches 108 users, then followed by E-Learning with an average usage of 90 users, then followed by Facebook and YouTube which average -Average use of 71 users. So from the research that has been done it is necessary to apply layer 7 protocol firewall security to limit social media access so that internet network access at SMK Wahidin Kota Cirebon is used according to the initial idea, namely facilitating the academic community to optimize E-learning learning.

4. Conclusion

The conclusions in this study are:

a. WhatsApp users get the highest average usage of 126 users, while the lowest users are Facebook and YouTube users with 71 users.

b. E-Learning users average 90 users per day

c. E-Learning users are less than WhatsApp users so it is necessary to implement a Layer 7 Protocol Firewall to cover network usage other than E-Learning.

REFERENCES

[1] B. Ababio, “Nature of Teaching : What Teachers Need to Know and Do 2 . The Concept of Teaching,” *Int. J. Innov. Educ. Res.*, vol. 1, no. 2002, pp. 37–48, 2016.

[2] OECD and ADB, *Education in Indonesia: Rising to the Challenge*, vol. 20, no. 15. 2015.

[3] D. L. Srinivasulu, A. V. Reddy, and V. S. G. Akula, “Improving The Scalability And Efficiency Of K-Medoids By Map Reduce,” *Int. J. Eng. Appl. Sci.*, no. 4, pp. 88–90, 2015.

[4] I. I. P. Damanik, S. Solikhun, I. S. Saragih, I. Parlina, D. Suhendro, and A. Wanto, “Algoritma K-Medoids untuk Mengelompokkan Desa yang Memiliki Fasilitas Sekolah di Indonesia,” *Pro. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 520, 2019.

[5] D. A. Silitonga, A. P. Windarto, and D. Hartama, “Penerapan Metode K-Medoid pada Pengelompokan Rumah Tangga Dalam Perlakuan Memilah Sampah Menurut Provinsi,” *Semin. Nas. Sains Teknol. Inf.* SENSAISI 2019 ISBN, pp. 313–318, 2019.

[6] S. Defiyanti, M. Jajuji, and N. Rohmawati, “Optimalisasi K-MEDOID dalam Pengklasteran Mahasiswa Pelamar Beasiswa dengan CUBIC CLUSTERING CRITERION,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 3, no. 1, pp. 211–218, 2017.

[7] D. Marliana, N. Lina, A. Fernando, and A. Ramadhan, “Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokkan Wilayah Sebaran Cacat pada Anak,” *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, p. 64, 2018.

[8] I. Kamila, U. Khairennisa, and M. Mustakim, “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokkan Data Transaksi Bongkar Muat di Provinsi Riau,” *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 119, 2019.