Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5850

Submitted : 29 April 2025
Accepted : 15 May 2025
Published : 18 May 2025

Data Analysis of E-Journal Usage in UPM Library with K-Means Clustering Method

Susi Rachmadhani Sugiyarto[1)*], Rahmi[2)]
[1)*,2)]Universitas Indonesia, Indonesia
[1)*]susi.rachmadhani31@ui.ac.id, [2)]rahmi.ami@ui.ac.id

ABSTRACT

This study aims to evaluate the usage patterns of Emerald and WileyOnline Library e-journals from January to December 2023. By employing the K-Means clustering method, the data were classified to analyze usage characteristics and efficiency. The clustering results indicate that journals in clusters C1 and C2 have higher relevance compared to those in C3, based on download and access numbers. Evaluation using three metrics—average cost per e-journal, average cost per access, and appropriate content usage—revealed that e-journal usage at the UPM library is not yet efficient, with high average costs per access and content usage needing improvement. This study recommends strategies to enhance the efficiency of e-journal usage to better support academic activities and research at UPM.

Keywords: E-Journal; K-Means; Efficiency; Clustering; Evaluation

INTRODUCTION

In this era of digitalization, the development of information technology is very fast, making people's lives easier, including the use of gadgets that are increasingly widespread throughout the world, including Indonesia. In response to this, libraries are required to align their services with the latest technology to remain relevant in providing reliable information. One of these adaptation efforts is to provide electronic information sources such as e-journals, which contain scientific journals. E-Journals are journals available through electronic media or the web that have been formatted in an easy way for users who need scientific information (Sawitry, 2023). In the International Encyclopedia of Information and Library Science (Feather and Struges, 1997) electronic journal is defined as a term used to describe a journal published in digital form to be displayed on a computer monitor screen.

UPM, as a private higher education institution, actively facilitates access to accurate reference sources for students and lecturers, by subscribing to e-journals as an initiative to support research activities. These subscriptions not only expedite the research process but also broaden access to the latest information. The university library allocated IDR 273,130,000 for an annual subscription to Emerald e-journal and IDR 382,950,000 for an ongoing subscription to WileyOnline Library. This investment, although large, is seen as an essential strategic move to enrich academic and research activities.

When UPM libraries subscribe to e-journals from Emerald, they get access rights for the duration of the subscription period. After that period ends, access is terminated as the e-journal does not belong to the library. This is in contrast to WileyOnline Library's perpetual subscription, where the initial payment provides continued access even after the first subscription year, in line with the subscription model established for Emerald.

Based on previous studies, it was found that in terms of cost, EBSCO subscriptions provide benefits according to the cost evaluation. However, when viewed in terms of usage efficiency, only 6% of the total journals were accessed, indicating that the users were not efficient. Meanwhile, an evaluation of e-journal and e-book subscriptions from ScienceDirect by the Islamic University of Indonesia library showed significant benefits. These studies have reviewed the elements of electronic information services, cost evaluation and usage, but have not explored usage patterns in detail. Therefore, a more comprehensive study is planned to be conducted with the aim of analyzing usage patterns, measuring efficiency, as well as evaluating the associated costs.

Referring to the context that has been described, there is an urgent need to evaluate how e-journals are used and how efficient their utilization is when compared to the relatively expensive subscription costs. Research questions that arise include: how to characterize e-journal usage patterns, assess the level of usage efficiency against the financial

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5850

Submitted : 29 April 2025
Accepted : 15 May 2025
Published : 18 May 2025

investment made, and conduct an in-depth evaluation of subscription costs in detail. The results of this study are expected to provide greater insight into the use of e-journals in the library environment and provide a basis for strategic decisions regarding the continuation of e-journal subscriptions. This study will focus on analyzing e-journal usage data from January to December 2023 provided by UPM library, specifically for e-journals provided by Emerald and WileyOnline Library

LITERATURE REVIEW

Analysis of e-journal usage data has become the focus of research in the field of library and information in order to support evidence-based decision making. In the context of higher education libraries, e-journal usage data has a strategic role in evaluating service effectiveness, assessing the efficiency of digital collection investment, and formulating service development policies that are oriented to user needs.

Monitoring and analyzing electronic resource usage data provides a significant understanding of the value and benefits of digital collections in supporting academic activities (Tenopir et al., 2012). However, the approach has been descriptive and has not utilized in-depth data exploration techniques, such as clustering algorithms, to identify more complex usage patterns.

In an effort to look at user behavior, the K-Means Clustering algorithm has been used in a number of studies to group users based on behavioral attributes obtained from access log data. The application of K-Means algorithm to classify digital library users based on the frequency of access to electronic resources (Rahman & Farhana, 2020). The results show that the clustering approach is effective in identifying active and inactive user groups, which in turn can be used to develop more targeted service strategies. However, the study only considered one dimension of user behavior and did not utilize other more complex variables, such as the type of resources accessed, access time, and duration of use.

Previous research on the addition of data visualization dimensions in the application of K-Means to analyze user behavior in digital libraries (Liu et al., 2019). This study shows that the combination of clustering and visualization can strengthen the data-based decision-making process. However, the context of the study was a large-scale university library in China, so the generalizability of the findings to the context of private higher education institutions in Indonesia, such as UPM, needs to be further examined.

Another study examined the utilization of electronic resources in private university libraries in Indonesia (Setiawan & Yulianingsih, 2021). The library has provided various electronic resources but their utilization is still not optimal. Therefore, an analytical system is needed that is able to explore and present usage data more comprehensively. Although this research contributes to the understanding of e-resources usage, there is no application of advanced analytical methods such as K-Means Clustering in the user data segmentation process.

Based on previous studies, it was found that in terms of cost, EBSCO subscriptions provide benefits according to the cost evaluation. However, when viewed in terms of usage efficiency, only 6% of the total journals were accessed, indicating that the users were not efficient (Maulidya, 2023). Meanwhile, an evaluation of *e-journal* and *e-book* subscriptions from ScienceDirect by the Islamic University of Indonesia library showed significant benefits (Dewi, 2018). These studies have reviewed the elements of electronic information services, cost evaluation and usage, but have not explored usage patterns in detail.

Based on the review of the literature, there are several research gaps that form the basis for this study. First, there are still limited studies that examine the use of e-journals with a comprehensive K-Means Clustering approach, especially in the context of private university libraries in Indonesia. Second, there are not many studies that explore the use of e-journals by considering various multidimensional user behavior variables. Third, there is still a lack of integration between the results of cluster analysis and digital service development policies that are strategic and responsive to user needs.

Therefore, this research aims to fill the void by applying the K-Means Clustering method to e-journal usage data at UPM Library, in order to produce user behavior segmentation that can support the development of data-based library services more effectively and adaptively.

METHOD

Evaluating the effectiveness of e-journals requires in-depth data analysis. Therefore, the library must be able to analyze the pattern of e-journal usage, which can be realized through the application of the K-Means clustering method (Altbach, 2015). This analysis will be complemented by a statistical assessment based on efficiency to determine how

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5850

Submitted : 29 April 2025
Accepted : 15 May 2025
Published : 18 May 2025

effective the use of e-journals is compared to the costs incurred for subscriptions. The approach used in this study is ratio analysis, by comparing the results obtained (output) with the resources used (input).

K-Means, as a technique in data mining, serves to group data into one or more groups based on the similarity of their characteristics. With this method, data that has different characteristics can be placed in different groups, through the process of determining the group center (centroid) and identifying members of the group (Zamroni, 2016). In addition, there are two processes in the implementation of the K-Means algorithm, namely determining the location of the center of each cluster and identifying each cluster member (Septiani, 2023). Clustering is a commonly used method in data analysis to categorize a collection of data objects into separate groups based on their qualities or characteristics. In the context of measuring the distance between data, the Euclidean Distance formula is used as follows:

$$d(x_i, x_j)\sqrt{\left(\left|x_{i1} - x_{j2}\right|^2 + \left|x_{i2} - x_{j2}\right|^2 + \cdots + \left|x_{ip} - x_{ip}\right|^2\right)} \quad (1)$$

RESULT

Emerald and WileyOnline Library are a collection of databases containing the latest scholarly articles that can be accessed *online*. In Emerald and WileyOnline Library subscriptions, scientific articles can be accessed in *fulltext* or only in bibliographic and abstract form if not subscribed. The following is data on the use of Emerald and WileyOnline Library *e-journals* for the period January - December 2023 which will be used for usage data analysis.

1. Emerald and WileyOnline Library Usage Data

Emerald and WileyOnline Library provide usage data statistics to UPM Library every 3 months. Usage data from January to December 2023 can be seen in the table below:

Table 1. Emerald and WileyOnline Library Usage Data Table

No.	Journal Title	Subject	Platform	Print_ISSN	Total Access / Usability	Number of downloads
1	AEM Education and Training	Education	Wiley	2472-5390	1	2
2	Abacus	Accounting ; Finance ; Business	Wiley	0001-3072	96	6
3	Academic Emergency Medicine	Medicine	Wiley	1069-6563	3	6
4	Accounting & Finance	Accounting ; Finance	Wiley	0810-5391	719	20
5	Accounting Perspectives	Accounting	Wiley	1911-382X	155	4

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5850

Submitted : 29 April 2025
Accepted : 15 May 2025
Published : 18 May 2025

652	World Journal of Entrepreneurship, Management and Sustainable Development	Management ; Entrepreneurship	Emerald Insight	2042-5961	48	5
653	World Journal of Science, Technology and Sustainable Development	Technology	Emerald Insight	2042-5945	12	7
654	Worldwide Hospitality and Tourism Themes	Tourism	Emerald Insight	1755-4217	97	16
655	Young Consumers	Consumer	Emerald Insight	1747-3616	426	113
656	Critical perspectives on international business	Business	Emerald Insight	1742-2043	10	4

From Table 1, it can be seen that there is a difference between the number of accesses and the number of downloads on Emerald and WileyOnline Library e-journals. The number of accesses shows usage data in terms of how many times the e-journal is clicked by users, while the number of downloads shows data on the number of articles downloaded from each e-journal.

2. Implementation of K-Means Clustering

The implementation of the k-means clustering algorithm is done after data collection, followed by preprocessing which aims to clean the data so that it is ready to be used in the clustering and classification stages. The following are the results of preprocessing:

Table 2. Data Preprocessing Results

Data To-	Journal Title	Print_ISSN	Cluster	Silhouette	Subject	Platform	Total Access / Usability	Number of downloads
1	AEM Education and Training	2472-5390	C3	0,63787	Education	Wiley	0	0,00262

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5850

Submitted : 29 April 2025
Accepted : 15 May 2025
Published : 18 May 2025

2	Abacus	0001-3072	C3	0,63609	Accounting ; Finance ; Business	Wiley	0,09223	0,01312
3	Academic Emergency Medicine	1069-6563	C3	0,63799	Medicine	Wiley	0,00194	0,01312
371	Direct Marketing: An International Journal	1750-5933	C2	0,58196	Marketing	Emerald Insight	0	0
372	Disaster Prevention and Management: An International Journal	0965-3562	C2	0,60574	Management	Emerald Insight	0,00291	0
373	Education + Training	0040-0912	C2	0,58892	Education	Emerald Insight	0,0301	0,01575
507	Journal of Financial Reporting and Accounting	1985-2517	C1	0,57999	Accounting	Emerald Insight	0,41651	0,24409
502	Journal of Fashion Marketing and Management: An International Journal	1361-2026	C1	0,62447	Management ; Marketing	Emerald Insight	0,63981	0,31234
503	Journal of Financial Crime	1359-0790	C1	0,61031	Finance	Emerald Insight	0,92621	0,22572

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5850

Submitted : 29 April 2025
Accepted : 15 May 2025
Published : 18 May 2025

In this process, the "replace with random value" method was used because if "remove rows with missing values" was used, the data would be reduced and would not meet the research needs. The scatter plot below shows the relationship between the number of downloads and the number of accesses.

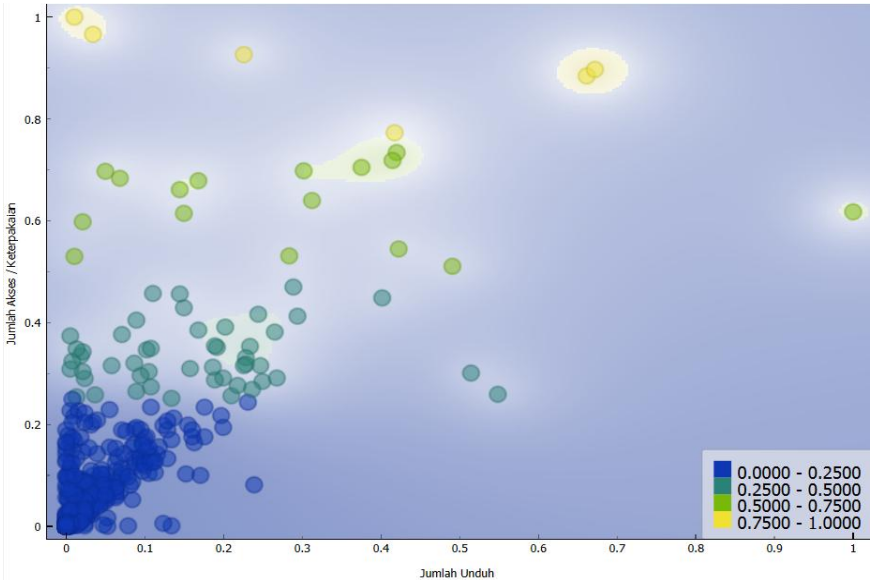

Fig. 1 Scatter Plot of Number of Downloads against Number of Accesses

The scatter plot in figure 1 above shows that most of the data is concentrated at low download and access values.

3. Testing Number of Clusters

In determining the number of clusters, researchers conducted tests using the silhouette coefficient to test the quality of the clusters. This process aims to determine the optimal number of clusters in the k-means algorithm. This research uses 657 data with testing the number of clusters from 2 to 8. The graph below illustrates the silhouette value for each cluster.

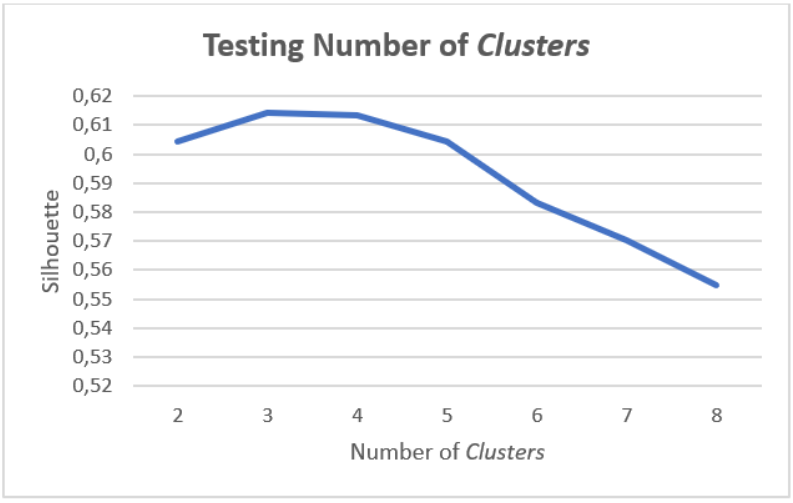

Fig. 2 Graph of Testing Results for the Number of Clusters

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5850

Submitted : 29 April 2025
Accepted : 15 May 2025
Published : 18 May 2025

Figure 2 shows that the number of clusters = 3 can produce a good average silhouette. The most optimal number of clusters is obtained when the number of clusters is 3 with an average silhouette value of 0.614105. This shows that with the number of clusters = 3, the level of similarity of data in one cluster is getting higher, while the level of similarity between clusters is getting lower.

4. Iteration Result

Iteration is performed until the cluster does not change. In this study, iterations were performed 10 times until no more clusters changed.

Table 3. Iteration Results

Number of Iterations	Number of Data Changes
Iteration - 1	374
Iteration - 2	72
Iteration - 3	8
Iteration - 4	7
Iteration - 5	8
Iteration - 6	4
Iteration - 7	3
Iteration - 8	1
Iteration - 9	1
Iteration - 10	0

From table 3 above, it can be seen that in the first iteration, the amount of data that has changed reaches the highest number, which is 374 data. This indicates that at the beginning of the iteration process, there were many changes in the processed data. In the second iteration, the number of changed data decreased drastically to 72 data, indicating an initial stabilization after the first significant iteration.

In the third iteration, the number of changed data again decreased sharply to 8 data. This downward trend continued in the fourth, fifth, and so on iterations, with the number of changed data being 7, 8, 4, and 3 in the fourth to seventh iterations, respectively. Starting from the eighth to the tenth iteration, the amount of data that changed was minimal, with 1 data in the eighth and ninth iterations, and 0 data in the tenth iteration. This indicates that the iteration process has reached a point of stability or convergence, where almost no more data changes.

Overall, the data shows a decreasing trend in the number of data changes as the number of iterations increases, with the most significant decrease occurring in the early iterations. The tenth iteration marks the point where there are no more changes in the data, indicating that the iteration process has reached a steady state or converged.

5. Clustering Result

Here are the clustering results summarized in tabular form:

Table 4. Clustering Results

Cluster	Subject	Description
C1	Accounting, finance, business, communications, economics, events, marketing, management and social sciences.	The number of downloads varied significantly, with some journals reaching up to 382 downloads, demonstrating the popularity and relevance of the topics covered in the journals.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5850

Submitted : 29 April 2025
Accepted : 15 May 2025
Published : 18 May 2025

C2	Business, management, accounting, finance, marketing, technology and tourism.	Cluster C2 shows the difference in the number of accesses and downloads of each journal. Journals with the highest usage tend to have a greater impact in the academic and professional community.
C3	More diverse subjects	The number of accesses and the number of downloads are much lower than C1 and C2.

6. Evaluation of E-Journal Collection with Three Metrics

E-journal subscriptions that provide collection usage statistics are helpful in evaluating collection usage. One way of evaluating a subscribed e-journal collection is by using three metrics: average cost per e-journal, average cost per access, and calculating appropriate content usage.

a. Calculating the average cost per e-journal

The formula used to calculate the mean e-journal price is as follows:
Mean = price : number of e-journals
The number of titles subscribed to the Emerald e-journal is 332 titles, while the number of titles subscribed to the WileyOnline Library e-journal is 325 titles. So that the mean of each article downloaded:
Article price - Emerald:
Subscription price: Number of e-journal titles
273.130.000 : 332 = 822,681 / e-journal
Article price - WileyOnline Library :
Subscription price: Number of e-journal titles
382.950.000 : 325 = 1,178,308/e-journal

b. Calculating the average cost per access

The formula used to calculate the mean price per e-journal access is as follows:
Mean = price / number of e-journal accesses
The number of collections accessed in the Emerald e-journal is 35,716, while the number of collections accessed in the WileyOnline Library e-journal is 22,664. Price per access - Emerald:
Subscription price: Number of e-journal accesses
273.130.000 : 35,716 = 7,647/access
Price per access - WileyOnline Library :
Subscription price: Number of e-journal accesses
382.950.000 : 22,664 = 16,897/access

c. Calculating the use of appropriate content

The formula used to calculate the appropriate use of content is as follows:
Content usage = number of accesses : number of article downloads
Content usage - Emerald:
35.716 : 9197 = 3.88 / access e-journal
Content usage - WileyOnline Library :
22.664 : 1,926 = 11,767 / e-journal access

In addition to evaluation through these three metrics, the efficiency of e-journal usage can also be seen with the following formula:

Efficiency = output/input
Efficiency - Emerald = 35.716 / 9197 = 3,88 = 3,88 %
Efficiency - WileyOnline Library = 22.664 / 1.926 = 11,767 = 11,77 %

When viewed from the total efficiency of using Emerald and WileyOnline Library e-journals of 3.88% and 11.77%, it can be concluded that the use of Emerald and WileyOnline Library e-journals in UPM libraries is not yet efficient.

DISCUSSIONS

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5850

Submitted : 29 April 2025
Accepted : 15 May 2025
Published : 18 May 2025

An analysis of Emerald and WileyOnline Library e-journal usage at Prasetiya Mulya University provides a clear picture of user behavior, resource effectiveness, and collection efficiency. The application of the K-Means Clustering algorithm successfully grouped the journal usage data into three different clusters. Cluster 1 and Cluster 2 contain journals with moderate to high usage, while Cluster 3 is a group of journals with very low usage.

The clustering results show that a large number of journals are in Cluster 3, which indicates a low level of usage. This can be caused by several factors such as lack of socialization to users, mismatch of journal topics with academic needs, or suboptimal access experience. On the other hand, journals in Clusters 1 and 2-especially in the fields of business, management, accounting, and technology-show a higher level of usage, reflecting the alignment between journal content and the focus of academic activities in the university environment.

In terms of subscription costs, the calculation of the average cost per journal and per access shows a difference between the two platforms. Emerald is more cost-efficient per access than Wiley. However, Wiley shows a higher level of content usage, as evidenced by its higher access-to-download ratio. This suggests that Wiley users tend to browse more content before downloading, which could indicate higher content relevance or broader search behavior.

Efficiency metrics show that the use of e-journals is still not optimal. Both Emerald and Wiley have efficiency levels below 15%, indicating the need for strategies to improve digital resource utilization. Efforts to increase efficiency can be made by promoting relevant and frequently used journals, reducing subscriptions to underused journals, and improving user information literacy through continuous training and promotion.

CONCLUSION

This study shows that data analysis and clustering techniques can be used as an effective approach in evaluating the usage of e-journal collections. Through the application of the K-Means Clustering algorithm and cost-benefit analysis, this study was able to identify usage patterns and group journals based on their performance, which can be used as a basis for making collection management decisions. The main findings of this research are: a) Most of the journals are in the low-usability group. b) Business, accounting, and technology are fields with high journal usage. c) The Emerald platform is more efficient in terms of cost per access, while Wiley shows a higher content utilization rate. d) In general, the efficiency level of e-journal usage is still relatively low and requires improvement. Based on these results, recommendations that can be given to the UPM Library include: 1) Conduct periodic usage evaluations to support subscription decision-making. 2)Promote journals that have high potential but are underutilized. 3)Customize the e-journal collection with the curriculum and research focus of lecturers. 4) Improve information literacy and e-journal access training for the academic community. For future research, it is recommended to combine this quantitative data with qualitative approaches such as user surveys or interviews, to gain a more thorough understanding of the behavior and information needs of e-journal users.

REFERENCES

Affiat, R. A. (2016, Maret). Perpustakaan Harvard sebagai contoh kemajuan universitas berbasis riset dan bertaraf internasional: Sebuah refleksi bagi UI. http://152.118.24.150/files/Rizki_Amalia_Affiat.pdf

Altbach, P. (2015). The costs and benefits of world-class universities. International Higher Education, (33). http://ejournals.bc.edu/ojs/index.php/ihe/article/viewFile/7381/6578

Arianto, S. (2012). Modul instalasi EPrint 3 pada Ubuntu (Matakuliah Perpustakaan Digital). Prodi Ilmu Perpustakaan, Fakultas Adab dan Ilmu Budaya, UIN Sunan Kalijaga.

Azra, A. (2006, April 20). World class university. Republika.

Dewi, M. (2019). Analisis statistik keterpakaian database online ScienceDirect (Januari–Juni 2018) dalam mengambil kebijakan pengadaan bahan pustaka elektronik pada Perpustakaan Universitas Islam Indonesia. Al-Kuttab: Jurnal Kajian Perpustakaan, Informasi dan Kearsipan, 1(1), 23–34.

Liu, Y., Zhang, X., & Fan, H. (2019). User behavior analysis in digital libraries using clustering and data visualization. Library Hi Tech, 37(4), 742–758. https://doi.org/10.1108/LHT-10-2017-0211

Maulidya, B. S. (2023). Analisis statistik penggunaan database online EBSCO dalam membuat kebijakan pengadaan bahan pustaka elektronik pada Perpustakaan ACD. Libraria: Jurnal Perpustakaan, 11(2), 323–346.

Rahman, M., & Farhana, M. (2020). User segmentation in digital library using K-means clustering algorithm: A case study. Journal of Information Science Theory and Practice, 8(2), 25–38. https://doi.org/10.1633/JISTAP.2020.8.2.3

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5850

Submitted : 29 April 2025
Accepted : 15 May 2025
Published : 18 May 2025

Sawitry, N. D. (2023). Pemanfaatan koleksi e-journal bidang ekonomi dan bisnis oleh civitas akademika pada Perpustakaan Fakultas Ekonomi dan Bisnis (FEB) UIN Syarif Hidayatullah Jakarta [Skripsi, UIN Syarif Hidayatullah Jakarta].

Septiani, N., Anwar, S., & Herdiana, R. (2023). Penerapan algoritma K-means clustering untuk harga rumah. Trending: Jurnal Ekonomi, Akuntansi dan Manajemen, 1(2), 35–47.

Setiawan, R., & Yulianingsih, W. (2021). Analisis penggunaan sumber daya elektronik di perpustakaan perguruan tinggi: Studi kasus pada perguruan tinggi swasta di Jakarta. Jurnal Ilmu Perpustakaan dan Informasi, 6(1), 45–58.

Tenopir, C., King, D. W., Edwards, S., & Wu, L. (2022). Use and users of electronic library resources: An overview and analysis of recent research studies. Council on Library and Information Resources.

Zamroni, Z. (2016, Maret 7). Wawancara oleh Mubasyaroh [Wawancara]. Sekolah Tinggi Agama Islam Kudus.

* Corresponding author