

Decision support system for lecturer publication mapping using k-means clustering method

Sri Sumarlinda¹, Wijiyanto², Wiji Lestari³

^{1,2,3}Faculty of Computer Science, University of Duta Bangsa Surakarta, Jawa Tengah 57154, Indonesia

Article Info

Article history:

Received Nov 2, 2022

Revised Dec 19, 2022

Accepted Dec 26, 2022

Keywords:

decision support system
clustering
lecturers publication
K-Means

ABSTRACT

Publication is an important tridharma activity for lecturers. This study aims to produce a clustering model using the K-Means algorithm which was built for ease of operation of publications. The method used is research and development which includes the stages of data collection, data preprocessing, clustering process and cluster analysis. The input data consists of 87 with 8 attributes, namely the number of journal articles indexed by Sinta, the number of journal articles indexed by Scopus, the number of citations in Scopus, the H-index in Scopus, the number of articles in indexed journals in Google Scholar, the number of citations in Google Scholar, the H-index in Google Scholar and H-index10 in Google Scholar. The K-Means algorithm is used with 3 clusters and 100 epochs. The clustering results are divided into 3 clusters, namely cluster 1 with 17 members, cluster 2 with 32 members and cluster 3 with 38 members. Clustering with 5 clusters produces cluster 1 with 5 members, cluster 2 with 12 members, cluster 3 with 20 members, cluster 4 with 18 and cluster 5 with 32 members. The results of the cluster analysis show that the clustering process with 3 clusters is improved and the academic application is better than clustering with 5 clusters.

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license.



Corresponding Author:

Sri Sumarlinda,
Faculty Of Computer Science,
University of Duta Bangsa Surakarta,
Jl. Bhayangkara No.55, Tipes, Kec. Serengan, Kota Surakarta, Jawa Tengah 57154
Email: sri_sumarlinda@udb.ac.id

1. Introduction

Publication is an activity related to research. Research is part of a lecturer's duties. Mapping and researching the performance of lecturers in research, especially publications, is very much needed for the development of lecturer careers and the development of higher education institutions. The basis used in the assessment is based on the tri dharma of higher education. Law Number 20 of 2003 concerning the National Education System states that universities are obliged to carry out education, research and community service. Lecturers are an important component of the education system in tertiary institutions (Lestari & Sumarlinda, 2020), (Hastuti et al, 2018). The role, duties and responsibilities of lecturers are very important in achieving the goals of national education, namely educating the life of the nation, improving the quality of Indonesian people which includes the quality of faith/piety, noble character, and mastery of science, technology and art, as well as realizing an advanced Indonesian society. fair, prosperous, and civilized (Sugiharti & Muslim, 2016), (Purnomo et al, 2016), (Purnomo et al, 2017). The quality of lecturer performance can be reflected in the productivity and quality of implementing the Tridharma. The Directorate General of Higher Education has implemented

a lecturer performance load program and the evaluation of lecturer assignments is running according to the criteria stipulated in the laws and regulations. (Sugiharti & Muslim, 2016) (Mabu & Muhammad, 2016), (Soriano et al, 2018).

Publication is part of the lecturer's research assignments. Research is an important task for lecturers besides teaching and carrying out community development. Some lecturers consider research, including publication, to be a difficult activity. Publication is a means to disseminate research results or thoughts from researchers or lecturers. Publications can be made in scientific journals and proceedings, both on a national and international scale. Scientific publications are also linked to indexing institutions such as Scopus, Web of Science (WoS), Copernicus, DOAJ, Google Scholar, and others. Scopus indexed or WoS journals are prestigious journals for publication. Lecturer publication performance is based on the number of publications, the number of citations and the quality of the journal for publication. Journal quality is determined by journal indexing institutions (Sugiyarti & Muslim, 2016), (Nursyahputra et al, 2017).

Grouping is used to analyze different data groupings, similar to classification, but these groupings are not defined before data mining tools are implemented. Clustering divides the data into groups based on what the data mining tool finds. The principle of clustering is to maximize the similarities between class members and minimize the similarities between clusters (Lestari, 2014), (Jipkate & Gohokar, 2012). Clustering can be done on data that has several attributes which are mapped as a multidimensional space and are usually of numeric type. Clustering tries to find the components of the group naturally, based on the proximity of the data, usually measured by Euclidean distance. Data grouping usually uses unsupervised learning and statistical data analysis techniques. The model is used to classify the same data into homogeneous groups. It is used to operate large data sets to find hidden patterns and relationships to help make decisions quickly and efficiently (Sugiharti & Muslim, 2016), (Lestari, 2014), (Jipkate & Gohokar, 2012).

The era of the industrial revolution 4.0 and Society 5.0 was marked by the development of an information technology paradigm for automation and better performance. Big data, artificial intelligence, and internet of things (IoT) are 3 new paradigms of computing that play an important role in this era. Clustering is part of data mining, clustering is usually used for numeric data without a label attribute. The application of clustering in education includes mapping academic institutions, mapping the performance of lecturers and employees, student achievements and others (Sugiharti & Muslim, 2016), (Lestari, 2015), (Sumarlinda & Lestari, 2019), (Shovon & Haque, 2012), (Oyelade et al, 2010), (Widiyaningtyas, 2017), (Nursyahputra et al, 2017).

The clustering method used in this study is the K-Means method. K-Means can also be interpreted as a clustering method which is included in the partitioning approach. The K-Means algorithm is a center of mass model. Centroid mode is a model that uses centroids to create clusters. Centroid is the midpoint of a cluster. Centroid is used to calculate the distance of the data object to the centroid. A data object is included in the cluster if it has the shortest distance from the cluster centroid. The K-Means algorithm can be interpreted as a simple learning algorithm for solving grouping problems that aim to minimize double errors (Oyelade et al., 2010), (Velmurugan, 2012), (Cebeci & Yildiz, 2015). Clustering was chosen because the lecturer's publication performance data is not valid if it is labeled, for example low, medium and high, considering that the input data is homogeneous with a low tendency to publish. It is hoped that the results and clustering analysis can be used as input for the decision support management process for campus institutions, especially in assessing lecturer research performance.

2. Methods

This study implements the K-Means algorithm for classifying lecturer performance data in publications. Data collection is the first step to collect lecturer publication performance data that will be clustered. Input data was taken from an indexing institution with the subject as a lecturer at Duta Bangsa University, Surakarta. Data accessed for the period October 2020. The input data totaled 84 data taken from various journal indexing institutions, especially Google Scholar, Sinta and Scopus. The data attributes for clustering are the number of Sinta indexed journal articles, the number of Scopus indexed journal articles, the number of citations in Scopus, the H-index in Scopus, the number of articles in Google

Scholar indexed journals, the number of citations in Google Scholar, H – index in Google Scholar and H - index10 on Google Scholar.

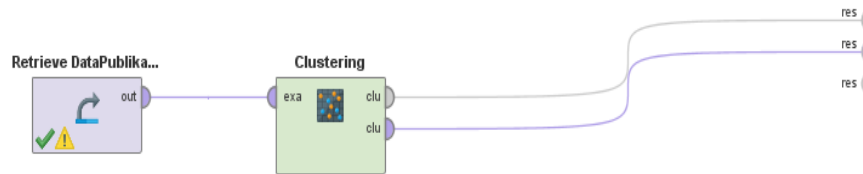


Figure 1. Clustering Using the K-Means Method

Data preprocessing consists of data cleaning and data transformation. Data cleansing is to remove missing data and attributes. Some data is not valid, such as not many articles because articles from other parties are included or articles that have not been indexed in one indexer but have been indexed in another indexer. Data transformation is changing data so that it can be used as input data for the clustering process using K-Means.



Figure 2. Clustering Process

The clustering process uses the K-Means algorithm. The clustering parameters used are the number of clusters 3 and 5, while the number of epochs is 100. After the clustering process, the results of the clustering analysis are carried out . Analysis of clustering results can be used to map lecturer performance as well as decision support data.

3. Results and Analysis

The clustering model with the K-Means algorithm uses 87 objects as input data with 8 numeric variables. The data attributes for clustering are the number of articles in Sinta indexed journals, the number of articles in Scopus indexed journals, the number of citations in Scopus, the H-index in Scopus, the number of articles in Google Scholar indexed journals, the number of citations in Google Scholar, H – index in Google Scholar and H-index10 in Google Scholar. The K-Means algorithm for this clustering process produces 3 clusters and 5 clusters with 100 epochs . The clustering process uses Rapidminer. The results of clustering with 3 clusters are distributed into 3 clusters namely cluster 1, cluster 2 and cluster 3. The cluster distribution is shown in Figure 3. Below.

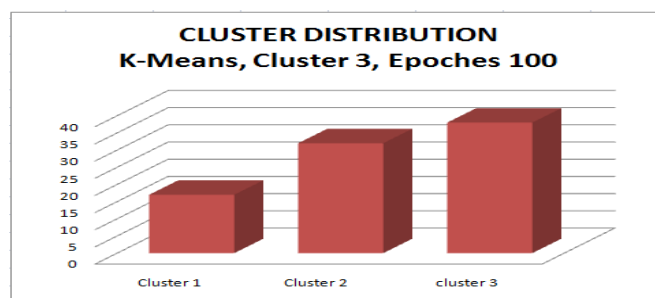


Figure 3. Cluster Distribution with K-means: 3 Clusters , 100 epochs

Cluster 1 has 17 lecturers, cluster 2 has 32 lecturers, and cluster 3 has 38 lecturers. Details of the membership of each cluster are shown in table 1.

Table 1
Number of Cluster Members (3 Clusters)

No	Clusters	Number of Cluster Members
1	Clusters 1	Data no.1, 3, 4, 5, 7, 9, 11, 12, 13, 14, 20, 21, 32, 34, 36, 39, 60. (17)
2	Clusters 2	data no. 2, 6, 8, 16, 17, 18, 19, 22, 23, 24, 25, 27, 28, 29, 30, 31, 38, 41, 45, 46, 47, 48, 49, 50, 51, 52, 54, 55, 56, 57, 61, 78. (32)
3	Clusters 3	data no. 10, 15, 26, 33, 35, 37, 40, 42, 43, 44, 53, 58, 59, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 79, 80, 81, 82, 83, 84, 85, 86, 87. (38)

Cluster analysis it can be seen that cluster 1 consists of lecturers whose publication performance is relatively high. Cluster 2 is being and cluster 3 consists of lecturers with relatively low performance. Cluster analysts can also view cluster centroids per cluster. Cluster 1, centroid mean value cluster is higher than cluster 2 and cluster 3. The average centroid of cluster 3 is the lowest. The distribution of centroid clusters is shown in Table 2

Table 2. Centroids Clusters (3 Clusters)

Attribute	Clusters 1	Clusters 2	Clusters 3
Number of Documents in Sinta	1,412	1,344	0.263
Number Documents in Scopus	1.118	0.219	0
Number of Citation in Scopus	1941	0.188	0
H-Index in Scopus	0.529	0.125	0
Number of Documents in Google Scholar	20,059	14,750	5053
Number of citations in Google scholar	37	14,375	2,553
H-Index in Google Scholar	3	2	0.711
H-Index10 in Google Scholar	0.882	0.344	0.026

The clustering process using 5 clusters with 100 epochs produces the output as shown in Figure 4.

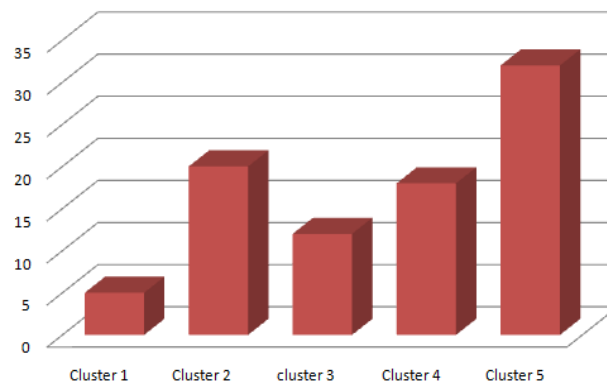


Figure 4. Cluster Distribution with K-means: 5 Clusters, 100 epochs

In detail the number of members per cluster for this process is as shown in table 3 below.

Table 3. Number of Cluster Members (5 Clusters)

No	Clusters	Number of Cluster Members
1	Clusters 1	data no. 1 3 5 12 13 (5)
2	Clusters 2	data no. 4 7 9 11 14 20 21 32 34 36 39 60 (12)
3	Clusters 3	data no. 2 8 15 16 17 19 22 23 24 25 28 30 38 42 48 50 56 57 61 78 (20)
4	Clusters 4	data no. 6 18 26 27 29 31 40 41 45 46 47 49 51 52 53 54 55 58 (18)

5	Clusters 5	data no. 10 33 35 37 43 44 59 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 79 80 81 82 83 84 85 86 87 (32)
---	------------	--

Values per cluster for 5 cluster processes with 100 epochs are shown in table 4 below.

Table 4. Centroid Clusters (5 Clusters)

Attribute	Clusters 1	Clusters 2	Clusters 3	Clusters 4	Clusters 5
Number of Documents in Sinta	2,600	0917	1,700	0.556	0.281
Number Documents in Scopus	2,600	0.500	0.250	0.111	0
Number of Citation in Scopus	5,200	0.583	0.250	0.056	0
H-Index in Scopus	1,200	0.250	0.150	0.056	0
Number of Documents in Google Scholar	30,400	15,750	17,350	9,722	4,438
Number of citations in Google scholar	33	38,667	11,400	15,944	1,312
H-Index in Google Scholar	3	3	1,900	2	0.531
H-Index10 in Google Scholar	0.400	1,083	0.150	0.500	0

Cluster analysis it can be seen that cluster 1 consists of lecturers whose publication performance is relatively high. Clusters 2 and 3 are and clusters 4 and 5 consist of lecturers with relatively low performance. The results of the clustering process are expected to be input data for the decision support system process for determining lecturer publication performance. Advanced model development can involve factors related to the tridharma, namely teaching, research and community service.

4. Conclusion

The clustering model using the K-Means algorithm was built to map lecturer publication performance. The input data consists of 87 with 8 attributes, namely the number of journal articles indexed by Sinta, the number of journal articles indexed by Scopus, the number of citations in Scopus, the H-index in Scopus, the number of articles in indexed journals in Google Scholar, the number of citations in Google Scholar, the H-index in Google Scholar and H-index10 in Google Scholar. The K-Means algorithm is used with 3 clusters and 5 clusters with 100 epoch s. Clustering results can map lecturers with high, medium and low publication performance. The clustering process with 3 clusters is computationally and academically better than clustering with 5 clusters.

References

- Cebeci, Z. and Yildiz, F., 2015, Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures, *Journal of Agricultural Informatics* (ISSN 2061-862X) 2015 Vol. 6, No. 3.
- Ghosh, S. and Dubey, SK, 2013, Comparative Analysis of K-Means and Fuzzy C-Means Algorithms, ((IJACSA) *International Journal of Advanced Computer Science and Applications*, Vol. 4, No.4.
- Hastuti, I., Purnomo, S. and Lestari, W., 2018, The guidance of Ttechnopreneurship using Expert System Computing Approach Based on Entrepreneurial Values and Multiple Intelligence, *International Journal of Economics, Business and Accounting Research (IJEBAR)* , Vol.2, Issue 3.
- Jipkate, B. and Gohokar, VV, 2012, A Comparative Analysis of Fuzzy C-Means Clustering and K Means Clustering Algorithms, *International Journal Of Computational Engineering* , Vol. 02, Issue 03.
- Lestari, Wiji. 2014. Student Multiple Intelligence Clustering System Using Self Organizing Maps (SOM) Algorithm. *Journal of Saintechn, Journal of Research and Community Service* , ISSN: 2355-5009, pages: 52 – 58.
- Lestari, W. and Sumarlinda, S., 2020, Implementation of Artificial Neural Network for Clustering of Interested Students in Computing Fields , *International Journal of Multi Science* , Vol. 1, No. 5.
- Mabu, AM, Muhammad, FA 2016. An Admission Decision Support System for Nigerian Universities. *International Journal of Computer Applications* (0975 - 8887) Volume 133 - No.2, January 2016.

- Nurzahputra, A., Much Aziz Muslim, MA and a Khusniati, M., 2017, Application of the K-Means Algorithm for Lecturer Assessment Clustering Based on the Student Satisfaction Index, *Techno.COM*, Vol. 16, No. 1.
- Oyelade, OJ, Oladipupo, OO and Obagbuwa, IC, 2010, Application of K-Means Clustering algorithm for prediction of Students' Academic Performance, (*IJCSIS*) *International Journal of Computer Science and Information Security*, Vol. 7, No.1.
- Purnomo, S., Lestari, W., Hastuti, I. 2017. Expert System Design for Guidance of Information Technology (IT) Technopreneurships Based on Students Personal Characteristics. *International Journal of Engineering Research and Technology (IJERT)*, ISSN : 2278-0181.
- Purnomo, S., Lestari, W., Hastuti, I. 2016. The Mapping The Information Technology (IT) Technopreneurship Based on Learning Styles and Multiple Intelligence. *Paper Proceedings International Conference of Education and Technology for Empowering Agents of Change* ISSN : 2541-1578 November 2016. Proceedings of International Seminar, Faculty of Teaching and Education, Satya Wacana Christian University Salatiga, Central Java, Indonesia.
- Shovon, MHI and Haque, M., 2012, An Approach of Improving Student's Academic Performance by using K-means Clustering Algorithm and Decision Tree, (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, Vol.3, No. 8.
- Soriano, LT, Capucac, JNB, Palaoag, TD 2018. Decision Support Model for Managing Academic Workload in State Universities and Colleges. *International Journal of Information and Education Technology*, Vol. 8, No. 7, July 2018.
- Sugiyarti, E. and Muslim, MA, 2016, On-line Clustering of Lecturers Performance of Computer Science Department of Semarang State University using K-Means Algorithm, *Journal of Theoretical and Applied Information Technology*, Vol.83. Number 1.
- Sumarlinda, S. and Lestari, W., 2020, Design of Decision Support System for Mapping Student Interest in Information Technology and Computing Specialization, *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, Volume 6, Issue 1.
- Velmurugan, T., 2012, Performance Comparison between k-Means and Fuzzy C-Means Algorithms using Arbitrary Data Points, *Wulfenia Journal*, Vol 19, No. 8.
- Widyaningtyas, T., Prabowo, MIW and Pratama, MAM, Implementation of K-Means Clustering Method to Distribution of High School Teachers, Proc. EECSI 2017, Yogyakarta, Indonesia, 19-21 September 2017.