JurnalMantik
Volume 4, Number 3, Nov 2020, pp. 2225-2231 E-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik

Book Data Grouping in the Library Using the K-Means Clustering Method

Heru Sugara[1], M. Ali Hanafiah[2]

[1]Prodi Sekretari, Politeknik Bisnis Indonesia, Jl. Sriwijaya No. 9 C-E, Pematangsiantar, Indonesia
[2]Prodi Komputerisasi Akuntasi, Politeknik Bisnis Indonesia, Jl. Sriwijaya No. 9 C-E, Pematangsiantar, Indonesia

E-mail: thesugara77@gmail.com, ikh.alie84@gmail.com

ARTICLEINFO

Article history:
Received: 01/11/2020
Revised: 18/11/2020
Accepted: 30/11/2020

Keywords:
Book, Grouping, Clustering, Ouput, Results

ABSTRACT

This study aims to classify the book information contained in the Business Indonesia Polytechnic library. Book grouping is done using the K-Means Clustering method. In this K-means clustering algorithm, the variables used as input are: book id, book title, total loan and copies. The resulting output consists of 3 clusters, namely books that are borrowed most frequently, books that are borrowed frequently, and books that are rarely borrowed. With the use of the K-means clustering method, the final results of the grouping are obtained up to the 6th iteration, where the center point no longer changes and no data moves between clusters. The final results obtained consisted of: members of cluster 1 consisting of 119 book titles, cluster2 of 8 books, and cluster3 of 21 books. From the lending data, the data contained in cluster 3 is the book group with the highest loan amount among the other 2 clusters. In addition, books contained in cluster 3 have the fewest copies. From the results of this data analysis, it can be seen that the book titles contained in cluster 3 are the most recommended books to be added to the Business Indonesia Polytechnic library

1. Introduction

The development of information technology is increasingly sophisticated and is increasingly being used in all areas of life. Lots of data are processed using computer-based technology[1], [2], starting from the industrial sector [3], economy [4] - [7], health[8] - [10], education [11], [12], as well as various other areas of life [12] - [14].

Data mining is the process of searching for patterns or information in selected data through the use of certain methods [11]. Techniques, methods or algorithms in data mining vary widely. The choice of the right method or algorithm is very much dependent on the objectives and process of extracting information as a whole. One of the data mining methods used in this research is clustering where the method can identify objects that have certain characteristics in common, and then use these characteristics as "centroids"[15], [16].

The library is a book warehouse, as well as the Indonesian Business Polytechnic library, so many books are contained in it, some books are stacked and not neatly arranged on bookshelves due to limited space..The average number of visitors who come to borrow books from the Business Polytechnic library is increasing every day. In the Indonesian Business Polytechnic library the admin always inputs data on books borrowed by students or lecturers and stored in the database. From existing historical data, the data is processed into new information which functions to find out which books are most often borrowed, often borrowed and books that are rarely borrowed..

To obtain this knowledge, data mining is carried out using the K-Means method. Through this method, data is grouped to determine which books are most often borrowed from the Indonesian Business Polytechnic library.

2. Research Method

The method used for grouping the books that are most often borrowed at the Indonesian Business Polytechnic library is carried out in several stages, namely, conducting the analysis required by the system consisting of: the need for input data systems to be built, namely data for borrowing books at the Indonesian Business Polytechnic library which can be sampled for analysis. The next stage is carried out by processing or processing the input data, namely using the K-means clustering technique. And the expected output is to

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

produce 3 clusters, namely books that are borrowed most often, books that are often borrowed, and books that are rarely borrowed.

The data material to be used in grouping this book data consists of book lending data from December 2018 to the end of January 2020, which consists of 548 records of book lending as in table 1

TABLE 1
SAMPLING DATA

No.	Member ID	Title	Ex	Borrow Date	Must Return Date
1	18310026	Introduction to Accounting: Concepts & Techniques for Preparing Financial Statements	1	12/18/2018	12/25/2018
2	18410012	Introduction to Accounting: Based on SAK ETAP and IFRS	1	12/17/2018	12/24/2018
3	16310002	Practical Chinese For Business: Complete With Phonetics & Vocabulary Lists	1	12/13/2018	12/20/2018
4	16210011	Auditing: Accounting Examination	1	12/4/2018	12/11/2018
5	16410023	Human Resource Management	4	12/3/2018	12/10/2018
6	16310001	Bank Fund Management Ed. 2	1	12/3/2018	12/10/2018
7	18310033	The 7 Habits Of Highly Effective People: 7 Habits of Highly Effective People	3	11/29/2018	12/06/2018
8	16310013	Managing the Quality of Banking Services	1	11/28/2018	12/5/2018
9	16410020	Management Secretary	1	11/27/2018	12/4/2018
10	18510017	The 7 Habits Of Highly Effective People: 7 Habits of Highly Effective People	3	11/27/2018	12/4/2018
…	…	…	…	…	…
540	18310025	Managing the Quality of Banking Services	1	8/1/2020	1/15/2020
541	17310001	Bank Fund Management Ed. 2	1	8/1/2020	1/15/2020
542	17310002	Islamic Banking: Theory, Policy, and Empirical Studies in Indonesia	1	8/1/2020	1/15/2020
543	18110041	13 Concepts of Beyond Leadership: Continued More About Beyond Leadership: 12 Concepts of Leadership	1	7/1/2020	1/22/2020
544	19210013	Accounting Basics Volume 1 Ed. 6	1	6/1/2020	1/20/2020
545	19110021	Java: Theory, Algorithms, and Applications	3	6/1/2020	13/1/2020
546	17310020	Capital Market & Portfolio Management Ed.2	1	6/1/2020	1/20/2020
547	17310033	Capital Market Business Law	1	6/1/2020	1/20/2020
548	17310031	Introduction to Capital Markets: Designed to Study Capital Markets Easily and Practically	1	6/1/2020	1/20/2020

After the book data is collected, the next step is to recapitulate the book loan data. At this stage, the process of calculating the amount of borrowing is carried out based on the same book, which aims to process the data using the K-means Clustering algorithm. The data variables of book lending consist of book id, book title, total loan and copies. The book recapitulation results consisted of 148 book titles borrowed along with the total loan and copies of each book. The results of the recitation of borrowing books at the Indonesian Business Polytechnic library can be seen in table 2.

TABLE 2
DATA TRANSFORMATION RESULTS

Id	Title	Total Loans	Copies
Book1	10 Exams In Preparation & Practice Exam TOEFL: Test of English As a Foreign Language	4	3
Book2	13 Concepts of Beyond Leadership: Continued More About Beyond Leadership: 12 Concepts of Leadership	2	1
Book3	Records Administration: An Introduction	4	1
Book4	Modern Office Administration Ed.4	2	1
Book5	Accounting: Basis for Business Decision Making Volume II	2	1
Book6	Cost accounting	8	1
Book7	Cost Accounting Ed. 3	4	1
Book8	Intermediate Accounting Ji. 2 ed. 12	2	1
…	…	…	…
Book140	Procedures for the Establishment and Management of Village-Owned Enterprises	5	1
Book141	Regional Financial Economic Governance	8	1
Book142	Modern Data Communication Technology	2	1
Book143	Theory and Applications of Digital Systems	2	1
Book144	The 7 Habits Of Highly Effective People: 7 Habits of Highly Effective People	7	3
Book145	The Cashflow Quadant: Rich Dad's Guide to Financial Freedom	2	1
Book146	The Power of OwnCloud: Build and Manage Your Own Cloud Technology in Your Environment	2	1
Book147	Programming Guide: Visual dBase 5.5	2	1
Book148	10 Day Tutorial: Building Database Applications With: Vb.Net	1	1

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

3. Result and Discussion

3.1 Application of K-means Clustering

After the transformation process, the next step is the data processing by applying the k-means clustering algorithm.

The first step that must be done is to determine the number of clusters to be formed. In this study, the clusters to be formed were 3 clusters. Then determine the starting center point of each cluster. In this study, the initial center point is determined randomly and the center point of each cluster is obtained based on the data as follows:

TABLE 3
THE STARTING POINT OF THE CLUSTER

Cluster	Id	Total Loans	Copies
Cluster 1	Book68	7	4
Cluster 2	Book85	5	2
Cluster 3	Book144	7	3

After the center point is obtained, then the distance from each data book to the cluster center point will be calculated. The calculation of each data to each cluster center point for iteration 1 is as follows:

The distance from the 1st data to the cluster center

Cluster 1 $= \sqrt{(4-7)^2 + (3-4)^2} = 3.1623$

Cluster 2 $= \sqrt{(4-5)^2 + (3-2)^2} = 1.4142$

Cluster3 $= \sqrt{(4-7)^2 + (3-3)^2} = 3.0000$

The distance from the 2nd data to the cluster center

Cluster 1 $= \sqrt{(2-7)^2 + (1-4)^2} = 5.8310$

Cluster 2 $= \sqrt{(2-5)^2 + (1-2)^2} = 3.1623$

Cluster3 $= \sqrt{(2-7)^2 + (1-3)^2} = 5.3852$

The distance from the 3rd data to the cluster center

Cluster 1 $= \sqrt{(4-7)^2 + (4-4)^2} = 4.2426$

Cluster 2 $= \sqrt{(4-5)^2 + (1-2)^2} = 1.4142$

*Cluster*3 $= \sqrt{(4-7)^2 + (1-3)^2} = 3.6056$

The calculation continues until the last data, namely the 148th data and the distance of all data after being calculated to each cluster can be seen in table 4.

TABLE 4
THE RESULTS OF EACH DATA CALCULATION IN ITERATION 1

Id	Total Loans	Copies	C1	C2	C3	Shortest Distance
Book1	4	3	3.1623	1.4142	3,000	1.4142
Book2	2	1	5.8310	3.1623	5.3852	3.1623
Book3	4	1	4.2426	1.4142	3.6056	1.4142
Book4	2	1	5.8310	3.1623	5.3852	3.1623
Book5	2	1	5.8310	3.1623	5.3852	3.1623
Book6	8	1	3.1623	3.1623	2.2361	2.2361
Book7	4	1	4.2426	1.4142	3.6056	1.4142
Book8	2	1	5.8310	3.1623	5.3852	3.1623
….	….	….	….	….	….	….
Book140	5	1	3.6056	1.0000	2.8284	1.0000
Book141	8	1	3.1623	3.1623	2.2361	2.2361
Book142	2	1	5.8310	3.1623	5.3852	3.1623
Book143	2	1	5.8310	3.1623	5.3852	3.1623
Book144	7	3	1.0000	2.2361	-	-
Book145	2	1	5.8310	3.1623	5.3852	3.1623
Book146	2	1	5.8310	3.1623	5.3852	3.1623
Book147	2	1	5.8310	3.1623	5.3852	3.1623
Book148	1	1	6,7082	4.1231	6.3246	4.1231

From table 4 above, then choose the smallest value because that value is the least distanceclose to the cluster. For more details, see table 5.

TABLE 5
CLUSTER POSITION IN ITERATION 1

Id	Total Loans	Copies	Closest Cluster Distance		
			C1	C2	C3
Book1	4	3	0	1	0
Book2	2	1	0	1	0
Book3	4	1	0	1	0
Book4	2	1	0	1	0
Book5	2	1	0	1	0

Id	Total Loans	Copies	Closest Cluster Distance		
			C1	C2	C3
Book6	8	1	0	0	1
Book7	4	1	0	1	0
Book8	2	1	0	1	0
….	….	….	….	….	….
Book140	5	1	0	1	0
Book141	8	1	0	0	1
Book142	2	1	0	1	0
Book143	2	1	0	1	0
Book144	7	3	0	0	1
Book145	2	1	0	1	0
Book146	2	1	0	1	0
Book147	2	1	0	1	0
Book148	1	1	0	1	0

The results of the iteration 1 cluster consist of 6 data on C1, 129 data on C2 and 13 data on C3.

After all the data are in the closest cluster, then a recalculation is carried out to determine the new cluster center based on the average member of the cluster. Calculations to determine the new cluster center point from the new data, namely by adding up the values of all cluster members then divided by the total number of cluster members. After calculating the new cluster center point, it can be seen in table 6.

TABLE 6
THE CENTER POINT OF THE NEW CLUSTER

Cluster	Id	Total Loans	Copies
Cluster 1	Book68	7.1667	6.6667
Cluster 2	Book85	2.8062	1.3047
Cluster 3	Book144	8.7692	1.3846

After the new center point is obtained, then perform the second iteration by performing the same calculation, namely between the data and the new cluster center.

The results of calculating each data to each cluster center point for the 2nd iteration can be seen in table 7.

TABLE 7
THE RESULTS OF THE CALCULATION OF EACH DATA TO EACH CLUSTER IN THE 2ND ITERATION

Id	Total Loans	Copies	C1	C2	C3	Shortest Distance
Book1	4	3	4.8448	2.0735	5.0354	2.0735
Book2	2	1	7.6685	0.8619	6,7801	0.8619
Book3	4	1	6.4914	1.2321	4,7847	1.2321
Book4	2	1	7.6685	0.8619	6,7801	0.8619
Book5	2	1	7.6685	0.8619	6,7801	0.8619
Book6	8	1	5,7276	5.2027	0.8600	0.8600
Book7	4	1	6.4914	1.2321	4,7847	1.2321
Book8	2	1	7.6685	0.8619	6,7801	0.8619
….	….	….	….	….	….	….
Book140	5	1	6.0668	2.2149	3,7888	2.2149
Book141	8	1	5,7276	5.2027	0.8600	0.8600
Book142	2	1	7.6685	0.8619	6,7801	0.8619
Book143	2	1	7.6685	0.8619	6,7801	0.8619
Book144	7	3	3.6705	4.5235	2.3958	2.3958
Book145	2	1	7.6685	0.8619	6,7801	0.8619
Book146	2	1	7.6685	0.8619	6,7801	0.8619
Book147	2	1	7.6685	0.8619	6,7801	0.8619
Book148	1	1	8.3749	1.8317	7,7787	1.8317

The next step is to calculate the closest distance to the cluster. Compare the members of each iteration 1 and iteration 2 clusters, if the position of the cluster members is still changing, then continue to iteration 3. If the position of the cluster members does not change, then the iteration is terminated. The calculation to determine the new cluster center point from the new data is done by adding up the values of all cluster members and then dividing by the total number of cluster members.

Based on the results of grouping all data using the k-means clustering method, the final results of the grouping are obtained until the 6th iteration, where the central point no longer changes and no data moves between clusters. The results of the clusters that were formed in the 6th iteration consisted of: members of cluster 1 as many as 119 books, cluster2 as many as 8 books, and cluster3 as many as 21 books.

From the borrowing data, the data contained in cluster 3 is the book group with the highest loan amount among the other 2 clusters. In addition, books contained in cluster 3 have the fewest copies. From the results of this data analysis, it can be seen that the book titles contained in cluster 3 are the most recommended

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

books to be added to the library. So, this information can be used by the library in the selection of books that should be added to the library. The members of the cluster results can be seen in table 8.

TABLE 8
MEMBERS OF THE CLUSTER RESULT C3

Id	Total Loans	Copies	Cluster
Book6	8	1	C3
Book27	9	2	C3
Book31	9	1	C3
Book34	6	1	C3
Book44	11	3	C3
Book64	10	1	C3
Book66	6	1	C3
Book68	7	4	C3
Book71	6	1	C3
Book74	10	1	C3
Book80	8	1	C3
Book83	6	1	C3
Book89	9	1	C3
Book109	21	6	C3
Book110	8	1	C3
Book112	6	1	C3
Book125	7	1	C3
Book131	10	1	C3
Book132	6	1	C3
Book141	8	1	C3
Book144	7	3	C3

Testing the use of the K-Means Clustering Method was carried out using Orange software to speed up the data processing. The stages carried out in the Orange software can be seen in Figure 1.

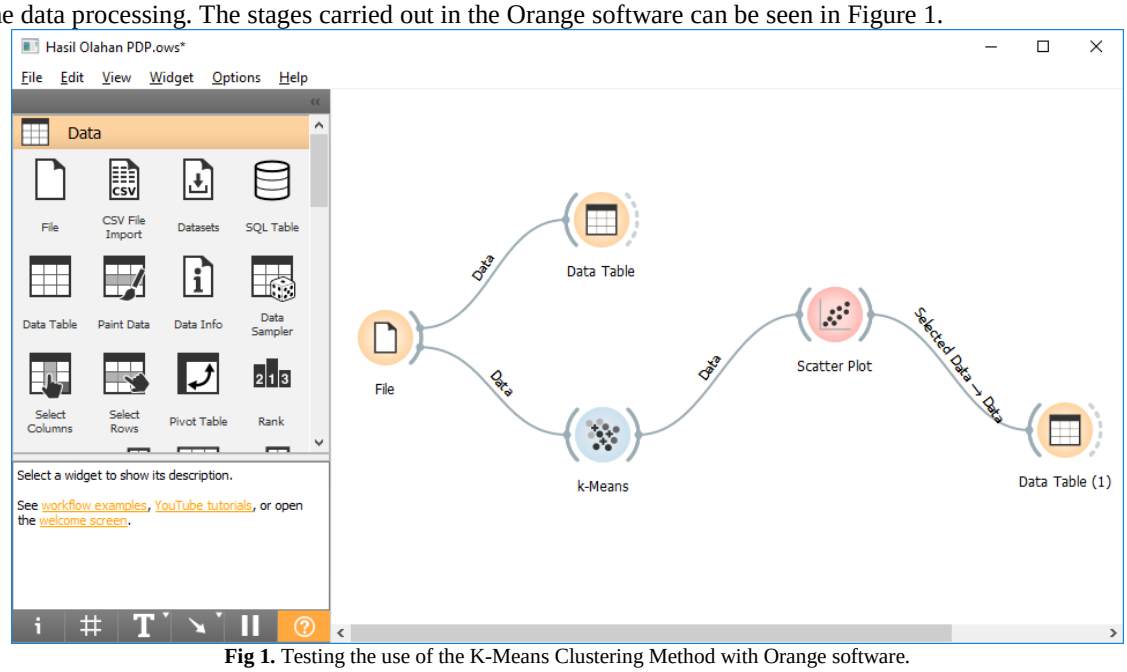

Fig 1. Testing the use of the K-Means Clustering Method with Orange software.

The results of the clusters formed from book lending data can be seen in Figure 2.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

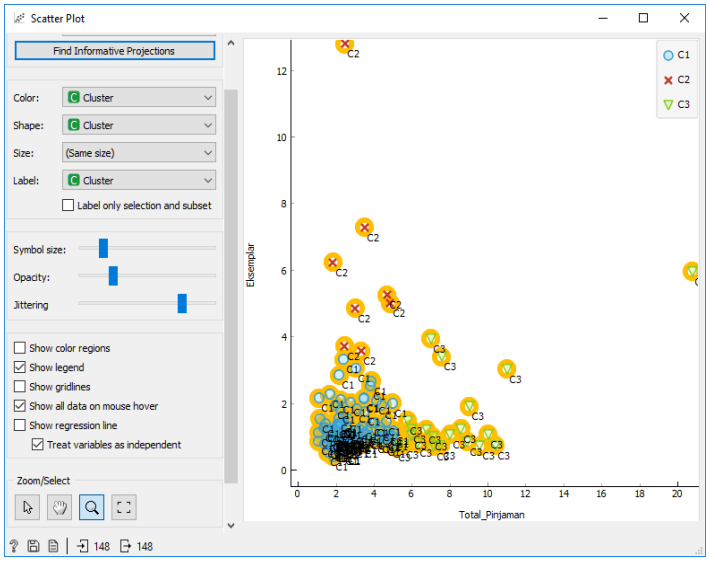

Fig 2. Clustering Results

4. Conclusion

The use of data mining applications using the K-Means Clustering method can be used to group data on books that are most often borrowed, books that are frequently borrowed and books that are rarely borrowed from the library, so that this information can be used by library management to determine which books should be added. in the library, and can also be used to sort books that are rarely borrowed in order to minimize the availability of bookshelf space. The results of the clustering process can also be used by the campus to determine restrictions on the types of books and book titles that will be donated to the library.

Acknowledgments

The authors would like to thank Politeknik Bisnis Indonesia and the Ministry of Research, Technology and Higher Education of Indonesia for supporting this research.

5. References

[1] A. Ardian, I. Purnama, and V. Sihombing, "DESIGN OF ANDROID-BASED STUDENT DATA PROCESSING APPLICATION (CASE STUDY: MIS NURUL HUDA LABUHAN BATU SELATAN)," *J. Servant. Kpd. Masy. IKA BINA EN PABOLO*, vol. 1, no. 1, pp. 40–53, 2019.

[2] VMM Siregar, H. Sugara, and IM Siregar, "DESIGN OF GOODS DATA COLLECTION INFORMATION SYSTEM IN PT. ATTACK HULU, "*J. Comput. Business*, vol. 12, no. 2, pp. 111–117, 2018.

[3] PDP Adi and A. Kitagawa, "Quality of Service and power consumption optimization on the IEEE 802.15.4 pulse sensor node based on Internet of Things," *Int. J. Adv. Comput. Sci. Appl.*, 2019.

[4] VMM Siregar, "Website Designing as a Media for Promotion and Product Sales," *TAM (Technology Accept. Model.*, vol. 9, no. 1, pp. 15–21, 2018.

[5] VMM Siregar, H. Sugara, and GA Purba, "Web-Based Application of Our Sales Report Recording --Kita.Net," *TEKINKOM*, vol. 2, no. 1, pp. 80–86, 2019.

[6] VMM Siregar, "Information System for Purchasing and Selling Clothing at Galoenk Pematangsiantar Distro," *JurTI (Jurnal Teknol. Information)*, vol. 1, no. 2, pp. 219–227, 2017.

[7] HA Simbolon and VMM Siregar, "Designing E-Commerce Based Information Systems to Increase Sales of Jersey Sports Products," *J. Tech. Inf. and Komput.*, vol. 1, no. 2, pp. 49–54, 2018.

[8] PDP Adi and A. Kitagawa, "ZigBee Radio Frequency (RF) performance on Raspberry Pi 3 for Internet of Things (IoT) based blood pressure sensors monitoring," *Int. J. Adv. Comput. Sci. Appl.*, 2019.

[9] M. Niswar *et al.*, "Performance evaluation of ZigBee-based wireless sensor network for monitoring patients' pulse status," in Proceedings - 2013 International Conference on Information Technology and Electrical Engineering: "Intelligent and Green Technologies for Sustainable Development", ICITEE 2013, 2013.

[10] PDP Adi and A. Kitagawa, "A Study of LoRa Performance in Monitoring of Patient's SPO2 and Heart Rate based IoT," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 2, 2020.

[11] W. Purba, S. Tamba, and J. Saragih, "The effect of k-means data mining is clustering toward the student profile model of drop out potential," *J. Phys. Conf. Ser.*, vol. 1007, no. 1, p. 12049, 2018.

[12] VM Mulia Siregar and H. Sugara, "Implementation of artificial neural networks to assess the lecturer's performance," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 420, no. 1, p. 012112, Oct. 2018.

[13] VMM Siregar, "Decision Support System for Determining Employee Monthly Incentives Using the Naïve Bayes

Method," *SYSTEMATION*, vol. 7, no. 2, pp. 87–94, 2018.
[14] VM Siregar and H. Sugara, "Design and Implementation of Desktop-Based Payroll Applications in Murni Sadar English Course," *J. Tech. Inf. and Komput.*, vol. 1, no. 2, pp. 42–48, 2018.
[15] SS S, AT Purba, and FOI Pardede, "GROUPING THE NUMBER OF POPULATION BY AGE CATEGORY WITH K-MEANS METHOD," *TEKINKOM*, vol. 2, no. 2, pp. 166–172, 2019.
[16] SP Tamba, MD Batubara, W. Purba, M. Sihombing, VM Mulia Siregar, and J. Banjarnahor, "Book data grouping in libraries using the k-means clustering method," *J. Phys. Conf. Ser.*, vol. 1230, no. 1, p. 012074, Jul. 2019.