JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering

Volume 9, Issue 2, October 2022, pp. 183-191
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3327

Implementation of Data Mining Clustering Using the K-Medoids Method in Grouping Library Books Politeknik Negeri Balikpapan

Maria Ulfah[1], Andi Sri Irtawaty[2]
[1,2] Electrical Engineering Department, Balikpapan State Polytechnic, Balikpapan, Indonesia

ARTICLE INFO

Article historys:

Received : 23/09/2022
Revised : 06/10/2022
Accepted : 11/10/2022

Keywords:

Clustering, Data Mining, K-Medoids, Rapid Miner

ABSTRACT

In order to procure the book according to the needs of users, Politeknik Negeri Balikpapan's Library needs information on the collection of books. To answer this problem, it is necessary to have a clustering system for existing books in the library by looking at the aspect of borrowing frequency. The Clustering system is made using the K-Medoids method with the selection of 3 Clusters, namely, very attractive, desirable and less desirable. From the results of data processing through the Rapid Miner Application with K = 3 , the results obtained are cluster_0 (low) consisting of 97 book titles with the frequency of borrowing books in the rare category in other words less desirable to borrow, cluster_1 (high) consists of 2 book titles which are the most in-demand group of books namely Teknologi Bahan dan Teori dan Praktik Hotel Front Office. Cluster_2 (medium) consisting of 8 titles which are books with moderate borrowing frequency. The results of the performance obtained Davies Bouldin index value of 0.287. The results of grouping the data of these books can be used as input for library managers in procuring book collections based on the frequency of borrowing books.

Corresponding Author:

Maria Ulfah
Electrical Engineering Dept., Balikpapan State Polytechnic, Balikpapan. Indonesia
Email: maria.ulfah@poltekba.ac.id

1. INTRODUCTION

The existence of the library cannot be separated from human culture. The high and low civilization of a nation can be seen from the condition of the library it has. In essence, the library is a cultural product in the form of an institution that collects, stores, organizes both printed and recorded works as a source of information and learning from generation to generation. In Indonesia there are five types of libraries and these five types of libraries are national libraries, public libraries, special libraries and university libraries.

The college library collections are held through a selection that refers to the needs of study programs that are organized and organized in such a way as to ensure the effectiveness and efficiency of services to the needs of the academic community. Even in every academic community, their needs cannot be equated because they have different needs in the literature. The literature used by students in each department is also different. With the organized grouping of each user in this case is a borrower from various majors, it can be seen what literature groups are most often borrowed by students.[1],[2],

Politeknik Negeri Balikpapan's Library aims to support information needs, the condition of library reading interest can be seen from the percentage of visitors, borrowers and borrowed books. Book borrowing data is used as a report. The library every year procures new book collections to be

JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 9, Issue 2, October 2022, pp. 183-191
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3327

reproduced so that they must know the types of priority book collections to be reproduced. So it is necessary to analyze the data processing of book lending transactions that explain the distribution of the intensity of book lending, the processing information helps to add to the collection.

To find out which books are the most popular, a cluster technique is used using the K-Medoids method. The K-Medoids algorithm is an iterative clustering algorithm that partitions the data set into a number of K Clusters that have been defined at the beginning. Calculations on K-Medoidss do not refer to the average value (mean) of all data on each cluster, the goal is to reduce outliers or the sensitivity of the cluster that generated on the dataset [3], [4]. The K-Means and K-Medoids algorithms both work by minimizing the distance between data points that are labeled to be in one cluster and the specified data points. as the center of the cluster. The K-Medoids algorithm chooses a data point as the center of the cluster (medoids). done randomly at the data points of each cluster not based on the average value [4].

Data mining is a term used to describe the discovery of knowledge in databases. Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and assembled knowledge from various large databases [5].

Data mining is a series of processes to explore added value from a data set in the form of knowledge that has not been known manually [4]. Where the results of the excavation process will form patterns from the data set, which is often referred to as pattern recognition. Pattern recognition is part of data mining. Data mining is often also called Knowledge Discovery in Database (KDD), which is an activity that includes collecting, using historical data to find regularities, patterns or relationships in large data sets [6]. There have been many studies that have conducted clustering using the K-Medoid algorithm [7],[8],[9],[10]

Therefore, the authors in this study took the title "Implementation of Data Mining Clustering Using the K-Medoids Method in Clustering Politeknik Negeri Balikpapan's Library Books" to help classify loan book data of interest in supporting the Library Manager in procuring book collections.

2. RESEARCH METHOD

The following are the methods of carrying out the research:

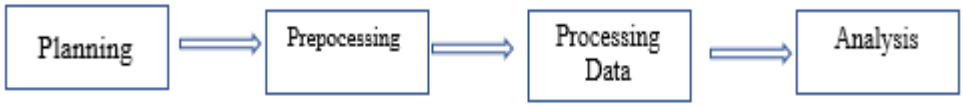

Figure 1. Research Method

2.1. Planning Stage

The first step in this research is to do research planning. There are four activities in planning, namely determining research objectives, identifying problems, determining problem boundaries and studying literature.

In the data collection stage, it is carried out in two ways, namely:

1. Library study. In this activity, researchers search for and study books, ebooks and journals to support writing related to the topics raised. From the literature study, the theories and methods in solving research problems related to the topics raised are obtained.
2. Observations and interviews, From the results of interviews obtained interview data which will later be used for report writing and from observations obtained documentation data and library book borrowing data.

2.2 Preprocessing Stage

In the preprocessing stage, the activities carried out are retyping the data obtained into Microsoft Excel to record all book lending transactions. After the data is copied at this stage, data cleaning will also be carried out, namely deleting data that is not clearly written or data that cannot be read. This cleaning process is carried out in order to get the correct calculation results.

JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 9, Issue 2, October 2022, pp. 183-191
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3327

2.3. Data Processing Stage

Processing data with rapidminer software and also processing data based on problem identification in research. Using the K-Means method, and through the application of the k-means clustering algorithm, it is hoped that in processing this data, it is expected to get good results for grouping books based on the frequency of borrowing. learning outcomes and produce new knowledge. Processing data with Rapidminer software.

2.4. Analysis Stage

After all the data has been collected, then the analysis stage is carried out.

3. RESULTS AND DISCUSSION

3.1. Data Set

The dataset used is data on borrowing books at the Politeknik Negeri Balikpapan's Library from the period 2019-June 2022. The number of types of books recorded is 107 titles with a total borrowing frequency of 231, shown in Table 1 which contains records of book titles and the frequency of borrowing each year.

Table 1. Data set

Number	Book Title	2019	2020	2021	2022	Total
1	101 Amazing Public Relation Ideas				1	1
2	9 Days Umratan	1				1
3	Akuntansi Dasar 1 Dan 2				2	2
4	Akuntansi Manajerial Suatu Orientasi Praktis				2	2
5	Akuntansi Perbankan Syariah				2	2
6	Akuntansi Perbankan Syariah Teori Dan Praktik Kontemporer Berdasarkan PAPSI 2013 Edisi Kedua			2	1	3
7	Akuntansi Syariah				1	1
8	Akuntasi Perusahaan Manufaktur				2	2
9	Algoritma & Pemrograman Dengan C++ Edisi Kedua		1		1	2
10	All About Corporate Valuation				1	1
11	Anak Rantau		1			1
12	Analisa Laporan Keuangan Edisi Keempat		2		4	6
13	Analisis Kritis Atas Laporan Keuangan				7	7
14	Aneka Kue		1			1
15	Aplikasi Logika Fuzzy Edisi Kedua				1	1
16	Arduino Belajar Cepat Dan Pemrograman		1		2	3
17	Aroma Karsa		1			1
18	Auditing Buku 1 Edisi Keenam				5	5
19	Bahan Produk Bakery		1			1
20	Belajar Jaringan Komputer Berbasis Mikrotik OS	1				1
21	Budidaya Unggul Lele Phyton			1		1
22	Build Up Your English Reading Skill				1	1
23	Catatan Harian Anne Frank		1			1
24	Dahsyatnya Bisnis Hotel Di Indonesia				1	1
25	Dasar - Dasar Pengolahan Makanan	1				1
26	Dasar - Dasar Perbankan Edisi Revisi 2014				6	6
27	Drainase Perkotaan				3	3
28	Ekonomi Manajerial Dengan Pendekatan Matematis Edisi Revisi		1			1
29	Ekonomi Moneter				1	1
30	Emotional Intelligence	1	1			2
31	Ensiklopedia Mini - Hotel				2	2
32	Food And Beverage Service Operational				2	2
33	Front Office Management				5	5
34	Hidrologi				1	1

JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 9, Issue 2, October 2022, pp. 183-191
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3327

Number	Book Title	2019	2020	2021	2022	Total
35	Hidrologi Jilid 1				5	5
36	Hidrologi untuk Pengairan				1	1
37	Hotel Courtesy				1	1
38	Housekeeping Hotel				1	1
39	Housekeeping Hotel Edisi Kedua				2	2
40	Hujan		2			2
41	Investasi Dan Pasar Modal Indonesia				3	3
42	Jurus Sukses Beternak Lele Sangkuriang			1		1
43	Kamus Populer Istilah Komputer dan Informatika		2			2
44	Khasiat Buah Dan Sayur				1	1
45	Kitab Khasiat Buah Dan Sayur				1	1
46	Kitab Kue Super Yummy		1			1
47	Konfigurasi Wireless Routerboard Mikrotik	1				1
48	Lintang				1	1
49	Manajemen Divisi Kamar				2	2
50	Manajemen Keuangan (Finance Management)				2	2
51	Manajemen Keuangan Modern				2	2
52	Manajemen Keuangan Sebagai Dasar Pengambilan Keputusan Bisnis				2	2
53	Manajemen Keuangan Teori Konsep & Aplikasi Edisi Revisi				2	2
54	Manajemen Kinerja Edisi Kelima				1	1
55	Manajemen Pemasaran Bank				5	5
56	Manajemen Pemasaran Jilid 1 Edisi 12				4	4
57	Manajemen Penyelenggaraan Hotel				2	2
58	Manajemen Perkreditan				1	1
59	Manajemen Risiko				2	2
60	Mengelola Kualitas Layanan Perbankan (Cover Baru)				3	3
61	Metode Penelitian Kuantitatif				2	2
62	Metode Penelitian Pendidikan - Pendekatan Kuantitatif, Kualitatif Dan R&D				4	4
63	Metodologi Penelitian Bisnis & Ekonomi				6	6
64	Metodologi Penelitian Bisnis Untuk Akuntansi & Manajemen Edisi Pertama				4	4
65	Metodologi Penelitian Dan Statistika				3	3
66	Mikrotik Kung Fu Kitab 1		1			1
67	Mikrotik Untuk Pemula		1			1
68	Modul Pemrograman Web (HTML, PHP, & MySWL) Edisi Ketiga				1	1
69	Napoleon Hill's Keys To Success				1	1
70	Nikmatnya Bangun Pagi, Tahajud, Subuh dan Dhuha				1	1
71	Panduan Praktis Dasar Analisa Laporan Keuangan				4	4
72	Paradigma, Metodologi & Aplikasi Ekonomi Syari'ah				1	1
73	Pedoman Pengerjaan Beton				3	3
74	Penerapan Soft Computing Dengan Matlab Edisi Revisi				1	1
75	Pengantar Akuntansi Edisi IFRS				1	1
76	Pengantar Akuntansi Lengkap Dengan Kumpulan Soal Dan Solusinya				1	1
77	Pengantar Bisnis Respon Terhadap Dinamika Global Ed.2				1	1
78	Pengantar Falsafah Ekonomi Dan Keuangan Syariah				1	1
79	Pengantar Ilmu Bisnis				4	4
80	Pengantar Metodologi Penelitian				1	1

Volume 9, Issue 2, October 2022, pp. 183-191
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3327

Number	Book Title	2019	2020	2021	2022	Total
81	Pengantar Statistika Sosial				3	3
82	Perempuan Kedua				1	1
83	Perencanaan dan Pemeliharaan Sistam Plambing				1	1
84	Perencanaan Struktur Baja Dengan Metode LRFD			1		1
85	Perilaku Konsumen Di Era Internet Implikasinya Pada Strategi Pemasaran				2	2
86	Praktikum Akuntansi Keuangan Manual Kasus Perusahaan Dagang Ed. Revisi				2	2
87	Rahasia Sukses Bisnis Dan Budi Daya Lele Unggul			1		1
88	Ranah 3 Warna		1			1
89	Rancang Bangun 3D dengan AutoCad 2012				1	1
90	Rangkaian Listrik		1			1
91	Santri Cangkir		2			2
92	Smart Traders Not Gamblers				3	3
93	Soal & Jawab Ekonomika Untuk Manajer		1			1
94	Statistik Deskriptif Untuk Ekonomi				2	2
95	Statistika Ekonomi & Bisnis				1	1
96	Sukses Beternak Lele Dumbo & Lele Lokal			1		1
97	Teknik Dan Prosedur Divisi Kamar Pada Bidang Hotel				3	3
98	Teknik Perhitungan Debit Rencana Bangunan Air				2	2
99	Teknologi Beton				14	14
100	Teknologi Perbankan				2	2
101	Teori Dan Praktik Hotel Front Office				13	13
102	The Greats On Leadership			1		1
103	The Maxwell Daily Reader	1		1		2
104	Ubur - Ubur Lembur				1	1
105	UMKM Aspek Hukum Dan Manajemen Pemasaran Produk				1	1
106	Who Am I				1	1
107	Who Moved My Cheese		2			2

3.2. Data Processing

At this stage, using RapidMiner tools with the K-Means method. The New Process view is to prepare a worksheet on RapidMiner. Import the data to be tested in .xls or .xlsx format. Next is the model for importing Microsoft Excel files. To enter data to be executed, you need to right click -> Insert Operator -> Data Access -> File -> Read Ecxel as shown in Figure 2

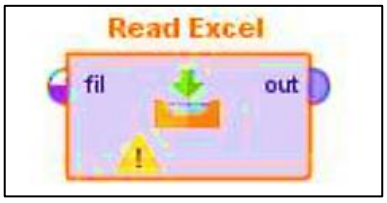

Figure 2. Import data

3.3. Modeling and Evaluation

At this stage, add the K-Medoids operator. via right click -> Insert Operator -> Modeling -> Segmentation -> K-Medoids.

JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 9, Issue 2, October 2022, pp. 183-191
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3327

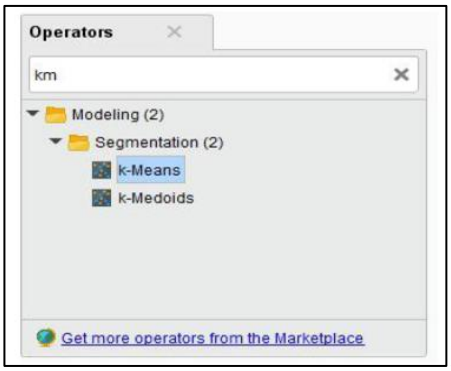

Figure 3. Addition of K-Medoids Operators

Next, work on the settings on the Parameters Clustering K-Medoids menu, setting the value of k, where k is the value that will be used to determine the number of clusters to be created. Here the number of clusters to be created is 3 clusters (low, medium, high) as shown in Figure 4.

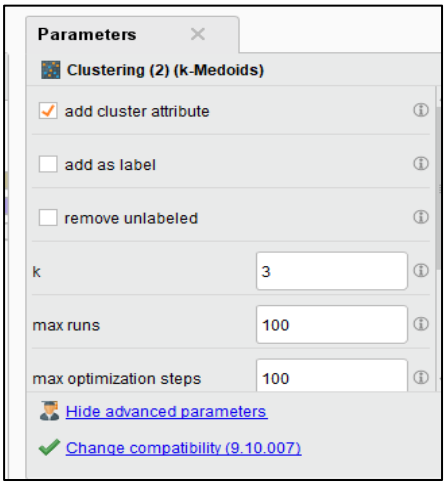

Figure 4. Number of clusters

The modeling form is as follows:

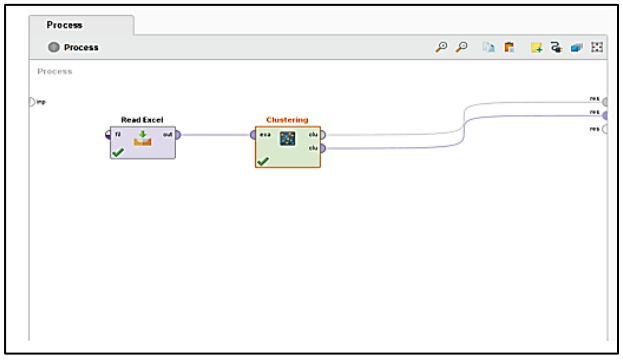

Figure 5. Clustering modeling with K-Medoids

In ExampleSet, you can see some of the cluster results contained in the Data View which is a display of the results of the data cluster that has been entered.

JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 9, Issue 2, October 2022, pp. 183-191
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3327

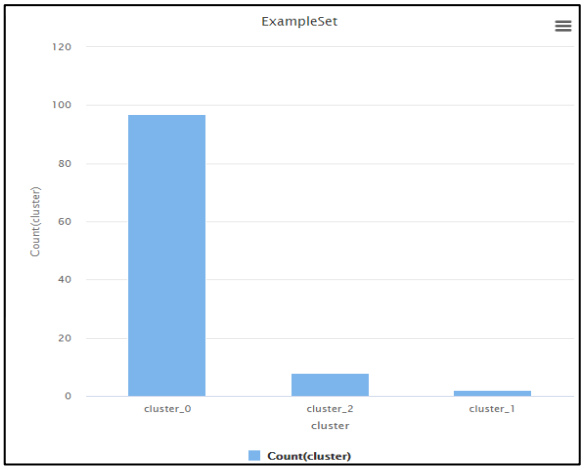

Figure 6. Data example set/cluster results in statistic

In Figure 6 the data results are broken down into 3 clusters, namely cluster 0, cluster 1, cluster 2 with each cluster pocketing the results of clustering data cluster_0 consisting of 97 book titles, cluster_1 consisting of 2 book titles, cluster_2 consisting of 8 book titles.

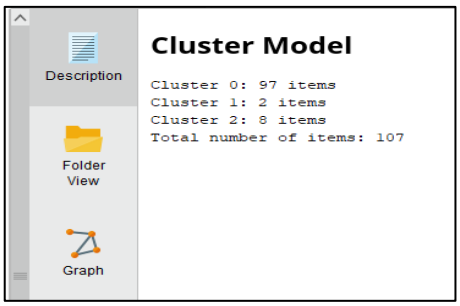

Figure 7. Cluster Model

The following are members of Cluster 0

Figure 8. Members of cluster 0

The following are members of Cluster 1

Figure 9. Members of cluster 1

JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 9, Issue 2, October 2022, pp. 183-191
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3327

The following are members of Cluster 2

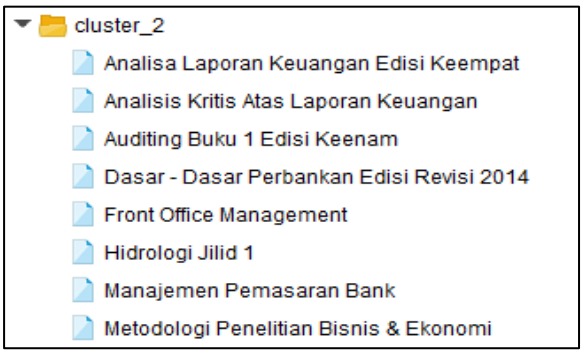

Figure 10. Members of Cluster 2

From the results of clustering with K = 3, it can be seen that in Cluster 1 (high) is the most popular group of books consisting of 2 titles, namely Teknologi Beton, Teori dan Praktik Front Office Hotel in the period 2019 to June 2022 as much as 14 and 13 times the frequency of borrowing.

Cluster 2 (medium) consists of 8 titles, which are books with a moderate frequency of borrowing, in other words, they are categorized as desirable for borrowing. While Cluster 0 (low) consists of 97 titles with the frequency of borrowing books in the rare category, in other words, they are less desirable to borrow.

3.4. Perfomance

For clustering performance with K=3 using the Davies Bouldin Index (DBI) value. Evaluation using the Davies Bouldin Index has an evaluation scheme from an internal cluster, where the good or bad results of the cluster are seen from the quantity and proximity of the data from the cluster results. From the performance results, the Davies Bouldin index value is 0.287 as shown below.

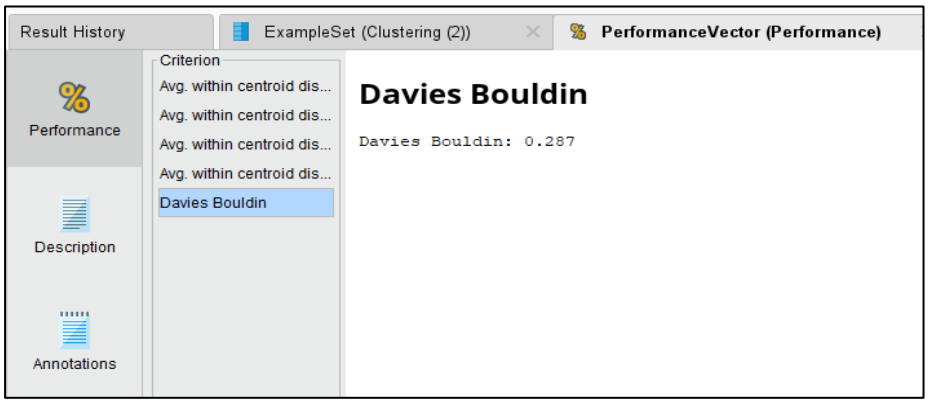

Figure 11. Davies Bouldin

4. CONCLUSION

From the results of data processing through the Rapid Miner Application with K=3, the results obtained are cluster_0 (low) consisting of 8 book titles with the frequency of borrowing books in the rare category in other words less desirable to borrow, cluster_1 (high) consists of 2 book titles which are books with moderate borrowing frequency, in other words, are categorized as desirable for borrowing, cluster_2 (high) consisting of 2 book titles, is the most popular group of books consisting of 2 titles, namely Teknologi Beton and Teori dan Praktik Hotel Front Office. The results of the performance obtained Davies Bouldin index value of 0.287

JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 9, Issue 2, October 2022, pp. 183-191
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3327

REFERENCES

[1] Intan fitri andyni, "Grouping of book borrowers using the k-means method at the Central Library of Upn veterans", East Java, 2013.

[2] D. Dwinavinta, "Klasterisasi judul buku dengan Metode K-Means". Universitas Islam Indonesia, 2014.

[3] Y. Hilda, "Perbandiangan K-Means dan k-Medoidss Clustering terhadap Kelayakan puskesmas di DIY Tahun 2015". Universitas Islam Indonesia, 2016.

[4] V. Aditya, "Perbandingan Algoritma K-Means dan K-Medoids dalam Pengelompokan komoditas peternakan Jateng", 2015.

[5] Kusrini and Lutfi, ET, "Data Mining Algorithms", Yogyakarta: Andi Offset, 2009.

[6] Pramudiono, What is Data Mining ?, [online], accessed on March 15, 2, 2006.

[7] M. Ainur, *et all,* "Pengelompokan Kategori Buku Berdasarkan Judul Menggunakan Algoritma Agglomerative Hierarchical Clustering Dan K-Medoids", JINACS Vol 02 No.03.

[8] R. Kristini, "Implementasi Algoritma K-Medoids dalam Pengelompokan Mahasiswa yang Layak Mendapat Bantuan UKT", Insologi Vol 1 No 2, 2022.

[9] S. Atmiatun, *et all "*Penerapan Metode K-Medoids Untuk Pengelompokan Jalan di Kota Semarang", Jurnal Teknik Informatika dan Sistem informasi, Vol 6 No 2 2020.

[10] A. Supriyadi. *et all,* "Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas", JIPI Vol 6 No 2 2021