INCREASE SALES FOR THE BEST SEED CLUSTERING USING RAPID MINER

by - -

Submission date: 20-Jan-2024 07:10AM (UTC+0300)
Submission ID: 2274348143
File name: 850-Article_Text-2866-2-10-20230810.pdf (724.57K)
Word count: 2597
Character count: 13015

ICEBIT

The 4th International Conference On Economics , Business and Information Technology

June 10, 2023 pp. 906-915

INCREASE SALES FOR THE BEST SEED CLUSTERING USING RAPID MINER

Fauzan Azim
Institute Technology and Science Mandala
Jl. Sumatra No. 118 - 120
Jember, Indonesia
082385584973
fauzan@stie-mandala.ac.id

Yuli Hartati
Institute Technology and Social Sciences Khatulistiwa
Jl. Mayor Said Zamzam, Simpang Tiga Ophir, Kabupaten Pasaman Barat
082283073518
Yulihartati802.yh@gmail.com

Eko Afrianto
Institute Technology and Science Mandala
Jl. Sumatra No. 118 - 120
Jember, Indonesia
082329319788
eko.afrianto@itsm.ac.id

ABSTRACT

The superiority of seeds is one of the main factors for farmers. UD. Tiara Bersaudara is a shop that sells seeds and agricultural needs. Poor sales of seeds and fertilizers can hurt UD. Tiara Brothers. To maintain the stock of seeds that are in demand by farmers, sellers must be able to analyze data on the sale of seeds that are in great demand by farmers so as not to cause losses in business. This process is difficult for UD Tiara Brothers to do because it has a lot sales data. Existing problems can be solved by clustering seed sales data according to most farmers' interests. Clustering is the grouping of data into several clusters based on the level of data similarity. The purpose of the study was to group the best-selling seed data at UD. Tiara Brothers in increasing sales. The clustering method uses the K-Means algorithm by partitioning data into clusters based on the centroid closest to the data. Furthermore, testing by comparing the calculation results with RapidMiner studio software 9.7. The results of this study can later be used as a benchmark for decision support by UD. Tiara Brothers to set marketing strategies for increasing sales.

Keywords: Clustering, Data mining, RapidMiner, K-Means

1. INTRODUCTION

Trading Business (UD) is a form of business whose main activity is to buy and resell goods or services to make a profit. One form of a trading business is UD. Tiara Brothers are engaged in the sale of seeds and agricultural needs. In the process of selling UD. Tiara Brothers often experience obstacles, especially in managing seed supplies for farmers' needs. This obstacle causes a frequent shortage of seed stocks that are in demand because sales are very high. So there is a buildup of seeds that are not in demand in warehouses because sales are low. In this study, seed sales transaction data selection is used as an object for analysis. In sales transaction data, facts are stored that can be extracted and processed to become useful information for trading businesses.

Data mining is a technology that automates the process of finding interesting and sensitive patterns from large sets of data. This allows human understanding to be able to find patterns and scalability of techniques. Data mining techniques are used to perform descriptive mining (describing a grouping of common traits) or predictive mining (trying to predict based on data classification conclusions) on large volumes of data (Siregar, 2018). Data mining is often used in the form of wider applications and can be applied in various areas of life such as business, industry, science, research, and other fields (S, 2019).

Clustering is a technique of grouping several data on a large set of data based on certain criteria. The clustering results are given to the end user to be able to give an idea of what is happening to the database (Ali, 2019). The cluster technique has two methods of grouping data, namely hierarchical clustering and non-hierarchical clustering. Hierarchical clustering is a method of grouping data by grouping two or more data that have similarities or similarities, then forwarding to other objects that have close data, the process will last until a cluster forms a kind of tree where levels or clear hierarchy between objects from the least similar to the most similar (Alkhairi & Windarto, 2019).

Grouping data using the K-Means algorithm is very simple, at the initial stage determine in advance the number of data groups to be determined. After the data grouping is determined, select the document that will be used as a centroid point cluster. Next, iterate until there is stability for all groups of objects that have converged (Yunita, 2018).

In this study, seed sales data will be grouped by calculating the closest distance between the data and the midpoint (centroid) of a cluster. So that the grouping can produce several seed clusters, namely best-selling and non-best-selling seeds. The application of the K-Means Clustering algorithm is expected to help UD. Tiara Brothers in increasing sales.

The application of the K-Means algorithm was once used to group sales data at one of the outdoor equipment retail stores by producing three clusters. The grouping of data is used to improve stock management and sales strategies (Liu et al., 2018). The application of the K-Means algorithm has also been used as a form of application of pre-consumption monthly load data analysis from electricity consumers from certain domestic areas in China (Bhargava, 2019)

2. RESEARCH METHODOLOGY

In this study, the research method used can be seen in Figure 1. The process will be described starting from data collecti and literature study until the research objectives a chieved and by the initial plan that has been determined. The purpose of this study is to analyze and group data using the K-Means algorithm to design strategies to increase sales.

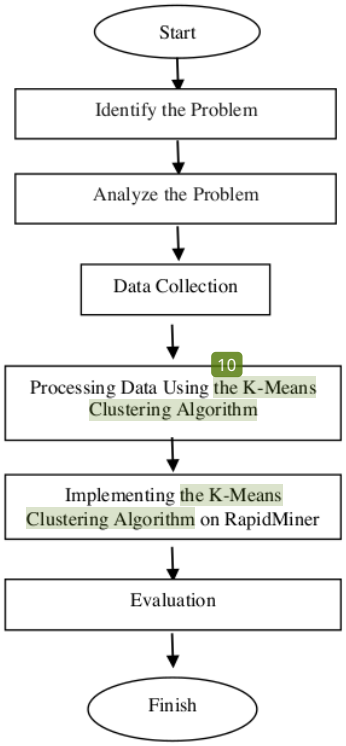


Figure 1 Method Research

2.1. Identify The Problem

Identifying problems is the initial stage in determining and formulating problems that exist in an object of research. The purpose of determining the problem formulation is so that the research carried out is directed and does not go out of the limits of the problem to be studied. This stage is the first step of a study

2.2. Analyze The Problem

Analyzing problems is a stage to understand problems in the object of research with a predetermined scope so that the results of the research can be as expected.

2.3. Data Collection

Collect the data needed to conduct research. The data collection process in this study is as follows:

a. Observation

Observation is the direct observation made on the object of research. The purpose of observation is so that the existing problems can be known.

b. Interview

The interview is a data collection technique carried out with direct questions and answers between the researcher collector and the resource person. In this study, researchers will conduct interviews with UD. Tiara Brothers. The purpose of the interview is to obtain information related to the existing problem.

c. Literature Study

A literature study is analyzing data and information by reading books, literature, journals, and articles that support research.

2.4. Processing Data Using The K-Means Clustering Algorithm

After all, the data is collected, the data will be analyzed and processed using the K-Means Clustering algorithm. The data to be processed at this stage is seed sales data from January-April 2022 at UD. Tiara Brothers. Data management is done by following the flow chart of the K-Means Algorithm process as seen in Figure 2:

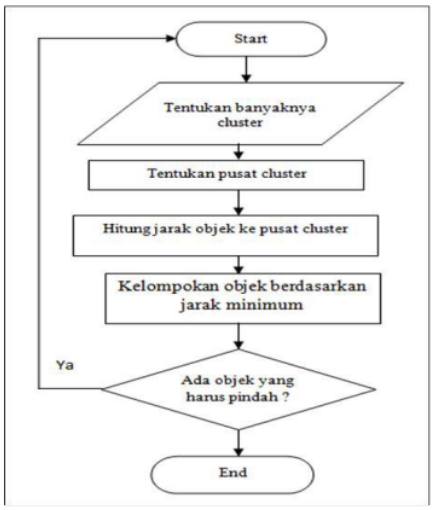


Figure 2 K-Means Clustering Algorithm Flow Diagram

From the flow chart, you can describe the steps of data processing with the K-Means algorithm

a. Determine the data to be in the cluster
   The data to be processed at this stage is seed sales data from January-April 2022 at UD. Tiara Bersaudara.
b. Determine the number of clusters (k) to be formed to perform K-Means calculations
c. Determine the center of the cluster (centroid) randomly.
d. Calculates the distance of each data to the center of the cluster by using the Euclidean formula (1).

$$D(j,k) = \sqrt{(J1 - K1)^2 + (Ji - K1)^2} \quad \text{..... (1)}$$

   Where:

   D (J, K) = Data distance to **j** to cluster center **k**

   Ji = Data Record

   Yi = Data Centroid

e. Grouping data into clusters with the shortest distance to the center of the cluster.
f. Calculate the new cluster center by using formula (2) :

$$CI = \frac{J1 + J2 + J3 + \ldots + Jn}{\sum J} \quad \text{..... (2)}$$

   Where:
   CI = *Centroid* data
   J1 = Data to "i" on attribute data to "i"
   ∑J = Amount of data

g. If the calculation of the new cluster center has been completed, then some data moves from the results of the previous iteration and repeat steps 2 to 4 until no more data is moving to another cluster.

2.5. Implementing The K-Means Clustering Algorithm On Rapidminer

After the manual calculation process using K-Mean is implemented. The next step is to apply the manual calculation using RapidMiner software. Here's an overview of the calculation steps using the Rapid Miner application:

2.6. Evaluation

The evaluation is carried out to see whether the results of manual calculations with the system testing process that has been made are by user requests. The purpose of the evaluation is to see to what extent the system is functioning and to identify problems that occur

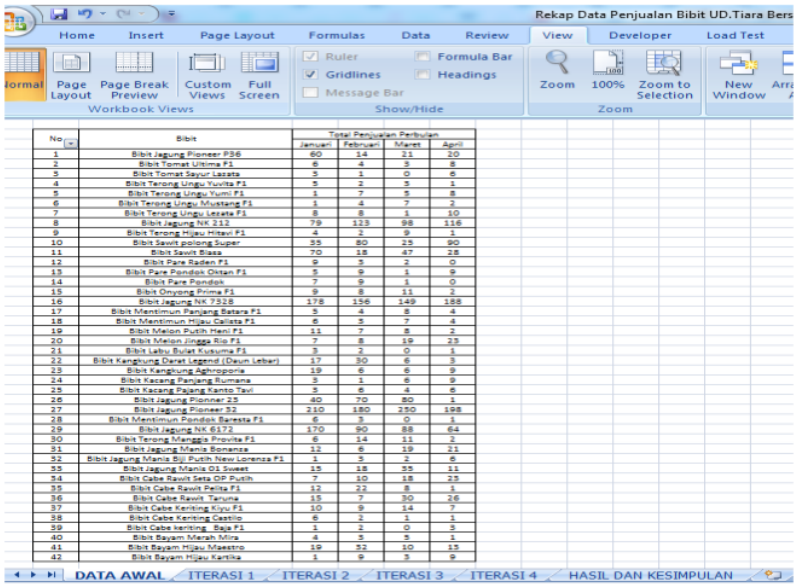
3. RESULT AND DISCUSSION

In this discussion, several steps were taken to analyze the data with software that aims to determine the best seed sales data clustering at UD. Tiara Brothers using Rapid Miner software.

3.1 Data Collection

After the manual calculation process using K-Mean is implemented. The next step is to apply the manual calculation using RapidMiner software. Here's an overview of the calculation steps using the Rapid Miner application.

1. The display of sales data to be tested can be seen in Figure 3:



No	Bibit	Total Penjualan Perbulan			
		Januari	Februari	Maret	April
1	Bibit Jagung Pioneer P36	60	14	31	20
2	Bibit Tomat Ultima F1	6	4	3	8
3	Bibit Tomat Sayur Lezata	3	1	0	6
4	Bibit Terong Ungu Yuvita F1	5	2	5	1
5	Bibit Terong Ungu Yumi F1	1	7	5	8
6	Bibit Terong Ungu Mustang F1	1	4	7	2
7	Bibit Terong Ungu Lezata F1	8	8	1	10
8	Bibit Jagung NK 212	79	123	98	116
9	Bibit Terong Hijau Hitavi F1	4	2	9	1
10	Bibit Sawit polong Super	35	80	25	90
11	Bibit Sawit Biasa	70	18	47	28
12	Bibit Pare Raden F1	9	3	2	0
13	Bibit Pare Pondok Oktan F1	5	9	1	9
14	Bibit Pare Pondok	7	9	1	0
15	Bibit Onyong Prima F1	9	8	11	2
16	Bibit Jagung NK 7328	178	156	149	188
17	Bibit Mentimun Panjang Batara F1	5	4	8	4
18	Bibit Mentimun Hijau Calista F1	6	5	7	4
19	Bibit Melon Putih Heni F1	11	7	8	2
20	Bibit Melon Jingga Rio F1	7	8	19	23
21	Bibit Labu Bulat Kuzuma F1	3	2	0	1
22	Bibit Kangkung Daret Legend (Daun Lebar)	17	30	6	3
23	Bibit Kangkung Aghroponik	19	6	6	9
24	Bibit Kacang Panjang Rumana	3	1	6	9
25	Bibit Kacang Pajang Kanto Tavi	5	6	4	6
26	Bibit Jagung Pionner 23	40	70	80	1
27	Bibit Jagung Pioneer 32	210	180	250	198
28	Bibit Mentimun Pondok Baresta F1	6	3	0	1
29	Bibit Jagung NK 6172	170	90	88	64
30	Bibit Terong Manggis Provita F1	6	14	11	2
31	Bibit Jagung Mania Bonanza	12	6	19	21
32	Bibit Jagung Mania Biji Putih New Lorenza F1	1	3	2	6
33	Bibit Jagung Mania 01 Sweet	15	18	55	11
34	Bibit Cabe Rawit Seta OP Putih	7	10	18	23
35	Bibit Cabe Rawit Pelita F1	12	22	8	1
36	Bibit Cabe Rawit Taruna	15	7	30	26
37	Bibit Cabe Keriting Kiyu F1	10	9	14	7
38	Bibit Cabe Kerting Castilo	6	2	1	1
39	Bibit Cabe keriting Baja F1	1	2	0	3
40	Bibit Bayam Merah Mira	4	3	3	1
41	Bibit Bayam Hijau Maestro	19	52	10	15
42	Bibit Bayam Hijau Kartika	1	9	3	9

Figure 3 Prepared Sales Data

2. Import repository on Rapid Miner software

After the data in Microsoft Excel is prepared, the next step is to test the results by importing the repository into the RapidMiner software. Import repository in RapidMiner software can be done by selecting Menu import data as seen in Figure 4:

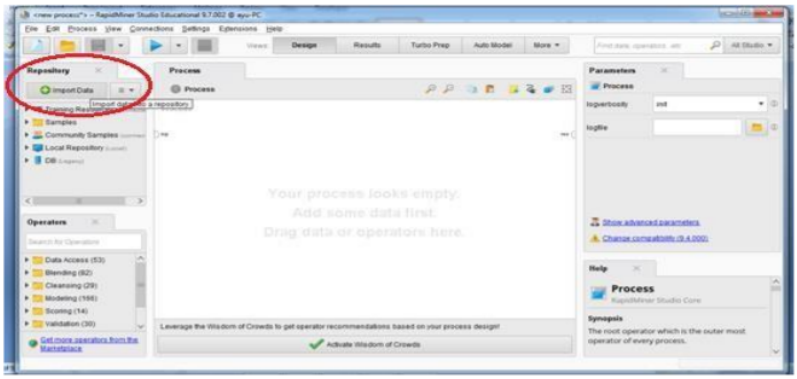


Figure 4 Import Data

3. Processing Import Data

After selecting the Import Data menu option, then select the data file that has been prepared previously as seen in Figure 5. The Sales Data file is taken From the" SALES DATA" folder with the file name UD Seed Sales Data Recap. Tiara Brothers

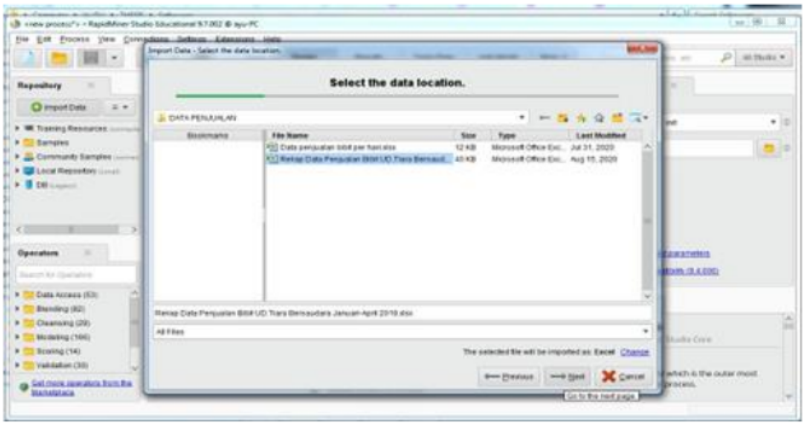


Figure 5 Import UD Seed Sales Data. Tiara Bersaudara

4. Select Data and Specify Id to Process

After the data import is carried out, the next process is to determine which cells will be used for testing the data that will be used in the Rapid Miner application. Please select the sheet, select INITIAL DATA, and Block cell range from (B6:G49) as seen in Figure 6 and Figure 7:

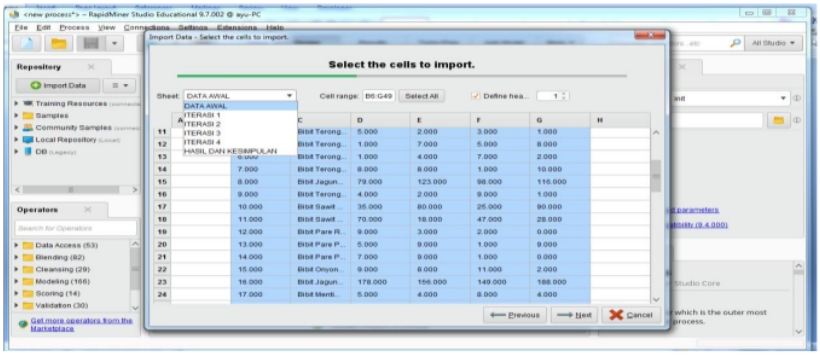


Figure 6 Select Data to Process

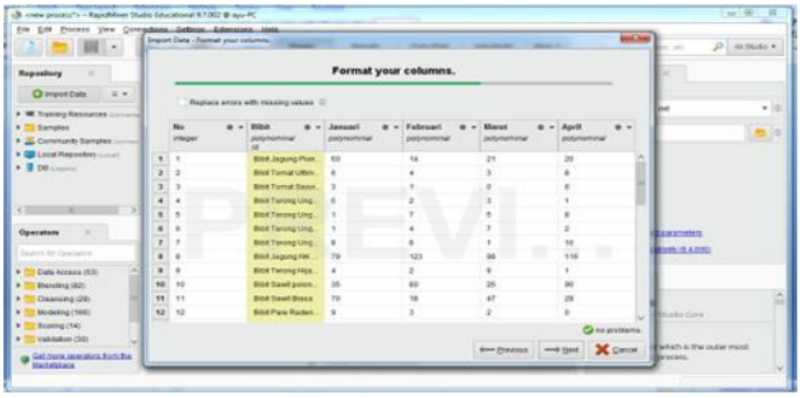


Figure 7 Specify ID Data

5. Save the Process to Repository

   Save data that has been implemented can be saved to the repository. The data storage process can be seen in Figure 8:

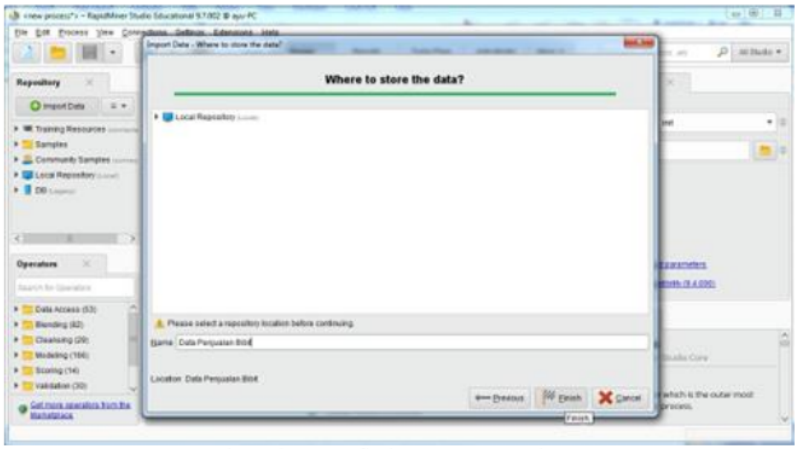


Figure 8 Process Saving Data to Repository

3.2 Data Testing Process with Rapid Miner studio 9.7

After the data storage process is carried out, the next process is testing the data that has been stored using the Rapid Miner application

1. Testing Data

   Testing data process is done by dragging and dropping data that has been stored in the repository to the process as shown in figure 9. Then on the operator interface select the operator K-Means then drag and drop it also to the process as seen in the picture 10.

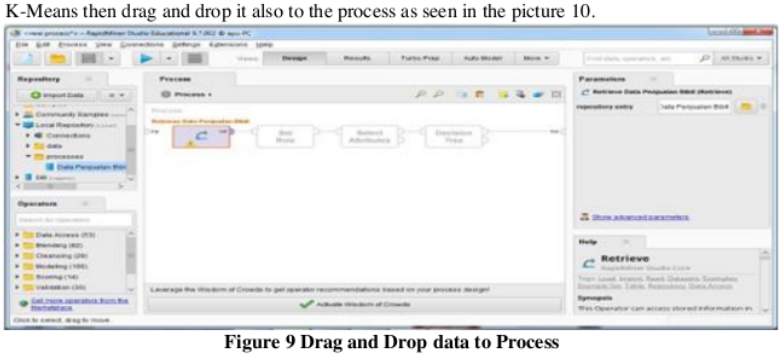


Figure 9 Drag and Drop data to Process

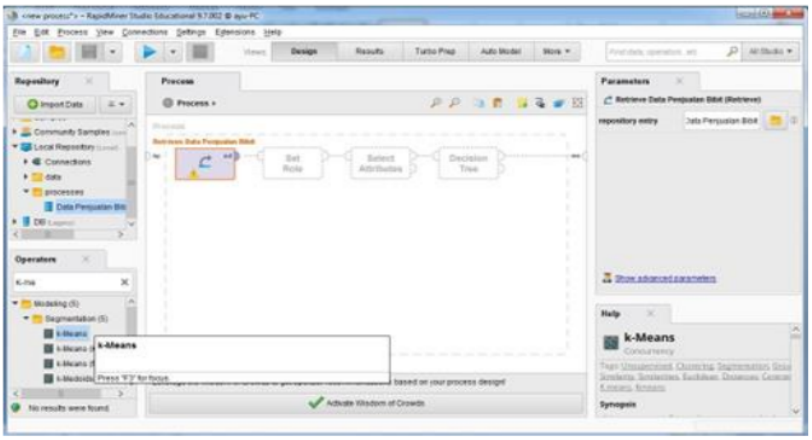


Figure 10 Specify Operators

2. Specify Parameter and Doing the Testing Process

After the operator determination process is carried out, the next process is to determine the parameters in the process to be carried out, parameter determination can be carried out by Double clicking the clustering operator then in the parameters section fill in the K value with the number "2" because we will form 2 clusters. Under measure types, select Numerical Measures. In numerical measure select Euclidean Distance as shown in figure 11. To perform the data grouping process with K-Means, click the star as shown in figure 12:

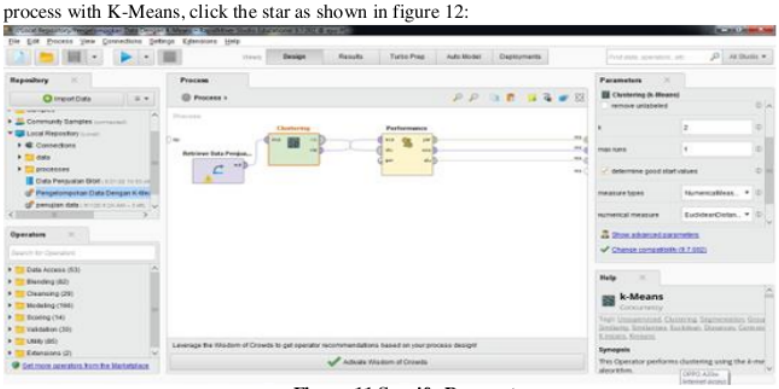


Figure 11 Specify Parameter

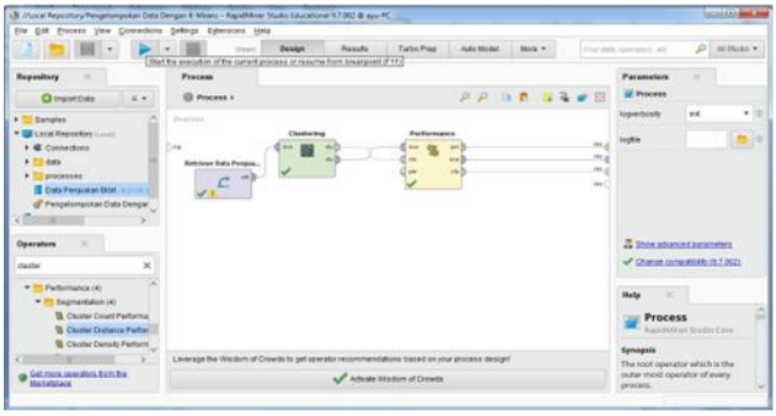


Figure 12 Testing Process

3.3 System Results and Evaluation

The results of system testing using Rapid Miner software on previously prepared seed sales data can be seen in figure 13, figure 14, figure 15 and 16. The picture shows the results of grouping data that has been carried out in the Rapid Miner software. In figure 17 can also be seen Grampic and the data is included in cluster 1. In figure 18 can also be seen the Grampic and data included in cluster 2.

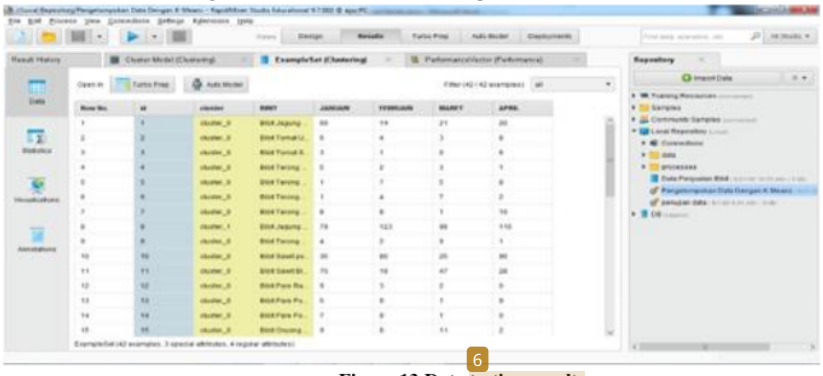


Figure 13 Data testing results

Figure 14 Next Data testing results

Figure 15 Next Data testing results

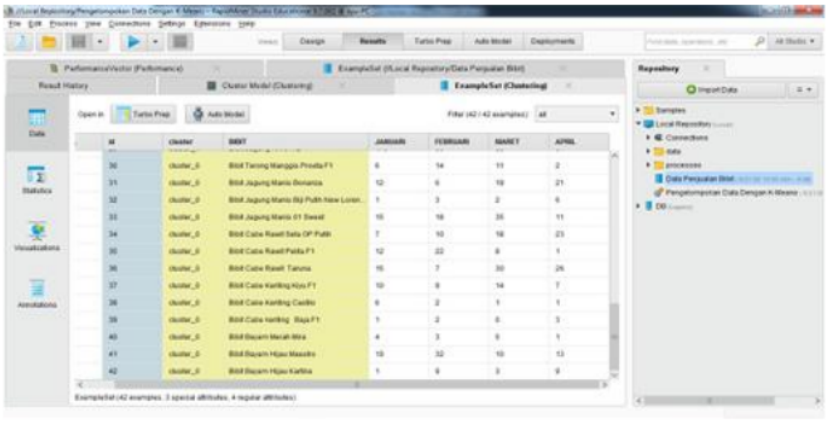


Figure 16 Next Data testing results

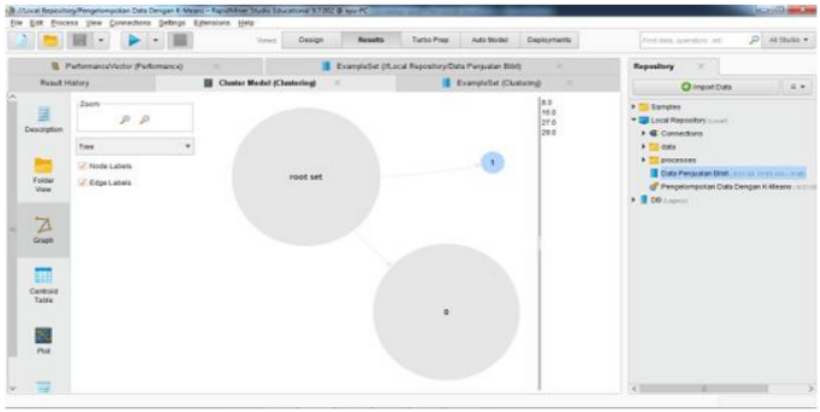


Figure 17 Graph and Data Views Included in Cluster 1

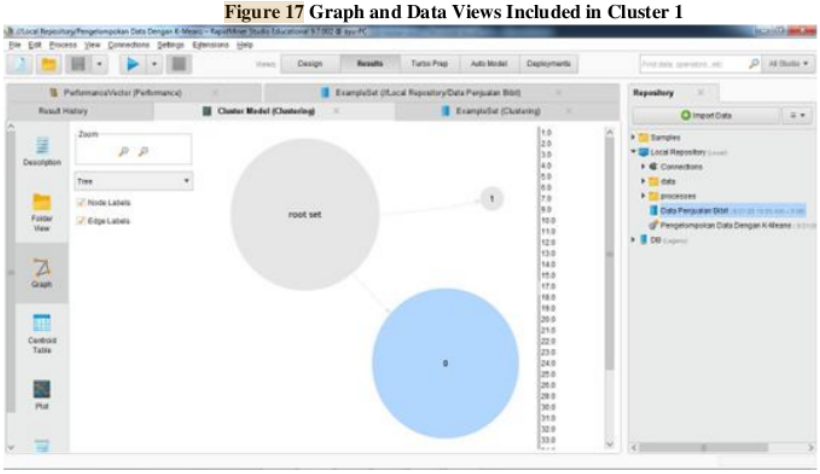


Figure 18 Graph and Data Views Included in Cluster 2

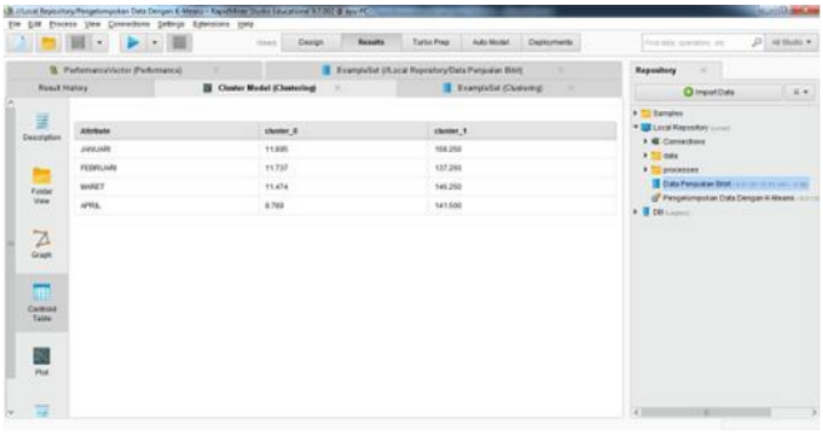


Figure 19 Center Cluster for the Last Iteration

Conclusion from the results of tests that have been carried out using RapidMining software, a type of seed that is in great demand by farmers at UD. The Tiara Brothers are NK 212 Corn Seeds, NK 7328 Corn Seeds, Pioneer 32 Corn Seeds, NK 617232 Corn Seeds

4. CONCLUSION

The conclusions that can be generated from research on the best seed clustering using Rapid Miner in increasing sales are as follows: With the application of the K-Means Algorithm to Rapid Miner, it can be applied in clustering and increasing seed sales based on seed sales data for 4 months, from January to April 2022. Grouping seed sales data by dividing seed sales data into 2 clusters. From the results of clustering, 4 types of seeds are in great demand by farmers, namely in cluster 1. NK 212 Corn Seeds, NK 7328 Corn Seeds, Pioneer 32 Corn Seeds, NK 617232 Corn Seeds are types of seeds that are in great demand by farmers. Cluster 2 shows seedlings that farmers are less interested in. With the application of the K-Means algorithm in Rapid Miner, the availability of seed stock can be managed properly, namely by maintaining the availability of seed stock that is most requested by farmers so that it can increase sales.

5. REFERENCES

Ali, A. (2019). Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *19*(1), 186–195. https://doi.org/10.30812/matrik.v19i1.529

Alkhairi, P., & Windarto, A. P. (2019). Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara. *Seminar Nasional Teknologi Komputer & Sains*, 762–767.

Bhargava, A. (2019). Grouping of Medicinal Drugs Used for Similar Symptoms by Mining Clusters from Drug Benefits Reviews. *SSRN Electronic Journal*, 1049–1058. https://doi.org/10.2139/ssrn.3356314

Liu, R.-Q., Lee, Y.-C., & Mu, H.-L. (2018). Customer Classification and Market Basket Analysis Using K-Means Clustering and Association Rules: Evidence from Distribution Big Data of Korean Retailing Company. *Knowledge Management Review*, *19*(4), 59–76. https://doi.org/10.15813/kmr.2018.19.4.004

S, I. T. J. (2019). Survey of Data Mining Algorithms for Intelligent Computing System. *Journal of Trends in Computer Science and Smart Technology*, *01*(01), 14–23. https://doi.org/10.36548/jtcsst.2019.1.002

Siregar, M. H. (2018). Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan). *Jurnal Teknologi Dan Open Source*, *1*(2), 83–91. https://doi.org/10.36378/jtos.v1i2.24

Yunita, F. (2018). Penerapan Data Mining Menggunkan Algoritma K-Means Clustring Pada Penerimaan Mahasiswa Baru. *Sistemasi*, *7*(3), 238. https://doi.org/10.32520/stmsi.v7i3.388

INCREASE SALES FOR THE BEST SEED CLUSTERING USING RAPID MINER

ORIGINALITY REPORT

9%	5%	6%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ojs.unud.ac.id Internet Source	1%
2	Fitri Nuraeni, N. Nelis Febriani SM, Lina Listiani, Eka Rahmawati. "Implementation of K-Means Algorithm with Distance of Euclidean Proximity in Clustering Cases of Violence Against Women and Children", 2019 1st International Conference on Cybernetics and Intelligent System (ICORIS), 2019 Publication	1%
3	Nurahman Nurahman, Agung Purwanto, Sigit Mulyanto. "Klasterisasi Sekolah Menggunakan Algoritma K-Means berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik", MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, 2022 Publication	1%
4	Submitted to National College of Ireland Student Paper	1%
	medikom.iocspublisher.org	

5	Internet Source	1 %
6	Submitted to University of Wales Swansea Student Paper	1 %
7	online-journals.org Internet Source	1 %
8	www.hindawi.com Internet Source	1 %
9	Submitted to RDI Distance Learning Student Paper	<1 %
10	ia803409.us.archive.org Internet Source	<1 %
11	www.researchgate.net Internet Source	<1 %
12	Lia Fitrianingrum, Taufik Hidayat. "Analysis of Energy Policy in Indonesia (Intermestic Model Approach)", 2020 International Conference on Sustainable Energy Engineering and Application (ICSEEA), 2020 Publication	<1 %
13	doaj.org Internet Source	<1 %
14	ebin.pub Internet Source	<1 %

15	Dony Novaliendry, Tegar Wibowo, Noper Ardi, Tiolina Evi, Dwi Admojo. "Optimizing Patient Medical Records Grouping through Data Mining and K-Means Clustering Algorithm: A Case Study at RSUD Mohammad Natsir Solok", International Journal of Online and Biomedical Engineering (iJOE), 2023 Publication	<1%
16	Relita Buaton, Solikhun Solikhun. "The Application of Numerical Measure Variations in K-Means Clustering for Grouping Data", MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, 2023 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		