Journal of Applied Engineering and Technological Science
Vol 1(1) 2019 : 17-29

IMPLEMENTATION OF DATA MINING FOR DETERMINING MAJORS USING K-MEANS ALGORITHM IN STUDENTS OF SMA NEGERI 1 PANGKALAN KERINCI

Yuda Irawan
STMIK Hang Tuah Pekanbaru
Yudairawan89@gmail.com

ABSTRACT

SMA Negeri 1 Pangkalan kerinci is one of the middle schools located at Jalan Lintas Timur Kerinci Pelalawan Indonesia which currently has 2 majors namely Science and IPS. This student majors can lead learners to focus more on developing their own abilities and interests. Selection of inappropriate majors can be very detrimental to students of their interests and careers in the future. With the majors are expected to maximize the potential, talent or individual talents, so as to maximize academic value. Based on the background, then by applying data mining techniques are expected to help students to determine the correct majors in accordance with the criteria set. Data mining techniques used in the determination of this department using K-Means Algorithm.

Keywords: *Data Mining, K-Means Algorithm, Senior High School*

1. INTRODUCTION

State High School (SMA) 1 Pangkalan Kerinci is a government agency whose basic purpose is to hone the students' abilities during the education process. The function of education is not only in the teaching and learning process, but also includes guidance, selection and placement of student majors in accordance with their individual capacities. Therefore, the school plays an important role in being able to develop the potential that is owned by students.

Each student has different abilities and academic qualifications, so the school is responsible for directing students in the selection of the right majors. However, there are problems faced, the school difficulties in grouping majors students who have different abilities and academic qualifications, because so far the selection of majors in SMA Negeri 1 Pangkalan Kerinci is done based on the wishes of each student. Data mining is a method for finding new information that is useful from a large amount of data collection and can help in making decisions. Data mining can be used for several things, namely estimation, prediction, classification, clustering and association methods. In this research, clustering method with the k-means algorithm will be applied. Clustering method is a technique of grouping data by separating data into a number of groups according to certain characteristics desired where the identity of the group from each data is unknown. With this grouping it is expected to find out which group of data can be identified and then be given an identity according to the problem at hand.

2. LITERATURE REVIEW

In this case the researchers conducted data collection from previous studies, namely research related to data mining clustering methods as a useful reference material to support this research. Below there are several studies relating to the problem to be investigated, namely:

Nurhayati and Pratiwi (2015) With the title "The Application of K-Means Algorithm in Data Mining for Department Specialization for Class X Students (Case Study: SMA Negeri 29 Jakarta)". The purpose of this study is to determine the majors for students grouped into two clusters. And then Putra & Wadisman (2018) in their research study about data mining using K-Means algorithm and Tanagra software to implemented in selection of potential customers in MC Laundry. And then Rahayu (2019) discuss about IQ Expert System In Grouping Mental Reterdation Children With K-Means Algorithm.

Sulistiyani, et. al., (2015) with the title "Determination System of State High School 1 Karangmojo". This study will analyze the application of the K-means and TOPSIS algorithms to classify and rank high school students (SMA) based on UN scores, report cards, and placement test scores.

Purnamaningsih, et. al., (2014) with the title "Utilization of the K-Means Clustering Method in Determining High School Student Management". The results of this study were carried out individually clustering for Natural Sciences / Social Sciences and presented the comparison of the results of K-Means clustering criteria for academic values, IQ scores, student interest with clustering K-Means academic grades. Student data is grouped individually according to their respective majors. The Natural Sciences majors are grouped into two, they are accepted by Natural Sciences and rejected by Natural Sciences. IPS majors are grouped into two, they are IPS and rejected IPS. Then each cluster is classified based on which criteria are prioritized. The cluster with the largest value in the final centroid is the cluster that IPA / IPS receives, while the cluster with the smallest value in the final centroid is the cluster that is rejected by IPA / IPS. It is almost same like Yunefri, et al., (2019) using K-Means Clustering, from the software that was built to help instructors in the subject of data structure in the process of grouping tutoring students. Grouping methods can be implemented to build valid student guidance grouping software.

Josi Aranda & Wirda Astari Galvani Natasya (2016) with the title "The Application of the K-Means Cluster Analysis Method in the Concentration Selection Decision Support System for Yogyakarta Amikom Stmik International Class Students". The results of this study are to produce a decision support system for choosing concentration for students. Through this decision support system, it is expected to help direct students in choosing concentrations that are in accordance with their interests and abilities. Determination of concentration is based on the number of credits that have been taken during lectures and the weighting of the related course values available in the curriculum offered.

Yuda Irawan (2019) with the research title: Application Of Data Mining For Evaluation Of Sales Data Using Clustering And Divisive Hirarki Algorithm Methods In Pekanbaru World Media Company. The main purpose of this research using clustering method is grouping a number of data / objects into a cluster (group) so that the cluster will contain the same data as each group. In this study, the Divisive Hierarchy Algorithm is used to form clusters. The pattern obtained is expected to provide knowledge for Media World Pekanbaru companies as a supporting tool for making policy.

Johan Oscar Ong (2013) with the title "Implementation of K-means Clustering Algorithm to Determine President University's Marketing Strategy". The problem of this research is that because president university students come from various regions and even countries, special marketing strategies are needed in conducting marketing to find prospective students so that promotions are carried out more effectively and efficiently. And the results of this study are strategies to promote cities in regions in Indonesia with majors that are in great demand and the level of academic ability of prospective students.

3. Research Methods

3.1 CRISP-DM method

3.1.1. Business Understanding Phase

Understanding of business is done by studying the research object, namely at SMA Negeri 1 Pangkalan Kerinci. In this phase, the determination of business goals is to group students based on the grades achieved by each student in class X (Ten) in semester II.

3.1.2. Data Understanding Phase

The data used for the K-Means method analysis process is value data. The grades taken as the basis for the clustering process are semester II grades in class X (ten). After the data is obtained, the next process is to understand the data.

3.1.3. Data Processing Phase (Data Preperation Phase)

Data preparation is done by selecting the attributes that will be used for the modeling process taken from student data. The attributes that have been selected will be saved back into a new dataset that is ready to be processed into the modeling process.

3.1.4 Modeling Phase

The modeling stage will use clustering method with algorithm-means. In applying the clustering method, it will be divided into 2 clusters which will cluster the second semester grades in class X (ten). The modeling tool used is Rapidminer.

3.1.5 The Evaluation Phase

At the evaluation stage the model will be assessed whether the results obtained from the clustering process have met the objectives set in the business understanding stage. At the stage of business understanding the goal has been determined, namely to determine the majors of students through semester II grades.

3.1.6 Deployment Phase

It is the application phase of the clastering technique in accordance with the goals of the objectives to be achieved in the first phase, which is to group students based on the grades achieved by each student in class X (ten) in semester II. From these results can also be known which students are entitled to occupy majoring in Natural Sciences and Social Sciences.

4. RESEARCH RESULTS AND DISCUSSION

4.1 Problem Analysis

To identify problems that exist in SMK Negeri 1 Pangkalan Kerinci in processing student majors data, an analysis of performance, information, economy, security or security, efficiency and service will be carried out. This analysis is also called PIECES Analysis. Where this analysis is as follows:

Table 1. Table 1 Pieces Analysis

No	Pieces Analysis	Old System Analysis	New System Analysis
1	Performance	Reporting services to those who are responsible for SMK Negeri 1 Pangkalan Kerinci have not been maximized and are not efficient.	In the form of a supporting software in processing data so that later it can smooth its performance.
2	Information	During this time the information generated has not been able to produce specific and efficient information, so the processing of student majors data has not been carried out properly.	In this research the supporting software that will be used is Rapidminer software, where the information obtained is more specific and efficient.
3	Control	There is no security system at SMK Negeri 1 Pangkalan Kerinci.	Development of the system is safer and there are no disturbances so that later it can smooth the performance of the responsible department.
4	Efficiency	In the process of determining the direction of students who are currently still determined manually by their desires. This is not optimal.	More efficient because it can help in majors.
5	Service	In the service of parties in SMK 1 Pangkalan Kerinci have not been able to effectively control student majors, so that there is a specialization in majors.	With the application of data mining is expected to effectively control student direction.

4.2 CRISP-DM

This research was conducted using the CRISP-DM (Cross Industry Standard Process for Data Mining) method. In this method there are 6 stages used, namely:

4.2.1 Business Understanding Phase

Vocational High School 4 Vocational School located on Jalan Hang Tuah V Pangkalan Kerinci, is a school whose basic purpose is to hone students' abilities during the education process.

a. Determine Business Objectives (Determine Business Objectives)
   The business objective of doing this research is to determine the pattern of student majors so that they can be used as one of the basic decision makers for determining the direction of students for the future.
b. Assess the situation (Asses The Situation)
   Understand student majors at SMK Negeri 1 Pangkalan Kerinci in business goals and then translate them into data mining purposes.
c. Determine the Initial Data Mining Strategy
   The initial strategy in this research is to review and request student data in advance at SMA Negeri 1 Pangkalan Kerinci.

4.2.2. Data Understanding Phase

The student dataset obtained from SMA Negeri 1 Pangkalan Kerinci is in the form of excel documents totaling 141 items in 2018. Following are the stages in the data understanding phase.

a. Initial Data Collection
   The data collected and used in this data mining process are student data at SMK Negeri 1 Pangkalan Kerinci in 2017. Where the data collected consists of student names, classes, Nisn, student grades namely mathematics, physics, chemistry, biology, geography, sociology, economics and history. (Appendix 1)
b. The collected data will later be processed in the data mining stage using clustering.
   Previously, the data that was collected had to be selected before entering the clustering process. So that in the grouping later it will be known which parts should be grouped in student majors data.
c. Describe Data
   The student dataset consists of the attributes of the student's name, mathematics, physics, chemistry, biology, geography, sociology, economics and history. The amount of data in the attribute is 141 items. (Appendix 1).
d. Data Quality Evaluation
   The results of the evaluation of the quality of the data are finding many empty or null values called missing values on the attributes in the student dataset.
e. Attribute Selection
   In this study, the attributes used are student names, mathematics, physics, chemistry, biology, geography, sociology, economics and history. Which has 141 items. (Appendix 2)

4.2.3. Data Processing Phase (Data Preparation Phase)

Data preparation includes all activities to build student datasets that will be applied into the modeling tool, from the initial raw data in the form of student datasets and subsequently will do the data mining process.

a. Data Selection
   Where the raw data to be selected in this study uses the attributes of student names, mathematics, physics, chemistry, biology, geography, sociology, economics and history in student data at SMK Negeri 1 Pangkalan Kerinci. (Appendix 2)
b. Raw Data Processing (Preprosessing Data)
   At this stage it is a stage to ensure that the selected student's data is suitable for processing. After checking the data one by one and there is no problematic data after cleaning, then the usage data is combined in one group. The data groups will be tested in the K-means Algorithm process to form 2 groups in use.

4.2.4. Data Modeling Phase

Modeling is a phase that directly involves data mining techniques by selecting data mining techniques and determining the algorithm to be used. Data mining modeling in this study uses Rapid Miner software version 7.0. In this application a clustering algorithm is

available in the form of the K-Means algorithm. Where is the K-Means flowchart algorithm as follows:

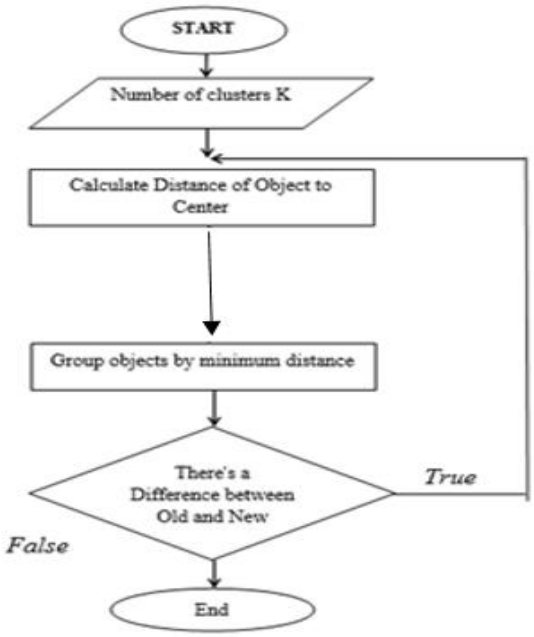

Figure 1. Flowchart Rapidminer.

4.2.5. Calculations with the K-Means Algorithm

a. Determination of Cluster Start Center
First of all the existing data will be grouped into 1 group. The initial center of a cluster or centroid is determined by the highest and lowest values. Here are the initial cluster centers for cluster 2, cluster 1 (89, 88, 86, 87, 88, 89, 90, 90), cluster 2 (69, 68, 68, 70, 70, 70, 69, 73).

b. Calculate The Distance Of Each Data To Each Centroids
To calculate the distance of data to each centroids using the formula 1 =

$$\sqrt{(C_{ij} - C_{kj})^2 + (C_{ij} - C_{kj})^2 + \cdots + (C_{ij} - C_{kj})^2}$$

Table 2. Student Data

No	Student's Name	Math	physics	chemistry	biology	geography	Sosiology	Economy	History
1	Ariaman	79	69	65	70	70	73	70	70
2	Corin Lamsihar Sianturi	79	67	70	67	88	75	82	80
3	Dandi Wahyudi	70	75	75	75	83	80	79	75
4	Daniel Parlindungan H	72	70	70	75	80	81	80	80
5	Diki	70	67	70	77	84	75	80	82
6	Eka Rianti Br. Malau	85	84	80	87	89	88	88	90
7	Fajar Bakti	80	75	70	73	85	80	79	80
8	Firman Syahroni	79	79	80	70	80	81	88	80
9	Iklas Gabariela	80	79	80	80	85	83	80	80
10	Irsan Azi Isnawan	75	70	70	75	80	80	88	80
11	Irwan Juanda	70	80	75	83	88	85	82	80

	Simare-Mare								
12	Irwanto	71	70	75	70	79	80	88	80
13	Jadiaman	78	77	77	75	80	81	80	80
14	Jumaiza Maherni	77	75	73	80	79	80	80	80
15	Karina Natalia	89	88	86	87	88	89	90	90
16	Kristina	80	75	83	77	88	85	79	80
17	Lastri	83	80	77	79	88	85	80	83
18	Lismawati Siagian	79	75	75	80	80	83	80	80
19	Lorida Tampubolon	79	76	70	77	73	77	79	80
20	Madison Abibu Sibarani	70	75	70	75	80	79	75	80
21	Melisa	70	70	77	85	88	85	80	88
22	Pianda Setiawan	69	70	75	80	80	80	75	79
23	Pika Talia	80	73	70	79	75	79	79	80
24	Rahul Arif F	70	70	73	67	80	77	75	80
25	Ramadani	70	70	75	77	88	89	85	90
26	Rati Mantili Situmorang	71	70	77	75	81	81	80	79
27	Rindi Antika	77	70	80	70	85	87	80	85
28	Suandrian	77	74	79	80	78	80	80	83
29	Sahman Tawar Tanjung	70	71	77	75	76	75	78	80
30	Septiani Pratiwi	84	85	81	80	80	81	80	83
31	Siti Aminah Hsb	75	79	77	80	80	83	79	81
32	Slamat Harianto	70	75	70	71	77	75	70	75
33	Toni Kurniawan	75	70	70	69	79	80	80	88
34	Tri Wulandari	80	70	77	79	80	83	88	88
35	Vera Yanti Br. Boang Manalu	85	80	80	84	88	85	87	90
36	Afriyana Mart	84	85	80	80	88	80	85	90
37	Agnes Heppi Monika Rumapea	80	79	75	78	80	79	88	85
38	Agung Priyanto	79	70	76	75	80	79	80	80
39	Amelia Permata Hati Marbun	70	73	70	75	88	85	79	85
40	Arnold Finesti Tampubolon	75	73	77	75	80	79	78	80
41	Asniar Nainggolan	77	78	79	80	80	79	79	80
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
140	Selvia Dewi Br Pasaribu	80	79	75	79	83	81	79	85
141	Siska Hariani	79	77	75	79	80	80	75	79

Calculation of the distance in each data to get the closest centroid with the following formula:
Cluster 1

$$D1 = \sqrt{\begin{matrix}(79-89)^2+(69-88)^2+(65-86)^2+(70-87)^2+\\(70-88)^2+(73-89)^2+(70-90)^2+(70-90)^2\\ \square \\ = 50{,}70502934\end{matrix}}$$

Cluster 2

$$D1= \sqrt{\begin{matrix}(79-69)^2+(69-68)^2+(65-68)^2+70-70)^2+\\((70-70)^2+(73-70)^2+(70-69)^2+(70-73)^2\\ \square \\ = 11{,}35781669\end{matrix}}$$

Here are the results of the calculation of the distance of each data created in the form of a table.

Table 3. Calculation of Cluster Center Distance

No	Distance to		Closest Distance
	C1	C2	
1	50,70502934	11,35781669	11,35781669
2	39,45883931	26,0959767	26,0959767
3	35,31288717	22,29349681	22,29349681
4	36,61966685	20,80865205	20,80865205
5	39,16631206	21,86321111	21,86321111
6	8,602325267	47,74934555	8,602325267
7	31,82766093	25,63201124	25,63201124
8	27,16615541	31,55946768	27,16615541
9	22,18107301	32,40370349	22,18107301
10	34,19064199	26,05762844	26,05762844
11	27,23967694	33,61547263	27,23967694
12	36,3868108	25,45584412	25,45584412
13	28,19574436	25,67099531	25,67099531
14	29,88310559	24,2693222	24,2693222
15	0	53,1789432	0
16	24,41311123	33,7934905	24,41311123
17	20,24845673	35,665109	20,24845673
18	27,18455444	27,14774392	27,14774392
19	34,49637662	20,59126028	20,59126028
20	37,6696164	18,57417562	18,57417562
21	29,83286778	34,72751071	29,83286778
22	37,21558813	20,61552813	20,61552813
23	33,40658618	22,04540769	22,04540769
24	42,27292278	16,52271164	16,52271164
25	30,5122926	36,51027253	30,5122926
26	34,74190553	22,64950331	22,64950331
27	30,5122926	31,48015248	30,5122926

28	27,71281292	26,57066051	26,57066051
29	38,19685851	17,52141547	17,52141547
30	19,62141687	35	19,62141687
31	26,62705391	27,76688675	26,62705391
32	44,5421149	11,5758369	11,5758369
33	36,95943723	23,91652149	23,91652149
34	25,65151068	33,7934905	25,65151068
35	12,24744871	43,61192497	12,24744871
36	15	41,55718951	15
37	23,60084744	32,26453161	23,60084744
38	32,12475681	23,32380758	23,32380758
39	33,88214869	29,12043956	29,12043956
40	32,46536616	21,86321111	21,86321111
41	26,96293753	26,73948391	26,73948391
42	31,4960315	28,31960452	28,31960452
43	36,70149861	19,31320792	19,31320792
44	39,66106403	18,02775638	18,02775638
45	45,46427169	10,90871211	10,90871211
46	47	13	13
47	23,49468025	34,92849839	23,49468025
48	29,22327839	34,40930107	29,22327839
49	50,11985634	6,633249581	6,633249581
.	.	.	.
.	.	.	.
.	.	.	.
140	24,12467616	30,09983383	24,12467616
141	29.94995826	24,14539294	24,14539294

c. Data Grouping

Distance calculation results at point 2 will be compared and selected the closest distance between the data with the center of the cluster, this distance will show that the data that has the closest distance is in one group with the closest cluster center, the grouping of data can be seen in table 4. below this, a value of 1 means that the data is in groups.

Table 4. Grouping Data in Iteration 1

No	Iteration 1	
	C1	C2
1		1
2		1
3		1
4		1
5		1
6	1	
7		1
8	1	
9	1	
10		1
11	1	
12		1
13		1
14		1
15	1	
16	1	
17	1	
18		1
19		1
20		1
21	1	
22		1
23		1
24		1
25	1	
26		1
27	1	
28		1
29		1
30	1	
31	1	
32		1
33		1
34	1	
35	1	
36	1	
37	1	
38		1
39		1
40		1
41		1
42		1
43		1
44		1
45		1
.		1
.		
.		
140	1	
141		1

d. Determination of a new Cluster Center

After obtaining members from each cluster then the new cluster center is calculated based on the data of each cluster member that has been obtained. The following results are explained in the table.

Table 5. Determination of New Cluster Centers

No	New Cluster	
	C1	C2
1	79,59574468	73,79787234
2	78,40425532	72,20212766
3	77,53191489	71,61702128
4	80,38297872	76,15957447
5	83,80851064	79,9893617
6	83,34042553	79,31914894
7	82,61702128	78,55319149
8	85,5106383	79,06976744

The next iteration is done in the same way until there is no change in data in a cluster.

4.2.6. Implementation by using RapidMiner

Enter the 2 clusters specified in the clustering operator in the parameters. After that, click run to conduct the test, it will produce the test in 2 clusters.

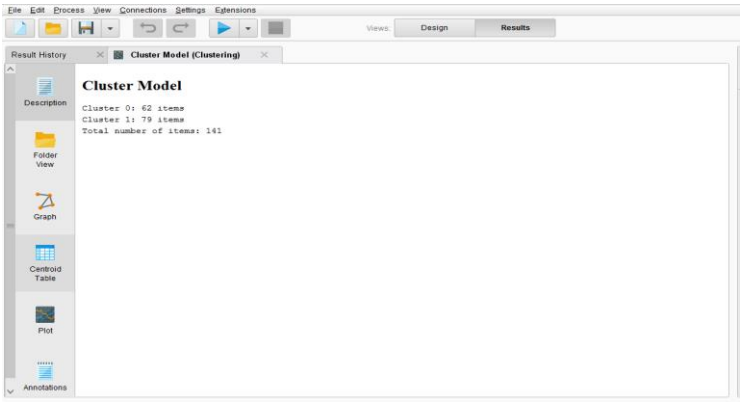

Figure 2. Display of test results

To see the Clustering results, click on the Description in the display as shown above. Description is to see the number of items contained in CO and C1.

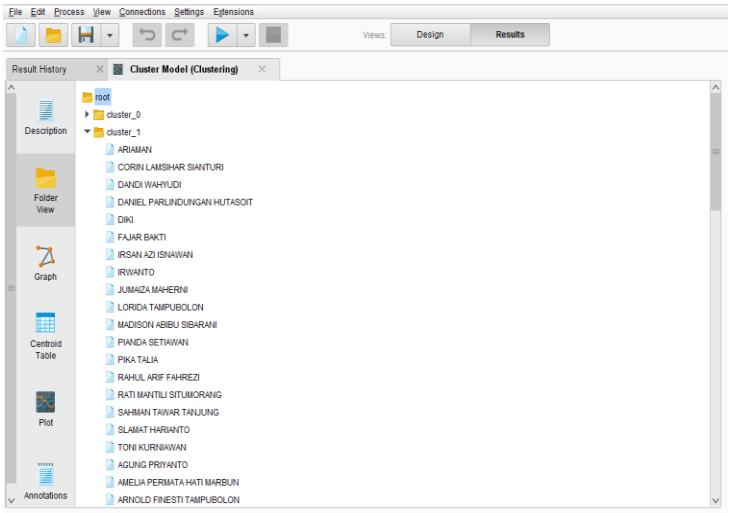

Figure 3. Cluster Display

The picture above is a display to see the data contained in each cluster. From Folder View we can see which data belongs to the data clusters group 0 and cluster 1.

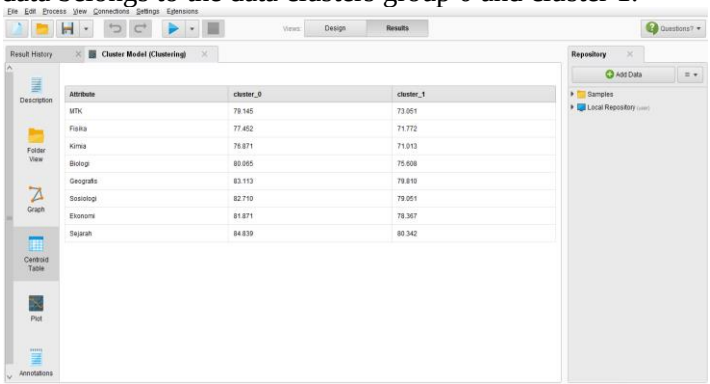

Figure 4. Display of Centroid Table

The picture above is the result of the centroid table. In the centroid table there are attributes.

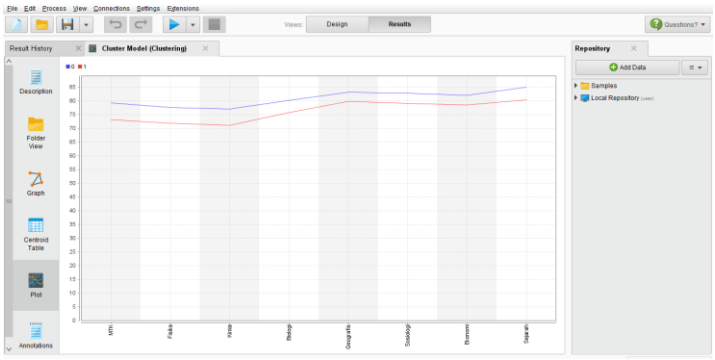

Figure 5. Plot Views

The picture above is the result of the plot display, which is a graph where the red line is C1 with the lowest value, which is around 71, while the blue line is C0, with the highest value, 76 above and is included in the science department.

4.2.7. Evaluation Phase

Evaluation is an advanced phase of the purpose of data mining. Evaluation is carried out in depth with the aim that the results at the modeling stage are in line with the objectives to be achieved in the business understanding stage.

4.2.8. Evaluation of Results

This stage assesses the extent to which the results of data mining modeling meet the objectives of data mining that have been determined at the business understanding stage. Where in testing data that has been processed in Rapid Miner. testing the data in Rapid Miner used in this study is the index.

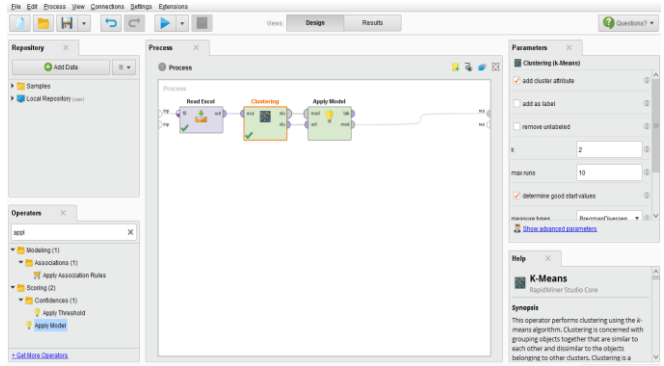

Figure 6. Display of apply models in the testing process

Take the apply model in the operators, then drag it into the process. Then connect the K-means with the apply model so that the testing process is successful, as shown above.

After the connection between K-means and apply models is carried out, the next step is to choose the cluster count performance in operators. Then drag the cluster performance count into the process.

The Apply model is used to read data to be clustered based on data that has been previously processed. Performance, this operator creates a performance vector that contains criteria for the number of clusters and the cluster number index of the cluster model.

After that click run, the results of the test data that will appear will appear.

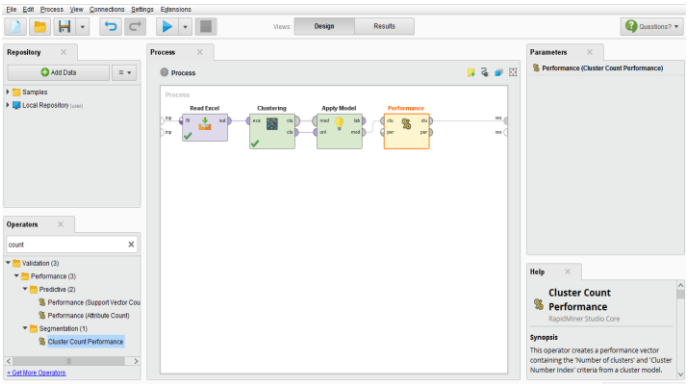

Figure 7. Display of Test Results

4.2.9. Deployment Phase

Deployment is the final stage in making reports on the results of data mining activities. The final report which contains the knowledge gained or pattern recognition in the data mining process.

Cluster Analysis

There are 8 attributes used in the clustering process namely Mathematics, physics, chemistry, biology, geography, sociology, economics and history. There are also 2 clusters used, cluster 0 taken from the highest value, cluster 1 taken from the lowest value. Here are the results of clustering:

1. Cluster 0, students with the highest category score, there are 62 students namely Eka rianti BR. Malau, Firman Syahroni, Iklas Gabariela, Karina Natalia, Kristina, Lastri, Irwan Juanda Simare-mare, Jadiaman, Lismawati Siagian, Melisa, Ramadhani, etc.
2. Cluster 1, students with low category grades, there are 79 namely Ariaman, Dandi wahyudi, Daniel parlindungan hutasoit, diki, Fajar devoted, Irsan azi isnaini, Pika talia, Amelia jewel of marbun heart, Rika rosa pertiwi, Riska dwiana untari, Devi anjani , etc.

5. CONCLUSION & SUGGESTION

Conclusion

From the analysis and testing that has been done, the authors can draw conclusions including:

1. Can help in processing student data in high school, so that the school can be easier and faster in processing data.
2. With this system can help in determining school majors.

Suggestions

From the conclusions drawn, suggestions that writers can provide for improvements in data processing are:

1. In processing data majors students of Pangkalan Kerinci 1 High School for the future can be even better.
2. Because this system is new for the SMA Negeri 1 Pangkalan Kerinci course, it must be adjusted between the data and reports with the system used at this time.

REFERENCES

Aranda, J., & Natasya, W. A. G. (2016). Penerapan Metode K-Means Cluster Analysis pada Sistem Pendukung Keputusan Pemilihan Konsentrasi Untuk Mahasiswa International Class STMIK Amikom Yogyakarta. *Seminar Nasional Teknologi Informasi dan Multimedia 2016*, di Yogyakarta, 6-7 Februari 2016.

Irawan, Y. (2019). Penerapan Data Mining Untuk Evaluasi Data Penjualan Menggunakan Metode Clustering Dan Algoritma Hirarki Divisive Di Perusahaan Media World Pekanbaru. *Jurnal Teknologi Informasi Universitas Lambung Mangkurat (Jtiulm)*, *4*(1), 13-20.

Nurhayati & Pratiwi, L. A. (2015). Penerapan Algoritma K-Means dalam Data Mining untuk Peminatan Jurusan Bagi Siswa Kelas X (Studi Kasus: SMA Negeri 29 Jakarta). *Prosiding Seminar Ilmiah Nasional Teknologi Komputer (SENATKOM 2015) Oktober 2015*, 1, 9-13.

Ong, J. O. (2013). Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University. *Jurnal Ilmiah Teknik Industri, 12*(1), 10-20.

Purnamaningsih, C., Saptono, R., & Azis, A. (2014). Pemanfaatan Metode K-Means Clustering dalam Penentuan Penjurusan Siswa SMA. *ITSMART: Jurnal Teknologi dan Informasi*, *3*(1), 27-33.

Putra, R., & Wadisman, C. (2018). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K Means. INTECOMS: *Journal of Information Technology and Computer Science*, *1*(1), 72-77. https://doi.org/https://doi.org/10.31539/intecoms.v1i1.141

Rahayu, N. (2019). Sistem Pakar IQ Dalam Pengelompokan Anak Reterdasi Mental Dengan Algoritma K-Means. INTECOMS: Journal of Information Technology and Computer Science, 2(1), 17-32. https://doi.org/https://doi.org/10.31539/intecoms.v2i1.659

Sulistiyani, M. E., Soedijono, B., & Syahdan, S. A. (2015). Sistem Penentuan Jurusan Sekolah Menengah Atas Negeri 1 Karangmojo. *Seminar Nasional Teknologi Informasi dan Multimedia 2015, 3*(1), 247-252.

Yunefri, Y., Pane, E., & Sutejo, S. (2019). Pengembangan Sistem Pengelompokan Belajar Mahasiswa Pada Matakuliah Struktur Data Dengan Metode K-Means. INTECOMS: Journal of Information Technology and Computer Science, 2(2), 59-66. https://doi.org/https://doi.org/10.31539/intecoms.v2i2.812