JURNAL IPTEKS TERAPAN
Research of Applied Science and Education V14.i2 (106-112)

ISSN : 1979-9292

E-ISSN : 2460-5611

ANALYSIS OF DETERMINATION OF TRY OUT MATHEMATICS SKILLS OF STUDENTS IN IPA 1 SMA N 1 KECAMATAN WITH CLUSTERING METHODS - MEANS

Sri Tria Siska[1)], Rahayu Mayang Sari[2)]
[1]Program Studi Teknik Komputer, Sekolah Tinggi Teknologi Payakumbuh
Email : siska6830@gmail.com
[2]Fakultas Sains dan Teknologi, Universitas Pembangunan Panca Budi
Email : rahayu@dosen.pancabudi.ac.id

Abstract

In this study aims to analyze the determination of the ability of TRY OUT Mathematics students majoring in Natural Sciences in SMA N.1 Kecamatan to get an information that is useful for schools in determining students based on their grades. In processing these data we use data mining with the K-Means Algorithm Clustering Method. Data Mining is extracting information from a large amount of data. The information that was obtained was in the form of groups of students with high grades, medium grades and low grades. The test that we do uses the existing application, RapidMiner, so that it can be determined by the desired student clusters.

Keywords : Data Mining, Clustering Methods, K – Means, Mathematics and Try Out,

INTRODUCTION

Within the system of planning for the 2018/2019 National Last Examination. Each school has in progress arranged the stabilization exercises for their understudies by extending the fabric and Attempt Out. Drawing closer the execution of the UN, arrangements are progressively being made strides both authoritative arrangements and tryouts (Attempt Out) of the UN. The National Tall School Final Examination Try Out may be a exceptionally imperative motivation for a school to extend graduation rates for its understudies. In learning exercises utilizing learning media such as tuning in, taking notes, and for preparing questions. SMAN.1 Payakumbuh Locale, is one of the state schools found in Koto Tangah Simalanggang, Payakumbuh, Lima Puluh Kota Rule. Within the system of conducting national examinations at the tall school level and identical, it is essential to hold a Attempt Out for subjects to be tried within the national exams, one of which is Science. Arithmetic is one of the subjects that will be tried in national examinations and entrance examinations in state and private colleges. Information Mining is extricating data from a expansive sum of information. Information gathering procedures can utilize K-Means Clustering. Agreeing to this calculation we must to begin with select the information esteem k as the beginning cluster center, at that point calculate the remove between each information esteem at the center of the cluster and decide the closest cluster, at that point upgrade the normal of all bunches, rehash this handle until the criteria don't

Submited : 03 Jul 2020 – Accepted : 11 Jul 2020 – Published : 12 Jul 2020

JURNAL IPTEKS TERAPAN
Research of Applied Science and Education V14.i2 (106-112)

ISSN : 1979-9292

E-ISSN : 2460-5611

coordinate . . There are numerous strategies for conducting Information Mining in its application, among others clustering, classification, affiliation run the show mining, neural systems, hereditary calculations and others. One of the clustering calculations that can be utilized in deciding numerical attempt out capacity in SMA N.1 Kecamatan is K-Means.

LITERATURE

Information Mining, regularly moreover called information revelation in database (KDD), is an action that includes collecting, utilizing chronicled information to discover regularities, designs or connections in expansive information sets. The yield of information mining can be utilized to make strides choice making within the future[1]..

Information Mining may be a strategy of handling information to discover covered up designs from the information. The comes about of information handling with the Data Mining method can be utilized to create choices within the future. Information Mining is additionally known as design acknowledgment. Information Mining may be a large-scale information preparing strategy subsequently Information Mining has an vital part within the areas of industry, back, climate, science and innovation. In common, the Information Mining consider examines strategies such as clustering, classification, relapse, variable determination, and advertise wicker container investigation[2.

Clustering is one of the strategies of Information Mining that's unsupervised, meaning that this strategy is connected without preparing and without a instructor and does not require target yield[3].K-Means may be a dreary clustering calculation. The K-Means calculation begins with arbitrary choice K, K here is the number of clusters that need to be shaped. At that point allot K values randomly, briefly these values become the center of the cluster or commonly alluded to as centroid, cruel or "implies". Calculate the remove of each existing information to each centroid utilizing the Euclidian equation until the closest separate from each information is found with a centroid. Classify each information based on its vicinity to centroids. Perform these steps until the centroid esteem does not alter [4]. K-Means Clustering is a method that belongs to the Partitioning Clustering algorithm. The steps of the K-Means method are as follows, Decide k as the number of clusters shaped. To decide the number of clusters k is done with a few contemplations such as hypothetical and conceptual contemplations that may well be proposed to decide how numerous clusters[5]. Stir the beginning Centroid k (cluster center point) haphazardly. The beginning centroid assurance is done haphazardly from the accessible objects as numerous as k clusters, at that point to calculate the following centroid cluster i, utilizing the taking after equation:

$$v = \frac{\sum_{i=1}^{n} x_i}{n} \quad ;i = 1,2,3,...n$$

Where :

xi	: object x to-i
yi	: source y to-i
n	: Summary of object

Calculate the distance of each object to each centroid of each cluster. To calculate the distance between objects and centroids, the writer uses Euclidian Distance.

Submited : 03 Jul 2020 – Accepted : 11 Jul 2020 – Published : 12 Jul 2020

JURNAL IPTEKS TERAPAN
Research of Applied Science and Education V14.i2 (106-112)

ISSN : 1979-9292

E-ISSN : 2460-5611

$$d(x,y) = \|x - y\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad ;i = 1,2,3,\ldots,n \qquad (2)$$

Where :

xi	: object x to-i
yi	: source y to-i
n	: Summary of object

Allocate each object into the closest centroid. To allocate objects into each cluster during iteration can generally be done in two ways, namely by hard k-means, where each object is expressly declared as a member of the cluster by measuring the distance of its proximity to the center point of the cluster, other ways can be done with fuzzy C-Means[6]. Iterate, then determine the new centroid position using the equation[7]. Repeat step 3 if the new centroid positions are not the same [8]. The merging check is done by comparing the gather task network within the past emphasis with the group task lattice within the current cycle. In case the comes about are the same at that point the k-means cluster examination calculation is converging, but on the off chance that it is distinctive at that point it isn't however focalizing so the another cycle has to be done. Within the application of the K-Means Cluster Examination strategy, the information that can be prepared in calculations is numeric information within the shape of numbers. Whereas information other than numbers can also be connected, it must to begin with be encoded to form it simpler to calculate the remove / similitude of characteristics had by each protest. Each question is calculated based on the vicinity of the characters claimed by the center of the cluster that has been foreordained, the littlest (2) separate between the question and each cluster is the closest cluster part. After the number of clusters is decided, at that point 3 objects are arbitrarily chosen agreeing to the number of clusters shaped as the center of the starting cluster to calculate the vicinity of all existing objects[9]. One strategy of gathering archives is K-Means Clustering. K-Means Clustering is the best gathering strategy that bunches information into k bunches based on the centroid of each gather. It's fair that the comes about of K-Means are enormously impacted by the k parameters and centroid initialization. In common, K-Means initializes centroid haphazardly. Be that as it may, the proposed strategy will alter K-Means in centroid initialization particularly in progressing execution in report gathering[10].

In contrast to the existing K-Means algorithm in general, K-Means Clustering in this study performs centroid initialization using Jaccard Distance measurement with the formula :

$$J_\delta(A,B) = 1 - J(A,B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|} \qquad (3)$$

This measurement is done to find the difference between two vectors in n dimensional space. We first measure the matrix differences between all pairs of documents, where this document is represented as a vector in n-dimensional space using formulas:

$$r'(d_1, d_2) = \left(\frac{|T(d_1) \cap T(d_2)|}{|T(d_1) \cup T(d_2)|}\right) \qquad (4)$$

Submited : 03 Jul 2020 – Accepted : 11 Jul 2020 – Published : 12 Jul 2020

JURNAL IPTEKS TERAPAN
Research of Applied Science and Education V14.i2 (106-112)

ISSN : 1979-9292

E-ISSN : 2460-5611

where T (d) is a feature that appears from documents d and $r\ '\ d1,\ d2\ =\ 1$ and $r\ '\ d1,\ d2\ =\ r\ '\ (d2,\ d1)$[11].

By using the complementary of the Jaccard similarity matrix $'\ 1\ =\ 1\ -\ r$ matriks, a matrix that indicates the differences between documents will be obtained. From this difference matrix will later be chosen k which is the most different document and initialize it as k center of the cluster. The next step is to place the document in the correct group. We use Cosine similarity to measure the similarity between documents and group centroids with formulas:

$$similaritas = \cos(\theta) = \left(\frac{|A.B|}{\|A\|\|B\|}\right) \quad (5)$$

then place each of these documents into the group that is most similar. If all documents have been distributed to all groups and there are no more additions to the centroid cluster, the iteration of the cluster will end [12].

RESULT

The data collection used is by conducting a TRY OUT Mathematics test for Natural Sciences Students and analyzing the results of the TRY OUT Mathematics test in SMA N.1 Payakumbuh District. From the Analysis Data the value will be processed to get information where the data to be used is 104 data

Table 1. Data Analysis Results TRY OUT exam students in SMA N.1 Subdistrict Payakumbuh.

NO	NAMA SISWA	JURUSAN	JENIS KELAMIN	JMLH BENAR	NILAI
1	RIZAL JASRIL	IPA-1	L	14	35
2	ARYA DZAKY ADHA	IPA-1	L	6	15
3	MAULIDA ANJELA	IPA-1	L	13	32.5
4	WINDI AULIA	IPA-1	L	11	27.5
5	ANDRA FADILA	IPA-1	L	11	27.5
6	SRI FITRIANI	IPA-1	P	16	40
7	WIDYA GRAWENITA	IPA-1	P	10	25
8	ZULAIKA WIDIA ERITA	IPA-1	P	11	27.5
9	FAJRIANI RATI	IPA-1	P	11	27.5
10	ICLIMA YESIKA	IPA-1	P	9	22.5
11	RETNO SASI P.H	IPA-1	P	12	30
12	MISA RAMADHANI	IPA-1	P	12	30
13	WENI NUR ANISA	IPA-1	P	11	27.5
14	MARDIYAH MUTIARA	IPA-1	P	7	17.5
15	SHINTA CHANIA P	IPA-1	P	10	25
16	SAFIRA NURUL FADILA	IPA-1	P	13	32.5
17	M.SYAKHI FADILA	IPA-1	L	13	32.5
18	AHMAD ADITIA	IPA-1	L	7	17.5
19	DESTY NESTALIA	IPA-1	P	10	25
20	MUTIARA BALVI	IPA-1	P	7	17.5
21	MUHAMMAD HANDAWAH	IPA-1	L	16	40
22	ANGGI OKTAVIANUS	IPA-1	L	6	15
23	REVINA ALFIANTI	IPA-1	P	15	37.5
24	DODI SATRIA	IPA-1	L	8	20
25	ARRIVA MARDATILLAH	IPA-1	P	10	25
26	DEFI JUITA	IPA-2	P	8	20
27	DINA YULIANTI	IPA-2	P	6	15
28	LARA OKTARI	IPA-2	P	5	12.5
29	DESMI FETRI IDRIANA	IPA-2	P	10	25
30	JENNYSA RABBANA. D	IPA-2	P	15	37.5

This data transformation is done because the data type is not numeric so the data must be transformed in advance by doing the frequency with the most data appearing by sorting the highest frequency to the lowest and initialing the data.

Table 2. Gender Transformation

Transformasi Jenis Kelamin		
Jenis Kelamin	**Frekuensi**	**Inisial**
Perempuan	70	1
Laki - laki	34	2

Table 3. Student Majors Transformation

Jurusan		
Jurusan	**Frekuensi**	**Inisial**
IPA - 2	27	1
IPA - 3	26	2

Submited : 03 Jul 2020 – Accepted : 11 Jul 2020 – Published : 12 Jul 2020

IPA - 4	26	3
IPA - 1	25	4

In this study, the determination of the number of groups or clusters is determined based on data from student test results in SMA N.1, Payakumbuh sub-district consisting of the names of students, classes and majors, gender and the number of results obtained in TRY OUT mathematics. Clusters to be formed are 3 clusters or k = 3. Where the attributes that will be used later are as many as 4 pieces of attributes.

The initial centroid value in this study was randomly selected, where the number of initial centroids was carried out as many as three initial centroids, the value for C1 was taken from the 6th data row, the C2 value was taken from the 69th row, the C3 value was taken from the 45th data row . The following initial centroid values in the study:C1 = (4, 1, 16, 40)
C2 = (2, 1, 2, 5)
C3 = (1, 1, 11, 27,5)

Calculate the distance of each existing data to each cluster center. To calculate the distance of each existing data to the center of the cluster there are several ways, namely by using the Manhattan / City Block formula, and Euclidean Distance. Whereas in this study the authors used the Euclidean Distance formula to calculate the distance of each data to the cluster center point. Allocate each data into the closest centroid. In reallocating objects into each cluster based on the comparison of the distance between the data with the centroid of each existing cluster, the object is explicitly allocated into a cluster that has the distance to the closest centroid to the data. Iterate,
Then determine the position of the new centroid by calculating the average of the data in the same centroid,

Results of the 0th DataIteration Grouping to the 1st Iteration

ITERASI - 0			ITERASI - 1		
Cluster 0	Cluster 1	Cluster 2	Cluster 0	Cluster 1	Cluster 2
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
0	1	1	0	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1

Submited : 03 Jul 2020 – Accepted : 11 Jul 2020 – Published : 12 Jul 2020

ISSN : 1979-9292
E-ISSN : 2460-5611

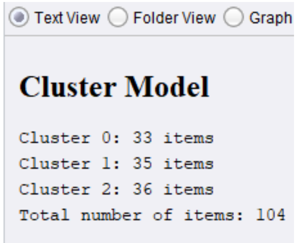
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	0	1	1	0
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	0	1	1	0	1
1	1	1	1	1	1
1	1	1	1	1	1

It can be seen in the 0 iteration and the 1 iteration no longer changes at the cluster point, so it can be concluded that the iteration can be stopped at the 1 iteration with the results:

Cluster 1 = 33 Students
Cluster 2 = 35 Students
Cluster 3 = 36 Students

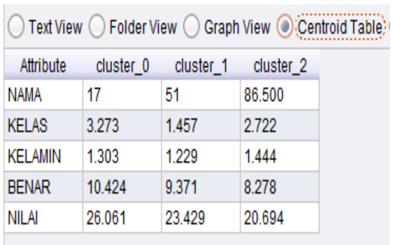
In Cluster Model (Clustering), it can be seen that the number of the 3 clusters is cluster 0 with 33 items. Cluster 1 has 35 items and Cluster 2 has 36 items. The total data is 104 items

Picture 1 Cluster Model Display (Text View)



Pada centroid table dapat kita lihat bahwa cluster 0 merupakan kategori nilai tinggi, cluster 1 kategori nilai sedang, dan cluster 2 merupakan kategori nilai rendah. Hal ini apat dibaca pada centroid masing-masing cluster.

Picture 2 . Display of Centroid Table



Attribute	cluster_0	cluster_1	cluster_2
NAMA	17	51	86.500
KELAS	3.273	1.457	2.722
KELAMIN	1.303	1.229	1.444
BENAR	10.424	9.371	8.278
NILAI	26.061	23.429	20.694

CONCLUTION

From the discussion that has been stated, it can be concluded that:
K-means algorithm clustering method can be applied to the Analysis of Determination of TRY OUT Mathematics Ability in Mathematics Department students at SMAN. 1 Payakumbuh District. so this method is very helpful for the School in determining students who get High, Medium and Low Scores.

Submited : 03 Jul 2020 – Accepted : 11 Jul 2020 – Published : 12 Jul 2020

ISSN : 1979-9292

E-ISSN : 2460-5611

The results of the K-Means Clustering manual data with the Data Mining RapidMiner application get the same results ie cluster 0 = 33 people, cluster 1 = 35 people and cluster 2 = 36 people.
Based on 3 clusters that have been tested using RapidMiner that most students are in the second cluster which is classified as the highest value.
In the Analysis of the Determination of TRY OUT Mathematics Ability of Science Department Students At SMA N.1, Payakumbuh District, the highest score is in cluster 0 with a total of 33 students female students majoring in Natural Sciences. In Cluster 1, the average score was 35 students, female students majoring in Natural Sciences-1. In Cluster 2 the number of students was 36 students of female gender majoring in Natural Sciences.

REFERENCE

[1] Anita, Y. Muharmi, and S. T. Siska, "Penentuan Tingkat Minat Belanja Online Melalui Media Sosial Menggunakan Metode Clustering K-Means," vol. I, no. 2, pp. 126-134, 2018.

[2] Ediyanto, dkk. 2013. Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis. Buletin Ilmiah Mat, Stat, dan Terapannya (Bimaster). Volume 02, No.2(2013), hal 133-136.

[3] Handoyo, R, dkk. 2014. Perbandingan Metode Clustering Menggunakan Metode Single Linkage Dan K-Means Pada Pengelompokan Dokumen. ISSN. 1412-0100 Vol 15, No, 2, OKTOBER 2014.

[4] Ong, Oscar. 2013. Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University. Jurnal Ilmiah Teknik Industri. Vol. 12, No 1, 2010. ISSN 1412-6869.

[5] Sari, R. M., & Tasril, V. (2020). Prediksi Jumlah APBD Kota Payakumbuh dengan metode K-Means. Jurnal Ipteks Terapan, 14(1), 45-50.

[6] Sari, R. M. (2015). Prediksi Data Anggaran Pendapatan Belanja Daerah Menggunakan Algoritma K-Means. SATIN-Sains dan Teknologi Informasi, 1(2), 73-78.

[7] S. T. Siska, "ANALISA DAN PENERAPAN DATA MINING UNTUK MENENTUKAN KUBIKASI AIR TERJUAL BERDASARKAN PENGELOMPOKAN PELANGGAN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING," vol. 9, no. 1, pp. 86-93, 2016.

[8] S. T. Siska, "Sistem Informasi Pemasaran Perumahan dan Pembayaran Konsumen pada CV Mandiri Utama Cabang Payakumbuh Menggunakan Visual Basic 6.0," vol. I, no. 2, pp. 296-278, 2018.

Submited : 03 Jul 2020 – Accepted : 11 Jul 2020 – Published : 12 Jul 2020