K-MEANS CLUSTERING CALCULATION TO DETERMINE MAINSTREAM DOMINATION OF COURSES

Dewi Anggraeni[1*], Rizaldi[1]

[1]Information System, Sekolah Tinggi Manajemen Informatika dan Komputer Royal
email: *anggraeni1987@gmail.com

Abstract : In Semester 6 students are required to choose mainstream elective courses at the STMIK Royal campus. In this study, researchers chose 2 mainstraim elective courses, namely Logic and Programming, and Database and Information Management. The purpose of choosing this elective course is for students to focus more on their majors. The process of selecting the data, researchers carry out an analysis process of student learning outcomes. This research data is semester 1 to semester 5 grades. All data on the value of student learning outcomes are divided into 2 parts, namely theory and practice and divided into the number of courses. For many uses, a data mining method using K-means cluster is needed. K-Means Clustering is a data analysis method or Data Mining method that performs an unsupervised learning modeling process and uses methods that group data from various partitions. The results of the k means clutering grouping with 2 clusters, namely cluster 0 of the mainstream group of logic and programming courses with a total of 28 students, and cluster 1 of the mainstream group of informatics database and management courses with a total of 72 students with a total of 100 students.

Keywords: data; K-means; student

Abstrak : Mahasiswa semester 6 diwajibkan memilih mainstream matakuliah pilihan yang ada di kampus STMIK Royal. Pada penelitian ini, peneliti memilih 2 mainstraim matakuliah pilihan yaitu Logika dan Pemprograman, serta Basis Data dan Manajemen Informatika. Tujuan pemilihan matakuliah pilihan ini adalah agar mahasiswa lebih terfokus pada jurusannya. Proses pemilihan datanya, peneliti melakukan proses analisa dari hasil belajar mahasiswa. Data penelitian ini nilai semester 1 sampai semester 5. Seluruh data nilai hasil belajar mahasiswa dibagi menjadi 2 bagian yaitu teori dan praktek serta dibagi jumlah banyaknya matakuliah. untuk penggunaan yang banyak, maka dperlukan metode data mining dengan menggunakan k-means cluster. K-Means Clustering adalah metode analisis data atau metode Data Mining yang melakukan proses pemodelan pembelajaran tanpa pengawasan dan menggunakan metode yang mengelompokkan data dari berbagai partisi. Hasil pengelompokan k means clutering dengan 2 cluster yaitu cluster 0 kelompok mainstream matakuliah logika dan pemprograman dengan jumlah mahasiswa 28 orang, dan cluster 1 kelompok mainstream matakuliah basis data dan manajemen informatika dengan jumlah mahasiswa 72 orang dengan total keseluruhan 100 mahasiswa.

Keyword : data; K-means; mahasiswa;

INTRODUCTION

Purpose of determining maintream This course is so that students can know which courses to take based on assessment during lectures. As well as directing students to the output that expected by STMIK Royal graduates.

Elective courses are non-compulsory courses but must be taken by students, whether in accordance with the bi-dang of interest or not. Matriculation courses are courses that must be taken First as a requirement for strengthening the courses to be taken. K-Means Clustering method is by grouping data based on age and ability (reading, writing, counting, and memorization) [1], Grouping analysis using the K-Means method. Based on the experiment, the grouping of TA research areas was carried out with 3 clusters (C), namely C1 is a few, consisting of 1 MKW; C2 is medium, 6 MKW members; and C3 is a lot, 3 MKW members have been acquired [2]. There are several factors that influence the prediction of student graduation in accordance with the time of study, including: average last GPA, number of credits, activeness in organizations, scholarships, and regional origin [3].

Disruption data clustering model designed with the k-means algorithm method, this application model can display an overview and show the pattern of distribution of customer complaint data [4]. The clustering method uses algorithm-ma k- means in grouping data on prospective new students at the University of Muhammadiyah Yogyakarta. The k-means cluster analysis method can be a solution for classifying the characteristics of objects [5]. K-Means Clustering to map oil palm plantations at PT Surya Intisari Raya which applies Davies Bouldin

Index with the RapidMiner tool as an evaluation of the optimal number of clusters [6]. Grouping data on Covid19 cases in Indonesian provinces using clustering techniques using the K-mean algorithm [7]. clustering graduate student data with attributes of address, major, and GPA into three clusters based on distance (Euclidean). The data processed is graduate data for 2016-2017[8]. Stock management that is carried out inaccurately and carelessly will cause high and uneconomical storage costs, because there can be vacancies or excess products of course. This will certainly greatly excite all business actors as well as online shops [9]. The study program is a unity of educational and learning activities that has a curriculum and learning methods [10].

In this study, researchers only used 2 mainstream elective courses, namely Logic and Programming, and Database and Information Management. The mainstream of this course is taken for students when students enter semester 6.
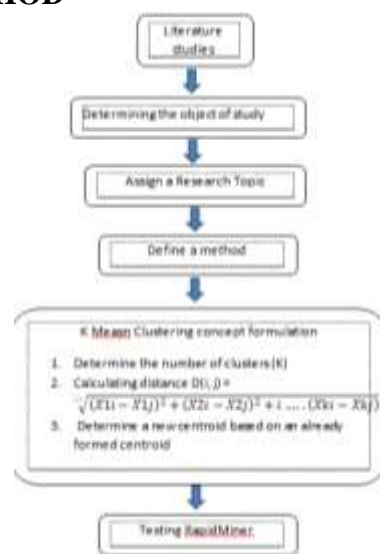
METHOD

Image 1. State Of The Art

In the early stages of this study, researchers first look for several references regarding the object to be studied. Researchers get these references from various sources such as journals, books, and other print media.

After that the researcher determines the object to be studied. The object of this study is STMIK Royal Students. After determining the object to be studied, the next stage is to determine the topic to be studied, where in this research topic, is to group the mainstream student courses based on semester 1 to semester 5 grades. . Next is to determine the method that is in accordance with the object studied, in this study, we will group linear course data with mainstream elective courses. The data that will be used is as many as 100 students and 24 courses that are aligned with the mainstream. With so much data used, we need a method for mainstream grouping of elective courses. The appropriate method with the object under study is measn clustering. The appropriate method with the object under study is measn clustering, where, the concept of k means clustering is to search and group data. After the data determines the data to be carried out for training, where this training data there are 24 courses and 100 students. . Implementation of k measn clustering calculation trial using rapid miner 5 application

RESULTS AND DISCUSSION

The data used is STMIK Royal Student data in semester 1 to semester 5 as many as 100 students. Determination of the number of clusters is done dynamically (according to the wishes of the user), below is a display of determining the number of centroids (classes). The first

step taken after determining the number of clusters is to determine the initial centroid or ten-gah point of each cluster in determining student data classes. The method of determining the initial centroid or midpoint of each cluster is to select one document from the entire data set used in the study that is randomly selected.

Researchers determined the number of clusters as many as 2 clusters. The cluster consists of programming logic and Database and Information Management. For clusters, the initial centroid programming logic group used is C1, while the initial centroid Database and Information Management cluster used is C2 as it is next.

Perform the calculation process to find the closest distance using the formula.

$$D(i, j) = \sqrt{\begin{array}{c}(X1i - X1j)^2 + (X2i - X2j)^2 \\ +i \ldots \ldots (Xki - Xkj)^2\end{array}} \quad (1)$$

Where :
$D(i,j)$ = Distance of data i to center of cluster j
Xki = Data to i on attribute data to k
Xkj = distance between center points to j on attribute k.

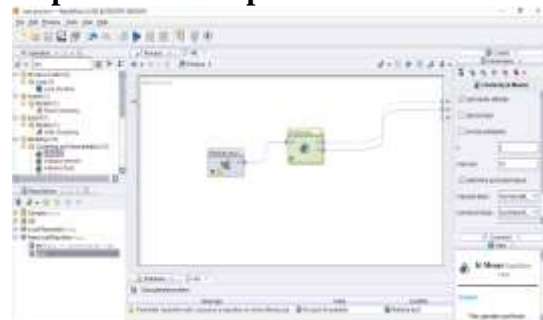### Implenemtasi Rapid Miner 5

Image 2. Student Mainsteram Implementation.

In Image 2 we can see the need for operators that we will use to manage the mainstream of the course. first we need a Retrieve that serves as a place for the data we will input. In this research, researchers used student grade data from semester 1 to semester 5 which were stored in Excel format. Researchers set the desired number of clusters as many as 2 clusters, namely logic and programming courses and Database and Information Management.Cluster Model
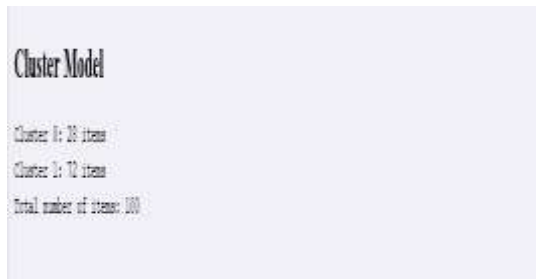

Image 3. Cluster Model

In Image 3, it is explained that the number of data or items that we input is 100 students, cluster 0 student groups who choose logic and programming courses with a total of 28 people, cluster 1 student groups who choose database and informatics management courses as many as 72 people.

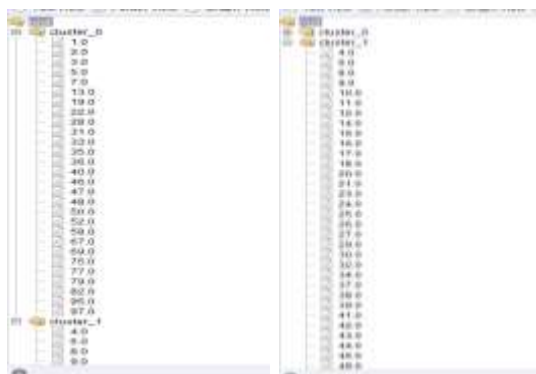Image 4 describes the appearance of each item in the cluster.

Image 4. Folder View

Displays the number of centroids with a total of 24 courses.

Image 5. Centroid Table

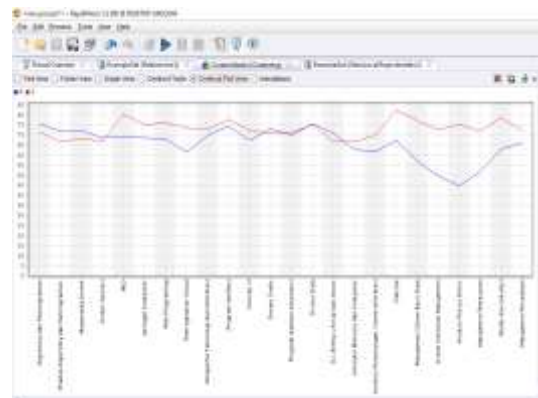Display the shape of the course mainsteram graph.

Image 6. Centroid Plot

CONCLUSION

Grouping mainstream elective courses using the k means clustering method in the rapidminer application with a total of 100 students and predetermined clusters, namely 2 clusters including cluster 0 mainstream logic and programming groups with a total of 28 students, and cluster 1 mainstream group of databases and informatics management with a total of 72 students, with a

total of 24 courses that are liner with mainstream.

BIBLIOGRAPHY

[1] L. A. Shofiana and L. A. Muharom, "Aplikasi K-Means Untuk Pengelompokan Siswa Pendidikan Anak Usia Dini (Paud) Berdasarkan Data Siswa Di Kecamatan Ledokombo," no. 1110651055, pp. 0–2, 2009.

[2] H. Haviluddin, S. J. Patandianan, G. M. Putra, N. Puspitasari, and H. S. Pakpahan, "Implementasi Metode K-Means Untuk Pengelompokkan Rekomendasi Tugas Akhir," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 16, no. 1, p. 13, 2021, doi: 10.30872/jim.v16i1.5182.

[3] A. R. Jannah, D. Arifianto, and M. Kom, "Penerapan Metode Clustering dengan Algoritma K-Means untuk Prediksi Kelulusan Mahasiswa Jurusan Teknik Informatika di Universitas Muhammadiyah Jember," *J. Manaj. Sist. Inf. dan Teknol.*, vol. 1, no. 1210651237, pp. 1–10, 2015.

[4] D. Hartanti, D. Mining, A. K-means, and D. Mining, "Model Clustering Menggunakan Algoritma K-Means Pada Data Keluhan Pelanggan Pt . Pln (Studi Kasus : Pt . Pln (Persero) Distribusi Jakarta Dan Tangerang)," *Model Clust. Menggunakan Algoritm. K-Means Pada Data Keluhan Pelangg. Pt . Pln (Stud. Kasus Pt . Pln (Persero) Distrib. Jakarta Dan*, vol. 4, p. 119, 2015.

[5] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018, doi: 10.18196/st.211211.

[6] A. Al Masykur, S. K. Gusti, S. Sanjaya, F. Yanto, and F. Syafria, "Penerapan Metode K-Means Clustering untuk Pemetaan Pengelompokan Lahan Produksi Tandan Buah Segar," *J. Inform.*, vol. 10, no. 1, 2023, doi: 10.31294/inf.v10i1.15621.

[7] N. Sari, H. H. Handayani, and A. M. Siregar, "Implementasi Clustering Data Kasus Covid 19 Di Indonesia Menggunakan Algoritma K-Means," *Bianglala Inform. J. Komput. dan Inform.*, vol. 11, no. 1, pp. 7–12, 2023, [Online]. Available: https://ejournal.bsi.ac.id/ejurnal/index.php/Bianglala/article/view/14762.

[8] R. Sibarani and O. Omby, "Algorithma K-Means Clustering Strategi Pemasaran Penerimaan Mahasisswa Baru Universitas Satya Negara Indonesia," *J. Algoritm. Log. dan Komputasi*, vol. 1, no. 2, pp. 44–50, 2018, doi: 10.30813/j-alu.v1i2.1367.

[9] A. Wicaksana and T. Rachman, " 済無No Title No Title No Title," *Angew. Chemie Int. Ed. 6(11), 951–952.*, vol. 3, no. 1, pp. 10–27, 2018, [Online]. Available: https://medium.com/@arifwicaksa

naa/pengertian-use-case-a7e576e1b6bf.

[10] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.