Nusantara Jurnal of Computer Application (2022), 1(1): 1-6
DOI: 10.47679/njca.v1i1.1
https://ukinstitute.org/journals/3/njca

RESEARCH ARTICLE

# Application of K-Means Clustering Method in Semester I Value Data Clustering Until Semester IV

Mardin Aprianto [a,1*]; Liza Yulianti [a,2]; Dewi Suranti [a,3]

Published online:1 February 2022

**Abstract**

In the Informatics Study Program there is a concentration which is divided into 2 parts which students will choose when entering semester V. The problem that often occurs is that some students sometimes choose concentration instead of seeing the results of the scores obtained during semesters I to IV, and there are some who choose majors because he followed his friends. This is what causes students to often be unprepared at the end of the semester. Therefore, it is necessary to analyze the value of semester I to semester IV obtained by students to help recommend to students which concentration to choose. The analysis is carried out by applying the K-Means Clustering Method, which will produce 2 groups, namely Software Engineering and Network Infrastructure. The K-Means Clustering method has the ability to group large amounts of data with relatively fast and efficient computation time. The Semester I to Semester IV Value Data Grouping application was made using Visual Basic .Net programming language and SQL Server 2008 database by applying one of the data mining methods used was K-Means Clustering. The grouping is done based on student score data obtained from Semester I to Semester IV (data attached). From this data, grouping is done into 2 groups, namely Cluster C1 Software Engineering and Cluster C2 Network Infrastructure. With this application, it can assist the Study Program in providing recommendations and also material for consideration to students in choosing a major whether Software Engineering or Network Infrastructure. Based on the results of the tests that have been carried out, the application of the K-Means Clustering Method in Grouping Semester Value Data has been successfully carried out, and can provide information based on 2 groups, namely Cluster C1 (Software Engineering) and Cluster C2 (Network Infrastructure), and the functionality of the application has been running. as expected.

**Keyword :** Application; K-Means Clustering; Semester Value Data

## Introduction

Currently, we are in an era that is full of communication and information technology. Advances in technology have provided a very broad source of information and communication from what humans already have. Although the role of information in recent decades has received less attention, the need for information and communication is no less important than the need for human food and clothing.

Dehasen Bengkulu University (UNIVED) is one of the private universities in Bengkulu Province. UNIVED already has 9 (nine) Faculties, one of which is the Faculty of Computer Science. The Faculty of Computer Science has three study programs, namely Informatics, Information Systems, and Computer Systems. The Informatics Study Program is a study program with the most demand.

In the Informatics Study Program there is a concentration which is divided into 2 parts which students will choose when entering semester V. The problem that often occurs is that some students sometimes choose concentration instead of seeing the results of the scores obtained during semesters I to IV, and there are some who choose majors because he followed his friends. This is what causes students to often be unprepared at the end of the semester.

[1-3] Information Study Program, Faculty of Computer Science, Dehasen University, Bengkulu

*) corresponding author

Mardin Aprianto

Information Study Program, Faculty Of Computer Science, Dehasen University, Bengkulu

## Ease of Use

### *A. Data Mining*

Data mining is the process of finding meaningful relationships, patterns and trends by examining large data in storage using pattern recognition technology, such as statistics and mathematics. This is in accordance with the definition put forward by the Gartner Group in Larose (2005:2) quoted by Abdurrahman, 2016:72, namely "Data mining is the process of discovering meaningful new correlations, patterns and trends by shifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques".

Data mining is an activity to find interesting patterns from large amounts of data, data can be stored in databases, data warehouses, or other information storage (Asriningtias, 2014: 837).

### *B. Algoritma K-Means Clustering*

Dalam statistik dan mesin pembelajaran, pengelompokan K-Means merupakan metode analisa kelompok yang mengarah pada pemartisian N objek pengamatan ke dalam K kelompok (cluster) dimana setiap objek pengamatan dimiliki oleh sebuah kelompok dengan mean (rata-rata) terdekat, mirip dengan algoritma Expectation-Maximazition untuk Gaussian Mixture dimana keduanya mencoba untuk menemukan pusat dari kelompok dalam data sebanyak iterasi perbaikan yang dilakukan oleh kedua algoritma (Prasetyo, 2012:178).

Algoritma K-means adalah algoritma yang terbaik dalam algoritma partitional clustering dan yang paling sering digunakan diantara algoritma clustering lainnya karena kesederhanaan dan efisiensinya. Kelebihan Algoritma K-means itu sendiri menurut K. Arai and A. R. Barakbah, merupakan algoritma clustering yang paling sederhana dan umum, hal ini dikarenakan K-means mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien. Namun, K-means mempunyai kelemahan yang diakibatkan oleh penentuan pusat awal cluster (Sulastri, 2017:301).

### *C. Visual Studio 2010*

Microsoft Visual Basic 2010 Express is a part of the Microsoft Visual Studio 2010 Express Family. A free tool used by windows developers of various levels to develop and build applications that run on the .NET Framework system, using the BASIC language. Visual Basic provides a fast and easy way to create applications (Aswan, 2012:20).

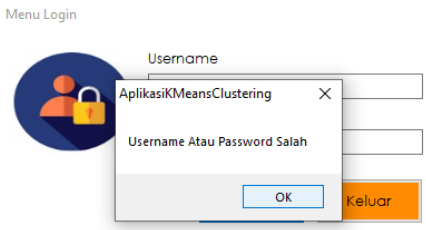
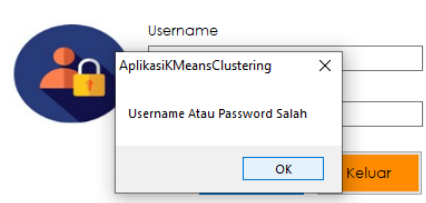
## Prepare Your Paper Before Styling

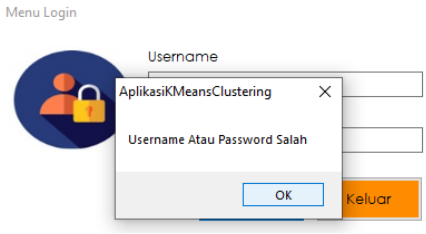
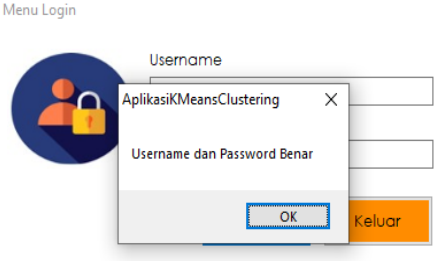
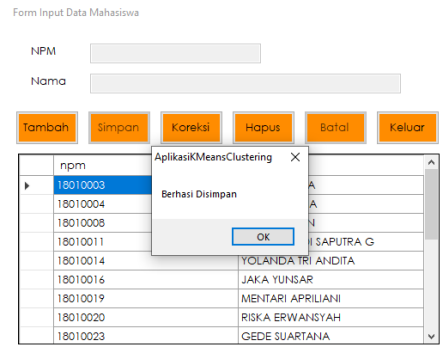
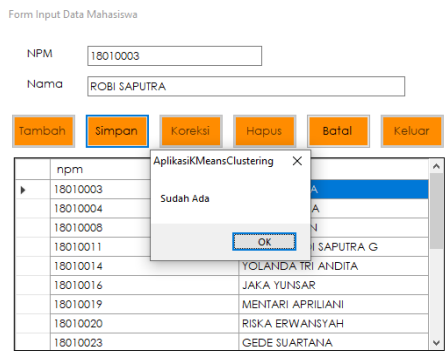
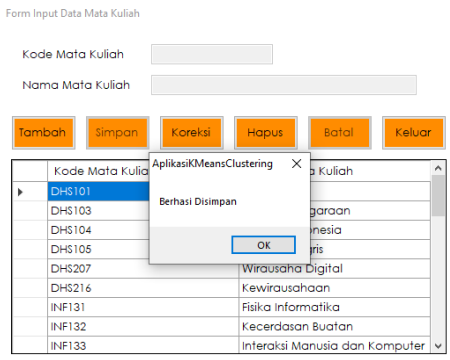
The research method applied in this research is the development of the waterfall method. Data collection methods are used to obtain data that can support the problems to be discussed.

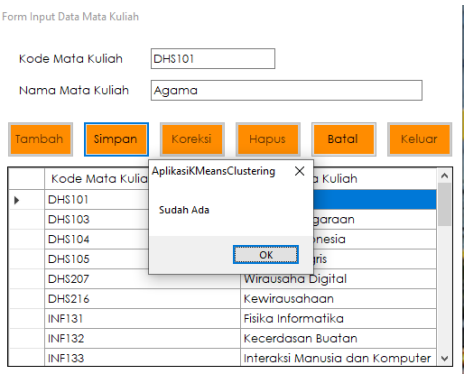
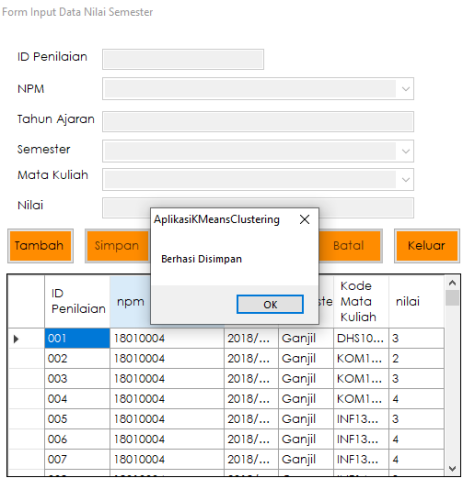
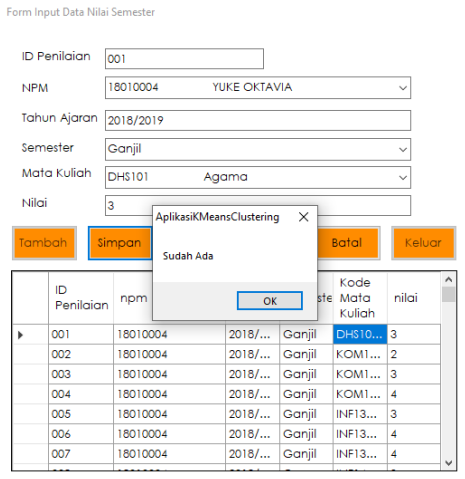
## Using the Template

Blackbox testing (blackbox testing) is one of the software testing methods that focuses on the functionality side, especially on the input and output of the Semester I to Semester IV Value Data Grouping Applications. The results of the black box testing that have been carried out are shown in table 1.

**Table I. Blackbox Test Results**

| Test Form | Tested Components | Black Box Test Results | Conclusion |
|---|---|---|---|
| Login | the username and password fields are left blank | system deny login access   | succeed |
| | fill in the wrong username | system deny login access   | succeed |

| | | | |
|---|---|---|---|
| | fill in the wrong password | system deny login access<br>  | succeed |
| | fill in the correct username and password | system deny login access<br>  | succeed |
| Student | save different student data | the system has successfully saved the data<br>  | succeed |
| | save the same student data | the system refuses access to save the data<br>  | succeed |
| Subject | save data for different courses | The system has successfully saved the course data<br>  | succeed |

| | | | |
|---|---|---|---|
| | save the same course data | the system refuses access to save the data<br>  | succeed |
| Semester Grades | save different semester value data | The system has successfully saved semester grade data<br>  | succeed |
| | save the same semester value data | the system refuses access to save the data<br>  | succeed |

| | | | |
|---|---|---|---|
| Clustering Analysis | carry out the process of clustering student value data into 2 groups | the system successfully runs the clustering process and displays the results of 2 groups  | succeed |

Based on the results of the tests that have been carried out, the application of the K-Means Clustering Method in Grouping Semester Value Data has been successfully carried out, and can provide information based on 2 groups, namely Cluster C1 (Software Engineering) and Cluster C2 (Network Infrastructure), and the functionality of the application has been running. as expected.

## Acknowledgment

1. Applications for Classification of Semester I to Semester IV Data are made using Visual Basic .Net programming language and SQL Server 2008 database by applying one of the data mining methods used is K-Means Clustering.
2. Grouping is done based on student score data obtained from Semester I to Semester IV (data attached). From this data, grouping is done into 2 groups, namely Cluster C1 Software Engineering and Cluster C2 Network Infrastructure.
3. With this application, it can assist the Study Program in providing recommendations and also material for consideration to students in choosing a major whether Software Engineering or Network Infrastructure.
4. Based on the results of the tests that have been carried out, the application of the K-Means Clustering Method in Grouping Semester Value Data has been successfully carried out, and can provide information based on 2 groups, namely Cluster C1 (Software Engineering) and Cluster C2 (Network Infrastructure), as well as the functionality of the application. has worked as expected.

## Supplementary Material

Supplementary material that may be helpful in the review process should be prepared and provided as a separate electronic file. That file can then be transformed into PDF format and submitted along with the manuscript and graphic files to the appropriate editorial office.

## References

[1] Abdurrahman, Ginanjar. 2016. Clustering Data Ujian Tengah Semester (UTS) Data Mining Menggunakan Algoritma K-Means. Jurnal Sistem dan Teknologi Informasi Indonesia Vol.1 No.2 Agustus 2016.

[2] Asriningtias, Yuli. Mardhiyah, Rodhyah. 2014. Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa. Jurnal Informatika Vol.8 No.1 Januari 2014.

[3] Asroni. Adrian, Ronald. 2015. Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang. Jurnal Ilmiah Semesta Teknika Vol.18. No.1 Mei 2015.

[4] Aswan. 2012. Kumpulan Program Kreatif Dengan Visual Basic .NET. Penerbit Informatika: Bandung.

[5] Lubis, Adyanata. 2016. Basis Data Dasar Untuk Mahasiswa Ilmu Komputer. Penerbit Deepublish: Yogyakarta.

[6] Merliana, Ni Putu Eka, dkk. 2015. Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering. Prosiding Seminar Nasional Multi Disiplin Ilmu & Call for Papers UNISBANK (SENDI_U). ISSN: 978-979-3649-81-8.

[7] Novanto, Yusak. 2015. Faktor-faktor Yang Berkaitan Dengan Prestasi Akademik Mahasiswa Sekolah Tinggi Teologi X. https://www.researchgate.net /publication/283352540

[8] Prasetyo, Eko. 2012. Data Mining Konsep dan Aplikasi Menggunakan Matlab. Penerbit Andi: Yogyakarta.

[9] Putri, Nisya Aldilla Hariza. Desrianty, Arie. Yuniar. 2014. Strategi Peningkatan Prestasi Akademik Mahasiswa Berdasarkan Variabel-Variabel Yang Mempengaruhinya. Jurnal Reka Integra Vol.02. No.1 Juli 2014. ISSN 2338-5081.

[10] Sudharmono, Neil Casaandra. Ayub, Mewati. 2015. Analisis Prestasi Akademik Mahasiswa Yang Mengikuti Kegiatan Kemahasiswaan. Jurnal Teknik Informatika dan Sistem Informasi Vol.1 No.2 Agustus 2015. e-ISSN: 2443-2229.

[11] Sulastri, Heni. Gufroni, Acep Irham. 2017. Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia. Jurnal Teknologi dan Sistem Informasi Vol.3 No.2 2017. ISSN 2476-8812.

[12] Sutabri, Tata. 2012. Analisis Sistem Informasi. Penerbit Andi: Yogyakarta.