

ON-LINE CLUSTERING OF LECTURERS PERFORMANCE OF COMPUTER SCIENCE DEPARTMENT OF SEMARANG STATE UNIVERSITY USING K-MEANS ALGORITHM

ENDANG SUGIHARTI¹, MUCH. AZIZ MUSLIM²

¹Lecturer of Computer Science Department of Semarang State University, Semarang Indonesia

²Lecturer of Computer Science Department of Semarang State University, Semarang Indonesia

E-mail: ¹endanghs02@yahoo.com, ²a212muslim@yahoo.com

ABSTRACT

Lecturer is one of the essential components of an educational system in universities. Quality of lecturer performance can be reflected in the productivity and quality of the implementation of the Three Responsibilities that include activities in education, research, community services, and other supporting activities. Clustering of lecturer performance is required in order to gain a knowledge (knowledge discovery) of the pattern of the Three Responsibilities implementation, in this case in the Department of Computer Science Unnes. Clustering is also necessary to control the balance of lecturer performance in implementing the Three Responsibilities. This is why the research was conducted. The problem is how to create an innovative program in the form of Clustering On-line of Computer Science Department for lecturer performance based on the implementation of the Three Responsibilities using of K-Means Algorithm? The objective of this institution research in order to obtain on-line information system for clustering of Computer Science Department for lecturer performance using the K-Means Algorithm. To resolve this problem, researchers divided into two stages of completion. The first stage was literature review. The second stage was making a clustering program for Lecturer performance of Department of Computer Science, based-on the implementation of College three responsibilities. The works sequence were: (1) Data Collection, (2) Building a data mining system, (3) Data Processing, (4) Conducting the clustering process with *K-Means algorithm*. Analysis of data/results included: (1) data reduction; (2) data display; and (3) conclusion (verification). As the results of this research, now the Department of Computer Science Unnes already has a clustering program that is capable of classifying and monitors all activities of lecturers performance in accordance with the implementation of College three responsibilities.

Keywords: *Clustering, K-Means Algorithm, Lecturer performance, College three responsibilities*

1. INTRODUCTION

Lecturer is one of the essential components of an educational system in universities. Role, duties and responsibilities of the lecturers are very important in achieving national education goals, namely the intellectual life of the nation, improve the quality of Indonesian human, which includes the quality of faith/piety, noble character, and mastery of science, technology, and art, and to realize Indonesian society that advanced, just, prosperous, and civilized.

In Act No. 37 of 2009 on Lecturers, lecturers expressed as professional educators and scientists

with the main task of transforming, develop and disseminate science, technology, and art through education, research, and community service. The main task of the lecturer is to carry out three responsibilities with a workload of at least commensurate with 12 credits and a maximum of 16 credits in each semester according to the academic qualification.

Quality of lecturers performance can be reflected in the productivity and quality in implementing the three responsibilities. Directorate General for Higher Education has imposed program of lecturers performance load and the evaluation of the three responsibilities (known as BKD and EPT-PT) to ensure the implementation of the lecturer

tasks running accordance with the criteria set out in the legislation.

Nowadays, one of the development trends of information technology (IT) is Data Mining which is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from a variety of large databases [1]. Data mining, often referred to Knowledge Discovery in Databases (KDD). Delavari[2], in a journal, revealed several examples of the application of data mining in higher education institutions. Explained that data mining can be used to evaluate the students assessment, lecturer assessment, course assessment, industrial training assessment, evaluation and registration students.

Clustering is used to analyze the different grouping of data, similar to the classification, but the grouping has not been defined before data mining tool is executed. Clustering divides the data into groups based on those found by data mining tool. The principle of clustering is to maximize the similarity between members of the classes and minimize the similarity between clusters. Clustering can be performed on the data that has some attributes that are mapped as a multidimensional space. In short, clustering try to find a component of the group naturally, based on the proximity of data. According to Shovon and Haque[5], Data Clustering is unsupervised and statistical data analysis technique. It is used to classify the same data into a homogeneous group. It is used to operate on a large data-set to discover hidden pattern and relationship helps to make decision quickly and efficiently.

In an article published in Elsevier, Jain [6] revealed that organizing data into a cluster is a model that is most fundamental to understand and learn. Cluster analysis is a formal study for grouping, or clustering of objects according to the characteristics which measured based on similarity to one another. Clustering is grouping using unsupervised learning techniques which do not require training on these methods, or in other words, there is no learning phase, and does not use labeling on each group.

At the same journal also expressed regarding the use of K-means as an algorithm which always even until now reliable enough to use for the clustering process. In the development of K-means clustering has also gone through several modifications, among others Fuzzy C-Means. However, the journal "Performance Evaluation of K-means and Fuzzy C-Means clustering Algorithms for Statistical Distributions of Input Data Points" [3] revealed that through a series of experiments it was concluded that the performance of the K-means algorithm is better than the algorithm of fuzzy C-means.

K-means algorithm is an algorithm that is best in partitioning clustering algorithms and the most commonly used among the clustering algorithms because of its simplicity and efficiency [4]. The basic concept of the K-means algorithm is a search of cluster center (centroid points) iteratively. Cluster centers are set based on the distance of each data to the cluster center. Clustering process begins by identifying the data to be clustered, x_{ij} ($i = 1, \dots, n; j = 1, \dots, m$) and n is the number of data to be clustered, and m is the number of variables. K-means clustering algorithm is described as figure 1.

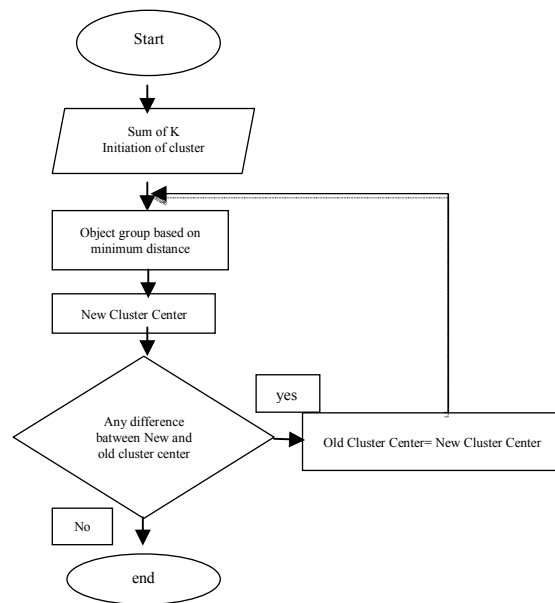


Figure 1. K-Means Algorithm

On the other hand, the Department of Computer Science Unnes trying to find a variety of new breakthroughs based on science and technology to develop an on-line program. One way is to design a

system development program in the form of on-line information to be able to perform clustering of lecturers performance based on the three responsibilities using the K-Means

Algorithm. Clustering of lecturers performance is required in order to gain a knowledge (knowledge discovery) of the pattern of three responsibilities implementation at college, in this case in the Department of Computer Science. It is necessary to control the balance of lecturers performance in implementing the three responsibilities. These results are expected to be used and developed by the institution to make the right policy and directed in accordance with the conditions and needs of the college in achieving its vision and mission. This is one reason why this institutional research was done.

Based on the above background then the formulated problems are how to set the pattern of three responsibilities implementation for lecturers of Computer Science Department by using the K-Means Algorithm, how clustering of lecturer performance based on implementation of three responsibilities using the K-Means Algorithm, and how does the development of new systems in the form of on-line information to be able to do clustering of lecturer performance based on three responsibilities using the K-Means Algorithm.

The purpose of this study were getting a description of the pattern of the three responsibilities implementation by the lecturer of the Computer Science Department using the K-Means Algorithm, clustering of Lecturer performance based on the three responsibilities implementation, and obtaining a new on-line information system for clustering of lecturer performance of Computer Science Department of Unnes by using K-Means Algorithm.

2. RESEARCH METHODS

Research was conducted at the Computer Laboratory of Computer Science Department Semarang State University. Selection of the location because it has devices that support for this research.

The clustering using K-Means algorithm is used to cluster the lecturer performance based on implementation of three responsibilities of college. Stages of research are described in Figure 2 with the following explanation.

- 1) Stage of data collection. The data collected is the data relating to process of lecturers

performance and data relating to the evaluation of lecturers performance.

- 2) Stage of data mining systems building. In this step, the researchers process information system development include database development, application development, testing of applications, and implementation of applications.
- 3) Stage of Data Processing. Once the system is built, the data obtained on the collection of data is processed on a system that has been built.
- 4) Conducting the process with K-Means clustering. This process is performed on the data that has been entered to the system. The system is given the K-Means Algorithm to be able to do the clustering process.
- 5) Analysis of the results. This process is carried out to determine the pattern of lecturer performance clustering.
- 6) Documentation.

The data collected in this research is in the form of facts about the settlement process of extracting clustering data of Computer Science Department lecturer performance by using the K-Means Algorithm. Regarding to it, the collection of data in this study used a thorough documentation and observation when the research activities were carried out. This data collection process using the guidance of data documentation at each stage of research conducted. How that is done is by recorded through the audio-visual.

Moleong [7] describes the definition of data analysis as a process to organize and sort the data into patterns, categories, and a unit of basic description so that it can be found a theme and working hypothesis can be formulated as suggested by the data.

Based on the understanding of the above data analysis, then the data from observations at each research step was narrated, identified related to clustering using K-Means Algorithm. While test of data clustering was documented and categorization was made.

Data in this research were analyzed by using a technical of data analysis of model of Miles and Huberman [8], which includes data reduction, data exposure/categorization (data display), and conclusion (verification).

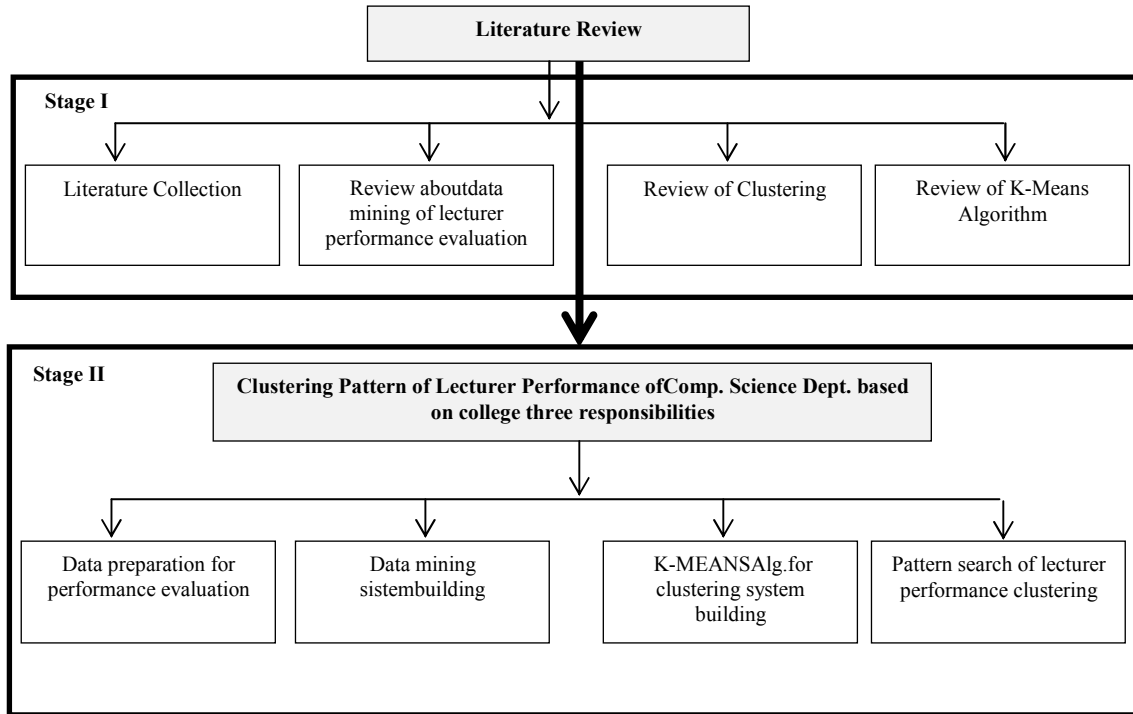


Figure 2. Research Stages

3. RESULTS AND DISCUSSION Setting of Implementation Patterns

For lecturers, the three responsibilities of college are included to its function as a lecturer. The three components (Teaching, Research, and Community Service) can be carried out in synergy between lecturers and students to improve their knowledge and optimize the implementation on the ground in order to reach the expected goals. Optimal results of the implementation of the three responsibilities of college can be synergized in an activity of expertise field group (KBK). The role of the KBK can be used to enhance the conducive academic atmosphere to lecturer and students, which in turn can be used as a spearhead for the development of institutions based on the three responsibilities of college.

clustering of the Department of Computer Science, Semarang State University. As base to do the clustering is based on research results. From this

Expertise field group (KBK) is a research group and working groups - as a forum for lecturers - who have expertise and interest in the manufacture of the significance of each process and the results of science and technology that are beneficial to society. With the KBK, a lecturer in the KBK can discuss and perform research for the development of science.

Computer Science Department, Semarang State University until 2015 had formed three expertise field groups namely Software Engineering, Networking, and E-Learning. Each lecturer is distributed in accordance with the expertise to get in on one of these three expertise fields. These rules serve as a basis for lecturer

process can be described that clustering formed as illustrated in the following figure.

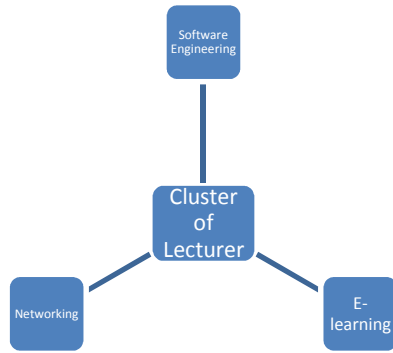


Figure 3. Setting of Pattern clustering.

Related to the method used, namely the K-Means Algorithm, then the number of clusters determined are three clusters. Lecturers formed clusters are clusters of software engineering, network cluster, and the cluster of e-learning. These clusters that are later used as a consideration in the process of K-Means Algorithm.

Lecturer Performance Clustering Process Using K-Means Algorithm

To make the process of lecturer performance clustering, the steps are as follows. (1) Stage of Data collection, is the initial stage after the

preparation phase is completed. The data collected is the data related to the provision of lecturer performance and data related to the evaluation of lecturer performance. (2) Stage of data mining system building. In this step, the researchers process the information system development including database development, application development, testing of applications, and implementation of applications. (3) Data Processing Stage. Once the system is built, the data obtained on the collection of data is processed on a system that has been built. (4) To process the K-Means clustering. This process is performed on the data that has been entered to the system. The system is given Algorithm K-Means so that be able to do the clustering process.

Development of Lecturer clustering Systems with K-Means Algorithm

The development of Lecturer clustering system with K-Means algorithm using the steps of extreme programming (XP). Software development cycle that adopting this method will pass through five phases consisting of the exploration phase, planning phase, iterations to release phase, productionizing phase, maintenance phase and death phase. Stages of XP method is described in more detail in the following diagram.

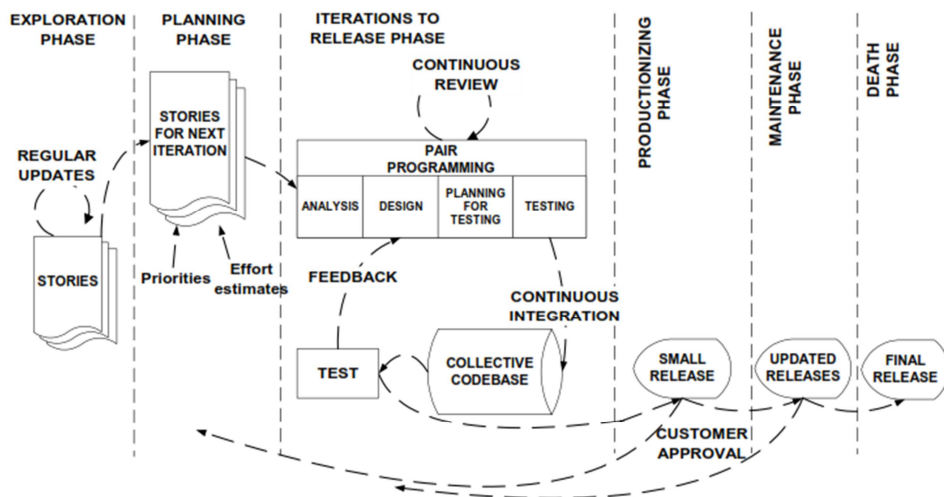


Figure 4. Process of Extreme Programming

Exploration Phase is a stage for the users in the software development process to formulate and write down the needs of the most fundamental of the system called the user story.

Planning phase is a continuation of the previous phase which used for sorting of a number of existing systems needs. The most important needs and most describe the system is chosen and then the construction is begun.

The next phase is the iteration release phase. In this phase there are four stages to be repeated several times to complete the development of the system. Repetition of the existing measures covering analysis, design, test planning, and test.

The next phase is the stage of doing the first release of the software that have been produced. During the release of course there will be criticism and suggestions to improve the software.

The last phase is the final release. Final release is done if the system requirements have been met and are functioning as required. The next development phase can be done by starting from the start.

Clustering Process

In the clustering process carried out the following steps:

1) ETL Process of Lecturer Data

ETL processes (extract transfer and load) is performed to perform extraction of existing lecturer data. The next step is the process of transfer to the new system.

2) Selection of Stop Word

Stop word is used to eliminate words that declared as a word which has no meaning in clustering.

3) The process of keywords selection related to KBK

KBK to perform clustering is used as the basis for the determination of the lecturers in the field of clustering. At this stage, according to the KBK owned by Computer Science Department, there are 3 clustering used namely software engineering, networking, and e-learning .

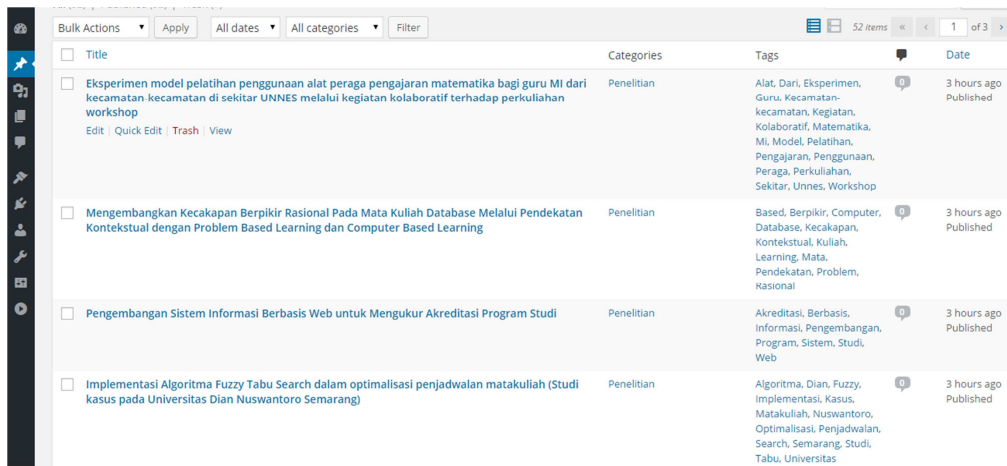
4) The process of data processing with stop word and keyword

This stage includes the following process

a) Deleting the word on the research data that is associated with a stop word/words that are not useful.

b) Perform calculations for keywords related to the research.

Figure 5 below shows the results of data transfer.



Title	Categories	Tags	Date
Eksperimen model pelatihan penggunaan alat peraga pengajaran matematika bagi guru MI dari kecamatan-kecamatan di sekitar UNNES melalui kegiatan kolaboratif terhadap perkuliahan workshop	Penelitian	Alat, Dari, Eksperimen, Guru, Kecamatan-kecamatan, Kegiatan, Kolaboratif, Matematika, MI, Model, Pelatihan, Pengajaran, Penggunaan, Peraga, Perkuliahan, Sekitar, Unnes, Workshop	3 hours ago Published
Mengembangkan Kecakapan Berpikir Rasional Pada Mata Kuliah Database Melalui Pendekatan Kontesktual dengan Problem Based Learning dan Computer Based Learning	Penelitian	Based, Berpikir, Computer, Database, Kecakapan, Kontesktual, Kuliah, Learning, Mata, Pendekatan, Problem, Rasional	3 hours ago Published
Pengembangan Sistem Informasi Berbasis Web untuk Mengukur Akreditasi Program Studi	Penelitian	Akreditasi, Berbasis, Informasi, Pengembangan, Program, Sistem, Studi, Web	3 hours ago Published
Implementasi Algoritma Fuzzy Tabu Search dalam optimalisasi penjadwalan matakuliah (Studi kasus pada Universitas Dian Nuswantoro Semarang)	Penelitian	Algoritma, Dian, Fuzzy, Implementasi, Kasus, Matakuliah, Nuswantoro, Optimalisasi, Penjadwalan, Search, Semarang, Studi, Tabu, Universitas	3 hours ago Published

Figure 5. Data transfer results

Keyword calculation process related to the research conducted is shown in Figure 6 below.

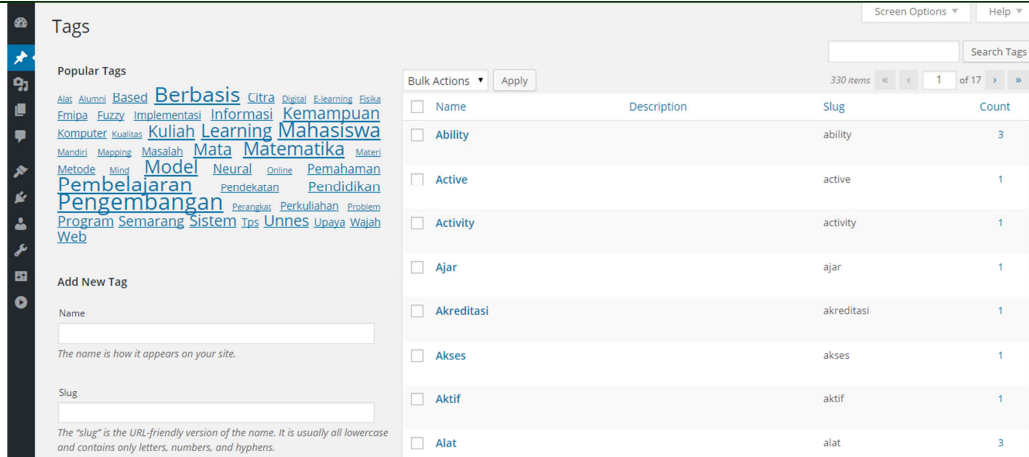


Figure 6. The Process Of Keyword Counting

After the stop word and keyword process, it is then performed clustering so that it can display the lecturer based on the existing KBK automatically.

After clustering, the display that will be generated is like Figure 7 below.

Lecturer Clustering, Computer Science Department

Cluster	Lecturer
Networking	Much Aziz Muslim
Software Engineering	Riza Arifudin
	Alamsyah
	Zaenal Abidin
Elearning	Isa Akhlis
	Endang Sugiharti

Figure 7. Results Of Lecturer Clustering Of Computer Science Department

The benefits of this clustering, then the quality of lecturer performance can be reflected in the productivity and quality of the implementation of the three responsibilities (education, research, community services, and other supporting activities) can be accessed at any time. Clustering program is also appropriate with the program of the Directorate General for Higher Education which has imposed program of lecturers performance load and the evaluation of the three responsibilities (known as BKD and EPT-PT) to ensure the implementation of the lecturer tasks running accordance with the criteria set out in the legislation.

In implementing the three responsibilities, each lecturer may differ from one another. There is a lecturer in implementing the three responsibilities more in education so that the field of research and community service became a little. Some are focused on areas of research while education and community service neglected. There is also a lecturer whose implementation of three responsibilities field is relatively balanced between each other.

On the other hand with this clustering program, Computer Science Department can find a variety of new ways to develop the on-line program. One way is to design a New Development System

program in the form of on-line information to be able to perform clustering of lecturer performance based on the three responsibilities using K-Means Algorithm. Other benefit, these results are expected to be used and developed by other departments in Unnes to create the right policy and directed.

4. CONCLUSION

Based on the results of research and discussion above, it can be concluded that the implementation pattern of the three responsibilities of Computer Science Department's lecturers, the expertise field groups namely Software Engineering, Networking, and E-Learning. Besides, the development of Lecturer clustering system with K-Means algorithm using the steps of extreme programming (XP), and the clustering process of lecturer performance based on implementation of the three responsibilities using the K-Means Algorithm has been completed, which includes the 5 phases of manufacture, namely: exploration phase, planning phase, iterations to release phase, productionizing phase, maintenance phase and death phase.

REFERENCES:

- [1] Turban, E., J.E. Aronson dan T.P. Liang. *Decision Support System and Intelligent Systems – 7th ed.* Pearson Education, Inc. Pearson Education, Inc. 2005.
- [2] Delavari, Naeimeh. Data Mining Application in Higher Learning. In *Informatics in Education*. VilniusUniversitetoLeidykla.2008.
- [3] Velmurugan, T. Performance Evaluation of K-Means and Fuzzy C-Means Clustering Algorithms for Statistical Distributions of Input Data Points. In *European Journal of Scientific Research*. EuroJournals Publishing.2010.
- [4] Liu, Bing. Top 10 Algorithms in Data Mining. In *Knowledge Information System*. Springer.2007.
- [5] Shovon, Md. Hedayetul Islam and Haque, Mahfuza. Prediction of Student Academic Performance by an Application of K-Means Clustering Algorithm. In *International Journal of Advanced Research in Computer Science and Software Engineering, Vol 2, Issue 7.2012*.

- [6] Jain, Anil K. Data Clustering: 50 Years Beyond K-Means. In *Pattern Recognition Letters*. 2009.
- [8] Miles, M. B., & Huberman, A. M.. *Qualitative Data Analysis an Expanded Sourcebook*. California: Sage Publishing, Inc.1994
- [9] Regulation of Indonesian Republic Government No 37 year 2009 about lecturers.