Decision Support System to Deciding Thesis Topic

Erna Daniati

Study Program of Information System

Universitas Nusantara PGRI Kediri

Kediri, Jawa Timur, Indonesia

ernadaniati@unpkediri.ac.id

Abstract—**By the end of semester, students are required to take thesis with certain subject. It is the final exam for students for their graduation in college. Taking this thesis must take some certain criteria such as the number of credits, GPA values, and compulsory courses that must be taken. At the Faculty of Engineering, Universitas Nusantara PGRI Kediri, there are several concentration courses to help students takes topic thesis. This course has several credits. However, things that happen in fact, students who take concentration courses are not appropriate in taking the topic of thesis even some feel confused in taking the thesis topic. One of expected solution is to provide decision support system to determine thesis topic for students. This system uses a combination of K-Means and SAW methods. The system initially clusters topic based on grade semester with K-Means. Then, it calculates the alternative value using the SAW method, so resulting alternative decisions. The system is expected to support student decisions in determining the topic of the thesis.**

Keywords—*Thesis; Decision Support System; Clustering; K-Means; SAW*

I. INTRODUCTION

Thesis is a scientific paper in the form of explanation writing of research results on bachelor students which discusses a problem in a particular field of science. The thesis is also a mandatory requirement to obtain a Bachelor's degree. However, in reality of the Information Systems of Engineering Faculty of UNP Kediri, there are many students who have confusion when preparing their thesis. Especially, they are confused, what group/topic they should take for their thesis.

They do not understand what kind of group or topic of thesis that suits their expertise. Even, some are seem desperate, disappeared and a few new months, they just come to their lecture with the reason of confused what group/topic that they should take for their thesis. Basically, they can actually determine what group/topics thesis according to their abilities while attending course in the 1st through 7th semester.

From these previous grade semester, they can actually determine what group/thesis topic is suitable for them. The best grade they get when attending course in 1st to 7th semester, so they can predict which thesis topic they should take.

With such conditions, the researcher develops web-based decision system, which can be used by students to help decide what topic of the thesis is appropriate for them, in accordance with the values in 1st to 7th semester. The clustering involes K-Means algorithm to process it. Otherwise, the method for calculating the weight and ranking of each thesis topic is used Simple Additive Weighting (SAW).

II. RESEARCH METHOD

The method of research to done includes the steps that shown in Figure 1. The steps in Figure 1 have a system development phase to be built. Development of this system using process model (Software Development Life Cycle) Extreme Programming. This process model is part of Agile Development commonly used in the software development process. The advantages, the research process is done to be more flexible and quickly better than other process model/conventional SDLC.

A. Literature Review

At this phase is conducted a literature review of several references relevant to the research topic. The references referred to the basics of data mining, K-Means Algorithm, Decision Support System, Decision Making Phases, Simple Additive Weighting, Extreme Programming, Agile Development, object oriented analysis, and design.

This phase is done by analyzing the problems and requirements of this research. Problem analysis is done to study and understand problem domains quite well for the overall analysis of problems, opportunities, and limitations [1]. The analysis step consists of problem analysis and its solution based on the reference in the previous stage. Problem analysis and its solutions explain the focus of the problem to be resolved in the study and determine the likelihood of success of the proposed solution. While the needs analysis is done to produce a specification of the requirements of a detailed specification of things that will be done when the system is implemented.

This requirement analysis is needed to determine the output produced by the system, the required input of the system, and the process description used to process the input into output. The problem to be solved in this research is to provide a website to support student decisions in determining the type of topic group thesis. Initially, the preference model is entered by the department of Information Systems. Then, the student can input the provided criteria. The final result of this system is an alternative choice in the form of thesis topic rangking so that students can use the alternative choice to determine the topic of the thesis. The displayed rangking is arranged based on weight of each defined topic. Every topic has value to give describing information for student . This value can be a suggestion for

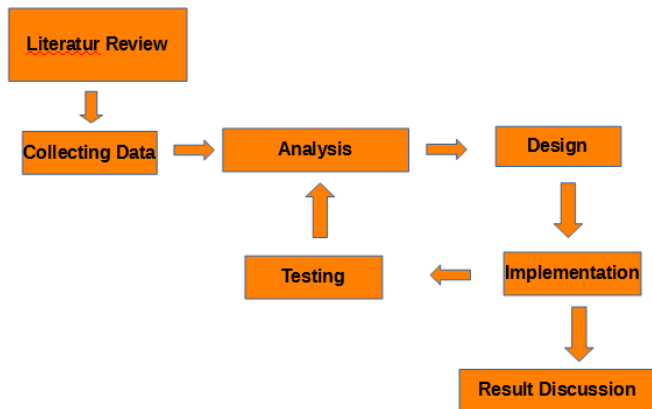student. So, students are not too much use their intuition to decide their decision.

Fig. 1.   Phases of The Reseach

B. Design Phase

Phase of design which is a continuation of the analysis phase is divided into 2 sub phase namely logical design and physical design. The final target of this phase is the resulting design that meets the specified requirements during the system analysis phase. The final result is a detailed design specification that is easily embodied in programming. In the conceptual design, user requirements and identified problem solving during the analysis phase are initiated to be implemented.

C. Implementation Phase

In the implementation phase includes programming and testing activities in level of the program code. Programming is conducted based on the results of physical design. Programming uses PHP programming language with IDE NetBeans 7.0.

D. Testing Phase

Testing at this stage is a further testing or testing when running the application after the application system is complete. The purpose of this test is to determine the overall system performance and to perform data retrieval that will be discussed in the next stage.

E. Result Discussion Phase

The results of some test parameters is obtained at the testing phase for each scenario will be analyzed at this phase

III.   Implementation of Decision Support System

Decision Support Systems (DSS) can be defined as computer systems designed to provide assistance in defining and evaluating alternative action [2]. DSS obtains data from routine transactions on a large scale in a company. Then, it is analyzed with advanced statistical techniques to gain meaningful information, and narrow the range of choices by applying rules based on decision theory. The goal is to facilitate a 'what if' analysis but not replacing manager's

decision. Therefore, DSS facilitates decision-making and does not replace decision-making of employee/manager.

As stated earlier, DSS is a computer-based system that supports choice by helping decision makers in organizing information and modeling outcomes [3]. In Figure 2, illustrates the range of available information systems products. In the diagram, the conventional management information system (MIS) or transaction processing system (TPS) is shown the furthest. MIS is intended for routine, structural, and anticipatory decisions. In such cases, the system may take extract or extract data, integrating, and generating reports. The system is not analysis-oriented and tends to be sluggish, batch processing system. Thus, they are not good to support the decision.
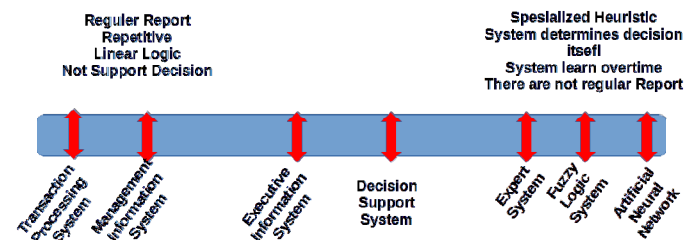

Fig. 2.   The Progress of Information System Product

Into generating alternative decision, there is proces to grouping the data. It is comonly called clustering. Clustering data is an important and powerful technique in machine learning. Clustering data has been used extensively in many areas, such as image processing, document categorization and bioinformatics, and others. The clustering algorithm aims to define a limited set of categories to describe the data set according to the similarity between the objects [4]. Many clustering algorithms have been proposed in the literature. Among them, K-Means has been the most popular algorithm since the 1960s because it is simple, scalable and adaptable to various application domains. Especially in recent years, K-Means can be implemented efficiently on parallel and distributed computers to solve large-scale practical problems. However, K-Means suffers from three major drawback. The K-Means algorithm is very sensitive to the initial starting point that designated as the center of the cluster or centroid. The number of cluster must be determined first; The main drawback of K-Means is that it can not solve a nonlinear clustering manifold.

K-Means clustering is a vector quantization method, which aims to divide n data points into k clusters where each point is in the cluster with the shortest distance to the mean value, functioning as the basic rule of grouping [5]. Specific algorithms are shown as below:

a)   Randomly, select the point k to be the initial centroid;

b)   Repeat the following steps until convergence or achieve max iteration time.

For each sample i, compute the cluster:

$$c^i := arg\,min\|x^i - \mu_j\|^2 \qquad (1)$$

For each cluster j, recalculate the centroid of the cluster or group according to the following equation:

$$\mu_j := \frac{\sum_{i=1}^{m} 1\{c^i = j\}x^i}{\sum_{i=1}^{m} 1\{c^i = j\}} \tag{2}$$

For the equation (1) and (2), c^i represents the group label where the centroid is closest to the ith sequence sample, and the value is a number in the range of 1 to k [5]. The average value represents the guess for the centroid of each group. It is also important to learn the parameter options to get good results:

a) Number of cluster: one of way to cluster is to use the hierarchy clustering method to test the data first and then set the number of K-means clusters to be the same. In addition, this can be done resampling the data to get two subsets, and then performing a similar grouping on two data sets to get the result with k clusters. Equal distribution of cluster point is important to determine whether the value of k can be used or not and it also has to try many times to obtain the best results.

b) Place of centroid: this can select the centroid randomly and perform the process several times. Finally, it should select the best start condition. In alternative way, it can use hierarchy method to calculate in advance and use result of calculation. Maybe, cross validation is also a good way to analyze.

Simple Additive Weighting (SAW) is one of the multi-attribute decision making methods [5]. This method is used to determine the best alternative of many alternatives. Phase of making decision in this research are as follows: make specific criteria, determine alternatives and make a normal matrix, and calculate the total integral value to perform the value of the results. When quality attributes are divided into two groups, for example, benefits and cost, the matrix element r can be normalized by using equation (3).

$$r_{ij} = \frac{a_{ij}}{\max_{i} a_{ij}} \tag{3}$$

The first equation is used to attribute the benefit quality of c_j, and the second equation to determine the cost quality attribute c_j. Based on the normal decision matrix r, the overall value of the architectural preferences of x_i can be expressed by the additive preferences of the coefficients obtained on the operator shown in equation (4).

$$Z_i(w) = \sum_{j=1}^{n} r_{ij} . w_j \; ; i = 1,2, ..., m \tag{4}$$

The attribute value of the overall equation (4) larger than Z_i (w), the better of architecture corresponding to x_i.

An overview of this system is shown in Figure 3. Initially there were several samples of the graduated Student Study Results Card [6]. The sample contains the grade point of achievement (GPA) or index prestation (IP) of students from each course. Then, the GPA value is averaged over each semester resulting in a cumulative grade index per semester. So, every student has a GPA per semester that is from 1st to 2nd semester.

Once the GPA is generated per semester, then the next step is to determine the number of groups and their members. Determination of this group using the K-Means algorithm. The algorithm will be repeatedly performed or past several iterations until the group members do not change. After the group is generated, the next step is to give name the resulting group. The resulting groups have members in the form of students with a GPA score of 1st to 7th semester. Then, it is calculated an average of each group so as to produce the average value of each group. This average value comparison is used to determine the weight of each group from the proportion of each group.

Then, students can also use this system to support their decision in choosing thesis topic. Before using this system, the operator or department staff must determine the criteria in the form of the course value which is the prerequisite of each thesis topic. Then, students fill this prerequisite with the GPA score of each course. Furthermore, the system will calculate the weight and ranking each group of thesis so that resulted in the ranking election thesis topic. Students can freely choose the topic of thesis group to be selected with decision support in the form of thesis topic group rankings. Furthermore, the title of the thesis that corresponds to the topic can also be input as well.
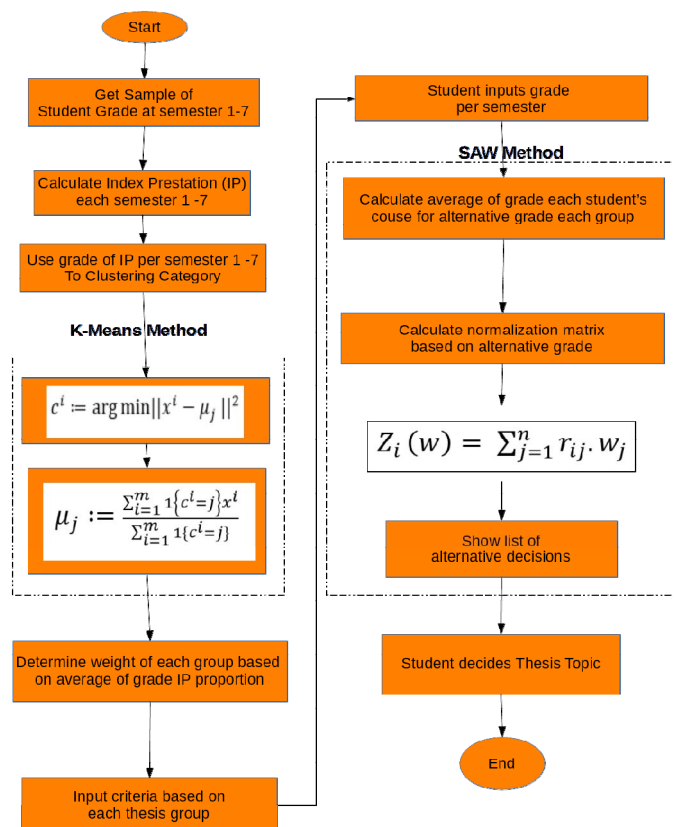

Fig. 3. Overview of System

IV. SYSTEM DEVELOPMENT

The phase of requirement analysis defines the business requirement of the system to be built. Functional requirements are a description of the activities and system services that must be available. Some functional requirements are defined in Table I. The defined requirements are some of the important requirement of the system which interacting with users. Users of the defined system are students and operators who work as department staff. In addition, the requirement for system security is also required.

TABLE I.   LIST OF FUNCTIONAL REQUIREMENT

No.	Fungsional Requirement
1.	Students must input personal data for account registration.
2.	Students can input the GPA score and generate an alternative decision.
3.	Department staff or operators can process student data.
4.	Operators can process the course data.
5.	Operators can process Student Study Results Card data.
6.	Operators can group students by GPA semester.
7.	Operators can process data on thesis topic list and topic weight.

Once the general description is defined, functional requirements can be documented using the use case. Use case describes the functionanl system from the point of view of external users and in the way and terminology that has been understood. Figure 4 shows the use case diagram for documentation of functional requirements. In Figure 4 there are 2 actors namely Operator (Department staff) and Student. Each actor has their own case. Case is a representation of the functional requirements from the requirement analysis. Cases associated with the Operator are account registration, processing student data, processing of course data, processing Study Results data, and clustering students based on GPA semester. The next case that is connected with the student is the account registration, input the GPA score, and generate an alternative decision.

Logical design is the stage that will reveal the business requirements on the system. This design still uses UML modeling (Unified Modeling language) which is object oriented. The beginning of this stage is to model the step of process or activity of the system. Activity modeling of this system uses activity diagram. The activity diagram illustrates the steps of a predefined use case. Figure 5 shows the activity diagram of Student clustering based on the GPA semester. This diagram is a description of the procedure of functional requirements that have been described in the use case. The process that occurs only involves the Operator and the Computer System. The operator selects the clustering feature and approves the bundling process. Then the system calculates the clustering using K-Means method so as to produce sample data of students who have been clustered based on GPA semester.
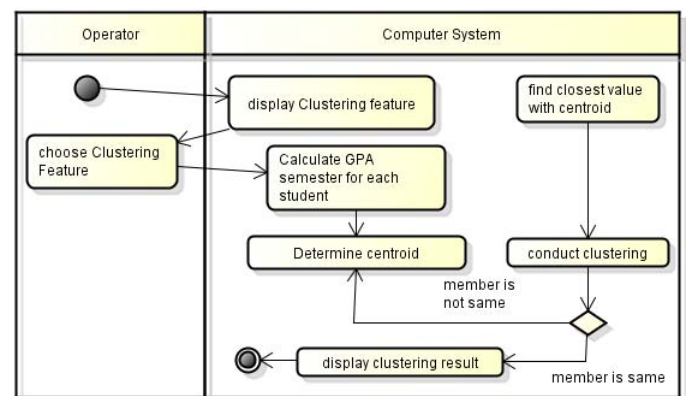

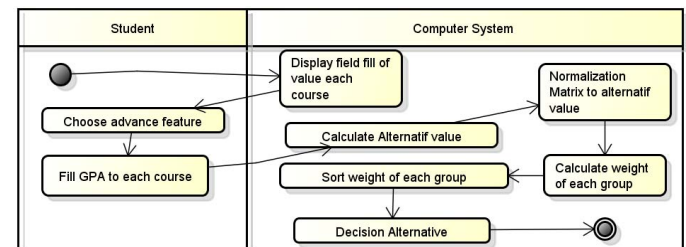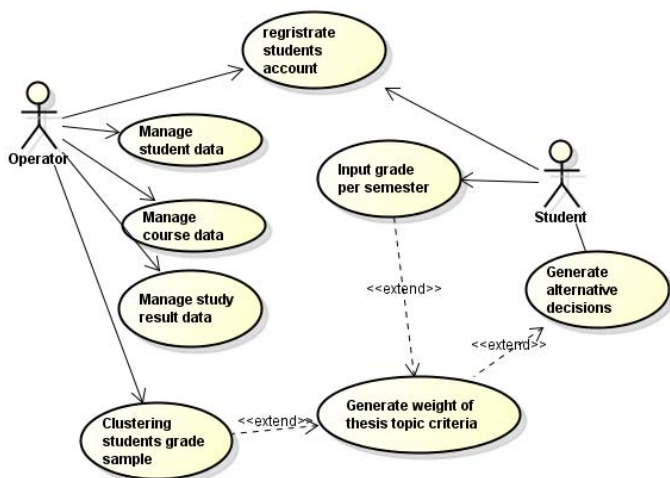Fig. 5.   Activity Diagram of Clustering students grade sample

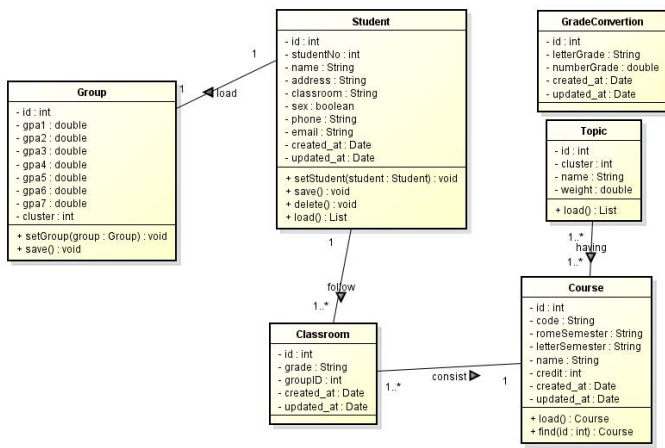Fig. 6.   Activity Diagram of generating decisions alternative

Fig. 7.   Class Diagram

In Figure 6, there is a procedure in the process of generating an alternative decision. This step is a further step of clustering student grades sample. This process involves students with the computer system. Students input the GPA score per semester then agree on the decision support decision process. Then, the system will perform the process of generating an alternative decision in the form of a weighting value for each assigned thesis topic.
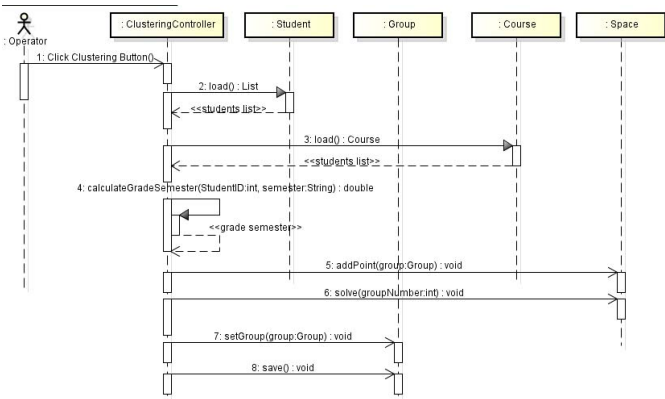

Fig. 8.   Sequence Diagram of Clustering Students Based on grade Semester
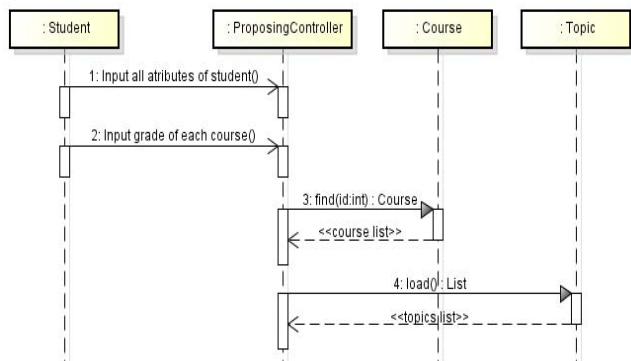
Fig. 9.   Sequence Diagram of Generating Decision Alternatives

The next step after designing using an activity diagram is to design a class diagram. This diagram is used as the initialization of the required data in the system. These data are represented in a class. Then, the class is used for object instanization. The objects that have been formed are useful for interacting and accommodating the interrelation of the process with the data. The class diagram is shown in Figure 7 contains classes of Student, Group, Course, GradeConversion, Topic and Classroom. The class is useful for generating data abstraction objects.

In Figure 8, there is a sequence diagram that is useful for technical designing of some descriptions in the activity diagram. This diagram illustrates the system flow that is looked by a cosntructor point of view. It is presented to define the responsibilitas of objects that exist on the system. This responsiveness is represented by sending messages to other objects. This message delivery is a representation of object behavior. Sequence diagrams are also shown in Figure 9 which is a representation of the technical design of alternative decision procedures.

The result of this system implementation is realized in the form of website. In Figure 10, there is a thesis topic clustering page. This page displays a list of students along with a GPA score per semester. On this page, there is also a field of number of groups. If the number of groups has been filled then clicked Clustering button will be displayed grouping table. This page is used by department staff or operators to clustering topic based on GPA semester students.

Fig. 10. The Webpage of Clustering Thesis Topic Result

The results of calculations and some alternative decisions are shown in the form of the webpage shown in Figure 11. This page displays the weight of the value of each defined interest. The interest is arranged sequentially based on its weight ranking. Topic with the highest weight value is above, while topic with the lowest weight is at the bottom. The results shown are expected to help students in determining the interest

of thesis and decision making has a strong basis and not based on the overall intuition randomly.

Fig. 11. Alternative Decisions Webpage

V. CONCLUSION

In this reseach, there are several conclusions. Some of these conclusions are described as follows:

1. Decision support system used to build a system that aims to help students determine thesis topic.

2. The result of clustering by the K-Means method is used for the preference value component which will produce the Criteria Weight. This leaves the decision maker unnecessary to assign further value to generate the weight of the criteria.

3. Development of this system use object-oriented approach and system design uses Unified Modeling Language.

4. Decision Support System that is implemented produces several alternative topic by having the weight of each interest.

5. The system has been built based on the website with the operator and student users as decision makers.

REFERENCE

[1] R. Pressman, Software Engineering : a Practitioner's Approach 7th Edition, 7th ed., New York, USA: Mc Graw Hill, 2010.

[2] V. L. Sauter, Decision Support System For Business Intelligence, Canada: John Wiley & Sons, 2010.

[3] A. D. Silva and T. D. Rupasinghe, "A Decision Support System for Demand Planning: A Case Study from Manufacturing Industry," in *Moratuwa Engineering Research Conference (MERCon)*, Moratuwa, Srilanka, 2017.

[4] E. Daniati and A. Nugroho, "K Means Clustering With Decision Support Sistem using SAW: Determining Thesis Topic," in *6th IEEE International Conference on Control System, Computing and Engineering*, Penang, Malaysia, 2016.

[5] C. Lu, Y. Shi, Y. Chen, S. Bao and L. Tang, "Data Mining Applied to Oil Well Using K-means and DBSCAN," in *7th International Conference on Cloud Computing and Big Data*, Beijing, China, 2016.

[6] A. Pranolo and S. M. Widyastuti, "Simple Additive Weighting Method on Intelligent Agent for Urban Forest Health Monitoring," in *International Conference on Computer, Control, Informatics and Its Applications*, Bandung, Indonesia, 2014.

[7] W. Xue, R.-l. Yang, X.-y. Hong, N. Zhao and S.-g. Ren, "A Novel k-Means Based on Spatial Density Similarity Measurement," *in Chinese Control And Decision Conference (CCDC)*, Chongqing, China, 2017.