Science Harmony: International Journal of Computer Science, Mathematics and Natural Sciences
Volume. 1 No. 3 August 2024
Page. 11-21
Available online at: https://science.ifrel.org/index.php/ScienceHarmony

Management System For Determining Promotional Strategies Cut Nya' Dien Vocational School With Data Mining Using K-Means Clustering Algorithm

Muhammad Abdur Rouf
Computer Systems Study Program ,College of Electronics and Computers **(STEKOM),**

Jl. Majapahit 605, Semarang, tel/fax : (024) 6723456
abdurrouf90@yahoo.com

ABSTRACT : *The development of the number of new students at Cut Nya' Dien Vocational School experiences ups and downs every year. The reason is because the data was not processed appropriately based on historical data. The aim of this research is to determine the right promotional strategy based on the results of the groupings formed. One solution that can be used is data mining techniques using the Kmeans Clustering algorithm. Clustering is a method for searching and grouping data that has similar characteristics between one data and another. K-Means is a non-hierarchical data clustering method that attempts to divide data into one or more clusters. The results of this research will form 3 clusters and are expected to be one of the basic considerations in making promotion strategy decisions based on the clusters formed by the school principal.*

Key words: *Promotion strategy, data mining, clustering, k-means, Cut Nya' Dien Vocational School*

1. INTRODUCTION

Advances in information technology are growing rapidly in all areas of life. A lot of data is produced by sophisticated information technology, starting from the fields of industry, economics, science and technology as well as various other fields of life which produce abundant data (Johar, 2013). To help make it easier for decision makers to analyze and extract abundant data, a new branch of science was born called *Data Mining* (Yuli & Rodhyah, 2014)

SMK (SMEA) CUT NYA' DIEN Semarang is located at Jl. Wolter Monginsidi No. 99 Genuk District. It has various facilities and quite spacious space with the number of classes reaching 14 classes which are divided into 4 departments and the number of students in the 2014-2015 academic year reached 399 students, whereas in 2014-2015 the number of new students entering was 170 students.

The following is data on the number of new student admissions for the last 5 years r

Table 1.1 Data on Number of New Student Admissions

School year	The number of students
2010/2011	190 students
2011/2012	132 students
2012/2013	143 students
2013/2014	124 students
2014/2015	170 students

(Source: TU SMK (SMEA) Cut Nya' Dien Semarang 2010 to 2015)

Judging from the data on new student admissions in the last 5 year period, the number of new students has experienced ups and downs. One of the causes of the fluctuating number of

Received: May 10, 2024; Revised: June 15, 2024; Accepted: July 08, 2024; Online Available: July 13, 2024;
* Muhammad Abdur Rouf , abdurrouf90@yahoo.com

new students is that student data is not processed appropriately based on historical data by the CUT NYA' DIEN Vocational School. This can influence decision making in determining the right target promotion area. Student data processing should be carried out in order to determine the right target promotion areas so that there is no decrease in the number of students in the following year.

Based on the existing problems, one way to solve them is by doing *data mining* . *Data mining* is a data processing method to find hidden patterns in the data. The results of data processing with *data mining* can be used to make decisions in the future (Santosa, 2007). There are several algorithms that can be used to complete *data mining* , including: *K-Median Algorithm, K-Means Algorithm, C45 Algorithm, and Naive Bayes Classifier Algorithm* .

The algorithm that will be applied for this case is *k-means clustering* . *K-means clustering* is a popular method used to obtain a description of a set of data by revealing the tendency of each individual data to group with other individual data. This grouping tendency is based on the similarity of the characteristics of the individuals in the existing data. The basic idea of this technique is to find the center of each possible data group and then group each individual data into one of these groups based on distance (Turban, 2005)

The results of data mining will later form 7 clusters. Some promotional strategies that will be carried out based on the results of the cluster that will be formed are: distributing brochures, placing banners in strategic places, making door to door presentations to each target school, making cooperation proposals with the target school, holding events (workshops, exhibitions, etc.), social activities, providing discounts and scholarships, promotions via blogs or school websites and other social media.

2. THEORETICAL FOUNDATION

Theoretical Description

2.1. Understanding Systems

A system is an organized set of components or variables that interact with each other, depend on each other and are integrated (Lucas, 2012).

2.2. Understanding Information

Information is data that is processed into a form that is more useful and more meaningful for those who receive it (Jakub, 2012).

2.3. Understanding Information Systems

An information system is an organized combination of people, hardware , *software* , communications networks, and data resources that collect, transform, and disseminate information within an organization. (Jakub, 2012)

2.4. Understanding Promotion Strategy

Philip Kotler (2001) states that "promotion is a variety of activities carried out by companies to communicate the benefits of their products and to convince target consumers to buy them"

In general, forms of promotion have the same function, but these forms can be differentiated based on their specific tasks. These special tasks are often referred to as the promotion *mix* . The promotional mix as mentioned by Philip Kotler is "a special combination of advertising, personal selling, sales promotion and public relations that a company uses to achieve its advertising and marketing objectives".

2.5. *Data Mining*

2.5.2. Understanding *Data Mining*

Data mining is a data processing method to find hidden patterns in the data. The results of data processing using this *data mining method* can be used to make decisions in the future. This *data mining* is also known as *pattern recognition* (Santosa, 2007) .

2.5.3. Data Mining Clustering

Data mining is divided into several groups based on the tasks that can be carried out, namely (Turban, 2005):

- Description

  Description is describing the patterns and trends contained in the data which makes it possible to provide an explanation of the pattern or trend.

- Estimate

  Estimation is almost the same as classification, except that the estimated target variable is more numerical than categorical. The model is built using complete records that provide the target variable values as predicted values.

- Prediction

  Prediction is almost the same as classification and estimation, but in predicting the value of the results will be in the future.

- Classification

  Classification is the process of finding a model or function that describes and differentiates data classes or concepts with the aim of predicting classes for data whose class is unknown.

- Clustering

  Clustering is grouping records, observations, or paying attention and forming classes of objects that are similar. A cluster is a collection of records that are similar to each other and dissimilar to records in other clusters.

- Association

  Association in data mining is finding attributes that appear at one time. In the business world it is more commonly called shopping cart analysis.

2.6. *Understanding Clustering*

Basically, *clustering* is a method for searching and grouping data that has similar characteristics *(similarity)* between one data and another data. *Clustering* is a *data mining method* that is unsupervised , meaning that this method is applied without training *and* without a teacher *and* does not require an output target. In data mining there are two types of clustering methods used in grouping data, namely *hierarchical clustering* and *non-hierarchical clustering* (Santosa, 2007) .

2.7. Understanding the K – Means Algorithm

K-Means is a non-hierarchical data grouping method that attempts to partition existing data into two or more groups. This method partitions data into groups so that data with different characteristics is grouped into other groups. The purpose of grouping this data is to minimize variation within a group and maximize variation between groups (Eko Prasetyo, 2012).

According to Santosa (2007), the steps for clustering using the *K-Means method* are as follows:

2.7.1. Choose the number of *clusters k* .

2.7.2. Initialization *of k cluster* centers can be done in various ways. However, what is most often done is in a random way. Cluster centers are given initial values with random numbers.

2.7.3. Allocate all data/objects to the nearest cluster. The proximity of two objects is determined based on the distance between the two objects. Likewise, the proximity of data to a particular *cluster* is determined by the distance between the data and the *cluster center* .

In this stage it is necessary to calculate the distance of each data to each *cluster center* . The maximum distance between data and a particular *cluster* will determine which data falls into which *cluster* . To calculate the distance of all data to each cluster center point, you can use Euclidean distance theory which is formulated as follows:

$$\mathbf{D\ (i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + ... + (X_{ki} - X_{kj})^2}}$$

Where:

D (i,j) = Distance of data *i* to the center of cluster *j*

X_{ki} = The ith data *in* the kth data *attribute*

X_{kj} = The *jth center point* on the *kth attribute*

2.7.4. Recalculate the *cluster center* with the current *cluster membership. The cluster* center is the average of all data/objects in a particular *cluster* . If desired, you can also use the median of the cluster. So the average (mean) is not the only measurement that can be used.

2.7.5. Reassign each object to use the new *cluster center.* If the cluster center does not change again then the *clustering process* is complete. Or, return to step number 3 until the *cluster center* does not change anymore.

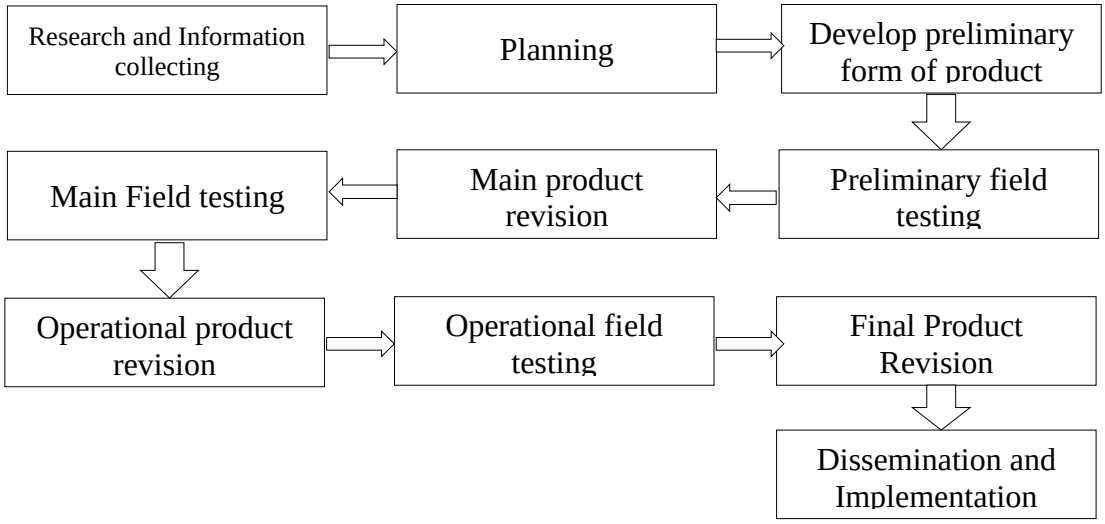
3. METHODOLOGY

To create a data mining information system requires careful preparation and planning. In this development, a development model will be put forward as the basis for product development . The model used is *the Research and Development* (R & D) model approach from Borg and Gall with 10 stages, but the author uses only 6 stages. The development plan with R & D design from Borg and Gall has the aim of developing and validating the product. This model approach has the following steps:

- *Research and Information Collecting* ,
- *Planning* ,
- *Develop Preliminary form of Product* ,
- *Preliminary Field Testing* ,
- *Main Product Revision* ,
- *Main Field Testing* .

The selection of the Borg and Gall model was based on consideration of the development model which was prepared programmatically with careful preparation and planning steps.

Procedurally, there are 6 research steps in the Borg and Gall (1983) R & D model as shown in the following picture:

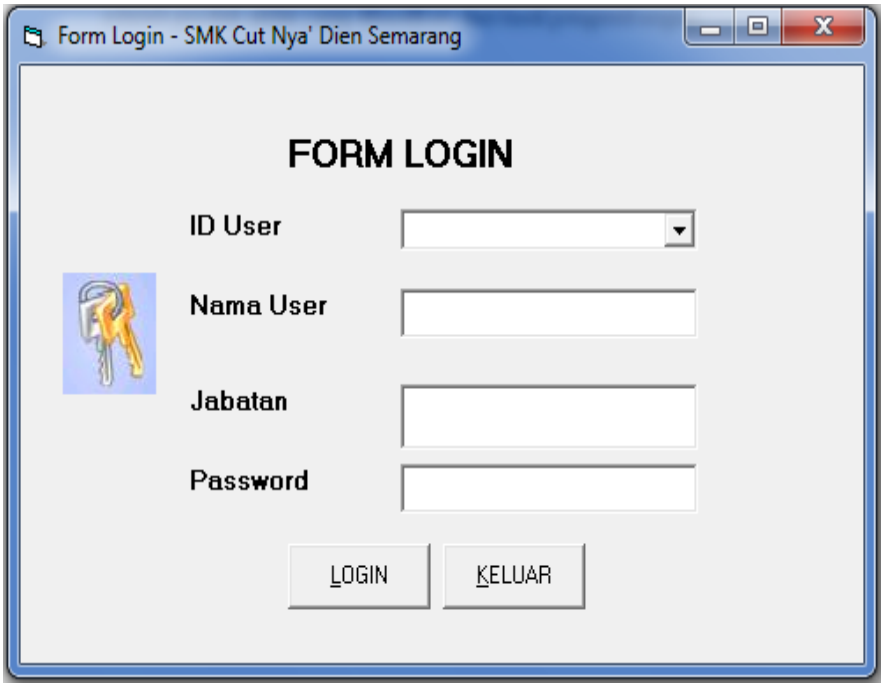


Borg and Gall R&D Model Drawing (1983)

4. RESULTS AND DISCUSSION

After conducting field trials involving users, the following is the final product resulting from the development results.

4.1. Login Form



Login Form Image

In this login form, the user is required to fill in the identifier and password. If the password is correct, the user can enter the main menu and if the password is incorrect, the user cannot enter the main menu.

4.2. Master Menu Form

Master Menu Form Image

The menu form can be accessed by registered users, accompanied by access rights that have been set by the administrator.

4.3. Student Data Form

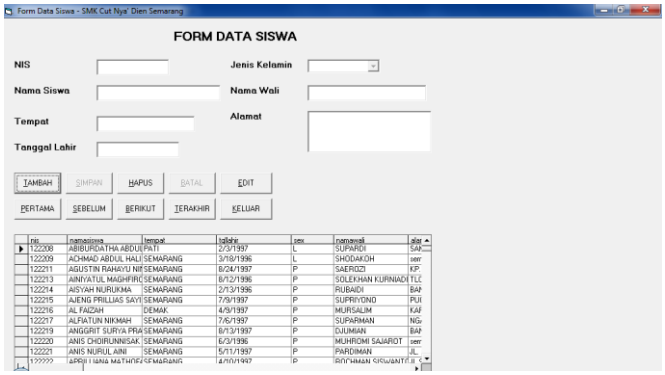


Image of Student Data Form

On the student data form, the nis will be given automatically by the system. If there is data that has not been filled in, then the data cannot be saved. The form also applies to data on the student's major, salary and school of origin.

4.4. Profile Data Form

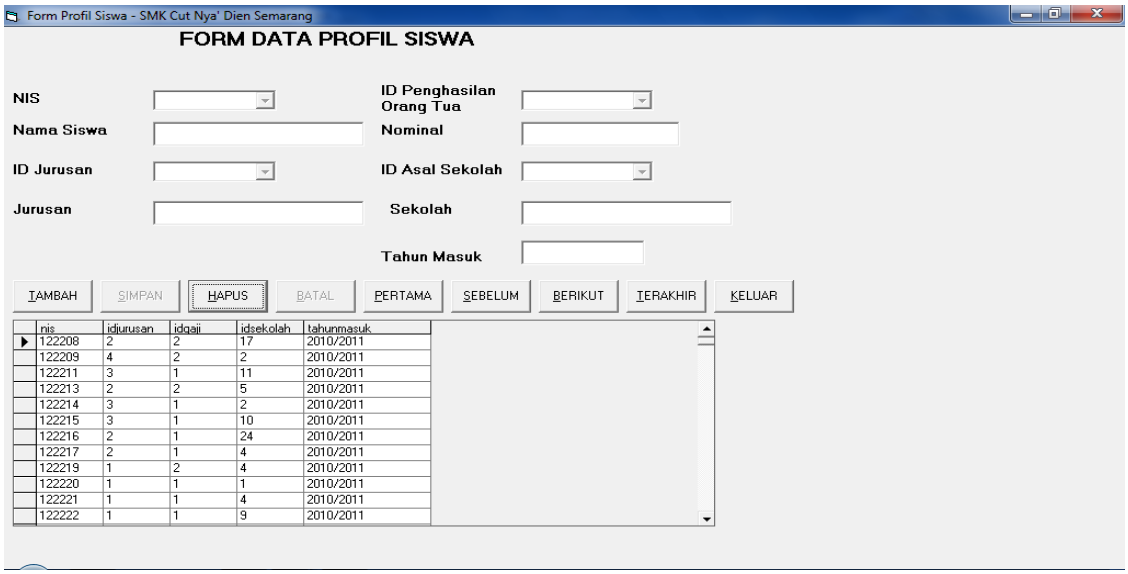


Image of Student Profile Data Form

The profile data form is used to fulfill the criteria that have been set for carrying out data mining. Each student will have a major, parent's salary, and the student's school of origin.

4.5. Data Mining Forms

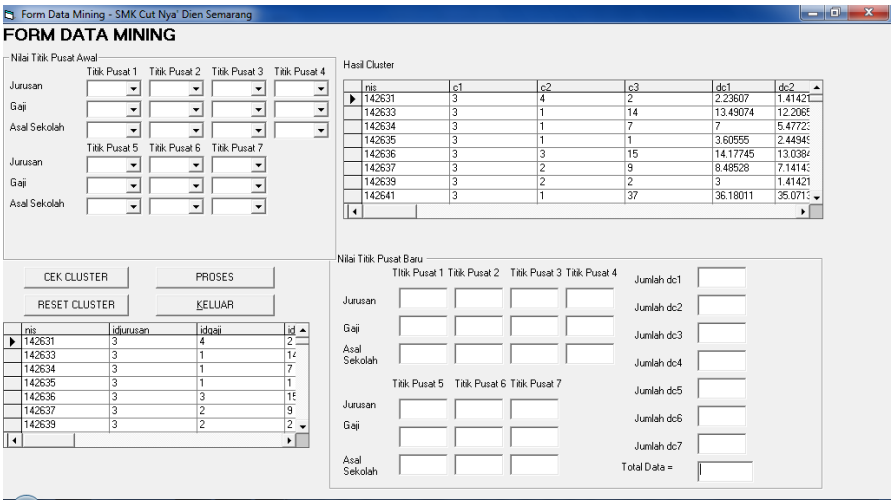


Image of Data Mining Form

Data mining forms are used to process student profile data into information that can be used to make decisions. How to use it is the first step, click reset cluster then click check cluster. After that, input all the initial center point values, then click process. If the process stops, then click the exit button.

4.6. Profile Data Report

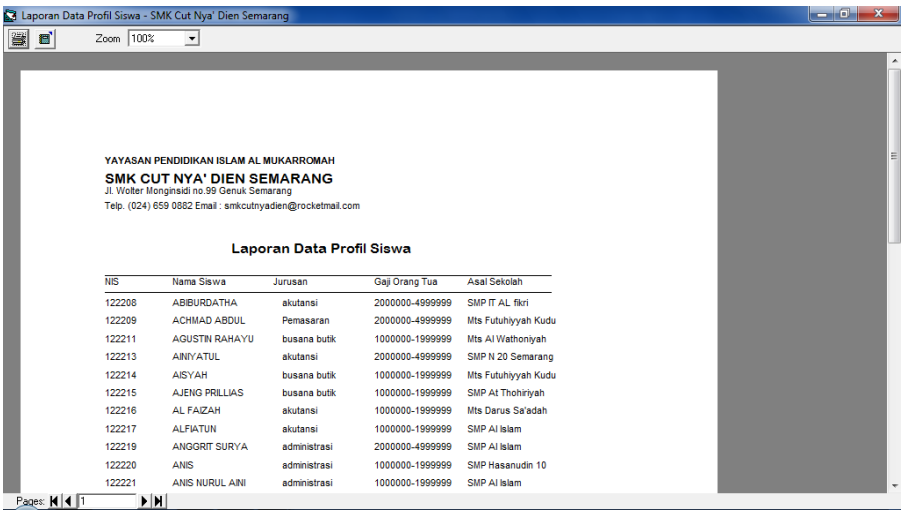


Figure 5.6 Profile Data Report

The profile data report is obtained from the input results in the student profile data form

4.7. Data Mining Report

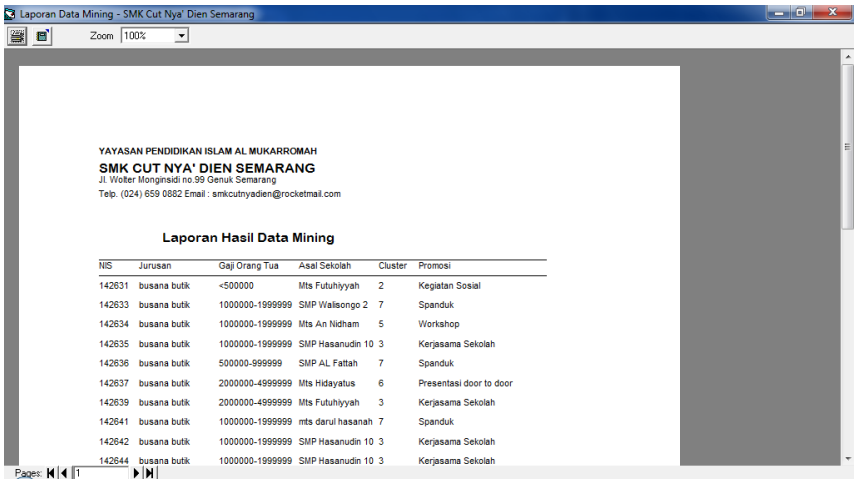


Figure 5.7 Data Mining Data Results Report

Data mining reports are obtained from the input results in the data mining form accompanied by promotions used according to the clusters formed.

5. CONCLUSION

From the results of research that the author has conducted on information systems for determining promotional strategies using data mining, the author can draw the following conclusions:

5.1. The results of determining the promotional strategy using data mining using the k-means clustering algorithm were successful.

5.2. After grouping student data, the results of the clusters formed are:

- Cluster 1 is filled by administration majors with salaries of 500000-999999, originating from Hasanuddin 10 Middle School with the same frequency, namely 4. The promotional strategy used is brochures.
- Cluster 2 is dominated by administration and accounting majors with a frequency of 45 and 40, with salaries between 1000000 – 1999999 as many as 33 and 2000000 – 4999999 as many as 41, and school origins from SMP Hasanuddin 10 and MTs Futuhiyyah Kudu with frequencies 57 and 27. The promotion strategy is by holding social activities.
- Cluster 3 is dominated by marketing and fashion design majors with frequencies of 42 and 36, with salaries between 1000000 – 1999999 as many as 62 and 2000000 – 4999999 as many as 40, and schools from MTs Futuhiyyah Kudu and SMP Hasanuddin 10 with frequencies 46 and 27. For promotional strategies namely by holding school collaborations.

- Cluster 4 is dominated by marketing and fashion design majors with frequencies of 19 and 3, with salaries between 1000000 – 1999999 as many as 17 and 2000000 – 4999999 as many as 5, and schools from SMP Al Islam and SMP N 20 Semarang with frequencies 14 and 2. For strategy promotion is by using online media.
- Cluster 5 is dominated by marketing and fashion design majors with a frequency of both 15, with salaries between 1000000 – 1999999 as many as 26 and 2000000 – 4999999 as many as 24, and schools from SMP N 20 Semarang and SMP Islam Siti Sulaechah with frequencies 27 and 14. For strategy The promotion is holding a workshop.
- Cluster 6 is dominated by administration and accounting majors with frequencies of 18 and 17, with salaries between 1000000 – 1999999 as many as 21 and 2000000 – 4999999 as many as 33, and school origins from MTs Hidayatus Syubban and MTs An Nidham with frequencies 17 and 11. The promotion strategy is by conducting *door-to-door presentations.*
- Cluster 7 is dominated by administration and accounting majors with a frequency of 114 and 70, with salaries between 1000000 – 1999999 with a frequency of 142 and 2000000 – 4999999 with a frequency of 116, and school origin from SMP Al Islam and MTs Al Wathoniyyah with a frequency of 28 and 16. The promotion strategy is by using banners.

BIBLIOGRAPHY

Asriningtias, Y., & Mardhiyah, R. (2014). Application of data mining to display information on student graduation levels. *Journal of Informatics*, 8(1).

Elvira, et al. (2015). Analysis of graduate data using data mining to support Lancang Kuning University's promotion strategy. *Journal of Information and Communication Technology Digital Zone*, 6(2).

Jogiyanto. (2010). *Information technology systems*. Yogyakarta: Andi.

Kotler, P., & Armstrong, G. (2001). *Marketing principles* (8th ed., Vol. 1). Jakarta: Erlangga.

Kusrini, & Taufiq, L. E. (2009). *Data mining algorithms*. Yogyakarta: Andi.

Lucas, H. (2012). *Design analysis and implementation of information systems*. Jakarta: Erlangga.

McLeod, R., Jr., & Schell, G. P. (2008). *Management information systems*. Jakarta: Salemba Empat.

Mulyanto, A. (2009). *Information systems concepts & applications*. Yogyakarta: Pustaka Pelajar.

Ong, J. O. (2013). Implementation of the K-means clustering algorithm to determine President University's marketing strategy. *Industrial Engineering Scientific Journal*, 12(1).

Prasetyo, E. (2012). *Data mining concepts and applications using Matlab*. Yogyakarta: Andi.

Santosa, B. (2007). *Data mining: Techniques for using data for business needs*. Yogyakarta: Graha Ilmu.

Santoso, S. (2010). *Multivariate statistics*. Jakarta: Elex Media Komputindo.

Turban, E., et al. (2005). *Introduction to data mining*. Pearson.

Yakub. (2012). *Introduction to information systems*. Yogyakarta: Graha Ilmu.