The Implementation of Data Mining to Analyze the Consumer which is divided Into Class to Support the Decision Support System (DSS) in TB. 80 Majalengka

D. Susanti
Universitas Majalengka, Jl. Kh Abdul Halim 103, Majalengka, Jawa Barat, Indonesia
Email: deflysusanti@gmail.com

Abstract. Consumers are an important asset to the company. This is the reason why a company must design and uses a good strategy to serve the consumer. With numerous consumers in a company, thus, the problem which the company face is to decide which one is the potential consumer. Using clustering method on data mining, the company can identify a potential consumer by grouping the consumers. Aims of the grouping the consumers are to know the consumers’ behaviour and implementing an appropriate marketing strategy thus make the company earn a profit. This research discusses how the process of data mining from data of the consumers in TB. 80 which is a company run in selling building materials and has a goal to find a potential consumer, therefore, the company profit will increase. The process of Data Mining was started with preprocessing data (selecting, cleaning and transformation), then, on clustering phase use algorithm-means by deciding the number of clusters. The result of clustering from algorithm-means was used to group the consumers and formed a class based on frequency and monetary.

1. Introduction
The need for an information system in these days is very essential. Without the help of the information system make a company cannot survive in the competition. Trusted support of reliable data processing is very important in supporting the daily operation of the company [1].
Many companies do not hesitate to spend so much money to build a good, trusted and reliable information system. In the beginning, computerization system in the company still focuses on processing data of daily operation of the company, which is a transaction in the company.
TB. 80 realise that the sales part is the spearhead of the company and need the help of data analysis that can be used to make a decision which is precise, trusted and reliable. The good and accurate decision will support the company strategy in nurturing a loyal consumer, thus, help in increasing the income of the company.
Today, the sales part of TB. 80 are still running an information system that only supports daily operational data. The information system needed by the company was meant to speed up producing a receipt for the consumers and also for making a simple report such as sales report per day, list of
consumers and list of sold out goods. The needs of data analysis were first realized when analyzing the sales data such as grouping the consumers[2].

One of the applications that popular in company management realm is the Decision Support System (DSS). DSS is an information system which uses for helping management in the process of taking a decision. One thing should be emphasized is that DSS is not going to replace every job of management, but become one of the tools for them[3].

In the process of taking a decision required some methods to complete it. K-means method is one of an algorithm that easy and often used in grouping technique because it involves efficient calculation and needs no parameter to form a group of consumers, only based on 3 aspect dimensions, which is Recency, Frequency and Monetary.

Research question

Based on the problem, the researcher formulates the research question as follows;

1. How to analyze the consumers in TB. 80 Majalengka?
2. How to group the consumers to help the Decision Support System (DSS) in TB. 80 Majalengka?
   1) Data mining

   Data Mining is a term used to describe a data discovery inside the database. Data Mining is a process which uses a static technique, mathematics, artificially intelligence and machine learning to extract and identify beneficial information and information which is constructed from several big databases.

   2) Description of Decision Support System

   Decision Support System (DSS), at first, was identified as a system which was expected to help in taking a decision on semi-structured or unstructured decisions. DSS is expected to be adding value for decision taker to expand their capability while not replace their role and their policy. Based on the explanation, the researcher concludes that the system will be a computer-based and operate online, and the result will be in graphic form while this time is simplified through a web server and browser[4].

3) K-Means method

   Based on the algorithm, K-means is one of the algorithms that easy and often used in grouping technique because it involves efficient calculation whereas it does not need any parameters[5].

   Until it found the stable result, K-means algorithm was conducted in a way as follows:
   a) Decide the centroid data, in this system, it was decided that centroid data is the first n data from data that will be clustered.
   b) Calculate the distance between centroid data with every each data.
   c) Grouping the data based on the minimum distance.
   d) If the placement of the data has been the same with the previous data, thus the process stop. If not, back to the second way (see Figure 1).
Figure 1. K-means algorithm way of working

Picture 1 explain that's the base of the K-means algorithm[5] as follows:
1) Specify the K as the total cluster which wants to be formed
2) Generate the K centroid (cluster centre point) randomly.
3) Calculate the distance between every data to each centroid using the formula of correlation between two objects that are Euclidean and Cosine similarity.
4) Grouping every data based on the closest distance between data and its centroid.
5) Specify the new centroid position (k C) by calculating the mean value from the data which is located on the same centroid.
   \[ c_k = \left( \frac{1}{n_k} \right) \sum d_i \]
   Where \( k \) is total of the document in cluster \( k \) and 1, \( d \) is a document inside cluster \( k \).
6) Back to step 3 if the new centroid position is not the same as the previous centroid.

2. Method
2.1. Data Mining
Data Mining is a term used to describe a data discovery inside the database. Data Mining is a process which uses a static technique, mathematics, artificially intelligence and machine learning to extract and identify beneficial information and information which is constructed from several big databases [2]. Knowledge discovery data (KDD) is a whole process of non-trivial to find and identify the pattern inside the data, where is legal and can bring benefit and can be understood [6]. Figure 2 explain about transformation of data
Figure 2. The process of knowledge discovery in database

Where the phase of the knowledge discovery process [6] as follows (see Figure 2):
1) Data selection
Create a set of data target, a choice of a set of the data or focus on a subset variable or data sample, where the discovery will be conducted.
2) Preprocessing/cleaning
Preliminary and clean up the data is a basic operation, such as noise elimination. Before conducting data mining, it needs to conduct a process of cleaning on the data which will be the focus of KDD
3) Transformation
This process is the creative process and depends on the type or pattern of the information that will be searched in the basic data.
4) Data mining
Selection of data mining algorithm searching process in the data mining is technique, method or inside the data mining is vary. Selecting method or the right algorithm dependent on the goals and process of KDD overall.
5) Interpretation/evaluation
This phase is part of KDD which cover the examination whether the pattern or the information that is found is contradictory with the fact or the previous hypotheses.

2.2. Classification
Data classification has two processes[7]. The first step is to build a model based on a series of data, which is called a learned model. This model was built by analyzing database tuple. Every tuple assumed to be a predefined class which determined by one attribute called class label attribute.

2.3. Research method
The research method used by the researcher is action research [8]. The model of action research used based on a model by Kurt Lewin (Picture 3) which developed action research with a basic concept that action research consists of 4 components as follow:
1) Planning
Plan in action cover all of the steps in detail
2) Acting  
   This step is the implementation of all of the plans  
3) Observing  
   The observation was conducted together with the acting  
4) Reflecting  
   This step is the step to process the data obtained when conducting the observation then the data will be analyzed.

\[ 
\text{Planning} \quad \text{Acting} \quad \text{Observing} \quad \text{Reflecting} 
\]

**Figure 3. Kurt Lewin Method 1990**

2.4. Data collection technique  
   1) Interview  
      The interview is one of the data collection technique by asking several questions directly to the informant, authority, or an expert in the field.  
   2) Questionnaire  
      The questionnaire is a data collection technique conducted by giving a set of questions or statements to the respondents to be answered.

3. Result and Discussion  
   This section will explain the analysis along with the results of the analysis obtained, so that the requirements can be designed regarding the customer's desires  
3.1. Findings and interpretation  
   Based on grouping and analyzing the data, the researcher concludes that several finding that became the base to solve the problems which the researcher propose, as follows:  
   3) Many costumers order in small or large scale  
   4) Many loyal costumers purchase with cash  
   5) Many costumers only order once in a while  
   6) There is no marketing strategy to defend and increase the costumers’ loyalties.  
3.2. System design  
   1) Use case diagram  
      Use case diagram at the picture below describing the interaction between the actor, which are staff marketing and business and development manager with the system [9]. You can see that in Figure 4.
2) Activity diagram

Figure 5 explains that an activity diagram is a decision-supporting system in determining the best consumers shown in the figure below:

**Figure 4. Use Case diagram system**

**Figure 5. Activity diagram**
3) Class diagram
The class diagram at the picture below describes a class or blueprint object on the system running and the system that will be built. At the picture, the diagram also describes the interaction between classes inside the construction of software. The analysis of the establishment class diagram is the main activity which influences the frame of the software design until the coding phase (Figure 6).

![Class Diagram](image)

**Figure 6. Class Diagram**

4) Flow chart
The flow of customer grouping with the K-Means algorithm on the system which will be developed is described using a flowchart [10]. By describing each step in the flowchart, it is expected to make it easier to implement, the following is a flowchart of the customer grouping process with the K-Means algorithm that will be developed on the system (Figure 7)
4. Conclusion

Based on the result of the research, the researcher draws a conclusion related to the process of the research, as follows:

1) The process of consumer analysis on TB. 80 Majalengka was done by applying data mining with the Knowledge Discovery Process in Database (KDD) using the K-Means method.

2) The process of grouping consumers using the K-Means method requires a fairly long process depending on the number of consumers in the selected transaction period. But with the construction of the Decision Support System (DSS) on TB. 80 Majalengka to group consumers can be done more easily and quickly so that the time needed is very efficient compared to doing grouping manually. Because speed and accuracy in decision making are very decisive opportunities in business.

References

[1] Chang, C, Wu, Y., Hou, S., 2009, Preparation and Characterization of Superparamagnetic Nanocomposites of Aluminosilicate/Silica/Magnetite, Coll. Surf. A336:159,166.

[2] Fayyad, Usama, 1996, Advances in Knowledge Discovery and Data Mining, MIT Press.

[3] Han, J, Kamber, M., 2001, Data Mining: Concepts and Techniue. Morgan Kaufmann Publishers: San Fransisco.
[4] Indrajit, Richardus Eko, 2013, Decision Support System, www.eko-indrajit.com. (Diakses 9 Maret 2015).
[5] Kusrini, Emha T. Luthfi, 2009, Algoritma Data Mining. Andi, Yogyakarta.
[6] Turban, 2005, Decision Support System and Intelligent System (Terjemahan: Sistem Pendukung Keputusan dan Sistem Cerdas)Jilid 1, Andi Offset, Yogyakarta
[7] Turban, Efraim. Aronson, Jay E. Liang, Ting-Peng. Sharda, Ramesh, 2007, Decision Support and Business Intelligence Systems, Pearson Education,
[8] Wulandari, 2015, Model Data Clustering Untuk Menentukan Strategi Promosi Dengan Metode KMeans Berbasis Framework Codeigniter : Studi Kasus Universitas Budi Luhur.
[9] Sepsugiarto, Datawarehouse Sebagai Alat Penunjang Pengambil Keputusan Bisnis Dalam Analisis Penjualan: Studi Kasus Pada Bagian Penjualan PT XYZ.
[10] Imanuel Ndaumanu, Ricky, dkk., 2014, Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode K-Nearest Neighbor, Jatisi, 1(1) September 2014 ISSN : 2407-4322.