Application model of k-means clustering: insights into promotion strategy of vocational high school

Satria Abadi 1, Kamarul Shukri Mat The 2, Badlilikham Mohd Nasir 3, Miftachul Huda 4, Natalie L. Ivanova 4, Thia Indra Sari 1, Andino Maselelo 1,5, Fiqih Satria 6, Muhamad Muslihudin 1

1Department of Information System, STMIK Pringsewu, Lampung, Indonesia
2Universiti Sultan Zainal Abidin Malaysia, Malaysia
3Universiti Teknologi Malaysia, Malaysia
4Tyumen Industrial University, Russia
5Institute of Informatics and Computing Energy, Universiti Tenaga Nasional, Malaysia
6Department of Informatics Management, STMIK Pringsewu, Lampung, Indonesia
*Corresponding author E-mail: andimaselelo@gmail.com

Abstract

Admission process is required in promoting the strategy to achieve the target. Through determining the strategic promotion, minimizing the cost in the marketing process could be reached with determining the appropriate promotion strategy. Data mining techniques in this initiative were applied to achieve in determining the promotional strategy. Using Clustering K-Means algorithm, it is one method of non-hierarchical clustering data in classifying student data into multiple clusters based on similarity of the data, so that student data that have the same characteristics are grouped in one cluster and that have different characteristics grouped in another cluster. Implementation using Weka Software is used to help find accurate values where the attributes include home address, school of origin, transportation, and reasons for choosing a school. The cluster of students was classified into five clusters in the following: the first cluster 22 students, the second cluster 10 students, the third cluster 10 students, the fourth cluster a total of 33 students, and the fifth cluster 25 students. The pattern of this result is supposed to contribute to enhance the significant data mining to support the strategic promotion in gaining new prospective students.

Keywords: Data Mining; Clustering; K-Means; K-Means Clustering; K-Means Application; Promotion Strategy

1. Introduction

The quality assurance in education is important in order to continuing the existing growth of enhanced standard, one of which is using data mining [1-3]. The advancement of education level will lead to enhance the nation widely on preparing into global competitiveness [4-6]. Attempts to achieve this attainment, the strategic plan should be prepared in promoting the school entries, including Private Schools. With providing the facilities, good strategy is required to meet this demand in order to successfully achieve the market goal and target in promoting the school [7-9]. In addition, the number of data classification towards an insurance customer using Algorithm C4.5 refers to data mining customers of an insurance company to enhance the smooth or not smooth among the customer [10-12]. Existing data is analyzed using decision tree approach (decision tree) that is algorithm C4.5. With this algorithm can be known which customer data are grouped into the class smoothly and which customer data are classified into the class is not smooth. Then the pattern can be used to estimate customers who join, so the company can take satisfaction accept or reject the prospective customer [13-15]. Implementation of Data mining Classification is to find rainfall prediction pattern using C4.5 Algorithm. About this study, the results would be always dependent into the application that can facilitate Fund Section of Bank XY in obtaining target marketing of credit in the future [16-18]. This research is planning the promotion of data mining-based campus by classification method on Pringsewu Lampung by way of Distributing Form to Respondents to determine five (5) attributes are Address, School Origin, Study Program, Parent Work and Time to get Information, and one (1) Target data consisting of banner, brochure and seminar. A total of 100 students [19-21]. This research will be done data processing students SMK Muhammadiyah 1 Pringsewu. From this student data, processing aims to assist the school in doing marketing and looking for new students. Because students of SMK Muhammadiyah 1 Pringsewu come from various regions, it takes a special strategy by the marketing department in doing marketing to find prospective students for the promotion is done more effectively and efficiently. In this research, data mining analysis is done by using K-Means clustering method. Using this method, the data obtained can be grouped into several clusters based on similarities of the data, so that the data having the same characteristics are grouped in one cluster and that possess different characteristics are grouped in another cluster that has the same characteristics. With the groupings of these data, it is expected that the marketing department can do marketing with the right strategy to get new students.

2. Literature review

2.1. Promotion
Promotion, the fourth marketing mix tools, stand for various activities, the company undertakes to communicate its products merits and to persuade target customers to buy them. The definition has the understanding that the promotion includes all the tools contained in the promotion mix whose main role is to persuade communication.

2.2. Promotion strategy

The variables present in the promotional mix are five [4], that is:

a) advertising
b) personal selling
c) sales promotion
d) public relation
e) direct marketing

2.3. Data mining

Data Mining is a term used to describe the discovery of knowledge in a database. Data Mining is a process that uses statistic, math, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from large databases [5]. The term data mining has the essence as a discipline whose primary purpose is to discover, dig, or mine the knowledge of the data or information that we have. Data mining often referred to as Knowledge Discovery in Database (KDD). KDD is a process consisting of a series of ordered sequence processes, and data mining is one step in the KDD process [6].

2.4. Definition of data mining

Data Mining is a series of processes to explore the added value of information that has not been known manually from a database by extracting patterns from the data in order to manipulate data into more valuable information obtained by extracting and recognizing important patterns or pulling from the data contained in the database.

2.5. Stages in data mining

There are four stages passed in data mining, among others:

1) The first stage: Precise statement of the problem before accessing data mining software, an analyst must have clarity on 'what questions will he want answered'. If there is no right formula for the problematic then you will only waste time and money in making the solution.
2) The second stage: Initial exploration phase begins by preparing data that also includes data mining "cleaning" (eg identifying and thinking about wrongly coded data), transforming data, selecting subset records, data sets, initial selection steps. Describing and visualizing data is the key of this stage.
3) Phase three: The building and validation model of this phase involves consideration of the variety of models and choosing the best for predictive performance.
4) Fourth stage: Deployment selects the right application and modeling to generate predictions. Next we look at the details of the stages of data mining [7].

2.6. Clustering

Basically clustering is a method to search and classify data that have similarity between one data with other data. Clustering is one method of data mining that is unsupervised, meaning that this method is applied without any training and without any teacher (teacher) and does not require target output. In data mining there are two types of clustering methods used in data clustering, is hierarchical clustering and non-hierarchical clustering [8].

2.7. K-means clustering algorithm

K-means clustering is one of the non-hierarchical data clustering methods that classify data in the form of one or more clusters/groups. The data that have the same characteristics are grouped in one cluster/group and the data having different characteristics are grouped with other clusters so that the data in one cluster has a small variation level [9]. According to [10] the steps of clustering with K-Means method are as follows:

a) Select the number of clusters k.
b) The initialization of this cluster center k can be done in various ways. But the most often done is by random. Cluster centers are assigned initial values with random numbers.
c) Allocate all data objects to the nearest cluster. The proximity of two objects is determined by the distance of the two objects. Likewise the proximity of a data to a particular cluster is determined the distance between the data with the cluster center. In this stage it is necessary to calculate the distance of each data to each cluster center. The most distance between one data and one particular cluster will determine which data to enter in which cluster. To distance all data to each cluster center point can use Euclidean distance theory formulated as follows:

\[ D(i,j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2} \]  

(1)

Where:

\[ D(i,j) = \text{Distance of data to } i \text{ to center cluster } j \]

\[ X_{ki} = \text{Data to } i \text{ on attribute data to k} \]

\[ X_{kj} = \text{Center point to } j \text{ attribute to k} \]

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

b) Reassign each object using the new cluster center. If the cluster center does not change again then the clustering process is complete. Alternatively, return to step number 3 until the center of the cluster does not change anymore.

c) SMK Muhammadiyah 1 Pringsewu

SMK Muhammadiyah 1 Pringsewu was established on July 12, 1988. The location of SMK Muhammadiyah building was originally located at Jl. Cambodia Pringkumpul Pringsewu, since 1992 the location of SMK Muhammadiyah Pringsewu moved to Jalan Pemuda no. 56 Pringsewu. SMK Muhammadiyah 1 Pringsewu has several expertise programs: Accounting, Marketing, Software Engineering (RPL), Computer and Network Engineering (TKJ), and Light Vehicle Engineering (TKR).

3. Analysis

3.1. Result and analysis

Testing of the analysis, it is important to determine whether the results of the analysis have been in accordance with the expected decision. To test the truth of the data processing done manually, it can use one of Weka application software 3.8.2.

3.2. Testing data

Test data used is a table that has the following constituent components:

1) Has attributes of student's home address, student's home school, student transportation, and student's reason for choosing school [4].
2) The number of instances is 100.
3.3. Implementation steps

All variables consisting of the attributes used to determine the promotion strategy in SMK Muhammadiyah 1 Pringsewu is stored on Microsoft excel with the name of the SMK M.xls Data Survey (which contains the case or criteria in generating the rule). Furthermore, the process of data transformation where the Data Survey file SMK M.xls then stored with CSV extension, then file opened with notepad or other text editor and data has changed in comma delimited format.

Then the data is adjusted by adding the initial information and the data can already be used as input in Weka 3.8.2. Next on the Weka app, click Explorer, open file (choose CSV file), click Choose, select Cluster and click Start. At this stage the process of data mining is done by choosing the algorithm that will be used in generating cluster centroids.

3.4. Implementation and results

3.4.1. Value of cluster centroids and cluster instances

Figure 1 shows results cluster centroids and clustered instances with Weka.

![Fig. 1: Results Cluster Centroids and Clustered Instances with Weka.](image1)

Figure 2 shows results cluster centroids and clustered instances with Weka.

![Fig. 2: Results Cluster Centroids and Clustered Instances with Weka.](image2)

Testing data with Software Weka based on the reasons students produce data form:

1) Students who choose to attend SMK Muhammadiyah 1 Pringsewu because they saw brochure for cluster 2, as many as 22 students from 100 students (22%).

2) Students who choose to attend SMK Muhammadiyah 1 Pringsewu because of annual competition for cluster 1, 10 students from 100 students (10%).

3) Students who choose to attend SMK Muhammadiyah 1 Pringsewu because they see banner for cluster 2, as many as 10 students from 100 students (10%).

4) Students who choose to attend SMK Muhammadiyah 1 Pringsewu because of friend invitation for cluster 3, 33 students from 100 students (33%).

5) Students who choose to attend SMK Muhammadiyah 1 Pringsewu because of direct socialization to the school of origin for cluster 4, 25 students from 100 students (25%).

Then cluster 3 with friend invitation strategy and cluster 5 with direct socialization from the SMK Muhammadiyah 1 Pringsewu to the origin school can be used as a promotion strategy.

Graph clustering and plot of clustering graphs.

![Fig. 3: Graph Clustering Results with Weka.](image3)

![Fig. 4: Plot Graph Clustering with Weka.](image4)

The clustering graph and the clustered chart plot indicate the exposure of strategic promotion as shown in figure 3 and figure 4. Strategic exposure on educational entrance [22-24] should be considered in particular as an attempt to promote the enhanced standard of learning with emphasizing personal behavior [25-27] with adopting technology innovation [28-30] to improve personalized enhancement [31-33]. In the way to engage the application procedural stage [34-36], the initial value of addressing wide range of constructing strategic initiation on the process of educational entrance examination [37-39]. It needs to deal with incorporating the student’s position on each cluster. The pattern of this result is supposed to contribute to enhance the significant data mining to support the strategic promotion in gaining new prospective students. In particular, the application strategy with the number of approaches [40-42] has to be involved with the wise procedural stage [43-45] in order to achieve the good appointment towards the planning well [46-48]. The significant attribution using an innovative approach here could be determined as an insightful performance [49-51] with expanding the valuable contribution into the school promotion [52-54]. The comprehensive approach should be underlined with gathering the entire process itself [55-
to best achieve the core point of the target set up through conducting the application strategy procedure stage [58-60]. In this view, the innovative performance should be more incorporated with applying such initiative [61-63], and thus, the promotion achievement could be obtained appropriately enhanced [64-66] in the way which determines the strategic initiative performance [67-68]. Moreover, this effort would contribute into performing promotional strategies with socialization directly to school.

4. Conclusion

The result of this study was found that there are two promotional strategies that can be applied by the SMK Muhammadiyah 1 Pringsewu. Those are: 1) Undertaking promotional strategies to invite friends from the school to provide incentives to people who are able to bring or invite prospective students to attend school at SMK Muhammadiyah 1 Pringsewu. 2) Performing promotional strategies with socialization directly to school.

References


Aminudin, N., Gumanti, M. Design of online transaction model on traditional industry in order to increase turnover and benefits. *International Journal of Engineering and Technology (UAE)* (In Press).