Implementation of the K-Means Clustering Method in Data Grouping Sales In Asia Africa Dentures Dental

Sonibe Halawa¹, Rita Hamdani²

¹²Engineering Informatics Program, STMIK Pelita Nusantara Medan, Jl. Iskandar Muda 1 Medan, North Sumatra 20154, Indonesia

E-mail: sonibehal@gmail.com

Abstract- Data mining can be applied to explore the added value of a set of data in the form of knowledge that had been unknown to them manually. There are several techniques used data mining eyes, one satuteknik data mining is clustering. Clustering can be used for grouping to something. As can group sales data that is most desirable, and others. Examples of companies engaged in the sale is a dental African Asia. Asia Africa Dental is one area of business engaged in the sale of false teeth. Asia Africa Dental these every day to meet the needs of consumers. But Asia Africa Dental lacking in reviewing products sold. What products are needed consumer and data storage is less effective. Thus the need for a system that can support the company in taking decisions quickly and precisely. So in this study, the authors used the application of K-Means Clustering method. To facilitate the author in analyzing the K-Means Clustering The author using the application Weka (Waikato Environment for Knowledge Analysis) .. The result of the calculation Weka (Waikato Environment for Knowledge Analysis) is inserted into the Visual Basic.Net.

Keywords: Data Mininng, Clustering, Weka, Visual Basic .Net

1. Introduction

Data mining has some of the techniques used in data mining, one of the data mining technique is clustering. There are two types of clustering methods used in data classification, ie non-hierarchical clustering and hierarchical clustering. K-means clustering as one method of data clustering non-hierarchical partitioning the data into the form of one or more clusters or groups, so that the data which has the same characteristics are grouped into the same cluster and data that have different characteristics are grouped into groups other. Group or cluster acquired a knowledge / information useful for policy users in the decision making process.

Asia Africa dental is one line of business in the sale of false teeth. Asia Africa Dental these every day to meet the needs of consumers. But Asia Africa Dental lacking in reviewing products sold. What products are needed consumer and data storage is less effective. Thus perluadanya a system that can support companies in taking decisions quickly and precisely. In this case analysis was used to overcome the above problems is to use the standard method of K-Means Clustering. To facilitate the author in analyzing the K-Means Clustering The author using the application Weka (Waikato Environment for Knowledge Analysis). Results of the calculations Weka (Waikato Environment for Knowledge Analysis) is inserted into the Visual Basic.Net.

2. Theory

A. Data Mining

By dawn astuti Hermawati (2018: 3), data mining is a process that employs one or more machine learning techniques (machine learning) to analyze and extract knowledge (knowledge) automatically. Data mining is an iterative and interactive process to find a pattern for the new model are valid (perfectly), useful and understandable in a very large database (massive databases)
- Sahih: can be generalized for the future
- New: what is unknown
- Helpful: can be used to perform an action
- Iterative: requires a process that is repeated
- Interactive: require human interaction in the process

B. K-Means Clustering
By dawn astuti herawati (2018: 127), K-Means Clustering using partitional clustering approach. Each cluster is associated with a centroid (center point). Each dot is placed into the cluster with the nearest centroid. The number of clusters, K, Must be specified. The algorithm is basically very simple, namely:

1) Select the K point as the initial centroid  
2) Repeat 
3) The shape of the K cluster by placing all nearest point, 
4) Repeat the calculation of the centroid of each cluster,  

Until centroid unchanged

C. Weka (Waikato Environment for Knowledge Analysis)

According SlametPujiono TIE et al in Journal ISSN: 1411-3201 Vol. 14 No. 2 JUNI 2013, entitled Analysis of Public Satisfaction Using Weka In Achieve Good Governance in Yogyakarta said Weka is a practical machine learning tools package. Weka is an abbreviation of Waikato Environment for Knowledge Analysis, created at the University of Waikato, New Zealand for research, education and a wide range of applications. WEKA is able to solve problems in real-world data mining, in particular the classification of the underlying machine learning approaches. The software is written in Java class hierarchy with object-oriented methods and can run almost all platforms.

3. Discussion And Testing

3.1. Discussion

A. Data analysis

Of the many sales data obtained, taken 9 types of dentures and selling products based on the years 2015-2018. This data is to be sampled k-means algorithm implementation in determining what products are in demand by customers.

<table>
<thead>
<tr>
<th>Table 1.</th>
<th>Sales report 2015-2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Produk</td>
<td>2015</td>
</tr>
<tr>
<td>1</td>
<td>PFM</td>
</tr>
<tr>
<td>2</td>
<td>EMAX</td>
</tr>
<tr>
<td>3</td>
<td>ZIRCONIA</td>
</tr>
<tr>
<td>4</td>
<td>FRS</td>
</tr>
<tr>
<td>5</td>
<td>VERTEX</td>
</tr>
<tr>
<td>6</td>
<td>VALPLAST</td>
</tr>
<tr>
<td>7</td>
<td>ACRYLAST</td>
</tr>
<tr>
<td>8</td>
<td>RODHUSA</td>
</tr>
<tr>
<td>9</td>
<td>FRAME</td>
</tr>
</tbody>
</table>

B. Method Analysis K-Means Clustering

<table>
<thead>
<tr>
<th>Table 2.</th>
<th>Sales data from 2015 until 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Produk</td>
<td>2015</td>
</tr>
<tr>
<td>1</td>
<td>P001</td>
</tr>
<tr>
<td>2</td>
<td>E001</td>
</tr>
<tr>
<td>3</td>
<td>Z001</td>
</tr>
<tr>
<td>4</td>
<td>F001</td>
</tr>
<tr>
<td>5</td>
<td>V001</td>
</tr>
<tr>
<td>6</td>
<td>V001</td>
</tr>
<tr>
<td>7</td>
<td>A001</td>
</tr>
<tr>
<td>8</td>
<td>R001</td>
</tr>
<tr>
<td>9</td>
<td>F001</td>
</tr>
</tbody>
</table>

a. Specify the number of clusters  
   Based on existing sales data annually 4. Cluster that will be the initial centroid with 4 clusters  
b. Allocate data into groups randomly  
   The sales data are allocated based on product codes in the table 2  
c. Calculate the cluster center (centroid) using the K-Means for each group  

Center cluster random sampling is

| Table 3. | Cluster Center |
The formula to calculate the cluster center Euclidean distance is as follows:

\[ d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \]

a. Distance Object A

1) Center cluster 1

\[ D = \sqrt{(1556 - 1566)^2 + (2130 - 2118)^2 + (2064 - 2064)^2 + (1639 - 1639)^2} = 0 \]

2) Center Cluster 2

\[ D = \sqrt{(1556 - 1536)^2 + (2130 - 2130)^2 + (2064 - 2064)^2 + (1639 - 1639)^2} = 0 \]

3) Center Cluster 3

\[ D = \sqrt{(1556 - 1556)^2 + (2130 - 2130)^2 + (2064 - 2064)^2 + (1639 - 1639)^2} = 0 \]

4) Center Cluster 4

\[ D = \sqrt{(1556 - 1536)^2 + (2130 - 2130)^2 + (2064 - 2064)^2 + (1639 - 1639)^2} = 0 \]

On the table is done grouping clusters according to the cluster by using the formula:

\[ \text{IF} \ (\text{AND} \ (C4 < D4, C4 < E4, C4 < F4), 1, \ \text{IF} \ (\text{AND} \ (D4 < C4, D4 < E4, D4 < F4), 2, \ \text{IF} \ (\text{AND} \ (E4 < C4, E4 < D4, E4 < F4), 3, 4))) \]

The process of the above calculation is done for each item sold so can result in the calculation of literacy table 1

Table 4. Literacy 1

Once can literacy first, then the cluster grouping to get a new cluster center value by using means

a. Grouping value of 1

1) If the cluster = 1, then write said first (2015), otherwise it will be 0, so based on the first number, then the result is 1556

2) If the cluster = 1, then write said second (2016), otherwise it will be 0, so based on the second number, then the result is 2130

3) If the cluster = 1, then write said third (2017), otherwise it will be 0, so by the third number, then the result is 2064

4) If the cluster = 1, then write said fourth (2018), otherwise it will be 0, so by the fourth number, then the result is 1639

b. Grouping cluster 2

1) If the cluster = 2, then write said first (2015), otherwise it will be 0, so based on the first number, then the result is 0

2) If the cluster = 1, then write said second (2016), otherwise it will be 0, so based on the second number, then the result is 0