Data Mining in Sales Data Grouping

Supriyati1*, S R Abdillah2
1Departemen Komputerisasi Akuntansi, Universitas Komputer Indonesia, Indonesia
2Departemen Sistem Informasi, Universitas Komputer Indonesia, Indonesia
Email: *supriyati@email.unikom.ac.id

Abstract. The purpose of making data mining is to make sales data neatly organized and based on sales data can see sales potential that can increase company profits. Selling goods requires a new method for analyzing data in sales and data mining is one way to analyze sales data created in an increasingly advanced business world. This study used a qualitative method and data collection techniques using literature reviews and observations with secondary data types. The results of this study are when processing data, there are many ways in data processing both in terms of data and in terms of processing. The compilation in subsequent processing becomes information that is easy to read and the data becomes the beginning of the information that we previously processed from several existing methods and data developed into information that is easy to read.

1. Introduction
The development of the technology nowadays has many characteristics in each of its fields so that everyone can compete in any field that is related to technology that is getting more sophisticated, many of us do not know different ways in processing the information and the data that we have in the modern world. Ways to process the data can be done manually and digitally or computerized. Therefore, we also know what differences and comparisons are easier to process because not every data we get can be processed digitally or vice versa.

Data mining usually has a characteristic with the process of finding a quite large data and many are grouped into neatly arranged data. In this case, we took sales data and processed it with clustering techniques [1]. The set cluster variable is a set of variables that represent the characteristics used by the object [2]. Marketing is a social process where individuals and groups get what they need and want through creating, offering, and exchanging free value products and services with others. As a managerial definition, marketing is often done to promote products or services they offer [3]. RAPIDMINER is the most popular open-source software in the world for data mining, and strongly supports text mining and other data mining techniques that are applied in combination with text mining. [4]. K-Means is an algorithm that can be integrated into suboptimal partitions since the stochastic optimization approach is good at avoiding convergence for locally optimal solutions [5]. Marketing research is an objective and systematic process in gathering information used to help in making decisions related to marketing [6].

The purpose of making data mining is to make sales data neatly organized and based on sales data can see sales potential that can increase company profits. This study used a qualitative method and data collection techniques using literature reviews and observations with secondary data types.

2. Method
The method used in this research is a qualitative method using literature reviews and observations with secondary data types. Research data collection from this method is obtained by searching from relevant sources.
3. Results and Discussion

There are many ways in data processing both in terms of data and processing. In the subsequent processing, everything becomes information that is easy to read, the data becomes the beginning of the information that we previously processed from several existing methods and makes the data developed into information that is easy to read (see Table 1).

Table 1. Sales of computer equipment

| NO | Item          | Value | Purchase Price | Total Buy   | Selling Price | TOTAL Sales | DISCO UNT | Selling Value | Gross     |
|----|---------------|-------|----------------|-------------|---------------|-------------|-----------|---------------|-----------|
| 1  | Hard disk     | 6     | Rp600,000.00   | Rp3,600,000.00 | Rp700,000.00 | Rp4,200,000.00 | 5%        | Rp670,000.00   | Rp70,000.00|
| 2  | RAM           | 2     | Rp200,000.00   | Rp400,000.00  | Rp220,000.00  | Rp440,000.00  | 2%        | Rp216,000.00   | Rp16,000.00|
| 3  | Keyboard      | 1     | Rp100,000.00   | Rp100,000.00  | Rp125,000.00  | Rp125,000.00  | 2%        | Rp123,000.00   | Rp23,000.00|
| 4  | Monitor       | 4     | Rp1,000,000.00 | Rp4,000,000.00| Rp1,500,000.00| Rp6,000,000.00| 15%       | Rp1,350,000.00 | Rp350,000.00|
| 5  | Webcam        | 2     | Rp50,000.00    | Rp100,000.00  | Rp80,000.00   | Rp160,000.00  | 5%        | Rp77,500.00    | Rp27,500.00|
| 6  | Mouse game    | 15    | Rp100,000.00   | Rp1,500,000.00| Rp150,000.00  | Rp2,250,000.00| 2%        | Rp148,000.00   | Rp48,000.00|
| 7  | VGA           | 3     | Rp50,000.00    | Rp150,000.00  | Rp70,000.00   | Rp210,000.00  | 10%       | Rp65,000.00    | Rp15,000.00|
| 8  | Speaker       | 4     | Rp75,000.00    | Rp300,000.00  | Rp85,000.00   | Rp340,000.00  | 15%       | Rp73,750.00    | Rp27,500.00|
| 9  | CPU           | 2     | Rp1,000,000.00 | Rp2,000,000.00| Rp1,500,000.00| Rp3,000,000.00| 10%       | Rp1,400,000.00 | Rp400,000.00|
| 10 | Motherboard   | 1     | Rp2,000,000.00 | Rp2,000,000.00| Rp3,500,000.00| Rp3,500,000.00| 5%        | Rp3,400,000.00 | Rp1,400,000.00|
| 11 | Flashdisk     | 4     | Rp40,000.00    | Rp160,000.00  | Rp50,000.00   | Rp200,000.00  | 15%       | Rp44,000.00    | Rp4,000.00|
| 12 | Earphone      | 2     | Rp30,000.00    | Rp60,000.00   | Rp40,000.00   | Rp80,000.00   | 20%       | Rp34,000.00    | Rp4,000.00|
| 13 | Modem         | 2     | Rp125,000.00   | Rp250,000.00  | Rp135,000.00  | Rp270,000.00  | 15%       | Rp116,250.00   | Rp8,750.00|
| 14 | Wi-fi         | 1     | Rp250,000.00   | Rp250,000.00  | Rp260,000.00  | Rp260,000.00  | 5%        | Rp247,500.00   | Rp2,500.00|
| 15 | Repeater      | 2     | Rp150,000.00   | Rp300,000.00  | Rp160,000.00  | Rp320,000.00  | 10%       | Rp145,000.00   | Rp5,000.00|
|    | Total         |       | Rp15,170,000.00| Rp21,355,000.00| Rp8,110,000.00| Rp2,340,000.00|           |               |           |

3.1 Process

In processing the existing data, it requires a formula in the rapid miner software that supports data processing and connected to some syntax that makes the data start processing the data. The read excel section explains which excel data section and which excel data will be inputted into the processing so that it can make new results that can be processed further in the clustering. The clustering start grouping based on the values needed in the process. The performance part is to get results so that it makes it easier to read the results (Figure 1).

Figure 1. Process
3.2 Data Cluster
The cluster data contains images containing data that has been grouped with their respective clusters so that we can know which parts and groups are related to each other. The process of identifying natural groupings in multidimensional data is based on several similarity [7] (Figure 2).

![Figure 2. Data Cluster](image)

3.3 Statistic
The statistics show the smallest, the largest, and the average data itself so that we can know directly there is no need to manually search for these data. The row tables all of these classifications can be retrieved and data processing is performed periodically. The correlation coefficient (r) between the results of two measurement methods as an indicator of the agreement [8] (Figure 3).
3.4 Table Centroid

The centroid table functions as a way of division within a group so that it can be a known benchmark that approaches within its group. A lot of data that is more or less close to the centroid will enter into the centroid itself. In the centroid-based classification algorithm, documents are represented using a model vector space [9] (Figure 4).

---

**Figure 3. Statistic**

**Figure 4. Table Centroid**
3.5 Visual Data

The visual reasoning approaches to extract and understand information encoded in data sets are the focus of what is called visual data mining [10] (Figure 5).

![Figure 5. Visual Data](image)

4. Conclusion

From the results, it can be concluded that using the clustering method in the data can be grouped in several clusters and the targeted value in the cluster finds many new factors that make the data easier to read. Besides, it can be valued just like the average of the values even all the specifications needed in the form of graphics and information.

Acknowledgement

We would like to thank all people who have helped, shared knowledge, also provided many inputs to our research.

References

[1] Nasari, F., & Darma, S. 2013. Penerapan k-means clustering pada data penerimaan mahasiswa baru (studi kasus: universitas potensi utama). SEMNASTEKNOMEDIA ONLINE, 3(1), pp. 21.
[2] Javadi, S., Hashemy, S. M., Mohammadi, K., Howard, K. W. F., & Neshat, A. 2017. Classification of aquifer vulnerability using K-means cluster analysis. Journal of hydrology, 549, pp. 27-37.
[3] Basalingappa, A., & Kumar, K. 2018. Marketing, communication and democracy: Towards a dialogue between the disciplines of marketing and communication. The Marketing Review, 18(2), pp. 149-160.
[4] Osman, C. C., & Zalhan, P. G. 2016. From natural language text to visual models: A survey of issues and approaches. Informatica Economica, 20(4), p. 44.
[5] Soleimany, G., & Abessi, M. 2019. A New Similarity Measure for Time Series Data Mining Based on Longest Common Subsequence. *American Journal of Data Mining and Knowledge Discovery*, 4(1), pp. 32.

[6] Soegoto, E. S. 2014. Entrepreneurship Becomes a Revised Edition Businessman. *Elex Media Komputindo*, 45(3), pp. 80-98.

[7] Teichgraeber, H., & Brandt, A. R. 2019. Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison. *Applied energy*, 239, pp. 1283-1293.

[8] Dawber, R., Armour, K., Ferry, P., Mukherjee, B., Carter, C., & Meystre, C. 2019. Comparison of informal caregiver and named nurse assessment of symptoms in elderly patients dying in hospital using the palliative outcome scale. *BMJ supportive & palliative care*, 9(2), pp. 175-182.

[9] Liu, C., Wang, W., Tu, G., Xiang, Y., Wang, S., & Lv, F. 2017. A new Centroid-Based Classification model for text categorization. *Knowledge-Based Systems*, 136, pp. 15-26.

[10] Abdelhafiez, H. A., & Amer, A. A. 2019. The Challenges of Big Data Visual Analytics and Recent Platforms. *World of Computer Science & Information Technology Journal*, 9(6), pp. 1-17.