

Pastry Sales Data Clustering with K-Means Clustering Approach for Product Grouping

M. Faiz Firdausi^{1*}, Eko Afrianto², Fery Wiranto³
Institut Teknologi dan Sains Mandala, Jember, Indonesia¹²³

Corresponding Author:

Corresponding Author: M. Faiz Firdausi, Institut Teknologi dan Sains Mandala, Jember, 68121, Indonesia
Email: faizfirdausi@itsm.ac.id

Abstract

In the business world that is run by many people today, we are required to always develop our business so that it always develops to make a profit. To achieve this, there are several things that can be done, namely by increasing product quality, adding product types, and reducing company operational costs by using company data analysis. Mr Dosi's Pastry Shop is a shop engaged in the sale of packaged dry food. Data on the purchase of MSME packaged pastry stock was not taken into account, which then led to a buildup of packaged food which made sales turnover less effective. Packaged pastries are a type of food that is packaged in packs per kilogram. Efforts are needed to find out the causes of sales that have decreased so that sales targets cannot be achieved. Related to the process of grouping the data, a data mining technique is used to perform clustering. This clustering process can be done using K-Means Clustering. This algorithm performs grouping of data sets into a predetermined cluster which aims to form separate data groups that have similarities. The results of the clustering of product sales levels are then used as a reference in warehouse management. Based on the grouping results, the grouping with 2 clusters is the most optimal grouping result with the smallest Davies-Bouldin Index (DBI) value, namely 0.125.

Keywords : data mining, clustering, K-Means, Davies-Bouldin Index (DBI)

1 INTRODUCTION

Technological developments from year to year are increasingly advanced and rapid, one of which is the internet. The internet is a means of information and communication that is fast and accurate. This makes many parties take advantage of internet media for various kinds of interests, one of which is for business purposes [1]. In the business world, competition between business people is common. In order to survive the competition, companies are inevitably required to develop and take advantage of existing opportunities in various fields and business needs. Judging from the times, the biggest opportunities that can support the company's business today are in the field of technology and information systems. Business needs also still have a big influence in supporting the company, there are three business needs that can be fulfilled, namely increasing product capacity, reducing operational costs, expanding marketing and profitability [2].

In the business world that is run by many people today, we are required to always develop our business so that it always develops to make a profit. To achieve this, there are several things that can be done, namely by increasing product quality, adding product types, and reducing company operational costs by using company data analysis. However, this is rarely or even not used by some MSMEs, one of which is the Pak Dosi Pastry Shop in East Kasiyan Village, Puger District, Jember Regency. Pak Dosi's Pastry Shop is a shop engaged in the sale of packaged dry food. The purchasing data for the stock of MSME packaged pastries was not taken into account, which then led to a buildup of packaged food which made sales turnover less effective. Packaged pastries are a type of food that is packaged in packs per kilogram. Efforts are needed to find out the causes of sales that have decreased so that sales targets cannot be achieved [3]. Store owners will be interested to know that there are several groups of pastries stock consistently on the shopping list. Store owners can use this data in setting store layouts to optimally place stock of pastries in relation to one another [4]. One effort to be able to balance consumer demand is to know consumer developments and changing consumer interests. To deal with this problem, we need a strategy to sell goods that are good for processing, one way is to use data mining. Systems development methods are methods, procedures, concepts work that will be used to develop

an information system [5]. Data mining is a process of extracting or extracting large amounts of data and information, which were not previously known, but understandable and useful from large databases and used to make very important business decisions [6].

The application of data mining is one way to overcome this, because data mining can provide knowledge that was previously hidden in the data warehouse so that it can become valuable information, and from this information can be used to determine a decision to achieve the desired goal. Classification is a process used to categorize a group of objects into certain groups [7]. The use of data mining science has been carried out by several companies to process data owned to produce information that can be used to determine strategies to be carried out to improve company performance, one of which is in the fashion industry, data mining is the exploration and evaluation of large information to find patterns. and significant rules. This science refers to the process of finding hidden information from large amounts of data through algorithms [8].

Related to the process of grouping the data, a technique is used to perform clustering. This clustering process can be done using K-Means Clustering. This algorithm performs grouping of data sets into a predetermined cluster which aims to form separate data groups that have similarities. The results of the clustering of product sales levels are then used as a reference in warehouse management. This concept manages products that are less desirable and unsold which causes the company's income to not be optimal, one of which is by giving discounts on products that are not selling well and making offers for products that are cheaper. [9]. From this discussion, this research was carried out to be able to assist companies in obtaining information about products based on their sales level using data mining techniques using the K-Means Clustering method. The grouping results will be used as a reference in implementing the down-selling strategy.

2 RESEARCH METHOD

In this study, researchers used a quantitative approach, namely data taken in the form of nominal / exact numbers or data that can be measured on a numerical scale. Data mining development method that used to analyze the data in the application This data mining uses a staged process knowledge discovery in databases (KDD) which consists of Data, Data Cleaning, Data transformation, datamining [1]. Researchers have divided this research into several stages, the following are the stages of the research methodology process shown in Figure 1 [2].

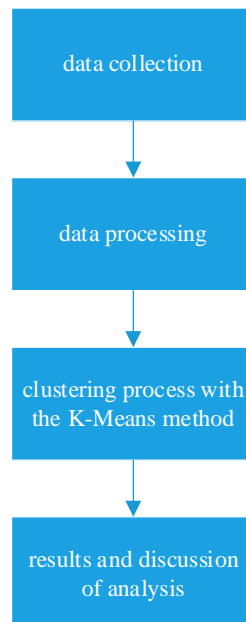


Figure 1. Research Stages

2.1 Data Collection Stage

In carrying out the data mining process in the case study of Mr. Dosi's pastry shop, real and valid data is needed to get the best analysis. The data was taken from the transaction data of Mr. Dosi's pastry shop during the sales period during the fasting month of Ramadan ahead of Hari Raya. The data retrieval process is done manually because the data is still in hardfile / bookkeeping form and will be entered into Microsoft Excel. At

this stage, the resulting output is raw data (data collected directly from the source without any changes) and is stored in the Rapidminer Application database and is shown in Figure 2..

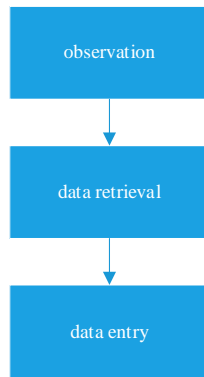


Figure 2. Data Collection Stage

2.2 Data Processing Stage

In the data processing process, the data is differentiated based on the type of pastry product, initial stock, sold, remaining stock, and price per kilogram, which is then followed by data cleansing, data selection and data transformation. Data cleansing is the process of cleaning missing data, the data is empty and inconsistent, data selection is the process of selecting data to be used for analysis in the next stage and data transformation is the data that has been selected to be changed in a certain form for the analysis stage [10] and shown in Figure 3. In this stage, the resulting output is data that is ready to be tested and analyzed and stored in Microsoft Excel format so that it can be processed at the K-Means clustering stage (shown in Figure 3).

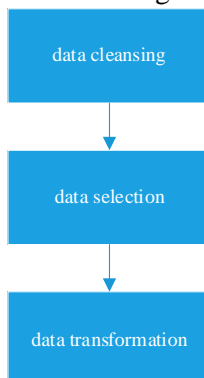


Figure 3. Data Processing Stage

2.3 K-Means Clustering Stage

At this stage, the data processing uses the Rapidminer application which functions to help perform data processing or conduct data training especially to take the K-Means Clustering function. [11]. The resulting output is a visualization of clustering results and presented in a table to obtain the results of data analysis. The stages of K-Means Clustering can be seen in Figure 4.

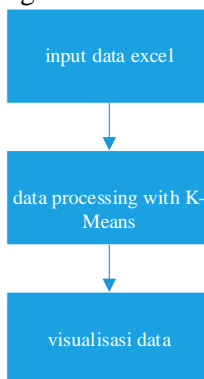


Figure 4. The K-Means Clustering Stage

2.4 Analysis Phase

At this stage data analysis of pastry products that are of public interest is carried out. The data will be differentiated based on the amount of initial stock, sold, remaining stock, and price per kilogram. The data will be analyzed from visual data in the form of tables from the results that have been carried out at the clustering stage. At this analysis stage, the output produced is that the researcher obtains the results of this study..

3 RESULTS AND ANALYSIS

The data in this study were sourced from Pak Dosi's pastry shop transaction data during the sales period during the fasting month of Ramadan ahead of Hari Raya. The data goes through a data pre-processing stage before being used to carry out the clustering process on several attributes. Of the 4 attributes used in the calculation process, namely initial stock, sold, remaining stock, and price per kilogram.

Cleansing data is to reduce noise that can affect calculations. In the data cleansing process, data that has vacant values on 4 attributes are not used, so the data used is 40 types of pastries out of 46 pastries. The data that has been processed for data cleansing can be seen in Table 1.

Table 1. Data on the Cleansing Results of Pastry Products

| No. | Cake Name | First Stock | Unit | Sale | Leftover Stok | Price/Kg |
|-----|------------------|-------------|--------|------|---------------|----------|
| 1. | Inaco | 10 | Karton | 10 | 0 | 28.000 |
| 2. | Untir-Untir Mini | 30 | Kg | 29 | 1 | 32.000 |
| 3. | Koro Kupas Pedas | 30 | Kg | 25 | 5 | 35.000 |
| 4. | Kacang Bali | 10 | Kg | 5 | 5 | 65.000 |
| ... | ... | ... | ... | ... | ... | ... |
| 39. | Mente Matang | 10 | Kg | 5 | 5 | 140.000 |
| 40. | Keripik Pisang | 20 | Kg | 18 | 2 | 55.000 |

This study aims to find the best cluster distribution by measuring the DBI value for grouping sales products. Furthermore, the mining process is carried out with the aim of finding information or patterns for clustering using the K-Means algorithm. The implementation of the K-Means algorithm in this study uses the Rapidminer software. Based on research [12] K-Means clustering modeling with Rapidminer can be seen in Figure 5.

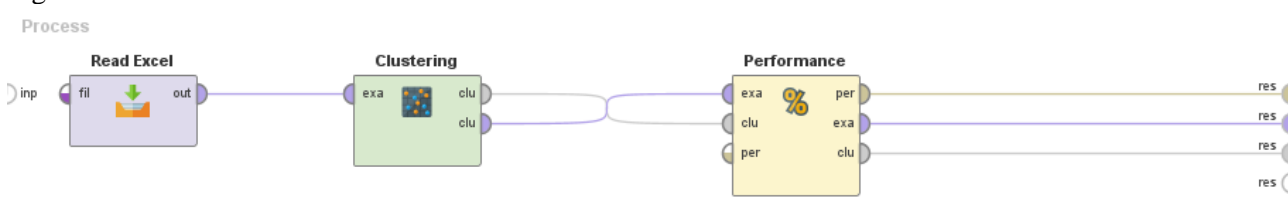


Figure 5. The K-Means Algorithm Model

This study aims to compare which group is the most optimal with the K-Means method in grouping product data selling pastries in clusters 1 to cluster 10 with the comparison results shown in table 2.

Table 2. K-Means Modeling Results

| Cluster | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---------------------|----|----|---|----|----|----|----|---|---|----|
| Amount of data k=2 | 39 | 1 | - | - | - | - | - | - | - | - |
| Amount of data k=3 | 27 | 12 | 1 | - | - | - | - | - | - | - |
| Amount of data k=4 | 12 | 1 | 4 | 23 | - | - | - | - | - | - |
| Amount of data k=5 | 16 | 1 | 4 | 9 | 10 | - | - | - | - | - |
| Amount of data k=6 | 9 | 1 | 5 | 1 | 9 | 15 | - | - | - | - |
| Amount of data k=7 | 14 | 1 | 1 | 5 | 9 | 7 | 3 | - | - | - |
| Amount of data k=8 | 9 | 1 | 1 | 5 | 3 | 6 | 10 | 5 | - | - |
| Amount of data k=9 | 3 | 1 | 5 | 1 | 3 | 9 | 5 | 9 | 4 | - |
| Amount of data k=10 | 9 | 1 | 1 | 4 | 2 | 5 | 6 | 5 | 4 | 3 |

After clustering using the K-Means method, the next stage of this research is to determine the most optimal number of clusters with the Rapidminer application seen from the Davies-Bouldin Index (DBI) value. This

process is carried out to find out the DBI value in each cluster. The test was carried out from clusters k=2 to k=10. The following results of the comparison of the DBI values of each cluster can be seen in table 3.

Table 3. DBI Comparison Results

| Cluster | Nilai DBI |
|---------------------|-----------|
| Amount of data k=2 | 0,125 |
| Amount of data k=3 | 0,368 |
| Amount of data k=4 | 0,410 |
| Amount of data k=5 | 0,442 |
| Amount of data k=6 | 0,400 |
| Amount of data k=7 | 0,363 |
| Amount of data k=8 | 0,366 |
| Amount of data k=9 | 0,384 |
| Amount of data k=10 | 0,378 |

To see the results of the comparison of the DBI values of the test results for each cluster on the K-Means method can be seen in Figure 6.

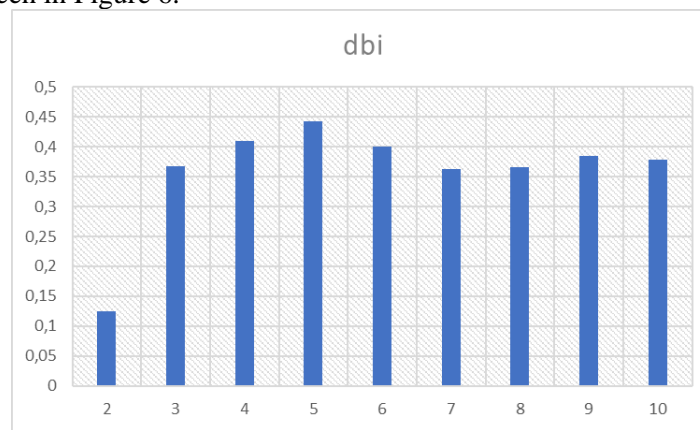


Figure 6. Comparison of the DBI cluster values of the K-Means Algorithm

After making a comparison of the DBI with the K-Means method from k=2 to k=10 in table 3, the smallest DBI value in K-Means is located in the 2nd K-Means cluster with a DBI value of 0.125. The number of 2 clusters produces better quality clusters compared to the number of clusters 3 to 10

Based on table 2 and table 3, the results of this comparison get an analysis that product sales with the 2 best clusters group the best-selling pastry product clusters as many as 39 types and the least-selling cake cluster is 1 type of cake. The results of grouping the clusters are used as a reference for grouping in managing inventory of sales products to minimize losses.

4 CONCLUSION

The application of the K-Means method in grouping product data on sales of pastries can produce the 2 best cluster groups, namely the best-selling product group and the least-selling product group. The clustering results can become Mr. Dosi to manage inventory in the warehouse optimally because he already knows what products are in the salable and unsold categories, so that later he can minimize the level of losses in sales. Based on the grouping results, the grouping with 2 clusters is the most optimal grouping result with the smallest DBI value, namely 0.125.

REFERENCES

- [1] A. Nugraha, O. Nurdiawan, and G. Dwilestari, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 849–855, 2022, doi: 10.36040/jati.v6i2.5755.
- [2] W. W. Kristianto and C. Rudianto, "Penerapan Data Mining Pada Penjualan Produk Menggunakan Metode K- Means Clustering (Studi Kasus Toko Sepatu Kakikaki)," *J. Pendidik. Teknol. Inf.*, vol. 5, no. 2, pp. 90–98, 2022.

- [3] N. Kadek Sri Wangi, I. Nuridja Made, and W. Bagia, "Analisis Belum Tercapainya Target Penjualan Elektronik Dan Furniture (Sebuah Kajian Dari Perspektif Manajemen Pemasaran)," *J. Pendidik. Ekon.*, vol. 4, no. 1, 2014.
- [4] M. H. Siregar, "Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan)," *J. Teknol. Dan Open Source*, vol. 1, no. 2, pp. 83–91, 2018, doi: 10.36378/jtos.v1i2.24.
- [5] A. Wahid and R. Agustina, "Rancang Bangun Pembayaran Eklektrik di Kantin Ibbrahimy Menggunakan Model Waterfall," *Semin. Nas. FST Unikama*, vol. 1, pp. 4–13, 2018.
- [6] C. Purnama, W. Witanti, and P. Nurul Sabrina, "Klasterisasi Penjualan Pakaian untuk Meningkatkan Strategi Penjualan Barang Menggunakan K-Means," *J. Inf. Technol.*, vol. 4, no. 1, pp. 35–38, 2022, doi: 10.47292/joint.v4i1.79.
- [7] S. Aulia, "Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja)," *Djtechno J. Teknol. Inf.*, vol. 1, no. 1, pp. 1–5, 2021, doi: 10.46576/djtechno.v1i1.964.
- [8] F. P. Samudra, T. H. Pudjiantoro, and I. Santikarama, "Klasterisasi Tingkat Penjualan Produk Menggunakan Metode K-Means Untuk Penerapan Konsep Down- Selling," *Semin. Nas. Inform. dan Apl.*, pp. 9–15, 2021.
- [9] E. Eva, B. Daniawan, and A. Hermawan, "Sales Information System of Mobile Marketing at PT. World Innovative Telecommunication Using Up Selling, Cross Selling and Down Selling Methods," *J. Tech-E*, vol. 4, no. 1, pp. 1–5, 2020, doi: 10.31253/te.v4i1.411.
- [10] T. D. Astuti, T. I. Hermanto, and I. Kaniawulan, "Analisa Data Mining Menggunakan Algoritma Apriori untuk Meningkatkan Cross Selling dan Up Selling (Studi Kasus Rumah Makan Mas Nur Purwakarta)," *J. Teknol. dan Inf.*, vol. 6, no. 2, pp. 69–77, 2016.
- [11] M. Hermansyah, R. A. Hamdan, F. Sidik, and A. Wibowo, "Klasterisasi Data Travel Umroh di Marketplace Umroh.com Menggunakan Metode K-Means," *J. Ilmu Komput.*, vol. 13, no. 2, p. 8, 2020, doi: 10.24843/jik.2020.v13.i02.p06.
- [12] M. Hermansyah, N. A. Prasetyo, Y. Ansori, M. F. Firdausi, and A. Wahid, "Implementation of K-Means Clustering for Analysis Students English Proficiency," *J. Educ. Sci. Technol.*, vol. 1, no. 6, pp. 31–35, 2023.