International Journal of Engineering, Management and Humanities (IJEMH)
Volume 2, Issue 4, pp: 447-451 www.ijemh.com

Implementation of K-means in Clustering Coffee Beans Based on Sales to Fulfill Market Needs

Ari Ramadani Mubarok, Yaddarabullah, Silvester Dian Handy Permana
Department of Informatics, Trilogy University, Jakarta, Indonesia

Date of Submission: 01-08-2021 Date of Acceptance: 14-08-2021

ABSTRACT, Coffee beans are one of Indonesia's most popular exports, with a volume of 189.6 thousand tons expected in 2020. However, numerous factors influence the increase and reduction in the value of exports, one of which is unmet market needs, which can cause the export value to fall. To reduce the likelihood of this happening, data should be clustered using the K-means algorithm. Testing the data grouping using the K-means algorithm on the WEKA application generates results such as export values with three clustering, namely salable, highly salable, and unsold based on one year, according to the findings of the research. For a complete year, the export value is broken down by month for classes 0, 1, and 2. The test findings show that class 0 has one data point with a percentage of 10%, class 1 has five data points with a percentage of 50%, and class 2 has four data points with a percentage of 40%.

KEYWORDS: Clustering, Coffee Beans, K-means

I. INTRODUCTION

The basic element in the production of the world-famous beverage is coffee beans. Indonesia is one of the world's largest coffee bean producers, producing a variety of coffee beans. Given the vast number of well-managed coffee plantations in Indonesia, coffee beans are also one of the most important export commodities. Coffee bean exports totaled 186.8 thousand tons in 2020, according to the Central Statistics Agency (BPS). In comparison to the previous year, this number climbed by 10.69 percent.

This growth in number has a positive influence on coffee producers, but the rising demand for coffee beans has not yet resulted in a balanced increase in the demand for other varieties of coffee beans. [1] Indonesia produces a wide variety of coffee beans, including the well-known Arabica coffee beans, such as Gayo coffee, Java coffee, and Lintong coffee. Coffee bean sales can also be influenced by market conditions. Following the needs of the market might lead to outstanding sales outcomes. To avoid a drop in shop sales turnover, an analysis of market needs is required. To avoid this, one option for assessing the market needs for coffee beans is to group data on items sold.

Unorganized sales data might make it difficult for business actors to track the evolution of the coffee bean market. As a result, data grouping is used to lessen the impact of imperfect markets and give business actors with more coordinated data on the results of purchasing and selling coffee beans [2]. Uncoordinated purchasing and selling operations might stymie a store's growth. Because the market demand for coffee beans is so high around the world, imperfect markets oblige all business actors to do market needs analyses in order to avoid losses on the products they offer [3].

in previous research conducted by [4] the usage of the K-means algorithm with the clustering method on a store product. This strategy is intended to avoid stockpiling unsold items and to reduce the selling of expired commodities. This strategy is also useful for reducing shop losses, which might result in lower income. Other studies conducted by[5] In Indonesia, the K-means algorithm was used to determine the online business model. in numerous Indonesian provinces by grouping which consists of three types, namely vendors, resellers, and dropshippers. The goal of this study is to analyze the sales model by province based on the results of its grouping in order to increase commercial potential in all of Indonesia's provinces. Subsequent research conducted by[6] Specifically, the K-means clustering method was used to classify laptop sales levels as selling, extremely selling, and not selling. Business players can assess products sold and unsold by categorizing data depending on the results of laptop sales, allowing businesses to readily address market needs. This strategy can also help the store reduce losses by providing products that are in high demand by consumers. Previous research[7]

International Journal of Engineering, Management and Humanities (IJEMH)
Volume 2, Issue 4, pp: 447-451 www.ijemh.com

focuses on determining the existing menu packages in the bakery using a combination of K-means and FP-Growth Algorithms. The purpose of the grouping in this study is to establish which products sell and which do not, in order to minimise the losses incurred by the store as a result of providing products that do not sell. Furthermore, the fp-growth algorithm searches for links to current products that can be utilized as a suitable recommendation menu, allowing retailers to identify consumer interest based on the grouping results. Further research[8] focuses on partitioning the load curve of residential electricity consumers by group using K-means. So that the power company can assess electricity users' use and regulate electricity sales and purchases.

The researcher's goal in this study is to cluster coffee bean sales in order to meet market demand. To calculate huge data, the researcher employs the K-means clustering approach. The K-means algorithm separates data into a number of clusters using a distance-based clustering method. [9]. To get group data and clustering accuracy, the calculated results will be clustered using the K-means algorithm. On the basis of the foregoing, a study was undertaken on the use of K-means in clustering coffee beans based on sales to meet market needs.

II. RESEARCH METHODOLOGY

This study's approach will be outlined in terms of numerous steps that will be completed. The stages of the research technique will be discussed in the following order.

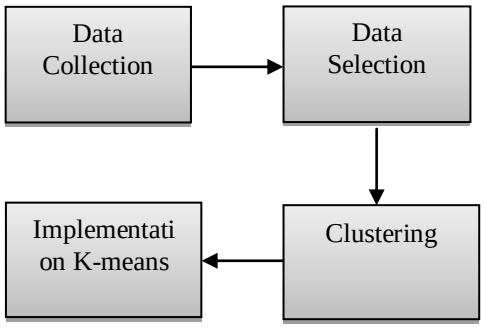

Figure1. Reasearch Stages

2.1 Data collection

The purpose of this stage of data collecting is to determine the basic information required for this study. The information gathered is based on coffee bean export sales by product type in 2017 and 2018. The information was gathered from the statistical center's website. The data utilized in this analysis is the value of coffee bean exports in Indonesia for the entire year of 2018. The categories of goods and the value of exports each month for a year are the variables in this data. To manage the data, the researcher used Microsoft Excel and the WEKA application. After that, the WEKA program will use agrotima K-means clustering to assess whether the product belongs in large, medium, or small exports.

2.2 Data Selection

The purpose of data selection is to sort the data that will be used in research. The chosen data consists of numerous variables such as commodity type and export value for each month during a one-year period, which will be processed for clustering using the K-means technique.

2.3 Clustering

Clustering by [10] is a mechanism for categorizing data into many groups. Clustering is the process of measuring two or more objects using a variety of data types such as interval-scale variables, binary variables, nominal, ordinal, and ratio variables [11]. The ability of the clustering algorithm to uncover hidden patterns in the data being processed must be measured. The clustering method using the K-means algorithm is thought to be capable of assisting in the grouping of enormous amounts of data [12]. By generating classes based on data features, the K-means method makes it simple to group data.

2.4 K-means algorithm

K-means is a method of grouping data to partition data in one or more clusters[13]. The K-means method divides data into many groups to allow data with similar characteristics be grouped together, while data with different characteristics is integrated into other groups with similar features in other data [14]. There are various basic algorithm steps in finding the group of data using the K-means algorithm, as follows [15]:

1. Specifies the number of clusters.
2. Allocating data into clusters randomly.
3. Calculate the centroidb of the data in each cluster.
4. Allocate each data to the nearest centroid.
5. Return to stage 3, if there is still data moving clusters.

It can be seen the following formula 1:

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

Provided is
D = distance
X = Data

International Journal of Engineering, Management and Humanities (IJEMH)
Volume 2, Issue 4, pp: 447-451 www.ijemh.com

Y = Center
I = Lots of Data

When the distance variable is D and the data value is X, the data value must be calculated by multiplying it by the distance variable, which is Y, which is the centroid or center point of a class picked at random.

III. Results and Discussion

Researchers used ten datasets of sample data on the value of coffee exports from Indonesia in this study. The dataset includes information on the type of coffee bean product, as well as monthly statistics on the product's export value. The WEKA program will process the export value provided in the dataset as a test variable in the K-means clustering method. The K-means clustering technique was tested using the WEKA application by the researcher. The WEKA program will maintain the following table of data on the export value of coffee bean products:

Table1. Data on the export value of coffee bean products

No	Commodity Type	January	February
1.	Arabica WIB/robusta OIB, not roasted, not decaffeinated	50745	44585
2.	Coffee oth than Arabica WIB/ robusta OIB, not roasted, not decaffeinated	354	535
3.	Arabica WIB/robusta OIB, not roasted, decaffeinated	33	37
4.	Coffee oth than Arabica WIB/ robusta OIB, not roasted, decaffeinated	10	5
5.	Coffee, roasted, not decaffeinated, unground	83	7
6.	Coffee, roasted, not decaffeinated, ground	675	703
7.	Coffee, roasted, decaffeinated, unground	1	0
8.	Coffee, roasted, decaffeinated, ground	1	0
9.	Coffee husks and skins	0	0
10.	Coffee substitutes contain coffee	0	0

The table above shows several coffee bean products with export values in US dollars. Processed coffee beans are distinguished from coffee beans that have not been processed or are still in the form of raw beans by the type of commodity. Then there's the monthly export value of items, which is known from January to December, but there's also an unknown monthly export value. Based on the value of exports each month, the WEKA program will use the K-means algorithm clustering method to identify things that sell, sell well, and do not sell.

Table2. Results of data clustering

Attribute	Full Data (10.0)	Cluster 0 (4.0)	Cluster 1 (5.0)	Cluster 2 (1.0)
Commodity Type	Arabica WIB/robusta OIB, not roasted, not decaffeinated	Coffee oth than Arabica WIB/ robusta OIB, not roasted, not decaffeinated	Coffee, roasted, not decaffeinated, ground	Arabica WIB/robusta OIB, not roasted, not decaffeinated
January	5190.2	120	135.4	50745
February	4587.2	146	140.6	44585
March	5907.2	62.5	140.6	58119
April	6281.1	37	103.8	62144
May	7729.9	207.75	10.4	76416
June	7005.1111	97.25	1420.0222	62552
July	8688.6	322.75	26.4	85463
August	8224.4	126	97.6	81252
September	6794.6	338.5	155.8	65813
October	7075	166.5	184.6	68981

International Journal of Engineering, Management and Humanities (IJEMH)
Volume 2, Issue 4, pp: 447-451 www.ijemh.com

November	7264.2	158	155.8	71231
December	81560.1	2035	1331.4	800804

From the table above, we get a result such as the value of exports with three clustering, namely selling, very selling and not selling based on one year. The export value is broken down based on each month for a full year with classes 0, 1 and 2. In the table there are types of commodities that have export values according to the observations for one full year. The results in the test are class 0 contains 1 data with a percentage of 10%, class 1 contains 5 data with a percentage of 50% and class 2 contains 4 data with a percentage of 40%. These results can be used to follow the development flow of the coffee bean market in the world so that the products to be exported can be prepared properly to meet market needs.

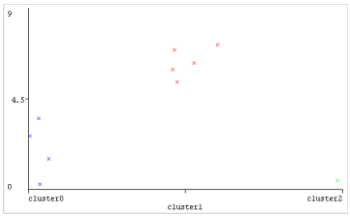

Figure2. Graph of clustering result

The picture above is a visualization of the grouping results using the WEKA application. The export value is broken down by three classes and types of commodities. The results in the test are class 0 contains 1 data with a percentage of 10%, class 1 contains 5 data with a percentage of 50% and class 2 contains 4 data with a percentage of 40%.

IV. CONCLUSION

Based on the results of the research above, a conclusion can be drawn that testing data grouping using the K-means algorithm on the WEKA application produces results such as export values with three clusterings, namely selling, very selling and not selling based on one year. The export value is broken down by month for a full year with classes 0, 1 and 2. The results in the test are class 0 contains 1 data with a percentage of 10%, class 1 contains 5 data with a percentage of 50% and class 2 contains 4 data with a percentage of 40 %. With these results, it can be proven that the clustering using the K-means algorithm was successful with the results as above.

References

[1]. ADDIN Mendeley Bibliography CSL_BIBLIOGRAPHY [1] Y. B. S. Panggabean, M. Arsyad, Mahyuddin, and Nasaruddin, "Coffee farming business development: E-commerce technology utilization," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 807, no. 3, p. 032011, 2021, doi: 10.1088/1755-1315/807/3/032011.

[2]. T. Karyani, E. Djuwendah, A. Hermita Sa, S. Kirana, and N. Risti Muti, "Arabica Coffee Agroindustry Cost Requirement Analysis at Margamulya Coffee Producers Cooperative," *Asian J. Agric. Res.*, vol. 12, no. 1, pp. 1–9, 2017, doi: 10.3923/ajar.2018.1.9.

[3]. E. Djuwendah, T. Karyani, E. Rasmikayati, S. Fatimah, and Deliana, "The Supporting and Impeding Factors of Java Preanger Coffee Agribusiness on Margamulya of Pangalengan Sub District of Bandung Regency," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 166, no. 1, 2018, doi: 10.1088/1755-1315/166/1/012043.

[4]. M. Imron, U. Hasanah, and B. Humaidi, "Analysis of Data Mining Using K-Means Clustering Algorithm for Product Grouping," *IJIIS Int. J. Informatics Inf. Syst.*, vol. 3, no. 1, pp. 12–22, 2020, doi: 10.47738/ijiis.v3i1.3.

[5]. R. R. Aria, "K-Means to Determine the e-commerce Sales Model in Indonesia," *Int. J. Inf. Syst. Technol. Akreditasi*, vol. 3, no. 36, pp. 166–172, 2020.

[6]. U. Hasanah and C. Latiffani, "Laptop Sales Level Using the K-Means Clustering Method," vol. 4509, no. 1, pp. 1–7, 2020.

[7]. N. P. Dharshinni, E. Bangun, S. Karunia, R. Damayanti, G. Rophe, and R. Pandapotan, "Menu Package Recommendation using Combination of K-Means and FP- Growth Algorithms at Bakery Stores," *J. Manik*, vol. 3, no. 2, pp. 10–19, 2019.

[8]. Y. He, Z. Jiao, F. Chen, F. Guang, P. You, and Q. He, "Residential power user segmentation based on k-means clustering method in the context of big data," *E3S Web Conf.*, vol. 53, pp. 1–4, 2018, doi: 10.1051/e3sconf/20185302006.

[9]. A. S. Ahmar, D. Napitupulu, R. Rahim, R. Hidayat, Y. Sonatha, and M. Azmi, "Using K-Means Clustering to Cluster Provinces in Indonesia," *J. Phys. Conf. Ser.*, vol. 1028, no. 1, 2018, doi: 10.1088/1742-6596/1028/1/012006.

[10]. J. Xu and K. Lange, "Power k -Means Clustering," 2019.
[11]. I. O. P. C. Series and M. Science, "Recommendation Product Based on Customer Categorization with K-Means Clustering Method Recommendation Product Based on Customer Categorization with K-Means Clustering Method," pp. 1–7, 2019, doi: 10.1088/1757-899X/508/1/012123.
[12]. V. Cohen-addad, B. Guedj, and G. Rom, "Online k -means Clustering."
[13]. W. Luyao, F. Hong, and G. Tianren, "The sales behavior analysis and precise marketing recommendations of FMCG retails based on geography methods," no. November, pp. 1–14, 2017, doi: 10.20944/preprints201711.0115.v1.
[14]. K. P. Sinaga and M. Yang, "Unsupervised K-Means Clustering Algorithm," vol. 8, 2020, doi: 10.1109/ACCESS.2020.2988796.
[15]. O. Access, "Developing cluster strategy of apples dodol SMEs by integration K-means clustering and analytical hierarchy process method Developing cluster strategy of apples dodol SMEs by integration K-means clustering and analytical hierarchy process method," 2018.