Implementation of K-Means Clustering for Inventory Projection

Zulham Sitorus[1*], Irwan Syahputra[2], Chairul Indra Angkat[3], Dewi Sartika[4]

[1,2,3,4] Faculty of Postgraduate Program, Panca Budi University of Development, Medan, Indonesia
*Corresponding Author:
Email: zulhamsitorus@dosen.pancabudi.ac.id

Abstract.

Inventory forecasting is crucial in effective supply chain management and cost reduction. However, traditional forecasting techniques face significant challenges due to the complexity and variability of demand patterns. This study explores the use of K-means clustering, a data-driven approach that can improve inventory forecasting accuracy. By grouping inventory items based on their unique demand profiles, we can create personalized forecasting models for each cluster. This technique enhances demand estimation, helping businesses make informed decisions and optimize their inventory. Our research delves into the use of K-means clustering to identify patterns and similarities within historical demand data. This clustering process divides inventory items into groups with similar demand characteristics. By applying specific forecasting models to each cluster, we achieve greater precision in our predictions. The proposed methodology is rigorously evaluated using real-world inventory datasets, and the results demonstrate its significant superiority in forecasting accuracy compared to conventional non-clustered methods. This study offers compelling evidence and valuable insights for practitioners seeking to improve their inventory management practices through data-driven techniques.

Keywords: *Inventory forecasting, K-means clustering, demand patterns, supply chain management, data-driven decision making and optimization.*

I. INTRODUCTION

Inventory of goods is an important component in a company's supply chain management. Effective inventory management can help trading companies optimize operational performance and increase customer satisfaction[1].Dapur Uniq is a trading company that focuses on selling household appliances in the form of glass. The right sales strategy is crucial for Dapur Uniq owners to remain competitive in the market. The prediction results from the recommended model can be a useful tool in supporting business decision-making. One of the keys to success is improving the quality of goods and having sufficient inventory to meet customer demand. Currently, software is already used to manage the stock of goods, but there are still discrepancies that cause the inventory of goods to not align with current data sales[2]. The fluctuating demand for goods from consumers causes the stock that must be prepared to become unstable. Apart from that, the types of goods are quite diverse, making stock management less accurate, thus ineffective and often disappointing consumers because an item is missing. Availability of goods significantly influences the level of customer satisfaction[3].Furthermore, in terms of grouping goods that are selling well and those that are not, there are still difficulties.

For instance, goods that should be selling well might face shortages due to high consumer demand, while goods that are not selling well continue to accumulate because of low sales.The problems that occurred were caused by difficulties in determining the minimum stock limits for the types of goods that must be met based on consumer interest. To address this challenge, effective inventory data management is needed, involving the processing of historical transaction data using data mining techniques such as the K-Means Clustering algorithm[4].Indeed, with data that has been grouped through clustering, it is anticipated to facilitate companies in determining inventory levels. This allows companies to conduct careful procurement of goods, aiming for efficiency. Clustering methods prove highly beneficial in determining the appropriate number of clusters for sales data, aiding in segmenting products effectively based on demand patterns and optimizing inventory management strategies accordingly[5][6].In this context, clustering analysis emerges as a useful tool for grouping inventory items. Achieving optimal inventory levels presents a significant

challenge, as excessive inventory can lead to high costs, while inventory that is too low can result in stockouts, potentially leading to lost sales and dissatisfied customers[7]. Dapur Uniq is committed to managing inventory wisely to ensure product availability and customer satisfaction.

In anticipation of potential sales forecast errors that could lead to customer loss, it's essential to employ forecasting techniques supported by data mining. These techniques aid in accurate predictions, facilitating correct decision-making processes[8]. Data mining is a concept utilized to extract knowledge from within a database[9]. The data mining process encompasses the application of statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify valuable information and hidden knowledge from extensive databases[10].This analysis will concentrate on clustering inventory data for goods in the form of glass using the K-Means method. K-Means is capable of efficiently and swiftly grouping data on a large scale. However, its weakness lies in the initial selection of cluster centers, as the results of K-Means clustering are significantly influenced by the initial value of the selected cluster center.[11] The primary objective is to identify emerging patterns in glass inventory data and categorize them into significant clusters based on similar characteristics[12].

II. METHODS

2.1 Research Stages

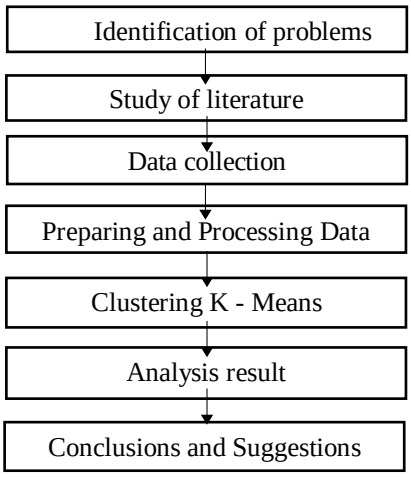

Fig 1. Research Stages

1. Identification of problems
   In this context, it is crucial to analyze the problems that occur, particularly how to effectively determine the stock of goods in Dapur Uniq based on evaluations and available data. This analysis aims to prevent errors in inventory management and accurately identify which goods need restocking. By understanding this problem thoroughly, we can devise appropriate solutions tailored to address Dapur Uniq's specific inventory challenges.
2. Study of literature
   Absolutely, conducting a literature review is indeed a crucial step in the research process. It helps in identifying theoretical references relevant to the case or problem at hand. By searching through various sources such as books, academic journals, and the internet, researchers can enrich their understanding of relevant concepts and theories. This process provides a strong foundation and in-depth knowledge essential for developing informed strategies and solutions to address the identified problem effectively.
3. Data collection
   At this stage, data and information required for research are collected through several methods, including:
   a. Observation: This method involves the direct observation of the object under study. In this case, observations were made of sales data at Dapur Uniq to obtain the necessary data.
   b. Interview: Short interviews were conducted directly with the authorities at Dapur Uniq to obtain more accurate data and information regarding sales data.

c. Library Research: To support the research, data was searched from various sources such as books and research journals that were relevant to the method used in this research, namely the K-means method.

4. Data Preparation and Processing
   The sales data is prepared for processing, requiring a data cleansing process to ensure its validity[13]. Subsequently, the data is converted to CSV format for further analysis using the Orange Data Mining application. Orange Data Mining offers sufficient features and is easily learnable for performing clustering using the K-Means method[11].
5. Clustering with K-Means
   The steps of the K-Means clustering algorithm include[14]:
   a. Determine the desired number of clusters.
   b. Random selection of initial centroids.
   c. Calculating the distance between data and the centroid to determine cluster membership.
   d. Centroid updates by calculating the average value in each cluster.
   e. Iterate the process until there are no significant cluster changes.
6. Analysis results
   Evaluation of model accuracy based on analysis results using the Clustering method with K-Means algorithm.
7. Conclusions and suggestions
   The conclusion summarizes the main findings of the study, provides a clear summary of what is known, and provides a better understanding of the topic. Meanwhile, the suggestions provide guidance on steps that can be taken based on research findings, both for further research and practical implementation.

2.2 K-Means steps

Steps for K-Means algorithm according[15].

1. Determine K as the number of clusters to be formed
2. Determine the initial k centroids (cluster center points) randomly.

$$v = \frac{\sum_{i=1}^{n} x_i}{n}; i = 1,2,3 \ldots n$$

   Where;
   v : centroid of the cluster[2]
   xi : ith object
   n : many objects/number of objects that are members of the cluster.
3. Calculate the distance of each object to each centroid in each cluster. To calculate the distance between an object and the centroid, you can use Euclidian Distance

$$d(\mathrm{x,y}) = \|x - y\| = \sqrt{\sum_{i}^{n} = 1(x_i - y_i)^2}\ \ i = 1,2,3, \ldots . n$$

   Where;
   xi : ith x object
   yi : i-th y power
   n : number of objects
4. Allocate each object to the closest centroid.
5. Perform iterations then determine the position of the new centroid using the equation.
6. Repeat step 3 if the new centroid position is not the same.

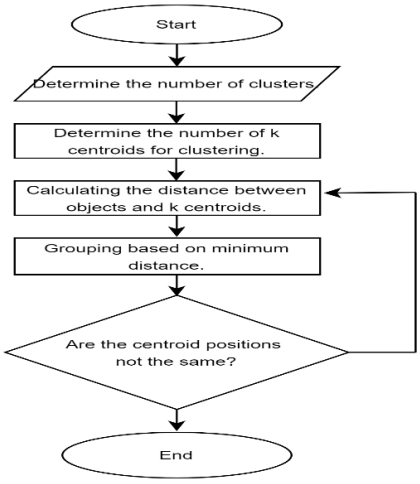

Fig 2. Flowchart of K-Means algorithm

III. RESULTS AND DISCUSSIONS

3.1 Research Scenario

The design of the K-Means clustering widget design using Orange Data Mining is as shown in the following image.

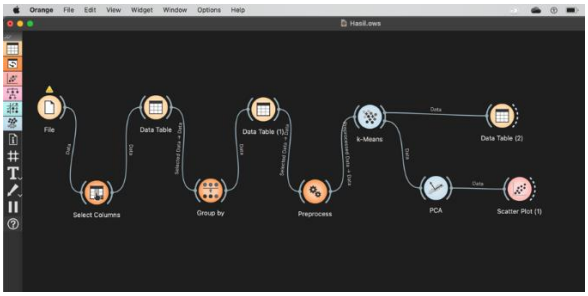

Fig 3. K-Means Clustering Widget

The data that is owned is immediately processed using the Orange application using the appropriate widget as in Picture 3 above.

3.2 Data Selection

At this selection stage the author selects the data needed to carry out the analysis process into a two column table which from the start, this data table has 4320 rows and 6 columns.

3.3 Table Data

Table data is the result of displaying which attributes have been selected for further processing as output using the group by widget.

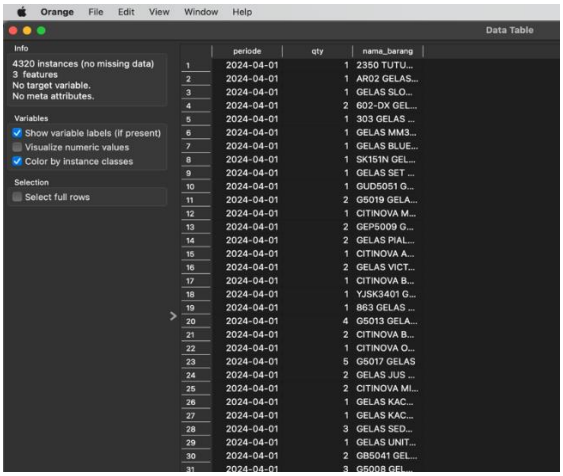

Fig 4. Table data

3.4 Group By

Group by is needed in this analysis process to calculate total sales that have the same characteristics in the Dapur Uniq store sales table.

3.5 Preprocessing

The data preprocessing stages in this research, the author uses the data preprocessing widget by Normalizing features in the orange tools then selecting standardize to$u = 0,\sigma^2=1$

3.6 K-Means

The k-means widget in this analysis process functions to determine the number of clusters needed, in this research the author used 3 clusters.

3.7 Scatter Plots

The scatter plot is used to visually display the results of the Dapur Uniq shop data process using K-Means method process to create three clusters that have been determined in the previous K-Means widget.

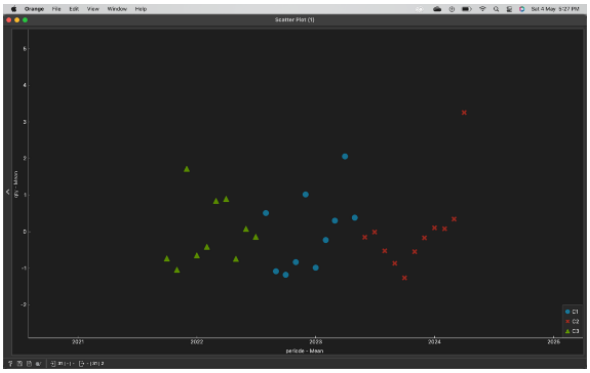

Fig 5. Scatter Plot

From Picture 5, the results of the scatter plot process show even cluster differences in K-Means process using the orange application. The results of this cluster can be seen as cluster = 1 in 2022 to 2023, cluster = 2 in 2023 to 2024, and cluster 3 in 2021 to 2022.

3.8 Results Table Data

From each process that has been carried out in the orange application, get the results which are entered back into the table widget provided in the orange application, the results of this process produce a total of 31 rows.

	nama_barang - Mode	periode	Cluster	Silhouette	periode - Mean	qty - Mean
11	804 GELAS MELAMIN PINK,UNGU,DLL	2022-08-01	C1	0.528228	2022-08-01	0.50805
12	804 GELAS MELAMIN PINK,UNGU,DLL	2022-09-01	C1	0.612927	2022-09-01	-1.08351
13	804 GELAS MELAMIN PINK,UNGU,DLL	2022-10-01	C1	0.657815	2022-10-01	-1.18035
14	804 GELAS MELAMIN PINK,UNGU,DLL	2022-11-01	C1	0.683224	2022-11-01	-0.83320
15	710 GELAS PESTA GOLDEN DRAGON	2022-12-01	C1	0.696186	2022-12-01	1.01565
16	804 GELAS MELAMIN PINK,UNGU,DLL	2023-01-01	C1	0.699105	2023-01-01	-0.98607
17	804 GELAS MELAMIN PINK,UNGU,DLL	2023-02-01	C1	0.686889	2023-02-01	-0.22960
18	804 GELAS MELAMIN PINK,UNGU,DLL	2023-03-01	C1	0.666402	2023-03-01	0.29995
19	804 GELAS MELAMIN PINK,UNGU,DLL	2023-04-01	C1	0.626031	2023-04-01	2.05602
20	31-4 ESKAN+4 GELAS VENICE	2023-05-01	C1	0.555986	2023-05-01	0.38116
21	28-4 ESKAN+4 GELAS VENICE	2023-06-01	C2	0.498104	2023-06-01	-0.15299
22	31-6 ESAKN+6 GELAS VENICE	2023-07-01	C2	0.588201	2023-07-01	-0.01153
23	804 GELAS MELAMIN PINK,UNGU,DLL	2023-08-01	C2	0.641664	2023-08-01	-0.52570
24	28-4 ESKAN+4 GELAS VENICE	2023-09-01	C2	0.671894	2023-09-01	-0.86667
25	808 GELAS MELAMIN	2023-10-01	C2	0.688384	2023-10-01	-1.26360
26	576 HOOVER GELAS BIR+TUTUP "M"	2023-11-01	C2	0.6973	2023-11-01	-0.54878
27	28-4 ESKAN+4 GELAS VENICE	2023-12-01	C2	0.700644	2023-12-01	-0.16970
28	250 GELAS MELAMIN UNICA	2024-01-01	C2	0.700142	2024-01-01	0.10344
29	8C8 GELAS MELAMIN	2024-02-01	C2	0.696553	2024-02-01	0.07916
30	J3612 GELAS 40CL VENICE/PCS	2024-03-01	C2	0.690771	2024-03-01	0.34340
31	8C8 GELAS MELAMIN	2024-04-01	C2	0.682309	2024-04-01	3.25421
1	2350 TUTUP GELAS SS 10"	2021-10-01	C3	0.684673	2021-10-01	-0.73294
2	804 GELAS MELAMIN PINK,UNGU,DLL	2021-11-01	C3	0.693662	2021-11-01	-1.04239
3	804 GELAS MELAMIN PINK,UNGU,DLL	2021-12-01	C3	0.699724	2021-12-01	1.71979
4	804 GELAS MELAMIN PINK,UNGU,DLL	2022-01-01	C3	0.702844	2022-01-01	-0.64856
5	804 GELAS MELAMIN PINK,UNGU,DLL	2022-02-01	C3	0.702092	2022-02-01	-0.41362
6	804 GELAS MELAMIN PINK,UNGU,DLL	2022-03-01	C3	0.696782	2022-03-01	0.84086
7	804 GELAS MELAMIN PINK,UNGU,DLL	2022-04-01	C3	0.683482	2022-04-01	0.89525
8	714 GELAS MINI GOBLET GOLDEN DRAG...	2022-05-01	C3	0.659093	2022-05-01	-0.74258
9	804 GELAS MELAMIN PINK,UNGU,DLL	2022-06-01	C3	0.612995	2022-06-01	0.07638
10	804 GELAS MELAMIN PINK,UNGU,DLL	2022-07-01	C3	0.531644	2022-07-01	-0.14152

Fig 6. Datatable results

IV. CONCLUSION AND SUGESSION

Conclusion

From the results of research in the application of data mining to group Dapur Uniq data using K-Means algorithm, the following conclusions are obtained:

1. The orange application really helps the author in analyzing glass sales data at the Dapur Uniq store, where the orange application has a widget that is needed in the data analysis process

2. Dapur Uniq sales data grouping uses K-Means algorithm with the results of grouping sales of goods that are easy to sell, namely the melamine glass group and those that are difficult to sell, namely mini goblet glasses.
3. The results of clustering using the Orange application produced 3 clusters with the following details: cluster 1 (C1) 10 transactions, cluster 2 (C2) 11 transactions, cluster 3 (C3) 10 transactions.

Suggestion

1. To enhance the validity and generalizability of research results, it's advisable to utilize a larger and more comprehensive dataset. Increasing the dataset size enables a more thorough analysis of clustering patterns in glassware inventories. Additionally, utilizing a dataset with more complete attributes can offer a better understanding of the factors influencing clustering patterns.
2. Following the testing of various clustering methods, it's crucial to conduct further analysis of the obtained results and their implications. This involves comprehending the disparities in clustering patterns among methods, interpreting the resulting groups, and discussing the practical application of research findings within the context of inventory management.
3. Upon analyzing glass sales data at the Dapur Uniq store, it is hoped that future researchers can delve into data analysis encompassing broader issues such as geopolitics, interest rate fluctuations, policy changes, and other relevant aspects. This broader analysis can provide deeper insights into the complexities influencing inventory management and sales dynamics.

REFERENCES

[1] J. V. A. Haryanto and W. Santosa, "Pengaruh Strategi Manajemen Rantai Pasok Terhadap Kinerja Operasional Pada Usaha Mikro Kecil Dan Menengah," ***Sains: Jurnal Manajemen dan Bisnis***, vol.15, no.1, p.63,Nov. 2022.

[2] I. Nawangsih and R. Puspita, "Implementasi Algoritma K-Means Dalam Mengkategorikan Produk Terlaris Dan Kurang Laris Pada Toko Alfamart Cikarang," ***Jurnal Pelita Teknologi***, vol. 16, no. 1, pp. 79–87.

[3] Vanessa, Lucia, Hellen Junvellen, and Meiliana Geraldin Hutagaol, "Analisis Ketersediaan Produk Terhadap Kepuasan Konsumen Berbelanja Di Maju Bersama Cabang Thamrin Plaza Medan." [Online]. Available: https://ejurnal.eka-prasetya.ac.id/index.php/

[4] S. S. Nagari and L. Inayati, "Implementation Of Clustering Using K-Means Method To Determine Nutritional Status,"***Jurnal Biometrika dan Kependudukan***,vol. 9,no.1,pp.62–68,Jun. 2020.

[5] R. Gustriansyah, N. Suhandi, and F. Antony, "Clustering optimization in RFM analysis based on k-means," ***Indonesian Journal of Electrical Engineering and Computer Science***, vol. 18, no. 1, pp. 470–477, 2019.

[6] F. Indriyani and E. Irfiani, "Clustering Data Penjualan," 2019.

[7] Subekti and Y. Nursyanti, "Optimasi Persediaan Dengan Pendekatan Deterministik Dinamis Pada Industri Manufaktur**," *Jurnal Teknologi dan Manajemen Industri Terapan (JTMIT),*** vol. 2, 2023.

[8] J. Nasir, "PENERAPAN DATA MINING CLUSTERING DALAM MENGELOMPOKAN BUKU DENGAN METODE K-MEANS**," *Jurnal SIMETRIS,*** vol. 11, no. 2, 2020.

[9] I. Anggriawan and W. Gunawan, "Implementation of Data Mining Using K-Means Algorithm for Bicycle Sales Prediction**," *ILKOM Jurnal Ilmiah,*** vol. 14, no. 3, pp. 284–293, Dec. 2022.

[10] Y. Yulianti, D. Y. Utami, and N. Hikmah, "Implementasi Data Mining Menentukan Game Android Paling Diminati Dengan Algoritma Apriori**," *Paradigma - Jurnal Komputer dan Informatika,*** vol. 21, no. 1, pp. 29–34, Mar. 2019.

[11] H. Suroyo, *Penerapan Machine Learning dengan Aplikasi Orange Data Mining Untuk Menentukan Jenis Buah Mangga*. Seminar Nasional Teknologi Komputer & Sains (SAINTEKS), 2019

[12] T. Akhir, "Analisis Cluster Dengan Metode K-Means Pada Persebaran Kasus Covid-19 Berdasarkan Provinsi Di Indonesia."

[13] A. P. Joshi and B. V.Patel, "Data Preprocessing: The Techniques for Preparing Clean and Quality Data for Data Analytics Process," ***Oriental journal of computer science and technology***,vol. 13,no.0203,pp.78–81,Jan. 2021.

[14] M. Zubair, M. A. Iqbal, A. Shil, M. J. M. Chowdhury, M. A. Moni, and I. H. Sarker, "An Improved K-means Clustering Algorithm Towards an Efficient Data-Driven Modeling," *Annals of Data Science*, 2022.

[15] S. Setiawan, "ANALISIS CLUSTER MENGGUNAKAN ALGORITMA K-MEANS UNTUK MENGETAHUI KEMAMPUAN PEGAWAI DIBIDANG IT PADA CV. ROXED LTD**," *Jurnal Pelita Informatika***, vol. 7, no. 3, 2019.