Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 712−720
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5524

Implementation of K-Means Clustering Algorithm to Determine the Best-Selling Snack in MSMEs

Ratih Ayuni[1*], Berlilana[2]

[1] Faculty of Computer Science, Study Program of Informatics, Universitas Amikom, Purwokerto, Indonesia
[2] Faculty of Computer Science, Study Program of System Information, Universitas Amikom, Purwokerto, Indonesia
Email: [1*]ratihayuni0@gmail.com, [2]berli_amikom@yahoo.co.id
Correspondence Author Email: ratihayuni0@gmail.com
Submitted: **09/07/2024**; Accepted: **08/09/2024**; Published**: 09/09/2024**

Abstract−This research aimed to provide detailed insights into the sales performance of MSME snacks, identifying which products are popular among buyers and which are not. The study also examined whether these snacks should continue to be sold in their current form or if new innovations are required to attract more customers. To achieve this, a k-means algorithm was employed for clustering the data. The analysis was based on data collected from January to March 2024, including the name of the snack, the number of stocks, and the number of sales. Interim results revealed two distinct clusters. Cluster one, representing best-selling products, included: Cimol with 500 units sold from a stock of 550 units, Noodle Skewers with 450 units sold from a stock of 500 units, Dimsum with 400 units sold from a stock of 450 units, Egg Rolls with 380 units sold from a stock of 400 units, and White Bread with 360 units sold from a stock of 370 units. Cluster two, representing under-selling products, included Seblak Snacks with 150 units sold from a stock of 300 units. These findings provide a preliminary understanding of the current market performance of various snacks. The results indicate that Seblak Snacks, in particular, require innovation to improve their market performance and compete with other popular products. This research successfully categorized the products and provided a clear numerical overview of their sales performance. Further research with an expanded dataset is recommended to validate and enhance these insights.

Keywords: Algorithm K-Means; MSMEs; Snack; Clustering; Purwokerto

1. INTRODUCTION

MSMEs (Micro, Small, and Medium Enterprises) play a very important role in the Indonesian economy, including in Purwokerto. They contribute significantly to job creation, local product development, and improving people's welfare. One of the MSME sectors that has great potential to develop is the snack industry. However, MSME players in this sector face various challenges, such as fierce competition and rapid and dynamic changes in consumer preferences [1]. To survive and compete in the market, MSMEs need to continue to innovate and have a deep understanding of market trends and consumer preferences. In an effort to understand consumer preferences and market trends, data analysis is critical. The K-Means Clustering algorithm is one of the data analysis methods that can be used to identify consumption patterns and product sales trends. This algorithm works by grouping data into multiple clusters based on similarity in characteristics. The implementation of the K-Means Clustering algorithm in the context of snack MSMEs in Purwokerto can help identify which snacks are the most popular and have the highest sales [2].

The main purpose of this study is to provide more in-depth insights for MSME actors in Purwokerto regarding the snack products that are most in demand by consumers. With this information, MSMEs can develop more targeted marketing strategies, improve production efficiency, and optimize sales. In addition, this research also aims to help MSMEs in making better business decisions based on existing data, so that they can be more responsive to changing trends and market demands. This research aims to achieve several main things. First, this study wants to identify the best-selling snack products that are most in demand by consumers in Purwokerto. By knowing the best-selling snacks, MSMEs can focus on these products to increase production and marketing. Second, this study aims to group snack products into several clusters based on the similarity of sales characteristics and consumer preferences. Thus, MSMEs can develop different strategies for each cluster, making them more efficient and effective in meeting market needs [3]. Third, this study also aims to provide recommendations for marketing strategies that are more targeted based on the results of data analysis. This is expected to help MSMEs in increasing their sales and competitiveness in the market. Fourth, this study wants to provide insight to MSMEs about the importance of using data analysis technology in business decision-making, so that they can be more adaptive and responsive to changes in market trends [4].

Assessing employee performance quality is crucial for companies, but PT Clariant Adsorbents Indonesia currently lacks such a system. The first previous research conducted by Regina aims to evaluate employee productivity and performance effectiveness using the K-means clustering method, which is popular for its optimal data partitioning. Employee performance is categorized into highly productive, moderately productive, and less productive clusters. Variables considered in this study include employee names, work quality scores, responsibility values, cooperation values, attendance values, and discipline values. Using Rapidminer Version 7.6.0.0.1, the study identified 0.42% of employees (16) as highly productive, 0.47% (18) as moderately productive, and 0.11% (4) as less productive. The application of the K-means clustering method provided a clear and structured way to assess and categorize employee performance, offering valuable insights into workforce productivity. These findings can help PT Clariant Adsorbents Indonesia implement targeted strategies to enhance employee performance, allocate resources more efficiently, and foster a more productive work environment [5].

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 712−720
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5524

Second, reisandi Employee performance assessment is crucial for companies as it impacts their goals. This research aims to identify employees with high competence and loyalty by categorizing their performance into best, very good, and average. The study employs the K-Means Clustering method, known for its fast and efficient computing and effective data partitioning. Variables include knowledge and skill value, quantity value, and quality value. The study's outcome categorizes employee performance into three groups: best, very good, and average. [6]

The three goddesses Lecturer performance appraisal evaluates their performance and work output. This research aims to classify lecturer performance using data mining techniques to provide information and evaluation for decision-making. The study follows the Knowledge Discovery in Database (KDD) method, including Data Selection, Preprocessing/Cleaning, Data Transformation, Data Mining, and Interpretation/Evaluation stages. The K-Means Clustering algorithm was used, starting with random centroid values. The process stops when centroids no longer change between iterations. Testing with RapidMiner Studio 9.10 and the Davies-Bouldin Index (DBI) on 983 data points revealed the following lecturer performance based on student satisfaction: 31.74% (312 students) rated very good, 40.79% (401 students) rated good, 19.23% (189 students) rated quite good, and 8.24% (81 students) rated not good. The DBI result of 0.270 indicates good clustering accuracy, as it is close to zero [7].

Febrianto Library at Institut Teknologi Nasional Malang provides a range of academic information services, including access to research papers, articles, theses, and dissertations. Despite this comprehensive collection, transactional data has not been effectively utilized to evaluate the library's holdings and usage patterns. To address this gap, researchers developed a web-based application using the K-Means clustering method. This application categorizes visitors and borrowers based on their usage patterns, enabling library staff to better understand and assess data related to library visits and book borrowing. The developed application has undergone rigorous functional testing and has received positive feedback from users. According to user ratings, 50% of respondents rated the application as Excellent, 48% as Good, and 2% as Fair. This feedback underscores its effectiveness and user satisfaction. Moreover, the application is designed to be compatible with Mozilla Firefox 83.0 and Google Chrome 87.0.4280.88, ensuring usability across popular web browsers. This compatibility enhances its accessibility and usability for library staff and users alike. Ultimately, the application not only aids in categorizing visitors and borrowers but also supports computerized recommendations for book acquisitions based on analyzed usage data. This functionality contributes to improving the overall efficiency and effectiveness of library operations and services at Institut Teknologi Nasional Malang [8].

The study conducted by The Five Syahputras aimed to identify severe malnutrition among toddlers across Indonesian provinces, utilizing data sourced from the Ministry of Health of the Republic of Indonesia. Employing the K-means algorithm for data mining, the research successfully pinpointed 15 provinces characterized by high malnutrition clusters, including Aceh, North Sumatra, and West Nusa Tenggara, among others. Conversely, the study identified 19 provinces, such as West Sumatra, Riau, and Jakarta, that exhibited low malnutrition clusters. The primary objective of these findings is to provide actionable insights to the central government to support efforts aimed at enhancing infant nutrition, thereby fostering optimal growth and development among toddlers throughout Indonesia. Such initiatives are crucial for addressing and mitigating the impacts of severe malnutrition on the nation's youngest population. By identifying specific provinces with high and low malnutrition clusters, the study enables targeted interventions, promotes the sharing of best practices, and contributes to the overall health and well-being of the country's future generations [9].

Therefore, this research is very necessary considering the competition in the snack industry is getting tighter and consumer preferences are changing so quickly. Without a proper understanding of market trends and consumer behavior, MSMEs risk losing market share and experiencing a decline in sales. By conducting this research, MSMEs can take advantage of data analysis technology to obtain accurate and relevant information, so that they can make better and strategic decisions. In addition, this research can also be an example for other MSMEs in different sectors to apply a data-based approach in developing their businesses. The implementation of the K-Means Clustering algorithm to determine the best-selling snacks for MSMEs in Purwokerto is a strategic step that can provide various competitive advantages. By leveraging data analytics technology, MSMEs can make better decisions, improve operational efficiency, and ultimately, improve their competitiveness in an increasingly dynamic and competitive market.

The K-Means algorithm is a clustering method that divides data into a number of clusters based on similarity in characteristics. First, determine the number of clusters and randomly select the points as the initial centroid. Each data point is then calculated its distance to each centroid and assigned to the cluster with the nearest centroid. Once all data points are assigned, the centroid position is updated by taking the average of all the data points in the cluster. The cluster assignment and centroid update steps are repeated until the centroid position no longer changes significantly or the maximum number of iterations is reached. The end result is that each data point is clustered into one of the clusters [10].

The K-Means Clustering algorithm is very effective for clustering because it has several advantages. First, the algorithm is simple and easy to implement, making it a popular choice for many applications across various domains. Its straightforward approach allows for quick understanding and deployment, even for those with limited experience in data science. Second, K-Means is very fast and efficient when it comes to computing, especially when working with large datasets, due to its linear time complexity. This efficiency is crucial in handling extensive data, enabling quick iterations and timely results without demanding excessive computational resources.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 712−720
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5524

Third, the algorithm tends to give good results when the data has clear clusters with spherical shapes and similar sizes. This characteristic makes it particularly suitable for datasets where natural groupings are apparent, ensuring accurate and meaningful clustering outcomes. Fourth, K-Means is flexible and can be used in a variety of fields, such as customer segmentation, image grouping, and pattern analysis. Its versatility extends its applicability to numerous industries, including marketing, healthcare, finance, and more. Finally, the results of K-Means clustering are easy to interpret and visualize, aiding in further decision-making and analysis. The clear visual representation of clusters helps stakeholders understand complex data patterns and derive actionable insights, making K-Means a valuable tool in data-driven decision-making processes [11].

2. RESEARCH METHODOLOGY

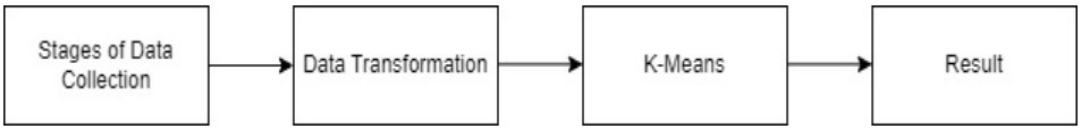

Figure 1. Research Flow

Data collection is a critical stage in the research process that requires careful planning and implementation. This stage begins by designing a data collection method that is in accordance with the research objectives. Researchers should consider methods such as interviews, questionnaires, observations, experiments, or the use of secondary data sources such as documents and reports, depending on the type of information needed and the availability of resources. Furthermore, the selection of tools or instruments to collect data becomes important; This can include effective questionnaire designs, accurate physical measurement instruments, or structured observation procedures. After planning, the implementation stage involves the execution of the data collection plan [12]. It involves a systematic process to ensure that data is collected with high consistency and accuracy according to established methodologies. During the data collection process, researchers also need to consider factors such as respondent compliance, reliability of data collection instruments, and sustainability of the data collection process within a set period of time.

The success of the data collection stage is pivotal as it significantly impacts the quality, reliability, and scientific validity of any study. A robust data collection process ensures that the data gathered is accurate, comprehensive, and representative of the phenomenon being studied. This, in turn, lays a solid foundation for conducting thorough analysis and drawing meaningful conclusions. Effective data collection involves careful planning, employing appropriate methods such as direct observation, interviews, surveys, or data records, depending on the research objectives and constraints. Rigorous adherence to data collection protocols helps minimize biases and errors, thereby enhancing the credibility of the findings. Ultimately, the quality of data collected influences not only the accuracy of analysis but also the reliability of interpretations and conclusions derived from the study. Researchers rely on well-collected data to support their hypotheses, make informed decisions, and contribute valuable insights to their respective fields of study.

2.1 Data Mining

Data mining is a process used to dig, analyze, and find hidden patterns, relationships, and information in large and complex datasets. The main goal is to transform raw data into meaningful and useful information that can support decision-making [13]. The data mining process involves several main stages, starting with the extraction of data from various sources. This extracted data is then stored in a format that is ready for further processing. The next step in data mining is data cleaning and pre-processing. At this stage, the data that has been collected is cleaned of errors, shortcomings, and inconsistencies. This involves filling in missing values, removing inaccurate data, and normalizing data formats [14]. This process ensures that the data used in the analysis is of good quality and reliable. Once the data is ready, various data mining techniques are applied to identify patterns and trends in the data. These techniques include clustering, classification, association, and regression. Clustering is a technique used to group data based on certain similarities, such as the K-Means algorithm. Classification involves creating predictive models that can categorize data into specific classes or categories based on specific attributes. Examples of algorithms used in the classification include decision trees, naïve Bayes, and support vector machines. Associations look for relationships and correlations between variables in data, as a priori algorithms do in finding association rules in purchase transactions. Regression is used to find functional relationships between independent and dependent variables in data, helping to predict the value of dependent variables based on independent variables [15].

Data mining has a wide range of applications in various industries such as business, healthcare, finance, marketing, and science. In business, for example, data mining can help identify customer buying patterns and design more effective marketing strategies [16]. The data mining process provides the ability to analyze data in depth and gain insights that were previously unseen, allowing organizations to make better and more informed decisions based on the information generated from deep data analysis. Figure 2 is a summary of the flow of the process in data mining.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 712−720
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5524

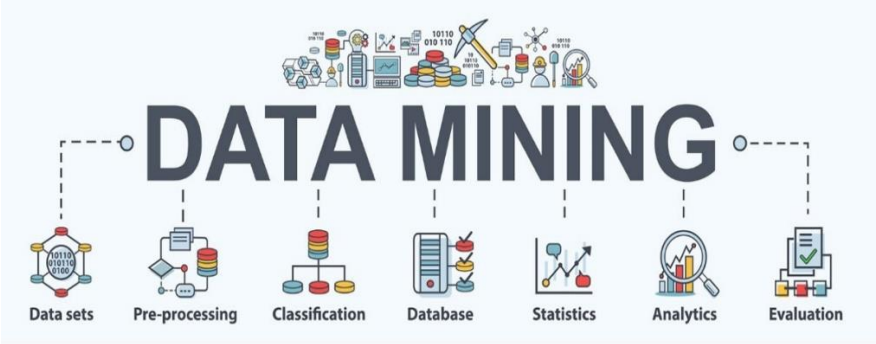

Figure 2. Data Mining Flow

2.2 Algoritma K-Means

The K-Means algorithm is one of the popular clustering methods in data mining and machine learning, which is used to divide a set of data into several groups or clusters based on certain similarities between the data. The main goal of the K-Means algorithm is to minimize variation within each cluster and maximize variation between different clusters [17]. The process begins by determining the desired number of clusters (K), which are usually predetermined based on domain knowledge or a specific method such as the Elbow Method. After that, the starting point K is selected as the initial centroid, which can be randomly selected or use other methods for a better initial distribution. Each data point in the dataset is then assigned to a cluster with the nearest centroid based on Euclidean distance or other distance metric. After all data points are assigned, the centroids of each cluster are recalculated as the average of all the data points in that cluster. The steps of data grouping and recalculating the centroid are repeated until there is no significant change in the position of the centroid or data grouping, which indicates that the algorithm has reached convergence [18]. The end result is a K cluster with each centroid representing the center of each cluster [19]. The K-Means algorithm is well-known for its speed and simplicity, but it also has weaknesses such as sensitivity to early selection of centroids and difficulty in determining the optimal number of clusters [20]. These algorithms are often used in a variety of applications, including market segmentation, pattern recognition, and image grouping. Figure 3 is an illustration of the application of the K-Means algorithm.

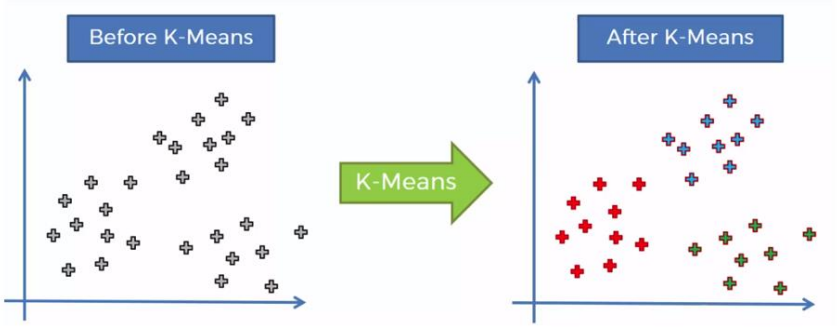

Figure 3. Illustration of K-Means algorithm

2.3 Clustering

Clustering in the realm of data mining and data analysis involves the process of grouping or classifying a set of objects or data points based on similarities in their characteristics. The primary objective of clustering is to form clusters or groups that are internally homogeneous, meaning objects within the same cluster are more similar to each other compared to those in different clusters [21]. This technique is particularly useful for uncovering hidden patterns and structures within data that may not be immediately apparent. By organizing data into clusters, it facilitates simplified data representation, enabling analysts to discern internal relationships and structures without relying on predefined class labels [22]. Various clustering methods are employed across different domains such as data mining, machine learning, statistical analysis, and image processing. Some common approaches include Hierarchical Clustering, which arranges objects in a tree-like hierarchy based on similarity; Partitioning Clustering, exemplified by algorithms like K-Means, which partition data into distinct clusters without forming a hierarchy; Density-Based Clustering, such as DBSCAN, which groups objects based on their density in attribute space; and Model-Based Clustering, like Expectation-Maximization (EM), which uses probability models to define clusters. The application of clustering is broad and diverse, encompassing tasks such as customer segmentation, social network analysis, identifying purchasing behaviors, and enhancing image processing techniques. Its effectiveness hinges on selecting appropriate methods that align with the specific characteristics of the dataset and the objectives of the analysis. As a result, clustering serves as

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 712−720
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5524

a powerful tool for revealing intricate structures and patterns within complex datasets, providing valuable insights that support decision-making across various disciplines and industries [23].

3. RESULT AND DISCUSSION

3.1 Data Collection

The first step that researchers take before starting the grouping process is to comprehensively collect the necessary data. This important data includes several key variables, including the name of each snack, the amount of stock available, and the number of sales recorded. Due to the short duration allocated for the study, the researchers focused on a concise time frame, collecting data from January to March 2024. To ensure the accuracy and reliability of the data, the researchers used several data collection methods. These methods include direct observation, which allows them to gather first-hand insights into snack stock and sales patterns. Additionally, the researchers conducted detailed interviews with sellers, providing an opportunity to gain in-depth information and context on sales trends and stock levels. Next, check the seller's records to corroborate the data obtained from observations and interviews. By utilizing this combination of data collection techniques, the researchers aim to compile a robust dataset that will form the basis for subsequent clustering analysis. This comprehensive approach ensures that the data is not only accurate but also reflects real-world sales and snack stock dynamics over a specified period.

3.2 Process K-Means Algorithm

In the second step, it is to carry out the calculation process with the K-Means algorithm. The calculation is carried out in stages. For details of the K-Means algorithm process, please see the following.

3.2.1 Selection of Number of Clusters (K)

Determining the desired number of clusters is the first step in this study. At this stage, the researcher will establish a cluster center, which in this case is represented by data 1 and 5. The data in question refers to product sales information and transactions that occurred from January to March 2024. Snack sales data are presented in Table 1, then only a few snacks are used due to time constraints in this study, so the selection of the number of clusters is as follows.

Table 1. Data Beginning

Transaction Month	Snack Name	Available quantity	Snack Sold
Jan 2024	Dimsum	100	75
Jan 2024	Mie Lidi	100	94
Jan 2024	Seblak	100	67
Jan 2024	Roti Tawar	100	88
Jan 2024	Telur Gulung	100	85
Jan 2024	Cimol	100	94
Feb 2024	Dimsum	100	94
Feb 2024	Mie Lidi	100	78
Feb 2024	Seblak	100	77
Feb 2024	Roti Tawar	100	70
Feb 2024	Telur Gulung	100	90
Feb 2024	Cimol	100	85
March 2024	Dimsum	100	89
March 2024	Mie Lidi	100	96
March 2024	Seblak	100	75
March 2024	Roti Tawar	100	84
March 2024	Telur Gulung	100	90
March 2024	Cimol	100	85

In Table 1, the initial data that will be used in this study is presented. The data obtained is classified into five groups of snacks, with each group representing a distinct category based on certain characteristics. This classification helps in organizing the dataset for further analysis. Additionally, the table includes information about the stock levels of each snack, with a standard value of 100 units per snack type. This stock level provides a baseline for comparing the availability and consumption rates of different snacks. Furthermore, the table also records the number of snacks sold out, which is crucial for understanding the demand and popularity of each snack category. By tracking the sales data, the study aims to identify trends and patterns in consumer preferences. To enhance data readability and user comprehension, Table 2 presents a streamlined version of the data grouping. This table condenses the information into a more concise format, making it easier for users to quickly grasp the key findings of the research. By simplifying the data presentation, Table 2 ensures that users can efficiently interpret the results and draw meaningful conclusions from the study.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 712−720
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5524

Table 2. Summary of Data Collection

Snack Name	Jan		Feb		March	
	Available quantity	Snack Sold	Available quantity	Snack Sold	Available quantity	Snack Sold
Dimsum	100	75	100	94	100	89
Mie Lidi	100	94	100	78	100	96
Seblak	100	67	100	77	100	75
Roti Tawar	100	88	100	70	100	84
Telur Gulung	100	85	100	90	100	90
Cimol	100	94	100	85	100	85

The information in tables 1 and 2 for jan is January, for feb is an abbreviation for February and for march is the month of March. In tables 1 and 2, the data components used are the same, only the data presentation is different.

3.2.2 Centroid Inisialitation

For the initial steps in centroid initialization, the data is randomly selected by selecting data point K as the initial centroid. In the context of this study, from Table 2, data 1 and data 3 were selected to form a cluster. After that, the centroid calculation is done using the formula shown in Table 3. This table shows the K1 and K2 data points that have been identified as the initial centroid. K1 specifically refers to products with the highest sales rate, while K2 reflects products that are less in demand. This process provides a foundation to analyze sales patterns and understand consumer preferences for products in Purwokerto MSMEs more deeply. Details of the data used are provided in Table 3.

Table 3. Inisilitation Centroid

Iterasi 1		Number		
No 1	C1	75	94	89
No 2	C2	67	77	75

After randomly determining K1 and K2, the next step is to perform calculations to get the values of C1 and C2. In this calculation, the Euclidean distance formula is used, which can be found in more detail in Table 4. This process is a critical stage in the analysis to determine the proximity between the data points and the specified centroid, helping in a deeper understanding of patterns in product sales data in Purwokerto MSMEs. of C1 and C2. In this calculation used Euclidean distance formula. For more details can be found in Table 4.

Table 4. Initial Results

Snack Name	Jan		Feb		March			
	Available quantity	Snack Sold	Available quantity	Snack Sold	Available quantity	Snack Sold	C1	C2
Dimsum	100	75	100	94	100	89	0	23.4
Mie Lidi	100	94	100	78	100	96	25.8	34.2
Seblak	100	67	100	77	100	75	23.4	0
Roti Tawar	100	88	100	70	100	84	27.7	23.9
Telur Gulung	100	85	100	90	100	90	10.8	26.8
Cimol	100	94	100	85	100	85	21.4	29.9

In table 1 is the initial calculation carried out in this study, this calculation obtained the results of cluster 1 and cluster 2.

3.2.3 New Centroid Calculation

In the K-Means algorithm, recalculations are carried out on the existing results to validate the data. In this study, a recalculation of centroids was carried out based on existing data points, as seen in Table 5. The results of the centroid update and conclusions can be seen in Table 5. This process is an important step in the analysis to ensure the compatibility of the cluster with the patterns in product sales data in Purwokerto MSMEs. This calculation is also one of the steps in calculating the K-Means Clustering Algorithm.

Table 5. New Cluster

Iterasi 1		Number		
No 1	C1	75	94	89
No 2	C2	67	77	75

In this study, the function of the new cluster is to reorganize data iteratively through reflection, ensuring convergence with each cluster's configuration from the previous iteration, particularly as observed in Iteration 1. The tabular representation in Table 6 visually depicts the updated data grouping process. This iterative approach helps refine the data clustering, leading to more accurate and meaningful groupings over time. The reflection mechanism

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 712−720
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5524

ensures that each iteration builds upon the previous one, aligning closer to the desired cluster configuration and improving overall data organization.

Table 6. Results After Seeding a new cluster

Snack Name	Jan		Feb		March					
	Available quantity	Snack Sold	Available quantity	Snack Sold	Available quantity	Snack Sold	C1	C2	Closest Distance	Item Group
Dimsum	100	75	100	94	100	89	16.2	23.4	16.2	Cls 1
Mie Lidi	100	94	100	78	100	96	11.3	34.2	11.3	Cls 1
Seblak	100	67	100	77	100	75	25.3	0	0	Cls 2
Roti Tawar	100	88	100	70	100	84	14.3	23.9	14.3	Cls 1
Telur Gulung	100	85	100	90	100	90	7.1	26.8	7.1	Cls 1
Cimol	100	94	100	85	100	85	8.0	29.9	8.0	Cls 1

The results that have been described in detail in table 6 are the results of a recalculation that has been carried out using a new cluster. Then because this is the second calculation, the results of the calculation are immediately grouped according to the class.

3.2.3 Iteration Process

In this stage, researchers repeat the steps of updating the centroids until there is no significant change in the centroids or the predetermined iteration limit is reached. The process involves recalculating the centroids based on the current cluster memberships and then reassigning data points to the nearest centroids. This iterative method ensures that the clustering converges to a stable solution where data points no longer switch between clusters. From the data grouping process in Iterations satu and dua, it was concluded that the resulting clusters remained the same. In Iteration satu, Cluster satu consists of data points satu, dua, tiga, empat, lima dan enam, while Cluster dua contains data point tiga. The same grouping occurred in Iteration dua, indicating that the clusters had stabilized and further iterations would not change the clustering outcome. This consistency across iterations signifies that the algorithm has successfully converged. The iterative process is crucial as it allows for refinement of the clusters, ensuring that the final groupings are accurate and meaningful. By stopping the process when a consistent clustering pattern is observed, researchers can be confident that the data has been effectively grouped according to the K-Means algorithm's criteria.

3.2.4 Final Results

In this step there is an overall result that is final, there is no calculation anymore, written in detail in table 7. This table presents the stable clusters obtained after the iterations, providing a clear and concise summary of the data groupings. The final centroids and the composition of each cluster are detailed, reflecting the outcome of the convergence process. The final results highlight the effectiveness of the iterative approach in achieving reliable and consistent clustering, demonstrating the robustness of the K-Means algorithm in organizing data into meaningful clusters. These results can now be used for further analysis, interpretation, and application in the relevant domain, ensuring that the data is accurately represented and grouped.

Table 7. Final Result of Calculation

Snack Name	Snack Sold 2024			Iterasi 1				Iterasi 2			
	Jan	Feb	March	C1	C2	Closest Distance	Item Group	C1	C2	Closest Distance	Item Group
Dimsum	75	94	89	0	23.4	0	Cls 1	16.2	23.4	16.2	Cls 1
Mie Lidi	94	78	96	25.8	34.2	25.8	Cls 1	11.3	34.2	11.3	Cls 1
Seblak	67	77	75	23.4	0	0	Cls 2	25.3	0	0	Cls 2
Roti Tawar	88	70	84	27.7	23.9	27.7	Cls 1	14.3	23.9	14.3	Cls 1
Telur Gulung	85	90	90	10.8	26.8	10.8	Cls 1	7.1	26.8	7.1	Cls 1
Cimol	94	85		21.4	29.9	21.4	Cls 1	8.0	29.9	8.0	Cls 1

Based on the steps of the K-means algorithm used to analyze sales data of various MSME snack products, the results of Iterasi I and Iterasi II show consistent results. In Iteration I, Cluster 1 consists of data 1, 2, 4, 5, and 6, which depicts products with high demand from consumers and strong sales performance. Cluster 2 only consists of data 3, namely Seblak products, which experienced a significant decrease in sales. In Iteration II, the grouping remained the same, with Cluster 1 continuing to show products that were in high demand and Cluster 2 confirming low demand for Seblak products. Based on this analysis, it is recommended not to increase the stock of Seblak products for the

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 712−720
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5524

following months, while products in Cluster 1 should be continuously monitored and may need to be replenished to meet high consumer demand.

4 CONCLUSION

After various stages of research ranging from the initial data collection process to the calculation using the K-Means algorithm, it can be concluded that the data taken in this study comes from MSMEs. The data range used starts from January to February in 2024, then for snacks taken for data such as cimol, white bread, egg rolls, seblak noodle skewers and also dimsum. This data was randomly selected due to the brevity of the research so that the researcher only used some snack data. Then after the data was collected, calculations were carried out with the k-means algorithm and the result was obtained that there was one snack whose sales were not in demand, this snack was seblak. So that seblak sellers are expected to make new innovations so that their sales are crowded. For other snacks, sales results are quite good and stable. The calculation of the K-Means algorithm is carried out twice until it produces the same result as the unchanged cluster. In this study, there are cluster one and cluster two, for cluster one product whose sales are stable while cluster 2 products are not selling. The benefit of this research is to provide a deeper insight into consumer purchasing patterns in Purwokerto MSMEs, which can help managers in making more informed decisions regarding product procurement and sales. With more accurate information about the products that are most in demand, MSMEs can optimize inventory, reduce the risk of shortage or excess stock, and increase customer satisfaction. The contribution of this research is to provide a data-driven analytical approach for inventory management and marketing strategies in MSMEs. By applying the K-Means algorithm, this study shows how data analysis techniques can be applied to small and medium-sized businesses to improve operational efficiency and competitiveness in the market. In addition, the results of this study can be used as a reference for other MSMEs who want to apply a similar approach to optimize their business performance. For further research, the use of more data can be done to obtain more comprehensive and accurate results. For further research, it is recommended to expand the scope of sales data used, for example by including data for a full year or more to get a more complete picture of sales trends and seasonal patterns that may exist. Additionally, considering the addition of variables such as product price, type of promotion used, sales location, and customer feedback can provide a deeper understanding of the factors that affect consumer preferences and sales success. In addition, adopting and comparing the results of the K-Means Algorithm with other clustering approaches such as Hierarchical Clustering or DBSCAN can provide a broader perspective on the suitability and advantages of each method in the context of Purwokerto MSMEs.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to all parties who have supported the implementation of this research. We would like to express our deepest gratitude to Amikom University Purwokerto for the support of facilities and an academic environment that is conducive to this research. Thank you to the supervisors for their invaluable guidance, advice, and support during the research process. We also appreciate the support from our peers who have provided morale encouragement and technical assistance. Finally, thank you to family and friends who have provided emotional support and understanding during difficult times. This research would not have been possible without the help of all of you.

REFERENCES

[1] D. Aulia *Et Al.*, “Nomor 1 Februari, Pp 47-60 Penerapan Algoritma K-Means Dalam Proses Clustering Penilaian Kinerja Asn,” Jurasik, Vol. 6, 2023

[2] D. Praseptian M, A. Fadlil, And H. Herman, “Penerapan Clustering K-Means Untuk Pengelompokan Tingkat Kepuasan Pengguna Lulusan Perguruan Tinggi,” *Jurnal Media Informatika Budidarma*, Vol. 6, No. 3, P. 1693, Jul. 2022, Doi: 10.30865/Mib.V6i3.4191.

[3] A. Septiarini, I. A. Thaher, And N. Puspitasari, “Pengelompokan Kualitas Kinerja Pegawai Menggunakan Metode K-Means Clustering,” *Komputika : Jurnal Sistem Komputer*, Vol. 11, No. 2, Pp. 131–141, Jul. 2022, Doi: 10.34010/Komputika.V11i2.5518.

[4] J. Penerapan, T. Informasi, D. Komunikasi, G. B. Kaligis, And S. Yulianto, “It-Explore Analisa Perbandingan Algoritma K-Means, K-Medoids, Dan X-Means Untuk Pengelompokkan Kinerja Pegawai (Studi Kasus: Sekretariat Dprd Provinsi Sulawesi Utara),” 2022.

[5] S. Regina, E. Sutinah, And N. Agustina, “Clustering Kualitas Kinerja Karyawan Pada Perusahaan Bahan Kimia Menggunakan Algoritma K-Means,” *Jurnal Media Informatika Budidarma*, Vol. 5, No. 2, P. 573, Apr. 2021, Doi: 10.30865/Mib.V5i2.2909.

[6] K. P. Karyawan Xyz, I. Reisandi, F. Sri Mulyati, And Dan Muchammad Fauzi, “Implementasi Clustering K-Means Terhadap Penilaian”, 2021, [Online]. Available: Http://Sostech.Greenvest.Co.Id.2021.

[7] N. Luh, P. P. Dewi, I. Nyoman Purnama, And N. W. Utami, “Penerapan Data Mining Untuk Clustering Penilaian Kinerja Dosen Menggunakan Algoritma K-Means (Studi Kasus: Stmik Primakara),” *Jurnal Ilmiah Teknologi Informasi Asia*, Vol. 16, No. 2, 2022.

[8] A. Febrianto, S. Achmadi, And A. P. Sasmito, “Penerapan Metode K-Means Untuk Clustering Pengunjung Perpustakaan Itn Malang,” 2021.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 712−720
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5524

[9] H. Syahputra, "Clustering Tingkat Penjualan Menu (Food And Beverage) Menggunakan Algoritma K-Means," *Jurnal Komtekinfo*, Pp. 29–33, Mar. 2022, Doi: 10.35134/Komtekinfo.V9i1.274.
[10] A. Lili, H. Cipta, And S. Widodo, "Pengelompokan Hasil Panen Kelapa Sawit Dalam Produksi Per Blok Menggunakan Algoritma K-Means," 2022.
[11] T. M. M. Tyas And A. I. Purnamasari, "Penerapan Algoritma K-Means Dalam Mengelompokkan Demam Berdarah Dengue Berdasarkan Kabupaten," *Blend Sains Jurnal Teknik*, Vol. 1, No. 4, Pp. 277–283, Mar. 2023, Doi: 10.56211/Blendsains.V1i4.231.
[12] M. Harahap, Y. Lubis, And Z. Situmorang, "Analisis Pemasaran Bisnis Dengan Data Science : Segmentasi Kepribadian Pelanggan Berdasarkan Algoritma K-Means Clustering," *Data Sciences Indonesia (Dsi)*, Vol. 1, No. 2, Pp. 76–88, Jan. 2022, Doi: 10.47709/Dsi.V1i2.1348.
[13] A. E. Pramitasari And Y. Nataliani, "Perbandingan Clustering Karyawan Berdasarkan Nilai Kinerja Dengan Algoritma K-Means Dan Fuzzy C-Means," *Jurnal Teknik Informatika Dan Sistem Informasi*, Vol. 8, No. 3, 2021, [Online]. Available: Http://Jurnal.Mdp.Ac.Id
[14] R. Riadi And Mesran, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Analisa Penjualan Parfume," *Journal Of Informatics, Electrical And Electronics Engineering*, Vol. 2, No. 4, Pp. 138–145, Jun. 2023, Doi: 10.47065/Jieee.V2i4.1181.
[15] M. D. Chandra, E. Irawan, I. S. Saragih, A. P. Windarto, And D. Suhendro, "Penerapan Algoritma K-Means Dalam Mengelompokkan Balita Yang Mengalami Gizi Buruk Menurut Provinsi," *Bios : Jurnal Teknologi Informasi Dan Rekayasa Komputer*, Vol. 2, No. 1, Pp. 30–38, Mar. 2021, Doi: 10.37148/Bios.V2i1.19.
[16] W. Tri Wulandari, T. Handayani, A. Sellyana, "Penerapan Metode K-Means Clustering Dan Perbandingan Eksponensial Dalam Pemilihan Tenaga Kerja Untuk Pengerjaan Proyek Pada Pt. Teluk Makmur Sakti Kota Dumai," 2023.
[17] D. Ariyanto, "Data Mining Menggunakan Algoritma K-Means Untuk Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut," *Jurnal Sistim Informasi Dan Teknologi*, Pp. 13–18, Feb. 2022, Doi: 10.37034/Jsisfotek.V4i1.117.
[18] T. Asy Aria, M. Julkarnain, And F. Hamdani, "Klik: Kajian Ilmiah Informatika Dan Komputer Penerapan Algoritma K-Means Clustering Untuk Data Obat," *Media Online*, Vol. 4, No. 1, Pp. 649–657, 2023, Doi: 10.30865/Klik.V4i1.1117.
[19] R. Taufiq, A. Hambali, And A. Saifudin, "Rancang Bangun Sistem Pendukung Keputusan Penilaian Karyawan Terbaik Pada Pt. Surya Toto Indonesia, Tbk Menggunakan Metode Simple Additive Weighting (Saw) Berbasis Web," *Jurnal Informatika Universitas Pamulang*, Vol. 6, No. 1, P. 188, Mar. 2021, Doi: 10.32493/Informatika.V6i1.9836.
[20] R. Fauziah And A. I. Purnamasari, "Implementasi Algoritma K-Means Pada Kasus Kekerasan Anak Dan Perempuan Berdasarkan Usia," *Hello World Jurnal Ilmu Komputer*, Vol. 2, No. 1, Pp. 34–41, Mar. 2023, Doi: 10.56211/Helloworld.V2i1.232.
[21] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *Jurnal ilmiah informatika dan ilmu computer*, 2024, Doi: 10.58602/Jima-Ilkom.V3i1.26.
[22] P. Apriyani, A. R. Dikananda, And I. Ali, "Penerapan Algoritma K-Means Dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi," *Hello World Jurnal Ilmu Komputer*, Vol. 2, No. 1, Pp. 20–33, Mar. 2023, Doi: 10.56211/Helloworld.V2i1.230.
[23] Preddy, P. Marpaung, I. Pebrian, And W. Putri, "Penerapan Data Mining Untuk Pengelompokan Kepadatan Penduduk Kabupaten Deli Serdang Menggunakan Algoritma K-Means," *Jurnal Ilmu Komputer Dan Sistem Informasi (Jikomsi)*, Vol. 6, No. 2, Pp. 64–70, 2023.