Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2025. Vol. 4. No. 2; e-ISSN: 2808-4519

# Application of K-Means for Product Grouping Best Sellers at Planet Tire Jatibarang Branch

**Risnawati[1]*, Rini Astuti[2], Willy Prihartono[3]**

*[1] Program Studi Teknik Informatika, STMIK IKMI Cirebon*
*J.l Perjuangan No. 10 B Majasem Kesambi Kota Cirebon*
*[2,3] Program studi Teknik Informatika, STMIK LIKMI Bandung*
*Jl. Ir H Juanda No 96, Lebahgede, Coblong, Bandung*
*jujurisnawati04@gmail.com[1], riniastuti@likmi.ac.id [2], willy@likmi.ac.id [3]*

**Abstract**

This research aims to identify the best-selling products at Planet Tire Workshop Jatibarang Branch using the K-Means Clustering method. Understanding product sales patterns is important in designing effective marketing strategies and managing stock efficiently. This research uses sales transaction data for one year, including the number of sales, product types, and total transaction value. The analysis process includes data preprocessing, selection of relevant attributes, application of the K-Means algorithm, and validation of the optimal number of clusters with the Elbow method. As a result, products were grouped into three categories: high, medium, and low sales. The high sales cluster contributes significantly to revenue, while the medium sales cluster shows potential for improvement through promotion, and the low sales cluster requires further evaluation. This research helps management manage stock, prioritize promotions, and optimize resource allocation. However, the research has limitations as it has not considered external factors such as seasonal trends and promotions, and focuses on one branch. Development of the research in other branches can expand its benefits. The results of this study are expected to improve operational efficiency, support data-driven strategies, and enrich academic literature related to the application of K-Means in retail management and sales data analysis.

***Keywords***: *K-Means Clustering, Sales Pattern, Best Selling Products, Marketing Strategy, Data Mining*.

## 1. Introduction

Business development in the automotive industry, especially repair services and tyre sales, is becoming increasingly competitive as the need for private vehicles grows. In Indonesia, this sector continues to show significant growth, driven by the increasing number of motorised vehicles and public awareness of routine vehicle maintenance. To remain competitive, companies must have a deep understanding of customer preferences and product sales patterns. As one of the major players in the provision of tyres and related services in Indonesia, Planet Ban workshops strive to meet the needs of this market by offering different types of products that meet consumer demand and preferences [1].

In today's competitive environment, it is important for companies to identify which products are the best sellers or "bestsellers". This understanding not only helps to optimise inventory levels, but also to develop more focused and efficient marketing strategies. However, without the right way to categorise and analyse sales data, companies can struggle to understand consumer buying patterns [2].

Based on the problem statement, this research aims to help Planet Tire workshops understand product sales patterns at the Jatibarang store and identify the product groups that customers are most interested in.

## 2. Literature Review

### 2.1. Data Mining

Data mining is an integral part of the Knowledge Discovery in Database (KDD) process, which aims to transform raw data into meaningful information. The process involves a number of transformation stages, including data pre-processing, which focuses on cleaning up dirty or duplicate data and selecting relevant records and features for further analysis. Data pre-processing aims to transform raw data into a format suitable for the next stage of management and analysis [3].

### 2.2. Clustering

Clustering is a data analysis technique that aims to group objects or data that have similar characteristics or attributes. It is used to identify hidden structures or patterns in data without labels or prior information about the classes or groups available. As such, clustering enables exploratory data analysis to discover deep new insights [4].

### 2.3. K-Means

The K-means algorithm is a data clustering technique that divides data into k clusters. This approach works by grouping data based on the degree of similarity between data in a cluster and separating data that has differences from other clusters. Each cluster has a centre, called the centroid or centre of gravity, which is calculated as the average of the values of the cluster members [5].

The following are the stages in clustering or grouping with the K-Means algorithm:
1. Determine the number of clusters (k)
2. Specifies the centroid (the midpoint coordinates of each cluster), for the first iteration it is randomly retrieved
3. Calculate the distance of an object to the centroid by using the dEuclidean or Manhattan distance formula
4. Determining the distance of each object to the midpoint coordinates
5. Group these objects based on their closest distance

## 3. Research Method

### 3.1. Research Methods

Research methodology is a systematic approach used to investigate and explore a problem using scientific methods. This approach involves the careful and objective collection, processing and analysis of data with the aim of systematically drawing conclusions. This process is aimed at solving problems or testing hypotheses, while producing knowledge that is useful for human life. The stages of research methodology in completing this research are [6].

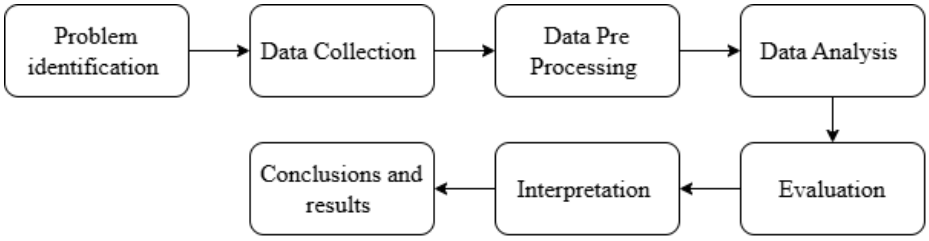


**Fig. 1:** Research Workflow

1. Problem identification
   Problem formulation, problem limitation, and research objectives are important first steps in the research process. Problem formulation aims to identify and formulate the main issues to be solved specifically and relevantly, while problem limitation is designed to narrow the scope of the research so that it remains focused and can be completed with limited time and resources.
2. Data collection
   Making observations and reviewing documents are strategic steps in collecting relevant data to support research. Observation is done by directly observing the phenomenon or object of research, so that researchers can obtain factual and objective empirical data. On the other hand, document review involves collecting and analyzing relevant documents, such as reports, archives, records, or other written data sources, to understand the research context in depth..
3. Data pre-processing
   The process of collecting and processing data in research involves several important steps, including downloading datasets, selecting relevant attributes, and cleansing, crawling, or scraping data. Dataset downloads are usually done from trusted sources, such as public data repositories or research platforms, to ensure data quality and accuracy.
4. Data analysis
   The application of Knowledge Discovery in Databases (KDD) methods and algorithms such as K-Means is an important step in the data analysis process to find meaningful patterns. KDD methods involve a series of systematic stages, from data selection, pre-processing, transformation, application of data mining algorithms, to interpretation of results.
5. Evaluation
   Analyzing the test results is an important step in ensuring that the model performs in line with expectations and research objectives. This analysis is done by evaluating the test results based on relevant performance metrics, such as accuracy, precision, recall, or silhouette value for clustering algorithms such as K-Means.
6. Interpretation
   Describing the parameters of the evaluation results is an important step to provide a clear insight into the performance of the model and its relevance to the research objectives. Evaluation parameters such as accuracy, precision, recall, F1-score, and silhouette value provide a quantitative overview of how well the model processes data and produces appropriate outputs.
7. Conclusions and findings
   Drawing conclusions and recommendations is an important final stage in the research to summarize the main findings and provide practical direction based on the results of data analysis. Conclusions are drawn by summarizing the evaluation results and significant

patterns found during the research, as well as answering the formulated problems. Conclusions should be concise, clear and data-driven to demonstrate the research's contribution to science or practical application.

# 4. Result and Discussion

## 4.1. Result

Penelitian ini memanfaatkan penerapan Algoritma K-Means pada bidang penjualan produk untuk membantu menentukan produk terlaris dan mengoptimalkan penjualan. Dalam hal ini, Algoritma K-Means dapat digunakan untuk melakukan analisis cluster pada data penjualan produk di Planet Ban.

This data collection, Data obtained by researchers directly from the planet tire workshop, jatibarang branch Data taken one day the amount of product expenditure.

**Table 1:** Sales dataset of planet tire jatibarang branch

| No. | Tanggal Transaksi | Nomer Transaksi | Qty | Nama Barang | Total Harga |
|---|---|---|---|---|---|
| 1 | 05-Agu-24 | 1020416013 | 2 | PS-588 TL 110/70-13 | 760.000.00 |
| 2 | 05-Agu-24 | 30901844001 | 3 | X-TGEAR OIL 120 ML | 60.000.00 |
| 3 | 05-Agu-24 | 309173040018 | 2 | X-T10W30 MTC 0.8L | 110.000.00 |
| ... | ... | ... | ... | ... | ... |
| 217 | 05-Okt-24 | 3091844001 | 3 | X-TGEAR OIL 120 ML | 60.000.00 |
| 218 | 05-Okt-24 | 2071495002 | 1 | GIPEN PNDK | 11.000.00 |
| 219 | 05-Okt-24 | 3236499001 | 1 | X-TTB1C 40 ML | 35.000.00 |

This data selection stage aims to select data that will be used in the Clustering process through the RapidMiner tool, the selection stage is carried out in Excel in the research process.

**Table 2:** attributes used for clustering

| No | Atribut |
|---|---|
| 1. | ID |
| 2. | Nomer Transaksi |
| 3. | QTY |
| 4. | Nama Barang |
| 5. | Total Harga |

Data preprocessing, the main attention is focused on identifying and correcting various aspects of the data that may contain incomplete, incorrect, irrelevant, inaccurate, or even missing information.

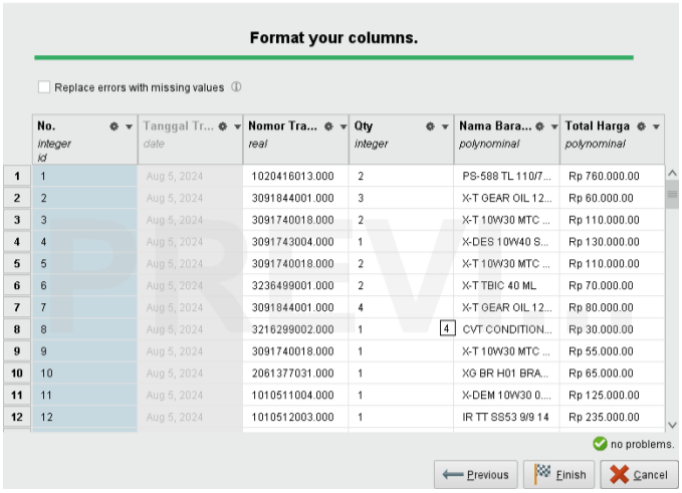


**Fig. 2:** Attributes used for clustering

This research uses the K-Means Clustering algorithm in RapidMiner to build a grouping of the best-selling products in the planet tire workshop jatibarang branch. Starting the Read Excel operator process which is the most basic used in the RapidMiner application functions as an Excel file reader.


**Fig. 3:** Operator Read Excel

The Nominal To Numeric operator is used to convert a non-numeric attribute type into a numeric type so that it can be processed further.

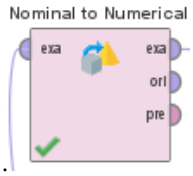


**Fig. 4:** Operator Nominal To Numerik

Next, what will be input into the rapid miner worksheet is the set role. In the Set Role operator there are parameters that need to be adjusted. The parameters used in this operator are Attribute Name and Target Role.


**Fig. 5:** Operator Set Role

After setting the set role or labeling the data that we want to process, the next operator that we will call is K-Means which will group or create clusters.


**Fig. 6:** Operator K-Means Clustering

The Cluster Distance Performance operator used to assess clustering performance is shown in Figure 4.10. This operator has two types of parameter criteria, namely Average Within Centroid Distance and Davies Bouldin Index (DBI). The Maximize parameter is applied to produce a non-negative DBI value (>= 0). Cluster evaluation is done to determine the number of clusters that produce the smallest or best DBI value.

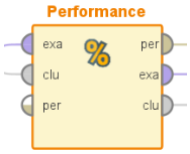


**Fig. 7:** Operator Performance

In the data mining stage, researchers will use the K-Means Clustering algorithm in RapidMiner to build a grouping of the best-selling products in the planet tire workshop jatibarang branch.

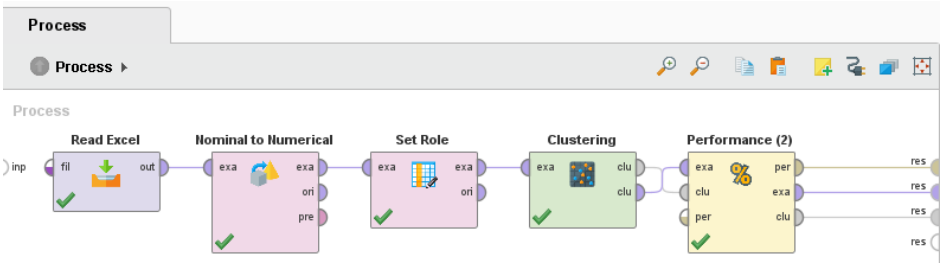


**Fig. 8:** Data Mining

After processing data using the K-Means Clustering algorithm in the Rapidminer application, the next step is to evaluate the results that have been made. The evaluation stage uses the Performance operator (Cluster Distance Performance) with the application of Davies Bouldin's main criteria on the parameters.

**PerformanceVector**

```
PerformanceVector:
Avg. within centroid distance: 383998085546.970
Avg. within centroid distance_cluster_0: 2515174798.393
Avg. within centroid distance_cluster_1: 17050163630.044
Avg. within centroid distance_cluster_2: 0.000
Avg. within centroid distance_cluster_3: 1.511
Avg. within centroid distance_cluster_4: 0.667
Avg. within centroid distance_cluster_5: 3473587359386.568
Avg. within centroid distance_cluster_6: 0.000
Avg. within centroid distance_cluster_7: 221779.243
Avg. within centroid distance_cluster_8: 0.000
Avg. within centroid distance_cluster_9: 1.312
Davies Bouldin: 0.028
```

**Fig. 9:** Performance Model

The application of the Performance operator (Cluster Distance Performance) for the evaluation stage in Rapidminer has one of the outputs in the form of the Davies Bouldin index. 10 trials were conducted to find the optimal DBI value, starting from the application into clusters of 2 to the application into clusters of 10.

**Table 3:** Cluster result and DBI value

| No | Jumlah Cluster | Nilai DBI |
|---|---|---|
| 1 | 2 | 0,204 |
| 2 | 3 | 0,067 |
| 3 | 4 | 0,039 |
| 4 | 5 | 0,039 |
| 5 | 6 | 0,147 |
| 6 | 7 | 0,133 |
| 7 | 8 | 0,117 |
| 8 | 9 | 0,032 |
| 9 | 10 | 0,028 |

Based on the evaluation results above, it is known that the most optimal DBI value is owned by an experiment with a total of 10 (ten) clusters with a DBI value of 0.028.

## 4.2. Discussion

Based on the evaluation results above, it is known that the most optimal DBI value is owned by an experiment with a total of 10 (ten) clusters with a DBI value of 0.028.

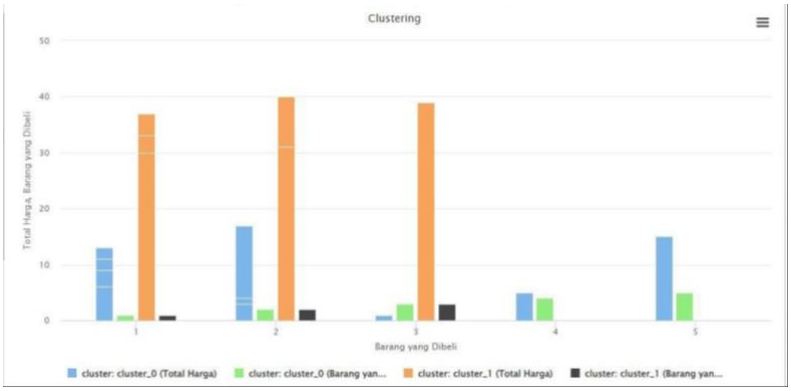


**Fig. 10:** Hasil Visualisasi K-Means Clustering

Cluster Quality Evaluation:

1. Best Selling Product Cluster
   This category includes products with very high sales rates, making a major contribution to workshop revenue. Examples are X-T GEAR OIL 120ML and X-T 10W30 MTC 0.8L, which show consistent demand from customers. These products are the main focus of stock management to ensure availability.
2. Medium Sales Product Cluster
   Products in this category have a stable sales rate, although not as high as the best-selling clusters. Examples are X-DEM 10W30 0.8L and X-DES 10W40 SPO 1LT. The potential for increased sales can be achieved through marketing strategies such as seasonal discounts or package promotions.
3. Low Sales Product Cluster
   This cluster includes products with low sales that require further evaluation. For example, CVT CONDITIONER-4 requires additional analysis to determine whether the product is still relevant in the market or needs to be promoted more.

# 5. Conclusion And Suggestions

## 5.1. Conclusion

This research successfully achieved the set objectives by applying the K-Means Clustering method through several stages, namely the collection of product sales data from Planet Ban Jatibarang branch in Excel format which includes attributes such as product ID, transaction number, number of items sold (QTY), item name, and total price, followed by data processing to overcome deficiencies or inconsistencies through normalization of values, filling in missing data, and conversion to an appropriate numerical format. The K-Means algorithm was applied using RapidMiner to cluster products based on numerical attributes such as sales amount and revenue, resulting in high, medium, and low sales performance clusters. Evaluation of cluster quality was conducted using the Davies-Bouldin Index (DBI), with an optimal value of 0.028 at a cluster count of 10, indicating good clustering results. The results of this analysis provide strategic insights, such as stock management for high-selling products, promotional focus on medium-selling products, evaluation for low-selling products, as well as development of new marketing strategies such as discounts or bundling offers. If this method is applied in other branches, the necessary steps include collecting local data with similar attributes, followed by preprocessing processes such as normalization, data conversion to numeric format, and missing data handling to ensure consistent results.

## 5.2. Suggestions

Based on the research results, there are several suggestions that can be applied for further development, namely adding attributes such as product category or sales time to increase the level of detail of clustering results, checking data regularly to maintain consistency and accuracy before the clustering process, and using additional methods such as Elbow or Silhouette Score to ensure the most appropriate number of clusters. In addition, it is recommended to apply the K-Means method to other branches by adjusting the local dataset to suit the characteristics of each branch.

# References

[1] D. Fitriyani, M. Jajuli, and G. Garno, "Implementasi Algoritma K-Means Untuk Klasterisasi Dalam Pengelolaan Persediaan Obat (Studi Kasus : Apotek Naza)," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 2841–2848, 2024, doi: 10.23960/jitet.v12i3.4921.

[2] M. Aji Dian Permana, M. Martanto, and U. Hayati, "Analisis Pengelompokan Penjualan Menggunakan Metode K-Means Pada Bisnis Center Smk Wahidin," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 925–936, 2024, doi: 10.36040/jati.v8i1.8695.

[3] Taufik Hidayat, Mohamad Jajuli, and Susilawati, "Clustering daerah rawan stunting di Jawa Barat menggunakan algoritma K-Means," *INFOTECH J. Inform. Teknol.*, vol. 4, no. 2, pp. 137–146, 2023, doi: 10.37373/infotech.v4i2.642.

[4] B. A. Pangestu, N. A. Kristiawan, and N. Sulistiyowati, "Clustering Obat Untuk Menentukan Pola Pemasaran Efektif di Apotek Amarta Sehat," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 16, pp. 115–126, 2022, [Online]. Available: https://doi.org/10.5281/zenodo.7058995

[5] S. Ega Yolanda, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Data Pasien Rehabilitasi Narkoba," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 182-`191, 2023, doi: 10.30865/klik.v4i1.1107.

[6] L. ' Izzah and A. Jananto, "Penerapan Algoritma K-Means Clustering Untuk Perencanaan Kebutuhan Obat Di Klinik Citra Medika," *Progresif J. Ilm. Komput.*, vol. 18, no. 1, p. 69, 2022, doi: 10.35889/progresif.v18i1.769.