Contents lists available at https://citedness.com/index.php/jdsi

Data Science Insights

Journal Page is available to https://citedness.com/index.php/jdsi

Research article

Cluster Analysis on Laptop Sales Data and Specifications using k-Means and k-Medoids Methods

Ibnu Fajar

Department of Information System, Pelita Indonesia Institute of business and Technology, Pekanbaru City, Indonesia
email: ibnu.fajar@student.pelitaindonesia.ac.id

ARTICLE INFO

Article history:
Received June 23, 2024
Revised July 03,2024
Accepted July 15, 2024
Available online August 01, 2024

Keywords:
Clustering
K-Means
K-Medoids
Laptop Sales
Laptop Specification

Please cite this article in IEEE style as:

I. Fajar, "Cluster Analysis on Laptop Sales Data and Specifications Using K-Means and K-Medoids Methods", Data Science Insights, vol. 2, no. 2, pp. 105–113, Aug. 2024.

ABSTRACT

This research aims to address the challenges in understanding the relationship between laptop specifications and sales prices and to enhance product segmentation based on cluster analysis. By using available laptop specifications and sales price data, this study aims to identify patterns in laptop specifications that influence sales prices using K-Means and K-Medoids cluster analysis. This research employs the K-Means and K-Medoids clustering methods to categorize laptops into several categories based on specifications such as screen size (inches), price, RAM capacity, and weight. The data transformation process, exploratory analysis, model building, and cluster performance evaluation were conducted using the RapidMiner analysis tool. The research results show that the K-Medoids algorithm provides more accurate clustering performance compared to K-Means, with a Davies-Bouldin Index value of -0.665 for K-Medoids and -0.487 for K-Means at configurations k=4 and k=5. A lower Davies-Bouldin Index value indicates that K-Medoids better represents the characteristics of the existing data. The clustering results identify laptop categories based on a combination of specifications and prices, which can be used by manufacturers and sellers to develop more targeted marketing strategies. This research is expected to provide useful insights for the laptop industry in understanding consumer preferences and needs, and to assist in making more informative decisions to improve sales and customer satisfaction.

Correspondence:
Ibnu Fajar
Department of Information System,
Pelita Indonesia Institute of business and Technology, Pekanbaru City, Indonesia

1. Introduction

The information technology industry, especially the laptop market, continues to grow rapidly as consumer demand for advanced and efficient devices increases. Laptops are not only used by professionals and students but also by the general public for various daily needs. With a variety of specifications and prices offered, a deep understanding of the factors influencing laptop sales prices is crucial for manufacturers and sellers in developing effective marketing strategies.

Market segmentation is one approach used to understand consumer preferences and market trends. By categorizing products based on specifications and prices, companies can identify different market segments and design appropriate products and marketing strategies. Clustering methods, such as K-Means and K-Medoids, are effective data analysis techniques for market segmentation because they can group data into homogeneous clusters based on certain characteristics. This research aims to identify patterns in laptop specifications and sales prices using K-Means and K-Medoids clustering methods. The data used includes screen size, price, RAM capacity, and laptop weight. Through this analysis, it is expected to gain better insights into laptop market segmentation, enabling manufacturers and sellers to make more informed decisions to improve operational efficiency and customer satisfaction.

2. Research Methods

Research methods are the steps taken by researchers to collect research data. This data is processed and analyzed scientifically until the research objectives are achieved. The steps in this research include data collection techniques, data transformation, exploratory data analysis, model building, performance clustering, and clustering results (Figure 1).

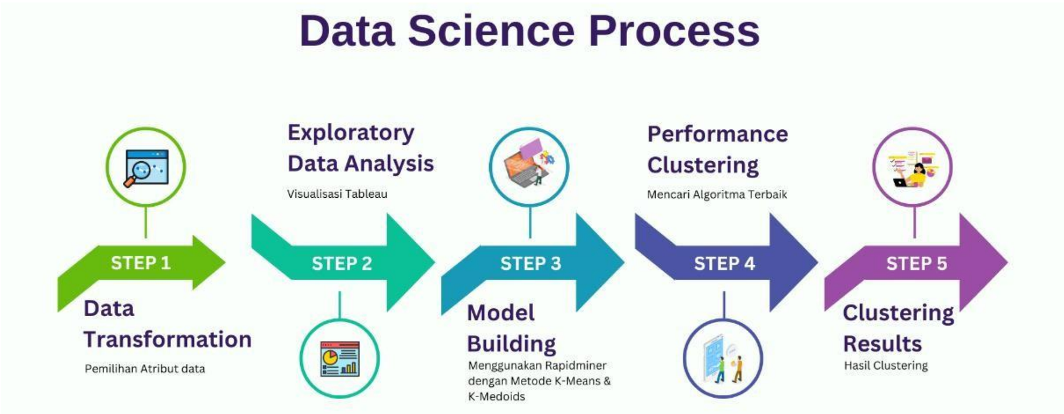

Figure 1 Proses Data Science

2.1 Data Collection Techniques

The data for this analysis was obtained through data download techniques from the Kaggle website (https://www.kaggle.com/datasets/muhammetvarl/laptop-price). The dataset used is the Laptop Price Dataset downloaded on September 6, 2024. This dataset was chosen because it contains laptop sales data that can be used to identify patterns in laptop specifications that influence sales prices. This data will be used to build a machine learning model to predict future sales.

2.2 Data Transformation

Data transformation is used to convert data into a suitable form for the data mining process. Data transformation is used to change the attribute data type from nominal to numerical with the Nominal To Numerical tool to match the required data type. Several techniques for data transformation are normalization, attribute selection, and discretization. Data normalization is a technique used in data mining to convert data values into a common scale. Attribute selection aims to select the most relevant subset of attributes from the original data. This technique reduces the number of features involved in determining a target class value. Discretization refers to the process of converting continuous data into a set of intervals. Discretization can help identify patterns that are not visible in the original data. This entire process improves data quality, making the analysis or model more accurate and efficient.

In this research, the data transformation used is attribute selection. Attribute selection is the process of selecting the most relevant and informative subset of attributes from a larger data set. The goal is to reduce data dimensionality and improve machine learning model performance.

2.3 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is an analytical approach to data to create a comprehensive summary that is easy to understand. Researchers can apply it with various statistical techniques and visualizations to uncover hidden patterns, trends, and relationships in the data. In this phase, it is essential to remain critical and not immediately draw conclusions from what is visible. Researchers must consider various possibilities and seek relevant evidence according to their research objectives. Hasty analysis can result in erroneous conclusions. Therefore, every finding must be carefully examined and validated.

In this process, the authors visualize data from the superstore to gain useful knowledge to better understand market trends. We use Tableau to facilitate the completion of this research. This tool makes it easier for researchers to process and analyze data effectively. Using Tableau as a data visualization tool has several significant advantages, including high flexibility in processing and visualizing data from various sources.

2.4 Model Building

The model-building process requires deep thought and understanding of the data and the algorithms to be used. In this process, researchers will use the K-Means and K-Medoids algorithms. In the context of the K-means algorithm, clusters usually refer to groups of data that are similar to each other. K-means is one of the most commonly used clustering methods in data analysis and machine learning. This method works by randomly defining cluster centers (centroids) and then calculating the distance between each data point and the centroid. Each data point will be grouped into the cluster with the nearest centroid. After all data points are grouped, new centroids are calculated based on the average of the data points in one cluster [10].

In addition to using the K-means algorithm, this research also compares it with other algorithms such as K-Medoids. K-Medoids is similar to K-means but uses actual data points as cluster centers, making it more robust against outliers. This research establishes three groups: most sold, sold, and not sold. These groups aim to categorize products based on sales levels. By comparing the two algorithms, it is expected to find the most effective clustering method for this sales analysis.

2.5 Performance Clustering

Based on the implementation results of several algorithms, performance results are obtained to compare which algorithm is more accurate. This performance evaluation is conducted using metrics such as the Davies-Bouldin Index for both algorithms, K-means and K-Medoids. The Davies-Bouldin Index metric is defined as the average ratio of within-cluster and between-cluster distances for each cluster with its nearest neighboring cluster. Using the same metric for both algorithms allows for a fair and consistent comparison in evaluating clustering quality.

From these results, it can be seen if there are significant differences in how the two algorithms group data. The best performance is determined not only by the metric value but also by the ease of interpreting the clustering results. Thus, the final decision on which algorithm is more accurate and effective can be made by considering various evaluation factors.

2.6 Clustering Results

This final stage is where the clustering results will be obtained. These clustering results will categorize laptops into several categories based on existing specifications: inches (screen size), price, RAM (capacity), and weight. With this information, companies can know which laptop categories have prices that match their specifications. This helps companies identify more specific market segments and develop more targeted marketing strategies. This knowledge is invaluable for making more effective business decisions.

Market trends inform about which laptop specifications are most favored by consumers and how prices affect purchase decisions. By knowing this, companies can focus their efforts on developing and marketing laptops with the most desirable specifications and price combinations. The most favored laptop subcategories can receive more attention in terms of promotion and inventory. Meanwhile, less favored laptops can be re-evaluated for improvements or additional promotions. Thus, companies can improve overall sales performance.

3. Result and Discussion

3.1 Data Transformation Results

In this research, we use the feature selection technique (attribute selection) to choose the most relevant subset of attributes from the original data (Figure 2 & Figure 3), reducing data dimensionality and improving model efficiency.

1	Company;Product;TypeName;Inches;ScreenResolution;CPU_Company;CPU_Type;CPU_Frequency (GHz);RAM (GB);Memory;GPU_Company;GPU_Type;OpSys;Weight (kg);Price (
2	Apple;MacBook Pro;Ultrabook;13.3;IPS Panel Retina Display 2560x1600;Intel;Core i5;2.3;8;128GB SSD;Intel;Iris Plus Graphics 640;macOS;1.37;1339.69
3	Apple;Macbook Air;Ultrabook;13.3;1440x900;Intel;Core i5;1.8;8;128GB Flash Storage;Intel;HD Graphics 6000;macOS;1.34;898.94
4	HP;250 G6;Notebook;15.6;Full HD 1920x1080;Intel;Core i5 7200U;2.5;8;256GB SSD;Intel;HD Graphics 620;No OS;1.86;575.0
5	Apple;MacBook Pro;Ultrabook;15.4;IPS Panel Retina Display 2880x1800;Intel;Core i7;2.7;16;512GB SSD;AMD;Radeon Pro 455;macOS;1.83;2537.45
6	Apple;MacBook Pro;Ultrabook;13.3;IPS Panel Retina Display 2560x1600;Intel;Core i5;3.1;8;256GB SSD;Intel;Iris Plus Graphics 650;macOS;1.37;1803.6
7	Acer;Aspire 3;Notebook;15.6;1366x768;AMD;A9-Series 9420;3.0;4;500GB HDD;AMD;Radeon R5;Windows 10;2.1;400.0
8	Apple;MacBook Pro;Ultrabook;15.4;IPS Panel Retina Display 2880x1800;Intel;Core i7;2.2;16;256GB Flash Storage;Intel;Iris Pro Graphics;Mac OS X;2.04;2139.97
9	Apple;Macbook Air;Ultrabook;13.3;1440x900;Intel;Core i5;1.8;8;256GB Flash Storage;Intel;HD Graphics 6000;macOS;1.34;1158.7
10	Asus;ZenBook UX430UN;Ultrabook;14.0;Full HD 1920x1080;Intel;Core i7 8550U;1.8;16;512GB SSD;Nvidia;GeForce MX150;Windows 10;1.3;1495.0
11	Acer;Swift 3;Ultrabook;14.0;IPS Panel Full HD 1920x1080;Intel;Core i5 8250U;1.6;8;256GB SSD;Intel;UHD Graphics 620;Windows 10;1.6;770.0
12	HP;250 G6;Notebook;15.6;1366x768;Intel;Core i5 7200U;2.5;4;500GB HDD;Intel;HD Graphics 620;No OS;1.86;393.9
13	HP;250 G6;Notebook;15.6;Full HD 1920x1080;Intel;Core i3 6006U;2.0;4;500GB HDD;Intel;HD Graphics 520;No OS;1.86;344.99
14	Apple;MacBook Pro;Ultrabook;15.4;IPS Panel Retina Display 2880x1800;Intel;Core i7;2.8;16;256GB SSD;AMD;Radeon Pro 555;macOS;1.83;2439.97
15	Dell;Inspiron 3567;Notebook;15.6;Full HD 1920x1080;Intel;Core i3 6006U;2.0;4;256GB SSD;AMD;Radeon R5 M430;Windows 10;2.2;498.9
16	Apple;"""MacBook 12"""""""";Ultrabook;12.0;IPS Panel Retina Display 2304x1440;Intel;Core M m3;1.2;8;256GB SSD;Intel;HD Graphics 615;macOS;0.92;1262.4
17	Apple;MacBook Pro;Ultrabook;13.3;IPS Panel Retina Display 2560x1600;Intel;Core i5;2.3;8;256GB SSD;Intel;Iris Plus Graphics 640;macOS;1.37;1518.55
18	Dell;Inspiron 3567;Notebook;15.6;Full HD 1920x1080;Intel;Core i7 7500U;2.7;8;256GB SSD;AMD;Radeon R5 M430;Windows 10;2.2;745.0
19	Apple;MacBook Pro;Ultrabook;15.4;IPS Panel Retina Display 2880x1800;Intel;Core i7;2.9;16;512GB SSD;AMD;Radeon Pro 560;macOS;1.83;2858.0
20	Lenovo;IdeaPad 320-15IKB;Notebook;15.6;Full HD 1920x1080;Intel;Core i3 7100U;2.4;8;1TB HDD;Nvidia;GeForce 940MX;No OS;2.2;499.0
21	Dell;XPS 13;Ultrabook;13.3;IPS Panel Full HD / Touchscreen 1920x1080;Intel;Core i5 8250U;1.6;8;128GB SSD;Intel;UHD Graphics 620;Windows 10;1.22;979.0
22	Asus;Vivobook E200HA;Netbook;11.6;1366x768;Intel;Atom x5-Z8350;1.44;2;32GB Flash Storage;Intel;HD Graphics 400;Windows 10;0.98;191.9
23	Lenovo;Legion Y520-15IKBN;Gaming;15.6;IPS Panel Full HD 1920x1080;Intel;Core i5 7300HQ;2.5;8;128GB SSD + 1TB HDD;Nvidia;GeForce GTX 1050;Windows 10;2.5;999.0
24	HP;255 G6;Notebook;15.6;1366x768;AMD;E-Series E2-9000e;1.5;4;500GB HDD;AMD;Radeon R2;No OS;1.86;258.0
25	Dell;Inspiron 5379;2 in 1 Convertible;13.3;Full HD / Touchscreen 1920x1080;Intel;Core i5 8250U;1.6;8;256GB SSD;Intel;UHD Graphics 620;Windows 10;1.62;819.0
26	HP;15-BS101nv (i7-8550U/8GB/256GB/FHD/W10);Ultrabook;15.6;Full HD 1920x1080;Intel;Core i7 8550U;1.8;8;256GB SSD;Intel;HD Graphics 620;Windows 10;1.91;659.0

Figure 2 Laptop Price Dataset before choosing variables.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Company	Product	TypeName	Inches	ScreenResolution	CPU_Company	CPU_Type	CPU_Frequency (GHz	RAM (GB	Memory	GPU_Company	GPU_Type	OpSys	Weight (kg	Price (Euro
2	Apple	MacBook Pro	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel	Core i5	2.3	8	128GB SSD	Intel	Iris Plus Graphics 640	macOS	1.37	1339.69
3	Apple	Macbook Air	Ultrabook	13.3	1440x900	Intel	Core i5	1.8	8	128GB Flash Storage	Intel	HD Graphics 6000	macOS	1.34	898.94
4	HP	250 G6	Notebook	15.6	Full HD 1920x1080	Intel	Core i5 7200U	2.5	8	256GB SSD	Intel	HD Graphics 620	No OS	1.86	575.0
5	Apple	MacBook Pro	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel	Core i7	2.7	16	512GB SSD	AMD	Radeon Pro 455	macOS	1.83	2537.45
6	Apple	MacBook Pro	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel	Core i5	3.1	8	256GB SSD	Intel	Iris Plus Graphics 650	macOS	1.37	1803.6
7	Acer	Aspire 3	Notebook	15.6	1366x768	AMD	A9-Series 9420	3.0	4	500GB HDD	AMD	Radeon R5	Windows 10	2.1	400.0
8	Apple	MacBook Pro	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel	Core i7	2.2	16	256GB Flash Storage	Intel	Iris Pro Graphics	Mac OS X	2.04	2139.97
9	Apple	Macbook Air	Ultrabook	13.3	1440x900	Intel	Core i5	1.8	8	256GB Flash Storage	Intel	HD Graphics 6000	macOS	1.34	1158.7
10	Asus	ZenBook UX430UN	Ultrabook	14.0	Full HD 1920x1080	Intel	Core i7 8550U	1.8	16	512GB SSD	Nvidia	GeForce MX150	Windows 10	1.3	1495.0
11	Acer	Swift 3	Ultrabook	14.0	IPS Panel Full HD 1920x1080	Intel	Core i5 8250U	1.6	8	256GB SSD	Intel	UHD Graphics 620	Windows 10	1.6	770.0
12	HP	250 G6	Notebook	15.6	1366x768	Intel	Core i5 7200U	2.5	4	500GB HDD	Intel	HD Graphics 620	No OS	1.86	393.9
13	HP	250 G6	Notebook	15.6	Full HD 1920x1080	Intel	Core i3 6006U	2.0	4	500GB HDD	Intel	HD Graphics 520	No OS	1.86	344.99
14	Apple	MacBook Pro	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel	Core i7	2.8	16	256GB SSD	AMD	Radeon Pro 555	macOS	1.83	2439.97
15	Dell	Inspiron 3567	Notebook	15.6	Full HD 1920x1080	Intel	Core i3 6006U	2.0	4	256GB SSD	AMD	Radeon R5 M430	Windows 10	2.2	498.9
16	Apple	"MacBook 12"""	Ultrabook	12.0	IPS Panel Retina Display 2304x1440	Intel	Core M m3	1.2	8	256GB SSD	Intel	HD Graphics 615	macOS	0.92	1262.4
17	Apple	MacBook Pro	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel	Core i5	2.3	8	256GB SSD	Intel	Iris Plus Graphics 640	macOS	1.37	1518.55
18	Dell	Inspiron 3567	Notebook	15.6	Full HD 1920x1080	Intel	Core i7 7500U	2.7	8	256GB SSD	AMD	Radeon R5 M430	Windows 10	2.2	745.0
19	Apple	MacBook Pro	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel	Core i7	2.9	16	512GB SSD	AMD	Radeon Pro 560	macOS	1.83	2858.0
20	Lenovo	IdeaPad 320-15IKB	Notebook	15.6	Full HD 1920x1080	Intel	Core i3 7100U	2.4	8	1TB HDD	Nvidia	GeForce 940MX	No OS	2.2	499.0
21	Dell	XPS 13	Ultrabook	13.3	IPS Panel Full HD / Touchscreen 1920x1080	Intel	Core i5 8250U	1.6	8	128GB SSD	Intel	UHD Graphics 620	Windows 10	1.22	979.0
22	Asus	Vivobook E200HA	Netbook	11.6	1366x768	Intel	Atom x5-Z8350	1.44	2	32GB Flash Storage	Intel	HD Graphics 400	Windows 10	0.98	191.9
23	Lenovo	Legion Y520-15IKBN	Gaming	15.6	IPS Panel Full HD 1920x1080	Intel	Core i5 7300HQ	2.5	8	128GB SSD + 1TB HDD	Nvidia	GeForce GTX 1050	Windows 10	2.5	999.0
24	HP	255 G6	Notebook	15.6	1366x768	AMD	E-Series E2-9000e	1.5	4	500GB HDD	AMD	Radeon R2	No OS	1.86	258.0
25	Dell	Inspiron 5379	2 in 1 Convertibl	13.3	Full HD / Touchscreen 1920x1080	Intel	Core i5 8250U	1.6	8	256GB SSD	Intel	UHD Graphics 620	Windows 10	1.62	819.0
26	HP	15-BS101nv (i7-8550U/8GB/256GB/FHD/W10)	Ultrabook	15.6	Full HD 1920x1080	Intel	Core i7 8550U	1.8	8	256GB SSD	Intel	HD Graphics 620	Windows 10	1.91	659.0
27	Dell	Inspiron 3567	Notebook	15.6	1366x768	Intel	Core i3 6006U	2.0	4	1TB HDD	Intel	HD Graphics 520	Windows 10	2.3	418.64
28	Apple	MacBook Air	Ultrabook	13.3	1440x900	Intel	Core i5	1.6	8	128GB Flash Storage	Intel	HD Graphics 6000	Mac OS X	1.35	1099.0
29	Dell	Inspiron 5570	Notebook	15.6	Full HD 1920x1080	Intel	Core i5 8250U	1.6	8	256GB SSD	AMD	Radeon 530	Windows 10	2.2	800.0
30	Dell	Latitude 5590	Ultrabook	15.6	Full HD 1920x1080	Intel	Core i7 8650U	1.9	8	256GB SSD + 256GB SSD	Intel	UHD Graphics 620	Windows 10	1.88	1298.0
31	HP	ProBook 470	Notebook	17.3	Full HD 1920x1080	Intel	Core i5 8250U	1.6	8	1TB HDD	Nvidia	GeForce 930MX	Windows 10	2.5	896.0
32	Chuwi	"LapBook 15.6"""	Notebook	15.6	Full HD 1920x1080	Intel	Atom x5-Z8300	1.44	4	64GB Flash Storage	Intel	HD Graphics	Windows 10	1.89	244.99
33	Asus	E402WA-GA010T (E2-6110/2GB/32GB/W10)	Notebook	14.0	1366x768	AMD	E-Series E2-6110	1.5	2	32GB Flash Storage	AMD	Radeon R2	Windows 10	1.65	199.0
34	HP	17-ak001nv (A6-9220/4GB/500GB/Radeon	Notebook	17.3	Full HD 1920x1080	AMD	A6-Series 9220	2.5	4	500GB HDD	AMD	Radeon 530	Windows 10	2.71	439.0
35	Dell	XPS 13	Ultrabook	13.3	Touchscreen / Quad HD+ 3200x1800	Intel	Core i7 8550U	1.8	16	512GB SSD	Intel	UHD Graphics 620	Windows 10	1.2	1869.0

Figure 3 Dataset Laptop Price

In this research use 4 attributes (Figure 4), here are some visualizations:

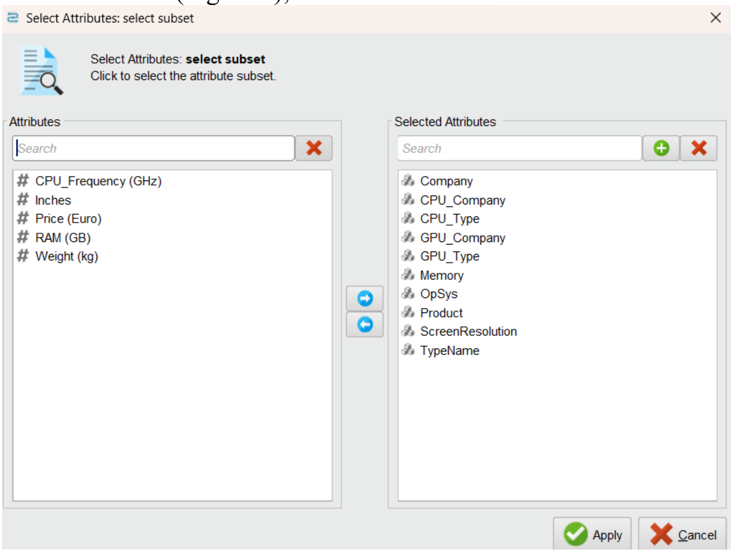

Figure 4 Attribute Dataset

3.2 Exploratory Data Analysis Results

The results of exploratory data analysis will display visualizations using Tableau. This aims to understand market trends and is useful for analyzing the most effective sales methods in general laptop sales stores. Here are some visualizations.

3.2.1 Distribution of Number of Laptops per Company

Figure 5 shows the distribution of the number of laptops available from various companies. Each bar in the graph represents a company, and the height of the bar reflects the number of laptop models offered by that company. From this graph, we can see which companies have the most laptop models on the market. This is useful for identifying the dominance of certain companies in the laptop industry and seeing the variety of products they offer.

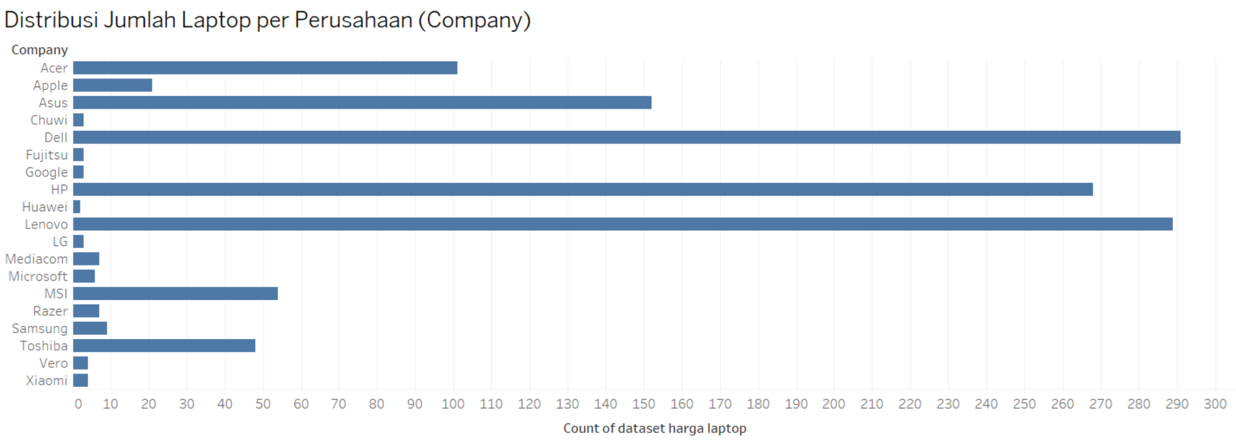

Figure 5 Distribution of Number of Laptops per Company

Figure 5 shows the number of laptops available in the dataset for each company.

- **HP**, **Lenovo**, and **Dell** have the highest number of laptops in the dataset. This indicates that products from these companies are more dominant in the market or that your dataset includes more laptops from these brands.
- **Apple**, **Razer**, and several other companies such as Mediacom have fewer laptops, indicating a more limited market coverage or a lack of product diversity recorded in the dataset.

3.2.2 Average Laptop Price per Company

This section visualizes the average laptop price per product company. Using the attributes Price (EURO) and Company. Knowing the average price helps companies in assessing pricing strategies.

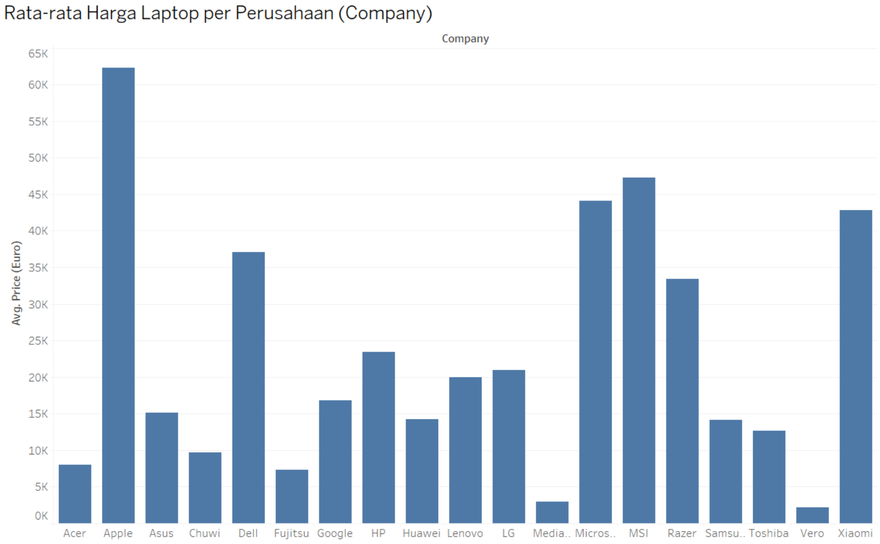

Figure 6 Average Laptop Price per Company

Figure 6 This graph shows the average price of laptops (in Euros) for each company.

- **Apple** has a much higher average price than other companies, indicating that their laptops are in the premium category**.**
- **Razer** and **Microsoft** also have fairly high average prices, but not as high as Apple.
- • **Acer**, **Asus**, and several other brands have lower average prices, indicating that they tend to produce more affordable laptops or are in the entry-level and mid-range categories.

3.2.3 Average Laptop Price by CPU Type

This section visualizes sales data by CPU Type. Using the attributes Company, CPU Type, and Price (Euro).

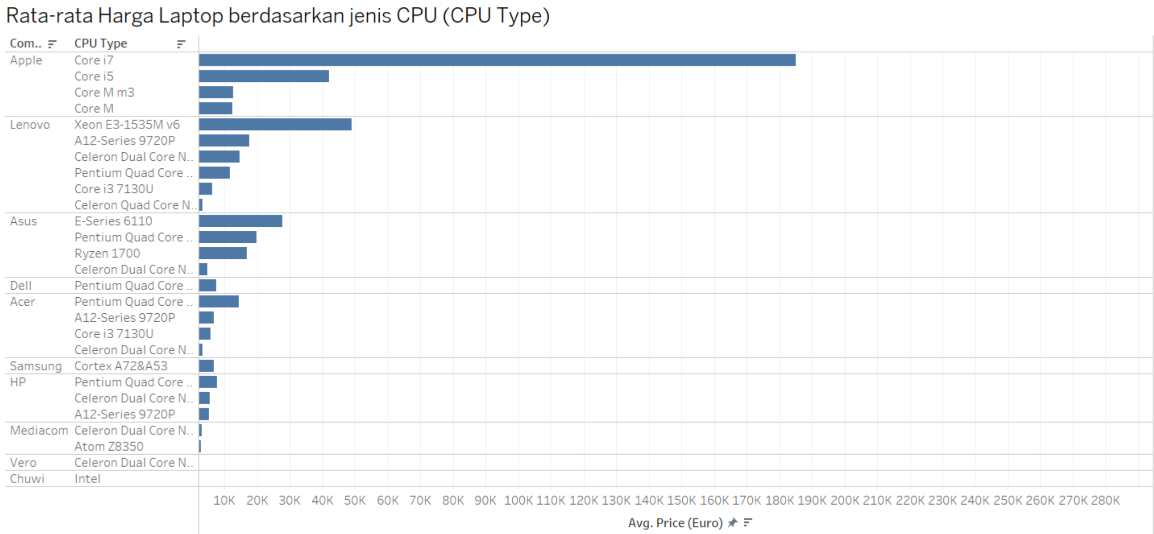

Figure 7 Average Laptop Price by CPU Type

Figure 7 This chart breaks down the average price of a laptop by brand and type of CPU used.

- Laptops with **Core i7 CPUs** from **Apple** have the highest average price, indicating that premium brands and components like these are driving the price.
- **Lenovo** has a wider range of prices with CPUs like the **Xeon**, **A12-Series**, and **Core i3**, which have lower average prices than premium CPUs.
- **Asus** and other brands like **HP** and **Acer** offer laptops with entry-level CPUs like the **Pentium** or **Celeron**, so the average price is more affordable.

3.2.4 Laptop Price Range and Quantity by Company

This section visualizes the laptop price range and quantity by Company. Using the attributes Number of dataset laptops, Price (Euro), and Company as a divider between the number of laptop models.

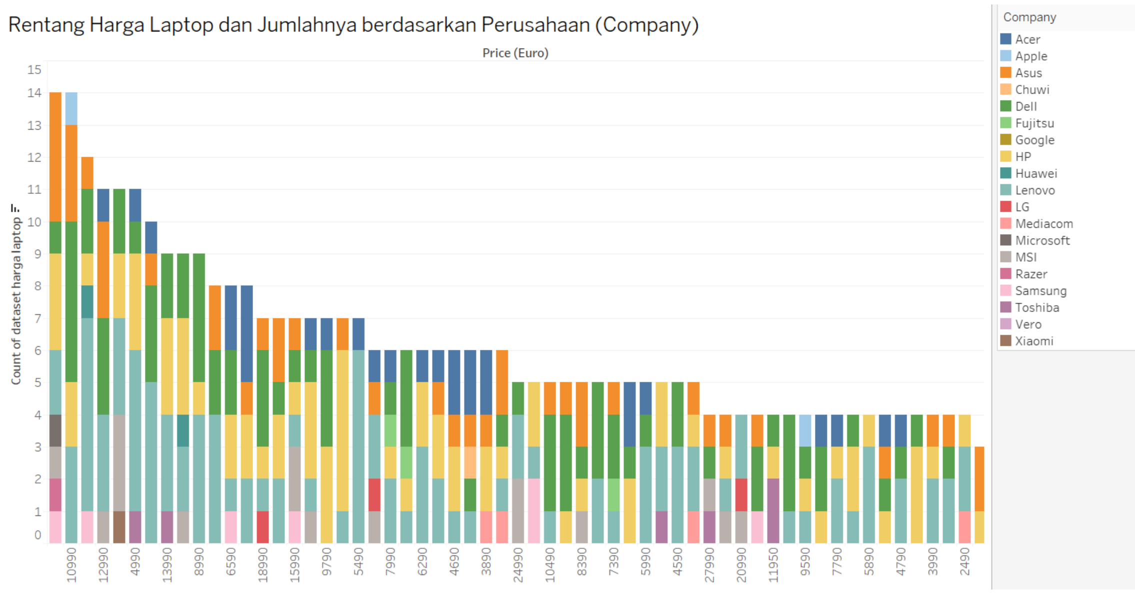

Figure 8 Laptop Price Range and Quantity by Company

Figure 8 shows the price ranges of laptops offered by different companies and the number of laptop models in each price range. Each company is represented by one section on the graph, which shows the distribution of laptop prices from lowest to highest. In this graph, we can see how widely each company varies in price and how many laptop models are available in each price category. This is useful for understanding the price segmentation carried out by companies and seeing their market strategies in reaching consumers from various economic segments.

3.3 Model Building Results

After visualizing the results from Tableau, proceed to model building using the K-Means and K-Medoids algorithms. Figure 9 explains that the operators used, namely retrieve dataset, are used to read the dataset in this research. Then add Select Attribute to choose the attribute that uses numerical data. Here, I use K values from 2 to 5 in one process, so I use Multiply for 4 types of clustering. K-Means and K-Medoids are used to model the existing dataset, and cluster distance performance is used to test the best clustering results. This research establishes 4 laptop groups based on Price, namely Low-End, Mid-Range, Mid to High-Range, and High-End.

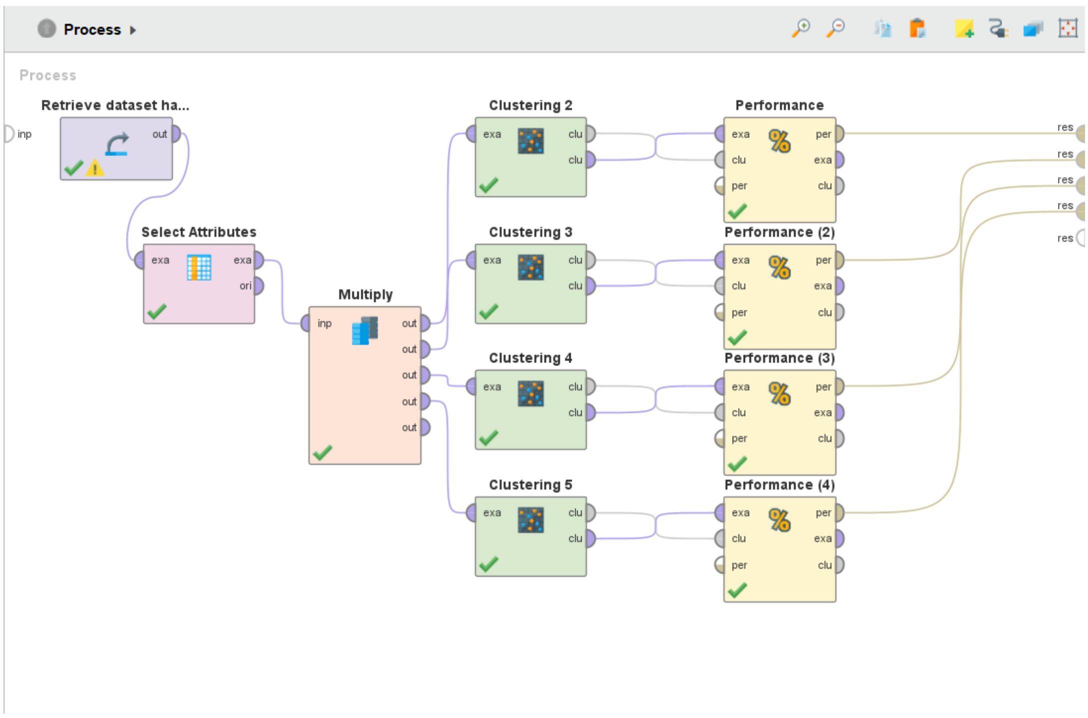

Figure 9 k-Mean - k-Medoids Models using RapidMiner

3.4 Performance Clustering Results

Based on the implementation results of the K-Means algorithm on RapidMiner, the performance results are as follows:

Performance Vector (Performance (4))
Result not stored in repository.

```
PerformanceVector:
Avg. within centroid distance: -43255.865
Avg. within centroid distance_cluster_0: -30303.570
Avg. within centroid distance_cluster_1: -146244.538
Avg. within centroid distance_cluster_2: -411067.468
Avg. within centroid distance_cluster_3: -35422.640
Avg. within centroid distance_cluster_4: -45645.278
Davies Bouldin: -0.487
```

Figure 10 Data Perfomance K-Means

Performance Vector (Performance (3))
Result not stored in repository.

```
PerformanceVector:
Avg. within centroid distance: -82737.154
Avg. within centroid distance_cluster_0: -56763.818
Avg. within centroid distance_cluster_1: -415087.973
Avg. within centroid distance_cluster_2: -10868.150
Avg. within centroid distance_cluster_3: -32414.850
Davies Bouldin: -0.665
```

Figure 11 Data Perfomance K-Medoids

Based on Figure 10 on data performance using the K-Means algorithm, the performance result is -0.487 Davies Bouldin. Meanwhile, Figure 11 using the K-Medoids algorithm, the performance result is -0.665. It can be concluded that the clustering performed by the K-Means algorithm is considered better than K-Medoids. K-Means shows more accurate and better performance compared to K-Medoids because according to the Davies-Bouldin Index (DBI) principle, the desired value in data mining is getting smaller or closer to zero. Thus, it shows better clustering quality.

No	Metode	K	Davies Bouldin
1	K-Means	2	-0.559
2	*null*	3	-0.553
3	*null*	4	-0.531
4	*null*	5	-0.487
5	K-Medoids	2	-0.616
6	*null*	3	-0.532
7	*null*	4	-0.665
8	*null*	5	-0.616

Figure 12 Comparison of Performance K-Means & K-Medoids Results

It can also be seen from Figure 12 that the K-Means performance results table is closer to zero or has a smaller value compared to the K-Medoids performance results.

3.5 Clustering Results

Based on the above performance values, Cluster 0 with 447 items is categorized as Mid-Range, Cluster 1 with 12 items is categorized as High-End, Cluster 2 with 583 items is categorized as Low-End, and Cluster 3 with 225 items is categorized as Mid to High Range. This can be seen from Figure 13 below:

Cluster Model

```
Cluster 0: 447 items
Cluster 1: 12 items
Cluster 2: 583 items
Cluster 3: 225 items
Total number of items: 1267
```

Figure 13 Cluster Model

Furthermore, to identify Laptop Prices included in the 4 previously mentioned clusters, visualization can be done using RapidMiner. This platform allows in-depth analysis of cluster data, enabling companies to quickly identify hidden patterns and trends in product sales. With this information, companies can direct marketing strategies and inventory management more precisely, ensuring that each Laptop Price receives the appropriate attention according to market potential and customer demand.

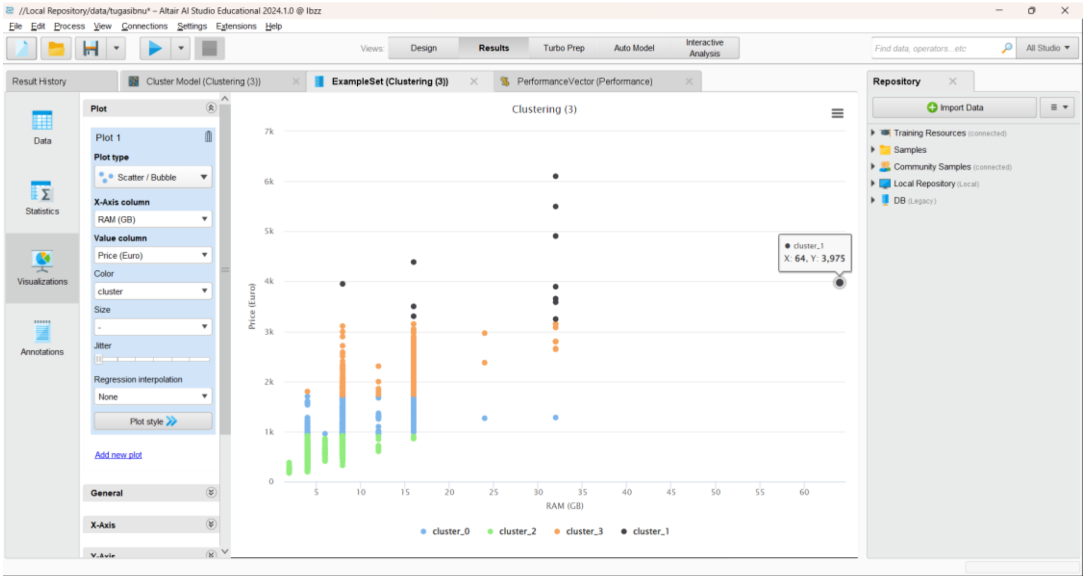

Figure 14 RapidMiner Visualization

From the visualization results using RapidMiner above (Figure 14), it can be seen (Table 1) that laptops with 32-64GB RAM dominate in cluster 1, which is the High-End category, with the highest price reaching 6,099 EURO. In the second position, 8-32GB RAM records the most laptops, which is cluster 3 in the Mid to High-End category. In cluster 0, which is the Mid-Range category, laptops with 4-16GB RAM are quite popular, with prices ranging from 9.19 EURO to 1,713 EURO. Meanwhile, in cluster 2, which is the Low-End category, this cluster is the cheapest because based on its specifications, it has 2-8GB RAM, with the lowest price being 174 EURO. This information provides a clear view of the sales performance of each Laptop, enabling companies to take strategic steps in managing inventory and directing marketing efforts more effectively according to the identified market conditions.

Table 1 Clustering Results

No.	Low-End(Cluster 2)	Harga	Mid-Range (Cluster 0)	Harga	Mid to High-Range(Cluster 3)	Harga	High-End(Cluster 1)	Harga
1	2-16GB	174-915 EURO	4-16GB	919-1.713 EURO	8-32GB	1.725-3.154 EURO	16-64 GB	3.240-6.099 EURO

4. Conclusion

The conclusion of the report "Cluster Analysis on Laptop Sales Data and Specifications Using K-Means and K-Medoids Methods" is that this research successfully identifies prices based on modeling laptop types by clustering products into 4 categories: Low-End, Mid-Range, Mid to High-Range, and High-End. Using data from the Kaggle site and analysis tools such as Microsoft Excel, Tableau, and RapidMiner, this research finds that the K-Means algorithm provides more accurate clustering performance compared to K-Medoids. This finding shows that the K-Means performance value is more accurate, at -0.487. K-Means is more suitable for the characteristics of the existing data, so this information can be used by companies to optimize marketing strategies and inventory management, focusing on products with the best sales performance. This conclusion emphasizes the importance of structured data analysis in supporting better and timely business decision-making.

References

[1] W. Aprilyani *et al.*, “Klasterisasi Data Penjualan Alat Transportasi Dengan Rapidminer Menggunakan Metode K-Medoid," vol. 8, no. 2, pp. 1348–1353, 2024.

[2] H. Syahputra, “Clustering Tingkat Penjualan Menu (Food and Beverage) Menggunakan Algoritma K-Means,” *J. KomtekInfo*, vol. 9, pp. 29–33, 2022, doi: 10.35134/komtekinfo.v9i1.274.

[3] A. Nugraha, O. Nurdiawan, and G. Dwilestari, “Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 849–855, 2022, doi: 10.36040/jati.v6i2.5755.

[4] M. N. Setiawan, Purwono, and I. A. Ashari, “Terakreditasi SINTA Peringkat 4 Analisa Cluster Data Transaksi Penjualan Minimarket Selama Pandemi Covid-19 dengan Algoritma K-means,” vol. 3, no. 1, pp. 153–160, 2018.

[5] R. Gustrianda and D. I. Mulyana, “Penerapan Data Mining Dalam Pemilihan Produk Unggulan dengan Metode Algoritma K-Means Dan K-Medoids,” *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 27, 2022, doi: 10.30865/mib.v6i1.3294.

[6] L. ANWER, “SuperStore Sales Dataset,” www.kaggle.com. Accessed: Jun. 25, 2024. [Online]. Available: https://www.kaggle.com/datasets/laibaanwer/superstore-sales-dataset

[7] N. Suarna and Y. Arie Wijaya, “Analisa Penerapan Metode Clustering K-Means Untuk Pengelompokan Data Transakasi Konsumen (Studi Kasus: Cv. Mitra Indexindo Pratama),” *J. Mhs. Tek. Inform.*, vol. 7, no.2, pp. 1322–1328, 2023.

[8] geeksforgeeks, “Normalisasi Data dalam Data Mining,” www.geeksforgeeks.org. Accessed: Jun. 25, 2024.[Online]. Available: https://www.geeksforgeeks.org/data-normalization-in-data-mining/

[9] Trivusi, “Metode-Metode dalam Feature Selection,” www.trivusi.web.id. Accessed: Jun. 25, 2024. [Online]. Available: https://www.trivusi.web.id/2019/02/metode-metode-dalam-feature-selection.html

[10] D. Vogiatzis, “Panduan Anda tentang Teknik Transformasi Data,” blog.coupler.io. Accessed: Jun. 25, 2024. [Online]. Available: https://blog.coupler.io/data-transformation-techniques/

[11] Y. Ws, “Mengenal Exploratory Data Analysis," eksplorasidata.mipa.ugm.ac.id. Accessed: Jun. 25, 2024. [Online]. Available: https://eksplorasidata.mipa.ugm.ac.id/2021/08/16/mengenal-exploratory-data-analysis/

[12] R. Maoulana, B. Irawan, and A. Bahtiar, “Data Mining Dalam Konteks Transaksi Penjualan Hijab Dengan Menggunakan Algoritma Clustering K-Means,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 515– 521, 2024, doi: 10.36040/jati.v8i1.8504.

[13] Y. Sopyan, A. D. Lesmana, and C. Juliane, “Analisis Algoritma K-Means dan Davies Bouldin Index dalamMencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan,” *Build. Informatics, Technol. Sci.*, vol.4, no. 3, pp. 1464–1470, 2022, doi: 10.47065/bits.v4i3.2697.

[14] Youtap, “Cara Jitu Mengetahui Produk Paling Laris Sesuai Tren Pasar,” www.youtap.id. Accessed: Jun. 25, 2024. [Online]. Available: https://www.youtap.id/blog/cara-mengetahui-produk-paling-laris

[15] N. Ameliana, N. Suarna, and W. Prihartono, “Analisis Data Mining Pengelompokkan Umkm Menggunakan Algoritma K-Means Clustering Di Provinsi Jawa Barat,” vol. 8, no. 3, pp. 3261–3268, 2024.