



# Smartphone Price Grouping by Specifications using K-Means Clustering Method

Ahmad Agung Zefi Syahputra<sup>1\*</sup>, Annisa Dwi Atika<sup>2</sup>, Muhammad Adam Aslamsyah<sup>3</sup>, Meida Cahyo Untoro<sup>4</sup>, Winda Yulita<sup>5</sup>

<sup>1,2,3,4,5</sup> Teknik Informatika, Institut Teknologi Sumatera, Jl. Terusan Ryacudu Lampung Selatan 35365, Indonesia

Email: <sup>1</sup> [ahmad.118140083@student.itera.ac.id](mailto:ahmad.118140083@student.itera.ac.id) \*; <sup>2</sup> [annisa.118140082@student.itera.ac.id](mailto:annisa.118140082@student.itera.ac.id);  
<sup>3</sup> [muhhammad.118140154@student.itera.ac.id](mailto:muhhammad.118140154@student.itera.ac.id); <sup>4</sup> [cahyo.untoro@if.itera.ac.id](mailto:cahyo.untoro@if.itera.ac.id); <sup>5</sup> [winda.yulita@if.itera.ac.id](mailto:winda.yulita@if.itera.ac.id)  
\*correspondence author

## ARTICLE INFO

## ABSTRACT

### Article history:

Received: Jun 03, 2021;  
Revised: Jun 14, 2021;  
Accepted: Jul 22, 2021;  
Available online: Sep 30, 2021.

### Keywords:

Smartphone;  
K-Means Clustering,  
Price,  
Data Mining.

The use of smartphones in the industrial era 4.0 had become more frequent and widespread in various circles of Indonesian society. In addition, the COVID-19 pandemic that had not end yet also made high school and college students obliged to carry out online learning. This research aimed to cluster the price from smartphones using the specifications of the smartphone. K-Means Clustering was used as a method in this research. This algorithm was a data mining algorithm with unsupervised learning as data grouping and could group the price of a smartphone into several clusters based on the similarity of the characteristics by one data with other data, which is `memory_size` and `best_price`. The results of this research indicated that the right clustering of smartphone prices was within 3 different clusters, which was cluster 0 has centroid of Rp2.000.000,00, cluster 1 has centroid of Rp18.000.000,00, and cluster 2 has centroid of Rp9.000.000,00. The results of the evaluation used a confusion matrix, summary of prediction result, indicated that the clustering process had 100% of accuracy that could be seen on the table which showed the results of clustering. The conclusion from this research was that K-Means Clustering could form clusters in determining the price of a smartphone in relation to the specifications used as the attribute determining the price cluster for a smartphone.

© 2021 JTI C.I.T. All rights reserved.

## 1. Introduction

The use of the internet in this era has become a habit for everyone. This also makes people easier now to get the information they need quickly and save time because the internet can be used as a search engine, file delivery media, electronic mail delivery, marketing media, etc [1]. This can be proved easily the circulation of information about events or events in a distant place but can be accessed on the Internet by anyone. Communication between humans also becoming easier and more efficient because it is not obstructed by distance and time.

Reviewed from the data released by Asosiasi Penyelenggara Jasa Internet Indonesia, the second quarter in 2020, there were 196.7 million internet users in Indonesia, or close to 73.7% of the population [2]. This is a large number with the results of monitoring data published by Badan Pusat Statistik RI in 2020, Indonesia's population has reached 270.2 million people [3]. That way, many Indonesians are "internet literated" and can use it in finding information and implementing long-distance communication.

Alongside with that, the pandemic COVID-19 that began in early March 2020 in Indonesia has led to the education process changes. Students are required to carry out Long Distance Learning



independently from home using the internet media [4]. This also applies to teachers and lecturers who had time to teach from their homes named Work from Home (WFH) [5]. During the learning is done online, the students will use their respective devices such as smartphone, laptop, tablet, and PC to access media that is provided by schools and colleges.

The number of students in Indonesia is 68,729,037 [6] people all students will use online learning media following the official rules written by the Ministry of Education and Culture in 2020. The rules are about regarding the implementation of the learning process carried out from home online during the COVID-19 pandemic. With these rules, the use of devices in carrying out the learning process becomes important as a distance learning tool. Devices used to attend in distance learning such as smartphones, laptops, computers, tablets, and others so that information can be reached easily and quickly at any location and unlimited time [7][8]. The device used must have an internet connection to properly access online learning media.

Related to the device options used for online media learning, one of the options that can be selected is a smartphone. In this case, the smartphones that students can use are phones that can connect to the internet and have accessibility to other media, such as the ability to take pictures and a long battery life. This is necessary so that students can study optimally and use the tools available on their smartphones so that learning can run more easily.

Circulation of smartphones in the market with various specifications has an impact such as many choices for students to use smartphones according to their respective needs. This also makes the prices of smartphones have a variety of price ranges. The dataset from Kaggle shows that there are 12 variables that determine the price range of a smartphone [9]. Therefore, it is necessary to classify the price of a smart phone based on appropriate and precise specifications, especially for students, making it easier for them to choose a smartphone as a media for distance learning.

To do the research become appropriate in a scientific way, the researchers plan to do it by using Jupyter Notebook to create a Python program and do some data preprocessing in it, so the data can be used effectively before the method is applied to the dataset. This research's goal is to cluster the price from smartphones using the dataset that contains specifications of the smartphones. This research is expected to create the right cluster where the grouping of smartphone specifications is similar in each cluster. With these clusters, it has the benefit to help readers in choosing a smartphone based on the price and specifications needed.

One of the studies on price prediction that has been done was a study that aims to find out the algorithm with the best accuracy in predicting a mobile phone price using datasets from Kaggle. This study applied the use of WEKA application in processing existing datasets. The comparison in this study used several methods, such as ZeroR algorithm, Naive Bayes, and J48 Decision Tree. The results explained that J48 algorithm could predict the price of mobile phones with 84% accuracy with a dataset of 104 records that trained with various attributes on the dataset using WEKA tool [10].

Price prediction research had also been conducted to made a stock price prediction of PT. Telekomunikasi Indonesia with the Support Vector Machine (SVM) method used a dataset from Yahoo Finance [11]. The dataset contained as many as 1265 records per day from 2015 to 2020, with a split of 20% testing data and 80% training data. Gaussian Radial basic Function (RBF) was the SVM kernel used in this study and be able to got the highest accuracy value of 0.9635 and RMSE of 0.091. In this study, testing was also carried out using the KNN algorithm with K values of 2, 4, and 6. The result of the accuracy value obtained was 0.9456 with the RMSE magnitude of 0.1162, so it was concluded that compared to the KNN method, the SVM method had a higher accuracy score.

Related to research for price prediction, Fatkhuroji, Sefanus Santosa, Ricardus Anggi Pramunendar [12] had been done some research that aimed to predict the price of local soybeans and imported soybeans using a Support Vector Machine (SVM). The first stage was the SVM method without optimization, tested with Dot, Radial, Polynomial, Neural, and Anova as parameter types and tested with 5 to 40 inputs. The C values used in this test were 0.1, 0.3, and 0.5 using Forward Selection as the basis, which also applied the 10 and 15-fold cross validation evaluation methods. The dataset used in this research was data that sourced from the Ministry of Trade with total 1089 records from 2011 to 2015. The results of this study obtained 79,749 +/-16,051 for the RMSE value based on input 4 days previously. Then, the RMSE increased to 74,276 after an optimization was carried out using Forward Selection.

Then, based on input 5 days before, the prediction of local soybeans obtained an RMSE value of 122,270 +/-56,049. This value increased by 3,738 after being optimized using Forward Selection.

The next research was conducted to predict commodity prices on the characteristics of consumers of national local coffee sales products [13]. The method used in this study was the Backpropagation algorithm, method that used to reduce error by adjusting weight based on differences output and targets, algorithm that used on an artificial neural network using primary data from user input and other primary data, such as from the Medan City Industry Service. The main target in this study was to predict the price of Arabica and Robusta coffee and to predict the commodity prices of these coffee products locally and nationally using the Matlab R2013a application. The variables used in building a network in this study were the iteration (epoch), the rate of learning (learning rate), and the number of neurons in each hidden layer which was determined by looking at the Mean Square Error (MSE) during training was performed. After doing the research, the Backpropagation algorithm managed to get an accuracy above 99.99%.

## 2. Research Method

### 2.1 Research Flow Chart

This research was conducted with several stages wrote in Figure 1. This flowchart showed the stages in this research that started from the identification of problems. The next stages were data collection, dataset preprocessing, data visualization, application of methods (K-Means Clustering) [16] [17] [18], and also evaluation and results.

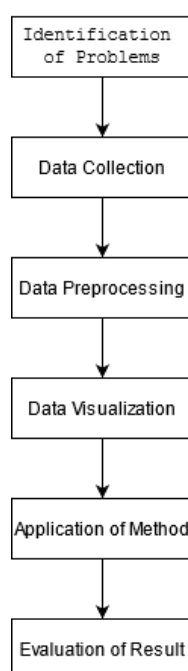


Fig 1. Flow Chart Diagram in Smartphone Prices Clustering Research

### 2.2 Data Collection

The research was conducted in March 2021 using a dataset from Kaggle named Mobile Phones Data which was updated in February 2021 with a total of 1224 records. The attributes in the dataset contained 12 attributes, such as brand\_name, model\_name, os, popularity, best\_price, lowest\_price, highest\_price, seller\_amount, screen\_size, memory\_size, battery\_size, and release\_date. Table 1 with the contents of the dataset from the initial five data that used as research materials.

**Table 1**  
Several Contents Of Dataset

Brand_name	Model_name	Os	Popularity	Best_price	Lowest_price	Highest_price	Sellers_amount	Screen_size	Memory_size	Battery_size	Release_date
ALCATEL	1 1/8GB Bluish Black (5033D - 2JALUA A)	Android	422	1690	1529	1819	36	5	8	2000	10-2020
ALCATEL	1 5033D 1/16G B Volcano Black (5033D - 2LALU AF)	Android	323	1803	1659	2489	36	5	16	2000	9-2020
ALCATEL	1 5033D 1/16G B Volcano Black (5033D - 2LALU AF)	Android	299	1803	1659	2489	36	5	16	2000	9-2020
ALCATEL	1 5033D 1/16G B Volcano Black (5033D - 2LALU AF)	Android	287	1803	1659	2489	36	5	16	2000	9-2020
Nokia	1.3 1/16G B Charcoal	Android	1047	1999	NaN		10	5,71	16	3000	4-2020

Table 1 shows the dataset with five initial data, it can be seen that there are two brand names, Nokia and Alcatel. In the dataset, there are 64 brand names including Samsung, Apple, OPPO, ASUS, LG, and others.

### 2.3 Data Preprocessing

The existing dataset will get preprocessing stage before the K-Means Clustering method applies in processing the dataset [26]. In this dataset, preprocessing removes unnecessary columns, converts the release\_date data type to datetime, fills the missing values, and converts smartphone prices into Rupiah exchange rates to facilitate researchers in applying method, reading classification results, and making this research in accordance with the conditions in Indonesia [15].

### 2.4 Data Visualization

This process is carried out to seek insights from the preprocessed dataset so that the dataset can be processed using the K-Means Clustering method [24]. Data visualization is done by using data analysis methods named Exploratory Data Analysis (EDA) [25]. EDA is a part of the data science process that allows analysts to understand the content of the data used, starting from distribution, frequency,

correlation and others. The results of the data visualization will get processed by applying the K-Means Clustering method to obtain smartphones price grouping.

## 2.5 Application of Methods

Next, the preprocessing dataset was tested using the proposed method, K-Means Clustering. At this stage, the application of the method was carried out using the Jupyter Notebook on the Acer Aspire 3 A315-41 model with an AMD Ryzen 3 3200U processor. This K-Means method was suitable for grouping small data, so the researchers chose K-Means Clustering as the method because the data that the researchers uses is included in the category of small data.

Prior to using the K-Means Clustering method, the researchers used the Elbow method to determine the correct value of cluster. Elbow method is used to determine the best number of clusters by looking at the percentage of the comparison results between the number of clusters that will form an elbow at a point. After obtained the correct value of cluster, the K-Means method could be used and visualized to display the clusters and centroids [15][16][17][18].

## 2.6 K-Means Clustering

K-Means Clustering is the most popular and most commonly used method of partitioning data [19] [20]. This method divides the dataset into clusters with a center point (centroid). The way the K-Means algorithm works is to determine the centroid as the average value of a cluster and then partition the data available for research as much as  $k$  [14][21]. In addition, this method classifies data based on the same traits or characteristics into the same cluster, and vice versa [22]. The absence of research conducted in predicting the price of a smartphone using this method motivates the authors to conduct this research. The K-Means algorithm has the following stages [23].

- 1) Determine the quantity value of  $k$  as the number of clusters and form this later in the process.
- 2) Determine the random value as the initial center/centroid. Then, using the Euclidean Distance formula, calculate the distance between each input data point and each centroid.

$$d(x_i, \mu_j) = \sqrt{\sum (x_i, \mu_j)^2} \quad (1)$$

with

$x_i$  : data that has certain criteria

$\mu_j$  : center point in cluster  $j$

$\sum$  : summation

- 3) Determine the cluster of each data according to the smallest distance from a centroid.
- 4) With the formula, update the value in the centroid new obtained from a cluster's average

$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in s_j} x_j \quad (2)$$

with

$\mu_j(t+1)$  : the new center point in the iteration  $(t+1)$

$N_{sj}$  : existing data in the  $S_j$  group

$\sum$  : summation

If the data for each *cluster* has not stopped changing, repeat steps 2 to 5, until there is a change from the members of each *cluster*.

## 2.7 Evaluation

Evaluation was carried out to get insights on whether the results of applying the method were good and had a good level of accuracy. The evaluation carried out in this study consisted of the following steps.

- 1) Evaluation of Results  
At this stage, the researchers provided an assessment of the results to determine how well the results of creating a model data mining in completing the study's objectives.
- 2) Re-checking the Research Process  
At this stage, the researchers double-checked all the stages that were passed in the research, especially in preprocessing so that no steps were missed.
- 3) Determining Advanced Steps

This stage was carried out to determined the next step when the research results had been obtained, returning to the goal-setting step or analyzing the results obtained.

4) Implementing the Confusion Matrix

This stage was carried out by researchers to got the level of accuracy of the clustering process that had been carried out.

### 3. Result and Discussion

#### 3.1 Data Analysis

From the results of data visualization, researchers get information that the most brands of smartphones on the dataset is Samsung, followed by Xiaomi and Apple. The data visualization is shown in the Figure 2.

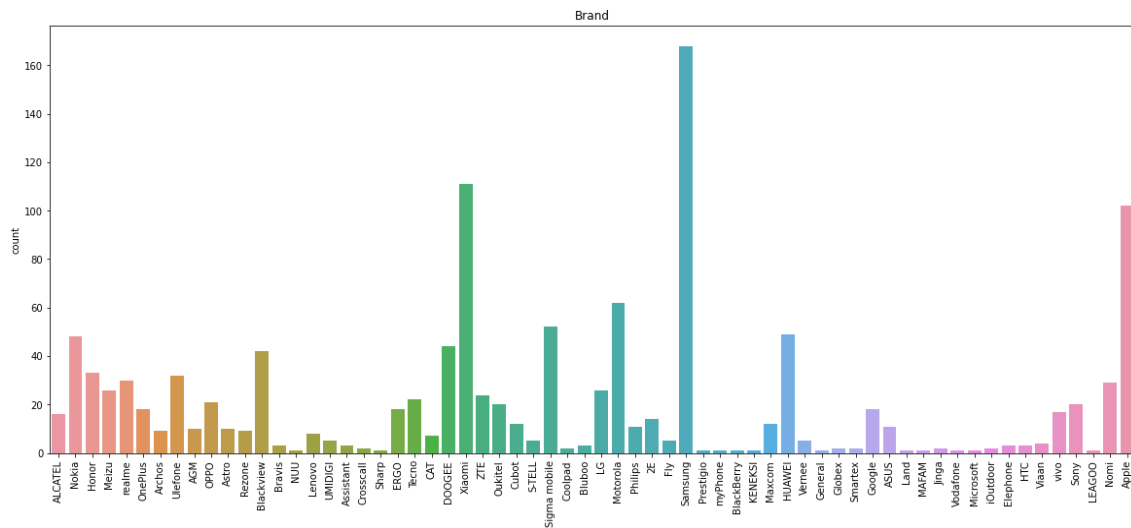


Fig 2. Visualization from most brand of smartphones

Next, the researchers look for a correlation between an attribute and another attribute in the dataset. The search results showed that the most influential attribute or high correlation with price is the memory size (memory\_size) with a correlation rate of 69%. It is shown in the Figure 3.

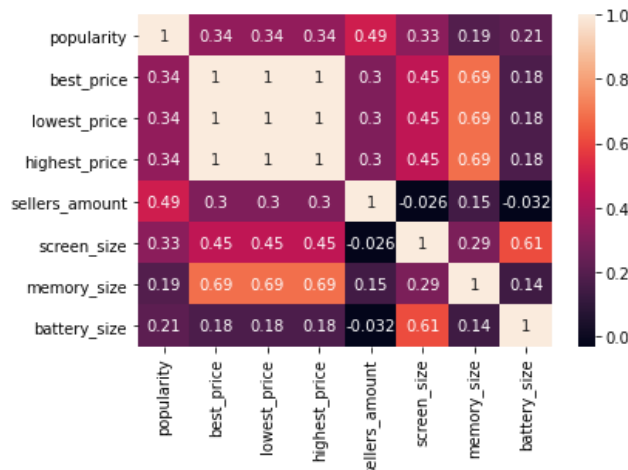


Fig 3. Visualization from Smartphone Attribute Correlation

In principle, researchers use the `best_price` attribute as the main price attribute of a smartphone because it is also used to fill in missing values in the `lowest_price` column and `highest_price` whose value is empty.

### 3.2 Grouping (Clustering)

By applying the Elbow Method, researchers obtained that the optimal number of clusters is three clusters as shown in the Figure 4. The three clusters named Cluster 0, Cluster 1, and Cluster 2.

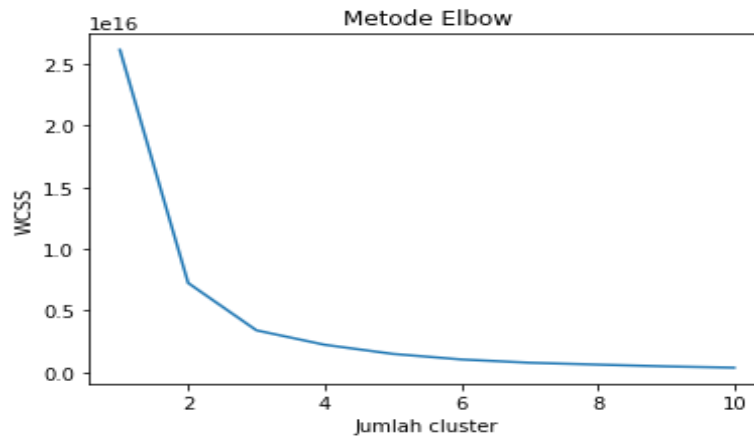


Fig 4. Cluster Determination Results by Elbow Method

After looking for the right number of clusters, which number is three clusters, researchers apply the K-Means Clustering method to the dataset that has been preprocessed with the following source code.

```
kmeans = KMeans(n_clusters = 3, init = 'k-means++', random_state = 42)
y_kmeans = kmeans.fit_predict(X)
y_kmeans
```

Clustering is done based on two attributes that have the highest correlation by the result of searching for insights in the data visualization stage. Then to show the clustering results, the researchers used the following source code.

```
plt.scatter(X[y_kmeans == 0, 0], X[y_kmeans == 0, 1], s = 100, c = 'blue', label = 'Cluster 0')
plt.scatter(X[y_kmeans == 1, 0], X[y_kmeans == 1, 1], s = 100, c = 'red', label = 'Cluster 1')
plt.scatter(X[y_kmeans == 2, 0], X[y_kmeans == 2, 1], s = 100, c = 'yellow', label = 'Cluster 2')
plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s = 200, c = 'green', label =
'Centroids')
plt.title('Cluster Harga')
plt.xlabel('Best Price')
plt.ylabel('Memory Size')
plt.legend()
plt.show()
```

When the source code is successfully executed, a clustering result will appear in the form of scatterplot with centroid as shown in the Figure 5.

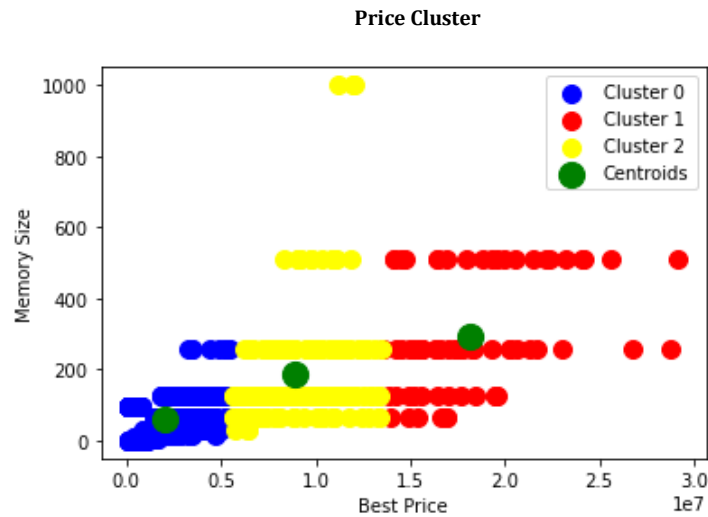


Fig 5. Clustering Result Visualization

From the visualization, it can be known that there are three clusters, namely cluster 0, where the centroid on the cluster is close to the number of Rp2,000,000.00 indicated by the blue scatterplot. Furthermore, there is a cluster 1 with centroids approaching the Rp18,000,000.00 shown with the red scatterplot. Then the last one, cluster 2 has a centroid that is close to Rp9,000,000.00 indicated by a yellow scatterplot. The researchers created a table with the contents of the results from clustering several smartphones in the dataset which can be seen in Table 2.

Table 2

Last 5 Records Clustering Results On Dataset

Brand_name	Model_name	OS	Best_price	Memory_size	Cluster
Apple	iPhone XS Max 64GB Gold (MT522)	iOS	11796200	64	2
Apple	iPhone XS Max Dual Sim 64GB Gold (MT732)	iOS	12792000	64	2
HUAWEI	nova 5T 6/128GB Black (51094MEU)	Android	4578080	128	0
ZTE	nubia Red Magic 5G 8/128GB Black	Android	9752600	128	2
Sigma mobile	x-style 35 Screen	Other	471640	95	0

Then, Table 3 is a table containing clusters of smartphones in the order of 353 to 358 from researchers notebook.

Table 3

Clustering Result In 353-358<sup>th</sup> Records

Brand_name	Model_name	OS	Best_price	Memory_size	Cluster
Fly	FF250 Black	Other	163280	95	0
Fly	FF281 Dual Sim (Black)	Other	284440	0	0
Fly	FF281 Silver	Other	283400	0	0
OPPO	Find X2 12/256GB Ocean Black	Android	15599480	256	1
OPPO	Find X2 Pro 12/512GB Orange	Android	16915600	512	1

To know the details, the researchers show the number of smartphones included on cluster 0 is 210 obtained from the number of data rows in the cluster. The initial five data of cluster 0 as shown in Table 4.

Table 4

First 5 Records Smartphone On Cluster 0

Brand_name	Model_name	OS	Best_price	Memory_size	Cluster
Honor	10 6/64GB Black	Android	5649800	64	0

Brand_name	Model_name	OS	Best_price	Memory_size	Cluster
OnePlus	6T 8/256GB Midnight Black	Android	8125000	256	0
OnePlus	7 8/256GB Mirror Gray	Android	7510360	256	0
OnePlus	8 12/256GB Onyx Black	Android	8929440	256	0
OnePlus	8 8/128GB Onyx Black	Android	8236800	128	0

Next, the number of smartphones on cluster 1 is 944 obtained from the number of data rows in the cluster. The initial five data of cluster 1 can be seen in Table 5.

**Table 5**  
First 5 Records Smartphone On Cluster 1

Brand_name	Model_name	OS	Best_price	Memory_size	Cluster
ALCATEL	1 1/8GB Bluish Black (5033D-2JALUAA)	Android	878800	8	1
ALCATEL	1 5033D 1/16GB Volcano Black (5033D-2LALUAF)	Android	937560	16	1
ALCATEL	1 5033D 1/16GB Volcano Black (5033D-2LALUAF)	Android	937560	16	1
ALCATEL	1 5033D 1/16GB Volcano Black (5033D-2LALUAF)	Android	937560	16	1
Nokia	1.3 1/16GB Charcoal	Android	1039480	16	1

Lastly, the number of smartphones in cluster 2 as many as 70 obtained from the number of data rows in the cluster. The initial five data of cluster 2 is shown in the Table 6.

**Table 6**  
First 5 Records Smartphone On Cluster 2

Brand_name	Model_name	OS	Best_price	Memory_size	Cluster
OPPO	Find X2 12/256GB Ocean Black	Android	15599480	256	2
OPPO	Find X2 Pro 12/512GB Orange	Android	16915600	512	2
Samsung	Galaxy Fold 12/256GB Black	Android	28775760	256	2
Samsung	Galaxy Fold 12/512GB Black (SM-F900FZKD)	Android	21481200	512	2
Samsung	Galaxy Fold 5G SM-F907B 12/512GB Black	Android	19572800	512	2

### 3.3 Evaluation

To find out the accuracy of the results that have been obtained, researchers applied splitting data from datasets that have been produced before. The resulting data is broken down into testing data and training data with a ratio of 80:20 for training and testing. After that, by reapplying the K-Means Clustering method on training data and data testing then evaluating using Confusion Matrix, the following results in Table 7 are obtained.

**Table 7**  
Confusion Matrix From Data Train Respective to Data Test

		Y_PRED		
		Cluster	0	1
Y_TEST	0	42	0	0
	1	0	188	0
	2	0	0	15

Afterward, by using the existing Confusion Matrix, the researchers produced the results in Table 8.

**Table 8**  
Evaluation Results Using Confusion Matrix

	Precision	Recall	F1-score	Support
0	1.00	1.00	1.00	42
1	1.00	1.00	1.00	188
2	1.00	1.00	1.00	15

	Precision	Recall	F1-score	Support
Accuracy			1.00	245
Macro Avg	1.00	1.00	1.00	245
Weighted Avg	1.00	1.00	1.00	245

Based on the results of the evaluation using the confusion matrix, the successful clustering process has an accuracy rate of 1.00 or around 100%.

#### 4. Conclusion

After this research is successfully conducted to cluster smartphones price using K-Means Clustering, the optimal number of clusters is obtained, which is as many as three clusters, Cluster 0 amounted to 944 smartphones with a centroid value of Rp2,000,000.00, cluster 1 amounted to 70 smartphones with a centroid value of Rp18,000,000.00, and cluster 2 amounted to 210 smartphones with a centroid value of Rp9,000,000.00. The evaluation results using the confusion matrix showed that the clustering process that has been done has an accuracy rate of 100%. In that way, students can know the price group of a smartphone according to the specifications needed for the online learning process that can make them learn more optimal. This research is limited to only using one dataset that the authors get from Kaggle and exchange the price from Hryvnia Ukraina to Rupiah Indonesia using Rupiah rate exchange in April 2021. This research has a scientific impact in grouping smartphone prices using the K-Means Clustering method for smartphone consumers in Indonesia and is also expected to be useful as a reference in related research in the future. For future research, it is expected that this research can be developed by creating an interface that shows the results of visualization of smartphone price grouping that has been made so that it is easy to understand by readers. In addition, it is expected that further research can be conducted using datasets with a greater number of records in accordance with the increasing variants of smartphones released.

#### References

- [1] N. R, "PEMANFAATAN INTERNET SEBAGAI MEDIA E-LEARNING UNTUK SISWA DAN GURU SMPN 3 TARAKAN," *J. Pengabd. Masy. Borneo*, vol. 2, no. 2, pp. 21–24, 2018, doi: 10.35334/jpmb.v2i2.518.
- [2] APJII, "Laporan Survei Internet APJII 2019 – 2020," *Asos. Penyelenggara Jasa Internet Indones.*, pp. 1–146, 2020, [Online]. Available: <https://apjii.or.id/survei>.
- [3] BPS, "Hasil Sensus Penduduk 2020," *Sensus Pendud. 2020*, vol. 6, pp. 1–18, 2020.
- [4] K. Pendidikan dan Kebudayaan, "Kementerian Pendidikan dan Kebudayaan » Republik Indonesia." <https://www.kemdikbud.go.id/main/blog/2020/05/kemendikbud-terbitkan-pedoman-penyelenggaraan-belajar-dari-rumah> (accessed Apr. 09, 2021).
- [5] L. D. Herliandry, N. Nurhasanah, M. E. Suban, and H. Kuswanto, "Pembelajaran Pada Masa Pandemi Covid-19," *JTP - J. Teknol. Pendidik.*, vol. 22, no. 1, pp. 65–70, 2020, doi: 10.21009/jtp.v22i1.15286.
- [6] Y. Pusparisa, "Pandemi Covid-19, Puluhan Juta Murid Belajar di Rumah," *Katadata.Co.Id*, no. April, p. 2020, 2020, [Online]. Available: <https://databoks.katadata.co.id/datapublish/2020/05/05/pandemi-covid-19-puluhan-juta-siswa-belajar-di-rumah>
- [7] L. H. A. Mana, "RESPON GURU BAHASA INDONESIA TERHADAP PEMBELAJARAN DARING DI ERA COVID-19," *JIRA J. Inov. dan Ris. Akad.*, vol. 2, no. 1, pp. 93–106, 2021, doi: 10.47387/jira.v2i1.76.
- [8] Z. N, Nurmayanti, and H. Ferdiansyah, "Efektifitas Media Pembelajaran Daring di masa Pandemi Covid-19," *Edumaspul J. Pendidik.*, vol. 5, no. 1, pp. 71–77, 2021.
- [9] Artem Pozdniakov, "Mobile Phones Data | Kaggle," Feb. 05, 2021. <https://www.kaggle.com/artempozdniakov/ukrainian-market-mobile-phones-data> (accessed Jun. 10, 2021).
- [10] P. Arora, S. Srivastava, and B. Garg, "Mobile price prediction using weka 1," vol. 5, no. 4, pp. 330–333, 2020.
- [11] W. R. U. Fadilah, D. Agfiannisa, and Y. Azhar, "Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode Support Vector Machine," *Fountain Informatics J.*, vol. 5, no. 2, pp. 45–51, 2020, doi: 10.21111/fij.v5i2.4449.
- [12] Fatkhuroji, S. Santosa, and R. Anggi Pramunendar, "PREDIKSI HARGA KEDELAI LOKAL DAN KEDELAI IMPOR DENGAN METODE SUPPORT VECTOR MACHINE BERBASIS FORWARD SELECTION," *J. Teknol. Inf.*, vol. 15, no. 1, pp. 61–76, 2019.
- [13] P. Indrayati Sijabat, Yuhandri, G. Widi Nurcahyo, and A. Sindar, "Algoritma Backpropagation Prediksi Harga

- Komoditi terhadap Karakteristik Konsumen Produk Kopi Lokal Nasional," *J. Teknol. Inf. dan Komun. Digit. Zo.*, vol. 11, no. 1, pp. 96–107, 2020, doi: 10.31849/digitalzone.v11i1.3880.
- [14] J. Han, M. Kamber, and J. Pei, "Third Edition : Data Mining Concepts and Techniques," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2012, [Online]. Available: <http://library.books24x7.com/toc.aspx?bkid=44712>.
- [15] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- [16] M. Cahyanti, M. R. D. Septian, E. R. Swedia, and R. A. Salim, "PENGELOMPOKKAN CITRA KENDARAAN (MOTOR DAN MOBIL) BERDASARKAN BENTUK MENGGUNAKAN ALGORITMA K-MEANS," *Sebatik*, vol. 22, no. 2, pp. 153–160, 2018, doi: 10.46984/sebatik.v22i2.322.
- [17] P.-N. Tan, M. Steinbach, and V. Kumar, "Introduction to Data Mining," pp. 51–56, 2006, doi: 10.1016/0022-4405(81)90007-8.
- [18] M. C. Untoro, L. Anggraini, M. Andini, H. Retnosari, and M. A. Nasrulloh, "Penerapan metode k-means clustering data COVID-19 di Provinsi Jakarta," *Teknologi*, vol. 11, no. 2, pp. 59–68, 2021, doi: 10.26594/teknologi.v11i2.2323.
- [19] Intan Trivena Maria Daeng, N. . Mewengkang, and E. R. Kalesaran, "Penggunaan Smartphone Dalam Menunjang Aktivitas Perkuliahan Oleh Mahasiswa Fispol Unsrat Manado," *Acta Diurna*, vol. 6, no. 1, pp. 1–15, 2017.
- [20] E. S. Pandia, , N., and , N., "Pengaruh Penggunaan Smartphone Sebagai Perangkat Pembelajaran Terhadap Pendidikan Karakter Siswa," *BEST J. (Biology Educ. Sains Technol.)*, vol. 3, no. 2, pp. 44–50, 2020, doi: 10.30743/best.v3i2.2805.
- [21] E. M. - AMIK BSI Yogyakarta, "Komparasi Metode Clustering K-Means Dan K-Medoids Dengan Model Fuzzy Rfm Untuk Pengelompokan Pelanggan," *Evolusi J. Sains dan Manaj.*, vol. 6, no. 2, 2018, doi: 10.31294/evolusi.v6i2.4600.
- [22] R. Yuliani, "Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K-Means di Dusun Bagik Endep Sukamulia Timur," vol. 4, no. 1, pp. 39–50, 2021.
- [23] R. K. Dinata, S. Safwandi, N. Hasdyna, and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," *INFORMAL Informatics J.*, vol. 5, no. 1, p. 10, 2020, doi: 10.19184/isj.v5i1.17071.
- [24] Adnan, M. M. J., Hemmje, M. L., & Kaufmann, M. A. (2021). Social Media Mining to Study Social User Group by Visualizing Tweet Clusters using Word2Vec, PCA and K-Means. In BIRDS+ WEPIR@ CHIIR (pp. 40-51).
- [25] Li, Q., Lin, H., Tang, C., Wei, X., Peng, Z., Ma, X., & Chen, T. (2021). Exploring the "Double-Edged Sword" Effect of Auto-Insight Recommendation in Exploratory Data Analysis.
- [26] Kumar, N., Yadav, S. K., & Yadav, D. S. (2021). An Approach for Documents Clustering Using K-Means Algorithm. In Innovations in Information and Communication Technologies (IICT-2020) (pp. 453-460). Springer, Cham.