Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Nomor 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14170
e- ISSN : 2541-2019
p- ISSN : 2541-044X

Clustering Analysis of Cadet Profiles of Age, Recency, Frequency and Monetary Variables Using K-Means and K-Medoids Algorithms

Muhamad Nursyi[1)*], Presma Dana Scendi Sumarna[2)], Arief Wibowo[3)]
[1,2,3)] Universitas Budi Luhur
[1)]mnursyi@gmail.com, [2)]zed99zandy@gmail.com, [3)]arief.wibowo@budiluhur.ac.id

Submitted:Oct 7, 2024 | **Accepted** : Oct 25, 2024 | **Published** : Oct 30, 2024

Abstract: Banten Maritime Polytechnic is a new academic school established in 2019 so that the formulation of data management is still being sought to be suitable and optimal, there are many obstacles if the data is not managed properly, starting from the recruitment of prospective cadets in taking sailor competency training such as not optimal socialization. According to data from the 2021 Transportation Human Resource Development Agency, it explains that there are still few enthusiasts, especially at the Banten Maritime Polytechnic. The purpose of this study is to analyze the profile of cadets in taking sailor competency training using the age, recency, frequency and monetary methods in categorizing data and clustering with the k-means and k-medoids algorithms so that the data can be used for cadet services and related parties in the Banten Maritime Polytechnic database. This analysis can also be used for mapping in recruiting prospective cadets in taking sailor competency training so that they can see opportunities and optimize target markets. This research was conducted in 2023 based on the latest data on the 2022-2023 academic year cadet profile at the Banten Maritime Polytechnic. The results of this analysis data can be used for cadets who have not graduated and have graduated in finding work partners and channeling cadets to the shipping industry. So it is very important to manage and cluster cadet profile data in taking this sailor competency training. The use of the K-means and K-medoids algorithms helps in compiling data groupings that have large data. It works by looking at the number of small groups or groups whose numbers are represented by the variable K. To be able to group the existing data, the K-means algorithm runs iteratively from each existing data point to the K group that has been created. The results of the study are cadet profile grouping data that can be managed again for strategies and management formulations at the Banten Maritime Polytechnic, especially in increasing the recruitment of prospective cadets in taking sailor competency training.

Keywords: Analysis, Cluster, K-means, K-medoids, Cadet Profile

INTRODUCTION

Banten Shipping Polytechnic, or referred to as Poltekpol Banten, is an educational institution specializing in maritime studies and operates under the auspices of the Ministry of Transportation. This institution is basically oriented towards transportation services and is managed by the Transportation Human Resources Development Agency(Amala, 2023). Located on an area of approximately eighteen thousand square meters in the Mauk subdivision at the western tip of Java Island in the Banten region, specifically located in Sukadiri District, Tangerang Regency, Banten Province(Admin, 2024). According to information obtained from https://pddikti.kemdikbud.go.id/, it was opened on July 22, 2023. 816 students enrolled at the Banten Maritime Polytechnic, which consists of three departments: Marine Transportation Management with 181 students, Nautical with 370 students, Ship Engineering with 265 students, and tens of thousands of alumni who have graduated before or after this school.

Vocational education, especially the education and training division at the Banten Maritime Polytechnic, plays an important role in preparing prospective maritime professionals. As a higher education institution that focuses on the field of shipping, the Banten Maritime Polytechnic is responsible for providing the knowledge, skills, and work ethics needed by students so that they can succeed and contribute to the maritime world.(Ministry of Transportation of the Republic of Indonesia, 2019).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Nomor 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14170
e- ISSN : 2541-2019
p- ISSN : 2541-044X

In recent years, namely in 2023, the recapitulation results of transaction and training documents taken from the training and education and training division, there were 10,787 transactions, in 2022, there were 11,500 transactions, in 2021, there were 11,798 transactions, in 2020, there were 12,303 transactions, and in 2019 there were 13,031 transactions.

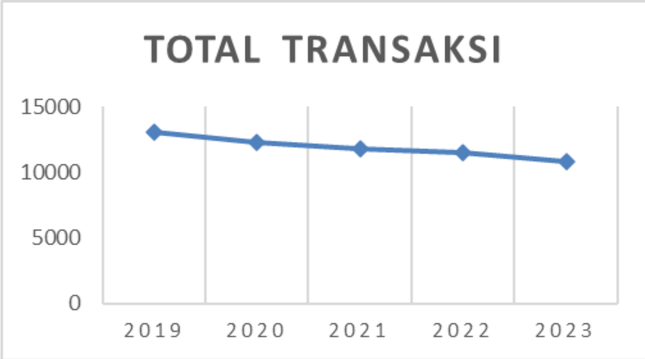

Fig. 1Total Training and Education Transactions

Fig. 1 shows a decrease in total transactions that is not significant enough. However, this decrease is not accompanied by an increase in the number of training and education participants that continues to increase.

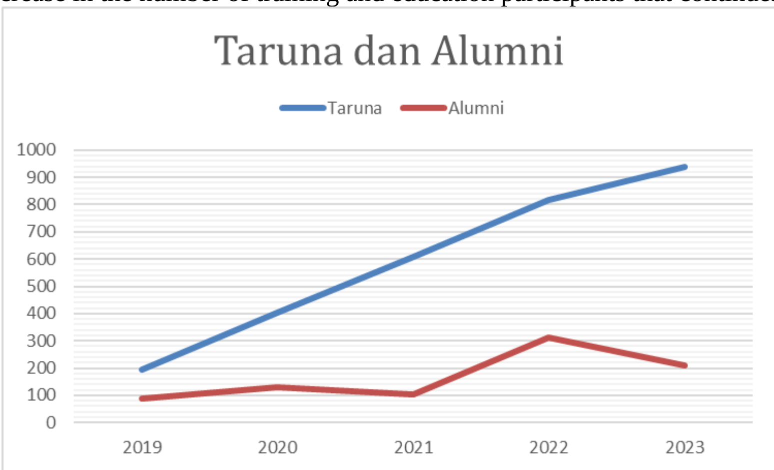

Fig. 2Training and Education Participant Graph

Fig. 1 shows a small decrease, but Fig. 2 shows an annual increase in the number of course and training participants. The above problems may hamper the ability of Banten Maritime Polytechnic to produce graduates who are ready to face the challenges of the ever-growing maritime workforce.

LITERATURE REVIEW

Data mining is the process of mining large data sets to find hidden patterns in databases. This process is known as Knowledge Discovery in Databases (KDD), which includes several stages such as data selection, cleaning, transformation, and data evaluation.(Alinafiah, Octariadi, & Sucipto, 2024). Data mining is a part of the overall KDD process, which aims to find hidden information in a database.(Jannah & Martanto, 2024). Data mining is becoming important in various sectors including education to understand patterns from large amounts of data.(Ampadu, 2023).

K-Means and K-Medoids are the two most commonly used algorithms in clustering.(Salsabila, Azise, & Ridla, 2024). K-Means uses the average to determine cluster centers.(Marini & Suhendra, 2023), while K-Medoids uses real objects as medoids or cluster centers.(Kamilah, Rohana, Rahmat, & Fauzi, 2024). The advantage of K-Medoids is the ability to handle data with outliers, which often affect the results of K-Means. K-Means is more efficient for larger datasets, but K-Medoids is more stable under some conditions.(Bumartaduri, Gusti, Syafria, Haerani, & Ramadhani, 2023). Recency, Frequency, Monetary (ARFM) Analysis is a process of analyzing the behavior of Cadets and Alumni. In determining the segmentation of Cadets and Alumni, the ARFM model is used based on three variables, namely the last recency of making a transaction, the frequency of the transaction, and the monetary value of the number of transactions for each Cadet and Alumni.(Djun, Gunadi, & Sariyasa, 2024). The age variable is obtained from the date of birth filled in by the cadet.

Research conducted by(Annas, Poerwanto, Sapriani, & S, 2022), the problem that occurs isThere is no segmentation of indicators for areas experiencing poverty gaps. The way to solve this problem is by using the

*name of corresponding author

Clustering method with the K-means Algorithm and the results obtained are the identification of two different clusters for the gap in Indonesia. The K-means algorithm is used for Clustering with 2 variables from the problem. There is no segment that shows the unevenness of the percentage of habitable houses throughout Indonesia. So that 4 clusters are the results of mapping areas with indicators of habitable houses(Septianingsih, 2022). Clustering with the K-means Algorithm is used to solve the problem of the absence of segmented poverty data to separate poverty alleviation data. The result is that there are 3 clusters of poverty indicator mapping.(Bahauddin, Fatmawati, & Sari, 2021). The Use of the K-means Algorithm in Research Conducted by(Ramadhani & Muslih, 2022)solve the problem of the absence of data on the level of segmentation of laying hens in Indonesia and the results produce 3 clusters of chicken populations, namely high, medium and low in Indonesia. The K-means algorithm is also used to solve the problem of the absence of mapping to determine the level of Covid 19 with K-means clustering. The result is that the K-means algorithm can map and cluster the spread of Covid-19 data with 3 clusters(Abdullah, Susilo, Ahmar, Rusli, & Hidayat, 2021). In the research initiated by(Sari & Sukestiyarno, 2021)using the K-means Algorithm to solve the problem of the absence of mapping to determine the level of Covid 19 with K-means clustering. The result is that the K-means Algorithm can map and cluster the spread of Covid-19 data with 3 clusters. K-means Algorithm, Hierarchical Algomerative and K-medoids Method in research(Luthfi & Wijayanto, 2021)solve the problem of the absence of human development index cluster data with the hierarchical, k-means, and k-medoids clustering methods. The results are the determination of clusters with 2 groups for Hierarchical Algorithm, and 4 groups for K-means, and 5 groups for K-medoids. Research conducted by(Rejito, Atthariq, & Abdullah, 2021), stated that Data mining analysis with rapidminer using the K-means method solves the problem There is no data for text mining with K-means clustering of tweets from Tokopedia accounts. The results obtained the most retweet frequency segmentation related to valuable quizzes, while the type of content that was least retweeted was active lifestyle. The Recency, Frequency and Monatary methods using the K-means Algorithm were used(Widiyanto & Witanti, 2021)in his research to produce Clustering of Cadet and Alumni groups called Consumers are at 226, Ordinary there are 186, Big Consumers are 299 and Top Class 25 from the problem of not grouping Taruna and Alumni PT Coversuper Indonesia Global. Based on previous research,Current research places more emphasis on transaction data in the education sector and comparisons by addingAge variable in the preprocessing process and then usingK-means and K-mendoids algorithms.

The initial hypothesis is thatIt is suspected that the segmentation of cadets produced using Age, Recency, Frequency, Monetary (ARFM) data and a combination of the K-means and K-methods algorithms along with the development of a data mining prototype will help management in determining the segmentation and levels of cadets so that they can later create strategies related to promotion, marketing, and the best service for the campus.Therefore, this study aims to analyze the profile of cadets using data from training transactions during the 2022-2023 academic year. The analysis was carried out using the method*Recency*, Frequency, Monetary(RFM) with the addition of the Age variable to find out what age range is most likely to enter education and training held by Poltek Pelayaran Banten and implemented using the K-Means and K-Medoids algorithms.

METHOD

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Nomor 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14170
e- ISSN : 2541-2019
p- ISSN : 2541-044X

This study uses an experimental quantitative research method. Primary data collection is carried out from the resultsfield observation and interviewsnamely (1) cadet profile datawho have transacted at the Banten Maritime Polytechnic, especially in the education and training section. (2) Data on ongoing businesses, namely data from cadets who have started to register, transact until they receive the education and training services they want. (3) Cadet data related to age, recency, frequency and monetary (ARFM) from 10,787 data on cadets and alumni who have participated in the education and training program. Secondary data collection was carried out by collecting data and information through secondary literature studies, namely data obtained through studying scientific articles that are related to the problems being studied.

The business process at the Banten Maritime Polytechnic includes several things in the world of education in general, starting from the registration of prospective cadets, polytechnic entrance selection, filing, learning until graduation. Before graduating from college, cadets are asked to take training on skills that support their profession. The data for this study were taken from the results of transactions carried out by the education and training division. The data is transaction data as proof of payment for cadets to take part in the training. Several field tables printed on the payment receipt are name, NRT, Training Program, Department/Semester, information for payment, nominal payment per training, total payment, in words from the total payment, deposit transaction date, bank party for approval and approval of the depositor.

The data modeling process goes through the preprocessing stage and will produce ARFM transaction model data (Recency, frequency, monetary) and RF payment training data (Recency, Frequency). Furthermore, the dataset will be used for the modeling process using the K-means and K-medoids algorithms, with the distance calculation used being the Euclidean Distance calculation, while for the evaluation process the number of recommended clusters uses the Sum of Square Error and Silhoutte Score.

RESULT

Data Selection

There are several stages carried out in this data selection process, namely (1) Data collection, (2) Data cleaning, (3) Data reduction, (4) Data aggregation.

Data collection

In this process, data from several tables related to ARFM (Age, Recency, Frequency, and Monetary) processing are collected by dividing them into several relevant parts, so that the data functions as a support for the education and training data that is combined as follows.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Nomor 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14170
e- ISSN : 2541-2019
p- ISSN : 2541-044X

Table 1. Transactions for Cadet Training and Education

NRT/Seaman Code	Date of birth	Age	Graduation year	Name	Transaction Date	Training	Transaction Amount
2115211013	2004-01-30	21	NOT YET PASSED	ANDRY IRMANSYAH ANGGAWIJAYA	2023-07-26	ATCTCO	985000
2115211013	2004-01-30	21	NOT YET PASSED	ANDRY IRMANSYAH ANGGAWIJAYA	2023-07-17	AFF	1025000
2115211013	2004-01-30	21	NOT YET PASSED	ANDRY IRMANSYAH ANGGAWIJAYA	2023-05-27	CMT	785000
2115211013	2004-01-30	21	NOT YET PASSED	ANDRY IRMANSYAH ANGGAWIJAYA	2023-01-25	GMDSS	1710000
2115211013	2004-01-30	21	NOT YET PASSED	ANDRY IRMANSYAH ANGGAWIJAYA	2022-11-13	ATOTCO	985000
2122221391	2003-04-30	22	NOT YET PASSED	A AIMAN FADIL FAD	2022-07-28	PSCRB	985000
2122221391	2003-04-30	22	NOT YET PASSED	A AIMAN FADIL FAD	2022-11-20	ECDIS	1785000
2122221391	2003-04-30	22	NOT YET PASSED	A AIMAN FADIL FAD	2022-08-21	BST-KLM	935000
2148376142	2002-03-15	23	NOT YET PASSED	A GILANG WIRA WIRAWAN	2023-04-09	PSCRB	985000
2148376142	2002-03-15	23	NOT YET PASSED	A GILANG WIRA WIRAWAN	2023-07-29	SAT	665000
2148376142	2002-03-15	23	NOT YET PASSED	A GILANG WIRA WIRAWAN	2023-05-12	SSO	835000
2148376142	2002-03-15	23	NOT YET PASSED	A GILANG WIRA WIRAWAN	2022-12-08	BRM	1185000
2148376142	2002-03-15	23	NOT YET PASSED	A GILANG WIRA WIRAWAN	2023-05-06	ERM	1185000
2108212401	2004-05-09	20	NOT YET PASSED	A WIRPIAN IRVAN	2022-06-11	US / ARPA SIMULATOR	985000
2108212401	2004-05-09	20	NOT YET PASSED	A WIRPIAN IRVAN	2022-08-10	RS / RADAR SIMULATOR	985000
2108212401	2004-05-09	20	NOT YET PASSED	A WIRPIAN IRVAN	2023-06-13	ATOTCO	985000
2108212401	2004-05-09	20	NOT YET PASSED	A WIRPIAN IRVAN	2023-04-30	PSCRB	985000

INFORMATION	STATUS	VALIDATION
FOR SEAMAN SKILLS TRAINING FEES	PRINTED	YES
FOR SEAMAN SKILLS TRAINING FEES	PRINTED	YES
FOR SEAMAN SKILLS TRAINING FEES	PRINTED	YES
FOR SEAMAN SKILLS TRAINING FEES	PRINTED	YES
FOR SEAMAN SKILLS TRAINING FEES	PRINTED	YES
FOR SEAMAN SKILLS TRAINING FEES	PRINTED	YES
FOR SEAMAN SKILLS TRAINING FEES	PRINTED	YES
FOR SEAMAN SKILLS TRAINING FEES	PRINTED	YES
FOR SEAMAN SKILLS TRAINING FEES	PRINTED	YES
FOR SEAMAN SKILLS TRAINING FEES	PRINTED	YES

Table 1 includes transaction sequence number, NRT, date of birth, name, age, year of graduation, transaction date, time, for, transaction nominal, amount, in words, total, description, status, and validation, which is a combination of cashier data with manual printed receipts for cadets or alumni.

Data Cleansing

This transaction data has been cleaned of canceled or failed transactions, so that only successful transactions are presented. This cleaning process is done in two stages: first, the data is manually cleaned by deleting unsuccessful transaction data until it is complete; second, successful but canceled transactions.

Data Reduction

Data reduction by eliminating irrelevant tables in the clustering process, as exemplified below.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Nomor 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14170
e- ISSN : 2541-2019
p- ISSN : 2541-044X

Table 2. Overall table

Transaction data column	
No_Transaction	Nominal_transaction1
Transaction_Date	For_payment1
Depositor_Name	Nominal_transaction2
NRT	For_payment2
Training Program	Nominal_transaction3
Major	For_payment3
Semester	Nominal_transaction4
Date of birth	For_payment4
Age	Nominal_transaction5
Graduation year	For_payment5
Validation	Total_nominal
Signed	Bank_approver

In the table it is stated that there is a number of data that is not needed for the cluster analysis calculation, so the data in the table is deleted and the table becomes as follows.

Table 3 Table after subtraction

Transaction data column	
No_Transaction	Nominal_transaction
Transaction_Date	For payment
Depositor_Name	Total_nominal
NRT	
Date of birth	
Age	
Graduation year	

Aggregation Data

From the data that has been prepared, an aggregation will be carried out for each Cadet and Alumni who are given an identity with a unique code called NRT, so that each cadet or alumni will have a new data dimension, calculating the date of birth by filling in the age (Age), the number of days from today to the last date of attending training (Recency), the number of transactions (Frequency), the total cost of attending training (Monetary). While for other data, adjust the names such as for payments to the name of the training table and so on so that they are easy to understand. An example of data that has been successfully aggregated can be seen in the following table.

Table 4Cluster Calculation Preparation Table

index	NRT/Seaman Code	Date of birth	Age	Graduation year	Name	Transaction Date	Training	Transaction Amount	Age
0	2115211013	01/30/04 00.00	1,985,753,424,657,530	NOT YET PASSED	ANDRY IRMANSYAH ANGGAWIJAYA	07/01/23 00.00	SAT	665000	19
1	2115211013	01/30/04 00.00	1,985,753,424,657,530	NOT YET PASSED	ANDRY IRMANSYAH ANGGAWIJAYA	07/26/23 00.00	ATCTCO	985000	19
2	2115211013	01/30/04 00.00	1,985,753,424,657,530	NOT YET PASSED	ANDRY IRMANSYAH ANGGAWIJAYA	07/17/23 00.00	AFF	1025000	19
3	2115211013	01/30/04 00.00	1,985,753,424,657,530	NOT YET PASSED	ANDRY IRMANSYAH ANGGAWIJAYA	05/27/23 00.00	CMT	785000	19
4	2115211013	01/30/04 00.00	1,985,753,424,657,530	NOT YET PASSED	ANDRY IRMANSYAH ANGGAWIJAYA	01/25/23 00.00	GMDSS	1710000	19
5	2115211013	01/30/04 00.00	1,985,753,424,657,530	NOT YET PASSED	ANDRY IRMANSYAH ANGGAWIJAYA	11/13/22 00.00	ATOTCO	985000	19
6	2122221391	04/30/03 00.00	20,610,958,904,109,500	NOT YET PASSED	A AIMAN FADIL FAD	07/28/22 00.00	PSCRB	985000	20
7	2122221391	04/30/03 00.00	20,610,958,904,109,500	NOT YET PASSED	A AIMAN FADIL FAD	11/20/22 00.00	ECDIS	1785000	20
8	2122221391	04/30/03 00.00	20,610,958,904,109,500	NOT YET PASSED	A AIMAN FADIL FAD	08/21/22 00.00	BST-KLM	935000	20
9	2148376142	03/15/02 00.00	21,736,986,301,369,800	NOT YET PASSED	A GILANG WIRA WIRAWAN	09/04/23 00.00	PSCRB	985000	21
10	2148376142	03/15/02 00.00	21,736,986,301,369,800	NOT YET PASSED	A GILANG WIRA WIRAWAN	07/29/23 00.00	SAT	665000	21
11	2148376142	03/15/02 00.00	21,736,986,301,369,800	NOT YET PASSED	A GILANG WIRA WIRAWAN	12/05/23 00.00	SSO	835000	21
12	2148376142	03/15/02 00.00	21,736,986,301,369,800	NOT YET PASSED	A GILANG WIRA WIRAWAN	12/08/22 00.00	BRM	1185000	21
13	2148376142	03/15/02 00.00	21,736,986,301,369,800	NOT YET PASSED	A GILANG WIRA WIRAWAN	06/05/23 00.00	ERM	1185000	21

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Nomor 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14170
e- ISSN : 2541-2019
p- ISSN : 2541-044X

14	2108212401	09/05/04 00.00	19,583,561,643,835,600	NOT YET PASSED	A WIRPIAN IRVAN	11/06/22 00.00	US / ARPA SIMULATOR	985000	19
15	2108212401	09/05/04 00.00	19,583,561,643,835,600	NOT YET PASSED	A WIRPIAN IRVAN	10/08/22 00.00	RS / RADAR SIMULATOR	985000	19
16	2108212401	09/05/04 00.00	19,583,561,643,835,600	NOT YET PASSED	A WIRPIAN IRVAN	06/13/23 00.00	ATOTCO	985000	19
17	2108212401	09/05/04 00.00	19,583,561,643,835,600	NOT YET PASSED	A WIRPIAN IRVAN	04/30/23 00.00	PSCRB	985000	19
18	2146933415	11/21/01 00.00	2,204,931,506,849,310	NOT YET PASSED	A.M. IRFAN	08/07/22 00.00	CMT	785000	22
0	2115211013	01/30/04 00.00	1,985,753,424,657,530	NOT YET PASSED	ANDRY IRMANSYAH ANGGAWIJAYA	07/01/23 00.00	SAT	665000	19

Data Transformation (Preprocessing)

At this stage, the data is prepared to be processed in the clustering process, including determining the ARFM for each data, followed by discretisation, which is the process of giving weight to each data feature that continues until it becomes categorical discrete data. Thus, a dataset of cadets and alumni is obtained whose arithmetic functions can be calculated precisely.

Changes to ARFM

The selection process is carried out based on the variables age, recency, frequency, monetary from the data obtained. The age variable is obtained from the date of birth filled in, the recency variable is obtained from the research date minus the last transaction date, the frequency variable is obtained from the number of transactions, and the monetary variable is obtained from the total transactions made by the cadet or alumni. The following is a sample ARFM table showing some examples of data that have become ARFM bound by NRT.

Table 5 ARFM Table

No	NRT/Seaman Code	Age	Recency	Frequency	Monetary
0	2001825302	20	12	5	4165000
1	2002405484	22	141	5	4735000
2	2002626173	21	57	3	2955000
3	2002627960	21	14	5	5315000
4	2003193739	21	109	4	3980000
5	2003279470	19	89	5	4625000
6	2003652256	21	9	6	5380000
7	2003679915	22	93	5	4705000
8	2003770058	19	2	4	3260000
9	2003929386	21	74	4	4450000
10	2004170573	21	45	6	6545000
11	2004196596	20	51	6	5700000
12	2004322996	22	126	6	5460000
13	2004552054	21	33	3	2955000
14	2004648796	20	108	3	3955000
15	2004899101	19	153	5	4575000
16	2004971698	20	151	3	3680000
17	2005188598	21	45	5	5380000
18	2006010753	20	161	4	3420000
19	2006532723	20	44	4	3540000
20	2006637664	20	11	5	4885000
21	2007053774	21	17	6	6385000
22	2007125991	19	61	6	5660000

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Nomor 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14170
e- ISSN : 2541-2019
p- ISSN : 2541-044X

Modeling

All data that has gone through the pre-processing process will be processed with the K-Means or K-Medoids algorithm using the SkLearn python library. The first stage in the K-Means or K-Medoids method is to determine the K value which will be the number of segmentations or clusters to be formed, the K value can come from company considerations or trial and error results, the K value will also mean the number of centroids or midpoints of each cluster, derived from the existing centroid points then all data will be calculated the distance for each centroid. The distance calculation in this study uses the Euclidean Distance calculation, in this first process each data will produce a distance value for each centroid, then the data will be categorized into the centroid cluster if the distance value is the smallest compared to the other centroid values. After all the data has obtained its cluster position, the process will continue by appointing a new centroid by averaging the entire data column (data features) of each cluster, the average value will then become the new centroid, and this process will continue to repeat itself such as calculating the distance of all data to each centroid, then concluding which cluster the data is in by choosing the smallest distance between the data and the centroid. Until the calculation for the new centroid from the average data results in each cluster, this process will stop running when the new data centroid no longer changes, which indicates that the centroid has reached the most optimum distance for the k value that was determined at the beginning. To get the most optimal number of segmentations or K values based on the characteristics of the data that has been prepared can be done in several ways, in this study each possible K value will be calculated for its segmentation quality based on the optimum distance and the number of segments that are not too large, the process of determining this K value will be discussed during the evaluation process.

At the learning or lecture stage before graduating, cadets take part in training and education and training carried out by the campus according to their wishes and interests in the chosen skills. The training can also be attended by prospective participants from outside as long as they meet the provisions and requirements. After taking part in several trainings and completing lectures, cadets are allowed to practice in the field either at sea or on land according to the policies of the department. Then after that, cadets prepare reports and final assignments to be tried as graduation requirements.

Algorithm Implementation

Implementation processusing stages that are directly implemented from data into a prototype using the help of the Python programming language on a Google Cloud-based engine computer.

K-means algorithm

There are several steps in implementing using this algorithm, namely:

Initiating a Cluster

OnFig. 4 explainsThe code related to the cluster calculation prototype by initializing and defining the number of clusters to be entered into the system, so that the system will process it to continue the calculation process.

```
class KMeansClustering:
    def __init__(self, X, num_clusters):
        self.K = num_clusters # jumlah cluster (bisa input sendiri)
        self.max_iterations = 100 # batas merubah centroid sampai iterasi ke 100 setelah itu berhenti
        self.num_examples, self.num_features = X.shape # jumlah fitur (ARFM -- > 4 fitur)
        self.plot_figure = True # plot figure
```

Fig. 4K-Mean Cluster Initiation

Determining the Centroid

Explanation of the code related to the centroid distance prototype by initializing and defining the centroid points that will be entered into the system, so that the system will process it to continue the next calculation process.

```
# randomly initialize centroids
def initialize_random_centroids(self, X):
    centroids = np.zeros((self.K, self.num_features)) # row , column full with zero
    for k in range(self.K): # iterations of
        centroid = X[np.random.choice(range(self.num_examples))] # random centroids
        centroids[k] = centroid
    return centroids # return random centroids
```

Fig. 5Determining the K-mean centroid

Creating a Cluster

Create a new cluster by calculating the distance of the centroid that will be entered into the system, so that the system will process it to continue the next calculation process.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Nomor 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14170
e- ISSN : 2541-2019
p- ISSN : 2541-044X

```
def create_cluster(self, X, centroids):
    clusters = [[] for _ in range(self.K)] # set tempat untuk tiap cluster --> [[], [], []]  <--- contoh 3 cluster
    for point_idx, point in enumerate(X):
        closest_centroid = np.argmin(
            np.sqrt(np.sum((point-centroids)**2, axis=1)) # cari jarak euclid seperti euclideanDistance(x, y) di kmedoids
        ) # closest centroid using euler distance equation(calculate distance of every point from centroid)
        clusters[closest_centroid].append(point_idx)
    return clusters
```

Fig. 6Creating a Cluster

Calculating New Centroid

Explanation of the code related to the centroid distance prototype by repeating the centroid points that will be entered into the system, so that the system will process it to continue the next calculation process.

```
def calculate_new_centroids(self, cluster, X):
    centroids = np.zeros((self.K, self.num_features)) # row , column full with zero
    for idx, cluster in enumerate(cluster):
        new_centroid = np.mean(X[cluster], axis=0) # find the value for new centroids
        centroids[idx] = new_centroid
    return centroids
```

Fig. 7Calculating New Centroid

Final clustering

```
def predict_cluster(self, clusters, X):
    y_pred = np.zeros(self.num_examples) # row1 fillup with zero
    for cluster_idx, cluster in enumerate(clusters):
        for sample_idx in cluster:
            y_pred[sample_idx] = cluster_idx
    return y_pred
```

Fig. 8Calculating the final clustering

K-medoids algorithm

Following are the steps for implementing this algorithm

Determining the Eccludian distance

Explains the code related to the ecludian distance prototype by determining the x and y points entered into the system, so that the system will process it to continue the next calculation process.

```
def euclideanDistance(x, y):
    '''
    Euclidean distance between x, y
    --------
    Return
    d: float
    '''
    squared_d = 0
    for i in range(len(x)):
        squared_d += (x[i] - y[i])**2
    d = np.sqrt(squared_d)
    return d
```

Fig. 9Determining the Eccludian distance

Determining Clusters

Explains the code related to the cluster calculation prototype by initializing and defining the number of clusters to be entered into the system, so that the system will process it to continue the calculation process.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Nomor 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14170
e- ISSN : 2541-2019
p- ISSN : 2541-044X

```
def __init__(self, k = 2, max_iter = 300, has_converged = False):
    '''
    Class constructor
    Parameters
    ----------
    - k: number of clusters.
    - max_iter: number of times centroids will move
    - has_converged: to check if the algorithm stop or not
    '''
    # inisialisasi nilai parameter
    self.k = k # jumlah cluster
    self.max_iter = max_iter # iterasi maksimal (kalau has_converged=True iterasi berhenti batasnya max_iter)
    self.has_converged = has_converged # konvergensi
    self.medoids_cost = [] # semacam rugi-rugi
```

Fig. 10Determining Clusters

Initiating index value

Explains the code related to the index value calculation prototype by initializing and defining the number of clusters to be entered into the system, so that the system will process it to continue the calculation process.

```
def initMedoids(self, X): # inisiasi
    '''
    Parameters
    ----------
    X: input data.
    '''
    self.medoids = [] # set awal medoid kosong (tidak ada nilai)

    #Starting medoids will be random members from data set X
    indexes = np.random.randint(0, len(X)-1,self.k) # inisiasi nilai index tabel random
    self.medoids = X[indexes] # inisiasi medoid dengan nilai index baris (sebanyak jumlah cluster) random dari 10788 baris

    for i in range(0,self.k):
        self.medoids_cost.append(0) # inisiasi rugi-rugi
```

Fig. 11Initiating index value

Calculating convergence

Explains the code related to the cluster calculation prototype by initializing and defining the number of clusters to be entered into the system, so that the system will process it to continue the calculation process.

```
def isConverged(self, new_medoids): # menilai konvergensi (ketika nilai medoids sudah tidak berubah ) ---> ini nanti yang menentukan
has_converge itu False atau True
        '''
        Parameters
        ----------
        new_medoids: the recently calculated medoids to be compared with the current medoids stored in the class
        '''
        return set([tuple(x) for x in self.medoids]) == set([tuple(x) for x in new_medoids])

    def updateMedoids(self, X, labels): # ganti ganti medoid
        '''
        Parameters
        ----------
        labels: a list contains labels of data points
        '''
        self.has_converged = True # ketika nilai medoid sudah tidak berubah maka has_converged berubah menjadi True

        #Store data points to the current cluster they belong to
        clusters = []
        for i in range(0,self.k):
            cluster = []   # tiga cluster. --->. [cluster 1:[tabel ARFM], cluster2: [tabel ARFM], cluster3:[tabel ARFM]]
            for j in range(len(X)):
                if (labels[j] == i):
                    cluster.append(X[j])
            clusters.append(cluster)
```

Fig. 12Calculating convergence

Counting New Medoids

Explains the code related to the new medoids calculation prototype by calculating the old medoids that will be entered into the system, so that the system will process it to continue the calculation process.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Nomor 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14170
e- ISSN : 2541-2019
p- ISSN : 2541-044X

```
#Calculate the new medoids
    new_medoids = [] # bikin medoid baru dari medoid awal --> self.medoids = X[indexes]
    for i in range(0, self.k):
        new_medoid = self.medoids[i] # update medoid baru
        old_medoids_cost = self.medoids_cost[i] # update medoid cost
        print(f"medoids {i}: {new_medoid}")
        for j in range(len(clusters[i])): # untuk menghitung nilai medoid baru dan cost baru

            #Cost of the current data points to be compared with the current optimal cost
            cur_medoids_cost = 0
            for dpoint_index in range(len(clusters[i])):
                cur_medoids_cost += euclideanDistance(clusters[i][j], clusters[i][dpoint_index]) # cari jarak euclid total

            #If current cost is less than current optimal cost,
            #make th c.e current data point new medoid of the cluster
            if cur_medoids_cost < old_medoids_cost: # kalo misalkan jarak totalnya lebih kecil dari jarak total lama (jarak total
            jadi jarak baru)
                new_medoid = clusters[i][j]
                old_medoids_cost = cur_medoids_cost

        #Now we have the optimal medoid of the current cluster
        new_medoids.append(new_medoid)
    print(f"medoids cost: {cur_medoids_cost}")

    #If not converged yet, accept the new medoids
    if not self.isConverged(new_medoids): # check apakah medoid masih berubah, kalau iya lanjut
        self.medoids = new_medoids
        self.has_converged = False
```

Fig. 13Counting New Medoids

Cluster final result calculation

explains the code related to the final cluster calculation prototype by defining the ecludian and medoids results that will be entered into the system, so that the system will process it to continue the calculation process.

```
def predict(self,data): # fungsi untuk prediksi cluster
    '''
    Parameters
    ----------
    data: input data.

    Returns:
    ----------
    pred: list cluster indexes of input data
    '''

    pred = []
    for i in range(len(data)):
        #Distances from a data point to each of the medoids
        d_list = []
        for j in range(len(self.medoids)):
            d_list.append(euclideanDistance(self.medoids[j],data[i])) # nyari jarak dari medoid

        pred.append(d_list.index(min(d_list))) # cari jarak minimum masukknya ke cluster mana

    return np.array(pred) # hasil keluaran berupa pembagian cluster tiap baris data ARFM
```

Fig. 14Cluster final result calculation

Evaluation of Results

The quality of a cluster or segmentation can be measured by looking at the proximity between the data in one cluster (cohesion) and also the distance between the data of one cluster and another cluster (separation). In this study, the quality of the cluster distance is measured by performing calculations. All clusters will be calculated and then combined so that each K value will have its own value. The K value that will be selected is the one with the smallest value, but not only that because it is necessary to find a balance point so that the number of segmentations or K values formed is not too large and by looking at the Davies Bouldin score.

Comparison of Results

Age, Recency, Frequency, Monetary Diagram Results

The number of age range applicants for the Banten Maritime Polytechnic training, it can be seen that the age group between 19-24 years is the largest, who are cadets who are domiciled and carrying out educational activities, while those over 25 years of age are alumni.

*name of corresponding author

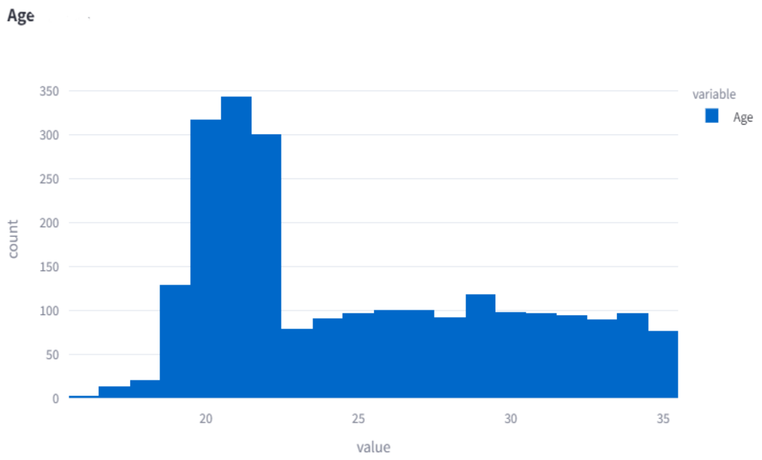

Fig. 15Age Result Diagram

The last distance of the registration time for the Banten Maritime Polytechnic training, can be seen in the table, indicating that when this data was taken, there were many registrations for the Banten Maritime Polytechnic training.

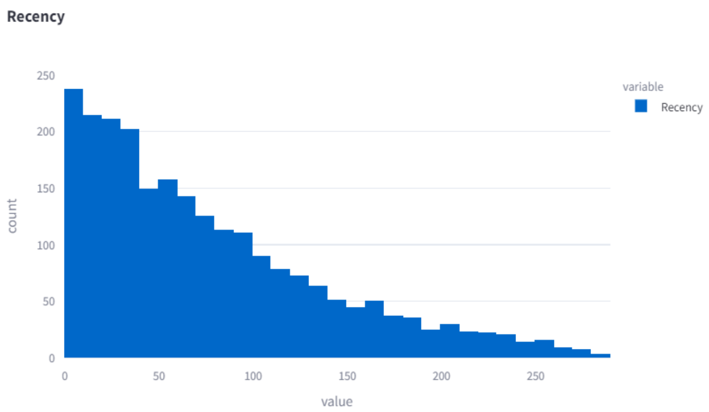

Fig. 16 Recency Result Diagram

The large number of applicants for the Banten Maritime Polytechnic training shows that registration has a relatively stable frequency.

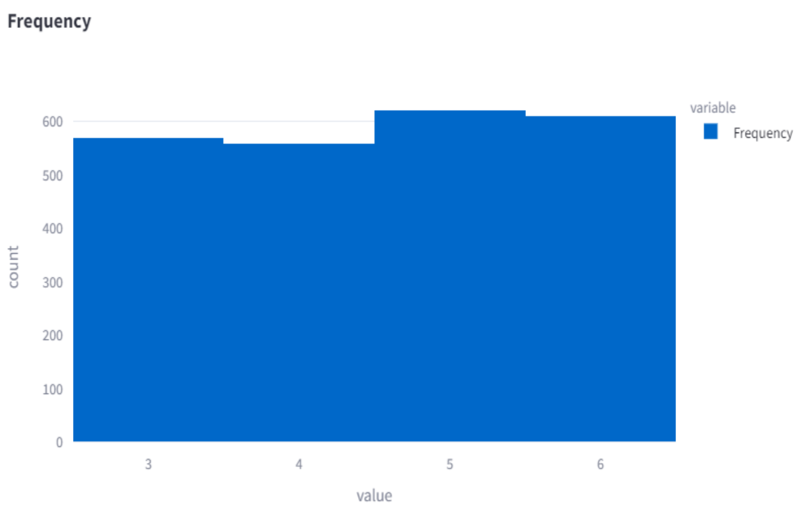

Fig. 17Frequency Result Diagram

*name of corresponding author

e- ISSN : 2541-2019
p- ISSN : 2541-044X

The amount of money spent on registration for the Banten Maritime Polytechnic training, on average, appears to have expenses that fluctuate between 2.5 million and 6.5 million.

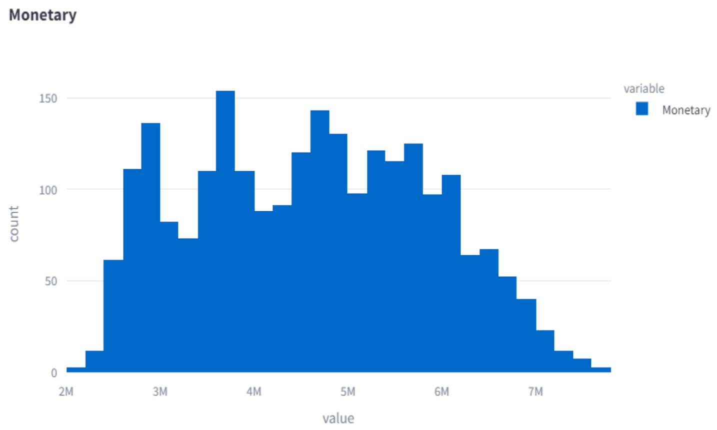

Fig. 18Monetary Yield Diagram

Clustering results table using the K-means method and Davies Bouldin score K-means graph

In Table 6, the sample results of K-means clustering are shown as 9 samples from 10787 data with 3 cluster results 0,1, and 2.

Table 6
Clustering Results with the K-Means Method

No	NRT/Seaman Code	Age	Recency	Frequency	Monetary	cluster
0	2001825302	20.00	12	5	4165000	00.00
1	2002405484	22.00	141	5	4735000	00.00
2	2002626173	21.00	57	3	2955000	02.00
3	2002627960	21.00	14	5	5315000	00.00
4	2003193739	21.00	109	4	3980000	00.00
5	2003279470	19.00	89	5	4625000	00.00
6	2003652256	21.00	9	6	5380000	00.00
7	2003679915	22.00	93	5	4705000	00.00
8	2003770058	19.00	2	4	3260000	02.00
9	2003929386	21.00	74	4	4450000	00.00

Validation of the calculated clusters shows the suitability of how many clusters will be used as shown inFig. 19

Fig. 19 Davies bouldin score graph results

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Nomor 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14170
e- ISSN : 2541-2019
p- ISSN : 2541-044X

Comparison of K-mean cluster results with age, recency, frequency, monetary data

The comparison between age in cluster 0, cluster 1 and cluster 2 with close differences is shown in table 7.

Table 7. Comparison of age results in the K-means method

index	Age Cluster 0	Age Cluster 1	Age Cluster 2
mean	24	25	24
min	17	16	16
max	35	35	35

Comparison between recency in cluster 0, cluster 1 and cluster 2 with significant differences is shown in table 8.

Table 8 Comparison of cluster recency results in the K-means method

index	Recency Cluster 0	Recency-Cluster 1	Recency-Cluster 2
mean	73	59	94
min	0	0	0
max	274	263	282

The comparison between frequencies in cluster 0, cluster 1 and cluster 2 with close differences is shown in table 9.

Table 9. Comparison of cluster frequency results in the K-means method

index	Frequency Cluster 0	Frequency Cluster 1	Frequency Cluster 2
mean	4	5	3
min	3	4	3
max	6	6	4

The comparison between monetary in cluster 0, cluster 1 and cluster 2 with significant differences is shown in table 10.

Table 10. *Comparison of monetary cluster results on the K-means method*

index	Monetary Cluster 0	Monetary Cluster 1	Monetary Cluster 2
mean	4681970	6117247	3197518
min	3940000	5400000	2095000
max	5390000	7675000	3930000

Clustering results table using the K-medoids method and Davies Bouldin score K-medoids graph

In Table 11, the sample results of the K-medoids cluster are shown as 19 samples from 10787 data with 3 cluster results 0,1, and 2.

Table 11 *Cluster results with the K-medoids method*

No	NRT/Seaman Code	Age	Recency	Frequency	Monetary	cluster
0	2001825302	20.00	12	5	4165000	1
1	2002405484	22.00	141	5	4735000	1
2	2002626173	21.00	57	3	2955000	0
3	2002627960	21.00	14	5	5315000	1
4	2003193739	21.00	109	4	3980000	1
5	2003279470	19.00	89	5	4625000	1
6	2003652256	21.00	9	6	5380000	2
7	2003679915	22.00	93	5	4705000	1

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Nomor 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14170
e- ISSN : 2541-2019
p- ISSN : 2541-044X

8	2003770058	19.00	2	4	3260000	0
9	2003929386	21.00	74	4	4450000	1
10	2004170573	21.00	45	6	6545000	2
11	2004196596	20.00	51	6	5700000	2
12	2004322996	22.00	126	6	5460000	2
13	2004552054	21.00	33	3	2955000	0
14	2004648796	20.00	108	3	3955000	1
15	2004899101	19.00	153	5	4575000	1
16	2004971698	20.00	151	3	3680000	0
17	2005188598	21.00	45	5	5380000	2
18	2006010753	20.00	161	4	3420000	0
19	2006532723	20.00	44	4	3540000	0

Validation of the calculated clusters shows the suitability of how many clusters will be used as shown inFig. 20

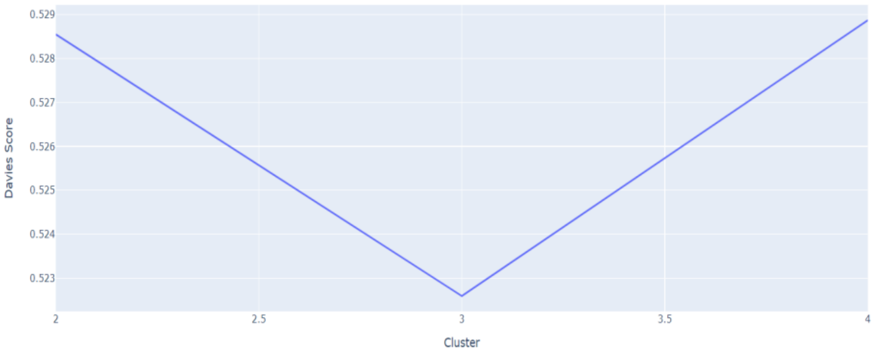

Fig. 20Davies Bouldin Score K-medoids graphic results

Comparison of K-medoids cluster results with age, recency, frequency, monetary data
The comparison between age in cluster 0, cluster 1 and cluster 2 with close differences is shown in table 12.

Table 12. K-medoids cluster results with age data

index	Age Cluster 0	Age Cluster 1	Age Cluster 2
mean	24	25	25
min	16	17	16
max	35	35	35

Comparison between recency in cluster 0, cluster 1 and cluster 2 with significant differences in table 13.

Table 13 K-medoids cluster results with recency data

index	Recency Cluster 0	Recency-Cluster 1	Recency-Cluster 2
mean	94	73	58
min	0	0	0
max	282	274	263

Comparison between frequencies in cluster 0, cluster 1 and cluster 2 with close differences in table 14.

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Nomor 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14170
e- ISSN : 2541-2019
p- ISSN : 2541-044X

Table 14. K-medoids cluster results with frequency data

index	Frequency Cluster 0	Frequency Cluster 1	Frequency Cluster 2
mean	3	4	5
min	3	3	4
max	4	6	6

Comparison between monetary in cluster 0, cluster 1 and cluster 2 with significant differences in table 15.

Table 15 Comparison of K-medoids cluster results with monetary data

index	Monetary Cluster 0	Monetary Cluster 1	Monetary Cluster 2
mean	3214810	4685765	6100863
min	2095000	3955000	5360000
max	3950000	5355000	7675000

Interpretation of Research Results

Comparison table of age, recency, frequency, monetary K-means and K-medoids

Comparison of K-means method with K-medoids between age in cluster 0, cluster 1 and cluster 2 with close differences in table 16.

Table 16. Comparison table of K-means and K-medoids age

index	K-Means Cluster 0	K-Medoids Cluster 0	K-Means Cluster 1	K-Medoids Cluster 1	K-Means Cluster 2	K-Medoids Cluster 2
mean	24	24	25	25	24	25
min	17	16	16	17	16	16
max	35	35	35	35	35	35

Comparison of K-means method with K-medoids between recency in cluster 0, cluster 1 and cluster 2 with close differences in table 17.

Table 17 Comparison table of K-means and K-medoids recency

index	K-Means Cluster 0	K-Medoids Cluster 0	K-Means Cluster 1	K-Medoids Cluster 1	K-Means Cluster 2	K-Medoids Cluster 2
mean	73	94	59	73	94	58
min	0	0	0	0	0	0
max	247	282	293	274	284	263

Comparison of the K-means method with K-medoids between frequencies in cluster 0, cluster 1 and cluster 2 with close differences in table 18.

Table 18 Comparison table frequency K-means and K-medoids

index	K-Means Cluster 0	K-Medoids Cluster 0	K-Means Cluster 1	K-Medoids Cluster 1	K-Means Cluster 2	K-Medoids Cluster 2
mean	4	3	5	4	3	5
min	3	3	4	3	3	4
max	6	4	6	6	4	6

Comparison of K-means method with K-medoids between monetary in cluster 0, cluster 1 and cluster 2 with close differences in table 19.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Nomor 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14170
e- ISSN : 2541-2019
p- ISSN : 2541-044X

Table 19
Comparison table*monetary*K-means and K-medoids

index	K-Means	K-Medoids	K-Means	K-Medoids	K-Means	K-Medoids
	Cluster 0	**Cluster 0**	**Cluster 1**	**Cluster 1**	**Cluster 2**	**Cluster 2**
mean	4681970	3214810	6117247	4685765	3197518	6100863
min	3940000	2095000	5400000	3955000	2095000	5360000
max	5390000	3950000	7675000	5355000	3930000	7675000

DISCUSSION

Clusterization Results

Table 21shows that cluster 0 is the most active and loyal cluster in participating in training and spending the most, while cluster 1 is a cluster that participates in training based on obligations with medium spending and cluster 2 is a cluster that participates in the least training and education programs among other clusters. So that the results of Clusterization using ARFM data and a combination of K-means and K-medoids algorithms are expected to help management in determining the best promotion, marketing, and service strategies for the campus.

Table 20 Clustering results table

Alias	Cadet/alumni type	Character	Strategy suggestions
cluster-0	platinum	Training participants in this cluster are more active and diligent in participating in training and education at the Banten Maritime Polytechnic.	Provide the latest information about the latest training, and monitor the success and benefits after attending training and coaching to provide testimonials.
cluster-1	gold	Training participants in this cluster are active in participating in mandatory training and education at the Banten Maritime Polytechnic.	Providing the latest information on the latest training and promotions of education and training programs.
cluster-2	silver	Training participants in this cluster have not been active in participating in mandatory training and education at the Banten Maritime Polytechnic.	Providing the latest information on campus policies related to education and training programs and displaying promotions and benefits of participating in education and training.

CONCLUSION

From the results of the study entitled clustering analysis of cadet profiles with age, recency, frequency and monetary methods using the k-means and k-medoids algorithms using 10787 data on cadets and alumni who participated in the education and training program. shows that cluster 0 is the most active and loyal cluster in participating in training and the most expenditure, while cluster 1 is a cluster that participates in training based on obligations with medium expenditure and cluster 2 is a cluster that participates in the least education and training programs among other clusters.

It is expected that the cadet segmentation produced using ARFM data and a combination of the K-means and K-methods algorithms along with the development of a data mining prototype will help management in determining the best promotion, marketing, and service strategies for the campus.

REFERENCES

Abdullah, D., Susilo, S., Ahmar, A. S., Rusli, R., & Hidayat, R. (2021). The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data. *Quality & Quantity*, *56*(3), 1283–1291. https://doi.org/10.1007/s11135-021-01176-w

Admin, A. (2024, September 21). Sejarah—Politeknik Pelayaran Banten. Retrieved September 27, 2024, from https://poltekpel-banten.ac.id/sejarah/

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Nomor 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14170
e- ISSN : 2541-2019
p- ISSN : 2541-044X

Alinafiah, A. M., Octariadi, B. C., & Sucipto, S. (2024). Impementasi Data Mining Dalam Pengelolaan Stok Obat Menggunakan Metode K-Means Clustering dan Asosociation Rules Apriori: Analisis Pola Pembelian dan Hubungan Antar Obat dalam Pengelolaan Stok menggunakan K-Means Clustering dan Association Rules Apriori. *Jurnal Informatika Polinema*, *10*(4), 551–558. https://doi.org/10.33795/jip.v10i4.5523

Amala, R. N. (2023). Pengaruh Budaya Organisasi, Kompetensi, dan Motivasi Berprestasi Terhadap Kinerja Dosen (Studi Kasus pada Politeknik Pelayaran Banten). *Innovative: Journal Of Social Science Research*, *3*(5), 4397–4410. https://doi.org/10.31004/innovative.v3i5.5398

Ampadu, Y. B. (2023). Handling Big Data in Education: A Review of Educational Data Mining Techniques for Specific Educational Problems. *Https://Www.Intechopen.Com/Journals/1/Articles/185*. https://doi.org/10.5772/acrt.17

Annas, S., Poerwanto, B., Sapriani, S., & S, M. F. (2022). Implementation of K-Means Clustering on Poverty Indicators in Indonesia. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *21*(2), 257–266. https://doi.org/10.30812/matrik.v21i2.1289

Bahauddin, A., Fatmawati, A., & Sari, F. P. (2021). Analisis Clustering Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means. *Jurnal Manajemen Informatika Dan Sistem Informasi*, *4*(1), 1–8. https://doi.org/10.36595/misi.v4i1.216

Bumartaduri, S., Gusti, S. K., Syafria, F., Haerani, E., & Ramadhani, S. (2023). Penerapan Metode Clustering Dalam Pengelompokan Kasus Perceraian Pada Pengadilan Agama di Kota Pekanbaru Menggunakan Algoritma K-Medoids. *JURIKOM (Jurnal Riset Komputer)*, *10*(1), 257–265. https://doi.org/10.30865/jurikom.v10i1.5560

Djun, S. F., Gunadi, I. G. A., & Sariyasa, S. (2024). Analisis Segmentasi Pelanggan pada Bisnis dengan Menggunakan Metode K-Means Clustering pada Model Data RFM. *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, *5*(4), 354–364. https://doi.org/10.35746/jtim.v5i4.434

Jannah, E. R., & Martanto. (2024). Klasterisasi Data Penduduk Berdasarkan Pekerjaan Menggunakan Metode K-Means Pada Wilayah Jawa Barat. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(2), 1709–1716. https://doi.org/10.36040/jati.v8i2.9055

Kamilah, N. A., Rohana, T., Rahmat, R., & Fauzi, A. (2024). Implementasi Algoritma K-Means dan K-Medoids Dalam Klasterisasi Kasus Kekerasan Terhadap Perempuan. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *8*(2), 810–819. https://doi.org/10.30865/mib.v8i2.7558

Kementerian Perhubungan RI. *Peraturan Menteri Perhubungan Republik Indonesia Nomor PM 25 Tahun 2019 Tentang Organisasi Dan Tata Kerja Politeknik Pelayaran Banten*. , PM 25 Tahun 2019 § (2019).

Luthfi, E., & Wijayanto, A. W. (2021). Analisis perbandingan metode hirearchical, k-means, dan k-medoids clustering dalam pengelompokkan indeks pembangunan manusia Indonesia. *INOVASI*, *17*, 761–773. https://doi.org/10.30872/jinv.v17i4.10106

Marini, L. F., & Suhendra, C. D. (2023). Penggunaan Algoritma K-Means Pada Aplikasi Pemetaaan Klaster Daerah Pariwisata. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *7*(2), 707–713. https://doi.org/10.30865/mib.v7i2.5558

Ramadanti, E., & Muslih, M. (2022). Penerapan Data Mining Algoritma K-Means Clustering pada Populasi Ayam Petelur di Indonesia. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, *7*(1), 1–7. https://doi.org/10.36341/rabit.v7i1.2155

Rejito, J., Atthariq, A., & Abdullah, A. S. (2021). Application of text mining employing k-means algorithms for clustering tweets of Tokopedia. *Journal of Physics: Conference Series*, *1722*(1), 012019. https://doi.org/10.1088/1742-6596/1722/1/012019

Salsabila, F., Azise, N., & Ridla, M. A. (2024). Perbandingan Kinerja Algoritma Clustering K-Means dan K-Medoids pada Popularitas Line Webtoon. *Prosiding Seminar Nasional Sains Dan Teknologi "SainTek,"* *1*(2), 411–416.

Sari, D. N. P., & Sukestiyarno, Y. (2021). Analisis Cluster Dengan Metode K-Means Pada Persebaran Kasus COVID-19 Berdasarkan Provinsi di Indonesia. *PRISMA, Prosiding Seminar Nasional Matematika*, *4*, 602–610.

Septianingsih, A. (2022). Analisis K-Means Clustering pada Pemetaan Provinsi Indonesia Berdasarkan Indikator Rumah Layak Huni. *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, *3*(1), 224–241. https://doi.org/10.46306/lb.v3i1.116

Widiyanto, A. T., & Witanti, A. (2021). Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global). *KONSTELASI: Konvergensi Teknologi Dan Sistem Informasi*, *1*(1), 204–215. https://doi.org/10.24002/konstelasi.v1i1.4293

*name of corresponding author