Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 295−307
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3630

K-Means and AHC Methods for Classifying Crime Victims by Indonesian Provinces: A Comparative Analysis

Ridha Maya Faza Lubis[1*], Jen-Peng Huang[2], Pai-Chou Wang[3], Nurafni Damanik[4], Ade Clinton Sitepu[5], Ceria D. Simanullang[6]

[1,6]Department of Business and Management, Southern Taiwan University of Science and Technology, Taiwan
[2,3]Department of Information Management, Southern Taiwan University of Science and Technology, Taiwan
[4,5]International Electrical Engineering and Computer Science, National Taipei University of Technology, Taiwan
Email: [1]db01g208@stust.edu.tw, [2]jehuang@stust.edu.tw, [3]pwang@stust.edu.tw, [4]t111999407@ntut.org.tw, [5]t110999406@ntut.org.tw, [6]db01g210@stust.edu.tw
Correspondence Author Email: db01g208@stust.edu.tw
Submitted: **13/06/2023**; Accepted: **29/06/2023**; Published**: 29/06/2023**

Abstract−Crime is a common phenomenon that often occurs in society and has a negative impact both individually and collectively. Gaining a deeper understanding of crime can help us tackle the problem more efficiently. In an era that is increasingly complex and globally connected as it is now, crime has undergone significant developments and changes. Crime remains a serious threat to our security, integrity, and well-being. Some common types of crime include theft, robbery, fraud, physical abuse, and murder. Crime can happen anytime and anywhere. To tackle crime, data mining techniques can be used to analyze the surrounding situation and gain new knowledge. One approach is to classify provinces based on crime data from previous years so that crime-prone areas can be identified and security measures can be increased. In this study, two grouping methods were used, namely K-Means and AHC using the complete linkage mode. There are 34 provinces in Indonesia which are grouped based on the number of victims of crime from 2019 to 2021. The grouping results using the K-Means method yield two groups with 17 provinces each. However, using the AHC complete linkage method, there is a difference in the number of provinces between cluster 0 and cluster 1 compared to the K-Means results. In addition, there are differences in the location of the province in the cluster between the two methods. In the K-Means method, provincial data is located in cluster 0, while in the AHC method, the province's data is in cluster 1. Thus, this study provides insight into crime in Indonesia and provides information about the grouping of provinces based on crime rates using the K-Means method. Means and AHC.

Keywords: Data Mining; Clustering; K-Means Method; AHC Method, Crime

1. INTRODUCTION

Everyone in society is aware that crime exists and frequently has a detrimental influence on individuals and society. Developing a greater understanding of crime can help us solve the issue more successfully. Crime has significantly changed and evolved in the modern era of an increasingly complex and internationally interconnected society. Crime continues to pose a major threat to our safety, integrity, and well-being in both its conventional and modern forms. Theft, robbery, and fraud are among the crimes kinds that we frequently come across. These crimes can financially hurt us and even jeopardize our safety through physical violence and murder.

Residents of the Kepayang Village road, Peninjauan District, Ogan Komering Ulu Regency, South Sumatra, penalized Rendi Novriansyah, age 22, for engaging in an immoral act against a woman who was driving on Sunday, April 6, 2023, about 17:50 WIB. The victim and the suspect were traveling together on a village road when the suspect abruptly rode up on a motorcycle and confronted the victim. Both of them were traveling in the same way when the suspect abruptly pinched the victim's breasts with his left hand as their motorcycles drew near to one another. The perpetrator panicked and attempted to flee as the startled victim yelled instantly. The victim, however, refused to remain silent and pursued the culprit while pleading for assistance. As a result, neighborhood neighbors beat the culprit till he was gravely hurt. This instance demonstrates that crimes can happen at any time and in any location.

Due to the aforementioned issues, a data mining technique is required, in which the work process analyzes the surrounding circumstances to learn new information, such as grouping locations in Indonesia so that they can be avoided during specific times and enhancing security in provinces that are more likely to experience crime. Measures to lessen the danger of crime include placing security posts in all areas that are prone to crime as well as the requirement for community outreach to provide guidance. Several categories, including prediction, classification, clustering, association, and estimate, can be utilized in data mining[1]–[5]. The researchers used data from the previous year to categorize the provinces where crime was prevalent based on the issues that happened. The researchers will use two techniques, K-Means, and AHC, in clustering.

A study comparable to this one, comparing the K-Means and AHC approaches used by Ellang Putro Priambodo and Arief Jananto, will be conducted in 2022. The two researchers believe that future inventory can be anticipated using data from recent sales. Based on the quantity of inventory that has been sold, the amount of inventory is anticipated and grouped. The test results demonstrate that the K-Means algorithm and the AHC algorithm are both capable of classifying sold items according to the degree of similarity of the average number of sales. But the outcomes generated by the two differ from one another. Therefore, more research must be done to ascertain which algorithm can generate more precise inventory predictions. This can be accomplished by contrasting the anticipated outcomes with historical sales data[6].

Aceng Supriyadi and colleagues carried out research in 2021. This study analyzed the fleet performance evaluation procedure, which was still carried out manually, complicating data processing and leading to unreliable evaluations. As

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 295−307
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3630

a result, we require a data processing methodology that is both quicker and more precise. One such technique involves combining data mining methods with the clustering technique. In this study, the K-Means and K-Medoids algorithms were compared, and the reliability of the clusters created was examined. The K-Means Algorithm gives a validity value of 0.67 for clusters using the Davies Bouldin Index, whereas the K-Medoids Algorithm produces a validity value of 1.78. Because it has a lower DBI validity value than K-Medoids, the K-Means Algorithm was chosen to be utilized in creating web-based vehicle fleet cluster applications based on this validity value. Testing on web application cluster results reveals a 97% conformity level when utilizing the Rapidminer tool and hand computations. As a result, the K-Means Algorithm is effective in this application for creating precise clusters[7].

2020 saw the completion of studies by Hotma Dame Tampubolon et al. In this study, they employed the K-Means algorithm to pinpoint regions in the Pematangsiantar Region with high and low crime rates. thievery, rape, drug use, and even murder are common crimes. As a result, it is important to use the K-Means algorithm to group regions based on documents that describe crimes committed in violation of Pematangsiantar Police law. Data from 2019 that cover 6 districts were used in this study. Using the K-Means algorithm, results show that Siantar Martoba, East Siantar, West Siantar, and North Siantar are the four sub-districts with the highest crime rates (C1), while Siantar Selatan and Siantar Marihat are the two sub-districts with the lowest crime rates (C2)[8].

Nyoman Gde Prajnawiweka Ratmasa Taram, et al. did research in 2019 that sought to categorize crime rates in 32 provinces using the agglomerative and K-Means methodologies. The Single Linkage approach was shown to be the most effective method for classifying crime rates after comparing the level of standard deviation ratio of the three AHC methods (Single, Complete, and Average Linkage) and K-Means. Consequently, a comparison between the Single Linkage AHC and the K-Means approach was done. The Single Linkage technique has three groups, according to the findings. The first group includes 29 provinces. Provinces in East Java and North Sumatra make up the second and third groupings, respectively. The following are the identifying factors for the three groups: The crimes that fall into the first group are those that involve domestic violence, the crimes that fall into the second group involve minor mistreatment, and the crimes that fall into the third group involve corruption[9].

2. RESEARCH METHODOLOGY

2.1 Research Stages

The term "stages of research" refers to a sequence of actions or processes that must be taken during a study to guarantee its success and smooth operation as well as the full consideration of all relevant factors. The following is a summary of the steps taken during the research.

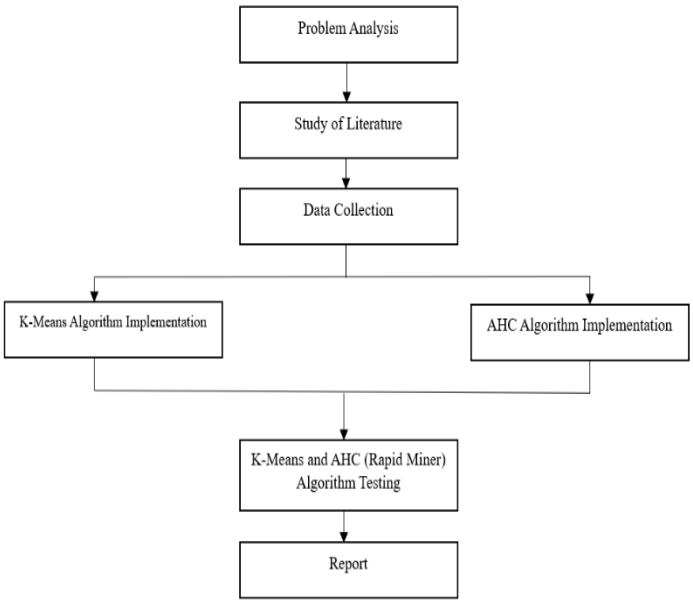

Figure 1. Research Stages

From Figure 1 can be explained as follows

a. Problem Analysis Stage
During the problem analysis stage, actions are performed to comprehend and pinpoint the underlying reasons for the issues being faced. Finding the root of the issue is the goal of this phase to come up with a sensible and workable solution. Data gathering, analysis, and interpretation are all steps in the problem analysis process that aim to pinpoint the root causes of issues and their effects on the systems they influence. We can comprehend the issue at hand and the factors that led to its formation by performing a thorough analysis of the issue. The problem can be resolved or overcome by taking appropriate action after having a thorough awareness of the core reasons. By focusing on treating

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 295−307
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3630

the problem's root rather than just its outward symptoms, problem analysis plays a crucial role in guiding suitable and effective problem-solving activities.

b. Literature Study Stage
The following step is to carry out a literature review or study, which is crucial for producing high-caliber research. This method entails gathering, dissecting, assessing, and assembling material from many sources of literature that are pertinent to the topic or issue being studied, including books, journals, articles, reports, and papers. The goal of the literature study stage is to get a deeper grasp of the subject or issue under investigation and to spot any knowledge gaps or areas in need of additional study.

c. Data Collection
Researchers will collect data using approaches that have been defined during the data-collecting phase. Depending on the type of research being done, methods utilized for data collecting may include interviews, surveys, observation, experiments, or document analysis. Additionally, data can be gotten from governmental organizations that offer datasets for information extraction so they can be used as sample data in a study.

d. K-Means and AHC Algorithm Implementation Phase
The next stage is to apply the algorithm to the data that has been acquired after doing problem analysis and data collecting. Data is processed using algorithms according to established rules. The K-Means and AHC clustering methods are employed in this work to address the issues that are present.

e. K-Means and AHC Algorithm Testing Phase
The K-Means and AHC algorithms must first be implemented before their performance can be assessed. The Rapid Miner platform was used to conduct the evaluation. The evaluation is deemed effective if the outcomes are in line with how the prior algorithm was put into practice.

f. Report
The researcher will complete producing a research report that concisely summarizes all phases and research findings in the last stage. An introduction, methods, findings, analysis, and conclusions are required in the research report. Additionally, the report must be written in the structure and writing style mandated by the appropriate research institution or publication.

2.2 Data Mining

Information retrieval, statistics, databases, machine learning, and other fields are combined in data mining to create a new science (new knowledge). Data mining is a step in the Knowledge Discovery in Databases (KDD) process that is used to find hidden information in databases. The process of examining and analyzing data using various techniques for diverse goals is also known as data mining. Using methods and techniques from the domains of statistics and database system administration, data mining is the process of extracting information from sizable data sets[10]–[15]. The following list of data mining groupings can be utilized as necessary.

a. Prediction
Based on the information that is currently available, prediction entails estimating or approximating future values or events. The objective is to offer an accurate or highly probable assessment of what might occur in the future. Utilizing statistical methods, machine learning, or other predictive algorithms to find patterns or trends in data that can be used to forecast future values or events is known as data mining. KNN, Naive Bayes, C4.5, Rough Set, SVM, and other popular algorithms for prediction[4], [16]–[18]

b. Clustering
Data can be grouped into groups or clusters based on various traits or characteristics using the clustering or grouping approach in data mining. Without employing preexisting categories or labels, clustering is used to find innate patterns or hidden structures in data. Algorithms that examine similarities or contrasts between data points are used in the clustering process. This algorithm seeks to cluster data so that data within a cluster are highly similar and data between clusters are significantly different. K-Means, AHC, K-Medoids, and other popular clustering techniques are only a few examples [19]–[21].

c. Classification
Data mining uses the classification process to classify data into predefined categories or classes. Finding patterns that categorize data into different classes based on pertinent properties is the aim. When data doesn't yet have a class assigned to it, classification models are used to forecast the correct class or label. Cart, C4.5, ID3, K-NN, Naive Bayes, and other methods of categorization that are frequently employed include these[3][22][23][24].

d. Association
In data mining, the idea of association refers to the connection or relationship between objects or attributes in a dataset. The goal of association analysis is to find hidden patterns in data that point to a relationship or pattern between these things. Searching for an itemset—a group of items that commonly appear together in transactions or events—is a step in the association analysis method. Apriori, Fp-Growth, and other techniques are some of those utilized in association analysis.[5], [25], [26].

e. Estimation
The process of estimation entails computing or approximating unknown values or sums using the information or data at hand. Even if there is no exact certainty, the objective is to offer an estimate that is close to the genuine number. Utilizing statistical methods or algorithms to compute or forecast unknown values based on the information at hand

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 295−307
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3630

is known as estimation in data mining. Variables or characteristics like income, population, expenses, and more can all be estimated. In numerous disciplines, including economics, finance, the social sciences, and other sciences, estimation is also commonly utilized. Expectation Maximization, Multiple Linear Regression, Simple Linear Regression, and other methods are utilized in the estimating process[1], [2], [27], [28].

2.3 Algoritma K-Means

The K-Means method is a technique for clustering and data analysis that seeks to divide data into groups according to the attributes given. Finding the centroid, or the center of a bunch of data that have a particular degree of similarity, is the aim of the K-Means algorithm. This algorithm's first step is to choose the desired number of groups (k) and randomly establish the initial centroid point. The program will then determine the separation between each data point and the centroid before grouping the data to that centroid. The centroid position is then updated by averaging the data from each group. Up until the termination requirements are reached or there is no change in the data placement, this process is repeated[29]–[31]. The steps that can be taken to implement the K-Means approach for clustering are as follows:

a. Determine how many clusters there are.
b. Calculate the centroid center in the first iteration using a random data value as the starting centroid center.

$$Ki = \frac{1}{M}\sum_{j=1}^{M} X_j \quad (1)$$

c. To get the shortest distance, use the starting centroid.

$$d_{Euclidean}(X,Y) = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2} \quad (2)$$

formula:

d(x,y) = Data separation from cluster center y to x
Xi = the n-th data attribute's i-th data
Yi = the n-th data attribute's j-th data

d. While data with a bigger distance will be placed in a distinct group, data with the closest value will be clustered together in one group (cluster).
e. Use the new centroid position, which is established based on the group that is closest to the data in the previous iteration. From the first step through the final iteration, this process is repeated. The iteration will proceed if the group's centroid position changes. The iteration process will be stopped, though, if the group's centroid position does not change.

2.4 Algorithm for Agglomerative Hierarchical Clustering (AHC)

Data are grouped into hierarchical clusters using the AHC algorithm or method in cluster analysis. This method begins by treating each cluster of data as a separate entity before eventually combining the clusters that are most similar to one another in terms of a certain metric or distance. The first step of the AHC procedure is to calculate the separation between each pair of data and combine the two clusters that have the shortest separation to create a new cluster. Up until all the data is integrated into one major cluster, this procedure keeps combining larger and larger clusters. As a result of this procedure, a hierarchical structure that typically takes the shape of a dendrogram tree or graph and explains the connections between clusters is created. The AHC method's adaptability in terms of selecting the metric or distance used to gauge data similarity is one of its benefits. Depending on the properties of the data and the goal of the study, the metric can be either the Euclidean distance, the Manhattan distance, or the Mahalanobis distance. Additionally, the number of clusters to be used can be chosen, and the cluster merging process can be stopped by utilizing cut-off criteria. Using the appropriate formula, the Manhattan or Euclidean distance between two clusters can be computed[21][32][33][34][35].

a. Does the distance matrix calculation

Manhattan Distance:

$$D_{man}(x,y) = \sum_{j=1}^{d}|x_{2j} - y_{1j}| \quad (1)$$

Euclidean Distance:

$$d(x_2,x_1) = \sqrt{\sum_{j=1}^{d}|x_j - y_j|^2} \quad (2)$$

b. AHC grouping

There are three main ways that the Hierarchical Agglomerative Clustering (AHC) method can be used to group data. Single linkage is the initial method, which divides data into clusters based on the shortest (smallest) distance between them. Complete linkage, which organizes data based on the greatest (maximum) distance between various items in the cluster, is the second mode. This method focuses on finding the cluster with the biggest difference between two clusters by measuring the greatest distance between them. The average linkage mode, which categorizes data based on the average value of the two associated clusters, is the third mode. This option focuses on the average distance between the two clusters that will be combined[36][37][38][39]. You can utilize the AHC grouping modes listed below.

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 295−307
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3630

a. Single Linkage Grouping:

$$d_{uv} = \min\{d_{uv}\}, d_{uv} \in D \quad (3)$$

b. Complate linkage grouping:

$$d_{uv} = \max\{d_{uv}\}, d_{uv} \in D \quad (4)$$

c. Average Linkage grouping:

$$d_{uv} = \text{average}\{d_{uv}\}, d_{uv} \in D \quad (5)$$

3. RESULTS AND DISCUSSION

The following link page contains the 2022 criminal statistics dataset that will be used in this study to classify Indonesian provinces that are crime victims: https://www.bps.go.id/publication/2022/11/30/4022d3351bf3a05aa6198065/statistik-kriminal-2022.html, The 34 Indonesian provinces that will be classified based on the number of victims of crime from 2019 to 2021 (the previous three years) are 34 provinces. The K-Means and AHC algorithms will be used to do clustering, with the result that two clusters will arise that will be examined. An example data table that will be grouped in this study can be seen in table 1.

Table 1. Data on crime victims

Province	2019	2020	2021
Aceh	0,71	0,61	0,32
North Sumatra	1,18	0,97	0,74
West Sumatra	0,98	0,97	0,48
Riau	1,24	0,73	0,56
Jambi	0,83	0,79	0,5
South Sumatra	1,35	1,07	0,57
Bengkulu	1,42	0,95	0,97
Lampung	1,42	1	0,5
Bangka Belitung Islands	0,66	0,72	0,41
Riau islands	1,02	1,44	0,55
DKI Jakarta	1,15	0,86	0,4
West Java	1,15	0,9	0,48
Central Java	0,8	0,57	0,37
In Yogyakarta	1,35	0,79	0,41
East Java	0,82	0,66	0,35
...	...	...	...
...	...	...	...
West Papua	1,66	1,04	0,73
Papua	1,36	0,73	0,55

3.1 K-Means Algorithm Implementation

By grouping the provinces of Indonesia into clusters and finding commonalities between each cluster, the K-Means method was used to categorize crimes committed in each province over the previous three years. Using the K-Means approach, a new cluster is first formed by calculating the initial centroid through a series of calculations from iteration one through iteration n. The K-Means approach performs the subsequent phases in the cluster-building process:

a. The number of clusters to be formed is 2 clusters (C1 and C2)

Iteration Step 1:

b. The following table displays the original centroid center, which was randomly selected from sample data of crime victims.

Table 2. Initial Centroid Centers

Province	2019	2020	2021
Bali	0,42	0,23	0,2
West Papua	1,66	1,04	0,73

c. Using the euclidean distance to determine the shortest distance between clusters

$$d_{Euclidean}(X, Y) = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2}$$

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 295−307
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3630

Data 1:

$C_1 = \sqrt{(0{,}71 - 0{,}42)^2 + (0{,}61 - 0{,}23)^2 + (0{,}32 - 0{,}2)^2} = 0{,}4488$

$C_2 = \sqrt{(0{,}71 - 1{,}66)^2 + (0{,}61 - 1{,}04)^2 + (0{,}32 - 0{,}73)^2} = 1{,}303$

Data 2:

$C_1 = \sqrt{(1{,}18 - 0{,}42)^2 + (0{,}97 - 0{,}23)^2 + (0{,}74 - 0{,}2)^2} = 1{,}5992$

$C_2 = \sqrt{(1{,}18 - 1{,}66)^2 + (0{,}97 - 1{,}04)^2 + (0{,}74 - 0{,}73)^2} = 0{,}485$

Data 3:

$C_1 = \sqrt{(0{,}98 - 0{,}42)^2 + (0{,}97 - 0{,}23)^2 + (0{,}74 - 0{,}2)^2} = 1{,}186$

$C_2 = \sqrt{(0{,}98 - 1{,}66)^2 + (0{,}97 - 1{,}04)^2 + (0{,}74 - 0{,}73)^2} = 0{,}7474$

In the search for the shortest distance between clusters, data from the fourth province to the 34th province are also calculated using the aforementioned Euclidean distance calculation technique. The distance calculation's outcome after accounting for all provinces can be seen in table 3.

Table 3. Euclidean distance in iteration 1

Province	C1	C2	Closest Distance	Cluster
Aceh	0,4488	1,303	0,4488	C1
North Sumatra	1,5992	0,485	0,485	C2
West Sumatra	1,186	0,7474	0,7474	C2
Riau	1,1996	0,545	0,545	C2
Jambi	0,8136	0,9454	0,8136	C1
South Sumatra	1,7725	0,3365	0,3365	C2
Bengkulu	2,1113	0,3057	0,3057	C2
Lampung	1,6829	0,2945	0,2945	C2
Bangka Belitung Islands	0,5242	1,2048	0,5242	C1
Riau islands	2,1866	0,8324	0,8324	C2
DKI Jakarta	1,1669	0,6513	0,6513	C2
West Java	1,2573	0,5921	0,5921	C2
Central Java	0,4488	1,303	0,4488	C1
In Yogyakarta	1,5992	0,485	0,485	C2
East Java	0,6074	1,1288	0,6074	C1
...	...	...	...	...
...	...	...	...	...
West Papua	2,177	0	0	C2
Papua	1,3125	0,4285	0,4285	C2

d. While data with a bigger distance will be placed in a distinct group, data with the closest value will be clustered together in one group (cluster).
e. Use the new centroid position for the subsequent iteration, which is chosen based on the group that is closest to the data.

$Ki = \frac{1}{M}\sum_{j=1}^{M} X_j$

C1

$2019 = \frac{1}{14}(0{,}71 + 0{,}83 + 0{,}66 + 0{,}8 + 0{,}82 + 0{,}8 + 0{,}42 + 0{,}84 + 0{,}63 + 0{,}8 + 0{,}84 + 0{,}78 + 0{,}83 + 0{,}74)$
$= 0{,}75$

$2020 = \frac{1}{14}(0{,}61 + 0{,}79 + 0{,}72 + 0{,}57 + 0{,}66 + 0{,}46 + 0{,}23 + 0{,}73 + 0{,}69 + 0{,}55 + 0{,}59 + 0{,}63 + 0{,}59 + 0{,}58) = 0{,}6$

$2021 = \frac{1}{14}(0{,}32 + 0{,}5 + 0{,}41 + 0{,}37 + 0{,}35 + 0{,}36 + 0{,}2 + 0{,}57 + 0{,}44 + 0{,}47 + 0{,}42 + 0{,}57 + 0{,}46 + 0{,}4)$
$= 0{,}4171$

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 295−307
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3630

C2

$$2019 = \frac{1}{20}(1{,}18 + 0{,}98 + 1{,}24 + 1{,}35 + 1{,}42 + 1{,}42 + 1{,}02 + 1{,}15 + 1{,}15 + 1{,}35 + 1{,}63 + 1{,}19 + 1{,}53 + 1{,}22 + 0{,}96 + 0{,}95 + 1{,}19 + 1{,}16 + 1{,}66 + 1{,}36) = 1{,}2555$$

$$2020 = \frac{1}{20}(0{,}97 + 0{,}97 + 0{,}73 + 1{,}07 + 0{,}95 + 1 + 1{,}44 + 0{,}86 + 0{,}9 + 0{,}79 + 1{,}49 + 0{,}6 + 1{,}36 + 1{,}17 + 0{,}73 + 0{,}9 + 1{,}09 + 1{,}06 + 1{,}04 + 0{,}73\ ) = 0{,}9925$$

$$2021 = \frac{1}{20}(0{,}74 + 0{,}48 + 0{,}56 + 0{,}57 + 0{,}97 + 0{,}5 + 0{,}55 + 0{,}4 + 0{,}48 + 0{,}41 + 0{,}98 + 0{,}47 + 0{,}57 + 0{,}63 + 0{,}41 + 0{,}5 + 0{,}78 + 0{,}55 + 0{,}73 + 0{,}55\ = 0{,}5915$$

Table 4. Iteration Centroid Center 2

Cluster	2019	2020	2021
C1	0,75	0,6	0,4171
C2	1,2555	0,9925	0,5915

From the initial step until the last iteration (the nth iteration), this process is repeated. The iteration will proceed if the group's centroid position changes. The iteration process will be stopped, though, if the group's centroid position does not change.

3.2 AHC Algorithm Implementation

a. Apply the following Euclidean formula to determine the matrix distance.

$$d(x_2, x_1) = \sqrt{\sum_{j=1}^{d} |x_j - y_j|^2}$$

$$d(Aceh, North\ Sumatera) = \sqrt{|0{,}71 - 1{,}18|^2 + |0{,}61 - 0{,}97|^2 + |0{,}32 - 0{,}74|^2}$$

$$d(Aceh, West\ Sumatera) = \sqrt{|0{,}71 - 0{,}98|^2 + |0{,}61 - 0{,}97|^2 + |0{,}32 - 0{,}48|^2}$$

$$d(Aceh, Riau) = \sqrt{|0{,}71 - 1{,}24|^2 + |0{,}61 - 0m73|^2 + |0{,}32 - 0{,}56|^2}$$

Find the Euclidean distance from the following matrix's formation in Aceh, Jambi, too (West Papua, Papua). Following are the calculations' findings for all province-to-province Euclidean distances with an order of 34*34.

Table 5. Euclidean Distance Matrix

Province	Aceh	North Sumatera	West Sumatera	Riau	...	...	West Papua	Papua
Aceh	0				...	...		
North Sumatra	0,726	0			...	...		
West Sumatra	0,478	0,328	0		...	...		
Riau	0,594	0,306	0,363	0	...	...		
Jambi	0,281	0,461	0,235	0,419	...	...		
South Sumatra	0,827	0,26	0,394	0,357	...	...		
Bengkulu	1,021	0,333	0,659	0,499	...	...		
Lampung	0,830	0,341	0,441	0,33	...	...		
Bangka Belitung Islands	0,151	0,665	0,412	0,599	...	...		
Riau islands	0,915	0,532	0,477	0,743	...	...		
DKI Jakarta	0,512	0,359	0,218	0,225	...	...		
West Java	0,551	0,271	0,184	0,208	...	...		
Central Java	0,110	0,664	0,452	0,505	...	...		
In Yogyakarta	0,671	0,413	0,417	0,195	...	...		
East Java	0,124	0,615	0,372	0,475	...	...		
...	...	...	...	...	...	...		
...	...	...	...	...	...	...		
West Papua	1,120	0,485	0,728	0,549	...	...	0	
Papua	0,700	0,355	0,455	0,120	...	...	0,5	0

The first step is to select the initial cluster with the shortest Euclidean distance based on Table 5. The South Kalimantan and Gorontalo matrices in this instance were determined to have the smallest (minimum) value of 0.041. As a result, the South Kalimantan and Gorontalo clusters, which are also the first clusters, can be classified as regions with relatively low crime rates or safer provinces.

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 295−307
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3630

b. The predetermined clusters will now be combined into a single cluster known as the South Kalimantan, Gorontalo cluster. The complete linkage method is used for subsequent grouping, in which the distance between each cluster and the next is calculated based on their distance from one another. In this procedure, the greatest value between each cluster is determined, then the subsequent clusters are joined to create a larger cluster.

3.3 K-Means and AHC algorithms are put to the test

Users can develop data analysis processes using Rapid Miner's drag-and-drop functionality and user-friendly interface without having to manually write code. Even users without programming experience can simply conduct data analysis with such a simple interface. In addition, RapidMiner offers a wide range of potent data analysis techniques, such as clustering, association analysis, regression, classification, and more. Users can compare the outcomes of the K-Means algorithm and the AHC algorithm to determine which method best meets their needs. RapidMiner offers several preprocessing tools to get data ready for analysis. With the help of these tools, users can normalize data, combine and split columns, clean data, remove missing values, and do other data transformations. Users can quickly prepare data for additional analysis with the help of this effective preparation procedure. In addition, RapidMiner offers a wide range of potent visualization tools to help customers analyze and comprehend data. Users can quickly and easily identify patterns and correlations in data using graphs, charts, and other representations. Users benefit from a better understanding of the data and improved decision-making as a result. The operator input for the two algorithms (K-Means and AHC) required for identifying Indonesian provinces based on crime victim data is shown in the figure 2.

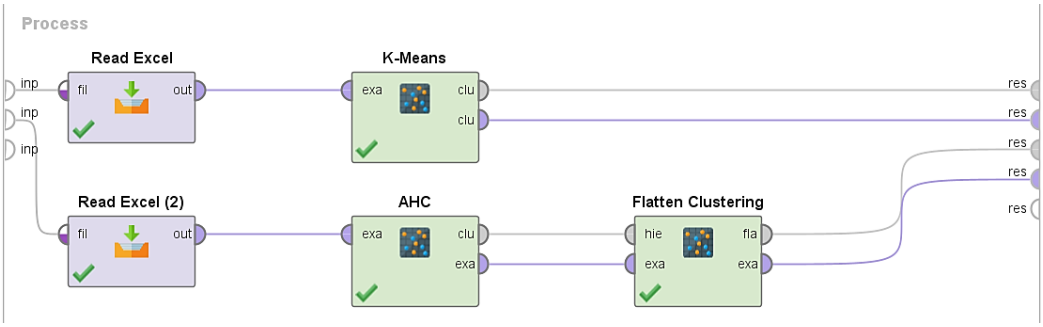

Figure 2. The K-Means and AHC file input dataset process

The procedures for using Rapid Miner are shown in Figure 2. Data files for processing are first entered as input; these data files can be duplicate Excel files so that they can be connected to the two clustering methods. Then enter the AHC grouping operator and connect it to the flatten clustering operator to determine the number of clusters to be formed as well as the maximum distance calculation method known as the complete linked method. Then enter the K-Means clustering operator by specifying the number of clusters to be formed as 2 clusters. Hierarchical distances between each pair of provinces are computed using this method. The RapidMiner procedure is launched after all operators based on each algorithm have been connected. The researchers created 33 clusters to test the Rapid Miner software and show that the manual computations performed previously and its use are compatible. The results of the manual computations and those obtained by utilizing the Rapid Miner application must agree with the first cluster that is generated. The cluster model created with 33 clusters is shown in the figure 3.

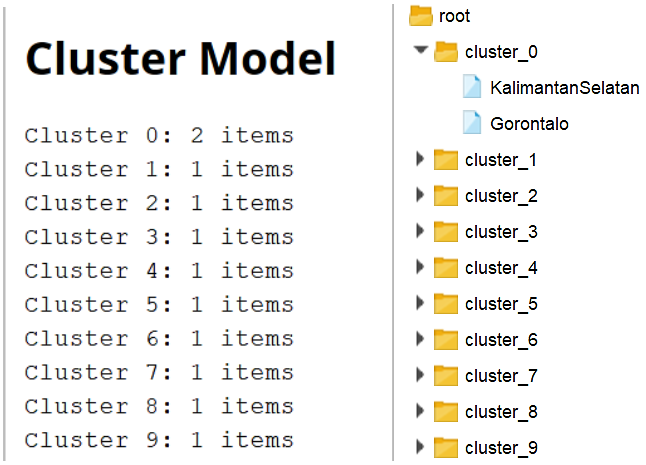

Figure 3. Formation of the first cluster

The provinces of South Kalimantan and Gorontalo make up the first cluster (cluster 0), as can be seen in Figure 3. South Kalimantan and Gorontalo are also the first cluster according to manual calculations using the AHC algorithm (for proof, you can only use the AHC algorithm since the K-Means algorithm requires many iterations and the results of

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 295−307
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3630

cluster placement in the first iteration may differ from those in subsequent iterations). This demonstrates how the manual procedure and the use of the RapidMiner application are compatible. As a result, it can be said that the outcomes of the two approaches are comparable, and RapidMiner can continue to be used. The emphasis will now be on creating two clusters.

b.2 K-Means Algorithm Testing Results

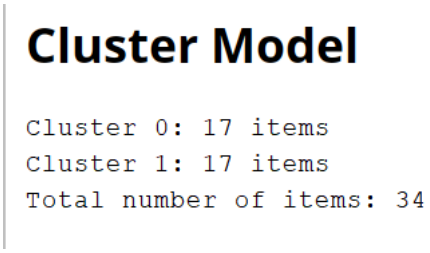

Figure 4. Cluster Models

In the first cluster (shown in Figure 4), there are 17 provinces (items), and there are also 17 provinces (items) aggregated in the second cluster (shown in Figure 4). As a result, the first and second clusters are in balance. The provinces included in clusters 1 and 2 are discussed in greater detail in the table 6.

Table 6. K-Means Formed Clusters

Province	2019	2020	2021	Cluster
Aceh	0.71	0.61	0.32	1
North Sumatra	1.18	0.97	0.74	0
West Sumatra	0.98	0.97	0.48	0
Riau	1.24	0.73	0.56	0
Jambi	0.83	0.79	0.5	1
South Sumatra	1.35	1.07	0.57	0
Bengkulu	1.42	0.95	0.97	0
Lampung	1.42	1	0.5	0
Bangka Belitung Islands	0.66	0.72	0.41	1
Riau islands	1.02	1.44	0.55	0
DKI Jakarta	1.15	0.86	0.4	0
West Java	1.15	0.9	0.48	0
Central Java	0.8	0.57	0.37	1
In Yogyakarta	1.35	0.79	0.41	0
East Java	0.82	0.66	0.35	1
...	...	...	...	...
...	...	...	...	...
West Papua	1.66	1.04	0.73	0
Papua	1.36	0.73	0.55	0

According to Table 6, there are 17 provinces divided into clusters 0 and 1, with cluster 0 consisting of North Sumatra, West Sumatra, Riau, South Sumatra, Bengkulu, Lampung, Riau Islands, DKI Jakarta, Java West, DI Yogyakarta, West Nusa Tenggara, North Kalimantan, Central Sulawesi, Maluku, North Maluku, West Papua, and Papua. Aceh, Jambi, the Bangka Belitung Islands, Central Java, East Java, Banten, Bali, East Nusa Tenggara, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Sulawesi, South Sulawesi, Southeast Sulawesi, Gorontalo, and West Sulawesi make up the provinces that are grouped into cluster 1.

Attribute	cluster_0	cluster_1
2019.0	1.295	0.800
2020.0	1.036	0.625
2021.0	0.615	0.425

Figure 5. Final Centroids

According to figure 5, the final centroid derived following the grouping procedure selection, namely cluster 0, has a centroid value of 1.295 in 2019, 1.036 in 2020, and 0.615 in 2021. While the centroid value for cluster 1's final year is 0.800 in 2019, 0.625 in 2020, and only 0.425 in 2021. The following graph shows the percentage of each cluster.

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 295−307
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3630

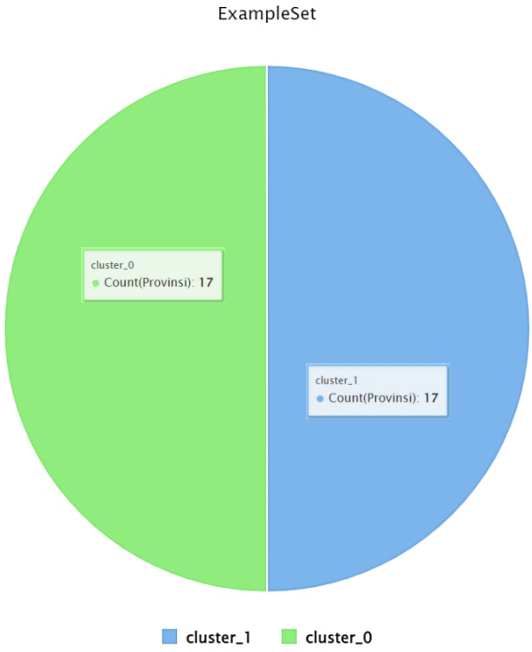

Figure 6. K-Means Cluster Visualization

According to Figure 6's representation of the clusters, each cluster's percentage is 50%. This percentage is also visible in blue for cluster 1, which has a total of 17 provinces, and green for cluster 0, which also has a group of 17 provinces.

b.3 AHC Algorithm Testing Results

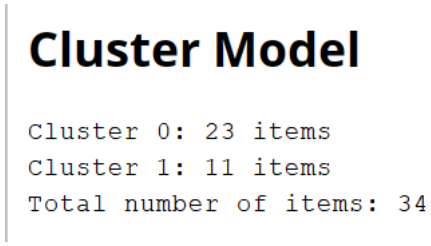

Figure 7. AHC Model Clusters

According to Figure 7, of the 34 provinces that are divided into two clusters, 23 items (provinces) are grouped into cluster 0 while only 11 things (provinces) are grouped into cluster 1. To see the categorization of provinces based on victims of the following crimes in further detail, see more information.

Table 8. AHC Formed Clusters

Province	2019	2020	2021	Cluster
Aceh	0.71	0.61	0.32	0
North Sumatra	1.18	0.97	0.74	1
West Sumatra	0.98	0.97	0.48	0
Riau	1.24	0.73	0.56	0
Jambi	0.83	0.79	0.5	0
South Sumatra	1.35	1.07	0.57	1
Bengkulu	1.42	0.95	0.97	1
Lampung	1.42	1	0.5	1
Bangka Belitung Islands	0.66	0.72	0.41	0
Riau islands	1.02	1.44	0.55	1
DKI Jakarta	1.15	0.86	0.4	0
West Java	1.15	0.9	0.48	0
Central Java	0.8	0.57	0.37	0
In Yogyakarta	1.35	0.79	0.41	0
East Java	0.82	0.66	0.35	0
...	...	...	...	...

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 295−307
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3630

...	...	...	...	...
West Papua	1.66	1.04	0.73	1
Papua	1.36	0.73	0.55	0

Based on Table 8, the provinces grouped into cluster 0 are 23 provinces consisting of Aceh, West Sumatra, Riau, Jambi, Bangka Belitung Islands, DKI Jakarta, West Java, Central Java, DI Yogyakarta, East Java, Banten, Bali, Nusa East Southeast, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Sulawesi, South Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, and Papua. North Sumatra, South Sumatra, Bengkulu, Lampung, Riau Islands, West Nusa Tenggara, North Kalimantan, Central Sulawesi, Maluku, North Maluku, and West Papua are the only 11 provinces in cluster 1, in contrast.

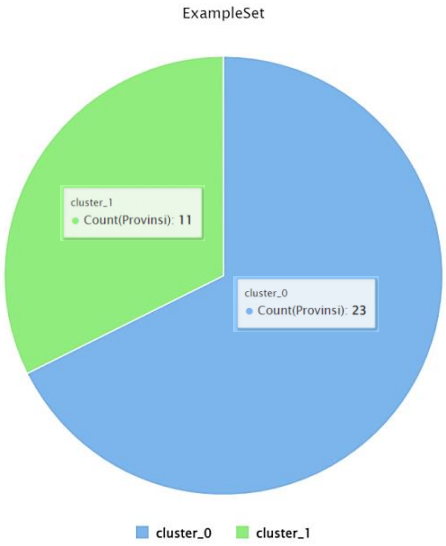

Figure 7. AHC Cluster Visualization

According to Figure 7, just 32.35% of the total number of provinces, or 34 provinces, are represented by cluster 0, which is much less than cluster 1's share of 67.65%.

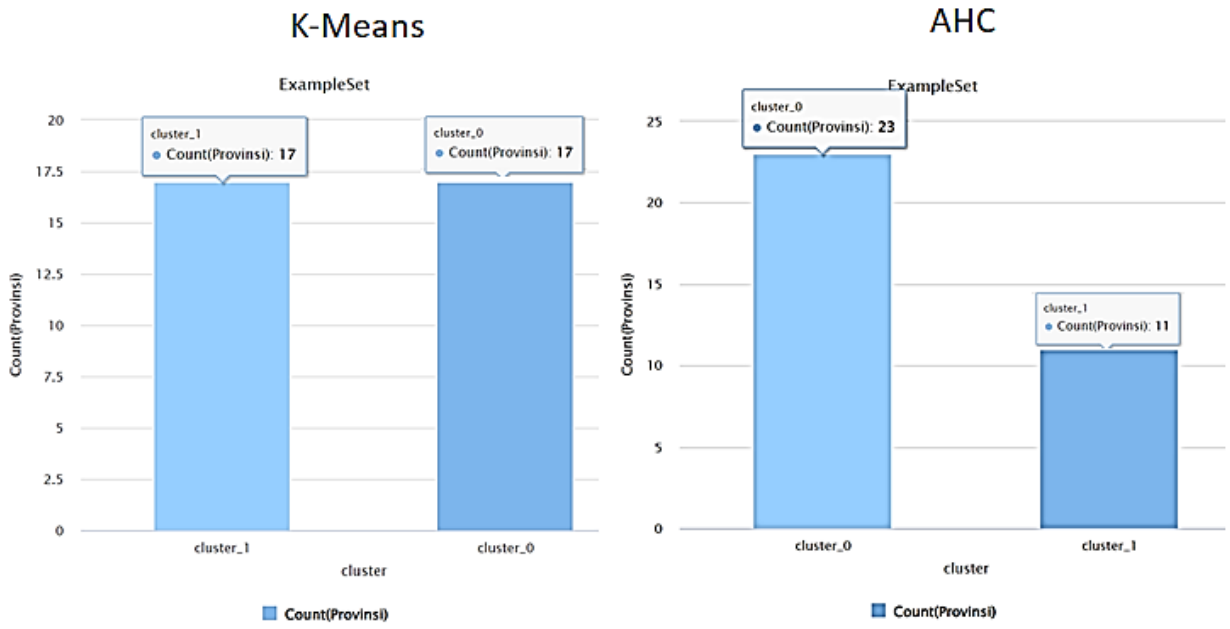

Figure 8. Comparison of the number of K-Means and AHC groups

Based on Figure 8, it is possible to analyze the grouping results and see that the K-Means method groups the provinces into equal groups of 17, whereas the AHC complete linkage method groups the provinces into clusters of 23, with a difference of 8 provinces for cluster 0 and a difference of 6 provinces for cluster 1. The location of the province inside the clusters is also different, with the K-Means technique placing the province's data in cluster 0 and the AHC placing it in cluster 1.

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 295−307
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3630

4. CONCLUSION

Using information from the 2022 crime statistics, this study seeks to categorize the Indonesian provinces that are the targets of crime based on the discussion that was conducted in the previous sub-chapter. Based on the number of victims of crime from 2019 to 2021, 34 provinces will be grouped. K-Means and AHC (Agglomerative Hierarchical Clustering) using complete linkage mode are two algorithms used in clustering. K-Means grouping results in two groups, each of which has 17 provinces. In contrast, using the AHC complete linkage, there are 6 distinct provinces in cluster 1 and 23 different provinces in cluster 0, which differ by 8 provinces from the K-Means results. The position of the province within the cluster varies between the two ways as well. The K-Means approach places provincial data in cluster 0, but the AHC method places it in cluster 1.

REFERENCES

[1] A. S. L. T. T. H. Hafizah, "Data Mining Estimasi Biaya Produksi Ikan Kembung Rebus Dengan Regresi Linier Berganda," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, no. Vol 1, No 6 (2022): EDISI NOVEMBER 2022, pp. 888–897, 2022.
[2] Y. L. Nainel, E. Buulolo, and I. Lubis, "Penerapan Data Mining Untuk Estimasi Penjualan Obat Berdasarkan Pengaruh Brand Image Dengan Algoritma Expectation Maximization (Studi Kasus: PT. Pyridam Farma Tbk)," *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 2, p. 214, 2020.
[3] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 640, 2021.
[4] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019.
[5] H. Maulidiya and A. Jananto, "Asosiasi Data Mining Menggunakan Algoritma Apriori dan FP-Growth sebagai Dasar Pertimbangan Penentuan Paket Sembako," *Proceeding SENDIU 2020*, vol. 6, pp. 36–42, 2020.
[6] E. P. Priambodo and A. Jananto, "Perbandingan Analisis Cluster Algoritma K-Means Dan AHC Dalam Perencanaan Persediaan Barang Pada Perusahaan Manufaktur," *Progresif J. Ilm. Komput.*, vol. 18, no. 2, p. 257, 2022.
[7] A. Supriyadi, A. Triayudi, and I. D. Sholihati, "Perbandingan algoritma k-means dengan k-medoids pada pengelompokan armada kendaraan truk berdasarkan produktivitas," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 6, no. 2, pp. 229–240, 2021.
[8] H. D. Tampubolon, D. Gultom, L. Y. Hutabarat, F. I. R. H. Zer, and D. Hartama, "Penerapan Algoritma K-Means Untuk Mengetahui Tingkat Tindak Kejahatan Daerah Pematangsiantar," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 146–151, 2020.
[9] N. G. P. R. TARAM, I. K. G. SUKARSA, and I. G. A. M. SRINADI, "Pengelompokan Tingkat Kriminalitas Dengan Metode Agglomerative Dan K-Means Serta Peubah Pencirinya," *E-Jurnal Mat.*, vol. 8, no. 2, p. 102, 2019.
[10] R. H. Sukarna and Y. Ansori, "Implementasi Data Mining Menggunakan Metode Naive Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu," *J. Ilm. Sains dan Teknol.*, vol. 6, no. 1, pp. 50–61, 2022.
[11] F. O. Lusiana, I. Fatma, and A. P. Windarto, "Estimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Simalungun," *J. Informatics Manag. Inf. Technol.*, vol. 1, no. 2, pp. 79–84, 2021.
[12] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021.
[13] G. Gunadi and D. I. Sensuse, "Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori Dan Frequent Pattern Growth (Fp-Growth) :," *Telematika*, vol. 4, no. 1, pp. 118–132, 2012.
[14] A. Z. Siregar, "Implementasi Metode Regresi Linier Berganda Dalam Estimasi Tingkat Pendaftaran Mahasiswa Baru," *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 2, no. 3, pp. 133–137, 2021.
[15] S. S. S, A. T. Purba, V. Marudut, M. Siregar, T. Komputer, and P. B. Indonesia, "SISTEM PENDUKUNG KEPUTUSAN KELAYAKAN PEMBERIAN PINJAMAN," vol. 3, pp. 25–30, 2020.
[16] M. M. Effendi, "Menentukan Prediksi Kelulusan Siswa Dengan Membandingkan Algoritma C4. 5 Dan Naive Bayes Studi Kasus SMKN. 1 Cikarang Selatan," *J. SIGMA*, vol. 11, no. 3, pp. 143–148, 2020.
[17] S. U. Putri, E. Irawan, and F. Rizky, "Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4. 5," *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 2, no. 1, pp. 39–46, 2021.
[18] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4, 5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019.
[19] F. Harahap, "Perbandingan Algoritma K Means dan K Medoids Untuk Clustering Kelas Siswa Tunagrahita," *TIN Terap. Inform. Nusant.*, vol. 2, no. 4, pp. 191–197, 2021.
[20] M. A. Rofiq, A. Qoiriah, S. Kom, and M. Kom, "Pengelompokan Kategori Buku Berdasarkan Judul Menggunakan Algoritma Agglomerative Hierarchical Clustering Dan K-Medoids," *J. Informatics Comput. Sci.*, vol. 2, no. 03, pp. 220–227, 2021.
[21] B. Harli Trimulya Suandi As and L. Zahrotun, "PENERAPAN DATA MINING DALAM MENGELOMPOKKAN DATA RIWAYAT AKADEMIK SEBELUM KULIAH DAN DATA KELULUSAN MAHASISWA MENGGUNAKAN METODE AGGLOMERATIVE HIERARCHICAL CLUSTERING (Implementation Of Data Mining In Grouping Academic History Data Before Students And Stud," *J. Teknol. Informasi, Komput. dan Apl.*, vol. 3, no. 1, pp. 62–71, 2021.
[22] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp. 219–225, 2021.
[23] I. A. Nikmatun and I. Waspada, "Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 10, no. 2, pp. 421–432, 2019.
[24] H. Hozairi, A. Anwari, and S. Alim, "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes," *Netw. Eng. Res. Oper.*, vol. 6, no. 2, pp. 133–144, 2021.
[25] H. Maulidiya and A. Jananto, "Asosiasi Data Mining Menggunakan Algoritma Apriori Dan Fpgrowth Sebagai Dasar Pertimbangan Penentuan Paket Sembako," 2020.
[26] K. Erwansyah, B. Andika, and R. Gunawan, "Implementasi Data Mining Menggunakan Asosiasi Dengan Algoritma Apriori

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 295−307
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3630

Untuk Mendapatkan Pola Rekomendasi Belanja Produk Pada Toko Avis Mobile," *J. Teknol. Sist. Inf. dan Sist. Komput. TGD*, vol. 4, no. 1, pp. 148–161, 2021.

[27] A. Rivandi, E. Bu'ulolo, and N. Silalahi, "Penerapan Metode Regresi Linier Berganda Dalam Estimasi Biaya Pencetakan Spanduk (Studi Kasus: PT. Hansindo Setiapratama)," *Pelita Inform. Inf. dan Inform.*, vol. 7, no. 3, pp. 263–268, 2019.

[28] P. Purwadi, P. S. Ramadhan, and N. Safitri, "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 18, no. 1, pp. 55–61, 2019.

[29] . F., F. T. Kesuma, and S. P. Tamba, "Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 67–72, 2020.

[30] S. A. Rahmah, "KLASTERISASI POLA PENJUALAN PESTISIDA MENGGUNAKAN METODE K-MEANS CLUSTERING (STUDI KASUS DI TOKO JUANDA TANI KECAMATAN HUTABAYU RAJA)," vol. 1, no. 1, pp. 1–5, 2020.

[31] W. Purba, W. Siawin, and . H., "Implementasi Data Mining Untuk Pengelompokkan Dan Prediksi Karyawan Yang Berpotensi Phk Dengan Algoritma K-Means Clustering," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 85–90, 2019.

[32] R. A. Setyawan and R. M. Fadilla, "Klasterisasi media pembelajaran daring di era pandemi COVID-19 menggunakan metode Agglomerative," *Inf. Interaktif*, vol. 5, no. 3, 2020.

[33] Marjiyono, "Penerapan Algoritma Ahc Algorithm Dalam Aplikasi Ppembagian Kelas Siswa Baru," *Semin. Nas. Teknol. Inf. dan Multimed. 2015*, pp. 6–8, 2015.

[34] T. Li, A. Rezaeipanah, and E. M. T. El Din, "An ensemble agglomerative hierarchical clustering algorithm based on clusters clustering technique and the novel similarity measurement," *J. King Saud Univ. Inf. Sci.*, vol. 34, no. 6, pp. 3828–3842, 2022.

[35] R. T. Adek, R. K. Dinata, and A. Ditha, "Online Newspaper Clustering in Aceh using the Agglomerative Hierarchical Clustering Method," *Int. J. Eng. Sci. Inf. Technol.*, vol. 2, no. 1, pp. 70–75, 2022.

[36] P. Govender and V. Sivakumar, "Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)," *Atmos. Pollut. Res.*, vol. 11, no. 1, pp. 40–56, 2020.

[37] C. Briggs, Z. Fan, and P. Andras, "Federated learning with hierarchical clustering of local updates to improve training on non-IID data," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–9.

[38] K. Zeng, M. Ning, Y. Wang, and Y. Guo, "Hierarchical clustering with hard-batch triplet loss for person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 13657–13665.

[39] N. K. Zuhal, "Study Comparison K-Means Clustering dengan Algoritma Hierarchical Clustering," *Pros. Semin. Nas. Teknol. dan Sains*, vol. 1, pp. 200–205, 2022.