ComTech: Computer, Mathematics and Engineering Applications, 15(2), December 2024, 119−127 P-ISSN: 2087-1244
DOI: 10.21512/comtech.v15i2.11933 E-ISSN: 2476-907X

Implementation of Clustering and Association for Early Warning of Disasters in Bojonegoro Regency

Denny Nurdiansyah[1]*, Erna Hayati[2], Ika Purnamasari[3], Anna Apriana Hidayanti[4], and Yuliana Fuji Rahayu[5]

[1,5]Statistics Study Program, Faculty of Science and Technology, Universitas Nahdlatul Ulama Sunan Giri
Bojonegoro, Indonesia 62115
[2]Department of Accounting, Faculty of Economics, Universitas Islam Lamongan
Lamongan, Indonesia 62211
[3]Statistics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Mulawarman
Kota Samarinda, Indonesia 75119
[4]Agribusiness Study Program, Faculty of Agriculture, Universitas Mataram
Kota Mataram, Indonesia 83115
[1]denny.nur@unugiri.ac.id; [2]ernahayati@unisla.ac.id; [3]ika.purnamasari@fmipa.unmul.ac.id;
[4]anna_apriana@unram.ac.id; [5]yuliana.fuji@unugiri.ac.id

Received: 22nd July 2024/ **Revised**: 10th November 2024/ **Accepted**: 11th November 2024

How to Cite: Nurdiansyah, D., Hayati, E., Purnamasari, I., Hidayanti, A. A., & Rahayu, Y. F. (2024). Implementation of Clustering and Association for Early Warning of Disasters in Bojonegoro Regency. *ComTech: Computer, Mathematics and Engineering Applications, 15*(2), 119−127. https://doi.org/10.21512/comtech.v15i2.11933

Abstract - The research aimed to analyze the relationships between different types of disasters, assess the likelihood of disaster occurrences, and enhance knowledge and understanding of disaster patterns in Bojonegoro Regency. The goal was to enable better disaster prediction and preparedness in the future. The methods applied included mapping, clustering using the K-means algorithm, and association rule mining with the Apriori algorithm. Secondary data were obtained from the National Disaster Management Agency and the Bojonegoro Regency Regional Disaster Management Agency Office, covering eight types of disasters. The results reveal that the K-means model groups the data into 5 clusters from 28 sub-districts in Bojonegoro. There are 13 sub-districts in Cluster 0, 1 sub-district in Cluster 1, 4 sub-districts in Cluster 2, 6 sub-districts in Cluster 3, and 4 sub-districts in Cluster 4. The association rule analysis produces four association rules using a minimum support of 10% and a minimum confidence of 50%. The findings highlight that the Ngasem and Bojonegoro sub-districts require more focused disaster management. The fourth association rule has the highest confidence level at 78.79%, indicating that forest and land fires are likely to follow when drought occurs. The research implies that it can support more targeted disaster management focusing on high-risk sub-districts such as Ngasem and Bojonegoro. The originality of the research lies in its novel application of clustering and association rules to analyze disaster patterns in the region, with implications for more targeted disaster mitigation strategies.

***Keywords*:** clustering and association, early warning, disasters, Bojonegoro Regency

I. INTRODUCTION

A disaster is a sequence of occurrences that impact people's lives and means of subsistence while causing misery to humans, property loss, environmental harm, and damage to infrastructure and facilities (Ahmed et al., 2022). It also refers to events or series of events that disrupt and endanger people's lives and means of subsistence caused by natural and man-made factors. They can result in fatalities, destruction of the environment, loss of property, and psychological repercussions (Morganstein & Ursano, 2020; Udori & Miranti, 2019). This definition is based on Article 1, Paragraph 1 of Law No. 24 of 2007 concerning disaster management. In Law No. 24 of 2007 concerning disaster management, Article 1, with paragraphs 2, 3, and 4, lists three categories of disasters: natural, non-natural, and social. Natural disasters include earthquakes, tsunamis, volcanic eruptions, floods, droughts, hurricanes, and landslides. Hence, planning is essential for any potential calamity that can happen at any time (Udori & Miranti, 2019). Due to its strategically located location at the meeting point of two continents and oceans, as well as the mountain paths that cut through it, Indonesia is frequently hit by natural disasters like earthquakes and volcanic eruptions. The disasters cause fatalities and extensive damage to the surrounding area, including infrastructure destruction and property losses (Murdiaty et al., 2020).

By analyzing disaster events and patterns discovered from disaster data, a process known as data mining, the impact of disasters may be lessened or avoided. Data mining is the process of gathering and analyzing historical data to identify patterns, norms, or connections in large amounts of data. This data mining output can help in making future decisions (Gupta & Chandra, 2020). Data mining can process disaster data, such as clustering and association rules.

A data mining method that is often used is clustering. It groups data points into two or more groups so that the data points belong to groups that have similarities rather than different groups based on the information available with the data points (Aggarwal, 2015). The application of clustering methods is easier with the help of RapidMiner software (Nurdiansyah et al., 2024). Many disaster studies use clustering methods to analyze natural disaster data using the K-means algorithm (Riasetiawan et al., 2022). Another research also uses K-means to analyze the distribution of disaster-prone points in disaster management (Prasetyadi et al., 2022). K-means is also commonly used in health studies such as clustering infectious diseases (Furuse et al., 2020). In previous research, the K-means method is superior to other clustering methods in extensive data studies, such as data on learning capacity and facilities for elementary schools in Bojonegoro Regency (Nurdiansyah et al., 2023) and population data in Bojonegoro Regency (Sholikhah, 2022).

Another data mining method that is rarely used is the association rule, which looks for relationships between one item and another item from a market basket analysis (Alawadh & Barnawi, 2022). The algorithm commonly used in association rule methods is the Apriori method. The Apriori method is one of the methods in data mining that find relationships between data based on their characteristics so that rules are formed based on clusters. The Apriori algorithm is a data mining algorithm with association rules to determine the association relationship of a combination of elements whose application meets the minimum support requirements specified (Das et al., 2021). The research with association rules with disaster data has been done with another algorithm, Frequent Pattern (FP)-growth (Wu & Zhang, 2023).

The research objective is to ascertain how sub-districts in Bojonegoro Regency are grouped according to the occurrence of disasters and to comprehend the implications of the association rule using the Apriori algorithm on disasters in Bojonegoro Regency. The research results are expected for the community to learn about disasters in Bojonegoro Regency, to prepare for future disasters, and to take part in environmental protection to reduce tragedies. The Bojonegoro Regency Regional Disaster Management Office can also take into consideration the research benefits when deciding on the next step in creating policies related to disaster management in Bojonegoro Regency.

Bojonegoro is a relevant location for disaster-related research due to its vulnerability to various types of natural disasters, particularly those triggered by geological activities, such as earthquakes and landslides, as well as hydrometeorological disasters like floods and droughts. The region’s geographic position, situated between mountainous areas and lowland plains, along with the increasingly unpredictable effects of climate change, heightens its susceptibility to such disasters. Additionally, Bojonegoro is rich in natural resources, including energy reserves and water sources, which can influence the intensity of disasters, making it an essential area of study for understanding the dynamics of risk mitigation and adaptation to natural disasters in regions, lying at the intersection of rural and urban areas.

In the research, descriptive statistics using frequency distribution tables and mapping are carried out as an update, and two data mining methods are applied, namely the clustering method (K-means) and the associative method (Apriori) in extracting disaster data in Bojonegoro Regency. Although it is impossible to predict when a disaster will strike, it is possible to determine when a disaster is likely to occur by examining the relationships between different disasters. When people, like in Bojonegoro Regency, need more information and awareness of disasters, it may be difficult to foresee when disasters will occur. Thus, the research is necessary to calculate disasters more easily.

II. METHODS

A quantitative approach is applied in the research. The proposed statistical methods are frequency distribution, distribution mapping, clustering with the K-means algorithm, and association rule with the Apriori algorithm. Data are analyzed with the help of RapidMiner and Quantum Geographic Information System (QGIS) software. Data collection is carried out using a random sampling technique. The research uses secondary data of disaster data originating from the National Disaster Management Agency and the Bojonegoro Regency Regional Disaster Management Agency Office. There are eight types of disasters in Bojonegoro Regency in the range of January 5, 2019, to January 18, 2023, with a total of 2,078 events occurring in 589 days. It includes house fires, extreme weather, overflowing floods, landslides, flash floods, forest and land fires, drought, and other events.

The research uses variables with different measurement scales for each analysis. The variables used are presented in Tables 1 and 2. Table 1 presents the definitions of the research variables applied in the Apriori Algorithm, including the variable names, data measurement scales, and descriptions provided for each research variable. Meanwhile, Table 2 outlines the definitions of the research variables used in the K-means algorithm and mapping. It also includes the variable names, data measurement scales, and descriptions provided for each research variable.

Then, the research uses three different analysis methods to obtain complete results. At the beginning of the analysis, a frequency distribution table is provided to simplify the data, followed by a visualization of the disasters that occur in Bojonegoro Regency. This approach allows for a clear understanding of the patterns and trends within the dataset. The research aims to identify significant correlations and insights that can guide disaster management strategies using various analysis techniques. The combination of quantitative and qualitative methods ensures a comprehensive analysis of the disaster data, contributing to more effective policy recommendations.

Table 1 Definition of Research Variables for the Apriori Algorithm

Variable	Scale Measurement	Description
Sub-districts	Nominal	The sub-districts in Bojonegoro Regency include Ngraho, Tambakrejo, Ngambon, Ngasem, Bubulan, Dander, Sugihwaras, Kedungadem, Kepohbaru, Baureno, Kanor, Sumberrejo, Balen, Kapas, Bojonegoro, Kalitidu, Malo, Purwosari, Padangan, Kasiman, Temayang, Margomulyo, and Trucuk.
Disaster_1	Ratio	Number of house fires
Disaster_2	Ratio	Number of extreme weather events
Disaster_3	Ratio	Number of overflow flood events
Disaster_4	Ratio	Number of landslide events
Disaster_5	Ratio	Number of flash flood events
Disaster_6	Ratio	Number of forest and land fire events
Disaster_7	Ratio	Number of drought events
Disaster_8	Ratio	Number of other events

Table 2 Definition of Research Variables for K-Means Algorithm and Mapping

Variable	Scale Measurement	Description
Date	Nominal	Date of disaster
Disaster_1	Ratio/Binary	0 = No house fire 1 = House fire occurs
Disaster_2	Ratio/Binary	0 = No extreme weather 1 = Extreme weather occurs
Disaster_3	Ratio/Binary	0 = No overflow flooding 1 = Overflow flooding occurs
Disaster_4	Ratio/Binary	0 = No landslide 1 = Landslide occurs
Disaster_5	Ratio/Binary	0 = No flash flooding 1 = Flash flood occurs
Disaster_6	Ratio/Binary	0 = No forest and land fires 1 = Forest and land fires occur
Disaster_7	Ratio/Binary	0 = No drought 1 = Drought occurs
Disaster_8	Ratio/Binary	0 = No various events 1 = Various events occur

The clustering method used is the K-means method. Based on the Elbow method approach, it is commonly used to generate information to determine the best number of clusters. The data clustering analysis follows several structured steps. First, data transformation is performed to prepare the dataset for clustering. Second, the number of clusters or groups to be formed is determined by setting the value of *k*, where the initial cluster centers (centroids) are randomly selected from the data. Third, the Euclidean distance formula is then applied to calculate the distance between each data point and the centroids, allowing for the identification of the closest centroid. Each data point is classified into its corresponding cluster based on the smallest distance. Last, after classification, the centroid values are updated using the average of the data points in each cluster, with the updated centroid calculated using Equation (1). The μ_k represents the centroid of the *k*-th cluster. Then, N_k is the number of data points in the *k*-th group, and x_q is the *q*-th data point in that cluster. These steps (2 through 5) are repeated iteratively until the cluster memberships remain unchanged, indicating that the clustering process has converged. Once the iteration stops, the final data classification is determined using the cluster structure formed in the last iteration.

$$\mu_k = \frac{1}{N_k}\sum_{q=1}^{N_k} x_q. \quad (1)$$

The association rule method in the research utilizes the Apriori algorithm, following several key steps. First, the data are transformed into a binary format. Then, observational data containing only one event per observation are cleaned. Afterward, the minimum support and confidence values are determined. The next step involves calculating the support value for itemsets of size $k = 1$ in the database and generating candidate rules. The support value is calculated using the Equation (2).

$$Support(A) = (\text{Number of observations containing } A)/(\text{Total observations}) \times 100\%. \quad (2)$$

Equation (2) is used in the Apriori algorithm to measure the frequency of an itemset (*A*) within a dataset. Support indicates how often an itemset appears in the dataset, which is related to the total number of observations. Specifically, it calculates the proportion of transactions that contain the itemset *A*, expressed as a percentage. A higher support value means that the itemset is more frequent or popular in the dataset, making it more significant for identifying association rules. This metric helps to filter out less frequent patterns and focus on those that are more likely to be relevant for decision-making.

Next, candidate sets with support values below the specified threshold are pruned. The most frequently occurring itemsets are then combined to form larger sets (size of $k + 1$). This process is repeated until no further itemsets are formed. Once the more significant itemsets are created, the confidence value is calculated using Equation (3). Candidate sets with lower confidence values than the threshold is pruned. Finally, conditional association rules are defined by setting the conditions and results.

$$Confidence(AB) = (Support(AB))/(Support(A)). \quad (3)$$

Equation (3) is also used in the Apriori algorithm to assess the strength of an association rule between two itemsets, *A* and *B*. It calculates the likelihood that itemset *B* will appear in a transaction given that itemset *A* is already present. The *Support(AB)* represents the proportion of transactions containing both *A* and *B*, while *Support(A)* is the proportion of transactions containing *A*. Confidence reflects the conditional probability of observing *B* when *A* occurs. A higher confidence value indicates a stronger association between the two itemsets, helping to identify more reliable and relevant patterns in the dataset.

III. RESULTS AND DISCUSSIONS

Before proceeding with further analysis using clustering and association methods, a frequency distribution analysis is conducted. The data displayed on the Bojonegoro Regency Regional Disaster Management Agency Office website comprises a series of disaster events that occurred in Bojonegoro from January 2019 to January 2023. However, the data presented are not very informative. It is challenging to be analyzed further. Therefore, a frequency distribution table is necessary to simplify data presentation, making it easier to read and analyze. The results of the frequency distribution are shown in Table 3.

In Table 3, the highest number of house fire disasters occurs in the Ngasem Sub-District, with 156 incidents, and the lowest number of house fire disasters is in the Kedewan Sub-District, with 13 incidents. Then, the highest number of extreme weather events is in the Ngasem Sub-District, with 146 events, and the lowest occurs in the Kedewan Sub-District, with 12 events. Meanwhile, the highest number of floods occurs in the Ngasem Sub-District, with 161 events, and the lowest is in the Kedewan Sub-District, with 14 events. Next, the highest number of landslides is in the Ngasem Sub-District, with 148, and the lowest number is in the Kedewan Sub-District, with 10.

Moreover, the highest number of flash floods is in the Ngasem Sub-District, with 162 incidents, and the lowest flash floods occur in the Kedewan Sub-District, with 14 incidents. Then, the highest incidence of forest and land fires happens in the Ngasem Sub-District, with 155 incidents, and the lowest incidence is in the Kedewan Sub-District, with 13 incidents. Meanwhile, the highest incidence of drought happens in the Bojonegoro Sub-District, with 142 events, and the lowest incidence occurs in the Kedewan Sub-District, with 8 events. The highest incidence of other disasters occurs in the Ngasem Sub-District, with 159

Table 3 Results of Frequency Distribution of Disaster Events in Bojonegoro Regency

Sub-District	Disaster_1	Disaster _2	Disaster_3	Disaster_4	Disaster_5	Disaster_6	Disaster_7	Disaster_8
Balen	54	40	60	61	68	67	68	60
Baureno	48	41	54	48	57	48	57	51
Bojonegoro	128	73	124	132	142	133	142	135
Bubulan	57	50	56	59	59	56	16	59
Dander	95	74	82	96	102	96	84	99
Gayam	17	12	25	25	25	24	25	24
Gondang	36	35	25	23	34	38	38	37
Kalitidu	40	24	46	50	50	49	50	42
Kanor	48	36	47	40	53	50	53	41
Kapas	80	52	75	85	87	83	64	84
Kasiman	64	67	68	61	68	66	18	66
Kedewan	13	13	14	10	14	13	8	13
Kedungadem	65	67	72	58	72	59	46	71
Kepohbaru	46	41	45	49	51	48	30	49
Malo	29	25	32	24	33	29	27	30
Margomulyo	14	12	16	13	16	14	14	13
Ngambon	37	25	37	33	39	37	27	38
Ngasem	156	146	161	148	162	155	46	159
Ngraho	126	115	130	118	132	122	43	128
Padangan	54	43	59	57	60	52	37	58
Purwosari	64	54	64	62	64	61	21	60
Sekar	45	36	40	17	45	45	45	42
Sugihwaras	105	105	110	102	110	101	22	109
Sukosewu	65	58	60	66	71	62	40	69
Sumberrejo	148	120	146	147	148	142	34	147
Tambakrejo	138	114	133	140	141	134	42	139
Temayang	108	102	107	107	113	103	40	113
Trucuk	50	29	53	56	56	48	54	48

events. The lowest incidence of other tragedies is in Kedewan and Margomulyo Sub-Districts, with 13 events.

Based on the data obtained, the Ngasem Sub-District has the highest number of different types of disasters, such as house fires (156 events), extreme weather (146 events), flash floods (161 events), and landslides (148 events). In contrast, Kedewan Sub-District records the lowest number of all these disasters, with 13 house fires, 12 extreme weather events, 14 flash floods, and 10 landslides. Similar trends are seen in forest and land fires and drought, with the Ngasem Sub-District again as the highest number and Kedewan Sub-District as the lowest. Drought is highest in Bojonegoro Sub-District (142 events), while the other disasters are most prevalent in Ngasem Sub-District (159 events). It shows that Ngasem Sub-District is more vulnerable to various types of disasters than Kedewan sub-district.

Research that classifies disaster data by region shows that clustering approaches such as K-means can be used to group regions based on the frequency and type of disaster (Murdiaty et al., 2020; Nurdiansyah et al., 2024). Modeling using this clustering algorithm is very helpful in disaster mitigation and regional policy planning related to natural disaster management (Prasetyadi et al., 2022). The data can support preventive measures in the most vulnerable areas, such as Ngasem Sub-District, where disaster events occur more frequently.

Next, the research uses clustering to group all disasters in each Sub-District in Bojonegoro Regency. The Elbow technique, which looks at the proportion of results between the number of groups that will form an elbow at a point (first sloping), is used to identify the number of clusters. The comparison of each cluster is to get the best group using the average within the cluster size. The number of sets to be tested is from two classes to ten classes. The results of determining the number of clusters with the Elbow method are presented in Table 4. Table 4 shows that the first significant change in the Average Within Cluster value, indicating a first sloping, occurs at 4 clusters. The difference in the Average Within Cluster values of -25.806 and -12.158 is much larger compared to the differences observed between the values of other clusters.

Table 4 Results of Average Within Cluster

Number of Clusters	Average Within Cluster
2	-29.016
3	-25.806
4	-12.158
5	-9.187
6	-8.395
7	-4.016
8	-2.88
9	-2.422
10	-1.708

The research uses clustering techniques of the K-means model to group disasters in 28 sub-districts in Bojonegoro Regency. The Elbow technique determines the best number of clusters by looking at the "elbow" point on the average within the cluster graph. The calculation results show that the optimal number of clusters is five, corresponding to the fifth cluster's first sloping line analysis. Thus, the number of groups chosen is five classes, which is the best number of sets. Furthermore, the operation is carried out with the help of RapidMiner. The results are shown in Table 5. Of the 28 sub-districts, the clustering results show that Cluster 0 consists of 13 sub-districts, Cluster 1 consists of 6 sub-districts, Cluster 2 consists of only 1 sub-district, and Cluster 3 and Cluster 4 have 4 sub-districts each.

The data in Bojonegoro Regency is divided into 5 groups of 28 sub-districts using the K-means model. Cluster 0 has the most group members, namely 13 sub-districts: Balen, Baureno, Bubulan, Kalitidu, Kanor, Kasiman, Kedungadem, Kepohbaru, Padangan, Purwosari, Sekar, Sukosewu, and Trucuk. Then, Cluster 1 has 6 sub-district members: Gayam, Gondang, Kedewan, Malo, Margomulyo, and Ngambon. Cluster 2 only has 1 member: Bojonegoro sub-district. Then, Cluster 3 has 4 members: Ngasem, Ngraho, Sumberrejo, and Tambakrejo sub-districts. Similarly, Cluster 4 has 4 members: Dander, Kapas, Sugihwaras, and Temayang sub-districts. The thematic map of the cluster results is shown in Figure 1.

Table 5 Output of Cluster Model Using K-Means

Cluster Model	
Cluster 0	13 items
Cluster 1	6 items
Cluster 2	1 item
Cluster 3	4 items
Cluster 4	4 items
Total number of items: 28	

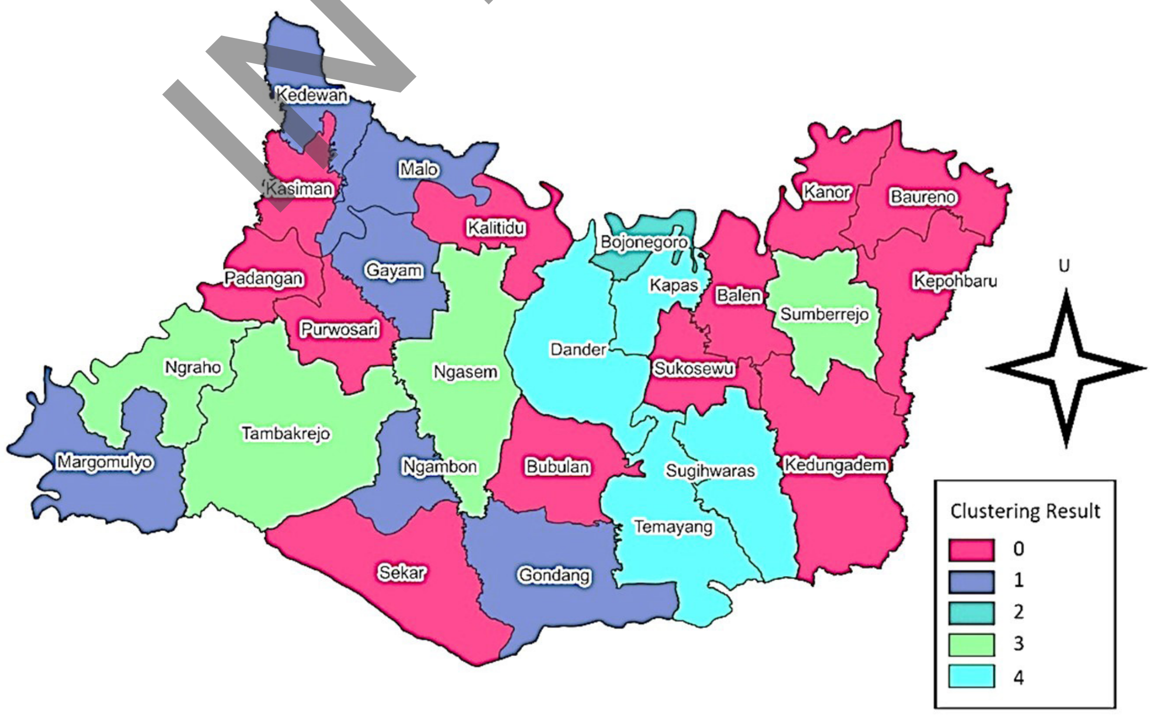

Figure 1 Mapping of Clustering Results

The approach used is in line with previous research that uses the K-means algorithm and Elbow technique to identify the best number of clusters in the classification of disaster data or other natural phenomena (Murdiaty et al., 2020; Nurdiansyah et al., 2024). Grouping areas based on the occurrence of disasters can help to formulate disaster mitigation policies and resource planning at the regional level. The utilization of Elbow and Average Within Cluster techniques to determine the optimal number of clusters is also in line with the approach used in various studies on disaster risk mitigation and data-driven clustering (Prasetyadi et al., 2022).

In association rule with the Apriori algorithm, the formation of one itemset of cleaned disaster data with a total of 118 observations is calculated with a minimum support of 10%. From the eight types of disasters in Table 6, there is one type of disaster that does not meet the minimum support, namely Disaster_5, with a minimum clearance of 0.85%.

The formation of two itemsets is done by a cross-item process on one itemset, which is calculated as the support value. Eight pairs of itemsets meet the minimum support, as shown in Table 7. Table 7 shows that the support values of the two itemsets meet the required support threshold. These itemsets indicate significant patterns that appear frequently within the dataset. The high support values suggest that the items in these sets are strongly associated, which can be useful for generating reliable association rules.

The potential of association rule forming is then ascertained by searching each itemset's confidence value for sets of items that have fulfilled the support. The value of 50% is the minimal level of confidence that has been established. As shown in Table 8, four association rules have been developed.

In Table 8, the association rules that are developed are as follows. First, if Disaster_3 (flash flood) occurs, there will be Disaster_4 (landslide) with a probability level of 56.52%. Second, if Disaster_3

Table 6 Support Value of One Itemset

Type of Disaster	Number of Events	Support
Disaster_1	29	24.58%
Disaster_2	47	39.83%
Disaster_3	26	22.03%
Disaster_4	34	28.81%
Disaster_6	49	41.53%
Disaster_7	33	27.97%
Disaster_8	28	23.73%
Total of Events	118	

Table 7 Support Value of Two Itemsets

Names of Itemset	Support Item	Support Itemset
Disaster_2, Disaster_3	34.75%	11.02%
Disaster_2, Disaster_4	34.75%	11.86%
Disaster_4, Disaster_2	26.27%	11.86%
Disaster_4, Disaster_3	26.27%	11.02%
Disaster_3, Disaster_4	19.49%	11.02%
Disaster_3, Disaster_2	19.49%	11.02%
Disaster_6, Disaster_7	38.98%	22.03%
Disaster_7, Disaster_6	27.97%	22.03%

Table 8 Results of Confidence Calculation

Association Rule	Support	Confidence
If Disaster_3 occurs, then Disaster_4 occurs	11.02%	56.52%
If Disaster_3 occurs, then Disaster_2 occurs	11.02%	56.52%
If Disaster_6 occurs, then Disaster_7 occurs	22.03%	56.52%
If Disaster_7 occurs, then Disaster_6 occurs	22.03%	78.79%

(flash flood) occurs, there will be a probability level of 56.52% that Disaster_2 (extreme weather) will occur. Next, Disaster_7 (drought) will occur 56.52% of the time if Disaster_6 (forest and land fires) occurs. Disaster_7 (drought) will occur 78.79% of the time if Disaster_6 (forest and land fires) occurs. It is the fourth association rule. Drought and forest and land fires are two disasters that frequently occur jointly. They have a support value of 22.03% and a confidence level of 78.79%, indicating a 22.03% dominance rate for both itemsets. The odds of two itemsets happening are 78.79%.

The association rule mining approach used is similar to previous studies that utilize the Apriori algorithm to find association patterns in disaster data. For example, the Apriori algorithm is applied in e-commerce (Das et al., 2021). Then, association rules are used for market basket analysis (Alawadh & Barnawi, 2022). Both studies demonstrate how this technique can be applied to find associations between items or events, including in the context of natural disasters.

IV. CONCLUSIONS

Based on the research results, the Bojonegoro Sub-District has the highest frequency of drought disasters, while the Kedewan Sub-District has the lowest frequency. Then, the Ngasem Sub-District has the highest frequency of other tragedies, while the Kedewan and Margomulo Sub-Districts have the lowest frequency of six other disaster events. The Kedewan Sub-District has the lowest occurrence, and the Ngasem Sub-District has the highest. Then, the data in the Bojonegoro Regency is divided into five groups of 28 sub-districts based on the results of the K-means model. There are 13 sub-districts in Cluster 0, 6 sub-districts in Cluster 1, 1 sub-district in Cluster 2, 1 sub-district in Cluster 3, and 4 sub-districts in Cluster 4. The research results are limited to the scope of disaster and mitigation analysis in Bojonegoro Regency. Further research can be conducted in other regions for future studies to explore similar aspects.

The association analysis produces four association rules results with a minimum support of 10% and a minimum confidence of 50%. From the data of the results of the rules that have been obtained, it can be seen that disasters often occur simultaneously. The findings obtained in disaster events in Bojonegoro Regency conclude that flash floods have a significant probability of occurring together with landslides and extreme weather. In addition, forest and land fires have a strong relationship with drought, which reflects the pattern of frequent co-occurrence.

Future research can use observation data with the latest year to provide more up-to-date information related to disaster events that occur in Bojonegoro Regency. It can also use village-scale data to provide more accurate information related to disaster events that occur in Bojonegoro Regency. Moreover, the Bojonegoro District Regional Disaster Management Agency should pay special attention to the sub-districts with a high disaster level so that it can be anticipated. When a disaster occurs, it can be handled immediately so as not to cause another disaster. Future research can also be carried out in a similar way with data from every district in Indonesia because the findings obtained later may be the same or different.

REFERENCES

Aggarwal, C. C. (2015). Data classification. In *Data mining* (pp. 285–344). Springer. https://doi.org/10.1007/978-3-319-14142-8_10

Ahmed, M. A., Mohamed, M. H., Parvin, M. M., & Ilić, P. (2022). The recurrence of natural disasters in Jowhar, Middle Shabelle Region, Somalia: The causes and impacts. *Journal of Environmental Protection*, *13*(9), 657–670. https://doi.org/10.4236/jep.2022.139042

Alawadh, M. M., & Barnawi, A. M. (2022). A survey on methods and applications of intelligent market basket analysis based on association rule. *Journal on Big Data*, *4*(1), 1–25. https://doi.org/10.32604/jbd.2022.021744

Das, A., Jana, S., Ganguly, P., & Chakraborty, N. (2021). Application of association rule: Apriori algorithm in e-commerce. In *2021 Innovations in Energy Management and Renewable Resources* (pp. 1–7). IEEE. https://doi.org/10.1109/IEMRE52042.2021.9386737

Furuse, Y., Sando, E., Tsuchiya, N., Miyahara, R., Yasuda, I., Ko, Y. K., ... & Oshitani, H. (2020). Clusters of coronavirus disease in communities, Japan, January–April 2020. *Emerging Infectious Diseases*, *26*(9), 2176–2179. https://doi.org/10.3201/eid2609.202272

Gupta, M. K., & Chandra, P. (2020). A comprehensive survey of data mining. *International Journal of Information Technology*, *12*, 1243–1257. https://doi.org/10.1007/s41870-020-00427-7

Morganstein, J. C., & Ursano, R. J. (2020). Ecological disasters and mental health: Causes, consequences, and interventions. *Frontiers in Psychiatry*, *11*, 1–15. https://doi.org/10.3389/fpsyt.2020.00001

Murdiaty, Angela, & Sylvia, C. (2020). Pengelompokkan data bencana alam berdasarkan wilayah, waktu, jumlah korban dan kerusakan fasilitas dengan algoritma K-Means. *Jurnal Media Informatika Budidarma*, *4*(3), 744–752. https://doi.org/10.30865/mib.v4i3.2213

Nurdiansyah, D., Ma'ady, M. N. P., Sukmawaty, Y., Utomo, M. C. C., & Mutiani, T. (2024). Clustering analysis for grouping sub-districts in Bojonegoro District with the K-Means method with a variety of approaches. *BAREKENG: Journal of Mathematics and Its Application, 18*(2), 1095–1104. https://doi.org/10.30598/barekengvol18iss2pp1095-1104

Nurdiansyah, D., Saidah, S., & Cahyani, N. (2023). Data mining study for grouping elementary schools in Bojonegoro Regency based on capacity and educational facilities. *BAREKENG: Journal of Mathematics and Its Applications*, *17*(2), 1081–1092. https://doi.org/10.30598/barekengvol17iss2pp1081-1092

Prasetyadi, A., Nugroho, B., & Tohari, A. (2022). A hybrid K-Means hierarchical algorithm for natural disaster mitigation clustering. *Journal of Information and Communication Technology*, *21*(2), 175–200. https://doi.org/10.32890/jict2022.21.2.2

Riasetiawan, M., Ashari, A., & Wahyu, P. (2022). The performance evaluation of K-Means and agglomerative hierarchical clustering for rainfall patterns and modelling. In *2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)* (pp. 431–436). https://doi.org/10.1109/ICITISEE57756.2022.10057729

Sholikhah, N. A. (2022). Studi perbandingan clustering kecamatan di Kabupaten Bojonegoro berdasarkan keaktifan penduduk dalam kepemilikan dokumen kependudukan. *Jurnal Statistika dan Komputasi*, *1*(1), 42–53. https://doi.org/10.32665/statkom.v1i1.443

Udori, A., & Miranti. (2019). Upaya Badan Penanggulangan Bencana Daerah (BPBD) dalam penanggulangan bencana banjir. *Jurnal Politik dan Pemerintahan Daerah*, *1*(2), 85–94. https://doi.org/10.36355/jppd.v1i2.8

Wu, Y., & Zhang, J. (2023). Retraction note: Building the electronic evidence analysis model based on association rule mining and FP-growth algorithm. *Soft Computing*, *27*. https://doi.org/10.1007/s00500-022-07709-1