JURNAL UPI

Jurnal Pendidikan Multimedia (EDSENCE)

Journal homepage: https://ejournal.upi.edu/index.php/Edsence

Comparative Analysis of K-Means and K-Medoids Clustering Methods on Weather Data of Denpasar City

Bima Prihasto [1], Darmansyah [2], Dastin Pranata Yuda [3], Fauzan Maftuh Alwafi [4], Herliana Nur Ekawati [5], and Yustika Perwita Sari [6]

[1,2,3,4,5,6] Informatics Department, Institute Kalimantan Technology, Balikpapan, Indonesia

Correspondence : E-mail: bima@lecturer.itk.ac.id

ABSTRACT

By applying data mining methods, particularly clustering techniques, the weather data of Denpasar city can be grouped based on similar characteristics. This provides deep insight into weather patterns, useful for more optimised travel planning. This research positively impacts tourism, helping stakeholders understand weather patterns in more detail. Furthermore, in-depth knowledge of weather conditions improves preparedness for potential global climate change. The clustering results can be visualised in a three-dimensional cartesian diagram, providing a clear picture of the various weather conditions using attributes such as temperature, precipitation, and humidity. Through Kaggle's Denpasar Weather Data dataset, with 264,924 data and 32 columns, this study illustrates that cloudy weather dominates in the K-Means and K-Medoids clustering process on rain data within one hour. At three hours, K-Means shows the dominance of cloudy weather and the possibility of rain, while K-means dominates all clusters. At six hours, K-Means dominate in sunny and rainy weather, while K-Medoids dominate evenly in all clusters.

ARTICLE INFO

Article History:
Submitted/Received 29 September 2023
First Revised 7 December 2023
Accepted 15 December 2023
First Available online 31 December 2023
Publication Date 31 December 2023

Keywords:
K-Means,
K-Medoids,
Clustering,
Weather.

1. BACKGROUND

Weather is an atmospheric phenomenon that occurs at certain times and locations and can change due to several factors. These factors include the amount of rainfall, humidity level, duration of sunlight, maximum temperature, minimum temperature, and average temperature. In other words, weather describes the state of the air at a particular time in an area and is one of the elements that affects climate. The Meteorology, Climatology and Geophysics Agency (BMKG) is a government agency in Indonesia tasked with monitoring the weather and providing important information for weather and climate-related purposes. Weather information is essential in many areas of life, including public safety, socioeconomics, agricultural production, plantations, fisheries, aviation, and other fields. BMKG is the primary source of weather information that significantly impacts many activities and sectors of life in Indonesia. Data Mining is an approach to data processing that aims to identify hidden patterns in the data being processed. Through data mining techniques, new information can be generated from existing data, and the results of this analysis can be used to make future decisions. Thus, data mining helps create knowledge rooted in historical data, which can then be applied for better decision-making in the future. One of the valuable data mining techniques to group weather data based on the level of similarity is the clustering method. Clustering is an approach used to identify and group data with similar characteristics. Weather data can be grouped using this technique based on similarity with other weather data.

Applying clustering techniques in weather analysis becomes even more relevant when looking at tourism, such as Denpasar and Bali. Denpasar is a popular tourist destination with weather that tends to be stable throughout the year. Still, certain variations in the weather can affect the tourist experience. With the application of data mining, namely the clustering method, the weather data of Denpasar City can be grouped based on the similarity of weather characteristics, such as rainfall, temperature and humidity. This is beneficial in the tourism sector as it can provide greater insight into weather patterns, which can be used for more optimised travel planning. Thus, data mining through the weather clustering method can be an essential tool to improve the experience and management of tourism destinations. K-Means is a popular clustering technique with many applications, including weather analysis. In the context of this research, K-Means clustering can be utilised to classify weather patterns based on historical data, including temperature, humidity, pressure, wind speed, wind direction, and percentage of cloud coverage. K-Medoids is a clustering algorithm technique used to divide data into several groups or clusters based on the similarity between the data elements. This algorithm is a partitioning algorithm family where data is divided into several groups. The main difference between K-Medoids and K-Means, a more common clustering algorithm, is how the group centres (centroids) are calculated. In K-Means, the group centre is the arithmetic mean of the elements in the group, whereas, in K-Medoids, the group centre is chosen as one of the data points, which is a more robust representation of outliers or noise.

The results of these two clustering methods can provide valuable information about weather conditions in Denpasar. This research is essential in the context of global climate change, which can affect regional weather patterns. By understanding more about the weather conditions in Denpasar and Bali, stakeholders and local governments will be better prepared to deal with potential weather changes that may occur in the future. Therefore, this

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

research is expected to positively impact the economic sustainability, comfort and safety of Denpasar City residents and tourists visiting the region.

2. METHODS

In this research, several stages will be carried out with the following image:

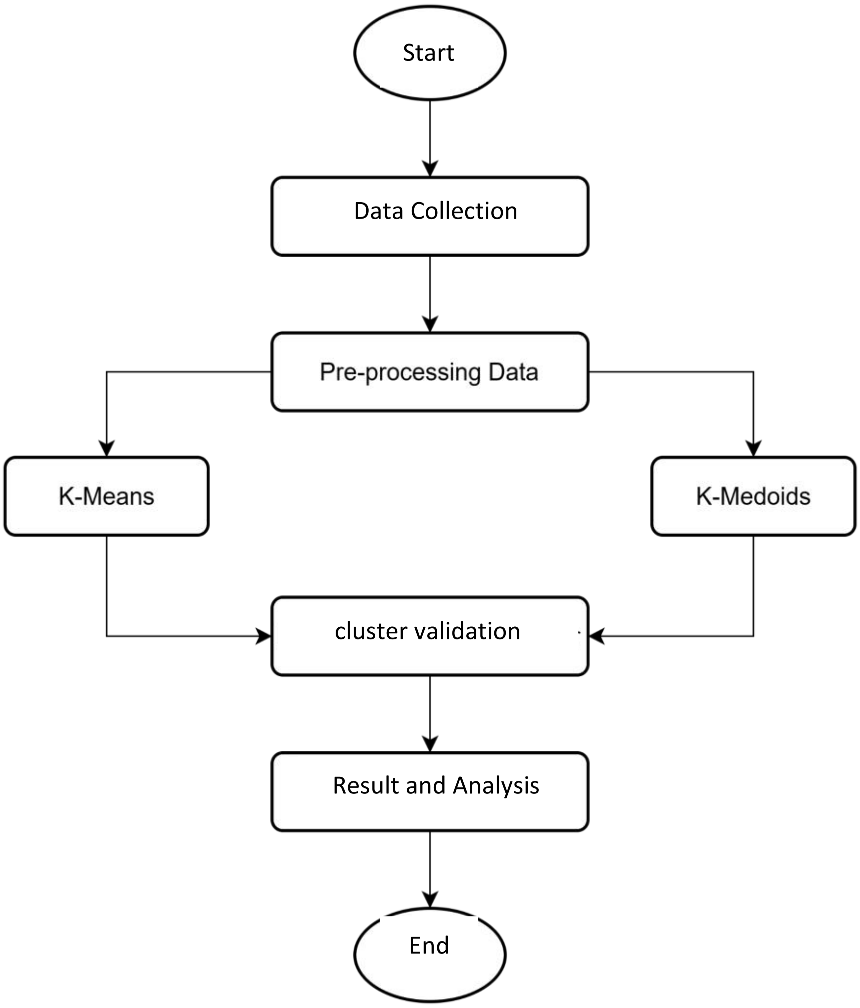

Figure 1. Research Methodology

2.1. Clustering

Clustering is one of the methods in data mining that uses similarity matrices to find naturally occurring groups or clusters in multidimensional data (Jain, 1999; Jain, 2000). It is essential in pattern recognition and machine learning (Hamerly & Elkan, 2002). In addition, data clustering is a critical technique in Artificial Intelligence (AI) (Hamerly, 2003). Image segmentation (Coleman & Andrews, 1979; Jain & Dubes, 1988), quantisation of vector and colour images (Kaukoranta, 1998; Baek, 1998; Xiang, 1997), data mining, compression, machine learning, and other applications all use clustering techniques. The cluster centre (or

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

centroid) is usually used to identify a cluster (Lee & Antonsson, 2000). Data clustering is a formidable challenge in unsupervised pattern recognition because data clusters can be of various shapes and sizes (Jain, 2000).

2.2. Data Preparation

Data Preparation is preparing raw data to be processed and analysed. This stage is critical as the first step in data analysis to eliminate bias caused by poor quality. Some necessary steps in data preparation are:

1. Acquiring data: Determining the data needed, collecting it, and ensuring consistent access
2. Discovering and assessing data: Understanding data, its quality, and its limitations.
3. Cleaning data: Removing or correcting errors, inconsistencies, and inaccuracies.
4. Transforming data: Transforming the data into a format that can be used for analysis.
5. Validating data: Ensuring data is accurate, complete, and consistent.
6. Integrating data: Combining data from multiple sources.

2.3. Algorithm ***K-Means Clustering***

K-Means Clustering Algorithm is a data analysis method for grouping data into several groups called clusters. Objective: The main thing about the K-Means algorithm is that for a shared data set, several such groups appear so that data points in the same group have similarities and inter-data points groups have significant differences. K-Means begins with determining the desired number of clusters. Algorithm: Then, find the position of the centre of each cluster, called the centroid. Election initial centroid can be done randomly or use a technique initialisation. Following is an explanation step by step of the K-Means algorithm:

1. Determine the random value for the initial cluster centre (centroid) with the Elbow Method, which determines an optimal number of clusters in the K-Means algorithm. The elbow Method helps determine an optimal number of clusters and avoids overfitting or underfitting the algorithm.
2. Classify each data based on its cosess with the centroid (distance smallest).
3. Repeat the steps above so that the members of each cluster do not change.

$$oneJ = \sum_{i=1}^{k}\sum_{j=1}^{n}(||x_i - v_j||)^2 = 1 \qquad (2.1)$$

$||x_i - v_j||$ is the Euclidean value between the distance of a point is $x_i$and a centroid is v_j, this used For determine the K value will be used on clusters. K-Means Clustering is a technique that divides data into one or more clusters, with the same data grouped in one cluster and diverse data shared becoming several clusters. Because it is an easily used algorithm, This is the most widely used. Disadvantages of the algorithm This is highly dependent on cluster initialisation.

2.4. K-Medoids Clustering Algorithm

The k-Medoids clustering algorithm is a technique for purposeful data grouping. It is used for partitioning a dataset into k groups (clusters) based on the similarity between data. Unlike

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

the K-Means algorithm that uses the centre gravity of deep data one cluster as cluster centre (centroid), K-Medoids uses actual data from a dataset called "medoid". A medoid is a member of the cluster that owns minimum distance or is nearest to all other members in the cluster. Distance between data points can usually be measured using metrics like Euclidean, Manhattan, or distance metrics other than data types and needs. K-medoids are often used in application grouping where necessary accurate cluster representation or when using centroids is inappropriate Because of the data.

$$C = \sum_{Ci} \sum_{Pi \epsilon Ci} |Pi - Ci| \qquad (2.2)$$

Equation 2.2 is an equation in the K-Medoids clustering algorithm, where C represent the centroid point on the cluster, C_i represents the centroid point in the ith cluster, cap P i. represents the data point in the ith cluster and absolute, and absolute values how distance between data points and centrescluster. Algorithm K-Medoids *Clustering* is related to groupings with K-Means clustering with share something data set becomes group or clusters. In the K-Medoids grouping, each cluster is personified by one existing point. That data point is considered a Medoid *Cluster.* The K-Medoids grouping is considered an alternative to K-Means clustering. Algorithm This is considered not prone enough to noise and outliers because, using medoid or point centre, the clusters are different from K-Means, which uses the mean as the centre.

3. RESULTS AND DISCUSSION

Research that has been carried out obtained the results of weather clustering in Denpasar City, Bali, using K-Means and K-Medoids methods. From the results and discussion, this shared becomes three parts in calculating the *clustering.* Calculation of the use reference from one The features in the dataset include rain_1h, rain_3h, and rain_6h. Calculation *clustering* uses reference temp, rain_3h, and humidity attributes. So, the results of the clustering visualisation will be plotted as a Cartesian diagram of three dimensions with points x, y, and z. Data and visualisation results will be obtained through datasets accessed on the Kaggle platform with Denpasar Weather Data, with 264924 data and 32 columns available, as seen in Figure 2.

With removing feature weather snow such as snow_1h, snow_3h, snow_6h, snow_12h, snow_24h, and snow_today, research focuses on more aspects relevant to determining *clustering* in Denpasar City. In context, the first clustering is carried out, namely clustering in bulk Rain in 1 hour; features used are on the index to 6, 7, 8, 9, 12, 13, 14, 15, 27. Where it only focuses on rain data in 1 hour. Next on context, the second *clustering* is carried out, namely *clustering* on rain data in 3 hours; features used are on the index to 6, 7, 8, 9, 12, 13, 14, 16, and 27, where only focuses on rain data period time for 3 hours. Then, deep context final *clustering,* i.e. clustering in bulk Rain in 6 hours, features used on the index to 6, 7, 8, 9, 12, 13, 14, 17, 27, which only focuses on rain data period time for 6 hours.

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 264924 entries, 0 to 264923
Data columns (total 32 columns):
 #   Column               Non-Null Count   Dtype  
---  ------               --------------   -----  
 0   dt                   264924 non-null  int64  
 1   dt_iso               264924 non-null  object 
 2   timezone             264924 non-null  int64  
 3   city_name            264924 non-null  object 
 4   lat                  264924 non-null  float64
 5   lon                  264924 non-null  float64
 6   temp                 264924 non-null  float64
 7   temp_min             264924 non-null  float64
 8   temp_max             264924 non-null  float64
 9   pressure             264924 non-null  float64
 10  sea_level            0 non-null       float64
 11  grnd_level           0 non-null       float64
 12  humidity             264924 non-null  int64  
 13  wind_speed           264924 non-null  float64
 14  wind_deg             264924 non-null  int64  
 15  rain_1h              16286 non-null   float64
 16  rain_3h              16156 non-null   float64
 17  rain_6h              36098 non-null   float64
 18  rain_12h             12 non-null      float64
 19  rain_24h             6817 non-null    float64
 20  rain_today           0 non-null       float64
 21  snow_1h              0 non-null       float64
 22  snow_3h              0 non-null       float64
 23  snow_6h              0 non-null       float64
 24  snow_12h             0 non-null       float64
 25  snow_24h             0 non-null       float64
 26  snow_today           0 non-null       float64
 27  clouds_all           264924 non-null  int64  
 28  weather_id           264924 non-null  int64  
 29  weather_main         264924 non-null  object 
 30  weather_description  264924 non-null  object 
 31  weather_icon         264924 non-null  object 
dtypes: float64(21), int64(6), object(5)
```

Figure 2. Denpasar Weather Data dataset

3.1 TEST K-Means and K-Medoids on The Rain_1H Attribute

visualisation results from the research K-Means clustering have been done based on the dataset; it will explained in the chapter. Next is the calculation of K -Means trials and Medoids trials.

3.1.1 Test K -Means on the rain_1h attribute

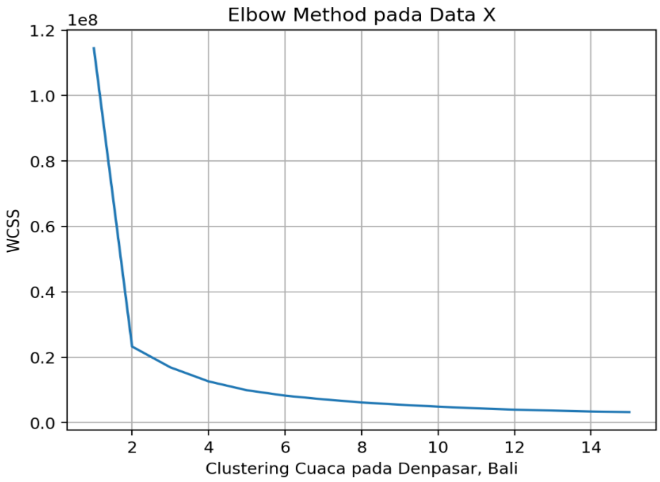

Figure 3. Visualization Chart *Elbow Method* in the rain Period 1 hour

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

The visualisation Elbow Method in Figure 3 aims to determine the *clusters* that fit the data at hand. From Figure 3, we can identify that the point elbow on the y-axis is <0.2. On point, visible decline inertia is not significant again and forms an elbow on the graph. More inertia is low, as seen in the numbers third *cluster*. So, using the number of clusters on the point of the elbow can become a choice between the complexity and quality of *clustering* results. Next in Figure 4 are the visualisation results of K-Means clustering in a scatter plot on rain data in 1 hour. In the K-Means calculation, three clustering points will be obtained and used as reference axes for this clustering method. This K-Means calculation uses three categories. Namely, cluster 1 is bright, cluster 2 is cloudy, and cluster 3 is rainy. Each cluster has a marked intensity of different rain; cluster 0 has a mark rain_1h intensity of 0, clustering 1 has a mark rain_1h intensity of 0.1 to 2, and clustering 2 has a mark rain_1h intensity above value 2. The amount of data from each cluster is as follows:

Table 1. Amount of Data for Each Cluster in rain_1h

Cluster Type	Amount of data
Cluster 1	2
Cluster 2	15817
Cluster 3	467

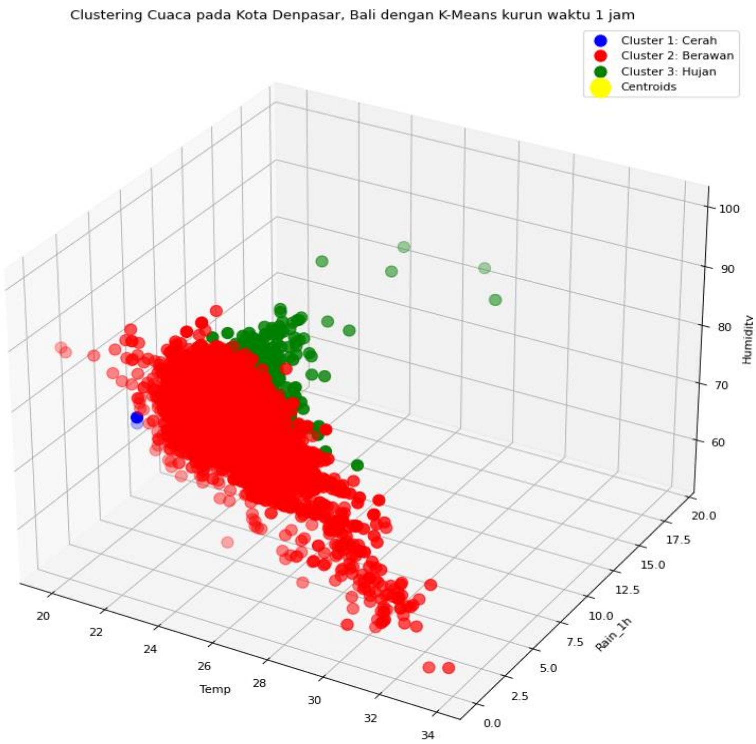

Figure 4. K-Means *Clustering* Weather in Denpasar City, Bali period 1 hour

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

Figure 4 shows that the cluster that dominates the most is cluster 2, which reflects the condition of cloudy weather. This is followed by cluster 3, which indicates the condition of weather rain, and clusters with the most negligible impact are cluster 1, which shows the condition of bright weather. On visualisation, several outliers are seen in cluster 2 and cluster 3, indicating the existence of some data that is significantly different from other data. Therefore, these data can not be categorised as a second *cluster.* So, that can concluded that in the *K-Means* weather clustering process on rain data in period 1 hour, the most dominant *cluster* is *cluster* 2, which illustrates weather cloudy.

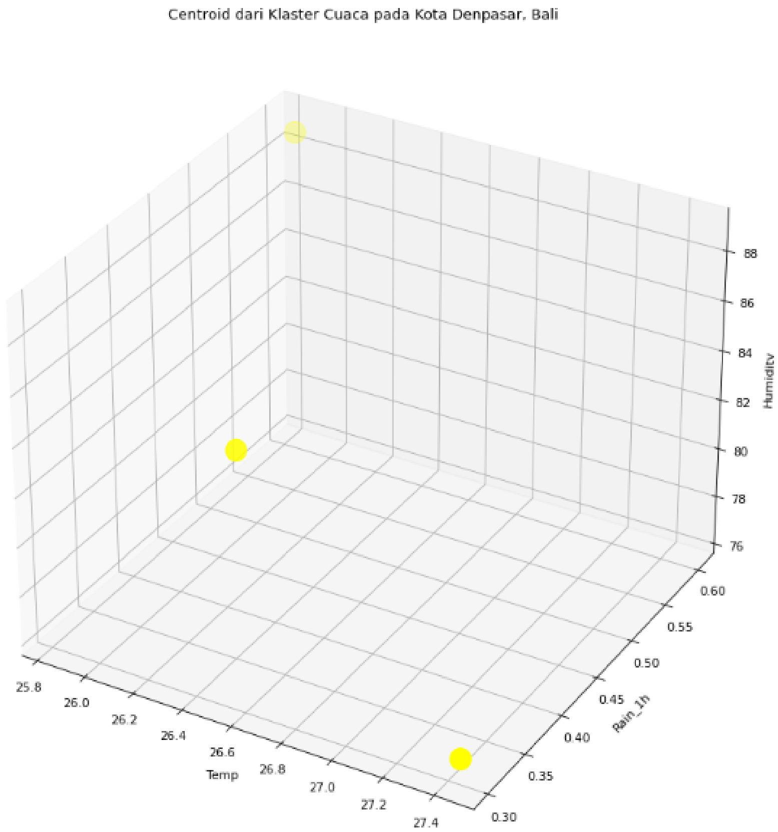

Figure 5. Visualization *Centroid* in *K-Means Clustering*

Figure 5 is a visualisation centroid in K-Means clustering, which plays a role as a point representing the centre of specific *clusters*. *Centroids* are calculated as the average of all included data points in *clusters*. This data marks the mean and median of each clustering on K-Means in bulk Rain in 1 hour, from Tables 1 to 3.

Table 2. *Mean* and Median Values *of K-Means Clustering* period 1 Hour (Sunny)

Information	temp	pressure	rain_1h	humidity	wind_speed
Median	22,780	1011,500	0,000	81,500	6,025
Mean	22,780	1011,500	0,000	81,500	6,025

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

Table 3. *Mean* and Median Values *of K-Means Clustering* period 1 hour (Cloudy)

Information	**temp**	**pressure**	**rain_1h**	**humidity**	**wind_speed**
Median	26.21	1010.00	0.23	86.00	3.60
Mean	26.213346	1010.174034	0.390455	85.996333	3.832903

Table 4. Mean and Median Values of K -*Means Clustering* period 1 hour (Rain)

Information	**temp**	**pressure**	**rain_1h**	**humidity**	**wind_speed**
Median	25.82	1009.00	2.63	89.00	4.48
Mean	25.876403	1009.045610	3.183940	88.518201	4.481435

3.1.2 Test K -Medoids on the rain_1h attribute

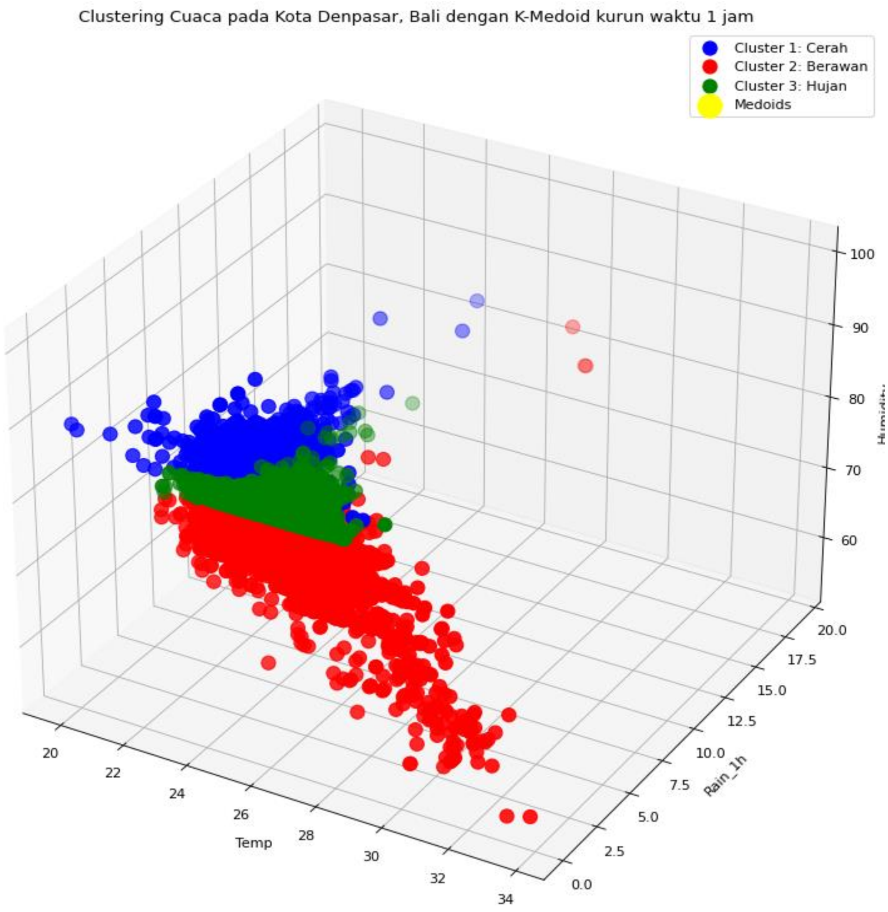

Figure 6. *Clustering* K-Medoids Weather in Denpasar City, Bali, for 1 hour

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

In Figure 6, it can be seen clearly that the cluster that dominates the most is cluster 2, which reflects the cloudy weather conditions. This is followed by cluster 1, which indicates weather conditions, clusters with the most negligible impact, and cluster 2, which shows cloudy weather conditions. On visualisation, several outliers are seen in cluster 1 and cluster 2, indicating the existence of some data that is significantly different from other data. Therefore _, these data No can categorised as in the two clusters. So, it can be concluded that in the K-Medoids clustering process, whether on rain data in period 1 hour, the most dominant cluster is cluster 2, which illustrates cloudy weather.

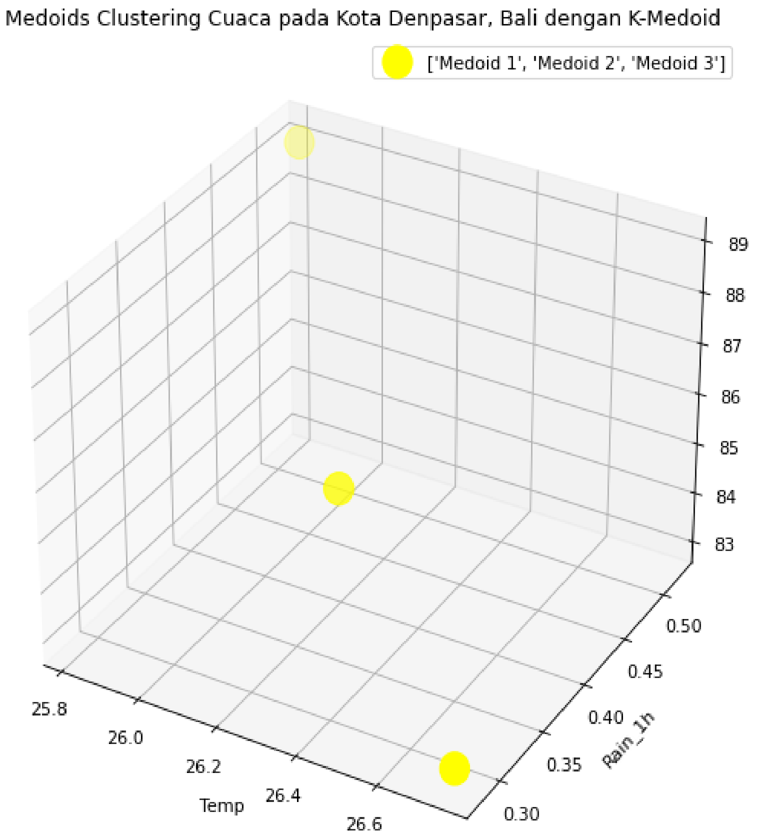

Figure 7. Visualization *Centroid* in K-Medoids *Clustering*

Figure 7 is a visualisation centroid in K-Medoids clustering, which plays a role as a point representing the centre of specific clusters. Centroids are calculated as the average of all included data points in clusters. From this amount of data, mark the mean and median of each clustering on K- Medoids on bulk Rain in a period 1 hour in time from Table 4 to Table 6.

Table 5. *Mean* and Median Values of K-Medoid *Clustering* for 1 Hour (Bright)

Information	temp	pressure	rain_1h	humidity	wind_speed
Median	25.80	1010.00	0.34	89.00	3.41
Mean	25.764608	1010.191962	0.652566	89.399901	3.623781

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

Table 6. *Mean* and Median Values of K-Medoids *Clustering period* 1 hour (Cloudy)

Information	temp	pressure	rain_1h	humidity	wind_speed
Median	26.86	1010.00	0.18	83.00	4.14
Mean	26.757906	1010.035351	0.308879	81.569155	4.321818

Table 7. *Mean* and Median Values of K-Medoid *Clustering* for 1 hour (Rain)

Information	temp	pressure	rain_1h	humidity	wind_speed
Median	26.33	1010.00	0.23	86.00	3.45
Mean	26.235095	1010.171557	0.404253	86.047693	3.727678

3.2 Test K-Means and K-Medoids on The Rain_3H Attribute

Upon testing, this calculation clustering will use the rain_3h attribute. As for the dataset that has the rain_3h attribute, it amounts to 16156 data. The following explanation about testing K-Means and K-Medoid calculations has been given.

3.2.1 K-Means method on rain_3h.

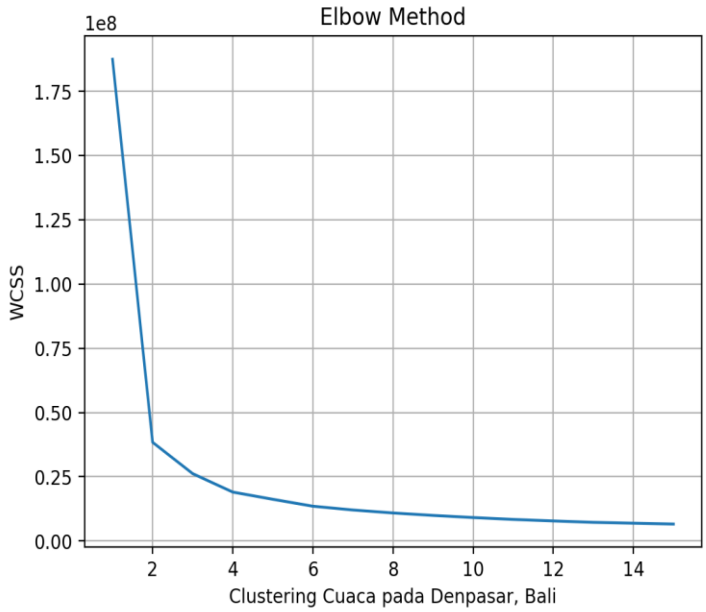

Figure 8. Visualization Chart *Elbow Method* in the rain Period 3 hours

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

The visualisation chart's elbow method in Figure 8 aims to determine the number of *clusters* that fit the data at hand. From Figure 8, we can identify that the point elbow on the y-axis is <0.2. On point, visible decline inertia is not significant again and forms an elbow on the graph. More inertia is low, as seen in the numbers third *cluster*. So, using the number of clusters on the point of the elbow can become a choice between the complexity and quality of *clustering* results. In the K-Means calculation, this time, we got three 3-point clustering that will used as a reference axis For this clustering method. This K-Means calculation uses three categories: *cluster* 1 is bright, *cluster* 2 is cloudy, and *cluster* 3 is Rain. Each cluster has a marked intensity of different rain; *cluster* 0 has a mark rain_3h intensity of 0, clustering 1 has a mark rain_3h intensity of 0.1 to 2, and clustering 2 has a mark rain_3h intensity above value 2. The amount of data from each *cluster* is as follows:

Table 8. Amount of Data for Each Cluster

Cluster Type	Amount of data
Cluster 1	349
Cluster 2	8132
Cluster 3	7675

The research results mark *clustering* rain_3 h in form visualisation as follows.

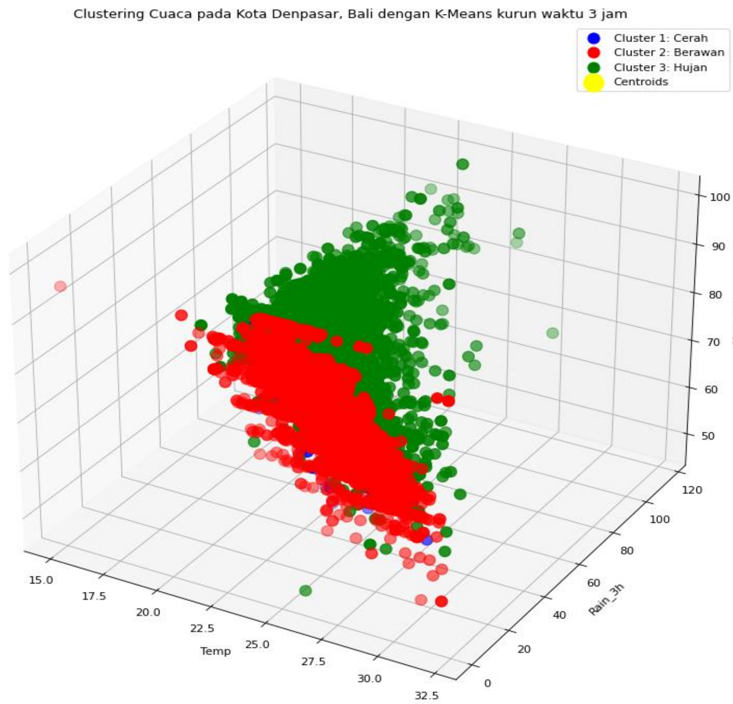

Figure 10. *Clustering* K-Means Weather in Denpasar City, Bali 3 hour time period

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

As pictured, each *cluster* also has a mark *centroid,* which, if visualised, will become as follows:

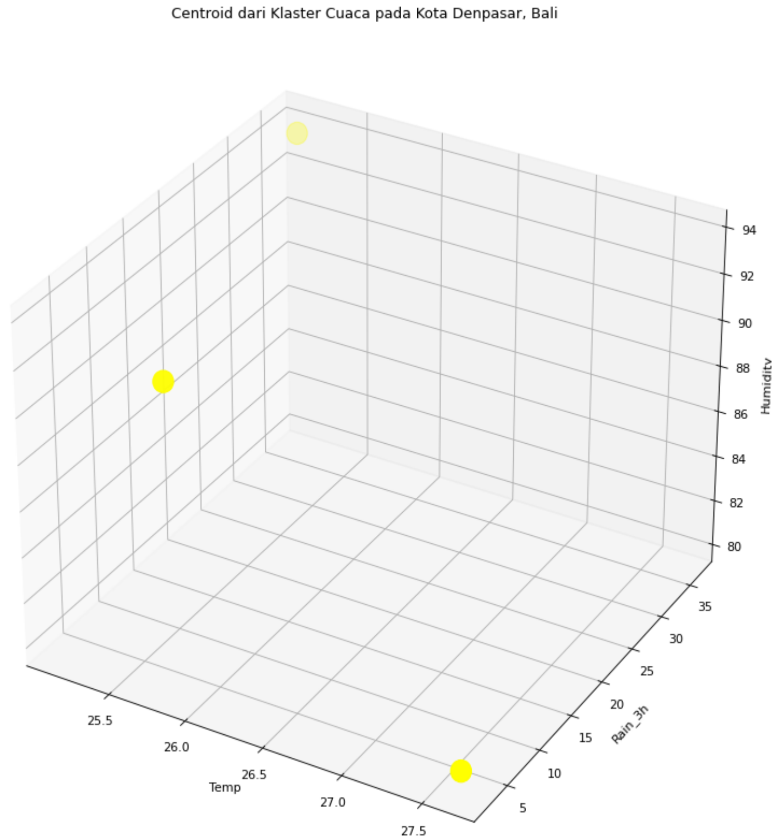

Figure 11. *Centroid* on Each K-Means *Clustering*

From the vast amount of data, we display the *mean* and median of each *clustering on* the K-Means in a form table like the following.

Table 9. *Mean* and Median Values of K-Means *Clustering* for 3 Hours (Bright)

Information	temp	pressure	rain_3h	humidity	wind_speed
Median	26.6	1009.0	0.0	85.0	3.6
Mean	26.705444	1009.105731	0.000000	84.587393	3.684957

Table 10. *Mean* and Median Values of K-Means *Clustering* for 3 Hours (Cloudy)

Information	temp	pressure	rain_3h	humidity	wind_speed
Median	26.4	1009.0	0.8	88.0	3.1
Mean	26.606524	1008.910145	0.933337	86.800172	3.639780

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

Table 11. *Mean* and Median Values of K-Means *Clustering* for 3 Hours (Rain)

Information	temp	pressure	rain_3h	humidity	wind_speed
Median	25.67	1009.00	8.00	93.00	3.10
Mean	25.813844	1008.832899	13.174710	90.630489	3.747368

3.2.2 K-Medoids method on rain_3h.

In the K-Medoids calculation, this time, we got three 3-point clustering that will used as a reference axis For this clustering method. In the K-Medoids calculation, three categories are used: cluster 1 is bright, *cluster* 2 is cloudy, and *cluster* 3 is Rain. The amount of data from each *cluster* is as follows:

Table 12. Amount of Data for Each Cluster

Cluster Type	Amount of data
Cluster 1	4862
Cluster 2	9054
Cluster 3	2240

Visualisation results of the calculations mark *clustering* rain_3 h in form visualisation as follows.

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

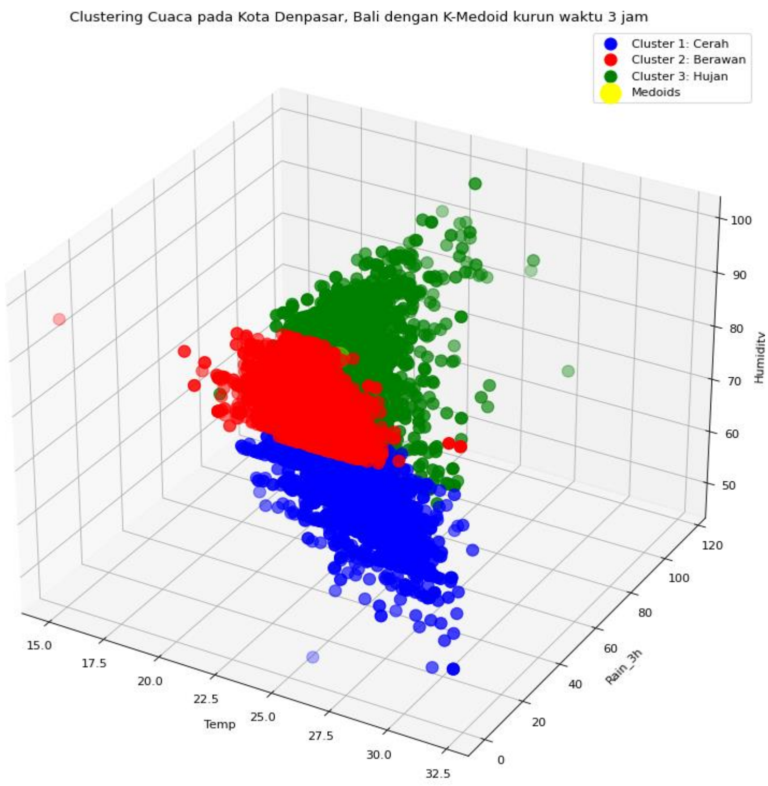

Figure 13. *Clustering* K-Medoids Weather in Denpasar City, Bali, for 3 hours

As pictured, each *cluster* also has a mark *centroid,* which, if visualised, will become as follows:

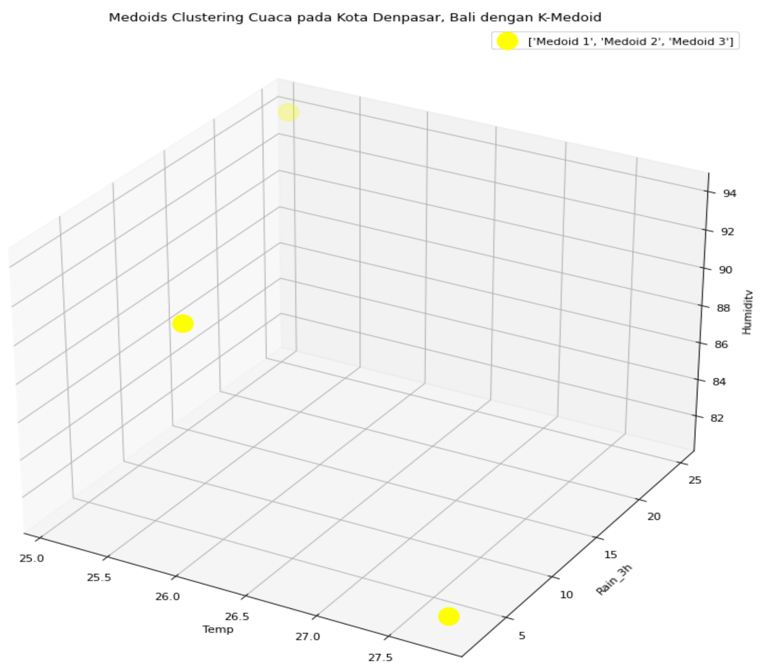

Figure 14. Centroid on Each *Clustering* K-Medoid

The vast data shows each cluster's marked mean and median on the K-Medoids in the following table.

Table 13. *Mean* and Median Values of K-Medoids *Clustering* for 3 Hours (Bright)

Information	temp	pressure	rain_3h	humidity	wind_speed
Median	27.6	1009.0	1.0	82.0	4.1
Mean	27.628149	1008.771678	2.341403	80.242698	4.438998

Table 14. *Mean* and Median Values of K-Medoids *Clustering* for 3 Hours (Cloudy)

Information	temp	pressure	rain_3h	humidity	wind_speed
Median	25.8	1009.0	2.0	93.0	2.6
Mean	25.739113	1008.963718	3.409421	91.822841	3.305225

Table 15. *Mean* and Median Values of K-Medoids *Clustering* for 3 Hours (Rain)

Information	temp	pressure	rain_3h	humidity	wind_speed
Median	25.06	1009.00	24.00	94.00	3.10
Mean	25.194513	1008.759955	29.666518	93.511161	3.632978

3.3 Test K-Means and K-Medoids on The Rain_6H Attribute

visualisation results from the research K-Means *clustering* has been done based on the dataset; it will explained in the chapter. Next is the calculation of K -Means trials and Medoids trials.

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

3.3.1 K-Means method on rain_6h.

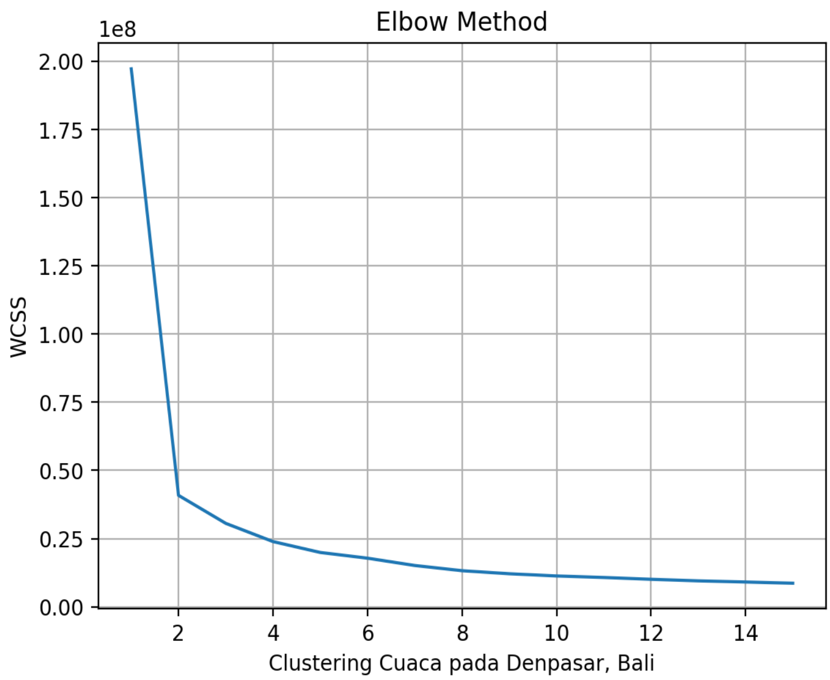

Figure 15. Visualization Chart *Elbow Method* in the rain Period 6 hours

The visualisation Elbow Method in Figure 15 aims to determine the *clusters* that fit the data at hand. From Figure 15, we can identify that the point elbow on the y-axis is <0.5. On point, visible decline inertia is not significant again and forms an elbow on the graph. More inertia is low, seen in the numbers *cluster* between 3. So, using the number of clusters on the point of the elbow can become a choice between the complexity and quality of *clustering* results. In the K-Means calculation, this time, we got three 3-point clustering that will used as a reference axis for this clustering method. In this K-Means calculation, we use three categories: cluster 1 is bright, cluster 2 is cloudy, and cluster 3 is rain. Each cluster has a marked intensity of different rainfall _, for *cluster* 0 has a mark rain_6h intensity of 0, clustering 1 has a mark rain_6h intensity of 0.1 to 2, and clustering 2 has a mark rain_3h intensity above value 2. The amount of data from each *cluster* is as follows:

Table 16. Amount of Data for Each Cluster

Cluster Type	Amount of data
Cluster 1	19769
Cluster 2	2338
Cluster 3	2893

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

Next, in Figure 15, is the visualisation result of deep K-Means clustering from scatter plots on rain data in 6 hours.

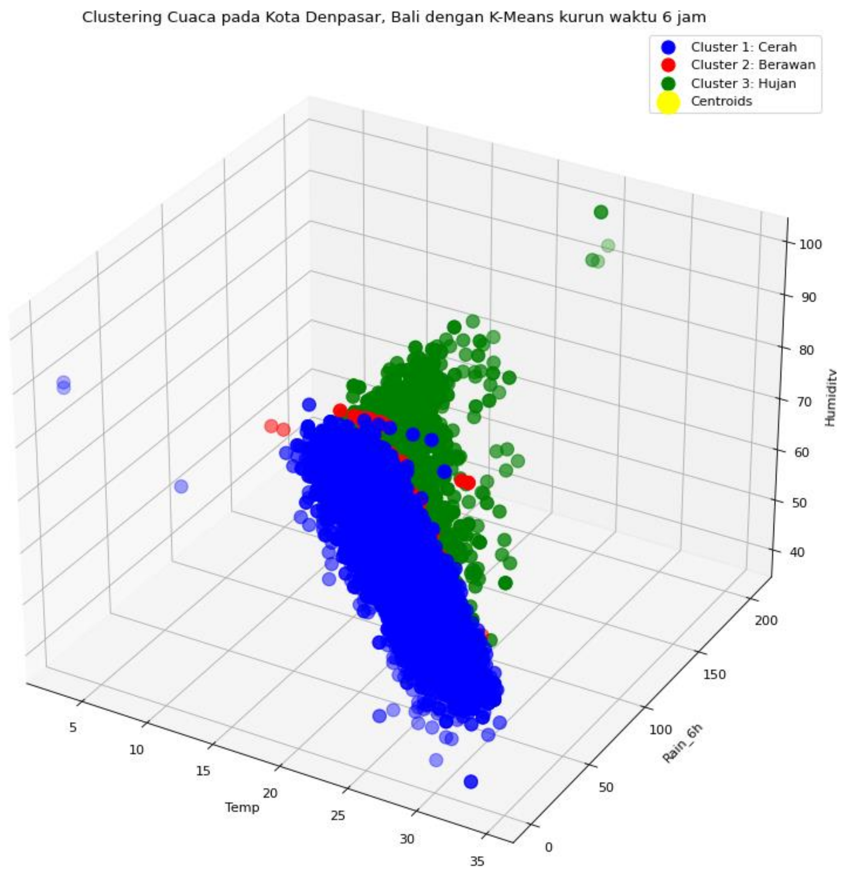

Figure 16. K-Means *Clustering* Weather in Denpasar City, Bali, 6 hour period

In Figure 16, it can be seen that the cluster that dominates the most among the others is cluster 1, which reflects bright weather conditions. This is followed by *cluster* 3, which indicates the condition of weather rain, and clusters with the most negligible impact are cluster 2, which shows the condition of cloudy weather. On visualisation, several outliers are seen in cluster 1 and cluster 3, indicating the existence of some data that is significantly different from other data. Therefore _, these data No can categorised as in second the *cluster*. So, it can be concluded that in the weather clustering process on rain data in 6 hours, the most dominant cluster is cluster 1, which illustrates bright weather.

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

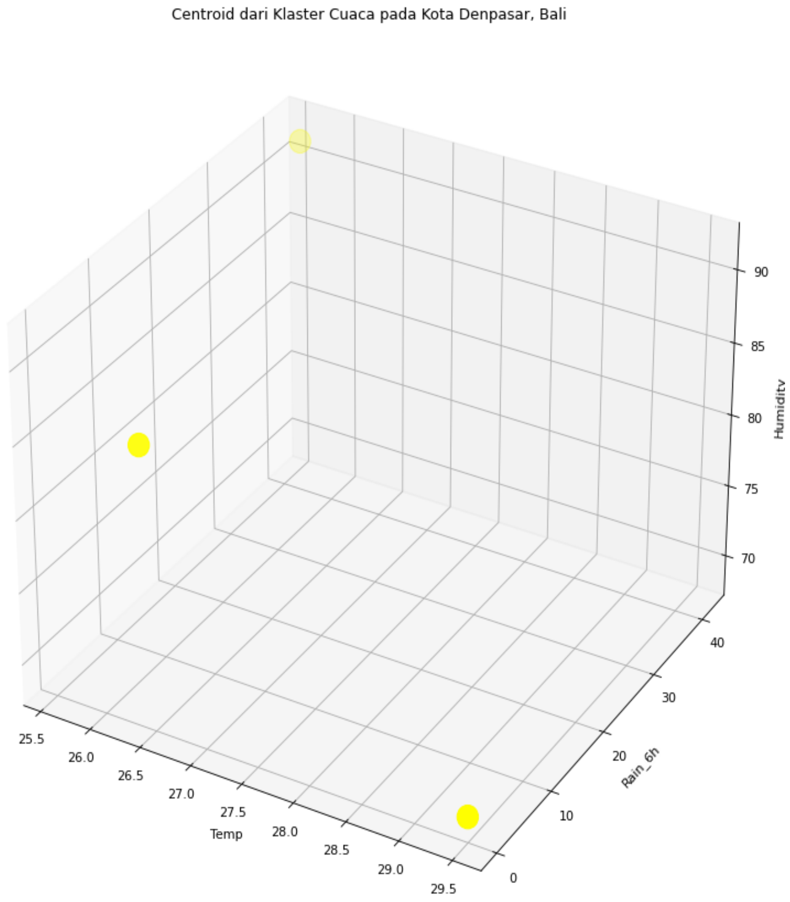

Figure 17. Visualization *Centroid* in K-Means *Clustering*

Figure 17 is centroid visualisation in K-Means clustering, which plays a role as a point representing the centre of specific *clusters*. *Centroids* are calculated as the average of all included data points in *clusters*. From this amount of data, displayed mark *mean,* and median of each *clustering* in K-Means exists bulk Rain in period 6 hours in time from table 13 to table 15

Table 16. *Mean* and Median Values of K-Means *Clustering* for 6 Hours (Bright)

Information	temp	pressure	rain_6h	humidity	wind_speed
Median	27.6	1010.3	0.0	79.0	3.1
Mean	27.757425	1010.349031	0.0	77.480652	3.198825

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

Table 17. *Mean* and Median Values of K-Means *Clustering* for 6 hours (Cloudy)

Information	temp	pressure	rain_6h	humidity	wind_speed
Median	27.0	1008.8	1.0	86.0	3.2
Mean	27.191074	1008.574166	1.022840	84.181352	3.701386

Table 18. *Mean* and Median Values of K -Means *Clustering* for 6 hours (Rain)

Information	temp	pressure	rain_6h	humidity	wind_speed
Median	26.0	1008.6	8.0	89.0	3.1
Mean	26.360100	1008.470377	15.160180	88.220878	3.837587

3.3.2. K-Medoids method on rain six h.

In the K-Medoids calculation, this time, we got 3-point clustering that will used as a reference axis For this clustering method. In the K-Medoids calculation, 3 categories are used: cluster 1 is bright, *cluster* 2 is cloudy, and *cluster* 3 is Rain. The amount of data from each *cluster* is as follows:

Table 19. Amount of Data for Each Cluster

Cluster Type	Amount of data
Cluster 1	8046
Cluster 2	9109
Cluster 3	7845

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

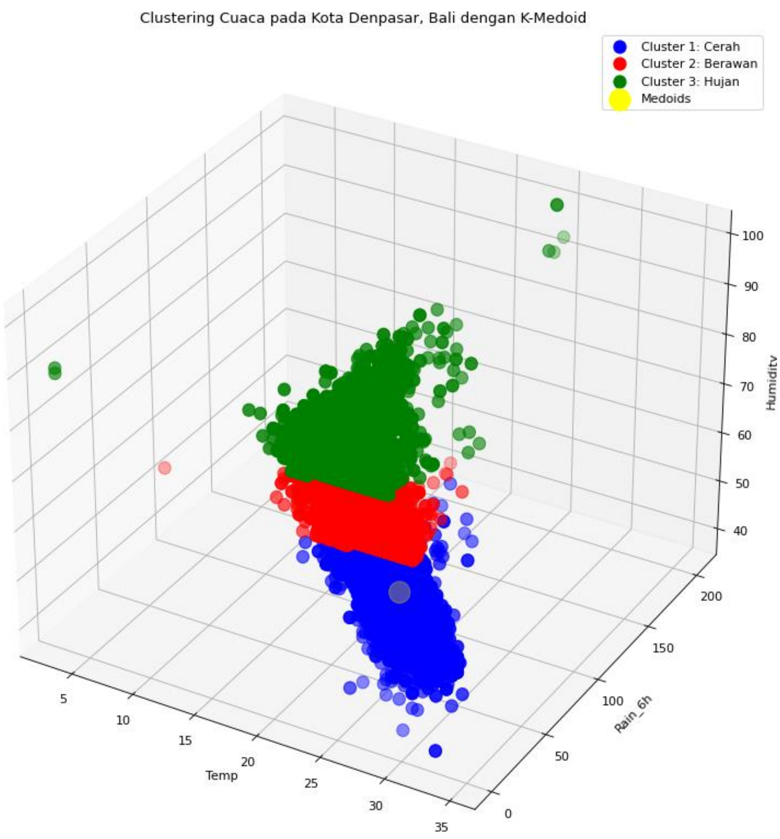

Figure 18. Weather K-Means *Clustering* in Denpasar City, Bali, for 6 hours

Figure 18 shows that the cluster that dominates the most is cluster 3, which reflects the rain condition. This is followed by cluster 1, which indicates weather conditions, and the cluster with the most negligible impact is cluster 2, which shows cloudy weather conditions. On visualisation, several outliers are seen in cluster 1 and cluster 3, indicating the existence of some data that is significantly different from other data. Therefore, these data No can categorised as in second the *cluster*. So, that can concluded that in the weather *clustering* process on rain data in period 6 hours, the most dominant *cluster* is *cluster* 3, which illustrates weather Rain.

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

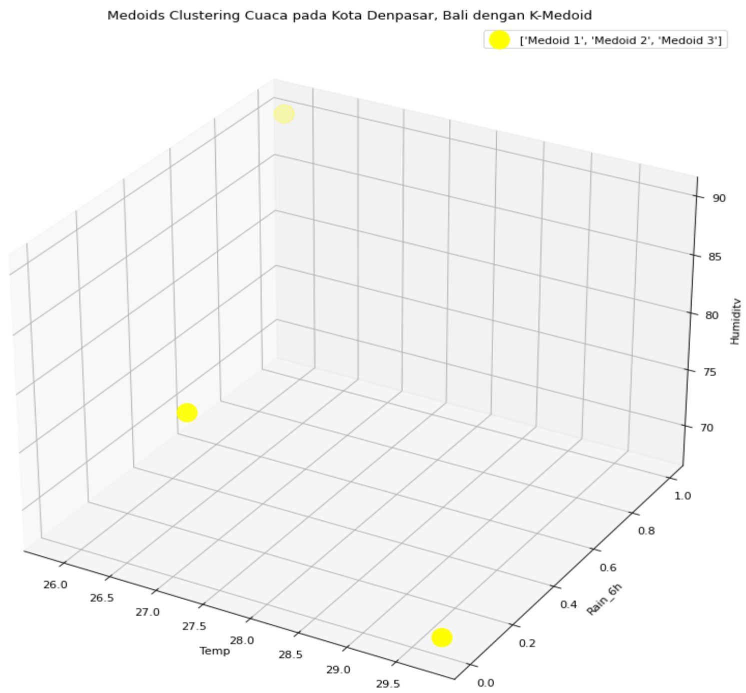

Figure 19. Visualization *Centroid* in K-Medoids *Clustering*

Figure 19 is a visualisation centroid in K-Medoids clustering, which plays a role as a point representing the centre of specific *clusters*. *Centroids* are calculated as the average of all included data points in *clusters*. From this amount of data, it is the mean and median of each clustering on Medoids on bulk Rain in 6 hours from Table 16 to Table 18.

Table 20. *Mean* and Median Values of K-Medoids *Clustering* for 6 Hours (Bright)

Information	temp	pressure	rain_6h	humidity	wind_speed
Median	30.0	1010.0	0.0	68.0	3.6
Mean	29.804044	1009.830910	0.322732	67.224087	3.934381

Table 21. *Mean* and Median Values of K-Medoids *Clustering period* 6 hours (Cloudy)

Information	temp	pressure	rain_6h	humidity	wind_speed
Median	27.2	1010.2	0.0	81.0	3.1
Mean	27.175277	1010.245603	0.664661	80.503019	3.245554

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

Table 22. *Mean* and Median Values of K-Medoids *Clustering* for 6 hours (Rain)

Information	temp	pressure	rain_6h	humidity	wind_speed
Median	25.80	1010.00	0.00	90.00	2.43
Mean	25.650236	1009.778776	4.792696	90.448311	2.775498

3.4. Model Evaluation

After doing a test with the use of the K-Means and K-Medoids cluster, the step furthermore is to determine the best cluster with the method comparing mark validity using the Davies-Bouldin Index (DBI) on each experiment carried out based on rain interval attributes of 1 hour, 3 hours, and 6 hours. The best DBI value is close to 0 (zero) (Risman et al., 2019, p. 20). The experiments carried out so results can be obtained in the following table:

Table 23. Comparison of DBI Values

Information	Rain for 1 hour	Rain for 3 hours	Rain 6 hours
K-Means	1,795	16,577	6,486
K-Medoids	27,599	7,602	11,485

Based on the comparison, The DBI value above is obtained as the best K-Means value, namely in the rain cluster with a 1-hour time interval of 1,795, whereas the best K-Medoids value in the cluster with a period of 3 hours is 7,602. Thereby, it is known that the K-Means value with a bulk Rain period of 1 hour time is more optimal than K-Medoids values with a bulk Rain period of 3 hours. Grouping will use the K-Means algorithm with bulk data with a Rain period time of 1 hour.

4. CONCLUSION

Based on the test results, the conclusion is that grouping the weather in Denpasar, Bali, can be done by grouping weather based on bulk rain with 1 hour, 3 hours, and 6 hours. Attribute variables that can cause grouping pattern characteristics in weather are temperature, humidity, and bulk Rain. As for clustering, we can use K-Means and K-Medoids algorithms. With the use of the algorithm so, we obtained the most optimal validity results for grouping weather using the K-Means algorithm viz amounting to 1,795 with bulk clusters Rain period time 1 hour meanwhile If the best K-Medoids value namely in the cluster with bulk Rain period 3 hours is 7,602. With three representative cluster conditions, sunny, cloudy, and rainy, usage of the K-Means algorithm can give an understanding of variation in weather in Denpasar. These clusters provide a richer picture of how the weather develops quickly,

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535

facilitating more monitoring, effective planning, and better planning, especially in a fast climate. So, the conclusion states that K-Means is more effective in grouping weather in Denpasar, Bali, and can give more accurate and relevant information for need planning and retrieval decisions regarding the weather in the region.

REFERENCES

Baek, S. J. (1998). Fast clustering algorithm for vector quantisation. *Electronics Letters*, *34* (2), 151-152.

Coleman, G., & Andrews, H. (1979). Image segmentation by clustering. *Proceedings of the IEEE*, pp. *67,* 773-785.

Hamerly, G. (2003). Learning the K in K-Means. *Neural Information Processing Systems*.

Hamerly, G., & Elkan, C. (2002). Alternatives to the k-means algorithm that find better clusterings. *Proceedings of the 2002 ACM CIKM International Conference on Information and Knowledge Management*, 600-607.

Jain, A. (2000, January). Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22* (1), 4–37.

Jain, A., & Dubes, R. (1988). Algorithms for Clustering Data. *Prentice Hall*.

Jain, A. K. (1999, September 1). Data Clustering: A Review,. *ACM Computing Surveys*, *31* (3), 264–323.

Kaukoranta, T. (1998). A new iterative algorithm for VQ codebook generation. *Proceedings 1998 International Conference on Image Processing*, 589-593.

Lee, C., & Antonsson, E. (2000). Dynamic partitional clustering using evolution strategies. *Conference: Industrial Electronics Society, 2000. IECON 2000. 26th Annual Conference of the IEEE., 4*.

Risman, Syaripuddin, & Suyitno. (2019, May). IMPLEMENTATION OF THE DBSCAN METHOD IN GROUPING DISTRICTS/CITIES ON THE ISLAND OF KALIMANTAN BASED ON PEOPLE'S WELFARE INDICATORS. *Proceedings of the National Seminar on Mathematics and Statistics*, *1*, 22-28.

Xiang, Z. (1997). Colour image quantisation by minimising the maximum intercluster distance. *ACM Transactions on Graphics*, *16* (3), 260–276.

DOI: https://doi.org/10.17509/edsence.v5i2.65925
p-ISSN:2685-2489 | e-ISSN:2685-2535