

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Journal

There are ten journals used in this project. In the ten journals used, it can be concluded that to determine the number of clusters to be used in the K-Means algorithm, it is still determined by the researcher, not using the Elbow Method, then the results of the Elbow Method have not been tested for cluster quality by the Silhouette Coefficient method. Therefore, the main topic in this research is to try to use the Elbow Method and the Silhouette Coefficient method to determine the quality of a cluster obtained from the Elbow Method.

3.2 Data

The data used in this study was obtained from the <http://dataonline.bmkg.go.id/> site for the cities of Semarang, Tegal and Cilacap for the period of January - February 2021 where the data used is monthly data. The parameters that will be used in the data are the highest, lowest, and average rainfall data.

3.3 Structure Design

The algorithm used in this study is the K-Means algorithm and will be assisted by the Elbow Method and the Silhouette Coefficient. The Elbow Method is a method that can be used to determine the number of clusters based on the WCSS (Within Cluster Sum of Squares) value, while the Silhouette Coefficient is a method that can be used to see the quality of a cluster. The Silhouette Coefficient has a range of values between -1 and 1, a good cluster quality value will be indicated by a value close to 1. To determine the distance of each data to the center of the centroid in each cluster, the Euclidian Distance formula in the K-Means algorithm will be used.

The first stage is to determine the number of k or clusters using the Elbow Method. The Elbow Method uses WCSS (Within Cluster Sum of Squares) or the distance between each data and each cluster at this stage will also determine the centroid point randomly.

After determining the number of k or clusters obtained from the Elbow Method, the Silhouette Coefficient method will be used to determine how much the quality of the cluster is compared to the number of other clusters. After finding a good number of clusters, the next step is to calculate the data with K-Means. The next step after determining the number of k or clusters using the Elbow Method and the Silhouette Coefficient, the K-Means algorithm will be calculated starting with randomly determining the centroid point, the first iteration process will be carried out, in the first iteration the Euclidian Distance formula will be used to determine the distance of each data with the center of the centroid.

3.4 Program

The programming language that will be used in this research is mysql, using laravel framework and vue js. The data source will come from a CSV file and will use Python's DataTable to process the existing data.

3.5 Implementation and analysis

The data that will be tested in this study is weather data from January to May 2021 originating from the city of Semarang, Tegal district, and Cilacap district, totaling around 112 data. It is also possible to add rainfall data from other months and years to compare research results.

3.6 Conclusion and writing report

Each existing data and attribute will form a cluster that has different levels of rainfall characteristics. After that, from the number of clusters formed, it will be concluded the condition of the level of rainfall owned by each city. The results of the Elbow Method and the Silhouette Coefficient will be compared whether it is true that the two methods in determining a quality cluster have the same number of k values. The better the quality of a cluster, the Silhouette Coefficient value will approach one.