CHAPTER 3
RESEARCH METHODOLOGY

3.1 Literature Study

In this process, the author takes and reads all journals related to the DBSCAN clustering algorithm. All journals collected by the author are used as a reference in working on this project. The first step is to determine the epsilon and minimum points for DBSCAN after getting the parameters to use DBSCAN, the next step is to run DBSCAN and visualize. After that, to calculate the DBSCAN performance, the author uses homogeneity, completeness, and V-Measure.

3.2 Gathering Data

The dataset used for this study was taken from Kaggle's website https://www.kaggle.com/justinas/nba-players-data in CSV format with 22 columns and 4 columns are used namely the height column or player_height, points or pts, rebound or reb, assist or ast. Only 4 columns were taken because there are only 3 columns that describe the overall performance in this dataset and the other 1 column is the height column.

3.3 Program Implementation

In the clustering process using the DBSCAN algorithm, the first thing to do is take the data to be processed. After taking the data to be processed, check whether the data is clean or not so that it can be used after that determine the parameters, namely epsilon and minimum points. To determine the minimum points, the formula is D+1 but because the data used is 2-dimensional data, the minimum points can be 4. Then to determine the epsilon the first step is to use the k-distance graph and then do the elbow method and then the next step is to see the results of the silhouette to optimize the parameters of the DBSCAN. After determining DBSCAN parameters to optimize the cluster results, do the calculations of the number of clusters to get the silhouette score. After that, the next step is to run the DBSCAN algorithm and visualize to get the results.