Phase2 – Clustering

In this part, we cluster the dataset and apply the classification algorithm to each cluster and check the results.

- Clustering methods: There are different types of clustering algorithms that handle all kinds of unique data.

  Density-based

  Distribution-based

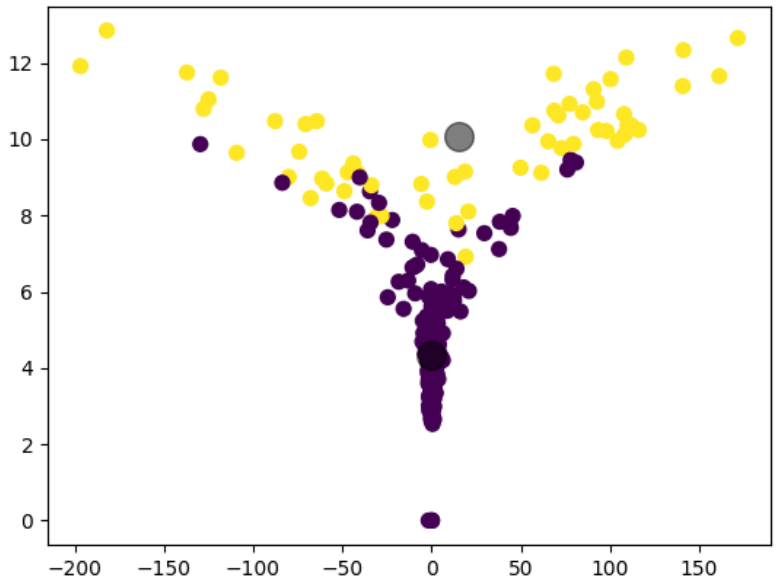
  Centroid-based : K-means

  Hierarchical-based

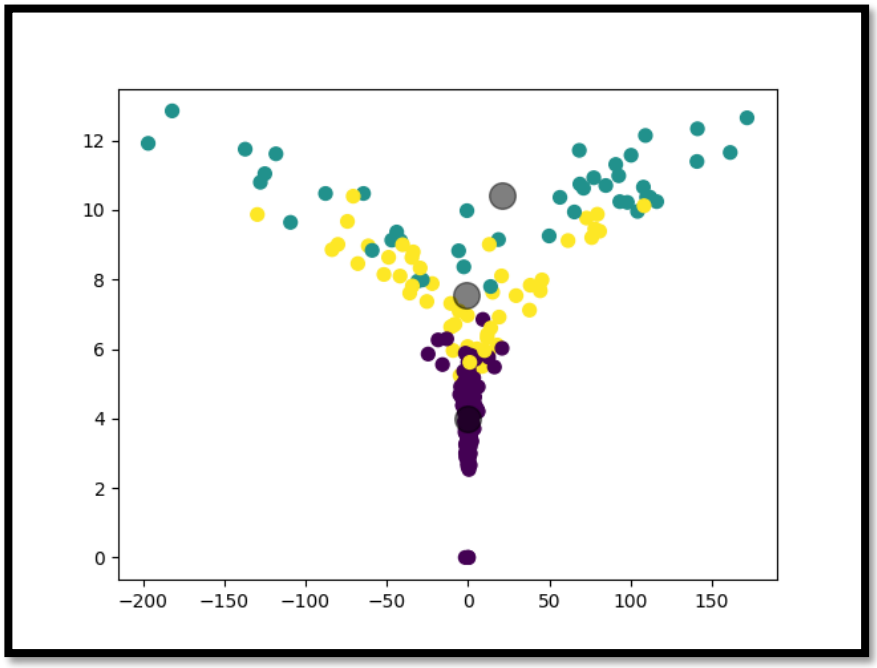
  K-means is a centroid-based algorithm, which tries to minimize the variance of data points within a cluster.

  We use K-means for clustering because it is best used on smaller data sets because it iterates over all of the data points also, in this algorithm, the number of clusters is not known from the beginning, and we tried several different clusters for better results.

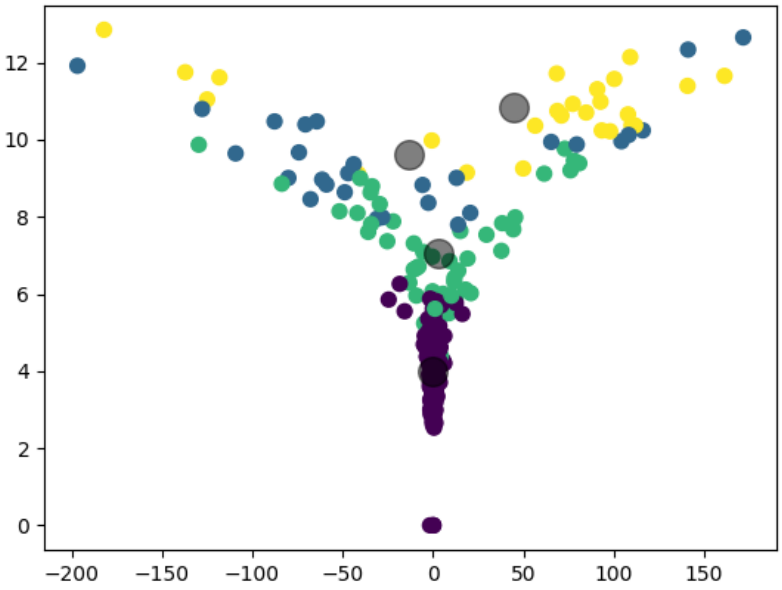
  - 2 clusters



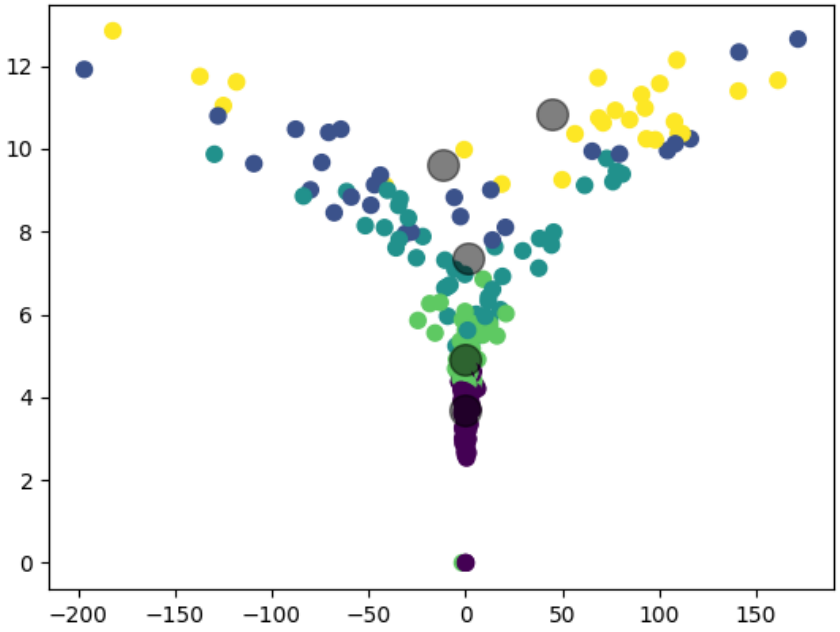
- 3 clusters



- 4 clusters



- 5 clusters



By examining the clustering results, it seems that choosing 3 clusters is suitable for the dataset.

- After clustering the data, we train a classifier (Decision Tree) with the training data for each cluster separately. Finally, we give each of the test data to the classifier that is closer to the center of the cluster and check the result.

  By doing this, the accuracy of seizure detection increased greatly because first the data with their features were placed in different clusters, which made the training data of each cluster used in the next step to train the algorithm have more similar features. In fact, each classifier is trained with similar data. During the test, the test data was given to a classifier that is more suitable for detecting that type of data.