Page No.
Date

BDA - Experiment 6

* **Aim:** To implement any one clustering algorithm

* **Theory:**

- Clustering is the process of dividing data points into a number of groups such that the data belongs to the same group.
- Types of clustering models:
  1) Connectivity models: The data points closer to each other exhibit more similarity to each other than the data points lying further away. The type of model, the data points can be classified as separate clusters and then aggregating them as the distance increases.

  2) Centroid models: These are iterative clustering algorithm where clusters are decided on the closeness of data point to the centroid of cluster. Eg. KMeans.

  3) Density models: Data points are assigned to a cluster based on the density of the points of the surrounding region. Eg. DBSCAN, OPTICS

* **Conclusion:** Clustering algorithm was implemented using PySpark and ML libraries.