Faculty of Computers and Data Science – Alexandria University

Course Name: Data Mining and Analytics

Semester: Spring 2021/2022

Final Project

The main goal of this project is to practice and apply clustering algorithms you have learned to real-world tasks.

Your task is to cluster the dataset into an optimal number of clusters, your steps will be as follows:

● Choose your preferred clustering real-world application

● Download any suitable datasets from (Kaggle, UCI repository, etc.)

Assess and clean your data if needed (you can use preprocessed data.)

Import the required libraries and perform:

1. **Agglomerative Hierarchical Clustering Algorithm**, find K clusters, using single linkage strategy, considering Euclidean distance as the distance measure.
2. **K-Medoids Clustering Algorithm** find K clusters, use the Manhattan distance.
3. **FP-growth or naïve Bayes** according to your data**.**

Visualize your results

- Conduct a comparison between the 2 Algorithms results.
- Team Members: only 2 students.

Please create a pdf report called "Project_Report", that:

- Describe your data, and your Target from this application.
- Explain your results and insight by describing your plotted graph.

● Submission Details:

- You should submit your code notebook containing your comments.
- Project_Report.pdf should contain your Name, ID and Group.