



European University of Bangladesh

Department of Computer Science and Engineering

Final Laboratory Examination, Fall-2023

B. Sc. in Computer Science and Engineering (Evening)

Course Title: **Artificial Intelligence Sessional**, Course Code: **CSE-312**

Semester: **4th Year 2nd Semester**

Course Teacher: **Promita Shabnam Khan**

Due: 09-Dec-2023

Full Marks: 40

This test will evaluate your understanding and practical skills on unsupervised learning, in applying clustering and segmenting customers.

You must complete all the tasks in the Jupyter Notebook under the Python 3 kernel.

All code must be written in one SINGLE “.ipynb” file.

Every person is different and hence is her/his behaviour as a customer. Imagine you are the owner of a shop. It doesn't matter if you own an e-commerce or a supermarket; it doesn't matter if it is a small shop or a huge company such as Amazon or Netflix, it's better to know your customers. Machine learning and specifically clustering is the most suitable unsupervised learning technique to create categories grouping similar individuals. Distance based clustering groups the points into some number of clusters such that distances within the cluster should be small while distances between clusters should be large.

Task 1 – Data Loading, Checking, Cleaning and Exploration

Dataset: I have given you a small dataset of customers shopping behaviour collected from Kaggle repository. The information gather is Customer ID, Gender, Age, Annual Income (k\$) and Spending Score (1-100). This task is about performing an exploratory analysis on the dataset including descriptive statistics and distribution, correlation between parameters, dimension reduction, etc. Please note that you need to support your analysis with appropriate visualisations.

Task 2 – Clustering

Now you’ve finished Task 1, you can apply suitable clustering algorithms to conduct unsupervised learning on the data. There are possible approaches for clustering, you must employ at least two clustering algorithms (K- means and DBSCAN are suggested, but you are free to apply other clustering algorithms) separately to do the task and compare their performance by appropriate evaluation tools. Before clustering, you need to check that the assumptions K-means and DBSCAN (other if applicable) make are fulfilled.



European University of Bangladesh

Department of Computer Science and Engineering

Final Laboratory Examination, Fall-2023

B. Sc. in Computer Science and Engineering (Evening)

Course Title: **Artificial Intelligence Sessional**, Course Code: **CSE-312**

Semester: **4th Year 2nd Semester**

Course Teacher: **Promita Shabnam Khan**

Due: 09-Dec-2023

Full Marks: 40

Task 3 – Report

In this task, you are asked to write a report to elaborate your analyses and findings in Tasks 1 and 2. Your report should contain the following sections:

- Why clustering is a right task to do for this goal?
- What clustering algorithm is appropriate to do the task and why?
- How does the Task1 help you to decide about clustering algorithm?
- Which algorithms work better and justify why is that the case?