The BRITISH UNIVERSITY IN EGYPT Informatics and Computer Science	**22CSAI03H** **Assignment 1** **2022-2023**
Module Title **Machine Learning**	
Module Leader **Associate Pro. Nahla Barakat**	Semester **One**
Assessment Weight **25% of the total course mark**	Due Date **04/11/2022@ 23:55**

Instructions to students:

1. This is a group assignment; each group consists of 3-4 students.
2. Submission: The submission is via the e-learning system only

3. Assessment: Assessment will be based on the code submitted, the report, in addition to scheduled discussion with team members if needed.

4. Feedback: Feedback for each team will be given through discussions.

5. Along with the submitted assignment, you need to submit: a fully completed and signed *Coursework submission form* *and* a *Statement of Academic Honesty Form*. You can only submit your own work. Any student suspected of plagiarism will be subject to the procedures set out in the academic university regulations.

Objectives:

This assignment objective is to demonstrate the knowledge and skills required to build an end-to-end machine learning project; that helps solving or improving solutions for real life problem domain, and report the obtained results. The scope of this assignment is the supervised machine learning algorithms.

Project resources

A repository of different data set links from different domains will be provided, where you can choose your project data set(s) from. You can also find your own project data set, however, you need to get the approval of the data set from one of the teaching team.

Python programming language should be used in all your implementations.

- ***For teams of three students, select one regression, and three classification data sets. For teams of four students, select two regression and four classification data sets***.

- ***Multiclass data sets can be performed as binary classification problem in one of the data sets***.

Assignment Tasks:

1- Describe your data set, and why you think it is interesting: (Domain, Features' types, percentage of missing values, outliers; if any, dimensionality, target class(es)

[10 Marks]

2- Visualize your data sets. **[10 Marks]**

3- Prepare a list of promising supervised learning algorithms from different types to be used to obtain different models; at least 3 models. **[12 Marks]**

4- Draw learning curves for your chosen ML methods **[16 Marks]**

6- Measure and compare different models' performances, using different measures. **[22 Marks]**

7- Analyze and comment on the obtained results. **[12 Marks]**

8- Document and report the above tasks as appropriate (1500-2000) words. **[18 Marks]**