Program: General/Intelligent Systems/Cybersecurity

Level: Third **Term: Fall 2023/2024**

Course Code: 02-24-01203 **Course Title: Data Science Tools &Software**

Total points: 15 **Professor name: Dr. Mohamed Abd El-Hafeez**

جامعة الإسكندرية
ALEXANDRIA UNIVERSITY
كلية الحاسبات وعلوم البيانات

Course Project

Download
https://github.com/mamahfouz66/EKNN_Ensemble_KNN_Based_Classifier/tree/main/EKNN

You will find the code and data of several cancer datasets such as Leukemia.csv

Before starting your project do the following:-

1) Select two datasets to experiment with
2) Identify the python library to be used
3) identify three different classification techniques to experiment with them such as KNN, SVM, and NN for example
4) Identify two feature selection (filter technique) techniques to be used
5) Identify two feature selection (wrapper technique) techniques to be used
6) Identify two dimension reduction technique to be used

Then,

1) start write your report with your selection above and show why you selected each option and expected advantages and disadvantages.
2) Sketch the pipelines and the layout of your approach

After that, read the attached code try to understand each part yourselves

Then create a python project to do the following such that each student in the group is assigned at least one of the following tasks:

1. Do cancer diagnosis using the above selected technique
2. fine tune the selected algorithms
3. Write down the result for each combination of technique
4. Cluster the genes (columns) using k-means, HC and DBSCAN after reducing the number of rows (samples) using PCA
5. Cluster the samples (rows) using HC after reducing the number of columns (genes) using PCA
6. Visualize the results of clustering
7. Compare and analyze the results of them.

A final report with code explanation and results should be submitted

Dr. Mohamed Mahfouz