
Assignment (2)
Naive Bayes & K-Means Implementation

1) Giving the following Data ("Weather New 2")

- a) In this dataset, there are 5 categorical Variables outlook, temperature, humidity, windy, and play, you must load our dataset "Weather_New_2.csv"
- b) We are interested in building a system which enable us to decide whether or not to play the game on the basis of weather conditions .
- c) Find the confusion matrix and calculate the Accuracy
- d) Calculate the Precision, F1 score, Error rate and Sensitivity
- e) Plot the ROC Graph (receiver operating characteristic curve)

2) Giving the following Data ("Mall Customers Dataset")

- a) This data set is created for the learning purpose of the customer segmentation you can do this by using (KMeans Clustering Algorithm) in the simplest form, load our dataset "Mall_Customers.csv"
- b) By the end of this case study , you would be able to answer below questions.
 - 1- How to achieve customer segmentation
 - 2- Who are your target customers with whom you can start marketing strategy
 - 3- How the marketing strategy works in real world
- c) Find the confusion matrix and calculate the Accuracy
- d) Calculate the Precision, F1 score, Error rate and Sensitivity
- e) Plot the ROC Graph (receiver operating characteristic curve)

What do I need to install?

You will need an installation of Python and also Anaconda from the following links

- [Python](#)
- [Anaconda](#)

You will need an installation of R and also RStudio from the following links

- [R](#)
- [RStudio](#)

Due Date (Submission Deadline)

May 6th at 11:59 PM

Assignment Discussion

May 7th on the same day of Data Mining Section

Assignment instructions

submit your Assignment to(mahmoud_samy@alexu.edu.eg). Get organized before you begin and you must submit it with your Full name and Academic ID before Deadline.

Avoid plagiarism !

Plagiarism is any act claiming or implying *another person's work is your own*, the below are some coding-specific examples of what constitutes plagiarism

Examples of plagiarism:

- Copying someone's code exactly, in whole or part.
- Combining code copied from multiple sources.
- Copying someone's code, or part, and making changes (e.g. changing variable or function names, comments, order of function definition).

If one of the previous points **Verified in your project**, your Grade will be **Zero**

Good Luck
M. Samy