Data Mining I: Final Examination (Spring 2020)
Friday, June 19, 12:00 AM - Tuesday, June 23, 23:59 PM (100%)
Wednesday, June 24, 12:00 AM - 23:59 PM (70%)
Thursday, June 25, 12:00 AM - 23:59 PM (40%)

- **Upload a single pdf file including all your work for all problems once.** If you upload multiple files, then only the last uploaded file will be counted. For example, you upload a file on June 23 and upload another file on June 24, then the file submitted on June 23 will be ignored.
- **You are fully responsible for uploading your work.** Try to upload your file as early as possible. There may be heavy internet traffic around the deadline. The above deadline is the time you finish uploading. It is not the time you start uploading.
- **Do not email your exam.**
- **Your pdf file must include this page with your name and CAU ID number.**
- **Show all work to obtain full credits. Grading is based on what you write, not on what you think.**

(10 points) This is to certify that I have used no other person as a resource. In addition, I have read the above instructions. I fully understand the instructions and agree on the terms.
Name: _____ ID# : _____

Please choose Accuracy or AUC to calculate your exam score.
I want to use _____.

Use the attached "Train.csv" data to classify the binary response variable y. Follow the writing instructions posted in eClass. There are two more files that are posted in eClass. "Xtest.csv" file has the predictor values that you need to calculate their predicted classes. The other file "Ans.csv" has two columns: one is "yhat" and the second is "prob". You fill the predicted classes and the posterior probabilities from your final model. Upload (1) your pdf file including all your work following the writing instructions and (2) your "Ans.csv" file that you fill two columns in the same ID order as "Xtest.csv" ID column. In addition, choose Accuracy or AUC to determine a part of your exam scores. Note that you are responsible for using the consistent coding and order as the original data when you fill the "Ans.csv" file.

grading items	points	policy
signed 1st page	10	Included as 1st page with name, ID, and your choice of Accuracy or AUC?
writing	10	Based on the writing instructions
methods and program	10	How many methods? Are they properly applied to the problem? Does the attached program run without error? Does the program produce the same results in your report? etc.
Accuracy or AUC	30	[(Accuracy or AUC) - 85] $\times 2$ based on your "Ans.csv" file If Accuracy or AUC is less than 85, then the score will be zero.