Problems Inclass 8_2. You can comment in this document and submit a pdf of your work. Please mark clearly all your answers and answer problems in the order provided.

1. Think through and answer the following problems to the best of your abilities.
   a) Valentine Day is approaching. A restaurant is trying to decide if to organize a singles' night or if to offer a special romantic menu. The restaurant has an established base of customers and collects demographic, income, social media and behavioral information on its customers. They decide to use the help of a data scientist to make sense of their Valentine's day menu in order to maximize sales (Valentine's days tend to be cash cows for restaurants). What algorithm would you use?

      I would use linear regression algorithm.

   b) Describe the type of information you would collect (what features) to decide if an email is spam or non-spam and what machine learning algorithm you would use.

      I would get the count of frequently occurring words that occur in most spam emails. Based on this, I'd use linear regression to get a probability if this is a spam email.

   c) Describe the type of information you would collect (what features) and from what sources to decide if to buy or sell a stock (financial investment). What machine learning algorithm can you use?

      I would collect data on the stock's recent trends, its mentions in the news, what experts are saying, and other stocks in the same market. I would use either logistic regression or linear regression to predict whether or to buy or sell the stock.

   d) How would you use Facebook to recommend certain products to people and what machine learning algorithm would you use?

      I would look at the pages they like, and the things they mention in their statuses. I would use linear regression.

2. A classification algorithm classifies emails into spam and non-spams. The following confusion matrix was returned by using the classifier on the testing set:

264	14
22	158

Consider "non-spam" = "positive" class. The matrix has the organization described in class. Calculate and interpret the following:

1) Accuracy rate
   422/458 = .92
2) Precision
   158/172 = .91
3) Recall
   158/180 = .88
4) F1
   1.82 * .49 = .89
5) Sensitivity
   158/180 = .88
6) Specificity
   264/278
7) In your opinion, is it more important to have good recall or precision?
   In this case, I think precision is more important.