Question 1:

* I ran the Naive Bayesian model and logistic regression model both using 4 fold cross validation methods.
* When I ran the Naive Bayesian model for the training data set with the name: `dataset.csv` the accuracy for 4 different iterations were:
  + 0.89
  + 0.94
  + 0.92 and
  + 0.86
* When I ran the logistic regression for the same dataset, the accuracy for 4 different iterations were:
  + 0.90
  + 0.93
  + 0.94 and
  + 0.92
* Therefore from the above observations, I was able to see that in each iteration Logistic regression was more accurate compared to the Naive Bayesian model. Therefore I would choose Logistic regression for this dataset.
* From the given file, following lines look to have inaccurate scores:
  + The programmer's service was so exceptional that I'm considering writing a book about it – How to Survive the Worst Programming Experience of Your Life.,1
    - Here, the sarcastic nature of the sentence might have been the problem. Here the service description seems to be exceptional but the score is low, so I think that’s inaccurate.
  + "I was disappointed with the programmer's service, primarily due to the expensive fees and insufficient support. The quality of work did not match the cost.",2
    - Here, the score I think is too high. It should have been 1 but 2 is given. I think the misclassification comes from the fact that there wasn’t strong adjective about the work
* Instructions on how to run the prediction based on my model:
  + I have a file named `predictQuestion1.py` that is located inside `question1` folder
  + I would like you to name your file to be `sample\_new.csv` which will contain 10,000 programmers and their data
  + I would prefer if you save the file in this directory: `​​question1/part\_1/sample\_new\_data/sample\_new.csv
  + But In case if you don’t want to do that, then remove this and just provide the name of the file inside the quotation mark.
  + After that you can run the python file and it will give you the prediction
  + This prediction is successful because I have `machine.pickle` which is saved after I created the model with the training data

Question 2:

* For the second question, I decided to go with the Naive Bayesian model as well. The structure of the dataset is the text along with the score. Since we are trying to create a classification model based on the profile picture given as a text, I think the Naive Bayesian model would be a good choice. Some of the reasons on why I chose Naive Bayesian model are:
  + The accuracy that I obtained for the training data is actually good:
    - 0.8,
    - 0.76,
    - 0.84 and
    - 0.79
  + Naive Bayesian model works really well with the small training data samples, therefore to fit the dataset this was another reason I went with Naive Bayesian
  + It handles both numerical and categorical values. Therefore these were the reasons I went with the Naive Bayesian model.
* Now I went ahead and saved the trained model using the name `machine.pickle`
* After creating the Naive Bayesian model, I created an image classifier:
  + For each of the folders: `buildings`, `faces` `dogs` I was able to iterate through the folders and grab each file in the jpg format
  + And using the code that was used in the class, I was able to parse the image to the numerical values with pixels being the column
  + Finally I saved them in the csv file titled as dataset\_extended.csv
* After saving the image data in the extended csv file, I went ahead and tried to run the neural network for this file.
  + I ended up creating the dense neural network using keras in python
  + The model I created was a sequential stack of layers.

Question 3:

* For the third question:
  + Combining Naive Bayesian model with Neural network as we did in both questions 1 and 2 might be the best idea
  + First of all for the text based classification:
    - I will have a model trained using the Naive Bayesian model
    - I will save it as a pickle
    - And then I will run the prediction on the new test dataset
  + And for the image based classification:
    - I would go ahead and train the model using Neural Network
    - I will save it as a pickle
    - And then I will run the prediction on the new test dataset
  + Then I will go ahead and create an algorithm to choose the score:
    - I will give some weight to the each score based on the accuracy of the model itself.