Assignment I – CS 6474 Social Computing

Grade	Max 75 points; 15% of overall grade (late policy applies)
Due	Oct 23, 2016, 11:59pm Eastern Time
What to hand in	A report (as a pdf file) with answers to the different questions; students choosing option A also need to include their code as a zipped folder
Where to submit	T-Square

Tasks (Choose ONE of the options A or B)

Option A. This assignment tests your technical ability in analyzing textual content shared in social media platforms. Refer to the two sets of text files enclosed in this assignment. 1) *Sentiment Dataset:* The first set contains 100,000 and 100,000 tweets in two files pos_examples_PosSentiment.txt and neg_examples_NegSentiment.txt, that are annotated to express positive and negative sentiment respectively. 2) *Emotional State Dataset.* The second set also contains two files – where the file pos_examples_happy.txt contains 17,740 tweets containing the hashtag #happy, while the file neg_examples_sad.txt contains 17,740 tweets with the hashtag #sad. Assume #happy and #sad to be noisy annotations of the positive and negative emotional state expressed in a tweet. In all the four files, each line represents a single tweet.

Based on these files, answer the following questions:

1) Build a binary classification framework that distinguishes between positive and negative sentiment tweets in the Sentiment Dataset. Specifically:
   a) *(10 points)* Extract n-gram (n=3) features from randomly sampled 80% of the positive and 80% of the negative sentiment posts, following lowercasing the text and excluding stopwords[1]. Filter for the top 5000 features (uni-, bi-, tri-grams combined) that occur most frequently in the above set of posts. Then, in your report, include a table with 50 of these top features and their respective frequencies in the sampled 80% positive and negative posts.
   b) *(10 points)* Build different classification models with three classifiers of your choice. In your report, justify your choice behind the choice of the three specific classifiers.
   c) *(15 points)* For each chosen classifier and using the top 5000 features extracted in 1(a), perform k-fold cross validation (k=5) on the above 80% sample of positive and negative posts. Calculate the mean accuracy, precision, recall, F1 score[2] across all of the cross validation folds for each classifier. In your report, include these results as a table with the measures as rows and the classifier types as columns.
   d) *(15 points)* For each classifier type, take the model that yielded the highest F1 score among the five cross validation folds. Apply this trained model for each classifier type to the remaining 20% of held out positive and negative sentiment posts. Compute the accuracy, precision, recall, F1 score for the three classifier types. Report these measures in a table in your report, with the measures as rows and the classifier types as columns.

---

[1] Ignore tokens that are "RT", Twitter usernames, or urls.

[2] Accuracy is the percentage of correct predictions made by your classification algorithm. Precision is defined as the number of true positives over the number of true positives plus the number of false positives. Recall is defined as the number of true positives over the number of true positives plus the number of false negatives. F1 score is the harmonic mean of precision and recall: 2 x ((precision x recall) / (precision + recall))

e) *(5 points)* Based on the validation and test results in 1(c) and 1(d) respectively, present a discussion of your findings. Are there any differences in classification performance between the validation and test tasks? If so, why? If not, why not?

2) Apply the classifier trained on tweets from the Sentiment Dataset to distinguish between the #happy and #sad posts in the Emotional State Dataset. Specifically,

    a) *(15 points)* Similar to the task in 1(d), take the model that yielded the highest F1 score among the five cross validation folds for each classifier type. Use the posts in files pos_examples_happy.txt and neg_examples_sad.txt as positive and negative test examples, and apply this trained model to this second test dataset. Compute the accuracy, precision, recall, F1 score for the three classifier types on this second test dataset. Report these measures in a table in your report, with the measures as rows and the classifier types as columns.

    b) *(5 points)* Based on the performance measures in 2(a), present a discussion on the performance of the three classifiers on this second test dataset. How does this performance compare to the first test dataset in 1(d)? Is it the same, better, or worse? Why?

Option B. This assignment tests your substantive understanding of the course material around the following topics covered in the class – performance, identity and deception, and social capital. Consider one of the first and one of the latest social computing technologies: MySpace and Snapchat. Although both platforms were/are very popular with teenagers of their time, they present significant differences in both their usage and affordances. Answer the following questions based on these similarities and differences between the platforms. You can use references to support your claims or observations (not included in the page limit given for each question). In writing your answer, you can also weave in your own experiences of the use of these or other social computing platforms.

1) *(20 points)* For each of these platforms, present an analysis of the various forms of identity curation we discussed in the class – performance, exhibition, and personal archiving. Specifically, what do performance, exhibition and personal archiving mean for MySpace and Snapchat? How did/do these platforms support each of these identity curation activities? Is one activity supported more extensively than others? Limit your answer to one page.

2) *(10 points)* Given the affordances of MySpace and Snapchat, discuss how identity deception may present itself on the platforms. In Donath's terminology, what are the sender and receiver costs of deception in these platforms? That is, is the cost of lying to the receiver high or low in one versus the other? Why? Limit your answer to one page.

3) *(25 points)* Describe "punishment models" that could be incorporated in the design of each of the platforms to discourage deception. Complementarily, how might you design a feature integrated in each of the platforms to encourage "authenticity" in interaction? For both the punishment model and the authentic interaction feature, include sketches of your design to support your argument, and describe your design choices. Limit your answer to three pages (including the sketches).

4) *(10 points)* Suppose you built a deception (or a lie) detector each for MySpace and Snapchat. Let us assume that a feature of this detector includes triggering a real-time notification to users on their phones. Discuss the sender and receiver costs of this signal in the context of the two platforms. For each, describe the costs when the detector incorrectly suggests a deception (or lie) and when they "miss" a lie. Limit your answer to one page.

5) *(10 points)* In Ellison's terms of bonding and bridging social capital on Facebook, discuss how MySpace and Snapchat allow(ed) users bond and bridge social capital with others. Is one platform more suited to facilitate one type of social capital building over the other? Provide your reasoning behind your answer based on the affordances of each platform, and the accepted community norms in them. Limit your answer to one page.