**Project Title:** title goes here  **Project Progress Report for CS 175, Winter 2016**

**List of Team Members:**Name1, StudentID1, uci\_email\_address  
Name2, StudentID2, uci\_email\_address  
[Name3, StudentID3, uci\_email\_address]

**[Maximum length should be 4 pages: if you go over a little that’s ok, but try to keep it to about 4 pages or less]**

**1. Problem Description and Background   
[This can be similar to (or even the same as) your proposal text unless you wish to improve it. Or if you have made a change to your project, from what you proposed a few weeks ago, please briefly note in this section that the current version of your project is different to what you proposed]**Write about 1 paragraph defining what problem your project is addressing. For example if your project is multi-label document classification then you would clearly define what multi-label document classification is. Mention what methods/algorithms have been used in the past to address this problem. Add one or more references to a text or a paper that describes the problem if you can.

**2. Description of Technical Approach   
[This should have more detail than your original proposal , ½ a page to 1 page maximum]**Provide a description of the methods and algorithms you are using on the project. For example, if your project involves comparing different classification algorithms for document classification then in this section you would list and briefly mention the classification algorithms you plan to use in your project (e.g., naïve Bayes, logistic regression, support-vector machines, neural networks, etc). Be as clear as you can about what versions of algorithms you plan to use, e.g., if you plan to use naïve Bayes, which version will you use (Bernoulli or Multinomial)? Or what type of neural network?

**3. Data Sets   
[This should have a more detail than what was in your proposal – make sure you include some details about the data – figures and tables are encouraged]**Describe what data set(s) you are using and/or investigating so far in your in the project – include references (e.g., URL) if you can. If for example you are doing document classification, you should mention how many documents are in the data set, average document length, how many classification labels. If you are using multiple data sets you could put this information in a table. Mention whether you plan to look at data with a predefined vocabulary or whether you plan to do investigate tokenization and define your own vocabulary. Feel free to also include figures in this section, e.g., a histogram of document lengths.

**4. Software  
[Separate this into subsections of (a) code or scripts you have written, (b) code or scripts you have successfully used on your data, and (c) software you plan to write or use (can be an updated version of your proposal text)]**Provide a list of the major pieces of project software where you (a) will use from publicly-available code, and (b) will write yourself. My expectation is that most students will use Python, given that we have been using Python in class and there are many useful publicly-available tools for text analysis in Python. However, if you prefer to use a language such as Java that is ok, but please indicate this clearly in this section.

**5. Experiments and Evaluation   
[Separate this into subsections of (a) experiments you have completed (if any) and (b) experiments that are planned (this can be an updated version of your proposal text)]**A brief description of how you will evaluate the results of your project, e.g., accuracy for classification, precision-recall for document ranking. Will you use cross-validation or does your data set(s) come with a fixed train-test partition? For tasks like clustering or topic modeling you may have to do some research to see how evaluation is done on these tasks. For some projects you may have to do some user studies for evaluation, e.g., present users with results from Algorithm A and Algorithm B, using the same input data for each algorithm, without telling the user which algorithm is which, and have them select the one they prefer.

**6. Challenges Identified**This is a new section. List 1 or more challenges that you have encountered so far in the project (e.g., algorithms are taking too long to run, or there is not enough labeled data available, or….) and for each challenge list how you plan to handle it.

**7. Updated Milestones**Your best estimate of what you aim to have completed at the end of each remaining week:

* End of Week 8
* End of Week 9
* End of Week 10

**8. Individual Student Accomplishments**Summarize briefly what each student has accomplished since the project started. There may be common elements here (i.e., items that multiple people worked on): if so provide a rough indication of what percentage of the task each person put into it (if its roughly even you can use 50% or 33% to indicate even contributions).

**Name 1:**   
**Name 2:**   
**Name 3:**