MET CS 677 Project Guidelines

One of the requirements for MET CS 677 is the final project (15% of the course grade). The project must be done individually. This is an opportunity for you to be creative in solving a data science problem that is of interest to you. The project should be challenging enough so that you could discuss it at future interviews with potential employers, and it will also give us an opportunity to provide you with a future reference if you ever need it.

You will have to record your (3-5 minute) presentation of your results using Kaltura video (similar to YouTube), then share the project presentation video at the "Media Gallery" section.

To to record a video and share in the "Media Gallery" section:

- Check out the directions to use Kaltura to capture and post or submit video.
- Inside the Kaltura directions, under the section of "Posting or Submitting Your Video in Your Course", follow the steps under "B. For sharing a video in the Media Gallery area".

In addition to a video presentation, you will submit the following:

1. one-to-two-page summary of what you have done and what were the main results
2. code source file(s)
3. instructions for how to run your code

Some of you are doing projects related to your work and we may not be able to run your code. In this case, please send us examples of runs with your summary. The presentations should be at the level that other students can understand what your project is all about. Imagine that you have an interview and you are asked to describe in a few words a data science project of your choice.

You are free to choose any topic that is of interest to you. The most important thing to keep in mind that this is a data science/machine learning project in Python. You will not be judged on the originality of your topic or the difficulty of implementation. We want you to show us that you can analyze a data science problem and write a short summary of your results. You are also free to pick up a data science/machine learning method or use any additional library that was not covered or used in class (e.g. deep learning or recommendation systems). If you use an algorithm or library we haven't discussed in class, **you must explain how the algorithm or library works in your writeup**. Major points will be lost if you utilize Neural Networks, for example, and do not explain how neural nets work under the hood (to the level of detail I give for algorithms we've learned in class). For new libraries, we just want to see that you understand the class structure and whatever functions you may be using.

In short, this is an opportunity for you to explore data science and do something that is of interest to you. We are asking you to send us a proposal so that we can assess if your project may be too difficult. Your proposal should be a few sentences to a paragraph long and **your proposal must include the following:**

1. Clear description of the data science problem you are trying to solve
2. Link to the dataset you will be using (see Kaggle, UCI Machine Learning repository, or any other publicly-available datasources) if not confidential employer data
3. List of algorithms you plan on using

Finally, your project should contain no confidential data from your employer or any non-public data source. Of course, all MET Academic Integrity policies apply to this project; **if anyone is found copying code from anywhere on the internet you will receive a grade of 0 for the project**.

Good luck to all of you, and have fun with this!