**Project Description**

This is an opportunity for you to explore an interesting machine learning problem of your choice.  Your project may be based on a real-world data set, or it may be theoretical in nature but grounded on a real problem.  
  
One of the best ways to identify a project topic is to choose a domain that interests you and identify problems in that domain.  Let the problem drive your choice of technique, rather than the other way around.

Topics can come from:

1. Health/Medical
2. Nature/Climate
3. Social/People
4. Sports/Gaming
5. Financial/Consumers
6. Miscellaneous (not from 1-5)   
     
   Your project will include three parts
7. A one-page project proposal, due in hard copy on January 2nd.
8. A presentation of your project to the class (5-10min) with power point slides starting the week of Jan. 8th
9. A final project report.

* Initial project proposal - 15%
* Presentation - 25%
* Final paper and project - 60%

You only have a month to complete the project, so keep the scope small and start early!!

Your paper should sufficiently describe your project, including:

* An abstract
* An introduction, describing the problem you are solving, the motivation for it, and a brief summary of your approach
* A description of your technical approach, using proper mathematical notation and formatting. This includes Normalization and Weighted KNN. Explain why you chose to increase/decrease the weight of certain categories.
* Your experimental methodology, description of your data set, the results you found (formatted as either tables or plots with proper labels and captions), and a discussion of your results.  If you did a theoretical project, you should have an expanded technical approach instead of this section.
* A brief conclusion

Here are some other sources of project ideas and data:

* Carlos Guestrin's site for the [CMU machine learning class](http://www.cs.cmu.edu/~guestrin/Class/10701/projects.html#datasets) includes further project ideas that will definitely need scaling down if you decide to use them for this class.  This site also has a lot of interesting data sets.
* Geoff Gordan’s class at CMU: http://www.cs.cmu.edu/~ggordon/10601/projects.html
* Ray Mooney's lists of project ideas:  [newer list](http://www.cs.utexas.edu/~mooney/cs391L/project-topics.html) and [older list](http://www.cs.utexas.edu/~mooney/cs391L/old-project-topics.html).
* Amy McGovern's [list of project ideas](http://www.cs.ou.edu/~amy/courses/cs5033_fall2009/project-ideas.html).
* The [UC Irvine ML Repository](http://archive.ics.uci.edu/ml/)
* Kaggle Data Sets: <https://www.kaggle.com/datasets>
* <https://www.data.gov>
* Stanford Project Ideas: <http://cs229.stanford.edu/projects2012.html>
* CalTech Projects: <http://courses.cms.caltech.edu/cs253/projects.html>
* Awesome Lists??