

# Class Projects

Prof. Xinhua Zhang

# Course Projects

## Overview

- 15% of your grade (graded out of 100)
- Each groups having at most 4 students
- Applying machine learning techniques to real-world data
- Very open-ended...

# Course Projects

## **Important dates**

- Oct 26: Proposal Due
- Nov 29: Video presentation due on VoiceThread
- Nov 30 to Dec 4: mutual comment and rebuttal
- Dec 14: Final Reports Due

# Project Proposal

## Project Proposal

- 1-2 pages
  - Who is in your group?
  - What dataset(s) do you plan to use?
  - What machine learning tasks?
  - What techniques do you plan to employ?
- 10 points (out of 100 of your project grade)
- Although the proposal itself doesn't require detailed description of the datasets, it'll be better to look at their properties. For example, is it too small or too large? Is it balanced in class? Does it have any attribute that is hard to handle (e.g., missing values for some instances). The better idea you have about the data, the less surprise you'll have in the future work (although it will be fine to change the dataset).

# Video Presentation

- 15-minute oral overview of the project
  - Problem addressed
  - Approaches used
  - Evaluation measures
  - Resulting performances (preliminary)
  - Lessons learned?
- 20 points

# Video Interaction

- Each student must post at least 1 question on one or more videos
- Each video will entertain at most 6 questions
  - VoiceThread shows the number of comments
  - Dividing it by 2 is the rough estimate of #questions
- Don't ask questions that have already been asked to a video
- Grade:  $10 * \max(0, x - 0.5 * y)$ , where
  - $x = 1$  if you've asked one question, and 0 otherwise
  - $y = \text{\#questions on your group's video that are not answered (by your group)}$

# Final Report

- Final Report
  - Use the Latex or WORD template from <https://nips.cc/Conferences/2015/PaperInformation/StyleFiles>
  - Max number of pages: eight (8), including figures and tables but excluding references.
  - One additional ninth page containing only the cited references is allowed.
  - At least 4 pages (excluding reference) recommended
  - Summary of problem, approach, evaluation, results
  - What did you learn from this project?
  - 60 points **(NOT the longer the better!)**

# Data Data Data

- Success depends on finding lots of data
  - Can't learn anything if you have no relevant data
- 30 Amazing (And Free) Public Data Sources [[link](#)]
- 100+ of the Best Free Data Sources For Your Next Project [[link](#)]
- Other sources
  - UIC projects (if you have data)
  - Kaggle.com
  - mldata.org
  - CORGIS Datasets (<https://think.cs.vt.edu/corgis/>)
  - US Census bureau (time use data)
  - Microsoft Research GPS dataset
  - Divvy!



# Evaluation

- Partially based on results
  - Does approach work well for problem?
- Partially based on effort
  - Did you put effort into evaluating different techniques, exploring the features used, etc.?
- Can I just implement an algorithm from a paper?
  - Yes, as long as the algorithm is not trivial and you add your own insights.
- Can I use the code of other authors? Yes, but more in other aspects
- Can I try something that's not covered in the course? Sure
- Can I writeup my own code? Sure
- Can I do a mathematical project – Yes!

**Implementation of algorithm + experiment on (big) dataset + insights from it + mathematical analysis**

# Example Projects

- Predicting Charged Off Loans on Lending Club
- Using Machine Learning Techniques to Predict Breast Cancer
- Sentiment Analysis of Twitter Corpus
- Predicting Audience Ratings for Movies with ML
- No Reference No Reference Image Quality Assessment with SVD, SVR, and General Regression Neural Networks
- Gentrification prediction in areas around Chicago using Socrata data based on business growth
- Survey of Techniques on Ames Housing Data
- Earning Potential Predictor based on demographic information
- Determining Mushroom Edibility for Human Consumption Using Classification