

# CS 542 Final Project Ideas (2017)

---

Please look at the following for data sets:

- 1) UC Irvine data sets: <http://archive.ics.uci.edu/ml/>
- 2) U Mich Predict data sets: [https://predict.org/default.aspx?cs\\_Category=2](https://predict.org/default.aspx?cs_Category=2)
- 3) Kaggle data sets: <https://www.kaggle.com/datasets>
- 4) Vast Data Set1: <http://vacommunity.org/VAST+Challenge+2015>
- 5) Vast Data Set2: <http://www.vacommunity.org/VAST+Challenge+2014>
- 6) C3E Challenge DataSets: <http://cps-vo.org/group/c3e/ccp>

Each of these the links above give you a data set to work on. (4) thru (6) give you a specific challenge problem to work on so if you choose to work on them, your project is to solve these challenge problems using the ML learning methods you learned in this class. (1) thru (3) give you data sets, list of papers that describe the research and ML tasks (classification, detection, clustering, etc.) people have done so far. Your project is to

- a) read some of the list of papers to determine what the state of the art currently is on that data set.
- b) Come up with a specific ML task you want to try on the data set
- c) Design the ML algorithms to perform this task.

You will work on this project with a team of 2, 3 or 4 (if you must) and will submit the result by the day of the final:

- The report will be in the NIPS submission format which can be found in: <http://nips.cc/Conferences/2013/PaperInformation/StyleFiles>
- The code of your algorithms
- Power point presentation of slides between 5 to 10.

If you find the somewhat open ended research like problem above too hard, you can work on the more specific class projects from the following two course websites:

- 7) <http://www.cs.cmu.edu/~gustrin/Class/10701/projects.html>
- 8) <http://www.cs.cmu.edu/~ggordon/10601/projects.html>

But I am hoping that most of you will choose a more research-y projects (1) – (6), as you may have the potential to turn it into a conference paper.

You may also choose topic from the videos we have watched in class:

- 9) Image Captioning - Li Fei-Fei
  - a) Object Detection
  - b) Sentence Generation

- c) Distance Estimation
- 10) Image Magnification
- 11) Voice Recovery from Image Magnification
  - a) Sound Generation
- 12) Seeing Around the Corner
- 13) Seeing Through the Books
- 14) Self-Driving Cars - Google
  - a) Velocity Detection
  - b) Auto Control
- 15) Intelligence Explosion

Some of the students have already sent me their individual topics, which are nice:

- 16) Deep Q-Learning Algorithm to Control a Simulated Car - Guang Yang  
Use simulated video streams taken by a camera to perform Q-function approximation. The output space contains Q values for each control actions (Accelerate, Break, Turn, etc.)
- 17) Recognition of American Sign Language using Kinect - Wei Jiang  
Kinect inputs do not need to be scaled before matching.
- 18) Dataset: [https://www.yelp.com/dataset\\_challenge](https://www.yelp.com/dataset_challenge) - Carrie Brown

If you are not yet satisfied...

- 19) Document Classification
- 20) Social Media Text Parser

Please let our TA know what your team is going to work on by 3/6/17. However, You can start now! 😊 Please submit to our TA 1 page summary of 1) your team (who is in your team), 2) problem you are going to work on, 3) a couple of rough ideas on your approach.