

# Data Ventures Workshops

Syllabus

Spring '18

## 1 Times

The workshop series for Spring 2018 will consist of 7 weekly sessions in SC216, alternating between labs and practicals. Practical sessions will run from 5:00pm - 6:30pm, and labs will run from 6:30pm - 8:00pm.

## 2 email

- [dataventures.harvard@gmail.com](mailto:dataventures.harvard@gmail.com)
- [russellkunes@college.harvard.edu](mailto:russellkunes@college.harvard.edu)

## 3 Labs

Labs will teach data analysis skills through guided workshops. There will be three labs. The topics are as follows:

- Introduction to Python, inference and prediction
- Overview of regression and classification techniques
- principal components analysis, k-means, and sliced inverse regression

Each lab will have a short coding assignment at the end.

## 4 Practicals

Practicals will be in the form of data analysis competitions. You will be presented with a task (for example, writing a program that identifies images of cats). You will make submissions in real time through a google form, and a leaderboard will be updated in real time. The winner of each practical will get a prize. There will be three practicals.

- Practical 1: Breast Cancer Detection, Preparation: Review basic pandas and numpy skills, pick out a few of your favorite classifiers from sklearn or another package to try out

- Practical 2: Toxic Comments Detection, Preparation: Figure out a good way to represent comments in a numerical basis (TF-IDF, word embeddings, etc.)
- Practical 3: Detecting Heart Abnormalities from Stethoscope Audio, Preparation: Figure out a good way to represent audio numerically. This will require a bit of research/ googling.

## 5 Final Project

The last session will be devoted to a short final project.

Final Project Topics:

- Option 1: super resolution - converting low resolution images to high resolutions images
- Option 2: recommendation - given a database of reviews, users, and businesses, predict how user  $i$  will rate business  $j$ .
- Option 3: anything else. You will have to submit a proposal (a 2-5 sentence email) before the last session for approval.

## 6 Office Hours

4:30pm - 5:00pm before workshops

## 7 Passing the workshops

Collaboration is encouraged on all the workshops. You can submit the labs, practical, and final project in groups of up to 5 people. Each lab or practical submission is worth 1 point, and the final project is worth 2 points. These are graded for completion. If the number of points you have at the end of the semester is greater than or equal to 6, then you pass the workshops.

## 8 Recommended Prep

Either CS50 or a strong interest in programming