

# COMP4332: Introduction to KDD project #1

[KDD Cup 2014 - Predicting Excitement at DonorsChoose.org](#)

# Project Overview

- task: predict exciting projects in [Donorschoose.org](https://donorschoose.org) (identify projects that are exceptionally exciting to people, at the time of posting)
- Donorschoose.org:
  - nonprofit organization that allows individuals to donate directly to public school classroom projects
  - Public school teachers can post their project requests
  - interested party can donate any amount of money to the project
  - a project may/may not reach its funding goal
- Kaggle website: <https://www.kaggle.com/c/kdd-cup-2014-predicting-excitement-at-donors-choose>

## Details of the available data(6 data files)

- donations.csv - contains information about the donations to each project. This is only provided for projects in the training set.
- essays.csv - contains project text posted by the teachers. This is provided for both the training and test set.
- projects.csv - contains information about each project. This is provided for both the training and test set.
- resources.csv - contains information about the resources requested for each project. This is provided for both the training and test set.
- outcomes.csv - contains information about the outcomes of projects in the training set.
- sampleSubmission.csv - contains the project ids of the test set and shows the submission format for the competition.

# Format of Submission

- format: csv

```
1:sampleSubmission.csv
1 projectid,is_exciting
2 ffff7266778f71242675416e600b94e1,0
3 fffeb510ee37a0bb01079f06bf141246,0
4 fff979abefa35a6bdd133b4e4150b737,0
5 fff8beec6de8c9411520d15d1f6979bf,0
6 fff745e9c0b8cc9e73e8c4c9a0ef4292,0
7 fff65a8a96697af9e6065451bd6d13d3,0
8 fff3b67b7818853d5bd13facc11449bb,0
9 fff1b626ea5fafa71b00369b3016c36d,0
10 fff005f999c0fb010ad97098366aec60,0
11 ffee6519c1e7e2423d3050df9fa29e69,0
```

# How to begin?

- formulate your own team, and name your team
- register your team on the [Kaggle Website](#)
- download the dataset from the [project page](#)
- do the experiment, and train your model
- submit the predicted result to the website
- get the rank and score of your team

# Score Criterion

- Codes of your application (30%)
- Documents describing your progress of building your model, features selection, model assessment and selection, etc.(40%)
- The score you got after your submission to Kaggle website, provide your score by screenshot. (30%)

# About submission

- Deadline: 11:59pm Apr 5<sup>th</sup>
- Submit to: [ywangby@connect.ust.hk](mailto:ywangby@connect.ust.hk)
- Title: **RMBI4310/COMP4332: Project 1, #Groupid**
- Content:
  - Group member:
    - Student id, email, name
  - Codes
  - Project report
  - Score in Kaggle

# Tools Recommended

- Anaconda(all included)
  - numpy
  - scipy
  - scikit-learn
  - matplotlib
  - and so on...