Capstone Project Ideas

Phai Phongthiengtham

Springboard Data Science Intensive Course

July 3, 2017

**Project 1: How do successful firms post their job vacancies?**

Potential Clients: Human resource department and job board websites

Data:

1. Job postings from CareerBuilder (all over the US from 2012-2016)
2. Compustat Database (Data on firms)

Broad Idea: A company’s success comes from each person’s individual contributions and unique talents. Putting the right people in the right jobs is very important and this starts with a firm posting vacancies. In this project, I intend to explore what features of job posts can predict a firm’s success (for example, profitability, Tobin’s Q, stock market movement). To narrow the question down, I intend to focus on management positions.

Possible factors (very preliminary):

1. Length of a post
2. Explicit employment benefit packages

Challenges and Concerns:

1. I already have access to data previously mentioned. However, it is limited for academic purpose only. Even if I manage to come up with a great model, I would not be able to sell my idea anywhere.
2. I still need to narrow down on what types of features I will be looking at. This requires considerable knowledge in natural language processing.

Table 1: Word Cloud from Management Positions



**Project 2: Software Learning Recommendation Product**

Potential Clients: Schools and job board websites

Data:

1. Job postings from CareerBuilder (all over the US from 2012-2016)

Idea: In this project, I intend to create an algorithm that recommends which up-to-date software(s) a person should know given education, location and job prospect. To do so, I plan to use job postings from all over the US from 2012-2016.

**Project 3: Predicting Video Game Sales and Ratings using Twitter and IMDB Data**

Potential Clients: Game companies

Data:

1. Video Game Sales and Ratings <https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings>
2. Twitter

Idea: In this project, I intend to use Twitter and IMDB data to build a model that can predict video game sales and ratings. Both sentimental analysis and numbers of comments should be main contributing factors.