Vince Laird

CAPSTONE 2 PROJECT PROPOSAL
DATA SCIENCE - CAREER TRACK

E-commerce: Predicting User Actions

The second capstone project will be based on the RecSys Challenge 2015. The website and details about this challenge are here:

http://2015.recsyschallenge.com/

This is a problem of coding a recommender system. The data used provides a sequence of click events / click sessions. For some sessions, there are also buying events. The goal is to predict whether the user is going to buy something or not, and if the user is buying, what would be the items the user would most likely buy.

This information is valuable to e-commerce businesses as it can indicate what items to suggest and how to encourage a user to buy (such as via promotions, discounts, etc.)

The data represents six months of activities of an e-commerce website selling random items such as garden tools, toys, clothing, electronics and more.

Approach to solving problem

Machine learning techniques will be used to solve this problem. Random forest is one technique that will be used. Depending on how effective or ineffective this approach is, other techniques such as deep learning may also be utilized.

The deliverable for this project will be a solutions file that contains two columns: one for session ID, one for item IDs. All session IDs will come from the test file (provided), and predictions will be made as to whether or not a “buy” will occur and if so, what will it be. Scoring will be determined using the formula at this site:
http://2015.recsyschallenge.com/challenge.html