CAPSTONE 2 Proposed IDEAS

Recommendation Engine with Yelp data

Yelp aims to help people find great local businesses, e.g. restaurants, salons etc. Their app is widely used by Yelpers to help other members of the Yelp community with a reliable review system that can help forming opinion about the business of interest before use or to identify the best suited business per the recommendations received from other users. Variety of work has already been done on their dataset. However not much of those have been tailored towards each customer's preferences. I would like to apply machine learning to predict a customer's star rating of a restaurant based on his/her reviews as well as other customers' reviews to recommend other restaurants to the customer. I would be interested in developing a restaurant recommendation algorithm that uses k-means clustering to suggest new restaurants to users based on keywords and other user preferences.

Dataset: https://www.yelp.com/dataset_challenge

It probably requires permission to download and disclose the results publicly.

HR Analytics/Employee attrition

Retaining valuable employees is essential for organization's success. Through this project I would like to identify, study and understand key drivers for employee attrition in any given organization's workforce and come up with a predictive model to help the companies retain good people by identifying most valuable employees who are at the highest risk of quitting.

Dataset: https://www.kaggle.com/ludobenistant/hr-analytics

Churn Analysis

According to Wikipedia, the definition of churn is: *"Churn rate (sometimes called attrition rate), in its broadest sense, is a measure of the number of individuals or items moving out of a collective group over a specific period of time."*

Churn is a critical measure for many businesses as it's often the case that acquiring new customers is a way expensive than retaining existing ones. "Churn" is a critical measure for many businesses yet very important for businesses especially with subscription models such as cell phone, cable, or merchant credit card processing plans. The term "churn modeling" can have different meanings depending upon the type of business and type of problem we are trying to address: For example, it can refer to calculating the proportion of customers who are churning, forecasting a future churn rate, or predicting the risk of churn for particular individuals. I would like to explore Orange data set in the link below and would like to see if some of the modelling techniques can be effectively used to predict churns for this telecom company.

Dataset: https://bigml.com/user/francisco/gallery/dataset/5163ad540c0b5e5b22000383