**Ray Gunawidjaja**
**Capstone project ideas**
**11.19.17**

1. Spectral analysis

**Motivation of study:**

Infrared (IR) spectroscopy is a simple, realiable technique that reveals the chemical structure of organic and inorganic materials. It is used widely in basic research and consumer and industrial products. Generally speaking, the peaks seen in an IR spectrum represents the resonance energy of the molecular vibration of the various functional groups present in a material. Hence, IR spectroscopy can be used to characterize materials by relying on the unique correlation between the spectral features in an IR spectrum and specific chemical functional groups.

Due to the frequent ambiguity in the spectral features the traditional approach to chemical identification using IR spectroscopy tends to rely on trained personnel. Fortunately, the National Institute of Science and Technology (NITS) has compiled a database that provide the IR spectra of over 40,000 known chemical compounds (http://webbook.nist.gov/chemistry/vib-ser/). This project will attempt to automate the spectral interpretation process by using the available reference spectra as the training data set.

**Goal:**

The purpose of this project is to utilize data science tools to first build calibration models using training spectra data sets and then test the model using three or more scenarios with varying ambiguity in the spectral quality (i.e., a spectrum randomly taken from the reference database, an unknown compound that is measured independently, and a material consisting of a mixture or a derivative of the compounds that are in the database).

**Scope:**

-Download the reference IR spectra from NIST website and be able to extract the spectral data.

-Employ the appropriate data science tools to correlate the chemical functional groups with the spectral features in the reference spectra.

-Obtain independently measured IR spectra of various materials, perform spectral processing routines (e.g., baseline subtraction, peaks normalization and peak identification), and test the model.

2. Investing in the S&P 500 composite index

**Motivation of study:**

In one way or another, all of us are tied to the financial stock market (e.g., 401K, college saving plan, banking system, etc.). The stock market is an appealing place to grow your idle money, but without being cautious many people have loss money doing so.

**Question:**

-Can we utilize data science to minimize risk investing in the stock market by selecting stock from the S&P 500 composite index?

-Is it possible to take samples from S&P 500 with certain criteria and beat its performance?

**Scope:**

-Download financial data from yahoo or google finance.

-Perform exploratory data analysis on data to classify stocks into various risk categories.

-Use basic financial tools to extract patterns in the price behavior.

-Establish a set of criteria using quantitative financial tools to minimize risks and determine if it is statistically feasible to make money in stock market exceeding the S&P 500 index benchmark.

3. Buying versus renting in Spokane, WA.

**Motivation of study:**

For a person with an average income in Spokane, determine whether renting or buying a home is more beneficial.

**Question:**

What is the break-even point of number of years for each of the two options?

**Scope:**

-Build a model that correlate housing price and renting price as a function of location, population density, crime rate, etc.

-Do web scraping to gather the relevant data (i.e., geographical location, population density, crime rate, etc.).