

## Capstone 1: Project Ideas

First idea:                    Classification Using PCA for Dimensionality Reduction

Satellite data will be increasing exponentially over the next decade, as many new projects come on line with advanced telescopes that acquire data at the scale of petabytes over their lifetimes. Classification of all these images will require enormous effort that can only be accomplished through computational automation. Pulsars are objects left over after certain types of stars have gone supernova. They produce unique emission patterns in the radio frequency. Unfortunately, radio frequency interference makes detecting these objects difficult.

This project will deliver a binary classification model to distinguish the signals of pulsar stars from those of the background radio noise.

The [dataset](#) will be obtained from the UCI machine learning repository.

Second idea:                Sentiment Analysis Using Network Theory

Product branding has always been part art and part science. A clever name can resonate with a customer to ensure product loyalty in the future. With the explosion of new craft breweries in the past decade, coming up with unique and memorable names for the millions of new beers being brewed has been a challenge.

This project will produce a model that will recommend beer names based on the words found in names of those beers with the highest ratings on a reputable beer rating platform.

As there is no database with the features needed for this task, it will be scraped and assembled from the [Beer Advocate](#) website.

Third idea:                    Recommender System Using Hybrid Filtering

Building recommender systems for movies using machine learning is all the rage today . Most rudimentary systems implement one of two approaches to filtering data, either collaborative filtering or content-based filtering. Both have their advantages and drawbacks. Collaborative filters can take advantage of a dataset with a large number of users and ratings to make recommendations based on the ratings of other users, ignoring other potentially useful features in the dataset. Content-based filters, on the other hand, can delve deeper into those features to give a richer recommendation based on what the user liked previously, but is hobbled when a movie has little or no reviews, known as the cold start problem.

This project will determine if there is an optimal method of combining these two filtering methods to produce a better recommender system than can be made by using either filter individually.

The dataset that will be used in this project is from [Grouplens](#).