Victor Ramirez

Project 1 Proposal

Springboard Data Science Career Track Bootcamp

**Project 1: Predicting Cryptocurrency Market\***

Problem: The cryptocurrency market is highly volatile. In this project, I will attempt to create an algorithm to predict future prices of cryptocurrencies based on their previous movements and sentiments from NLP.

Relevance: While volatile, cryptocurrency timing the market right could yield more lucrative gains.

Dataset and Acquisition: <https://www.cryptodatasets.com/platforms/Bitfinex/> has datasets on 21 cryptocurrencies. There’s also https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory coinmarketcap.com. For sentiment analysis, I can look at bitcointalk and r/crypto.

Approach: This is a supervised problem that attempts to predict price increase % the following day. Movement features may include the price fluctuations from previous days, the price trend from previous days, days of the week, etc. For NLP, I could perform a sentiment analysis on forums and news platforms. I can then use these features and plug them into an appropriate machine learning pipeline.

Deliverance: Code will be written in python and a paper will be written.

\*The complexity of this type of project might be better suited for capstone 2.

**Project 2: Tracking Political Polarization Over Time Using Twitter**

Problem: Social media has changed political discourse and is hypothesized to contribute to the growing political divide in the US. This project will aim to examine how political discourse has changed, by comparing tweets before, during, and after the 2016 election, particularly by focusing on polarization and how much ‘meaner’ people have been tweeting.

Relevance: The modern political climate is volatile but an interesting case-study.

Dataset and Acquisition: There are numerous ways to find and parse twitter data. The challenge is to categorize them as political then analyze the sentiment.

Approach: This requires classification using NLP.

Deliverance: Code will be written in python and a paper will be written.

\*The complexity of this type of project might be better suited for capstone 2.

**Project 3: Unpacking Reviews on Yelp into Smaller Subsets**

Problem: The aggregate score of a business tell only a small portion of the story. This project will attempt to unpack why a score is the way it is.

Relevance: Some aspects of a restaurants are more tolerable to others. Optimizing the review system to reveal more info can be helpful in helping users make more informative decisions.

Dataset and Acquisition: [Yelp has data on Kaggle](https://www.kaggle.com/yelp-dataset/yelp-dataset).

Approach: First I’ll identify some key features that users may find interesting, like vegetarian-friendly, kid-friendly, service, etc. that can be found on reviews. Then I’ll use NLP to divide reviews based on descriptions of the listed factors.

Deliverance: Python, paper

Alternative: Predicting genres based on lyrics. Lyrics data can be found on kaggle

**Previous Ideas:**

**Healthy Food Accessibility vs. Socioeconomic Status by Region**

Problem: Regions with low socioeconomic status tend to have poor access to healthy food. I’ll attempt to identify regions with the biggest disparity.

Relevance: Highlighting areas with the biggest disparities will help identify key areas in need of better food access.

Dataset and Acquisition: data.gov has datasets on various food data: [food prices](https://catalog.data.gov/dataset/food-price-outlook), [fruits and vegetable prices](https://catalog.data.gov/dataset/fruit-and-vegetable-prices), [food environment](https://catalog.data.gov/dataset/food-environment-atlas-f4a22). The food data will be crossed with [economic data by counties](https://catalog.data.gov/dataset/county-level-data-sets).

Approach: I’ll be merging the datasets by county and examine correlations between food accessibility and economic data.

Deliverance: Python, paper

Other ideas:

Examining use of buzzwords in scientific papers, Using Spotify’s API to optimize their recommendations to include local and underground artists (though this might