Springboard Project Ideation
Nicolas Wong (Feb 2021)

Public forum discussion analysis for identifying and predicting stock ticker behavior
Application - r/WallStreetBets is a popular “subreddit,” a channel that allows people to write and view comments of a particular topic, on the public discussion forum Reddit. Users on this subreddit were able to organize themselves and disrupt financial institutions by forcing a “short squeeze” on certain stocks. Since the discussions were public, is it possible to identify the stock tickers being converged on by analyzing the comments on subreddit?
Dataset - PRAW (Python Reddit API Wrapper) to scrape data and analyze

Identifying fruit health
Application - Identifying the health of a fruit is important for the agriculture industry at every stage: from growing, harvesting, processing, transporting, to selling. Using an annotated training set, one could build a model for identifying fruit by pictures (stretch goal: by video). An example of an application is monitoring changes in fruit health in response to changes in farming technique, climate, or storage. For consumers, one could use an app to get more information of the fruit of interest with a relative health check
Dataset - https://github.com/softwaremill/lemon-dataset

Building a model to predict if new code commits on Github will introduce a bug
Application - (Inspired by the Code Defect AI project) Bugs in programs are incredibly undesirable as they can either hurt the performance of the program, break the program or, in the worst cases, expose the user to vulnerabilities. As developers or programmers improve or add to their code, there is an increased likelihood of introducing a bug to the system. By analysing the bug reports and resulting bug fixes of programs, one could develop a framework of predicting how likely a new commit could introduce a bug or even suggest a test case for recognized patterns. The current offering available today is through a framework by Atlan (in conjunction with Microsoft) utilizing Azure (https://www.microsoft.com/en-us/ai/ai-lab-code-defect, https://github.com/aricent/codedefectai)
Dataset - Github API (https://docs.github.com/en/rest/reference/repos#commits)