**Capstone Project Ideas**

*Topic 1:*

Objective – To analyze Amazon review, specifically the fine foods category, to predict the number of stars for the product rating, based on the text review description.

Source – The dataset consists of reviews of fine foods from the Amazon website.

There are 568,454 reviews, 256,059 users, 74,258 products, etc.

The link to the data can be found here:

<http://snap.stanford.edu/data/web-FineFoods.html>.

*Topic 2:*

Objective – To analyze the news links on Facebook, from February 2016 to present when the emotion feature started, to predict the amount of emotional feedback and comments based on the keywords used in the headline of the article.

Source – The dataset contains 2012-2016 Facebook posts from 15 of the top mainstream media posts. The data will be further narrowed down into posts from February 2016-Present as the emotional feedback feature data can be found during the time-period. There are 15 different news sources, and the data will further be narrowed to one media.

The link to the data can be found here:

<https://data.world/martinchek/2012-2016-facebook-posts>.

*Topic 3:*

Objective – To predict the operating condition of the waterpoint for each record in the dataset. The water point will either be functional, functional needs repair, or non functional.

Source – The dataset is from Taarifa Platform who is currently working on an Innovation Project in Tanzania, and the Tanzanian Ministry of Water. There are 59,400 training data and 14,850 test dataset.

The link to the data can be found here:

<https://www.drivendata.org/competitions/7/page/25/>.