**Capstone 2 Project Idea**

**Idea 1:** **Stock portfolio performance Data Set**

**Description:** Three disadvantages of weighted scoring stock selection models. 1st, they cannot identify the relations between weights of stock-picking concepts and performances of portfolios. 2nd, they cannot systematically discover the optimal combination for weights of concepts to optimize the performances. 3rd, they are unable to meet various investors preferences. This study aims to more efficiently construct weighted scoring stock selection models to overcome these disadvantages.

To systematically discover the optimal combination for weights of concepts to optimize the stock portfolio performances, this study aims to analyze a more efficiently weighted scoring stock selection model to improve the relations between weights of stock-picking concepts and performances of portfolios as well as to meet various investors preferences.

**Link:** <https://archive.ics.uci.edu/ml/datasets/Stock+portfolio+performance>

#### Idea 2: SEER Cancer Incidence

**Description:** The U.S. government also has [data about cancer incidence](http://seer.cancer.gov/faststats/selections.php?series=cancer), again segmented by age, race, gender, year, and other factors. It comes from the National Cancer Institute’s Surveillance, Epidemiology, and End Results Program. The data goes back to 1975 and has 18 databases, so you’ll have plenty of options for analysis. The target is to predict Rate per 100,000 based on the attributes.

**Link:** <https://seer.cancer.gov/explorer/application.html?site=1&data_type=1&graph_type=3&compareBy=sex&chk_sex_3=3&chk_sex_2=2&race=1&rate_type=2&advopt_precision=1#tableWrap>

#### Idea 3: Dow Jones Weekly Returns

**Description:** Predicting stock prices is a major application of data analysis and machine learning. One relevant data set to explore is the [weekly returns of the Dow Jones Index](http://archive.ics.uci.edu/ml/datasets/Dow+Jones+Index) from the Center for Machine Learning and Intelligent Systems at the University of California, Irvine. This is one of the sets specially made for machine learning projects.

**Link:** <http://archive.ics.uci.edu/ml/datasets/Dow+Jones+Index>