Capstone Project 1: Project Proposal

Mehmet Erdogdu

1-What is the problem you want to solve?

If breast cancer left untreated, the cancer spreads out to other parts of the body since it is a malignant cell growth. Benign cells are usually localized and do not spread to other parts. I will predict if a given cancer cell is benign or malignant to be able to treat the cancer cells in a timely manner.

2-Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?

My client would be hospitals, medical institutions. They care this problem because using highly accurate prediction model can reduce the cases of life loss due to cancer by taking necessary preventive actions on malignant cancer cells.

3-What data are you using? How will you acquire the data?

I will be using University of California, Irvine Machine Learning repository Breast cancer diagnostic data set. I will acquire the data set through Kaggle website.

URL: https://www.kaggle.com/uciml/breast-cancer-wisconsin-data

4-Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.

The data set is relatively clean since I acquired from Kaggle website. I would still search for any missing values in the dataset. Secondly, I will do an exploratory data analysis for the feature selection. I will find the features with

high correlation. I will then decide which features I should drop from the data set to reduce my dimensionality.

5-What are your deliverables? Typically, this includes code, a paper, or a slide deck.

My deliverables will be the Jupyter notebook that includes codes and some notes. I will also prepare a paper that summarizes the key results from the project.