Capstone Project Proposal

1. What is the problem you want to solve?

Within the people of age group 65 or older, the average medical expenses for Dementia or Alzheimer's is reported being nine times higher than expenses for other Medicaid members in the same age group [1]. About one third of individuals 85-year-old or older in the U.S. are affected by Alzheimer's disease [2].

We need to first uncover the diagnosis process of the cognitive impairments so that the tests could be identified even the disease begin. Early discovery of potential cognitive decline may result in prevention of the disease as well as costs to treat the patients. We want to predict the signs of the disease as soon as possible to dry out or at least reduce the costs to both patient and the Medicaid agency.

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 otherwise?

My client is any Medicaid agency dealing with high costs of dementia. Especially, states with the high population of elderly people may be under potential risk of high rates of cognitive impairment type of diseases so the medical expenses likely to be higher in those states if age is a factor of the disease. To uncover this, we need to get population demographics data from states as well.

3. What data are you going to use for this? How will you acquire this data?

Although I am still searching for a good data to see the factors of dementia, I will use healthy aging data from Behavioral Risk Factor Surveillance System (BRFSS).

4. In brief, outline your approach to solving this problem (knowing that this might change later).

I will approach this machine learning project by following the steps below:

a. Create a repository in Github (in addition, create google driver to use shareable documents with my mentor)

b. Gather the data and load it into Python.

c. Analyze the data to determine the data quality

d. Prepare the data:

(1) Clean that which may require it (remove duplicates, deal with missing values, correct errors, normalization, data type conversions, etc.)

(2) Randomize data, which erases the effects of the particular order

(3) Visualize data to help detect relevant relationships between variables and perform exploratory analysis.

e. Feature engineering and selection

f. Split the data as training and test data

g. Choose a machine learning algorithm

h. Train the model

i. Evaluate the model

j. Parameter Tuning

k. Make predictions.

l. Prepare a report

5. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

My deliverables will be a paper, a PowerPoint presentation summarizing my paper, and the code associated with my project.