Springboard Data Science Career Track
Proposal for Capstone Project 1
By: Huong Ivy Nguyen, Ph.D.

1. What is the problem you want to solve?

   This capstone project is inspired by a Kaggle competition posted and hosted by Intel and MobileODT. The main goal of this project is to develop a new algorithm that can effectively identify the type of cervix a patient has based on images. This workflow will help health providers give proper cervical cancer treatment referral to their patients.

   Cervical cancer is classified as a easy-to-prevent cancer if caught in its precancerous stage. According to the American Cancer Society, cervical cancer is one of the most successfully treatable cancers if detected early. However, one of the most problematic issue in treating patients with this type of cancer is the ability to identify an appropriate treatment that works effectively and accordingly to the patient's physiological needs. In rural parts of the world, many women who are susceptible to cervical cancer are receiving treatment that will not work due to the position of their cervix. In order to solve this problem, it would be necessary to develop an algorithm that could help health providers identify the type of cervix that a patient has based on images.

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?

   The client for this problem is health providers as well as the patients that are at high risk for cervical cancers. The health providers can use this algorithm to have an effective real-time determination to provide an appropriate cervical cancer treatment for their patients. That way, patients don't have to face high-cost treatment as well as ineffective treatments. This will help both health providers and patients reduce their time in giving and receiving cervical cancer treatment.

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

   At this time, the data is provided by Intel and MobileODT via their posted Kaggle competition.

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.

In order to solve this problem, I will first have to get familiar with the given dataset to understand all the given facts.

In order to begin building my algorithm, I will have to look into different methods and packages that could be used to build my algorithm such as convolutional neural network or TensorFlow. After building my algorithm using the training dataset, I will test how precisely and accurately it works against the testing sample dataset.

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

   The deliverables will be in either R markdown document or Jupyter Notebook pdf detailing any exploratory data analysis and algorithm construction. All codes will be presented.