Capstone Project Proposal Template

Notes:

- This should take no more than one hour to complete – the clearer you are about the business problem you're working to solve with your ML-driven solution, the easier your proposal will be to complete
- This will be uploaded to your repo, which will be a part of your final submission
- Due date for submission is 12/9

Instructions:

1. Download this document as a Word Doc
2. Answer each question using a few sentences, at most
3. Save your completed proposal as a PDF
4. Create a project GitHub repo (if you have yet to do so)
5. Add your instructor as a collaborator (username nickmccarty) to your project repo
6. Add your mentor as a collaborator
7. Push your proposal PDF (created in Step 3) up to your repo
8. Copy the URL corresponding to the location of the PDF in your repo
9. Submit the copied URL using this link

Book Genres

Business Understanding

- What problem are you trying to solve, or what question are you trying to answer?
    - Which diseases/age groups/genders should hospitals, research facilities, and educational health programs focus on to help prevent mortality?
- What industry/realm/domain does this apply to?
    - Healthcare
- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)
    - The healthcare field has always interested me and while there has been a ton of research in COVID-19, I'd like to look into something other than COVID this time.

Data Understanding

- What data will you collect?
    - Gender, age, diseases, did they die,
- Is there a plan for how to get the data (API request, direct download, etc.)?
    - Kaggle data set
- Are the features that will be used described clearly?
    - Yes, it's a 0 or 1 answer in the data set.

Data Preparation

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
    - encoding
- What are some of the cleaning/pre-processing challenges for this data?
    - 0 or 1 type data is pretty easy to understand, so I don't see too many challenges other than making sure that the data is

Modeling

- What modeling techniques are most appropriate for your problem?
    - Hierarchical data model or Relational data model
- What is your target variable? (Remember - we require that you answer/solve a supervised problem for the capstone, thus you will need a target)
    - Mortality
- Is this a regression or classification problem?
    - Classification problem: understanding that 0 = false and 1 = true

Evaluation

- What metrics will you use to determine success (MAE, RMSE, etc.)?
    - MAE

Tools/Methodologies

- What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)?
    - Decision trees (yes and no modeling)