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 proposal submission is 3/12**

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 charles-rice) 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

Classifying Amazon Reviews

Business Understanding

- What problem are you trying to solve, or what question are you trying to answer?
  Analyze customer comments and reviews from Amazon's Echo family devices in order to predict if is a negative or positive review, algo identify the most popular applications or skills as well as the most common problems that users encounter.

- What industry/realm/domain does this apply to?
  Customer Feedback Analysis.

- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)
  Understand the main uses that are given to these devices and their common errors to improve the user experience in future system updates.

Data Understanding

- What data will you collect?
  Text Data from Amazon's Echo family devices reviews.
- Is there a plan for how to get the data (API request, direct download, etc.)?

Direct Download from Amazon.

- What are the features you'll be using in your model?
  Reviews, rating, feedback.

Data Preparation

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
  Cleaning, lower case words, ignore stop-words, tokenize, encode

- What are some of the cleaning/pre-processing challenges for this data?
  Structure data, find all stop words.

Modeling

- What modeling techniques are most appropriate for your problem?
  Decision Trees, Naive Bayes Classifier, Support Vector Machines, Random Forest, Convolutional Neural Networks.

- What is your target variable? (remember - we require that you answer/solve a supervised problem for the capstone, thus you will need a target)
  Rating

- Is this a regression or classification problem?
  Classification Problem

Evaluation

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

Tools/Methodologies

- What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)?

Decision tree.