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](#)

[project name]

Business Understanding
- What problem are you trying to solve, or what question are you trying to answer?
Predict whether a college basketball team will win march madness
- What industry/realm/domain does this apply to?
Sports /College Basketball
- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)
With March Madness coming up and Northwestern basketball is making it to the tournament for only the second time ever, I thought it would be interesting to look at college basketball data of how teams performed in the post season and predict success from that

Data Understanding
- What data will you collect?
Data from previous D1 college basketball schools
- Is there a plan for how to get the data (API request, direct download, etc.)?
Kaggle
- What are the features you'll be using in your model?
Wins/Losses, ratings, stats (field goal percentage, offensive/defense efficiency, etc)

Data Preparation
- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?

One Hot encoding
- What are some of the cleaning/pre-processing challenges for this data?

Checking for null values and deciding whether to remove or fill with average value (don't want to skew data)

Modeling
- What modeling techniques are most appropriate for your problem?
- What is your target variable? (remember - we require that you answer/solve a supervised problem for the capstone, thus you will need a target)

If a team won march madness or not
- Is this a regression or classification problem?

Classification

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

Accuracy, MAE

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

Random Forests

Dataset: College Basketball Dataset | Kaggle