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 end-of-day 3/13 for Cohort 3b

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 jvntra) 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 for Cohort 3b

NBA Salary Predictions

Business Understanding

- What problem are you trying to solve, or what question are you trying to answer?
    - Using player statistics, I will be trying to determine the key indicators of NBA career earnings. The reason being that owning an NBA team can be expensive. If you're able to reasonably predict how much a player will cost throughout their career, you're more able to understand how much a player should cost.
- What industry/realm/domain does this apply to?
    - Sports, specifically the NBA.
- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)
    - I really enjoy everything basketball. I initially was interested in predicting wins of college basketball teams. As I looked at the datasets available, I realized that it would be very hard to make accurate predictions in the timeframe allotted to this project. So, I decided to check out the NBA and figured that this project has a much more attainable scope within the allotted time.

Data Understanding

- What data will you collect?

    - Player statistics such as Points per game, assists per game, rebounds per game, etc. I also will collect player salary data.
- Is there a plan for how to get the data (API request, direct download, etc.)?
    - I found two datasets on Kaggle that references basketball_reference.com's players via unique ID's
- What are the features you'll be using in your model?
    - Some of the features I'll be using are the player's points per game, assists per game, rebounds, salaries, player ID's

Data Preparation

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
    - Most likely will need to use one hot encoding.
- What are some of the cleaning/pre-processing challenges for this data?
    - Connecting the two databases, filtering out empty data, as well as entries that don't match up or have an entry in the other database.

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)
    - The target variable is the player salary in the salaries database.
- Is this a regression or classification problem?
    - This is a regression problem.

Evaluation

- What metrics will you use to determine success (MAE, RMSE, etc.)?
    - RMSE and R^2

Tools/Methodologies

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