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

[Locational layoff projections]

Business Understanding

- What problem are you trying to solve, or what question are you trying to answer?
    - Future projection of Layoff across the labor industry and locations.

- What industry/realm/domain does this apply to?
    - The data provides information on the labor industry defining what area are having layoffs. I will target the labor industry and location

- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)
    - Projecting Career stability within the labor industry by location. As I work as US government consulting practitioner. The projection will provide me with knowledge on future career planning for myself. As well used to project business opportunities available for Deloitte.

Data Understanding

- What data will you collect?

    - Layoff data collected since 2020 – current which include company name, location of the layoff, industry, Number of employees laid off, percentage of employees laid off, date, stage of funding, and funds raised by the company

- Is there a plan for how to get the data (API request, direct download, etc.)?
    - Direct download through Kaggle- downloaded data set as a file, unzipped the file and downloaded into my Jupiter notebook

- Are the features that will be used described clearly?
    - Yes

Data Preparation

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
    - Missing value adjustments, encoding, data row/column manipulation.

- What are some of the cleaning/pre-processing challenges for this data?
    - A few columns have missing values, so I have the challenge of deciding weather to set them to null or drop the values.
    - Decision on weather I want to use total number or percentage of laid off employees since both are obtaining missing values as of now.
    - Linear regression to determine the relationship of certain aspects I would like to use in my model

Modeling

- What modeling techniques are most appropriate for your problem?
    - XGBoost regression

- What is your target variable? (Remember - we require that you answer/solve a supervised problem for the capstone, thus you will need a target)
    - Layoff percentage or number of layoffs for specific industries and location

- Is this a regression or classification problem?
    - Regression problem

Evaluation

- What metrics will you use to determine success (MAE, RMSE, etc.)?
    - MAE, RSME, MSE- will find all still need to decide which metric I will be using for regression analysis. Most likely RSME

Tools/Methodologies

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