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 June 23, 2023.

# 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
   1. File should be saved in the following format:
   2. GROUP NUMBER\_DATE OF SUBMISSION (example: GROUP 8\_MAY 2)
4. [Create a project GitHub repo](https://github.com/new) (if you have yet to do so)
5. [Add your instructor as a collaborator](https://docs.github.com/en/account-and-profile/setting-up-and-managing-your-personal-account-on-github/managing-access-to-your-personal-repositories/inviting-collaborators-to-a-personal-repository) to your project repo
6. Add your Deloitte mentor and VT Advisor (when assigned) 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](https://my.learn.co/courses/543/quizzes/6212)

**Market Research for videogame reception based on tweets**

# Business Understanding

* What problem are you trying to solve, or what question are you trying to answer?

Multiplayer games rely on playerbase retention to be profitable in the long run, which is why updates are rolled out periodically. We want to measure the response of the playerbase towards said updates to be able to gauge the reception of these by creating an AI model that analyzes the sentiment of a tweet.

* What industry/realm/domain does this apply to?

Videogames Industry.

* What is the motivation behind your project?

Some of our team members likes videogames and being able to know the reception of an update for a game you like seemed like an interesting idea.

# Data Understanding

* What data will you collect?

We will use a dataset from Kaggle that contains tweets of various videogames and is already tagged as positive, neutral or negative ([Twitter Sentiment Analysis | Kaggle](https://www.kaggle.com/datasets/jp797498e/twitter-entity-sentiment-analysis?resource=download)).

* Is there a plan for how to get the data (API request, direct download, etc.)?

The idea would be to use the Twitter API to collect data but will download an already tagged dataset from Kaggle.

* Are the features that will be used described clearly?

We will use text analysis techniques with Python and Keras

# Data Preparation

* What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)

EDA, Data cleaning, One hot encoding, word classification…

* What are some of the cleaning/pre-processing challenges for this data?

The amount of words and syntax one could have for creating a negative or positive connotation in a tweet.

# 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)

Tweet sentiment: Whether the discourse surrounding the game is overall positive or negative and whether it’s tied to an update for the game.

* Is this a regression or classification problem?

Classification because we’re classifying the tweets between three categories.

# Evaluation

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

MAE and R-squared

# Tools/Methodologies

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

Decision trees, K-nearest neighbors