# 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 1/16

**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 dodgy719) 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

## Predicting the Attrition of Employees Using Decision Trees

**Business Understanding**

- What problem are you trying to solve, or what question are you trying to answer?
    - Can you determine whether an employee will leave an organization based on factors such as an employee's satisfaction of their work environment, their monthly income, and their level of involvement in their job.
- What industry/realm/domain does this apply to?
    - This problem applies to employee satisfaction and how it affects retention. This can also be further applied to almost any industry because happy employees are generally more involved and productive in their organizations.
- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)
    - I am passionate about employee well-being because it has affected me, but also affects everyone in any working setting. I believe it is important for upper leadership to pay attention to their workers' attitudes towards work and make necessary changes when employees express dissatisfaction. An employee and a

company have a symbiotic relationship that depends on both an employee's well-being (which affects productivity).

**Data Understanding**

- What data will you collect?
    - Data has been collected by IBM human resources.
- Is there a plan for how to get the data (API request, direct download, etc.)?
    - Data will be directly downloaded from Kaggle ( https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset )
- What are the features you'll be using in your model?
    - Age
    - Department
    - Education
    - Environment Satisfaction
    - Gender
    - Job Involvement
    - Job Level
    - Job Role
    - Job Satisfaction
    - Monthly Income
    - Performance Rating
    - Work Life Balance
    - Years at Company

**Data Preparation**

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
    - I will have to encode binary values such as gender (male/female) and whether the employee left the organization or not (true/false). Most values, such as job level, education, and work life balance are already encoded.
- What are some of the cleaning/pre-processing challenges for this data?
    - As with any dataset, there are missing values in various columns. I will filter out these null values because I cannot train a model to predict attrition if values are missing in the above features.

**Modeling**

- What modeling techniques are most appropriate for your problem?
    - I believe that decision trees are the most appropriate for this issue.
- 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 for this project will be whether the employee left the organization (attrition – true/false).
- Is this a regression or classification problem?
    - This is a classification problem based on the binary nature of the attrition feature that we are trying to predict.

**Evaluation**

- What metrics will you use to determine success (MAE, RMSE, Accuracy, Precision etc.)?
    - I will be using a confusion matrix to visualize my model's performance versus the actual value of the attrition feature for each employee.
    - Additionally, I will be using both Accuracy and Precision to determine my model's success.

**Tools/Methodologies**

- What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)?
    - I will be using decision trees to predict whether an employee with leave an organization based on the features listed above.