

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
 - a. File should be saved in the following format:
 - b. GROUP NUMBER_DATE OF SUBMISSION (example: GROUP 8_MAY 2)
4. [Create a project GitHub repo](#) (if you have yet to do so)
5. [Add your instructor as a collaborator](#) 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](#)

Credit Card Churn Prediction

Business Understanding

- What problem are you trying to solve, or what question are you trying to answer?
 - We are trying to solve the problem: What is the reason for customer attrition, and can we predict customers who are likely to drop/cancel their card. If we can identify risk factors or customers who are likely to drop, then we may be able to address the cause and decrease our customer churn.
- What industry/realm/domain does this apply to?
 - This is going to apply to the banking, or finance industry. However, any store that provides credit cards could find this information valuable to decrease their churn. Churn models can be valuable and applicable to businesses within many different industries.
- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)
 - Churn models are frequently used in the Customer & Marketing business area, gaining experience in churn models is ideal for increasing our ability to delivery for clients. For the business, there are

benefits for understanding what causes customer attrition.

Data Understanding

- What data will you collect?
 - We are utilizing a dataset from Kaggle that collected information from credit card users including age, gender, education, etc.
- Is there a plan for how to get the data (API request, direct download, etc.)?
 - Kaggle provides a download for a csv of the dataset that we will use to read into a Jupyter notebook.
 - <https://www.kaggle.com/datasets/anwarsan/credit-card-bank-churn>
- Are the features that will be used described clearly?
 - The features that are going to be used are described clearly by the column name.

Data Preparation

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
 - Our preprocessing is going to include one-hot encoding for the categorical variables, as well as scaling of the data.
- What are some of the cleaning/pre-processing challenges for this data?
 - One of the challenges is going to be balancing the data set. There are an uneven number of current customers vs. exited customer observations.

Modeling

- What modeling techniques are most appropriate for your problem?
 - There are quite a few modeling techniques that are appropriate for our problem including logistic regression, random forest, XGBoost, and neural networks.
- 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 called "Attrition_Flag" and it is a binary variable relaying whether a customer is currently with the company or if they exited.
- Is this a regression or classification problem?
 - This is a classification problem.

Evaluation

- What metrics will you use to determine success (MAE, RMSE, etc.)?
 - The main metrics we will use to determine success are going to include precision, recall, F-1 score, and accuracy. Feature importance is also a necessary piece to evaluate as this can impact various business strategies going forward, particularly marketing, and focused customer acquisition.

Tools/Methodologies

- What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)?
 - As stated above, our modeling algorithms are going to include decision trees, random forests, neural networks, and logistic regression.