

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](#)

Book It: Hotel Reservation Cancellation Prediction

Business Understanding

- What problem are you trying to solve, or what question are you trying to answer?
 - The problem I am trying to solve is that hotels offer reservations that are convenient for the customer to book but also face lots of cancellations. I aim to predict if a customer will cancel a reservation to help guide projections on expected revenue for the hotel.
- What industry/realm/domain does this apply to?
 - This applies to the hotel industry.
- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)
 - I thought it would be interesting to see if I could use publicly-available data to predict possible customer actions, like cancelling a reservation. During the recent 2022 Christmas holiday, I noticed how many hotels offer no cost cancellation and wondered how much of an impact it makes.

Data Understanding

- What data will you collect?

- I will be using a dataset from Kaggle:
<https://www.kaggle.com/datasets/ahsan81/hotel-reservations-classification-dataset>. This dataset includes booking information from 2017-2018 with many relevant features that will be used to predict if a customer will cancel their reservation.
- Is there a plan for how to get the data (API request, direct download, etc.)?
 - The data will be sourced from a Kaggle dataset so this will be a direct download of a CSV file.
- What are the features you'll be using in your model?
 - Lead time, average price per room, room type reservation, meal plan type and booking status.

Data Preparation

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
 - I currently don't foresee many preprocessing steps as the data is clean and workable in its current state. I do see some potentially noisy data that might need removed but (type_of_meal_plan for example) but even that could be relevant.
- What are some of the cleaning/pre-processing challenges for this data?
 - As for cleaning, I mostly will be dropping features that I don't deem relevant, but I do see a lot of the features having great value for visualization. For example, showing number of previous cancellations, time of year, number of children, etc. to demonstrate any potential relevance.

Modeling

- What modeling techniques are most appropriate for your problem?
 - Multiple linear 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)
 - The target variable is the booking status.
- Is this a regression or classification problem?
 - This is a regression problem.

Evaluation

- What metrics will you use to determine success (MAE, RMSE, Accuracy, Precision etc.)?
 - Accuracy and RMSE.

Tools/Methodologies

- What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)?
 - I am planning to use multiple linear regression and random forests.