

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](#)

[Used Car Analysis- what types are used car are good to buy in this current market]

Business Understanding

- What problem are you trying to solve, or what question are you trying to answer?
 - What types are used car should car dealerships focus on purchasing in the current market
- What industry/realm/domain does this apply to?
 - Automotive Industry
- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)
 - I am in the market for a Used car, so I want to see what type of cars have the best price. Also understand what kind of cars would be more valuable to dealer if they were looking for used cars to purchase

Data Understanding

- What data will you collect?
 - Car models, engine types, body styles etc.
- Is there a plan for how to get the data (API request, direct download, etc.)?

- Used car data set from Kaggle
 - Direct download
- What are the features you'll be using in your model?
 - Types of engines, types of fuels used, body style

Data Preparation

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
 - One-hot encoding for the types, engines, fuels, and gas mileage
- What are some of the cleaning/pre-processing challenges for this data?
 - Dealing with missing values, making sure there are no leakage

Modeling

- What modeling techniques are most appropriate for your problem?
 - Regression Techniques including 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)
 - What types of engine would be the most cost efficient
- Is this a regression or classification problem?
 - Classification

Evaluation

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

Tools/Methodologies

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