Capstone Initial Proposal

Start Assignment

Due May 26 by 9:59pm   **Points** 0   **Submitting** a website url

It's time to start planning for Capstone in earnest!

Take a look at the **Capstone Project Requirements** **(https://github.com/learn-co-curriculum/dsc-capstone-project)** page for more details on what we're looking for in a capstone project

But...we are only at the beginning of Phase 4?

We fully recognize that you have only completed Phase 3 at this point, not Phase 4. Luckily, it is fully possible to complete a capstone project using only techniques from Phase 3, so long as it meets all of the listed requirements! In other words, ***it is acceptable to propose a project with a tabular dataset containing numeric and categorical data***, **similar to the datasets you have already worked with**. It is not a problem if you feel like you already have all of the knowledge you need to get started on this project.

But we also recognize that you might want to use one of the advanced techniques introduced in Phase 4! Unfortunately we have had students propose projects based on a misunderstanding of what will be covered in this just-starting phase. So here is a brief summary of techniques we will cover, which you can use in your capstone proposal and project:

☑ Time Series ☑

- Each record represents a point in time
- **In addition to a timestamp, each record has a label that is treated as the target variable. For example, the average price of a particular product or stock.**
- You are attempting to predict labels for new unseen records, with a training process that considers the time-based trend of past records rather than just the features of the new unseen records

📖 NLP (Natural Language Processing) 📖

- Each record is a "document", a collection of words (with potential additional features)
- **Each document has a label that is treated as the target variable. For example, a set of Supreme Court cases might have labels of "upheld" or "overturned". Or, a set of newspaper articles might be labeled with the section they appeared in.**
- You are attempting to predict labels for new unseen records (documents), using preprocessing techniques and models that will be introduced in Phase 4

✵ Recommendation Systems ✵

- Each record is a "review", ideally containing information about both the reviewer and the product being reviewed (e.g. a movie, a recipe, a game)
- **Each review is labeled with a rating. This can be a "thumbs up/thumbs down" rating, or a scale like 1-5 stars. If you only have review text without a numeric or categorical rating, the dataset is not suitable for this project.**
- You are attempting to predict labels for new unseen records (reviews), so that you can recommend products that are likely to receive high reviews from a particular customer

🖼 Image Processing and Neural Networks 🖼

- (Technically neural networks can be used for any kind of input data, but we mainly show how to use them for image classification)
- Each record is an image
- **Each image is labeled in some way. For example, a set of pictures of fruit might be labeled "apple", "melon", etc.**
- You are attempting to predict labels for new unseen records (images) with either a numeric or a categorical label

In all of the examples above, we have bolded the bullet point that talks about labels. This is because choosing the label (target) for your capstone project is one of the most challenging and most important aspects of the proposal phase! Unlike in previous contexts, you need to figure out what feature of a dataset would be interesting or useful to predict. Remember that this is primarily a *predictive* project, not an *inferential* project, so we are asking a question like "Can we predict...", not something like "What is the relationship between...". Although you are certainly welcome to address the latter in your EDA phase!

Proposal Submission Format

Make a copy of **this Google doc (https://docs.google.com/document/d/1B6Zv5xgWKcYLQ3S2vcbIxPbgDv7xzFpKTBZkhuQVJQE/edit)** , write your name at the top, and fill out an answer to every question to the best of your ability.

If there is a question you can't answer, explain why and document your plan to get it answered.

If it turns out that your idea won't work for a capstone project, we will let you know! There is still plenty of time before the Capstone phase begins, and we are asking you to submit this now so that we can make sure everyone is ready to hit the ground running when we actually start.

When you have filled out the proposal document, make sure you have the sharing settings so anyone can comment, and submit the link below.