UDACITY

Back to Machine Learning Engineer Nanodegree

Capstone Proposal

REVIEW

HISTORY

Requires Changes

3 SPECIFICATIONS REQUIRE CHANGES

Dear student

Great start on this proposal! I've noted a few areas where you should add a bit more detail, but I think that you've picked a great project and you're definitely on the right track. I think you'll see that most of these things shouldn't take long to update. Almost there...keep going!

Cheers!

Project Proposal

Student briefly details background information of the domain from which the project is proposed. Historical information relevant to the project should be included. It should be clear how or why a problem in the domain can or should be solved. Related academic research should be appropriately cited. A discussion of the student's personal motivation for investigating a particular problem in the domain is encouraged but not required.

Great job giving the reader an introduction to the problem domain!

Still required:

- Please cite or link to an academic paper where machine learning was applied to this type of problem.

Suggested:

- If you include a link to your data source in this section, you can directly lift it into the 'Project Overview' section of your capstone report.

Overview" section of your capstone report.

Student clearly describes the problem that is to be solved. The problem is well defined and has at least one relevant potential solution. Additionally, the problem is quantifiable, measurable, and replicable.

> In this project, I took a Kaggle competition(Bike Sharing Demand, https://www.kaggle.com/c/bike-sharing-demand), where participants are tasked to utilize historical usage patterns with weather data to forecast bike rental demand in the Capital Bikeshare program in Washington, D.C.

Great work overall! There are a couple of other things that I'd note you should be sure to include in the `Problem Statement` section next time:

- Please specify how the problem is structured. Is it a classification or a regression?
- Please explain what the inputs and outputs for the problem will be (no need to be super detailed about the features).

The dataset(s) and/or input(s) to be used in the project are thoroughly described. Information such as how the dataset or input is (was) obtained, and the characteristics of the dataset or input, should be included. It should be clear how the dataset(s) or input(s) will be used in the project and whether their use is appropriate given the context of the problem.

Nice start! Here are a few things to be sure to include in this section next time you submit your project:

- Please be sure to specify how many data points there are in the dataset and what the total number of features is.
- Will you need to handle the data chronologically as a time-series or can the data points be randomly shuffled/split?

Student clearly describes a solution to the problem. The solution is applicable to the project domain and appropriate for the dataset(s) or input(s) given. Additionally, the solution is quantifiable, measurable, and replicable.

A benchmark model is provided that relates to the domain, problem statement, and intended solution. Ideally, the student's benchmark model provides context for existing methods or known information in the domain and problem given, which can then be objectively compared to the student's solution. The benchmark model is clearly defined and measurable.

> According to competition, the result should be benchmarked with Random Forrest Model. Random Forrest Model is an ensemble method that combining several weak learner(i.e. decision tree) to enhance and improve the overall prediction accuracy. Compare to most ensemble model, Random Forrest offers relatively high performance with shorter training time; the model can also handle

> Forrest offers relatively high performance with shorter training time, the model can also handle missing values; it can be modeled with categorical inputs; given enough trees, the model tends not to overfit.

Great choice! This is a common default implementation that shouldn't be too hard to beat. Please note that you should be sure to implement this model yourself so that you can be sure it was trained/tested using the same data as your solution.

Student proposes at least one evaluation metric that can be used to quantify the performance of both the benchmark model and the solution model presented. The evaluation metric(s) proposed are appropriate given the context of the data, the problem statement, and the intended solution.

Looks good!

Student summarizes a theoretical workflow for approaching a solution given the problem. Discussion is made as to what strategies may be employed, what analysis of the data might be required, or which algorithms will be considered. The workflow and discussion provided align with the qualities of the project. Small visualizations, pseudocode, or diagrams are encouraged but not required.

Nice job here! I think that you're meeting the specifications. Keep in mind that this is a great opportunity to bounce lots of ideas off of the reviewers. You can get feedback without putting in much work. It's also a good idea to build some backup plans into your workflow in case something doesn't work. You don't want to get stuck...

Suggested:

- The XGBoost and LightGBM models could be good supervised learning approaches to try here.
- Since you're creating multiple supervised learning models, you could try combining them all together into a custom ensemble model:

http://blog.kaggle.com/2016/12/27/a-kagglers-guide-to-model-stacking-in-practice/
https://www.kaggle.com/arthurtok/introduction-to-ensembling-stacking-in-python

Proposal follows a well-organized structure and would be readily understood by its intended audience. Each section is written in a clear, concise and specific manner. Few grammatical and spelling mistakes are present. All resources used and referenced are properly cited.

The template format is followed and the proposal is well written.

RESUBMIT

DOWNLOAD PROJECT

Learn the best practices for revising and resubmitting your project.

RETURN TO PATH

Student FAQ