REVIEW

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!

Suggested:

- Since you've included a link to your data source in this section, you can directly lift it into the 'Project Overview' section of your capstone report (to save a bit of time).

✓

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.

> I want to take a subset of their data to test and verify how prediction by details help a user to get best price of their product.

Great job! I think that this is a pretty clear statement of the problem. I'd suggest also noting that you want to treat the problem as a regression.

🔄

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:

- How many examples or data points are there in the subset of data that you'll be using?
- What are the labels or outcome variables? If there are classes, are they balanced?
- How will you split the selected subset of data into training/validation/testing sets?

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.

Excellent job! I think that this is a nice high-level statement of a viable solution.

Just one small thing you should update:

- Please note that since this is a regression problem, you should be using the regression models (not the classifiers) for this project. Just update the description of each model to indicate that you'll be using regressors rather than classifiers.

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.

> As this is like a recommended system, I want to use Decision tree as with GridSearchCV as hyperparameter tuning to save its best model.

This should be a good baseline for the project. This is a common default implementation that won't be hard to beat with the three boosting algorithms you've indicated you'll use.

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.

RMSE is a good metric for this project. Nice job!

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.

This is a very good start on your `Project Design` section. When you resubmit, please add just a bit more detail to your workflow:

- For instance, what are all of the steps that you'll take during pre-processing? What methods will you take for dealing with missing values (imputing them, dropping the data points etc.)? Do you have any

approaches for detecting outliers?
- How will you vectorize or embed the text data for input into your models?
- What are all of the supervised learning techniques you plan on trying (any others besides AdaBoost, GradientBoosting, and XGBoost)?
- How will you tune or refine your supervised learning algorithms (hint: a randomized search could allow you to search larger parameter spaces)?

We understand that you can't know exactly what your final implementation will look like...you haven't done it yet! However, you'll need to give us a clearer picture of what you want to try so we can make sure that you're on the right track and won't end up wasting time on things that don't work out.

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...

Optional Suggestions:

> Normalization, if required

- Tree-based learning models do not require any scaling or normalization. You can certainly do this, but it wouldn't change the results.
- The LightGBM model could also be a good supervised learning approach 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.