PROJECT

Capstone Proposal

A part of the Machine Learning Engineer Nanodegree Program

PROJECT REVIEW | NOTES

Requires Changes

SHARE YOUR ACCOMPLISHMENT

2 SPECIFICATIONS REQUIRE CHANGES

Hi,
Good job so far! You have selected an interesting Kaggle challenge for your capstone and you have a sound approach for solving it. There are just a few aspects to be amended in your proposal, these should not require too much of your time.

Best Regards.

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.**

Awesome

- Good overview of supervised learning followed by an introduction to the allstate claims severity challenge.

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.

Required

- When you use any material from others make sure to include a proper reference. In this case, the first paragraph here is a good overview of the real world problem, still it was copied from the challenge page and should be properly referenced. Something among the lines: "According to the Kaggle challenge: "..." ". It is really important to always refer to others work, be it code, a paragraph, an image, etc.
- The first paragraph presents the problem from the real world perspective, still it is necessary that you clearly discuss here the problem from a machine learning point of view. For example, you should mention what are the inputs (images? text? structured data? ...), the output (1 out of 10 possible class labels? a continuous value?), the specific learning task (regression? classification?). Out of these 3 you already mention the output (total cost of a claim) and that you are dealing with supervised learning (could be more specific though, mention regression explicitly). There is only a lack of discussion about the format of the input to clearly define the scope of this problem.

✓ **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.**

Awesome

- Good job including a link to where the data can be acquired
- There is an overview about the data set, including relevant characteristics and statistics (e.g. number of claims and features).

✓ **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.**

Awesome

You describe your overall solution to the problem and the learning algorithms that you plan on using.

✓ **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.**

Awesome

An SVM model is a reasonable benchmark model for this project.

✓ **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.**

Awesome

Good job on selecting MSE and providing an overview of how it is calculated.

✓ **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.**

Awesome

- Good job on providing sufficient information about how you plan to apply your solution.
- PCA is a reasonable approach to reduce the dimensionality of your data.

Suggestion

- Besides PCA for step 2 you should also try feature selection, for example, using SelectKBest.
- Consider using GridSearch for step 5.

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.

Awesome

In general the proposal is well organized and can be easily understood.

Required

To completely pass this rubric you should remember to properly cite all resources used (see the first requirement in the Problem Statement comments).

RESUBMIT PROJECT

DOWNLOAD PROJECT

Learn the best practices for revising and resubmitting your project.

RETURN TO PATH

Student FAQ