UDACITY

< Return to Classroom

Capstone Proposal

REVIEW
HISTORY

Meets Specifications

Good selection of project for capstone proposal, challenging and not an easy choice!!! Its was fun and interesting read. Well researched problem domain. Excellent work explaining problem statement, dataset inputs, solution, evaluation metrics, benchmark and design approach to diagnose malignant breast cancer from images problem.

Well done on each section documentation, that's reflect the hard work you did to finish this proposal!! Want to congratulate you for getting this far and finishing important milestone towards completing this nanodegree program.

Wish you all the best for final capstone machine learning project!!

Project Proposal

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.

Well done!! In this section we have to define the problem statement in terms of machine learning task like whether its classification vs regression problem or unsupervised learning problem of clustering. So would recommend to mention its supervised binary classification task to predict malignant cancer presence from the xray images.

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.

Excellent work on problem background discussing how machine learning can be used to diagnose or predict disease from the the medical image. This point "increase survival rates from 56% to more than 86%" makes this use case strong and significant to purse further research on identification of malignant breast tumors.

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!! Well done justifying the evaluation metric for this case study sensitivity, Specificity and Accuracy.

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.

Well done choosing SVC model as presented in research as model benchmark. Its challenging benchmark score!! Nice selection of XGBoost algorithm. My recommendation would be leverage transfer learning approach on CNN based pretrained deep learning model like VGG architecture.

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.

Well done citing the source from where data can be obtained. Good work mentioning the size of dataset, listing 10 cytology features and output classification. Would recommend to also mention the distribution of images positive and negative samples in dataset. This will helps us to understand skewness in data.

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.

Good work discussing how cytology features enable ML model to provide an accurate diagnosis. Would recommend to mention at least one potential ML algorithm you plan to implement to solve this problem.

Student summarizes a theoretical workflow for approaching a solution given the problem. A 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.

Well done summarizing the key steps in this section. For EDA my recommendation would be also analyze pixel intensity distribution to check if you see any pattern for malignant cancer images. To analyze correlations probably you can use sns heatmaps. Do remember to scale or normalize the image data before feeding it into ML model. Good strategy applying PCA but ensure that maximum variance is captured in data, my recommendation would be to capture at least 90% variance.

Good work citing references on Bibliography section.

The 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.

Well organised proposal document!!

⬇ DOWNLOAD PROJECT

RETURN TO PATH