Udacity Mentors Review

Capstone Project:

Predicting Life Expectancy rate of a Country

Meets Specifications

This is a solid solution to an interesting problem, and your report is well written. You've clearly communicated your full understanding of the machine learning process, by discussing your methods and results in great detail. Overall, I have no major concerns here. Congratulations on passing your capstone!

Definition

Student provides a high-level overview of the project in layman’s terms. Background information such as the problem domain, the project origin, and related data sets or input data is given.

- Solid introduction to the problem you're solving, with a good focus on the practical importance of a good solution
- It's clear how machine learning is a viable solution in this situation, based on your discussion here
- The source and overall goal of your project are nicely addressed

The problem which needs to be solved is clearly defined. A strategy for solving the problem, including discussion of the expected solution, has been made.

- The input and output are well defined, which makes for a solid problem statement
- Your approach to the problem is clear, and certainly suitable. Having a concrete and detailed plan in place is the best way to start a project, since it gives you clear direction and helps you understand where you are in the process at any given time

Metrics used to measure the performance of a model or result are clearly defined. Metrics are justified based on the characteristics of the problem.

- This metric makes sense for the data type of the prediction, and is well defined here
- Your choice has been clearly justified based on its characteristics and how they align with your goals for the model

Analysis

If a dataset is present, features and calculated statistics relevant to the problem have been reported and discussed, along with a sampling of the data. In lieu of a dataset, a thorough description of the input space or input data has been made. Abnormalities or characteristics of the data or input that need to be addressed have been identified.

- The source of the data, its size, the features and their structure are all clearly described, which makes for a solid overview of the key characteristics

- The descriptive statistics give a good idea of the distributions contained in the dataset, which is an important property to explore before modelling, as this effects some assumptions in algorithms, among other things

A visualization has been provided that summarizes or extracts a relevant characteristic or feature about the dataset or input data with thorough discussion. Visual cues are clearly defined.

- This is an important data quality and certainly something well worth visualizing; good use of this section here
- The visualization itself is clean and well presented, with appropriate labels and identifiers, and the right visual encoding for the data type

Algorithms and techniques used in the project are thoroughly discussed and properly justified based on the characteristics of the problem.

- The algorithms you're using are clear, and you did a good job of justifying your choices
- This is a well written, detailed, and most importantly intuitive and easily understood explanation of the algorithms that you're using. It's most important that we give our readers a solid grasp on the techniques that we're presenting as our solution, and you've done that well

Student clearly defines a benchmark result or threshold for comparing performances of solutions obtained.

- Comparing to well defined, objective, concrete models / results is always the best approach for obtaining a baseline for our own work, and that's what you have here. Good choice

Methodology

All preprocessing steps have been clearly documented. Abnormalities or characteristics of the data or input that needed to be addressed have been corrected. If no data preprocessing is necessary, it has been clearly justified.

- This is a solid step by step overview of the work required to prepare your data for proper training, and again it's written in a way that's clear and detailed
- It's easy to see what the structure of the data would be before and after each of these transformations

The process for which metrics, algorithms, and techniques were implemented with the given datasets or input data has been thoroughly documented. Complications that occurred during the coding process are discussed.

- The goal of this section is to make our work as reproducible as possible; for any future researchers that read your work and wish to expand on it, they'll have to start by re-implementing what you have done, and they can only do that if your explanation of your work through this report is detailed and accurate. You've certainly met that requirement with your discussion here

The process of improving upon the algorithms and techniques used is clearly documented. Both the initial and final solutions are reported, along with intermediate solutions, if necessary.

- The hyperparameters tuned, the values tried, and the results obtained are all clearly and cleanly recorded, fully characterizing the refinement process

Results

The final results are compared to the benchmark result or threshold with some type of statistical analysis. Justification is made as to whether the final model and solution is significant enough to have adequately solved the problem.

The final model’s qualities—such as parameters—are evaluated in detail. Some type of analysis is used to validate the robustness of the model’s solution.

- Your final results are presented in a way that's easy to analyze and compare, with good surrounding discussion
- Robustness is explicitly addressed by means of an objective test, and the results are thoroughly discussed. Good work