UDACITY

‹ Back to Machine Learning Engineer Nanodegree

Finding Donors for CharityML

REVIEW
HISTORY

Requires Changes

1 SPECIFICATION REQUIRES CHANGES

Dear student,

Awesome submission 👍

You have corrected and re-ran most of the sections and I really appreciate your work.

You missed two points :

- Explanation of reason of selection of your selected models
- Algorithm suitability of your final model.

I am trying to help you in understanding the models really well and it will help you when you work on real world applications as well as on kaggle competitions.

I am sure the required changes won't take much time and it is worth your time and I hope your next submission will meet project requirements.

Keep up the good work! I look forward to next submission.

Exploring the Data

Student's implementation correctly calculates the following:

- **Number of records**

- **Number of individuals with income >$50,000**

- **Number of individuals with income <=$50,000**

- Number of individuals with income <=$50,000
- Percentage of individuals with income > $50,000

Preparing the Data

Student correctly implements one-hot encoding for the feature and income data.

Evaluating Model Performance

Student correctly calculates the benchmark score of the naive predictor for both accuracy and F1 scores.

The pros and cons or application for each model is provided with reasonable justification why each model was chosen to be explored.

Please list all the references you use while listing out your pros and cons.

Required changes

> RandomForest : the model is suitable for this problem since it can handle the categorical data input; it also returns feature importances which helps in identify the major features affecting the outcomes
> Adaboost : the model returns feature importances; it helps identifying the key factors in affecting the outcome.
> Gradient boosting : the model returns feature importances; it helps identifying the key factors in affecting the outcome.

It seems you just mentioned the same reason of selection for all the models. Does feature importance is just enough to select the models? If that is the only reason why not select only one of ensemble methods and select some other supervised models like svm, logistic regression etc.. to test their performance against ensemble method?

The reason of selection must be in both data set properties as well as in model strengths. The question is what makes you choose this particular model for the given dataset?

Student successfully implements a pipeline in code that will train and predict on the supervised learning algorithm given.

Nice implementation of pipeline!

Student correctly implements three supervised learning models and produces a performance visualization.

Improving Results

Justification is provided for which model appears to be the best to use given computational cost, model performance, and the characteristics of the data.

Nice explanation and I agree with your reason of selecton of your final model.

Suggestion

> the algorithm's suitability for the data.

Please answer this question as well.

Student is able to clearly and concisely describe how the optimal model works in layman's terms to someone who is not familiar with machine learning nor has a technical background.

Nice explanation!

The final model chosen is correctly tuned using grid search with at least one parameter using at least three settings. If the model does not need any parameter tuning it is explicitly stated with reasonable justification.

Nice implementation using grid search.

Student reports the accuracy and F1 score of the optimized, unoptimized, models correctly in the table provided. Student compares the final model results to previous results obtained.

Great work and you updated correctly for both optimized and unoptimized models.

Though you updated the scores correctly I am making this section as required due to you need to update previous sections with clear answers.

I can't revert previous review meets specification sections. I am trying to help you to understand the concepts clearly and it comes handy when you work on real world applications.

Feature Importance

Student ranks five features which they believe to be the most relevant for predicting an individual's' income. Discussion is provided for why these features were chosen.

Student correctly implements a supervised learning model that makes use of the `feature_importances_` attribute. Additionally, student discusses the differences or similarities between the features they considered relevant and the reported relevant features.

Student analyzes the final model's performance when only the top 5 features are used and compares this performance to the optimized model from Question 5.

☑ RESUBMIT

⤓ DOWNLOAD PROJECT

Learn the best practices for revising and resubmitting your project.

RETURN TO PATH

Student FAQ