UDACITY

PROJECT

Finding Donors for CharityML

A part of the Machine Learning Engineer Nanodegree Program

PROJECT REVIEW

NOTES

SHARE YOUR ACCOMPLISHMENT!

Requires Changes

8 SPECIFICATIONS REQUIRE CHANGES

Dear student,

Great submission

I am really impressed with your analysis of each of model pros, cons and reason of selection but you have not answered most of the sections.

Please make sure you read the instructions clearly before implementing the code. I have provided suggestions for each of required sections. I am sure the required changes won't take much time and it is worth your time.

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

You correctly calculated the dataset statistics. Good work!

If you observed the statistics it is clearly indicating the classes (individuals with income > $50k = 11208 and individuals with income atmost $50k = 34014) are imbalanced. Please check this link to understand how to deal with imbalanced data.

Preparing the Data

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

Nice implementation using `get_dummies` method.

Evaluating Model Performance

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

You wrongly calculated both accuracy and f-score. Please re-check and implement them correctly.

You could check this link for further understanding of precision and recall.

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.

Nice explanation of each of selected models pros, cons and reason of selection.

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.

> Use a 'random_state' for each model you use, if provided.

Make sure you set the random_state for your selected models if provided in sklearn library.

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.

please answer this section.

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.

please answer this section.

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.

If you observe the graphs decision tree is not the best model. Most of the times the final model selection depends on how well the model is doing on out of sample data (test data). The training and testing times are considered based on the context and computing resources.

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.

Please update this section after you correctly choose the final model.

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.

Nice selection of features but make sure you explain why you selected those features and how they relevant to individuals income?

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.

Please re-run this section after you correctly choose the final model and update the description as per results.

RESUBMIT

DOWNLOAD PROJECT

Best practices for your project resubmission

Ben shares 5 helpful tips to get you through revising and resubmitting your project.

Watch Video (3:01)

RETURN TO PATH

Rate this review

Student FAQ