UDACITY

Logout

A part of the Machine Learning Engineer Nanodegree Program

PROJECT REVIEW | CODE REVIEW | NOTES

Requires Changes

SHARE YOUR ACCOMPLISHMENT

1 SPECIFICATION REQUIRES CHANGES

Excellent work! Very good first submission!

Classification vs Regression

Student is able to correctly identify which type of prediction problem is required and provided reasonable justification.

Awesome

- Great job! Yes this is a classic classification problem!

Exploring the Data

Student response addresses the most important characteristics of the dataset and uses these characteristics to inform their decision making. Important characteristics must include:

- **Number of data points**
- **Number of features**
- **Number of graduates**
- **Number of non-graduates**
- **Graduation rate**

Preparing the Data

Code has been executed in the iPython notebook, with proper output and no errors.

Training and test sets have been generated by randomly sampling the overall dataset.

Awesome

- Very good job here implementing a train-test split.

Suggestion

- Note that an even better Validation method to use would be a stratified method for example, StartifiedShuffleSplit due to the class imbalance. Or this can also be done by using the 'stratify' parameter in train/test split. This is just something to consider.

Training and Evaluating Models

Three supervised models are chosen with reasonable justification. Pros and cons for the use of each model are provided, along with discussion of general applications for each model.

Awesome

- Great job providing the pros and cons of the models, as well as their applications and why they were selected.

All the required time and F1 scores for each model and training set sizes are provided within the chart given. The performance metrics are reasonable relative to other models measured.

Awesome

- Good job implementing the training for the models selected.

Required

- A random state should be set for the algorithms that accept them, this is to ensure that the same results are obtained every time the code is run.

Suggestion

- Note that this could also have been achieved with fewer code. Consider -

```
for clf in [GaussianNB(), RandomForestClassifier(),  SVC()]:
    for size in [100, 200, 300]:
        train_predict(clf, X_train[:size], y_train[:size], X_test, y_test)
```

Choosing the Best Model

Justification is provided for which model seems to be the best by comparing the computational cost and accuracy of each model.

Awesome

- Great job here! SVM is chosen after exploring the performance in terms of F1 score and train/prediction time

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

Awesome

- Well done! The chosen model is described in very simple terms.

Suggestion

- Note that some more discussion may be included, some very important concepts that could be discussed in layman include the kernel trick e.t.c

The final model chosen is correctly tuned using gridsearch 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.

Awesome

- Gridsearch run for 2 parameters, each with more than 3 settings.

The F1 score is provided from the tuned model and performs approximately as well or better than the default model chosen.

Awesome

- Great job! The same test F1 score does imply we already have optimal parameters.

Quality of Code

Code reflects the description in the documentation.

RESUBMIT PROJECT

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)

Have a question about your review? Email us at review-support@udacity.com and include the link to this review.

RETURN TO PATH

Student FAQ