

Identifying Fraud from Enron Email

Meets Specifications

[Code Review](#) [Project Feedback](#)

Quality of Code

Meets Specifications

Dataset, list of features and algorithm are exported using code in poi_id.py, so that it can be checked easily using tester.py.

Our Assessment

Just one little nitpick. A line of code like this:

```
new_feature.add_poi_interaction(my_dataset, features_list)
```

even though it's commented out, should really be removed from the final poi_id.py version. It's no longer used and the new_feature module wasn't included in the final submission.

Full rubric ▼

Meets Specifications

Code reflects the description in the documentation.

Our Assessment

This report is very good, well organized and easy to understand. The one thing I want to point out is sometimes you make some claims without

backing them up with numbers. Also it seems like some things from the first submission were mistakenly left in the report. You'll see I talk about these things in the review. It's always good to read your own work with a critical eye before turning it in. But in general, this submission is much more thorough and meticulous compared to the first submission, so I congratulate you.

Full rubric ▼

Pick and Tune an Algorithm

Meets Specifications

Response addresses what it means to perform parameter tuning and why it is important.

Our Assessment

Awesome Job!

Full rubric ▼

Meets Specifications

At least 2 different algorithms attempted, and their performance is compared with the more performant one used in the final analysis.

Our Assessment

Ideally, in the report you'd include a little bit on the precision and recall values you got from Gaussian Naive Bayes to convince me that it didn't work, and you needed to use a different algorithm.

Full rubric ▼

Meets Specifications

At least one important parameter tuned, with at least 3 settings investigated; or any of the following are true:

- GridCV used for parameter tuning
- Several parameters tuned
- Parameter tuning incorporated into algorithm selection (i.e. parameters tuned for more than one algorithm, and best algorithm-tune combination selected for final analysis).

Our Assessment

This was mentioned in the last review, but consider using GridSearchCV for your future work in machine learning. It's just a short cut to doing the manual parameter tuning done in poi_id.py

Full rubric ▼

Optimize Feature Selection/Engineering

Meets Specifications

At least one new feature implemented. Justification for that feature is provided in the written response, and the effect of that feature on the final algorithm performance is tested.

Our Assessment

It's great that you made a feature and it improved the algorithm. In the report, it would've been awesome if you had included precision/recall or some other indication of how it improved the algorithm.

Full rubric ▼

Meets Specifications

Univariate or recursive feature selection is deployed, or features are selected by hand (different combinations of features are attempted, and the performance is documented for each one). For an algorithm that supports getting the feature importances (e.g. decision tree), those are documented as well.

Our Assessment

Thanks for including the work you did in the ipython notebook as well as the `poiid.py` file. *Otherwise, I wouldn't have been able to see how you selected features by hand. You still didn't include feature importances of the decision tree, which was pointed out in the last review as well. I think the rest of the report meets specifications, so I won't mark it off just for not including feature_importances* Please read about feature importances in <http://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>

Full rubric ▼

Meets Specifications

If algorithm calls for scaled features, feature scaling is deployed.

Our Assessment

Nice job recognizing when feature scaling is necessary. When referring to Gaussian Naive Bayes, I wasn't sure what was meant by "Features in the model are scaled automatically depending on their assigned coefficients which renders feature scaling useless." That sounds more like linear or logistic regression to me. Here's a really good explanation of Naive Bayes <http://stackoverflow.com/questions/10059594/a-simple-explanation-of-naive-bayes-classification>

Full rubric ▼

Understanding the Dataset and Question

Meets Specifications

Student response addresses the most important characteristics of the dataset and uses these characteristics to inform analysis. Important characteristics include:

- total number of data points
- allocation across classes (POI/non-POI)
- number of features used

Our Assessment

Awesome Job!

Full rubric ▼

Meets Specifications

Student response identifies outlier(s) in the financial data, and explains how they are removed or otherwise handled.

Our Assessment

Great job finding the TOTAL outlier. I especially liked the visualization in the ipython notebook. There's also another row of bad data called 'TRAVEL AGENCY IN THE PARK' that isn't exactly an outlier, but it isn't a person either. You'd only find it by manually inspecting the data.

Full rubric ▼

Validate and Evaluate

Meets Specifications

Precision and recall are used to evaluate performance performance, and student articulates what those metrics measure.

Our Assessment

Awesome Job!

Full rubric ▼

Meets Specifications

The data is split into training and testing sets, with the testing data used for assessing overall analysis performance; or k-fold cross validation is deployed.

Our Assessment

Awesome Job!

Full rubric ▼

Meets Specifications

Precision and recall are both at least 0.3.

Our Assessment

I think you left your old values of precision and recall in the report from the first submission. The very last line of the report says "My precision is .40 My recall is .349" even though earlier in the report you correctly state that it's Precision: 0.50516 Recall: 0.41600. Make sure when resubmitting a project, the report really reflects work done for the second submission.

Full rubric ▼

Meets Specifications

Response addresses what validation is and why it is important.

Our Assessment

Awesome Job!

Full rubric ▼

You rocked it! [Provide feedback on your review](#)

INFORMATION

[Nanodegree Credentials](#)
[Udacity for Organizations](#)
[Help and FAQ](#)
[Feedback Program](#)

COMMUNITY

[Blog](#)
[News & Media](#)
[Developer API](#)

UDACITY

[About](#)
[Jobs](#)
[Contact Us](#)
[Legal](#)

FOLLOW US ON

MOBILE APPS



Nanodegree is a trademark of Udacity

© 2011-2015 Udacity, Inc.