UDACITY

PROJECT

Finding Donors for CharityML

A part of the Machine Learning Engineer Nanodegree Program

PROJECT REVIEW

NOTES

SHARE YOUR ACCOMPLISHMENT!

Requires Changes

9 SPECIFICATIONS REQUIRE CHANGES

Dear student,

Great first submission 👍

You have got great understanding on each of selected models and I really appreciate your explanations of pros, cons and reason of selection for selected models.

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**

Good job!

Preparing the Data

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

> Use pandas.get_dummies() to perform one-hot encoding on the 'features_log_minmax_transform' data.

You need to encode normalized and log transformed features. You could write like :

```
features_final = pd.get_dummies(features_log_minmax_transform)
```

Suggestion

After you correctly encode the features, please re-run the train test split function.

Evaluating Model Performance

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

Good job! You correctly calculated both accuracy and f-score.

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 only default models with random_state for your selected models.

Make sure you re-run this section after you correctly encode the features.

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.

If you observe the graphs, it clearly showing the decision tree is doing good on training data compared to test data and you can say the decision tree is overfitting to training data but in case of RandomForest it is better both on training as well as testing. So it is not overfitting to training data and predicts very well compared to decision tree.

Please re-check your model selection and re-run the models after you correctly encode the features.

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 update this section after you correctly choose your final model.

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.

Please re-run your final model after you correctly encode the features and make sure you correctly selected your final model.

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 re-run your final model.

Make sure you update the results in respective model boxes.

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.

You could explain hours per week and age like below :

- hours-per-week : If we considered two persons with same hourly rate then the person have higher hours-per-week will earn more compared to other one.
- age : Usually older people earn more compared to younger people

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.

> If you were not close, why do you think these features are more relevant?

Please answer this question as well. Why do you think martial status and capital gain are more relevant compared to your previous selected ones?

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.

I would recommend you to re-run this section acutally we are using the same final model here on both full and reduced data and I guess the results will change after you run your final model with correct encoded features.

Please update the description as per results.

RESUBMIT

DOWNLOAD PROJECT

Learn the best practices for revising and resubmitting your project.

RETURN TO PATH

Student FAQ