

[← Back to Machine Learning Engineer Nanodegree](#)

Finding Donors for CharityML

REVISÃO

REVISÃO DE CÓDIGO

HISTORY

Requires Changes

4 ESPECIFICAÇÕES NECESSITAM DE MUDANÇAS

Seu projeto está indo muito bem. Apenas mais alguns ajustes e estará completo. Bons estudos!

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

Ótima implementação dos cálculos.

Preparing the Data



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

Aqui é necessário revisar a implementação do `get_dummies()`. Leia a instrução novamente. Em que você deve usar o `get_dummies()`, segundo as instruções?

```
# TODO: Utilize o one-hot encoding nos dados em 'features_log_minmax_transform' utilizando pandas.get_dummies()
```

Boa implementação do mapeamento dos valões de 50K para 0 ou 1.

Você também poderia usar uma função lambda:

```
income = income_raw.apply(lambda n: 0 if n == '<=50K' else 1)
```

Evaluating Model Performance



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

Excelente. Os valores são esses mesmos.



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.

As suas descrições e detalhamento dos modelos estão ótimos. Entretanto aqui é necessário aprofundar um pouco mais no porque usar esses modelos. Existem muitos modelos de ensemble. Porque você escolheu esses 3 em particular? Que detalhes sobre eles leva você a crer que eles farão o trabalho?

Sugestão: Do ponto de vista acadêmico, talvez fosse interessante você testar ao menos uma técnica diferente de ensemble. Isso pode ajudar você a ganhar perspectiva sobre as diferenças de operação dos modelos (Velocidade de Train/test, dificuldade de Tuning...etc)

Para referencias futur:

Esse [Algorithm Selection Cheat sheet](#), e esse outro [chart aqui](#), são boas referências.

Lembre-se que são apenas guias gerais e não devem ser levados "a ferro e fogo". Mas ajudam se não sabemos por onde começar.



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

Correta implementação.



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

Muito boa implementação. Fez muito bem em usar o random state

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.

Ótima justificativa quanto ao custo computacional e performance.



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.

Está indo na direção certa. Mas ainda é necessário mais elaboração. Lembre-se que aqui apenas é necessário explicar o funcionamento do modelo em termos leigos. Uma pessoa sem background matemático(Além do segundo grau) ou de programação, deve ser capaz de com a sua explicação entender o funcionamento desse algoritmo. Descreva com mais detalhes o que é um weak learner. Como ele faz esse treinamento? Etc...



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.

Aqui é necessário usar ao menos 3 valores para ao menos 1 parâmetro na hora de afinar o modelo. Tem algumas linhas que parecem atender a esse requisito, mas estão como "Comment".



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.

Excelente avaliação.

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.

Boa intuição na escolha de features.



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.

Sua avaliação quanto ao Marital Status está muito boa. As features_importance variam de acordo com Random state e o algoritmos selecionado. Mas não variam muito. Hours-per-Week também é um parâmetro que aparece muito como importante. Mas o importante aqui é questionar sempre. Bom trabalho.



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.

Muito bom. Lembre-se que é sempre importante considerar o custo de processamento de um algoritmo. Muitas vezes, sacrificar um pouco de precisão para melhorar o custo do modelo é ideal. Afinal, nem sempre temos como prever que a base de dados vai triplicar em 6 meses e o modelo começará a "pesar".

 REENVIAR

 BAIXAR PROJETO



Melhores práticas para sua resubmissão do projeto

Ben compartilha 5 dicas úteis para a revisão e resubmissão do seu projeto.

 Assistir Vídeo (3:01)

RETORNAR

[FAQ do Estudante](#)