

[Return to "Machine Learning Engineer Nanodegree" in the classroom](#)

Finding Donors for CharityML

REVIEW

HISTORY

Meets Specifications

Hi there, it's Cláudio! Thanks for sending all the required files for the review process and for all code executing without any issue.
Congratulations for this project submission and for the quality presented in this project. You really did a great job.

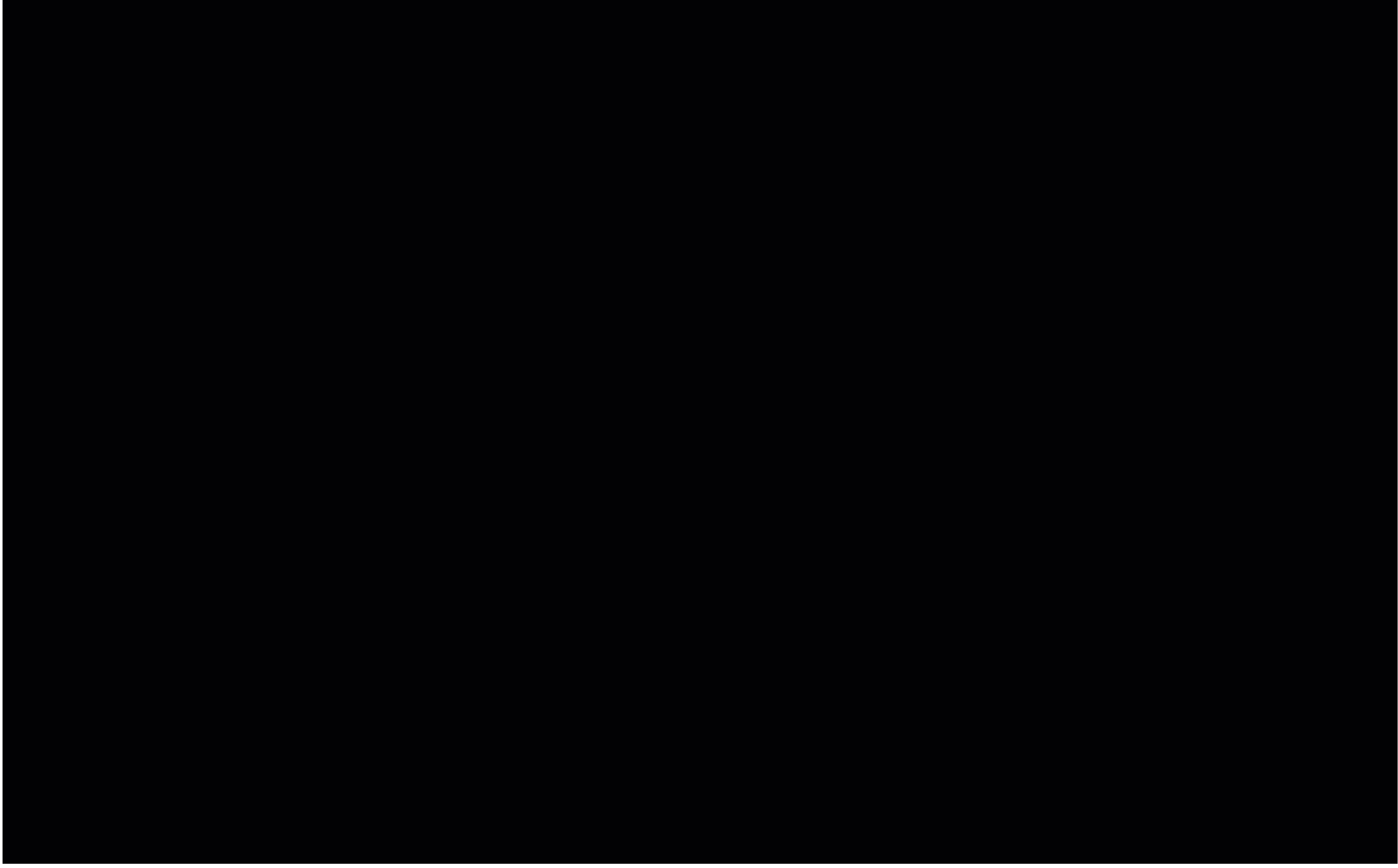
I hope you had enjoyed doing this project and put in practice good and important concepts from machine learning. I will leave my contact below in case you have any doubt about this review as well to get connected.

That's all. Enjoy learning and keep it up the great work!

Finally, I wanted to share a interesting tool from google that helps machine learning engineers to understand the data really fast and then make a decision on what type of algorithm it will fit better that data, called: Facets - Visualizations (<https://pair-code.github.io/facets/>). Definitely check it out.

The power of machine learning comes from its ability to learn patterns from large amounts of data. Understanding your data is critical to building a powerful machine learning system.

Facets contains two robust visualizations to aid in understanding and analyzing machine learning datasets. Get a sense of the shape of each feature of your dataset using Facets Overview, or explore individual observations using Facets Dive.



Thank you.
Cláudio

Email: cglmenest@uol.com.br
Linkedin: <https://www.linkedin.com/in/claudiogimenestoledo/>

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

Preparing the Data

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

Evaluating Model Performance

- ✓ Student correctly calculates the benchmark score of the naive predictor for both accuracy and F1 scores.
- ✓ 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.
- ✓ Student successfully implements a pipeline in code that will train and predict on the supervised learning algorithm given.
- ✓ Student correctly implements three supervised learning models and produces a performance visualization.

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.
- ✓ 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.

Good job. This is a very important step that we should carefully consider when explaining to all type of audience. Usually not all in the audience knows deeply the technical aspects and it's up to us to open the "black box" and convey a clear message on why we are choosing that model and that is the best solution we may have for the given problem.

Suggestion:
 - Usually it's a good idea to provide link for further reference as well images, diagrams and business cases when describing in layman's terms.
Bonus:
 - I will leave here some good resources about layman's terms in AI:
<https://www.quora.com/How-can-you-explain-artificial-intelligence-in-laymans-terms>
<https://stopad.io/blog/artificial-intelligence-facts>
<http://blog.aylien.com/10-machine-learning-terms-explained-in-simple/>
- ✓ 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.

Great job. You have correctly implemented the model tuning using the grid search. You have provided all the requirements from this part of the code for the model tuning.

Bonus:
 - GridSearch is not the only technique available to us though! Another similar technique worth looking is RandomizedSearchCV
 - Here I will leave some good articles about this subject:
https://en.wikipedia.org/wiki/Hyperparameter_optimization
<https://machinelearningmastery.com/grid-search-hyperparameters-deep-learning-models-python-keras/>
<https://www.quora.com/Machine-Learning-How-does-grid-search-work>
- ✓ 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.

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.

Great job.

You have done a great job indicating and ranking those features.

Suggestion:
 - It's always a good idea to put some science behind of our opinions when it's possible. You can try to reinforce your arguments using statistics and show the correlations between then or show a sample where, for instance, you have higher occupations making more money than others.
- ✓ 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.

 [DOWNLOAD PROJECT](#)

RETURN TO PATH