U D A C I T Y

Finding Donors for CharityML

REVIEW
HISTORY

Meets Specifications

Dear student,

Great re-submission 👍🏼

You have corrected and re-ran most of the sections and I really appreciate your work. I would recommend you to work on some kaggle competitions on supervised models. Initially you just observe the winning solutions and their reason of selection of model for given data set. Later you can test with your own choice and compare the results. And this is a good way to strengthen your understanding and easy to work on real world applications as well.

keep up the good work!

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.

Nice implementation using `get_dummies` method.

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.

Nice explanation!

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.

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.

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

Rate this review