UDACITY

‹ Return to Classroom

Finding Donors for CharityML

REVIEW

CODE REVIEW 5

HISTORY

Meets Specifications

Very impressive submission here! Congratulations on completing this project!
It seems that you have acquired great knowledge of supervised learning algorithms and are now ready to apply it on your own problems.
Keep doing this great job! Keep experimenting and be curious.
All the best for your future.
Happy Learning!!

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**

Well done !!
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.

Good !! correctly implements one-hot encoding
features_final = pd.get_dummies(features_log_minmax_transform)

Evaluating Model Performance

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

Wonderful !! correctly calculates the benchmark score of the naive predictor
Naive Predictor: [Accuracy score: 0.2478, F-score: 0.2917]

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.

Great !! pros and cons or application for each model is provided with reasonable justification

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

Good !! implementation of a pipeline in code

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

Wonderful !! Three supervised learning is implemented.
clf_A = LogisticRegression(random_state=42)
clf_B = RandomForestClassifier(random_state=42)
clf_C = GradientBoostingClassifier(random_state=42)

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.

Good conclusion that The test results and other indicators show that gradient boosting would be the best algorithm model of the three.

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.

Wonderful explanation starting Gradient boosting combines weak learners into a single strong learner (an ensemble of learners) that produces better predictions.

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.

Well done !! final model chosen is correctly tuned using grid search
grid_obj = GridSearchCV(clf, parameters, scoring=scorer)

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.

The accuracy score for the optimized model is 0.871 and the F-score is 0.754, both of which are slightly better than the un optimized model. Well done !!

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.

Five features , most relevant for predicting is identified. Well done.

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.

GradientBoostingClassifier is chosen for feature*importances* attribute. Great !!

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.

Great !! final model's performance when only the top 5 features are used

DOWNLOAD PROJECT

5 CODE REVIEW COMMENTS ❯

RETURN TO PATH

Rate this review