Module 19 Challenge

New Attempt

Due Feb 21 by 11:59pm **Points** 100 **Submitting** a text entry box or a website url

Background

Lending services companies allow individual investors to partially fund personal loans as well as buy and sell notes backing the loans on a secondary market.

You will be using this data to create machine learning models to classify the risk level of given loans. Specifically, you will be comparing the Logistic Regression model and Random Forest Classifier.

Before You Begin

1. Create a new repository for this project called `supervised-machine-learning-challenge`. **Do not add this Challenge to an existing repository**.
2. Clone the new repository to your computer.
3. Inside your local git repository, create a directory for the Supervised Machine Learning Challenge.
4. Push the above changes to GitHub.

Files

Download the following files to help you get started:

Module 19 Challenge files (https://static.bc-edx.com/data/dl-1-1/m19/lms/starter/Starter_Code.zip)

Instructions

Retrieve the Data

The data is located in the Challenge Files Folder:

- `lending_data.csv`

Import the data using Pandas. Display the resulting dataframe to confirm the import was successful.

Predict Model Performance

You will be creating and comparing two models on this data: a Logistic Regression, and a Random Forests Classifier. Before you create, fit, and score the models, make a prediction as to which model you think will perform better. You do not need to be correct!

Write down your prediction in the designated cells in your Jupyter Notebook, and provide justification for your educated guess.

Split the Data into Training and Testing Sets

Create the features DataFrame, `X`, by removing the `loan_status` column. Create `y`, the labels set, by using the `loan_status` column. Split the data into training and testing datasets by using the `train_test_split` function.

Create, Fit and Compare Models

Create a Logistic Regression model, fit it to the training data that you created in the previous step. Then, determine the model's score by using the `score` function and the testing data from the previous step. Do the same for a Random Forest Classifier. You may choose any starting hyperparameters you like.

Review the scores of each model. Which model performed better? How does that compare to your prediction? Write down your results and thoughts in the designated markdown cell.

Requirements

Retrieve the Data (5 points)

To receive all points for this section, you must

- Import the `lending_data.csv` file as a Pandas dataframe (3 points)
- Confirm that the import was successful by displaying the dataframe (2 points)

Predict Model Performance (15 points)

To receive all points for this section, you must

- Make a prediction on which model will perform better on the data (5 points)
- Justify the prediction with information about the models (10 points)

Split the Data into Training and Testing Sets (30 points)

To receive all points for this section, you must

- Create the features DataFrame, `X`, by removing the `loan_status` column (10 points)
- Create `y`, the labels set, by using the `loan_status` column (10 points)
- Split the data into training and testing datasets by using the `train_test_split` function (10 points)

Create, Fit and Compare Models (50 points)

To receive all points for this section, you must

- Create and train a Logistic Regression model (10 points)
- Score the Logistic Regression model (10 points)
- Create and train a Random Forest Classifier model (10 points)
- Score a Random Forest Classifier model (10 points)

- State which model performed better (5 points)
- Compare the actual model performance with your predictions (5 points)

Grading

This assignment will be evaluated against the requirements and assigned a grade according to the following table:

Grade	Points
A (+/-)	90+
B (+/-)	80–89
C (+/-)	70–79
D (+/-)	60–69
F (+/-)	< 60

Submission

To submit your Challenge assignment, click Submit, and then provide the URL of your GitHub repository for grading.

NOTE

You are allowed to miss up to two Challenge assignments and still earn your certificate. If you complete all Challenge assignments, your lowest two grades will be dropped. If you wish to skip this assignment, click Next, and move on to the next Module.

Comments are disabled for graded submissions in BootCamp Spot. If you have questions about your feedback, please notify your instructional staff or your Student Success Manager. If you would like to resubmit your work for an additional review, you can use the Resubmit Assignment button to upload new links. You may resubmit up to three times for a total of four submissions.

References

Loan Approval Dataset (2022). DData for this dataset was generated by edX Boot Camps LLC, and is intended for educational purposes only.