SCALE FOR PROJECT PISCINE PYTHON DATA SCIENCE (/PROJECTS/PISCINE-PYTHON-DATA-SCIENCE) / DAY 08 (/PROJECTS/PISCINE-PYTHON-DATA-SCIENCE-DAY-08)

You should evaluate 1 student in this team

★

Git repository

`git@vogsphere.kzn.21-school.ru:vogsphere/intra-uuid-9cca(` 📋

Introduction

The methodology of School 21 makes sense only if
peer-to-peer assessments are done seriously.
This document will help you to do it properly.

- Please, stay courteous, polite, respectful and constructive in all
communications during this assessment. The bond of trust
between community 21 and you depends on it.

- Highlight possible malfunctions of the work done by the person
and take the time to discuss and debate it.

- Keep in mind that sometimes there can be differences
in interpretation of the tasks and the scope of features.
Please, stay open-minded to the vision of the other.

Guidelines

- Evaluate only the files that are on the GIT
repository of the student or group.

- Doublecheck that the GIT repository is the one
corresponding to the student
or the group as long as to the project.

- Meticulously check that nothing malicious has been used
to mislead you and have you assess something except
the content of the official repository.

- If you have not finished the project yet, it is compulsory
to read the entire instruction before starting the review.

- Use the special flags in the scale to report an empty
or non-functional solution as long as a case of cheating.
In these cases, the assessment is completed
and the final grade is 0 (or in a case of cheating is -42).
However, except for a case of cheating, you are encouraged to continue
reviewing the project to identify the problems that caused the situation
in order to avoid them for the next assessment.

- You must stop giving points from the first wrong exercise
even if the following exercises are correct.

Attachments

Preliminaries

Respect the rules

- The repository contains the work of the student (or group).
- The student is able to explain their work
at any time during the assessment.
- The general rules and any rules specific to the day are
respected throughout the assessment.

☑ Yes   ✕ No

Piscine Python | Data Science D08

Any hardcoded result is worth zero for the exercise.

Exercise 00 – Binary classifier

- Run all the cells in the notebook, they should work without errors
- The result of df.info() contains:
0 date 35 non-null object
1 am 35 non-null int64
2 pm 35 non-null int64
3 target 35 non-null object
- The answer to the first question: no, it is not easy?
- The answer to the second question: no, it is not good?
- The accuracy of the prediction is 0.6285714285714286?
- The accuracy of the naive prediction is 0.7142857142857143?

- The answer to the third question: no, it is not good?
- There are two plots: the first shows the actual data
and the second – the forecasted data
In all other cases, the test is failed.

⊘ Yes   ✕ No

Exercise 01 – Decision boundaries

- Run all the cells in the notebook, they should work without errors
- All the plots contain decision boundaries?
- The accuracy of logistic regression with scaled features is 0.7428571428571429?
- The accuracy of SVC is 0.7428571428571429?
- The accuracy of the decision tree with max_depth=4 is 0.9428571428571428?
- There is a visualization of the tree itself using plot_tree?
- The answer to the questions about the number of leaves is 4?
In all other cases, the test is failed.

⊘ Yes   ✕ No

Exercise 02 – Multiclass

- Run all the cells in the notebook, they should work without errors
- The dataframe that was saved into a file has 1686 rows × 44 columns?
- The accuracy of the naive solution is 0.23487544483985764?
- The top-1 feature for the logistic regression is labname_code_rvw?
- The top-1 feature for the SVC is uid_user_2?
- The top-1 feature for the tree is labname_project1?
- The top-1 feature for the random forest is numTrials?
In all other cases, the test is failed.

⊘ Yes   ✕ No

Exercise 03 – Overfitting

- Run all the cells in the notebook, they should work without errors
- The accuracy of the logistic regression on the test dataset is 0.6272189349112426?
- The accuracy of the SVC (kernel = linear) on the test dataset is 0.7159763313609467?
- The accuracy of the tree on the test dataset is 0.5295857988165681?
- The accuracy of the random forest on the test dataset is 0.9289940828402367?
- The answer to the first question is random forest?
- The answer to the second question is c?
- The best and final model is RandomForestClassifier?
- The graph with the top-10 most important features for the final model exists?
- The most important feature for the final model is NumTrials?
In all other cases, the test is failed.

⊘ Yes   ✕ No

Exercise 04 – Regression

- Run all the cells in the notebook, they should work without errors
- The result of df.info() contains:
0 num_commits 29 non-null int64
1 pageviews 29 non-null float64
2 AVG(diff) 29 non-null float64
- The answer to the question is a diagonal?
- The function is written and used for cross-validation?
- There are the plots for all the three models that contain
the actual values and the predictions?
In all other cases, the test is failed.

          ⊘ Yes   ✕ No

Exercise 05 – Clustering

- Run all the cells in the notebook, they should work without errors
- The function shows both plots on the same graph using subplots?
- The function shows the results for all the three model classes?
- The best number of clusters for KMeans is 8?
- The best number of eps for DBScan is 22?
- The best number of min_samples for DBScan is 1?
- The best number of clusters for AgglomerativeClustering is 8?
- The dendrogram exists?
In all other cases, the test is failed.

          ⊘ Yes   ✕ No

Ratings

Don't forget to check the flag corresponding to the defense

    ✔ Ok   ★ Outstanding project

📄 Empty work   💬 No author file   ☠ Invalid compilation   📓 Norme   🗐 Cheat

♟ Crash   💧 Leaks   ⊘ Forbidden function

Conclusion

Leave a comment on this evaluation

Finish evaluation

Terms of use for video surveillance (https://profile.intra.42.fr/legal/terms/1)
Rules of procedure (https://profile.intra.42.fr/legal/terms/4)
Declaration on the use of cookies (https://profile.intra.42.fr/legal/terms/2)
General term of use of the site (https://profile.intra.42.fr/legal/terms/6)
Legal notices (https://profile.intra.42.fr/legal/terms/3)
Privacy policy (https://profile.intra.42.fr/legal/terms/5)