Data Science Lab: Process and methods
Politecnico di Torino

Project description
Winter Second Call, A.Y. 2020/2021

Last update: February 2, 2021

1 Competition dates

Start date: February 02, 2021 at 23:59 CET
Due date: February 17, 2021 at 23:59 CET

Due date is a **strict deadline**.

2 Problem description

For this project, you are provided with a list of songs, each described by several features in a tabular format. Your goal is to predict the mode of the song given the other descriptors. The mode defines whether a song starts with a Major or Minor chord.

Technically, you are required to build a classification model capable of inferring the song's mode given the other features at your disposal.

> ⬥ **Warning:** For this competition, you will not be allowed to use external datasets other than the one provided for the competition. The adoption of external resources will result in failure of the exam.

2.1 Dataset

The dataset for this project contains information on real-world published songs. The songs are described in tabular format. Specifically, the dataset contains 136522 entries.

Each song is characterized by several numerical attributes, the publication year and the name of the artist(s). The mode of the song is indeed reported on the feature named `mode`, and it is valued 1 or 0 if the song is in Major or Minor mode respectively. The dataset is located at:

`https://dbdmg.polito.it/wordpress/wp-content/uploads/2021/02/second_dataset.zip`

Within the archive, you will find the following files:

- **dev.tsv** (development set): a tab-separated values file containing the songs from the development set. This portion does have the `mode` feature, which you should use to train and validate your models.

- **eval.tsv** (evaluation set): a tab-separated values file containing the songs from the evaluation set. This portion does not have the `mode` feature.

- **sample_submission.csv**: a sample submission file.

2.2 Task

You are required to build a classification pipeline to infer the mode of each record in the Evaluation set.

2.3 Evaluation metric

Your submissions will be evaluated with the following lines:

```python
from sklearn.metrics import f1_score
def evaluate(y_true, y_pred):
    return f1_score(y_true, y_pred, average="macro")
```

3 Submit your result

Submission file   In order to get your results evaluated, you have to upload a result file on our submission competition platform. The submission file has to be a `.csv` file formatted as follow:

```
Id,Predicted
0,0
1,0
2,1
3,0
4,1
...
```

The submission file must contains a header line and a row for each record in the Evaluation collection. Each row must have two fields:

- the `Id` of the corresponding record in the Evaluation set, as an integer number.

  > 🛈 **Info:** Note that the `Ids` in the submission file must correspond to the positions of the records in the Evaluation set. **The first record in the Evaluation set has Id=0, the second has Id=1 ans so on.**

- the `Predicted` label for the corresponding record.

  You can find a sample submission file on the competition material.

Submission platform   At the beginning of the semester, we sent you a submission key which is not valid anymore. We updated the key right before the first call, so that is the one you want to use for submitting your results[1]. Please refer to the guide on the course website, to go through the submission procedure. You can find the competition platform at http://trinidad.polito.it:8888

4 Upload the report and the software

> ◆ **Warning:** The report and the software have to be submitted by the due date reported in Section 1. This is a **strict deadline**.

Submission   All the required files (i.e. for the report and the software) must be included in a **single _.zip_** file. The archive must be uploaded to the "Portale della Didattica", under the *Homework* section. Please use **report_exam_winter_2021_second_call.zip** for both the zip name and the description field.

Formatting rules   The formatting rules for both the report and the software are described in the exam rules document. You can find it on the course website.

---

[1]You should have received a new key recently: check your email inbox. If that is not the case, please send an email to giuseppe.attanasio@polito.it