

Assignment 1

12/1/2024

Attempt 1



In Progress

NEXT UP: Submit assignment



Add comment

Unlimited Attempts Allowed

11/5/2024 to 12/1/2024

Details

The aim of Assignment 1 is to train and test two classifiers of your choice for identifying wine quality and picking the best-performing one.

- **Group assignment:** Max 2 students
- **Prerequisite reading:** ML in Python Lab material, Lecture 2: ML foundations, and **Imbalanced Data**  (<https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data>)
- **Language:** Python (you are allowed to use Python libraries such as Pandas, NumPy, scikit-learn, imbalanced-learn, etc.)
- **Data:** The Wine Quality dataset is available at <https://archive.ics.uci.edu/dataset/186/wine+quality>  (<https://archive.ics.uci.edu/dataset/186/wine+quality>)
 - The dataset contains **two** CSV files, one for white wine and one for red wine. You decide which one to study.
- **Steps:**
 1. Inspect the dataset, e.g., check the ratio of the classes and discuss your findings.
 2. Divide the data into train and test sets.
 3. Perform scaling on the data.
 4. For each classifier, perform validation using Repeated k-Fold Cross-Validation ($n_splits = 3$, $n_repeats = 10$) on the train set.
 - Choose the best metric, such as Accuracy, Precision, Recall, etc., which can explain the task of *wine quality classification* adequately.
 - Evaluate the performance of the two classifiers based on the **average** and **standard deviation** of the chosen metric.
 - Identify the best classifier based on its performance.
 5. Use the whole training set to build the final model using the best-performed classifier from the previous step.
 6. Report the model's performance on the test set.

7. Consider using the [imbalanced-learn](#) library or similar libraries to balance the scaled [train set](#). Check the ratio of the classes and discuss your findings.
8. Perform steps 4 to 5 using the balanced train set.
9. Report the model's performance on the test set.
10. Discuss your findings.

- **Written report**

- *Template*: The IEEE conference template and citation style should be followed ([templates](#) → <https://www.ieee.org/conferences/publishing/templates.html>) in MS Word and LaTeX).
- *Language*: English without spelling mistakes.
- *Style*: Clear.
- *Content*: The report should explain the problem you have solved, the data and the performed data preprocessing, the used algorithms and motivation behind their selection, the training and testing process, model evaluation results, etc.
- *Format*: PDF.
- *Page limit*: 1 page excluding references (**no abstract should be included**).

- **Code**

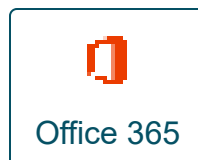
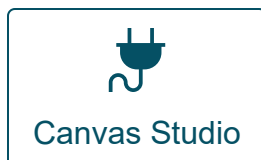
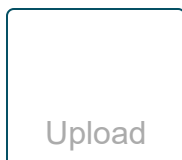
- Provide meaningful comments for different blocks of the code.
- A README.TXT file must clearly state exactly how to execute the code and any necessary setups.

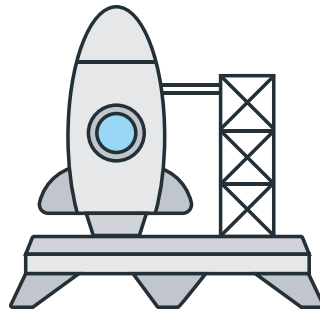
- **Submission**

- Make sure to include your names in the report and the code.
- The report must be submitted as a PDF separately (**not to be included in the ZIP file**).
- Code and additional files related to implementation must be archived using ZIP.

Keep in mind, this submission will count for everyone in your Assignment 1-Groups group.

Choose a submission type





Choose a file to upload

or

 Webcam Photo

 Canvas Files

☐ This assignment submission is my own, original work

Submit assignment