

Introduction to Artificial Intelligence

Summer 2021

EXERCISE 6: Regression and classification

deadline: **May 21st, 2021**

Teams assigned to teacher Tomasz Lehmann, tomasz.lehmann.dokt@pw.edu.pl :

Achermann Jan and Tchórzewski Michał
Gonzalez Herrera Lucas and Gómez Falcón Carlos
Jabłońska Karolina 4 and Marosek Wojciech
Kamieniarz Kacper and Szumski Jakub 2
Krishnamoorthy Kaushikram and Nath Akash
Miazga Tomasz 2 and Kurczak Michał
Orzełowski Konrad 2 and Nowakowski Dawid
Skrzyński Jan and Szpila Jakub
Wyrostkiewicz Aleksandra and Bara Wiktor

Teams assigned to teacher Paweł Wawrzyński pawel.wawrzynski@pw.edu.pl :

Aulad Thani Uhood Khalfan Said and Al-Maskari Fahad Farid Khamis
Baydur Beste and Merve Kizil
Capron Titouan and Yadan Warren
Caridi Flora Biblia Ina
Cataffo Francesca and Cricri Giuseppe
Cypcer Piotr and Kopiczko Dawid
Debeauvais Guillaume and Barbe Victor
Gałęcki Szymon 2 and Wojcieszuk Jakub
Glaser-Gallion Michael and Glaser-Gallion Michael
Gondouin Hadrien and Abdallah Serghine
Hącel Oskar and Lisowski Marcin 2
Jemielity Marcin and Bartosz Pacułt
Kaya Caner and Ekin Cakir
Khoury Wiktor and Tesliuk Illia
Kleinschmidt Thomas and Guinand Thomas
Kochański Piotr 2 and Korzeniewski Filip
Kos Szymon and Krupowicz Barnaba
Liu Yang and Liu Xuan
Mai Quoc Tuan and Wu Austin Dan
Markiewicz Paulina and Abu Hasan Sundus Maria
Mehra Tanya and Uzguc Yusuf
Pujol Higuera Micaela and Perdomo Barrios Ana Luisa
Rząd Maria and Abass Suliaman
Saif Hawra and Moustafa Ehab Hamdy Sayed
Świerczyński Jan and Kaniuk Michał 2
Szulc Oskar and Płatek Andrzej
Szwed Jakub and Zdanowski Arkadiusz
Viatteau Antoni and Kowalski Karol 5
Zhu Jihao and Wang Binjie

Task for teams assigned to teacher Tomasz Lehmann – wine

1. Exercise details

Write a program that predicts wine quality based on dataset:

<https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009>

(dataset is available after logging in).

Use at least two separated methods (e.g. SVM, Random Forrest, ...) and compare them to each other. The solution should be resolved as a classifier. The data should be properly divided into training, validation and test sets. You should also deliver report with comparison of metrics (e.g. F1, recall, ROC curve, confusion matrix, ...) for every used method and explanation which one do you consider your “*state-of-the-art*”.

Delivered solutions will be assessed by achieved results of the chosen models and correctness of research methods.

2. Technical details

- a. The preferred language to write your solutions is Python, however solutions in R, C, C++, C# or Java also will be accepted.
- b. Remember to adhere to basic standards of clean coding (with comments in the crucial parts).
- c. A clear instruction how to run the code should be attached.
- d. A brief report with methods comparison should be attached.

3. Handing-in guidelines

- a. You should submit the source code of your solutions to tomasz.lehmann.dokt@pw.edu.pl not later than **2021.05.21 23:59:59 CEST**. **Programs delivered after the deadline will not be assessed.**
- b. Please include "[EARIN] Exercise 6" in the title, and do not forget about adding names and emails of both team members in the email content.
- c. Details on the discussion part of the exercise will be given a few days after the deadline.
- d. You may get 0-8 pts for this assignment.
- e. In case of questions, please contact me via the given email address.

Task for teams assigned to teacher Paweł Wawrzyński – cancer

1. Exercise details

Write a program that distinguishes cancer versus normal patterns from mass-spectrometric data based on the dataset:

<https://archive.ics.uci.edu/ml/datasets/Arcene>

Use at least two separate methods (e.g. Random Forrest, Gradient Boosting) and compare them to each other. The solution should be resolved as a classifier. Compare accuracy on the training set and on the test set. You should also deliver report with comparison of metrics (e.g. F1, recall, ROC curve, confusion matrix, ...) for every used method.

Delivered solutions will be assessed by achieved results of the chosen models and correctness of research methods.

2. Technical details

- a. The preferred language to write your solutions is Python, however solutions in R, C, C++, C# or Java also will be accepted.
- b. You can take implementations of machine learning models (e.g. Random Forest, Gradient Boosting) from libraries.
- c. Remember to adhere to basic standards of clean coding (with comments in the crucial parts).
- d. A brief (max. 1 page) report should be attached with a clear instruction how to run the code, and discussion of the results.

3. Handing-in guidelines

- a. You should submit the source code and the report of your solutions to pawel.wawrzynski@pw.edu.pl not later than **2021.05.21 11:59:59 CEST**. **Programs delivered after the deadline will not be assessed.**
- b. The mail title should be "[EARIN] Exercise 6". Do not forget about adding names and emails of both team members in the email content.
- c. You may get 0-8 pts for this exercise. The necessary conditions for getting any points are to send the code with the report and present the work.
- d. The works will be presented on 2021.05.21 at 12.15-2.00 pm via MS Teams. Subsequent teams will be invited on the general channel of the EARIN team. Immediately after invitation the teams will be expected to call the teacher. The teams will be invited in the following order:

Szulc Oskar and Płatek Andrzej
Świerczyński Jan and Kaniuk Michał 2
Kochański Piotr 2 and Korzeniewski Filip
Rząd Maria and Abass Suliaman
Liu Yang and Liu Xuan
Kleinschmidt Thomas and Guinand Thomas
Capron Titouan and Yadan Warren
Saif Hawra and Moustafa Ehab Hamdy Sayed
Viatteau Antoni and Kowalski Karol 5
Markiewicz Paulina and Abu Hasan Sundus Maria

Mehra Tanya and Uzguc Yusuf
Szwed Jakub and Zdanowski Arkadiusz
Hacel Oskar and Lisowski Marcin 2
Cataffo Francesca and Cricri Giuseppe
Cypcer Piotr and Kopiczko Dawid
Khoury Wiktor and Tesliuk Illia
Pujol Higuera Micaela and Perdomo Barrios Ana Luisa
Kos Szymon and Krupowicz Barnaba
Caridi Flora Biblia Ina
Aulad Thani Uhood Khalfan Said and Al-Maskari Fahad Farid Khamis
Gondouin Hadrien and Abdallah Serghine
Galecki Szymon 2 and Wojcieszuk Jakub
Baydur Beste and Merve Kizil
Zhu Jihao and Wang Binjie
Mai Quoc Tuan and Wu Austin Dan
Jemielity Marcin and Bartosz Pacułt
Debeauvais Guillaume and Barbe Victor
Glaser-Gallion Michael and Glaser-Gallion Michael
Kaya Caner and Ekin Cakir