|  |
| --- |
|  |

Fundamentals of Data Science - project

## The aim of the project is to familiarize you with practical algorithms and problems of modern data analysis and tools used to implement selected task. The project is **individual**. It is also aimed at encouraging you to use some of the modern data science tools.

## The report along with any attachments (source code) should be submitted in electronic form using Microsoft Teams. A special assignment section will be prepared there.

## The report should not contain too much theoretical information - only the most important practical aspects of the problem under consideration. Particularly appreciated will be the ability to critically analyze and formulate valuable technical conclusions.

## The deadline for submitting the report is the end of **January 27th 2023**. After this date, the final evaluation will take into account the delay in the implementation of the task. Information about the exam will be published on MS Teams.

## Project meetings are not obligatory, but rather consultations for those interested.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | Monday. 16.01.2023 | 9:15 - 11:35 | MS Teams | | | Thursday 19.01.2023 | 9.15 – 11.35 | MS Teams | | | Friday 20.01.2023 | 12:10 – 14:25 | Onsite | | | Monday. 23.01.2023 | 16:00 – 18:15 | MS Teams | | Friday. 27.01.2022 | 12:10 – 14:25 | Onsite | |
|  |

## For MS Teams you can just drop me a message or call.

**The task:**

Please download *wine* dataset.

Please register yourself to Chat GPT (be stubborn and repeat the registration procedure as the server is usually busy). You can try different working hours.

Please perform the following steps

1. Do statistical summary of the data and discuss it.
2. Reduce data dimensionality
3. Visualize the reduced dataset
4. Cluster the dataset (and evaluate clustering results with classification labels).
5. Split the dataset into training and testing.
6. Perform classification and evaluate its result.

Please ask ChatGPT how to do these things and compare how this information differs from the one obtained during classes. The result of this test should be discussed in the report. In your report, apart from that please present your approach and obtained results, code should be sent as an attachment or a link to a repository. The report should be concise and do not exceed 10-15 pages.