



*Faculty of Mechanical Engineering
and Robotics*

*Department of Robotics and
Mechatronics*



Python for Machine Learning and Data Science
Course for Mechatronic Engineering

Project datasets

Course supervisor:

Ziemowit Dworakowski, zdw@agh.edu.pl

Instruction author:

Adam Machynia, machynia@agh.edu.pl

Introduction

This document contains exemplary datasets with short descriptions that you can use for your projects. However, feel free to browse the UCI ML repository (<https://archive.ics.uci.edu>) or the Kaggle website (<https://www.kaggle.com/datasets>) to choose a dataset that interests you.

Each team in a laboratory group works with a different dataset.

The choice of a dataset should be approved by the teacher.

Individual tasks

After selecting a dataset, you will have time to explore it and think about potential tasks to solve. Generally, most datasets have straightforward problems associated with them. You can view examples of these solutions on Kaggle. Then, you'll need to suggest your own alternative task that you'd like to work on. This task should typically be a classification or regression problem, or it may require a similar approach for solving. It may also involve other hypotheses to prove or questions to answer.

You should discuss your ideas for the task with the teacher. The final task will be assigned to the group by the teacher. During the process of solving your tasks, they may be reformulated if needed, but only by the teacher.

Projects will be completed through a combination of laboratory sessions, project classes, and independent work. Each component will focus on different aspects of the project. In laboratory sessions, you will present your implementations, engage in detailed discussions about your data, and explore various research directions. Project classes will emphasize team building, project management theory, and presenting key elements of your work. At home, you are responsible for tracking team management decisions, self-assignments, and documenting your interactions with generative AI (such as ChatGPT).

Exemplary datasets

1. Travel Reviews

<https://archive.ics.uci.edu/dataset/484/travel+reviews>

Reviews on destinations in 10 categories mentioned across East Asia. Each traveler rating is mapped as Excellent(4), Very Good(3), Average(2), Poor(1), and Terrible(0) and average rating is used.

2. Taiwanese Bankruptcy Prediction

<https://archive.ics.uci.edu/dataset/572/taiwanese+bankruptcy+prediction>

The data were collected from the Taiwan Economic Journal for the years 1999 to 2009. Company bankruptcy was defined based on the business regulations of the Taiwan Stock Exchange.

3. Secondary Mushroom

<https://archive.ics.uci.edu/dataset/848/secondary+mushroom+dataset>

Dataset of simulated mushrooms for binary classification into edible and poisonous.

4. Student Performance

<https://archive.ics.uci.edu/dataset/320/student+performance>

Predict student performance in secondary education (high school).

5. Estimation of Obesity Levels Based On Eating Habits and Physical Condition

<https://archive.ics.uci.edu/dataset/544/estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition>

This dataset include data for the estimation of obesity levels in individuals from the countries of Mexico, Peru and Colombia, based on their eating habits and physical condition.

6. Chronic Kidney Disease

<https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease>

This dataset can be used to predict the chronic kidney disease and it can be collected from the hospital nearly 2 months of period.

7. Appliances Energy Prediction

<https://archive.ics.uci.edu/dataset/374/appliances+energy+prediction>

Experimental data used to create regression models of appliances energy use in a low energy building.