

Machine Learning Course Project Guidelines A practical application for machine learning

We live in a world flooded with data. Globalization, widespread internet and mobile applications have converted us into a continuous source of information for big corporations, and also has given us the opportunity to access knowledge in a way that had never been possible before. What can you use artificial intelligence for? the answer is: pretty much the same that you can use your intelligence for.

So, the question would be: what do YOU want to use AI for?

In this project you will find a practical application of machine learning, something that motivates you and is relevant to people out there. Then, you will use the tools learnt during our course and provide a solution that satisfies both your appetite for knowledge and the needs of the potential users.

There will be two deliverables, the first one will be a “project proposal” which will be delivered on the 6th session. The second will be a scientific paper, in which you will communicate effectively the results of your research.

First deliverable: Project Proposal

You will make a presentation with the following sections:

1. Introduction and motivation:
 - What is the problem (application) you will tackle?
 - Why is it important?
2. Literature review about other solutions to the problem
3. Your solution/process description, including an schematic diagram
4. Dataset explanation and visualization

You will submit the slides of your presentation. You will have 15 minutes to present and 5 mins for questions.

Final deliverable: Results Publication

You will write a scientific paper in IEEE format (in English) that contains at least the following items:

1. Description of the phenomenon / process modeled and the problem to be addressed
2. Process of obtaining / generating the dataset.
3. Description of the machine learning problem, what kind of problem is it? What is the learning task to perform?
4. Data pre-processing used using at least three different methodologies
5. Training and evaluation of at least three models, including a complexity analysis

In addition to the article, you will provide all the files needed to run your experiment, and make a practical presentation (20 minutes) in which you will demonstrate how it works and what were the results.

All project ideas are welcome. Please first try to propose a project of your interest. If you don't find one, below are some suggested projects.

Suggested Projects

1. Classification on the 'credit-g' dataset.
You can download the dataset and see it's description at <https://www.openml.org/d/31>

2. Regression on Sydney Dataset.

You can load the Sydney housing dataset from <https://www.kaggle.com/shree1992/housedata> where you can also find a description. While you can find several kernels on kaggle already, I highly recommend you start your own solution from scratch.

3. Classification of the spam dataset.

You can load the spam dataset from <https://www.kaggle.com/c/hw5-spam-competition-cs189sp22>. While you can find several kernels on kaggle already, I highly recommend you start your own solution from scratch.

4. Classification of the titanic dataset.

You can load the spam dataset from <https://www.kaggle.com/competitions/hw5-titantic-competition-cs189sp22>. While you can find several kernels on kaggle already, I highly recommend you start your own solution from scratch.

5. Predicting wine quality from review texts and other properties of the wine.

You can find the dataset here: <https://www.kaggle.com/zynicide/wine-reviews>. While you can find several kernels on kaggle already, I highly recommend you start your own solution from scratch.