UNIVERSIDADE DE AVEIRO
DEPARTAMENTO DE ELECTRÓNICA TELECOMUNICAÇÕES E INFORMÀTICA

CAA PROJECT 2 2024/2025 - Instructions

Deadline for Project 1 submission: 10/June 2025
Submit to elearning: report (pdf+ source files); presentation slides; implementation code

You are free to propose a machine learning (ML) problem you would prefer to work. Please, discuss your idea with the instructor.

I. PROJECT PROPOSALS

A. Focused on specific application problem:

PROJECT PROPOSAL 1 ALZHEIMER MAGNETO RESONANCE IMAGES (MRI) DATASET
https://www.kaggle.com/datasets/sachinkumar413/alzheimer-mri-dataset

PROJECT PROPOSAL 2 FOREST FIRES DATA SET
https://www.kaggle.com/datasets/elikplim/forest-fires-data-set

PROJECT PROPOSAL 3 STAR DATASET TO PREDICT STAR TYPES
https://www.kaggle.com/datasets/deepu1109/star-dataset

Project proposal 4 DeepSig Dataset: RadioML 2018.01A
This dataset includes both synthetic simulated channel effects and over-the-air recordings of 24 digital and analog modulation types. Classify different modulations.
https://www.kaggle.com/datasets/pinxau1000/radioml2018

Project proposal 5 Fake News Detection (different data sets, choose one of them)
https://www.kaggle.com/c/fake-news/data
https://www.kaggle.com/datasets/emineyetm/fake-news-detection-datasets
https://www.kaggle.com/c/fake-news-pair-classification-challenge/data
https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset

B. Focused on new AI methods, models, frameworks :

Project proposal 6 - Data generation with Diffusion Models. Application for a basic machine learning data set.
Web sources:

- https://www.superannotate.com/blog/diffusion-models
- https://arxiv.org/abs/2209.00796 Diffusion Models: A Comprehensive Survey of Methods and Applications
- https://lilianweng.github.io/posts/2021-07-11-diffusion-models/
- Overview of diffusion model architectures: https://encord.com/blog/diffusion-models/

Project proposal 7 Overview of Graph Neural Networks (GNN). Application of GNN for a basic machine learning problem.

Web sources:

https://uvadlc-notebooks.readthedocs.io/en/latest/tutorial_notebooks/tutorial7/GNN_overview.html
https://colab.research.google.com/drive/16GBgwYR2ECiXVxA1BoLxYshKczNMeEAQ?usp=sharing
https://www.youtube.com/watch?v=cWIeTMklzNg

Project proposal 8. Overview of explainable AI (XAI). Application of XAI method (GradCam, Lime, Shap) for a basic machine learning problem.

Web sources:

https://arxiv.org/pdf/2211.06579
https://insights.sei.cmu.edu/blog/what-is-explainable-ai/

Project proposal 9 Create conversational AI-based Chatbot.
Goolge Dialogflow platform https://dialogflow.cloud.google.com/
Building a Simple Chatbot from Scratch in Python (using NLTK)
https://github.com/parulnith/Building-a-Simple-Chatbot-in-Python-using-NLTK

Project proposal 10 Open Source AI framework Haystack. Overview of Haystack. Illustrating it for a simple LLM application.

Web sources:
- https://www.deeplearning.ai/short-courses/building-ai-applications-with-haystack/
- Tutorials / Walkthroughs: https://haystack.deepset.ai/tutorials
- Cookbooks: https://haystack.deepset.ai/cookbook
- Discord / Github: https://haystack.deepset.ai/community
- Haystack Documentation: https://docs.haystack.deepset.ai/docs/intro
- deepset Studio: https://www.deepset.ai/deepset-studio
- deepset AI Platform Dokumentation: https://docs.cloud.deepset.ai/docs/getting-started

II.   PROJECT ASSESMENT (25 % of the final grade)

1. **Report.** The project is evaluated based on a submitted report/paper (IEEE Latex format). The work done by each student has to be explicitly specified.
   All project's files (pdf and Latex files of the report, the presentation slides and the code implementing the algorithms) are **submitted to the elearning.ua.pt page of the course unit** in section *SUBMISSION – PROJECT 2* in a compressed format having the following name: P2_CAA2025_XXXXX_YYYYY (where XXXXX and YYYYY are substituted by the student number of each student.

   The report should contain the following:
   - A short review of research papers related to your topic.
   - A description of the data set you will be working with .
   - A brief description of the machine learning method you will use. Don't re-explain basics such as how CNNs work, rather describe the properties of your approach.
   - Train the model and put the results in your report.
   - Describe successes and problems in your analysis.
   - Submit the code you produced for the project.