UNIVERSIDADE DE AVEIRO
DEPARTAMENTO DE ELECTRÓNICA TELECOMUNICAÇÕES E INFORMÀTICA

PROJECT 2 2024/2025 - Instructions

Deadline for Project 2 submission: 15/January 2025
Submit to eLearning: report (pdf+ source files), implementation code

You are free to propose a machine learning (ML) problem you would prefer to work. Please, discuss your idea with the instructor. **Choose a project with some importance for the society.**

I. PROJECT PROPOSALS

Project proposal 1. Skin Cancer Detection and Classification.
International Skin Imaging Collaboration (ISIC) : https://challenge.isic-archive.com/data
https://www.isic-archive.com/#!/topWithHeader/tightContentTop/challenges

Project proposal 2. Face Detection from images.
https://github.com/zlmo/Face-Detection
http://shuoyang1213.me/WIDERFACE/

Project proposal 3. Face Recognition with Olivetti dataset.
https://www.kaggle.com/serkanpeldek/face-recognition-on-olivetti-dataset

Project proposal 4 Machine Learning for cybersecurity
https://github.com/PacktPublishing/Hands-on-Machine-Learning-for-Cyber-Security

Project proposal 5 CNN for CIFAR-10 dataset : https://www.cs.toronto.edu/~kriz/cifar.html

Project proposal 6 Intel Image Scene Classification of Landscapes
Image Scene Classification of Multiclass (Buildings, Forest, Glacier, Mountain, Sea Street)
https://www.kaggle.com/datasets/puneet6060/intel-image-classification

Project proposal 7 AI4I 2020 Predictive Maintenance Data
https://archive.ics.uci.edu/ml/datasets/AI4I+2020+Predictive+Maintenance+Dataset
Synthetic dataset that reflects real predictive maintenance encountered in industry.

PROJECT PROPOSAL 8 AIRLINE PASSENGER SATISFACTION
https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction

Project proposal 9 Create AI-based Chatbot. Choose the purpose of the Chatbot, for example to help students learn Data Science subject, or some other topic.
https://github.com/parulnith/Building-a-Simple-Chatbot-in-Python-using NLTK/blob/master/Chatbot.ipynb
https://www.youtube.com/watch?v=EPzqKkjcnro
https://www.udemy.com/course/create-a-python-powered-chatbot-in-under-60-minutes

Project proposal 10 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
(Political Fake News Detection from Different News Source on Social Media using
Machine Learning Techniques, Article in AIUB Journal of Science and Engineering (AJSE) · Nov 2022)

II. WORK LOAD AND STRUCTURE OF THE PAPER

1. State of the art review
Search and review of at least 5-6 references (papers, reports, thesis, etc.) handling the same or similar problem. Make a review of different techniques used to solve the problem you want to explore.

2. Data description, visualization and statistical analysis
Describe the problem you want to solve, the features and visualize the data (if it is difficult due to high dimension, show only some samples). Provide some statistical analysis such as metadata (e.g. features range of variation), histograms, try to identify if there are some data quality problems, detect interesting subsets.

3. Data preprocessing (if relevant)
Describe possible preprocessing steps to construct the final input to the machine learning algorithm from the initial data, such as data normalization, feature selection or dimensionality reduction in case of redundant features.

4. Description of the applied machine learning algorithm(s)
Apply a suitable ML algorithm (learned in class or self-learned) to solve the problem with the chosen dataset. Introduce the method shortly, define its parameters. Make a selection of the most important model hyper parameters after their variation in a selected range. Show graphically the results of this search.

5. Presentation and discussion of results
Presentation of the results preferably in a graphical format. Analysis, discussion, interpretation. Compare your results with the results in the reviewed references or apply and compare at least two ML methods on the same problem. For new data sets apply at least one new ML method not applied in Project 1. For projects that reuse dataset from Project 1 apply new ML methods and make a comparison with the previous results.

6. Conclusions
Critical discussion of the gained knowledge regarding the advantages/disadvantages of the applied methods on the problem in hand. Suggestions for potential future directions of study.

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

1. **Report.** The project is evaluated based on a submitted report (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, and the code implementing the algorithms) are **submitted to the elearning.ua.pt page of the course** in section *SUBMISSION – PROJECT 2* in a compressed format having the following name: P2_FML2025_XXXXX_YYYYY (where XXXXX and YYYYY are substituted by the student number of each student.

IV. Evaluation criteria (total score 20)

1. Report content (14)

– State of the art review. Motivation for this problem.
– ML problem complexity
– Data Description.
– Data Preprocessing. Train/validation/test data devision. K-fold cross validation.
– Description of the Applied Machine Learning methods.
– Results, Conclusions.

2. Report formatting (3) :

– IEEE Latex format, affiliation (Department, University, subject, course instructor), abstract, keywords, work load per student.
– Sufficiently detailed report.
– References, reference citation in the report.
– Clear figures (title, legends, axis labels) and tables referred in the text.

3. Novelty and contributions (3)

- Compare your solution with results published in the literature. Try to improve something, propose a better idea, or just justify why your solution is better or worse than in the consulted references.