UNIVERSIDADE DE AVEIRO
DEPARTAMENTO DE ELECTRÓNICA TELECOMUNICAÇÕES E INFORMÀTICA

PROJECT 2 2021/2022 – Spring Semester

You are strongly encouraged to propose a machine learning problem you would prefer to work. Please, discuss your idea with the instructor.

I. PROJECT PROPOSALS

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

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
https://www.kaggle.com/puneet6060/intel-image-classification

Recommended Data Repositories:

- Kaggle Data Repository : https://www.kaggle.com/datasets
- UCI Machine Learning Repository : https://archive.ics.uci.edu/ml/index.php

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)

The project is evaluated based on a submitted paper according to the IEEE format. The paper should follow the structure of the Work Load. The work done by each student has to be explicitly specified. All project's files (pdf and Latex files of the paper, and the code implementing the algorithms) are sent to the course instructor (petia@ua.pt) in a compressed format having the following name: P2_ML2021_XXXXX_YYYYY (where XXXXX and YYYYY are substituted by the academic (mechanographic) number of each student. If the file is too big to email as an attached document, feel free to use any big file transfer option you may know (we transfer, dropbox, link in a cloud. etc.)

IV. Evaluation criteria (total score 20)

1. Report content (13)

- State of the art review.
- 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 (4) :

- 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)

- Based on the references and what has been done previously by other authors, propose a better solution, e.g. improve the performance of the ML model in solving the problem you work with.

Deadline for Project2 submission: 30/June, 2022.