# Machine Learning Projects

Team Size:

**Members: 5-6**

## Description:

**As a team you will be implementing:**

**1. Linear regression & KNN as regressors on a numerical dataset.**

**2. Logistic regression & KNN as classifiers on an image dataset (5 classes at maximum).**

Grading:

**Total: 20 marks**

* **10 marks: Numerical dataset.**
* **10 marks: Image dataset.**

Deliverables:

1. **Source Code + Datasets: Uploaded to GitHub, to be filled by the team leader later before discussion.**
2. **Project Cover Sheet:**

**It should include Faculty name, course name, team number, team members’ IDs and names.**

1. **Project Description Document:**

For each model, you should specify:

* 1. ***General Information on dataset:*** the name of dataset used, number of classes and their labels, the total number of samples in dataset and the size of each (in case of images), and finally the number of samples used in training, validation and testing.
  2. ***Implementation details:*** For each model you should show all these results for your model on testing data (loss curve, accuracy, confusion matrix, precision, recall, ROC, AUC graph)

**Datasets:**

1. **numerical dataset of your selection.**
2. **Choose image dataset from one of the following:**
3. Cell Images for Detecting Malaria   
    <https://www.kaggle.com/datasets/iarunava/cell-images-for-detecting-malaria>
4. Columbia Object Image Library (COIL-100) Dataset  
    http://www1.cs.columbia.edu/CAVE/software/softlib/coil-100.php
5. Flower Species Recognition  
    <http://www.robots.ox.ac.uk/~vgg/data/flowers/102/index.html>
6. Traffic Sign Recognition

http[://people.ee.ethz.ch/~timofter/traffic\_signs/index.html](http://people.ee.ethz.ch/~timofter/traffic_signs/index.html)

1. Fruits 360 Dataset  
    https://www.kaggle.com/moltean/fruits
2. Caltech-UCSD Birds-200 2011

<http://www.vision.caltech.edu/datasets/cub_200_2011/>

1. Character Recognition in Natural Images ( The Chars74K dataset )

http://www.ee.surrey.ac.uk/CVSSP/demos/chars74k/

1. Oxford-IIIT Pet Dataset  
    <https://www.robots.ox.ac.uk/~vgg/data/pets/>
2. STL-10 dataset

<https://www.kaggle.com/jessicali9530/stl10>

1. Stanford Dogs Dataset

<http://vision.stanford.edu/aditya86/ImageNetDogs/main.html>

1. Age estimation

<https://susanqq.github.io/UTKFace/>

1. Plant Pathology 2020 – FGVC7

<https://www.kaggle.com/c/plant-pathology-2020-fgvc7/data>

1. Plant Disease Classification:
   * **Dataset:** [PlantVillage Dataset](https://www.kaggle.com/emmarex/plantdisease)
2. Food Recognition:
   1. **Dataset:** [Food-101](https://www.kaggle.com/dansbecker/food-101)
3. Fine-Grained Bird Species Classification:
   1. **Dataset:** [CUB-200-2011](http://www.vision.caltech.edu/visipedia/CUB-200-2011.html)
4. Medical Image Diagnosis:
   1. **Dataset:** [Diabetic Retinopathy Detection](https://www.kaggle.com/c/diabetic-retinopathy-detection)
5. Vehicle Make and Model Recognition:
   1. **Dataset:** [Stanford Cars Dataset](https://ai.stanford.edu/~jkrause/cars/car_dataset.html)
6. Fashion Item Classification:
   1. **Dataset:** [Fashion MNIST](https://github.com/zalandoresearch/fashion-mnist)
7. Scene Classification:
   1. **Dataset:** [MIT Scene Parsing Benchmark](http://sceneparsing.csail.mit.edu/)
8. Traffic Sign Recognition:
   1. **Dataset:** [German Traffic Sign Recognition Benchmark](http://benchmark.ini.rub.de/?section=gtsrb&subsection=news)
9. Mammogram Classification for Breast Cancer Detection:
   1. **Dataset:** [Digital Database for Screening Mammography (DDSM)](https://www.kaggle.com/salim14/mammogram)
10. Art Style Classification:
    1. **Dataset:** [Painter by Numbers](https://www.kaggle.com/c/painter-by-numbers)
11. Facial Expression Recognition:Dataset:

[Facial Expression Recognition Challenge (FER2013)](https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data)