# Project 1

Choose a dataset from publicly available datasets such as:

UCI - Machine learning Repository:

<https://archive.ics.uci.edu/ml/datasets.html>

DATA.GOV.UK

<https://data.gov.uk/data/search>

US.GOV

<https://www.data.gov/>

KAGGLE

<https://www.kaggle.com/datasets>

WORLD BANK

https://data.worldbank.org/data-catalog

**The selected dataset must be at least order of 10,000 instances**.

Your task is to rank/filter/process/adjust features in the dataset **and** run one or more ML algorithms to predict a class (the class depends upon the dataset).

For this project, **max. 2 students each team.**

It is preferable not to use WEKA, scikit-learn or other tools more suitable for large datasets.

A good project should include:

* description of the dataset and of the ML task
* analysis and filtering of features,
* description of the algorithms used
* comparison of experiments,
* final comments on performance metrics.

See examples of good submitted projects on the course web site.

# Project 2 – Sentiment classification of Twitter concerning Italian Broadcasts

You need to know Italian or to work with at least one Italian-speaking student in the team.

You are provided with a set of Twitter messages mentioning a broadcast or including a broadcast hashtag.

* Your task is to classify tweets in positive/negative/neutral using ML algorithms;
* You could also try to predict the winner by considering only tweets prior to the nomination of the winner (Data have been collected during the week of the Festival)

You CAN start by reading hints on how to perform the task, from:

[https://www.ravikiranj.net/posts/2012/code/how-build-twitter-sentiment-analyzer/](https://www.ravikiranj.net/posts/2012/code/how-build-twitter-sentiment-analyzer/" \t "_blank)

You must release also the dataset of annotated tweets (if you use supervised learning) and set of positive/negative tokens/hashtags that you found

For this project, max. 2 students each team

(Those who like challenges, can also read – and be inspired by - :

<https://medium.com/@thoszymkowiak/how-to-implement-sentiment-analysis-using-word-embedding-and-convolutional-neural-networks-on-keras-163197aef623>

or

<https://blog.keras.io/using-pre-trained-word-embeddings-in-a-keras-model.html>

)