Individual Project: Introduction

Daniele Spiga, INFN-Perugia

spiga@pg.infn.it

28 November 2022 to 2 December 2022
INFN Perugia

Recap

The SOSC22 foresees the development of a personal project.

- Upon successful completion of the project the School will issue an official **"SOSC School of Computing Diploma**"

The project is structured with Jupyter Notebooks that will include also exercises. While the first part of the notebooks will be guided by tutors, the exercises will be developed just by you.

- On Tue: **Data sets exploration and feature engineering**
- On Wed: **machine learning**
- On Thu: **workload parallelization and distribution**.

The project foresees also **bonus tracks for advanced users**

NOTE: An **attendance certificate will be provided to the students attending the whole School programme**

What we do

We will provide you with 3 Jupyter Notebooks (you'll get them from git when we will tell you to do it)
-   One per day: Tuesday to Thursday

We will provide you with a dataset
-   Lucio will guide you through it. how to obtain and process the dataset

The dataset The dataset, prepared and distributed by kaggle.com
-   was obtained from the Kepler mission taking photos of a set of stars at regular intervals of 30 minutes for a few days.
-   Not HEP specific nor physics related data

Few words about the project

During the first part of the project you will deal with Pandas DataFrame objects.

You will learn how Pandas provides a valuable tool to get insight in the data, without any need for machine learning tools. A bit of feature engineering to get prepared for the next day

In the second part (read it as a second notebook), we will define a metric for the classification task and we will try to push our classification capabilities using Deep Neural Networks.

Finally in the third day we will start from a simple case to show how to parallelize and distribute work. We then try to use what we did on Thusday as a payload to manage

What you should do

As you will see starting from tomorrow you will be asked to complete few exercises each day

- Day 1 (Thuesday): we will complete a pre processing step. Which will be then used on Wednesday when we quickly describe the solution.
  - This is why on Wednesday you should give us what you did on Thuesday
- Day2 and 3: will be similar.

Where to work

Already today you started playing with JupyterHub (and Lab)
-   You did a quick tour with Diego. Please ask any question

P.s.
   You already know it: The Jupyter environment is available also from outside. Use it as much as you like.
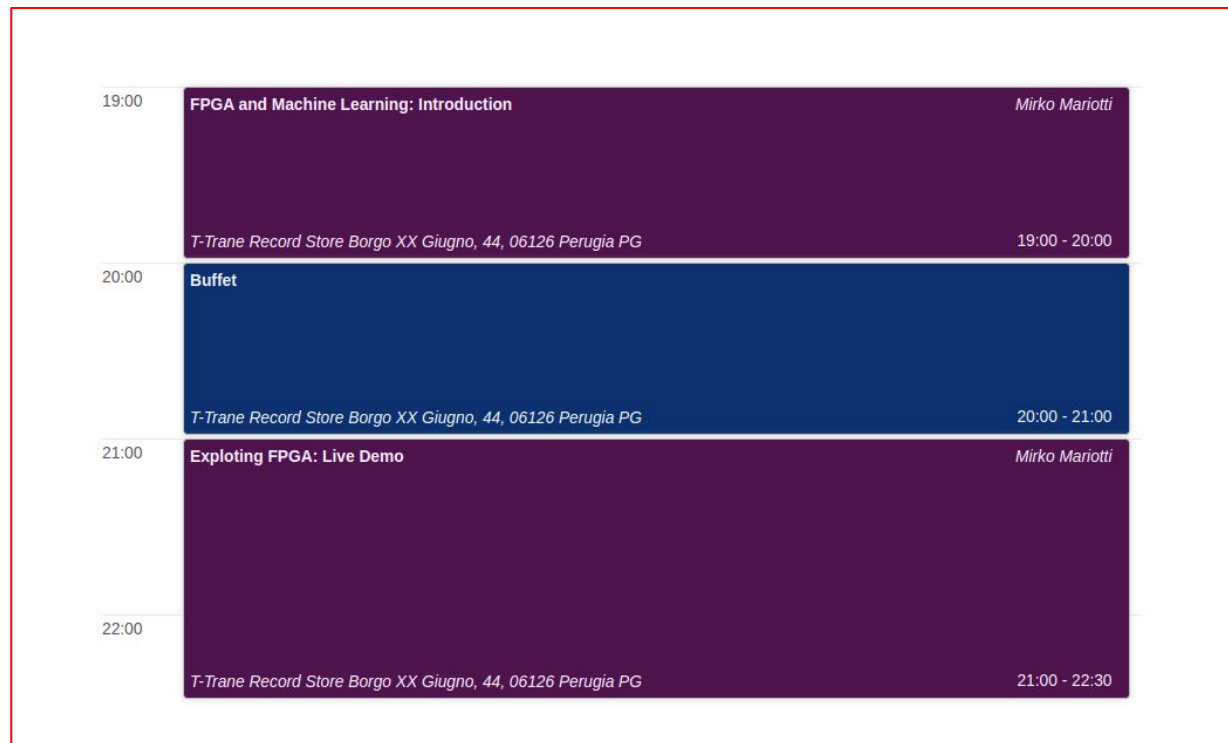   -   We will keep it running for the whole week after the school ends

And now, the first day is gone

But not yet time to call it a day…

Let's relax a bit with some wine and FPGA :)
- And socializing

How to get there

25 minutes from here, Google Maps will help

Borgo XX Giugno, 44, 06126 Perugia PG