DEEP LEARGNING

PROJECTS

L. BACCO – F. D'ANTONI – L. PETROSINO – A. MARCHETTI –
M. MERONE

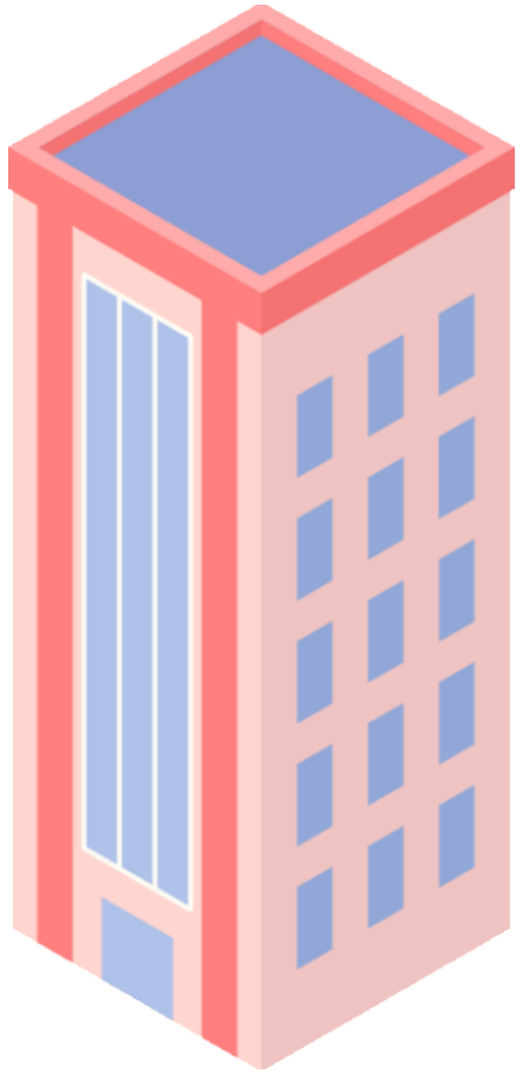
PROGETTO 1

Data: **Renewable Energy Community (REC) energy management** https://doi.org/10.1016/j.dib.2022.108590

The dataset reports data of the prosumers' (producers+consumers) energy dynamics for a period of 1 year.

- The consumption related to 50 households and 1 public building is reported, separated for each consumer and for different appliances.
- The energy production is reported for the entire REC, consisting in 15 photovoltaic panels.

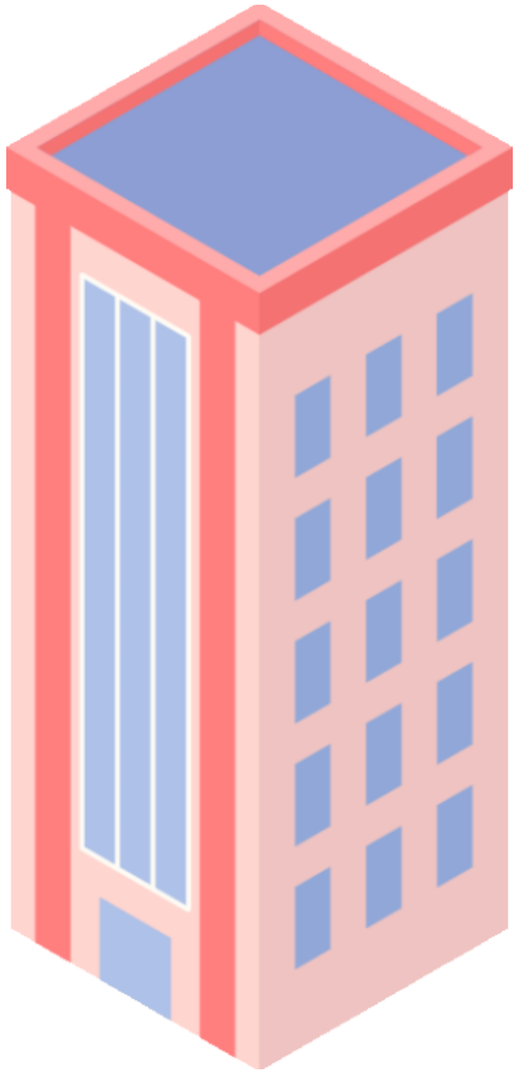
The REC is provided with a fuel cell to store the electricity charge or to provide power, and the buying/selling to the grid happens with a different pricing depending on the time of the day.



Task:
to develop a deep learning application to deal with the **prediction of energy consumption and photovoltaic generation** of the entire Renewable Energy Community (REC), exploiting a **Federated Learning** approach for the prediction of the photovoltaic production.

To exploit the above mentioned predictions to develop an optimal strategy for battery management using **Reinforcement Learning** in order to maximize the economical reward achieved buying energy from the provider or by selling the photovoltaic-produced energy to the grid.

Data from the **1st half of the year can be used to train** the predictive models, and data from the 2nd half of the year can be used to **evaluate the performance.**



NATURAL LANGUAGE PROCESSING

Task: Automatic text classification for categorization of patents into areas of Intellectual Property

Rationale: Patent classification is one of the areas in Intellectual Property Analytics (IPA), and a growing use case since the number of patent applications has been increasing worldwide.

Dataset: The dataset (--> https://huggingface.co/datasets/big_patent) consists of a large amount of documents, grouped by 8 (or 9) categories. Each document presents a long text and its abstract.

Challenges & ideas:

- Imbalancement of the dataset: there are classes with fewer examples than others.
- Data size: what happens at varying number of training samples.
- Are modern Large Language Models better than traditional NLP pipelines for the task? Can exploiting long texts lead to better models?
- Others...