Project: Predicting Climate Risks

General instructions

0. Pick your team members. Decide beforehand who will contribute what. Declare it clearly in your project git page, which files each person contributed. Pick carefully your team members so that each person has complementary skills.
1. Find first to which UN sustainability goals it addresses.
2. Do some
    a) literature search to see what has been done in this area in the last 5 years (up to 2 pages).
    b) Do a search in github for code that solves the problem (or parts of it)
    c) Are there any commercial solutions out there? Is there room for improvement in your solution/prototype?
3. Decide how you will model or solve the problem.
4. Collect data, clean data, annotate data.
5. Quantify your solution (eg accuracy, precision, recall, etc). Here, you also need to provide some baselines and explain in what aspects your technique is better. It may be faster, simpler, more accurate, more “visual”, more intuitive, etc, or maybe it is the first method to address and solve this problem.
6. **Deliverables:** A github page with a readme.md that is your report and your results. The git should contain the related code and experiments and also a youtube link to a video showcasing your solution (no more than 15mins). We expect to see a few commits by each person of the team as proof of (team) work.

Problem description

Last summer was the warmest ever recorded in Europe and was associated with droughts, several heatwaves, and wildfires. Increasing water stress due to climate change threatens agricultural production and energy supply. In addition, the intensity of extreme events is expected to increase, raising questions and risks on human infrastructure. In this project, you will try to predict the climate risks (wildfire, heatwaves, flood, water stress, etc.) faced by infrastructures depending on their location, category of activity, etc.. You can imagine being a company deciding on the location of your future facilities; or a farmer wondering about crops’ productivity. To do so, you will work with the TruCost Environmental data, an extensive database measuring environmental impact across key dimensions for over 16,800 companies. If needed, you could combine this dataset with other climate data, such as from the FAO.

Tasks

1. You will be provided access to the TruCost database and documentation, and will need to explore the database and download the data that is useful for the analysis you wish to conduct
2. Start with a good EDA first to understand the data and their quality. Should some data be removed? Do you need additional data from another source? What is the correlation between your variables?
3. Further discover your dataset by performing a graph/clustering analysis shedding some light on the network of companies/facilities in the dataset
4. Design an algorithm to predict climate risks faced by infrastructures based on their location, category of activity, etc. Assess the performance of your algorithm and try to improve it with various techniques learned during the class.
5. Create an interface in WL (or JS) that, given the location of a facility and category of activity, informs the user about the climate risks faced and can suggest alternative “safer” locations.

Note that this is a suggestion. As you further discover the literature, you can decide to focus on a specific topic (e.g., energy, agriculture) and deepen your analysis. Instead of a prediction problem, you could transform it to a classification one (e.g., what categories of assets are riskier? Which companies are more exposed? Etc,). And you should of course be creative in the design of your online interface.

Finally, remember to have fun!