Preparatory Exercise – Applied GIScience: Environment and Health

Please provide concise answers to the following questions, based on the preparatory reading material (and potentially further internet search):

1. What are the key ideas of machine learning, opposed to explicit programming and traditional statistics? Cite the article by Leo Breiman (2001) in your argumentation.
2. What is typically denoted as X and y in supervised learning? Make an example. What is different in an unsupervised setting?
3. How are regression models and classification models evaluated regarding the question whether they are "good models"?
4. How do Kondylatos et al. (2022) attempt to learn something about drivers of wildfires?
5. How do Dutta et al. (2025) come up with scenarios for urban heat in New York?
6. Bonus question: based on the first page of Thiagarajan et al. (2020), what is an emulator model, and why could it be useful?

Upload your answers to these questions in a single PDF file to Moodle until Thursday, July 3rd, 23:59 at the latest.

In the exercise, we will use Jupyter Notebooks. Please watch this short tutorial video before the seminar, to get to know the functionality and navigation elements of this software environment:

https://www.youtube.com/watch?v=LW2Rye_l8L0

Please also consider to make yourself a little bit familiar with Python, if you have not used it before. Don't worry: you will not be expected to code anything from scratch in the exercise. Still, knowing some basic aspects beforehand will be an advantage. On Moodle you will find the first slides from a Python course at HU – however, they are in German. Introductory material in English is everywhere on the internet. For example, you could look here:

https://developers.google.com/edu/python/introduction