The half semester project includes different tasks for a total of **50 project points** which will be converted to semester points and included in your final grade.

1. Organize your group (**5 pts**)

   https://cloud-new.gdb.tools/index.php/s/scRaYPtdqCe2JaG

2. Answer background questions, and upload them to your github (**5 pts**)

   1. Which packages are available for ML? Describe the pros and cons and document the availability.
   2. What is Chembl? How do you access it?
   3. What is machine learning, and how does it differ from traditional programming?
   4. What are the key concepts and techniques in machine learning?
   5. What are the different types of machine learning algorithms?
   6. What are the common applications of machine learning?
   7. How do you evaluate the performance of a machine learning model?
   8. How do you prepare data for use in a machine learning model?
   9. What are some common challenges in machine learning, and how can they be addressed?
   10. What are some resources and tools available to help you learn and practice machine learning?

3. Read tutorial in full before starting anything & answer questions and document the important steps of the process (**10 pts**)

   The tutorial you will follow is below, however we will **NOT** use their training set.

   https://projects.volkamerlab.org/teachopencadd/talktorials/T022_ligand_based_screening_neural_network.html

   The dataset for kinases can be found here (you might need to prepare or alter it):

   https://cloud-new.gdb.tools/index.php/s/ZfZM7itQf3rm6Sw

   1. What is in the training set, how big is it?
   2. What modifications do you need to do to the data set to perform the tutorial.
   3. What is a test set? Any other types of set?
   4. Before starting describe with 1-2 sentences, in your own words, what is done in each of the cells.

4. Perform the tutorial on either your computer or google colab (**10 pts**)

   Document with screenshots or notebook to get the points.

5. Gain access to Ubelix (**5 pts**)

   Read the documentation: https://hpc-unibe-ch.github.io/

1. What is Ubelix?
2. How do you gain access?
3. How do you submit a job?
4. Who can have access?
5. What resources are available there?

6. Reuse the code from the tutorial to run the relevant part (training) on Ubelix. Document how you do the code modification and transfer.
7. (**10 pts**)
8. Run the training step on Ubelix (**5 pts**)

   Provide the submission script and the code you run there.
9. Other considerations:

   You will be using **ssh** to access Ubelix. It is integrated in the windows power shell and mac and linux terminals.

   Filezilla download: https://filezilla-project.org

   Filezilla is used to easily transfer files between servers.

   Download the client Terminal cheatsheet (guide for terminal commands): https://cheatography.com/davechild/cheat-sheets/linux-command-line/