**Objective 1:** Extraction of data from clinicaltrials.gov and making it available in SQL database. Creating a GUI if possible.

* Workflow:

Extracting all the fields from clinicaltrials.gov with the API (beta)

Extract relevant info, like Drug name and phase for ML model

Clean, filter the data and Make it available in SQL database

**Objective 2:** ML model to predict small molecules phase prediction.

* Workflow:

Collecting molecular data of drug names extracted from clinicaltrials.gov from different website like pubchem, zinc and chembl

Extract useful properties from Pubchem API

Separating test and train data from the database created

Data cleaning and filtering

Deciding features to be used for model and creating a database for ML model

Testing different bayesian/ML models