Introduction

* Overview of the topic
* Why is it important / relevant
* What are current methods that are prominently used (“the gold standard”)
  + Machine learning in drug design / activity prediction
  + Deep learning in drug design / activity prediction
* Interactions (very shortly)
* Goals of the thesis

Methods

* Describe the data, how did we get interactions (docking, PLIP - only mention methods, don’t describe them), maybe some statistics about the data
* Data partitioning
  + Train, validation, test split (or cross validation)
* Different machine learning approaches (these should be described in detail, as they are going to be the main part of the thesis)
  + Nearest neighbour approaches
    - Differences in k
    - Differences in distance functions
  + Random forests
    - Lots of parameters to try
    - Feature importance
  + Decision trees
  + Gradient boosted trees
  + Artificial neural networks
  + Evolutionary algorithms
* Quality metrics
  + This will be mostly what I already used in my thesis, the functions should be re-usable for this
* Hyperparameter search
  + If something like grid search is used, should also be mentioned here
* Feature engineering
  + If done, also describe here

Results

* Results of the models, mostly tables but for best performing models in each category also do some plots like confusion matrices, AUC plots if available, bar plots for comparison between models

Discussion and conclusion

* Conclusion: re-iterate the most important points of the results
* Did it perform better than what the scoring model from my master thesis or not
* What could be done to possibly improve results
* Any other things that might be worth mentioning