Assignment#5

CH5650: Molecular Data Science and Informatic
Jan-May, 2025

Due Date: April 11, 2025 (Online Submission, https://courses.iitm.ac.in/)

Data: 3000 molecules in the SMILES representation are provided for learning and predicting new molecules.

Problem Statement: Develop an AI model capable of generating chemically realistic molecules that have not yet been synthesized. In your report, clearly describe the model, its parameters, and the procedure used for molecular generation. Additionally, outline any assumptions made during the process. Based on this exercise, provide a list of 100 new generated molecules, along with relevant structural information such as its uniqueness and similarity to the existing molecules.