**Overview of proposed course contents for the Data Chemistry OLCC**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Unit | Topic | Non-programming activity | Programming activity | Textbook materials *a* |
| 1 | Introduction to Data Chemistry and Course Overview | PubChem interactive tutorial (<https://www.nlm.nih.gov/oet/ed/pubchem/tutorial/index.html>) | Jupyter notebooks (data types, variables, conditions, loops) | Yes |
| 2 | Understanding Public Chemical Databases | PubChem tutorial (<https://doi.org/10.1002/cpz1.217>)  Review of technical aspects of PubChem | Jupyter notebooks (data visualization using PubChem Element data.) | Yes |
| 3 | Chemical representations | Searching PubChem (and a few other public databases) using chemical names and line notations. | No new programming topics. Catch-up week for students without programming background | Yes |
| 4 | Searching chemical databases | Learn advanced topics for data search and retrieval using PubChem, including PubChem Sketcher, classification browser, imaging services, etc. | Search PubChem using APIs  Decode/encode SMILES strings using RDKit | Yes |
| 5 | Exploratory data analysis of PubChem BioAssay data |  | Write a python script to perform exploratory data analysis of PubChem BioAssay data | No |
| 6 | Building an Interactive Dashboard using Tableau | Building an interactive dashboard for bioactivity data in PubChem |  | No |
| 7 | QSAR/QSPR |  | Build a QSAR model from public data. | Yes |
| 8 | Molecular Similarity | Perform 2-D/3-D similarity search using PubChem (revisiting Unit 4) | Python script that tests molecular fingerprints and similarity coefficient. | Yes |
| 9 | Computer-Aided Drug Discovery and Design | Perform molecular docking experiment. | Perform a ligand-based virtual screening using PubChem. | Yes |
| 10 | SQL with python | Build your own database using SQLite. | Access the database using SQL queries. | No |
| 11 | Linked Open Data and Knowledge Graph | Access PubChemRDF data through web browsers | Access PubChem RDF data through a python script. | No |
| 12 | Artificial Intelligence 1: Unsupervised machine learning | Use KNIME to perform a cluster analysis on molecules. | Perform cluster analysis on a group of molecules. | Yes |
| 13 | Artificial Intelligence 2: Supervised machine learning | Use KNIME to build a supervised bioactivity prediction model. | Build a bioactivity prediction model from PubChem’s bioactivity data. | Yes |
| 14 | Artificial Intelligence 3: Generative AI and LLMs in Chemistry | Brief review of all topics covered in the course before final exam. |  | No |

*a* Availability of textbook materials. **YES**: textbook materials from previous OLCCs will be reused for this course. They are currently available on LibreTexts, and we can move them to another platform. Some of these materials may need minor updates and revisions. **NO**: new textbook materials will be created from scratch.