The instructions provided in this document will give step-by-step instructions on how to run the MF-LOGP algorithm that was published in “Dimensionally Reduced Machine Learning Model for Predicting Single Component Octanol-Water Partition Coefficients”.

The instructions are designed for Windows operating systems. It is encouraged that users read through all steps, regardless of coding experience, to avoid confusion in later steps.

Some steps will have bolded text at the end of the statement. These texts are either directional instructions for the computer or code that can be directly copied into the terminal.

For questions about running this program, please contact the principal investigator:

Dr. Andrew R. Teixeira

[arteixeira@wpi.edu](mailto:arteixeira@wpi.edu)

**Download Python**

1. Download and install the newest version of Python ([here](https://www.python.org/downloads/)). Take note of the install file directory.

**Graphical user interface, text, application

Description automatically generated**

1. Open the install directory and confirm that an executable instance of python is located within the folder

**Graphical user interface

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated**

1. Open a Windows Command Prompt: **Start (Windows) > “Command Prompt”**

Application

Description automatically generated with low confidence

1. Once the prompt is opened, in the command, confirm that you are operating in the newest version of Python by running **py --version**
   1. As of Aug, 2022 🡪 Python 3.10.6
2. The pip install tool will be a critical aspect to running the codes. Please confirm that the pip install tool was downloaded with the Python application by running **py -m pip help**
   1. The help menu should appear if present. If the help menu does not appear, or an error message is presented, re-install Python and ensure that pip is included in the downloaded packages.

Graphical user interface, text

Description automatically generated

**Downloading MFLOGP Files from GitHub**

1. Go to the Teixeira Research Lab GitHub and enter the MF-LOGP\_Development- branch ([here](https://github.com/TeixeiraResearchLab/MF-LOGP_Development-))
2. Download a copy of the repository in a zip file format: **Code > Download ZIP**

Graphical user interface, text, application, email

Description automatically generated

1. The file will download as MF-LOGP\_Development—main. Unzip and extract the file to the same Python directory noted in **Step 1**. Please note the directory of this file as well.



1. Open the file and ensure you have the following files:
   1. MFLOGP.zip
   2. MFLOGP\_Run\_Code.py
   3. scale\_X.sav
   4. scale\_y.sav

NOTE: Additional files will be downloaded but the four listed above are the core files needed to predict partition coefficients

1. Due to size constrains the published MF-LOGP algorithm had to be compressed to be published on GitHub. It is important that the MFLOGP.zip be unzipped and contents extracted to be with the other files in MF-LOGP\_Development--main folder

**Download All Relevant Python Packages**

If you have just downloaded Python and are running code via the command prompt, you will need to download function packages to properly use this code. If you have these packages installed already you can skip this portion. If you are unsure if the packages are downloaded, you can run the commands below and it will inform you if the packages are already present.

1. pip update: **py -m pip install --upgrade pip**
2. NumPy ([Documentation](https://numpy.org/)): **py -m pip install numpy**
3. Pandas ([Documentation](https://pandas.pydata.org/)): **py -m pip install pandas**
4. Sklearn ([Documentation](https://scikit-learn.org/stable/)): **py -m pip install sklearn**
5. MatPlotLib ([Documentation](https://matplotlib.org/)): **py -m pip install matplotlib**
6. Chemparse ([Documentation](https://pypi.org/project/chemparse/)): **py -m pip install chemparse**
7. Joblib ([Documentation](https://joblib.readthedocs.io/en/latest/)): **py -m pip install joblib**

**Editing Python Scripts**

1. Using the File Explorer, enter the MF-LOGP\_Development--main folder that was placed in the Python directory. Right-click on the MFLOGP\_Run\_Code.py to open the script in a notepad

Text

Description automatically generated

1. Scroll down to the **User Inputs** section:

Text

Description automatically generated

* 1. First, choose if you would like to analyze single compound or multiple compounds via an excel spreadsheet. The 1 and 0 are logical statements **(1 = Yes, 0 = No)**
  2. If looking at a single compound (**single\_compound = 1)**, simply change the molecular formula to the compound of interest.

NOTE: This algorithm can only accept compounds with (C, H, N, O, S, P, F, Cl, Br, I) without further training.

Table

Description automatically generated

* 1. If looking at multiple compounds (**compound\_list = 1**), add the file directory for the excel document containing the list of compound formulas under **file\_dir**
     1. If running multiple compounds, ensure that the compound list is in Excel, the sheet name matches the Excel document, and the list has a header ‘**Formula’**
  2. Save and close the excel document

1. Save the document. The notepad can be closed or remain open.
2. Change the command prompt directory to that of the unique file path from **Step 8**: **cd *Enter File Path***
   1. Once entered, you will see a change in the text prior to the cursor location

 

1. Run the MF-LOGP algorithm: **py -m MFLOGP\_Run\_Code**
   1. The output might be led by warnings messages, but the final partition coefficient prediction(s) will be the final output in double brackets

Single compound:

Multiple compounds: Text

Description automatically generated

End of Instructions