

1) Objective wise brief module or learning outcomes

- a. Data retrieval
- b. Chemical Structure Representation via Descriptors or Features
- c. Descriptor Selection for Dimension Reduction
- d. Classification Using Machine Learning Methods
- e. Evaluating classifier performance

2) Datasets to be used

The high throughput bioassay dataset will be downloaded from the PubChem database [AID: 891]. The bioassay was conducted to identify inhibitors and substrates of Cytochrome P450 2D6.

3) Pre-requisite for participants

Laptops

4) Software which you can download or install

- a. Weka-3-8-3jre-x64.exe; 120.3 MB, Windows
Weka-3-8-3.zip; 51.4 MB, Other platforms (Linux, etc.)
URL: <https://www.cs.waikato.ac.nz/ml/weka/downloading.html>
- b. PaDEL-Descriptor
URL: <http://www.yapcwsoft.com/dd/padeldescriptor/>