ChemML: A Machine Learning and Informatics Program Package for the Analysis, Mining, and Modeling of Chemical and Materials Data

Nitin Murthy[12], Johannes Hachmann[13]

[1]) Department of Chemical and Biological Engineering, University at Buffalo SUNY
[2])nitinmad@buffalo.edu
[3]) hachmann@buffalo.edu

Recent advances in high-throughput screening (HTPS) studies have enabled the collection of large amounts of chemical data, with the primary motivation being modeling properties of molecules via data-driven approaches. However, the construction of data-driven models (such as machine learning models) can be a daunting prospect, with expertise needed in molecule representation, model parameter tuning and evaluating metrics and visualizing model performance.

We at the Hachmann Lab have developed ChemML, a software package written in Python that aims to simplify this process for those unfamiliar with the field of machine learning. ChemML features several modules to represent, model and visualize model results, given a set of molecules and a target property. In particular, a recent development is AutoML, a wrapper that aims to further simplify the process by automating representation and model optimization.

In this workshop, we aim to introduce ChemML and demonstrate these capabilities live. The source code as well as installation instructions can be found at https://github.com/hachmannlab/chemml