Molecular dynamics (MD) is a computer simulation technique for studying structure and dynamics of complex systems with extreme detail--literally on scales where the motion of individual atoms can be tracked. MD acts as a bridge between experiment and theory to develop a realistic model which describes experimental measurements. We can test a model by conducting a simulation and comparing the simulation result with experimental measurements. At the same time, the hidden detail behind the experiment can be revealed. The main strength of MD method is the ability to determine the time evolution of a system at atomic-level resolution which brings new insights into mechanisms and processes that are not directly accessible through experiment. In addition, we are able to carry out simulations that are too difficult or impossible to be treated with standard experimental tools. Providing information at atomic-level resolution, MD has subsequently become an invaluable tool to facilitate and complement experimental studies in different research fields such as chemistry, biochemistry, and material science.

MD simulations are required various input parameters and their results are validated against experimental data. The underlying physics of interaction of two atoms is too complex to describe with a function, and therefore optimizing a set of input parameter that would adequately mimic the real materials is nontrivial. Due to the complexity of the problem, an empirical approach is mostly used to optimize input parameter. In this technique, the optimization procedure is based on the reproduction of experimental measurements by simulating a small system. Via an iterative approach, the input parameters are primarily optimized to reproduce experimental data. This strategy is significantly demanding since MD simulations are very expensive and results are very sensitive to the input parameters.

The main aim of this project is to employ machine learning algorithms to find a relationship between input parameters and simulation results. It is a supervised learning problem, more specifically a regression problem. The input parameters that we want to predict are called the dependent variables or target variables. In contrast to most of the supervised learning problem, here we have more than one target variable. The MD simulation results that we will use to predict the target variables are called independent variables or regressor variables. We also would like to study the influence of the number of training sample in the accuracy of our prediction. MD simulations are very expensive, and developing a model to predict the input parameters with a limited number of training sample can significantly reduce the computational time.

Here, we will employ several algorithms for training, and we will evaluate the performance of each algorithm by calculating an appropriate score. To optimize the parameters of each algorithm, we will use grid search and cross-validation techniques. Finally, the optimized method will be used to predict the input parameters based on the experimental measurements. Using the predicted parameters, we will perform MD simulation and calculate the accuracy of MD simulation by comparing simulation results and experimental measurements.