**HOMO energy level**

Highest occupied molecular orbital

**LUMO energy level**

lowest unoccupied molecular orbital

**Artificial Neural Network**

A computational model that consists of several processing elements that receive inputs and deliver outputs based on their predefined activation functions

**Coefficient of determination**

The coefficient of determination (R²) is a number between 0 and 1 that measures how well a statistical model predicts an outcome.

**Root mean square error**

Root mean squared error (RMSE) is the square root of the mean of the square of all of the error.

**DFTB+ module**

DFTB+ is a versatile community developed open source software package offering fast and efficient methods for carrying out atomistic quantum mechanical simulations.

**Forcite module**

Forcite is an advanced classical molecular mechanics tool, that allows fast energy calculations and reliable geometry optimization of molecules and periodic systems.

**KFOLD**

Cross-validation is usually used in machine learning for improving model prediction when we don't have enough data to apply other more efficient methods like the 3-way split (train, validation and test)

**Determination coefficient**

It provides a measure of how well observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model.