

Code-based documentation of dataset generation using active learning:

Step 1:

Iteration == 0: Create training pool data

- a) Calculate train (80%) and test (20%) value in train_test_split(..) method.
- b) Apply split(..) method for creating train and test data in Training_pool_dataset folder.
- c) Save split data by using write_selected_files() method

Step 2:

Iteration == {1...n}: Query selection data

- a) Select number of percentage data from train folder in Training_pool_dataset.
- b) Create Query_selection_dataset folder by using query_based_selection(..) method.
- c) Calculate train (80%) and valid (20%) value in train_test_split(..) method.
- d) Apply split(..) method for creating train and valid data in Iteration_dataset folder.
- e) Save split data by using write_selected_files(..) method
- f) Create final dataset in Active_learning_datasets folder by applying file_transfer() method.
- g) Delete Query_selection_dataset folder.