AI-Disaccharide Classifier

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1. Introduction

AI-Disaccharide Classifier is meant to serve as an artificial intelligence (AI)-platform for analyzing the tunneling readouts of disaccharide isomers using Google Colab for the identification of complex disaccharide constitutional and regioisomers.

The program contains four steps as follows:

a. Feature generation: import the tunneling readouts transmission and energy in ‘.xlsx’ format and generate the features for each individual isomer and prepare the input datasets.

b. Import dataset: training dataset, testing dataset and validation dataset in ‘.xlsx’ format.

c. Train ‘XGBC Classifier’ using training dataset and testing dataset and get output training accuracy, testing accuracy, and classification report.

d. Using the trained XGBC classifier, make predictions on unknown dataset and get the output confusion matrix with labels.

2. Operating procedures:

-Unzip the demo\_tunneling.rar to local folder.

-Open the AI-Disaccharide Classifier in Google Colab.

-Enter the file name of the training datasets. // example: x\_train.xlsx and y\_train.xlsx

-Enter the file name of the testing datasets. // example: x\_test.xlsx and y\_test.xlsx

-Enter the file name of the unknown datasets. // example: x\_val.xlsx and y\_val.xlsx

-Run the code

3. Output

The value of the variable 'Accuracy' is the output in the command Widow. Confusion matrix with labels is plotted and saved as 'confusion\_matrix.jpg' in local folder.