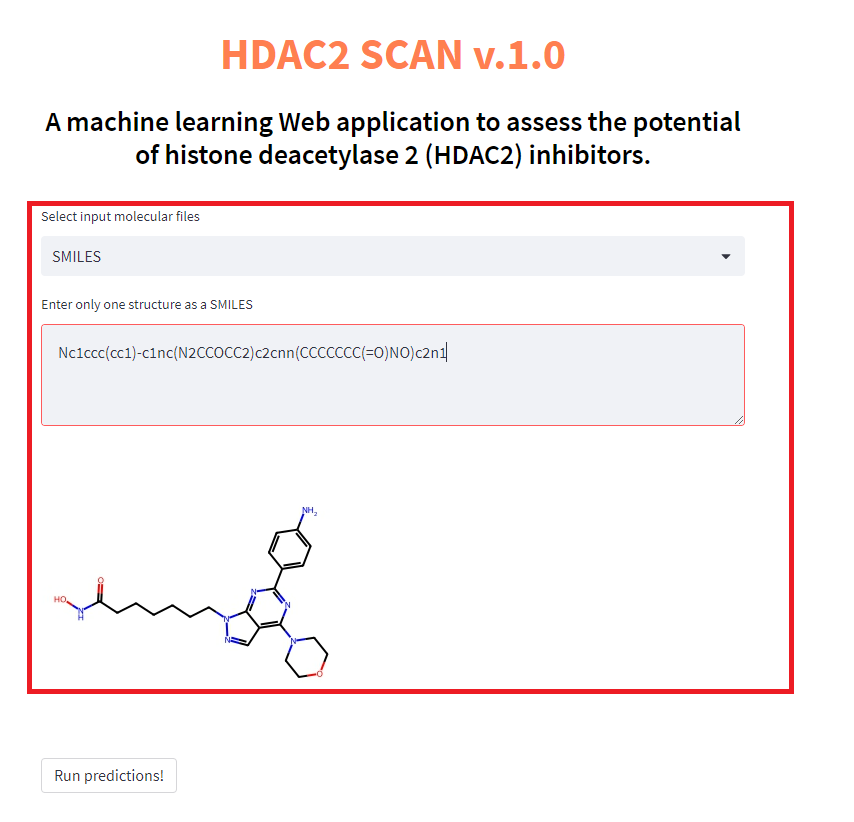
**Manual of HT\_PREDICT**

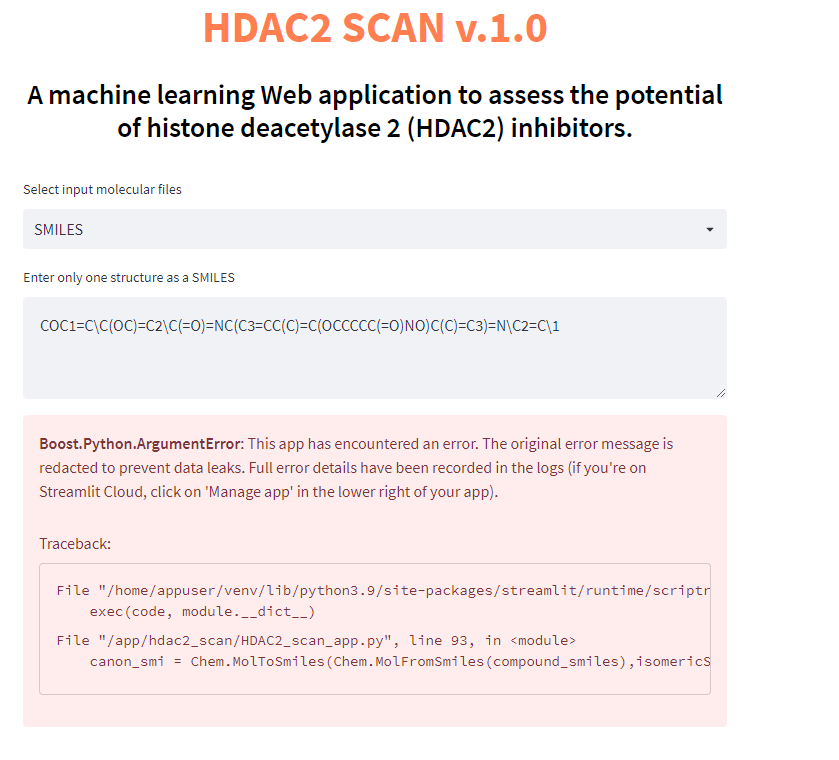
**open-source software**

**Step 1.** *Select input molecular files.*

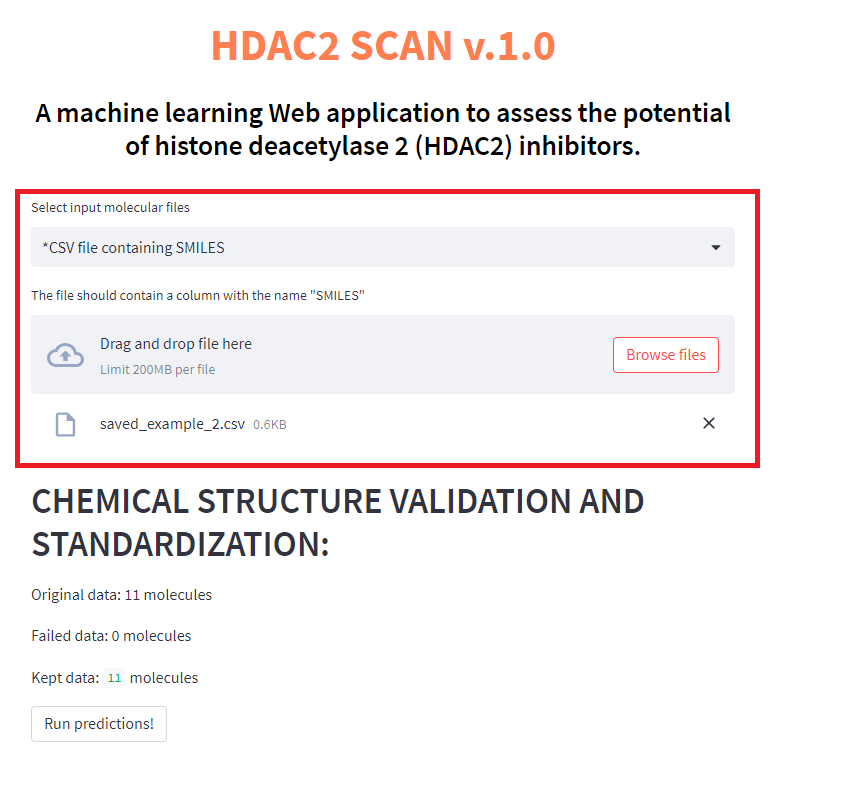
If you choose smiles, please, directly paste the SMILES representation of the desired chemical structure and press Ctrl+Enter. If the entered chemical structure is correct, the application will generate a 2D image of the studied compound.



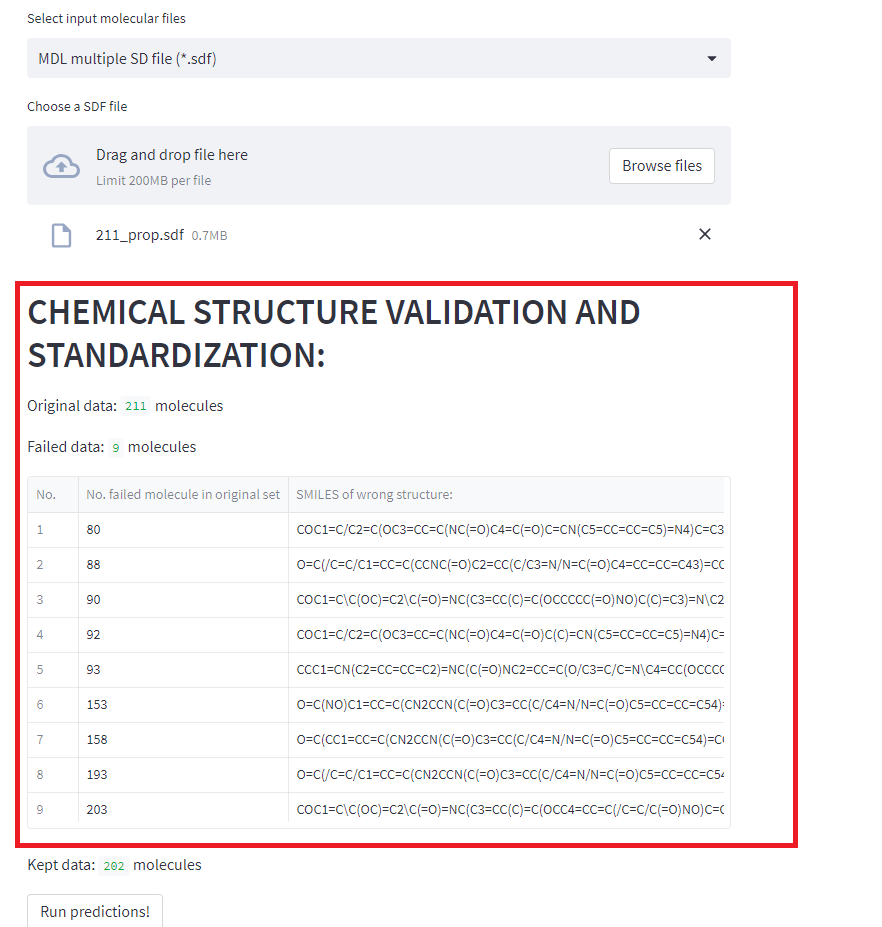
If the entered structure is incorrect, the application reports an error.



If you choose a file \*sdf or \*csv, that may contain a different number of chemical structures, please specify the path to this file on your computer's hard drive. In this case, you need to click the "Browse files" button. It is important to note that if you choose a file with the \*csv extension, the file should contain a column with the name "SMILES"



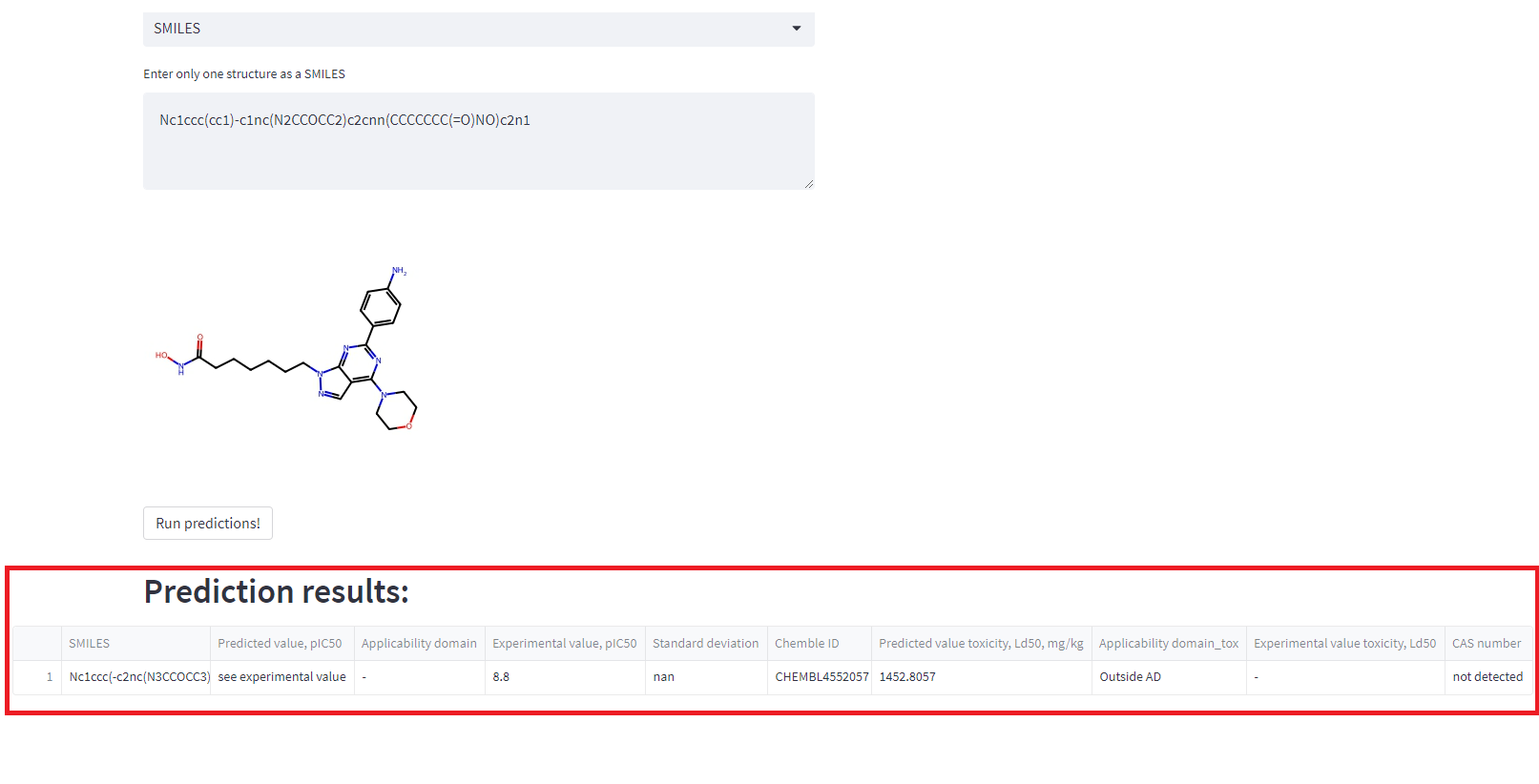
If incorrect structures are detected in the file \*sdf or \*csv, the corresponding information will appear in the section "CHEMICAL STRUCTURE VALIDATION AND STANDARDIZATION"



**Step 2**. Click on the “Run predictions!” button for prediction.

**Step 3**. *Prediction results*.

The form of presentation of the results depends on the type of descriptors selected, as well as the format of the input chemical data. For example, when selecting SMILES, the results will be displayed for a single molecule. When displaying the results on the screen, it is taken into account whether there are experimental values of activity and toxicity for the studied compound.



The final table contains the following columns:

1) ***SMILES*** - the chemical structure is displayed in the SMILES notation

2) ***Predicted value, plC50*** - the predicted value of the activity to inhibit the HDAC2 enzyme, expressed in pIC50, where pIC50 is the negative logarithm of IC50 in molar concentration. If experimental data is available in the ChEMBL database, the label "see experimental value" is displayed in this cell.

3) ***Applicability domain*** - compliance of the chemical compound with Applicability domain. If experimental data is available in the ChEMBL database, the label "-" is displayed in this cell.

4) ***Experimental value, plC50***- experimental data presented in the ChEMBL database. Where pIC50 is the negative logarithm of IC50 in molar concentration. If there is more than one value in the database, then the average value is given in the cell.

5) ***Standard deviation*** - еhe cell indicates the standard deviation of the experimental activity values presented in the ChEMBL database.

6) ***Chemble ID*** - identifier from the ChEMBL chemical database of molecule

7) ***Predicted value toxicity, Ld50, mg/kg*** - predicted value of acute toxicity when administered orally to rats. If experimental data is available in the PubChem database, the label "see experimental value" is displayed in this cell.

8) ***Applicability domain\_tox*** - compliance of the chemical compound with Applicability domain. If experimental data is available in the PubChem database, the label "-" is displayed in this cell.

9) ***Experimental value toxicity, Ld50*** - experimental data presented in the PubChem database. Toxicity was measured by a dose of LD50 when administered orally to rats

10) ***CAS number*** - a unique identification number assigned by the Chemical Abstracts Service (CAS)

If you choose a file \*sdf or \*csv, the prediction results for correct chemical structures are displayed in a table that can be downloaded.

