**Drug Interactions Toolkit (DITK)**

**Description**

This Python script uses data from DrugBank.com to identify interactions between drugs, and then analyzes a dataset of patients to determine which drugs are interacting with each other…

* Plotting histograms and CDF of each column in the dataframe
* Generating plots and saving them in the designated folder
* Calculating the correlation matrix based on different methodologies
* Plotting the heatmap of the correlation matrix and saving the file in the designated folder
* Filtering correlation coefficients based on a given threshold. List of weak, strong, and null correlations are generated. The threshold is requested as input from the user
* The script uses input files from the input folder, while saved files and plots can be found in the output folder.

**Prerequisites**

Make sure that the dataframe is in the correct format and encoding before running the script. The files must be placed in the input folder.

**Package requirements**

The script is tested on Python (Anaconda) V 3.10.13

The following packages are required for the correct execution of the script:

* numpy
* distfit
* random
* tqdm
* os
* sys
* pandas
* matplotlib
* math
* seaborn

**Usage**

Run the script by typing python … in the command line. The dataset folder is the designated directory to hold the target clinical dataset for analysis. It contains two subfolders (namely dictionary and interactions)

Follow the prompts to input the threshold value for filtering the correlation coefficients.

The script will generate plots and outputs that can be found in the designated output folder.

**Output**

Histograms and CDF plots of each column in the dataframe

Correlation matrix heatmap plot

List of weak, strong and null corre