Diversity Validation Program:

To observe structural diversity in selected dataset, in terms of distance scores (calculated by the Euclidean distance norm).

Note: To observe diversity among compounds present in dataset; plot a 'scatter plot' of Normalized Mean Distance Vs Respective Activity/Property.

Input File: Single *.csv file (saved as 'comma delimited' in Microsoft Excel).
Input File Format: Column containing descriptors and their values for each compound present in dataset. Header*(i.e. name of respective descriptor) for each column is must. To further clarify file format, sample input files are provided. (* *mandatory for correct calculations*)

To run the program: Download ***DiversityValidation.jar*** file (it is platform-independent) and just single/double click on it to run the program.
Note: Keep your input files in the same folder where you keep ***DiversityValidation.jar*** file.

Output File: [Name of input dataset]_Diversity.csv
Output File Format: Single file containing four columns titled Compound No.(are in the same order to the compounds present in input dataset file), Distance Score, Mean Distance, Normalized Mean Distance.
Note: Please close all input/output files before running this program.

*Snapshot: *_Diversity.csv in Microsoft Excel*

	A	B	C	D
1	Compound No.	Distance Score	Mean Distance	Normalized Mean Distance
2	1	3187134.76	28712.926	0.274
3	2	3305280.305	29777.3	0.293
4	3	3113012.95	28045.162	0.262
5	4	2978833.451	26836.337	0.24
6	5	3509414.619	31616.348	0.327
7	6	2340221.769	21083.079	0.136
8	7	1511959.281	13621.255	0
9	8	1527125.348	13757.886	0.003
10	9	2080292.248	18741.372	0.093
11	10	1907650.514	17186.041	0.065
12	11	2780917.02	25053.306	0.208
13	12	2735138.712	24640.889	0.2
14	13	1605046.258	14459.876	0.015
15	14	1886124.862	16992.116	0.061
16	15	1885900.209	16990.092	0.061
17	16	1845153.721	16623.006	0.055

Disclaimer

For academic purpose only;

The program "Diversity Validation" has been developed (in Java) and validated on known data sets by Pravin Ambure (ambure.pharmait@gmail.com) of Drug Theoretics & Cheminformatics (DTC) Laboratory, Jadavpur University (2013).