



# **USEDAT: USA-Europe Data Analysis Training School Workshop**



# **Batch processing in transformation of continuous variables for PTML Theory**

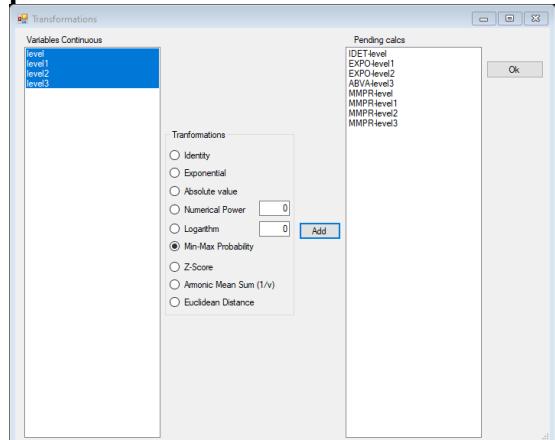
Bernabe Ortega-Tenezaca (bortega@uea.edu.ec)<sup>a,b</sup>,  
Viviana Quevedo-Tumailli (vquevedo@uea.edu.ec)<sup>a,c</sup>.

<sup>a</sup> *RNASA-IMEDIR, Computer Science Faculty, University of A Coruña, 15071, A Coruña, Spain.*

<sup>b</sup> Universidad Estatal Amazónica, UTIC, Puyo - Ecuador

<sup>c</sup> Universidad Estatal Amazónica, Science of earth Departament, Puyo – Ecuador

## Graphical Abstract



### **Abstract.**

In the present work, a software module has been developed that allows the selection of continuous variables from an Excel file, which are initially subjected to a process of verification and cleaning of the information, allowing the elimination of cases, or otherwise replacing outliers. With the average value, it is then possible to perform transformation operations such as Identity, exponential, absolute value, numerical power, logarithm, maximum and minimum probability, z-score, harmonic mean sum, Euclidean distance, in batch processing of continuous variables. In the end, the respective results can be obtained within a dataset that can be stored in CSV format, or in turn continue processing with PTMI.

## References

1. Vásquez-Domínguez E, Armijos-Jaramillo VD, Tejera E, González-Díaz H. Multioutput Perturbation-Theory Machine Learning (PTML) Model of ChEMBL Data for Antiretroviral Compounds. *Mol Pharm.* 2019 Aug 30. doi:10.1021/acs.molpharmaceut.9b00538. [Epub ahead of print] PubMed PMID: 31426639.
2. Concu R, D S Cordeiro MN, Munteanu CR, González-Díaz H. PTML Model of Enzyme Subclasses for Mining the Proteome of Biofuel Producing Microorganisms. *J Proteome Res.* 2019 Jul 5;18(7):2735-2746. doi:10.1021/acs.jproteome.8b00949. Epub 2019 Jun 4. PubMed PMID: 31081631.
3. Bediaga H, Arrasate S, González-Díaz H. PTML Combinatorial Model of ChEMBL Compounds Assays for Multiple Types of Cancer. *ACS Comb Sci.* 2018 Nov 12;20(11):621-632. doi: 10.1021/acscombsci.8b00090. Epub 2018 Oct 3. PubMed PMID: 30240186.
4. Arrasate S, Duardo-Sánchez A. Perturbation Theory Machine Learning Models: Theory, Regulatory Issues, and Applications to Organic Synthesis, Medicinal Chemistry, Protein Research, and Technology. *Curr Top Med Chem.* 2018;18(14):1203-1213. doi: 10.2174/15680266180810124031. Review. PubMed PMID: 30095052.
5. Simón-Vidal L, García-Calvo O, Oteo U, Arrasate S, Lete E, Sotomayor N, González-Díaz H. Perturbation-Theory and Machine Learning (PTML) Model for High-Throughput Screening of Parham Reactions: Experimental and Theoretical Studies. *J Chem Inf Model.* 2018 Jul 23;58(7):1384-1396. doi:10.1021/acs.jcim.8b00286. Epub 2018 Jun 27. PubMed PMID: 29898360.
6. Ferreira da Costa J, Silva D, Caamaño O, Brea JM, Loza MI, Munteanu CR, Pazos A, García-Mera X, González-Díaz H. Perturbation Theory/Machine Learning Model of ChEMBL Data for Dopamine Targets: Docking, Synthesis, and Assay of New I-Prolyl-I-leucyl-glycinamide Peptidomimetics. *ACS Chem Neurosci.* 2018 Nov 21;9(11):2572-2587. doi:10.1021/acschemneuro.8b00083. Epub 2018 Jun 25. PubMed PMID: 29791132.
7. Munteanu CR, Gestal M, Martínez-Acevedo YG, Pedreira N, Pazos A, Dorado J. Improvement of Epitope Prediction Using Peptide Sequence Descriptors and Machine Learning. *Int J Mol Sci.* 2019 Sep 5;20(18). pii: E4362. doi:10.3390/ijms20184362. PubMed PMID: 31491969.
8. Munteanu CVA, Chiritoiu GN, Petrescu AJ, Petrescu SM. Profiling Optimal Conditions for Capturing EDEM Proteins Complexes in Melanoma Using Mass Spectrometry. *Adv Exp Med Biol.* 2019;1140:155-167. doi:10.1007/978-3-030-15950-4\_9. PubMed PMID: 31347047.
9. Tenorio-Borroto E, Castañedo N, García-Mera X, Rivadeneira K, Vázquez Chagoyán JC, Barbabosa Pliego A, Munteanu CR, González-Díaz H. Perturbation Theory Machine Learning Modeling of Immunotoxicity for Drugs Targeting Inflammatory Cytokines and Study of the Antimicrobial G1 Using Cytometric Bead Arrays. *Chem Res Toxicol.* 2019 Aug 26. doi: 10.1021/acs.chemrestox.9b00154. [Epub ahead of print] PubMed PMID: 31327231.
10. Sarbu M, Ica R, Petrut A, Vukelić Ž, Munteanu CVA, Petrescu AJ, Zamfir AD. Gangliosidome of human anencephaly: A high resolution multistage mass spectrometry study. *Biochimie.* 2019 Aug;163:142-151. doi:10.1016/j.biochi.2019.05.017. Epub 2019 Jun 13. PubMed PMID: 3120184600.