Machine-Learning models to predict the antioxidant capacity of food

Estela Guardado Yordi [1,2]*, Raúl Koelig[1], Maria João Matos[2], Yailé Caballero Mota[1], Eugenio Uriarte[2], Enrique Molina[1,2], Amaury Pérez Martínez[1,3], Lourdes Santana[2] and Enrique Molina[1,2]

[1] Universidad de Camagüey Ignacio Agramonte Loynaz, Circunvalación Norte Km 5 ½, Camagüey, Cuba; email: estela.guardado@reduc.edu.cu; raul.koelig@reduc.edu.cu; yaile.caballero@reduc.edu.cu; enrique.molina@reduc.edu.cu

[2] Universidad de Santiago de Compostela, Address Avda. das Ciencias, s/n. Campus sur 15782 Santiago de Compostela, Spain; email: eugenio.uriarte@usc.es; lourdes.santana@usc.es; mcmatos@gmail.com

[3] Universidad Estatal Amazónica, Address Km 2 ½ vía Puyo a Tena (Paso Lateral). CP. 160150. Puyo, Ecuador; email: amperez@uea.edu.ec

* Author to whom correspondence should be addressed; estela.guardado@reduc.edu.cu; Tel.: +53 261192.

Abstract: The growing increase in the amount and type of nutrients in food created the necessity for a more efficient use in dietetics and nutrition. Flavonoids are exogenous dietary antioxidants and contribute to the total antioxidant capacity of the food. The current work aims to obtain optimal models to predict the total antioxidant properties of food by the ORAC method. A dataset based on the Database for the Flavonoid Content of Selected Foods and the Database for the Isoflavone Content of Selected Foods, was created. Different algorithms of artificial intelligence were applied, in particular Machine-Learning methods. They were employed using a R language. The performed study allowed to show the effectiveness of the models using structural-topologic features of Topological Substructural Molecular Design (TOPSMODE) in the databases. The proposed models can be considered, without overfitting, effective in predicting new values of ORAC, excepting the MultiLayer Perceptron (MLP) algorithm. The optimal model was obtained by the Random Forest (RF) algorithm, which presented the best R2 of the series (R2 = 0.9571313 for the training series and R2= 0.9247337 for the external prediction series).

Keywords: keyword; keyword; keyword (3–10 keywords separated by semi colons)

Mol2Net YouTube channel: *http://bit.do/mol2net-tube*
YouTube link: *please, paste here the link to your personal YouTube video, if any.*