Editors:
T.I. Madzhidov, V.A. Afonina

Editor-in-chief:
I.S. Antipin (corresponding member of the RAS)

T45 **Third International School-Seminar "From Empirical to Predictive Chemistry"** (April 5–7, 2018, Kazan, Russia): book of abstracts / ed.: T.I. Madzhidov, V.A. Afonina; I.S. Antipin. – Kazan: Kazan University Press, 2018. – 96 p.

This book contains abstracts, accepted for presentation in Third International School-Seminar "From Empirical to Predictive Chemistry" (E2PC 2018, 5–7 April 2018, Kazan).

V. Solov'ev [1]
A. Varnek [2]
G. Marcou [2]

THE ISIDA_NB PROGRAM: CONSENSUS NAÏVE BAYESIAN MODELING ON FRAGMENT DESCRIPTORS

[1] A.N. Frumkin Institute of Physical Chemistry and Electrochemistry, Russian Academy of Sciences, Leninskiy prosp., 31, 119071 Moscow, Russia;
[2] Laboratoire de Chémoinformatique, UMR 7140 CNRS, Université de Strasbourg, 1, rue Blaise Pascal, 67000 Strasbourg, France.
solovev-vp@mail.ru

In this presentation, we report implementation of consensus naïve Bayesian classifier ISIDA_NB into the ISIDA_QSPR program package [1, 2].

Ensemble modeling workflow implemented into the ISIDA_NB program includes (i) generation of ensemble of individual classification models, (ii) selection of the most predictive ones using internal and external cross-validations and (iii) consensus application of selected models to a test set compounds. Ensembles of individual models are generated by varying parameters of naïve Bayesian model (weight of Laplace correction, score threshold), variable selection techniques (variable filtering, excluding concatenated fragments) and using of multiple types of molecular fragments [3] as descriptors. The program runs under the Windows operating system. The graphical interface pilots this workflow and supports the analysis of the obtained results. The Consensus Predictor tool realizes property predictions and virtual screening using previously obtained consensus models.

The program was used to classify organic ligands (L) onto weak and strong binders of different metal cations (M) in water. Metal binders were classified according to stability constant ($\log K$) threshold for 1:1 (M:L) complexes. For strong binders, $\log K$ was assumed to be more than 5.5. The ISIDA_NB consensus models demonstrate a good predictive performance in the fivefold cross-validation procedure: the balanced accuracy on predictions vary from 0.8 to 0.9 for studied metal cations: Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}, Fe^{2+}, Fe^{3+}, Mn^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, La^{3+}, Pb^{2+}.

1. Solov'ev V. et al. *Chemoinformatics*, 2017: 135 – 162.
2. Solov'ev V. ISIDA (In Silico Design and Data Analysis) program; version 5.79; 2008–2018. http://vpsolovev.ru/programs/ (accessed 9 February 2018).
3. Varnek A. *Chemoinformatics and Computational Chemical Biology*, 2011: 213–243.