SHAASTRA 2023

Computational Epidemiology of SARS-Cov-2 (COVID-19)

Event Description

Computational Biology (or Bioinformatics) brings order to our understanding of life. And what's more important than being able to "debug" viruses, agents that operate in the grey area between living and non-living? They've baffled researchers for generations; now it's our turn to take a little peek into their world and figure out how to stop them.

Introduction to Round 3

Congratulations on making it to the final round of Biogen, Shaastra 2023!

Quantitative structure-activity relationship (QSAR) models are mathematical models that can predict compounds' physicochemical, biological and environmental fate properties from the knowledge of their chemical structure.
This final problem statement allows you to explore the essentials of drug design using QSAR modelling, the basis for developing anti-SARS-Cov-2 agents.

Problem Statement 3

Q) Develop a regression-based QSAR (quantitative structure-activity relationships) model to identify compounds that have 3-chymotrypsin-like protease (3CLpro) inhibitory activity.

- You will be given a dataset (provided below) containing compound SMILES notations and their 3CLpro inhibitory activity measured by its pIC50 value (in μM).
- Build a regression-based QSAR model using 2D descriptors.
- Provide (in a similar format to the dataset) the pIC50 values of all 'blind' compounds (ones with 'BLINDED' pIC50 scores). From these 'blind' compounds, choose the one you think is most promising for drug design and briefly explain your choice.
- **Brownie points**: Split your dataset into train and test sets and provide internal validation based on the test set compounds (R2, LOO-Q2 scores). Explain how your model can pass the Golbraikh and Tropsha acceptable model criteria.

Dataset

Computational Epidemiology Hackathon PS3 Dataset

References

Computational Epidemiology PS3 Tools