Methodology Report

Description of the Model

Describe the ML model or method you used. The following keypoints need to be included:
- MAP4 (r=3, 2048 bits), MACCS (167 fp) and Mordred (1658 descriptors after fixing) were used;
- PCA was applied independently on each feature space
- The principal components from each feature space were concatenated as features XGBoost models with 5-fold CV for hyperparameter optimization
- Both scaffold and random splits were used

Performance comments

Both scaffold and random splits were used; random splits showed marginally better performance on the internal holdout test set.

Team members:

Ashok Palaniappan, Ph.D
Ida Titus (Graduate student)