

Практикум 5

Лоскутова Софья, 615 группа, химический факультет

Train: [train](#)

Eval: [eval](#)

Была проведена тренировка и перетренировка генеративной модели, были прописаны функции-метрики, считающие желаемые свойства.

1) `def water_octanol_partition_coefficient_scores(mols, norm=False)`

Известно, что молекулы, которые потенциально могут использоваться в медицине, обладают logP от -0.4 до 5.6. <https://en.wikipedia.org/wiki/Druglikeness>

Поэтому функция возвращает единицу, если молекула удовлетворяет данному условию, если нет, то 0.

2) `def activity_coefficient(mols)`

В данной функции считается активность к herg с помощью модели-классификатора из второго практикума

`RandomForestClassifier(class_weight='balanced_subsample', criterion="entropy", n_estimators=500, warm_start=True)`

Модель была обучена на датасете herg.sdf.

Стоит заметить, данная модель не предсказывает ни одной активной к herg молекулы из исходного датасета.

В ноутбуке eval находятся сгенерированные молекулы. Ни одна не обладает нужной активностью, поэтому были посчитаны вероятности к активности. Только 11 молекул обладают вероятностью больше 2.00%. Для них были посчитаны logP. Только 8 имеют нужное значение. Данные молекулы представлены на рисунке.

