4. 結果

構成したデータセットを学習データ174個、テストデータを20個に分けた。学習において、k分割検証法(k:4)、機械学習モデルはXGBoostとLightGBMを用い1:1でアンサンブルした。結果、テストデータについてF1\_score: 70.6%の性能を示した。

　また各モデルが学習において重要と判断した特徴量を可視化した(図２)。その結果、各モデル共にMean alp. sph. Solvent access, Polarity score, Scoreが重要特徴量の上位であることがわかった。

グラフ

自動的に生成された説明グラフ

自動的に生成された説明

図2: 重要特徴量の可視化(Xgboost(左)とLightGBM(右))