

Pembuatan Model HKSA dengan Pembelajaran Mesin Menggunakan Pustaka PyCaret: Studi Kasus pada Penghambat DPP-4 = QSAR Model Building with Machine Learning Using the PyCaret Library: A Case Study on DPP-4 Inhibitor

Simangunsong, Andrew Jonathan Brahms, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920540449&lokasi=lokal>

Abstrak

Diabetes merupakan salah satu penyebab kematian terbesar di dunia. Salah satu obat untuk menangani diabetes adalah penghambat DPP-4. Obat penghambat DPP-4 berpotensi menjadi obat diabetes aksi panjang, menyelesaikan masalah tatalaksana diabetes terkait ketidakpatuhan pasien dalam mengonsumsi obat. Namun, belum ada obat penghambat DPP-4 aksi panjang dengan hasil memuaskan. Obat penghambat DPP-4 juga masih memiliki masalah berupa efek samping pankreatitis dan nyeri sendi. Oleh karena itu, penelitian ini bertujuan menemukan kandidat penghambat DPP-4. Penapisan virtual dilakukan dengan model HKSA pembelajaran mesin yang dilatih dengan data aktivitas DPP-4 menggunakan pustaka PyCaret. Hasilnya didapatkan tiga model yaitu model klasifikasi biner dan multiclass untuk menentukan aktivitas molekul dan model regresi untuk memprediksi tingkat aktivitas. Model dievaluasi dengan membandingkan kualitas model dan hasil dengan penelitian Hermansyah et al. Hasil penelitian menunjukkan bahwa model klasifikasi biner yang dibuat memiliki kualitas yang lebih baik, sementara model regresi sedikit lebih rendah. Perbedaan ini disebabkan oleh variasi data dan pendekatan penggunaan model. Kesepuluh senyawa yang diidentifikasi berbeda dengan hasil Hermansyah. Nilai pIC₅₀ tertinggi dari penelitian ini adalah 8,44, sedangkan Hermansyah mencapai 9,21. Namun, validitas model dapat dipertanggungjawabkan dengan hasil analisis SHAP yang menunjukkan peran substruktur farmakofor obat dalam keputusan model.

.....Diabetes is one of the biggest causes of death in the world. DPP-4 inhibitor is a drug to treat diabetes. DPP-4 inhibitors can potentially be long-acting diabetes drugs, solving diabetes management problems related to patient non-compliance with taking medication. However, there are no long-acting DPP-4 inhibitors with satisfactory results. DPP-4 inhibitors also still have problems in the form of side effects of pancreatitis and joint pain. Therefore, this study aims to find candidate DPP-4 inhibitors. Virtual screening was performed with machine learning QSAR model trained with DPP-4 activity data using PyCaret library. Three models were obtained: binary and multiclass classification models to determine molecular activity and regression models to predict activity levels. The model was evaluated by comparing the quality and results with Hermansyah et al's research. The results showed that binary classification model had better quality, while regression model was slightly lower. This difference is caused by variations in data and approaches to using the model. Ten compounds identified were different from Hermansyah's results. The highest pIC₅₀ value from this study was 8.44, while Hermansyah reached 9.21. However, validity of the model is justified by SHAP analysis results which shows the role of drug pharmacophore substructure in model decisions.