

Pengaruh Tipe-Tipe Deskriptor Pada Pembelajaran Mesin Terhadap Akurasi Prediksi Docking Score Senyawa-Senyawa Flavonoid Terhadap Virus Dengue-2 = Effect of Molecular Descriptors in Machine Learning on Prediction Accuracy of Docking Score in Flavonoid Compounds on Dengue Virus-2

Alifian Atras Timur, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920525836&lokasi=lokal>

Abstrak

Virus Dengue merupakan virus endemik yang telah ada sejak abad ke 15. Virus Dengue memiliki 4 varian, dan keempat varian tersebut beredar secara bebas di Indonesia. Kasus demam berdarah di Indonesia telah mencapai 131.265 kasus pada tahun 2022. Hingga saat ini, belum ditemukan obat ataupun vaksin yang efektif dalam mencegah persebaran dan mengobati penyakit demam berdarah. Senyawa flavonoid diketahui memiliki sifat antioksidan, antikanker, dan sifat inhibisi virus. Salah satu virus yang berpotensi untuk di inhibisi adalah virus Dengue. Namun, dengan ribuan senyawa yang termasuk ke dalam senyawa flavonoid, dibutuhkan waktu lama untuk melakukan pengujian baik secara *in vitro* ataupun *in vivo*. Salah satu cara untuk mempercepat proses penemuan senyawa inhibitor potensial adalah metode *in silico*, dengan metode yang umum digunakan adalah penambatan molekuler. Namun, proses penambatan molekuler juga masih tetap membutuhkan waktu yang lama. Pada penelitian ini, dilakukan prediksi skor penambatan menggunakan metode pembelajaran mesin (*machine learning*). Data skor penambatan didapat dengan penambatan molekuler menggunakan aplikasi Autodock Vina. Dikembangkan empat model pembelajaran mesin, yakni K-Nearest Neighbor, Xtra Trees, Xtreme Gradient Boosting, dan Artificial Neural Network. Agar komputer dapat mengenali senyawa flavonoid, digunakan deskriptor AlvaDesc dan SMILES. Dari model yang dikembangkan, diambil hasil berupa akurasi dan waktu pelatihan model, dan kemudian dilakukan perbandingan dari setiap model yang dikembangkan beserta deskriptor yang digunakan. Hasil penelitian menunjukkan deskriptor AlvaDesc memiliki akurasi yang lebih tinggi dan waktu pengembangan yang lebih singkat dibandingkan dengan deskriptor SMILES. Model Xtreme Gradient Boosting dengan deskriptor AlvaDesc memiliki performa terbaik dengan akurasi 88,06% dan waktu pelatihan selama 2 menit 40 detik. Model Xtreme Gradient Boosting dengan deskriptor AlvaDesc juga dapat memprediksi skor penambatan senyawa propolis dengan rata-rata perbedaan hasil prediksi dengan nilai aktual sebesar 0,3517428866.

.....

Dengue virus is an endemic virus that exist since the 15th century. Dengue Virus has 4 variants, and all of them exist in Indonesia. Dengue virus case reached over 131,265 cases in 2022 alone. Until now, there are no proven and effective drug or vaccine that can prevent the spread and infection of Dengue virus. Flavonoid compounds are known to have antioxidant and anticancer properties. In some cases, Flavonoid is also exhibit antiviral properties against some virus, one of them being Dengue virus. However, with thousands of compounds classified as flavonoid, it will take significant amount of time to discover and develop the drug for Dengue virus, both with *in vitro* and *in vivo* method. Alternative method to speed up the drug discovery process is to utilize *in silico* or simulation method. One of the method is to use molecular docking. Nevertheless, this method still take considerable amount of time to find the suitable compound. In

this research, we will examine the usage of machine learning to predict the docking score of flavonoid compound. Docking score data are retrieved from the molecular docking with Autodock Vina application. In this research, four models are developed, K-Nearest Neighbor, Xtra Trees, Xtreme Gradient Boosting, and Artificial Neural Network models. The molecular descriptors that used in this research are AlvaDesc and SMILES. From the developed model, the accuracy and training time of the model will be collected and will be analyzed in relation with their model and descriptor that is used. Result shows that AlvaDesc more suitable to be used as descriptor, because of the higher model accuracy and lower training time compared to the SMILES model. XGB Regressor model with AlvaDesc descriptor shows the best performance with accuracy of 88.02% and training time of 2 minutes and 40 seconds. The XGB Regressor model is also able to accurately predict the docking score of Propolis compounds with the mean value of the difference between the predicted and actual result value of 0,3517428866.