

# Analisis Hubungan Kuantitatif Struktur Aktivitas (QSAR) menggunakan Rotation Forest pada Desain Obat Diabetes Mellitus Tipe II = Quantitative Structure Activity Relationship (QSAR) Analysis using Rotation Forest in Drug Design of Diabetes Mellitus type II

Nadya Asanul Husna, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20509443&lokasi=lokal>

---

## Abstrak

Inhibitor DPP-4 adalah pendekatan baru yang menjanjikan untuk pengobatan diabetes tipe-2 dengan risiko rendah hipoglikemia. Pemodelan hubungan kuantitatif struktur-aktivitas (QSAR) adalah pemodelan yang digunakan untuk menyaring basis data besar suatu senyawa untuk menentukan sifat biologis molekul kimia berdasarkan struktur kimianya. Pada tesis ini pemodelan QSAR yang digunakan adalah QSAR klasifikasi dan QSAR regresi. Sebelum membuat model QSAR akan melakukan ekstraksi ciri pada struktur molekul (SMILES). Hasil ekstraksi ciri tersebut kemudian akan digunakan sebagai masukan untuk metode *rotation forest* kasus klasifikasi dan kasus regresi. Model QSAR klasifikasi akan memprediksi molekul aktif dan tidak aktif pada inhibitor DPP-IV. Sedangkan model QSAR regresi akan memprediksi nilai aktivitas IC50 inhibitor DPP-IV. Pada penelitian ini untuk kasus klasifikasi dan regresi juga membandingkan performa model *rotation forest* menggunakan matriks rotasi PCA dengan *rotation forest* menggunakan matriks rotasi Sparse PCA.

Hasil penelitian ini menunjukkan bahwa model QSAR regresi menggunakan *rotation forest* dengan matriks rotasi PCA (RFR(PCA)) memperoleh koefisien korelasi kuadrat 29.2% dengan RMSE 45%. Sementara itu, menggunakan *rotation forest* dengan matriks rotasi Sparse PCA (RFR(SPCA)) memperoleh koefisien korelasi kuadrat 27.1% dengan RMSE 45.6%. Pada QSAR klasifikasi persentase banyaknya molekul yang aktif sangat besar dibandingkan yang molekul tidak aktif, hal ini dapat menyebabkan nilai evaluasi berbeda. SMOTE (*Synthetic Minority Oversampling Technique*) merupakan salah satu metode untuk menangani data tidak seimbang tersebut dengan cara membangkitkan data buatan. Hasil penelitian ini menunjukkan bahwa model QSAR klasifikasi menggunakan *rotation forest* dengan matriks rotasi PCA (RFC(PCA)) memperoleh performa tertinggi dalam memprediksi molekul aktif dan tidak aktif, yaitu nilai MCC 77.7% dengan nilai akurasi sebesar 89%, sensitivitas 89.6%, dan spesifisitas 88.1%. Sementara itu, model QSAR klasifikasi menggunakan *rotation forest* dengan matriks rotasi SPCA (RFC(SPCA)) memperoleh performa tertinggi, yaitu nilai MCC 80.9% dengan nilai akurasi sebesar 90.5%, sensitivitas 90.8%, dan spesifisitas 90.2%.

.....DPP-4 inhibitors are a new approach for the treatment of type 2 diabetes with a low risk of hypoglycemia. The Quantitative Structure-Activity Relationship (QSAR) model is a model used to filter large databases of compounds to determine the biological properties of chemical molecules based on their chemical structure. The QSAR modeling that is used in this research is QSAR classification and QSAR regression. Before creating the model, QSAR will perform feature extraction on the molecular structure (SMILES). The results of the feature extraction will be used as inputs for the rotation forest method of the classification and regression cases. The QSAR classification model predicts active and inactive molecules in DPP-IV inhibitors, while the regression QSAR model predicts the value of IC50 DPP-IV

inhibitor activity. In this study, the classification and regression cases are also comparing the performances between the rotation forest model using the PCA rotation matrix and the rotation forest model using the Sparse PCA rotation matrix.

The results of this study indicate that the QSAR regression model using rotation forest with the rotation matrix PCA (RFR (PCA)) obtained a squared correlation coefficient of 29.2% with RMSE 45%. Meanwhile, using rotation forest regression with the Sparse PCA (RFR (SPCA)) rotation matrix obtained a quadratic correlation coefficient of 27.1% with RMSE 45.6%. In the QSAR classification, the percentage of active molecules is very large compared to inactive molecules, this can cause different evaluation values. SMOTE (Synthetic Minority Oversampling Technique) is one method for handling such unbalanced data by generating artificial data. The results of this study indicate that the classification QSAR model using rotation forest classification with PCA (RFC (PCA)) rotation matrix obtained the highest performance in predicting active and inactive molecules as follows: MCC value of 77.7% with an accuracy value of 89%, sensitivity value of 89.6% and specificity value of 88.1%. Meanwhile, the QSAR classification model using rotation forest classification with the SPCA rotation matrix (RFC (SPCA)) obtained the highest performance as follows: MCC value of 80.9% with an accuracy value of 90.5%, sensitivity value of 90.8%, and specificity value of 90.2%.