

Application of Ensemble Method on QSAR Study of PTP1B Inhibitor as Anti-diabetes Mellitus

Muhamad Al Rafi Bhadraka, Azka Khoirunnisa, Isman Kurniawan

School of Computing, Telkom University, Bandung
malrafi@student.telkomuniversity.ac.id, azkakhoi@telkomuniversity.ac.id, ismankrn@telkomuniversity.ac.id

Abstract

Diabetes Mellitus (DM) is one of the metabolic diseases whose symptom is abnormally elevated blood glucose levels. Diabetes is classified into several types, which are type 1, type 2, Maturity-Onset Diabetes of the Young (MODY), gestational diabetes, neonatal diabetes, and secondary causes. Some of the causes are endocrinopathies, steroid use, and other factors. The International Diabetes Federation (IDF) stated that the prevalence of diabetes in Indonesia was 4.8 percent in 2012. One solution to this problem is to create a QSAR model using the Ensemble Method for diabetes drug discovery. The quantitative structure activity relationship (QSAR) is a way to develop a computational or mathematical model that utilizes a chemometric technique to try to find a statistically significant correlation between structure and function. This research uses the ensemble method which aims to compare the performance between Random Forest, AdaBoost, and XGBoost. The ensemble method successfully predicts the activity of PTP1B inhibitor. Feature selection and parameter tuning is performed and resulting in XGBoost with 90% features used as the best algorithm. The accuracy, recall, precision, and F1-score of XGBoost are 94.84%, 94.74%, 94.86%, and 95.27%, respectively.

Keywords: QSAR, Diabetes, PTP1B, Ensemble Method

