Implementasi Metode Ensemble pada Klasifikasi Inhibitor CDK2 sebagai Agen Anti Kanker

Mela Mai Anggraini, Isman Kurniawan, Erwin Budi Setiawan

Fakultas Informatika, Universitas Telkom, Bandung
melamaianggraini@students.telkomuniversity.ac.id, ismankrn@telkomuniversity.ac.id,
erwinbudisetiawan@telkomuniversity.co.id

Abstract

Chemotherapy treatment has side effects and cell resistance to certain drugs. Therefore, new drugs are needed that can reduce side effects and provide a better treatment effect. In general, anti-cancer drugs developed by discussing Cyclin Dependent Kinase 2 (CDK2) as a target. Conventional drug designs have proven to be ineffective and inefficient for obtaining new compositions with biological activity that need to be synthesized beforehand to determine their activity. To design new drugs, we need a model that can predict the activity of drug candidates. This prediction process can be done using a mathematical model that can determine the structure and activity known as *Quantitative Structure-Activity Relationships* (QSAR). *XGBoost, Random Forest and Adaboost*. The study was conducted by applying the ensemble method on each *fingerprint (Estate, Extended, Maccs* and *Pubchem*). Based on the results obtained, *Random Forest* with *Pubchem fingerprint* is the best prediction model compared to other methods. This discusses the highest *Matthews Correlation Coefficient* (MCC) dan *Area Under the ROC Curve* (AUC) values, which are followed, MCC = 0.979 and AUC = 0.999

Keywords: QSAR, CDK2, *XGBoost, Random Forest, Adaboost*