TS1.6

Prediction of Organic Dyes Absorption Wavelength Using Different Machine Learning Boosting Models

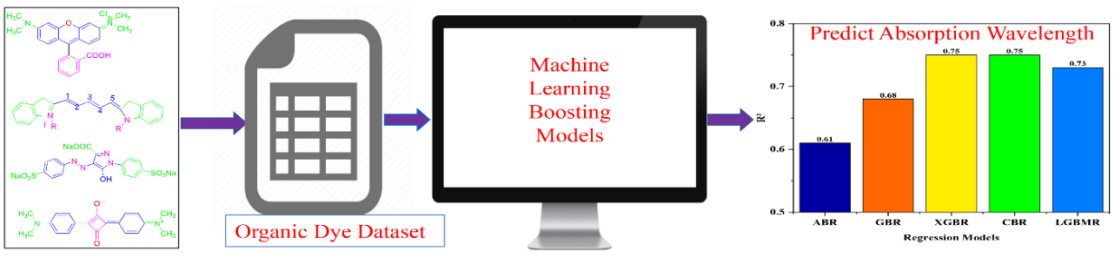
Kapil Dev Mahato[1*], S S Gourab Kumar Das[1], Chandrashekhar Azad[2], Uday Kumar[1*]

[1]Department of Physics, National Institute of Technology Jamshedpur, Jamshedpur 831014, India.
[2]Department of Computer Science & Engineering, National Institute of Technology Jamshedpur, Jamshedpur 831014, India.

* Corresponding Authors' Email IDs: 2018rsphy005@nitjsr.ac.in, uday.phy@nitjsr.ac.in

ABSTRACT

Fluorescent dye molecules have numerous applications in pharmaceutical industries, R & D, bioimaging, fluorescence imaging, light harvesting, drug delivery, and others, and several attempts have been made to develop new fluorescent dyes with desirable photophysical properties. The absorption wavelength is one of the most important photophysical properties of fluorescent dyes. The determination of the absorption properties of new fluorescent organic dyes at a low cost of time and money is a difficult task for experimentalists. For this purpose, various Machine Learning (ML) boosting regression models are used for estimating photophysical properties (particularly absorption wavelength) and may be an alternate approach to density functional theory or Time-Dependent density functional theory. For predicting the absorption wavelength, we examined 9% of the test size data of a given dataset of 3073 organic dyes using five different ML-based boosting regression models, such as AdaBoost Regression (ABR), Gradient Boosting Regression (GBR), XGBoost Regression (XGBR), CatBoost Regression (CBR), and LightGBM Regression (LGBMR). Before beginning the work, the chemical structures were converted into continuous values by their molecule weights using the RDKit library. Then, the proposed models were evaluated using three evaluation parameters: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). R^2 values were 0.61, 0.68, 0.75, 0.75, and 0.73 for ABR, GBR, XGBR, CBR, and LGBMR, respectively. XGBR was the best-performing model across all of these implemented models in terms of the three assessment parameters of RMSE-29.83, MAE-21.26, and R^2-0.75. The proposed boosting models can predict the absorption wavelength, which benefits scientists and industrialists by producing accurate drug designs and new organic dyes for large-scale manufacturing and material assessment in a short time.



Keywords: Machine Learning, Fluorescent organic dyes, Photophysical properties, Absorption wavelength, Density functional theory, XGBoost Regression

DOI: 10.21467/abstracts.164 ISBN: 978-81-965621-3-7 (eBook)