Joint Conference on
2nd International Conference of
AI and Data Science
October 26-27, 2022, Dubai, UAE

Predicting the Scalar Coupling Constants Between Atom Pairs in a Molecule Through Machine Learning Applications

[1]Sean Kim
[1]Fort Lee High School, New Jersey, USA

Abstract

In this paper, we used data on various molecular characteristics from the CHAMPS Kaggle competition (CHAMPS, 2019) to build a prediction model based on various machine learning approaches. We built models on the training set (N = 4,659,076 observations) and then used the best performing one to obtain and evaluate predictions on the testing set (N = 2,505,190 observations). We evaluated the performance of three models – linear regression, XGBoost, and Neural Net – on three metrics: R-squared, MAE, and RMSE. The XGBoost model resulted in a superior fit over Neural Nets and linear regression, with RMSE as lower as 2.75 on the test dataset. This result suggests that XGBoost is a viable approach for predicting the scalar coupling constant.

Biography

Sean is a senior at Fort Lee High School in New Jersey, USA. Sean is passionate about computer science and hopes to conduct more research in the field.