

7th Global Conference on

DATA SCIENCE AND MACHINE LEARNING

December 13, 2022 | Webinar

Received date:21-09-2022 | Accepted date:24-09-2022 | Published date: 20-12-2022

Predicting the scalar coupling constants between atom pairs in a molecule through machine learning applications

Sean Kim

Fort Lee High School, USA

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.

Keywords: Scalar coupling constants, atom pairs, machine learning, data science, prediction model

Recent Publications

1. Kim, S. (2022) "A Token-Based Voting System for Rating News Sources" presented and published at The Artificial Intelligence & Robotics Conference in Osaka, Japan.

Biography

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

e: seankim7038@gmail.com