Predicting biodiesel properties and its optimal fatty acid profile via explainable machine learning

ABSTRACT

The accurate prediction of biodiesel fuel properties and determination of its optimal fatty acid (FA) profiles is a non-trivial process. To this aim, machine learning (ML) based predictive models were developed for cetane number (CN) and cold filter plugging point (CFPP), where the extreme gradient boost (XGB) and random forest (RF) algorithms had the best performance with R2 of 0.89 and 0.91 on the test data, respectively. A classifier model for oxidative stability (OS) was devised to predict if it would pass or fail the ASTM and EU limits, where the support vector classifier (SVC) had the highest accuracy of 0.93 and 0.77 for ASTM and EU limits. Causal analysis via Shapley and Accumulated Local Effects revealed the significance and correlation of FAs with the fuel properties. This eventually aided the determination of the optimal FA composition via evolutionary optimization, such that the properties would meet the ASTM and EU standards. This study presents an end-to-end ML framework including descriptive, predictive, causal and prescriptive analytics to predict biodiesel fuel properties as a function of its FA composition; and eventually prescribes the optimal FA composition necessary to ensure that the fuel properties meet the regulatory standards.