Predictive Modelling of Biodiesel Production from Waste through Transesterification

Rutvik Kumar[a], Rejeti Venkata Srinadh[a], Neelancherry Remya[a]

[a]Indian Institute of Technology Bhubaneswar, Odisha, India – 752050

Corresponding author's email: remya@iitbbs.ac.in
Corresponding author's Ph. No.: +91-8280549054

ABSTRACT

Biodiesel production and use have been the topic of extensive research due to the ever-increasing emphasis on the creation of eco-friendly and low carbon footprint energy alternatives. Several process factors drive the biodiesel production process, which must be kept at optimal levels to ensure maximum productivity. Because biodiesel productivity and quality are also affected by the numerous raw materials used in transesterification, practical studies are required before making any assumptions about them. However, due to the high number of process parameters and the underlying non-linear relationship between the process parameters and responses, a brute force strategy of carrying out physical trials until the optimal process parameters are attained will fail. In this context, this research employs a machine learning-based prediction approach to quantify the response properties of the biodiesel manufacturing process as a function of process parameters. This paper investigates four powerful machine learning algorithms in depth: linear regression, random forest regression, AdaBoost regression and artificial neural network. Both random forest regression and AdaBoost regression show good accuracy in prediction modelling of biodiesel yield. However, random forest may be a better strategy for modelling biodiesel production because it gets the lowest error among the studied algorithms. Furthermore, random forest can be deployed more quickly because it was found to be insensitive to the amount of regressors used.

Keywords: Biodiesel; Machine learning; Linear regression; Random Forest regression; AdaBoost regression; Artificial neural network

1. Introduction

Extended Biodiesel, a sustainable and environmentally friendly fuel, is gaining prominence due to concerns over dwindling petroleum reserves and global warming. It is produced through transesterification, a reaction that converts waste oils with alcohols like methanol or ethanol. Biodiesel is non-toxic, biodegradable, and lacks the aromatic compounds and sulfur found in traditional fuels. Researchers are increasingly exploring the use of vegetable oils for biodiesel, which is less polluting than petroleum-based diesel [1]. This shift towards biofuels can benefit both the environment and local job markets, encouraging the cultivation of biomass sources like Calophyllum inophyllum in Asia or soybeans in Brazil and the United States.

However, the efficiency of biodiesel production depends on factors like reaction time, temperature, oil/alcohol ratio, and catalyst yield. To optimize these variables, researchers employ response surface methodology (RSM) and diverse designs of experiments (DoE), such as Taguchi, factorial, central composite, and box-benhken designs [2]. These statistical tools help predict how each factor and their interactions affect biodiesel yields [3]. The goal is to develop a reliable and sustainable alternative to fossil fuels without compromising on performance or environmental impact.

2. Materials and Methods

Predictive modelling of biodiesel production will require the use of machine learning algorithms and the process involved for experimentation. We have covered the method involved in the production of biodiesel. It has described the transesterification process, the different catalysts involved and the type of alcohol that can be used. This also provides the information about the basic machine learning algorithms like linear regression and some advanced algorithms like random forest, adaptive boosting regression and artificial neural network also supplies a detail overview of how this algorithms work.

Four machine learning algorithms are used in this work to predict biodiesel production process responses based on process parameters such as temperature, reaction time, catalyst weight and methanol/oil ratio: Linear Regression, Random Forest Regression, AdaBoost regression and Artificial Neural Network. Two separate examples from the literature are presented to illustrate the method and carry out a robust study. The biodiesel production yield is estimated in the first example based on the process parameters. Several error metrics are used to compare the machine learning regression predictive models.

3. Results and Discussion

Evaluating the performance of different experimentation with model is a must to get the best one for prediction. This includes the information about how much these parameters are correlated with each other and the performance of different machine learning algorithms using these process parameters for predicting the biodiesel yield.

Random forest regression model is used to get the best values of the methanol/oil molar ratio (Mr), catalyst weight (Wc), temperature (T) and reaction time (Tr) for maximum yield (%). Table 1 shows the 4 values of yield (%) which are maximum and obtained from the predictive model on various values of input parameters. The maximum value of yield (%) in the dataset was in the range of 95% to 96% while from the predictive model the highest value of yield (%) is around 99.5%. It has been seen that by keeping the methanol/oil molar ratio (Mr) between 6 and 8, the catalyst weight (Wc) in the range 0.7 to 0.8, temperature (T) between 45°C and 55°C and reaction time (Tr) between 40 min and 50 min while get the maximum values of yield (%).

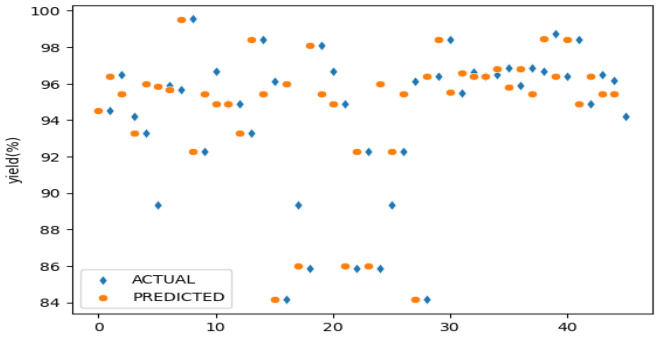

Figure 1. Predictive performance of the random forest regression predictive model

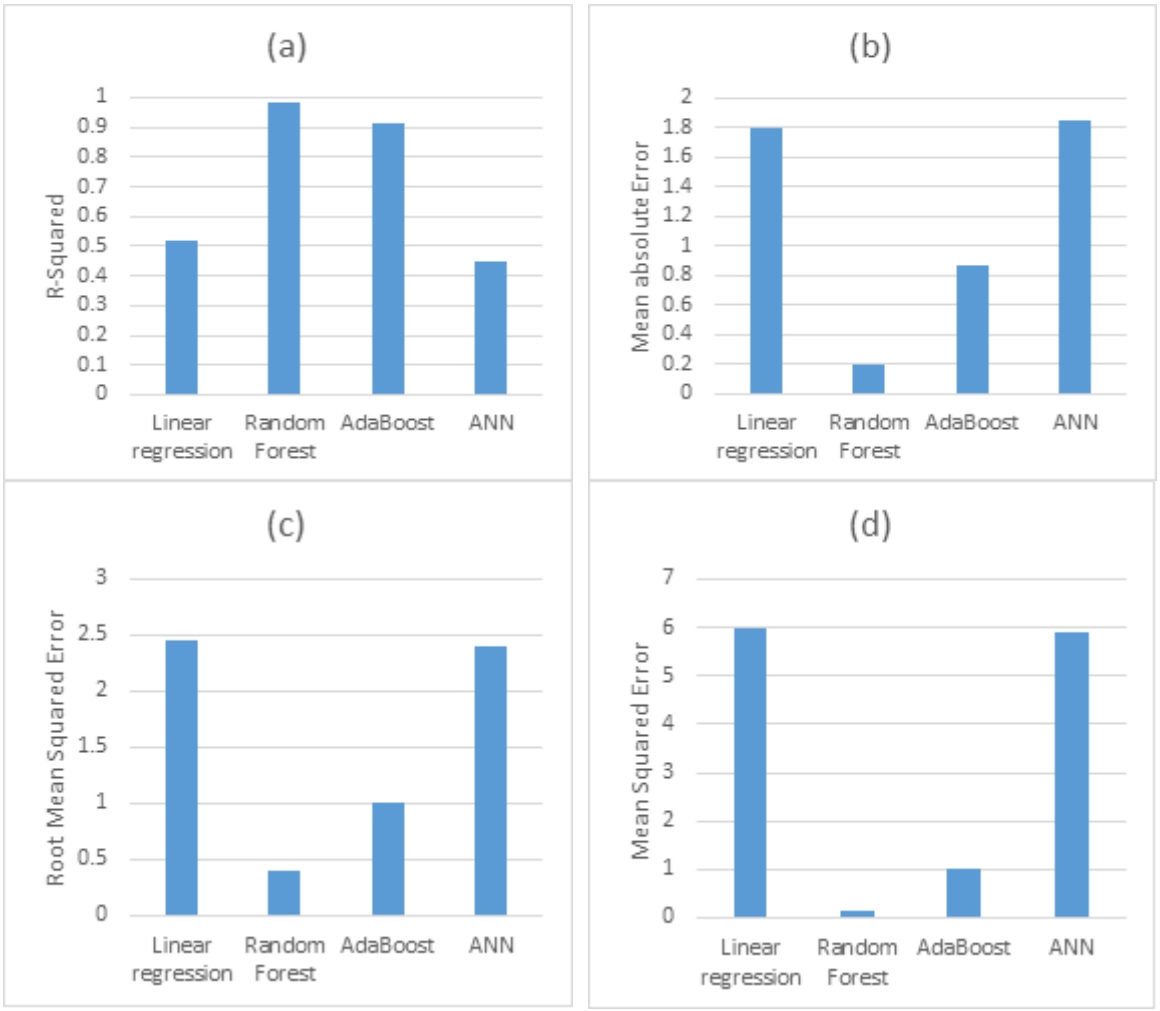

Figure 2 Performance of the models (a) R2; (b) MAE; (c) RMSE; (d) MSE.

R2	MAE	RMSE	MSE
0.971	0.415	0.641	0.411

Table 1. Validation results from the random forest

Catalyst	Methanol/ Oil Molar Ratio (M_r)	Catalyst Weight (W_c)	Reaction Temperature (T)	Reaction Time (T_r)	Yield (%)
	7	0.8	50	50	99.51
KOH	**7**	0.8	60	40	98.84
	9	0.4	35	50	97.13
	6	0.4	25	60	96.86

Table 2. Optimization results from the random forest

4. Conclusion

Predictive models for biodiesel yield based on process parameters like catalyst weight, temperature, and reaction time were tested using four machine learning techniques. Linear regression and artificial neural networks showed moderate accuracy, while random forest and AdaBoost regression performed extremely well. Random forest predicted a high yield of 99.51% with specific process parameters, making it a preferred choice for efficient biodiesel production optimization.

5. Acknowledgement

I express my sincere gratitude towards my guide Dr. Remya Neelancherry for guiding me through the project. Her consistent guidance helped me in understanding and effective execution of the work. I would like to acknowledge her support, timely advice, and suggestions regarding the project, which immensely helped me complete this project.

6. References

1.Bisheswar Karmakar, Sumit H.Dhawane, Gopinath Halder (2018), "Optimization of biodiesel production from castor oil by Taguchi design" ScienceDirect S2213343718301957

2.Monzur Alam Imteaz, A.B.M. Sharif Hossain, Maryam Bayatvarkeshi (2022), "A mathematical modelling framework for quantifying production of biofuel from waste banana"

3.Min-Yuan Cheng, Doddy Prayogo(2012), "Optimizing Biodiesel Production from Rice Bran Using Artificial Intelligence Approaches"

4.Rouf Ahmad Dar, Mudasir Yaqoob, Manisha Parmar, Urmila Gupta Phutela (2019), "Biofuels from Food Processing Wastes"