Data Science for Predicting and Optimizing Syngas Production in Fluidized Bed Biomass Gasification

October 30, 2023

Deadline for Case Study Submission: (Week 28) Friday, 5-Jan-2024 before 11:59:59 on FASER

1 Description

Due to escalating energy demand driven by economic and population growth, and a heightened focus on environmental concerns, there is an increasing emphasis on transitioning from fossil fuels to renewable alternatives. Biomass, an abundant natural resource, shows significant promise as an alternative to traditional fossil fuels. Biomass gasification, a prominent thermal conversion process, frequently employs fluidized bed reactors to generate syngas characterized by relatively low heating values [1]. Yet, despite the potential, there is currently no effective model available for predicting gasification yield across a broad spectrum of applications. In this case study, the objective is to employ machine learning techniques to predict syngas compositions and lower heating values (LHV) utilizing diverse lignocellulosic biomass feedstocks under a wide range of operating conditions. In essence, this research seeks to uncover the intricate and nonlinear correlations between input parameters—namely lignocellulosic compositions and operating conditions—and output variables, including syngas compositions and properties of the resulting byproducts. To achieve this, efficient machine learning models such as Random Forest and Support Vector Machine (SVM) were utilized [2].

For evaluating the models, key performance metrics such as regression coefficient (R2) and root mean square error (RMSE) will be useful to provide an accurate assessment of their predictive capabilities. Through this study, we aim to shed light on the potential of machine learning in optimizing biomass gasification processes, a critical step towards a sustainable and environmentally conscious energy landscape.

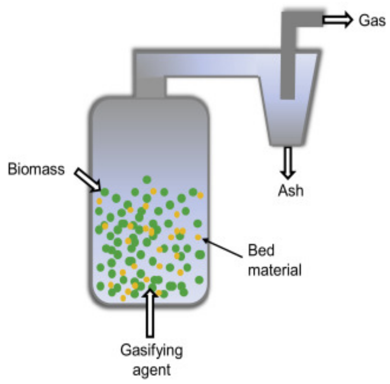

Figure 1: Fluidized Bed Gasifier

Attribute Information: The biomass gasifiers dataset involves predicting multiple gasses [1]. The dataset contains a total of 336 records, where 8 input columns contain Lignocellulose composition, temp, pressure, equivalence mass ratio, steam to biomass mass ratio, and Superficial gas velocity. The output variables are Syngas composition, lower heating value, char yield, and tar yield.

In our dataset, the input features (i.e. variables) and output are illustrated in Figure 2.

2 Dataset availability

Your dataset is available in the text format and to be downloaded Click me download.

Variables		Ranges
Input variables		
Lignocellulose composition [wt.%]	Cellulose (*Cell.*)	0.20–0.58
	Hemicellulose (*Hem.*)	0.08–0.63
	Lignin	0.10–0.49
Temperature (T) [$^{\circ}$C]		600–900
Pressure (P) [abar]		1–10
Equivalence ratio (ER) [−]		0–0.86
Steam to biomass ratio (SBR) [−]		0–8.03
Superficial gas velocity (U_g) [m/s]		0.02–9.59
Output variables		
Syngas composition [vol%]	H_2	5.39–66.03
	CO	5.01–55.44
	CO_2	6.78–62.56
	CH_4	1.31–20.1
Lower heating value (LHV) [MJ/Nm^3]		1.74–15.0
Char yield [wt.%]		0–45
Tar yield [g/Nm^3]		0–134.1

Figure 2: Input and output variables of the constructed database.

3 Aim

Your job is to create a multiple input and multiple output (MIMO) regression model[3]. Please don't forget to split your dataset into training and test sets and make sure your test set should be at least 20%.

4 Help

- Try different regression algorithms such as SVM, Random Forest, KNN, XGBoost, Neural Networks, and more.
- Split data into training, validation, and test sets to evaluate your model.
- Feel free to use any machine learning model. However, make sure you can justify that the approach is leading to better prediction.
- Please check which evaluation metrics to select for your problem. Selecting the wrong evaluation metrics can deduct your marks.
- Exploratory data analysis is quite important and please use pandas.
- Discuss with CE880 academic staff (Dr Haider Raza).

References

[1] J. Y. Kim, D. Kim, Z. J. Li, C. Dariva, Y. Cao, and N. Ellis, "Predicting and optimizing syngas production from fluidized bed biomass gasifiers: A machine learning approach," *Energy*, vol. 263, p. 125900, 2023.

[2] J. Y. Kim, U. H. Shin, and K. Kim, "Predicting biomass composition and operating conditions in fluidized bed biomass gasifiers: An automated machine learning approach combined with cooperative game theory," *Energy*, p. 128138, 2023.

[3] machinelearningmastery.com, "How to develop multi-output regression models with python," https://machinelearningmastery.com/multi-output-regression-models-with-python/, accessed: 27.04.2021.