Enhancing Hydrogen Consumption Predictions in Refinery Hydrotreating through Machine Learning Techniques

Domilies, Julongbayan, Laag

College of Engineering
University of the Philippines - Diliman

June 15, 2024

Outline

1. Introduction
2. Methodology
3. Results and Discussion
4. Conclusion

Introduction

Introduction

Refineries heavily depend on hydrogen for purifying petroleum products through processes like hydrotreating, where hydrogen plays a crucial role in removing impurities and enhancing product quality. This is particularly critical for meeting stringent emission standards such as Euro IV, which mandate reducing sulfur content to less than 50 parts per million (ppm).

However, accurately predicting hydrogen consumption remains a significant challenge due to its highly complex and non-linear nature, compounded by the dependency on numerous operational parameters.

Hydrotreater Process

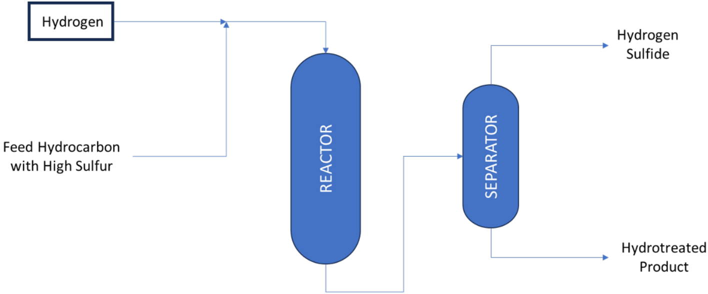

Figure: Hydrotreater Process Overview

Before adjusting the feed going into the reactor, hydrogen supply is first ensured. Thus, optimal hydrogen consumption prediction is crucial. High hydrogen consumption prediction leads to hydrogen waste, while low hydrogen consumption prediction leads to potential operational issues.

Objectives

The objectives of this study are:

- Employ outlier detection techniques to identify and remove outliers from the dataset.
- Utilize dimensionality reduction methods to reduce the number of features while maintaining prediction accuracy.
- Implement regression models to accurately forecast hydrogen consumption in refinery hydrotreating processes.

Methodology

Methodology

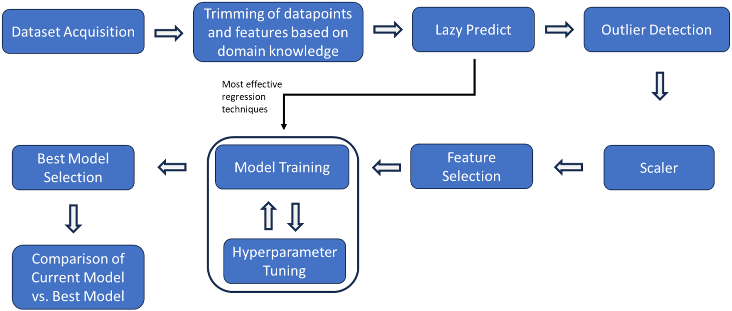

Figure: Overview of Machine Learning Methodology

EDA

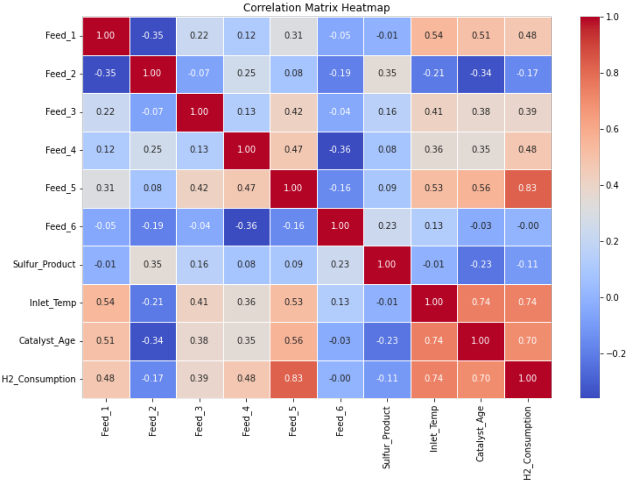

Figure: Feature-Target Correlation Heatmap

Results and Discussion

Lazy Predict Results

Model	Test RMSE	Model	Test RMSE
1. **ExtraTreesRegressor**	**329.39**	11. AdaBoostRegressor	1,052.65
2. **RandomForestRegressor**	**381.5**	12. LassoCV	1,190.74
3. XGBRegressor	390.8	13. Lasso	1,190.76
4. LGBMRegressor	401.34	14. LassoLars	1,190.76
5. HistGradientBoostingRegressor	403.55	15. BayesianRidge	1,190.78
6. BaggingRegressor	405.23	16. Ridge	1,190.80
7. KNeighborsRegressor	436.97	17. RidgeCV	1,190.80
8. ExtraTreeRegressor	531.08	18. LassoLarsCV	1,190.81
9. DecisionTreeRegressor	532.28	19. LassoLarsIC	1,190.81
10. GradientBoostingRegressor	584.67	20. LinearRegression	1,190.81

Random Forest vs. Extra Trees

Aspect	Random Forest	Extra Trees
Sampling Method	Uses bootstrap sampling (bagging).	Typically uses the whole training dataset (or bootstrap if specified).
Node Splitting	Selects the best split from a random subset of features.	Selects random splits from a random subset of features.
Randomness	Introduces randomness through feature subset selection and bootstrap samples.	Introduces more randomness by also selecting random split points.
Training Speed	Can be slower due to the need to find the best split at each node.	Generally faster because it selects splits randomly.
Variance vs. Bias	Typically reduces variance effectively with a possible risk of overfitting.	Tends to reduce variance more but may increase bias.

Results and Discussion

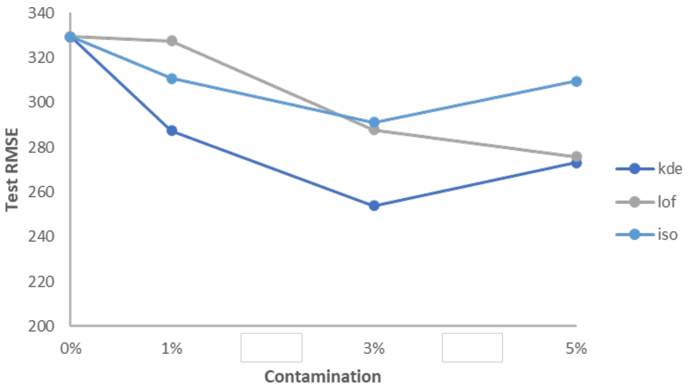

Figure: Effect of Outlier Removal on RSMEs at Various Contamination

Results and Discussion

Method	Outlier Removal	Train RMSE	Test RMSE
Extra Trees	None	430.2	488.6
	KDE	395.7	400.5
Random Forest	None	432.2	510.9
	KDE	385.4	390.9

Table: Comparison of RMSE Scores Before and After Outlier Removal

*Note that models are based on default hyperparameters.

Results and Discussion

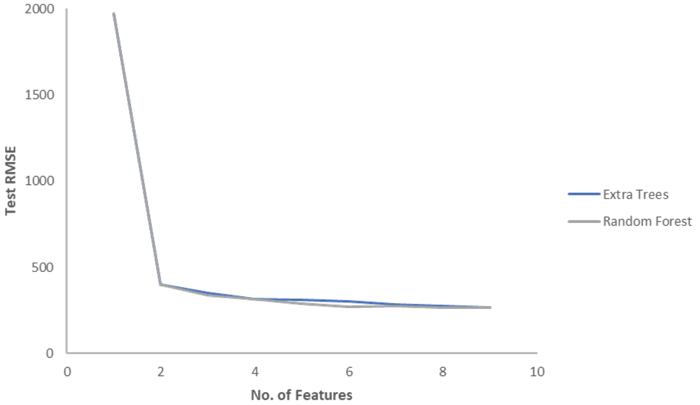

Figure: Feature Selection Using RFE

Results and Discussion

Hyperparameters	Search Values
n_estimators	50 - 150
max_depth	15 - 25
min_samples_leaf	1 - 5
min_samples_split	2 - 10
ccp_alpha	0 - 0.2

Table: Hyperparameters and Search Values

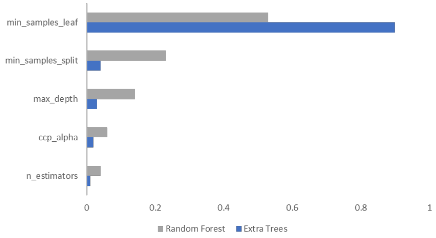

Figure: Hyperparameter Importance

Algorithm	n_estimators	max_depth	min_samples_leaf	min_samples_split	ccp_alpha
Extra Trees	112	20	2	5	0.007
Random Forest	82	15	1	5	0.055

Table: Comparison of Extra Trees and Random Forest Hyperparameters

Table: Model Performance Metrics

Model	Test R^2	Test RMSE	Training Time, sec
Extra Trees	0.988	356.2	517.7
Random Forest	0.986	385.8	1808.8

- The Extra Trees model is selected as the best model due to its superior performance metrics, including higher R^2, lower RMSE, and shorter training time.

Results and Discussion

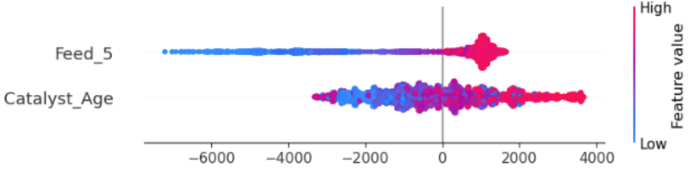

Figure: Feature Importance

Results and Discussion

Model Description	RMSE	R^2
Current Model Performance	2194	0.58
Refitted Linear Regression with Existing Parameters	1552	0.63
Linear Regression with Additional Critical Features	1268	0.82
Optimal Model Identified through Machine Learning	**356**	**0.99**

Table: Comparison of RMSE and R^2 for Different Models

Conclusion

Conclusion

The combination of outlier detection techniques, dimensionality reduction methods, and effective regression modeling has resulted in substantial improvements in the accuracy and robustness of the predictive models for forecasting hydrogen consumption in refinery hydrotreating processes.

- **Outlier Detection:** By removing outliers, the models became more reliable and showed reduced overfitting, as evidenced by the narrower gap between training and test RMSE.
- **Dimensionality Reduction:** The models were simplified by reducing the features from 9 to 2 while maintaining accuracy
- **Regression Models:** The decrease in RMSE to 356 from the initial 2100 demonstrates a significant boost in predictive capability.

Overall, the application of these techniques not only enhanced the precision of the predictions but also streamlined the modeling process, making it more efficient and interpretable for practical applications in refinery operations.

Enhancing Hydrogen Consumption Predictions in Refinery Hydrotreating through Machine Learning Techniques

Domilies, Julongbayan, Laag

College of Engineering
University of the Philippines - Diliman

June 15, 2024