Evaluating performance of Distillation column using different Machine learning algorithm

SATISH PRASAD

224107121

**Abstract:**

The chemical process industry involves various complex processes and requires accurate predictions of yield and energy requirements. In this project, we applied different regression models to predict the energy required for distillation processes. The dataset contained 160 samples with ten features, including feed composition of A and B , temperature, pressure, column height, column diameter, reflux ratio, number of trays, yield, and energy. We used linear regression, kernel regression, Lasso regression, Ridge regression, random forest, decision tree, and K-NN models.

The results show that the K-NN model gave the highest R2 score of 0.8684, indicating a good fit of the model to the data. The random forest and decision tree model also gave a good R2 score of 0.6791 and 0.8396 respectively. The other models performed poorly with negative or close to zero R2 scores. The RMSE values for all the models were within the range of 9,894 to 51,998, indicating moderate to high error rates.

**Introduction:**

The chemical process industry is critical for producing various chemicals, fuels, and materials used in everyday life. Accurate prediction of yield and energy requirements is essential for optimizing the process efficiency, reducing production costs, and minimizing environmental impacts. However, predicting these parameters is challenging due to the complex and nonlinear relationships among various process variables. Therefore, several machine learning algorithms have been developed to model these relationships and predict the yield and energy requirements.

**Methodology:**

We used a dataset with 160 samples and ten features, including feed composition of A and B, temperature, pressure, column height, column diameter, reflux ratio, number of trays, yield, and energy. We first performed data preprocessing, including feature scaling and feature selection. We then applied various regression models, including linear regression, kernel regression, Lasso regression, Ridge regression, random forest, decision tree, and K-NN. We used 10-fold cross-validation for model evaluation and hyperparameter tuning.

**Results:**

The K-NN model gave the highest R2 score of 0.8684, indicating a good fit of the model to the data. Decision tree also have good R2 value of 0.8396 close to K-NN model performace. The random forest model also gave a good R2 score of 0.6791. The other models performed poorly with negative or close to zero R2 scores. The RMSE values for all the models were within the range of 9,894 to 51,998, indicating moderate to high error rates.

**Conclusion:**

In this project, we compared various regression models for predicting energy requirements in the distillation process. The K-NN,decision Tree & random forest models showed good performance, while the other models performed poorly. However, the RMSE values for all the models were relatively high, indicating that further improvements are needed in the model development. Overall, this study provides a valuable insight into the applicability of machine learning algorithms for predicting energy requirements in the distillation process.

**References:**

Zhang, Y., Zheng, Y., Wang, Z., & Zhao, L. (2019). Predicting the yield and quality of gasoline in the fluid catalytic cracking process using a hybrid modeling approach. Industrial & Engineering Chemistry Research, 58(30), 14006-14014.

Kim, J., Kim, J., Park, H., & Ahn, J. (2020). Machine learning-based yield prediction for catalytic cracking process. Journal of Industrial and Engineering Chemistry, 91, 138-145.

Zhang, Y., Chen, X., Jiang, L., & Cheng, G. (2020). A comparison of machine learning algorithms for predicting the conversion rate of the gas-to-liquid process. Processes, 8(11),