TECHNICAL UNIVERSITY OF CLUJ-NAPOCA

ACTA TECHNICA NAPOCENSIS

Series: Applied Mathematics, Mechanics, and Engineering
Vol. 68, Issue Special I, May, 2025

APPLICATION OF MACHINE LEARNING TECHNIQUES FOR MODELING THE REACTOR OF THE CATALYTIC CRACKING PROCESS

Cristina Roxana POPA, Mădălina CĂRBUREANU, Bogdan DOICIN

Abstract: *This paper presents the research and results obtained by authors regarding the modeling of the reactor of the fluid catalytic cracking unit (FCCU), using various machine learning techniques based on supervised learning algorithms Compared with the analytic mathematic modeling methods, machine learning techniques (ML) can solve the difficulties in FCC process modeling, such as strong correlation between variables, nonlinearities, the complexity of kinetic models, and feedstock complexity. The ML method establishes a mathematical correlation between the input and output variables of the process using industrial data. Five supervised learning machine methods are used to model the reactor: random forest regression (RFR), decision trees regression (DTR), k-nearest neighbors (KNN), support vector machines (SVM) and gradient boosting regression (GBR).*
Keywords: *machine learning, modeling, fluid catalytic cracking, regression*

1. INTRODUCTION

In a refinery, one of the most important processes is fluid catalytic cracking, which can convert heavy hydrocarbons from the vacuum distillate process into light hydrocarbons, which constitute the feedstock for petrochemicals and gasoline with a high research octane number (COR). The fluid cracking catalytic contains four components: preheated furnace, riser, stripper, and regenerator as shown in Figure 1 [1, 2].

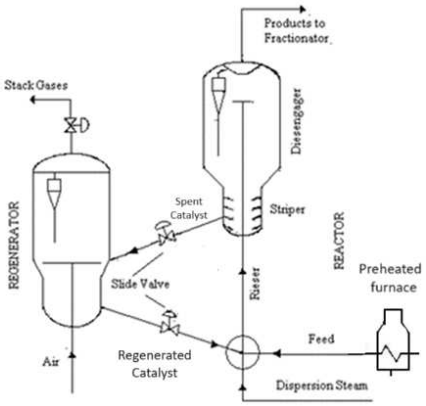

Fig. 1. Process cracking catalytic unit [2].

An important problem studied by researchers is modeling the catalytic cracking process for developing a simulator used to evaluate process control strategies and analyze the behavior of the process under varying operational conditions.

The first model in the literature is the steady-state models, which characterize the kinetic reactions taking place in riser. The starting point of the kinetic models is the 3-lumps model developed by Weekman and his coworkers. This model can be applied to any type of feedstock [3]. Interesting contributions to the basic 3-lump were made in the papers [4].

Starting from this model, other complex kinetic models were developed, based on 4-lumps [5], 5-lumps [6], 7-lump [7], 10 lumps [8], 11 lumps [9]. All kinetic models are integrated into the material and heat balances on the riser.

In recent years, machine learning techniques are a powerful predictive tool to solve complex chemical processes, such as fluid catalytic cracking processes.

In this paper, the authors present the conclusions drawn about the modeling the reactor using machine learning based on different supervised methods. The raw data used in this research are from a simulator of the

catalytic cracking process developed by Popa C. and presented in the paper [2]. The simulator was tested and validated using data from an industrial catalytic cracking process from Romania.

Figure 2 shows an input-output characterization of the reactor. The input variables are feedstock temperature - T_{mp}, regenerated catalyst temperature -T_{reg}, feedstock flow – Q_{mp}, and regenerated catalyst flow -Q_{cat}. The output variables are interfusion nod temperature -$T_{nod.}$, reactor temperature $T_{react,}$ yield gasoline- $Q_{gasoline}$, octane number research-COR [10].

Fig. 2. Input-output characterization of the reactor.

2. MACHINE LEARNING METHODS

Machine learning is a branch of artificial intelligence, that aims to develop programs that by accessing a dataset, can predict the output data based on input data. This prediction is due to the ability of machine learning to learn by themselves automatically but also to the opportunity to become better during the learning process [11,12]. Machine learning algorithms are categorized into supervised machine learning algorithms and unsupervised machine learning algorithms.

The supervised learning algorithms determined a correlation between output data and input data. The models obtained by supervised machine learning are the regression model and classification model. In a regression model, the output variables are predicted with a continuous function that sets a relation between input variables and output variables. In the classification model, the relationship between output variables and input variables is described by a discrete function [13, 14].

In this study, the data are obtained from a simulator of the catalytic cracking process and will be considered experimental data in this research. Therefore, the output variables of the catalytic cracking process (interfusion nod temperature -$T_{nod.}$, reactor temperature $T_{react,}$ yield gasoline- $Q_{gasoline}$, octane number research-COR) are estimated using five different supervised machine learning. These are support vector machines (SVM), random forest regression (RFR), decision trees regression (DTR), gradient boosting regression (GBR), and k-nearest neighbors (KNN). To implement these algorithms, specific functions for each method are from the Scikit-Learns library from Python software.

In the following subsequent, the used algorithms are described, together with a comparison of the experimental data and the estimated output variables. The evaluation of the efficiency of each algorithm was done using specific evolution metrics described by explained variance score (EVS), R squared error (R^2), mean absolute error (MA), mean absolute percentage error (MAPE), and median absolute error (MAE). The equations of performance criteria are presented in paper [13].

2.1. Random forest regression

Random forest regression (RFR) is a supervised ML algorithm that uses an ensemble of decision tree h (x; θ_k), k=$\overline{1,k}$, where x is the observed input (covariate) vector and θ_k are independent and identically distributed random vectors [10]. Instead of predicting a single value, such as in the multiple linear regression (MLR) case, RFR predicts a mean value of the entire set of decision trees, represented by formula (1) [15]:

$$\overline{h(x)} = \frac{1}{K} \cdot \sum_{i=1}^{K} h(x, \theta_k) \quad (1)$$

So, the final prediction using RFR is derived from the majority vote of the individual tree prediction [12].

Table 1 presents the obtained model evaluation metrics for the RFR algorithm.

Table 1

The model evaluation metrics for the RFR algorithm.

Performance criteria / Output variable	EVS	R^2	MA	MAPE	MAE
Tnod	0.74	0.74	8.23	0.014	4.56
Treact	0.82	0.81	5.088	0.010	2.61
$Q_{gasoline}$	0.87	0.87	0.004	0.010	0.002
COR	0.93	0.92	0.064	0.0006	0.037

The output variables predicted with the RFR are compared with experimental output dataset and are shown in Figures 3-6.

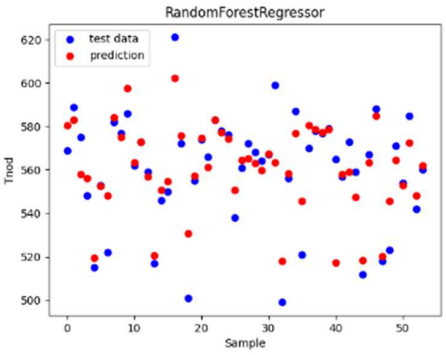

Fig. 3. Comparison between T_{nod_exp} and T_{nod_pred} using RFR.

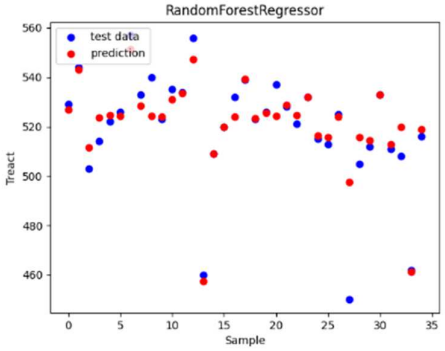

Fig. 4. Comparison between T_{react_exp} and T_{react_pred} using RFR.

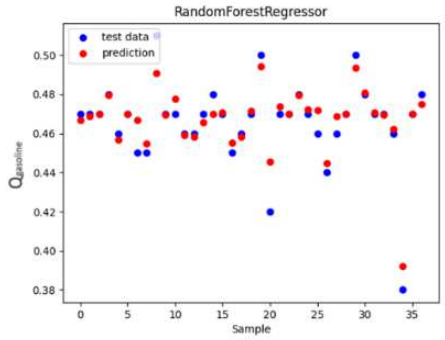

Fig. 5. Comparison between $Q_{gasoline\_exp}$ and $Q_{gasoline\_pred}$ using RFR.

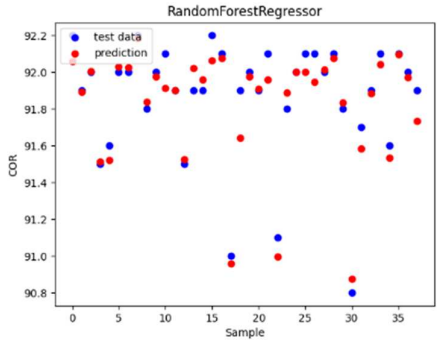

Fig. 6. Comparison between $COR_{\_exp}$ and COR_{pred} using RFR.

2.2. Decision trees regression

Decision trees regression (DTR) is a supervised ML algorithm used for regression, classification, and prediction, the methods that work for both continuous as well as for categorical output variables. It uses a tree structure in which the decision nodes are conditions, while the end nodes represent the obtained results [10]. In the decision-making process, the DT's learn from *if-then* rules, examining all possible tests to identify the one that provides the most relevant information about the target variable [15, 16].

Decision tree regression (DTR) predicts data to produce continuous output by observing an object feature and by training a model under a tree structure form. In the decision-making process, respectively for the prediction task, the DTR algorithm traverses the tree using each node test and finds the node leaf i8 which the new data point falls into [17]. So, a DTR is a decision tree used for regression tasks, respectively for predicting continuous outputs (instead of discrete ones).

The output variables calculated with DTR are compared with experimental output dataset and illustrate in Figures 7-10. Table 2 presents the obtained model evaluation metrics for the DTR algorithm.

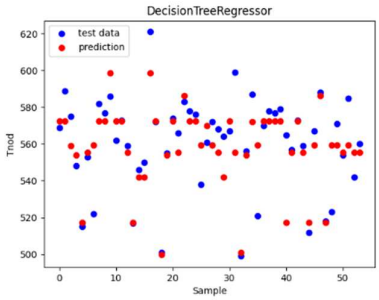

Fig. 7. Comparison between $Tnod_{\_exp}$ and T_{nod_pred} using DT.

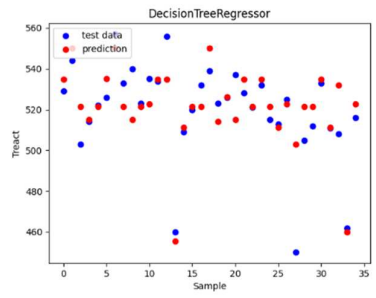

Fig. 8. Comparison between T_{react_exp} and T_{react_pred} using DTR.

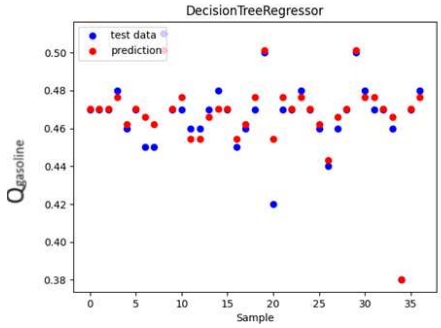

Fig. 9. Comparison between $Q_{gasoline\_exp}$ and $Q_{gasoline\_pred}$ using DTR.

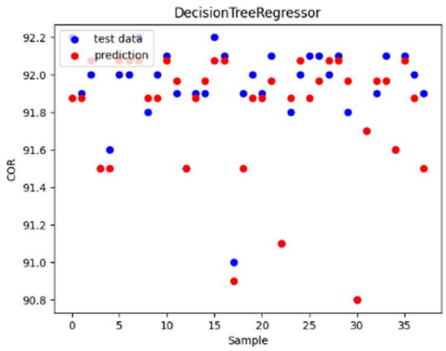

Fig. 10. Comparison between $COR_{\_exp}$ and COR_{pred} using DTR.

Table 2

The model ev□lu□tion metrics for the DTR □lgorithm.

Perform□nce criteri□ / Output v□ri□ble	EVS	R^2	M□	M□PE	M□E
Tnod	0.624	0.6	10.31	0.0183	5.483
Treact	0.646	0.64	9.00	0.0176	6.399
$Q_{gasoline}$	0.879	0.87	0.004	0.009	0.003
COR	0.837	0.81	0.095	0.001	0.077

2.3. The k-ne□rest neighbors' regression

The K-nearest neighbors (KNN) algorithm is a supervised ML used for prediction and classification tasks. When used for regression problems is known as K-Nearest Neighbors Regression (KNR), in which case the prediction result is the average target of the K nearest neighbors [17].

An advantage of using KNR is the fact that it allows the programmers to understand and then interpret what happens inside a model, being very fast to implement [18].

The output variables predicted using KNR are compared with experimental output dataset and illustrated in Figures 11-14.

Table 3 presents the obtained model evaluation metrics for the KNR algorithm.

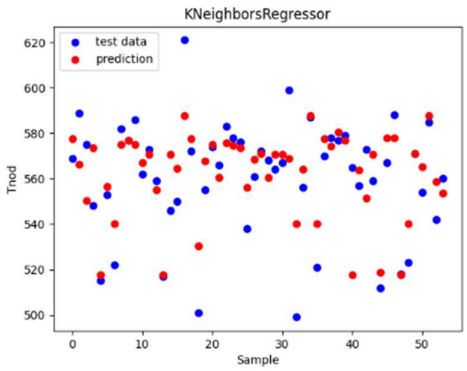

Fig. 11. Comparison between T_{nod_exp} and T_{nod_pred} using KNR.

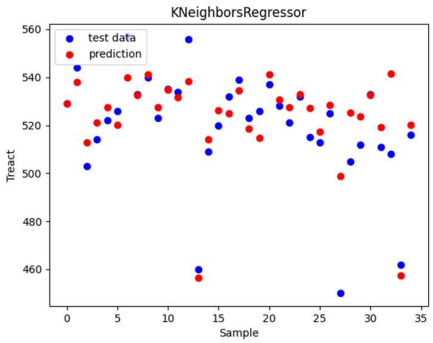

Fig. 12. Comparison between T_{react_exp} and T_{react_pred} using KNR.

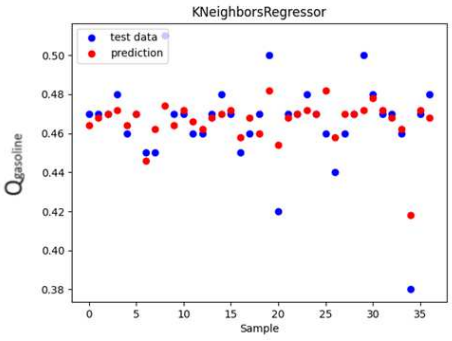

Fig. 13. Comparison between $Q_{gasoline\_exp}$ and $Q_{gasoline\_pred}$ using KNR.

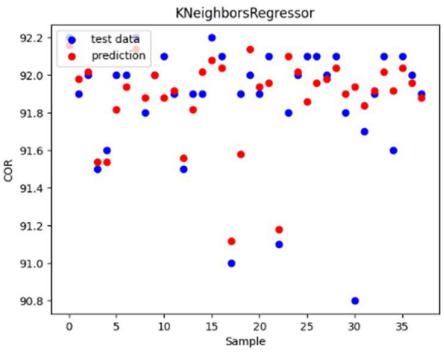

Fig. 14. Comparison between $COR_{\_exp}$ and COR_{pred} using KNR.

Table 3

The model ev□lu□tion metrics for the KNR □lgorithm.

Perform□nce criteri□ / Output v□ri□ble	EVC	R^2	M□	M□PE	M□E
Tnod	0.625	0.62	11.18	0.020	7.39
Treact	0.719	0.69	8.148	0.016	5.20
$Q_{gasoline}$	0.590	0.59	0.008	0.019	0.005
COR	0.503	0.49	0.127	0.001	0.079

2.4. Support Vector M□chine

Support Vector Machine (SVM) is a supervised ML method used for solving different linear problems (even with reduced training data sets), such as linear and nonlinear regression (Support Vector Regression- SVR), linear and nonlinear classification (Support Vector Classification- SVC) and outlier detection [19].

Support vector regression (SVR) is employed to address numerical regression problems, where the input vectors are mapped (using similar functions named kernels like the Gaussian (RBF) kernel) into higher dimensional spaces.

Unlike linear regression, SVR defines new margins that regress along the line with the margins (called SVR tube), so the points that lie within the boundaries of the SVR tube boundaries are not regarded as errors, while the outside points the SVR tube are considered errors19].

SVR generates precise regression models by finding an optimal hyperplane for sample segmentation, minimizing the difference between predicted and actual values [18, 19].

The experimental output datasets are compared with output variables predicted using SVR and are shown in Figures 15-18.

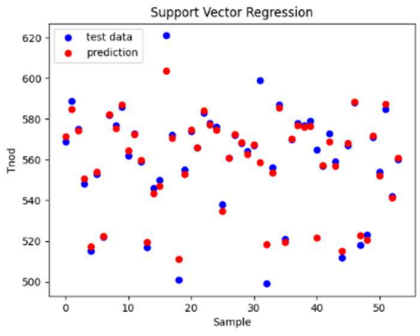

Fig. 15. Comparison between T_{nod_exp} and T_{nod_pred} using SVR.

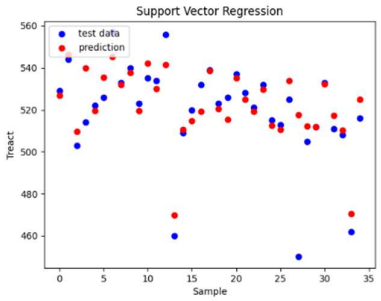

Fig. 16. Comparison between T_{react_exp} and T_{react_pred} using SVR.

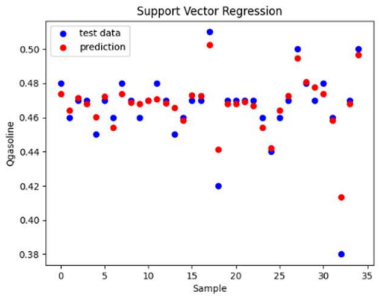

Fig. 17. Comparison between $Q_{gasoline\_exp}$ and $Q_{gasoline\_pred}$ using SVR.

Table 4 presents the obtained model evaluation metrics for SVR algorithm.

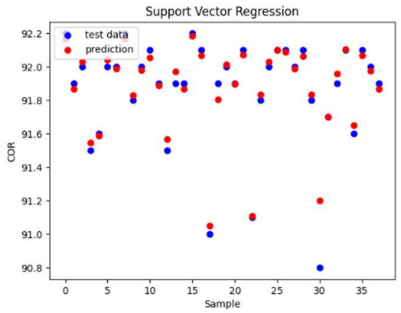

Fig. 18. Comparison between $COR_{\_exp}$ and COR_{pred} using SVR.

Table 4

The model ev□lu□tion metrics for the SVR □lgorithm.

Perform□nce criteri□ / Output v□ri□ble	EVC	R^2	M□	M□PE	M□E
Tnod	0.872	0.86	3.84	0.006	1.48
Treact	0.654	0.64	7.45	0.014	3.54
$Q_{gasoline}$	0.849	0.84	0.005	0.012	0.003
COR	0.946	0.94	0.040	0.0004	0.032

2.5. Gr□dient boosting regression

In ML boosting is understood as the combination of multiple simple models (weak

learners such as decision trees) into a single and complete final model (stronger predictor). For loss minimization the algorithm uses the gradient descent method, a fact that justifies the term "gradient" from the boosting method name [20]. It is a supervised learning algorithm, applied to classification (when gradient boosting is used to predict classes) and for regression (when gradient boosting is used to predict a continuous value).

The basic idea of the boosting methods is that the predictors are sequentially trained, each trained predictor correcting its predecessor. The most used boosting methods are Adaptive Boosting and Gradient Boosting [19].

According to [20, 21], the Gradient Boosting Regression (GBR) uses the algorithm Gradient Boost regression, utilizing decision trees technique, enhances prediction accuracy by adding a sequence of weak classifiers though iterative repetitions.

Figs. 19-21 presents the comparison between the output experimental data of the reactor and output variables predicted using the GBR algorithm.

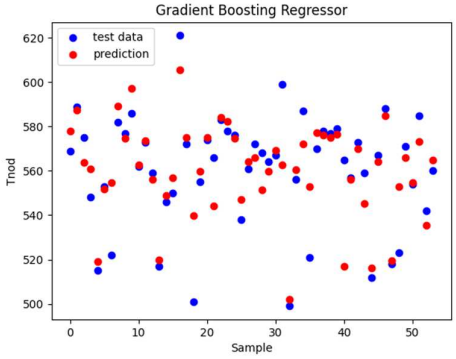

Fig. 19. Comparison between T_{nod_exp} and T_{nod_pred} using GBR.

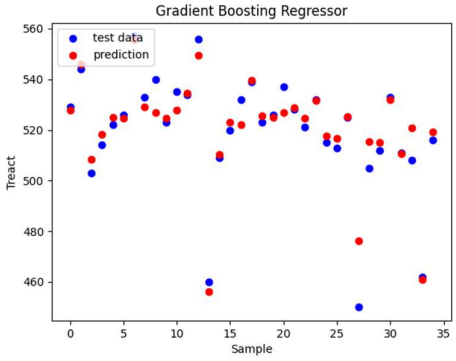

Fig. 20. Comparison between T_{react_exp} and T_{react_pred} using GBR.

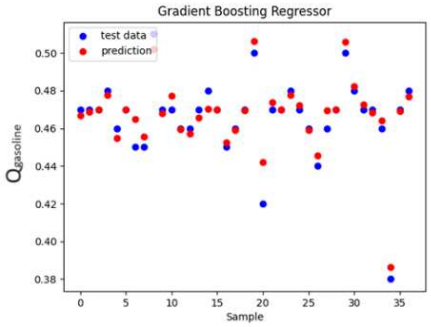

Fig. 21. Comparison between $Q_{gasoline\_exp}$ and $Q_{gasoline\_pred}$ using GBR.

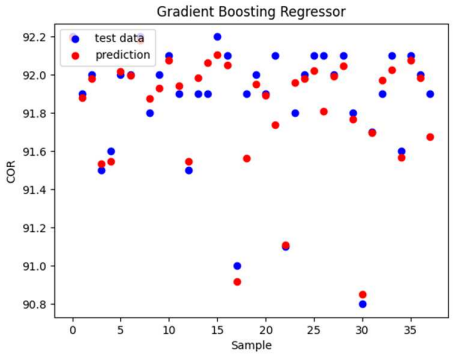

Fig. 22. Comparison between $COR_{\_exp}$ and COR_{pred} using KNR.

Table 5 presents the obtained model evaluation metrics for the GBR algorithm.

Table 5

The model ev□lu□tion metrics for the GBR □lgorithm.

Perform□nce criteri□ / Output v□ri□ble	EVS	R^2	M□	M□PE	M□E
Tnod	0.685	0.685	8.96	0.016	4.38
Treact	0.914	0.913	4.391	0.008	2.823
$Q_{gasoline}$	0.924	0.920	0.004	0.008	0.002
COR	0.880	0.869	0.073	0.0007	0.047

Summarizing the previously presented results, table 6 presents the parameters of the models obtained for each reactor output, respectively *Tnod*, *Treact, Qgasoline*, and *COR*. For each reactor output variable, the method with the smallest error was chosen.

Table 6

The model parameters were obtained for different ML algorithms.

Output p□r□meter	ML □lgorithms	M□PE
Tnod	SVR	**0.006**
COR	SVR	**0.0004**
Treact	GBR	**0.010**
Qgasoline	GBR	**0.008**

3. CONCLUSION

The purpose of this study was to establish what method of ML is more suitable for modeling the reactor of the fluid catalytic cracking process from an industrial refinery. ML can significantly improve optimization, maintenance and quality control in catalytic cracking process, by leveraging the advantages of regression method analyzed.

The methods that were applied for each output variable of the reactor. Analyzed the evolution metric for each method, the best prediction for interfusion nod and COR was obtained using Support Vector Machine regression, while for the reactor temperature and yield gasoline, the best prediction was obtained using Gradient Boosting Regression.

The ML model development can be used in modeling the catalytic cracking process and in developing data-driven applications useful by operators in decision-making. Additionally, ML significantly improves the optimization and yield of process.

4. REFERENCES

[1] Reza, S., *Fluid Catalytic Cracking Handbook,* Butterworth- Heinemann, ISBN 978-0-12-812663-9, Texas, 2020.

[2] Popa, C, Pătrășcioiu, C., *New Approach in Modelling, Simulation and Hierarchical Control of the Fluid Catalytic Cracking Process I - Process Modeling*, Revista de chimie, vol. 61, pp. 419-246, 2010.

[3] Weekman Jr., V.W., *A Model of Cracking Catalytic Conversion in Fixed, Moving, and Fluid Bed Reactors,* Industrial &Engineering Chemistry Process Design and Development, vol. 7, pp. 90-95, 1968.

[4] Wojciechowski, B.W., *The Kinetic Foundations and the Practical Application of the Time on Stream Theory of Catalyst Decay*, Industrial &Engineering Chemistry Process Design and Development, volume 9, pp. 79-113, 1974.

[5] Vorobev, A., Chuzlov, V., Elena, I., Artem, A., *Simple Model of Industrial Catalytic Cracking Riser Reactor*, Industrial & Engineering Chemistry Research, volume 52, pp. 22005-22016, 2023.

[6] Dragomir, R., Paul, R., Popa, C., *Five-kinetic model for Catalytic Cracking Process*, Revista de chimie, volume 59, pp. 2633-2638, 2018.

[7] Heydrei, M., Habib, A, Dabir, B., *Study of seven lump kinetic model in fluid catalytic cracking unit*, American Journal of Applied Sciences, volume 7, pp. 71-76, 2010.

[8] Jacob, S.M., Gross, B., Voltz, S.E., Weekman, V.M., *A Lumping and Reaction Scheme for Catalytic Cracking*, AIChE Journal, volume 2, pp. 701-713, 1976.

[9] Barbosa, A.C., Lopes, G.C., Rosa, L.M., Mori, M., Martignoni, W.P., T*hree Dimensional Simulation of Catalytic Cracking Reactions in an Industrial Scale Riser Using an 11-lump Kinetic Model*, AIDIC Conference, volume 11, pp. 31-41, ISBN 978-88-95608-55-6, 2013.

[10] Doicin, B., Carburenu, M., Popa, C.R, *Artificial Neuronal Network VS Linear Regression for Modeling Reactor of the Catalytic Cracking Process*, Proceeding of the 24th International Multidisciplinary Scientific GeoConferences: Informatics, SGEM, volume 24(2.1), pp. 27-34, ISBN 978-619-7603-69-9, 2024.

[11] Khalid. A.A, Omara, H., Lazzar, M., Achlab, M.A., *Computational Intelligence and Applications for Pandemic and Healthcare*, ISBN 9781799898313, 2022.

[12] Popescu, C., Cangea, O., Bucur, G., Mois,e A., *The Utility of Neuro-Fuzzy Hybrid Systems in Control*, Informatics, Geoinformatics and Remote Sensing Conference Proceeding, SGEM, volume 1, pp. 383-390, ISBN 1314-2704, Albena, Bulgaria, 2015.

[13] Sen, P. C., Hajra, M., Ghosh, M., *Supervised Classification Algorithms in Machine Learning: A Survey and Review,* Advances in Intelligent Systems and Computing, volume 27, pp. 99-111, 2020.

[14] Reza, R.R., Mohammadreza, B., Sajadi, S.M., Pirmoradian, M., Renani, M.R., Baghaei, S., Salahshour, S., *Prediction of the thermal behavior of multi-walled carbon nanotubes-CuO-CeO2 (20-40-40)/water*

hybrid nanofluid using different types of regressors and evolutionary algorithms for designing the best artificial neural network modeling, Alexandria Engineering Journal, volume 84, pp. 184–203, 2023.

[15] Cui, Z., *Machine learning and small data*, Educational measurement, volume 40, pp. 8-12, 2021.

[16] Carbureanu, M., Mihalache, S, F., Zamfir, F., *Machine Learning Methods Applied for Waster pH neutralization Process Modelling*, 14th International Conference on Electronics, Computers and Artificial Intelignece, pp. 101-107, ISBN:978-1-6654-9535-6, Romania, 2022.

[17] Muller, A.C., Guido, S., *Introduction to Machine Learning with Python, A Guide for Data Scientists*, Published by O'Reilly Media Inc., ISBN 1449369901, 2017.

[18] Huang, R., Ma, C., Ma, J., Huangfu, X., He, Q., *Machine Learning in Natural and Engineered Water Systems*, Water Research, volume 205, pp. 1-14, 2021.

[19] Yildiz, B., Bilbao, J.I., Sproul, A.B., *A Review and Analysis of Regression and Machine Learning Models on Commercial Building Electricity Load Forecasting*, Renewable and Sustainable Energy Reviews, volume 73, pp. 1104-1122, 2017.

[20] Bentéjac, C., Csörgő, A., Martínez-Muñoz, G., *A comparative analysis of gradient boosting algorithms,* Artificial Intelligence Review, volume 54, pp. 1937–1967, 2012.

[21] Shai, S., Shai, D., *Understanding machine learning. From Theory to Algorithms*, Cambridge University Press, ISBN 9781107298019, 2014.

Aplicarea tehnicilor de învățare automată pentru modelarea reactorului asociat procesului de cracare catalitică

Această lucrare prezintă cercetările și rezultatele obținute de autori privind modelarea reactorului din cadrul procesului de cracare catalitică, utilizând diverse metode ale tehnicilor de învățare automată bazate pe algoritmi de învățare supravegheată. Comparativ cu modelarea experimentală, tehnicile de învățare automate pot elimina din dezavantajele modelări experimentale, cum ar fi corelațiilor puternice între variabile, neliniaritățile modelului, complexitatea modelului cinetic și diversitatea materie prime. Tehnicile de învățare automată stabilește o relație matematică între variabilele de intrare și cele de ieșire ale procesului, pe baza unor date experimentale. În cadrul lucrări au fost aplicate cinci metode de învățare supravegheată pentru modelarea reactorului: algoritmul bazat pe randomizare(RFR), algoritmul bazat pe arbori de decizie(DTR), algoritmul celor mai apropriați vecini(KNN), algoritm bazat pe suport vectorial(SMV) și regresia cu crestară graduală.(GBR).

Cristina Roxana POPA, PhD., Associate Professor**,** Petroleum Gas University of Ploiesti, Automatic Control Computers &Electronics Department, ceftene@upg-ploiesti.ro, tel. 0744121152

Mădălina CĂRBUREANU, PhD., Assistant Professor, Petroleum Gas University of Ploiesti, Automatic Control Computers &Electronics Department, mcarbureanu@upg-ploiesti.ro

Bogdan DOICIN, PhD., Assistant Professor, Petroleum Gas University of Ploiesti, Automatic Control Computers &Electronics Department, bogdan.doicin@upg-ploiesti.ro