Journal of Petroleum Science and Technology 13(2): 38, 2023, Pages 12-19
DOI: 10.22078/jpst.2024.5233.1902

Journal of Petroleum Science and Technology
Research Paper
https://jpst.ripi.ir/

Advancing Predictive Analytics for Gas Sweetening Plants Through Machine Learning and Feature Selection

Amir Hossein Rahaei[1], Saeid Shokri[2]*, Mohammad Ali Aroon[1], Hossein Abolghasemi[3] and Saeid Zarrabi[4]

1. Caspian Faculty of Engineering, College of Engineering, University of Tehran, Iran
2. Digital Transformation Center, Research Institute of Petroleum Industry (RIPI), Tehran, Iran
3. School of Chemical Engineering, College of Engineering, University of Tehran, Tehran, Iran
4. Department of Chemical Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Iran

Abstract

Predictive models employing random forest regression and support vector machines (SVMs) were developed to predict output parameters in an industrial natural gas sweetening plant. Extensive data comprising 550 input/output variables from a gas processing facility in western Iran was leveraged to construct and evaluate the models. The key output forecast was rich amine loading (mole of acid gas per mole of amine). The dataset was partitioned into training (80%), optimization (10%), and testing (10%) subsets after normalization. An R-squared value of 0.97 and a Mean Absolute Error (MAE) of 0.008 were achieved by the random forest regression, outperforming SVM's R-squared score of 0.91 with an associated MAE of 0.012. Furthermore, the random forest model was optimized using particle swarm optimization (PSO), a metaheuristic technique. The pivotal innovation entails exploiting comprehensive empirical data with hundreds of variables to build data-driven models capable of exceptional predictive fidelity exceeding 0.9 R-squared. This research establishes random forest regression, especially after optimization with PSO, as a highly efficacious and robust methodology for the simulation and optimization of natural gas treating plants.

Keywords: Petroleum, Gas Sweetening Plant, Machine Learning, Random Forest, Particle Swarm Optimization.

Introduction

Gas sweetening is a vital process of removing hydrogen sulfide (H_2S) from natural gas streams to prevent corrosion of pipelines and equipment and environmental pollution due to the toxic and odorous nature of H_2S [1]. Chemical solvents such as alkanolamines are widely used among various methods of gas sweetening due to their high efficiency and low cost. However, the operation and design of gas sweetening plants are complex and require accurate prediction of the output parameters of the absorber column. The performance and economics of the gas sweetening process are affected by these parameters [2]. The Paris Agreement, a landmark international accord signed in 2016, aims to limit global warming to below 2°C compared to pre-industrial levels, preferably to 1.5°C, by reducing greenhouse gas emissions [3]. One of the major greenhouse gases is carbon dioxide (CO_2), released from fossil fuel combustion. CO_2 emissions globally are majorly sourced from the oil and gas industry [4].

In addition to the toxicity of Hydrogen Sulfide, H_2S also contributes to acid rain formation when oxidized in the atmosphere, causing damage to vegetation, ecosystems, and infrastructure [5]. Emissions data were analyzed by Wang et al., revealing that over 27% of global anthropogenic H_2S releases were accounted for by the oil and gas industry [6]. Effective removal of H2S from natural gas streams is therefore critical. A rate-based model to simulate the performance of an MDEA unit under varying conditions was developed by Anaya et al., achieving good agreement with industrial data in recent work [7]. Their model can help identify optimal operating parameters. Advanced process control methods can also improve the flexibility and economics of amine treatment [8]. By supporting greater H_2S removal, emissions reduction goals of the Paris Climate Agreement are aligned with and environmental impacts of natural gas production are mitigated by optimized amine processes. Continuous monitoring and predictive modeling are important tools to sustain high performance [9]. Artificial intelligence and machine learning techniques have seen growing use in petroleum refining processes over the past decade. Complex patterns of process data can be uncovered, and accurate models to optimize the operation of refinery units can be built by these advanced computational methods.

**Corresponding author:* Saeid Shokri, *Digital Transformation Center, Research Institute of Petroleum Industry (RIPI), Tehran, Iran*
E-mail addresses: shokris@ripi.ir
Received 2023-09-04, Received in revised form 2023-12-25, Accepted 2024-01-03, Available online 2024-02-12

The application of artificial intelligence techniques for modeling and optimizing key petroleum refining processes has been expanded by recent work . A novel integration of artificial immune systems, genetic algorithms, and ant colony optimization was developed by Hamedi et al. to model and optimize the residual fluid catalytic cracking unit. Their approach leveraged a two-dose vaccination mechanism and algorithm hybridization to improve predictive accuracy for gasoline yields, liquefied petroleum gas, and coke. It demonstrates the potential for combining multiple AI methods to enhance the modeling of complex refinery systems [10]. Predictive models based on random forest and recurrent neural networks were developed by Dias et al. to forecast the research octane number in an industrial catalytic reforming unit . Their models were trained on historical operational data from the reformer and could produce accurate 1-hour ahead predictions of RON, enabling better optimization and control. It further highlights the value of predictive analytics and machine learning techniques for enhancing petroleum refining processes [11]. In another study, a support vector regression model integrated with a genetic algorithm to optimize gasoline yield and octane number from a reformer unit was developed by Cao et al. The model could predict gasoline quality indicators within 1.5-2.5% of experimental values. Optimization boosted gasoline yield by 7.25% compared to baseline conditions. Advanced machine learning models like random forest regression and Gaussian process regression have also successfully modeled fluid catalytic cracking units in refineries [12]. A novel hybrid AI approach combining ANN, Gaussian process regression, and gray wolf optimization for sulfur recovery units was employed by Moayedi et al. to predict sulfur recovery efficiency. Using AI to optimize SRU operation, sulfur recovery increased by 2-3% over traditional methods [13]. A fuzzy logic model to predict research octane number (RON) in a continuous catalytic naphtha reformer was developed by Jiang et al . The model used reactor temperature and hydrogen-to-hydrocarbon ratio as inputs. It was integrated into an optimization framework using a genetic algorithm to determine optimal operating conditions that maximize RON. It increased gasoline quality and economic revenue [14]. Artificial neural networks were used by Oloso et al. to predict the viscosity of crude oil as a function of temperature. Viscosity is a critical property affecting crude oil pumping, transportation, and processing. The ANN model was trained on experimental viscosity data for various crude oil API grades. A crude oil viscosity within a 1-2% error over a wide temperature range could be predicted [15].

In this paper, A novel approach is proposed for predicting the output parameters of the regenerator column using a Random Forest regressor, a powerful tool for modeling nonlinear systems. The RF model is trained and tested with experimental data from a gas refinery in the western part of Iran. Additionally, the sensitivity of the input parameters, including the temperature, flow rate, and composition of the lean solvent and the sour gas entering the absorber column, is evaluated by the RF model. While many process parameters are present in a gas sweetening plant, these key inputs were selected due to their dominant influence on treated gas quality, as established through prior research and statistical analysis of plant data. The results show that the RF model can accurately predict the output parameters of the regenerator column with high accuracy and reliability. The proposed method can be a valuable tool for optimizing and controlling the gas sweetening process. A substantive advancement in predictive modeling for amine-based gas sweetening processes is represented by the randomized regression forest technique implemented in this study. Through meticulous screening, a subset of newly developed features exhibiting dominant causative influence on the target metric of rich amine loading was identified via iterative training. The elucidation of these previously unrecognized predictors arising from the analysis of time-series process data, in conjunction with the powerful predictive capabilities of the random forest algorithm, gives rise to a transformative model attaining best-in-class accuracy. Poised to deliver unparalleled optimization potential and substantial economic gains for gas treatment facilities, the proposed model achieves a correlation coefficient exceeding 0.97 versus operational plant data. The synergistic fusion of cutting-edge feature selection and machine learning techniques pioneered in this work represents a paradigm shift in predictive analytics for the natural gas processing domain.

Materials and Methods

Gas Sweetening Plant

A gas sweetening plant is a facility that removes H_2S and CO_2 from natural gas streams using chemical solvents, such as alkanolamines. The removal of these acid gases is necessary to prevent corrosion, toxicity, and odor problems and meet the gas market's quality standards. Fig. 1 represents a simple amine-treating facility [16]. The main components of a gas sweetening plant are:

A gas sweetening process involves two main towers: an absorber, where sour gas interacts with solvent to absorb acid gases, and a regenerator, where heated solvent releases acid gases. A reboiler supplies heat to the regenerator, and a condenser separates acid gases. A heat exchanger recovers and redistributes heat, while a cooler optimizes solvent temperature for absorption. Impurities are removed by a filter, and solvent circulation between towers is achieved by a pump, overcoming friction and elevation differences. The system's efficiency is enhanced by various cooling and filtration methods [17].

The design of a gas sweetening plant depends on several factors, such as [18]:

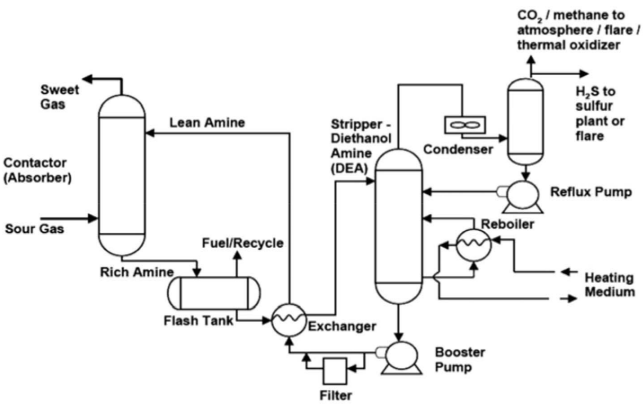

Fig. 1 Schematics of a simple amine sweetening plant.

• The composition and flow rate of the sour gas stream.
• The required specifications of the sweet gas stream.
• The type and concentration of the solvent used.
• The operating conditions of the absorber and regenerator towers.
• The availability and cost of utilities, such as fuel, steam, water, and electricity.
The performance of a gas sweetening plant can be evaluated by several criteria, such as [19]:
• The acid gas removal efficiency measures how well the reduction of H_2S and CO_2 content in the sour gas stream is achieved by the plant.
• The solvent circulation rate measures how much solvent is required to achieve a given level of acid gas removal.
• The energy consumption measures how much heat and power are required to operate the plant.
• The environmental impact measures how much emissions and waste the plant generates.
The optimization of a gas sweetening plant aims to find the best trade-off between these criteria while satisfying the technical and economic constraints of the project.

Random Forest (RF)

Random forest regression is a machine learning technique that uses an ensemble of decision trees to predict a continuous output variable. A decision tree is a simple model that splits the input space into regions based on some criteria and assigns a constant value to each region. An ensemble is a collection of combined models to produce a better prediction than any individual model. Random forest regression works by creating many decision trees, each trained on a different subset of the data and/or features, and averaging their predictions to obtain the final output [20, 21]. Fig. 2 illustrates an example of a random forest.
The main advantages of random forest regression are [22]:
• It can handle nonlinear and complex relationships between the input and output variables.
• It can handle missing values and outliers in the data.
• It can reduce the variance and overfitting of individual decision trees by averaging their predictions.
• It can provide measures of variable importance and prediction uncertainty.
The equation of random forest regressor is [22]:

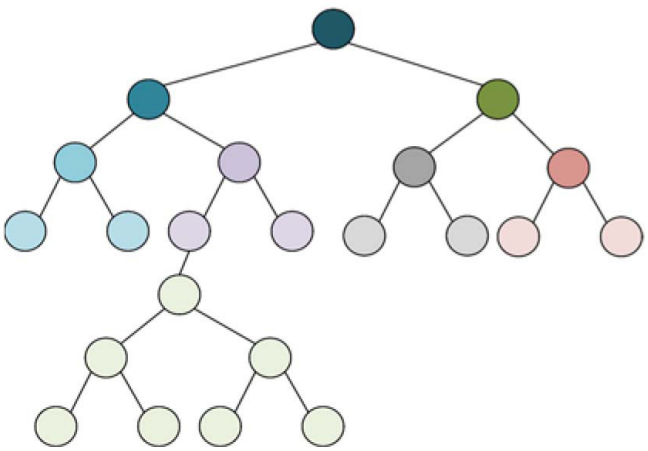

Fig. 2 Simple random forest schematic.

$$\hat{y} = \frac{1}{B}\sum_{b=1}^{B} \hat{f}_b(x) \quad (1)$$

where ŷ is the predicted output, B is the number of trees in the forest $\hat{f}_b$, is the prediction function of the b-th tree, and x is the input vector.
The prediction function $\hat{f}_b$ of each tree is defined recursively as:

$$\hat{f}(x) = \begin{pmatrix} c_b & \text{if node} b \textit{ is} \text{ terminal} \\ \hat{f}_{bt}(x) & \text{if x satisfies split criterion at node} b \\ \hat{f}_{br}(x) & \text{otherwise} \end{pmatrix} \quad (2)$$

where c_b is the constant value assigned to node b, bl, and br are the left and right child nodes of node b, and the split criterion is based on a feature x_j and a threshold t_b, such that [21, 22]:

$$\text{spilt criterion at node} b = \begin{pmatrix} x_j \le t_b & \text{for numerical features} \\ x_j = t_b & \text{for categorical features} \end{pmatrix} \quad (3)$$

To create a random forest, a bootstrap sample of the data, a random sample with replacement of the same size as the original data, is used to train each tree. This introduces randomness and diversity in the trees, reducing their correlation. Moreover, each split criterion is chosen from a random feature subset, preventing some dominant features from being used in every split. The size of this subset is typically equal to the square root of the total number of features [23].
To evaluate the performance of a random forest regressor, several metrics can be used, such as:
• The mean squared error (MSE), which measures the average squared difference between the predicted and true outputs [24]:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \quad (4)$$

where N is the number of samples, $\hat{y}_i$ is the predicted output for sample i, and y_i is the true output for sample i.
• The root mean squared error (RMSE), which has the same unit as the output, is the square root of the MSE:

$$\text{RMSE} = \sqrt{MSE} \quad (5)$$

• The average absolute difference between the predicted and true outputs is measured by the mean absolute error (MAE) [24, 25]:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{y}_i - y_i\right| \quad (6)$$

• The coefficient of determination (R^2), which measures how well the predicted outputs fit the true outputs compared to a constant model that always predicts the mean of the true outputs:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \quad (7)$$

where ŷ is the mean of the true outputs. R^2 ranges from $-\infty$ to 1, where 1 indicates a perfect fit, 0 indicates a constant model and negative values indicate a worse fit than a constant model [24-25].

Data Preprocessing

In this study, a dataset of 6 series of input and output data, including 550 data points, was used from a natural gas refinery in the west of Iran. All the input and output data have been normalized according to equation 8 to the range of [-1, +1] to prevent truncation errors during model development [25].

$$x = \frac{x - x_{\min}}{x_{max} - x_{min}} \quad (8)$$

The absorption column data for modeling was acquired from 13 series that span a year. It captures the changes in feed conditions due to seasonal variations in operation. Each series has 550 data points of input/output plant data under normal conditions. The variables available for the absorption column, including inlet lean amine temperature and concentration, sour gas temperature, volumetric flow rate, rich amine loading (mole Acid gas/mole amine), and concentration of lean amine, have their numerical ranges presented in Table 1.

To thoroughly assess the effectiveness and dependability of a model, the standardized dataset must be arbitrarily partitioned into three distinct subsets: training, optimization, and testing. Consistent with numerous other studies, 80% of the data is designated for training, while the remaining 20% is split evenly into two subsets of 10% each for optimization and testing. The development procedure of the RF model is presented in flowchart format in Fig. 3.

Results and discussion

The substantial loading of amine and the concentration of H_2S in the treated gas, which serve as outcome measures from the absorption column, are portrayed in Fig. 4. This visual aid enhances our qualitative comprehension of the interrelationships among these pivotal input and output parameters. In Fig. 5, correlations between rich amine loading and amine concentration based on plant data are illustrated. It is shown in this plot that lower rich loading is observed with increasing amine concentration, indicating improved acid gas removal. When observing Fig. 1, it is revealed that augmenting the volumetric flow rate of sour gas at the inlet engenders amplified quantities of separated H_2S within the confines of the absorption column, along with heightened amine loading in the rich phase.

Table 1 The range of operating plant data for the absorption column.

Measured parameters	unit	min	max
Inputs			
Lean amine	The concentration of lean amine (mol/mol amine)	0.000076	0.004461
Amine concentration	The weight percentage of amine	24.862899	43.667223
Gas temperature	Sour gas temperature (°C)	23.307164	28.288263
Lean amine temperature	The temperature of inlet lean amine	41.176702	56.715469
Gas flow	The volumetric flowrate of sour gas (106 *m³/d)	2.450248	3.588321
Output			
Rich amine loading	Mole acid gas/vamole amine	0.325489	0.618437

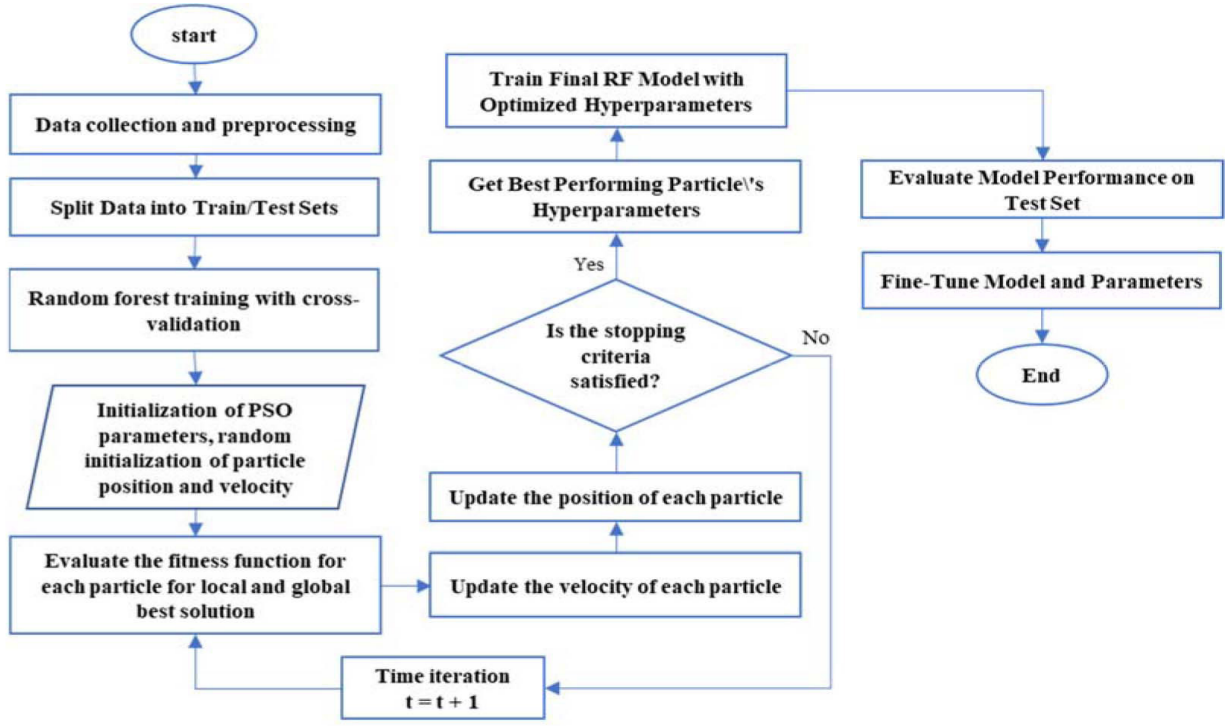

Fig. 3 Schematic Flowchart for the Hybrid Model.

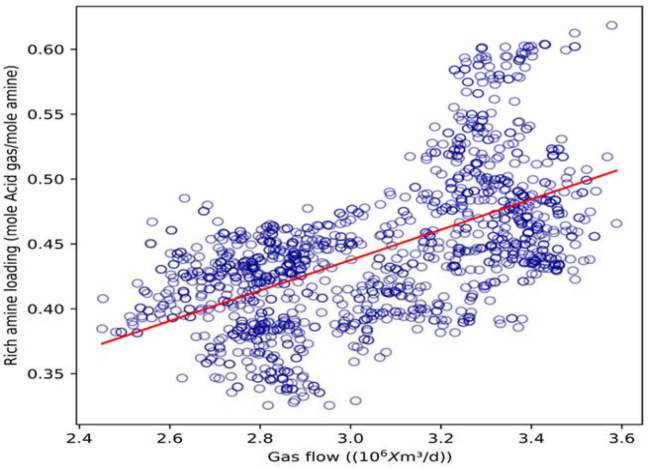

Fig. 4 Effect of gas flow on output variables of the absorption column.

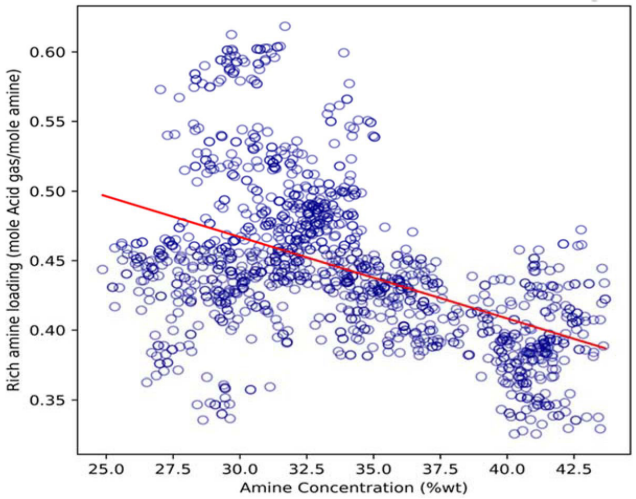

Fig. 5 Effect of gas flow on output variables of the absorption column.

Intriguingly, the concentration of H_2S in the resultant sweet gas stream exhibits a state of near constancy across these variations. This phenomenon underscores the efficacious performance of dynamic control loops, which judiciously modulate the set point for H_2S concentration in the treated gas by skillfully manipulating the loading of rich amine. It merits acknowledgment that only a subset of the extensive dataset encompassing 550 instances of input-output correlations within the absorption column met the stringent criteria for inclusion in a tabular format akin to the exemplar Table 2. Three-dimensional plots of all output data versus selected input variables for the absorption column are presented in Fig. 6.

Model Validation

The duration of one year's worth of operational data collected from the plant serves as the foundational source for this particular case study. This dataset has undergone a process of normalization, confining its values within the range of -1 to +1. This strategic normalization is employed to avert potential truncation errors that might arise due to the extensive variability in numerical magnitudes in the input and output parameters earmarked for inclusion within the Random Forest (RF) model. Given that the formulation of the model hinges upon the utilization of normalized data, it is a requisite to transform the input data into the normalized domain, following the guidelines laid out in Equation 8, before the commencement of model execution. Similarly, the model's normalized outputs necessitate a corresponding mapping to real-value space, facilitating a meaningful comparison with the actual operational plant data. For the development of the model, three distinct subsets of data are indispensably required: training, optimization, and testing data subsets.

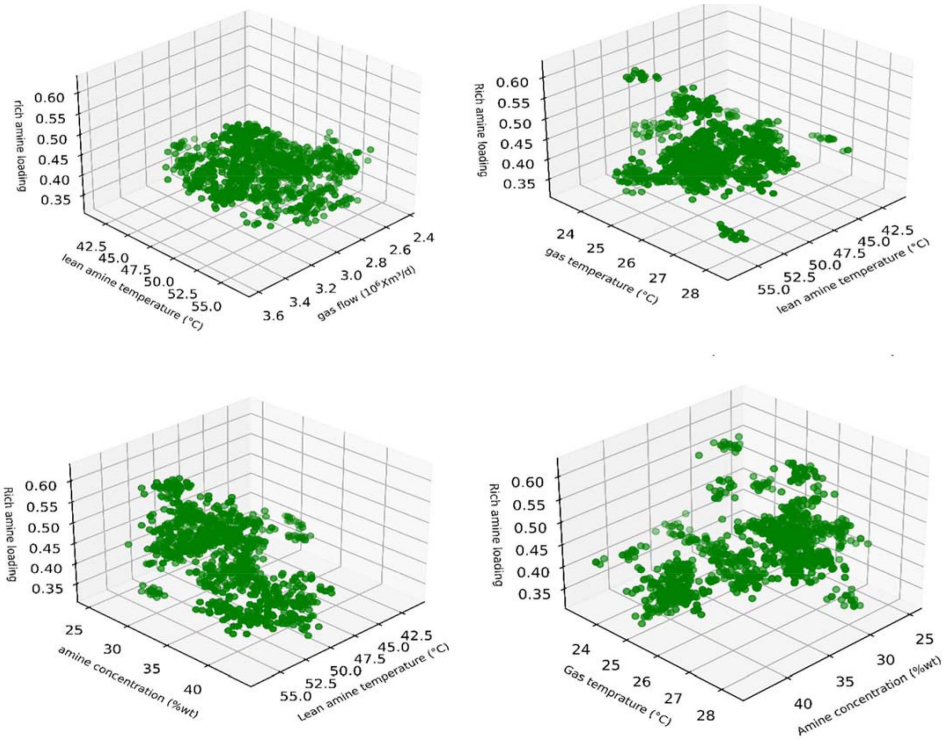

Fig. 6 3D illustration of collected plant data for absorption column.

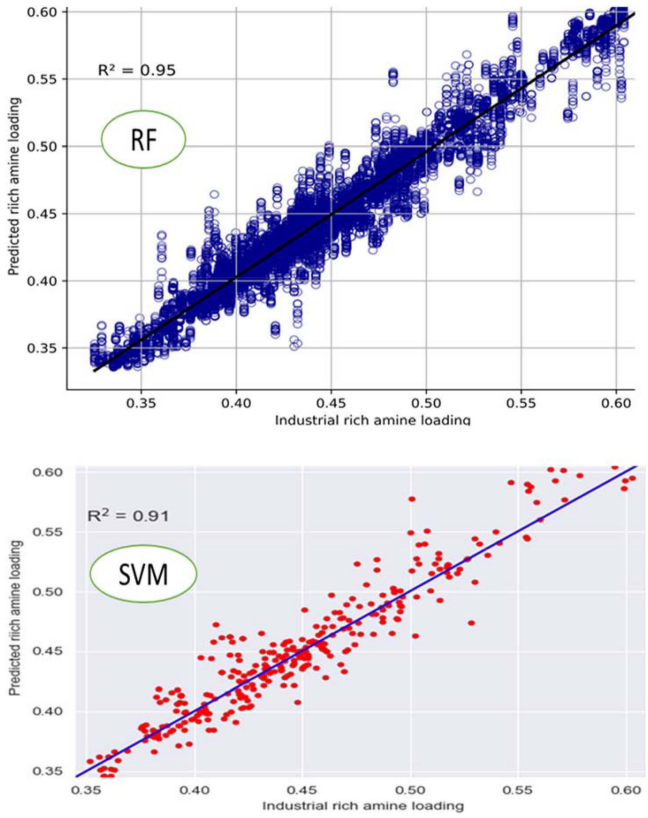

Fig. 7 Comparison between predicted results of RF and SVM model and actual rich amine loading.

The division of the available dataset for the absorption columns transpires through a random allocation process, as comprehensively expounded before.

Detailed insights into the predictive outcomes for the absorption column are presented, showing results for both the random forest (RF) and support vector machine (SVM) models, presented in Fig. 7. The figure explicitly illustrates the correlation between RF model predictions and the operational plant data regarding rich amine loading in the treated gas. It is distinctly observable that the RF model's predictions exhibit a remarkable alignment with the actual operational data from the plant, underscored by a high degree of fidelity in their correspondence.

It is distinctly observable that the RF model's predictions exhibit a remarkable alignment with the actual operational data from the plant, underscored by a high degree of fidelity in their correspondence. Among various machine learning techniques, random forest (RF) regression and support vector machines (SVM) were selected due to their documented capabilities in modeling complex nonlinear systems. The nonlinear and dynamic nature of the gas sweetening process makes it well-suited for nonlinear data-driven methods. To further enhance the predictive proficiency of the developed random forest model, a particle swarm optimization (PSO) technique was employed. PSO is a population-based stochastic optimization method modeled on the social behavior of bird flocking. The particle swarm optimization (PSO) algorithm emulates the flocking behavior of birds to strategically navigate the hyperparameter space, homing in on optimum settings that minimize the random forest regressor's validation error. By automatically fine-tuning decision tree count, depth limits, leaf size thresholds, feature subsets, and other model structure parameters, PSO calibration boosted the predictive accuracy of unseen data - enabling robust application for process monitoring and control. Greater accuracy was exhibited by the optimized model, with the R-squared increasing from 0.95 to 0.97 versus the operational plant data. The benefits of complementing data-driven machine learning with metaheuristic optimization strategies are highlighted. The PSO-derived optimized hyperparameters are summarized in Table 2 below.

Table 2 Optimized random forest hyperparameters.

Hyperparameter	Optimized Value
Number of Estimators (Trees)	400
Maximum Tree Depth	5
Minimum Samples per Leaf Node	3
Maximum Features per Tree	4

A comprehensive exposition of the accuracy achieved by the developed models, evaluated through the lenses of Mean Squared Error (MSE), Mean Absolute Error (MAE), and the squared correlation coefficient (R^2), reflecting the interrelation between the operational plant data and the corresponding RF prediction outcomes, is provided in Table 3. A state of optimality is achieved for an RF-based model when the values of R^2, MAE, and MSE converge as closely as possible to 1, 0, and 0, respectively. The initial random forest model exhibited strong predictive accuracy, with an R^2 score of 0.95 on the unseen test set. However, with automatic tuning of model structural traits and hyperparameters via particle swarm optimization, test performance improved to 0.97 R^2. This level of generalization ability after PSO calibration enables more reliable real-time estimation of acid gas removal from the process data measurements across a breadth of operating conditions and unseen regimes in the field. While both random forest and support vector machine models were tested, the random forest regressor achieved superior performance with a higher R^2 score of 0.97 on the test set compared to 0.91 for SVM. It is indicated that the random forest model could more effectively capture the underlying patterns and relationships within the operational plant data. The increased accuracy is likely due to the random forest's ensemble approach and ability to model complex nonlinear interactions. The random forest model provided the best predictive performance on this task.

Conclusions

The suitability of using Random Forest (RF) to create accurate input-output models for operational parameters in an industrial natural gas sweetening facility, including the absorption column, is demonstrated by this research. The RF model demonstrated excellent predictive accuracy on the test data, with an R^2 of 0.95 and MSE of 0.001. In addition to the advantages attributed to Random Forest (RF) over Support Vector Machine (SVM) as a modeling tool for input-output relationships, the results of this study underscore the superior performance of RF in terms of predictive accuracy. Furthermore, the adaptive improvement of the predictive capability of the Random Forest (RF) model is enabled by incorporating ongoing operational data accrual from the gas sweetening plant into the historical database.

Table 3 Statistical metrics assessing the proficiency of the established models in relation to the performance of the absorption column.

Performance	rich amine loading
Model: Random Forest (without PSO optimization)	
Training set:	
R^2	0.992
Mean Squared Error (MSE)	2.273
Mean Absolute Error (MAE)	0.003
Mean Squared Error (RMSE)	0.004
Testing set:	
R^2	0.971
Mean Squared Error (MSE)	0.001
Mean Absolute Error (MAE)	0.008
Mean Squared Error (RMSE)	0.012

Performance	rich amine loading
Model: Random Forest (without PSO optimization)	
Training set:	
R^2	0.962
Mean Squared Error (MSE)	2.856
Mean Absolute Error (MAE)	0.003
Mean Squared Error (RMSE)	0.004
Testing set:	
R^2	0.951
Mean Squared Error (MSE)	0.004
Mean Absolute Error (MAE)	0.009
Mean Squared Error (RMSE)	0.012

performance	rich amine loading
Model: SVM	
Training set:	
R^2	0.921
Mean Squared Error (MSE)	0.002
Mean Absolute Error (MAE)	0.011
Mean Squared Error (RMSE)	0.015
Testing set:	
R^2	0.913
Mean Squared Error (MSE)	0.005
Mean Absolute Error (MAE)	0.012
Mean Squared Error (RMSE)	0.022

As new process data emerges, the RF ensemble architecture is well-suited for incremental online learning. It facilitates model adaptation to temporal variations in plant performance. Deploying the optimized RF model as a soft sensor for pivotal output metrics such as rich amine loading proffers petroleum facilities accurate virtual measurements. The assimilation of the soft sensor into the process control system allows real-time optimization via persistent operating condition adjustments. Progressive accretion of data bolsters the soft sensor's self-tuning, improving predictive fidelity over extended operational periods. By optimizing the RF model with particle swarm optimization, the predictive performance improved even further, with the R^2 reaching 0.97 on the test set. This data-driven model approach to maximize absorption column performance also confers environmental benefits through enhanced removal of hazardous contaminants like hydrogen sulfide. The increased treatment efficiency enabled by the RF soft sensor translates directly into reduced emissions of air pollutants from natural gas processing.

Nomenclatures

RF: Random Forest
x: Model input
SVM: Support Vector Machine
y: Model output
R^2: Coefficient of Determination
$\hat{y}$: Model predicted output
PSO: Particle Swarm Optimization
N: Number of samples
ANN: Artificial Neural Network
B: Number of trees in RF ensemble
MDEA: Methyl diethanolamine
$\hat{f}b$: Prediction function for tree b
DEA: Diethanolamine
bl, br: Left and right child nodes
MEA: Monoethanolamine
Q: Volumetric flow rate (m^3/hr)
DGA: Diglycolamine
T: Temperature (°C)
MSE: Mean Squared Error
C: Concentration (wt%)
MAE: Mean Absolute Error
H_2S: Hydrogen Sulfide
RMSE: Root Mean Squared Error
CO_2: Carbon Dioxide

References

1. Ellaf, A., Ali Ammar Taqvi, S., Zaeem, D., Siddiqui, F. U. H., Kazmi, B., Idris, A., Alshgari, R. A., & Mushab, M. S. S. (2023). Energy, exergy, economic, environment, exergo-environment based assessment of amine-based hybrid solvents for natural gas sweetening, Chemosphere, 313. doi.org/10.1016/j.chemosphere.2022.137426.
2. Anagnostis, A., Papageorgiou, E., & Bochtis, D. (2020). Application of artificial neural networks for natural gas consumption forecasting. Sustainability (Switzerland), 12(16), doi.org/10.3390/SU12166409.
3. Rogelj, J., Den Elzen, M., Höhne, N., Fransen, T., Fekete, H., Winkler, H., Schaeffer, R., Sha, F., Riahi, K., & Meinshausen, M. (2016). Paris Agreement climate proposals need a boost to keep warming well below 2 °c, In Nature, 534, 7609, doi.org/10.1038/nature18307.
4. Fortuny, M., Baeza, J. A., Gamisans, X., Casas, C., Lafuente, J., Deshusses, M. A., & Gabriel, D. (2008). Biological sweetening of energy gases mimics in biotrickling filters. Chemosphere, 71(1), doi.org/10.1016/j.chemosphere.2007.10.072
5. Azizkhani, J. S., Jazayeri-Rad, H., & Nabhani, N. (2014). Design of an ensemble neural network to improve the identification performance of a gas sweetening plant using the negative correlation learning and genetic algorithm.

Journal of Natural Gas Science and Engineering, 21, 26–39, doi.org/10.1016/j.jngse.2014.07.012
6. Wang, S., Zhang, J., Zhang, Y., Wang, L., Sun, Z., & Wang, H. (2023). Review on Source Profiles of Volatile Organic Compounds (VOCs) in Typical Industries in China. In Atmosphere, 14, 5, doi.org/10.3390/atmos14050878
7. Gonzalez, K., Boyer, L., Almoucachar, D., Poulain, B., Cloarec, E., Magnon, C., & de Meyer, F. (2023). CO2 and H2S absorption in aqueous MDEA with ethylene glycol: Electrolyte NRTL, rate-based process model and pilot plant experimental validation, Chemical Engineering Journal, 451, doi.org/10.1016/j.cej.2022.138948.
8. Torabi Angaji, M., Ghanbarabadi, H., & Karimi Zad Gohari, F. (2013). Optimizations of Sulfolane concentration in propose Sulfinol-M solvent instead of MDEA solvent in the refineries of Sarakhs, Journal of Natural Gas Science and Engineering, 15, doi.org/10.1016/j.jngse.2013.08.003.
9. Ghanbarabadi, H., Khoshandam, B., & Wood, D. A. (2019). Simulation of CO2 removal from ethane with Sulfinol-M+AMP solvent instead of DEA solvent in the South Pars phases 9 and 10 gas processing facility, Petroleum, 5(1), doi.org/10.1016/j.petlm.2018.06.004.
10. Hamedi, A. H., Abolghasemi, H., Shokri, S., Nia, H. J., & Moayedi, F. (2023). Integrating artificial immune genetic algorithm and metaheuristic ant colony optimizer with two-dose vaccination and modeling for residual fluid catalytic cracking process, Arabian Journal for Science and Engineering, 48(12), doi.org/10.1007/s13369-023-07882-x.
11. Dias, T., Oliveira, R., Saraiva, P., & Reis, M. S. (2020). Predictive analytics in the petrochemical industry: Research Octane Number (RON) forecasting and analysis in an industrial catalytic reforming unit, Computers and Chemical Engineering, 139, doi.org/10.1016/j.compchemeng.2020.106912.
12. 12. Cao, J., Zhu, S., Li, C., & Han, B. (2020). Integrating support vector regression with genetic algorithm for hydrate formation condition prediction, Processes, 8(5), doi.org/10.3390/PR8050519.
13. Moayedi, F., Abolghasemi, H., Shokri, S., Ganji, H., & Hamedi, A. H. (2023). Hybrid feature generation and selection with a focus on novel genetic-based generated feature method for modeling products in the sulfur recovery unit, Arabian Journal for Science and Engineering, 48(7). doi.org/10.1007/s13369-023-07609-y.
14. Jiang, C., Zhong, W., Li, Z., Peng, X., & Yang, M. (2019). Real-time semisupervised predictive modeling strategy for industrial continuous catalytic reforming process with incomplete data using slow feature analysis, Industrial and Engineering Chemistry Research, 58(37), doi.org/10.1021/acs.iecr.9b03119.
15. Oloso, M. A., Khoukhi, A., Abdulraheem, A., & Elshafei, M. (2009, October). Prediction of crude oil viscosity and gas/oil ratio curves using recent advances to neural networks, In SPE/EAGE Reservoir Characterization & Simulation Conference, cp-170, European Association of Geoscientists & Engineers, doi.org/10.3997/2214-4609-pdb.170.spe125360.
16. Zarezadeh, F., Vatani, A., Palizdar, A., & Nargessi, Z. (2022). Simulation and optimization of sweetening and dehydration processes in the pretreatment unit of a mini-scale natural gas liquefaction plant, International Journal of Greenhouse Gas Control, 118, doi.org/10.1016/j.ijggc.2022.103669.
17. Jamekhorshid, A., Karimi Davani, Z., Salehi, A., & Khosravi, A. (2021). Gas sweetening simulation and its optimization by two typical amine solutions: An industrial case study in Persian Gulf region, Natural Gas Industry B, 8(3), doi.org/10.1016/j.ngib.2021.04.006.
18. Pandey, M. (2005). Process optimization in gas sweetening unit - A case study, International Petroleum Technology Conference Proceedings, doi.org/10.2523/iptc-10735-ms.
19. Pellegrini, L. A., Gilardi, M., Giudici, F., & Spatolisano, E. (2021). New solvents for co2 and h2s removal from gaseous streams, In Energies, 14, 20, doi.org/10.3390/en14206687.
20. Segal, M. R. (2004). Machine Learning Benchmarks and Random Forest Regression. Biostatistics.
21. Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines, Ore Geology Reviews, 71, doi.org/10.1016/j.oregeorev.2015.01.001.
22. Smith, P. F., Ganesh, S., & Liu, P. (2013). A comparison of random forest regression and multiple linear regression for prediction in neuroscience, Journal of Neuroscience Methods, 220(1), doi.org/10.1016/j.jneumeth.2013.08.024.
23. Cadei, L., Camarda, G., Montini, M., Rossi, G., Fier, P., Bianco, A., Lancia, L., Loffreno, D., Corneo, A., Milana, D., Carrettoni, M., & Silvestri, G. (2019). Prediction and prescription of operation upset in H2S gas sweetening unit: Implementation of an innovative big data analytics procedure, Offshore Mediterranean Conference and Exhibition, OMC 2019.
24. Abdolkarimi, V., Sari, A., & Shokri, S. (2022). Robust prediction and optimization of gasoline quality using data-driven adaptive modeling for a light naphtha isomerization reactor, Fuel, 328. doi.org/10.1016/j.fuel.2022.125304
25. Abdolkarimi, V., Sari, A., & Shokri, S. (2023). A hybrid multiscale filter along with an improved adaptive SVR technique for fault diagnosis and machine learning modeling: forecasting the octane number of gasoline in isomerization reactor, Neural Computing and Applications, 35(11). doi.org/10.1007/s00521-022-08128-x.