Artificial intelligence approach in crude distillation unit operation

Obumneme Onyeka Okwonna ^{1,*} and Amalate Ann Jonathan Obuebite ²

¹ Department of Chemical Engineering, University of Port Harcourt, PMB 5323 Port Harcourt, Rivers State, Nigeria.

² Department of Petroleum Engineering, Niger Delta University, PMB 071, Wilberforce Island, Bayelsa State, Nigeria.

Global Journal of Engineering and Technology Advances, 2021, 09(02), 075–082

Publication history: Received on 18 October 2021; revised on 22 November 2021; accepted on 24 November 2021

Article DOI: <https://doi.org/10.30574/gjeta.2021.9.2.0155>

Abstract

This study incorporates the use of Artificial Intelligence in the monitoring of atmospheric distillation unit of large scale refining operation using Google AutoML tables, Jupyter, and Python software. The process involved training, evaluation, improvement, and deployment of the models based on the input data. The predicted yield (vol %) for the models were: Auto ML model: liquefied petroleum gas (LPG) - 1.41, straight run gasoline (SRG) - 4.96, straight run naphtha (SRN) - 17.87, straight run kerosene (SRK) - 14.5, light diesel oil (LDO) - 26.47, heavy diesel oil (HDO) - 2.7, and atmospheric residue (AR) - 30.03; Jupyter Model: LPG - (0.93), SRG - (4.69), SRN - (17.24), SRK - (14.39), LDO - (26.43), HDO - (2.7), and AR - (30.18); and Python Model: LPG - (1.66), SRG - (7.58), SRN - (11.68), SRK - (14.92), LDO - (24.77), HDO - (4.59), and AR - (24.59). The coefficient of determination (R^2) values of 0.99981, 0.99943, and 0.93078 and Standard Error values of 0.240918, 0.419291, 3.536064, were obtained for the 3 models, respectively. All the software gave good predictions of the actual yield, although the Google Auto ML Table gave the best prediction. The training of the model is fundamental to its performance and precision.

Keywords: Artificial intelligence; Crude distillation unit; Mathematical modeling; Machine learning; Refining; Processing

1. Introduction

The oil and gas industry plays a crucial role in energy generation and is the mainstay and a major driver of most economies around the world. Despite the research and discoveries on other forms of energy, the petroleum industry remains crucial and hence the need for more inventions to enhance its processes. Crude distillation unit, being the first stage of the refining process is a fundamental process in the value chain. It is a highly energy intensive process and demands high efficiency and the constant demand on the need to improve its operation through necessary due diligence [1]. Al-Dunainawi & Abbod [2] evaluated the relationship between process performance and product composition of distillation columns, describing the process as nonlinearly transit and its complexity.

The need for process improvement in the refining process and its products implies the need for more sophisticated technologies and methods through the application of artificial intelligence (AI). Stephen [3] described AI as an ever evolving process at an ever increasing pace. It is imperative that AI can also have major breakthroughs in the oil and gas industry, especially the downstream sector to harness its efficiency and effectiveness. This approach will involve data observation, preparation, planning, and model building [4].

Frankenfield [5] described AI as any mechanism that displays traits related with a human mind like learning and solving problem. It is built on the principle that a machine can effortlessly imitate it and implement tasks, from the most-simple to those that are even more composite. A vital aspect of AI is machine learning which could be categorized into four

*Corresponding author: Obumneme O Okwonna

Department of Chemical Engineering, University of Port Harcourt, PMB 5323 Port Harcourt, Rivers State, Nigeria.

Copyright © 2021 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0.

algorithms namely: supervised, unsupervised, semi-supervised, and reinforcement machine learning involved with the use of classified, unclassified, semi-classified data, and data interaction with the environment, respectively. It has the ability to learn and improve from experience when there has not been any explicit programming [6].

Several models with have been applied in the study of the distillation process [2, 7 – 13]. Additionally, Abbasi et al.[14] highlighted the importance of machine learning in the prediction of solutions to large scale industrial processes. These models focused on the need to improve efficiency of several aspects of the distillation operation.

Although much improvement has been made in applying auto machine learning (ML) in distillation as cited in these studies, the rigor of integrating neural networks and other optimization systems are still inbound. This study evaluates the application of Auto ML in the prediction of the output of a commercial distillation process, and to draw a comparison on the performance of different software.

2. Methods

This study involved an analytical research design. Google Auto ML Console, Jupyter, and Python software were used to model the data obtained from the atmospheric distillation unit of a commercial process over a 6-month period of continuous run operation. The structured data comprised of numeric and alphabetic components which describe the input data.

2.1. Mathematical modeling

A multiple linear regression equation was used to model the data as follows. This model considered the relationship between the dependent variable and the independent variables as shown in Equation (1).

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 \quad (1)$$

Where: Y = product yield (vol %), x_1 = Temperature, x_2 = Flow rate (m^3/hr), x_3 = Specific Gravity, b_1 , b_2 , b_3 = regression coefficients.

2.2. Programming and software development

A Visual Studio Code (VS-CODE) version 1.49.3 designed by Microsoft was used to generate the linear regression models for the predictions. It involved installing a programme interpreter using the appropriate extension (`-py` = python and `ipynb` = Jupyter) which was generated for language coding, libraries created, data imported, graphical link of the variables created using matplotlib, and the linear regression model utilized. The developed model was used to make predictions on the input trained data and tested using a new input data to obtain a new set of data (predicted data).

The Google cloud autoML uses simple graphical user interface to train, evaluate, improve, and deploy models based on the data provided. Available custom models include: natural language, tables, vision intelligence, and translation and were dependent on the information provided. Google AutoML Tables was used for this study on the structured data following the approach (Fig. 1). Training the model on the data, involved normalization and encoding.

Figure 1 Flow diagram of AutoML Process

The following criteria were met for the use of the software:

- input data was ≤ 100 gb.
- Inclusion of the target column
- There were not more than 2 - 1,000 columns.
- One column was the target, with feature column available to train the model.
- There were at least 1,000 and no more than 100,000,000 rows.

The prepared data set was thereafter imported into AutoML using CSV files. For this study, the smaller dataset which ranged from 2 – 1000 columns was used. The imported data set was split into training set, validating set and test set. In splitting the data into columns, the time and target columns were included (Fig. 2).

Figure 2 Auto ML Split Page

The target column was numerical and not null. The target column is what the model was trained to predict. The data type determined the resulting model as a regression (Numeric) model. A representation of the numeric model is shown (Fig. 3).

Column name ? ↑	Data type ?	Nullability ?	Missing% (Count) ?	Invalid values ?	Distinct values ?
Gravity	Numeric ▼	☐ Nullable	0% (0)	0% (0)	7
Input	Numeric ▼	☐ Nullable	0% (0)	0% (0)	15
Target Target	Numeric ▼	☐ Nullable	0% (0)	0% (0)	21
Temperature	Categorical	☐ Nullable	0% (0)	0% (0)	7

Figure 3 AutoML data type column

2.3. Statistical analysis

The accuracy of the model was used to evaluate the precision of the model on the prediction of the test dataset. Precision of the regression evaluations was determined using the correlation (Eq. 2)

$$precision_{micro} = \frac{TP_1 + \dots + TP_n}{TP_1 + \dots + TP_n + FP_1 + \dots + FP_n} \quad (2)$$

Where: TP_n and FP_n are the true positives false positives for each of the n classes, respectively.

In addition, Google AutoML Tables provided two other ways to comprehend the model, the confusion matrix and a feature importance graph. The confusion matrix (Fig. 4) explained the misclassifications within the model whereas the feature importance (Fig. 5) indicated how much each element of the independent variable impacted on the model. The coefficient of determination (R^2) and Standard Error values from these models were also obtained.

Figure 4 AutoML Confusion Matrix

Figure 5 Feature importance

3. Results and discussion

The structured technical data (Table 1) highlights the product composition, the dependent variable (product yield) and the independent variables (temperature, flow rate, and specific gravity).

Table 1 Technical data of product yield

Product	Yield (Vol%)	Temp. (°C)	Flow rate (m ³ /hr)	Sp. gravity @ 60 °F
LPG	1.5	17	14.90625	0.55
SRG	5	64	49.6875	0.673
SRN	18	138	178.875	0.765
SRK	14.5	208	144.0938	0.828
LDO	27	283	268.3125	0.863
HDO	3.5	324	34.7813	0.898
AR	30.5	391	303.0938	0.944

Note: Values were a mean of 180 determinations. Mid range temperature values have been presented.

The actual yield of products in this unit were LPG – 1.5%, SRG – 5%, SRN – 18%, SRK – 14.5%, LDO – 27%, HDO – 3.5%, and AR – 30.5%. The predicted yield of these products (Table 2) with the empirical models (Equations 3 to 5) using the

AutoML, Jupyter, and Python software, respectively indicate the yields as follows : LPG - 1.41% , SRG – 4.96%, SRN – 17.87%, SRK – 14.5%, LDO – 26.47%, HDO – 2.7%, and AR –30.03%; LPG - 0.93% , SRG – 4.69%, SRN – 17.24%, SRK – 14.39%, LDO – 26.43%, HDO – 2.7%, and AR –30.18%; LPG - 1.66% , SRG – 7.58%, SRN – 11.68%, SRK – 14.92%, LDO – 24.77%, HDO – 4.59%, and AR –24.59%. All the software gave almost a similar range of product quality as that of the actual yield.

$$Y = -1.1465 - 0.0042x_1 + 0.1017x_2 + 1.9650x_3 \quad (3)$$

$$Y = -1.0114 - 0.0012x_1 + 0.1013x_2 + 0.8464x_3 \quad (4)$$

$$Y = -2.5938 - 0.0018x_1 + 0.0735x_2 + 6.7977x_3 \quad (5)$$

Where: Y = product yield (Vol %), x_1 = Temperature (°C), x_2 = Flow rate (m³/hr), x_3 = Specific gravity.

Al-Dunainawi and Abbod[2] highlighted the substantial impact of quality in the analysis of the products of the atmospheric distillation unit and the need for a reliable analysis of the product composition. The effectiveness of the Auto ML software in predicting the product yield could be attributed to the software training and its outstanding ability to learn the input – output relationship of nonlinear complex system of this unit which made it possible to effectively reduce the error between the predicted and actual yield (Table 3). Training of the AutoML software is fundamental to its performance and effectiveness and sufficient effort must be made to achieve this. The close range values of the predicted values with the actual values indicate the sufficiency of the model for this purpose. A multivariable linear model was chosen to give sufficient consideration of some of the key parameters guiding the process and for the sake of simplifying the complex process [15, 16]. Moreover, studies have shown that consideration of multiple variables improves model predictions [17] although the importance of maintaining a good start-up behavior for these model types have been highlighted and adequate care was taken to achieve this during the configuration process [10, 18].

Table 2 Product yield

Components	Product Yield (Vol %)			
	Actual	Predicted		
		AutoML	Jupyter	Python
LPG	1.5	1.41	0.93	1.66
SRG	5	4.96	4.69	7.58
SRN	18	17.87	17.24	11.68
SRK	14.5	14.5	14.39	14.92
LDO	27	26.47	26.43	24.77
HDO	3.5	2.7	2.7	4.59
AR	30.5	30.03	30.18	24.59

The statistical data (Table 3) highlights the precision of the model prediction to the actual yield. The Google AutoML Tables model gave the best coefficient of determination of 0.99981 an indication of how well this model fitted the actual yield. This model also had the least standard error value of 0.240918 nearest to zero. Jupyter model also had a high correlation of determination and low standard error value. However, the Python model had a coefficient of determination of 0.93078 and Standard Error of 3.536064. Hence, Google AutoML Tables model gave the best fit based on its R² value in line with the work of Ramachandran &Tsokos [19] whereas the standard error values shows that the Auto ML model gave more accuracy than the Jupyter and Python models, and therefore an indication that the Google AutoML Tables is an improved model. Further training of the model with the test dataset improved its prediction.

Iterations and training of the software to develop the best regression model was automatic. Additionally, all the identified variables were considered for the process. This high accuracy can be attributed to the high precision associated with automatic machine learning in agreement with the work of Al-Dunainawi & Abbod [2].

Table 3 Statistical Analysis

Model	R ²	SE
Auto ML	0.99981	0.240918
Jupyter	0.99943	0.419291
Python	0.93078	3.536064

The correlation of the Target with Temperature and Specific Gravity (Figs 6 & 7) observed through the line of best fit highlights a linear relationship and the fitness of these variables with the target. The feature importance (Fig. 3), however, shows the impact of these variables on the process and on the outcome of the prediction as shown by the VS Code software.

Figure 6 Target vs Temperature **Figure 7** Target Vs Sp. Gravity

4. Conclusion

The prediction model from the Auto ML software was successfully established as the most accurate method of prediction in this study based on the predicted yield and statistical analysis. The introduction of artificial intelligence, as proposed, would help mitigate losses during operation, improve process efficiency and enhance the process. Being able to predict what is to be expected before the actual process can go a long way in harnessing the accurate logistics planning of refining operations. In the advancement into the world of the big data and Internet of Things, the development of this method and incorporation of same in the downstream sector is imperative towards achieving this purpose.

Compliance with ethical standards

Acknowledgments

The authors are grateful to the management and staff of Ballantyne analytical research for their technical support

Disclosure of conflict of interest

The authors declare that there they have no competing interest.

References

- [1] K Liebmann, VR Dhole, M Jobson. Integrated design of a conventional crude oil distillation tower using pinch analysis, J. Chemical Engineering Research and Design. 1998; 76: 335-347.
- [2] Y Al-Dunainawi, MF Abbod. Hybrid intelligent approach for predicting product composition of distillation column, Int. J. Adv. Res. Artif. Intell. 2016; 5: 28-34.
- [3] H Stephen. BBC News. 2021.
- [4] N Srivastav, SL Agrwal, SK Gupta, SR Srivastava, B Chacko, H Sharma. Hybrid object detection using improved three frame differencing and background subtraction, 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence. 2017; 613-617.
- [5] J Frankenfield. Artificial intelligence (AI), J. Investopedia udgivet. 2019; 13: 6-19.
- [6] F Lardinois. Microsoft launches visual studio code, a free cross-platform code editor for os x, linux and windows. 2015.
- [7] MA Durrani, I Ahmad, M Kano, S Hasebe, An artificial intelligence method for energy efficient operation of crude distillation units under uncertain feed composition. J. Energies.2018; 11: 2993- 2997.
- [8] Y Chetouani. Detecting changes in a distillation column by using a sequential probability ratio test, J. Systems Engineering Procedia. 2011; 1: 473-480.
- [9] JK Prakash, DS Patle, AK Jana. Neuro-estimator based GMC control of a batch reactive distillation, J. ISA transactions. 2011; 50: 357-363.
- [10] TR Biyanto, BL Widjiantoro, AA Jabal, T Budiati. Artificial neural network based modeling and controlling of distillation column system. International J of Engineering, Science and Technology. 2010; 2: 177 – 188.
- [11] P Kathel, AK Jana. Dynamic simulation and nonlinear control of a rigorous batch reactive distillation, J. ISA transactions. 2010; 49: 130-137.
- [12] K Ramesh, SR AbdShukor, N Aziz. Nonlinear model predictive control of a distillation column using NARX model, J. Computer Aided Chemical Engineering. 2009; 27: 1575-1580.
- [13] V Singh, I Gupta, HO Gupta. ANN-based estimator for distillation using Levenberg–Marquardt approach, J. Engineering applications of artificial intelligence. 2007; 20: 249-259.
- [14] B Abbasi, T Babaei, Z Hosseinifard, K Smith-Miles, M Dehghani. Predicting solutions of large-scale optimization problems via machine learning: A case study in blood supply chain management, J. Computers & Operations Research. 2020; 119: 104941.

- [15] MA Durrani, A Avila, J Rafael, I Ahmad, March. An integrated mechanism of genetic algorithm and Taguchi method for cut-point temperatures optimization of crude distillation unit, international conference on computing, mathematics and Engineering Technologies. 2018; 1-6.
- [16] E Larsen, S Lachapelle, Y Bengio, E Frejinger, S Lacoste-Julien, A Lodi. Predicting tactical solutions to operational planning problems under imperfect information, J. arXiv preprint arXiv. 2018; 11807-76.
- [17] D Trierwiler, RL Tan. Advances in Crude Oil LP Modelling-Handbook of National Petrochemical & Refiners Association, Singapore. 2001; 52–58.
- [18] A Rogina, I Šiško, I Mohler, Ž Ujević, N Bolf. Soft sensor for continuous product quality estimation (in crude distillation unit), J. Chemical Engineering Research and Design. 2011; 89: 2070-77.
- [19] KM Ramachandran, CP Tsokos. Mathematical statistics with applications in R. Academic Press. 2020.