National Multidisciplinary Sciences
UMJember Proceeding Series (2023) Vol. 2, No. 3: 146-152

SEMARTANI 2

Machine Learning Approach to Design of Biodiesel Production Extraction Equipment from Tamanu Seed Oil

Achri Isnan Khamil[1] , Eko Saputra Widarianto[1], Anandya Zulham Valensyah[1], Maktum Muharja[1*], Rizki Fitria Darmayanti[2], Riza Umami[3], Khofifah Shinta Mamnukha[3], M. Zikrillah[1]

[1] Jember University; achriar@gmail.com, eko.8d.08@gmail.com, azulham03@gmail.com, maktum@unej.ac.id, 211910401004@mail.unej.ac.id
[2]University of Muhammadiyah Jember; r.f.darmayanti@gmail.com
[3] State University of Semarang; Umamiriza.mail@gmail.com, shinta.mamnukha@gmail.com

DOI: https://doi.org/10.32528/nms.v2i3.278
*Correspondensi: Maktum Muharja
Email: maktum@unej.ac.id

Published: Mei, 2023

Abstrak: Tamanu oil is one of the sources of vegetable oil needed on an industrial scale, so it is widely developed for its production. In designing essential oil processing on an in-dustrial scale requires mathematical calculations in measuring parameters in the re-actor, where this requires quite a long time in manual calculations. Utilization of ar-tificial intelligence technology, especially Machine Learning, That is very useful in the calculation process in the Tamanu oil extraction reactor. Machine Learning method used in this research is linear regression for data prediction and mean absolute error is used to measure absolute error. The model is used to predict the results of the outer diameter of the connecting pipe between tools with the parameters used, namely the target output in tons per year from cooking oil products made from tamanu seeds. R2 value of the graph of the CCIO Training Nominal Size data; Cooking oil; Methanol; Petroleum Ether is close to a value of 1, which means that the Input Feed value required to achieve the desired Output is close to 100% accuracy. Meanwhile, the Mean Absolute Error (MAE) value is the result of determining the Nominal Size with the Output value = 5000; 55000; and 150000 kg/hour shows a small difference in error values so that the design is acceptable.

Keywords: Tamanu seeds oil, Machine Learning, Linear Regression, Mean Absolute Error

PENDAHULUAN

Tamanu plant (Calophyllum inophyllum) used as vegetal oil source in Indonesia with 22% productivity of total amount mangrove tree in the world (Hapsari et al., 2019). This plant classified in Calophyllaceae famili and mangrove plant species which contain high vegetal oil between 40 – 70% (Gunawan et al., 2020). Another research shows 33,39% yield of vegetal oil in tamanu seed with 1,2% moisture content by using press extraction method (Fadhlullah et al., 2015). The prospect to be vegetal cooking oil due to Oleic Fatty Acid (OFA) contents and could be produced become biodiesel by reduce 19,82% Free Fat Acid (FFA) contents (Kurniati et al., 2018). Advanced research shows tamanu seed oil as a goods material for green bio-energy production (Ansori et al., 2019).

Biodiesel industry gain serious global concern due to limited source and global oil reserves. Hence, demand of alternative source expected could increase diverse material to replace crude oil (Agarwal et al., 2017). Industrial scale production solve the current circumtances about high demand for tamanu seeds oil (Damanik et al., 2017). Tamanu seeds oil extraction factory could be build to meet the current circumtances as for industrial scale of biodiesel production (Chahal & Gulia, 2019). There are many modern methods for extraction such as : Solvent extraction, Microwaved-assissted extraction, Ultrasound-assissted extraction, Supercritical-assissted extraction, Supercritical fluid extraction, Ionic liquids, Enzyme-assisted extraction, Pressurized liquid/fluid extraction that can be used in industrial process [9], (Geow et al., 2021). Paid software often used in biodiesel industry to predict outer diameter pipe and industrial simulation such as : Aspen HYSYS, DWSIM, Aspen Plus, CHEMCAD, dan PROSIM for pilot scale nor industrial scale [11], (Parvatker & Eckelman, 2019). Yet, expensive licensed and high requirements for device specification turn out to commercially used limitation (Taimoor, 2016).

Machine Learning (ML) approach can be used as an alternative method to assisst in industrial equipment design such as in oil and gas industry (Hanga & Kovalchuk, 2019; Lotfian et al., 2021; Makhotin et al., 2022; Sircar et al., 2021). The advantages for using ML including : the method is easy to operate, low cost operation, and standard device requirements. Google colab chosen as Integrated Development Environment (IDE) to compute python language because of simple usage and availability of library in it (Sharifi et al., 2015). Python language used because of similiarity with human language (Alves & Machado Vieira, 2019). The usage of ML also can predict outer diameter of pipe in industrial process (Maulud & Abdulazeez, 2020).

As described above, this research focus on possibility to predict outer diameter for pipe by using ML approach. The selected ML model are Linear regression and Mean Absolute Error (MAE). This design parameter including, input feed, output target, and outer diameter of pipe in process. Furthermore, comparison of result between ML prediction and conventional method for determine error value.

METODE

Tamanu plant (Calophyllum inophyllum) used as vegetal oil source in Indonesia with 22% productivity of total amount mangrove tree in the world (Hapsari et al., 2019). This plant classified in Calophyllaceae famili and mangrove plant species which contain high vegetal oil between 40 – 70% (Gunawan et al., 2020). Another research shows 33,39% yield of vegetal oil in tamanu seed with 1,2% moisture content by using press extraction method (Fadhlullah et al., 2015). The prospect to be vegetal cooking oil due to Oleic Fatty Acid (OFA) contents and could be produced become biodiesel by reduce 19,82% Free Fat Acid (FFA) contents (Kurniati et al., 2018). Advanced research shows tamanu seed oil as a goods material for green bio-energy production (Ansori et al., 2019).

Biodiesel industry gain serious global concern due to limited source and global oil reserves. Hence, demand of alternative source expected could increase diverse material to replace crude oil (Agarwal et al., 2017). Industrial scale production solve the current circumtances about high demand for tamanu seeds oil (Damanik et al., 2017). Tamanu seeds oil extraction factory could be build to meet the current circumtances as for industrial scale of biodiesel production (Chahal & Gulia, 2019). There are many modern methods for extraction such as : Solvent extraction, Microwaved-assissted extraction, Ultrasound-assissted extraction, Supercritical-assissted extraction, Supercritical fluid extraction, Ionic liquids, Enzyme-assisted extraction, Pressurized liquid/fluid extraction that can be used in industrial process [9], (Geow et al., 2021). Paid software often used in biodiesel industry to predict outer diameter pipe and industrial simulation such as : Aspen HYSYS, DWSIM, Aspen Plus, CHEMCAD, dan PROSIM for pilot scale nor industrial scale [11], (Parvatker

& Eckelman, 2019). Yet, expensive licensed and high requirements for device specification turn out to commercially used limitation (Taimoor, 2016).

Machine Learning (ML) approach can be used as an alternative method to assisst in industrial equipment design such as in oil and gas industry (Hanga & Kovalchuk, 2019; Lotfian et al., 2021; Makhotin et al., 2022; Sircar et al., 2021). The advantages for using ML including : the method is easy to operate, low cost operation, and standard device requirements. Google colab chosen as Integrated Development Environment (IDE) to compute python language because of simple usage and availability of library in it (Sharifi et al., 2015). Python language used because of similiarity with human language (Alves & Machado Vieira, 2019). The usage of ML also can predict outer diameter of pipe in industrial process (Maulud & Abdulazeez, 2020).

As described above, this research focus on possibility to predict outer diameter for pipe by using ML approach. The selected ML model are Linear regression and Mean Absolute Error (MAE). This design parameter including, input feed, output target, and outer diameter of pipe in process. Furthermore, comparison of result between ML prediction and conventional method for determine error value.

HASIL DAN PEMBAHASAN

ML model shows more convenient way to determine input feed for tamanu seed and nominal size of process pipe. The result shown by ML could be renewed by addition new source for datasets.

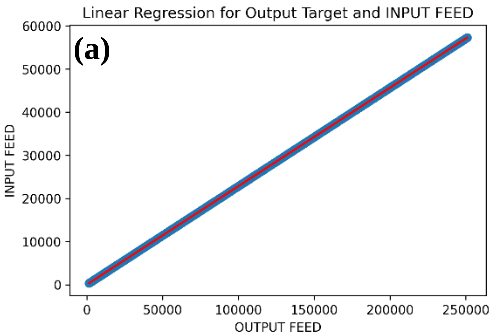

Figure 1a. Linear Regression for Output Target and Input Feed

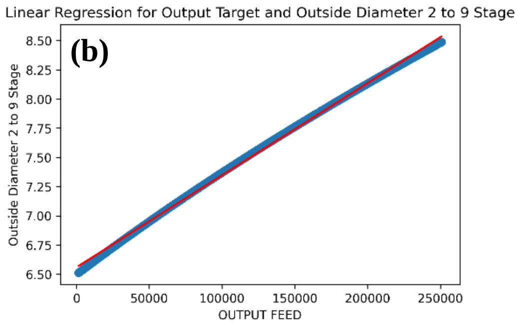

Figure 1b. Linear Regression for Output Target and Outside Diameter 2 to 9 Stage

As shown in Fig.1a, the correlation between output and input feed presented by an equation y = 0,228x – 1,09139e^{-11} and R^2 value 1. In Fig.1b, the correlation between output feed and nominal size of stage 2 to 9 pipe presented in y = 7,879e^{-06}x – 6,5596 and R^2 value 0,998. It shown by two parallel line that overlap each other.As for shown in fig.3a to fig.3d, each prediction shows same R2 value 0,953. Then, ML model predict the actual size of pipe by using MAE library as shown in table.1.

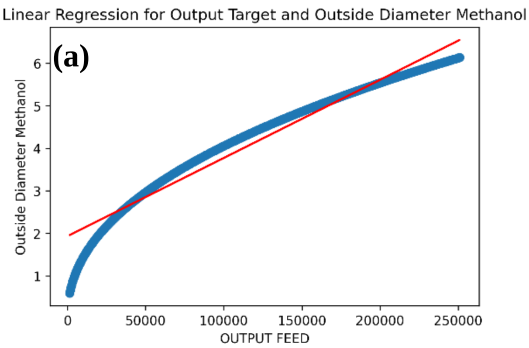

Figure 2a. Linear Regression for Output Target and Outside Diameter Methanol

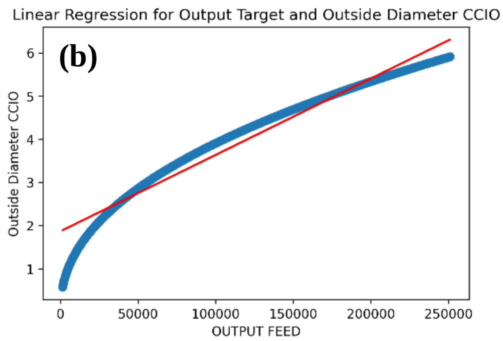

Figure 2b. Linear Regression for Output Target and Outside Diameter CCIO

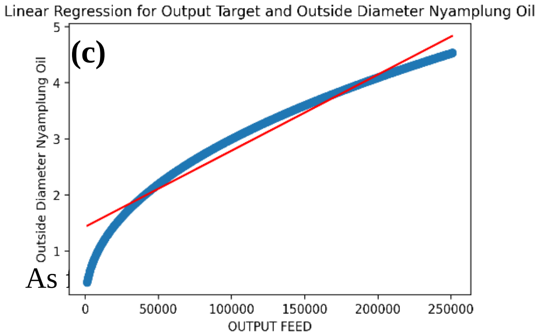

Figure 2c. Linear Regression for Output Target and Outside Diameter Tamanu Seed Oil

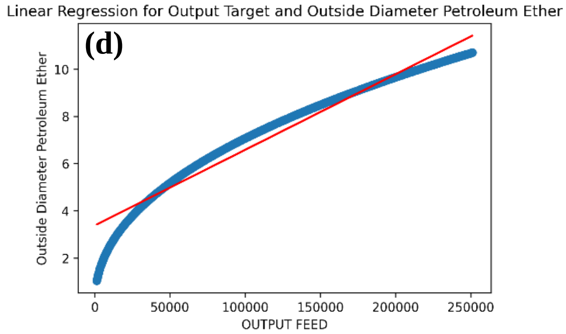

Figure 2d. Linear Regression for Output Target and Outside Diameter Petroleum Ether

MAE Result 5000 for Output Prediction			**MAE Result 55000 for Output Prediction**			**MAE Result 150000 for Output Prediction**		
Pipe type	**MAE Value**	**Pipe size**	**Pipe type**	**MAE Value**	**Pipe size**	**Pipe type**	**MAE Value**	**Pipe size**
Stage 2-9 pipe	0.599	6	Stage 2-9 pipe	0.992	6	Stage 2-9 pipe	0.258	8
CCIO Pipe	0.046	2	CCIO Pipe	0.161	3	CCIO Pipe	0.520	4
Cooking oil Pipe	0.496	1	Cooking oil Pipe	0.174	2	Cooking oil Pipe	0.462	3
Methanol pipe	0.027	2	Methanol pipe	0.053	3	Methanol pipe	0.692	4
PE Pipe	0.464	4	PE Pipe	0.861	6	PE Pipe	0.182	8

Table 1. MAE RESULT 5000, 55000, 150000 For Output Prediction

Afterwards comparisson between ML prediction and conventional method calculation are shown in table.2. By using output prediction for 5000, 55000, and 150000 per years, ML prediction present a splendid result by showing low error value at 55000 output prediction. However, ML prediction show high error value at 5000 output due to underfitting on ML model. Underfitting happens because of a low variance and high bias in linear regression library. This library used to process 2 interrelated variables, bias will occur in processing data if there's more interrelated variable in datasets (Austin & Steyerberg, 2015). Based on fig 2a. to fig 2d., machine learning model show high error prediction in less than output feed 50000 and more than 150000.

	Machine Learning predictions			**Conventional Calculations**			**Percentage of Error (%)**		
	5000	55000	150000	5000	55000	150000	5000	55000	150000
Nominal Size of Stage Pipe 2-9	6,546	6,995	7,763	6,599	6,993	7,742	0,806	0,034	0,282
Nominal Size of CCIO Pipe	1,015	2,986	4,690	1,954	2,839	4,521	92,470	4,928	3,604
Nominal Size of Cooking Oil Pipe	0,777	2,287	3,592	1,496	2,175	3,463	92,471	4,928	3,604
Nominal Size Methanol	1,053	3,099	4,868	2,028	2,946	4,692	92,470	4,928	3,604
Nominal Size PE	1,837	5,405	8,489	3,536	5,138	8,183	92,470	4,928	3,604

Table 2. Error Value Between ML Prediction and Conventional Method

SIMPULAN

ML model could be an alternative method for outside diameter of pipe prediction. These method shorten process time than conventional method. R2 value in ML prediction shown as high accuracy by value close to 1. By using input output prediction for 5000, 55000, and 150000 per years, high accuracy happens at 55000 output prediction. However, ML prediction show low accuracy below 50000 output prediction and over 55000 output prediction due to underfitting on ML model. Underfitting happens because linear regression library used to process 2 interrelated variables, bias will occur in processing data if there’s more interrelated variable in datasets. ML method could be more effective by using different library and input more variable and parameter. Simple model such as linear regression, SVM, Knearest Neighbour, K-Means, etc, are lack of accuracy to be used in industrial process prediction with many variables and parameters.

DAFTAR PUSTAKA

Agarwal, A. K., Gupta, J. G., & Dhar, A. (2017). Potential and challenges for large-scale application of biodiesel in automotive sector. Progress in Energy and Combustion Science, 61, 113–149. https://doi.org/10.1016/j.pecs.2017.03.002

Alves, F. R. V., & Machado Vieira, R. P. (2019). The Newton Fractal’s Leonardo Sequence Study with the Google Colab. International Electronic Journal of Mathematics Education, 15(2). https://doi.org/10.29333/iejme/6440

Ansori, A., Wibowo, S. A., Kusuma, H. S., Bhuana, D. S., & Mahfud, M. (2019). Production of Biodiesel from Nyamplung (Calophyllum inophyllum L.) using Microwave with CaO Catalyst from Eggshell Waste: Optimization of Transesterification Process Parameters. Open Chemistry, 17(1), 1185–1197. https://doi.org/10.1515/chem-2019-0128

Austin, P. C., & Steyerberg, E. W. (2015). The number of subjects per variable required in linear regression analyses. Journal of Clinical Epidemiology, 68(6), 627–636. https://doi.org/10.1016/j.jclinepi.2014.12.014

Chahal, A., & Gulia, P. (2019). Machine learning and deep learning. International Journal of Innovative Technology and Exploring Engineering, 8(12), 4910–4914. https://doi.org/10.35940/ijitee.L3550.1081219

Damanik, N., Ong, H. C., Chong, W. T., & Silitonga, A. S. (2017). Biodiesel production from Calophyllum inophyllum−palm mixed oil. Energy Sources, Part A: Recovery, Utilization and Environmental Effects, 39(12), 1283–1289. https://doi.org/10.1080/15567036.2017.1324537

De Monte, C., Carradori, S., Granese, A., Di Pierro, G. B., Leonardo, C., & De Nunzio, C. (2014). Modern extraction techniques and their impact on the pharmacological profile of Serenoa repens extracts for the treatment of lower urinary tract symptoms. BMC Urology, 14(1), 1–11. https://doi.org/10.1186/1471-2490-14-63

Fadhlullah, M., Widiyanto, S. N. B., & Restiawaty, E. (2015). The potential of nyamplung (Calophyllum inophyllum L.) seed oil as biodiesel feedstock: Effect of seed moisture content and particle size on oil yield. Energy Procedia, 68, 177–185. https://doi.org/10.1016/j.egypro.2015.03.246

Geow, C. H., Tan, M. C., Yeap, S. P., & Chin, N. L. (2021). A Review on Extraction Techniques and Its Future Applications in Industry. European Journal of Lipid Science and Technology, 123(4), 1–10. https://doi.org/10.1002/ejlt.202000302

Gholami, A., Pourfayaz, F., & Saifoddin, A. (2021). Techno-economic assessment and sensitivity analysis of biodiesel production intensified through hydrodynamic cavitation. Energy Science and Engineering, 9(11), 1997–2018. https://doi.org/10.1002/ese3.941

Gunawan, S., Aparamarta, H. W., Taufany, F., Prayogo, A., Putri, S. H. A., & Wijaya, C. J. (2020). Separation and purification of triglyceride from nyamplung (Calophyllum inophyllum) seed oil as biodiesel feedstock by using continuous countercurrent extraction. Malaysian Journal of Fundamental and Applied Sciences, 16(1), 18–22. https://doi.org/10.11113/mjfas.v16n1.1439

Hanga, K. M., & Kovalchuk, Y. (2019). Machine learning and multi-agent systems in oil and gas industry applications: A survey. Computer Science Review, 34, 100191. https://doi.org/10.1016/j.cosrev.2019.08.002

Hapsari, S., Susanto, D. F., Aparamarta, H. W., Widjaja, A., & Gunawan, S. (2019). Separation and purification of wax from nyamplung (Calophyllum inophyllum) seed oil. Materials Science Forum, 964 MSF, 1–6. https://doi.org/10.4028/www.scientific.net/MSF.964.1

Kurniati, S., Soeparman, S., Yuwono, S. S., & Hakim, L. (2018). Characteristics and Potential of Nyamplung (Calophyllum inophyllum L .) Seed Oil from Kebumen , Central Java , as a Biodiesel Feedstock. International Research Journal of Advanced Engineering and Science, 3(4), 148–152.

Lotfian, M., Ingensand, J., & Brovelli, M. A. (2021). The partnership of citizen science and machine learning: Benefits, risks and future challenges for engagement, data collection and data quality. Sustainability (Switzerland), 13(14). https://doi.org/10.3390/su13148087

Makhotin, I., Orlov, D., & Koroteev, D. (2022). Machine Learning to Rate and Predict the Efficiency of Waterflooding for Oil Production. Energies, 15(3). https://doi.org/10.3390/en15031199

Maulud, D., & Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. Journal of Applied Science and Technology Trends, 1(4), 140–147. https://doi.org/10.38094/jastt1457

Parvatker, A. G., & Eckelman, M. J. (2019). Comparative Evaluation of Chemical Life Cycle Inventory Generation Methods and Implications for Life Cycle Assessment Results. ACS Sustainable Chemistry and Engineering, 7(1), 350–367. https://doi.org/10.1021/acssuschemeng.8b03656

Sharifi, M., Yao, K. T., Raghavendra, S., Ershaghi, I., House, R., & Blouin, J. (2015). Prediction of remaining life in pipes using machine learning from thickness measurements. SPE Western Regional Meeting 2015: Old Horizons, New Horizons Through Enabling Technology, 48–56. https://doi.org/10.2118/173995-ms

Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., & Oza, H. (2021). Application of machine learning and artificial intelligence in oil and gas industry. Petroleum Research, 6(4), 379–391. https://doi.org/10.1016/j.ptlrs.2021.05.009

Taimoor, A. A. (2016). Virtualization of the process control laboratory using ASPEN HYSYS. Computer Applications in Engineering Education, 24(6), 887–898. https://doi.org/10.1002/cae.21758

Wang, W., & Lu, Y. (2018). Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model. IOP Conference Series: Materials Science and Engineering, 324(1). https://doi.org/10.1088/1757-899X/324/1/012049