

A Decision Tree Approach to Estimate the Microalgae Production in Open Raceway Pond

^{1,2}Supriyanto, ¹Ryozo Noguchi*, ¹Tofael Ahamed, ³Demura Mikihide, Makoto M. Watanabe

¹ Faculty of Life and Environmental Sciences, University of Tsukuba, Japan

² Department of Mechanical and Biosystem Engineering, Bogor Agricultural University, Bogor, Indonesia

³ Algae Biomass and Energy System Research and Development Centre, University of Tsukuba, Japan

corresponding author: noguchi.ryozo.gm@u.tsukuba.ac.jp

Abstract. Microalgae are one of the potential biomass to produce biofuel energy. The popular cultivation method of microalgae is using open raceway pond. The open raceway pond is waterway included with paddlewheel. Cultivation microalgae in open raceway pond using water, nutrient, CO₂, and maintaining the pH condition. To avoid the contamination from unwanted microalgae, the Algae Biomass and Energy System Research and Development (ABES), University of Tsukuba cultivated the polyculture microalgae to minimize the contamination that make unstable ecosystem in culture. The disadvantages of microalgae culture in open raceway pond is difficult to control the environmental parameter such as temperature and solar radiation. However, the environmental parameter is important in growing microalgae and potential to use as parameter to estimate the growth of microalgae. In This research, we proposed the growth estimation model of microalgae in Open Raceway Pond using the temperature and solar radiation parameter. The growth estimation model developed to estimate the production rate (g/m²/day) using the decision tree model using previous dataset from Minamisoma microalgae pilot plant. The result of this study indicated that decision tree algorithm able to estimate the microalgae production using the temperature and solar irradiance parameter with correlation coefficient 0.89 and RMSE 0.085.

Keyword: decision tree, microalgae, open raceway pond

1. Introduction

Microalgae are one of the potential biomass to produce bioenergy and environmentally friendly compared to fossil fuel [1][2]. The microalgae potential to reduce the carbon dioxide (CO₂) emission [3]. Compare to another plant; microalgae are higher productivity (per ha). The lipid productivity of microalgae is 4.5-7.5-ton ha⁻¹ y⁻¹ that estimated can produce 136,900 L ha⁻¹ oil [4], higher than another crop such as Soybean (446 L ha⁻¹) and Oil Palm (5,950 L ha⁻¹) [5]. The microalgae can utilize the fresh water, seawater, and wastewater [6]. The cultivation of microalgae in sewage commonly using the native polyculture microalgae to improve stability and productivity of biomass [7]. The Algae Biomass and Energy System (ABES) R&D centre, the University of Tsukuba as a member of Algae Industry Incubation Consortium (AAIC) built the microalgae pilot plant in Minamisoma, Fukushima Prefecture Japan [8]. The primary target of the pilot plant is to produce the biofuel. In this pilot plant, the open raceway pond used to cultivate the microalgae. In this pilot plant, the native polyculture microalgae



grown in ORP. The advantages of the polyculture microalgae are the stability of biomass production caused by the diversity of the species in one culture [9] [10].

The cultivation of microalgae in ORP is lower productivity than photobioreactor, but lower capital and operational cost [11]. The production optimization conducted by adding the sufficient nutrient, injected the additional CO₂ and enough solar radiation to improve the microalgae productivity. However, the ORP is easy to contaminate with unwanted microalgae and another organism. Utilization the polyculture microalgae is one of the solution to minimizing the contaminant from unwanted microalgae. The polyculture make the microalgae culture more stable like natural ecosystem. However, the temperature and solar irradiance are fluctuated seasonally, also affected to the microalgae growth [12] [13]. In ORP, solar radiation and temperature is difficult to control.

However, the temperature and solar radiation parameter are potential to estimate the microalgae growth. In This research, we proposed the growth estimation of microalgae in Open Raceway Pond using the temperature and solar radiation parameter. Previously, the kinetics models are commonly used in microalgae population dynamics but only focus on monoculture microalgae [14]. In this research, we proposed the decision tree model to estimate the polyculture microalgae in Open Raceway Pond (ORP). Machine learning method are algorithm that allow computer to develop model from previous dataset (training set). Our unique contribution solving the estimation model for complex combination of species in culture seasonally. The model developed using the machine learning model (decision tree), and environmental parameter (solar irradiance and temperature) to estimate the microalgae production in ORP.

2. Material and Method

2.1. Study Area

In this research data collected from pilot plant microalgae cultivation in Minamisoma, Fukushima Prefecture, Japan (37.644673°N, 141.009490°E). The location of microalgae cultivation shown in Figure 1. The large area microalgae cultivation facility was builds to demonstrate the microalgae fuel production. The conversion method to crude using Hydrothermal Liquefaction (HTL) process.



Figure 1. Microalgae Pilot Plant in Minamisoma, Fukushima Prefecture, Japan

2.2. Data Collection

The data collected from the pilot scale ORP facilitated with the paddlewheel for mixing the culture. The cultivation in 1 m² area, 120 L volume, and 0.3 m depth of culture. The cultivation method using the semi-continuous culture with harvesting rate was 30 L day⁻¹. The nutrient added in sufficient condition and additional CO₂ injected to the culture. The microalgae species used in this culture was polyculture consist of *Scenedesmus* spp., *Desmodesmus* spp., *Dictyosphaerium* spp., and *Klebsormidium* sp [6]. The Figure 2 shown the schematic of the ORP used in this culture.

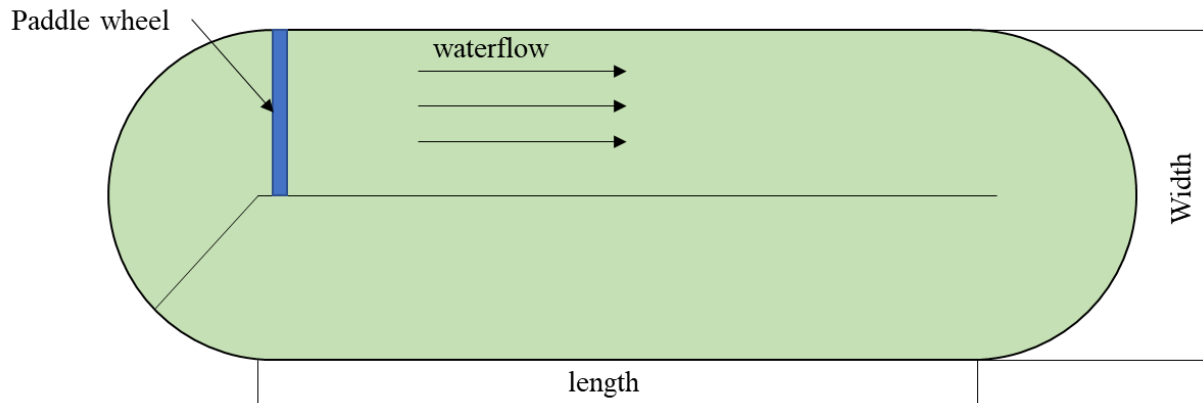


Figure 2. The schematic of the Open Raceway Pond

2.3. Model Development

For developing the estimation model using the tree model, data from the cultivation pond collected during January to December 2017. The water temperature (T) in °C, solar irradiance (I) in $\mu\text{mol s}^{-1} \text{m}^{-2}$ and production rate of microalgae (P) in $\text{g m}^{-2} \text{day}^{-1}$ were collected during culture. The temperature and solar irradiance chosen caused by dynamically change during culture. The temperature and solar irradiance collected using the sensor. The volumetric production ($\text{g m}^{-2} \text{day}^{-1}$) calculated by the Dry Cell Weight (DCW) during culture. Table 1 shown the dataset used in this research.

2.3.1. Data Preparation

In this step data that was collected were checked the consistency, calculated the production rate in area ($\text{g m}^{-2} \text{day}^{-1}$). Finally, the data reformat into the input-output form. The input variable used in this research was Temperature (T) and Solar Irradiance (I) for the input and Production Rate (P) for the output variable. Table 1 shown the dataset used in in this research.

Table 1. The dataset for the microalgae cultivation in Open Raceway Pond (ORP)

No	Day of year		Average Outdoor Temperature (°C)	Solar Irradiance ($\mu\text{mol s}^{-1} \text{m}^{-2}$)	Volumetric Productivity ($\text{g L}^{-1} \text{day}^{-1}$)
	Begin	Harvest			
1	2	9	4.03	98	1.26
2	9	16	2.08	93	1.03
3	16	23	2.58	109	0.71
4	23	30	2.91	128	0.77
5	30	37	4.5	137	1.11
...
89	359	363	3.98	85	2.83

2.3.2. Model Development

Machine learning algorithm is a process to use the dataset to fit a model through training or learning. The learning model will observe the relation between the input and output in several dataset to generate the model [15]. The decision tree is one of the machine learning algorithms that usually used by researchers in many application such as agriculture [16], hydrology [17], and energy [18]. In this research, Random Forest (RF) tree used to train the dataset to build the model. The Random Forest (RF) introduced by Breiman et al., (1984) [19][20] [21].

The RF algorithm is consist of several activity such as [22]:

1. Data preparation
2. Creating the tree using the Algorithm 1.
 - a. The *createNode* () is a function to create the tree.
 - b. The *find_best_split* () is a function to determine the attribute should be selected as the test condition for split the training data.
 - c. The *classify* () is a function to determine the class determined to a leaf node.
 - d. The *stopping_Cond* () is functioning to terminate the growing tree.

3. Evaluation

Algorithm 1. Decision tree Induction

ThreeGrowth (T,I; P)

```

1: if stopping_Cond (T,I;P) = true then
2:   leaf = createNode ().
3:   leaf.label = Classify (T,I)
4:   return leaf.
5: else
6:   root = createNode ()
7:   root.test_cond = find_best_split (T,I; P).]
8:   let V = {v|v is a possible outcome of root.test_cond}
9:   for each v do
10:    Ev = {e| root.test_cond (e) = v and e ∈ T,I }
11:    child = TreeGrowth (T,Iv; P).
12:    add child as descendent of root and label the edge (root → child) as v
13:   end for
14: end if
15: return root.

```

2.3.3. Model Validation (Evaluation)

In this step model evaluated using Root Mean Squared Error (RMSE) and coefficient correlation (R^2). The prediction accuracy calculated for each training sample set using Equation 1.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \gamma_i)^2} \quad (1)$$

where: MSE is Mean Square Error, n is number of sample, y_i is the actual number, γ_i is predicted number.

3. Result and Discussion

The Random forest algorithm in WEKA 3.9 software was used to develop the tree model. *At first*, data were reformatting and prepare in input variable (T and I) and output (P: g L⁻¹ day⁻¹). *Second*, the dataset was trained using Random Forest (RF) algorithm. The algorithm was built a tree model based on the training dataset. *Third*, the output of the training is a decision tree with 25 decision output. The size of the tree was 49 leaf's with the depth of tree is 6. The tree result from this training shown in Table 2.

Table 2. The Tree Result Using Random Forest

T < 16
I < 61 : 0.5 (3/0)
I ≥ 61
T < 4
I < 105
T < 3 : P = 1 (1/0)
T ≥ 3 : P = 3.2 (1/0)
I ≥ 105 : P = 1 (3/0)
T ≥ 4
I < 257
I < 108
T < 5 : P = 2.3 (3/0)
T ≥ 5 : P = 3.3 (9/0)
I ≥ 108
I < 145 : P = 1.9 (4/1)
I ≥ 145 : P = 3.3 (11/0)
I ≥ 257
T < 14 : P = 3.7 (1/0)
T ≥ 14 : P = 4.2 (3/0)
T ≥ 16
T < 25
T < 23
T < 22
T < 21
I < 209 : P = 4.6 (9/1)
I ≥ 209 : P = 3.7 (12/0)
T ≥ 21
I < 169 : P = 2.6 (3/0)
I ≥ 169 : P = 3.5 (2/0)
T ≥ 22
I < 146 : P = 4.7 (1/0)
I ≥ 146
I < 222 : P = 6.8 (3/-0)
I ≥ 222 : P = 3.5 (2/0)
T ≥ 23
I < 96 : 4 (1/1)
I ≥ 96
I < 178
I < 135 : P = 2.5 (1/0)
I ≥ 135 : P = 3.1 (3/0)
I ≥ 178
I < 207 : P = 1.9 (1/0)
I ≥ 207 : P = 2.4 (3/0)
T ≥ 25
I < 291
I < 222 : P = 6.1 (4/0)
I ≥ 222 : P = 5.3 (2/0)
I ≥ 291 : 4 P = (3/0)

The performance of model was evaluated using coefficient correlation (R^2) and RMSE. The result of the correlation coefficient (R^2) and RMSE was 0.89 and 0.085, respectively. The training was performed in 89 iterations. Based on this result, the correlation coefficient is good enough for predicting the areal growth of microalgae in Open Raceway Pond. In this research, we assumed that another

condition such as pH and nutrient in good condition. In this research also assumed the polyculture condition is good condition and not count the composition of microalgae in culture.

The advantages of using this tree models are: Easy to understand by expert, easy to evaluate the tree, expert can modify the tree, and easy to use if the previous data is available. The disadvantages of the tree algorithm are cannot build without sufficient dataset. The prediction of future condition based on the previous condition that represented as dataset for training and validation.

4. Conclusion and Future Work

Based on the experiment, we concluded that the RF CART algorithm able to use for estimating the microalgae production ($\text{g L}^{-1} \text{ day}^{-1}$) using the temperature parameter. The correlation coefficient was 0.89 and RMSE 0.085, that's indicate that the model has the good performance. The result of the tree is easy to use by expert to evaluate and improve the model that build from dataset during training phase. The advantages of the machine learning method are easy to use if the time series dataset from previous experiment is available. In the future, addition more dataset with several condition is essential to improve the model capability. The modification of model or utilization another model may improve the accuracy of the model.

References

- [1] R. P. Rastogi, A. Pandey, C. Larroche, and D. Madamwar, "Algal Green Energy – R&D and technological perspectives for biodiesel production," *Renew. Sustain. Energy Rev.*, vol. 82, pp. 2946–2969, Feb. 2018.
- [2] E. S. Shuba and D. Kifle, "Microalgae to biofuels: 'Promising' alternative and renewable energy, review," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 743–755, Jan. 2018.
- [3] S. Banerjee and S. Ramaswamy, "Dynamic process model and economic analysis of microalgae cultivation in open raceway ponds," *Algal Res.*, vol. 26, pp. 330–340, Sep. 2017.
- [4] R. P. Rastogi, A. Pandey, C. Larroche, and D. Madamwar, "Algal Green Energy – R&D and technological perspectives for biodiesel production," *Renew. Sustain. Energy Rev.*, no. October, pp. 1–24, 2017.
- [5] Y. Chisti, "Biodiesel from microalgae," *Biotechnol. Adv.*, vol. 25, pp. 294–306, 2007.
- [6] P. J. le B. Williams and L. M. L. Laurens, "Microalgae as biodiesel & biomass feedstocks: Review & analysis of the biochemistry, energetics & economics," *Energy Environ. Sci.*, vol. 3, no. 5, p. 554, 2010.
- [7] P. Kumar, S. K. Prajapati, A. Malik, and V. K. Vijay, "Cultivation of native algal consortium in semi-continuous pilot scale raceway pond for greywater treatment coupled with potential methane production," *J. Environ. Chem. Eng.*, vol. 5, no. 6, pp. 5581–5587, 2017.
- [8] M. Demura *et al.*, "Biomass productivity of native algal communities in Minamisoma city, Fukushima Prefecture, Japan," *Algal Res.*, vol. 29, pp. 22–35, Jan. 2018.
- [9] A. A. Corcoran and W. J. Boeing, "Biodiversity Increases the Productivity and Stability of Phytoplankton Communities," *PLoS One*, vol. 7, no. 11, pp. 1–9, 2012.
- [10] J. B. Shurin *et al.*, "Industrial-strength ecology: Trade-offs and opportunities in algal biofuel production," *Ecology Letters*, vol. 16, no. 11, pp. 1393–1404, 2013.
- [11] J. W. Richardson, M. D. Johnson, X. Zhang, P. Zemke, W. Chen, and Q. Hu, "A financial assessment of two alternative cultivation systems and their contributions to algae biofuel economic viability," *Algal Res.*, vol. 4, no. 1, pp. 96–104, 2014.
- [12] M. El-Sheekh, A. El-Fatah Abomohra, M. Abd El-Azim, and R. Abou-Shanab, "Effect of temperature on growth and fatty acids profile of the biodiesel producing microalga *Scenedesmus acutus*," *Biotechnol. Agron. Soc. Env.*, vol. 21, no. 4, pp. 233–239, 2017.
- [13] Q. Béchet, P. Moussion, and O. Bernard, "Calibration of a productivity model for the microalgae *Dunaliella salina* accounting for light and temperature," *Algal Res.*, vol. 21, pp. 156–160, 2017.
- [14] A. Solimeno and J. García, "Microalgae-bacteria models evolution: From microalgae steady-state to integrated microalgae-bacteria wastewater treatment models – A comparative review,"

- Sci. Total Environ.*, vol. 607–608, pp. 1136–1150, Dec. 2017.
- [15] E. Witten, I. H. and Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, no. 1. 2014.
- [16] K. Tatsumi, Y. Yamashiki, M. A. Canales Torres, and C. L. R. Taípe, “Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data,” *Comput. Electron. Agric.*, vol. 115, pp. 171–179, 2015.
- [17] J. Shiri, “Improving the performance of the mass transfer-based reference evapotranspiration estimation approaches through a coupled wavelet-random forest methodology,” *J. Hydrol.*, vol. 561, pp. 737–750, Jun. 2018.
- [18] D. Assouline, N. Mohajeri, and J. L. Scartezzini, “Large-scale rooftop solar photovoltaic technical potential estimation using Random Forests,” *Appl. Energy*, vol. 217, pp. 189–211, 2018.
- [19] J. R. Quinlan, “Induction of Decision Trees,” *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.
- [20] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, vol. 19. 1984.
- [21] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [22] C.-C. Yeh, D.-J. Chi, and Y.-R. Lin, “Going-concern prediction using hybrid random forests and rough set approach,” *Inf. Sci. (Ny)*, vol. 254, pp. 98–110, 2014.