
Taxation of Palm Oil Production Results Using on Regression Model (PT. Menthobi Makmur Sustainable)

Wandha Atmaka Aji^{1*}, Bagus Nugroho¹, Toyyibah¹²

¹ Akademi Komunitas Perkebunan Yogyakarta, Indonesia

² Magister of Agricultural Economics, Universitas Gadjah Mada Yogyakarta, Indonesia

*Corresponding author: wandha.aji@gmail.com

Article History: Received: 21 December 2022, Final Revision: 3 January 2023, Accepted: 10 January 2023, Online Publication: 20 January 2023

ABSTRACT

PT Menthobi Makmur Lestari is a growing industrial company in the field of palm oil production. The company is targeting increased palm oil production to forecast capacity plans and manufacturing facilities. One of the prediction methods used is multiple linear regression. The free variables used for prediction are the age of the tree, land area (ha), number of trees, number of vines, and the bound variable is the yield of oil palm production. The results of the correlation test using multiple regression showed a significant correlation of 0.05 or less. Hypothesis testing includes multiple linear regression and correlation using the t test and f test with a significance level of $\alpha = 0.05$. The value of multiple correlation analysis (R) is 0.947 and the coefficient of determination is 90%. The performance of the multiple regression equation is the accuracy of predictions with an average absolute error percentage (MAPE) of 21%, which is formed from the validation of training and testing data.

Keywords: Palm Oil Production Prediction, Multiple Linear Regression

INTRODUCTION

The demand for palm oil as a vegetable oil continues to increase worldwide [1]. This is because palm oil is not only for human consumption, but also for use as fuel and as a raw material in the chemical industry [2]. This increase in global consumption has resulted in continuous land expansion [3]. There are several factors that affect the productivity of oil palm crops, namely climate, regional shape, soil conditions, planting material, and cultivation techniques [4]. That the age of the plant, the number of plant populations per hectare, the pollination system, the coordination system of harvest-processing, and so on also affect the productivity of oil palm [5]. To

support these efforts, a technique is needed such as supervising and determining policies in predicting future palm oil production results [6]. Data mining contains the search for desired trends or patterns in a large database to assist in making decisions in the future [7]. One of the algorithms in data mining is Multiple Linear Regression which will be used in predicting palm oil production yields and finding patterns of relationships between factors that affect production yields for the next year [8]. Based on the discussion that has been previously presented, this study will estimate the yield of palm oil production and look for patterns of relationships between factors that affect palm oil production

using the multiple linear regression method [9]. Therefore, this study raised a theme related to "the application of the multiple Linear Regression method in predicting palm oil production results" [10].

Palm oil is an important source of vegetable oil [11]. The use of palm oil has begun since the 15th century, while for marketing to Europe began in the 1800s [12]. The palm oil used comes from the flesh of the fruit (mesocarp) and from the palm

MATERIALS AND METHODS

The following research method contains the stages of the research flowchart carried out

Predictions

Prediction is a process of systematically estimating something that is most likely to happen in the future based on past and present information owned, so that the difference between something that happened and the forecast result) can be reduced [16]. According to the big Indonesian dictionary, prediction is the result of the activity of predicting or predicting or estimating future values using past data [17]. Predictions show what will happen in a given situation and are inputs to the planning and decision-making process.

Multiple Linear Regression

In the case of multiple regression, there is one dependent variable and more than one independent variable. An analysis that has more than one free variable is called multiple Linear Regression analysis. The multiple Linear Regression technique is used to determine whether or not there is a significant influence of two or more free variables ($X_1, X_2, X_3, \dots, X_k$) on the bound variable (Y).

Mathematically the equation Multiple

Linear Regression can be written on equation 1:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_nX_n$$

Information:

Y = Predicted variable (variable dependent)

$X_1, X_2, X_3, \dots, X_n$ = Known variables

(variable Independent)

$b_1, b_2, b_3, \dots, b_n$ = Regression coefficient

kernel or kernel (endosperm) [13]. The increasing state of the world's population has an impact on the demand for CPO (crude palm oil) which is also increasing rapidly [14]. To meet these needs, several countries, especially Indonesia, are increasing oil palm production through the expansion of oil palm plantations throughout Indonesia [15].

After obtaining F calculate, the next comparison with F table how to find F table is by the formula :

$$F_{tabel} = \frac{K}{n-K-1}$$

Information:

K = free variable

n = number of observations

F = F table on the adjusted α

Criterion:

H_0 is accepted when F counts $< F$ table

H_0 is rejected when F counts $> F$ table

That is to say: if F counts $< F$ table, then the double correlation coefficient tested is insignificant, but vice versa if F calculates $> F$ the table then the double correlation coefficient tested is significant and shows that there is a simultaneous influence.

After obtaining the calculated T , looking for the T table can use the formula:

$$T_{tabel} = n - 2$$

Information:

K = free variable

n = many samples

t = t table on customized α

Test Criteria

H_0 is accepted if $-t$ table $< t$ calculate $< t$ table

H_0 is rejected if $-t$ count $< -t$ table or t count $> t$ table

This means that if t count $> t$ table or $-t$ count $< -t$ table then the partial correlation coefficient is significant(real) and indicates the existence of a partial influence between the bound variable and

the free variable, and vice versa.

RESULTS AND DISCUSSION

Data collection is carried out by conducting interviews and secondary data collection. The data used in this study is data on palm oil production at the Kujan Estate for the last 10 years precisely from 2009 to 2019. And there are as many as 450 records. Independent attributes, namely age, principal amount, number of bunches while dependent is the production of palm oil. Furthermore, predictions will be made to find patterns of relationships between variables for

palm oil production in 2019.

Next, do the cleaning or cleaning of data. Data cleaning is done to reduce noise or confusion that can affect calculations. The initial data amounted to 460 records after the data cleaning process was carried out, then it became as many as 450 records. The following are shown some data attachments from the 450 data records presented in table 1 below:

No	Age (X1)	Area (Ha) (X2)	Number of Trees (X3)	Number of Bunches (X4)	Yield (Kg) (Y)
1	5	24	3.209	6.120	8,5
2	5	10	1.534	2.643	7,3
3	5	12	1.658	2.920	7,5
4	5	4	798	1.049	9,0
.					
.					
.					
450	14	25	3.300	6.897	6,6

In this study, the *dependent* variable is the yield of palm oil production (Y). While *the independent* variables are age (X1), area (Ha) (X2), principal amount (X3), number of bunches

(X4). A description of the data regarding the lowest score, highest score, mean, standard deviation (SD), and number of data (N) can be seen in table 2 below: Table 2. *Descriptive Statistics*

Statistics

		Yield	Age	Area	Number of Trees	Number of Bunches	BJR
N	Valid	495	495	495	495	495	495
	Missing	0	0	0	0	0	0
Mean		39.1207	10.0000	17.5778	144.5957	75.2656	5.9636
Std. Deviation		88.72585	3.16548	7.46852	285.74647	211.13333	1.92330

To find the regression equation, we first calculate its regression coefficients by looking for the multiplication of a variable by another variable.

Here are the results of the Multiple Linear Regression coefficients in table 3 as follows:

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients		Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance VIF
1	(Constant)	87.863	25.360		3.465	.001	
	Age	-.962	1.294	-.034	-.743	.458	.846 1.182
	Area	1.191	.655	.100	1.818	.070	.593 1.686
	Number of Trees	.060	.017	.192	3.450	.001	.581 1.722
	Number of Bunches	.033	.019	.080	1.758	.079	.882 1.133
	BJR	-11.940	2.235	-.259	-5.341	.000	.768 1.302

Dependent Variable: Yield

Production Results The regression equation is as follows:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5$$

$$Y = 87.863 - 0.962X_1 + 1.191X_2 + 0.060X_3 + 0.033X_4 - 11.940X_5$$

Results A significant correlation between free variables and production yields is worth 0.00. Which means that each free variable has an Model Summary^b

influence on the bound variable, namely the yield of palm oil products. The results of the regression analysis can be seen in table 4 below:

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.343 ^a	.118	.109	83.75774	1.156

a. Predictors: (Constant), BJR, Number of Bunches, Area, Age, Number of Trees

b. Dependent Variable: Yeild

Based on the table above, an R number of 0.343 is obtained. This shows that there is a very strong relationship between age, area, principal

amount, and number of bunches to palm oil production

From the results of the regression analysis, see that the output of the *summary model* obtained the number R^2 (*R Square*) of 0.118 or 12%. *Standard Error* of the Estimate is a measure of the number of errors of the regression model in predicting the value of Y. From the regression results in the value of 83.75774 (units of production results), this means that the number of errors in the prediction of kernel production is

83.75 kg.

The performance of multiple linear regression models in the prediction of palm oil production yields resulted in a level of accuracy with a *Mean Absolute Percentage Error* (MAPE) value of 21%.

Testing the hypothesis, the first step taken is to see if F counts. This F test is carried out to see whether independent variables together affect

dependent variables. The results of the F test in the study can be seen in table 5 below: Table 5. F Test Results ANOVAb

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	458393.461	5	91678.692	13.068	.000 ^a
	Residual	3430510.737	489	7015.359		
	Total	3888904.198	494			

a. Predictors: (Constant), BJR, Number of Bunches, Area, Age, Number of Trees

b. Dependent Variable: Yeild

The calculated F value is 13.068 with a *significant* value of 0.000 less than the significance level of 0.05. Using a 95% confidence level, $\alpha = 5\%$ is obtained for F of 2.4. F values count > F table (13.068>.....), then H₀ is rejected. Since F

counts > F table (13.068> 2.4), then H₀ is rejected, meaning that there is a significant influence between age, area, principal amount, and number of marks together on palm oil production.

CONCLUSION

Based on the results of research conducted at PT Menthobi Makmur Lestari using satsitic analysis of multiple regression, the regression equation $Y = 33548.405 - 3036.110X_1 + 16642.878X_2 + 26,069X_3 - 0.246X_4$ was obtained. The resulting regression equation is positive if the variable increases by 1% on the contrary, if it is negative, if it decreases by 1%. From the results of the hypothesis test above, it can

be seen that both simultaneously and partially the Independent variable has a significant effect on the Dependent variable. The next research can predict the results of palm oil production using other methods, in this study using poor datasets so that the regression prediction results have a high error rate of 21% can be seen from the prediction results in the actual data.

REFERENCES

- [1] W. A. Aji, A. D. Utami, and B. Nugroho, "FERTILIZATION EFFECTIVENESS ON THE PRODUCTIVITY OF FRESH FRUIT," vol. 4270, 2020.
- [2] W. A. Aji, S. A. Saloko, I. Kurniawan, and M. A. Nurcahya, "The Effect of Rainfall and Rain Days On Palm Oil Production," *Trop. Plant. J.*, vol. 1, no. 1, pp. 16–20, 2022, doi: 10.56125/tpj.v1i1.3.
- [3] K. Setiawan and Hartono, "EFEK EKSTRAK ALELOPATI TERHADAP PEMBIBITAN KELAPA SAWIT (PRE NURSERY)," *J. AHLI MUDA Indones.*, vol. 1, no. 51, 2020, doi: <https://doi.org/10.46510/jami.v1i1.10>.
- [4] W. Y. Wong, S. Lim, Y. L. Pang, S. H. Shuit, W. H. Chen, and K. T. Lee, "Synthesis of renewable heterogeneous acid catalyst from oil palm empty fruit bunch for glycerol-free biodiesel production," *Sci. Total Environ.*, vol. 727, p. 138534, 2020, doi: 10.1016/j.scitotenv.2020.138534.
- [5] M. A. Heredia Salgado, J. A. Coba S, and L. A. C. Tarelho, "Simultaneous production of biochar and thermal energy using palm oil residual biomass as feedstock in an auto-thermal prototype reactor," *J. Clean. Prod.*, vol. 266, p. 121804, 2020, doi: 10.1016/j.jclepro.2020.121804.
- [6] A. A. Lawal *et al.*, "One-step steam pyrolysis for the production of mesoporous biochar from oil palm frond to effectively remove phenol in facultatively treated palm oil mill effluent," *Environ. Technol. Innov.*,

- vol. 18, p. 100730, 2020, doi: 10.1016/j.eti.2020.100730.
- [7] P. R. Furumo, X. Rueda, J. S. Rodríguez, and I. K. Parés Ramos, "Field evidence for positive certification outcomes on oil palm smallholder management practices in Colombia," *J. Clean. Prod.*, vol. 245, no. xxxx, 2020, doi: 10.1016/j.jclepro.2019.118891.
- [8] A. A. Hanieh, A. Hasan, and M. Assi, "Date Palm Trees Supply Chain and Sustainable Model," *J. Clean. Prod.*, 2020, doi: 10.1016/j.jclepro.2020.120951.
- [9] W. N. S. Mohd Luthfi, H. Alias, G. S. Tay, and C. K. Lee, *Production and characterization of bioflocculant via solid state fermentation process using oil palm empty fruit bunch fibers as substrate*, vol. 23. Elsevier Ltd, 2020.
- [10] [10] B. K. Zaied, M. Nasrullah, M. N. Islam Siddique, A. W. Zularisam, L. Singh, and S. Krishnan, "Enhanced bioenergy production from palm oil mill effluent by co-digestion in solar assisted bioreactor: Effects of hydrogen peroxide pretreatment," *J. Environ. Chem. Eng.*, vol. 8, no. 2, p. 103551, 2020, doi: 10.1016/j.jece.2019.103551.
- [11] S. K. Loh, M. E. Lai, and M. Ngatiman, "Vegetative growth enhancement of organic fertilizer from anaerobically-treated palm oil mill effluent (POME) supplemented with chicken manure in food-energy-water nexus challenge," *Food Bioprod. Process.*, vol. 117, pp. 95–104, 2019, doi: 10.1016/j.fbp.2019.06.016.
- [12] S. Raharja *et al.*, "Institutional strengthening model of oil palm independent smallholder in Riau and Jambi Provinces, Indonesia," *Heliyon*, vol. 6, no. 5, May 2020, doi: 10.1016/j.heliyon.2020.e03875.
- [13] V. P. Singh and L. Zhang, "Copula–entropy theory for multivariate stochastic modeling in water engineering," *Geosci. Lett.*, vol. 5, no. 1, 2018, doi: 10.1186/s40562-018-0105-z.
- [14] A. Promraksa and N. Rakmak, "Biochar production from palm oil mill residues and application of the biochar to adsorb carbon dioxide," *Heliyon*, vol. 6, no. 5, May 2020, doi: 10.1016/j.heliyon.2020.e04019.
- [15] P. Panen Kelapa *et al.*, "Harvest Management of Oil Palm (*Elaeis guineensis* Jacq.) in Pelantaran Agro Estate, Kota Waringin Timur, Center Borneo," 2017.
- [16] R. Dong *et al.*, "Oil palm plantation mapping from high-resolution remote sensing images using deep learning," *Int. J. Remote Sens.*, vol. 41, no. 5, pp. 2022–2046, 2020, doi: 10.1080/01431161.2019.1681604.
- [17] C. Mirande-Ney *et al.*, "Metabolic leaf responses to potassium availability in oil palm (*Elaeis guineensis* Jacq.) trees grown in the field," *Environ. Exp. Bot.*, vol. 175, 2020, doi: 10.1016/j.envexpbot.2020.104062.