

## Calibration of Proximate Content of Arabica Sidikalang Green Bean Coffee Using NIRS and PLS

Melva Lianur Ritonga, Achwil Putra Munir, Adian Rindang and Putri Chandra Ayu

Department of Agricultural and Biosystems Engineering, Universitas Sumatera Utara, Medan, 20155, Indonesia

\*Corresponding Author: [pputricandra@usu.ac.id](mailto:pputricandra@usu.ac.id)

### ARTICLE INFO

#### Article history:

Received 06 September 2022

Revised 03 November 2023

Accepted 25 August 2023

#### How to cite :

M L Ritonga, A P Munir, A Rindang and P C Ayu, "Calibration of Proximate Content of Arabica Sidikalang Green Bean Coffee Using NIRS and PLS", Journal of Sustainable Agriculture and Biosystem Engineering, Vol. 01, No. 01. 2023.



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International.

<http://doi.org/10.26594/register.v6i1.Lidarticle>

### ABSTRACT

Generally, coffee bean's content measurement is conducted destructively, where it is known to be less effective. Measurement using the Near Infrared Spectroscopy (NIRS) method could be an alternative to a rapid determination of the chemical content non-destructively. The study aimed to obtain a partial least square (PLS) calibration model of water, lipid and carbohydrate content in Sidikalang Arabica green bean coffee. This method can predict the chemical content assisted with several data pretreatments, namely multiple scatter correction (MSC), standard normal variative (SNV), normalization, first derivative (dg1), second derivative (dg2), combination of dg 1 + MSC and combination of dg2 + MSC. Results showed that the best calibration model for water content was obtained using normalization with 6 PLS factor with r value of 0.574. moreover, the best calibration model in lipid content was obtained using MSC with 5 PLS factor with r value of 0.647, and the best calibration model for carbohydrate was using MSC with 6 PLS factors with r value of 0.563. The model obtained from this study is categorized into a moderate model and could be use in simulations with predictive values with moderate accuracy.

**Keywords:** NIRS, PLS, pretreatment method, Sidikalang Coffee

## 1. Introduction

Sidikalang is an area in the highland in the Dairi Regency, Sumatera Utara Province with low humidity. Sidikalang has an average altitude of 700 – 1,100 m above sea level (asl). The topography of this area consists of mountains, hills and cool air which make it suitable for coffee plantation. Nowadays, Sidikalang coffee is very well known because it has a distinctive and bitter taste.

Coffee quality is a crucial aspect in the implementation of the coffee business. This can be seen from the percentage of damaged seeds and from the aroma content. The components that make up the aroma of coffee, namely chemical content, caffeine, trigonelin, chlorogenic acid (CGA), lipid, protein, carbohydrate also water have been predicted using destructive chemical analysis methods. Of course this method provides efficiency that is less than optimal in terms of time and cost, beside that, it also does not produce good quality uniformity [1]. Therefore, the application of non-destructive measurement methods is needed, for example the Near Infrared Spectroscopy (NIRS) method which is more effective and efficient.

The advantages of this method include a method that does not damage the material (non-destructive) which is used to obtain an ingredient in the material without taking a long time and does not use other chemical additives or ingredients. NIRS applications and research on coffee

have often been carried out. Among them are determining the concentration of caffeine in roasted coffee using FT-NIRS (Zhang et al., 2013), determining the levels of caffeine, theobromine, and theophylline from roasted coffee (Huck et al., 2004), determining the levels of chlorogenic acid from roasted coffee beans (Shan et al., 2014), determination of caffeine content in Gayo Arabica coffee beans (Rosita et al., 2016), non-destructive determination of the main chemical components of Arabica Java Preanger coffee beans (Naripati 2017), development of the identification model for the flavor forming constituents of Java Preanger coffee using NIRS (Ayu, 2017), estimation of moisture and lipid content of single green coffee beans using hyperspectral images (Caporaso et al., 2018), and analysis of the effect of variations in the number of bean layers on the accuracy of predicting the minor content of Bondowoso Green Arabica Coffee beans using NIRS (Madi et al., 2018).

The content of water, lipid and carbohydrates (part of the proximate composition of coffee) can be determined using NIRS method, where the material will be exposed to NIR light and the reflectance, absorbance and transmittance spectrum will be obtained. Ayu [2] stated that the chemical information obtained from the diffuse reflectance is useful in obtaining chemical information on materials which will later be displayed on a wavelength spectrum of 1000 to 2500 nm and will be processed using the NIRS method.

Research using NIR has been widely used in determining the content contained in food ingredients. Near infrared Spectroscopy (NIRS) is a method that applies spectroscopic waves which nondestructively could determine the content of a material without requiring a long time. This method is safe because it does not use chemicals or other chemical additives. NIRS can be used to determine the protein content, water content, carbohydrates and lipid content of an ingredient. The use of this spectroscopic wave has been widely used in the food industry and agriculture [3], also in green bean coffees [4] – [8].

Partial least square (PLS) is a structural equation modeling (SEM) model based on components or various kinds of data. The PLS equation was first developed by Herman O. A. Wold in the field of econometrics in the 1960s. PLS is a presumptive technique that can be used in various independent variables even though there is multicollinearity [9]. The PLS model can relate the existing information on two variables [10]. Determination of the calibration of the proximate content of Sidikalang green bean Arabica coffee has not been studied, so it is hoped that this research will be useful for the coffee industry in Indonesia in order to determine the characteristics and authenticity of coffee effectively and efficiently to avoid counterfeiting the origin of a variety of coffee beans.

## **2. Materials and Methods**

### *2.1 Material preparation*

The materials used during the study included 5 kg of Sidikalang Arabica green bean coffee which was obtained from Dairi Regency, North Sumatra. After that 90 samples of Sidikalang Arabica green bean coffee were prepared which were divided into 60 samples of calibration data and 30 samples for validation. The NIR reflectance was measured from a sample of coffee beans which was weighed at 96 grams using NIRFlex N-500 Merk Buchi and then for the chemical measurement the sample was weighed at 10 grams.

### *2.2 Chemical analysis*

Chemical analysis was carried out to obtain the proximate content of coffee beans, namely water content (thermogravimetric method), lipid (AOAC, 2006) and carbohydrates (by difference).

### *2.3 Research data analysis*

The spectroscopic NIR measurement obtained spectral data in the form of reflectance data. In order to obtain a linear relationship between NIR absorption values and chemical data, it is necessary to change the reflectance into absorbance spectra. Then the absorbance spectra and chemical data were divided into calibration data and validation data. According to [11] for calibration data obtained from 2/3 of the sample, namely 60 samples of data and validation data of 1/3 of the sample, which is 30 samples. Then the pretreatment data is divided into several methods in order to get a good calibration model. As mentioned [12] in [11] that to build a good model there are no rules about what method should be used in spectra pretreatment and data processing. There are 7 data processing used, namely the first derivative, second derivative, MSC,

SNV, normalization, first derivative and MSC and second derivative and MSC.

In this study, the PLS calibration method was used to build a model for determining the proximate content of water, lipid, and carbohydrate content using NIRS data and chemical data as references, where spectral data were taken randomly. In forming a model to get a good model, the first step is to determine the number of PLS factors [13]. The selection of factors will have an impact on the model to be built. If the PLS factor is too large, this can cause the prediction value to be low and cause the data to be overfitting [14], but if the PLS factor is too small it can cause the model to be underfitting [2].

#### 2.4 Evaluation of calibration value

Evaluation of the calibration value and validation of the model obtained in estimating the proximate content of water, lipid and carbohydrates using the NIRS method can be evaluated using several parameters, namely the correlation coefficient ( $r$ ), calibration standard error (SEC), standard prediction error (SEP), coefficient of variation (CV) and residual predictive deviation (RPD) [15].

### 3. Results and Discussion

#### 3.1 Sidikalang coffee bean proximate content

This study was conducted to find a model that can be used to non-destructively predict proximate content. The value of water, lipid and carbohydrate content of Sidikalang coffee beans by destructive analysis was obtained by 10.83% for the water content, for lipid 16.06% and for the carbohydrate value of 25.33% which was carried out on 90 samples with 30 validation samples and 60 samples. calibration samples using 3 repetitions

Table 1. The content of water, lipid and carbohydrates of Sidikalang Arabica coffee beans

Chemical Components	Min (%)	Max (%)	Mean (%)	Standard Deviation (%)
Moisture content	8.97	11.57	10.83	0.67
Lipid	12.20	19.50	16.06	1.72
Carbohydrate	18.94	30.67	25.33	2.93

#### 3.2 Loading plot of Sidikalang Arabica coffee beans

As demonstrated in [16] the wavelength which indicates the value of the water content is at the wavelength of 1934 nm, the value of the lipid content is in the waves of 1477 nm, 1726 nm, and 2128 nm, and the value of the carbohydrate content is at the wavelength of 2128. nm. Furthermore, according to [5] the value of water content is at wavelengths of 1433 to 1450 nm and 1940 nm, the value of lipid content is at wavelengths of 1410 nm, 1700 nm and 1891 to 1892 nm and the value of carbohydrate content at wavelengths of 1477 nm and between 2127 up to 2129 nm. In Figure 1 the known wavelengths indicate that there are moisture content at 1470 nm and 1980 nm. The wavelength value on the water content loading plot using normalization pretreatment is the same as Zulfahrizal's study [17] which shows that the water content value is at wavelengths of 1400 to 1480 nm and 1900 to 2000 nm.

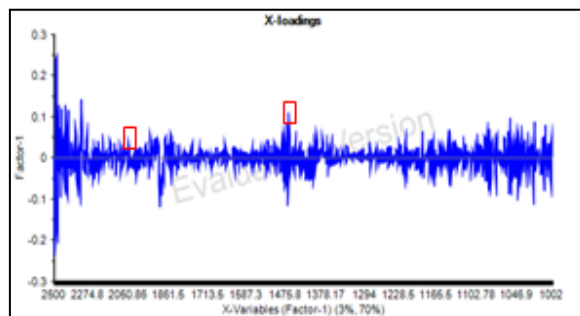


Figure 1. Loading plot of moisture content using normalization pretreatment

Athfiah [18] conducted a study to determine the chemical content of Bondowoso coffee beans with the lipid content value located between the wavelengths of 1410 nm, 1700 nm and 1892 nm.

Based on the lipid loading plot image using MSC pretreatment in Figure 2, the content values indicating the presence of lipid were located at wavelengths of 1703 nm and 1870 nm. This shows that the value of lipid content is not much different from previous studies.

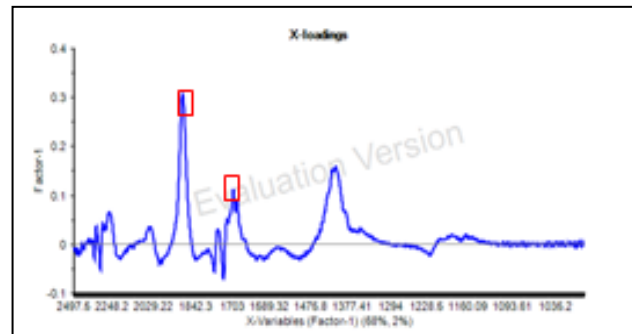


Figure 2. Loading plot of lipid using MSC pretreatment

In Figure 3, can be seen that the peak wavelength on the carbohydrate loading plot using MSC pretreatment which indicates the presence of carbohydrates is in the range of 1870 nm and 2130 nm. This is the same as the previous study by [16] where the wavelength indicating the presence of carbohydrates was at 2128 nm and [18] in determining the chemical content of Arabica coffee using NIRS.

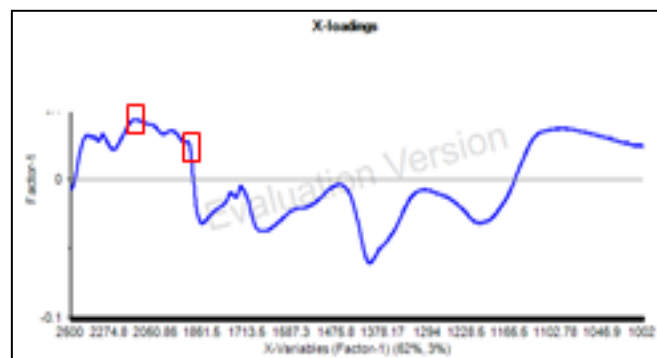


Figure 3. Loading plot of carbohydrate using MSC pretreatment

### 3.3 Results of calibration and validation with the PLS method

The value of calibration and validation in determining the water content using normalized pretreatment (Table 5) in this study can be said to be pretreatment with a model that has a moderate correlation with  $r$  value of 0.574%; SEC value of 0.54%; SEP value of 0.63%; CV value of 5.02% and RPD of 1.23% assisted by a consistency value of 85.73%. The results of this calibration and validation show that  $r$  is in the range 0.41 - 0.70 which is categorized as a moderate relationship according to Razak [19] in Mutaqin [20], the difference between the SEC and SEP values is 0.091 and the RPD value is in the range of 1. The value of this parameter is lower when compared to Sahfitri's research [5] where the  $r$  value is 0.80; SEC = 0.12%; SEP = 0.14% and RPD = 2.09.

Table 2. The results of calibration and validation of the moisture content of Sidikalang Arabica coffee beans

Type of pretreatment	Factor	r	SEC (%)	SEP (%)	CV (%)	RPD	Consistency (%)
Original	8 ; 6	0.622	0.51	0.62	4.80	1.29	83.05
MSC	5 ; 1	0.574	0.54	0.66	5.06	1.22	82.51
SNV	6 ; 2	0.671	0.49	0.66	4.55	1.36	74.09
Derivative 1	3 ; 1	0.768	0.42	0.65	3.93	1.57	65.41
Derivative 2	5 ; 1	0.839	0.36	0.69	3.33	1.86	51.85
D <sub>1</sub> + MSC	2	0.494	0.57	0.80	5.33	1.16	72.24
D <sub>2</sub> + MSC	1	0.845	0.36	0.68	3.34	1.85	53.05
<b>Normalization</b>	<b>6 ; 2</b>	<b>0.574</b>	<b>0.54</b>	<b>0.63</b>	<b>5.02</b>	<b>1.23</b>	<b>85.73</b>

Table 2 shows the results of calibration and validation values in determining lipid content using several pretreatments including MSC, SNV, normalization, derivative 1, derivative 2, a combination of derivative 1 and MSC and a combination of derivative 2 and MSC. From some of these treatments, it can be seen that a good model is MSC pretreatment. From the pretreatment, the value of r was 0.647%; SEC value of 1.30%; SEP value of 1.55%; CV value of 8.10%; RPD value of 1.32; and the consistency value is 83.87%.

Table 3. The results of calibration and validation of the lipid content of Sidikalang Arabica coffee beans

Type of pretreatment	Factor	r	SEC (%)	SEP (%)	CV (%)	RPD	Consistency (%)
Original	7 ; 6	0.696	1.22	1.49	7.63	1.40	82.08
<b>MSC</b>	<b>5</b>	<b>0.647</b>	<b>1.30</b>	<b>1.55</b>	<b>8.10</b>	<b>1.32</b>	<b>83.87</b>
SNV	6 ; 4	0.724	1.17	1.62	7.33	1.46	72.26
Derivative 1	3	0.805	1.01	1.57	6.30	1.70	64.21
Derivative 2	1	0.857	0.88	1.79	5.48	1.96	48.90
D <sub>1</sub> + MSC	1	0.680	1.25	1.63	7.79	1.38	76.52
D <sub>2</sub> + MSC	1	0.858	0.87	1.75	5.46	1.96	49.87
Normalization	8 ; 6	0.758	1.11	1.47	6.93	1.55	75.27

Based on Table 3, it can be seen that using MSC pretreatment showed better results than some other pretreatments. In MSC pretreatment, the value of r was 0.563%; SEC value of 2.42%; SEP value of 2.83%; CV value of 9.56%; RPD value of 1.22; and the consistency value is 85.38%.

Table 4. The results of calibration and validation of the carbohydrate content of Sidikalang Arabica coffee beans

Type of pretreatment	Factor	r	SEC (%)	SEP (%)	CV (%)	RPD	Consistency (%)
Original	8 ; 5	0.610	2.32	2.88	9.16	1.27	80.52
<b>MSC</b>	<b>6 ; 2</b>	<b>0.563</b>	<b>2.42</b>	<b>2.83</b>	<b>9.56</b>	<b>1.22</b>	<b>85.38</b>
SNV	7 ; 4	0.610	2.32	2.89	9.16	1.27	80.09
Derivative 1	3	0.721	2.03	2.88	8.02	1.45	70.34
Derivative 2	1	0.680	2.15	2.97	8.49	1.37	72.34
D <sub>1</sub> + MSC	2 ; 1	0.654	2.21	2.80	8.75	1.33	79.19
D <sub>2</sub> + MSC	1	0.797	1.77	2.72	6.99	1.66	64.91
Normalization	7 ; 3	0.589	2.36	2.797	9.35	1.25	84.71

### 3.4 Calibration and validation data plot

Figure 4 is a plot of calibration and validation data using normalized pretreatment on water content. It can be seen that the chemical data did not have results that were much different from the predicted data from the NIRS although there were still some data that showed inaccurate results. This can be said to have the ability to estimate the moisture content of Sidikalang Arabica coffee beans because it has a correlation coefficient value close to 1. In this data plot, the PLS factor used is 6 factors, if the PLS factor is greater than 6 then the value between calibration and

validation has a wide gap that causes overfitting. So to prevent this inequality, 6 PLS factors are used.

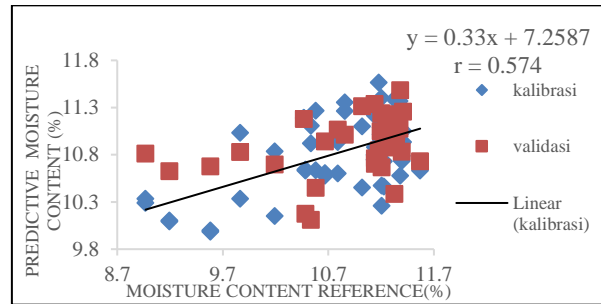


Figure 4. Data plot of moisture content prediction in Sidikalang Arabica coffee beans

The data plot of calibration and validation results for lipid using MSC pretreatment can be seen in Figure 5. It can be seen that the chemical data and NIRS predicted data are close to the regression line, with an  $r$  value of 0.647 which is almost close to 1 and the PLS factor used is different from the PLS factor on the water content, namely 6 factors. The use of 5 PLS factors is a better prediction for lipid content because if you use a factor above 5 you will get a value between calibration and validation that is much different which causes the distribution of the data to be less representative of the regression line.

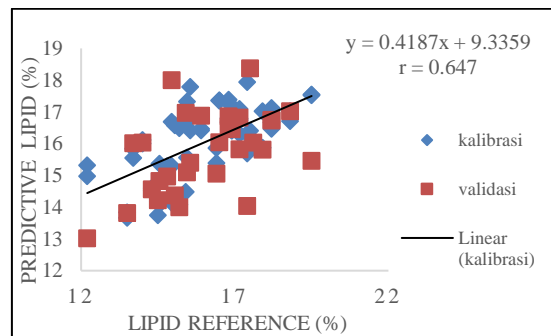


Figure 5. Data plot of lipid content prediction in Sidikalang Arabica coffee beans

In Figure 6, can be seen the data plot of the calibration and validation results on the proximate carbohydrate content of Sidikalang Arabica coffee beans using MSC pretreatment. It can be seen that the distribution of the data is close to the regression line so that it can be used to predict the presence of carbohydrates in Sidikalang Arabica coffee beans. This can be seen from the value of  $r$  obtained by 0.563 using the same 6 calibration factors as the use of the calibration factor for water content.

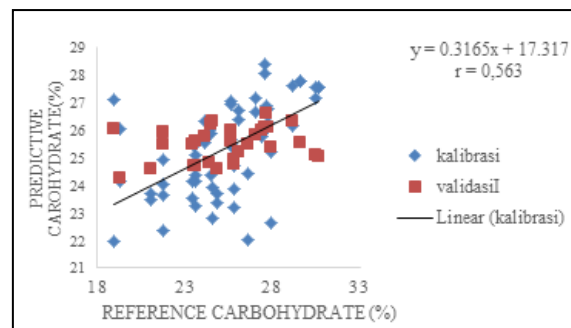


Figure 6. Data plot of carbohydrate prediction in Sidikalang Arabica coffee beans

From the results of the research conducted, the correlation coefficient value on the water content was obtained at 0.574; lipid and carbohydrates respectively 0.647; 0.563. The use of PLS factors can affect the percentage of chemical content in the material as well as a lot of data used. With the increase in the PLS factor used, it is easier to obtain a good calibration model. However,

if the difference in results between calibration and validation causes a large gap, this can be indicated as overfitting. Lengkey [11] stated that in developing a model, overfitting should be avoided in order to obtain a good model. Therefore, the selection of factors between calibration and validation is not expected to have much different result values so that a good predictive model can be obtained.

Standard error (SE) is the standard deviation of a sample distribution which states how much accuracy the data has. If the standard error value is small, it means that the accuracy of the data is good. Conversely, if the greater the standard error value obtained, the accuracy of the data is not good [21]. SE can be used to determine the standard error of calibration and standard error of prediction by comparing the conventional method with the predicted results of NIR because SE is the result of the difference between the actual value and the predicted value. The calibration model obtained in this study in predicting the water content is by using normalization pretreatment with 6 PLS factors, lipid content with MSC pretreatment with 5 PLS factors and carbohydrate content using MSC pretreatment also using 6 PLS factors.

In this study, the magnitude of  $r$  obtained falls into a moderate model category according to [19] in [20]. This is due to the deviation of the average value with a high difference between the data on the chemical data obtained from the average value of the Sidikalang coffee bean proximate data, namely water, lipid and carbohydrate content. The standard deviation value in this study makes it clear that there is a difference in the standard deviation value in the previous study conducted by Sahfitri [5], with the standard deviation value for the water content of 0.26; on lipid by 0.44 and on carbohydrates by 4.33. While the results of this study obtained the value of the standard deviation of the water content of 0.67; in lipid by 1.72 and in carbohydrates by 2.93. According to [21] the bigger the SE value in the variation of a data group, the data can be categorized as inaccurate, on the contrary, the smaller the SE value in the variation of a data group, the data is said to be accurate.

#### 4. Conclusion

Result showed that NIR spectroscopy could be used to predict chemical content in coffee beans. Moreover, the model obtained from this study could be categorized into a moderate model. This model can be used in simulations with predictive values with moderate accuracy with R value of 0.574, 0.647 and 0.563 for water content, lipid and carbohydrate content, respectively.

#### References

- [1] A. Rindang, P.C. Ayu, Sahfitri, "NIR spectroscopy as an environmentally friendly method to determine the lipid and carbohydrate concentration in Lintong Arabica coffee bean," *IOP Conf. Ser.: Earth Environ. Sci.*, 2022.
- [2] P.C. Ayu, "Pengembangan model penentuan kandungan kimia utama pembentuk flavor biji kopi Java Preanger menggunakan FT-NIRS", Department of Mechanical and Biosystem Engineering, IPB University, Bogor, 2017.
- [3] H. Cen and Y. He, "Theory and application of near infrared reflectance spectroscopy in determination of food equality," *Trends in Food Science and Technology*, 2007.
- [4] P.C. Ayu, I.W. Budiastara, A. Rindang, presented at International Conference on Agriculture, Environment and Food Security (AEFS), "NIR spectroscopy application for determination caffeine content of Arabica green bean coffee," May 2020.
- [5] Sahfitri, A. Rindang, P.C. Ayu, "Prediction of water content in Lintong green bean coffee using FT-NIRS and PLS method," *IOP Conf. Ser.: Earth Environ. Sci.*, 2020.
- [6] S. D. Siregar, A. Rindang, P.C Ayu, "Principle component analysis (PCA) classification of Arabica green bean coffee of North Sumatera using FT-NIRS," *IOP Conf. Ser.: Earth Environ. Sci.*, 2020.
- [7] I.W. Budiastara, M. Sutrisno, S. Widyaotomo, P.C. Ayu, "Determination of trigonelline and chlorogenic acid (CGA) concentration in intact coffee beans by NIR spectroscopy," *CIGR Journal*, vol. 22, no.1, Maret 2020.

- [8] J. Akhir, A. Rindang, P.C Ayu, "Caffeine content calibration model on green beans Arabica Mandailing Natal using NIRS and artificial neural network," *IOP Conf. Ser.: Earth Environ. Sci.*, 2020.
- [9] S. Ramzan and M.I. Khan, "Dimension reduction and remedy of multicollinearity using latent variable regression methods," *World Applied Sciences Journal*, vol. 8, no. 4, 2010.
- [10] H. Abdi, L.J. Williams, D. Valentin, "Multiple factor analysis: Principal component analysis for multi-table and multi-block data sets," *Article in Wiley Interdisciplinary Reviews: Computation Statistic*, February 2013.
- [11] L.C.E.C.H. Lengkey, I.W. Budiastra, K.B. Seminar, B.S. Purwoko, "Model pendugaan kandungan air, lemak dan asam lemak bebas pada tiga provenan biji jarak pagar (*Jatropha curcas l.*) menggunakan spektroskopi inframerah dekat dengan metode partial least square (PLS)," *Jurnal Littri*, vol. 19, no. 4, Desember 2013.
- [12] W. Tiaprasit and C. Sangpithukwong, "BUCHI NIRFlex N-500 Training Course", BUCHI NIR Application Support, Bangkok, Thailand. 2010.
- [13] D. M. Haaland and E.V Thomas, "Partial least square methods for spectral analyses relative to other quantitative calibration methods and the extraction of quantitative information," *Anal. Chem.*, vol. 60, no.11, February 1988.
- [14] R. Rosita, I.W. Budiastra, Sutrisno, "Prediksi kandungan kafein biji kopi Arabika Gayo dengan near infrared spectroscopy," *Jurnal Keteknik Pertanian*, vol. 4, no. 2, October 2016.
- [15] A.M. Mouazen, W. Saeys, J. Xing, J. Debaerdemaeker and H. Ramon, "Near infrared spectroscopy for agricultural materials: an instrument comparison," *Journal Near Infrared Spectroscopy*, vol. 13, no. 2, 2005.
- [16] J.S. Ribeiro, M.M.C. Ferreira, T.J.G. Salva, "Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using NIR Spectroscopy," *Talanta*, 2011.
- [17] Zulfahrizal, "Pengembangan metode pengukuran non destruktif untuk menentukan mutu dan fermentasi biji kakao utuh menggunakan NIR spectroscopy," thesis, Department of Mechanical and Biosystem Engineering, IPB University, Bogor, 2014.
- [18] S.N Athfiah, "Penentuan kandungan kimia utama biji kopi Arabika Bondowoso secara nondestruktif dengan near infrared spectroscopy," undergraduate thesis, Department of Mechanical and Biosystem Engineering, IPB University, Bogor, 2017.
- [19] A. Razak, "Statistika Bidang Pendidikan," Faculty of Teacher Training and Education, Universitas Riau, Pekanbaru, 1991.
- [20] A.S. Mutaqin, D.M. Hendri, H.R. Septina, K.C. Nawaji, N.R.A Sari. R.S. Mumupuni, Y. Kristy, "Ukuran Penilaian Kemampuan/Kesesuaian Model R-square," *Makalah Analisis Regresi*, STIS, Jakarta, 2014.
- [21] D.N. Sari, Zulfahrizal, A.A. Munawar, "Non-destruktive test menggunakan NIRS dengan metode partial least square untuk bubuk biji kakao (dengan pretreatment baseline shift dan derivatif-2)," *Jurnal Ilmiah Mahasiswa Pertanian Unsyiah*, vol. 1, no. 1, November 2016.