

THE OCCUPANCY RATE MODELING OF KENDARI HOTEL ROOM USING MEXICAN HAT TRANSFORMATION AND PARTIAL LEAST SQUARES

Margaretha Ohyver¹; Herena Pudjihastuti²

^{1,2}Mathematics and Statistics Department, School of Computer Science, Bina Nusantara University
Jln. KH Syahdan no.9, Kemanggisan, Palmerah, Jakarta Barat 11480, Indonesia
¹mohyver@binus.edu

ABSTRACT

Partial Least Squares (PLS) method was developed in 1960 by Herman Wold. The method particularly suits with construct a regression model when the number of independent variables is many and highly collinear. The PLS can be combined with other methods, one of which is a Continuous Wavelet Transformation (CWT). By considering that the presence of outliers can lead to a less reliable model, and this kind of transformation may be required at a stage of pre-processing, the data is free of noise or outliers. Based on the previous study, Kendari hotel room occupancy rate was affected by the outlier, and it had a low value of R^2 . Therefore, this research aimed to obtain a good model by combining the PLS method and CWT transformation using the Mexican Hats them other wavelet of CWT. The research concludes that merging the PLS and the Mexican Hat transformation has resulted in a better model compared to the model that combined the PLS and the Haar wavelet transformation as shown in the previous study. The research shows that by changing the mother of the wavelet, the value of R^2 can be improved significantly. The result provides information on how to increase the value of R^2 . The other advantage is the information for hotel managements to notice the age of the hotel, the maximum rates, the facilities, and the number of rooms to increase the number of visitors.

Keywords: occupancy rate modeling, hotel room modeling, Mexican hat, partial least squares

INTRODUCTION

Partial Least Squares (PLS) method was developed in 1960 by Herman Wold. The method particularly suits with constructing a regression model when the numbers of independent variables have much and high collinear.

There are several definitions regarding the PLS. According to Yeniay and Goktas (2002), the PLS is one of the methods used to overcome the problem of multicollinearity, regardless the number of independent variables even for more than the number of observations. Ramzan and Khan (2010) defined the PLS as a modeling technique that can be used to handle many independent variables that are high collinear. Likewise, according to Carrascal, Galvan, and Gordo (2009), PLS is a method that can be used when the number of independent variables is equal to or more than the number of the dependent variables, or there is a problem of multicollinearity between the independent variables.

The PLS can be combined with other methods, one of which is a Continuous Wavelet Transformation (CWT). The rationale of this kind of transformation is the need to do pre - process data to get noise-free data. In here, the noise means the presence of the outlier that can obtain less reliable models (Ohyver & Tanty, 2012). The wavelet is often used to detect the outlier. The outlier can be found in data mining and detected based on transform wavelet (Yu *et al.*, 2002). The wavelet transform can be used to analyze the location of the outlier and the strength of singularity in financial

data (Zong *et al.*, 2013). With the wavelet-based algorithm, an outlier can be detected (Mercorelli, 2014).

This research is made to apply the CWT and the PLS methods into the hospitality data. The data used is hotel visitor data in Southeast Sulawesi Province that have been used in the previous study by Ohyver and Pudjihastuti (2014). The previous study has applied the combined methods of PLS and CWT with the Haar wavelet in 1-5 scales as other wavelet. However, the model generated has a low value of R^2 (30%). The R^2 is a measure of how well the regression line fits the data. It can also be interpreted as the proportion of the variation in Y that is explained by the regression relationship of Y with X. The low value of R^2 shows the model cannot predict the value of Y very well. Adding independent variables is a common step that people choose to increase the value of R^2 but this step requires more work. Therefore, people need another one to overcome that problem. The low value of R^2 can be solved by constructing the model using available variables. Ohyver and Pudjihastuti (2014) constructed the model using CWT as the pre-processing stage combined with PLS as the modeling stage. They stated that one of the causes of the low value of R^2 was a discrepancy of used mother wavelet. Therefore, this research intends to use another mother wavelet namely the Mexican hat in the pre-processing stage. It is expected that this wavelet can improve the value of R^2 of the model.

There are several goals in the research. First, it is to obtain the noise-free data from pre-processing data using the Mexican hat mother wavelet. Secondly, it is to generate the CWT-PLS models for hotel visitor data in Kendari, Southeast Sulawesi. This research is expected to help the hotel managements in Kendari to increase the number of the customers. By knowing the factors that affect the number of the visitors, the hotel managements can improve the service in the quality and quantity. Hopefully, the improvements in the service can enhance the number of visitors that leads to higher incomes and economic growth of the hotel employees.

METHODS

The data used in the study is secondary data obtained from the catalog of the Central Bureau of Statistics (BPS) Southeast Sulawesi in 2011. The sample covers 90 hotels or inns located in Kendari (Katalog BPS, 2011). Table 1 shows the variables used in the research including seven independent variables (X_1 - X_7) and a dependent variable or response variable (Y).

Table 1 Variables Used In The Research

No	Variables	Symbol
1.	The age of the hotel/inn	(X_1)
2.	The minimum of the hotel/inn rates	(X_2)
3.	The maximum of the hotel/inn rates	(X_3)
4.	The facility in the hotel/inn	(X_4)
5.	The number of workers in the hotel/inn	(X_5)
6.	The number of rooms in the hotel/inn	(X_6)
7.	The number of beds in the hotel/inn	(X_7)
8.	The number of visitors	(Y)

The research applies the Continuous Wavelet Transformation (CWT) and Partial Least Squares (PLS) methods. The mother wavelet used is the Mexican hat. The steps of this research are applying the continuous wavelet transformation to the 90 observations involving only the independent variables, and establishing a regression model using the transformed data.

The theory of Wavelet is relatively a new concept. The wavelet was first introduced by Alfred Haar in 1909. The term "wavelet" was invented by Jean Morlet and Alex Grossmann in early 1980. It was derived from French word "*ondelette*" which means a small wave. The word "*onde*" was translated into English as "wave" and combined with the original word to form the new word "wavelet".

The function $\psi(t)$ is defined as a wavelet if it is satisfying:

$$\int_{-\infty}^{\infty} \psi(t) dt = 0 \quad (1)$$

$$\int_{-\infty}^{\infty} \psi^2(t) dt = 1 \quad (2)$$

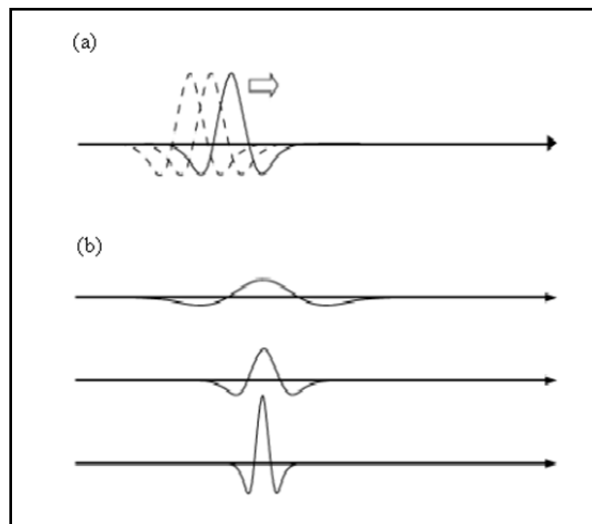


Figure 1 (a) Translation and (b) Dilation
(Source: Addison, 2002)

When there is a wavelet function or called as the mother wavelet, the other functions that become the basis functions of space $L^2(R)$ can be raised using dilation and translation. Figure 1 shows the examples of translation and dilation wavelet.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right), a \neq 0, a, b \in R \quad (3)$$

The "a" is a dilation parameter or a scale that measures the degree of compression. If $|a|$ is smaller than 1, the wavelet will condense, and if $|a|$ is bigger than 1, the wavelet will be widened. The "b" is a translation parameter that determines the location of the time of the wavelets.

The function or signal can be transformed to the elements of the wavelet. This transformation is known as wavelet transformation. Mathematically, wavelet transformation is a convolution between the wavelet function and the signal (Addison, 2002). There are two types of wavelet transformation which are the Discrete Wavelet Transformation (DWT) and the Continuous Wavelet Transformation

(CWT). The difference between them is the value of a and b , where the value of a and b is limited in discrete values for DWT.

$$a = a_0^m, b = n b_0 a_0^m, m, n \in Z, \text{ dan } a_0 > 1, b_0 > 0 \text{ (Daubechies, 1992).}$$

DWT:

$$C_f(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (4)$$

CWT:

$$C_f(a,b) = a_0^{-m/2} \int_{-\infty}^{\infty} f(t) \psi(a_0^{-m} t - n b_0) dt \quad (5)$$

Many events in the world dealing with signals can be analyzed, such as stream or, human voice, machine vibration, financial data, and music. These signals can be analyzed using wavelet analysis. As noted earlier that there are two types of wavelet transformation, DWT, and CWT, which can be distinguished by their scale and translation. In CWT, the transformation can be performed at many scales and translations.

$$C_f(a,b) = \int_{-\infty}^{\infty} f(t) \psi_{a,b}^*(t) dt, \quad (6)$$

or

$$C_f(a,b) = \langle f(t), \psi_{a,b}(t) \rangle \quad (7)$$

Where $C_f(a,b)$ is a wavelet coefficient in different scales and $f(t)$ refers to the data to be transformed. The sign “*” states the complex conjugate of wavelet function. However, this complex conjugate is only necessary if the mother wavelet used is the complex wavelet.

Partial Least Squares (PLS) is one of the methods that can be used to overcome the problem of multicollinearity. PLS is the combination of Principal Component Analysis (PCA) and multiple linear regression (Abdi, 2003). PCA is a method used to reduce the number of independent variables to some new variables that are not correlated each other and can explain both the diversity of data and new variables.

To establish the relationship between the response variable and the independent variables, the PLS method forms new independent variables called factors or latent variables or components, where each component formed is a linear combination of the independent variables. The main objective of PLS is to form components that can capture the information from independent variables to predict the response variable (Hoskuldsson in Garthwaite, 1994).

The method of least squares can not be used if n is smaller than p because the matrix $\mathbf{X}^T \mathbf{X}$ is singular (Naes *et al.*, 2002). Instead, the Partial Least Square (PLS) method can be used if the case n is smaller than p since the PLS regression is based on the decomposition of components.

$$\mathbf{Y} = \mathbf{TQ}' + \mathbf{F} \quad (8)$$

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} \quad (9)$$

Where \mathbf{T} is the matrix of components, \mathbf{P} and \mathbf{Q} are the loading of the matrix. Moreover, \mathbf{X} and \mathbf{Y} , \mathbf{E} and \mathbf{F} are the error vectors (Boulesteix and Strimmer, 2006).

The PLS method can be considered as the method to form the matrix component \mathbf{T} as a linear transformation of \mathbf{X} .

$$\mathbf{T} = \mathbf{XW} \quad (10)$$

Where, \mathbf{W} is the weight matrix. Therefore, the equation can be written as following:

$$\begin{aligned} T_1 &= w_{11}x_1 + w_{21}x_2 + \dots + w_{p1}x_p \\ T_2 &= w_{12}x_1 + w_{22}x_2 + \dots + w_{p2}x_p \\ &\vdots \\ T_c &= w_{1c}x_1 + w_{2c}x_2 + \dots + w_{pc}x_p \end{aligned} \quad (11)$$

The components are used to estimate by substituting \mathbf{X} to obtain the least squares estimator. The equation can be seen below.

$$\hat{\mathbf{Q}}' = (\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\mathbf{Y} \quad (12)$$

RESULTS AND DISCUSSIONS

Southeast Sulawesi is one of the provinces in Indonesia. It has been established as an autonomous region by decree No. 2 1964 junto Law 13 of 1964. At first, it consists of four districts and recently it has been expanded into 10 districts and 2 cities. The capital of Southeast Sulawesi is located in Kendari. 27 April 1964 is the birthday of the Southeast Sulawesi (Biro Humas & PDE Pemerintah Provinsi Sulawesi Tenggara, 2013).

Southeast Sulawesi is one of the developing provinces in Indonesia since the economic growth of the region is relatively high above 8%. The poverty rate decreased from 21% to 14,6% in 2011 (Investor Daily Indonesia, 2012). One of the factors that can be considered in the economic growth is the development of the hotel sector. Based on manpower and transmigration official data in the Southeast Sulawesi in December 2011, the hotel and restaurant sectors absorbed about 8.791 local work forces (Media Sultra, 2011).

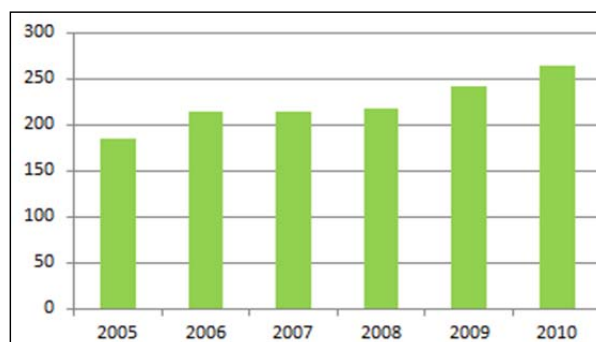


Figure 2 The Growth of the Number of Hotels in Southeast Sulawesi

Figure 2 shows that the number of hotels in Southeast Sulawesi has continued to grow since 2005. In 2006, South East Sulawesi hosted Al-Quran reading competition or Musabaqah Tilawatil Quran (MTQ). This competition made people provide lodging for the participants. From Figure 2, it can be seen that there is a significant growth in the number of hotels in 2006. Mining in Southeast Sulawesi also shows an improvement lately. Because of that, there are many people or companies investing their money in South East Sulawesi. The development of mining become one of the attractions for people to come. Moreover, the tourism also has become tourist attention. Wakatobi Island is one of the favorite choices for vacation by local and foreign tourists. All the factors above have become challenges to set up a hotel with the best performance. This implies that there is a high competition between the owners of the hotel. Therefore, they should be aware of the factors that influence the number of their customers. This research gives the information to the owners of the hotel, and they can use it to improve their services.

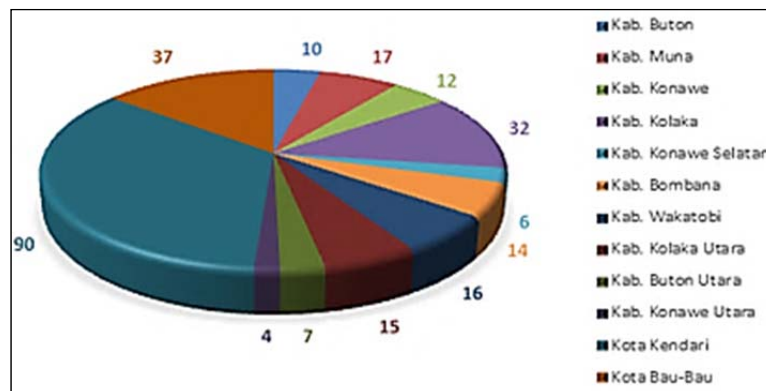


Figure 3 The Number of Hotels in Southeast Sulawesi

Figure 3 shows the number of hotels in each city in Southeast Sulawesi. It can be seen that Kendari has the most number of hotels because it is the capital of the province, and it has more and better facilities and infrastructure compared to the other cities in Southeast Sulawesi. On the contrary, North Konawe is a city with the fewest number of hotels since it is a new city established in 2007 and it is still in the stage of constructing the facilities and infrastructure in various sectors including the hotel. From Figure 3, it can be seen that all of the cities has the hotel to serve the visitors.

Kendari is the city with the highest average number of hotel visitors in Southeast Sulawesi. It makes sense that the hotels in Kendari have the highest average number of rooms. The hotels in Kendari also have the highest average number of beds. Thus, it makes the hotels need more employees. If the researchers investigate the age of the hotels, Kabupaten Muna has the average age of over ten years. Moreover, hotels in Kendari have the highest maximum and minimum rate. The status as a capital city can be the reason of it. The other reason is this city has a strategic location, and the hotels have many good facilities.

A fairly large number of hotels in Kendari make the competition high. It is also added by the fact that Kendari is not big. Various factors can lead to the increase or decrease in the number of visitors. The researchers select 7 factors to be analyzed. The seven factors are the age of the hotel, the minimum rates, the maximum rates, the number of existing facilities in the hotel, the number of workers, the number of rooms, and the number of beds. The other factors that might be related are not analyzed in this research because of the difficulty to obtain the data or information.

Ohyver (2013) showed that this data have an outlier. The outlier is detected in all variables. To overcome it, Ohyver (2013) used logarithm natural transformation. To gain the model, the logarithm natural is combined with PLS like this research. The result is quite bad because the value of R^2 is only 34%. In the previous study by Ohyver and Pudjihastuti (2014), the model resulted was in Equation (13). The model is obtained by using the Haar wavelet on the scale of 6. In addition, the value of R^2 is 30%, which indicates that the contribution of the six variables to the diversity of the number of hotel visitor is too small. Based on the theory, another variable should be added to increase the value of R^2 . However, in this research, the researchers try to replace the mother wavelet rather than adding the variables.

$$\text{Visitors} = -11,492 - 1,415 \text{ age} - 1,494 \text{ minimumrate} - 3,440 \text{ maximumrate} + 1,123 \text{ facilities} + 1,090 \text{ workers} + 2,113 \text{ rooms} + 219,596 \text{ beds} \quad (13)$$

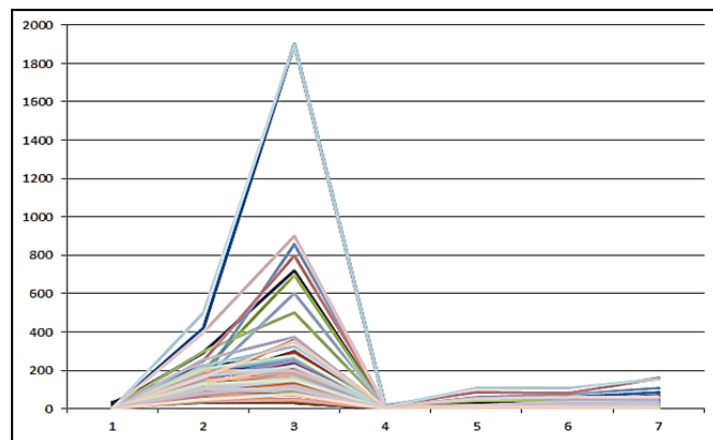


Figure 4 The Results of Data Transformation

The research transforms the six factors using Mexican hat on a scale of 1-5. Figure 4 shows the results of the transformation. It suggests that on the scale of 3, it already has a smooth pattern. Hence, it is considered that the scale above 2 has not depicted the actual data patterns. The figure also shows that the results obtained are not too smooth, but it is enough to shrink the distance between the points. The next step is to create the model using Partial Least Square (PLS) into the results of the transformation on the scale of 1, and 2. The resulted in the model using the scale of 1 is a model in six components as seen in Equation (14).

$$\text{Visitors} = 591,574 + 20,437 \text{ age} - 8,852 \text{ minimumrate} + 31,516 \text{ maximumrate} + 49,646 \text{ facilities} - 54,280 \text{ workers} + 258,451 \text{ rooms} - 52,121 \text{ beds} \quad (14)$$

The model shows that the variables of the minimum rate, worker, and the number of beds are led to the decrease in the number of hotel visitors. Figure 5 is the result of ANOVA which indicates that the model in (13) is a significant model.

Analysis of Variance for y					
Source	DF	SS	MS	F	P
Regression	5	496763032	99352606	47,23	0,000
Residual Error	84	176683614	2103376		
Total	89	673446647			

Figure 5 ANOVA using Scale 1

On the other hand, the result obtained by using the scale of 2 is a model in six components as following:

$$\begin{aligned} \text{Visitors} = & \\ & 722,920 - 167,280 \textit{ age} + 205,876 \textit{ minimumrate} - 15,379 \textit{ maximumrate} - 97,686 \textit{ facilities} - \\ & 146,065 \textit{ workers} + 636,473 \textit{ rooms} - 466,125 \textit{ beds} \end{aligned} \quad (15)$$

The model shows that the variables of age, the maximum rates, facilities, the number of workers, and the number of beds decrease the number of hotel visitors. First, the visitor prefers to stay in the “old” hotel. A hotel that can survive for a long time means that people are satisfied with the service of the hotel. Second, the negative sign for maximum rate shows that the visitor prefers to stay in the hotel with low price. This is related to the positive sign of minimum rate. Third, the hotel with a few facility is liked by hotel visitors. It probably is because the visitor only wants to stay, and for the entertainment, they can find it outside of the hotel. Fourth, the negative sign for workers shows that visitor prefers to choose the hotel with a few worker. Fifth, the negative sign for beds, states that the visitor prefers to stay in the hotel with a few beds. From the variables which have the negative sign, it can be said that the hotel visitors prefer to stay in the “old” hotel that is not crowded. Figure 6 shows the result of ANOVA that indicates that the model (15) is also a significant model.

Analysis of Variance for y					
Source	DF	SS	MS	F	P
Regression	5	496392469	99278494	47,10	0,000
Residual Error	84	177054178	2107788		
Total	89	673446647			

Figure 6 ANOVA using Scale 2

The value of R^2 of both models is nearly equal. The value of R^2 of the model (14) with the scale of 1 is 73,76%, and the model (15) with the scale of 2 is 73,71%. It suggests that both models are good models. However, by comparing the PRESS values of both models, it shows that the best model is the model with the scale of 1 since it has the smallest PRESS value. By comparing the R^2 values, it also shows that the model (14) is better than the model resulted in the previous studies using the Haar wavelet as seen in equation (12). Without adding variables or factors to the model, the value of R^2 can be improved significantly from 30% to 73.76%. Therefore, the using of wavelet as a transformation tool is very helpful in increasing the value of R^2 , and Mexican hat is one of the appropriate mother wavelets for it.

CONCLUSIONS

There are several conclusions in this research. First, the model transformed by using a continuous wavelet Mexican Hat has a pretty good value of R^2 . Second, the variables of the minimum rates, the number of workers, and the number of beds cause a decrease in the number of hotel visitors. Third, the variables of the age of the hotel, the maximum rates, the facilities, and the number of rooms can increase the number of visitors. Fourth, the model resulted by using the wavelet transformation with Mexican Hat and the Partial Least Square method is a significant model. The R^2 of the model is 73,76 %, and it is better than what Ohyver and Pudjihastuti (2014) have achieved.

REFERENCES

- Abdi, H. (2003). Partial Least Square regression (PLS regression). *Encyclopedia for Research Methods for the Social Sciences*, 792-795.
- Addison, P. S. (2002). *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance* London: Institute of Physics Publishing.
- Biro Humas & PDE Pemerintah Provinsi Sulawesi Tenggara. (2013). *Profil Sulawesi Tenggara*. Retrieved from <http://www.sulawesitenggaraprov.go.id/ensiklopedia.php>
- Boulesteix, A., & Strimmer, K. (2006). Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics*, 8(1), 32-44.
- Carrascal, L. M., Galvan, I., & Gordo, O. (2009). Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos*, 118(5), 681-690.
- Daubechies, I. (1992). *Ten lectures on wavelets*. Philadelphia: Society for industrial and applied mathematics.
- Garthwaite, P. H. (1994). An interpretation of partial least squares. *Journal of the American Statistical Association*, 89(425), 122-127.
- Investor Daily Indonesia. (2012). *Pertumbuhan Ekonomi Sultra Lampau Nasional*. Retrieved from <http://www.investor.co.id/home/pertumbuhan-ekonomi-sultra-lampau-nasional/28618>
- Katalog BPS. (2011). *Direktori dan Tingkat Penghunian Kamar Hotel Provinsi Sulawesi Tenggara Tahun 2011*. Kendari: Badan Pusat Statistik Provinsi Sulawesi Tenggara.
- Mercorelli, P. (2014). A wavelet-based algorithm without a prior knowledge of noise level for gross errors detection. *WIT Transactions on Information and Communication Technologies*, 53, 9-16.
- Media Sultra. (2011). *Sektor Perhotelan Sultra Serap 8791 Tenaga Kerja*. Retrieved from <http://sindikasi.inilah.com/read/detail/1812732/sektor-perhotelan-sultra-serap-8791-tenaga-kerja>.
- Naes, T., Isaksson, T., Fearn, T., & Davies, T. (2002). *A user friendly guide to multivariate calibration and classification*. Chichester: NIR publications.
- Ohyver, M. (2013). Penerapan metode transformasi logaritma natural dan partial least squares untuk memperoleh model bebas multikolinier dan outlier. *MatStat*, 13(1), 42-51.
- Ohyver, M., & Pudjihastuti, H. (2014). Pemodelan tingkat penghunian kamar hotel di Kendari dengan transformasi wavelet kontinu dan partial least squares. *ComTech*, 5(2), 1178-1189.
- Ohyver, M., & Tanty, H. (2012). Pendeteksian outlier pada model regresi ganda studi kasus: Tingkat penghunian kamar hotel di Kendari. *MatStat*, 12(2), 114-122.
- Ramzan, S., & Khan, I. M. (2010). Dimension reduction and remedy of multicollinearity using latent variable regression methods. *World Applied Science Journal*, 8(4), 404-410.

- Yu, D., Sheikholeslami, G., & Zhang, A. (2002). Find Out: Finding outliers in very large datasets. *Knowledge and Information Systems*, 4(4), 387-412.
- Yeniay, O., & Goktas, A. (2002). A comparison of partial least squares regression with other prediction methods. *Hacettepe Journal of Mathematics and Statistics*, 31(99), 99-101.
- Zong, N. N., Che, E. G., Ji, T., & Xiao, Y. (2013). Detection of outliers of financial time series based on wavelet transform modulus maximum. In *Computer Science and Network Technology (ICCSNT), 2013 3rd International Conference on* (pp. 529-533). IEEE.