An adjustment degree of fitting on fuzzy linear regression model toward manufacturing income

Nurfarawahida Ramly[1], Mohd Saifullah Rusiman[1], Shuhaida Ismail[1], Suparman[2], Firdaus Mohamad Hamzah[3], Ozlem Gurunlu Alma[4]

[1]Faculty of Applied Sciences and Technology, Universiti Tun Hussien Onn Malaysia, UTHM Pagoh, Johor, Malaysia
[2]Faculty of Teacher Training and Education, Ahmad Dahlan University (UAD), Yogyakarta, Indonesia
[3]Centre for Defence Foundation Studies, Universiti Pertahanan Nasional Malaysia, Kuala Lumpur, Malaysia
[4]Faculty of Sciences, Mugla Sitki Kocman University, Mugla, Turkey

Article Info

Article history:

Received Sep 21, 2021
Revised Sep 25, 2022
Accepted Oct 25, 2022

Keywords:

Degree of fitting
Fuzzy linear regression
Manufacturing income
Mean square error
Multiple linear regression

ABSTRACT

The regression analysis is a common tool in data analysis, while fuzzy regression can be used to analyze uncertain or imprecise data. Manufacturing companies often having difficulty predicting their future income. Thus, a new approach is required for the prediction of future company income. This article analyzed the manufacturing income by using the multiple linear regression (MLR) model and two fuzzy linear regression (FLR) model proposed by Tanaka and Zolfaghari, respectively. In order to find the optimum of the FLR model, the degree of fitting (H) was adjusted in between 0 to 1. The performance of three models has been measured by using mean square error (MSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). Detailed analysis proved that Zolfaghari's FLR model with the degree of fitting of 0.025 outperformed the MLR and FLR with Tanaka's model with the smallest error value. In conclusion, the manufacturing income is directly correlated with six independent variables. Furthermore, three independent variables are inversely related to manufacturing income. Based on the results of this model, it appears to be suitable for predicting future manufacturing income.

Corresponding Author:

Mohd Saifullah Rusiman
Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology
Universiti Tun Hussien Onn Malaysia Cawangan Pagoh
84600 Muar, Johor, Malaysia
Email: saifulah@uthm.edu.my

1. INTRODUCTION

A regression model is one of the most popular statistical models used to determine association between multivariate data [1]. This model is commonly used in the fields of applied sciences, computer science, social sciences, engineering, and economics [2]. Regression analysis was used when an explanatory variable is dependent on a response variable. It shows the value of the response variable changing when one of the explanatory variables varies while the rest remain unchanged [3].

However, statistical modelling cannot be used on all data. The traditional method would not be able to accurately determine a result when vague data exist. Thus, for studies related to the association between dependent and independent variables, the regression method unable to predict precisely because of an unpredicted event. Some data may not necessarily be normal in some situations, particularly when predicting income. In some cases, data are not normal due to outliers or missing values. Due to this, the existing method was unable to estimate and find out the results accurately. To deal with that situation, alternative approaches are necessary.

Many fields, especially engineering and technology, use the fuzzy method in the analysis of uncertain data [4]. In complex systems involving human estimation, fuzzy methods can be applied to analyze uncertain or imprecise data between response and explanatory variables [5]. First introduced by Tanaka, fuzzy linear regression (FLR) has demonstrated its usefulness in solving complex problems where many cases are difficult to quantify [6]. FLR can provide an approximation between variables with insufficient uncertainty information [7].

Zolfaghari on the other hand has introduced an extension model that involved triangular fuzzy numbers (TFNs). This model considers either symmetrical or asymmetrical with its membership function (MF). The model also considered two parameter estimation factors, which is the degree of fitting [8]. The factors of parameter estimation can be transformed into two ways which are linear programming and fuzzy least squares method. Previous studies have proved the multiple linear regression method with the fuzzy regression technique according to various fields of study [9]–[13].

Previous researcher proposed the least squares method as a common FLR model [14]. However, the model showed an influenced towards the outliers, which led to inaccuracies. Another fuzzy model, based on least absolute deviation was introduced as an alternative to cater an outlier issue [15]. In addition, it works well on both symmetrical and non-symmetrical data. By applying least absolute deviation approach, a model from fuzzy numbers of matrices is created accordingly. A study by [16] has proven that fuzzy model based on least absolute deviation performed better and more structured compared to least squares method.

The aim of this study is to propose a Zolfaghari's FLR model with adjusting the degree of fitting (*H*) for estimating future manufacturing income. It is expected that the proposed model will prove to be the most optimal model that can be applied to the industries sector. Moreover, there are no assumptions to be considered before the model can be analyzed. The optimal model can be obtained by adjusting the degree of fitting (*H*) in order to find out the smallest error value.

2. PROPOSED METHOD

The research framework of this study is shown in Figure 1. The data were obtained from the Department of Statistics Malaysia (DoSM) from various industry sectors which include farming, fishing, mining, quarrying, manufacturing, construction, transport, and others [17]. Data filtering were performed accordingly and only manufacturing sector has been chosen for detailed analysis. The dataset has a total of nine explanatory variables including legal status (individual proprietorship, partnership, private, public, co-operative, others), ownership (Malaysian residents, non-Malaysian residents, joints), value of assets (total net book value), total employment, total salaries and wages paid, number of degree and above holder, number of diploma holder, number of Malaysian Certificate of Education (MCE) and below holder and total expenditure, while income is the dependent variable [17].

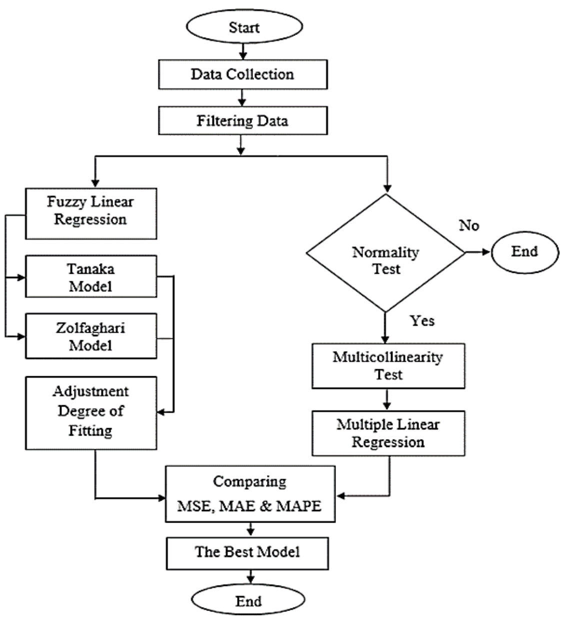

Figure 1. Research framework of proposed model

Once data filtering was performed, data pre-processing such as outlier treatment, normality and multicollinearity test were done accordingly. The tests are compulsory as it needed to fulfill the first assumption of the model. Next, the dataset was then fed into various algorithm such as multi linear regression (MLR), Tanaka's FLR and Zolfaghari's FLR algorithm with adjusted degree of fitting (*H*). Then, the obtained error values were calculated using several performance indicators such as mean square error (MSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) [18]. The resulting MSE, MAE and MAPE were then compared in order to find out the best predicting model for the industries sector.

3. RESEARCH METHOD

3.1. Multiple linear regression (MLR)

MLR is an extension of simple linear regression [19] and is usually applied for many statistical analyses and is also known as a method that can evaluate the association among the dependent and independent variables [20]. MLR model has two key assumptions which are normality distribution and multicollinearity among explanatory variables. The Q-Q plot is used to identify normality distribution among the response variable and explanatory variables [21]. The next test is multicollinearity checking. It should be tested among explanatory variables by using variance inflation factor (VIF) to avoid any dependency between variables. The MLR model can be detailed as in (1).

$$\hat{Y}_r = \beta_0 + \beta_1 X_{r1} + \beta_2 X_{r2} + \ldots + \beta_p X_{rp} + \varepsilon_r(\beta) \tag{1}$$

Where *r* = 1, 2, *N*, the response variable is *Y*, the explanatory variable is X_1 to X_p, regression coefficient is β_1 to β_p. The least square method (LSM) is shown in (2).

$$S(\beta_0, \beta_1, \beta_2, \ldots, \beta_p) = S(\beta) = \sum_{j=1}^{d} \varepsilon_j^2 \tag{2}$$

From (1), $\varepsilon(\beta) = Y - X\beta$, Then, *S*(*β*) is shown in (3).

$$S(\beta) = (Y - X\beta)^T (Y - X\beta)$$

$$= Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta \tag{3}$$

In LSM, the best fitting data is computed by minimizing *S*(*β*). Then, differentiate *S*(*β*) with respect to β where $\left.\frac{\delta S}{\delta \beta}\right|_{\hat{\beta}}$ is equal to zero as in (4),

$$\left.\frac{\delta S}{\delta \beta}\right|_{\hat{\beta}} = -2X^T Y + 2X^T X\beta = 0 \tag{4}$$

then, LS estimator is shown in (5),

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{5}$$

the value of dependent of $\beta_0 + \beta_1 X_{r1} + \beta_2 X_{r2} + \ldots + \beta_p X_{rp}$ is represented by $\hat{Y}$, and the residual $\varepsilon_r = Y_r - \hat{Y}_r$. Few previous researchers could explain the detailed of least square estimator method [19], [22].

3.2. Fuzzy linear regression (Tanaka)

In 1982, FLR was first proposed by Tanaka. FLR analysis aims to explore the potential models that fit observed fuzzy data. The difference of model is usually based on the fitting's formula. The final model of FLR is shown in (6). In order to obtain the fuzzy model, the estimation of the fuzzy parameters is done by solving a linear programming problem as in (7) [6], [23].

$$\hat{Y} = A_0(\alpha_0, c_0) + A_1(\alpha_1, c_1)x_1 + A_2(\alpha_2, c_2)x_2 + \ldots + A_k(\alpha_k, c_k)x_k \tag{6}$$

Linear programming problem,

$$\min_{\alpha, c} = c_1 + \ldots + c_N$$

subject to c ≥0 and,

$$\alpha^s x_q + (1-H)\sum_r c_r \left|x_{qr}\right| \geq y_q + (1-H)\varepsilon_q$$

$$-\alpha^s x_q + (1-H)\sum_r c_r \left|x_{qr}\right| \geq -y_q + (1-H)\varepsilon_q \tag{7}$$

where H is a degree of fitting, α is fuzzy center, c is fuzzy width and ε_q is the error. If the linear programming problem in (7) could be solved, it is considered that the fitted model is satisfied.

3.3. Fuzzy linear regression (Zolfaghari)

In 2014, Zolfaghari proposed a new extension of FLR toward Tanaka's model. There are two parameters that need to be considered in this model, either symmetric or asymmetric parameters. In addition, to determine the fuzzy parameters, the objectives of regression were done by following the linear programming method. This study will focus on the symmetric parameter. Under symmetric parameters, fuzzy coefficients are assumed as a triangular fuzzy number. The final model of FLR is also shown in (6). The linear programming problem is shown as in (8) [8].

Linear programming problem,

$$min = 2ms_0 + 2\sum_{q=1}^{N}[s_q \sum_{r=1}^{m}\left|x_{qr}\right|]$$

subject to c ≥0 and,

$$(1-H)s_0 + (1-H)\sum_{q=1}^{N}(s_q\left|x_{qr}\right|) - a_0 - \sum_{q=1}^{N}(a_q x_{qr}) \geq -y_r$$

$$(1-H)s_0 + (1-H)\sum_{q=1}^{N}(s_q\left|x_{qr}\right|) + a_0 - \sum_{q=1}^{N}(a_q x_{qr}) \geq y_r \tag{8}$$

where H is a degree of fitting, s is the spread and α is the center from triangular fuzzy numbers in symmetric parameter.

3.4. Statistical performance measurements

A statistical formula is used to evaluate the results from analysis to generalize the ability in prediction models and prevent overfitting. The main use of the statistical performance measurements is to compute how precisely a projecting model will occur in real life [20]. There are various types of performance measurements that could be used in statistical analysis. In this study, three methods are shown.

Mean square error (MSE) is represented as in (9),

$$\text{MSE} = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N} \tag{9}$$

mean absolute error (MAE) is represented as in (10),

$$\text{MAE} = \frac{\sum_{i=1}^{N}|y_i - \hat{y}_i|}{N} \tag{10}$$

mean absolute percentage error (MAPE) is represented as in (11),

$$\text{MAPE} = \frac{\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \text{x}100}{N} \tag{11}$$

where y_i is the real data
$\hat{y}_i$ is the predicted data
N is observations number

4. RESULTS AND DISCUSSION

Based on this research, the proposed and other existing methods have been modeled by including all the significant variables. Besides, the error value of MSE, MAE and MAPE for each model was calculated by adjusting the degree of fitting (*H*) from 0 to 1. Then *H* value was selected by obtaining the smallest error value. Furthermore, the details discussion of the analysis will be elaborated by the following sections.

4.1. Multiple linear regression

MLR model was analyzed toward nine explanatory variables that contribute in predicting manufacturing income. Based on two early assumption tests, the first normality result is shown in Figure 2. The Q-Q plot in Figure 2 showed the data were normally distributed since the linear line is nearly straight. Next, the variance inflation factor (VIF) values were shown in Table 1 as a result for multicollinearity test. The result of the analysis indicates that all VIFF value of the explanatory variable is less than 10, which specifies that dependencies are not severe enough for multicollinearity situation [24], [25].

Futhermore, an analysis of MLR indicated that only six explanatory variables are significant toward manufacturing income as shown in Table 1. The significant of independent variables is determined if the *p*-value < 0.05. Meanwhile, the correlation coefficient (*r*) is 0.993 and the determination coefficient (r^2) is 0.987, which indicates a strong positive linear correlation between *X*'s and *Y* variables. The error values are shown in Table 2 where MSE = 6620192158000, MAE = 191340.3873 and MAPE = 290.0394 accordingly.
The potential MLR model chosen is indicated as in (12),

$$\widehat{Y} = 1689.652 + 166.250x_1 - 1046.874x_2 + 0.056x_3 - 196.453x_7 - 23.868x_8 + 1.055x_9 \qquad (12)$$

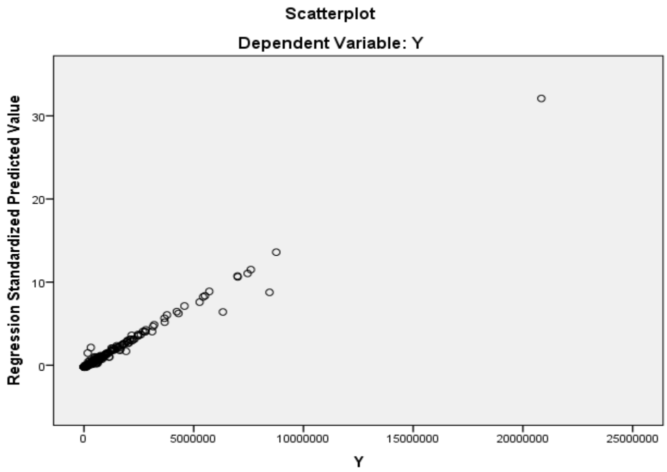

Figure 2. Normal Q-Q plot of MLR model

Table 1. Coefficient ($\boldsymbol{\beta}$) and variance inflation factor (VIF) values

Variables	Coefficient (β)	VIF
(Constant)	1689.652	-
x_1	166.250	1.018
x_2	-1046.874	1.051
x_3	0.056	1.226
x_7	-196.453	3.055
x_8	-23.868	3.237
x_9	1.055	1.936

Table 2. Result of errors for MLR model

MSE	MAE	MAPE
6620192158000	191340.3873	290.0394

4.2. Fuzzy linear regression (Tanaka)

In Tanaka's model, the degree of fitting in the FLR model has been adjusted between 0 and 1 to obtain the least error value as shown in Table 3. The best of MSE, MAE and MAPE values are 829657000000, 185519.8663 and 126.5645 respectively. The best model is shown as in (13) with *H* = 0.95 involving all explanatory variables. Table 4 shows the α_q as fuzzy centre of a parameter and c_q as the fuzziness of its parameter.

$$\hat{Y} = (4505, 0)+(553, 0)\ x_1+(-5702, 0)\ x_2+(-0.5711, 0)\ x_3+(-37137, 0)\ x_4+(1.7581, 0)\ x_5$$

$$+(37309, 0)\ x_6+(37036,0)\ x_7+(37134, 0)\ x_8+(2.5560, 43.8499)\ x_9 \quad (13)$$

Table 3. Result of measurement error for FLR (Tanaka model)

H	MSE	MAE	MAPE
0.1	1081480000000	213432.401	148.1297
0.2	1049280000000	210523.225	146.6079
0.3	1028260000000	207442.1272	144.5161
0.4	987919000000	204217.4769	143.5276
0.5	958462000000	201065.4216	138.6407
0.6	929897000000	197840.8844	136.2766
0.7	899143000000	193323.8679	134.4598
0.8	872273000000	190428.6030	132.8196
0.825	864476000000	189695.2996	131.4377
0.85	856749000000	188407.1895	129.2915
0.9	843124000000	186978.1420	128.2866
0.925	837978000000	186529.5707	127.7575
***0.95**	**829657000000**	**185519.8663**	**126.5645**
0.99	26747600000000	2346763.969	5091.5502

*Bold indicating best results

Table 4. Detailed of the antecedent parameter for FLR (Tanaka model)

Fuzzy Parameter	Fuzzy Center, α_q	Fuzzy Width, c_q
A_0	4505	0
A_1	553	0
A_2	-5702	0
A_3	-0.5711	0
A_4	-37137	0
A_5	1.7581	0
A_6	37309	0
A_7	37036	0
A_8	37134	0
A_9	2.5560	43.8499

4.3. Fuzzy linear regression (Zolfaghari)

Zolfaghari's model is also used in predicting manufacturing income. The MSE, MAE and MAPE values are 575629000000, 154153.8335 and 104.6929 respectively as shown in Table 5. The best model is shown as in (14) where H = 0.025. Table 6 shows that the α_q is fuzzy centre of a parameter and c_q is the fuzziness of its parameter.

$$\hat{Y} = (2961, 0) +(378, 0)\ x_1+(-3507, 0)\ x_2+(-0.4862, 0)\ x_3+(-28925, 0)\ x_4+(1.9951, 0)\ x_5$$

$$+(29029, 0)\ x_6+(28664, 0)x_7+(28918, 0)\ x_8 +(2.3061, 1.9699)\ x_9 \quad (14)$$

Table 5. Result of measurement error for FLR (Zolfaghari model)

H	MSE	MAE	MAPE
***0.025**	**575629000000**	**154153.8335**	**104.6929**
0.05	580762000000	154660.2431	106.7944
0.1	592114000000	156315.3180	107.9295
0.15	603390000000	157780.7371	107.6545
0.175	608365000000	158271.2320	108.8825
0.2	614717000000	159256.9142	109.7747
0.3	638015000000	162232.2573	111.6189
0.4	660062000000	164442.1358	111.5769
0.5	686064000000	168189.6212	112.8879
0.6	715225000000	173957.0441	121.1385
0.7	740641000000	176370.6562	121.7880
0.8	765726000000	179421.2831	124.6240
0.9	792142000000	182630.8518	127.3443

*Bold indicating best results

Table 6. Detailed of the antecedent parameter for FLR (Zolfaghari) model

Fuzzy Parameter	Fuzzy Center, α_q	Fuzzy Width, c_q
A_0	2962	0
A_1	388	0
A_2	-3517	0
A_3	-0.4863	0
A_4	-28935	0
A_5	1.9961	0
A_6	29129	0
A_7	28665	0
A_8	28928	0
A_9	2.3161	1.9599

4.4. Summary of results

Table 7 summarizes the experimental results for MLR, FLR of Tanaka and FLR of Zolfaghari models. The performance of three models was evaluated by using MSE, MAE and MAPE values. Among other, FLR proposed by Zolfaghari exhibit the best results with the lowest MSE, MAE and MAPE. Figure 3 shows the plot of real and expected data for manufacturing income.

Table 7. Summary of error measurement for three models

Models of Linear Regression	*H*	MSE	MAE	MAPE
MLR	-	6620192158000	191340.3873	290.0394
FLR (Tanaka)	0.95	829657000000	185519.8661	125.5645
FLR (Zolfaghari)	0.025	575619000000	154163.8335	105.6919

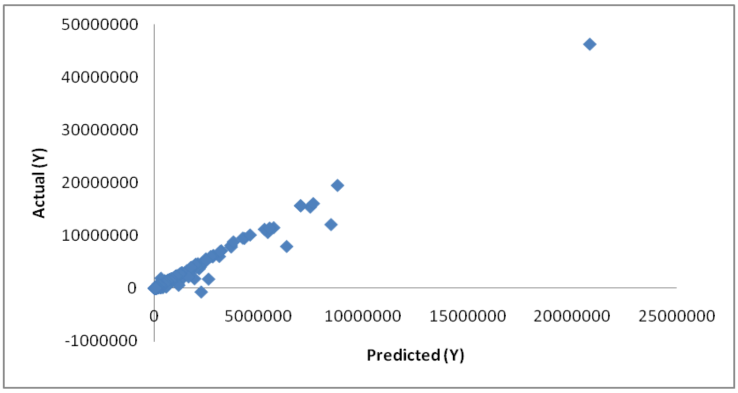

Figure 3. The actual and predicted values for manufacturing income

5. CONCLUSION

For the purpose of predicting manufacturing incomes, three models were applied: the MLR, the FLR proposed by Tanaka, and the FLR proposed by Zolfaghari. It appears that the FLR model introduced by Zolfaghari with H = 0.025 is the optimal model based on the MSE, MAE, and MAPE derived from all nine explanatory variables. By contrast, only six explanatory variables were significant according to the MLR. Since the MLR has the highest error values in comparison to the other methods, it cannot be used as a guide. Zolfaghari's FLR model indicates that manufacturing income is directly proportional to legal status, total salaries and wages paid, number of degree and diploma holders, number of SPM and below holders, and total expenditure. Additionally, manufacturing income is inversely proportional to ownership, asset value and number of employees. A manufacturer can use this output as a guide to improve their earnings.

ACKNOWLEDGEMENTS

The research work was supported by Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme Vot K297 (FRGS/1/2020/STG06/UTHM/02/4), Universiti Tun Hussein Onn Malaysia and Department of Statistics Malaysia (DOSM) for data collection.

REFERENCES

[1] S. M. Patil and R. Mudholkar, "Fuzzy regression model for knee osteoarthritis disease diagnosis," *International Research Journal of Engineering and Technology (IRJET)*, vol. 3, no. 2, pp. 366–369, 2016.

[2] A. Agresti, *An introduction to categorical data analysis*, Seciond. John Wiley & Sons, Inc., 2007.

[3] H. Thoni, J. Neter, W. Wasserman, and M. H. Kutner, "Applied linear regression models," *Biometrics*, vol. 46, no. 1, pp. 282–283, 1990, doi: 10.2307/2531657.

[4] K. D. Atalay, E. Eraslan, and M. O. Çinar, "A hybrid algorithm based on fuzzy linear regression analysis by quadratic programming for time estimation: an experimental study in manufacturing industry," *Journal of Manufacturing Systems*, vol. 36, pp. 182–188, 2015, doi: 10.1016/j.jmsy.2014.06.005.

[5] A. Chaudhuri and K. De, "Achieving greater explanatory power and forecasting accuracy with non-uniform spread fuzzy linear regression," *arXiv preprint arXiv:1307.1903*, 2013, doi: 10.48550/arXiv.1307.1903.

[6] H. Tanaka, S. Uejima, and K. Asai, "Linear regression analysis with fuzzy model," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-12, no. 6, pp. 903–907, 1982, doi: 10.1109/tsmc.1982.4308925.

[7] F. Chen, Y. Chen, J. Zhou, and Y. Liu, "Optimizing h value for fuzzy linear regression with asymmetric triangular fuzzy coefficients," *Engineering Applications of Artificial Intelligence*, vol. 47, pp. 16–24, 2016, doi: 10.1016/j.engappai.2015.02.011.

[8] Z. S. Zolfaghari, M. Mohebbi, and M. Najariyan, "Application of fuzzy linear regression method for sensory evaluation of fried donut," *Applied Soft Computing Journal*, vol. 22, pp. 417–423, 2014, doi: 10.1016/j.asoc.2014.03.010.

[9] M. A. Shafi and M. S. Rusiman, "The use of fuzzy linear regression models for tumor size in colorectal cancer in hospital of Malaysia," *Applied Mathematical Sciences*, vol. 9, pp. 2749–2759, 2015, doi: 10.12988/ams.2015.5175.

[10] M. A. Shafi, M. Saifullah, S. Ismail, and M. Ghazali, "A hybrid of multiple linear regression clustering model with support vector machine for colorectal cancer tumor size prediction," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 4, pp. 323–328, 2019, doi: 10.14569/IJACSA.2019.0100439.

[11] N. A. Samat and M. N. M. Salleh, "A study of data imputation using fuzzy c-means with particle swarm optimization," in *International Conference on Soft Computing and Data Mining*, 2016, pp. 91–100, doi: 10.1007/978-3-319-51281-5_10.

[12] K. Hussain and M. N. M. Salleh, "Analysis of techniques for ANFIS rule-base minimization and accuracy maximization," *ARPN Journal of Engineering and Applied Sciences*, vol. 10, no. 20, pp. 9739–9746, 2015, doi: 10.13140/RG.2.1.1324.5525.

[13] I. Mohamed, K. Khalid, and M. S. Yahya, "Combined estimating function for random coefficient models with correlated errors," *Communications in Statistics - Theory and Methods*, vol. 45, no. 4, pp. 967–975, Feb. 2016, doi: 10.1080/03610926.2013.853794.

[14] S. H. Choi and J. H. Yoon, "General fuzzy regression using least squares method," *International Journal of Systems Science*, vol. 41, no. 5, pp. 477–485, 2010, doi: 10.1080/00207720902774813.

[15] S. M. Taheri and M. Kelkinnama, "Fuzzy linear regression based on least absolutes deviations," *Iranian Journal of Fuzzy Systems*, vol. 9, no. 1, pp. 121–140, 2012.

[16] J. Li, W. Zeng, J. Xie, and Q. Yin, "A new fuzzy regression model based on least absolute deviation," *Engineering Applications of Artificial Intelligence*, vol. 52, pp. 54–64, Jun. 2016, doi: 10.1016/j.engappai.2016.02.009.

[17] J. P. Malaysia, "Malaysia standard industrial clasiification," *Jabatan Perangkaan Malaysia*, no. ISSN 1511-6824, pp. 1–203, 2000.

[18] F. Wahid and D. H. Kim, "Short-term energy consumption prediction in Korean residential buildings using optimized multi-layer perceptron," *Kuwait Journal of Science (KJS)*, vol. 44, no. 2, pp. 67–77, 2017.

[19] L. L. Nathans, F. L. Oswald, and K. Nimon, "Interpreting multiple linear regression: a guidebook of variable importance," *Practical Assessment, Research and Evaluation*, vol. 17, no. 9, pp. 1–19, 2012, doi: 10.7275/5fex-b874.

[20] M. S. Rusiman, S. N. M. Nor, S. Suparman, and S. N. A. M. Razali, "Robust method in multiple linear regression model on diabetes patients," *Mathematics and Statistics*, vol. 8, no. 2A, pp. 36–39, Mar. 2020, doi: 10.13189/ms.2020.081306.

[21] K. R. Das and A. Imon, "A brief review of tests for normality," *American Journal of Theoretical and Applied Statistics*, vol. 5, no. 1, pp. 5–12, 2016, doi: 10.11648/j.ajtas.20160501.12.

[22] L. Fahrmeir, T. Kneib, S. Lang, and M. Brian, *Regression models, methods and applications*. Springer, 2013.

[23] M. A. Shafi, M. S. Rusiman, N. S. A. Hamzah, S. F. Sufahani, A. Khamis, and N. A. H. M. Azmi, "Fuzzification of quantitative data to predict tumour size of colorectal cancer," *Far East Journal of Mathematical Sciences (FJMS)*, vol. 103, no. 5, pp. 951–959, 2018, doi: 10.17654/ms103050951.

[24] N. Shrestha, "Detecting multicollinearity in regression analysis," *American Journal of Applied Mathematics and Statistics*, vol. 8, no. 2, pp. 39–42, 2020, doi: 10.12691/ajams-8-2-1.

[25] M. J. Rezaee, M. Jozmaleki, and M. Valipour, "Integrating dynamic fuzzy C-means, data envelopment analysis and artificial neural network to online prediction performance of companies in stock exchange," *Physica A: Statistical Mechanics and its Applications*, vol. 489, pp. 78–93, 2018, doi: 10.1016/j.physa.2017.07.017.

BIOGRAPHIES OF AUTHORS

Nurfarawahida Ramly involves in decision-making in statistical data analysis. At the University Tun Hussein Onn Malaysia (UTHM), she learns the importance of applying data analysis, visualization, and mining on various sorts of data from experts in the field, including her supervisor, Assoc. Prof. Dr. Mohd Saifullah Bin Rusiman. Currently, she is pursuing her PhD program. She can be contacted at email: hw200002@siswa.uthm.edu.my.

Assoc. Prof. Dr. Mohd Saifullah Rusiman is a lecturer with 17 years of teaching experience at the University Tun Hussein Onn Malaysia (UTHM) under the Department of Mathematics and Statistics. He majored in mathematics at the University Teknologi Malaysia (UTM) and Dokuz Eylul University (DEU), Turkey, where he also earned his Doctor of Philosophy in Mathematics. His area of research interest was focused on applied statistics specifically statistics modeling, fuzzy statistics, and time series. He has published more than 90 articles where 52 of it has been indexed by scopus. He is also Chief Editor for the international journal of Enhanced Knowledge of Science and Technology (EKST) which was discovered in 2020. He can be contacted at email: saifulah@uthm.edu.my.

Shuhaida Ismail is a lecturer at the Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia (UTHM). She obtained her first degree in Computer Sciences majoring from UTM. She also obtained a Master degree and PhD from the same university. Throughout her studies, she developed an interest in Machine Learning research area, specifically in predictive modelling, classification, and clustering. Her current research areas are in Hydrological Modelling, Big Data Analytics and Deep Learning. She can be contacted at email: shuhaida@uthm.edu.my.

Assoc. Prof. Dr. Suparman is a lecturer at the Department of Mathematics Education, Ahmad Dahlan University (UAD), Indonesia. He obtained a Doctor of Philosophy in Applied Mathematics at Paul Sabatier University, France. His research interests are focused on applied statistics, especially reversible jump MCMC, bootstrap, simulated annealing, and its application to time series, signal processing, and finance. He can be contacted at email: suparman@pmat.uad.ac.id.

Assoc. Prof. Dr. Firdaus Mohamad Hamzah has been a lecturer at the Faculty of Engineering and Built Environment in the Universiti Kebangsaan Malaysia (UKM) since 2000. He obtained his BSc (Statistics) in 1998 and MSc (Quality Management and Productivity) in 1999 from UKM. He was conferred a PhD (Environmental Statistics) from the University of Glasgow, United Kingdom in 2012. He is currently a Senior Research Fellow at the Institute of Climate Change, UKM. He has published 50 journal articles, more than 90 proceeding papers, several books and technical papers. He has presented research papers in more than 30 conferences locally and internationally. He can be contacted at email: fir@ukm.edu.my.

Assoc. Prof. Dr. Özlem Gürünlü Alma is a lecturer at the Department of Statistics, University of Muğla Sıtkı Koçman University. She got a Ph.D. degree in Statistics in 2009. She finished post-doctoral studies at McMaster University Department of Mathematics and Statistics, Canada, under the supervision of Prof. N. Balakrishnan in 2013-2014. The main topic of the post-doctoral studies was the estimation problem in the censored data. She is teaching and has an expert in probability, censored models, robust data analysis, machine learning, and statistics courses, along with numerical analysis and coding in R. She can be contacted at email: ozlem.gurunlu@gmail.com.