Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

The Comparison of Fuzzy Regression Approaches with and without Clustering Method in Predicting Manufacturing Income

Nurfarawahida Ramly[1], Mohd Saifullah Rusiman[1,*], Efendi Nasibov[2], Resmiye Nasiboglu[2], Suparman[3]

[1] Department of Mathematics and Statistics, Faculty of Applied Science and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh, 84600 Muar, Johor, Malaysia
[2] Department of Computer Science, Faculty of Science, Dokuz Eylul University, 35220 Konak/Izmir, Turkiye
[3] Department of Mathematics Education, Faculty of Teacher Training and Education, Ahmad Dahlan University (UAD), Yogyakarta 55166, Indonesia

ARTICLE INFO

Article history:
Received 18 May 2023
Received in revised form 1 October 2023
Accepted 16 April 2024
Available online 22 May 2024

Keywords:

Fuzzy linear regression; Fuzzy least square regression; Fuzzy C-means; Manufacturing income; Root mean square error

ABSTRACT

In the manufacturing area, predicting future income is more important to keep maintain their industry profits. In addition to this, most of the manufacturing company having a bit problem in predicting their manufacturing income, especially in terms of data and method used. Hence, this paper proposed another improvise method of fuzzy regression approach with and without clustering method for uses of predicting manufacturing income. Then, one of the widely uses of statistical analysis are fuzzy regression approach such as fuzzy linear regression (FLR) and fuzzy least squares regression (FLSR). Furthermore, clustering is one of the most common methods for grouping data based on its similarity. Apart from this, fuzzy c-means (FCM) recognised as the best clustering method. This study's model was evaluated by three measurements errors: root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). Based on numerical calculations, it was determined that the proposed fuzzy least square regression with fuzzy c-means clustering model is superior to others, with RMSE = 59756.78229, MAE = 2948.616554, and MAPE = 13.34916083. Therefore, this model indicates as the robust method and suitable use for prediction analysis, especially in handling uncertain and imprecise data.

1. Introduction

Rapid growth in technology in recent years has made the global market more competitive. It is time for manufacturing companies to adapt to the changes in order to compete for the survival of their business. To acquire a competitive advantage in the market, manufacturers must constantly improve and come up with new ideas. The success of an organisation can be measured using a variety of criteria, including financial income performance, customer and staff satisfaction [1]. Although financial income performance plays an essential role in maintaining and improving a company's

* *Corresponding author.*
E-mail address: saifulah@uthm.edu.my

https://doi.org/10.37934/araset.46.1.218236

growth. This study will use the clustering approach for predicting manufacturing income, which is adaptable and important in data analysis.

Cluster analysis is defined as a process of separating a data or observation object into subsets and also known as unsupervised learning method [2]. The basic goal of the clustering method is to group objects with similar characteristics together in one cluster group. Currently, cluster analysis widely appears in various disciplines, including medicine, marketing, biology, image processing, data mining, psychology and other fields [3]. Furthermore, there are three main characteristics of clustering algorithms. First characteristic is either cluster algorithm in partitional or hierarchical. Aside from that, it means dividing the object or observation data into simple groups (partitional) or groups and subgroups (hierarchical). Second characteristic is either hard or soft clustering algorithms. Hard clustering is defined as assigning each item or observation data to a single class, whereas soft clustering assigns a probability of belonging to each class. The third feature is either centroid-based or density-based. In other words, centroid-based refers to objects or observation data based on their distance from the centre of the cluster, whereas density-based assigns items based on the local density around the object [4]. This main characteristic is important, particularly in establishing and constructing more predictable and flexible mechanisms for clustering data objects [3].

However, conducting a clustering analysis is not always possible, especially in complicated datasets with overlapping clusters. Therefore, the fuzzy c-means (FCM) clustering technique could be applied to solve the problem of overlapping structure. The FCM method is one of the most popular and is already widely used by many researchers. Because it produces better results for overlapped data sets, the fuzzy c-means clustering approach outperforms other common clustering methods such as k-means [5]. Moreover, Dunn [6] proposed FCM clustering, which was later improved by Hathaway *et al*., [7]. The membership degree is used in the FCM clustering analysis to measure the uncertainty that a data object or observation belongs to two or more classes [8]. In addition, cluster validation refers to the process of evaluating the outcomes of a clustering technique. Cluster validation is provided as a method for assessing clustering results and identifying the optimal number of clusters [2]. External and internal validation are the two primary approaches to partition validation. [9]. The difference is whether or not prior information is used in the validation process.

Besides of clustering, there is another method widely used in data analysis particularly for modelling and estimating the data such fuzzy regression approach. Fuzzy regression approach is defined as a way of approaching problems involving ambiguity and vagueness. Currently, it is a popular method used by mathematicians, researchers, engineers, medical and natural scientists, business analysts, and computer software developers [10]. Zadeh was the first to propose a fuzzy approach to computing problems [11]. It's allowing for the inclusion of ambiguous human assessments and controls complex systems more effectively than classic methods [3]. Therefore, fuzzy regression is a useful tool for assessing the ambiguous relationship between dependent and independent variables. This paper explores two types of fuzzy regression models: fuzzy linear programming and fuzzy least squares.

Tanaka *et al*., [12] introduced fuzzy linear programming, commonly referred to as fuzzy linear regression (FLR), for resolving problems with fuzzy structures and human estimation. The model was modified by Zolfaghari *et al*., [13], who introduced the assumption of triangular fuzzy numbers (TFNs) that are either symmetrical or asymmetrical and are then represented by their own membership function. Other than that, fuzzy least squares regression (FLSR) techniques are an extension to the least squares method to solve for fuzzy linear regression [14]. The purpose of the fuzzy least squares regression approach is to obtain values for fuzzy parameters by minimizing the sum of the squared distance between the observed and estimated responses. As well as that, the fuzzy least squares approach contains the concept of goodness of fit, and the residuals can be implemented for evaluating model accuracy [15].

According to previous researchers, the majority of them evaluated the analysis using a single existing method. The performance of the method can be measured by computing the error value of RMSE, MAE, MAPE and comparing it to another method to find the lowest one. However, when the error values are not far apart, there is a notable research gap. This means that the best prediction model should produce a lowest error value that is significantly different from another. To address this research gap, a combination of fuzzy regression and clustering algorithms would be used. This combination method will produce the most optimal model that can be used to the industry sector. Furthermore, these models with smaller error can be used to maintain industry profit and estimate manufacturing income more accurately. Hence, industrial planning can be improved for future income prediction. In addition, the aim of this study is to refine the fuzzy regression approach with clusters and non-clusters in predicting manufacturing income.

2. Material and Methods

The current study was carried out according with the research framework shown in Figure 1. The datasets were gathered from the Department of Statistics Malaysia (DOSM), which is involved in a variety of industry sectors such as manufacturing, farming, transportation, quarrying, fishing, building, and many others [16]. The industries data were filtered, and it was chosen to focus just on the manufacturing sector in order to enhance the analysis of this study. There are nine exploratory variables: 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 & wages paid, number of degree & above holders, number of diploma holders, number of Malaysian Certificate of Education (MCE) & below holders, and total expenditure, while total income is the dependent variable.

In addition, unlike other classic regression methods, there are no assumptions to consider before proceeding with the study. After removing any missing values, the data was transmitted into the main approach, which included fuzzy linear regression (FLR) and fuzzy least square regression (FLSR) with and without clustering. The error values were analysed afterwards using a variety of performance measures such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) [17]. The RMSE, MAE, and MAPE data were then compared to determine the best prediction model that may be used in the industry sector. Finally, Section 2 describes the recommended methods. Section 3 provides numerical results and a discussion of the proposed approaches, followed by conclusions in Section 4.

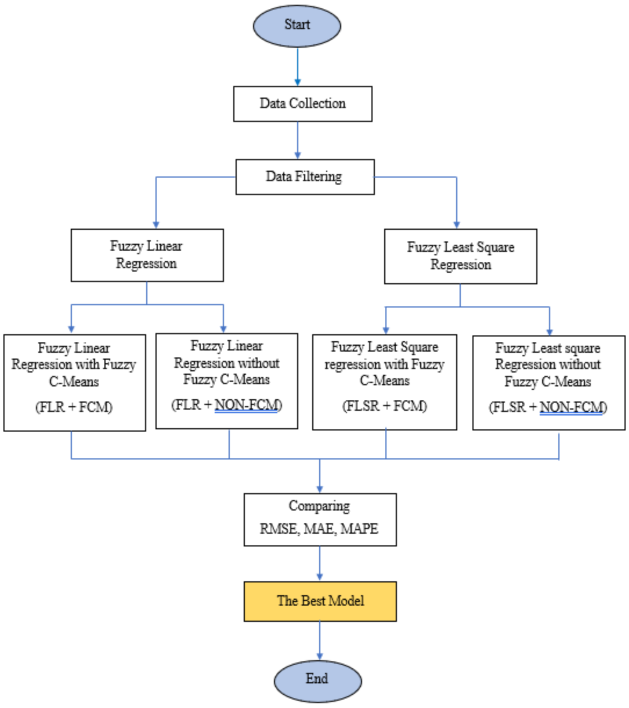

Fig. 1. Research framework of proposed model

2.1 Fuzzy C-Means (FCM)

Clustering is a method for unsupervised learning. It is also defined as the process of separating a data or observation object. Furthermore, the main goal of the clustering method is to group objects that are similar to each other [2]. Clustering is also frequently used in a wide range of industries, including business, image analysis, data mining, data processing, pattern recognition, and many others. Fuzzy c-means (FCM) is the most extensively used clustering approach, often known as the soft clustering method. Dunn [6] proposed FCM at first, and Hathway *et al.*, [7] later modified it.

In addition, the FCM algorithm divides cluster partitioning into the degree to which data points belong to a specific class. Meanwhile, other clustering methods, such as hard clustering, only consider data points that correspond to a specific class of divisions. The majority of clustering algorithms are based on minimizing an objective function to obtain the most compact clusters placed in dense data regions [18]. FCM is defined as an objective function and is shown as.

$$Jm(U,V)=\sum_{k=1}^{C}\sum_{l=1}^{N}u_{kl}^{m}d_{kl}^{2}$$

$$=\sum_{k=1}^{C}\sum_{l=1}^{N}u_{kl}^{m}\left\|x_l-v_k\right\|^2 \tag{1}$$

where U is the membership matrix, V is represent as cluster matrix, N is total number of data objects in the data set, C is the number of clusters, m is the fuzziness parameter in FCM clustering, v_k is the cluster centre of k, u_{kl} is the membership degree of lth data object x_l to cluster v_k and d_{kl}^2 is the Euclidean distance [8]. The fuzzy c-means algorithm is then summarised as follows.

i. Start
ii. Randomly initialise the cluster centre 'v_k' and fuzzy membership matrix 'u_{kl}'. Choose fuzziness parameter, $m > 1$ (usually $m = 2$).
iii. Update the membership matrix $U = [u_{kl}]$:

$$u_{kl} = \frac{1}{\sum_{q=1}^{C}\left(\frac{d_{kl}}{d_{kq}}\right)^{\frac{2}{m-1}}} \tag{2}$$

iv. where $u_{kl} \in [0,1]$, $\sum_{k=1}^{C} u_{kl} = 1$, $0 < \sum_{l=1}^{N} u_{kl} < N$
v. Update the cluster centre $V = [v_k]$:

$$v_k = \frac{\sum_{l=1}^{N}(u_{kl}^{m} x_l)}{\sum_{l=1}^{N} u_{kl}^{m}} \tag{3}$$

vi. Calculate the objective function Jm by continuous iteration until reaches a minimum value points or target value.
vii. Determine the findings for the final cluster centre.
viii. End

2.2 Cluster Validation

The cluster validation technique can be used to measure clustering results, which has been recognized as an essential component for the success of clustering applications. Moreover, as an unsupervised learning technique, a method for evaluating the validity of partitions after clustering is required [18]. Aside from that, external and internal clustering validation techniques are available. External validation methods assess the quality of clusters using external criteria obtained from prior knowledge or external data. The internal validation is then based on intrinsic criteria generated from the data used to develop the clustering algorithm [19].

This paper only concerns internal validation. Internal validation evaluates the fuzzy c-means clustering algorithm, and cluster validity is measured by the Xie-Beni index (*XB*). Furthermore, Xie *et al.,* [20] proposed the *XB* index, which takes both membership degree values and data information into account. Besides, the *XB* index information is based on the compactness of data objects within the same cluster and the separation across clusters [8]. The *XB* index formula is then applied, as indicated,

$$XB = \frac{\sum_{k=1}^{C}\sum_{l=1}^{N} u_{kl}^{2} d_{kl}^{2}}{N \times \min_{k \neq l} \left\| v_k - v_l \right\|^2} \tag{4}$$

where, $d_{kl}^2 = \|x_l - v_k\|^2$, the numerator of the *XB* index represents compactness validity, which assesses how tightly connected the objects in a cluster are, while the denominator represents separation validity. It evaluates how distinct a cluster is from other clusters. As a result, the best cluster number is identified when the *XB* index value is minimised.

2.3 Fuzzy Linear Regression (FLR)

In the modelling of some systems that are influenced by human estimation, it is necessary to deal with a fuzzy structure that doesn't take any assumptions into consideration before conducting the analysis. In general, this structure is recognized as a fuzzy linear function with fuzzy set parameters. Tanaka *et al.*, [12] then formulates and first proposes fuzzy linear regression. Classic regression models can only fit crisp data and depend on probability theory. Meanwhile, in the fuzzy regression model, it is possible to fit both fuzzy and crisp data into the classic regression model. Moreover, these analyses are based on possibility and fuzzy set theory, as well as the fuzziness of uncertainty [3]. As a result, it revealed that the fuzzy regression model is a more acceptable and adaptable method of dealing with uncertainty concerns.

Besides, there are two approaches for estimating the common factor parameter of fuzzy linear regression (FLR):

i. linear programming
ii. fuzzy least squares.

The FLR model is illustrated as follows:

$$\tilde{Y}_k = f(x,\beta) = (\tilde{\beta}_0 + \tilde{\beta}_1 x_{kl} + \ldots + \tilde{\beta}_l x_{kl} + \ldots + \tilde{\beta}_n x_{kn}) \quad (5)$$

where $\tilde{Y}_k$ represents as fuzzy output, $\tilde{\beta}_k$ for $k = 1,\ldots,n$ are fuzzy coefficients, $x = (x_1, x_2, \ldots, x_n)$ is an n-dimensional crisp input vector and x_{kl} is the jth observed value of ith input variable.

In 2014, Zolfaghari *et al.*, [13] developed an extension model of Tanaka's fuzzy linear regression model that utilized fuzzy numbers and their membership functions. The fuzzy coefficients are considered to be triangular fuzzy numbers (TFNs) based on the FLR model. Apart from this, TFNs can be represented with triple $\tilde{\beta} = (a, s^L, s^R)$ where a, s^L and s^R are mode, left spread and right spread respectively. Hence, the coefficients can be characterised by membership function $\tilde{\beta}(x)$. Furthermore, there are two types of TFNs in this model: symmetrical and asymmetrical parameters, each with its own membership function [21]. This study concentrated on symmetrical parameters, which occur when the two spreads are equal. The FLR symmetrical model is defined as follows:

$$\min = 2ms_0 + 2\sum_{k=1}^{N}[s_k \sum_{l=1}^{m}|x_{kl}|] \quad (6)$$

Subject to

$$\begin{aligned}&(1-H)s_0+(1-H)\sum_{k=1}^{N}(s_k|x_{kl}|)-a_0-\sum_{k=1}^{N}(a_k x_{kl})\geq -y_l\\&(1-H)s_0+(1-H)\sum_{k=1}^{N}(s_k|x_{kl}|)+a_0-\sum_{k=1}^{N}(a_k x_{kl})\geq y_l\end{aligned} \tag{7}$$

where x_{kl} is the lth observed value of the kth input variable. Then, the membership function as shows as Eq. (8) and Figure 2 illustrated the symmetrical TFNs.

$$\tilde{Y}(x)=\begin{cases}1-\dfrac{a-x}{s}, a-s\leq x\leq a\\1-\dfrac{x-a}{s}, a<x\leq a+s\end{cases} \tag{8}$$

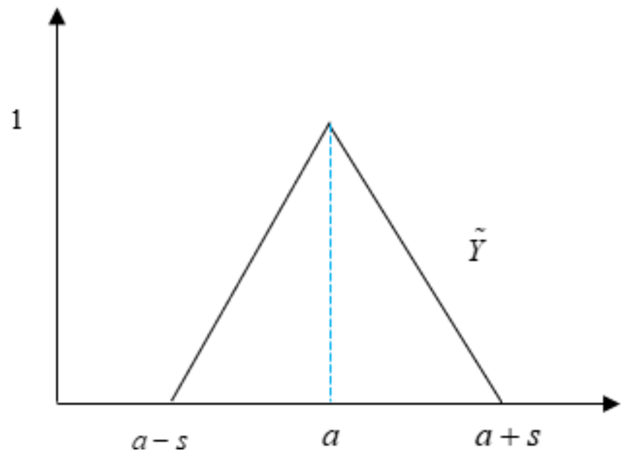

Fig. 2. Membership function of symmetrical TFNs

2.4 Fuzzy Least Square Regression (FLSR)

Aside from linear programming, another method for estimating common factor parameters in fuzzy linear regression is the fuzzy least squares method. Furthermore, in standard linear regression, the least squares method is always used to identify the β parameters that minimize the sum of squared errors, as shown below:

$$S(\beta_0,\beta_1,\beta_2,...,\beta_n)=S(\beta)=\sum\varepsilon^2 \tag{9}$$

Alternatively, the Fuzzy-Least Square Regression (FLSR) technique was introduced as an extension of the traditional least square approach to fuzzy linear regression [14]. The FLSR model is written like this:

$$\tilde{Y}=\tilde{\beta}_0+\tilde{\beta}_1x_1+...+\tilde{\beta}_nx_n+\tilde{\varepsilon} \tag{10}$$

where $\tilde{Y}$ is the fuzzy response, n is the number of variables, $\tilde{\beta}_0...\beta_n$ are represents as fuzzy parameters of model, $x_1...x_n$ are input variables and $\tilde{\varepsilon}$ is the fuzzy error. Then, the aim of these FLSR is to find the fuzzy parameter values $\tilde{\beta}$ that minimize the sum of the squared distance [22] between observed and estimated response $\sum_{k=1}^{m} d^2(y,\tilde{y})$.

In addition to this, triangular fuzzy numbers (TFNs) are used to represent the fuzzy parameters $\tilde{\beta}$ as follows:

$$\tilde{\beta} = (\beta_L, \beta_A, \beta_R) \tag{11}$$

where β_A is the middle, β_L is the left point and β_R is the right point of triangular fuzzy numbers. Moreover, these distance metric adopted by Diamond [22] to measure the distance between of two TFNs $(\tilde{y}_1 = (a_{1L}, a_{1A}, a_{1R})$ and $(\tilde{y}_2 = (a_{2L}, a_{2A}, a_{2R}))$. Then, the Diamond distance formulated as:

$$d^2(\tilde{y}_1,\tilde{y}_2) = (a_{1L}-a_{2L})^2 + (a_{1A}-a_{2A})^2 + (a_{1R}-a_{2R})^2 \tag{12}$$

Based on $S(\tilde{\beta}_0,...,\tilde{\beta}_n) = \sum_{k=1}^{m} d^2(y,\tilde{y})$,

$$S(\tilde{\beta}_0,...,\tilde{\beta}_n) = \sum_{k=1}^{m} d^2(y_k, \tilde{\beta}_0 + \tilde{\beta}_1 x_{k1} +,...,+ \tilde{\beta}_n x_{kn}) \tag{13}$$

$$= \sum_{k=1}^{m} [(y_{kL} - \beta_{0L} - \beta_{1L}x_{k1} - ... - \beta_{nL}x_{kn})^2 + (y_{kA} - \beta_{0A} - \beta_{1A}x_{k1} - ... - \beta_{nA}x_{kn})^2$$

$$+ (y_{kR} - \beta_{0R} - \beta_{1R}x_{k1} - ... - \beta_{nR}x_{kn})^2] \tag{14}$$

2.5 Cross Validation Technique

There are numerous formulas that can be used to evaluate the effectiveness and performance of fuzzy approaches for predicting manufacturing income [23,24]. The formula is used to compute the error value for each model in this study. The formula is recognized as root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), as illustrated below.

i. Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}} \tag{15}$$

ii. Mean Absolute Error (MAE)

$$MAE = \frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|}{N} \tag{16}$$

iii. Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{\sum_{i=1}^{N} \frac{|y_i - \hat{y}_i|}{y_i}}{N} \times 100\% \tag{17}$$

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

3. Numerical Results

In Section 2, proposed methods for conducting and modelling manufacturing data were described. In this section, some numerical findings are shown, and a comparison of error numbers for each improvisation model is made in the table summary at the end.

3.1 Improvisation of Fuzzy Linear Regression with Fuzzy C-Means

This section illustrates how to improve the fuzzy approach in order to produce a new model suitable for predicting manufacturing income. Apart from that, no assumptions are required to proceed with this fuzzy analysis. In addition, four improved models are used in this study: fuzzy linear regression with fuzzy c-means (FLR + FCM), fuzzy linear regression without fuzzy c-means (FLR + NON-FCM), fuzzy least squares regression with fuzzy c-means (FLSR + FCM), and fuzzy least squares regression without fuzzy c-means (FLSR + NON-FCM).

First and foremost, in the first model of improvisational fuzzy linear regression with fuzzy c-means, the correlation value for all variables must be measured, as shown in Table 1. According to Table 1, a few variables have the highest correlation values, such as x_5, x_6 and x_9. The highest correlation value is chosen from among the others. This is because the higher the correlation value, the stronger the linear correlation.

Table 1
Correlation Values

Variables	Correlation value
x_1	0.360
x_2	0.122
x_3	0.364
x_4	0.494
x_5	0.601
x_6	0.681
x_7	0.539
x_8	0.416
x_9	0.993

*Highlight indicating best result

The three variables were compared using the error values: root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), as shown in Table 2. According to the error value results in Table 2, the x_9 variable obviously achieved the minimum error value as compared to the other variables, with RMSE = 6153011311, MAE = 11967.229, and MAPE = 9.644248998. The value here was chosen because the smallest the error value, the better the result of the model.

Table 2

Comparison error values between x_5, x_6 and x_9

Variables	RMSE	MAE	MAPE
x_5	279722000000	129374.655	209.6246094
x_6	234122000000	110190.567	143.1104971
x_9	6153011311	11967.229	9.644248998

*Highlight indicating best result

Next, based on variable x_9 as a smallest error value then proceed to calculated for determine the number of clusters. The number of clusters were obtained by using Xie-Beni index as shows in Eq. (4). Table 3 represents the results number of clusters and XB-value. Addition to this, the number of clusters chosen is two ($c = 2$) because of the XB-value reached the minimum value which close to zero with 0.0260 compared to others value.

Table 3

The value of c and XB for x_9

Number of clusters, c	XB -value
2	0.0260
3	0.0457
4	0.1218
5	46.3006

*Highlight indicating best result

Table 4 states the details of x_9 variable in two clusters. Apart from this, the amount of data for cluster 1 is 1727 with minimum value 1835 and maximum value 19962439. Meanwhile, the amount of data for cluster 2 is 1129 with minimum value 395 and maximum value is 69030. Then, fuzzy c-means clustering method is used in this study to obtain the better model with smaller error and goes to contribute the best prediction model for manufacturer sector.

Table 4

The details of cluster for x_9 variable

	Cluster 1	Cluster 2
Number of data	1727	1129
Minimum value	1835	395
Maximum value	19962439	69030

3.1.1 Fuzzy linear regression model on cluster 1

While finished of clustered the data then proceed to construct the model by using fuzzy linear regression method. In method of FLR model, it is constructed by involvement of all independent's variable includes $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9$. Next, this FLR model were evaluated by three measuring performance of error which are RMSE, MAE and MAPE. Addition to this, the three method of measuring performance error also evaluated by the degree of fitting (H) which adjusted between 0 to 1 as shown in Table 5. Based on the measurement error results in Table 5, the error value for (H=0.1) was chosen since it attained the minimum error value when compared to others. This is because the FLR model with the lowest error value is the best.

Table 5
Measurement error of FLR with FCM on cluster 1

H	RMSE	MAE	MAPE
0.1	577827.6493	139823.9623	55.83590993
0.2	657486.8036	160115.6377	64.10465592
0.3	734609.4506	181364.2182	73.54123016
0.4	812010.8187	202875.2327	83.39534963
0.5	890502.9914	224696.1487	93.56038054
0.6	971561.8260	244844.7727	101.3474867
0.7	1054619.750	266782.6035	111.0015318
0.8	1137066.212	287990.5583	119.8740060
0.9	1225313.707	308624.7297	127.0950877

*Highlight indicating best result

Table 6 represents the value of symmetrical fuzzy parameter for each variable together with mode value and spread value. According to the results of mode and spread value from Table 6, the best fuzzy linear regression model for cluster 1 was obtained and it is denoted as Eq. (18).

Table 6
Symmetrical fuzzy parameter of FLR with FCM on cluster 1 (*H*=0.1)

Fuzzy parameter	Mode (a)	Spread (s)
x_1	2971	0.0000
x_2	-6706	0.0000
x_3	-0.2812	0.0000
x_4	99830	0.0000
x_5	2.9078	0.0000
x_6	-99774	0.0000
x_7	-99999	0.0000
x_8	-99890	0.0000
x_9	1.7887	2.6149

$$\hat{Y} = (2971,0)x_1 + (-6706,0)x_2 + (-0.2812,0)x_3 + (99830,0)x_4 + (2.9078,0)x_5 + (99774,0)x_6 + (99999,0)x_7 + (99890,0)x_8 + (1.7887,2.6149)x_9 \quad (18)$$

3.1.2 Fuzzy linear regression model on cluster 2

Moreover, this part is FLR model for cluster 2 and it is also involvement of all independent variable. As well as that, this FLR model were evaluated by three measuring performance of error which are RMSE, MAE and MAPE. Then, the three method of measuring performance error also evaluated by the degree of fitting (H) which adjusted between 0 to 1 as shown in Table 7. Based on the measurement error results in Table 7, the error value for (H=0.5) was chosen since it attained the minimum error value when compared to others.

Table 7
Measurement error of FLR with FCM on cluster 2

H	RMSE	MAE	MAPE
0.1	4751.349331	3167.154190	22.48345327
0.2	4757.150044	3199.510172	22.85063026
0.3	4734.875931	3183.247502	22.89043187
0.4	4628.844404	3102.916158	22.59961100
0.5	4516.655735	3080.405814	22.10813925
0.6	4625.153004	3117.518008	23.03844838
0.7	4563.620050	3115.478781	23.43269981
0.8	4520.839218	3133.266266	24.50647495
0.9	4799.104807	3336.045642	25.47829029

*Highlight indicating best result

Table 8 represents the value of symmetrical fuzzy parameter for each variable together with mode value and spread value. According to the results of mode and spread value from Table 8, the best fuzzy linear regression model for cluster 2 was obtained and it is denoted as Eq. (19).

Table 8
Symmetrical fuzzy parameter of FLR with FCM on cluster 2 (*H*=0.5)

Fuzzy parameter	Mode (a)	Spread (s)
x_1	-595	0.0000
x_2	1354	0.0000
x_3	-0.1879	0.0000
x_4	99843	0.0000
x_5	0.4117	0.0000
x_6	-99713	0.0000
x_7	-99999	0.0000
x_8	-99831	0.0000
x_9	1.2063	2.8136

$$\hat{Y} = (-595,0)x_1 + (1354,0)x_2 + (-0.1879,0)x_3 + (99843,0)x_4 + (0.4117,0)x_5 + (-99713,0)x_6 + (-99999,0)x_7 + (-99831,0)x_8 + (1.2063,2.8136)x_9 \quad (19)$$

3.1.3 Total overall error (TOE) value for improvisation of FLR with FCM

This section displays the total overall error (TOE) computation of improvised FLR with FCM for both cluster 1 and cluster 2. The formula of TOE is shown in Eq. (20). The main goal of this TOE formula is to produce one perfect error value while having several clusters. Based on TOE formula, the results as shown in Table 9 with RMSE = 64658.700842, MAE = 3014.556413 and MAPE = 20.84784438 respectively.

$$TOE_{new} = \frac{(e_{c1})(n_1) + (e_{c2})(n_2)}{n_1 + n_2} \quad (20)$$

where: e_{c1} = smallest error value for cluster 1
e_{c2} = smallest error value for cluster 2
n_1 = total number of data in cluster 1
n_2 = total number of data in cluster 2

Table 9
The details value of total overall value for FLR with FCM

	RMSE	MAE	MAPE
Total overall error (TOE)	64658.700842	3014.556413	20.84784438

3.2 Improvisation of Fuzzy Linear Regression without Fuzzy C-Means

In this part, the model constructed without clustering method and obtained the model by using fuzzy linear regression method only. The way is same goes as FLR model within cluster FCM in terms of involvement of all independent variable. Besides, this FLR model were evaluated by three measuring performance of error which are RMSE, MAE and MAPE. Then, the three method of measuring performance error also evaluated by the degree of fitting (H) which adjusted between 0 to 1 as shown in Table 10. Based on the measurement error results in Table 10, the error value for (H=0.1) was chosen since it attained the minimum error value when compared to others.

Table 10
Measurement error of FLR without FCM

H	RMSE	MAE	MAPE
0.1	141345.3206	29064.79551	21.16394758
0.2	151430.3405	29616.40936	21.46012203
0.3	155255.6679	30477.42920	22.64868876
0.4	145381.4602	31848.07017	28.60358976
0.5	278227.8651	58012.71986	44.68049384
0.6	411286.8224	84127.49343	61.44023242
0.7	531830.5499	108325.7055	76.66431895
0.8	655363.7705	133732.6701	93.60839277
0.9	780042.8384	159584.4594	111.36039500

*Highlight indicating best result

Table 11 represents the value of symmetrical fuzzy parameter for each variable together with mode value and spread value. According to the results of mode and spread value from Table 11, the best fuzzy linear regression model was obtained and it is denoted as Eq. (21).

Table 11
Symmetrical fuzzy parameter of FLR without FCM (*H*=0.1)

Fuzzy parameter	Mode (a)	Spread (s)
x_1	-322	0.0000
x_2	910	0.0000
x_3	-0.0986	0.0000
x_4	99903	0.0000
x_5	-2.4304	0.0000
x_6	-99727	0.0000
x_7	-99999	0.0000
x_8	-99862	0.0000
x_9	1.2792	1.8697

$$\hat{Y} = (-322,0)x_1 + (910,0)x_2 + (-0.0986,0)x_3 + (99903,0)x_4 + (-2.4304,0)x_5 + (-99727,0)x_6 + (-99999,0)x_7 + (-99862,0)x_8 + (1.2792,1.8697)x_9 \quad (21)$$

3.3 Improvisation of Fuzzy Least Squares Regression with Fuzzy C-Means

In constructed of this model, the process of clustering method has been used same as model of improvisation fuzzy linear regression with fuzzy c-means at section 3.1 previously. Apart from this, the difference is how to modelled it for each cluster by evaluating the performance of error RMSE, MAE and MAPE. Then, for further analysis by clustered has been shown in following section 3.3.1 and 3.3.2.

3.3.1 Fuzzy least squares regression on cluster 1

In this model of cluster 1, has been involves of all nine (9) independent variables same as shows in previous section. Besides, it is was evaluated by three measurements error of RMSE, MAE and MAPE as represented in Table 12.

Table 12
Measurement error of FLSR with FCM on cluster 1

RMSE	MAE	MAPE
96830.72503	20356.44033	13.15705263

Moreover, in Table 13 represents the value of fuzzy parameter for each variable together with diamond distance value. According to the results of diamond distance value from Table 13, the best fuzzy least squares regression model for cluster 1 was obtained and it is denoted as Eq. (22).

Table 13
Fuzzy parameter of FLSR with FCM on cluster 1

Fuzzy parameter	Diamond distance (d^2)
x_1	1044
x_2	-1284
x_3	0.0557
x_4	-203.3631
x_5	1.8372
x_6	192.1549
x_7	0
x_8	178.9803
x_9	1.0554

$$\hat{Y} = (1044)x_1 + (-1284)x_2 + (0.0557)x_3 + (-203.3631)x_4 + (1.8372)x_5 + (192.1549)x_6 + (0)x_7 + (178.9803)x_8 + (1.0554)x_9 \tag{22}$$

3.3.2 Fuzzy least squares regression on cluster 2

In this part describes details fuzzy least squares regression on cluster 2 which has been involves of all nine (9) independent variables. In addition to this, it is also evaluated by three measurements error of RMSE, MAE and MAPE for cluster 2 as shown in Table 14.

Table 14
Measurement error of FLSR with FCM on cluster 2

RMSE	MAE	MAPE
3045.799905	1617.07389	13.64302341

Furthermore, in Table 15 represents the value of fuzzy parameter for each variable together with diamond distance value. According to the results of diamond distance value from Table 15, the best fuzzy least squares regression model for cluster 2 was obtained and it is denoted as Eq. (23).

Table 15
Fuzzy parameter of FLSR with FCM on cluster 2

Fuzzy parameter	Diamond distance (d^2)
x_1	219.0100
x_2	181.7799
x_3	-0.0471
x_4	27.6961
x_5	0.5261
x_6	0
x_7	-46.5326
x_8	-31.9116
x_9	0.9638

$$\hat{Y} = (219.0100)x_1 + (181.7799)x_2 + (-0.0471)x_3 + (27.6961)x_4 + (0.5261)x_5 + (0)x_6 + (-46.5326)x_7 + (-31.9116)x_8 + (0.9638)x_9 \quad (23)$$

3.3.3 Total overall of error for improvisation of FLSR with FCM

This Section illustrates the total overall error (TOE) computation of improvising FLSR with FCM for both clusters, and the TOE formula is shown in Eq. (20). Based on TOE formula, the results as shown in Table 16 with RMSE = 59756.78229, MAE = 2948.616554 and MAPE = 13.34916083 respectively.

Table 16
The details value of total overall value for FLSR with FCM

	RMSE	MAE	MAPE
Total overall error (TOE)	59756.78229	2948.616554	13.34916083

3.4 Improvisation of Fuzzy Least Squares Regression without Fuzzy C-Means

This part was explained the details of model improvisation fuzzy least squares regression without clustering method. This model constructed by involves of all nine (9) independent variables like previously. Three measurements error of RMSE, MAE and MAPE was applied to evaluates the data and the results as shown in Table 17.

Table 17
Measurement error of FLSR without FCM

RMSE	MAE	MAPE
75339.01187	13161.03283	15.05751384

Then in Table 18 represents the value of fuzzy parameter for each variable together with diamond distance value. According to the results of diamond distance value from Table 18, the best improvisation of fuzzy least squares regression without FCM model was obtained and it is denoted as Eq. (24).

Table 18
Fuzzy parameter of FLSR without FCM

Fuzzy parameter	Diamond distance (d^2)
x_1	639.1437
x_2	-803.646
x_3	0.0558
x_4	-201.9910
x_5	1.8310
x_6	190.8872
x_7	0
x_8	177.8652
x_9	1.0554

$$\hat{Y} = (639.1437)x_1 + (-803.646)x_2 + (0.0558)x_3 + (-201.9910)x_4 + (1.8310)x_5 + (190.8872)x_6 + (0)x_7 + (177.8652)x_8 + (1.0554)x_9 \quad (24)$$

3.5 Table Summary

Based on results for all model at previous part, this section was summaries and compared it each other for choose which one is the best model. Table 19 summarises the error values for each model used in the current study. In addition, each model was tested using three methods of error measurement: RMSE, MAE, and MAPE. According to the results, the model of improvisation FLSR with FCM exhibits the best model by attaining the lowest error value when compared to other models.

Table 19
Summary of measurements error for all improvisation model

	RMSE	MAE	MAPE
FLR + FCM	64658.7008	3014.5564	20.8478
FLR + NON-FCM	141345.3206	29064.7955	21.1639
FLSR + FCM	59756.7823	2948.6166	13.3492
FLSR + NON-FCM	75339.0119	13161.0328	15.0575

*Highlight indicating best result

4. Conclusions

Manufacturer sector demand models that is capable to use for prediction of manufacturing income. Four different improve modelling approaches namely Fuzzy Linear Regression with Fuzzy C-Means (FLR + FCM), Fuzzy Linear Regression without Fuzzy C-Means (FLR + NON-FCM), Fuzzy Least Squares Regression with Fuzzy C-Means (FLSR + FCM) and Fuzzy Least Squares Regression without Fuzzy C-Means (FLSR + NON-FCM) were applied to the manufacturing income prediction.

According to the numerical results, the superior performance of the model improvisation FLSR with FCM shows the best model and is capable of being used in prediction. This model was picked because it has the lowest error value when compared to another model. The model with the lowest error value is regarded as the best model and is recommended for use in prediction operations. Therefore, this improvised model is found to be superior to models in predicting future manufacturing income for industry firms.

Acknowledgement
The research work was supported by Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS/1/2020/STG06/UTHM/02/4). We also want to thank to the Department of Statistics Malaysia (DOSM) for the manufacturing data collection.

References
[1] Chandren, Thiagu P., Fatimah Salim, Noorlizawati Abd Rahim, Wan Normeza Wan Zakaria, and Siti Hasliah Salleh. "Determinants of Turnover Intention Among Employees: A Case Study of Manufacturing Company." *Journal of Advanced Research in Business and Management Studies* 25, no. 1 (2021): 26-37.
[2] Zhu, Ling-Feng, Jie-Sheng Wang, and H. Y. Wang. "A novel clustering validity function of FCM clustering algorithm." *IEEE Access* 7 (2019): 152289-152315. https://doi.org/10.1109/ACCESS.2019.2946599
[3] Ramly, Nurfarawahida. "Improvisation of fuzzy c-means method and fuzzy linear regression model in predicting manufacturing income." PhD diss., Universiti Tun Hussein Onn Malaysia, 2018.
[4] Garcia-Dias, Rafael, Sandra Vieira, Walter Hugo Lopez Pinaya, and Andrea Mechelli. "Clustering analysis." In *machine learning*, pp. 227-247. Academic Press, 2020. https://doi.org/10.1016/B978-0-12-815739-8.00013-4
[5] Kumar, Manoj, Mohammed Alshehri, Rayed AlGhamdi, Purushottam Sharma, and Vikas Deep. "A de-ann inspired skin cancer detection approach using fuzzy c-means clustering." *Mobile Networks and Applications* 25 (2020): 1319-1329. https://doi.org/10.1007/s11036-020-01550-2
[6] Dunn, Joseph C. "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters." (1973): 32-57. https://doi.org/10.1080/01969727308546046
[7] Hathaway, Richard J., and James C. Bezdek. "Recent convergence results for the fuzzy c-means clustering algorithms." *Journal of Classification* 5, no. 2 (1988): 237-247. https://doi.org/10.1007/BF01897166
[8] Zhou, KaiLe, Chao Fu, and ShanLin Yang. "Fuzziness parameter selection in fuzzy c-means: The perspective of cluster validation." *Science China Information Sciences* 57 (2014): 1-8. https://doi.org/10.1007/s11432-014-5146-0
[9] Misuraca, Michelangelo, Maria Spano, and Simona Balbi. "BMS: An improved Dunn index for Document Clustering validation." *Communications in statistics-theory and methods* 48, no. 20 (2019): 5036-5049. https://doi.org/10.1080/03610926.2018.1504968
[10] Singh, Harpreet, Madan M. Gupta, Thomas Meitzler, Zeng-Guang Hou, Kum Kum Garg, Ashu MG Solo, and Lotfi A. Zadeh. "Real-life applications of fuzzy logic." *Advances in Fuzzy Systems* 2013 (2013): 3-3. https://doi.org/10.1155/2013/581879
[11] Makkar, Renu, and C. Renu Makkar. "Application of fuzzy logic: A literature review." *Int. J. Stat. Appl. Math* 3, no. 1 (2018): 357-359.
[12] Asai, H. T. S. U. K., S. Tanaka, and K. Uegima. "Linear regression analysis with fuzzy model." *IEEE Trans. Systems Man Cybern* 12 (1982): 903-907. https://doi.org/10.1109/TSMC.1982.4308925
[13] Zolfaghari, Zahra Sadat, Mohebbat Mohebbi, and Marzieh Najariyan. "Application of fuzzy linear regression method for sensory evaluation of fried donut." *Applied Soft Computing* 22 (2014): 417-423. https://doi.org/10.1016/j.asoc.2014.03.010
[14] Sarip, Abdul Ghani, Muhammad Burhan Hafez, and Md Nasir Daud. "Application of fuzzy regression model for real estate price prediction." *Malaysian Journal of Computer Science* 29, no. 1 (2016): 15-27. https://doi.org/10.22452/mjcs.vol29no1.2
[15] Chukhrova, Nataliya, and Arne Johannssen. "Fuzzy regression analysis: systematic review and bibliography." *Applied Soft Computing* 84 (2019): 105708. https://doi.org/10.1016/j.asoc.2019.105708
[16] Kim, Henry H., and Freydoon Shahidi. "Symmetric cube L-functions for GL2 are entire." *Annals of Mathematics* (1999): 645-662. https://doi.org/10.2307/121091
[17] Wahid, Fazli, and Do Hyeun Kim. "Short-term energy consumption prediction in Korean residential buildings using optimized multi-layer perceptron." *Kuwait Journal of Science* 44, no. 2 (2017).
[18] Liu, Yanchi, Zhongmou Li, Hui Xiong, Xuedong Gao, and Junjie Wu. "Understanding of internal clustering validation measures." In *2010 IEEE international conference on data mining*, pp. 911-916. IEEE, 2010. https://doi.org/10.1109/ICDM.2010.35
[19] Rendón, Eréndira, Itzel Abundez, Alejandra Arizmendi, and Elvia M. Quiroz. "Internal versus external cluster validation indexes." *International Journal of computers and communications* 5, no. 1 (2011): 27-34.

[20] Xie, Xuanli Lisa, and Gerardo Beni. "A validity measure for fuzzy clustering." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 13, no. 08 (1991): 841-847. https://doi.org/10.1109/34.85677

[21] Ramly, Nurfarawahida, Mohd Saifullah Rusiman, Shuhaida Ismail, Firdaus Mohamad Hamzah, and Özlem Gürünlü Alma. "An adjustment degree of fitting on fuzzy linear regression model toward manufacturing income." (2023). https://doi.org/10.11591/ijai.v12.i2.pp543-551

[22] Diamond, Phil. "Fuzzy least squares." *Information sciences* 46, no. 3 (1988): 141-157. https://doi.org/10.1016/0020-0255(88)90047-3

[23] Rusiman, Mohd Saifullah, Efendi Nasibov, and Robiah Adnan. "The optimal fuzzy c-regression models (OFCRM) in miles per gallon of cars prediction." In *2011 IEEE Student Conference on Research and Development*, pp. 333-338. IEEE, 2011. https://doi.org/10.1109/SCOReD.2011.6148760

[24] Zahari, Fatin Z., Kamil Khalid, Rozaini Roslan, Suliadi Sufahani, Mahathir Mohamad, Mohd Saifullah Rusiman, and Maselan Ali. "Forecasting natural rubber price in Malaysia using ARIMA." In *Journal of Physics: Conference Series*, vol. 995, no. 1, p. 012013. IOP Publishing, 2018. https://doi.org/10.1088/1742-6596/995/1/012013