The Fuzzy Inference System with Rule Bases Generated by Fuzzy C-Means Clustering

Article *in* Thai Journal of Mathematics · December 2018

1 author:

Samingun Handoyo
Brawijaya University

The Fuzzy Inference System with Rule Bases Generated by Fuzzy C-Means Clustering

Samingun Handoyo[1, a)]

[1]*Department of Statistics, Faculty of Mathematics and Sciences, University of Brawijaya, Jl. Veteran Malang, Indonesia*

[a)]Corresponding author: samistat@ub.ac.id

Abstract. The aims of the research are to build the Fuzzy Inference System (FIS) Takagi-Sugeno of zero order and to investigate the influence number of linguistic value on FIS performance evaluated by using MAPE and R^2 indicator. The fuzzy rule bases are generated by using the Fuzzy C-Means (FCM) clustering. The regional minimum wage in 2015 and 2016 dataset of East Java province of Indonesia are consecutively used as the training and the testing data. The number of clusters (n) that become FCM input including n = 3, 5, 7, and 9. The fuzzification process uses the Gaussian membership function with the mean parameter is the cluster center of the FCM output. Based on the values of MAPE and R^2 on the training and the testing data is obtained the system with the number of cluster of 9 which is the best performance. Furthermore, the increasing number of cluster is followed by the decreasing MAPE and the increasing R^2 value in both training and testing data.

Key word: *FuzzyC-Means; Fuzzy rule base; Membership function; Prediction or forecasting*

1. INTRODUCTION

Forecasting time series is a field of study that continues to grow because of the need for information in the future and supported by the availability of time series data are abundant. There are two groups of time series modeling involving fuzzy logic ie Fuzzy Time Series (FTS) and Fuzzy Inference System (FIS). The recent implementation of FTS was conducted by Ab Mutalib et al. [1] to predict the number of unemployed, and by Azahari and Othman [2], FTS to predict the distribution of rainfall. FIS Implementation for Seasonal short-term electricity demand forecasting under tropical conditions is done by Akil and Mitani [3]. The advantage of both FTS and FIS is the number of observation values that are not a constraint because the main component of FTS ie Fuzzy Logic Relationship (FLR) is easy to establish, and the main component of FIS ie Fuzzy Rule Base (FRB) is obtained from experts.

The more available observation values allow for more patterns of data that can be captured by the model, but it does make the FTS implementation less feasible due to the complex preparation of FLR and the slow computing process. FIS implementation for time series forecasting is usually constrained by the unavailability of FRB from experts. Some attempts were made to allow the FIS to be implemented without the FRB provided by the expert ie FIS hybridization with Neural Networks (NN) that gave rise to a model called Neuro-fuzzy or ANFIS. Some of the ANFIS forecasting time series implementations are performed by Bakirtzis et al. [4] ie Short-term load forecasting, prediction of water level in reservoir by Chang and Chang [5], Efendigil et al. [6] developing a decision support system based on NN and ANFIS, and the development of adaptive ANFIS method by Boyacioglu and Avci [7], and Osório et al. [8]. The implementation of ANFIS hybrid model produces a system with less satisfactory performance and slow computing process.

The generation of FRB based on the input-output data pair was first performed by Wang and Mendel [9] using a lookup table scheme, Cordon et al. [10] using genetic learning method, and evolution strategies method by Jin et al. [11]. Currently, Arif et al. [12] generated FRB using a priori algorithm of the association rule, while

Mansor et al. [13] creating a framework for extracting the FRB on the Mamdani fuzzy system. The above attempts result in an ever-increasing number of rules with an increasing number of input variables and linguistic values, as well as complicated computational framework processing. However, Priyono et al. [14] once evoked FRB with the Fuzzy Subtractive Clustering (FSC) method. The FRB generated by clustering method has the number of rules the same as the number of clusters or number of linguistic values.

Based on the previous explanation, the research proposed the development of Takagi-Sugeno FIS of zero order with FRB generated by using Fuzzy C-Means (FCM) clustering method. The FCM method which cluster the data to be some number of groups is used to generate the FRB so that the influence of the number of linguistic values against the performance of the system can be investigated. The regional minimum wage in East Java province of Indonesia in 2015 was used as the training data, while the testing data are the regional minimum wage in East Java province of Indonesia in 2016. The system performance will be evaluated based on MAPE and R^2 value indicators on both the training and the testing data.

2. LITERATURE REVIEW

Fuzzy clustering is included in distance-based clustering in which there are several algorithms including Fuzzy Subtractive and Fuzzy C-means clustering. Fuzzy Subtractive is a clustering algorithm in which the cluster number information is not an input parameter, while Fuzzy C-Means is a supervised clustering algorithm so that the number of cluster information is the input parameter whose magnitude must be determined first.

A. The Fuzzy C-Means Clustering

Fuzzy C-Means (FCM) is a clustering technique to a dataset in the form of a collection of attributes of an object in which the existence of each object in a cluster is determined by its membership value on a cluster center [15]. The basic concept of FCM is to determine the cluster center which will be the average location position of each cluster. Each object will have a degree of membership on each cluster. Clustering process occurs by repairing the cluster center and the membership value of each object repeatedly [16]. The process leads to the acquisition of a cluster center that moves in the right direction. This loop is based on the minimization of the objective function as the distance from each object to the center of the cluster which is weighted by the degree of membership of the object [17].

The FCM method gives smooth results because the weighting used is based on the fuzzy set [ ], smoothing here means that an object is not absolute to belong to a single group, but that object can also be a member of another group with a different memberships degree. An object will tend to be a member of a group with the greatest degree of membership when compared with the degree of membership in the other group [18].

The objective function used in FCM is [15]:

$$P_t = \sum_{k=1}^{n} \sum_{i=1}^{c} ((\mu_{ik})^w D(x_k, v_i)^2) \tag{1}$$

where Pt is the objective function at the t-th iteration, c is the number of clusters desired, n is the number of objects or records, and w is weight. The input data structure is expressed in matrix format as follows:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & \cdots & x_{1j} \\ \vdots & & \vdots \\ x_{k1} & \cdots & x_{kj} \end{bmatrix} \tag{2}$$

where, x_{kj} is the k-object on the j-th attribute, while the cluster center is declared as

$$\boldsymbol{V} = \begin{bmatrix} v_{11} & \cdots & v_{1j} \\ \vdots & & \vdots \\ v_{i1} & \cdots & v_{ij} \end{bmatrix} \tag{3}$$

v_{ij} is the cluster center of the i-th object on the j-th attribute, the partition matrix is expressed as

$$\boldsymbol{U} = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1k} \\ \vdots & & \vdots \\ \mu_{i1} & \cdots & \mu_{ik} \end{bmatrix} \tag{4}$$

where, μ_{ik} is the membership degree of the k-th object in the i-th cluster, and the distance between an object and the cluster center is expressed in formula (5) as follows:

$$D(x_k, v_i) = \left[\sum_{j=1}^{m} \left(x_{kj} - v_{ij}\right)^2\right]^{1/2} \quad (5)$$

The following is given in brief computation process of FCM clustering method as follows [17]:

1. Setting input data (X_{ij}) in the matrix data structure where index i is the number of objects or records, while the number of columns declared by index j is the number of features or attributes.
2. Specifying some parameters needed in the application of FCM method ie many clusters (c), weight factor (w), maximum iteration (t), and stopping condition (ξ).
3. Generating the initial elements of partition matrix $\mathbf{U}_0$ (μ_{0ik}) for i = 1, 2, ..., c and k = 1, 2, ..., n. Based on the initial partition matrix, then compute the degree of membership of each object in the each cluster uses the formula follows:
$$\mu_{ik} = \frac{\mu_{0ik}}{Q_i} \text{ , where } Q_i = \sum_{i=1}^{c} \mu_{0ik}$$
4. Calculating the center of the i-th cluster (v_{ij}) with the formula:
$$v_{ij} = \frac{\sum_{k=1}^{n}\left((\mu_{ik})^w * x_{kj}\right)}{\sum_{k=1}^{n}(\mu_{ik})^w} \quad (6)$$
5. Calculating the value of the objective function Pt at the t-th iteration by using equation (1)
6. Computing the new partition matrix using the following formula:
$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m}(x_{kj}-v_{ij})^2\right]^{\frac{-1}{w-1}}}{\sum_{i=1}^{c}\left[\sum_{j=1}^{m}(x_{kj}-v_{ij})^2\right]^{\frac{-1}{w-1}}} \quad (7)$$
7. Stopping the iteration if ($|P_t - P_{t-1}| < \xi$) or ($t > t_{max}$), in other conditions, iteration is continued with t=t+1 and repeat from step 4.

B. Rules Bases Generate and Inference Process

The fuzzy rule base is a collection of propositions in the form of implications that have a central role in the inference process of the FIS [19]. Suppose a rule is composed of an input and an output ie “IF x is A THEN y is B”, with x and y expressing linguistic variables, whereas A and B are linguistic values in the form of fuzzy sets defined in the domain range of variables X and Y. The statement "x is A" is called antecedence or the premise, whereas the statement "y is B" is called consequence or conclusions. This implication can be extended by using fuzzy operators, as follows [19]:

$$IF\ (x_1\ is\ A_1) \bullet (x_2\ is\ A_2) \bullet \ldots\ldots \bullet (x_n\ is\ A_n)\ THEN\ y\ is\ B \quad (8)$$

The notation dot (•) denotes disjunction or conjunction operator.

The inference process in the FIS with some rules (a fuzzy rule bases) is done on every rule and also the correlation between rules. The method used in the fuzzy inference process is the Max (Maximum) method. In the Max method, a fuzzy set of solution is obtained by taking the maximum membership degree of a rule, then using it to modify the fuzzy region and applying it to the output using the OR (Union) operator. If all the rules have been evaluated then the output will contain a fuzzy set that reflects the contribution of each rule [19].

Suppose a system has two input variables ie production cost (X_1), and demand (X_2), and one output variable of production quantity (Y). The linguistic variable of X_1 has three linguistic values (low, standard, high), X_2 consists of two linguistic values (down, up), while the linguistic variable Y with three linguistic values (reduced, normal, increased). The fuzzy rule bases with 3 rules that may be formed are as follows [20]:

$$[R_1]: IF\ (X_1\ is\ Low)\ and\ (X_2\ is\ Up)\ THEN\ (Y\ is\ Increased)$$
$$[R_2]: IF\ (X_1\ is\ Standard)\ THEN (Y\ is\ Normal)$$
$$[R_3]: IF\ (X_1\ is\ High\ )\ and\ (X_2\ is\ Down)\ THEN (Y\ is\ Reduced\ )$$

Based on the above fuzzy rule bases, if there are the input values of (X_1, X_2) then the input values are inserted into [R_1], this input will be fuzzified, then calculate the alpha cut value on [R_1] by MIN principle (minimum membership degree between X_1 and X_2 on each linguistic value) to produce the output of [R_1]. The same input values are inserted into [R_2], then undergoes a fuzzification process, calculate alpha cut value on [R_2] to get the output of [R_2], the same process is done on [R_3]. The outputs of each rule are aggregated to produce the output of the system in the form of a fuzzy set. The output of the system in the form of crips number is obtained through the defuzzification process against the fuzzy set of aggregation results.

The format of the fuzzy rule bases and the inference method described above is a Mamdani FIS model that is characterized by both the antecedents and consequence parts having the linguistic values all in the form of a fuzzy set. The Sugeno FIS model is characterized by the consequence part of each rule is a constant or a linear equation. In addition, in the Sugeno FIS is also not needed the defuzzification process, because the results of the aggregation of the output of each rule is a number of crips generated from the weighted average of output in all rules. According to Palit and proponic [21], there are two type of models for Sugeno's FIS, the Sugeno zero order model with the form of the rule as follows:

$$IF\ (x_1 \text{ is } \mathrm{A}_1)\bullet(x_2 \text{ is } \mathrm{A}_2)\bullet \ldots \bullet(x_\mathrm{n} \text{ is } \mathrm{A}_\mathrm{n}) \text{ THEN } (\mathrm{z} = \mathrm{k})$$

and the Sugeno first order model with the following rule form:

$$IF\ (x_1 \text{ is } \mathrm{A}_1)\bullet(x_2 \text{ is } \mathrm{A}_2)\bullet \ldots \bullet(x_\mathrm{n} \text{ is } \mathrm{A}_\mathrm{n}) \text{ THEN } (z = p_1 * x_1 + p_2 * x_2 + \ldots + p_n * x_n + q)$$

In the Sugeno FIS model, the fuzzification process occurs only on the input variables.

C. The performance of the System for Forecasting

System performance measures refer to the system's ability to generate an output based on a given input. The purpose of assessing system performance is to select the best performing system. There are several performance measures, among which are

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| x100\% \tag{9}$$

$$R^2 = \frac{\left(\sum_{i=1}^{n}(y_i - \bar{y}_i)(\hat{y}_i - \bar{\hat{y}}_i)\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2 \sum_{i=1}^{n}\left(\hat{y}_i - \bar{\hat{y}}_i\right)^2} \tag{10}$$

Where y_i is the i-th actual value and $\hat{y}_i$ is the i-th prediction value. The Mean Absolute Percentage Error (MAPE) is a measure of system performance that belongs to a relative measure because the value of MAPE which can be said well is subjective [22]. According to Gujarati [23], R^2 value explains the suitability between prediction and actual values. The intended value of R^2 is the squared correlation value that is the correlation between the actual and the prediction values. The magnitude of the relationship between the actual and the predicted values can be viewed as the precision of the predicted values in predicting the actual values. If the prediction values are close to the actual values then the two groups of values will form a strong linear relationship shown by the two curves almost coincide and form a straight line, so based on this state, R^2 is defined as the measure of the match between the actual and the predicted values [20].

3. DATA AND RESEARCH PROCEDURES

In this study used dataset about the factors that affect the amount of regional minimum wage (RMW) in East Java in 2015 and in 2016. The data consists of three independent variables (predictors) and one dependent variable (response). The three predictor variables are inflation rate(IFR1), economic growth (EG1), and decent living component (DLC1) in 38 cities or districts in East Java, while the dependent variable is regional minimum wage(RMW1) in 38 cities or districts in East Java. There are two datasets, dataset1 as training data are the dataset with RMW in the year 2015, and dataset2 as testing data are the dataset with RMW in the year 2016. The attributes of both datasets are presented in table 1. The training data are used to build system, while the testing data are used to perform cross-validation of system performance.

Table 1. Predictor and Response Varibles for Training and Testing Data

Training data		**Testing data**	
Predictor	**Response**	**Predictor**	**Response**
IFR14 EG14 DLC14	RMW15	IFR15 EG15 DLC15	RMW16

In the paper, FIS development uses R software that is a semi programming language and a very popular statistical computing software currently characterized by free and open source. The overall computation process can be divided into three stages as follows:

a. Clustering of training datasets with FCM. Computational processes occurring at this stage include: setting some FCM parameters (maximum iteration, large of threshold error, number of clusters being formed). The number of clusters used as inputs is 3.5, 7, and 9.
b. Prepare the fuzzy rule bases. The outputs of FCM are the cluster center, the degree of membership, and the number of members of each cluster. The fuzzy rule bases are arranged based on the cluster centers that are formed. The cluster centers of each attribute are ranked to produce the linguistic values of each input variable.
c. FIS Implementation for prediction. At this stage, we first determine the function of fuzzy membership, the linguistic values of each variable, and the setting of its parameters. the fuzzy rule bases that have been obtained in the previous step is used as a reference for fuzzification and inference process to generate the output of each system. the next steps are Comparison of actual value versus prediction value, error calculation, MAPE and R^2 calculation to evaluate system performance both in training data and data

4. RESULT AND DISCUSSION

One of the advantages of a heuristic approach such as FIS is that it does not require some assumptions to be met such as conventional statistical modeling. Since the number of observations on the training dataset is quite small, only 38 records with 3 input variables and one output variable, so in this study, implemented the FIS Sugeno method of zero order based on FCM clustering method.

A. Clustering of Training Data with FCM

One of the advantages of FIS based on the clustering method is the number of rules that are formed as much as the number of clusters so that the inference process can run faster. Bases of fuzzy rules are obtained without going through the fuzzification process of observational values, and without having to calculate the degree of each rule. In addition, the number of linguistic or fuzzy set values involved is also as much as the number of clusters. In this study the training dataset is grouped into many groups c = 3,5,7,9, setting the parameters of threshold error is 0.00005, while the maximum parameter iteration, tmax = 100. The output of the clustering FCM results includes information about the cluster center and the number of members of each cluster as stated in table 2 below:

Table 2. Cluster Centers and Its Members for Number of Cluster c=3,5,7, and 9

Cluster center for cluster c=3 and its elements numbers				
EG14	**IFR14**	**DLC14**	**RMW15**	**Elements**
6.23	6.64	2048447	2705000	5
5.43	5.7	1086494	1253500	17
4.53	5.71	1138571	1311000	16
Cluster center for cluster c=5 and its elements numbers				
IFR14	**EG14**	**DLC14**	**RMW15**	**Elements**
6.2	5.7	1521259	1962000	3
3.72	5.71	1293800	1575000	8
5.58	5.28	941068	1150000	7
7.05	6.64	2092026	2705000	5
5.51	5.81	1088392	1265000	15
Cluster center for cluster c=7 and its elements numbers				
IFR14	**EG14**	**DLC14**	**RMW15**	**Elements**
3.6	5.24	1353746	1566800	7
6.2	5.7	1521259	1962000	1
5.58	5.45	1106926	1267300	9
4.49	6.93	1577513	1817000	1
6.23	6.36	1680000	2695000	5
5.39	5.81	1086494	1250000	14
4.57	5.82	1472425	1882250	1

Cluster center for cluster c=9 and its elements numbers				
IFR14	**EG14**	**DLC14**	**RMW15**	**Elements**
3.6	5.93	1248369	1437500	5
7.05	7.06	2101608	2707500	1
7.72	6.18	2048447	2705000	3
3.76	5.21	941068	1150000	5
3.22	5.71	1224683	1460500	3
6.09	5.28	934527	1150000	2
5.51	5.45	1102409	1265000	14
3.72	6.64	2092026	2700000	1
4.49	6.93	1577513	1817000	4

Table 2. explains the cluster centers and the number of members of each cluster. Suppose that on the number of clusters c = 9, the last row of table 2 can be seen that there is a cluster center with variables IFR14 = 4.49%, EG14 = 6.93%, DLC14 = IDR 1577513, and RMW15 = IDR 1817000 which in this cluster there are 4 districts or city as its members.

B. Fuzzy Rule Bases of Sugeno FIS Model

Once the information about the center of the cluster has been obtained to generate the rule bases, we must know the linguistic value of each input variable. The many linguistic values of an input are as numerous as many clusters of k = 3,5,7, and 9. Consider the number of clusters c = 3 with the linguistic values of Small (S), Medium (M), and Big (B). Since the positions of each value of S, M, and B are sequentially in each domain of the crip variable, the linguistic value can be obtained by ranking the cluster centers of each related variable. Rank r = 1 means S, r = 2 means M, and r rank of 3 means B. The naming of linguistic value in this study uses the guidance that the rank plotted as median is always named M, to the left of M is S, and on the other side of M is B. Suppose that in many clusters k = 9, then at r = 5 means M, r = 1 is S4, r = 4 is S1, r = 6 is B1, and r = 9 is B4. The ranking of each input-output variables are presented in columns 1 through 4 of Table 3.

Table 3. Linguistic Values and Fuzzy Rule Bases for Number of Cluster c=3,5,7, and 9

Rules formed for cluster c=3 and its consequence part				
IFR14	**EG14**	**DLC14**	**RMW15**	**Consequence**
3	3	3	3	2705000
2	1	1	1	1253500
1	2	2	2	1311000
Rules formed for cluster c=5 and its consequence part				
IFR14	**EG14**	**DLC14**	**RMW15**	**Consequence**
4	2	4	4	1962000
1	3	3	3	1575000
3	1	1	1	1150000
5	5	5	5	2705000
2	4	2	2	1265000
Rules formed for cluster c=7 and its consequence part				
IFR14	**EG14**	**DLC14**	**RMW15**	**Consequence**
1	1	3	3	1566800
6	3	5	6	1962000
5	2	2	2	1267300
2	7	6	4	1817000
7	6	7	7	2695000
4	4	1	1	1250000
3	5	4	5	1882250
Rules formed for cluster c=9 and its consequence part				
IFR14	**EG14**	**DLC14**	**RMW15**	**Consequence**
2	5	5	4	1437500

8	9	9	9	2707500
9	6	7	8	2705000
4	1	2	1.5	1150000
1	4	4	5	1460500
7	2	1	1.5	1150000
6	3	3	3	1265000
3	7	8	7	2700000
5	8	6	6	1817000

Based on the linguistic value terminology described above, the last line of Table 3 can be expressed as a fuzzy rule as follows:

IF ((IFR14 is M) and (EG14 is B3) and (DLC14 is B2)) THEN (RMW15 is B2).

Since the Sugeno model FIS is to be implemented, the fuzzy rule bases must have a constant value on the consequent part. In this study the constants used are the cluster center of the output variable ie the 5th column of table 3, so the fuzzy rule of the last row of Table 3 becomes as follows

IF ((IFR14 is M) and (EG14 is B3) and (DLC14 is B2)) THEN (RMW15 is 1817000)

C. Implementation of FIS Sugeno Zero Order on Data Training

The fuzzy rules bases have a central role in generating FIS output. Fuzzification process to input data based on the linguistic or fuzzy set values that exist in the bases of the rules. This process can be done if the membership function and its parameters have been determined. In this study is used the Gaussian membership function. The parameters of the Gaussian function are the standard deviation (sb) and mean. The cluster center obtained by FCM is used as the mean parameter, whereas the sb parameter is determined based on the data of each input variable by calculating the range of each input then divided in the same interval. In this study, every linguistic value of each input variable is considered to have the same sb. In Table 4 is shown the parameters of the Gaussian membership function parameters of 9 linguistic values in the three input variables.

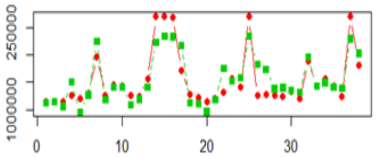
Table 4.Parameters of Gaussian Membership Function of All Input Variables

Ling. value	**Param. IFR14**		**Param. EG14**		**Param. DLC14**	
	Std	**Mean**	**Std**	**Mean**	**Std**	**Mean**
S4	1.32	3.22	0.545	5.21	291986	934527
S3	1.32	3.6	0.545	5.28	291986	941068
S2	1.32	3.72	0.545	5.45	291986	1102409
S1	1.32	3.76	0.545	5.71	291986	1224683
Z	1.32	4.49	0.545	5.93	291986	1248369
B1	1.32	5.51	0.545	6.18	291986	1577513
B2	1.32	6.09	0.545	6.64	291986	2048447
B3	1.32	7.05	0.545	6.93	291986	2092026
B4	1.32	7.72	0.545	7.06	291986	2101608

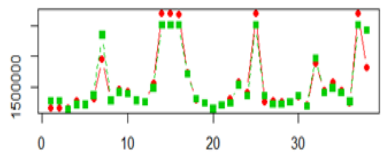
In table 4 it is shown that for each input variable at all linguistic values have the same sb parameter value and the mean parameter is the center of the cluster of FCM results. The condition of the two parameters causes the interval width between the two cluster centers on the two linguistic values of each input variable to vary. Suppose that the input variable IFR14 on the linguistic value of B4 has sb = 1.32 and the parameter mean = 4.49. EG14 input variable at linguistic value B4 with parameter sb = 0.55 and mean = 6.93, while input variable DLC14 at linguistic value B4 with parameter sb = 291986 and mean = 1577513.

After all linguistic value parameters have been set, each record of the input variable is processed into the rule bases to generate the output of the system. The fuzzification process is only imposed on each input pairs according to the linguistic values contained in each rule bases. The input pairs are inserted into each fuzzy rule with the output of each rule in the form of a constant value. The output of each fuzzy rule is multiplied by the alpha cut of each fuzzy rule, then aggregated for the output of all rules so that a new constant is produced that is the FIS output.

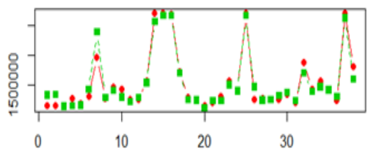
All input pairs are processed one at a time, and finally, the set of all output values of the system is called predicted value. The difference between the actual output value (response variable data) and the output of the system is called error. Figure 1 shows plots for all actual values versus predicted values of 4 FIS.



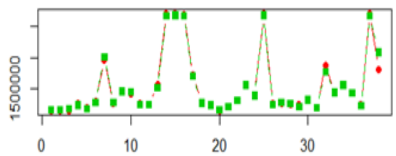
a. System with 3 Clusters(FIS3)



b. System with 5 Clusters(FIS5)



c. System with 7 Clusters(FIS7)



d. System with 9 Clusters(FIS9)

Figure 1: Plot Actual versus Predicted values of RMW in 2015 of East Java (the training data)

Based on Figure 1 exposed that the more clusters or more rules, the outputs of the system are close to the actual value very closely even in the number of clusters c = 9, only some points have error rather large.

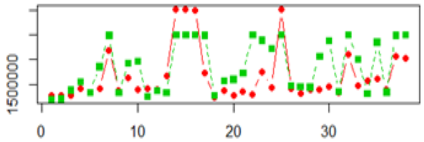
D. Prediction of Testing Data by Using FIS Models

An important step in system modeling that is often left behind is the system test with.the out-sample data or the testing data. The Knowledge of system performance in the out-sample data will increase the confidence of all parties both developers and users to apply the system in the real world. The input pairs of the testing data IFR15, EG15, and DLC15 are used as input data for FIS3, FIS5, FIS7, and FIS9. The output of each FIS model is RMW16 which is the regional minimum wage of 2016 for the districts or cities in East Java, Indonesia. Table 6 shows the last 5 records for actual values, predictions, and errors. Visualization of all actual values and predicted values of RMW 2016 is given in Figure 2.

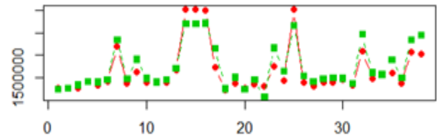
Table 5. Five Last Records Contained Respectively Actual, Predicted, and Error Values in the Testing Data

System with 3 clusters				**System with 5 clusters**			
Rec	Actual	Predict	Error	Rec	Actual	Predict	Error
[34,]	1566296	1317279	249017	[34,]	1566296	1577696	11400
[35,]	1602813	2356996	754183	[35,]	1602813	1907913	305100
[36,]	1386168	1343694	42474	[36,]	1386168	1500202	114031
[37,]	2076580	2495159	418579	[37,]	2076580	2348585	272005
[38,]	2025955	2500886	474931	[38,]	2025955	2444793	418838
System with 7 clusters				**System with 9 clusters**			
Rec	Actual	Predict	Error	Rec	Actual	Predict	Error
[34,]	1566296	1823028	256732	[34,]	1566296	1600568	34272
[35,]	1602813	1323776	279037	[35,]	1602813	1584949	17864
[36,]	1386168	1300285	85883	[36,]	1386168	1283876	102292
[37,]	2076580	2305633	229053	[37,]	2076580	2271702	195122
[38,]	2025955	2068161	42	[38,]	2025955	2142447	116492

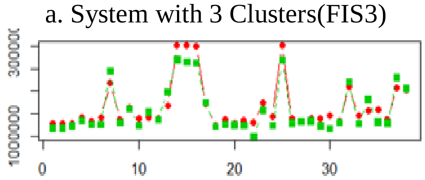
Based on the 5 error values of the four systems in table 6, it is clear that there is an increasing trend of error values. In FIS3 the smallest error is IDR 42,473 and the highest error is IDR 754,183, FIS5 has the smallest error IDR 11,400 and highest error IDR 418,873, FIS7 smallest error is IDR 42 and highest error is IDR 279,027, while FIS9 has smallest error IDR 17,864 and the highest error IDR 195,122. The sample of 5 error values in table 5 above indicates that FIS9 is the best system with the most narrow error range. Overall comparison of actual value and predicted value of RMW in 2016 is given in Figure 2 below:



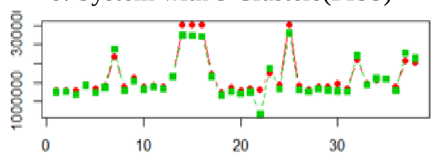
a. System with 3 Clusters(FIS3)



b. System with 5 Clusters(FIS5)



c. System with 7 Clusters(FIS7)



d. System with 9 Clusters(FIS9)

Figure 2. Plot Actual versus Predicted values of RMW in 2016 of East Java (the testing data)

In general, if the four graphs are compared, then there is a consistent information that is the greater the number of clusters, the plot of actual value and the predicted value more squeeze. This may be understandable because the increasing number of clusters will automatically be followed by increasing the number of rules in the fuzzy rule bases. Thus the fuzzy rule bases with a large number of rules will be able to capture the data pattern better.

E. Performance Evaluation of FIS Models on Training and Testing Data

In the training data, the MAPE performance indicator decreases from 11.76% in FIS3 to 1.61 in FIS9, although it slightly increases from 4.85% in FIS5 to 5.23% in FIS7. On the other hand, the R^2 performance indicator consistently increases from 0.75 in FIS3 to 0.99 in FIS9 of the training data as exposed in table 6.

Table 6. Performance of the FIS on the training and the testing data.

System	Training Data		Testing Data	
	MAPE(%)	R^2	MAPE(%)	R^2
FIS3	11.76	0.75	21.88	0.45
FIS5	4.83	0.91	8.84	0.87
FIS7	5.23	0.95	8.5	0.9
FIS9	1.61	0.99	6.72	0.93

Based on the results in Table 6 above, it is known that the FIS9 has the best performance on the training data that it has MAPE = 1.61% and R^2= 0.99. The value of MAPE = 1.61% means that the accuracy of FIS9 is 98.39%, while R^2=0.99 means that the actual values match prediction values are 99% or theoretically can be stated that FIS9 is able to explain the existing pattern in the training data of 99%. The performance indicator of MAPE and R^2 in testing data also tend to be similar to those of both performance indicator in training data. When the increasing number of clusters also obtained better system performance. In this case, the MAPE value ranges between 6.72% in FIS9 and 21.88% in FIS3, whereas the R^2 value ranges between 0.45 in FIS3 and 0.93 in FIS9. Thus FIS9 is a system that has the best performance when it is compared to FIS3, FIS5, and FIS7, both of on the training and on the testing data.

5. CONCLUSIONS

The FCM clustering method can be used to generate the fuzzy rule bases from dataset in the form of input-output pairs data structure and can yield two advantages ie the number of rules generated equal to the number of clusters and the mean parameters of the Gaussian membership function automatically obtained from the central value of the cluster. The FIS9 system has the best performance which the value of MAPE=98% and the value of R^2=0.99 in the training data, while in the testing data, the value of MAPE=6.72% and the value of R^2=0.93. The system performance indicator of R^2 is more robust than the system performance indicator of MAPE.

REFERENCES

1. S.M. Ab Mutalib, N. Ramli, and D. Mohamad, “Forecasting Unemployment based on Fuzzy Time Series with Different Degree of Confidence” , *Journal of Telecommunication, Electronic and Computer Engineering (JTEC),* vol. *9, no.* 1-4, pp. 21-24, 2017.
2. S. N. F. Azahari, and M. Othman, “Forecasting Rainfall Distribution Based on Deseasonalising Fuzzy Time Series Model”, *Journal of Telecommunication, Electronic and Computer Engineering (JTEC),* vol. *9, no.* 2-4, pp. 89-93, 2017.
3. Y. S. Akil, and Y. Mitani, “Seasonal short-term electricity demand forecasting under tropical condition using fuzzy approach model”, *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. *9, no.*1-3, pp. 77-82, 2017.
4. A. G. Bakirtzis, J. B Theocharis, S. J. Kiartzis, and K.J. Satsios,”Short term load forecasting using fuzzy neural networks”, *IEEE Transactions on Power Systems*, vol. *10, no.* 3, pp. 1518-1524, 1995.
5. F. J. Chang, and Y.T. Chang, “Adaptive neuro-fuzzy inference system for prediction of water level in reservoir”, *Advances in water resources*, vol. *29, no.*1, pp. 1-10, 2006.
6. T. Efendigil, S. Önüt, and C. Kahraman, “ A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: A comparative analysis”, *Expert Systems with Applications,*vol. *36,* no.3, pp. 6697-6707, 2009.
7. M. A. Boyacioglu, and D. Avci, “ An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: the case of the Istanbul stock exchange”, *Expert Systems with Applications*, vol. *37, no.* 12, pp. 7908-7912, 2010.
8. G. J. Osório, J. C. O. Matias, and J. P. S. Catalão, "Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information", *Renewable Energy* , vol. 75, pp. 301-307, 2015.
9. L. X , Wang, and J.M. Mendel, ”Generating fuzzy rules by learning from examples”, *IEEE Transactions on systems, man. and cybernetics*, vol. *22, no.* 6, pp. 1414-1427, 1992.
10. O. Cordón, F. Herrera, and P. Villar, “Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base”, *IEEE Transactions on fuzzy systems*, vol. *9, no.* 4, pp. 667-674, 2001.
11. Y. Jin, W. Von Seelen, and B. Sendhoff, “ On generating FC/sup 3/fuzzy rule systems from data using evolution strategies”, *IEEE Transactions on Systems, Man. and Cybernetics, Part B (Cybernetics),* vol. *29, no.* 6, pp. 829-845, 1999.
12. M.F. Arif, B. Anoraga, S. Handoyo, and H. Nasir, “ Algorithm Apriori Association Rule in Determination of Fuzzy Rule Based on Comparison of Fuzzy Inference System (FIS) Mamdani Method and Sugeno Method”, *Business Management and Strategy*, vol. *7, no.* 1, pp. 103-124, 2016.
13. R. Mansor, M. M. Kasim, and M. Othman, “A Weighted Subsethood Mamdani Fuzzy Rules Based System Rule Extraction (MFRB-WSBA) for Forecasting Electricity Load Demand-A Framework”, *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. *8, no.* 8, pp. 97-102, 2016.
14. A. Priyono, M. Ridwan, A. J. Alias, R. Atiq, O. K. Rahmat, A. Hassan, and M. Ali, “Generation of fuzzy rules with subtractive clustering”. *Jurnal Teknologi*, vol. *43*(D), pp. 143-153, 2005.
15. J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm" , *Computers & Geosciences,* vol. 10, no. 2-3, pp. 191-203,1984.
16. R. L. Cannon, J. V. Dave, and J. C. Bezdek, "Efficient implementation of the fuzzy c-means clustering algorithms" , *IEEE transactions on pattern analysis and machine intelligence* , vol. 2, pp. 248-255, 1986.
17. W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation", *Pattern recognition*, vol. 40, no. 3, pp. 825-838, 2007.
18. T.C. Havens, J. C. Bezdek, C. Leckie, L.O. Hall, and M. Palaniswani, "Fuzzy c-means algorithms for very large data", *IEEE Transactions on Fuzzy Systems, vol.* 20, no. 6, pp. 1130-1146, 2012.
19. L. X. Wang, *A Course Fuzzy Systems and Control,* Upper Saddle River, NJ: Prentice Hall PTR, 1997.
20. S. Handoyo, and A. P. S. Prasojo, ”Applied Fuzzy System with R Software (*Sistem Fuzzy Terapan dengan Software R*)”, UBpress, Malang, Indonesia, 2017.
21. A. K. Palit, and D. Popovic, Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications, London: Springer, 2005.
22. S. Makridakis, S. C. Wheelwright, and V. E. McGee, Forecasting Methods and Applications, Second Edition, Jakarta: Erlangga, 1999.
23. D. N. Gujarati, and D. C. Porter, *Basic Econometrics,* Fifth Edition, New York: McGraw-Hill, 2008