Indonesia Scholarship Selection Model Using a Combination of Back-Propagation Neural Network and Fuzzy Inference System Approaches

Luther Alexander Latumakulita[1*]   Tsuyoshi Usagawa[1]

[1]*Graduate School of Science and Technology, Kumamoto University, Kumamoto, Japan*
* Corresponding author's Email: latumakulitala@unsrat.ac.id

Abstract: "Bidik Misi" (BM) is a famous scholarship offered by Indonesia Government which has two main qualification aspects: economic needs and academic performance. Since BM scholarship was initiated in 2010, the number of candidates has increased year by year despite limited quota, causing difficulties in selecting the most qualified candidates. This research is proposing a new selection model by combining Back-Propagation Neural Network (BPNN) as candidate's predictor and Fuzzy Inference System (FIS) as candidate's selector. BPNN is used to classify candidates into two, and three recommendation classes then perform FIS on priority set of candidates which is derived from the intersection of classification's result of two and three recommendation classes to make candidate's ranking. Closed test results show that for assuming quota of 50% from total 625 candidates or quota = 312, system can choose 267 highly recommended and reject 268 non-recommended candidates while 45 are identified as moderate recommended, and 45 other are identified as candidates of lower priority with selection accuracy of 85.6% better than five previous similar works. The open test results provide accuracy average of 86.88% and the lowest accuracy of 84% which is close enough to the closed test's accuracy of 85.6% with the difference around 1%, and a small standard deviation value of 2.37 indicates stability of system performance in the selection process. The lowest and highest accuracies of the open test results indicate the proposed system can predict and select BM scholarship's candidates in the future with satisfactory accuracies from 84% up to 90%; thus the proposed system can be trusted as a proper model for BM scholarship selection process.

Keywords: Indonesia scholarship, Bidik misi, BPNN, FIS.

1. Introduction

"Bidik Misi" (BM) scholarship is a scholarship offered by the Indonesia Government which has two qualification criteria: economic needs and academic performance. The most qualified candidates are those from low economic families with good academic performance [1]. Since the BM scholarship was initiated in 2010, it has become the main target of candidates from low economic families. While candidates for BM scholarships increased year by year, the quota remains limited, and this imbalance causes difficulty for the selection committee to select the most qualified candidates. Previous work [2] has provided a "fuzzy" framework for the BM scholarship selection process. The framework behavior is similar to a fuzzy logic controller (FLC), with two inputs

producing one output. The elbow method is used to obtain the optimal number of clusters for each input data set, then defines the number of corresponding linguistics. K-Mean clustering is used to design the fuzzy membership function of input datasets [3]. Mamdani mechanism is used to produce the output of each FLC with the center of gravity method performing defuzzification process. The proposed method can overcome the problem in making a clear decision by the selection committee due to limited quota with 71.43% accuracy rate. Previous work [4] redesigning the framework by considering correlation coefficient of input parameters. Selection accuracy of 84.9% was achieved using the modified framework, indicates increasing accuracy around 13%. The experiment also conducted on the dataset with incomplete parameters to observe system stability

when there is parameter lost in the dataset, and the accuracy of 82.9% was achieved. Although the accuracy was decreased around 2% compared to the experiment on the dataset with complete parameters, it is still sufficient, indicates system performance remain stable when parameter lost in datasets. Back-Propagation Neural Network (BPNN) is used to classified and eliminate non-recommended candidates before the selection process. Accuracy average of 88.7%, with 83% the lowest accuracy was reported in testing data while the highest accuracy of 90.1% was reported on all data. Parameters of house area, floor area per capita, and land area reported as parameters most influence classification results of two recommendation class by observing misclassified data.

BPNN also used in many research to solve the problem of data classification and prediction with good accuracies. In a research to classify breast tumors [5] an accuracy of 84.6% was reported using BPNN. Another research [6], which combined BPNN and fuzzy analytical hierarchy process (Fuzzy_AHP) to predict the risk of heart failure, achieved a prediction accuracy of 91.10%. Prediction accuracy of 91.4% for stress concentration factors of cracked concrete structures was enabled by ascertaining relationships between stress concentration factors and crack parameters [7]. Classification accuracy of 94.17% for males and 93.75% for females in research about human age classification using BPNN classifier and facial skin aging features [8]. BPNN was used to predict the relationship between the content in *Melilotus alba* roots and elevated concentration of polycyclic aromatic hydrocarbons carcinogen (PAH_{carc}) in soils [9]. The method precisely predicted levels of PAH_{carc} in *Melilotus alba* roots with an average R^2 of 0.9994, indicating the strong level of determination. An accuracy of 99.96% on training and 65% on testing of classification results was reported using multilayer back-propagation for osteoporosis screening [10]. Multilayer back-propagation also reported successfully in detecting and classifying masses into benign and malignant on digital mammograms data with achieved accuracy of 98.9% [11]; and classifying the Indian folk dance song into four classes with the accuracy more than 90% [12]; and designing the receiver for Multi Carrier-Code Division Multiple access (MC–CDMA) system which were more efficient and effective compared to the classical receiver [13]. Another research [14] investigated the ability of feed-forward BPNN by utilizing various training algorithm to predict the nanofluids relative viscosity. In this case, the Levenberg-Marquardt training algorithm was more precise than others. The results revealed that the

proposed network was able to correlate and predict the relative viscosity accurately on training dataset with an overall mean square error value of 0.00901 and correlation coefficient of 0.9954. Several research combined BPNN with other methods in order to improve the forecasting ability; [15] combine wavelet transform called Wavelet De-noising-based Back Propagation (WDBP) neural network to forecast the stock prices; [16] introduced an exponent back propagation neural network (EBPNN) which contributes to financial market cross-correlations forecasting with better accuracy; [17] proposed a hybrid intelligent approach named ADE-BPNN where BPNN was supported by an adaptive differential evolution algorithm, to estimate energy consumption; [18] combined the rough set, least square, and BPNN methods for shield machine fault diagnosis where results showed that the predicting accuracy of fault diagnosis was improved; [19] proposed a novel water temperature forecasting model based on empirical mode decomposition (EMD) and BPNN which named EMD-BPNN. The results showed that the hybrid EMD-BPNN model has better prediction performance compared to the standard BPNN, therefore EMD-BPNN method is suitable for water quality prediction.

On the other hand, the Fuzzy Inferences System (FIS) approach has been used to measure problems of both poverty and academic performance levels. A research measured indexes of poverty in a fuzzy environment and found that not only income which affects poverty level but there are other attributes since poverty is a complicated phenomenon [20]. Poverty measurement using fuzzy logic approach is also described in [21, 22]. Saleh and Kim [23] proposed a fuzzy logic system, represented in an FLC diagram, to evaluate students' learning achievement and was improved by [24] by automatically generating weights for input attributes. The previous works cited above demonstrate the high performance in solving data classification problems provided by BPNN; besides, high performance in solving poverty measurement problems and academic performance measurements are provided by fuzzy logic.

The objective of this research is to increase accuracy and efficiency of the selection process by extending classification into three recommendation classes to propose the best selection model for BM scholarship.

This study focuses only on the case of BM scholarship selection. The hypothesis is by doing classification with two and three recommendation classes and then matching those results before conduct selection process can increase the performance of selection process itself. Three processes will be

discussed here; First is classification process with two and three recommendation classes using BPNN; Second is matching process between those classification results to obtain the priority set of candidates; Third is conduct selection process using FIS sequentially on the priority set of candidates to make candidate's ranking as the final result. The high performance of the proposed method is proven to be able to predict and select the most qualified candidates of BM scholarship in the future which will greatly contribute to the effectiveness of the selection committee's works.

This study is also compared with five similar works include authors' previous works. A selection model has been proposed for academic and non-academic scholarships using Unified Modeling Language (UML) in Fuzzy Multi-Attribute Decision Making (FMADM) [25] which incorporated both Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) and Weighted Product (WP) methods. Qualification criteria for both scholarships were grade point average (GPA), household income, family member number, and other social aspects. Qualified candidates for academic scholarships were those from low economic families with high academic achievement; for non-academic awards, candidates with significant activities in social aspects and moderate academic achievement. This work shows that TOPSIS and WP methods provide similar results. The other work [26] compared herein also used TOPSIS method in FMADM for academic excellence and underprivileged scholarships selection process at Udayana University Bali, Indonesia. [27] the work combined Analytic Hierarchy Process (AHP), Support Vector Machine (SVM), and TOPSIS as a selection model for the "Gerakan Orang Tua Asuh" scholarship selection process. AHP was used to reduce criteria from 50 to 33; SVM, to classify processes; and TOPSIS, to recommend rankings. In this work, SVM achieved the highest accuracy rate of 89.94%.

The rest of this paper is organized as follows. Section 2 presents details of BM scholarship parameters, condition of datasets, and the methodology of the proposed method based on combination of BPNN and FIS approaches. Section 3 presents experiments for closed and open test. Section 4 discusses comparison and analysis between closed and open test results, and discusses the comparison between this work and previous similar works. The conclusion and future work are given in section 5.

2. Methodology

BM scholarship parameters are described in Table 1 where P# denotes parameter identification number.

Table 1. Description of BM scholarship's parameters

P#	Parameter Name	Description	Data Type
P_1	Parent's income	Total household's income	Indonesia Currency (Rp)
P_2	Dependent	Total of parent's income over total of family members	Indonesia Currency (Rp)
P_3	House area	Total of house area	Meter square
P_4	Floor area per capita	Total of house area over total of family members	Meter square
P_5	Land area	Total of land area	Meter square
P_6	Distance from downtown	The Distance between home resident and downtown. Further distance to downtown indicates that the residential home is more isolated, which means the poverty level is lower. Contrary, closer distance to downtown means higher poverty level.	Kilo meter
P_7	Water source	Options for family water source are pet water, PDAM (Indonesia Government water company), well, and river or spring water.	[0..100]
P_8	Sanitation source	Options for family sanitation source are private indoor, private outdoor, and public sanitation	[0..100]
P_9	Electricity source	Options for family electricity source are PLN (Indonesia Government electricity company), solar panel, PLN and generator, and no electricity source	[0..100]
P_{10}	Average values of high school report	Average of report values from semester 1 to semester 6	[0..100]
P_{11}	Extracurricular achievement	Student's extracurricular activities	[0..100]

Table 2. The used symbols

Symbol	Description
ms, ac, at	Manual score, Actual class, Actual target
b_{middle}, b_{lower}, b_{upper}	Boundaries to decide actual class and target for two and three recommendation classes
W_1^2, W_2^2	Weighed matrices of classification with two classes
W_1^3, W_2^3	Weighed matrices of classification with three classes
C_21, C_20	A set of recommended and non-recommended candidates provided by classification with two recommendation classes
C_31, $C_30.5$, C_30	A set of Highly, moderate, and low recommended candidates provided by classification with three recommendation classes
S_1 to S_5	Priority set of candidates which will be processed in selection process

Table 3. Abbreviations

Symbol	Description
BPNN	Back-Propagation Neural Network,
FMADM	Fuzzy Multi-Attribute Decision Making
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
AHP	Analytic Hierarchy Process
SVM	Support Vector Machine
Rp	Indonesia Currency Rupiahs

Table 2 describes the used symbols while Table 3 shows abbreviations in this paper.

2.1 Condition of data set

BM scholarship only allows the candidates from two years of high school graduates who will graduate in the current year and had graduated in the previous year. However, submission period of BM scholarship has to be done before the period of national examination which causes the candidates who graduate in the current year will be lost the national high school test score parameter while candidates from the previous year have a complete set of parameters. This condition causes the parameter of the national high school test score to be lost in part of the dataset. This work use dataset with incomplete parameters.

The human evaluation procedure is based on scoring techniques by defined scoring interval for each parameter, then calculates a total score and make candidate's ranking. Table 4 shows scoring interval for parent's income parameter.

Table 4. Scoring Interval for Parent's Income

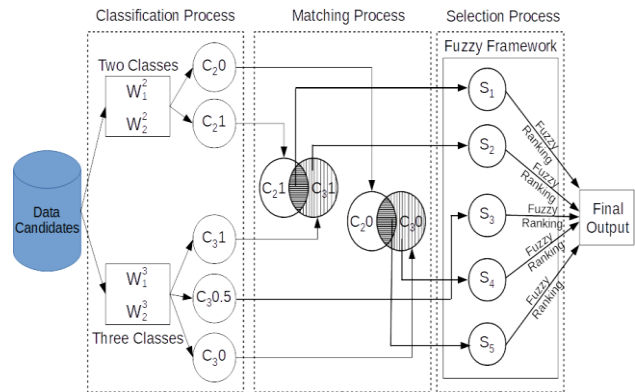| Interval	Score
Parent's income $\leq$ Rp250,000	100
Rp250,000 < Parent's income $\leq$ Rp500,000	90
..	..
Parent's income $\geq$ Rp3,000,000	10

Figure. 1 BM scholarship's selection model based on BPNN and FIS approach.

The following rules are used to define the actual class and target of two and three recommendation classes with assuming the given quota of 50% of total N candidates.

Two classes:
  b_{middle} = (maximum(ms) + minimum(ms)) / 2;
  IF ms< b_{middle} THEN ac=non-recommended; at=0
  ELSE ac=recommended; at=1.

Three classes:
  b_{lower} = (maximum(ms) + minimum(ms)) / 3;
  b_{upper} = 2 * b_{lower};
  IF ms $\leq$ b_{lower} THEN ac=non-recommended; at=0
  ELSE IF ms $\leq$ b_{upper} THEN ac=recommended; at=0.5 ELSE ac=highly-recommended; at=1.

ms, ac, and at are a manual score, actual class, and actual target; and b_{middle}, b_{lower}, b_{upper} are boundaries to decide actual class and target for two and three recommendation classes.

2.2 BM Scholarship's selection model

This research proposes a selection model for BM scholarship based on BPNN and FIS approaches. The model consists of three processes: classification, matching, and selection processes as shown in Fig.1.

In the classification process, candidates will be classified into two and three recommendation classes using weighted matrices W_1^2, W_2^2 for two classes, and W_1^3, W_2^3 for three classes, where those weighted matrices are derived from the corresponding training processes. The results are a set of recommended and non-recommended candidates indicated by C_21 and

Table 5. Defined priority set of candidates

$S_1 = C_31 \cap C_21$
$S_2 = C_31 - (C_31 \cap C_21)$
$S_3 = C_30.5$
$S_4 = C_30 - (C_30 \cap C_20)$
$S_5 = C_30 \cap C_20$

C_20, and a set of high, moderate, and low recommended candidates indicate by C_31, $C_30.5$, and C_30.

In the matching process, the operation of intersection and subtraction set will be applied to obtain the priority set of candidates indicated by S_1, S_2, S_3, S_4, and S_5 as shown in Table 5. The highest priority set is to S_1, and the lowest is to S_5. After obtaining the priority set of candidates then perform selection process sequentially from S_1 to S_5 using the fuzzy framework to make candidate's ranking and then combine them to obtain the final result. The fuzzy framework will be discussed in section 2.4.

2.3 Design multi-layer perceptron (MLP)

Fundamental theorem of Multi-Layer Perceptron (MLP), Back-Propagation (BP) algorithm, and activation function described in this section are refer to [28].

MLP with one hidden layer and one output layer is designed in this work to running the classification process with two and three recommendation classes as shown in Fig.2 (a) and (b). Eleven input parameters indicated by P_1 to P_{11}, referring to the input parameters of BM scholarship as described in table 1 above.

Classification with two recommendation classes has fifteen neurons in the hidden layer while classification with three classes has eight neurons. Both classifications have one neuron in output layer indicated by vo while b_1 and b_2 are biased. W_1^2, W_2^2, and W_1^3, W_2^3 indicate two pairs of weighted matrices of corresponding classification classes. The target of the training process is to obtain the best-weighted matrices which will be used to predict BM scholarship's candidates into corresponding recommendation classes. The values of W_1^2, W_2^2, W_1^3, and W_2^3 are randomly generated in the interval [-1,+1] at the beginning of training process and will be adjusted during the training process until the error between actual and predicted values was sufficiently small. This mechanism is known as BP algorithm which consists of two passing processes: pass forward and pass backward. Pass forward calculates the output network and compares this with actual output to obtain the error. Pass backward propagates the error back to each neuron in the previous layer to update the weighted matrices to minimize the error. A cycle of one pass forward and one pass backward over the network is known as one epoch. The training process can be terminated by initialized the accepted error or by defined the number of epochs for the training process.

Each neuron has a summing and an activation function. Fig. 3 shows a mathematical model of a neuron where P_1 to P_{11} are input parameters, W_1 to W_{11} are weight values, x_0 is biased, f(.) is activation function, and μ_i is the output of neuron v_i. Sigmoid used as activation function as shown in Eq. (1).

$$\varphi(yi) = \frac{1}{1+e^{yi}} \qquad (1)$$
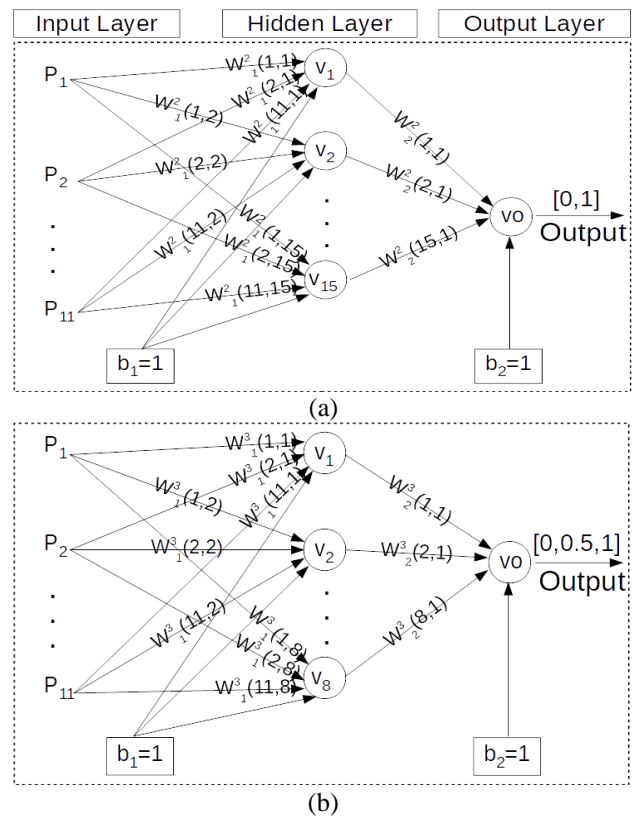

(a)

(b)

Figure. 2 MLP model for classification process with two (a) and three (b) recommendation classes
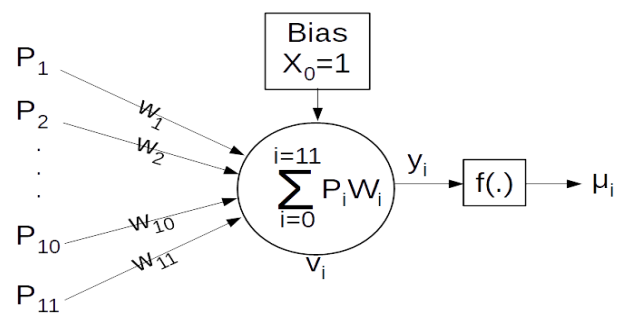

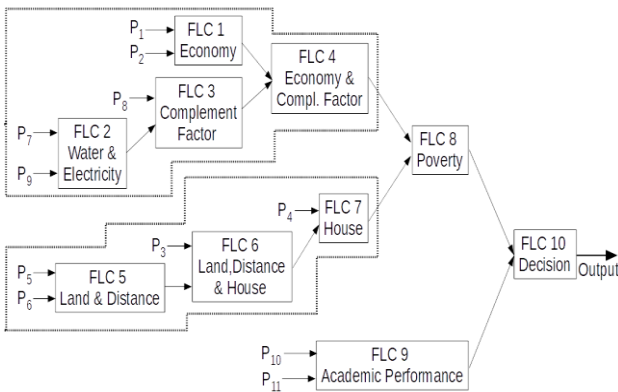Figure. 3 Mathematical model of a neuron

Figure. 4 Modified version of BM scholarship fuzzy framework based on R value of input parameters

Table 6. Design rule bases for FLC 1

Consequent part: 1="Poor";2="Average"; 3="Above average"		Parent's Income (p)		
		Low (1)	Average (2)	High (3)
Dependent (d)	Low (1)	1	2	2
	Average (2)	1	2	3
	Many (3)	2	3	3

2.4 Redesigning BM scholarship framework

An accuracy of 71.43% was achieved used the fuzzy framework proposed in [2]. The framework was redesigned as described in [4] by considering correlation coefficient (R) of input parameters as shown in Fig. 4. Eq. (2) use to calculate R value between input parameters and total manual score.

$$R_j = \frac{\sum_{i=1}^n \left(X_i^j - \overline{X_j}\right)(Y_i - \overline{Y})}{\sqrt{\sum_1^n \left(X_i^j - \overline{X_j}\right)^2 \sum_1^n (Y_i - \overline{Y})^2}} \tag{2}$$

R_j denotes R values of jth parameter; X_i^j denotes j^{th} parameter value belonging to i^{th} candidate; Y_i is a total manual score of ith candidate. $\overline{X_j}$ denotes average value of jth parameter; Y and $\overline{Y}$ are total manual score and its average; j refers to P# in Table 1, and n is the number of candidates.

Input parameters with lower R value will be processed first, and its output will be paired to another parameter with higher R value to be processed into next FLC until derive the final output from FLC decision. R values which were calculated in [4] used 119 datasets with complete parameters. Results show that R value of parameter P_8 is greater than P_7 and P_9 then P_7 and P_9 will be processed first in FLC 2, and its output will be paired with P_8 to be processed in FLC 3 as shown in the economy and complement factor block in Fig. 4. The same way was used to determine the input pairs of FLC 5, 6, and 7.

Figure. 5 5-fold cross-validation of BM scholarship data

This work is also using the modified framework for the selection process.

R values of input parameters also considered in designing the rule bases for each FLC. The parameter with higher R value contributes more to defining the consequent part of IF-THEN rules. Table 6 shows designing rule bases for FLC 1. The number of membership function is based on the optimal number of cluster in the input parameter which is calculated by using the Elbow method. Because of the optimal number of cluster of parent's income and dependent parameters are three then the number of fuzzy membership function of both parameters are set to 3. The type of triangular membership function is used in antecedent as well as consequent parts.

Rule #2, #4, and #6 can be read as follows:
Rule #2: IF (p is *Average*) AND (d is **Low**) THEN q is ***Average***
Rule #4: IF (p is **Low**) AND (d is ***Average***) THEN q is ***Low***
Rule #6: IF (p is **High**) AND (d is ***Average***) THEN q is ***High,***
where q is consequent. In defining the consequent part of IF-THEN rules, the relative weight is to parent's income because of higher R value than the dependent parameter. Designed rule bases for other FLCs used the same method. Other methods as well as the algorithm used to design the membership function for each input data set described in ref. [2].

3. Experimental results

In these experiments using total 625 datasets excluding the national high school test score parameter. 5-fold cross-validation as shown in Fig. 5 is used to observe the accuracy of the classification process.

3.1 Closed test

3.1.1. Classification with two classes

The best system performance to classify the candidates into two recommendation classes is achieved by using architecture MLP as shown in Fig.2

(a) and a combination of network parameters: learning rate = 0.01, epoch number = 10,000, and maximum of Mean Square Error (MSE) = 0.01. Accuracy average on testing data is 88.48% where the lowest accuracy is 81.6%.

After training process, five pairs of weighted matrices W_1^2 and W_2^2 are obtained for each fold. To choose the best as system weighted matrices to predict BM scholarship's candidates into two recommendation classes in the future then classification process performs for all data using those weighted matrices. The highest accuracy of 91.36% was achieved from fold two; thus the system weighted matrices W_1^2 and W_2^2 were chosen from fold two. Table 7 shows classification result on all data over 5-fold cross-validation while Table 8 shows predicting result using those weighted matrices on total 625 candidates.

Table 7. Classification result with two recommendation classes on all data

625 Data	Fold#				
	1	2	3	4	5
Predicted True	564	571	556	561	560
Predicted False	61	54	69	64	65
Accuracy (%)	90.24	91.36	88.96	89.76	89.6
Highest Accuracy: **91.36%**					

Table 8. Prediction result on 625 candidates into two recommendation classes

Recommendation Classes	
$C_21 = 328$ candidates	$C_20 = 297$ candidates

Table 9. Confusion matrix of final selection result on classification with two classes

n = 625 (Total Candidates)		Fuzzy Ranking	
		<=312	*>312*
Manual Ranking	*<=312*	TP = 282	FN = 30
	>312	FP = 30	TN = 283
Accuracy = **90.4%**		312	313

Table 10. Classification result with three recommendation classes on all data

625 Data	Fold#				
	1	2	3	4	5
Predicted True	473	471	485	481	476
Predicted False	152	154	140	144	149
Accuracy (%)	75.7	75.4	77.6	77	76.2
Highest Accuracy: **77.6%**					

Table 11. Prediction result on 625 candidates into three recommendation classes

Recommendation Classes		
$C_31 = 225$	$C_30.5 = 204$	$C_30 = 196$

Next, selection process performs first on 328 recommended candidates in the set of C_21 follows by 297 non-recommended candidates in C_20 and then combine their fuzzy ranking to obtain the final result. By using Eq. (3) an accuracy of 90.4% is obtained, indicates increasing accuracy around 9% compared to the accuracy of 81.76% if selection process performs directly without classification process, means the proposed framework and the classification process are helpful to increase selection accuracy.

$$Accuracy = \frac{(TP+TN)}{n} \times 100\% \qquad (3)$$

n is total data. Table 9 shows confusion matrix of selection result on classification with two classes. TP, FP, FN, and TN are true positive, false positive, false negative, and true negative. TP=282 and TN=283 show that 565 candidates are correctly predicted by FIS while 60 candidates are incorrectly predicted indicated by FP+FN.

3.1.2. Classification with three classes

The architecture of MLP as shown in Fig.2 (b) is used to run the classification process with three recommendation classes. The best result is achieved by using the combination of network parameter: Learning rate = 0.001; Number of epoch = 10,000; maximum MSE = 0.001; and the thresholds to decide prediction class is 0.25 and 0.70. If the network's output is less than 0.25, then prediction class decided to low recommended class and if network's output is greater than 0.70 to highly-recommended class, otherwise to moderate recommended class. Classification's result on testing data over 5-fold cross validation shows the accuracy average, and the lowest accuracy is 74.88% and 68.8% while the highest accuracy of 77.6% is obtained from fold 3 when classification process performed on all data. Table 10 shows classification result with three recommendation classes on all data.

Based on the results above, system weighted matrices W_1^3 and W_2^3 are chosen from fold three to predict BM scholarship's candidates into three recommendation classes. Table 11 shows prediction result using system weighted matrices W_1^3 and W_2^3 on total 625 candidates.

Table 12. Confusion matrix of final selection result on classification with three classes

n = 625		Fuzzy Ranking	
		<=312	>312
Manual Ranking	<=312	TP = 267	FN = 45
	>312	FP = 45	TN = 268
Accuracy = **85.6%**		312	313

Table 13. Confusion matrix of selection result on priority set S1

Number of candidates n = 225		Fuzzy Ranking	
		<=312	>312
Manual Ranking	<=312	TP = 212	FN = 0
	>312	FP = 13	TN = 0
Accuracy = **94.22%**		225	0

Table 14. Confusion matrix of selection result on priority set S5

Number of candidates n = 192		Fuzzy Ranking	
		<=312	>312
Manual Ranking	<=312	TP = 0	FN = 1
	>312	FP = 0	TN = 191
Accuracy = **99.48%**		0	192

Table 15. Confusion matrix of final selection result of section 3.1.3

n = 625 (Total Candidates)		Fuzzy Ranking	
		<=312	>312
Manual Ranking	<=312	TP = 267	FN = 45
	>312	FP = 45	TN = 268
Accuracy = **85.6%**		312	313

Table 16. Comparison of selection accuracy between section 3.1.1, 3.1.2, and 3.1.3

Selection Accuracy		
Section 3.1.1	Section 3.1.2	Section 3.1.3
90.4%	85.6%	85.6%

Table 17. Comparison of ratio average between selected candidates and number of quota

Ratio Average		
Section 3.1.1	Section 3.1.2	Section 3.1.3
78.43%	80.58%	80.67%

The results show that 225 are predicted as highly recommended candidates while 204 and 196 are predicted as moderate and low recommended candidates.

Classification accuracy of 77.6% was decreased compared with the accuracy of two recommendation classes. It can be understood by the natural logic that it is easier to make a decision about two given choices

than three. However, classification of three recommendation classes can be used to determine a set of highly recommended candidates from recommended candidates, and a set of low recommended candidates from non-recommended candidates which is provided by classification with two recommendation classes by matching process between them.

Next, selection process performs first on 225 candidates in the set of $C_3 1$ follows by 204 candidates in $C_3 0.5$ and 196 candidates in $C_3 0$. Table 12 shows confusion matrix of selection result on classification with three recommendation classes. The accuracy of 85.6% is obtained using Eq. (3).

3.1.3. Matching process as the final results

Based on the methodology of the matching process as shown in Fig. 1 and assigned priority set as shown in table 5, number of candidates in priority set $S_1=225$, $S_2=0$, $S_3=204$, $S_4=4$, and $S_5=192$. Next, selection process performs first on 225 candidates in the priority set of S_1 follow by 204 candidates in S_3, four candidates in S_4, and 192 candidates in S_5. Table 13 and 14 show confusion matrix of selection result on the priority set S_1 and S_5.

The results show that system can able to select 225 as the best-recommended candidates and reject 192 the worst non-recommended candidates from total 625 candidates with satisfactory accuracies of 94.22% and 99.48% for the given quota of 312. It means that system workload is decreased around 66.7% calculated by $(225+192)/625 \times 100\%$ which is an advantage when selection process conducted on matching's result between classification with two and three classes.

Table 15 show confusion matrix of the final results in this section while table 16 shows a comparison of selection accuracies between section 3.1.1, 3.1.2, and 3.1.3.

Although section 3.1.1 provides the highest accuracy of 90.4%, the best ratio average between selected candidates and number of quotas is provided by section 3.1.3 as shown in Table 17. It is shown that the best system performance in the selection process is provided by section 3.1.3 thus selection model of section 3.1.3 is chosen as propose model for BM scholarship selection process.

Fig. 6 shows selected candidates according to percentages of given quotas where selection model of sections 3.1.2 and 3.1.3 are almost same, but section 3.1.3 provides better ratio average of 80.67%.
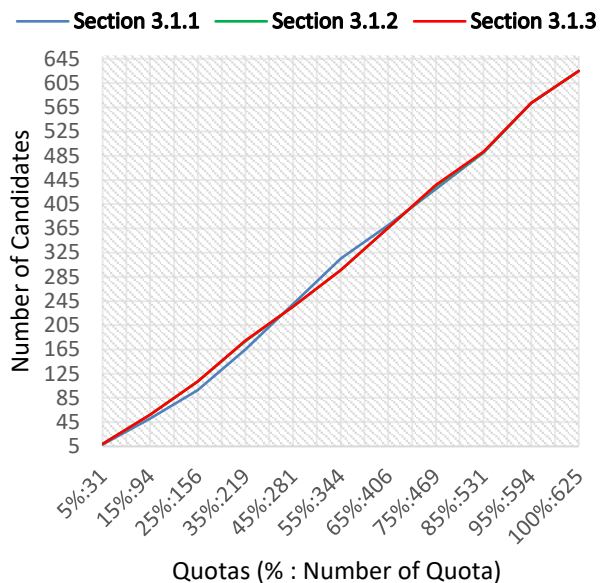
Figure. 6 Selected candidates per percentages of quota between selection model of section 3.1.1, 3.1.2, and 3.1.3
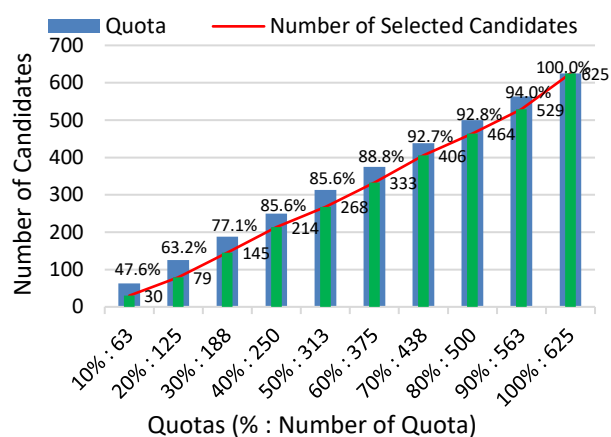

Figure. 7 Ratio between selected candidates and percentages of given quotas of section 3.1.3

Table 18. Candidate's fuzzy ranking

| Candidate # | FLC Decision | | | Ranking | |
| | Input | | | | |
	Output from FLC8 (Poverty)	Output from FLC10 (Academic)	Output	Fuzzy	Manual
89	13.3250405	50	86.6750	1	16
..	..	..		..	..
358	86.675	50	13.3250	625	601

Fig. 7 shows the ratio between selected candidates and number of the given quotas of section 3.1.3. Table 18 shows candidate's fuzzy ranking as the final result.

Table 19. Open test result of the proposed system over 5-fold cross-validation

Selection Accuracy per Fold				
1	**2**	**3**	**4**	**5**
90.40%	85.60%	84.00%	87.20%	87.20%
Accuracy average = 86.88%				
Standard deviation = 2.37				

3.2 Open test

This section reports the open test for the proposed model. Assuming training and testing data in 5-fold cross-validation as shown in Fig. 5 are current year data and coming year data, and the quota is 50% of total data, then for each fold contain 500 current year data with the quota of 250, and 125 coming year data with the quota of 62. Training process with two and three recommendation classes perform on training data in each fold to obtain their weighted matrices and then classify testing data using those weighted matrices then perform matching and selection processes follow the way in the proposed method. Table 19 shows open test result over 5-fold cross-validation.

4. Discussion

4.1 Comparison between closed and open tests

The results of closed and open tests revealed that accuracy average of 86.88% of the open test is not too far to the accuracy of 85.60% of the closed test result, where the difference around 1%. It means the proposed system will provide almost same accuracy for BM scholarship's data in the future. On the other hand, lowest and highest accuracies of the open test results indicate that the proposed system can predict and select BM scholarship's candidates in the future with satisfactory accuracies from 84% up to 90%, and a small standard deviation value of 2.37 indicates a stability of system performance in selection process; thus the proposed system can be trusted as a proper model for BM scholarship selection model.

4.1 Comparison with other similar works

This section reports the comparison of classification and selection accuracies between this work and similar works include author's previous works [2,4]. Selection model in this work is referred to selection model in section 3.1.3. Because of the accuracies of similar works in [25-27] are not presented in their paper then authors calculated them as described in [4]. Selection accuracy of ref. [4] will be compared in this section is taken from the

experiment on a dataset with incomplete parameters. Classification accuracy with two recommendation classes will be compared in this section because of other similar works also conducted classification process only on two recommendation classes. Table 20 shows comparison accuracy between this work and similar works. Comparison result shows that the proposed technique provides better accuracy in both classification and selection processes for the special case of BM scholarship selection in this research.

Fig. 8 shows a comparison of time complexity $T(n)$ with assumed quota ≤ 225 from total 625 candidates. It can be seen that this work provides lower $T(n)$ indicates higher efficiency thus this work can be said more efficient than other similar works include author's previous works. $T(n)$ of ref. [4] is 326 derived by using its weighted matrices for classify 625 data used in this work and its result shows that 326 data classified as recommended candidates, while $T(n)$ of the work in ref. [27] can be estimated in the range of [326,625] due to its accuracy of 89.9% is less than classification accuracy provided by [4].

Table 20. Comparison results between this work and other similar works

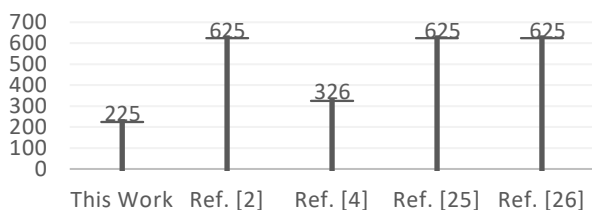| Ref.	Classification Process		Selection Process	
	Method	Accuracy	Method	Accuracy
[2]	-	-	Mamdani FIS with Elbow method, K-Mean clustering	71.4%
[4]	BPNN	90.1%	Mamdani FIS with Elbow method, K-Mean clustering, and Pearson's correlation	82.9%
[25]	-	-	FMADM with TOPSIS and WP	
[26]	-	-	FMADM with TOPSIS	75%
[27]	SVM	89.9%	FMADM with AHP and TOPSIS	
This work	BPNN	91.3%	Mamdani FIS with Elbow method, K-Mean clustering, Pearson's correlation, and additional technique refer to section 3.1.3	85.6%

Figure. 8 Comparison of time complexity $T(n)$

5. Conclusion

This paper combines two famous approaches in machine learning domain, namely BPNN and FIS, as a new selection model for BM scholarship selection process. BPNN used to classify candidates into two, and three recommendation classes then perform FIS on priority set of candidates which was derived from the intersection of classification's result of two and three recommendation classes to make candidate's ranking. The closed test result shows that for assuming quota of 50% from total 625 candidates, the system can choose 267 highly recommended candidates and reject 268 non-recommended candidates while 45 are identified as moderate recommended, and 45 other are identified as the lower priority with selection accuracy of 85.6% better than previous similar works. The open test results provide accuracy average of 86.88%, and the lowest accuracy of 84% which is close enough to the closed test's accuracy of 85.6% with the difference around 1% and a small standard deviation value of 2.37 indicate stability of system performance in the selection process. The lowest and highest accuracies of the open test results indicate the proposed system can predict and select BM scholarship's candidates in the future with satisfactory accuracies from 84% up to 90%; thus the proposed system can be trusted as a proper model for BM scholarship selection process.

In the future, an intelligent recommender system can be develop based on this proposed model for BM scholarship selection process.

References

[1] The National Education Ministry Republic of Indonesia, *"Bidik Misi" Scholarship Guidance Book*, Directorate of Learning and Student Affair, Jakarta, 2015.

[2] L. A. Latumakulita, F. Purnama, D. A. Prima, S. Paturusi, and T. Usagawa, "Indonesia scholarship selection framework using fuzzy inferences system approach. Case study: "Bidik Misi" scholarship selection", In: *Proc. of International Conf. on Information & Communication Technology and Systems (ICTS)*, Surabaya, Indonesia, pp.107-113, 2016.

[3] I. Mala, P. Akhtar, A. R. Memon, and T. J. Ali, "Automatic Generation of Fuzzy Membership Function based on K-Mean and EM Clustering", *International Journal of Sindh University*

Reseach Journal (Science Series), Vol.46, No.2, pp.123-128, 2014.

[4] L. A. Latumakulita and T. Usagawa, "A Combination of Backpropagation Neural Network on Fuzzy Inference System Approach in Indonesia Scholarship Selection Process. Case Study: "Bidik Misi" Scholarship Selection", In: *Proc. of International Conf. on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD 2017)*, Guilin, China, pp.1273-1278, 2017.

[5] B. K. Singh, K. Verma, and A. S. Thoke, "Adaptive Gradient Descent backpropagation for classification of breast tumors in ultrasound imaging", In: *Proc. of International Conf. on Information and Communication Technologies (ICICT 2014)*, Vol.46, Kochi, India, pp.1601-1609, 2015.

[6] O. W. Samuel, G. M. Asogbon, and A. K. Sangaiah, "An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction", *International Journal of Expert Systems With Application,* Vol.68, pp.163-172, 2017.

[7] Z. Lin, J. Minping, S. Guanglin, and S. Yanyou, "Estimation study of stress concentration factor of crack structure based on BP neural networks", In: *Proc. of International Conf. on Measuring Technology and Mechatronics Automation*, Macau, China, pp.270-273, 2016.

[8] J. Jagtap and M. Kokare, "Human age classification using facial skin aging features and artificial neural network", *International Journal of Cognitive System Research,* Vol.40, pp.116-128, 2016.

[9] R. Olawoyin, "Application of backpropagation artificial neural network prediction model for PAH bioremediation of polluted soil", *International Journal of Chemosphere,* Vol.161, pp.145-150, 2016.

[10] EI. Sela, S. Hartati, A. Harjoko, R. Wardoyo, and M. Mudjosemedi, "Segmentation On Periapical Dental X-Ray Images for Osteoporosis Screening", *International Journal of Advanced Computer Sciences and Applications (IJACSA),* Vol.4, No.7, pp.147-151, 2013.

[11] S. Thawkar and R. Ingolikar, "Automatic Detection and Classification of Masses in Digital Mammograms", *International Journal of Intelligent Engineering and Systems,* Vol.10, No.1, pp.65-74, 2017.

[12] M. Bhatt and T. Patalia, "Neural Network Based Indian Folk Dance Song Classification Using MFCC and LPC", *International Journal of Intelligent Engineering and Systems,* Vol.10, No.3, pp.173-183, 2017.

[13] R.K. Chinthaginjala and K. Bagadi, "Receiver Design Using Artificial Neural Network for Signal Detection in Multi Carrier-Code Devision Multiple Access System", *International Journal of Intelligent Engineering and Systems,* Vol.10, No.3, pp.66-74, 2017.

[14] H.R. Ansari, M.J. Zarei, S. Sabbaghi, and P. Keshavarz, "A New Comprehensive Model for Relative Viscosity of Various Nanofluids Using Feed-Forward Back-Propagation MLP Neural Networks", *International Journal of International Communications in Heat and Mass Transfer,* Vol.91, pp.158-164, 2018.

[15] J.Z. Wang, J.J. Wang, Z.G. Zhang, and S.P. Guo, "Forecasting Stock Indices with Back Propagation Neural Network", *International Journal of Expert Systems with Applications,* Vol.38, No.11, pp.14346-14355, 2011.

[16] H. Mo, J. Wang, and H. Niu, "Exponent Back Propagation Neural Network Forecasting for Financial Cross-Correlation Relationship", *International Journal of Expert Systems with Applications,* Vol.53, pp.106-116, 2016.

[17] Y.R. Zeng, Y. Zeng, B. Choi, and L. Wang, "Multifactor-influenced Energy Consumption Forecasting Using Enhanced Back-Propagation Neural Network", *International Journal of Energy,* Vol.127, pp.381-396, 2017.

[18] Y. Yu and C. Han, "Rough Sets – Least Square and Neural Networks in Fault Diagnosis Shield Applied Research", *International Journal of Intelligent Engineering and Systems,* Vol.3, No.3, pp.42-49, 2010.

[19] S. Liu, L. Xu, and D. Li, "Multi-Scale Prediction of Water Temperature Using Empirical Mode Decomposition With Back-Propagation Neural Networks", *International Journal of Computers and Electrical Engineering,* Vol.49, pp.1-8, 2016.

[20] A. Chatterjee, S. Mukherjee, and S. Kar, "Poverty level of households: A Multidimensional Approach Based on Fuzzy Mathematics," *International Journal of Fuzzy Information And Engineering*, Vol.6, No. 4, pp.463-487, 2014.

[21] P. Giordani and G.M Giorgi, "A Fuzzy Logic approach to Poverty Analysis Based on the Gini and Bonferroni Inequality Indices", *International Journal of Statistical Methods & Application*, Vol.19, No.4, pp.587-607, 2010.

[22] G. Maniu, "The Construction of the Membership Functions in the Fuzzy Measuring of Poverty", *International Journal of Economic Insights –*

Trends and Challenges Petroleum-Gas University of Ploiesti Bulletin, Vol.61, No.1, pp. 107-117, 2009.

[23] I. Saleh and S. Kim, "A Fuzzy System For Evaluating Students' Learning Achievement", *International Journal of Expert Systems With Application,* Vol.36, pp.6236–6243, 2009.

[24] S. Chen and T. Li, "Evaluating Students' Learning Based On Fuzzy Rules With Fuzzy Reasoning Capability", *International Journal of Expert Systems With Application,* Vol.38, pp.4368–4381**,** 2011.

[25] S. Uyun and I. Riadi, "A Fuzzy TOPSIS Multiple-Attribute Decision Making for Scholarship Selection", *International Journal of Telecommunication, Computing, Electronics and Control (TELKOMNIKA)*, Vol. 9 No.1, pp. 37–46, 2011.

[26] G. A. M. S. Wimatsari, I. K. G. D. Putra, and P. W. Buana, "Multi-Attribute Decision Making Scholarship Selection Using A Modified Fuzzy TOPSIS", *International Journal of Computer Science Issues (IJCSI),* Vol. 10, Issue 1, No.2, pp.309-317**,** 2013.

[27] M. G. Langgawan, W. Ariayanti, and I. Cholissodin, "Selection and Recommendation Scholarship Using AHP-SVM-TOPSIS", *International Journal of Information Technology and Computer Science*, Vol.1, No.1, pp.1-13, 2016

[28] S. Haykin, *Neural Networks A Comprehensive Foundation*, Second Edition, Pearson Education, Delhi, India. 1999.