Jurnal Varian
Vol. 8, No. 3, October 2025, pp 319–332
e-ISSN: 2581-2017

Evaluating Fisherman Insurance Participation using Bagging Multivariate Adaptive Regression Splines

Ulil Azmi[1], Soehardjoepri[1], Prilyandari Dina Saputri[1], Thalia Rizki Salsabila[1], Widya Iswara[1], Roslinazairimah Zakaria[2]

[1]Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2]Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, Malaysia

Article Info

Article history:

Received : 07-19-2025
Revised : 10-22-2025
Accepted : 10-28-2025

Keywords:

Bootstrap Aggregating;
Fishermen;
Independent Fishermen's Insurance;
Multivariate Adaptive Regression Spline;
Risk.

ABSTRACT

The Fishermen's Insurance Premium Assistance Program and the Independent Fishermen's Insurance Scheme are initiatives by the Indonesian government aimed at enhancing the protection of fishermen, whose occupations are considered high-risk compared to other professions. One of the regions actively participating in both programs is Lekok District, located in Pasuruan Regency, East Java Province. The objective of this research is to analyze the factors influencing fishermen's participation in self-funded insurance schemes using the Multivariate Adaptive Regression Spline method. The research is based on primary data collected through direct surveys and structured questionnaires distributed to fishermen in Lekok District. The results of this research are that five key variables significantly influence participation, with the most influential factor being participation in outreach or socialization activities. Other important factors include the number of family members (X_4), income (X_3), and age (X_1), while fishing experience (X_5) does not show a significant effect. The model's classification accuracy on the training data reached 82%, while on the test data it was 75.8%. Furthermore, applying the bootstrap aggregation technique to Multivariate Adaptive Regression Splines models significantly improved classification accuracy to 92% on the training data and 100% on the test data. The findings are expected to support stakeholders in formulating strategies to increase fishermen's engagement in independent insurance programs. Strengthening such participation is crucial for reducing occupational risks, ensuring the sustainability of fishing activities, and improving the welfare and resilience of the fishing community.

Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: https://doi.org/10.30812/varian.v8i3.5373

Corresponding Author:

Ulil Azmi,
Institut Teknologi Sepuluh Nopember, Surabaya,
Email: ulilazmi0211@gmail.com

How to Cite:
Azmi, U., Soehardjoepri, S., Saputri, P. D., Salsabila, T. R., Iswara, W., & Zakaria, R. (2025). Evaluating Fisherman Insurance Participation using Bagging Multivariate Adaptive Regression Splines. *Jurnal Varian*, *8*(3), 319–332.

A. INTRODUCTION

Indonesia is a maritime country with vast marine areas. Fishermen are one of the livelihoods of many coastal communities in Indonesia, with a total of 2,210,000, which are divided into 59.86% full-time fishermen, 26.20% main part-time fishermen, and 13.94% additional part-time fishermen (Statistik, 2016). According to the International Labor Organization (ILO) Convention No.

188 of 2007, work in the fisheries and marine sector is considered dangerous compared to other work. Destructive Fishing Watch (DFW) Indonesia recorded 42 accidents at sea from December 2020 to June 2021. One of the government's efforts to increase protection for fishermen against work-related accidents or fatalities is the Fishermen's Insurance Program. The Fishermen's Insurance Premium Assistance Program (FIPAP) is a government program that provides insurance assistance to fishermen, aiming to raise awareness of the importance of insurance, build fishermen's desire to participate in insurance independently, and provide guarantees against the risks individual fishermen face. According to the National Coordinator of DFW Indonesia, many fishermen are unaware of the insurance program. Out of a total of 2,643,902 fishermen, 1,014,498 fishermen have registered with FIPAP (Asyya & Agusta, 2021). This means that half of the total fishermen do not have a fishermen's insurance program.

The lack of fishermen's participation cannot be separated from the bureaucratic and procedural conditions that make it difficult for every fisherman to access insurance, as well as the lack of socialization from the Ministry of Maritime Affairs and Fisheries (MMAF), local governments, and insurance companies. One of the areas participating in the FIPAP and fishermen's self-insurance programs is Jatirejo Village, Lekok District, Pasuruan Regency, East Java Province, Indonesia. The area has 4 Joint Business Groups (KUB). However, not all the KUB members have fishermen's insurance. Many factors influence the ownership of fishermen's insurance in Jatirejo Village, including the lack of awareness and knowledge among fishermen, family dependents, income, and socialization in the area, encouraging participation in fishermen's insurance. To examine the relationship between the predictor and response variables, regression analysis can be used. However, if there is no information about the shape of the function and there is no clear pattern of relationship between the predictor variables and the response variable, the analysis can be done using a nonparametric regression approach (Eubank, 1999).

Previous studies applied MARS to domains such as bankruptcy forecasting, fraud detection, and health (Amin et al., 2020; De Andrés et al., 2011; Hlokoe et al., 2022). The MARS method has the advantage of being relatively flexible and innovative to investigate the pattern of relationships between variables without assumptions about their functional form (Al-Musaylh et al., 2018). MARS is focused on overcoming the problem of high dimensionality and discontinuity in data and involves many interactions between variables (Bose et al., 2021; Hastie et al., 2009; Seno et al., 2024). The classification accuracy of the MARS model can be improved using resampling, one of which is the Bootstrap Aggregating (Bagging) method which can improve stability, increase accuracy, and predictive power (Hasyim et al., 2018; Rupilu & Rosadi, 2024).

In the context of fishermen's insurance, most studies have relied on survey and regression-based approaches (Brahmantyo et al., 2021) but rarely adopted advanced machine learning models. However, no prior research has specifically examined fishermen's insurance participation in Indonesia using MARS combined with Bagging to enhance predictive performance. This study, therefore, differs by integrating Bagging into MARS to significantly improve classification accuracy, while also focusing on socio-economic factors influencing fishermen's participation in independent insurance schemes in Lekok District, Pasuruan Regency, as obtained from the Pasuruan Regency Fisheries Service. At the same time, primary data were collected through direct surveys and structured questionnaires administered to fishermen in the region. The contribution of the research will determine variables considered to influence fishermen's participation in independent fishermen's insurance, including age, education level, length of work, duration of fishing, income per month, number of family dependents, health, participation in socialization regarding independent fishermen's insurance, boat ownership status, and number of boats owned by fishermen. The type of fishermen's insurance studied is independent fishermen's insurance. Binary categorical responses are respondents who have independent fishermen insurance and respondents who do not have independent fishermen insurance.

B. RESEARCH METHOD

The data used in this study are primary data obtained through direct surveys on fishermen in Jatirejo Village, Lekok District, using a questionnaire. The total population of fishermen in Jatirejo Village is 687. The study used 162 respondents.

The variables used in this study comprise response and predictor variables, as presented in Table 1.

Table 1. Research Variable

Variable	Description	Category	Scale
		Response Variable	
Y	Ownership of fishermen's insurance	0 = Has independent fishermen insurance 1 = Does not have independent fishermen's insurance	Nominal
X_1	Age	-	Ratio

Variable	Description	Category	Scale
X_2	Education Level	0 = Below primary school/not in school	Nominal
		1 = Elementary school graduate	
		2 = Junior high school graduate	
		3 = High school graduate	
X_3	Income	-	Ratio
X_4	Number of Dependents	-	Ratio
X_5	Experience	-	Ratio
X_5	Experience	-	Ratio
X_6	Duration at sea	-	Ratio
X_7	Health	0 = Smoking	Nominal
		1 = Does not smoke	
X_8	Participation in Socialization	0 = active	Nominal
		1 = inactive	
X_9	Boat Ownership Status	0 = own/personal property	Nominal
		1 = someone else's	
X_{10}	Number of Ships	-	Ratio

The research methodology as shown in Figure 1:

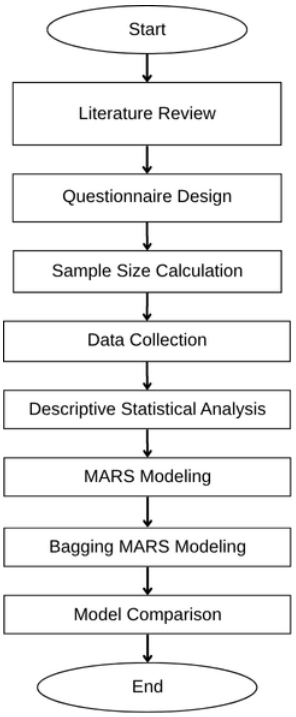

Figure 1. Research Workflow Diagram

Figure 1 illustrates the research stages, from literature review and data collection to statistical modelling using MARS and Bagging MARS, and ending with model evaluation and interpretation. MARS modelling is applied to explore key predictors and build the initial classification model. Furthermore, additional Bagging MARS is performed with multiple replications to improve model accuracy and stability. Then, to select the best model, both methods are compared using the 1-APER metric, specificity, and sensitivity. The study phases employed in the methodology mentioned in Figure 1 are as follows.

1. Multivariate Adaptive Regression Spline (MARS)

One of the nonparametric regression approaches is Multivariate Adaptive Regression Spline (MARS), which was introduced by Friedman in 1991 and is suitable for use on data patterns where the shape of the regression curve is unknown, and or there is no complete past information about the shape of the data pattern (Eubank, 1999). The MARS model is focused on solving high-dimensional problems and discontinuities in the data (Hasyim et al., 2018).

A high-dimensional problem is one with a large number of variables and a large sample size, requiring complex calculations. One of the advantages of using the MARS method is its ability to estimate the contribution of the basis function to the response variable, by not only being able to capture adaptive effects but also being able to capture interaction effects between predictors (Otok et al., 2023). The MARS method became popular because it does not specify special types, such as the relationship (linear, quadratic, and cubic) between the predictor and response variables The MARS model formation process does not require assumptions (Hastie et al., 2009).

In general, the MARS model, according to Friedman (1991) can be written in Equation (1) below:

$$y_i = \alpha_0 + \sum_{m=1}^{M} \alpha_m B_m(x) + \varepsilon_i \tag{1}$$

where, α_0 is the constant coefficient of the basis function B_0, and α_m is the coefficient of the mth basis function, and $B_m(x) = \prod_{k=1}^{Km} \left[s_{km}(x_{v(k,m)} - t_{km}) \right]$. So, if written in matrix form is Equation (2) below:

$$y = B\alpha + \varepsilon \tag{2}$$

where,

$$y = (y_1,\ y_2, \ldots,\ y_n)^T,\ \alpha = (\alpha_1,\ \alpha_2, \ldots,\ \alpha_M)^T,\ \varepsilon = (\varepsilon_1,\ \varepsilon_2, \ldots,\ \varepsilon_n)^T \tag{3}$$

$$B = \begin{bmatrix} 1 & \prod_{k=1}^{K_1} \left[s_{1m}(x_{1(1,m)} - t_{1m}) \right] & \cdots & \prod_{k=1}^{K_1} \left[s_{Mm}(x_{1(M,m)} - t_{Mm}) \right] \\ 1 & \prod_{k=1}^{K_1} \left[s_{1m}(x_{2(1,m)} - t_{1m}) \right] & \cdots & \prod_{k=1}^{K_1} \left[s_{Mm}(x_{2(M,m)} - t_{Mm}) \right] \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1} \left[s_{1m}(x_{n(1,m)} - t_{1m}) \right] & \cdots & \prod_{k=1}^{K_1} \left[s_{Mm}(x_{n(M,m)} - t_{Mm}) \right] \end{bmatrix} \tag{4}$$

Friedman's modification to estimate the MARS model is written in the following equation:

$$\hat{f}(x) = \alpha_0 + \sum_{m=1}^{M} \alpha_m \prod_{k=1}^{Km} \left[s_{km}(x_{v(k,m)} - t_{km}) \right]_+ \tag{5}$$

with the function,

$$\left(x_{v(k,m)} - t_{km}\right)_+ = \begin{cases} \left(x_{v(k,m)} - t_{km}\right), & x_{v(k,m)} - t_{km} > 0 \\ 0 & , \ x_{v(k,m)} - t_{km} \leq 0 \end{cases} \tag{6}$$

where α_0 is the constant coefficient of the B_0 basis function, and α_m is the coefficient of the m-th basis function, $x_{v(k,m)}$ is the independent variable, t_{km} is the knot value of the independent variable $x_{v(k,m)}$, with M is the number of basis functions, Km is the number of interactions in the m-th basis function. s_{km} is a value that is 1 if the data is to the right of the knot point or -1 if the data is to the left of the knot point, v is the number of predictor variables, and k is the number of interactions. So that Equation (5) can be described as follows.

$$\begin{aligned} \hat{f}(x) \quad = \quad & a_0 + \sum_{m=1}^{M} a_m \left[s_{1m} \cdot (x_{v(1,m)} - t_{1m}) \right]_+ + \sum_{m=1}^{M} a_m \left[s_{1m} \cdot (x_{v(1,m)} - t_{1m}) \right]_+ \\ & \cdot \left[s_{2m} \cdot \left(x_{v(2,m)} - t_{2m}\right) \right]_+ + \sum_{m=1}^{M} a_m \left[s_{1m} \cdot \left(x_{v(1,m)} - t_{1m}\right) \right]_+ \\ & \cdot \left[s_{2m} \cdot \left(x_{v(2,m)} - t_{2m}\right) \right]_+ \cdot \left[s_{3m} \cdot \left(x_{v(3,m)} - t_{3m}\right) \right]_+ + \ldots \end{aligned} \tag{7}$$

In general, Equation (7) can be written as follows:

$$\hat{f}(x) = a_0 + \sum_{i=1}^{v} f_i(x_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^{v} f_i(x_i,\ x_j) + \sum_{\substack{i,j,k=1 \\ i \neq j, i \neq k}}^{v} f_i(x_i,\ x_j,\ x_k) + \ldots \tag{8}$$

Equation (8) shows that the first summation includes all basis functions for one variable. The second summation includes all basis functions for the interaction between two variables. The third summation includes all basis functions for the interaction between three variables, and so on. To simplify the interpretation of the MARS model, the equation in (8) can be rewritten as follows.

$$\hat{f}(x) = a_0 + a_1 BF_1 + a_2 BF_2 + a_3 BF_3 + \ldots + a_M BF_M \tag{9}$$

where $\hat{f}(\mathbf{x})$ is the response variable, a_0 is a constant, a_M is the coefficient for the M-th basis function, and BF_M is the M-th basis function obtained from the calculation of the formula $B_m(\mathbf{x}) = \prod_{k=1}^{Km} \left[s_{km}(x_{v(k,m)} - t_{km}) \right]$, and the basis function (BF) used in the model is obtained via backward stepwise selection.

Classification in the MARS model is based on regression analysis. Classification of the response variable into two values is called regression with a binary response (Cox & Snell, 1989). The probability model used for classification is shown in the following equation:

$$\hat{\pi}(x) = \frac{e^{\hat{f}(x)}}{1 + e^{\hat{f}(x)}}, \; 1 - \hat{\pi}(x) = \frac{1}{1 + e^{\hat{f}(x)}} \tag{10}$$

Where, $\hat{f}(x)$: logit $\hat{\pi}(x)$, the probability of $(Y = 0)$ is $\hat{\pi}(x)$ and the probability of $(Y = 1)$ is $1 - \hat{\pi}(x)$. Y is a binary response variable (with codes 0 and 1 or codes 1 and 2) and m is the number of predictor variables, $x = (x1, \ldots, xm)$. The binary response MARS model for classification can be expressed as follows:

$$\begin{aligned} logit\, \hat{\pi}(x) &= \ln\left(\frac{\hat{\pi}(x)}{1 + \hat{\pi}(x)}\right) \\ &= \alpha_0 + \sum_{m=1}^{M} \alpha_m \prod_{k=1}^{Km} \left[s_{km}(x_{v(k,m)} - t_{km}) \right] \end{aligned} \tag{11}$$

The equation if written in matrix form is as follows:

$$logit\, \hat{\pi}(x) = B\alpha \tag{12}$$

where B found in Equation (4).

The classification of the binary response MARS model is used to assess the accuracy of grouping a set of data into the correct group. According to Holmes & Huber (2019), a good classification method will produce a small classification error or a low probability of misclassification (allocation).

2. Classification Accuracy and Classification Stability Testing

The Total Accuracy Rate (TAR) is used to calculate the classification accuracy of clustering results. The TAR value represents the proportion of samples correctly classified—determination of the accuracy of binary response MARS classification with calculations in Table 2.

Table 2. Binary Response MARS Classification

Observation Result	Prediction	
	y_0	y_1
y_0	n_{00}	n_{01}
y_1	n_{10}	n_{11}

The variable y_0 represents independent fishermen who do not have independent fishermen's insurance, while y_1 represents those who do have independent fishermen's insurance. The symbol n denotes the total number of observations. Furthermore, n_{00} refers to the number of observations of y_0 that are correctly classified as y_0, whereas n_{11} indicates the number of observations of y_1 that are correctly classified as y_1. Meanwhile, n_{01} represents the number of observations of y_0 that are misclassified as y_1, and n_{10} denotes the number of observations of y_1 that are misclassified as y_0.

The Total Accuracy Rate (TAR) value is obtained with the following calculation:

$$TAR\,(\%) = \frac{\text{number of correct prediction}}{\text{total number of prediction}} = \frac{n_{00} + n_{11}}{n_{00} + n_{01} + n_{10} + n_{11}} \times 100\% \tag{13}$$

While the APER value used to indicate the magnitude of the classification error is as follows:

$$APER = \frac{\text{number of incorrect prediction}}{\text{total number of prediction}} = \frac{n_{01} + n_{10}}{n_{00} + n_{01} + n_{10} + n_{11}} \times 100\% \tag{14}$$

In measuring classification, the sensitivity value that describes the accuracy on positive class samples is also considered. Specificity to describe the accuracy on negative class samples. The G-means value is able to describe a classification method through sensitivity and specificity simultaneously. the greater the G-means value indicates that the classification method is able to predict data in each class well and is suitable for unbalanced data. A good classification method should be able to measure sensitivity, specificity and G-means equally well.

$$\text{Sensitivity} = \frac{n_{00}}{n_{00} + n_{01}} \tag{15}$$

$$\text{Specificity} = \frac{n_{11}}{n_{10} + n_{11}} \tag{16}$$

$$\text{G} - \text{means} = \sqrt{\text{sensitifity} - \text{specificity}} \tag{17}$$

To evaluate the consistency and reliability of a classification model, it is necessary to compare its accuracy both across the entire sample and within specific groups. This comparison can be conducted by testing the following hypotheses: the null hypothesis $(\boldsymbol{H_0})$ states that the classification results are **not statistically stable and consistent**, while the alternative hypothesis $(\boldsymbol{H_1})$ asserts that the classification results are **statistically stable and consistent**. To assess these hypotheses, the Press's Q test statistic is employed, which is calculated using the following equation.

$$Press's\ Q\ = \frac{[N - (nK)]^2}{N(K-1)} \tag{18}$$

Where, N represents the total number of samples, n denotes the number of individuals correctly classified, and K refers to the number of groups in the analysis. The decision regarding the stability and consistency of the classification is based on the value of the Press's Q statistic. If the calculated Press's Q value exceeds the critical value, the null hypothesis (H_0) is rejected, indicating that the classification results are **stable and statistically consistent** (González Ariza et al., 2021).

C. RESULT AND DISCUSSION

The analysis and discussion in this research focus on the classification of fishermen's participation in independent fishermen insurance in Jatirejo Village, Lekok District, Pasuruan. Among the 162 fishermen surveyed in Jatirejo Village, 89 respondents (55%) were participants of the insurance program, while 73 respondents (45%) were not. In terms of age, most insured respondents were in the 36-50 and 50+ age groups, suggesting that older fishermen are more aware of the importance of insurance. On the other hand, uninsured respondents were dominated by the 21-35 years age group, showing that younger fishermen were less likely to join the program.

To better understand the characteristics of the respondents, a cross-tabulation was carried out based on several factors that may influence insurance ownership. The results are presented in Table 3.

Table 3. Respondent Characteristics Based on Factors Affecting Insurance Ownership

Variable	Category	Have Insurance		Do not Have Insurance	
		Frequency	Percentage	Frequency	Percentage
	Below Junior High School	79	49%	59	36%
X_2	Junior High School	9	7%	12	6%
(Education Level)	Senior High School	1	1%	2	1%
	Higher Education	0	0%	0	0%
X_7	Smoker	71	44%	53	33%
(Smoking Habits)	Non-Smoker	18	11%	20	12%
X_8	Active	82	51%	28	17%
(Social Participation)	Not Active	7	4%	45	28%
X_9	Own Boat	70	43%	41	25%
(Boat Ownership Status)	Other's Boat	19	12%	32	20%

Table 3 shows that the majority of respondents, both insured and uninsured, had low levels of education, with most completing only up to junior high school. The smoking habit was common among fishermen in both groups. A clear difference appears in social participation: almost all insured fishermen participated in social activities, whereas many uninsured fishermen did not. This indicates that participation in community activities may increase fishermen's awareness and willingness to join the insurance program.

1. Making of the MARS Model

The classification of fishermen insurance ownership in this study applies the Multivariate Adaptive Regression Splines (MARS) method. The software used to perform MARS analysis is RStudio, with the specific package used being the "earth" package. Before modeling, the data is divided into training and test sets. Training data is used to build a model, while test data is used to validate the best model and assess the model's predictive ability. The training and testing data split is 80:20.

Table 4. Comparison of Category Percentage on Training and Testing Data

Data	Categories of Having Insurance	Category Not Having Insurance
Actual Data	54.94%	45.06%
Training Data	54.26%	45.74%
Testing Data	57.58%	42.42%

Table 4 shows that dividing the categories into training and test data already yields equal proportions, making it suitable for training and testing models.

The MARS model in R requires two parameters: nprune and degree. Nprune is the maximum number of bases in the model (including constants), while *degree* is the maximum number of interactions in the model. If the degree is 1, there is no interaction between the variables in the model. If the degree is 2, then there is an interaction between two variables in the model. The nprune value is set to 2–4 times the number of predictor variables. In this study, 10 predictors are used, so the number of nprune combinations is 20, 30, and 40. For the degree, values of 1, 2, 3, and 4 are used, and the set.seed value is set to 50.

Based on the combination of these parameters, 16 possible models are generated, and the model with the highest accuracy is selected. The results of the model combinations are presented in Table 5.

Table 5. MARS Model Combination Results

Model	degree	nprune	Accuracy	Kappa
1	1	20	0.71474359	0.41455907
2*	**2***	**20***	**0.76089744**	**0.51128589**
3	3	20	0.74551282	0.4828521
4	4	20	0.76089744	0.51377719
5	1	30	0.71474359	0.41455907
6	2	30	0.76089744	0.51128589
7	3	30	0.74551282	0.4828521
8	4	30	0.76089744	0.51377719
9	1	40	0.71474359	0.41455907
10	2	40	0.76089744	0.51128589

DOI: https://doi.org/10.30812/varian.v8i3.5373

Model	degree	nprune	Accuracy	Kappa
11	3	40	0.74551282	0.4828521
12	4	40	0.76089744	0.51377719
13	1	50	0.71474359	0.41455907
14	2	50	0.76089744	0.51128589
15	3	50	0.74551282	0.4828521
16	4	50	0.76089744	0.51377719
* Selected model				

Table 5 is the result of the combination of MARS modeling for fishermen insurance ownership with 10 (ten) variables that are thought to affect it. Based on all possible models across the degree and nprune combinations, the best MARS model is the one with the highest accuracy, namely the first model with nprune=20 and degree=2, which achieves 76.09%. This means that the first model correctly classifies 76.09% of the data. Other model goodness-of-fit values are: GCV of 0.1522814; RSS of 15.74785; and R2 of 0.5081181. The following is the best MARS model obtained for classifying fishermen's insurance ownership, as presented in Table 6.

Table 6. Parameter Coefficients of the Best MARS Model

Parameter	Function Basis Code	Parameter Coefficient
Constant		-26.745.388
X81	BF1	33.821.421
h(1.5e+06-X3) * X81	BF2	-0.0000109
h(25-X5) * X81	BF3	0.2850533
h(43-X1) * h(X4-1)	BF4	-0.2257212
h(2.5e+06-X3) * h(X4-1)	BF5	0.0000014

Based on Table 6, the MARS equation model can be written with the following equation:

$$\hat{f}(x) = -2.6745 + 3,3821 * BF_1 - 0.0000109 * BF_2 + 0.2851 * BF_3 - 0.2257 * BF_4 + 0.0000014 * BF_5 \quad (19)$$

In the equation, $(BF_1 = (X_8 = 1))$ indicates that the first basis function equals one when $X_8 = 1$. Next, $(BF_2 = \max(0, 1{,}500{,}000 - X_3) * (X_8 = 1))$ shows that the second basis function has a positive value when X_3 is less than 1,500,000 and $X_8 = 1$. Then, $(BF_3 = \max(0, 0.25 - X_5) * (X_8 = 1))$ means that the third basis function is positive when X_5 is less than 0.25 and $X_8 = 1$. Furthermore, $(BF_4 = \max(0, 0.43 - X_1) * \max(0, X_4 - 1))$ shows that the fourth basis function becomes positive only when X_1 is less than 0.43 and X_4 is greater than 1. Finally, $(BF_5 = \max(0, 2{,}500{,}000 - X_3) * \max(0, X_4 - 1))$ indicates that the fifth basis function is positive when X_3 is less than 2,500,000 and X_4 is greater than 1.

Based on the model obtained, the probability value of fishermen in Lekok sub-district, Pasuruan Regency, having fisherman insurance and the probability of not having fisherman insurance as a whole is as follows:

$$\begin{aligned} \hat{f}(x) &= -2.6745 + 3.3821 * (1) - 0.0000109 * (1) + 0.2851 * (1) - 0.2257 * (1) + 0.0000014 * (1) \\ &= 0.766926 \end{aligned}$$

So that,

$$\hat{\pi}(x) = \frac{e^{(\hat{f}(x))}}{1 + e^{(\hat{f}(x))}} = \frac{e^{(0.766926)}}{1 + e^{(0.766926)}} = \frac{2.15314}{3.15314} = 0.6829$$

Thus, the probability of fishermen having insurance is 0.6829 and the probability of fishermen not having insurance is 0.3171. The interpretation of the MARS model and the probability of insurance ownership for each basis function assuming the other basis functions are constant in turn is as follows:

1. $BF_1 = (X_8 = 1)$

$$\hat{\pi}(x) = \frac{e^{(-2.6745 - 0.0000109 * (1))}}{1 + e^{(-2.6745 - 0.0000109 * (1))}} = 0.0645$$

DOI: https://doi.org/10.30812/varian.v8i3.5373

This means that the coefficient of BF_2 will mean that if the fisherman's income (X3) is less than 1,500,000 and actively participates in socialization, the chance of the fisherman having insurance is 6.45%.

2. $BF_3 = \max(0,\ 25 - X_5) * (X_8 = 1)$

$$\hat{\pi}(x) = \frac{e^{(-2.6745\ +0.2851*(1))}}{1 + e^{(-2.6745\ +0.2851*(1))}} = 0.0839$$

This means that fishermen with less than 25 years of fishing experience (X5) and who actively participate in socialization have a 8.39% chance of having insurance.

3. $BF_4 = \max(0,\ 43 - X_1) * \max(0,\ X_4 - 1)$

$$\hat{\pi}(x) = \frac{e^{(-2.6745\ -0.2257*(1))}}{1 + e^{(-2.6745\ -0.2257*(1))}} = 0.0521$$

This means that the coefficient of BF_4 will mean that if Age (X1) is less than 43 years old and the number of family dependents (X4) is less than 1, then the probability that a fisherman has insurance is 5.21%.

4. $BF_5 = \max(0,\ 2,500,000 - X_3) * \max(0, X_4 - 1)$

$$\hat{\pi}(x) = \frac{e^{(-2.6745\ +0.0000014*(1))}}{1 + e^{(-2.6745\ +0.0000014*(1))}} = 0.0645$$

This means that the coefficient of BF_5 will mean that if the Income (X3) is less than 2,500,000 and the Number of Family Dependents (X4) of fishermen is less than 1, then the probability that fishermen have insurance is 6.45%.

Furthermore, from the best MARS model derived from the equation above, it can be seen that five predictor variables enter the model: Age (X1), Income (X3), Number of Family Dependents (X4), Fishing Experience (X5), and Socialization Participation (X8). To see the extent to which these variables affect the formation of the MARS model, refer to the following variable importance graph.

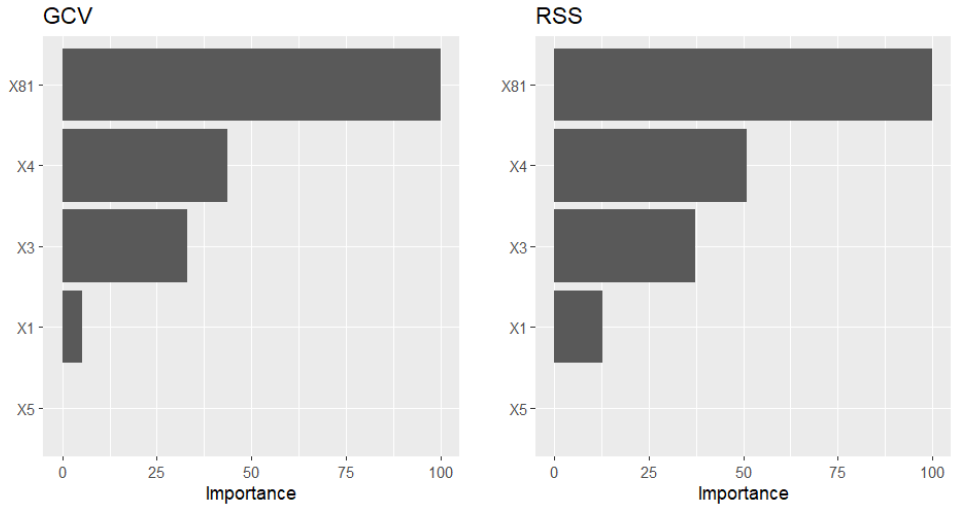

Figure 2. Variable Importance Graph

Figure 2 shows that the variable of socialization participation, X8(1), is the most important in the MARS model, with a level of importance of 100%, followed by the variable number of family dependents (X4), with a level of importance of 43.799%. The income variable (X3) has an importance level of 33.029%, and the age variable (X1) has an importance level of 5.32%. The variable for fishing experience (X5) is not significant (0%) because the previous four variables capture it.

The results of this study show that participation in socialization activities is the most influential factor in fishermen's participation in insurance. This finding is consistent with Brahmantyo et al. (2021), who emphasized that education and outreach significantly affect fishermen's willingness to join insurance schemes, thus strengthening the importance of socialization in this

study. The role of socio-economic factors, such as income and the number of dependents, identified here also aligns with earlier work, such as Al-Musaylh et al. (2018) and Hasyim et al. (2018), who found that income and household income responsibilities often serve as strong predictors in participation-related decisions. However, unlike some previous studies that highlighted work experience as a determinant, this study shows that fishing experience does not significantly affect participation, indicating that social and economic variables play a more dominant role.

2. MARS Classification Accuracy

Classification accuracy is the ratio of correctly classified observations to the total number of observations. To assess the model's classification performance, we need to examine classification accuracy metrics based on the 1-APER value, sensitivity, and specificity. The classification accuracy results on the training data are shown in Table 7.

Table 7. Classification of Insurance Ownership on training data

Actual Data	Prediction Data		Total of Actual
	0	1	
0	43	7	50
1	16	63	79
Total of Prediction	59	70	129

Table 7 shows that of 129 individuals, 43 were correctly classified as not having marine insurance, and 7 were incorrectly classified from not having insurance to having insurance. Meanwhile, 63 individuals are correctly classified as having insurance, and 16 are misclassified from having insurance to not having insurance. Based on the classification table, the 1-APER, sensitivity, and specificity values can be calculated, and are shown in Table 8.

Table 8. Model Goodness Value on Training Data

Criteria	Value
1-APER	82%
Sensitivity	90%
Specificity	73%

Table 8 shows that the classification accuracy generated from the training data is 82% and the classification error is 18%. Based on the sensitivity value, 90% of fishermen are correctly classified as having insurance, while, based on the specificity value, 73% are correctly classified as not having insurance. The large and balanced values of 1-APER, sensitivity, and specificity indicate that the MARS method is suitable for training data in classifying the incidence of fishermen insurance ownership. Furthermore, it can be continued to classify the testing data, and the classification accuracy results are shown in Table 9.

Table 9. Classification of Insurance Ownership on Testing Data

Actual Data	Prediction Data		Total of Actual
	0	1	
0	11	3	14
1	5	14	19
Total of Prediction	16	17	33

Table 9 shows that in the test data, 14 individuals are correctly classified as having marine insurance, and 5 are incorrectly classified from having insurance to not having insurance. Meanwhile, 11 individuals are correctly classified as not having insurance, and 3 are misclassified from not having insurance to having insurance. Based on the classification table, the values for 1-APER, sensitivity, and specificity can be calculated and are shown in Table 10.

Table 10. Model Goodness Value on Training Data

Criteria	Value
1-APER	75.8%
Sensitivity	89.6%
Specificity	66.7%

Table 10 shows that the MARS model trained on the test data achieves a classification accuracy of 75.8%. This indicates that the best MARS model is good enough for predicting whether a fisherman has marine insurance.

3. Classification by MARS Bagging Method

The MARS Bagging method is carried out with bootstrap replication on training data and testing data as many as 50, 100, 150, 200, and 500 times, so that the prediction results are obtained on the response variable and can be classified to determine the accuracy of the classification of fishermen's insurance ownership using the MARS Bagging method. The following are the classification accuracies on training and test data for various replication combinations using the <Rweka> Package in Software R.

Table 11. MARS Bagging Results on Training Data

Method	Combination of Repetition	1-APER
MARS	-	82%
Boosting MARS	50	83.72%
	100	84.49%
	250	89.15%
	500	93.02%
	600	94.57%
	700	96.12%
	800	96.89%
	900	98.45%
	1000	98.45%

The results in Table 11 show that, with 50-1000 replications, the MARS Bagging method can increase classification accuracy from 82% to 98.45%, and stability begins at the 900th replication. If the replication value is repeated several times, the results for accuracy and classification error remain stable. So, it can be said that the MARS bagging method increases MARS's classification accuracy.

Table 12. MARS Bagging Results on Testing Data

Method	Combination of Repetition	1-APER
MARS	-	75.8%
Boosting MARS	50	100%
	100	100%
	250	100%
	500	100%
	600	100%
Boosting MARS	700	100%
	800	100%
	900	100%
	1000	100%

Furthermore, the accuracy of the Bagging MARS classification on test data is shown in Table 12, indicating that applying Bagging MARS in this research improved classification accuracy from 75.8% to 100%. This result is in line with Hasyim et al. (2018) and Rupilu & Rosadi (2024), who demonstrated that Bagging enhances both the accuracy and stability of MARS models across different domains, thereby reinforcing the methodological robustness of this study.

In the MARS model, a parameter represents the true coefficient associated with each basis function in the underlying population. In contrast, a parameter estimate is the value calculated from the sample data to approximate the true coefficient. In the Bagging MARS method, the model is built repeatedly on multiple bootstrap samples, and each replication generates different knots and basis functions. Consequently, the parameter estimates vary across replications, and it is not possible to compute a single average parameter estimate for the final Bagging model. Therefore, the Bagging MARS model does not produce explicit parameter estimates; instead, it serves as an ensemble method to reduce the classification error of the original MARS model. Accordingly, the classification of fishermen's insurance ownership in Lekok Village, Pasuruan District, is based on the MARS model with parameters degree = 1 and nprune = 20. At the same time, Bagging is used solely to improve prediction accuracy.

D. CONCLUSION AND SUGGESTION

This study found that the MARS method, especially when combined with Bagging, can accurately classify and identify key factors influencing fishermen's participation in independent insurance. The five most important factors are participation in socialization, number of family dependents, income, age, and fishing experience. The original MARS model achieved 82% accuracy on the training data and 75.8% on the test data. After applying Bagging, the accuracy increased to 98.45% and 100%, respectively. The most influential factor is participation in socialization activities.

The basis functions (BFs) and knots in the MARS model revealed nonlinear relationships between these predictors and insurance participations. For example, the knot on social participation indicates that insurance ownership increases significantly after a certain level of activity. In contrast, the knots on income and family dependents indicate that these variables exceed a threshold. These turning points highlight behavioral thresholds in fishermen's decision-making.

Hence, the policy recommendations are directly linked to these findings. Since social participation shows a strong threshold effect, the government should focus on increasing fishermen's involvement in socialization and education programs to raise awareness about insurance benefits. The link between income and family dependents suggests that financial support or premium subsidies for low-income fishermen and family-based insurance schemes could encourage higher participation. Future research should refine the model by analyzing interaction effects between knots and by including behavioral variables, such as trust in insurance and risk perception, to improve interpretability.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Department of Actuarial Science, Institut Teknologi Sepuluh Nopember (ITS), and Universiti Malaysia Pahang Al-Sultan Abdullah for their support and collaboration throughout this research. We also thank the Fisheries Service of Pasuruan Regency and all the fishermen in Jatirejo Village, Lekok District, who participated in the survey and provided valuable information—special thanks to the supervisors and colleagues who contributed ideas and guidance during the preparation of this paper.

DECLARATIONS

AUTHOR CONTIBUTION

First Author: Conceptualization, supervision, funding acquisition, and provided key resources. Second Author: Conceptualization, Methodology, Data analysis, Software, Validation, and Writing of the original draft. Third Author: Data curation and analysis and contributed to validation and manuscript review. Fourth Author: Draft preparation, Writing of the original draft, and software development. Fifth Author: Validation, Visualization, and Writing – review and editing. Sixth author: Literature review process, assisted in data validation, and contributed to manuscript refinement. All authors discussed the results and contributed to the final manuscript.

FUNDING STATEMENT

This work is funded by the Directorate of Research and Community Service, Institut Teknologi Sepuluh Nopember (ITS), Surabaya for the programme of Scientific Research in ITS.

COMPETING INTEREST

The authors declare that they have no competing interests in this article.

REFERENCES

Al-Musaylh, M. S., Deo, R. C., Adamowski, J. F., & Li, Y. (2018). Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. *Advanced Engineering Informatics*, *35*, 1–16. https://doi.org/10.1016/j.aei.2017.11.002

Amin, M. M., Zainal, A., Azmi, N. F. M., & Ali, N. A. (2020). Feature Selection Using Multivariate Adaptive Regression Splines in Telecommunication Fraud Detection. *IOP Conference Series: Materials Science and Engineering*, *864*(1), 012059. https://doi.org/10.1088/1757-899X/864/1/012059

Asyya, M. F., & Agusta, I. (2021). Analisis Partisipasi Nelayan dalam Program Asuransi Nelayan. *Jurnal Sains Komunikasi dan Pengembangan Masyarakat [JSKPM]*, *5*(2), 294–311. https://doi.org/10.29244/jskpm.v5i2.818

Bose, A., Hsu, C.-H., Roy, S. S., Lee, K. C., Mohammadi-ivatloo, B., & Abimannan, S. (2021). Forecasting stock price by hybrid model of cascading Multivariate Adaptive Regression Splines and Deep Neural Network. *Computers and Electrical Engineering*, *95*, 107405. https://doi.org/10.1016/j.compeleceng.2021.107405

Brahmantyo, Y., Riaman, R., & Sukono, F. (2021). Willingness to Pay of Fishermen Insurance Using Logistic Regression with Parameter Estimated by Maximum Likelihood Estimation Based on Newton Raphson Iteration. *Jurnal Matematika Integratif*, *17*(1), 15. https://doi.org/10.24198/jmi.v17.n1.32037.15-21

Cox, D. R., & Snell, E. J. (1989, May 15). *Analysis of Binary Data, Second Edition*. CRC Press.

De Andrés, J., Lorca, P., De Cos Juez, F. J., & Sánchez-Lasheras, F. (2011). Bankruptcy forecasting: A hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS). *Expert Systems with Applications*, *38*(3), 1866–1875. https://doi.org/10.1016/j.eswa.2010.07.117

Eubank, R. L. (1999, February 9). *Nonparametric Regression and Spline Smoothing* (0th ed.). CRC Press. https://doi.org/10.1201/9781482273144

Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, *19*(1), 1–67. https://doi.org/10.1214/aos/1176347963

González Ariza, A., Arando Arbulu, A., Navas González, F. J., Delgado Bermejo, J. V., & Camacho Vallejo, M. E. (2021). Discriminant Canonical Analysis as a Validation Tool for Multivariety Native Breed Egg Commercial Quality Classification. *Foods*, *10*(3), 632. https://doi.org/10.3390/foods10030632

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer. https://doi.org/10.1007/978-0-387-84858-7

Hasyim, M., Rahayu, D. S., Muliawati, N. E., Hayuhantika, D., Puspasari, R., Anggreini, D., Hastari, R. C., Hartanto, S., & Utomo, F. H. (2018). Bootstrap Aggregating Multivariate Adaptive Regression Splines (Bagging MARS) to Analyse the Lecturer Research Performance in Private University. *Journal of Physics: Conference Series*, *1114*, 012117. https://doi.org/10.1088/1742-6596/1114/1/012117

Hlokoe, V. R., Mokoena, K., & Tyasi, T. L. (2022). Using multivariate adaptive regression splines and classification and regression tree data mining algorithms to predict body weight of Nguni cows. *Journal of Applied Animal Research*, *50*(1), 534–539. https://doi.org/10.1080/09712119.2022.2110498

Holmes, S., & Huber, W. (2019). *Modern statistics for modern biology*. Cambridge university press.

Otok, B. W., Rumiati, A. T., Ampulembang, A. P., & Azies, H. A. (2023). ANOVA Decomposition and Importance Variable Process in Multivariate Adaptive Regression Spline Model. *International Journal on Advanced Science, Engineering and Information Technology*, *13*(3), 928–934. https://doi.org/10.18517/ijaseit.13.3.17674

Rupilu, R. A. H. W., & Rosadi, D. (2024). Classification Analysis Using Bootstrap Aggregating Multivariate Adaptive Regression Spline (Bagging MARS). *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, *18*(3), 1381–1390. https://doi.org/10.30598/barekengvol18iss3pp1381-1390

Seno, M. E., Zeini, H. A., Imran, H., Noori, M., Henedy, S. N., & Ghazaly, N. M. (2024). Advancing in creep index of soil prediction: A groundbreaking machine learning approach with Multivariate Adaptive Regression Splines. *Results in Materials*, *24*, 100641. https://doi.org/10.1016/j.rinma.2024.100641

[This page intentionally left blank.]

Vol. 8, No. 3, October 2025, pp 319–332
DOI: https://doi.org/10.30812/varian.v8i3.5373