JURNAL STATISTIKA DAN APLIKASINYA

e-ISSN 2620 – 8369
VOLUME 9 (ISSUE 1) JUNE 2025: 112 – 125
https://doi.org/10.21009/JSA.09110

IDENTIFICATION OF LOCATION ALLOWANCE ZONE FOR BANK SYARIAH "X" OUTLETS USING ORDINAL LOGISTIC REGRESSION

Aldinda Albanna[1*], Yekti Widyaningsih[2]

[1,2]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Indonesia*
Jl. Lingkar Pondok Cina, Depok, 16424, Indonesia

Corresponding author's e-mail: ***aldindaalbanna28@gmail.com***

ABSTRACT

Article History:
Received: June 2, 2025
Revised: June 26, 2025
Accepted: June 29, 2025
Published: June 30, 2025
Available online.

Keywords:
Cost; Location Access; Remoteness.

Companies need good human resources to achieve their goals, one of which is by providing rewards, such as location allowances. Bank Syariah "X" is one of the institutions that provides location allowance, which is an allowance based on the employee's work location. This policy was last established in 2021, therefore adjustments are needed. This study aims to analyze the factors that explain the determination of location allowance zoning and predict the zoning of new outlet location allowances. Location allowance zoning is determined based on the factors of cost, remoteness, and location access. Factors that are thought to represent these three factors and influence the determination of location allowance zoning are the consumer price index (CPI), human development index (HDI), construction cost index (CCI), infrastructure pillar index (IPI), outlet distance to the nearest health center (ODHC), and outlet distance to the nearest primary school (ODPS). The location allowance zones consist of three categories with an ordered nature. Based on the research objectives and the type of dependent variable, the method used was ordinal logistic regression. The research identified three key factors influencing the determination of location allowance zoning, namely CCI, IPI, and ODHC. The model achieved an overall accuracy of 70%, with balanced accuracy scores of 81.2% for Zone 1, 70.8% for Zone 2, and 76.7% for Zones 3 & 4. Based on the initial policy data of Bank Syariah "X", the model misclassified 35.6% of outlets.

How to cite this article:
A. Albanna, Y. Widyaningsih, "IDENTIFICATION OF LOCATION ALLOWANCE ZONE FOR BANK SYARIAH "X" OUTLETS USING ORDINAL LOGISTIC REGRESSION", Jurnal Statistika dan Aplikasinya, vol. 9, iss. 1, pp. 112 – 125, June 2025

Journal e-mail: *jsa@unj.ac.id*
Research Article · Open Access

1. INTRODUCTION

Companies require good human resources to achieve their vision, mission, and goals. Good human resources will enhance the company's chances of performing well, as company performance is influenced by employee performance. One of the factors that influences employee performance is the reward system [1]. The goals of the reward system are to appreciate the employee, increase the employee's loyalty, and reduce the turnover rates [2]. Bank Syariah "X" is one of the companies that implements a reward system for its employees. One of the rewards provided by Bank Syariah "X" is a location allowance, an allowance the company gives to employees based on where the employee works. The location allowance is determined based on factors of cost, remoteness, location access, and other things decided by the board of directors, which result in differences in the cost of living and quality of life.

Bank Syariah "X" implements a three-zone system to determine location allowances. Zone 1 does not receive the location allowance, while Zone 2 and Zone 3 & 4 receive location allowances with different nominal amounts. The higher the zone category, the greater the nominal allowance given. The zoning policy for location allowances at Bank Syariah "X" was last established in 2021. However, changes in conditions may occur every year and cause conditions in each region to change. Therefore, it is necessary to update and adjust the policy to ensure the location allowance remains appropriate with the current circumstances in each outlet's area. In this study, the factors that influenced the zoning of location allowances at Bank Syariah "X" will be determined. The determination of these factors is chosen based on the criteria set by the company, namely the factors of cost, remoteness, and location access. The factors that are thought to influence the determination of location allowance zoning at Bank Syariah "X" and can represent the factors of cost, remoteness, and location access are the consumer price index (CPI), human development index (HDI), construction cost index (CCI), infrastructure pillar index (IPI), outlet distance to the nearest health center (ODHC), and outlet distance to the nearest primary school (ODPS).

The ordinal logistic regression method is widely used in research to determine the factors that can influence the dependent variables with more than two categories with ordered properties. Additionally, this method produced the probability of occurrence for each category, which could be used to predict new observations. The ordinal logistic regression method was used by Döringer et al. [3] to determine the relationship between the independent variables (number of visitors, the number of visitors passing each other during the visit, and estimate of the crowd level) and the dependent variable, namely the crowd level (not crowded, slightly crowded, crowded, very crowded, extremely crowded) in the Bavarian Forest National Park. The study found that the factors influencing the crowd level were the number of visitors, the number of visitors passing each other during the visit, and the estimate of the crowd level. Therefore, the final model was built using these three factors and resulted in an accuracy of 70.6%. The ordinal logistic regression method was also used by Zalaghaie et al. [4] to determine the relationship between the independent variables (grain size, porosity, moisture content, rainfall, climatic data, liquid limit, calcium carbonate content, vegetation, and soil saturation level) and the dependent variable, which is the collapse sensitivity of soil deposits consisting of four levels (slight, moderate, less severe, and severe). The study found that the factors influencing the sensitivity of soil sediment collapse were porosity, moisture content, rainfall, climate data, and liquid limit. Therefore, the final model was formed using these five factors, which resulted in an accuracy of 70%. Research conducted by Addini et al. [5] used the ordinal logistic regression method to determine the relationship between the independent variables (vehicle type, gender, education level, occupation, religion, and age) and the dependent variable, which is material loss (light, medium, and heavy). The study found that the factors influencing material loss were vehicle type, gender, education level, occupation, religion, and age. Therefore, the final model was formed using these six factors and resulted in an accuracy of 76.27%.

The studies conducted by Döringer et al. [3], Zalaghaie et al. [4], and Addini et al. [5] showed that ordinal logistic regression used to determine the factors affecting dependent variables of ordinal type and predict new observations based on the final model. The ordinal logistic regression model will be formed to determine what factors influence the zoning of location allowances at Bank Syariah "X" and to make predictions for new outlets where zoning categories have not been defined. The model is evaluated based on test data and the initial policy data of Bank Syariah "X". The ordinal logistic regression method was used because this method can model the relationship between the independent variable and the dependent variable (location allowance zoning), which consists of three levels with an ordered nature, namely Zone 1, Zone 2, and Zone 3 & 4.

2. METHODS

Material and Data

This study used data on the outlets of Bank Syariah "X" in Indonesia in 2023. The variables used in the study were shown in Table 1.

Table 1. Research Data

Variable	Variable Name	Type	Source
Y	Location allowance zoning	Ordinal (1, 2, 3 & 4)	Bank Syariah "X"
X_1	Consumer price index (CPI)	Ratio	Badan Pusat Statistik (BPS)
X_2	Human development index (HDI)		
X_3	Construction cost index (CCI)		
X_4	Infrastructure pillar index (IPI)		Badan Riset dan Inovasi Nasional (BRIN)
X_5	Outlet distance to the nearest health center (ODHC)		Bank Syariah "X" & Google Maps
X_6	Outlet distance to the nearest primary school (ODPS)		

Research Method

Ordinal Logistic Regression

The ordinal logistic regression method is a method to see the relationship between the independent variable and the categorical dependent variable, which has an ordered nature with the categories of the dependent variable numbering more than two. The probability of the independent variable Y falling into category k or $Y = k$ is as follows [6]:

$$P(Y = k|\boldsymbol{x}) = \pi_k(\boldsymbol{x}) = \frac{\exp\left(\beta_{k0} + \sum_{i=1}^{p} \beta_{ki} x_i\right)}{1 + \exp\left(\beta_{k0} + \sum_{i=1}^{p} \beta_{ki} x_i\right)} \tag{1}$$

where $k = 1,2, \dots, K-1$ and $1 - \pi_k(\boldsymbol{x})$ expressing the probability of Y not belonging to the kth category. In logistic regression, the logit model formation uses the cumulative probability denoted by $\gamma_k(\boldsymbol{x})$, which is as follows [7]:

$$\gamma_k(\boldsymbol{x}) = P(Y \leq k|\boldsymbol{x}) = \pi_1(\boldsymbol{x}) + \pi_2(\boldsymbol{x}) + \cdots + \pi_k(\boldsymbol{x});\ k = 1,2, \dots, K \tag{2}$$

The logit model of cumulative probability (cumulative logit model) is defined as follows [7]:

$$logit[P(Y \leq k|\boldsymbol{x})] = \beta_{k0} + \sum_{i=1}^{p} \beta_{ki} x_i\,; k = 1, \dots, K-1; i = 1, \dots, p \tag{3}$$

In logistic regression, there are three types of models commonly used, namely the adjacent-category, the continuation-ratio, and the proportional odds models. In this study, the proportional odds model will be used because this model is the most common and frequently used in the ordinal logistic regression model. The following is the logit form of the model [6]:

$$c_k(\boldsymbol{x}) = \ln\left[\frac{P(Y \le k|\boldsymbol{x})}{P(Y > k|\boldsymbol{x})}\right]$$
$$= \ln\left[\frac{\pi_0(\boldsymbol{x}) + \pi_1(\boldsymbol{x}) + \cdots + \pi_k(\boldsymbol{x})}{\pi_{k+1}(\boldsymbol{x}) + \pi_{k+2}(\boldsymbol{x}) + \cdots + \pi_K(\boldsymbol{x})}\right]$$
$$= \theta_k - \boldsymbol{x}'\boldsymbol{\beta}; k = 1, \dots K-1 \tag{4}$$

Parameter Estimation

The parameters θ_k and $\boldsymbol{\beta}$ are estimated using the maximum likelihood method [8]. Suppose for each ith observation with $i = 1,2,\dots,n$, $y_{ij} = 1$, when $j = 1,2,\dots,K$ and $y_{ij} = 0$ otherwise. The general form of the likelihood for a sample of n mutually independent observations, $(y_i, \boldsymbol{x}_i)$; $i = 1,2,\dots,n$ is as follows [6], [7]:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n}\left[\prod_{j=1}^{K}[\pi_j(\boldsymbol{x}_i)]^{y_{ij}}\right]$$
$$= \prod_{i=1}^{n}\left[\prod_{j=1}^{K}\left[\frac{exp(\theta_j + \boldsymbol{x}_i{}'\boldsymbol{\beta})}{1 + exp(\theta_j + \boldsymbol{x}_i{}'\boldsymbol{\beta})} - \frac{exp(\theta_{j-1} + \boldsymbol{x}_i{}'\boldsymbol{\beta})}{1 + exp(\theta_{j-1} + \boldsymbol{x}_i{}'\boldsymbol{\beta})}\right]^{y_{ij}}\right] \tag{5}$$

where "$\boldsymbol{\beta}$" denotes the coefficient for the pth independent variable (β) and the intercept for each category in the model (θ_k). The form of the log-likelihood equation is as follows:

$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{j=1}^{K} y_{ij} \ln\left[\frac{\exp(\theta_{\mathrm{j}} + \boldsymbol{x}_i{}'\boldsymbol{\beta})}{1 + \exp(\theta_{\mathrm{j}} + \boldsymbol{x}_i{}'\boldsymbol{\beta})} - \frac{\exp(\theta_{\mathrm{j}-1} + \boldsymbol{x}_i{}'\boldsymbol{\beta})}{1 + \exp(\theta_{\mathrm{j}-1} + \boldsymbol{x}_i{}'\boldsymbol{\beta})}\right] \tag{6}$$

Parameter estimates are obtained by deriving the log-likelihood equation for each parameter to be estimated and then equating the derivative to zero.

Multicollinearity Detection

Logistic regression models are sensitive to multicollinearity between independent variables in the model [6]. Variance inflation factors (VIF) value is used to detect multicollinearity with the following formula [9]:

$$VIF = \frac{1}{1 - R_p^2} \tag{7}$$

where R_p^2 is the coefficient of determination of the regression model with x_p as the dependent variable and other independent variables as independent variables. If the VIF value is more than 10, then there is multicollinearity between independent variables.

Assumption of Proportional Odds

There is an assumption of the proportional odds model that must be held, where the parameters do not change for different categories [10], [11]. In the proportional odds model assumption test, the following hypothesis is used [7]:

H_0: $\beta_{1p} = \beta_{2p} = \dots = \beta_{kp} = \beta$ (The coefficient β_k of the pth independent variable is the same for each model in different categories)

H_1: Not so (The coefficient β_k of the pth independent variable is not the same for each model in different categories)

for each $p = 1,2,\dots,P$ where P denotes the number of independent variables and $k = 1,2,\dots,K-1$ where K denotes the number of categories.

The test statistics used is the likelihood ratio test, which is as follows [6]:

$$LR = -2(L_0 - L_1) \tag{8}$$

where L_0: Log-likelihood of the model with the same independent variable coefficients for each model in different categories; L_1: Log-likelihood of the model with independent variable coefficients not equal for each model in different categories.

At a significance level of α, the null hypothesis (H_0) is rejected if the likelihood ratio (LR) exceeds the chi-square critical value $\chi^2_{\alpha;(K-2)p}$ or if the $p-value$ is less than α. If H_0 is rejected, the proportional odds assumption doesn't hold [6], [7].

Lipsitz Test for Goodness of Fit

The method used to test model fit is the Lipsitz method with the following hypothesis [7]:
H_0: The ordinal logistic regression model fits the data.
H_1: The ordinal logistic regression model does not fit the data.
The test statistics are as follows [12]:

$$LR = -2(L_0 - L_1) \tag{9}$$

where L_0: Log-likelihood of the model without indicator variables; L_1: Log-likelihood of the model with indicator variables.

At significance α, H_0 is rejected if $LR > \chi^2_{\alpha;(g-1))}$ with g being the number of groups. If H_0 is rejected, the ordinal logistic regression model formed does not fit the data.

Parameter Significant Test

The parameter significance test is carried out to test whether the independent variables in the model are significantly associated with the dependent variable [6]. The parameter significance test consists of a simultaneous test and a partial test.

The simultaneous parameter significance test is carried out using the likelihood ratio test method with the following hypothesis [6]:

H_0: $\beta_1 = \beta_2 = \ldots = \beta_p = 0$ (There is no significant effect of the independent variable on the dependent variable)

H_1: At least one $\beta_p \neq 0$ (There is at least one significant independent variable on the dependent variable)

The test statistics used are as follows:

$$G = -2(L_0 - L_1) \tag{10}$$

where L_0: Log-likelihood for a model without independent variables (only constants); L_1: Log-likelihood for the model with independent variables.

At significance α, H_0 is rejected if $G > \chi^2_{\alpha;p}$ with p denoting the number of parameters. If H_0 is rejected, at least one independent variable is significant to the dependent variable [6].

A partial parameter significance test was conducted using the Wald test method with the following hypothesis [6]:

H_0: $\beta_p = 0; p = 1,2,\ldots,P$ (The pth independent variable has no effect on the dependent variable)
H_1: Not so (The pth independent variable has an influence on the dependent variable)
The test statistics used are as follows:

$$W = \frac{\hat{\beta}_p}{\widehat{SE}(\hat{\beta}_p)} \tag{11}$$

where $\hat{\beta}_p$: Estimated coefficient of the pth independent variable; $\widehat{SE}(\hat{\beta}_p)$: Standard error of the estimated coefficient $\hat{\beta}_p$.

At α significance, H_0 is rejected if $W^2 > \chi^2_{\alpha;p}$ where p denotes the number of independent variables used. If H_0 is rejected, the independent variable is significant or influences the dependent variable [6].

Model Evaluation

The results of the ordinal logistic regression model must be evaluated. One of the methods that can be used to evaluate the model is the confusion matrix. The confusion matrix displays the performance of the modeling results. From the confusion matrix, accuracy, sensitivity, specificity, and balanced accuracy can be obtained. The display of the confusion matrix is shown in Table 2 [13].

Table 2. Confusion Matrix

Actual	Predicted		
	Class 1	Class 2	Class 3
Class 1	f_{11}	f_{12}	f_{13}
Class 2	f_{21}	f_{22}	f_{23}
Class 3	f_{31}	f_{32}	f_{33}

where f_{ij} for $i = j$ is the number of observations of the-i class that were correctly classified as the-i class and f_{ij} for $i \neq j$ is the number of observations of the-i class that are incorrectly classified as the j class.

Table 2 shows the comparison between actual data and predicted data. The confusion matrix consists of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). True positive (TP) is the number of positive classes predicted to be true positive. True negative (TN) is the number of negative classes predicted to be true negative. False positive (FP) is the number of negative classes predicted as positive. False negative (FN) is the number of positive classes predicted to be negative [14].

In this study, the model was evaluated using accuracy, sensitivity, specificity, and balanced accuracy with the following formula [15]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (13)$$

$$Specificity = \frac{TN}{TN + FP} \quad (14)$$

$$Balanced\ Accuracy = \frac{sensitivity + specificity}{2} \quad (15)$$

Accuracy is the proportion of correctly predicted observations, sensitivity is the proportion of correctly predicted observations as a positive class, specificity is the proportion of correctly predicted observations as a negative class, and balanced accuracy is the average of sensitivity and specificity.

Model evaluation is essential to assess classification performance. However, for cases with an unbalanced class distribution, it is necessary to add other measurement values, namely sensitivity and specificity [16]. This is because minority classes contribute less to the accuracy calculation than majority classes. The sensitivity value indicates the proportion of observations that are correctly predicted as positive classes, and the specificity measures the proportion of observations that are correctly predicted as negative classes. From the sensitivity and specificity values, a balanced accuracy value can be obtained, which is a value that shows how well the model correctly predicts positive and negative classes [15].

3. RESULTS

Descriptive Statistical Analysis

This section discusses descriptive statistics from observations totaling 225 outlets for each variable used in this study. Table 3 and Figure 1 show the distribution of outlets based on the location allowance zoning category.

Table 3. Percentage and Number of Outlets by Zoning Category

Zoning Category	Number of Outlets	Percentage	Description
Zone 1	106	47.1%	No Allowance
Zone 2	73	32.4%	Allowance of a certain amount (Allowance A)
Zone 3 & 4	46	20.4%	Allowance greater than allowance A (Allowance B)

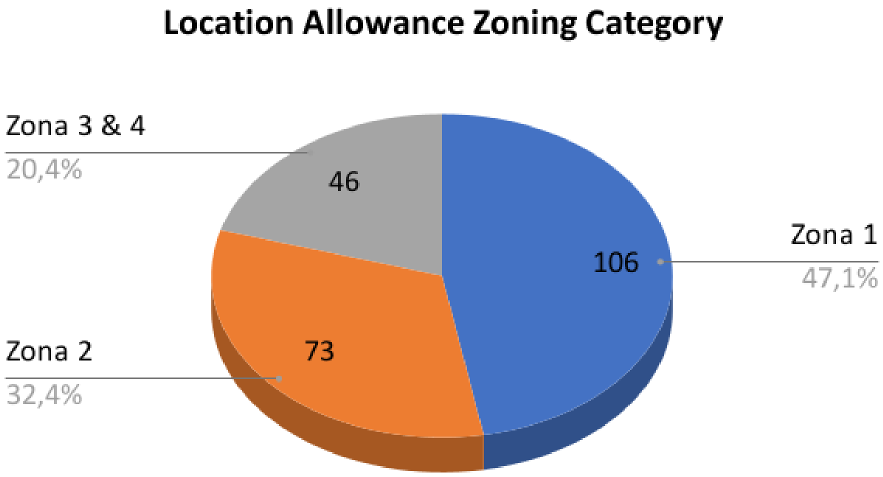

Figure 1. Location Allowance Zoning Category Diagram

Six variables are thought to influence the determination of location allowance zoning at Bank Syariah "X", namely the consumer price index (CPI), human development index (HDI), construction cost index (CCI), infrastructure pillar index (IPI), outlet distance to the nearest health center (ODHC), and outlet distance to the nearest primary school (ODPS). Each outlet in the same district or city is in the same location allowance zoning category. Therefore, the observation unit for the CPI, HDI, CCI, and IPI variables is for each district/city. Meanwhile, the observation unit for ODHC and ODPS data is the outlet location because it calculates the distance from the outlet to the nearest health center and primary school. Table 4 presents the descriptive statistics of each independent variable.

Table 4. Descriptive Statistics of Independent Variables

	CPI	HDI	CCI	IPI	ODHC	ODPS
Minimum	111.8	64.13	83.86	1.330	57.45	37.61
First Quartile	115.2	69.97	92.95	1.920	684.59	194.03
Median	116.1	71.75	97.60	2.260	1077.89	376.87
Mean	116.0	73.10	98.45	2.316	1206.40	404.74
Third Quartile	117.1	75.70	102.63	2.640	1528.87	545.81
Maximum	123.0	86.69	133.93	4.210	4114.07	1943.29

Data Preprocessing

Data preprocessing must be done before processing and modeling with ordinal logistic regression to ensure the data is ready to be processed. In this research, data preprocessing involves outlier detection, data scaling, and data splitting. In this study, one of the data scaling methods used is the Min Max method. This method is sensitive to outliers [17]. Therefore, outliers will be removed for the data used for scaling with the Min Max method.

Table 5. Comparison of Number of Outlets Before and After Outliers are Removed

Zoning Category	Number of Outlets before Outliers Removed	Number of Outlets after Outliers Removed
Zone 1	106	92
Zone 2	73	64
Zone 3 & 4	46	42

Based on Table 5, there is a reduction in the number of outlets in Zone 1, Zone 2, and Zone 3 & 4. The number of outlets in Zone 1 is reduced by 14 outlets, the number of outlets in Zone 2 is reduced by 9 outlets, and the number of outlets in Zone 3 & 4 is reduced by 4 outlets.

The distribution of data in this study has different units and quantities. Therefore, in this study, data scaling was carried out using the Min Max and Robust Scaler methods. Data scaling is done so that the data has uniform units and ranges to produce a good model. In addition, data scaling is done to avoid bias in the data processing process [18], [19].

Scaling data with the Min Max method is done by converting observed values into values ranging from 0 to 1. This method is sensitive to outliers [17]. Therefore, the data used is data without outliers, which is 198 observations. The second data scaling method is the Robust Scaler. This method is resistant to outliers [19]. Therefore, Robust Scaler uses the data before the outliers are removed, which is 225 observations.

This research divides the data into two parts, namely training data and testing data. The determination of the data split percentage is based on testing seven different data split percentages and two different data scaling methods. Table 6 and Table 7 show the comparison of balanced accuracy at each percentage of data division.

Table 6. Balanced Accuracy from Data Scaling Using the Min Max Method

Data Split	Training Data	Testing Data	Balanced Accuracy		
			1	2	3 & 4
60:40	118	80	75%	64.8%	76.9%
65:35	128	70	71.4%	61.8%	72.4%
70:30	138	60	69.8%	59.5%	69.6%
75:25	148	50	69.8%	62.1%	72.5%
80:20	158	40	72.6%	63.9%	71.8%
85:15	168	30	80.6%	70%	70.4%
90:10	**178**	**20**	**81.2%**	**70.8%**	**76.7%**

Based on Table 6, the highest balanced accuracy is the data split with 90% training data and 10% testing data. The balanced accuracy values for Zone 1, Zone 2, and Zones 3 & 4 are 81.2%, 70.8%, and 76.7%.

Table 7. Balanced Accuracy from Data Scaling Using the Robust Scaler Method

Data Split	Training Data	Testing Data	Balanced Accuracy		
			1	2	3 & 4
60:40	**135**	**90**	**73.6%**	**60.8%**	**66.7%**
65:35	146	79	69.6%	57%	67.6%
70:30	157	68	66.1%	51.8%	64.7%
75:25	168	57	66.7%	52.9%	68.4%
80:20	180	45	68.8%	55.7%	68.3%
85:15	191	34	67.6%	55.4%	67.7%
90:10	202	23	73.5%	55.4%	70%

Based on Table 7, the highest balanced accuracy is the data split with 60% training data and 40% testing data. The balanced accuracy values for Zone 1, Zone 2, and Zones 3 & 4 are 73.6%, 60.8%, and 66.7%. Based on the comparison between the highest balanced accuracy in Table 6 and Table 7, the highest balanced accuracy is the data split with 90% training data and 10% testing data on the data scaled using the Min Max method. Therefore, this study uses 178 training data and 20 testing data on the scaled data using the Min Max method without involving outliers.

Multicollinearity Detection

Multicollinearity detection uses the variance inflation factors (VIF) value with VIF values for each independent variable shown in Table 8.

Table 8. Multicollinearity Detection Results (VIF Value)

Variable	IHK	IPM	IKK	IPI	JOP	JOSD
VIF Value	1.054536	1.176913	1.086839	1.136817	1.019433	1.038699

Based on Table 8, the VIF value of all independent variables is less than 10, which means there is no multicollinearity between independent variables.

Assumption of Proportional Odds

The proportional odds model assumption test uses the likelihood ratio value test, and the test results are shown in Table 9.

Table 9. Proportional Odds Model Assumption Test Results

Model	Log Likelihood	Chi-square	Degrees of Freedom	P-Value
The model with the same independent variable coefficient for each model in different categories (L_0)	-139.90			
The model with unequal independent variable coefficients for each model in different categories (L_1)	-136.38	7.0463	6	0.3166

Based on Table 9, the values of $L_0 = -139.90$ and $L_1 = -136.38$ are obtained so that $LR = -2(L_0 - L_1) = -2(-139.90 - (-136.38)) = 7.04 < \chi^2_{0.05;6} = 12.59$ with $p - value = 0.3166 > \alpha = 0.05$ so H_0 fails to be rejected. Therefore, the coefficients of the independent variables are the same for each model with different categories, which means the proportional odds assumption is met.

Goodness of Fit Test

The method used to test model fit is the Lipsitz method, and the test results are shown in Table 10.

Table 10. Model Fit Test Results

Statistical Test Value	Degrees of Freedom	P-Value
3.147	9	0.9582

Based on Table 10, $LR = 3.147 < \chi^2_{0.05;9} = 16.919$ with $p - value = 0.9582 > \alpha = 0.05$ so H_0 fails to be rejected. Therefore, the ordinal logistic regression model fits the data.

Parameter Significant Test

The simultaneous parameter significance test was conducted using the likelihood ratio test, with the results shown in Table 11.

Table 11. Model Fit Test Results

Model	Log Likelihood	Chi-square	Degrees of Freedom	P-Value
Model without independent variables (L_0)	-186.11			
Model with independent variables (L_1)	-139.90	92.416	6	< 0.001

Based on Table 11, $G = -2(L_0 - L_1) = -2(-186.11 - (-139.90)) = 92.416 > \chi^2_{0.05;6} = 12.6$ with $p - value < \alpha = 0.05$ so H_0 is rejected. Therefore, at least one independent variable is significant to the dependent variable. To assess the effect of each independent variable on the dependent variable, a Wald test was performed, and the results are presented in Table 12.

Table 12. Partial Parameter Testing Results

Variable	Parameter Estimation	P-Value	Description
CPI	-0.1428	0.85016	Not Significant
HDI	-1.1089	0.16773	Not Significant
CCI	6.9557	1.6×10^{-12}	Significant
IPI	-2.4934	0.00363	Significant
ODHC	-1.4214	0.04491	Significant
ODPS	-0.7085	0.29481	Not Significant
Zone 1 | Zone 2	0.2973		
Zone 2 | Zone 3	2.5083		

Based on Table 12, the p-value for the CCI, IPI, and ODHC variables is smaller than $\alpha = 0.05$, so H_0 is rejected. Therefore, only the CCI, IPI, and ODHC variables influence the determination of location allowance zoning. After knowing what variables are significant in determining the zoning of location allowances, a new model was formed by involving only significant variables, namely the CCI, IPI, and ODHC variables, and the results are shown in Table 13.

Table 13. Partial Parameter Testing Results Significant Variables

Variable	Parameter Estimation	P-Value	Description
CCI	6.6466	2.1×10^{-12}	Significant

Variable	Parameter Estimation	P-Value	Description
IPI	−2.8949	0.000322	Significant
ODHC	−1.4938	0.033527	Significant
Zone 1 | Zone 2	0.8302		
Zone 2 | Zone 3	3.0211		

The final model of ordinal logistic regression is a model that only involves significant variables, namely the CCI, IPI, and ODHC variables. The final model of ordinal logistic regression is as follows.

$$ln\left(\frac{\widehat{\gamma}_1(\boldsymbol{x})}{1-\widehat{\gamma}_1(\boldsymbol{x})}\right) = 0.8302 - 6.6466x_3 + 2.8949x_4 + 1.4938x_5 \quad (16)$$

$$ln\left(\frac{\widehat{\gamma_2}(\boldsymbol{x})}{1-\widehat{\gamma_2}(\boldsymbol{x})}\right) = 3.0211 - 6.6466x_3 + 2.8949x_4 + 1.4938x_5 \quad (17)$$

Model Evaluation

The evaluation of the model was conducted using the test set data (10%) and the initial policy data of Bank Syariah "X". The results of model evaluation based on the test set data (10%) are shown in Table 14, the results of model evaluation based on the initial policy data of Bank Syariah "X" are shown in Table 15.

Table 14. Model Evaluation Based on Test Set Data (10%) Results

Evaluation Measurements	Zone Category	Percentage
Accuracy		70%
	Zone 1	87.5%
Sensitivity	Zone 2	57.1%
	Zone 3 & 4	60%
	Zone 1	75%
Specificity	Zone 2	84.6%
	Zone 3 & 4	93.3%
	Zone 1	81.2%
Balanced Accuracy	Zone 2	70.8%
	Zone 3 & 4	76.7%

Table 15. Model Evaluation Based on Initial Policy Data of Bank Syariah "X" Results

Evaluation Measurements	Zone Category	Percentage
Accuracy		64.4%
	Zone 1	82.1%
Sensitivity	Zone 2	45.2%
	Zone 3 & 4	54.3%
	Zone 1	66.4%
Specificity	Zone 2	79.6%
	Zone 3 & 4	94.9%

Evaluation Measurements	Zone Category	Percentage
Balanced Accuracy	Zone 1	74.2%
	Zone 2	62.4%
	Zone 3 & 4	74.6%

The final model was created from 90% of training data using three significant variables, namely CCI, IPI, and ODHC. Based on Table 9, accuracy from 10% of test data using the model is 70%, and the balanced accuracy for Zone 1 is 81.2%, Zone 2 is 70.8%, and Zona 3 & 4 is 76.7%. Based on Table 10, accuracy from initial policy data (225 outlets) using the model is 64.4%, and the balanced accuracy for Zone 1 is 74.2%, Zone 2 is 62.4%, and Zona 3 & 4 is 74.6%. Therefore, based on the formed ordinal logistic regression model, the model misclassified 80 outlets or 35.6% of outlets.

4. DISCUSSIONS

The analysis of factors explaining the determination of location allowance zoning for Bank Syariah "X" outlets was conducted using the ordinal logistic regression method because the dependent variable (location allowance zoning) is ordinal. The selection of data for analysis was based on a comparison of balanced accuracy results obtained from seven training–testing data splitting schemes and various data scaling methods. Based on the comparison of the balanced accuracy results of each zone, the selected data division scheme is 90% training data and 10% testing data with the Min Max scaling method. In this study, the proportional odds model is used because it is the most common and frequently used model. The proportional odds model can be built if there is no multicollinearity and it fulfills the proportional odds assumption. Parameter estimation is done using the maximum likelihood method. The final model was formed from 90% of the training data, using three significant variables such as the construction cost index (CCI), the infrastructure pillar index (IPI), and the outlet distance to the nearest health center (ODHC). Based on the obtained ordinal logistic regression model equation, the estimated coefficient value for the CCI variable is -6.6466, so $exp(-6.6466) = 0.0013$. It means that for every 1-point increase in the CCI, the likelihood of an outlet being classified in Zone 1 (compared to Zone 2 or Zones 3 & 4) or Zone 1 or Zone 2 (compared to Zones 3 & 4) decreases by a factor of 0.0013. The estimated coefficient value for the construction cost index (CCI) variable is 2.8949, so $exp(2.8949) = 18.0834$. It means that for every 1-point increase in the CCI, the likelihood of an outlet being categorized in Zone 1 (compared to Zone 2 or Zones 3 & 4) or Zone 1 or Zone 2 (compared to Zones 3 & 4) increases by 18.0834 times. The estimated coefficient value for the variable distance from the outlet to the nearest health center (ODHC) is 1.4938, so $exp(1.4938) = 4.4542$. It means that for every 1-meter increase in ODHC, the likelihood of an outlet being classified in Zone 1 (compared to Zone 2 or Zone 3 & 4) or Zone 1 or Zone 2 (compared to Zone 3 & 4) increases by 4.4542 times. The final proportional odds model produced an accuracy rate of 70%, and the balanced accuracy for Zone 1, Zone 2, and Zone 3 & 4 were 81.2%, 70.8%, and 76.7%, respectively. Based on the formed ordinal logistic regression model, 80 outlets, or 35.6% of the outlets, were misclassified by the model.

5. CONCLUSION

In this research, the following conclusions were obtained:

1. Using the ordinal logistic regression method, this study identified the construction cost index (CCI), infrastructure pillar index (IPI), and outlet distance to the nearest health center (ODHC) as significant factors influencing the zoning of location allowances for Bank Syariah "X" outlets. A decrease in the construction cost index (CCI), an increase in the infrastructure pillar index (IPI), and an increase in the outlet distance to the nearest health center (ODHC) increases the odds of an outlet being classified as Zone 1 (versus Zone 2 or Zones 3 & 4) or in Zone 1 or Zone 2 (versus Zones 3 & 4). The final proportional odds model produced an accuracy rate of 70%, and the balanced accuracy for Zone 1, Zone 2, and Zone 3 & 4 were 81.2%, 70.8%, and 76.7%, respectively.

2. Based on the results of the proportional odds ordinal logistic regression model, the predicted location allowance zoning for the new Bank Syariah "X" outlets was as follows: Zone 2 for Bengkayang Regency, Zone 1 for Bangli Regency, and Zone 3 for Tual City.

6. REFERENCES

[1] I. W. Pratama and G. Sukarno, "Analisis Penilaian Kinerja, Reward, dan Punishment terhadap Kinerja Pegawai pada Badan Kepegawaian Daerah Provinsi Jawa Timur," *Jurnal STEI Ekonomi*, vol. 30, no. 02, pp. 20–32, Nov. 2021, doi: 10.36406/jemi.v30i02.460.

[2] R. M. Rostandi and S. H. Senen, "Pengaruh Kepuasan Kerja dan Reward System Terhadap Turnover Intention," 2021. Accessed: Jun. 07, 2024. [Online]. Available: https://ejournal.upi.edu/index.php/mdb/article/view/38726

[3] S. Döringer, F. Porst, L. Stumpf, and M. Heurich, "The Relationship between Measured Visitor Density and Perceived Crowding Revisited: Predicting Perceived Crowding in Outdoor Recreation," *Leis Sci*, 2023, doi: 10.1080/01490400.2023.2265366.

[4] S. Zalaghaie, M. Heidari, M. R. Nikudel, S. H. Jalali, and A. Seif, "Application of Ordinal Logistic and Gene Expression Programming Methods to Predict The Collapse Sensitivity Classes of Loess Soils, A Case Study: Golestan Province, Northeastern Iran," *Environ Monit Assess*, vol. 195, no. 2, Feb. 2023, doi: 10.1007/s10661-023-10915-5.

[5] F. F. Addini, D. Haryanto, and R. A. Maolani, "Klasifikasi Tingkat Risiko Kerugian Kecelakaan berdasarkan Karakteristik Pengemudi dengan Analisis Regresi Logistik Ordinal," *Jurnal Matematika Integratif*, vol. 18, no. 2, p. 167, Dec. 2022, doi: 10.24198/jmi.v18.n2.41317.167-177.

[6] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, Third Edition. 2013.

[7] A. Agresti, *Analysis of Ordinal Categorical Data*. John Wiley & Sons, Incorporated, 2010.

[8] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*. John Wiley & Sons, Incorporated., 2012.

[9] D. N. Gujarati, *Basic Econometrics*, 4th ed. The MC Graw-Hil Companies, 2003.

[10] T. Bekelcho, A. B. Olani, A. Woldemeskel, M. Alemayehu, and G. Guta, "Identification of Determinant Factors for Crash Severity Levels Occurred in Addis Ababa City, Ethiopia, from 2017 to 2020: Using Ordinal Logistic Regression Model Approach," *BMC Public Health*, vol. 23, no. 1, Dec. 2023, doi: 10.1186/s12889-023-16785-3.

[11] P. Lu, H. Wang, and D. Tolliver, "Prediction of Bridge Component Ratings Using Ordinal Logistic Regression Model," *Math Probl Eng*, vol. 2019, 2019, doi: 10.1155/2019/9797584.

[12] M. W. Fagerland and D. W. Hosmer, "How to Test for Goodness of Fit in Ordinal Logistic Regression Models," *Stata J*, vol. 17, no. 3, pp. 668–686, 2017, Accessed: Jun. 07, 2024. [Online]. Available: https://journals.sagepub.com/doi/pdf/10.1177/1536867X1701700308

[13] P. A. Octaviani, Y. Wilandari, and D. Ispriyanti, "Penerapan Metode Klasifikasi Support Vector Machine (SVM) pada Data Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang," vol. 3, no. 4, pp. 811–820, 2014, [Online]. Available: http://ejournal-s1.undip.ac.id/index.php/gaussian

[14] A. Kulkarni, F. A. Batarseh, and D. Chong, "Foundations of Data Imbalance and Solutions for a Data Democracy," 2021. Accessed: May 20, 2024. [Online]. Available: https://arxiv.org/pdf/2108.00071

[15] M. Grandini, E. Bagli, and G. Visani, "Metrics for Multi-Class Classification: an Overview," Aug. 2020, [Online]. Available: http://arxiv.org/abs/2008.05756

[16] A. Handayani, A. Jamal, and A. A. Septiandri, "Evaluasi Tiga Jenis Algoritme Berbasis Pembelajaran Mesin untuk Klasifikasi Jenis Tumor Payudara," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 6, 2017, Accessed: May 10, 2024. [Online]. Available: https://jurnal.ugm.ac.id/v3/JNTETI/article/view/2806/843

[17] X. H. Cao, I. Stojkovic, and Z. Obradovic, "A Robust Data Scaling Algorithm to Improve Classification Accuracies in Biomedical Data," *BMC Bioinformatics*, vol. 17, no. 1, Sep. 2016, doi: 10.1186/s12859-016-1236-x.

[18] A. Ambarwari, Q. J. Adrian, and Y. Herdiyeni, "Analisis Pengaruh Data Scaling Terhadap Performa Algoritme Machine Learning untuk Identifikasi Tanaman," *Jurnal Rekayasa Sistem dan*

Teknologi Informasi, vol. 4, pp. 117–122, 2017, Accessed: May 07, 2024. [Online]. Available: https://jurnal.iaii.or.id/index.php/RESTI/article/view/1517/210

[19] A. Khoirunnisa and N. G. Ramadhan, “Improving Malaria Prediction with Ensemble Learning and Robust Scaler: An Integrated Approach for Enhanced Accuracy,” *JURNAL INFOTEL*, vol. 15, no. 4, pp. 326–334, Nov. 2023, doi: 10.20895/infotel.v15i4.1056.