BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 1, December 2025 | Hal 491-501
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i1.861

School Accreditation Prediction Based on Literacy and Numeracy: Ordinal Logistic Regression vs KNN

Nabila Syukri, Yani Prihantini Hiola, Mega Ramatika Putri, Budi Susetyo*

Sekolah Sains Data, Matematika, dan Informatika, Program Studi Statistika dan Sains Data, Institut Pertanian Bogor, Bogor, Indonesia

Email: [1]nabilasyukri@apps.ipb.ac.id, [2]yanihiola@apps.ipb.ac.id, [3]megaramatikaputri@apps.ipb.ac.id, [4,*]budisu@apps.ipb.ac.id

Email Penulis Korespondensi: budisu@apps.ipb.ac.id

Abstract−School accreditation in Indonesia has traditionally relied on administrative inputs and institutional documentation, which often fail to capture the actual quality of student learning. In contrast, the National Assessment provides direct evidence of student literacy and numeracy outcomes, offering a more objective and outcome-based measure of educational quality. Leveraging these results as predictors for accreditation rankings is therefore crucial, as they reflect the competencies most relevant to effective learning delivery. This study aims to develop and evaluate classification models for school accreditation rankings using literacy and numeracy results as predictor variables. The dataset consists of secondary data from the 2023 and 2024 National School Assessments, covering 789 schools across four provinces: DKI Jakarta, Yogyakarta, Bali, and Banten. Two methods were applied, Ordinal Logistic Regression and K-Nearest Neighbors (K-NN) under two scenarios: with and without class imbalance handling. To address imbalance, two techniques were employed: Synthetic Minority Oversampling Technique (SMOTE) and Class Weight. The results indicate that K-NN consistently outperformed Ordinal Logistic Regression in both scenarios. On data without imbalance handling, K-NN achieved Accuracy, Precision, Recall, and F1-Score of 0.803, 0.705, 0.587, and 0.619, respectively. with imbalance treatment using SMOTE, the values were 0.753, 0.619, 0.686, and 0.644. While class balancing did not significantly improve overall accuracy, it enhanced the model's ability to recognize minority classes. These findings highlight the strong relationship between literacy and numeracy outcomes and school accreditation status, demonstrating that outcome-based measures can complement traditional accreditation instruments, and that conventional statistical approaches are still relevant for modeling school accreditation.

Keywords: Accreditation; Literacy; Numeracy; Ordinal Logistic Regression; K-Nearest Neighbor (KNN)

1. INTRODUCTION

Education is a fundamental component in human resource development that needs more attention in order to improve national progress [1]. Quality education will produce a superior generation, therefore improving the quality of education is a top priority in human development efforts. One of the benchmarks in assessing the quality of education is school accreditation ranking, which reflects the quality of learning based on standards set by the accreditation agency [2]. Accreditation assessment covers various aspects such as curriculum, learning process, human resources (HR), facilities and infrastructure, school management, and student learning outcomes [3]. Thus, accreditation is not only a formal evaluation tool but also a reflection of the quality of education in schools.

One important aspect in assessing student learning outcomes is literacy and numeracy skills, as the two are interrelated and support the development of critical thinking in students [4]. Literacy is the ability to understand, use, evaluate, and create information through various forms of language, while numeracy is the ability to use mathematical concepts and skills in various contexts of life [5]. The application of literacy and numeracy skills will help students understand lessons more broadly and deeply [6]. Therefore, strengthening literacy and numeracy is an important effort in improving the quality of education, which should be integrated with external quality assurance systems such as accreditation.

Determining accreditation ranking is essentially a matter of classification. Previous studies highlight that accreditation results in Indonesia are often inconsistent with other quality indicators such as national exams and PISA, and recommend reform through predictive statistical models and performance based instruments [7]. In line with this, Ridho (2024) developed a predictive model using machine learning to strengthen school accreditation. His study explored the influence of gender, growth mindset, and education level on accreditation outcomes, finding significant relationships with educational quality [8]. These findings reinforce the importance of shifting accreditation towards outcome based indicators that directly reflect student learning. In line with this direction, the present study proposes a school quality assessment model that utilizes student literacy and numeracy outcomes derived from the National Assessment as predictor variables for accreditation classification. By focusing on literacy and numeracy, this approach provides a more direct reflection of learning effectiveness and strengthens the relevance of accreditation rankings to actual student performance.

The accreditation dataset is characterized by ordinal response categories (ranks: A, B, C) and predictor variables derived from literacy and numeracy scores, which exhibit non-linear relationships and potential multicollinearity. Ordinal Logistic Regression is employed as the baseline conventional method because it is specifically designed for ordinal outcomes, allowing estimation of cumulative probabilities across accreditation levels. However, its performance may be constrained when predictor variables violate assumptions of linearity in the logit or proportional odds, conditions that are frequently encountered in educational assessment data.

As a comparison, the K-Nearest Neighbor (K-NN) algorithm is utilized as a non-parametric alternative that does not rely on distributional assumptions. K-NN is particularly suitable for datasets where class boundaries are irregular and shaped by complex interactions among predictors, as is the case with literacy and numeracy scores. In addition, the accreditation dataset is imbalanced, with accreditation category A dominating compared to B and C. This imbalance can

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 1, December 2025 | Hal 491-501
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i1.861

bias classification models toward the majority class and reduce their ability to correctly identify schools in lower accreditation categories. To address this issue, two imbalance-handling techniques are applied: Class Weighting, which adjusts the influence of minority classes during model training, and the Synthetic Minority Oversampling Technique (SMOTE), which generates synthetic samples to balance class distribution. The integration of both classification models with imbalance-handling strategies is evaluated to rigorously assess the predictive strength of literacy and numeracy outcomes in determining school accreditation status.

School accreditation results in Indonesia show significant variations across provinces [9]. These differences or disparities illustrate the inequality of education at the national level. This condition can affect the performance of the classification model, because the relationship pattern between the predictor and response variables can differ in each region. Therefore, to maintain the uniformity of education quality between regions, this study focuses on senior high schools in four provinces: DKI Jakarta, Yogyakarta, Banten, and Bali, to obtain a more representative and stable model. These provinces were selected based on the consideration that they are in the highest cluster nationally according to 2023 data on high school equivalent education completion rates from www.bps.go.id, with all four provinces ranking in the top ten nationally.

Based on this background, this study aims to develop a model for classifying school accreditation rankings based on students literacy and numeracy skills, and to compare the performance of Ordinal Logistic Regression and KNN models. The results are expected to provide empirical evidence on the extent to which literacy and numeracy represent school quality and to inform policies for improving education quality in Indonesia.

2. RESEARCH METHODOLOGY

This research was conducted in a series of stages using two methods to develop a classification model. The stages began with data exploration, dividing the data into training data and test data, modeling with logistic regression and K-NN, model evaluation, and interpretation of the best model. The data exploration stage aims to understand the characteristics of the data, identify class distributions, and detect missing values as well as class imbalance. Following the data exploration stage, the dataset is split into training data, which are used to train the models, and testing data, which are used to generate predictions from the trained models. In the training phase, Logistic Regression and K-Nearest Neighbor (KNN) models employ different modeling procedures, as illustrated in Figure 1. After model development, the analysis is continued by evaluating the models using both training and testing data. The evaluation results obtained from these two datasets are expected to be relatively consistent in order to avoid indications of overfitting. Finally, the performance of the two models is compared, leading to the selection of the best model for classification.

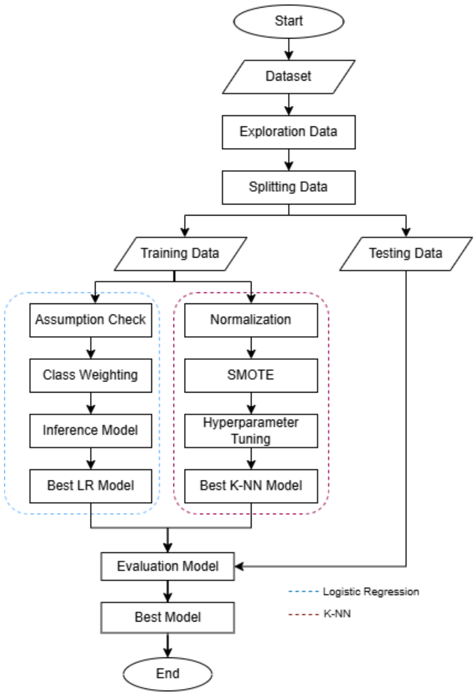

Figure 1. Flowchart Analysis

2.1 Data

The data used in this study is secondary data from the 2023 and 2024 National School Assessments conducted by the Ministry of Education and Culture, with a specific focus on four provinces: DKI Jakarta, Special Region of Yogyakarta, Bali, and Banten. The total observation consisted of 789 schools. The response variable is the school accreditation ranking (Y), categorized into three ordinal levels: A, B, and C. The predictor variables include four indicators derived from the

National Assessment results, namely Literacy 2023 (X1), Numeracy 2023 (X2), Literacy 2024 (X3), and Numeracy 2024 (X4). To provide a clearer picture of the dataset, Table 1 presents the distribution of schools across accreditation categories.

Table 1 Descriptive Statistics of Predictor Variables

Accreditation	A	B	C
Number of School (%)	542 (66.8%)	201 (25.5%)	46 (5.8%)

Table 1 shows that accreditation category A dominates the dataset, while categories B and C are relatively underrepresented, indicating the presence of class imbalance. This imbalance justifies the application of techniques such as Class Weighting and the Synthetic Minority Oversampling Technique (SMOTE) in the modeling process to improve the recognition of minority classes.

2.2 Data Exploration

Data exploration was conducted to understand the characteristics of the variables to be used in further modelling. This included understanding the data structure, patterns, and relationships in the dataset [10]. Exploration was also used to identify data quality issues that would affect the analysis results and the model built.

2.3 Splitting Data

The dataset was splitted into training data (80%) and testing data (20%). The training data was used to build the classification model, while the test data was used to evaluate the model. *Cross-validation* is performed on the training data with a total of 10 *folds*. This *cross-validation* process is used to obtain a more stable and robust estimate of the model's performance on data that has not been seen before, as well as to select the best set of hyperparameters before the final model is retrained on all training data and evaluated on the test data [11].

2.4 Handling Imbalance Class

In classification modeling, the most common problem is class imbalance in the response variable, so it is necessary to take measures to prevent the model from being biased in classifying the majority class [12]. To overcome this, various class balancing techniques have been developed, including the Synthetic Minority Oversampling Technique (SMOTE) and Class Weight. SMOTE balances the data by generating synthetic samples for the minority class, while Class Weight addresses imbalance by assigning greater weights to classes with fewer observations [13]. In the context of Ordinal Logistic Regression, the use of Class Weight is more appropriate because it preserves the natural order between categories, making it preferable to resampling-based methods such as SMOTE. In contrast, the K-Nearest Neighbor (K-NN) algorithm treats the response variable as nominal without considering ordinal structure. Since KNN relies heavily on distance metrics (commonly Euclidean), it is highly sensitive to the distribution of samples in feature space. Therefore, adding synthetic minority samples through SMOTE is more effective for KNN than class weight, as it directly reshapes the neighborhood structure and reduces bias toward majority classes.

2.4.1 *Class Weight*

Class weight is a handling method that gives greater weight to minority classes and smaller weight to majority classes so that the model will pay more attention to minority classes [14]. This approach makes the model more sensitive to minority classes without changing the original data distribution. Class weights are calculated using the following formula:

$$w_k = \frac{N}{K \times n_k} \quad (1)$$

Where N is total number of observations, K represent number of classes and n_k is number of observation in class k

2.4.2 Synthetic Minority Oversampling (SMOTE)

SMOTE is a data based class balancing technique that adds observations to the minority class so that it has a balanced number of observations with the majority class. SMOTE generates synthetic data by finding k nearest neighbors for each minority class data [15].

2.5 Classification Modeling

Classification modeling aims to build predictive models that can group or predict classes or categories of data based on their features or variables [16]. These classification models study the relationship between input data and predetermined class labels to then predict the class of new data that has not been seen before. This study uses conventional and machine learning approaches in building classification models.

2.5.1 Ordinal Logistic Regression

Ordinal logistic regression is a conventional statistical method used to model the relationship between an ordinal response variable (Y) and its predictor variables [17]. The ordinal nature is formed into cumulative logit or proportional odds to

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 1, December 2025 | Hal 491-501
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i1.861

estimate the cumulative probability of a particular category. The cumulative model of ordinal logistic regression is as follows:

$$logit\ [P(Y \leq J|X)] = \log\left[\frac{P(Y \leq j|x)}{1-P(Y \leq j|x)}\right] = \log\left[\frac{\pi_1(x)+\pi_2(x)+\cdots+\pi_j(x)}{\pi_{j+1}(x)+\pi_{j+2}(x)+\cdots+\pi_j(x)}\right] = \alpha_j + \beta_1 x_1 + \cdots \beta_p x_p \quad (2)$$

The resulting model has different intercept values (cutpoints) for each cumulative logit and has the same value for β for all cumulative logits. Parameter significance testing is performed to determine the effect of predictor variables on response variables, both simultaneously and partially. In performing simultaneous parameter significance testing, the G test statistic is used, which is a form of *the* Likelihood Ratio Test (LRT) resulting from a comparison between the likelihood function without independent variables and the likelihood function with predictor variables [18]. The general formula for this test is:

$$G = 2\ln\left[\frac{likelihood\ tidak\ dengan\ variabel\ bebas}{likelihood\ dengan\ variabel\ bebas}\right] \quad (3)$$

Theoretically, the G statistic follows a χ^2 distribution with k degrees of freedom. The hypotheses for the G test are as follows:
$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$
H_1 : minimal ada satu β yang tidak sama dengan 0
The rejection region for H0 is if$G > \chi^2_{(\alpha,p)}$, or the *p-value* is < α.

Meanwhile, for the partial parameter significance test, the Wald test is used to determine the effect of each predictor variable on the response variable with the following formula:

$$W = \frac{\widehat{\beta_i}}{SE(\widehat{\beta_i})} \quad (4)$$

Theoretically, the statistic W follows a standard normal distribution (Z) if H_0 is true.

To facilitate interpretation and adjustment to the more commonly used test distribution, the statistical value can be squared so that it follows the *Chi-square* distribution with one degree of freedom (df = 1), which is formulated as follows:

$$\chi^2_{wald} = \left(\frac{\widehat{\beta_i}}{SE(\widehat{\beta_i})}\right)^2 \quad (5)$$

The rejection region is if$\chi^2_{wald} > \chi^2_{(\alpha,1)}$ or the *p-value* is less than *α*.

Model coefficient interpretation is performed using odds ratios, which indicate changes *in* response category odds when the predictor variable increases by one unit [19]. The odds values of two individuals can be compared using the following formula :

$$\widehat{OR} = \exp[\beta_i(x_1, x_{i+1})] \quad (6)$$

Equation (6) above is the odds ratio value for continuous explanatory variables when the variable$x$ increases by one unit. This means that when there is an increase of one unit, the odds increase by $\exp[\beta_i(x_1, x_2)]$ s compared to before.

2.5.2 K-Nearest Neighbors (KNN)

K-NN is a supervised learning algorithm that can be used for regression and classification. In classification, K-NN determines the class of new data based on the class of its k nearest neighbors in feature space [20]. The main process of K-NN is to first determine the value of k , in this case the number of nearest neighbors used for classification. The k nearest neighbors are found by calculating the distance of observation points in the training data. After determining the distance of each point, the smallest value or closest distance is sought. The distance is generally measured using Euclidean distance, which can be represented as follows:

$$D(a,b) = \sqrt{\sum_{k=1}^{d}(a_k - b_k)^2} \quad (7)$$

The process of finding the k in machine learning can be done with hyperparameter tuning to produce the optimal k [21], with a range of candidate values was tested using cross-validation to identify the optimal classification performance. Since all literacy and numeracy predictors are measured on the same scale (0–100), normalization or standardization was not required. This ensures that each variable contributes proportionally to the Euclidean distance calculation without bias from differing measurement units. K-NN is a simple but effective non-parametric algorithm, with its strength in its ability to adapt classification based on proximity in the data.

2.6 Evaluasi Model

The evaluation of classification models aims to assess how well the model predicts the correct data class. One of the main tools in evaluation is the confusion matrix, which maps the model's prediction results with the actual conditions of the

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 1, December 2025 | Hal 491-501
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i1.861

data [22]. This matrix displays four main categories: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The metrics used in model evaluation are:

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (8)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (9)$$

$$Recall = \frac{TP}{(FN+TP)} \quad (10)$$

$$F1\ Score = 2 \times \frac{(precision \times recall)}{(precision+recall)} \quad (11)$$

3. RESULTS AND DISCUSSION

3.1 Data Exploration

Data exploration was conducted to understand the characteristics prior to modeling. This analysis included exploration of the response variable (Y), predictor variables (X), and the relationship between the two. Initial exploration was conducted on the response variable to identify class distribution. Of the total 789 schools in the dataset, the accreditation rating distribution was imbalanced.

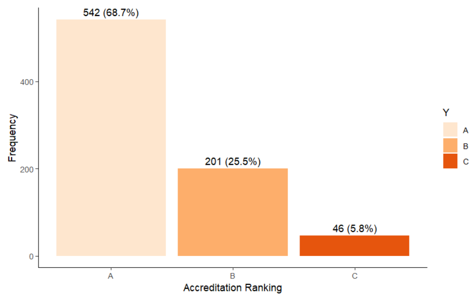

Figure 2 Distribution of Accreditation Ranking

As shown in Figure 2, the number of schools with a 'C' rating (46 schools, 5.8%) is significantly lower than those with an A rating (542, 68.7%) and B rating (201, 25.5%). This finding validates the need to apply class balancing techniques at the modeling stage to prevent models that are biased toward the majority class. The analysis continued on the four predictor variables. Summary statistics are shown in Table 2.

Table 2 Descriptive Statistics of Predictor Variables

Variable	Minimum	Median	Mean	Maximum	Standar Deviation
Literacy 2023	0	91.11	81.80	100	22.045
Numeracy 2023	6.67	83.33	77.17	100	20.334
Literacy 2024	0	93.33	83.46	100	22.083
Numeracy 2024	0	87.81	80.11	100	21.205

Based on Table 2, there are minimum values for literacy and numeracy variables recorded as 0 (zero). These values do not represent missing data or schools that did not participate in the assessment, but rather reflect the actual scores obtained by certain schools in the national evaluation. Therefore, the presence of zero scores should be considered part of the valid data distribution, even though they may be treated as outliers in statistical analysis. Boxplot visualizations in Figure 3 are used to examine the scale, distribution, and outliers.

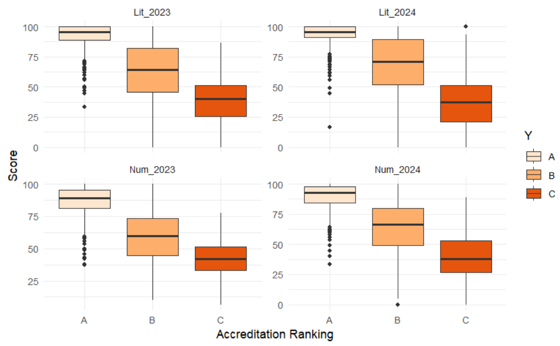

Figure 3 Distribution of Literacy and Numeracy Variables

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 1, December 2025 | Hal 491-501
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i1.861

Based on Figure 3, the boxplot for each variable shows a clear pattern in distinguishing accreditation rankings. Schools with an A ranking consistently have a higher median score (literacy and numeracy) than schools with a B ranking, which is also higher than those with a C ranking. Therefore, we will next examine the linear relationship between the four predictor variables.

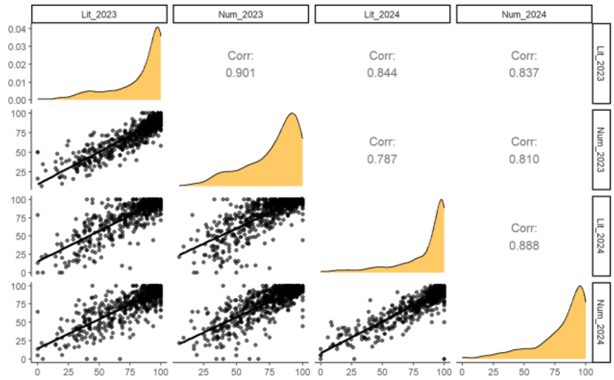

Figure 4 Correlation between Predictor Variables

Figure 4 displays the correlation matrix between predictor variables. The correlation of all variables is above 0.7, which means that every increase in one variable also increases the other variables. However, even though the correlation between predictor variables is high, it does not indicate multicollinearity, as shown in Table 2 below.

Table 3 VIF Value of Predictor Variables

Variable	Literacy 2023	Numeracy 2023	Literacy 2024	Numeracy 2024
VIF	7.24	5.64	5.63	5.65

Table 3 shows that the VIF values for each predictor variable are under 10, so it can be concluded that there is no multicollinearity in any of the predictor variables.

3.2 Ordinal Logistic Regression

In conducting classification analysis using ordinal logistic regression, there are several assumptions that must be met, including proportional odds and logit linearity. Fulfilling these assumptions is important so that the results of inferential analysis (such as testing the significance of parameters) and predictions (determining response categories) can be interpreted correctly and produce a valid model.

3.2.1 Inference

At the inference stage, the analysis focuses on ensuring that the ordinal logistic regression model meets the necessary assumptions so that the parameter estimation results can be interpreted correctly. Several analyses that are conducted at this stage include testing the *proportional odds* assumption, testing logit linearity, and testing the significance of model parameters.

a. *Proportional Odds*

In conducting *proportional odds* testing, a model is first developed to obtain the following two logit forms.

$$logit\left(P(Y < C|X|)\right) = 3.6160 - (0.0399Lit_{2023} + 0.0248Num_{2023} + 0.0176Lit_{2024} + 0.0260Num_{2024})$$

$$logit\left(P(Y < B|X|)\right) = 7.7592 - (0.0399Lit_{2023} + 0.0248Num_{2023} + 0.0176Lit_{2024} + 0.0260Num_{2024})$$

From the model formation above, testing is carried out using *the Brant test* to determine whether the resulting model meets the proportional odds assumption or not with the following hypothesis.

H_0 : The proportional odds assumption is satisfied (there is no significant difference in the coefficients between intercepts).

H_1 : The proportional odds assumption is not met (there are significant differences in the coefficients between intercepts).

The decision criterion is to reject the proportional odds hypothesis (H_0) if *the p-value* ≤ 0.05.

Table 4 Propotional Odds Test

Test	*p-value*
Omnibus	0.02
Lit 2023	0.21
Num 2023	0.60
Lit 2024	0.45
Num 2024	0.35

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 1, December 2025 | Hal 491-501
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i1.861

Table 4 shows the results of testing using the Brant test. The omnibus test indicates a violation of the proportional odds assumption, while testing per variable does not show a significant violation. This indicates that the potential violation is more global or cumulative in nature, rather than originating from a single variable. Thus, the proportional odds model can still be considered for use.

b. Logit Linearity

In logistic regression, one key assumption is linearity in the logit, meaning that each continuous predictor should have a linear relationship with the log-odds (logit) of the outcome variable. If this assumption is violated, the estimated coefficients may be biased, and the model's predictive performance can be compromised. Therefore, assessing logit linearity helps ensure that the functional form of the model is correctly specified and that the interpretation of regression coefficients remains valid.

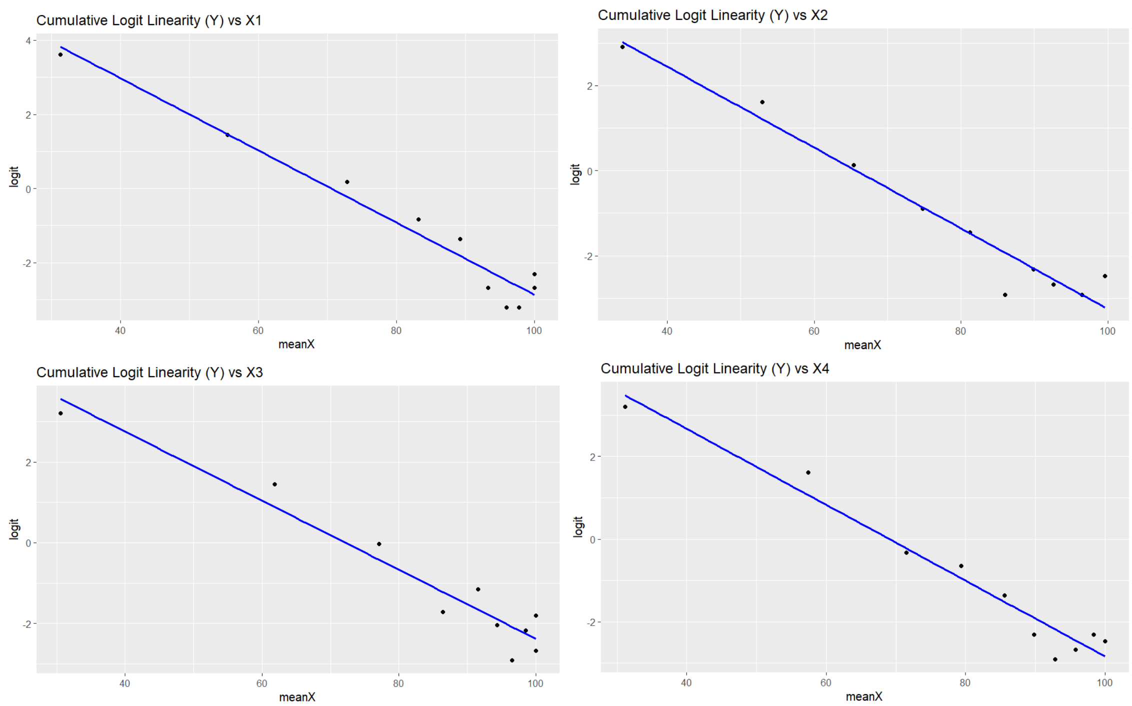

Figure 5. Logit Linearity Assumption

From the plot above, it can be seen that the four explanatory variables (X) have a linear relationship with the cumulative logit of the response variable (Y). Therefore, it can be said that the logit linearity assumption is satisfied

c. Parameter Significance Test

Simultaneous Test

Based on the results of the simultaneous test using the LRT test, the following results were obtained.

Table 5 Likelihood Ratio Test

LR Stat	*p-value*
497.32	0.00

From Table 5, it can be seen that at a significance level of 0.05, the H_0 are rejected because the p-value is < α. Therefore, the result obtained is that at least one β is not equal to 0, or in other words, there is a simultaneous relationship between the predictor variables and the response variables.

Partial Test

Based on the results of the *Wald* test, the following results are obtained.

Table 6 Partial Test

Variable	|t value|	t table	Decision
Lit 2023	3.782	1.96	Reject H0
Num 2023	2.365		Reject H0
Lit 2024	1.954		Fail to reject H0
Num 2024	2.801		Reject H0

From the partial test results using the Wald test, it is known that all variables except Literacy 2024 have a t-value of < 1.96. This means that the three predictor variables each have a significant effect on the response variable Y in the ordinal logistic regression model.

ISSN 2774-3659 (Media Online)
Vol 6, No 1, December 2025 | Hal 491-501
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i1.861

d. Interpretation
Based on the model obtained previously, it can be interpreted that each variable has a positive coefficient. This means that all predictor variables (literacy & numeracy in 2023 and 2024) have a positive effect on the increase in the Y category. Thus, the higher the literacy and numeracy scores, the greater the likelihood that respondents will be in a better category (e.g., A compared to B or C). Table 7 below is the interpretation of odds for each variable.

Table 7 Odds Ratio Interpretation

Variable	Interpretation
Lit 2023	For every 1-unit increase in the 2023 literacy score, the odds of entering a higher category increase by approximately exp(0.0399) ≈ 1.04 times
Num 2023	For every 1-unit increase in the 2023 numeracy score, the odds of entering a higher category increase by approximately exp(0.0248) ≈ 1.02 times.
Lit 2024	For every 1-unit increase in the 2024 literacy score, the odds of entering a higher category increase by approximately exp(0.0176) ≈ 1.02 times.
Num 2024	For every 1-unit increase in the 2024 numeracy score, the odds of entering a higher category increase by approximately exp(0.0260) ≈ 1.03 times.

3.2.2 Prediction

In making predictions, ordinal logistic regression modeling was performed using two approaches, namely class weight handling, to address imbalanced data and modeling without using handling techniques. In the initial step, the data was divided into training and test data in a ratio of 80:20. The number of observations in each class in the training data is as follows: Accreditation A (434 schools), Accreditation B (159 schools), and Accreditation C (38 schools). In modeling using imbalanced data handling, equation (1) is used, resulting in the following weights for each class.

$$w_A = \frac{631}{3\times 434} = 0.484$$

$$w_B = \frac{631}{3\times 159} = 1.322$$

$$w_C = \frac{631}{3\times 38} = 5.535$$

Based on these weights, it can be seen that for classes with few observations (minority class), the weights obtained are greater than those for other classes. Table 8 below shows the results of the evaluation of the test data using Ordinal Logistic Regression with two approaches, namely without handling imbalanced data and with handling imbalanced data.

Table 8 Evaluation of Ordinal Logistic Regression

Scenario	Evaluation Metric	Logistic Regression
Without Handling	*Accuracy*	0.778
	Precision	0.711
	Recall	0.534
	F1-Score	0.575
With Handling	*Accuracy*	0.753
	Precision	0.609
	Recall	0.686
	F1-Score	0.626

The evaluation results in Table 8 show that handling class *imbalance* in the Ordinal Logistic Regression model with *class weighting* changes the model's performance pattern. The accuracy value decreased slightly from 0.778 to 0.753, but the *recall* value increased significantly from 0.534 to 0.686. This demonstrates that imbalance handling makes the model more sensitive to minority classes, even at the expense of overall accuracy. Thus, the application of class balancing techniques in Ordinal Logistic Regression does not substantially increase total accuracy, but it does improve the model's ability to detect underrepresented accreditation categories.

3.3 K-Nearest Neighbors

K-NN is a non-parametric distance-based algorithm that ignores ordinal properties, so K-NN treats the response variable as nominal. The performance of the K-Nearest Neighbors (K-NN) model is highly dependent on the selection of the hyperparameter 'K' (number of neighbors). 'K' value that is too small can cause overfitting and sensitivity to noise, while a 'K' value that is too large can cause underfitting. To find the optimal 'K' value, a hyperparameter tuning process is performed using 10-Fold Cross Validation on the training data.

This process begins by defining a series of candidate 'K' values (from 2 to 50). For each 'K' value tested, the training data is divided into 10 "folds". The model is then trained and evaluated 10 times in each iteration, with 9 folds used to train the model and the remaining 1 fold used for validation, ensuring that each fold has been validation data at some point. Pre-processing is applied within each cross-validation iteration to prevent data leakage. The performance for

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 1, December 2025 | Hal 491-501
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i1.861

each candidate 'K' value is calculated as the average evaluation metric (e.g., Accuracy*)* of those 10 folds. The 'K' value that produces the highest average performance is then selected as the optimal 'K' for training the final K-NN model.

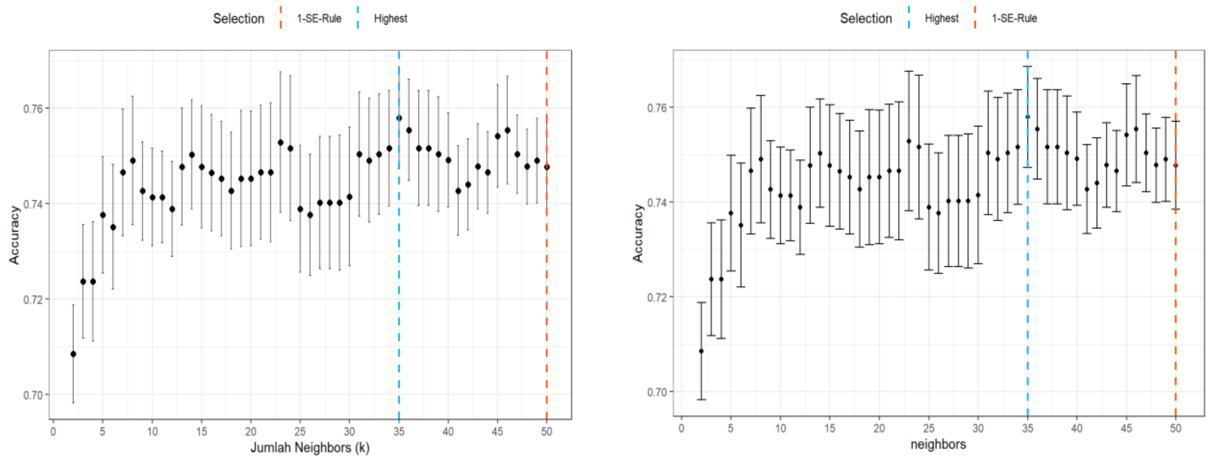

Figure 6 Number of Neighbors (k)

The left side of Figure 6 shows hyperparameter checking for$k$ without SMOTE handling, and the right side shows SMOTE handling. In both scenarios, it can be observed that the model accuracy performance increases sharply at very small K values (e.g., K < 10). After that , the model performance becomes relatively stable and flat in the higher range of$k$ with small fluctuations. The$k$ value that produces the highest absolute average accuracy is marked with a blue line, while the orange line represents the optimal$k$ value based on the 1-SE Rule. Table 8 below details the training data accuracy based on the optimal k value.

Table 9 Hyperparameter Tuning Results (k)

Scenario	Metric	k	Training Data Accuracy
No-SMOTE	Highest	35	0,834
No-SMOTE	1-SE Rule	50	0,829
SMOTE	Highest	35	0,781
SMOTE	1-SE Rule	50	0,781

Based on the given scenario, the model without imbalance handling (No SMOTE) shows better performance than the model using SMOTE. This may indicate that class imbalance handling through SMOTE does not always improve K-NN accuracy, especially if the data already has a fairly good distribution or if SMOTE produces synthetic samples that are less representative. Other factors may also influence this, such as data complexity and noise added during oversampling. The best hyperparameters in this case $k = 35$, were then used on the test data in each scenario. Table 10 below shows the evaluation results on the test data using K-Nearest Neighbor

Table 10 Evaluation of K-NN

Scenario	Evaluation Metrics	K-NN
Without Handling	Accuracy	0.803
	Precision	0.705
	Recall	0.587
	F1-Score	0.619
With Handling	Accuracy	0.753
	Precision	0.619
	Recall	0.680
	F1-Score	0.648

The evaluation results in Table 10 show that the K-NN model without SMOTE achieved the highest accuracy (0.803), but recall was limited (0.587). After applying SMOTE, recall increased to 0.680, showing improved recognition of minority classes, although accuracy declined slightly. Importantly, within the context of school accreditation, Recall is often more critical than Accuracy. Misclassifying a low-quality school (C) as high-quality (A) is more detrimental than misclassifying a high-quality school as lower-ranked. Therefore, models that improve recall even with a small reduction in accuracy are more valuable for accreditation purposes, as they better safeguard against overlooking schools that require quality improvement interventions. In theory, SMOTE aims to address class imbalance so that the model can better recognize minority classes. Therefore, in the context of classification with evaluation using accuracy metrics, the application of SMOTE does not necessarily improve model accuracy, but rather plays a role in improving recall related to model assessment in classifying minority classes. This finding is in line with the classification results in Table 10, which shows that the KNN model without SMOTE produces higher accuracy than KNN with SMOTE, but has a lower recall value than KNN with SMOTE.

3.4 Model Evaluation

After obtaining the best models from Logistic Regression and K-NN, each was evaluated on the test data using several metrics shown in the following table.

Table 11 Comparison Evaluation Results of Logistic Regression and K-NN

Scenario	Metric	Logistic Ordinal	K-NN	Best Model
Without Handling	Accuracy	0.778	0.803	KNN
	Precision	0.711	0.705	
	Recall	0.534	0.587	
	F1-Score	0.575	0.619	
With Handling	Accuracy	0.753	0.753	KNN
	Precision	0.609	0.619	
	Recall	0.686	0.680	
	F1-Score	0.626	0.648	

Based on Table 11, the model evaluation results show that, in general, the K-Nearest Neighbors (K-NN) method performs better than Ordinal Logistic Regression. In data without imbalance handling, K-NN produced the highest accuracy value of 0.803, while Logistic Regression produced 0.778. However, in data with imbalance handling, both models showed the same accuracy value (0.753), with K-NN slightly superior in F1-Score value (0.648 compared to 0.626).

This difference shows that K-NN is able to learn data patterns more flexibly without specific distribution assumptions, resulting in better predictions on data with natural distributions. Meanwhile, Ordinal Logistic Regression is more sensitive to handling imbalances and shows an increase in recall, which means it is better at recognizing minority classes. Thus, the selection of the best model depends on the purpose of the analysis. K-NN excels in overall accuracy, while Ordinal Logistic Regression is better at sensitivity to minority classes.

In addition to comparing methods based on evaluation metrics, the interpretation of the influence of Literacy and Numeracy variables on Accreditation Rankings also needs to be explained. Table 6 shows that the only insignificant variable is Literacy 2023. In essence, this shows the assumption that in 2023, the policies or curriculum implemented only had a small effect on the accreditation rating of students literacy skills, so that changes in policy and curriculum in 2024 can be said to have an effect on the accreditation rating. Conversely, if the 2023 Literacy variable turns out to have a significant effect compared to 2024 Literacy in partial testing, it means that there is an assumption that changes in the curriculum and policy cause students to experience a period of transition and adjustment so that 2024 Literacy skills do not have a significant effect on accreditation ratings. In addition to indications of policy and curriculum changes between 2023 and 2024, changes in assessment methods are also one of the suspicions that cause the two variables to provide different conclusions regarding the significance of their influence on accreditation ratings.

Based on the odds ratio (OR) obtained from Table 7, the variable with the highest OR is Literacy 2023. This result shows that compared to the other three variables, Literacy 2023 has the greatest effect on changes in accreditation rankings compared to other variables. However, the partial test results show that the effect of Literacy 2023 is not yet statistically significant. This finding indicates that although the effect of Literacy 2023 is relatively high, the level of statistical confidence in this effect is still limited. Thus, the odds ratio represents the magnitude of a variable's effect on changes in accreditation ratings, while the partial test describes the level of statistical certainty of this effect.

In addition, based on Table 11, learning outcomes (literacy and numeracy) are considered to be relevant indicators for accreditation rating assessment because the predictive ability of the best model (KNN) shows quite good evaluation results, especially in terms of accuracy and precision. Thus, learning outcomes can serve as valid indicators of institutional quality, but they are best employed in combination with complementary measures rather than as the sole benchmark, in line with the accreditation instrument outlined by the Ministry of Education, Culture, Research, and Technology [23].

4. CONCLUSION

This study demonstrates that students' literacy and numeracy outcomes are strongly correlated with school accreditation status, confirming their validity as indicators of institutional quality. While K-Nearest Neighbors showed stronger predictive performance and Ordinal Logistic Regression offered clearer interpretability, both methods highlighted the importance of addressing class imbalance to improve sensitivity toward underrepresented accreditation categories. The findings imply that accreditation assessments can be strengthened by integrating data driven models that prioritize recall, ensuring that schools with lower quality are accurately identified rather than overlooked. For the Ministry of Education and Culture, this provides evidence to support the use of national assessment results as a basis for accreditation monitoring and targeted interventions. For schools, the results underscore the need to improve literacy and numeracy competencies

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 1, December 2025 | Hal 491-501
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i1.861

as a pathway to higher accreditation, thereby aligning classroom practices with broader educational quality goals. Future research can explore alternative imbalance handling techniques and cost-sensitive approaches to further refine predictive accuracy while maintaining fairness across accreditation levels

REFERENCES

[1] T. Fajartriani, A. Habibi, D. Rosalina, and W. Karsiwan, "PERAN PENDIDIKAN DALAM MENINGKATKAN KUALITAS SUMBER DAYA MANUSIA," *Jurnal Administrasi Pendidikan*, vol. 6, no. 1, pp. 9–17, Jul. 2024, Accessed: Dec. 28, 2025. [Online]. Available: https://journal.umbogorraya.ac.id/index.php/jealo/article/view/318

[2] A. Nabila, Raihani, Yuliani, Aslamiah, and C. Cinantya, "AKREDITASI SEKOLAH SEBAGAI JAMINAN MUTU PENDIDIKAN PADA SMA NEGERI 1 DAN 2 PARINGIN," *Satya Widya*, vol. 41, no. 1, p. 2025, Jun. 2025, Accessed: Dec. 27, 2025. [Online]. Available: https://ejournal.uksw.edu/satyawidya

[3] Abdul Malik *et al.*, *PEDOMAN AKREDITASI SEKOLAH/MADRASAH*, 1st ed. Jakarta Selatan: BADAN AKREDITASI NASIONAL SEKOLAH/MADRASAH, 2021. Accessed: Dec. 28, 2025. [Online]. Available: https://infomadrasah.net/?p=677

[4] A. Nursyifa and S. Masyithoh, "ANALISIS HUBUNGAN LITERASI NUMERASI DAN HASIL BELAJAR SISWA," *Jurnal Pendidikan Dasar dan Keguruan*, vol. 8, 2023, doi: https://doi.org/10.47435/jpdk.v8i1.1798.

[5] L. Darmastuti, Meiliasari, and W. Rahayu, "Kemampuan Literasi Numerasi: Materi, Kondisi Siswa, dan Pendekatan Pembelajarannya," *Jurnal Riset Pembelajaran Matematika Sekolah*, vol. 8, no. 1, 2024, doi: https://doi.org/10.21009/jrpms.081.03.

[6] Tenny, A. Khairun Nisa, and Murtaplah, "Pengembangan Literasi dan Numerasi dalam Proses Belajar dan Mengajar Berbagai Mata Pelajaran," Jakarta Selatan, 2021. Accessed: Nov. 08, 2025. [Online]. Available: https://sma.dikdasmen.go.id/data/files/buku/Pengembangan_literasi_dan_numerasi_dalam_proses_belajar.pdf

[7] B. Susetyo and H. Muksin, "REFORMASI AKREDITASI SEKOLAH/MADRASAH: PENDEKATAN MODEL PREDIKSI," *Jurnal Pendidikan dan Kebudayaan*, vol. 7, no. 2, pp. 13–24, Jun. 2022, doi: https://doi.org/10.24832/jpnk.v7i1.2423.

[8] A. Ridho, "PENGUATAN AKREDITASI SEKOLAH/MADRASAH MELALUI PENGEMBANGAN MODEL PREDIKSI: PENDEKATAN INOVATIF UNTUK MENJAMIN MUTU PENDIDIKAN," vol. 2, no. 2, pp. 3025–7425, 2024, doi: https://doi.org/10.62288/creativity.v2i2.23.

[9] H. Karwono and B. Susetyo, "PETA MUTU SATUAN PENDIDIKAN DI INDONESIA (Studi Pilotting Project akreditasi 2020)," *Jurnal Penelitian Kebijakan Pendidikan*, vol. 14, no. 1, pp. 1–10, Aug. 2021, doi: 10.24832/jpkp.v14i1.434.

[10] M. Radhi, D. Ryan Hamonangan Sitompul, S. Hamonangan Sinurat, and E. Indra, "ANALISIS BIG DATA DENGAN METODE EXPLORATORY DATA ANALYSIS (EDA) DAN METODE VISUALISASI MENGGUNAKAN JUPYTER NOTEBOOK," *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, vol. 4, no. 2, 2021, doi: https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v4i2.2475.

[11] Y. Widyaningsih, G. P. Arum, and K. Prawira, "Application of k-fold cross validation in determining the best negative binomial regression model," *Barekeng*, vol. 15, no. 2, pp. 315–322, Jun. 2021, doi: 10.30598/barekengvol15iss2pp315-322.

[12] R. Dwi Fitriani, H. Yasin, D. Statistika, and F. Sains dan Matematika, "PENANGANAN KLASIFIKASI KELAS DATA TIDAK SEIMBANG DENGAN RANDOM OVERSAMPLING PADA NAIVE BAYES (Studi Kasus: Status Peserta KB IUD di Kabupaten Kendal)," vol. 10, no. 1, pp. 11–20, 2021, doi: https://doi.org/10.14710/j.gauss.v10i1.30243.

[13] C. A. Ferrer and E. Aragón, "Note on 'A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance,'" *Inf Sci (N Y)*, vol. 630, pp. 322–324, Jun. 2023, doi: 10.1016/j.ins.2022.10.005.

[14] M. Zhu *et al.*, "Class weights random forest algorithm for processing class imbalanced medical data," *IEEE Access*, vol. 6, pp. 4641–4652, Jan. 2018, doi: 10.1109/ACCESS.2018.2789428.

[15] E. Setiawan, B. Sartono, and K. A. Notodiputro, "SMOTE and Weighted Random Forest for Classification of Areas Based on Health Problems in Java," 2025. doi: https://doi.org/10.30871/jaic.v9i4.9933.

[16] B. Sartono and H. Dharmawan, *Pemodelan Prediktif Berbasis Pohon Klasifikasi*. Bogor: PT Penerbit IPB Press, 2023.

[17] M. Fathurahman, "Regresi Logistik Ordinal untuk Memodelkan Predikat Lulusan Perguran Tinggi," *Jurnal Statistika dan Aplikasinya*, vol. 7, no. 2, 2023, doi: https://doi.org/10.21009/jsa.07201.

[18] M. E. Lelisho, A. A. Wogi, and S. A. Tareke, "Ordinal Logistic Regression Analysis in Determining Factors Associated with Socioeconomic Status of Household in Tepi Town, Southwest Ethiopia," *Scientific World Journal*, vol. 2022, 2022, doi: 10.1155/2022/2415692.

[19] R. Bintang Vollencia and H. Sugiarti, "ANALISIS REGRESI LOGISTIK ORDINAL UNTUK MENENTUKAN FAKTOR-FAKTOR YANG MEMENGARUHI PERFORMA BELAJAR," *Prosiding Seminar Nasional Sains dan Teknologi Seri III Fakultas Sains dan Teknologi*, vol. 2, no. 1, 2025.

[20] N. T. Ujianto, Gunawan, H. Fadillah, A. P. Fanti, A. D. Saputra, and I. G. Ramadhan, "Penerapan algoritma K-Nearest Neighbors (KNN) untuk klasifikasi citra medis," *IT-Explore: Jurnal Penerapan Teknologi Informasi dan Komunikasi*, vol. 4, no. 1, pp. 33–43, Feb. 2025, doi: 10.24246/itexplore.v4i1.2025.pp33-43.

[21] R. C. Chen, C. Dewi, S. W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *J Big Data*, vol. 7, no. 1, Dec. 2020, doi: 10.1186/s40537-020-00327-4.

[22] K. N. Khikmah, B. Sartono, B. Susetyo, and G. A. Dito, "Performance Comparative Study of Machine Learning Classification Algorithms for Food Insecurity Experience by Households in West Java," *Jurnal Online Informatika*, vol. 9, no. 1, pp. 128–137, Apr. 2024, doi: 10.15575/join.v9i1.1012.

[23] Kementerian Pendidikan Kebudayaan Riset dan Teknologi Republik Indonesia, "Keputusan Menteri Nomor 246/O/2024 tentang Instrumen Akreditasi Pendidikan Anak Usia Dini, Pendidikan Dasar, dan Pendidikan Menengah," 2024.