Binary Logistic Regression Model of Parental Interest in Islamic Boarding Schools with R Program: A Case Study Islamic Boarding Schools Tahfidz Daarul Qur'an Tangerang

Gerry Sastro¹, Abdullah Syafiih²*, Ilmadi³

¹-³Department of Mathematics, Universitas Pamulang, South Tangerang, Indonesia

Abstract
This paper aimed to determine factors that affect significantly of parental interest in Islamic Boarding Schools. The subjects of the study were parents of students at Tahfidz Daarul Quran, an Islamic boarding school in Tangerang. Binary logistic regression model was used to analyze the association between parental interest and some covariates namely volition, needs, motivation, knowledge and school environment with used software R. The results showed that a binary logistic model combining knowledge and school environment provides the best model as the significance values of 5.02e-05 and 1.53e-04 are still 0.05 smaller and the wald values are 16.440 and 14.335 which are greater of 3.841. The coefficient estimation of the knowledge variable is 0.246 with an odds-ratio value exp (0.246) = 1.27. This means that parents who have a knowledge of Islamic boarding schools of 1.27 times more interested than do not have a knowledge. Furthermore, the coefficient estimation of school environment is 0.234 with an odds ratio value exp (0.234) = 1.26. It means that parents are more interested 1.26 times in a good school environment than in a bad school environment. So, knowledge and school environment are factors that greatly influence parents' concerns when choosing an Islamic boarding school.

Keywords: Binary Logistic Regression; Islamic Boarding School; Knowledge; Parental Interest; School Environment

Introduction
The progress of a country is greatly determined by the quality of education received by its citizens. Education is the primary foundation in shaping individuals' knowledge and skills, which in turn contributes to the overall advancement and well-being of society (Ilham, 2019; Solikah, 2019; Hafshah & Nugraheni, 2024). With good education, a person can acquire the knowledge and skills necessary to become a better and more productive individual (Bowen & Rieckmann, 2018; Madani, 2019; Hanushek, 2020; Nofika, 2021). Education also serves as a crucial means in helping individuals develop to face various challenges and changes in life with the right attitude (Hilton & Pellegrino; Shor, 2012) (Prihanto et al., 2013). In this context, the role of education is vital in creating an intelligent, innovative, and highly competitive society (Lee & Trimi, 2018; Lase, 2019; Sanga & Wangdra, Puspa et al., 2023).

Every parent certainly wants their children to grow up to be independent individuals, with noble character, a broad vision, and to be beneficial to society, religion, and the country in the future (Mitchall & Jaeger, Hill et al., 2018; Sukatin et al., 2020). Therefore, it is important for parents to guide their children towards good education from an early age (Hasanah, 2018; Puspytasari, 2022). Proper education not only provides general knowledge but also shapes the character and morality of the children (Chowdhury, 2018; Hirsch, 2019; Birhan et al., 2021). In this modern era, with many challenges such as drug abuse, fighting, alcohol consumption, sexual harassment, and even murder committed by young individuals, parents are increasingly worried about their children's future (Lickona, 2022; Muslich, 2022). This concern drives parents to be more selective in choosing educational institutions that not only provide knowledge but also nurture morals and ethics.

For these reasons, many parents are now considering pesantren as an alternative education for their children. Pesantren are believed to provide holistic education, encompassing the teaching of general knowledge as well as the
instillation of moral and religious values (Velasufah & Fahham, 2020; Silfiyasari & Zhafi, 2020; Astuti et al., 2023). With an environment conducive to learning and character development, pesantren have become a choice for parents who want their children to be not only intellectually intelligent but also possess good character (Rukhayati, 2019; Suparman et al., 2020; Tolchah, 2020). This trend shows that education is not only viewed from an academic perspective but also from how it can shape children's personality and morality, preparing them to face various challenges in the future with integrity and responsibility (Raharjo et al., 2023).

One of the very famous Islamic boarding schools in Tangerang city is Tahfidz Daarul Quran (DAQU) Islamic Boarding School Tangerang. DAQU is an Islamic boarding school under the leadership of famous cleric Ustad Yusuf Mansur. This Islamic boarding school not only provides formal education but also provides religious education and prints thousands of Hafidz Qurans in Indonesia. However, the data on the number of new students obtained in the past 4 years has decreased, specifically in 2020 there are 256 students, in 2021 there are 234 students, in 2022 there are 155 students and in 2023 there are 116 students, as can be seen in Figure 1 (Pesantren DAQU, 2023).

This problem needs to be resolved by investigating the causes, including the factors that influence parents’ interest in choosing Islamic boarding schools. To find these factors, one of the frequently used methods is binary logistic regression. According to Patel (2021), binary logistic regression is a method of measuring the relationship between a target variable with only two types of possibilities (binary) and one or more independent variables that help identify important factors and the nature of their relationship affects the target variable. Additionally, binary logistic regression is a method of analyzing factors that influence the dependent variable or dependent variable. These factors can be categorical or numeric data and the dependent variable is binary or dichotomous (Nurdiansah & Khikmah, 2020).

Several related studies aimed at identifying influencing factors using binary logistic regression include the study by Pratiwi & Dewi (2021) which used binary logistic regression to identify influencing factors on Customer satisfaction with the pick-up service in charge, which results in price and service significant impact on customer satisfaction with a ranking result of 92%. Annas et al. (2022), examined the significant factors for stroke type in stroke patients at Dadi Regional Hospital, Makassar City using binary logistic regression, in which the results showed that factors influencing Effects are blood sugar levels and previous medical history. The overall goal of this study is to compare different combination models and identify factors that significantly influence of parental interest in Islamic Boarding Schools.

From this explanation, it goes without saying that the binary logistic regression method can be applied to this problem. Anyone who wants to know what factors influence parents' interest when choosing Tahfidz Daarul Quran Islamic Boarding School in Tangerang. Therefore, the overall objective of this study is to compare different combination models and identify the factors that significantly influence parents' interest when choosing an Islamic boarding school.

Method

The data on expressed interest were obtained from parents of students enrolled at Tahfidz Darul Quran, an Islamic boarding school located in Tangerang, Banten Province. The data include information on the parents of children in classes X, XI, and XII. The survey consists of 29 questions. Consequently, the final 220 participants provided
responses in accordance with the required sample size determined by the Slovin formula. The collected data included violations, needs, motivation, knowledge, and school atmosphere. The dependent variable Y, representing the level of interest of parents, can be dichotomized into two categories: non-interest (0) and interest (1).

The method used in this study is binary logistic regression. Binary logistic regression is a method in regression analysis used when the dependent variable is binary or categorical with two values, such as 0 and 1 (Nick & Campbell, 2007; Hilbe, 2011; Cokluk, 2010). The primary goal of binary logistic regression is to model the probability that the dependent variable will take a value of 1 based on one or more independent variables (Bonney, 1987). In this study, it is used to investigate the correlation between parental interest and five covariates: willingness (X1), need (X2), motivation (X3), knowledge (X4), and school environment (X5). This model allows categorizing the outcomes of the variable of interest (varieties of parental interest) into two categories: non-interest and interest.

The binary logistic regression model is typically expressed as follows:

$$\pi(x_p) = P(Y|x) = \frac{e^{(\beta_0 + \beta_1x_1 + \ldots + \beta_px_p)}}{1 + e^{(\beta_0 + \beta_1x_1 + \ldots + \beta_px_p)}}$$  \hspace{1cm} (1)

Here, $\pi(x_p)$ is the likelihood of the binary outcome being present (Anugrawati et al., 2023). The logit function is a linearizing transformation derived from Equation 1. It is defined as follows:

$$g(x) = \beta_0 + \beta_1x_1 + \ldots + \beta_px_p + \sum_{i=1}^{i=p} \beta_ix_i$$  \hspace{1cm} (2)

The function $g(x)$ represents the logit, which is often referred to as the log chances of the outcome variable. The term $\beta_0$ represents the intercept, which indicates the estimated log chances or logit of the response variable when the independent variables of the model are evaluated at zero. In binary logistic regression, there is a single intercept estimate and i independent variables. The $\beta_1, \beta_2, \ldots, \beta_i$ are the coefficients used in logistic regression (Speelman, 2014; Pituch & Stevens, 2015). The term "Exp($\beta_i$)" represents the odds ratio of an explanatory variable, which is equivalent to the exponential function of the natural logarithm base e increased to the power of $\beta$ (Maroof, 2012). The odds ratio of an explanatory variable represents the magnitude by which the explanatory variable positively or negatively affects the logarithm of the probabilities of the dependent variable. Binary logistic regression models were executed using R version 3.6.1 (R Core Team, 2022). The analysis conducted in this study begins by performing parameter significance tests simultaneously using the G test for both dependent and independent variables, followed by inferential statistics. The best model was chosen by evaluating it using the Wald Test. The Wald test is a statistical test used to determine the suitability of include an independent variable (predictor) in a model (Juhl, 2021), as indicated by Equation 3.

$$W = \frac{\hat{\beta}^2}{se(\hat{\beta})}$$  \hspace{1cm} (3)

Where $\hat{\beta}$ is the estimator of $\beta$ and $se(\hat{\beta})$ is the standard error estimator of $\beta$. If $H_0$ is valid, the W statistic will adhere to the standard normal distribution. If $|W| > Z_{0.025}$, decision $H_1$ is rejected.

In the chosen model, variables were deemed significant if their p-value was less than or equal to 0.05. For a more comprehensive breakdown, the steps are as follows:

1. Correlation analysis was conducted on data collected through the distribution of questionnaires to guardians of students at the Tahfidz Daarul Qur'an Islamic Boarding School in Tangerang;
2. Determining the values of model parameters;
3. Perform parameter significance checks concurrently using the G test (Eq.4);
4. Conduct a Wald test to determine the statistical significance of certain parameters;
5. Employing a model's appropriateness test, the Chi-square value (Eq.5) and Hosmer-Lemeshow test (Eq.6) are used at a significance level of 5% to assess the model's fit to the measurements;
6. Analyze the model using Odds comparison.
Where $L_0$ is a model composed solely of constants, while $L_p$ is a comprehensive model that includes independent variables and follows a chi-square distribution (Fadjryani et al., 2022).

$$G = -2 \ln \frac{L_0}{L_p}$$  \hspace{1cm} (4)

$x^2 = \sum \frac{(fo - fe)^2}{fe}$  \hspace{1cm} (5)

Where $x^2$ represents the chi-square value. The observed frequency, denoted as " $fo$ ", represents the empirical frequency, while the predicted frequency, denoted as " $fe$ ", represents the theoretical frequency (Yuhadisi & Suliadi, 2021).

$$\hat{C} = \sum_{r} \frac{(c_r - n_r \bar{p}_r)^2}{n_r \bar{p}_r (1 - \bar{p}_r)}$$  \hspace{1cm} (6)

Where $g$ is the quantity of groupings. $n_r$ represents the total count of observations within the $r$ group. $C_r$ represents the $Y$ value in the $r$ group, while $\bar{p}_r$ represents the average predicted success probability of the same group.

Results and Discussion

Result

The data used in this study is primary data obtained from questionnaire surveys. This questionnaire was completed by 220 parents or guardians of students at High School Pesantren Tahfidz Daarul Qur'an Tangerang, serving as the sample. Prior to distributing the questionnaire, the researcher conducted validity and reliability tests for each item of the questions.

![Figure 2. Data on new students registering at Tahfidz Daarul Quran Islamic Boarding School from 2020-2023.](image-url)

The results of this research consist of several testing assessments as follows:

1. Simultaneous Test

Simultaneous testing is conducted to examine the influence of predictor variables on the response variable.

$H_0$: All predictor variables have no effect on the response variable.

$H_1$: There is at least one predictor variable that does not have an effect on the response variable.

Based on the simultaneous test results using the Likelihood Ratio test statistic as follows:

<table>
<thead>
<tr>
<th>Table 1. Likelihood Ratio Test (Test G).</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 Log likelihood ($G_2$)</td>
</tr>
<tr>
<td>76,07</td>
</tr>
</tbody>
</table>
The obtained value of G is 76.07, the McFadden coefficient is 0.29, and the Cox & Snell R square (r2ML) is also 0.29. Given \( \alpha = 5\% \), the value of \( \chi^2(\alpha,\nu) \) is obtained as 15.51, and since the value of G is greater than \( \chi^2(\alpha,\nu) \), \( H_0 \) is rejected, indicating that there is at least one independent variable that has a simultaneous effect on the dependent variable at \( \alpha = 0.05 \). Furthermore, it is known that the Nagelker R Square value (r2CU) or the coefficient of determination of logistic regression is 0.422, indicating that the contribution of independent variables to the dependent variable simultaneously is 42.2%.

2. Partial Test

Partial testing is conducted with the aim of determining the individual relationship between independent variables and the dependent variable. The hypotheses used in individual testing are:

- \( H_0 \): The k-th predictor variable does not have a significant effect on the response variable.
- \( H_1 \): The predictor variable at position k has a significant effect on the response variable.

Next, the results of the Wald test are presented in the following Table 2.

<table>
<thead>
<tr>
<th>Variabel</th>
<th>Coefficient (B)</th>
<th>Wald Test</th>
<th>P-value (Significance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constanta</td>
<td>-8,35999</td>
<td>22,182</td>
<td>2,48e-06</td>
</tr>
<tr>
<td>Volitions (X1)</td>
<td>0,02708</td>
<td>0,123</td>
<td>0,7256</td>
</tr>
<tr>
<td>Needs (X2)</td>
<td>0,13279</td>
<td>2,869</td>
<td>0,0903</td>
</tr>
<tr>
<td>Motivation (X3)</td>
<td>-0,23821</td>
<td>3,546</td>
<td>0,0585</td>
</tr>
<tr>
<td>Knowledge (X4)</td>
<td>0,26644</td>
<td>16,229</td>
<td>5,61e-05</td>
</tr>
<tr>
<td>School environment (X5)</td>
<td>0,25899</td>
<td>15,562</td>
<td>7,98e-05</td>
</tr>
</tbody>
</table>

Based on the Wald test in Table 2, the statistical test values for the variables of volitions (X1), need (X2), and motivation (X3) are smaller than 3.841, and the significance value (p-value) is greater than 0.05, indicating that they are not rejected. Meanwhile, the knowledge variable (X4) and the school environment variable (X5) have significance values smaller than 0.05, therefore rejecting \( H_0 \).

Table 3. The results of the partial test were reduced when the predictor had no effect

<table>
<thead>
<tr>
<th>Variabel</th>
<th>Coefficient (B)</th>
<th>Wald Test</th>
<th>P-value (Significance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constanta</td>
<td>-8,05979</td>
<td>35,779</td>
<td>2,21e-09</td>
</tr>
<tr>
<td>Knowledge (X4)</td>
<td>0,24587</td>
<td>16,440</td>
<td>5,02e-05</td>
</tr>
<tr>
<td>School environment (X5)</td>
<td>0,23464</td>
<td>14,335</td>
<td>1,53e-04</td>
</tr>
</tbody>
</table>

Based on Table 3, the process of predicting the variable involves removing insignificant variables and only including the influential variables in the new model, which consists of the independent variables of knowledge (X4) and school environment (X5). The Wald test result is greater than the chi-square test and the significance value is still less than 0.05. The meaning of the statement is that both variables, namely Knowledge and School Environment, have a significant influence on the Parents' Interest in Daarul Qur'an Tahfidz Boarding School in Tangerang.

In order to obtain the optimal logistic regression model, the variables included in the model should have a significance value of less than 0.05. These variables should have a direct impact on the timely completion of graduation. Based on the estimated coefficients, the following binary logistic regression model is obtained:

\[
g(x) = \beta_0 + \beta_4 X_4 + \beta_5 X_5
= -8,060 + 0,246X_4 + 0,23464X_5
\]

3. Goodness of fit Test

The model's goodness of fit is assessed to determine whether it is suitable for the data and meets the criteria for Goodness of Fit. The adequacy of this model can be assessed using the Hosmer and Lemeshow statistical test, with the following hypotheses:
H₀: The hypothesized model is a good fit for the data (there is no difference between the model and the data).
H₁: The hypothesized model does not fit the data (there is a discrepancy between the hypothesis and the data).

Table 4. Results of the overall partial test of predictor variables.

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Chi-Square</th>
<th>p-value (Sign. Value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9,455</td>
<td>0.3054</td>
</tr>
</tbody>
</table>

From table 4, it can be inferred that the p-value (sig.) obtained is 0.3054, which is greater than 0.05. Therefore, the decision made is to accept H₀, meaning that the binary logistic regression model generated is consistent with the data.

4. Model Interpretation Using Odds Ratios

The interpretation of the model using odds ratios can be done by examining the value of Exp (B) as presented in Table 5.

Table 5. Odds Ratio Value

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient (B)</th>
<th>Odds Ratio (Exp(B))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constanta</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Knowledge (X₄)</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>School environment (X₅)</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

It can be observed that the magnitude of the tendency of each independent variable is as follows:

1. Knowledge (X₄): The presence of knowledge input for children at Pesantren Tahfidz Daarul Qur’an will influence parents' interest in the Pesantren by 1,278 times more compared to the absence of knowledge,
2. School environment (X₅): The presence of a favorable school environment at the Tahfidz Daarul Qur’an Islamic Boarding School will influence parental interest in the boarding school by 1,264 times more than a poor school environment.

Discussion

Based on Figure 2, the research conducted on the guardians of the students at Tahfidz Daarul Qur’an Islamic Boarding School in Tangerang, with a sample of 220 respondents, showed that 159 people or 72% chose interest and 61 people or 28% chose disinterest. The variables tested in this study are volitions (X₁), needs (X₂), motivation (X₃), knowledge (X₄), and school environment (X₅). Table 3 demonstrates that both knowledge (X₄) and school environment (X₅) collectively exert a significant influence on parental interest in Pesantren. Based on the partial analysis results in Table 3, after reducing the non-significant predictor variables, the estimated parameter value for the binary logistic regression equation is obtained as -8.060.

Conclusions

Based on the findings of this study, several conclusions can be drawn:

1. The obtained binary logistic regression model for the interest of guardians of Pesantren Tahfidz Daarul Qur’an Tangerang is:
   \[ g(X) = -8.060 + 0.246X₄ + 0.2346X₅ \]
2. The significant factors influencing parents' interest in choosing the Tahfidz Daarul Qur’an Islamic Boarding School in Tangerang are knowledge(X₄) and the school environment(X₅).

References


