BAREKENG: Journal of Mathematics and Its Applications
June 2026 Volume 20 Issue 2 Page 1199–1214
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss2pp1199-1214

TREE-BASED MIXED EFFECTS MODELING OF TEACHER CERTIFICATION OUTCOMES IN MADRASAH ALIYAH: A COMPARATIVE STUDY OF GLMM TREES AND GMET

Dodi Irawan Syarip[1*], **Khairil Anwar Notodiputro**[2], **Bagus Sartono**[3]

[1,2,3] *Statistics and Data Science, School of Data Science, Mathematics and Informatics, IPB University*
Jln. Meranti, Kampus IPB Dramaga, Bogor, Jawa Barat, 16680, Indonesia

[1]*Directorate General of Islamic Education, Ministry of Religious Affairs*
Jln. Lapangan Banteng Barat No. 3-4, Jakarta, 10710, Indonesia

Corresponding author's e-mail: * **dodi.irawan.syarip@apps.ipb.ac.id**

Article Info

Article History:
Received: 6th May 2025
Revised: 28th July 2025
Accepted: 27th September 2025
Available online: 26th January 2026

Keywords:
GLMM trees;
GMET;
PPG program.

ABSTRACT

The Teacher Professional Education program, or "Pendidikan Profesi Guru" (PPG), is a continuing education program designed for prospective or in-service teachers to obtain a teaching certificate. PPG is a priority program of the Ministry of Religious Affairs in providing competent and professional madrasah teachers. This study is expected to identify the challenges encountered in the implementation of the Madrasah teacher certification program and provide valuable input to enhance the success rate of Madrasah Aliyah teachers in the PPG program. The main objective of this study is to find the most appropriate tree-based mixed effects model to analyze the effectiveness of PPG for Madrasah Aliyah teachers in 2022. This study applies two tree-based mixed effects modeling methods: generalized linear mixed model trees (GLMM trees) and generalized mixed effects trees (GMET). Both methods model variability across subjects as a random effect. Based on the performance indices measurement results, the GMET model shows superiority over the GLMM trees model. The GMET model has an accuracy index of 0.7653, higher than the GLMM trees model of 0.7306. Substantively, teachers of English and Indonesian Language exhibit higher probabilities of passing than those of other subjects, whereas Arabic and Islamic Cultural History have the lowest estimated probabilities of success. Analysis of the variable importance from both models indicates that teachers' age is the most influential predictor of PPG graduation among Madrasah Aliyah teachers. Based on these findings, to improve the effectiveness of PPG implementation for madrasah Aliyah teachers, policymakers at the Ministry of Religious Affairs are advised to implement a structured coaching and mentoring program for prospective PPG participants, with a special emphasis on support for senior teachers specializing in Arabic and Islamic Cultural History.

How to cite this article:

D. I. Syarip, K. A. Notodiputro, and B. Sartono, "TREE-BASED MIXED EFFECTS MODELING OF TEACHER CERTIFICATION OUTCOMES IN MADRASAH ALIYAH: A COMPARATIVE STUDY OF GLMM TREES AND GMET", *BAREKENG: J. Math. & App.*, vol. 20, no. 2, pp. 1199-1214, Jun, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · **Open Access**

1. INTRODUCTION

Generalized linear mixed models (GLMMs) are a statistical model that combines elements of both generalized linear models (GLMs) and linear mixed models (LMMs) [1]. It allows for the analysis of data with hierarchical or correlated structures, such as longitudinal data or data with repeated measurements on the same subjects. GLMMs integrate the flexibility of GLMs in handling non-normal distributions with the capability of mixed models in addressing hierarchical or correlated structures in the data. GLMMs estimate fixed and random effects and are especially useful when the response variable is binary, ordinal, count, or quantitative but not normally distributed [2]. GLMMs are employed to fit multilevel models for binary response variables, while constraining the covariates to exert linear effects on a transformed scale of the response variable [3]. In a GLMM, random effects can be modeled to account for variation between groups or subjects, while fixed effects represent the influence of predictor variables. Because GLMM is a combination of GLM and LMM, the GLMM component consists of the response variable (Y), the independent variable coefficient (β), the predictor variable (X), random effect (v), and model error (ε) [4].

Tree-based models find subgroups in different data regarding model parameters [5]. Tree-based models are useful for handling plenty of prospective predictor variables and automatically detecting relations among them [6]. The GLMM trees method is a tree-based algorithm developed to find relationships and non-linearities in GLMMs [6]. GLMM trees allow for the identification of treatment-subgroup interactions by considering the cluster structure of the dataset. This technique uses GLMM to estimate random effect parameters and model-based recursive partitioning to find treatment-subgroup interactions [6]. Model-based recursive partitioning performs automatic detection of treatment subgroups identified by predictive factors [7]. By considering potential relationships between observations in multilevel and/or longitudinal data sets, this approach extends the GLM trees algorithm. GLMM trees were originally developed for clustered cross-sectional data, but nowadays this method is also used for longitudinal data [8].

The application of GLMM trees to educational data has been conducted by [9], particularly in exploring potential variations in student learning outcomes on the 9th Grade On-Track to Graduation (9G-OTG) indicator and high school graduation rates. Other studies applying GLMM trees to health datasets include those conducted by [10], [11]. The application for socio-economic datasets has been carried out by [12] who examined poverty in Indonesia and [13] to model the employment status of residents of Bogor Regency and Pangandaran Regency, West Java Province, using the GLMM trees approach.

The application of tree-based mixed models to various classes of response variables in the exponential family is extended by the generalized mixed-effect trees (GMET) method [14]. This method could handle clustered data structures as GLMM does. In the GMET model, the response variable Y from the exponential family distribution makes up the random component. The fixed component of the GMET model is non-linear and is substituted by a function $f(X_i)$ estimated through a tree-based algorithm. The GMET method was first developed and applied to educational data by [14] to simulate dropout rates of students in a variety of bachelor's degree programs at Politecnico di Milano. The results show that GMET outperforms CART when there is a random effect. Moreover, GMET has been applied in research on the temporary unemployment rate in West Java [15] and a study on the classification of household poverty in West Java [16].

In this study, the author empirically analyzes the Madrasah Aliyah teacher certification obtained through the Teacher Professional Education program, or "*Pendidikan Profesi Guru*" (PPG), by applying the GLMM trees and GMET methods. These methods are particularly well-suited for the structure of the data, which involves hierarchical or grouped observations [6], [14]—such as teachers nested within different subject areas. Both methods enable the accommodation of variability across subjects through the inclusion of random effects. By treating subject area as a random effect, GLMM trees and GMET allow for the modeling of unobserved heterogeneity among different groups, accounting for the possibility that certification outcomes may systematically vary by subject. Additionally, both methods combine the strengths of mixed-effects modeling with decision tree algorithms, making them capable of capturing complex interactions and nonlinear relationships between predictors and certification outcomes. This analytical approach is expected to yield more robust and interpretable insights into the factors influencing teacher certification success, while also revealing whether certain predictors have different effects across subject groups.

The teacher certification program is one of the government's priority programs following the issuance of Law No. 14/2005 on Teachers and Lecturers. Between 2007 and 2017, teacher certification was conducted through the Teacher Professional Education and Training program, or "*Pendidikan dan Latihan Profesi*

Guru" (PLPG). However, the PPG program has been used to provide teacher certification since 2018. The teacher certification program continues to undergo improvements from year to year. These improvements include changes in policies, procedures, mechanisms for determining participants, implementing institutions, and certification patterns used [17].

The PPG program is an initiative program designed to provide S1/D-IV graduates with the necessary skills and motivation to become professional teachers by teaching the Teacher Education Standards. The objective of this program is to develop teachers who become professional educators, committed to God Almighty, have noble character, think critically, creatively, innovative, and competitively, with the main tasks of educating, teaching, guiding, directing, training, assessing, and evaluating students.

Based on SIMPATIKA, a management information system for data on madrasah teachers and education personnel under the Ministry of Religious Affairs, the number of madrasah teachers in 2023 is 793,174 teachers, with details: 102,372 or 12.91% of Raudhatul Athfal (RA) teachers, 285,954 or 36.05% of Madrasah Ibtidaiyah (MI) teachers, 264,195 or 33.31% of Madrasah Tsanawiyah (MTs) teachers, and 140,653 or 17.73% of Madrasah Aliyah (MA) teachers, as listed in Table 1.

Table 1. Number of Madrasah Teachers in 2023

Madrasah Level	Certified		Not Certified		Total
	Number	%	Number	%	
RA	28,869	28.20	73,503	71.80	102,372
MI	125,680	43.95	160,274	56.05	285,954
MTs	103,181	39.05	161,014	60.95	264,195
MA	47,858	34.03	92,795	65.97	140,653
Total	305,588	38.53	487,586	61.47	793,174

Data source: Ministry of Religious Affairs

In addition, Table 1 shows that out of 793,174 madrasah teachers throughout Indonesia, only 38.53% or 305,588 teachers already have a professional education certificate. The remaining 61.47% or 487,586 teachers have not passed the certification process. Specifically for teachers at the Madrasah Aliyah level, out of 140,653 teachers, only 34.03%, or 47,858 teachers, have passed the certification, and the remaining 92,795 teachers, or 65.97%, have not been certified as professional educators. The substantial proportion of uncertified madrasah teachers is a major challenge for the Directorate General of Islamic Education (DGIE), Ministry of Religious Affairs (MoRA), which is responsible for facilitating and providing guidance to madrasah teachers, especially those who have not been certified to participate in the PPG program. Thus, these teachers will have adequate competence as professional teachers in the future.

Competent and professional teachers will contribute greatly to improving the quality of teaching and learning activities. Therefore, teachers are required to improve their competence constantly [17]. Teacher professional competence is a set of competencies related to a profession that requires various expertise in the field of education or teaching [18]. For this reason, it is necessary to conduct a study to find methods to analyze the effectiveness of the implementation of teacher certification through the PPG program in producing professional madrasah teachers.

The main objective of this study is to find a better tree-based mixed effects model to analyze data on Madrasah Aliyah teachers who participated in the 2022 PPG program and provide recommendations to the Ministry of Religious Affairs, as policymakers, to improve the effectiveness of the implementation of the PPG program for Madrasah Aliyah teachers.

2. RESEARCH METHODS

2.1 Observation Data

The data that will be analyzed in this study are data on the results of the implementation of the PPG program for Madrasah Aliyah teachers in 2022. This data is secondary data sourced from the Directorate of Madrasah Teachers and Education Personnel, Directorate General of Islamic Education, Ministry of Religious Affairs.

Observation data totaled 2,451 teachers, consisting of 52.55% or 1,288 males and 47.45% or 1,163 females, as presented in Table 2. Based on the subject, Fiqh teachers have the highest number of Madrasah Aliyah teachers participating in the 2022 PPG program with 350 teachers, followed by Aqidah Akhlak subject with 311 teachers, then Arabic subject with 293 teachers, Natural Sciences with 292 teachers, Quran Hadith with 282 teachers, Social Sciences with 263 teachers, and English with 158 teachers. Meanwhile, the least number of subject teachers participating in the 2022 PPG program is in Physical and Health Education with 32 teachers, followed by Civic Education with 43 teachers, Indonesian Language with 87 teachers, and "Others" subject with 88 teachers.

Table 2. Total Observation Data

Subjects	Label	Male	Female	Total
Aqidah Akhlak	AA	186	125	311
Arabic	Arab	156	137	293
Civic Education	CE	26	17	43
English	Eng	40	118	158
Fiqh	Fiqh	254	96	350
Indonesian Language	Ind	31	56	87
Islamic Cultural History	ICH	89	51	140
Math	Math	48	64	112
Natural Science	Nat	84	208	292
Physical and Health Education	PHE	28	4	32
Quran Hadith	QH	210	72	282
Social Science	Soc	103	160	263
Others	Oth	33	55	88
Total		1,288	1,163	2,451
%		52.55	47.45	100.00

Furthermore, the variables used in this study are shown in Table 3 as follows:

Table 3. List of Variables in the Study

Variable Name	Label	Variable Type
Response Variable		
PPG Graduation Status	ppg_stat	Factor (Passed, Failed)
Predictor Variables: Fixed Effect		
Teachers' Age	age	Numeric
Teachers' Gender	sex	Factor (Male, Female)
Teachers' Highest Educational Qualification	educ	Factor (S1/D-IV, S2, S3)
Teachers' Employment Status	emp_stat	Factor (PNS, PPPK, Non-PNS)*
Teachers' Length of Service (Years)	los	Numeric
Institution Status of Madrasah	mad_stat	Factor (Public, Private)
Accreditation Status of Madrasah	accre	Factor (A, B, C, Not Accredited)
Predictor Variables: Random Effect		
Subject Group	subject	Factor (13 Subjects)

* PNS: *Pegawai Negeri Sipil* (Civil Servant); Non-PNS: Non-Civil Servant);
PPPK: *Pegawai Pemerintah dengan Perjanjian Kerja* (Government Employee with Working Agreement)

The predictor variable used as a random-effect component in this study is the subject groups. For analysis purposes, some subjects were combined into one group. Chemistry, Physics, and Biology are combined into the "Natural Science" subject group. Then, Economics, Geography, History, and Sociology are grouped into the "Social Science" subject group. Meanwhile, additional subjects, such as Cultural Arts, Information and Communication Technology (ICT), Fashion, and Guidance Counseling, are combined into the "Others" subject group.

2.2 Research Methodology

The data analysis for this study applied two tree-based mixed effects methods: GLMM trees and GMET. As shown in Fig. 1, the methodology in this study consists of the following steps:

1. Data collection: coordinate with the data suppliers.
2. Data exploration: checking for data completeness and data consistency.
3. Data splitting: the observation data is divided into two datasets, with 80% used for training and 20% for testing. This proportion ensures the model has sufficient data to learn underlying patterns while allowing for an unbiased evaluation of its performance on unseen data.
4. Model building using GLMM trees and GMET methods.
5. Model performance indices measurement using a confusion matrix.
6. Evaluate model performance by comparing the accuracy, precision, sensitivity, specificity, and F1 Score indices resulting from step (5).
7. Measurement of variable importance.
8. Data analysis and interpretation.

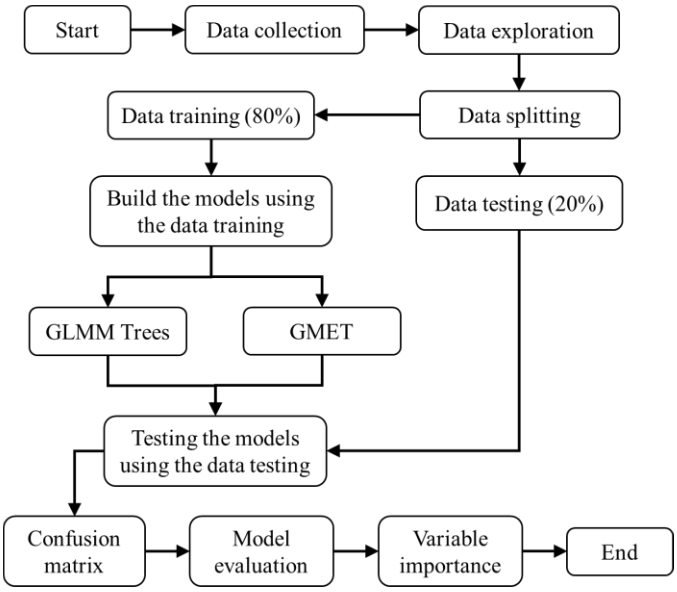

Figure 1. Research Methodology

2.2.1 Model Performance Evaluation

The performance evaluation of the GLMM trees and GMET models in this study used a confusion matrix measurement tool. The confusion matrix evaluates the performance of a classification model based on the accuracy, precision, sensitivity/recall, specificity, and F1 Score indices, which are obtained from the number of correctly and incorrectly predicted objects [19].

True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values shown in Table 4 are utilized to compute the classification model accuracy indicator. True Positive (TP) and True Negative (TN) are conditions when the predicted label matches the actual value, both for positive and negative results. Meanwhile, a False Positive (FP) occurs when the predicted label is positive, but the actual value is negative. On the other hand, it is referred to as a False Negative (FN) if the predicted label is negative even though the actual value is positive. The more TP and TN values generated by a model signify increased model accuracy. A positive event is a condition or event that is considered a “target” or “event to be detected” by the model. Meanwhile, a negative event is a condition or event that is opposite to the target, or the “absence” of a positive event.

Table 4. Two-Classes Confusion Matrix

Predicted Values	Actual Values	
	Positive	**Negative**
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

The following formulas are used to obtain the accuracy, precision, sensitivity/recall, specificity, and F1 Score indices:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\frac{\text{Sensitivity}}{\text{Recall}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{4}$$

$$\text{F1 Score} = \frac{2 \text{ x Precision x Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

2.2.2 GLMM Trees Model

The GLMM tree model constructed is as follows:

$$g(\mu_{ij}) = \boldsymbol{x}_{\boldsymbol{i}}^{\top} \beta_j + \boldsymbol{z}_{\boldsymbol{i}}^{\top} b \tag{6}$$

where $g(.)$ is a link function, $\boldsymbol{x}_{\boldsymbol{i}}$ is the vector of predictors for observation i, β_j is the local parameter of the fixed effects, $\boldsymbol{z}_{\boldsymbol{i}}$ is the vector of predictors associated with random effects, and b is the global parameter of random effects.

The algorithm of the GLMM trees is shown in Fig. 2, as follows [6]:

1. Initial value estimation, where the value of r and all values of $\hat{b}_{(r)}$ are set to 0.
2. Set $r = r + 1$. Estimate a GLM tree model with $\boldsymbol{z}_{\boldsymbol{i}}^{\top} \hat{b}_{(r-1)}$ as an offset.
3. Perform GLMM modelling: $g(\mu_{ij}) = \boldsymbol{x}_{\boldsymbol{i}}^{\top} \beta_j + \boldsymbol{z}_{\boldsymbol{i}}^{\top} b$ with the terminal node $j(r)$ derived from the GLM trees obtained in step (2). Then extract the posterior predicted value of $\hat{b}(r)$.
4. Repeat steps (2) and (3) until converged. Convergence is considered to have been achieved when no changes occur in the tree structure between consecutive iterations.

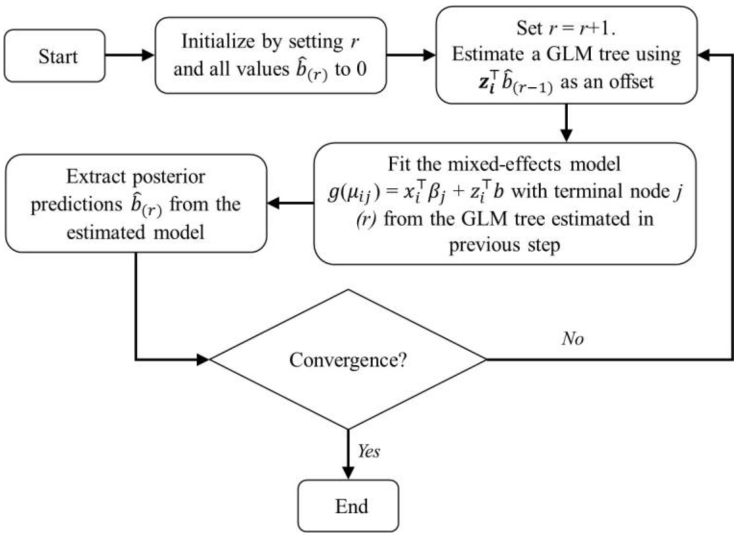

Figure 2. Algorithm of the GLMM Trees

2.2.3 GMET Model

The GMET model matrix is formulated as follows [14]:

$$\begin{aligned} \mu_i &= E[y_i | b_i], \ \ i = 1, 2, \dots, I, \\ g(\mu_i) &= \eta_i, \\ \eta_i &= f(X_i) + Z_i b_i, \\ b_i &\sim N_q(0, \Psi) \ ind. \end{aligned} \tag{7}$$

(Note: *ind.* denotes that b_i and $b_i{}'$ are independent for $i \neq i'$)

where:
i : index of group;
I : a total group number;
n_i : observations number in the group i;
η_i : linear predictor vector of dimension n_i;
$g(.)$: link function
X_i : fixed effects regressors matrix of observations in group i of dimension $n_i \times (p+1)$;
Z_i : regressors matrix of dimension $(n_i \times q)$ for the random effects;
b_i : Z_i components vector of dimension $(q+1)$;
Ψ : within-group covariance matrix of the random effects of dimension $(q \times q)$;
The canonical link function is the logit function in the following scenario: a binary random variable and a univariate random effect.

$$g(\mu_{ij}) = g(p_{ij}) = \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = logit(p_{ij}). \tag{8}$$

The structure of random effects is simplified to a random intercept. Thus, the observation model y_{ij} can be expressed with the following formula:

$$\begin{gathered} Y_{ij} \sim Bernoulli\,(p_{ij}),\ \ i = 1,\ldots,I \quad j = 1,\ldots,n_i\,, \\ p_{ij} = E[Y_{ij}|b_i], \\ \text{logit}(p_{ij}) = f(x_{ij}) + b_i, \\ b_i \sim N(0,\psi) \quad ind. \end{gathered} \tag{9}$$

(Note: *ind.* denotes that b_i and $b_i{}'$ are independent for $i \neq i'$)

where $x_{ij} = (x_{1ij}, \ldots, x_{ijp})^T$ is a vector of fixed-effects covariates of dimension $(p+1)$ for every observation j in group i.

The GMET parameters are estimated using the random effects expectation-maximization (RE-EM) tree approach. This algorithm's fundamental notion is to decipher the estimates of fixed effects and random effects [14]. Fig. 3 presents the algorithm of GMET with the following steps [14]:

1. Initialize the estimated random effects $b_i = 0$.
2. Estimate the target μ_{ij} using GLM. Get estimate $\hat{\mu}_{ij}$ of target variable μ_{ij}.
3. Build a CART regression tree with $\hat{\mu}_{ij}$ as the response variable and x_{ij} as the predictor. Through this regression tree, define a set of indicator variables $I(x_{ij} \in R_\ell)$, for $\ell = 1,\ldots,L$, where $I(x_{ij} \in R_\ell)$ is 1 if observation ij belongs to the terminal node ℓ-th and 0 otherwise.
4. Fit the mixed-effects model using: y_{ij} as the response variable, the set of indicator variables $I(x_{ij} \in R_\ell)$ as fixed-effects covariates (dummy variables), and the random-effects structure $z_{ij}^{\mathrm{T}} b_i$. Extract the estimator of $\hat{b}_i$.
5. Replace the predicted response at each terminal node R_ℓ of the tree with the estimated predicted response $g(\hat{\gamma}_\ell)$ from the mixed-effects model fitted in step (4).

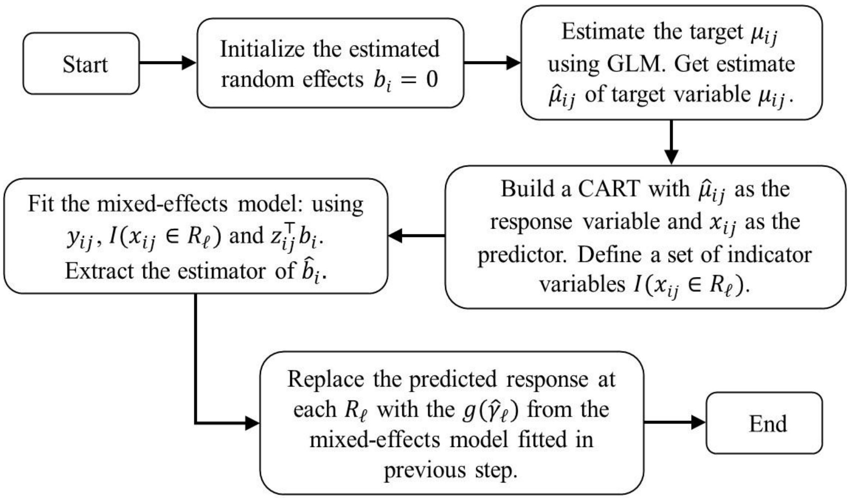

Figure 3. Algorithm of the GMET

At step 2, the GLM is fitted via maximum likelihood estimation, where the parameter estimates are commonly derived using either the iteratively reweighted least squares (IRLS) algorithm or the Newton–Raphson procedure [20]. In step 3, the fitting of the tree can be conducted using any suitable tree algorithm, depending on the chosen tree-growing rules. In this study, tree construction is carried out using the CART algorithm [21].

In step 4, the GLMMs are estimated using the maximum likelihood method. In GLMM, the likelihood function involves integrals that generally cannot be solved analytically, thus requiring a numerical approach. The most reliable approximation is adaptive Gauss–Hermite quadrature, which is currently only available for models with a single scalar random effect. Meanwhile, for models with higher random effect dimensions, Gaussian quadrature is used as an alternative [22], [23].

3. RESULTS AND DISCUSSION

3.1 Data Exploration

The observation data in this study amounted to 2,451 Madrasah Aliyah teachers who participated in the 2022 PPG program. Fig. 4 shows that based on the graduation status of the PPG program, 1,540 teachers, or 63%, successfully passed, while the remaining 911 teachers, or 37%, failed to pass.

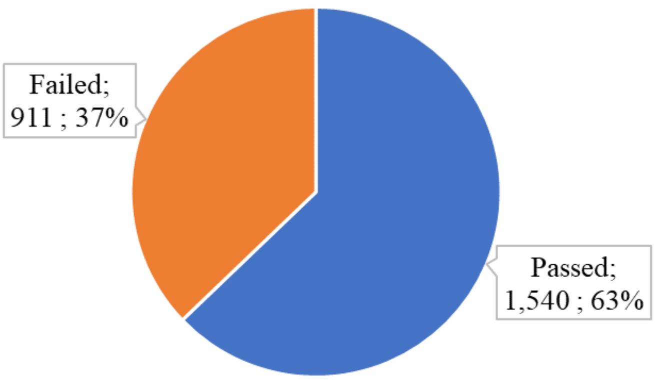

Figure 4. Proportion of MA Teachers by Graduation Status of the 2022 PPG Program

A comparison of pass–fail outcomes among MA teachers in the 2022 PPG program, disaggregated by subject area is presented in Fig. 5. English teachers recorded the highest pass rate (93.67%; 148 passed, 10 failed), followed by Indonesian language (86.21%; 75 passed, 12 failed), Social Science (84.41%; 222 passed, 41 failed), Physical and Health Education (84.38%; 27 passed, 5 failed), Natural Science (79.79%; 233 passed, 59 failed), Others subject (76.14%; 67 passed, 21 failed), Mathematics (67.86%; 76 passed, 36 failed), and Civic Education (67.44%; 29 passed, 14 failed). Conversely, teachers of Islamic Studies subjects, which are characteristic and strength of madrasahs, showed lower pass rates: Fiqh (64.29%; 225 passed, 125 failed), Arabic (47.10%; 138 passed, 155 failed), Aqidah Akhlak (44.37%; 138 passed, 173 failed), Quran Hadith (42.91%; 121 passed, 161 failed), and Islamic History and Culture (29.29%; 41 passed, 99 failed).

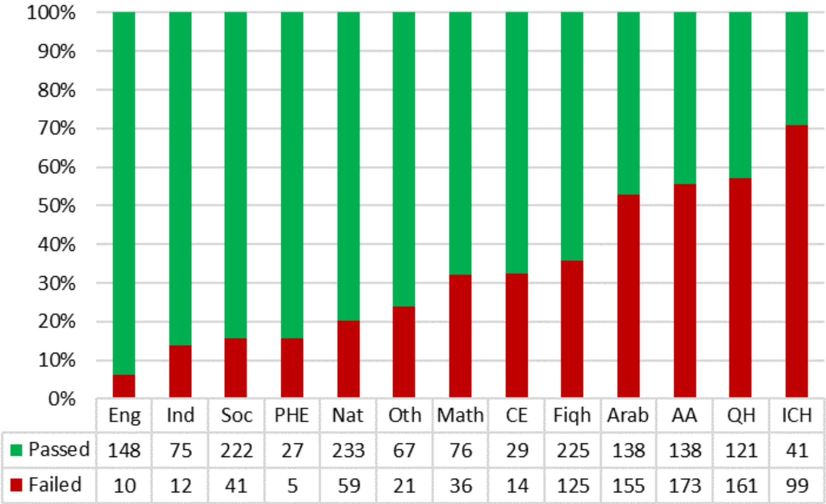

Figure 5. Graduation Status of MA Teachers in the 2022 PPG Program based on Subjects

Comparison of the age distributions of Madrasah Aliyah teachers by PPG graduation status is presented in Fig. 6. The "passed group" exhibits a narrower interquartile range (IQR) than the "failed group", indicating lower within-group age dispersion, and its median age is slightly lower. The third quartile (Q3) for the "passed group" lies below 40 years, implying that at least 75% of successful candidates are younger than 40. Several successful candidates aged 47 years or older are flagged as high-end outliers, and a single low-age outlier is also observed. In contrast, the "failed group" exhibits a broader age dispersion extending into the mid-50s, with no flagged outliers. These summaries are descriptive and should not be construed as evidence of causality.

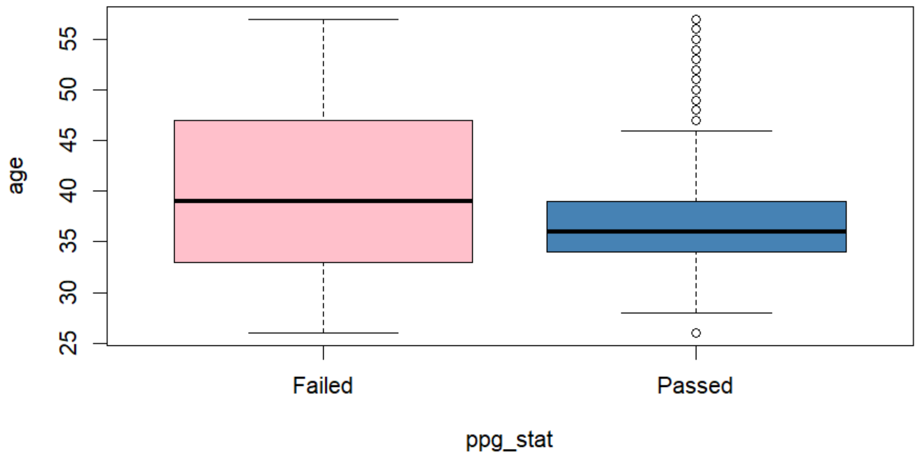

Figure 6. Boxplot Teacher Graduation Status by Age

3.2 GLMM Trees Model Analysis

The construction of the GLMM tree model begins by dividing the observation data into two parts: the training dataset and the testing dataset, with a ratio of 80% to 20%. The GLMM tree model is built using the training dataset, tested with the testing dataset, and then re-validated using the training dataset. Fig. 7 shows the mixed effects tree estimation of the GLMM tree model constructed.

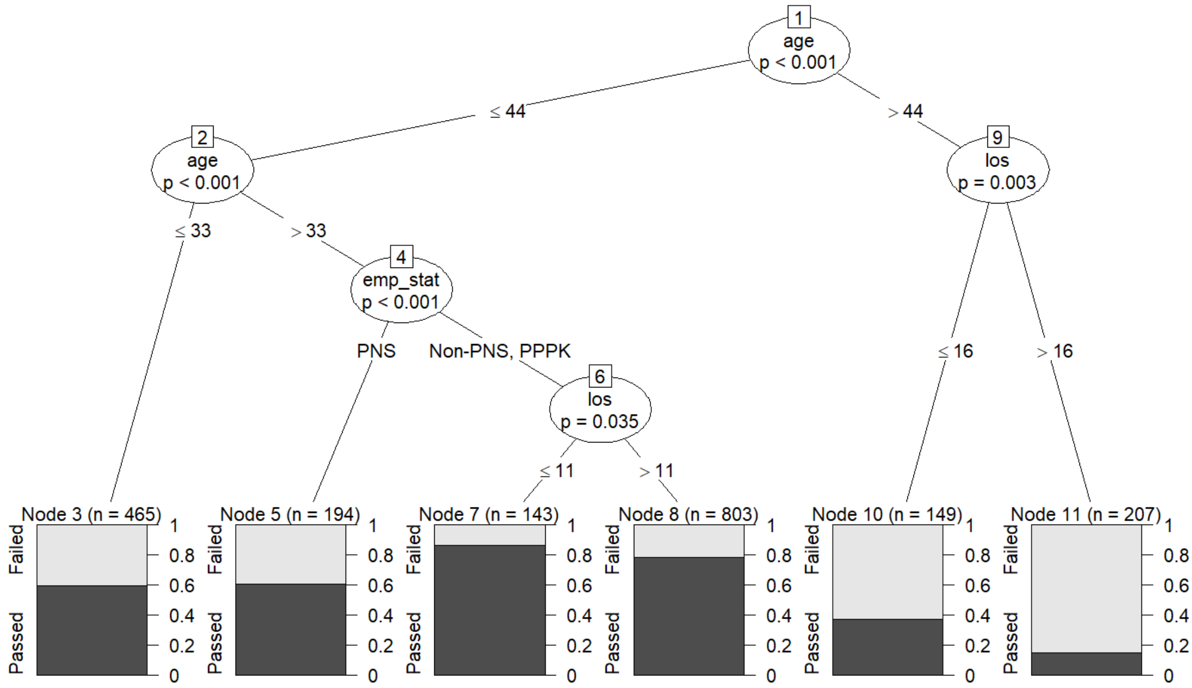

Figure 7. Mixed Effects Tree Estimation of GLMM Trees Model for PPG Graduation Probability

The estimation of the intercepts of the fixed effects of the GLMM tree model is shown in Table 5. The fixed effect intercept in GLMM trees is a constant on the linear predictor specific to each terminal node, representing the baseline value of the output when all covariates are at their reference level, and the random effects are assumed to be zero.

Node 7 is the node with the highest probability of predicting a teacher passing, with the criteria being teachers aged between 34 and 44 years old, having non-civil servant or PPPK employment status, and having a tenure of 11 years or less. As presented in Table 5, node 7 shows the highest fixed effects intercept contribution in the GLMM trees model in predicting teacher certification completion with an intercept of 2.3477079.

Table 5. Estimation of Intercepts of Fixed Effects of GLMM Trees Model

Nodes	(Intercept)
7	2.3477079
8	1.5795302
5	0.6163572
3	0.6097383
10	-0.1868425
11	-1.3328530

Meanwhile, Fig. 7 presents that the lowest probability of a teacher graduating from the PPG program is produced by node 11, provided that the teacher is over 44 years old and has more than 16 years of service. Based on Table 5, node 11 has the lowest fixed effect intercept value of -1.3328530.

Fig. 8 shows that, for most subject categories, the 95% confidence interval for the random intercept estimate is not equal to zero, indicating a deviation from the overall mean (on the log-odds scale). Thus, based on the GLMM trees model, the probability of passing the PPG program for most subjects differs from the mean. Subjects with intervals entirely to the right of zero are associated with above-average probabilities of passing, while those entirely to the left are associated with below-average probabilities.

As for the random effects component, Indonesian Language and English subjects are indicated to have a large positive effect on the intercept, which is 1.2164515 and 1.0903455, respectively. This means that teachers of these two subjects have a greater probability of passing the PPG program. On the other hand, Islamic Cultural History and Arabic have the smallest intercept values of -1.2693007 and -1.2804024, respectively, indicating that teachers of these subjects have the lowest probability of passing the PPG program. The above explanation is illustrated by the estimation of the random effect intercept confidence interval for each subject group in the GLMM trees model presented in Fig. 8.

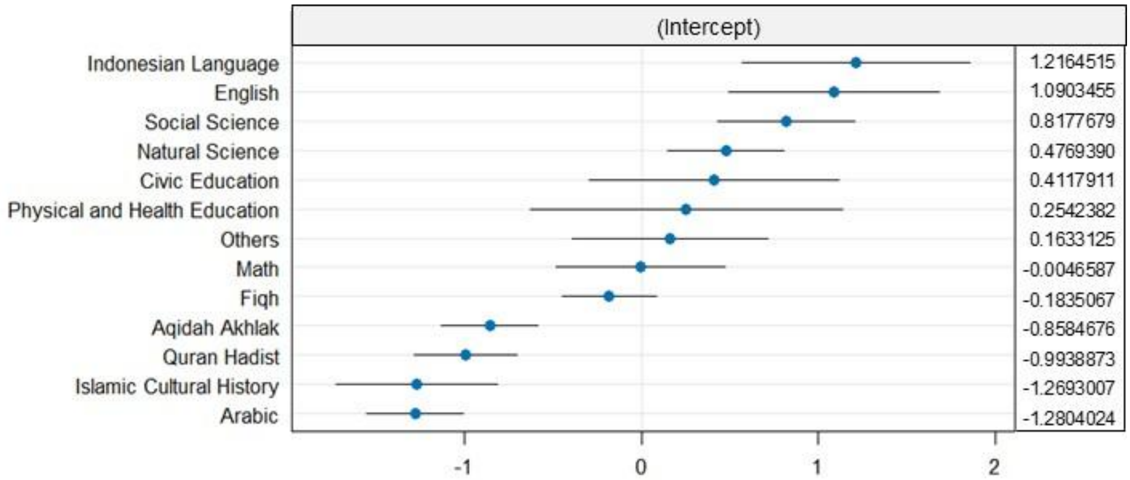

Figure 8. Estimation of Random Intercept Confidence Interval for each Subject Group in the GLMM Trees Model

Based on the confusion matrix shown in Table 6, the GLMM trees model has an accuracy index of 0.7306, a precision index of 0.7558, a sensitivity index of 0.8442, a specificity index of 0.5385, and an F1 Score of 0.7975. This profile suggests strong detection of the positive class (high sensitivity/F1) but comparatively weak discrimination of the negative class (lower specificity), implying a higher false-positive rate. Performance metrics are higher on the training set than on the test set—as expected—which reflects in-sample fit and may indicate mild overfitting rather than "stability". Overall, the model shows reasonable predictive utility.

Table 6. Confusion Matrix of GLMM Trees Model

Dataset	Prediction	Actual		Indices				
		Passed	Failed	Accuracy	Precision	Sensitivity	Specificity	F1 Score
Testing Dataset	Passed	260	84	0.7306	0.7558	0.8442	0.5385	0.7975
	Failed	48	98					
Training Dataset	Passed	1,077	276	0.7802	0.7960	0.8742	0.6214	0.8333
	Failed	155	453					

3.3 GMET Model Analysis

As with the construction of the GLMM trees model, the GMET model is built using a training dataset of as much as 80% of the total observation data, tested with a testing dataset of 20%, and then applied to the training dataset to revalidate the model. The mixed effects tree estimation of the GMET model in Fig. 9 shows that the greatest probability of Madrasah Aliyah teachers passing the PPG program is when the teacher is between 33 and 36 years old, teaches in a madrasah with a minimum accreditation of B, and has a non-civil servant (Non-PNS) or PPPK employment status.

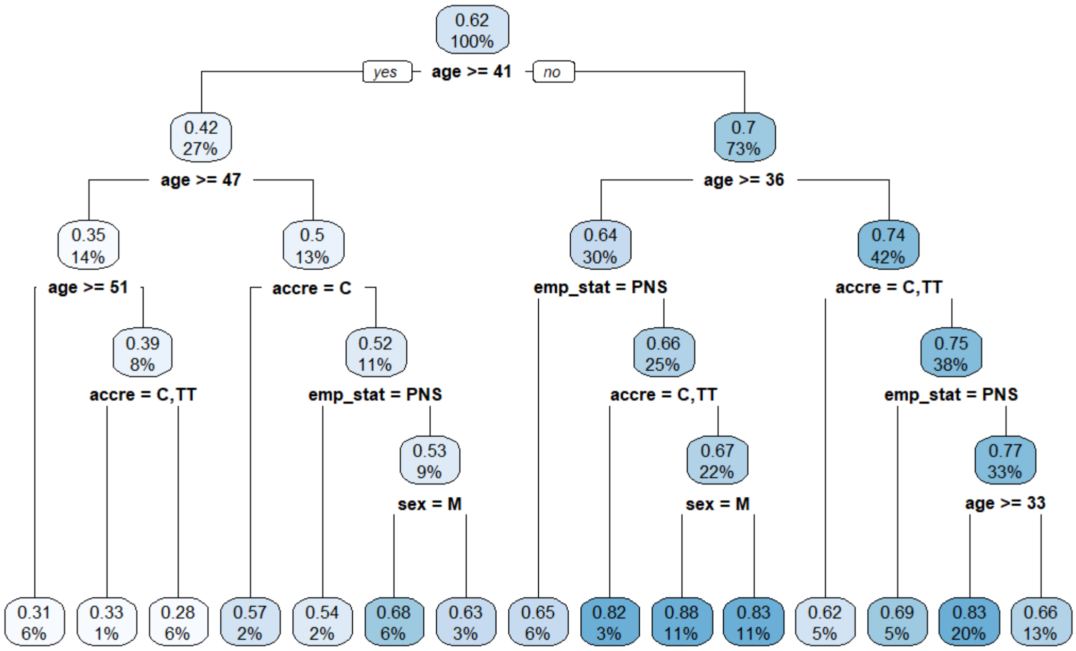

Figure 9. Mixed Effects Tree Estimation of GMET Model for PPG Graduation Probability

In the GMET model, random intercept confidence intervals are estimated for each subject following a distribution derived from the estimated model parameters. These random intercept confidence intervals are estimated based on quantiles $CI_{95\%} = [\hat{\theta}_{2.5\%}, \hat{\theta}_{97.5\%}]$. Fig. 10 presents that most subject groups have 95% confidence intervals that do not overlap with the zero vertical line, indicating that there are significant differences among subject groups. If the GMET model is used to estimate the likelihood of teachers passing the PPG program, then most subjects will give results that differ from the average. Subject groups that have confidence intervals to the right of the zero vertical line have a higher-than-average probability of passing. And vice versa.

Furthermore, Fig. 10 shows the random effects intercept estimation of the GMET model. The value of the random effects intercept describes the variability among subject groups that cannot be elucidated by the predictor variables in the model. English and Indonesian languages are two subjects that are indicated to have the largest positive effect on the intercept: a teacher for both subjects increases the probability of passing by 1.42729227 and 1.35321546, respectively. Conversely, an Arabic and Islamic Cultural History teacher decreases the probability of passing by 1.19596906 and 1.46832199, respectively.

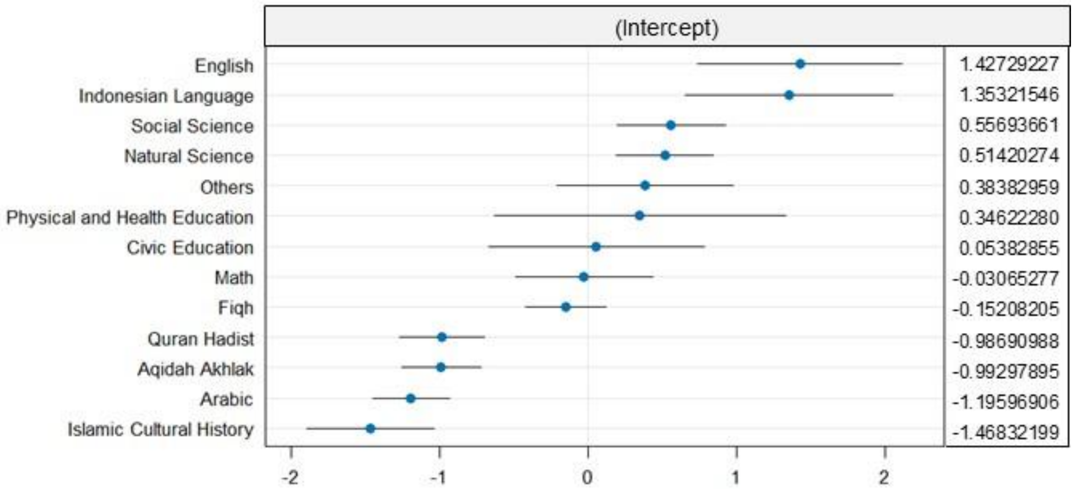

Figure 10. Estimation of Random Intercept Confidence Interval for Each Subject Group in the GMET Model

Referring to the confusion matrix in Table 7, the GMET model achieves, on the held-out test dataset, an accuracy of 0.7653, precision of 0.7849, sensitivity (recall) of 0.8809, specificity of 0.5497, and an F1-

score of 0.8301. This metric profile indicates strong detection of the positive class (high sensitivity and F1), but comparatively weaker discrimination of the negative class (lower specificity), implying a greater propensity for false positives. Revalidation on the training set yields slightly higher accuracy than on the test dataset —an expected pattern consistent with mild overfitting rather than improved generalization. Overall, the model exhibits reasonable predictive performance, though additional tuning (e.g., decision-threshold adjustment or class weighting) may help to better balance sensitivity and specificity.

Table 7. Confusion Matrix of GMET Model

Dataset	Prediction	Actual		Indices				
		Passed	Failed	Accuracy	Precision	Sensitivity	Specificity	F1 Score
Testing Dataset	Passed	281	77	0.7653	0.7849	0.8809	0.5497	0.8301
	Failed	38	94					
Training Dataset	Passed	1,052	285	0.7685	0.7868	0.8616	0.6149	0.8225
	Failed	169	455					

3.4 Discussion

Based on the results of the analysis of Madrasah teacher certification data using two methods, the GLMM tree and GMET, the performance results are shown in Table 8. All model performance indices indicate that the GMET model has superiority over the GLMM trees model in predicting the graduation of Madrasah Aliyah teachers in the PPG program. The accuracy index of the GMET model is 0.7653, 0.0347 higher than the GLMM trees model, which has an accuracy of 0.7306. The GMET model has a precision index of 0.7849, 0.0291 higher than the GLMM trees with a precision of 0.7558. Meanwhile, the F1 Score for GMET of 0.8301 is 0.0326 higher than the GLMM trees of 0.7975.

Table 8. Comparison of Performance Indices of GLMM Trees and GMET Models

Model	Prediction	Actual		Indices				
		Passed	Failed	Accuracy	Precision	Sensitivity	Specificity	F1 Score
GLMM Trees	Passed	260	84	0.7306	0.7558	0.8442	0.5385	0.7975
	Failed	48	98					
GMET	Passed	281	77	0.7653	0.7849	0.8809	0.5497	0.8301
	Failed	38	94					

Assessing variable importance is fundamental for quantifying the relative contribution of predictors to the modeled outcome [16]. In this study, we compute variable-importance scores that reflect each predictor's contribution to the overall predictive performance of both models, while accounting for the mixed-effects structure (fixed and random effects). These importance levels are obtained from the fitted GMET and GLMM mixed effects models.

Fig. 11 displays the importance of the predictor variables in the mixed effects tree estimation of the GMET model. Among the seven predictor variables in the model, the variable of *age* is indicated to have the highest level of variable importance, which is 37.35. This means that teachers' age is the most influential variable in the GMET model to predict the graduation of Madrasah Aliyah teachers participating in the PPG program. The next order is the variable of *los* (teachers' length of service), *accre* (accreditation status of madrasah), *emp_stat* (teachers' employment status), *sex* (teachers' gender), *mad_stat* (institution status of madrasah), and the lowest is the *educ* (teachers' highest educational qualification).

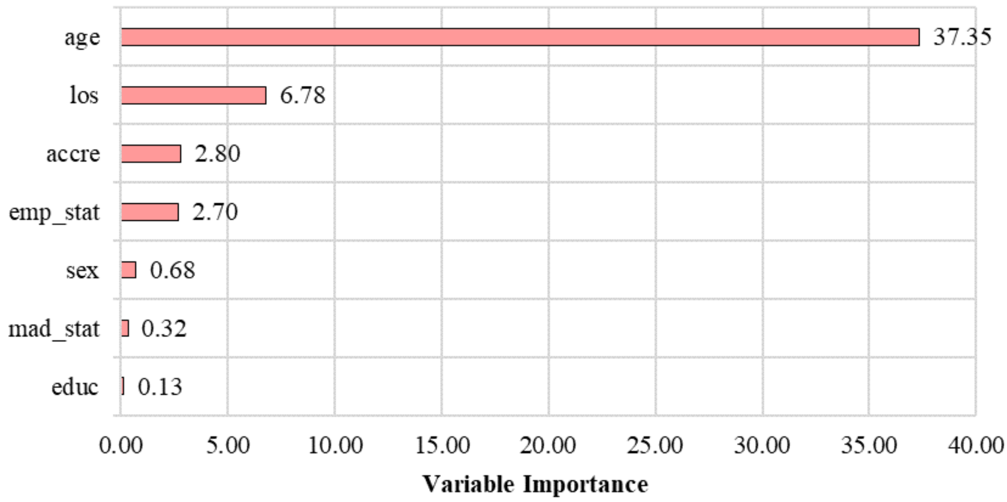

Figure 11. Variable Importance of the GMET Model

Fig. 12 reports the variable importance profile from the mixed effects tree estimation of the GLMM trees model. Of the seven candidate predictors, only three exhibit non-zero importance: *age*, *los* (length of service), and *emp_stat* (employment status). Consistent with the GMET model, teachers' age ranks first with an importance score of 0.1355, indicating the largest contribution to predictive performance for PPG graduation among Madrasah Aliyah teachers, followed by length of service (0.0173) and employment status (0.0128); the remaining predictors have negligible (≈0) importance under this model and metric.

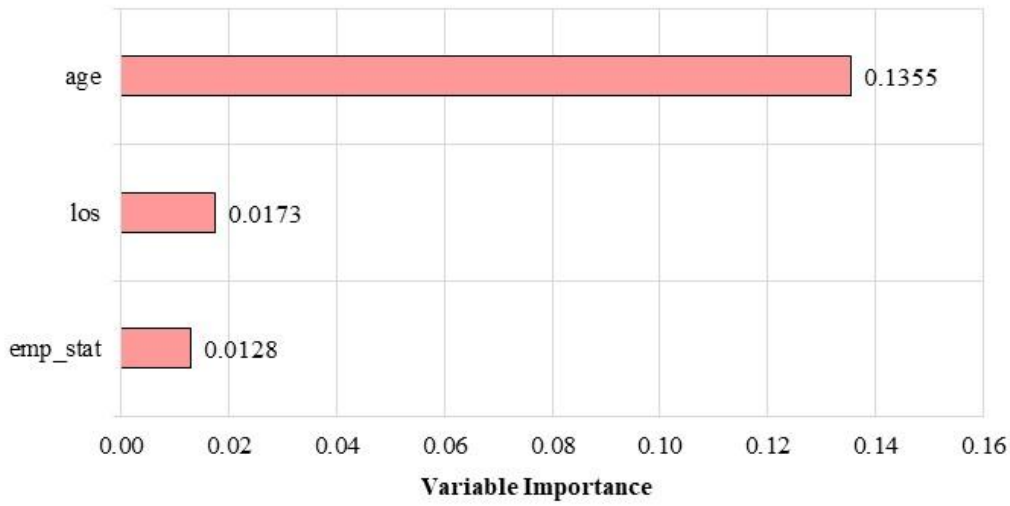

Figure 12. Variable Importance of the GLMM Trees Model

4. CONCLUSION

This study has applied two methods, GLMM trees and GMET, to predict the graduation of Madrasah Aliyah teachers who participated in the 2022 PPG program. The results of data modeling show that the GMET model performs better than the GLMM trees model. Based on the subjects taught, English and Indonesian language teachers are more likely to pass the PPG program compared to teachers of other subjects. Meanwhile, Arabic and Islamic Cultural History teachers have the lowest probability of passing the PPG program among all subject teachers. Based on the variable importance measurement of the GMET and GLMM tree models, teachers' age is indicated as the most influential variable on Madrasah Aliyah teachers' graduation in the 2022 PPG program. Based on these findings, the recommendation proposed to the Ministry of Religious Affairs as policymaker is to provide appropriate coaching and mentoring programs for teachers who are not yet certified professional educators, especially for senior teachers over 40 years old, as well as teachers of Arabic and Islamic Culture History subjects. The training required for teachers to prepare for the

PPG program includes material on innovative teaching methods, the use of technology in learning, and leadership development in the classroom. These conclusions pertain specifically to graduation outcomes in the Madrasah Aliyah (MA) teacher certification program and should not be assumed to generalize to other madrasah levels—Raudhatul Athfal (RA), Madrasah Ibtidaiyah (MI), or Madrasah Tsanawiyah (MTs)—without further evidence.

Author Contributions

Dodi Irawan Syarip: Conceptualization, Methodology, Data Curation, Writing-Original Draft, Software, Validation, Visualization. Khairil Anwar Notodiputro: Supervision, Methodology, Writing-Review and Editing. Bagus Sartono: Validation, Software, Writing of Reviews, and Editing. All authors discussed the results and contributed to the final manuscript.

Funding Statement

The first author expresses gratitude to the Ministry of Religious Affairs for supporting the author to continue his doctoral studies through the Beasiswa Indonesia Bangkit (BIB) Program.

Acknowledgment

The authors would like to express their gratitude to the Directorate General of Islamic Education, Ministry of Religious Affairs, particularly the Directorate of Madrasah Teachers and Educators, for granting the authors permission to access data from the 2022 PPG Program for Madrasah Aliyah Teachers.

Declarations

The authors declare that there are no conflicts of interest in this study.

Declaration of Generative AI and AI-assisted Technologies

AI tools (e.g., ChatGPT) were used solely for language refinement, including grammar, spelling, and clarity. The scientific content, analysis, interpretation, and conclusions were developed entirely by the authors. All final text was reviewed and approved by the authors.

REFERENCES

[1] B. M. Bolker, "LINEAR AND GENERALIZED LINEAR MIXED MODELS," in *Ecological Statistics*, Oxford University PressOxford, 2015, pp. 309–333. doi: https://doi.org/10.1093/acprof:oso/9780199672547.003.0014.

[2] R. Bono, R. Alarcón, and M. J. Blanca, "REPORT QUALITY OF GENERALIZED LINEAR MIXED MODELS IN PSYCHOLOGY: A SYSTEMATIC REVIEW," *Front Psychol*, vol. 12, Apr. 2021. doi: https://doi.org/10.3389/fpsyg.2021.666182.

[3] A. Agresti, *AN INTRODUCTION TO CATEGORICAL DATA ANALYSIS*. NJ, USA: Wiley: Hoboken, 2018.

[4] A. Rusyana, A. Kurnia, K. Sadik, A. H. Wigena, I. M. Sumertajaya, and B. Sartono, "COMPARISON OF GLM, GLMM AND HGLM IN IDENTIFYING FACTORS THAT INFLUENCE THE DISTRICT OR CITY POVERTY LEVEL IN ACEH PROVINCE," *J Phys Conf Ser*, vol. 1863, no. 1, p. 012023, Mar. 2021. doi: https://doi.org/10.1088/1742-6596/1863/1/012023.

[5] H. Seibold, T. Hothorn, and A. Zeileis, "GENERALISED LINEAR MODEL TREES WITH GLOBAL ADDITIVE EFFECTS," *Adv Data Anal Classif*, vol. 13, no. 3, pp. 703–725, Sep. 2019. doi: https://doi.org/10.1007/s11634-018-0342-1.

[6] M. Fokkema, N. Smits, A. Zeileis, T. Hothorn, and H. Kelderman, "DETECTING TREATMENT-SUBGROUP INTERACTIONS IN CLUSTERED DATA WITH GENERALIZED LINEAR MIXED-EFFECTS MODEL TREES," *Behav Res Methods*, vol. 50, no. 5, pp. 2016–2034, Oct. 2018. doi: https://doi.org/10.3758/s13428-017-0971-x

[7] H. Seibold, A. Zeileis, and T. Hothorn, "MODEL-BASED RECURSIVE PARTITIONING FOR SUBGROUP ANALYSES," *International Journal of Biostatistics*, vol. 12, no. 1, pp. 45–63, May 2016. doi: https://doi.org/10.1515/ijb-2015-0032.

[8] M. Fokkema and A. Zeileis, "SUBGROUP DETECTION IN LINEAR GROWTH CURVE MODELS WITH GENERALIZED LINEAR MIXED MODEL (GLMM) TREES," *Behav Res Methods*, vol. 56, no. 7, pp. 6759–6780, May 2024. doi: https://doi.org/10.3758/s13428-024-02389-1.

[9] C. M. Loan, G. Tindal, and E. Tanner-Smith, "EXPLORATORY DATA ANALYSIS WITH CLUSTERED DATA: SIMULATION AND APPLICATION WITH OREGON'S STATEWIDE LONGITUDINAL DATA SYSTEM USING GENERALIZED LINEAR MIXED-EFFECTS MODEL TREES," 2024.

[10] M. Fokkema, J. Edbrooke-Childs, and M. Wolpert, "GENERALIZED LINEAR MIXED-MODEL (GLMM) TREES: A FLEXIBLE DECISION-TREE METHOD FOR MULTILEVEL AND LONGITUDINAL DATA," *Psychotherapy Research*, vol. 31, no. 3, pp. 329–341, Apr. 2021. doi: https://doi.org/10.1080/10503307.2020.1785037.

[11] Y. Wei, L. Liu, X. Su, L. Zhao, and H. Jiang, "PRECISION MEDICINE: SUBGROUP IDENTIFICATION IN LONGITUDINAL TRAJECTORIES," *Stat Methods Med Res*, vol. 29, no. 9, pp. 2603–2616, Sep. 2020. doi: https://doi.org/10.1177/0962280220904114.

[12] S. Bayu, K. A. Notodiputro, and B. Sartono, "GLMM AND GLMMTREE FOR MODELLING POVERTY IN INDONESIA," *Proceedings of The International Conference on Data Science and Official Statistics*, vol. 2023, no. 1, pp. 121–131, Dec. 2023. doi: https://doi.org/10.34123/icdsos.v2023i1.333.

[13] D. A. N. Sirodj, K. A. Notodiputro, and B. Sartono, "COMPARISON OF BINOMIAL GLMM TREE AND BIMM FOREST FOR MODELING THE EMPLOYMENT STATUS OF THE POPULATION," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 95–106, Feb. 2024. doi: https://doi.org/10.25126/jtiik.20241117531.

[14] L. Fontana, C. Masci, F. Ieva, and A. M. Paganoni, "PERFORMING LEARNING ANALYTICS VIA GENERALISED MIXED-EFFECTS TREES," *Data (Basel)*, vol. 6, no. 7, p. 74, Jul. 2021. doi: https://doi.org/10.3390/data6070074.

[15] Sukarna, K. A. Notodiputro, and B. Sartono, "COMPARISON BETWEEN BINOMIAL GLMM AND BINOMIAL GMET FOR TEMPORARY UNEMPLOYMENT IN WEST JAVA, INDONESIA," 2023, pp. 198–209. doi: https://doi.org/10.2991/978-94-6463-332-0_22.

[16] F. Rahmawati, K. A. Notodiputro, and K. Sadik, "CLASSIFICATION OF HOUSEHOLD POVERTY IN WEST JAVA USING THE GENERALIZED MIXED-EFFECTS TREES MODEL," *Jurnal Natural*, vol. 23, no. 3, pp. 209–219, Oct. 2023. doi: https://doi.org/10.24815/jn.v23i3.33079.

[17] F. N. Arifa and U. S. Prayitno, "IMPROVING EDUCATION QUALITY: PRE-SERVICE TEACHER PROFESSIONAL EDUCATION PROGRAMS IN MEETING THE NEED FOR PROFESSIONAL TEACHERS IN INDONESIA," *Aspirasi: Jurnal Masalah-Masalah Sosial*, vol. 10, no. 1, Jun. 2019. doi: https://doi.org/10.46807/aspirasi.v10i1.1229

[18] A. Dudung, "TEACHER PROFESSIONAL COMPETENCE," *JKKP (Jurnal Kesejahteraan Keluarga dan Pendidikan)*, vol. 5, no. 1, pp. 9–19, Apr. 2018. doi: https://doi.org/10.21009/JKKP.051.02.

[19] D. Valero-Carreras, J. Alcaraz, and M. Landete, "COMPARING TWO SVM MODELS THROUGH DIFFERENT METRICS BASED ON THE CONFUSION MATRIX," *Comput Oper Res*, vol. 152, p. 106131, Apr. 2023. doi: https://doi.org/10.1016/j.cor.2022.106131.

[20] P. McCullagh and J. A. Nelder, "GENERALIZED LINEAR MODELS," *Statistical Modelling by Exponential Families*, pp. 271–291, 2019. doi: https://doi.org/10.1201/9780203753736

[21] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *CLASSIFICATION AND REGRESSION TREES*. Routledge, 2017. doi: https://doi.org/10.1201/9781315139470.

[22] R. Gueorguieva, "A MULTIVARIATE GENERALIZED LINEAR MIXED MODEL FOR JOINT MODELLING OF CLUSTERED OUTCOMES IN THE EXPONENTIAL FAMILY," *Stat Modelling*, vol. 1, no. 3, pp. 177–193, Oct. 2001. doi: https://doi.org/10.1177/1471082X0100100302

[23] D. Handayani, K. A. Notodiputro, K. Sadik, and A. Kurnia, "A COMPARATIVE STUDY OF APPROXIMATION METHODS FOR MAXIMUM LIKELIHOOD ESTIMATION IN GENERALIZED LINEAR MIXED MODELS (GLMM)," in *AIP conference proceedings*, AIP Publishing LLC, Mar. 2017, p. 020033. doi: https://doi.org/10.1063/1.4979449.