Comparison between binomial GLMM and binomial GMET for temporary unemployment in West Java, Indonesia

Sukarna[1,2] , Khairil Anwar Notodiputro[2], and Bagus Sartono[2]

[1] Universitas Negeri Makassar, Indonesia
[2] IPB University, Indonesia
khairil@apps.ipb.ac.id

Abstract. Unemployment poses a significant challenge to national development. In response to this concern, the government initiated a unique survey on unemployment known as the *Survey Angkatan Kerja Nasional* (*SAKerNas*, National Labor Force Survey) in August 2022. This study aims to assess the predictive performance of two models: the binomial Generalized Linear Mixed Model (GLMM) and the binomial Generalized Mixed Effect Tree (GMET) in the context of temporary unemployment. West Java Province was selected as the focal point of this study due to its high unemployment rates in February 2022 and February 2023. The dataset for this study is based on the 2022 *SAKerNas* results for West Java Province, encompassing 55,957 individuals. The analysis focuses on six independent variables: age, number of household members, gender, marital status, training attendance (whether individuals have attended a course), and the highest level of education achieved. The performance of these models is evaluated using five measures: Accuracy, Sensitivity, Precision, Recall, and the F1 score. The study's findings indicate that the binomial GMET model outperforms the binomial GLMM model in predicting temporary unemployment. However, the time required to run the R syntax of the GMET model is longer than the GLMM model.

Keywords: Binomial GLMM, Binomial GMET, Temporary Unemployment, West Java

1 Introduction

The LM (Linear Model) is performed to model statistics, with normality distribution as a basic assumption. This statistical model is dynamic and continuously adapted in response to advancements in information, technology, and computer power. This circumstance gives rise to a multitude of approaches, rules, and assumptions. Hence, it is important to do research on the correlation between models and data, irrespective of the amount or kind of data.

LM has recently been undergoing substantial and intricate expansion. The LM has transformed into the Generalized LM (GLM) and the Linear Mixed Model (LMM). LM is converted into GLM by transforming the prior response variable from a Gaussian

N. Djam'an et al. (eds.), *Proceedings of the 5th International Conference on Statistics, Mathematics, Teaching, and Research 2023 (ICSMTR 2023)*, Advances in Computer Science Research 109,
https://doi.org/10.2991/978-94-6463-332-0_22

distribution to a non-Gaussian distribution. By augmenting the random influence on the fixed variable, LM transitions into LMM. The integration of both the GLM and LMM enhances the Generalized Linear Mixed Model (GLMM). GLMs are linear models used when the response variable follows a random distribution, such as binomial, Poisson, or gamma distribution. GLMM is a composite model that combines the features of GLM and LMM. Its constituents consist of the response variable (Y), the predictor variables (X), the random effects (v), and the fixed effects (v). HGLM, conversely, is a GLMM when the random impact on independent variables does not follow a Gaussian distribution [1].

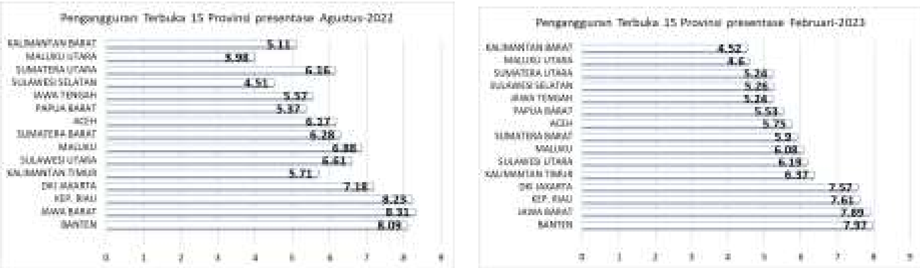
GLMM is a statistical model that enhances the performance of fixed and random effects. This effect is widely considered to have a natural look [2]. Lawson and Clark (2002) mentioned some of the GLMM principles used in spatial model building in their discussion of the potential risks of surface non-continuity, and Loh and Zhu (2007) calculated the spatial correlation of scan statistics with GLMM spatial models in an attempt to obtain more accurate analysis results [3].

GLMMs incorporate multilevel structures in binary response variables while imposing linear effects of covariates on the transformation of the response variable [4]. Tree-based approaches, such as the classification and regression tree (CART) model, are implemented to gain insights into the correlation between responsibility and predictability [5]. One notable aspect of tree-based techniques is their ability to convey and understand information using straightforward visual representations [4]. The approach was subsequently referred to as the Generalized Mixed Effect Tree (GMET).

Tree-based algorithms may also be utilized for analyzing longitudinal and clustered data. Sela and Simonoff (2012) offered a regression tree approach specifically designed for analyzing longitudinal or clustered data [6]. The name of this approach is the random-effects expectation-maximum tree (RE-EM). Ahlem Hajjem, Francois Bellavance, and Denis Larocque independently proposed the mixed-effects regression tree model (MERT) in 2011 [7]. When considering grouped observations, these approaches may be seen as expansions of the basic regression tree algorithm to handle situations where people are organized into groups. These techniques use observation-level variables during the splitting procedure and can accommodate random effects that may be linked to those covariates.

Ahlem Hajjem, Denis Larocque, and Francois Bellavance employed the MERT methodology for non-Gaussian data, as described in their publication from 2017 [8]. Additionally, they utilized the generalized mixed-effects regression tree (GMERT) approach. The penalized quasi-likelihood (PQL) method estimates GLMMs. Nevertheless, the weighted linear mixed-effects pseudo-model is replaced with a weighted MERT pseudo-model. Fokkema, Smits, Zeileis, Hothorn, and Kelderman (2018) proposed a method called the generalized linear mixed-effects model (GLMM) tree approach [9], which iteratively combines the GLM tree with mixed-effects model estimate until convergence. The critical difference between the GLMM tree method and the GMET methodology is that the GLMM tree algorithm is based on model-based recursive partitioning (MOB, [10]) instead of CART. Speiser et al. [11] introduced a decision tree method to analyze clustered and longitudinal binary outcomes. Their methodology encompasses binary products utilizing Bayesian GLMMs and incorporates a random intercept. Unlike GMET, GMET commences by setting the random effect to zero. It calculates the target variable through GLM, employing a suitable link function based on the response family distribution. A regression tree is constructed with the estimated target variable as the dependent variable. Finally, a mixed-effects model is fitted to estimate the random effect component, utilizing the fixed effect component estimated by the tree as an offset. GMET is a particular variant of GLMM Tree, as described by Fontana et al. [4]. Hence, this study aims to assess and compare the efficacy of GLMMs and GMETs in accurately describing temporary unemployment instances in the West Java region.

In 2022, West Java had the highest open unemployment rate, according to BPS, while in 2023, it had the second highest rate. According to the BPS 2023 report, West Java and Banten consistently rank first regarding open unemployment in Indonesia. Figure 1 illustrates the ranking and percentage of open unemployment in the 15 provinces of Indonesia [12].



(a) (b)

Fig. 1. Bar chart depicting the 15 provinces with the most open unemployment [12]

The provinces in Indonesia with the highest unemployment rates as of August 2022, according to BPS report [12], are as follows: West Java (8.31 percent), Riau Islands (8.23 percent), Banten (8.09 percent), DKI Jakarta (7.18 percent), Maluku (6.88 percent), North Sulawesi (6.61 percent), West Sumatra (6.28 percent), Aceh (6.17 percent), North Sumatra (6.16 percent), East Kalimantan (5.71 percent).

West Java, shown in Figure 2, is a province in Indonesia near the country's capital city, Jakarta. The population of West Java is 48,264,516, and its size is 35,378 km2. West Java has 18 regencies and nine cities, with 627 sub-districts and 5,957 villages.

The provinces in Indonesia with the highest unemployment rates as of August 2022, according to BPS report [12], are as follows: West Java with an unemployment rate of 8.31 percent, Riau Islands with rate of 8.23 percent, Banten with 8.09 percent, DKI Jakarta with 7.18 percent, Maluku with 6.88 percent, North Sulawesi with 6.61 percent, West Sumatra with 6.28 percent, Aceh with 6.17 percent, North Sumatra with 6.16 percent, and East Kalimantan with an unemployment rate of 5.71 percent.

West Java, shown in Figure 2, is a province in Indonesia near the country's capital, Jakarta. West Java has a population of 48,264,516 with an area of 35,378 km2. West Java encompasses 18 regencies and 9 cities, with 627 sub-districts and 5,957 villages.

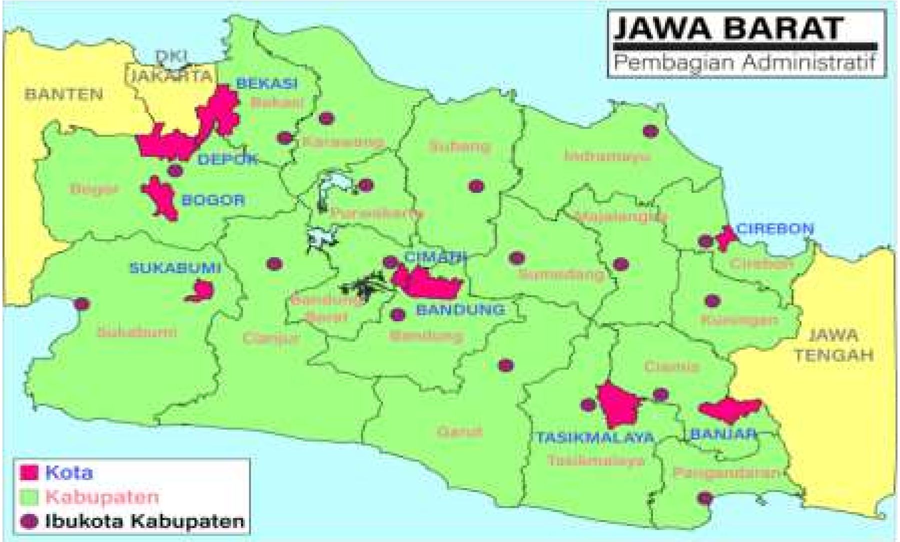


Fig. 2. The Map of West Java [13]

The ten districts/cities in West Java with the highest population [13, 14] are Bogor Regency (5,427,068 people), Bandung Regency (3,623,790), Bekasi Regency (3,113,071), Sukabumi Regency (2,725,450), Garut Regency (2,585,607), Bekasi Regency (2,543,676), Cianjur Regency (2,477,560), Bandung City (2,444,160), Karawang Regency (2,439,085), and Cirebon Regency (2,270,621).

The open unemployment rate in West Java Province [14] has decreased over the past three years. It was 10.46% in 2020, 9.82% in 2021, and 8.31% in 2022. According to the BPS West Java Province report in 2022, 14 districts/cities in West Java have consistently experienced high levels of open unemployment for three consecutive years. These districts/cities are Bogor Regency (14.29%, 12.22%, 10.64%), Cimahi City (13.30%, 13.07%, 10.77%), Bogor City (12.68%, 11.79%, 10.78%), West Bandung Regency (12.25%, 11.65%, 9.63%), Sukabumi City (12.17%, 10.78%, 8.83%), Bekasi Regency (11.54%, 10.09%, 10.31%), Karawang Regency (11.52%, 11.83%, 9.87%), Cirebon Regency (11.52%, 10.38%, 8.11%), Kuningan Regency (11.22%, 11.68%, 9.81%), Bandung City (11.19%, 11.46%, 9.55%), Purwakarta Regency (11.07%, 10.7%, 8.75%), Cianjur Regency (11.05%, 9.32%, 8.41%), Cirebon City (10.97%, 10.53%, 8.42%), and Bekasi City (10.68%, 10.88%, 8.81%). The percentages in parentheses represent the open unemployment rates for 2020, 2021, and 2022, respectively.

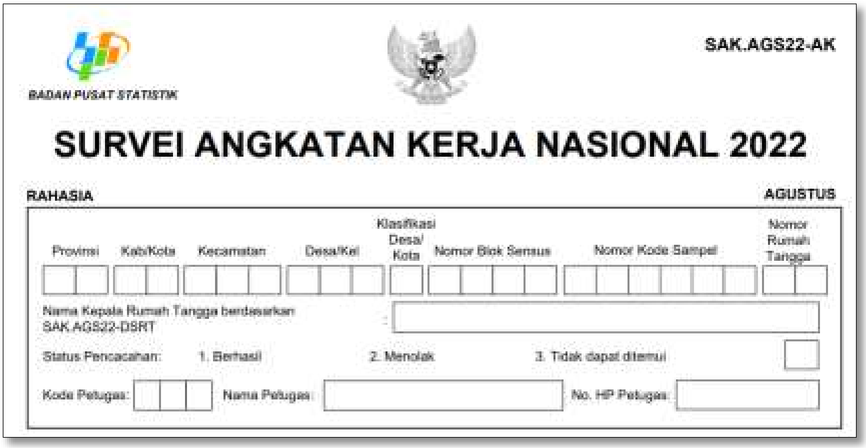
This study conducts a comparative analysis of the performance of the binomial-GLMM and binomial-GMET to identify the superior model for forecasting future instances of temporary unemployment in West Java. These findings are anticipated to be valuable for informing policy decisions on unemployment.

2 Data and Methods

The LM model is performed to model statistics, with normality distribution as a basic assumption. This statistical model is dynamic and continuously adapted in response to advancements in information, technology, and computer power. This circumstance gives rise to a multitude of approaches, rules, and assumptions. Hence, it is important to do research on the correlation between models and data, irrespective of the amount or kind of data.

2.1 Data

The data for this study is only derived from the findings of the National Labor Force Survey (*SAKerNas*) conducted in 2022, primarily focusing on the West Java Province. The data collecting equipment, depicted in Figure 3, was provided by the Central Statistics Agency (BPS).


BADAN PUSAT STATISTIK

SAK.AGS22-AK

SURVEI ANGKATAN KERJA NASIONAL 2022

RAHASIA

AGUSTUS

Provinsi | Kab/Kota | Kecamatan | Desa/Kel | Klasifikasi Desa/Kota | Nomor Blok Sensus | Nomor Kode Sampel | Nomor Rumah Tangga

Nama Kepala Rumah Tangga berdasarkan SAK.AGS22-DSRT :

Status Pencacahan: 1. Berhasil 2. Menolak 3. Tidak dapat ditemui

Kode Petugas: Nama Petugas: No. HP Petugas:

Fig. 3. Instrument paper of *SAKerNas* 2022 [15]

This survey included individuals residing in the selected homes. The resolute sample was nonetheless granted the prerogative to decline being utilized as a survey sample, with the intention of preventing respondents from feeling compelled and instead willingly offering precise information to BPS surveyors for the betterment of the nation and state. The overall sample obtained by *SAKerNas* amounted to 55,957 individuals.

2.2 Variables

The dependent variable in this study is the temporary unemployment status (abbreviated as "Work"), which is measured on a categorical scale. Specifically, a value of "1" represents being employed, while a value of "0" represents temporary unemployment [15]. Meanwhile, there are 7 independent variables with varying scales. The seven variables comprise two metric variables, age and number of family members (ART), one ordinal variable, highest education (Edu), and four categorical variables, gender, marital status, experience in attending training or courses, and the category of respondent's residence (Class = 0 Village and 1 City).

The *SAKerNas* 2022 instrument [15] provides detailed explanations of the household head, gender, and working, including their ideas and meanings. This explicit clarification is designed to provide guidance and instruction to surveyors to prevent any misinterpretation of the significance of these three points. The Head of Household is an individual who assumes responsibility for the daily necessities and is regarded as the leader of the household. home members refer to individuals who typically reside in the same dwelling as the head of the home. Work refers to the act of engaging in labor for a minimum of one hour throughout the previous week with the intention of obtaining financial compensation or gain. Work encompasses two categories: (a) Engaging in activities that generate money or profit for a minimum of one hour within the previous week, and (b) Individuals who possess a business or job but did not perform any work in the last week.

2.3 Methods

The procedure for evaluating regression tree and random forest models is as follows:

a. *Data analysis and investigation*. Perform data exploration by identifying missing data for each district/city, filtering data to assess data suitability (outliers), and verifying the accuracy of data entry.
b. *Split the data into training and testing sets*. The data is partitioned into two segments: 80% is allocated as training data for the purpose of constructing the model, while the remaining portion is designated as testing data for evaluating the model's performance.
c. *Constructing* the GLMMs with binomial distribution and GMETs with binomial distribution.
d. *Calculate* the magnitude of the assessment metric. Evaluate the efficacy of the binomial-GLMM and binomial-GMET models.
e. *Evaluating the optimal model*. The model is evaluated for three sets of analysis, specifically 10, 20, and 50 iterations, by varying the sample diversity while keeping the training data size constant at 80%. Every time the analysis is conducted, the model's effectiveness is evaluated. Assessing the quality of a model by comparing several types of training data, each comprising 80% of the total data, and the number of model iterations.
f. *Analysis of findings*.

2.3.1 GLMM

GLMM combines GLM and LMM [1, 16]. It has three components: a linear predictor, a link function, and a variance function. The GLMM, similar to the linear mixed model, has fixed effects (β), random effects ($\boldsymbol{u}$~N($\boldsymbol{0}$,$\boldsymbol{G}$)), a design matrix X for fixed effects, a design matrix Z for random effects, and an observation vector $\boldsymbol{y}|\boldsymbol{u}$ with an expected value and covariance matrix. In other words, it may be expressed as: The equation (1) states that $\boldsymbol{\eta}$ is equals the sum of $\boldsymbol{X}$Q and $\boldsymbol{Zu}$. Here, $\boldsymbol{\eta}$ represents a linear predictor with a link function denoted as (.) = ($\boldsymbol{y}|\boldsymbol{u}$) = $\boldsymbol{\mu}$. This link function g(.) relies on the linear predictor $\boldsymbol{\eta}$. The variance matrix $\boldsymbol{R}$ is dependent on $\boldsymbol{\mu}$ through the variance function. The often used linking functions in GLMM are as follows: identity, which assumes a normal distribution; logit and probit, which assume a binomial distribution; log, which assumes a Poisson distribution; and inverse and log, which assume a Gamma distribution. The variance function in the GLMM represents the random, unpredictable variation. In the context of the GLMM, the residual variance arises from two distinct sources: the inherent variability in the sample distribution and any extra variation or overdispersion. Overdispersion-induced variability may be represented by many modeling approaches, such as by setting the variance of $\boldsymbol{y}$ given $\boldsymbol{u}$ to be ϕ times $v(\boldsymbol{\mu})$, where ϕ represents the overdispersion parameter. Another approach involves including a stochastic component $\boldsymbol{ei}$~$(0,\phi)$ into the linear predictor for each observation or selecting a different distribution that better aligns with the characteristics of the data. Parameter estimation in GLMM may be accomplished by methods such as maximum likelihood, generalized estimating equations (GEE), Penalized quasi-likelihood, and Conditional likelihood [1].

2.3.2 GMET

The GMET model starts by constructing a GLMM. GMET develops a mathematical model based on observed data y_{ij}, which is represented as [4]:

$$Y_{ij} \sim Bernoulli(p_{ij});\ p_{ij} = E[Y_{ij}|\boldsymbol{b}_i];\ logit(p_{ij}) = f(x_{ij}) + \boldsymbol{b}_i;\ b_i \sim N(0,\psi);\ Ind.$$

where $\boldsymbol{x}_{ij} = (x_{1ij}, \dots, x_{ijp})^T$, covariate of fixed-effect (p+1) dimension for j-observation in i-group.

The GMET model's parameter estimation is conducted using the RE-EM tree technique. The fundamental concept of RE-EM is the segregation of fixed and random effects estimates. However, it should be noted that the GMET technique does not need iteration. The steps of the GMET algorithm can be succinctly stated as follows:

a. Set the initial estimation of random effects b_i to zero.
b. Utilize Generalized Linear Models (GLM) to estimate the target variable μ_{ij}, considering the fixed-effects variables x_{ij}, resulting in the estimation $\hat{\mu}_{ij}$.
c. Construct a regression tree f using the Classification and Regression Tree (CART) algorithm, where $\hat{\mu}_{ij}$ represents the dependent variable and x_{ij} represents the covariate vector.
d. Apply the mixed-effects model, with y_{ij} as the dependent variable.
e. Substituted the forecasted outcome at each final node for the TREE with the estimated forecasted outcome g($\hat{\gamma}_l$) from the mixed-effects model in step c.

3 Result and Discussion

3.1 Result

The analysis covered a total of 55,957 respondents, out of whom 24,155 individuals (43%) were classified as temporarily unemployed. The distribution of this data is rather equitable, with 57% of individuals being employed and 43% experiencing temporary unemployment. Hence, the number of employed respondents exceeds the number of unemployed respondents (Table 1).

Table 1. Worker distribution segmented by Gender, Marital Status, and Training

Worker	Gender			Marriage			Training		
	F	M	T	N	Y	T	N	Y	T
No	16,88	7,27	24,15	10,72	13,44	24,15	21,20	2,95	24,15
Yes	11,67	20,13	31,80	7,55	24,25	31,80	25,83	5,97	31,80
Total	28,56	27,40	55,96	18,27	37,69	55,96	47,04	8,92	55,96

Note: F=Female; M=Male; T=Total; No=N; Yes=Y

Table 2. Worker distribution based on the level of highest educational attainment.

Worker	1	2	3	4	5	6	7	8	Total
No	2,37	6,83	6,86	6,62	5,13	896	60	6	24,155
Yes	2,67	10,87	5,49	9,06	8,94	2,50	294	18	31,80
Total	5,03	17,70	12,36	15,68	14,07	3,39	354	24	55,96

Note: 1=No School; 2=SD/MI; 3=SMP/MTs; 4=SMA/MA; 5=Diploma; 6=Sarjana; 7=Magister; 8=Doctor

There is a higher proportion of females who are unemployed compared to those who are employed, but there is a higher proportion of males who are employed compared to those who are unemployed. Respondents without a spouse exhibit higher rates of unemployment, whereas those with a spouse have higher rates of employment. Conversely, the employment status of respondents is unaffected by their participation in training. This is due to the larger number of working respondents compared to the number of jobless respondents who have not undergone any training. Out of the total 12,360 respondents who completed junior high school or its equivalent, only 5,497 are employed while the remaining 6,863 are jobless. This pattern holds true for respondents with higher levels of education, where the number of employed individuals consistently exceeds the number of unemployed individuals.

3.2 Discussion

Table 4 illustrates the performance assessment outcomes of the binomial-GLMM and binomial-GMET models for a single iteration.

Binomial GLMM	Binomial GMET
pred_bin 0 1 0 2925 1594 1 1887 4786 Total = 7711	pred_bin 0 1 0 3686 2262 1 1126 4118 Total = 7804
Accuracy : 0.689 95% CI : (0.6803, 0.6975) No Information Rate : 0.5701 P-Value [Acc > NIR] : < 2.2e-16 Kappa : 0.3607 Mcnemar's Test P-Value : 7.454e-07 Sensitivity : 0.6079 Specificity : 0.7502 Pos Pred Value : 0.6473 Neg Pred Value : 0.7172 Precision : 0.6473 Recall : 0.6079 F1 : 0.6269 Prevalence : 0.4299 Detection Rate : 0.2613 Detection Prevalence : 0.4038 Balanced Accuracy : 0.6790 'Positive' Class : 0	Accuracy : 0.6973 95% CI : (0.6887, 0.7058) No Information Rate : 0.5701 P-Value [Acc > NIR] : < 2.2e-16 Kappa : 0.3999 Mcnemar's Test P-Value : < 2.2e-16 Sensitivity : 0.7660 Specificity : 0.6455 Pos Pred Value : 0.6197 Neg Pred Value : 0.7853 Precision : 0.6197 Recall : 0.7660 F1 : 0.6851 Prevalence : 0.4299 Detection Rate : 0.3293 Detection Prevalence : 0.5315 Balanced Accuracy : 0.7057 'Positive' Class : 0
Around 2 minutes	Around 35 minutes

Fig. 3. The R displays the result.

The results indicate that the binomial-GMET outperforms the binomial-GLMM in terms of accurately predicting employment and temporary unemployment. The performance metrics (accuracy, sensitivity, and F1) indicate that the binomial-GMET outperforms the binomial-GLMM.

The subsequent tables illustrate the replicability of the model analysis, conducted 10, 20, and 50 times, in order to examine the consistency of the performance evaluation metrics.

Table 1. Model assessment value (10 runs)

iteration	10			20			50		
Assessment	Acr	Stvy	F1	Acr	Stvy	F1	Acr	Stvy	F1
Mean GLMM	0.686	0.610	0.625	0.684	0.611	0.625	0.685	0.611	0.626
Mean GMET	0.693	0.765	0.682	0.692	0.768	0.683	0.692	0.765	0.682
p-value	0.0003	5E-13	1E-10	1E-08	2E-24	5E-22	1E-16	4E-61	3E-50
Conclusion	GMET	GMET	GMET	GMET	GMET	GMET	GMET	GMET	GMET

Note: acr=accuracy; stvy=sensitivity

After doing many iterations (10, 20, and 50) and evaluating three different model performance metrics, it was determined that the binomial GMET model outperformed the binomial GLMM model. However, it is worth noting that executing the R syntax for the GMET model took around 30 minutes longer.

The binomial-GLMM and binomial-GMET models are evaluated based on three key performance metrics: accuracy, sensitivity, and F1 score. The F1 score is a metric that quantifies the balance between accuracy and recall by taking a weighted average. This study examines five simultaneous measurements: Accuracy, Sensitivity, Precision, Recall, and F1. The three fundamental metrics indicate that the GMET binomial model outperforms the GLMM binomial model in forecasting temporary unemployment, particularly in the West Java Province.

4 Conclusion

There are 55,957 respondents participating in the *SAKerNas* 2022 survey in West Java, divided into 27 districts/cities. The unemployment rate was 43% of the total population of 55,957, corresponding to 24,155 individuals. Consequently, the percentage of participants who are employed is 57%. Therefore, the data may be steady or balanced.

GLMM and GMET have commonalities regarding fixed and random effects, but they diverge in the type of conclusions that may be drawn. GMET is a generalized linear mixed model that utilizes a tree-based approach, incorporating CART in constructing the tree.

The performance measurements utilized to evaluate binomial GMET and binomial GLMM are accuracy, sensitivity, and F1. The F1 score is calculated by taking the weighted average of accuracy and recall. The obtained result indicates that the binomial GMET model is superior. Based on all iterations (10, 20, and 50) and all three metrics of model quality, it can be concluded that the binomial GMET model is superior to the binomial GLMM model. However, it is important to note that executing the R syntax for the GMET model takes around 30 minutes longer.

Acknowledgments

The author acknowledges the *Beasiswa Pendidikan Indonesia* (BPI, Indonesian Education Scholarship) for their support in funding and the Institut Pertanian Bogor (IPB University) for facilitating the author's education.

References

1. Rusyana, A., Kurnia, A., Sadik, K., Wigena, A. H., Sumertajaya, I. M., & Sartono, B. Comparison of GLM, GLMM and HGLM in Identifying Factors that Influence the District or City Poverty Level in Aceh Province. Journal of Physics: Conference Series, 1863(1). https://doi.org/10.1088/1742-6596/1863/1/012023 (2021).
2. Hedeker, D., & Gibbons, R. D. A random-effects ordinal regression model for multilevel analysis. Biometrics, 50(4), 933–944. (1994).
3. Widyaningsih, Y., Saefuddin, A., Notodiputro, K. A., & Wigena, A. H. Nested Generalized Linear Mixed Model with Ordinal Response: Simulation and Application on Poverty Data in Java Island. AIP Conference Proceedings, 1450, 147–154. https://doi.org/10.1063/1.4724132 (2012).
4. Fontana, L., Masci, C., Ieva, F., & Paganoni, A. M. Performing learning analytics via generalised mixed-effects trees. Data, 6(7). https://doi.org/10.3390/data6070074 (2021).
5. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. Classification and Regression Trees. Biometrics, 40(3). (1984).
6. Sela, R. J., & Simonoff, J. S. RE-EM trees: A data mining approach for longitudinal and clustered data. Machine Learning, 86(2), 169–207. https://doi.org/10.1007/s10994-011-5258-3 (2012).
7. Hajjem, A., Bellavance, F., & Larocque, D. Mixed effects regression trees for clustered data. Statistics and Probability Letters, 81(4), 451–459. https://doi.org/10.1016/j.spl.2010.12.003 (2011).
8. Hajjem, A., Larocque, D., & Bellavance, F. Generalized Mixed Effects Regression Trees. Statistics and Probability Letters, 126, 114–118. https://doi.org/10.1016/j.spl.2017.02.033 (2017).
9. Fokkema, M., Smits, N., Zeileis, A., Hothorn, T., & Kelderman, H. Detecting Treatment-Subgroup Interactions in Clustered Data with Generalized Linear Mixed-Effects Model Trees. Behavior Research Methods, 50(5), 2016–2034. https://doi.org/10.3758/s13428-017-0971-x (2018).
10. Zeileis, A., Hothorn, T., & Hornik, K. Model-Based Recursive Partitioning. Journal of Computational and Graphical Statistics, 17(2), 492–514. https://doi.org/10.1198/106186008X319331 (2008).
11. Speiser, J. L., Wolf, B. J., Chung, D., Karvellas, C. J., Koch, D. G., & Durkalski, V. L. BiMM tree: a decision tree method for modeling clustered and longitudinal binary outcomes. Communications in Statistics: Simulation and Computation, 49(4), 1004–1023. https://doi.org/10.1080/03610918.2018.1490429 (2020).
12. BPS. Tingkat Pengangguran Terbuka Menurut Provinsi, 1986-2023. Jakarta. Retrieved from https://www.bps.go.id/indicator/6/543/1/tingkat-pengangguran-terbuka-menurut-provinsi.html (2023).
13. Wikipedia. West Java Province.id. Wikipedia.org. https://id.wikipedia.org/wiki/Berkas:West_Java_province.png last accessed 2023/06/14 (2020).

14. BPS Provinsi Jawa Barat. Keadaan Ketenagakerjaan Provinsi Jawa Barat Agustus 2022. Berita Resmi Statistik, (31), 1–11. (2022).
15. BPS. Instrumen: Survei Angkatan Kerja Nasional 2022. Jakarta. (2022).
16. Fokkema, M., Edbrooke-Childs, J., & Wolpert, M. Generalized Linear Mixed-Model (GLMM) Trees: a flexible decision-tree method for multilevel and longitudinal data. Psychotherapy Research, 31(3), 1–13. https://doi.org/10.1080/10503307.2020.1785037 (2020).