Bima Journal of Science and Technology, Vol. 8(1B) Apr, 2024 ISSN: 2536-6041

DOI: 10.56892/bima.v8i1B.648

Comparison of Bayesian and Classical Approches of Logistic Regression in Modeling Risk of Preterm Birth in Nasarawa State of Nigeria

Olugbo, E. J.*, Adenomon, M. O. and Nweze, N. O.

Department of Statistics, Nasarawa State University, Keffi

ABSTRACT

The accuracy of a predictive model is crucial in various fields, including classification tasks. One commonly used method for classification is logistic regression, which relies on maximum-likelihood estimation. However, there is growing interest in exploring the use of Bayesian logistic regression as an alternative approach. This interest stems from the advantages offered by the Bayesian network approach, which allows for explicit modeling of feature dependencies and the introduction of hidden nodes. Furthermore, Bayesian inference can be associated with cognitive processes, making it a potentially powerful tool for analyzing complex data. In a comparative analysis, both classical and Bayesian logistic regression models were evaluated for their performance in classification tasks using data collected from a hospital based retrospective study on postpartum mothers and their babies is confined to Two (2) Tertiary Facilities and Three (3) Secondary Facilities across the three (3) Senatorial Zones of Nasarawa state, Nigeria. The Cohort design is adopted for the study. Model prediction Measures such as R-Square, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) were used. Model Performance Measures such as Accuracy, Precision, Recall, F1 Score, Area Under Curve (AUC) were also used. Conclusively, The Bayesian Logistic Regression Model outperforms the Classical Logistic Regression Model across all evaluated metrics. It demonstrates higher accuracy, precision, recall, F1-Score, and AUC, indicating better overall predictive performance.

Keywords: Classical Logistic Regression, Bayesian Logistic Regression, Maximum Likelihood estimation.

INTRODUCTION

In exploring the methodologies for predicting preterm birth risks, the study delineates between two predominant statistical models: Bayesian logistic regression and classical logistic regression. Bayesian logistic regression, distinct for its integration of prior knowledge with current data, offers a dynamic approach to parameter estimation. This method enhances flexibility in modeling by updating parameter estimates with new data, presenting an advantage over the conventional method. Classical logistic regression, conversely, operates strictly on empirical data without incorporating prior beliefs, representing a more traditional form of analysis. Research, including Acquah (2013) and Weber-Schoendorfer et al. (2015), has assessed these methodologies in various contexts, highlighting their efficacy in predicting outcomes like disease severity and preterm birth risks.

Notably, logistic regression often utilizes maximum likelihood estimation for parameter determination. This classical approach, however, shows limitations in cases of small sample sizes or when the dependent variable exhibits imbalance, potentially leading to biased parameter estimates. Dumouchel (2012) advocates for the Bayesian approach to mitigate these issues, emphasizing its utility in small sample conditions and its foundation in probabilistic statements about unknown parameters.

DOI: 10.56892/bima.v8i1B.648

Classical binary logistic regression has been widely applied in identifying preterm birth risk factors, where Maximum Likelihood Estimation (MLE) is predominant. Yet, MLE's reliability dwindles with small samples, where Bayesian methods, leveraging posterior distributions that amalgamate observed data with prior information, exhibit superiority (Howson, 1993).

Gladence et al. (2015) compared logistic regression against Bayesian classification in machine learning, concluding logistic regression's superior performance across various metrics. This analysis, utilizing datasets from the UCI Repository, underscores logistic regression's applicability to diverse data types.

In a different vein, Santos et al. (n.d.) evaluated classical and Bayesian methods in structural models, particularly for intervention models. Their findings, from extensive Monte Carlo experiments and real-world data analysis, indicate the Bayesian method's enhanced accuracy in estimation and prediction, particularly with limited sample sizes or low signal-to-noise ratios. Yet, classical methods may prevail under conditions of large samples or high signal-to-noise ratios, with the choice of prior distribution significantly impacting Bayesian performance.

Ma et al. (2009) assessed Bayesian and classical methods in analyzing binary outcomes from cluster randomized trials, with their study revealing no significant difference in outcomes between groups. This comparison highlighted the Bayesian random-effects model's potential in such analyses, influenced markedly by the choice of prior distribution.

Kawo et al. (2018) explored anemia determinants among Ethiopian children, employing both classical and Bayesian approaches. Their findings underscore significant regional disparities in anemia prevalence, with Bayesian analysis providing consistent estimates and offering insights into parameter distribution.

Austin et al. (2000) discussed provider profiling and hospital performance, comparing frequentist and Bayesian hierarchical models. Their analysis revealed discrepancies in identifying outlier hospitals, spotlighting the need for further research into effective performance measurement methods.

RESEARCH DESIGN

This research is a retrospective hospital-based study focusing on mothers and their newborns across Nasarawa State, Nigeria, spanning two tertiary and three secondary healthcare facilities within its three senatorial zones. Utilizing a cohort design, this study specifically examines a subgroup of the population—mothers and their babies—sharing common characteristics pertinent to the research question. A cohort study, by definition, observes statistical occurrences within a particular segment of the population over time, distinguishing it from analyses that consider the broader population, thereby providing targeted insights into the factors influencing maternal and neonatal outcomes.

Sample Size

Data from maternal, fetal/neonatal, and obstetric records from the labor wards of all live births in the selected health facilities is obtained. All deliveries less than 22 weeks gestation and post term birth at greater or equal to 42 weeks gestation are excluded. All records of multiple and stillbirths are also excluded.

Study Area

Nasarawa, situated in the north-central region of Nigeria, was established on October 1, 1996, with Lafia as its capital. It shares borders with Kaduna State to the north, the

Federal Capital Territory to the west, Kogi and Benue States to the south, and Taraba and Plateau States to the east. The state spans an area of 27,117 square kilometers (10,470 square miles) and had an approximate population of 1,869,377 according to the 2006 Census. Nasarawa State is divided into three Senatorial districts: Nasarawa North, Nasarawa South, and Nasarawa West, encompassing 13 Local Government Areas (LGAs) and 16 Development Areas.

Sampling Strategy

The study utilized a cluster randomized trial approach, randomly selecting healthcare facilities based on their fulfillment of specific inclusion criteria. The selection encompassed two tertiary and three secondary healthcare facilities distributed across the state's three senatorial zones. The research utilized data concerning live births recorded between 2016 and 2020. These facilities were selected for their comprehensive databases on all live births, including detailed parameters stored in their records units. This extensive data collection on both mothers and their children provided a rich source of variables, enabling the identification of risk factors associated with preterm births in Nasarawa State.

Data Collection Procedure

Documents and records are used for data collection. Only relevant variables were extracted for the purpose of this research

Variables in the Study

The dependent variable is the live births among women in the selected facilities in the three senatorial zones in Nasarawa State of Nigeria. This is coded as '1' for those that are preterm and '0' for those that are term.

The independent variables are factors such as Age of mother, Booking status, parity, previous miscarriages, multiple or singleton gestation, gestational age, blood group, type of delivery, sex of infant, Education of parents, Body Mass Index (BMI), Income, smoking history, HIV status, Diabetes, Hypertension (Gestational or Chronic) etc. The choice of these variables is guided by literatures on the risk factors of preterm births. The table below summarizes the variables used in this research work.

Ethical Consideration

Ethical clearance was gotten from the Nasarawa State Ministry of Health and in particular the Health Research and Ethics Committee (HREC) before the commencement of data collection. A letter of permission is sought from the Hospitals Management Board, Primary Health Care Development Agency and all the Facilities that are visited for the purpose of this research. Names of the clients are not documented in the research. The protection of the privacy of research participants is ensured. Anonymity of individuals is also ensured. Adequate level of confidentiality of the research data is ensured. The information was not accessible to any other person outside the research team, and it is used only for the purpose of research.

Data Analysis

Data Analysis is performed using Stata. Bayesian Logistic regression, and Bivariate analysis is considered for identifying the risk factors associated with preterm births as a discrete and binary response variable.

MODEL SPECIFICATION

Logistic Regression

Logistic Regression Model originally developed for survival analysis that usually has output (y) in form 0 or 1 (binary) (William and Terry, 2012). Logistic regression model for a binary dependent variable is:

$$E(y) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i)} \quad (1)$$

Where $y = \begin{cases} 1; category\ A \\ 0; category\ B \end{cases}$

$x_1, x_2, \dots, x_i$ are the i predictors.

The model above can be expressed as follows:

$$E(y) = P(y = 1) = \pi$$

$$\pi = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i)} \quad (2)$$

Equation 2 can be expressed in the log-odds terms:

$$In\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i \quad (3)$$

Given some random samples $(Y_j, X_{1j}, \dots, X_{ij})$ where j=1,2,3,…,k and Y_j is a result of Bernoulli experiment with probability of success as we can see in equation (3); coefficient β_j from the model is a constant that we don't know the value and it will be estimated from the data.

Likelihood Function

The Likelihood function from the sample is:

$$L(\beta; Y) = \prod_{j=1}^{k} \pi_j^{Y_j} (1 - \pi_j)^{1-Y_j} \quad (4)$$

Or we can write it in another form,

$$L(\beta; Y) = \prod_{j=1}^{k} \left[\left(\frac{e^{\beta_0 + \beta_2 x_2 + \dots + \beta_i x_{ij}}}{1 + e^{\beta_0 + \beta_2 x_2 + \dots + \beta_i x_{ij}}} \right)^{y_j} \left(1 - \frac{e^{\beta_0 + \beta_2 x_2 + \dots + \beta_i x_{ij}}}{1 + e^{\beta_0 + \beta_2 x_2 + \dots + \beta_i x_{ij}}} \right)^{1-y_j} \right] \quad (5)$$

In Classical Logistic regression, model parameters can be determined by Maximum Likelihood Estimation (MLE) Method,

$$\sum_{j=1}^{k} \left[Y_j \log(\pi_j) + (1 - Y_j) \log(1 - \pi_j) \right] \quad (6)$$

The main foundation of the Bayesian method is the Bayes theorem. Bayes theorem can be stated as follows:

$$P(\theta/y) = \frac{P(y/\theta)P(\theta)}{P(y)}$$

Where $P(\theta/y)$ is posterior distribution.

P(θ) is prior distribution.

$P(y/\theta)$ is sampling distribution or likelihood function.

P(y) is marginal likelihood.

Prior Distribution

Prior Distribution is a distribution that gives information about the parameters. The prior ought to mirror the entire knowledge of the experts in terms of data variation and dynamic properties (Adenomon, 2017). There are several types of prior distribution.

i. Non-Informative Prior Distribution: For the selection of the prior distribution is not based on existing data.

ii. Informative Prior Distribution: This prior distribution is based on parameter value from the prior distribution. Parameter value from the previous distribution will affect the posterior distribution.

Posterior distribution

The conditional sample likelihood in equation (3) is combined with joint prior distribution using bayes theorem. Recall equation (9). The joint posterior distribution of the model parameters is:

Posterior =Prior x likelihood

$$\text{Posterior}=\prod_{p=1}^{i}\frac{1}{\sqrt{2\pi\sigma_p^2}}exp\left\{-\frac{1}{2}\left(\frac{\beta_p-\mu_p}{\sigma_p}\right)^2\right\}\times$$

$$\prod_{j=1}^{k}\left[\left(\frac{e^{\beta_0+\beta_2x_2+\ldots+\beta_ix_{ij}}}{1+e^{\beta_0+\beta_2x_2+\ldots+\beta_ix_{ij}}}\right)^{y_j}\left(1-\frac{e^{\beta_0+\beta_2x_2+\ldots+\beta_ix_{ij}}}{1+e^{\beta_0+\beta_2x_2+\ldots+\beta_ix_{ij}}}\right)^{1-y_j}\right]$$

Where the prior is the pdf of normal distribution. The marginal posterior distribution can be computed from the joint posterior distribution. The means of these distributions are the parameter estimates.

To evaluate and compare the Classical Logistic Regression Model with the Bayesian Logistic Regression Model, various predictive and performance metrics are utilized, including R-Square, Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) for model prediction measures. Model performance is assessed through Accuracy, Precision, Recall, F1 Score, and Area Under Curve (AUC), facilitating a comprehensive analysis of each model's efficacy.

Model Prediction Measures

R-Square

R-Square in logistic regression measures how well the independent variables explain variability in the dependent variable's log-odds ratios. It quantifies the proportion of variance in the response variable that can be attributed to variation in the predictors (Nhu et al. 2020).

One commonly used formula for calculating R-Square in logistic regression is Cox and Snell's pseudo-R-Squared:

$$R_{CS}^2 = 1-\left(\frac{L_0}{L_M}\right)^{\frac{2}{N}} = 1-e^{2(In(L_0)-In(L_M))/n}$$

Where L_M and L_0 are the likelihoods for the model being fitted and the null model respectively.

Mean Squared Error (MSE)

Mean Squared Error (MSE) is a commonly employed metric for evaluating the performance of predictive models, including logistic regression models. MSE measures the average squared difference between predicted and actual values. In the context of logistic regression analysis, MSE provides insights into the accuracy and reliability of model predictions (Zhang et al., 2021)

Mathematically, MSE can be expressed as:

$$MSE = \sum (y-p)^{\frac{2}{n}}$$

Where:

- y represents observed outcomes (0 or 1)

- p represents predicted probabilities

- n represents the total number of observations

DOI: 10.56892/bima.v8i1B.648

Root Mean Squared Error (RMSE)

RMSE is an evaluation metric that measures the average magnitude of errors between predicted values and actual observations in logistic regression models. It assesses how well predictions align with observed outcomes by considering both false positives and false negatives.

To calculate RMSE in logistic regression, we first obtain predicted probabilities for each observation using the fitted model equation. We then compare these probabilities with binary outcomes (0 or 1) to determine prediction errors for each observation. Finally, we take the square root of the mean squared errors to obtain RMSE (Nhu et al., 2020). Therefore, the formula of RMSE is

$$\sqrt{\frac{(\sum y_i - \hat{y}_i)^2}{N}}$$

Where, y_i is the ith measurement, $\hat{y}_i$ is its corresponding prediction and N is the number of datapoints.

Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is a measure that quantifies the average difference between predicted and observed values in logistic regression. It provides a numerical value that represents the magnitude of error in model predictions (Leffondre et al., 2013)

To calculate MAE in logistic regression, we first obtain predicted probabilities for each observation using the fitted logistic regression model. These predicted probabilities represent the estimated probability of success (e.g., occurrence of an event) given a set of predictor variables.

Next, we compare these predicted probabilities with the actual binary outcome values (0 or 1). The absolute differences between predicted probabilities and observed outcomes are calculated for each observation. Finally, we take the mean of these absolute differences to compute the MAE. (Leffondre et al., 2013).

$$MAE = \frac{1}{n}\sum_{1}^{n} |y_i - \hat{y}_i|$$

Where y_i =actual value, $\hat{y}_i$ =predicted value, n=sample size.

Akaike Information Criterion (AIC)

AIC, which stands for Akaike Information Criterion. Developed by Hirotugu Akaike, AIC quantifies the quality of a given model by balancing goodness-of-fit with parsimony using mathematical formulas (Jin et al.,2022). The formula for calculating AIC is as follows:

AIC = -2 * log-likelihood + 2 * k, where log-likelihood represents the maximized value of the logarithm of likelihood function for a particular model, and k represents the number of parameters estimated in that model.

Bayesian Information Criterion (BIC)

BIC (Bayesian Information Criterion) is a statistical measure commonly used in model selection and evaluation (Conca, et al., 2022) It is a criterion that balances the goodness of fit of a model with its complexity, aiming to find the most parsimonious yet accurate model.

Like AIC, it is appropriate for models fit under the maximum likelihood estimation framework.

The BIC statistic is calculated for logistic regression as follows:

BIC = -2 * LL + log(N) * k

Where log () has the base-e called the natural logarithm, LL is the log-likelihood of the model, N is the number of examples in the training dataset, and k is the number of parameters in the model. The score as defined above is minimized, e.g., the model with the lowest BIC is selected.

The quantity calculated is different from AIC, although can be shown to be proportional to the AIC. Unlike the AIC, the BIC penalizes the model more for its complexity, meaning that more complex models will have a worse (larger) score and will, in turn, be less likely to be selected.

Model Performance Measures

In the realm of statistical modeling, accurately evaluating model performance is pivotal. The precision of predictions from logistic regression models is indispensable for generating dependable forecasts and making well-informed choices, as highlighted by recent studies (Bokaba et al., 2022; Yacoob et al., 2019). Within logistic regression analysis, accuracy denotes the model's proficiency in correctly predicting outcomes and accurately categorizing observations into their designated groups, essentially reflecting the ratio of correct predictions to total predictions.

To gauge accuracy, a set of metrics including sensitivity (or true positive rate), specificity (true negative rate), precision (positive predictive value), and recall (also known as sensitivity) are frequently employed. These metrics shed light on various facets of the model's predictive performance, offering a nuanced understanding of the precision in the model's forecasts (Poitras and Lajoie, 2014).

Precision is defined as the model's capacity to accurately identify positive instances while limiting false positives. This metric is particularly critical in sectors like healthcare, finance, and marketing, where precise predictions are fundamental to decision-making processes. Precision is calculated as the quotient of true positives over the sum of true positives and false positives, indicating the model's efficacy in correctly identifying positive outcomes among those predicted as positive (Acharya et al., 2022).

Recall, on the other hand, gauges a model's ability to detect all pertinent positive cases from the actual positives available in the dataset, emphasizing the model's success in capturing true positives without overlooking false negatives. The recall metric, computed as the division of true positives by the aggregate of true positives and false negatives, reflects the model's competency in accurately identifying positive instances (Bokaba et al., 2022).

The F1-Score serves as a composite metric that evaluates the accuracy and efficiency of classification models, particularly beneficial in situations of class imbalance. By harmonizing precision and recall, the F1-Score delivers a comprehensive measure of model performance, calculated as 2 times the product of precision and recall divided by their sum. A higher F1-Score signifies superior performance, encapsulating both precision and recall considerations.

Lastly, the Area Under the Curve (AUC) metric assesses the logistic regression model's discriminative capacity, or its ability to distinguish between positive and negative cases. A higher AUC value, ranging from 0 to 1, suggests a model's enhanced ability to differentiate outcomes. AUC is determined by plotting the Receiver Operating Characteristic (ROC) curve, which illustrates the balance between sensitivity and specificity across various threshold settings, thus offering a graphical representation of model efficacy (Rasyid et al., 2016).

Bima Journal of Science and Technology, Vol. 8(1B) Apr, 2024 ISSN: 2536-6041

DOI: 10.56892/bima.v8i1B.648

RESULTS AND DISCUSSION

Table 1: Comparison on Prediction Performance

Criteria	Logistic model	Bayesian Logistic Model
R-square	0.2323	0.3539
MSE	0.9650	0.2812
RMSE	0.9824	0.1676
MAE	0.9651	0.5779
AIC	2716.274	2716.081
BIC	2776.945	2776.251

Source: Author's Computation 2023

Table 1 shows the prediction performance comparison between the frequentist logistic regression model and the Bayesian logistic regression model. The measures used are the R-square, Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Alkaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The model with the higher values of R-square performs better, while the one with the lowest value for the MSE, RMSE, MAE, AIC and BIC would be considered better than the other. Considering Table 4.10 above, the Bayesian satisfied the conditions stated above as it produces higher R-square and lower MSE, RMSE, MAE, AIC and BIC values compare to the Logistic regression model. In conclusion, the Bayesian Logistic regression outperform the classical logistic regression model.

Table 2: Comparison on Performance Measure

Performance measure	Logistic model	Bayesian Logistic model
Accuracy	0.3490	0.876
Precision	0.3240	0.3609
Recall	0.7000	0.9400
F1-score	0.6192	0.7085
AUC	0.35162	0.9485

Source: Author's Computation 2023

Table 2 shows the model performance measures using the accuracy, precision, recall, f1-score and the AUC value. In this section, the model with the higher value is considered better than the other. Looking at the table, Bayesian logistic model produced higher accuracy, precision, recall, f1-score and AUC values compared to the classical logistic regression model. In conclusion, Bayesian logistic regression model performs better than the classical logistic regression model. The density of the posterior means and trace plots are displayed below.

In contrasting the Bayesian Logistic Regression Model, boasting an accuracy of 87.6%, precision of 36.09%, recall of 94%, F1-Score of 70.85%, and AUC of 94.85%, against the Classical Logistic Regression Model, which exhibits an accuracy of 34.9%, precision of 32.4%, recall of 70%, F1-Score of 61.92%, and AUC of 35.16%, it's clear that each performance metric merits individual examination. The Bayesian approach significantly surpasses the Classical in terms of accuracy, showcasing superior overall prediction effectiveness for both positive and negative outcomes. In precision, the Bayesian model slightly outperforms the Classical, suggesting a higher likelihood of accurate positive predictions. Regarding recall, the Bayesian method proves more proficient at detecting true positives, capturing a larger fraction of actual positive cases. The F1-Score, representing the harmonic mean of precision and recall, is notably higher in the Bayesian model, indicating a superior balance between precision and recall. Moreover, with a considerably higher AUC, the Bayesian model demonstrates enhanced capability in differentiating between positive and negative instances. Overall, the Bayesian Logistic Regression Model exhibits superior performance across all metrics, evidencing more effective predictive accuracy, precision, recall, F1-Score, and AUC than its Classical counterpart.

DOI: 10.56892/bima.v8i1B.648

REFERENCES

Acharya, K. P., Khanal, S. P., and Chhetry, D. (2022). On the Use of Logistic Regression Model and its Comparison with Log-binomial Regression Model in the Analysis of Poverty Data of Nepal. Nepalese Journal of Statistics, 6(01), 63-79. https://doi.org/10.3126/njs.v6i01.50806

Acquah, H. D.G, (2013). Bayesian Logistic Regression Modeling via Markov Chain Monte Carlo algorithm. Journal of Social and Development Sciences., 4:193-197

Adenomon, M. O. (2017). Introduction to Univariate and Multivariate Time Series Analysis with Examples in R. Nigeria: University Press Plc.

Austin, P. C., Naylor, C. D., and Tu, J. V. (2001). A comparison of a Bayesian vs. a frequentist method for profiling hospital performance. Journal of Evaluation in Clinical Practice 7 (1):35-45.

Bokaba, T., Doorsamy, W., and Paul, B. S. (2022). Comparative Study of Machine Learning Classifiers for Modelling Road Traffic Accidents. Applied Sciences, 12(2), 828. https://doi.org/10.3390/app12020828

DuMouchel W. (2012) Multivariate Bayesian Logistic Regression for Analysis of Clinical Study Safety Issues. Statistical Science, Vol 27, No3, pp 319-339, 2012.

Gladence, L.M., Karthi, M. and Maria-Anu, V. (2015). A Statistical Comparison of Logistic Regression and Different Bayes Classification Methods for Machine Learning. ARPN Journal of Engineering and Applied Sciences. Vol. 10 No. 14, p. 5547-5953.

Howson C, Urbach P. (1993) Scientific reasoning: The Bayesian approach 2ed. Open Court Chicago.

Kawo, K.,N., Asfaw, Z.G., Yohanes, N. (2018). Multilevel Analysis of Determinants of Anemia Prevalence among Children Aged 6–59 Months in Ethiopia: Classical and Bayesian Approaches

Leffondre, K., Jager, K. J., Boucquemont, J., Stel, V. S., and Heinze, G. (2013). Representation of exposures in regression analysis and interpretation of regression coefficients: basic concepts and pitfalls. Nephrology Dialysis Transplantation, 29(10), 1806–1814. https://doi.org/10.1093/ndt/gft500

Ma, J., Thabane, L., Kaczorowski, J., Chambers, L., Dolovich, L., Karwalajtys, T., Levvitt, C. (2009). Comparison of Bayesian and classical methods in the analysis of cluster randomized controlled trials with a binary outcome: The Community Hypertension Assessment Trial (CHAT). BioMed Central Ltd.

Rasyid, A. R., Bhandary, N. P., and Yatabe, R. (2016). Performance of frequency ratio and logistic regression model in creating GIS based landslides susceptibility map at Lompobattang Mountain, Indonesia. Geoenvironmental Disasters, 3(1). https://doi.org/10.1186/s40677-016-0053-x

Santos, T.R., Franco, G.C., Gamerman, D. (nd). Comparison of classical and Bayesian approaches for intervention analysis in structural models.

Weber-Schoendorfer, C., Oppermann, M., Wacker, E., Bernard, N., Beghin, D., Cuppers- Maarschalkerweerd, B., Richardson, J L., Rothuizen, L E., Pistelli, A., Malm, H., Eleftheriou, G., Kennedy, D., Kadioglu, M., Meister, R.,

DOI: 10.56892/bima.v8i1B.648

and Schaefer, C. (2015, May 28). Pregnancy outcome after TNF-α inhibitor therapy during the first trimester: a prospective multicenter cohort study. https://scite.ai/reports/10.1111/bcp.12642

Yaacob, W. F. W., Nasir, S. A. M., Yaacob, W. F. W., and Sobri, N. M. (2019). Supervised data mining approach for predicting student performance. Indonesian Journal of Electrical Engineering and Computer Science, 16(3), 1584. https://doi.org/10.11591/ijeecs.v16.i3.pp1584-1592.

Zhang, L.X, Sun, Y., Zhao, H., Zhu, N., Sun, X.D., Jin, X., Zou, A.M., Mi, Y., Xu, J.R. (2017). A Bayesian Stepwise Discriminant Model for Predicting Risk Factors of Preterm Premature Rupture of Membranes: A Case-control Study. Chin Med J 2017;130:2416-22.