Zhongguo Kuangye Daxue Xuebao
Journal of China University of Mining and Technology

2025 | Vol 30 | Issue 1 | Page 48-59
Journal Homepage: https://zkdx.ch/
DOI: 10.1654/zkdx.2025.30.1-6

Bayesian Modelling of Weibull Distribution for Predicting Student Dropout Rates: A Case of Higher Education Institutions in Nepalese University

Govinda Prasad Dhungana*
Department of Statistics, Tribhuvan University, Birendra Multiple Campus, Nepal
*Corresponding author

Arun Kumar Chaudhary
Department of Management Science Nepal Commerce Campus, Tribhuwan University, Nepal

Ramesh Prasad Tharu
Department of Statistics, Tribhuvan University, Mahendra Multiple Campus, Nepalgunj, Nepal

Vijay Kumar
Deen Dayal Upadhyaya Gorakhpur University, Gorakhpur, India

Abstract
Higher education contributes to knowledge creation, and the enhancement of essential skills. Despite its crucial role, dropout rates in higher education affect all universities globally, impeding the effective transmission of knowledge and skills to learners. Hence, this study aims to predict the dropout rates in higher education institutions by using Bayesian modelling of Weibull distribution.

This study utilized a quantitative research design. To ensure representativeness across different university and colleges, 27 colleges were randomly selected. The Bayesian model of the Weibull distribution emerges as an acceptable model for predicting dropout rates. The model is validated using graphical methods such as P-P and Q-Q plots, Kolmogorov-Smirnov, Anderson-Darling, and Cramer-Von Mises tests. In Bayesian modelling, the MCMC simulation method is implemented using the Stan package. Model accuracy is evaluated through trace plots, ergodic mean plots, autocorrelation plots, BGR plots, n_effect, and Rhat. Ultimately, Bayesian modelling of the Weibull distribution emerged as an alternative model for prediction.

The study predicts a dropout rate of approximately 26%, indicating that one in every four students annually drops out of higher education in Nepal. This finding poses a significant challenge to both Nepalese universities and the higher education system of Nepal.

Keywords
Bayesian Modelling, Student Dropout Rate, Higher Education, Nepal, Weibull Distribution

1. Introduction
Higher education is crucial for the creation, dissemination, and development of knowledge and skills. Universities are deeply committed to fostering the development of these competencies in their students. However, during this process, some students permanently leave their educational institutions, which continues to be a significant challenge in the field of education (Lacave et al., 2018; Subedi, 2022;). Student dropout is a widespread concern across all universities, leading to economic losses, social consequences, and potential psychological challenges for students. Likewise, inadequate orientation; desired job opportunity; university admission pathway; academic adjustment; time management; study technique; family characteristics; inconsistency between prior knowledge and university studies (Chauhan, 2008; Lassibille & Navarro Gómez, 2008; Paura & Arhipova, 2014; Perchinunno et al., 2021; Stăiculescu & Elena Ramona, 2019). As a result, it is a key metric for evaluating higher education institutions and has drawn significant attention from both educators and policymakers (Aulck et al., 2017). This issue is particularly increasing in Nepalese university. Various factors have contributed to this issue, including dissatisfaction with the current educational system, limited availability of part-time employment opportunities during studies, and insufficient job prospects after graduation. To address the

prevalent issue of dropout rates, the University Grants Commission (UGC) has launched an initiative for entrepreneurship and self-employment. As part of this initiative, the UGC provides seed money as loans to 50 students annually, with a broader goal of reaching 500 students per year. Additionally, some colleges have established incubation centers to foster entrepreneurship and self-employment among their students (UGC, 2024). However, a significant number of students remain underserved by these programs. So that, it is emerging issue in Nepal therefore, various approaches and techniques have been applied to understand and predict dropout rates by different researcher. Some research efforts have relied on the creation of surveys or questionnaires for students, followed by the application of descriptive statistics to analyze the data. However, these methods often struggle to accurately predict students dropping out, resulting in inefficiencies and limited effectiveness (Kostopoulos et al., 2015). Even when at-risk students are identified, it is frequently too late to prevent dropout. Therefore, the prediction and prevention of dropout, particularly in Nepalese University, employed the Bayesian modelling of Weibull distribution.

Bayesian analysis has emerged as a valuable tool for understanding student dropout in higher education. Several studies have utilized Bayesian techniques to identify dropout rates and the factors contributing to them. A study like as Bayesian techniques to identify factors contributing to dropout rate in Mexico's higher education system (De La Cruz et al., 2024). This model employed to uncover patterns and influences on dropout rates. Similarly, Bayesian analysis of the Exponentiated Weibull Weibull distribution has been used to predict the student dropout rate. It is a probabilistic approach to analyzing complex datasets and providing educational institutions with valuable insights to address the decrease of student. Findings revealed that the median dropout time, estimated using this distribution, indicating that instructors should monitor student attendance closely during this period (Alzahrani, 2024).

In another study, Bayesian methods were applied to analyze factors influencing dropout rates in computer science programs, with results showing that the best-fitting model was obtained using these probabilistic approaches (Lacave et al., 2018). Additionally, a Bayesian network model was used to predict dropout times and graduation rates. The study suggested that graduation rates could be predicted based on factors such as religious affiliation, the proportion of students enrolled full-time, socioeconomic status, enrollment size, and institutional revenues and expenditures (Crisp et al., 2018).However, notable literature on modelling of student dropout in Nepal is scarce. Therefore, this study aims to apply Bayesian Modelling of Weibull distribution to predict student dropout rates.

The motivation for this study lies in the strength of Bayesian modelling and the flexibility of the Weibull distribution. The use of Bayesian methods allows for the incorporation of prior knowledge and uncertainty, offering a probabilistic approach that is well-suited to the complex nature of educational data. The study ultimately provides the critical methodology of predictive tools to better understand and manage dropout rates. Also, this study seeks to provide valuable insights into the student dropout rates in Nepalese universities. Furthermore, in sections 2 and 3, we present the methodology and analysis of data; in section 4, we discuss the findings of the study, while in section 5, we present the conclusions of the study.

2. Material and Method

2.1 Study Design

A quantitative, cross-sectional research design was adopted for this study. Bayesian modelling of the Weibull distribution was employed to predict student dropout rates within higher education institutions of a Nepalese university.

2.2 Weibull Distribution

The Weibull distribution, a positively skewed two-parameter distribution, is widely applicable due to its enormously versatile (Garrido et al., 2021). Such a skewed distribution has been used in modelling in various areas, including engineering, actuarial science, environmental science, medical science, biology, demography, economics, finance, survival data analysis, dentistry, telecommunications, and other application areas (Dhungana & Kumar, 2022).

The proposed distribution having the cdf, $F(x) = 1 - e^{-\left(\frac{x}{\beta}\right)^{\alpha}}$ (1)

and corresponding pdf, $f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}; x > 0, \alpha > 0, \beta > 0.$ (2)

Where, α is shape parameter and β is scale parameter. The pdf curve of proposed distribution has also positively skewed in shape. Similarly, hazard rate function is the probability of failure during very small interval, assuming that individual has survived to the beginning of the interval, which is $h(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}$ (3)

The common types of hazard function are monotonically increasing, bathtub shaped, monotonically decreasing, inverse bathtub shaped, J shape and constant hazard function (Lawless, 2011). (Figure 1)

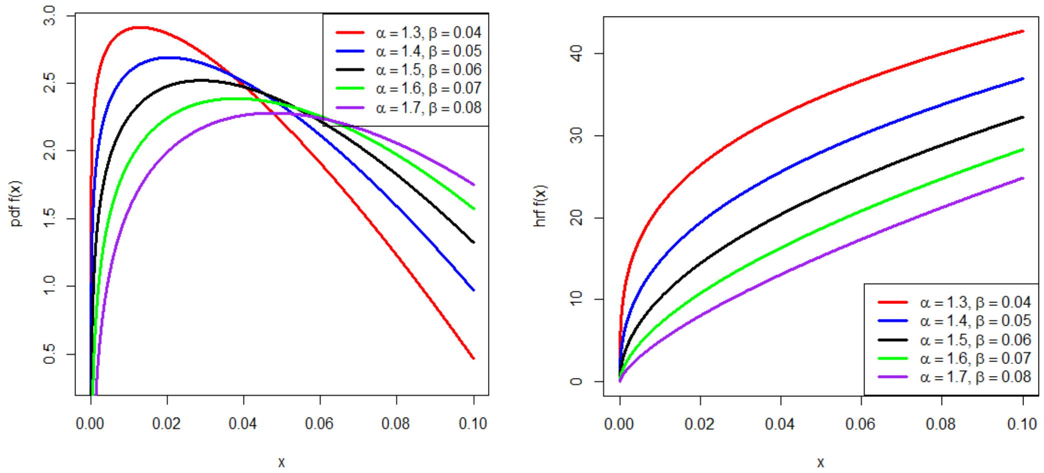


Fig. 1 Probability density function (left panel) and hazard rate function (right panel)

2.3 Parameter Estimation

We used the maximum likelihood estimates (MLEs) to determine the model parameters and asymptotic confidence intervals. Let, $\underset{\sim}{x} = (x_1, \ldots, x_n)$ observed sample from a distribution with probability density function (2), the likelihood function $\ell(\alpha, \beta)$ is given by,

$$\ell\left(\alpha, \beta / \underset{\sim}{x}\right) = \left(\frac{\alpha}{\beta}\right)^n \prod_{i=1}^{n} \left(\frac{x_i}{\beta}\right)^{\alpha-1} e^{-\sum_{i=1}^{n}\left(\frac{x_i}{\beta}\right)^{\alpha}} \tag{4}$$

Which is equivalent to log likelyhood function,

$$\ell\left(\alpha, \beta / \underset{\sim}{x}\right) = n \ln\left(\frac{\alpha}{\beta}\right) + (\alpha - 1)\ln\left(\frac{\sum_{i=1}^{n} x_i}{\beta}\right) - \sum_{i=1}^{n}\left(\frac{x_i}{\beta}\right)^{\alpha} \tag{5}$$

To estimate the parameters, the maximum likelihood estimation technique was employed by partially differentiating equation (5) with respect to the parameters and equating to zero.

$$\frac{\partial \ln(\ell)}{\partial \alpha} = \frac{n}{\alpha} + \ln\left(\frac{\sum_{i=1}^{n} x_i}{\beta}\right) - \sum_{i=1}^{n}\left(\frac{x_i}{\beta}\right)^{\alpha} \ln\left(\frac{\sum_{i=1}^{n} x_i}{\beta}\right) \tag{6}$$

$$\frac{\partial \ln(\ell)}{\partial \beta} = -\frac{n}{\beta} + \frac{1-\alpha}{\beta} - \frac{\alpha}{\beta}\sum_{i=1}^{n}\left(\frac{x_i}{\beta}\right)^{\alpha} \tag{7}$$

Finally, solve non-linear equations $\frac{\partial \ln(\ell)}{\partial \alpha} = 0$, $\frac{\partial \ln(\ell)}{\partial \beta} = 0$, and estimate ($\hat{\alpha}$, and $\hat{\beta}$) for parameters (α, and β). Furthermore, the asymptotic normality of MLEs, approximate $100(1-\gamma)\%$ confidence intervals of α, β can be constructed as: $\hat{\alpha} \pm z_{\gamma/2} SE(\hat{\alpha})$, $\hat{\beta} \pm z_{\gamma/2} SE(\hat{\beta})$, where, $z_{\gamma/2}$ is the upper percentile of standard normal variate.

2.4 Bayesian Modelling of Weibull distribution

The Bayesian model is constructed by specifying the prior distribution for the model parameters α and β, and then multiplying with the likeyhood function to obtain the posterior distribution function.

- Probability model: $f(x / \alpha, \beta)$
- Prior distribution: $p(\alpha, \beta)$
- Data: $\underset{\sim}{x} = (x_1, \ldots, x_n)$

The posterior distribution is,

$$p\left(\alpha,\beta / \underset{\sim}{x}\right) \propto \ell\left(\underset{\sim}{x}/ \alpha,\beta\right).p(\alpha)\,p(\beta) \tag{8}$$

Where, $p(\alpha)\,p(\beta)$ were prior distribution, $\ell\left(\underset{\sim}{x}/ \alpha,\beta\right)$ is the likelihood function of Weibull distribution.

2.5 Prior Distribution

In Bayesian approach, $p(\alpha)\,p(\beta)$ as prior distribution. Commonly, the prior distribution are categories as; non-informative priors, a subjective prior, informative prior, and conjugate priors (Depaoli et al., 2020). There was no unique method to select the best prior distribution (Udomvisawakal, 2021). However, the dataset is positively skewed, and the dropout rate shows an increasing trend in Nepalese higher education, as international universities are preferred by Nepalese students. Therefore, the researcher selects an informative prior in the form of a gamma distribution, which reflects the underlying characteristics of the dataset.

So that, $\alpha \sim G\left(k_1,\theta_1\right).\,\beta \sim G\left(k_2,\theta_2\right)$; having the pdf,

$$p(\alpha) = \frac{1}{\Gamma k_1 \theta_1^{k_1}} \alpha^{k_1-1} e^{-\frac{\alpha}{\theta_1}}; \alpha > 0, \theta_1 > 0, k_1 > 0 \tag{9}$$

$$p(\beta) = \frac{1}{\Gamma k_2 \theta_2^{k_2}} \beta^{k_2-1} e^{-\frac{\beta}{\theta_2}}; \beta > 0, \theta_2 > 0, k_2 > 0 \tag{10}$$

2.6 Posterior Distribution

Combining the likelihood function with prior via Baye's theorem yield the posterior proportionality as,

$$p\left(\alpha,\beta / \underset{\sim}{x}\right) \propto \frac{\alpha^{n+k_1-1}\beta^{k_2-n\alpha-1}}{\Gamma k_1 \Gamma k_2 \theta_1^{k_1}\theta_2^{k_2}} \prod_{i=1}^{n} x_1^{\alpha-1} e^{-\sum_{i=1}^{n}\left(\frac{xi}{\beta}\right)^{\alpha} - \frac{\alpha}{\theta_1} - \frac{\beta}{\theta_2}} \tag{11}$$

The posterior distribution is clearly complex for analytical inference. Therefore, we propose using the Makov Chain Monto Carlo **(**MCMC) method to simulate samples from the posterior, so that sample-based inference can be easily drawn.

2.7 Makov Chain Monto Carlo (MCMC)

MCMC is a widely used method for estimating parameters and obtaining better approximations of the target posterior distribution. Among its various techniques, we employ the No-U-Turn Sampler (NUTS), an extension of Hamiltonian Monte Carlo (HMC), for this study (Hoffman & Gelman, 2014). The NUTS algorithm is implemented in the Stan package in R, or *rstan* (Stan Development Team, 2023).

2.8 Data Collection Technique

This research was conducted in Chitwan district of Nepal, where 47 colleges are located. Among them, 7 are constituent colleges, 17 are community colleges, and 23 are private campuses affiliated with different universities (UGC, 2024). To ensure the representativeness of different universities and college types (constituent, community, and private) of Nepal, 27 (57.44%) colleges were randomly selected.

This study focused on investigating dropout rates in master's and bachelor's degree programs at various colleges. Data was collected from the administrative departments of these institutions for the academic years 2020/2022 and 2021/2023 for master's programs, and 2018/2022 and 2019/2023 for four-year bachelor's programs.

Dropout students were defined as those who enrolled in a college or university in the first intake but did not submit the final exam form at the end of the year/semester/trimester. This indicated a discontinuity in their academic pursuits. Hence, total dropout student was found of each year/semester/trimester and aggregated the total number of dropouts for each program. To ensure data accuracy, students who transferred from one college to another were excluded from the dropout analysis. Data was collected as following flow diagram (Figure 2).

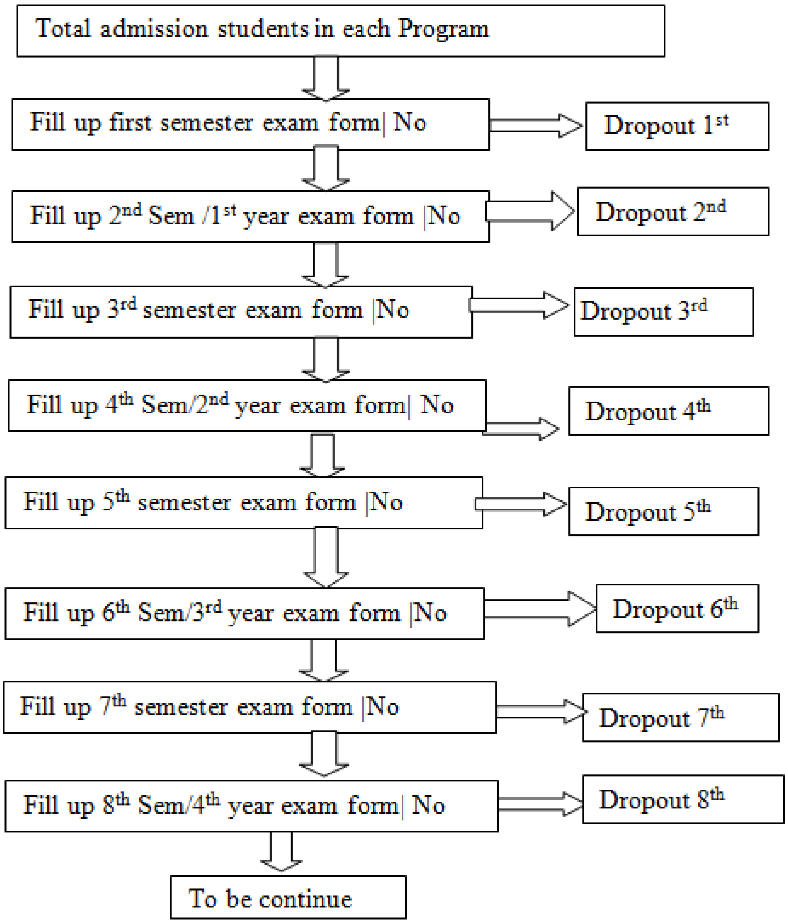


Fig. 2 Flow diagram of data collection procedure

The dropout rate ($\hat{\theta}$), was defined as the proportion of students who left the course before completing their degree. Therefor,

$$\hat{\theta} = \frac{\sum \text{Students who permanently left the college without completing their own degree}}{\text{Total number of intake of students in first year, semester, or trimester}} \times 100 \qquad (12)$$

3. Data Analysis

3.1 Descriptive Data Analysis

Based on analysis, the overall dropout rate was 37.84% (95% CI: 36.1% - 39.6%). Moreover, the findings revealed that the humanities discipline had a higher dropout rate of 40.04%, while bachelor's degree programs experienced a dropout rate of 39.50%. Specifically, Pokhara University exhibited a comparatively lower dropout rate of 29.94%. In contrast, the annual program showed a significantly higher dropout rate of 48.07% compared to semester and trimester programs. (Table 1).

Table 1 Dropout Rate of students according to program, university, academic level

Characteristics	Dropout Rate	95% Confidence Interval
Total Dropout Rate	37.84	36.1 - 39.6
University		
Tribhuvan	38.26	36.40-40.16
Pokhara	29.94	23.51-37.27
Purbanchhal	38.86	32.25-45.80
Level of Program		
Master	32.20	28.69-35.87
Bachelor	39.50	37.50-41.54
Faculty		
Science and Technology	31.61	26.88-36.64
Management	38.65	36.21-41.12
Humanities	40.07	35.98-44.27
Education	37.54	31.98-43.36
Program		
Annual	48.07	45.63- 50.50
Semester/trimester	25.08	22.76- 27.50

3.2 Analysis by Weibull Distribution

3.2.1 Exploratory Data Analysis

Exploratory Data Analysis (EDA) was a vital step in the analysis, involving the examination and summarization of the primary characteristics and patterns in the dataset (Camizuli & Carranza, 2018). The mean and median were found to be 28.34 and 23.75, respectively (Min = 2.13, Q_1 = 15.69, Q_3 = 41.55, Max = 58.68). The box plot revealed that dropout rates varied from the median to the third quartile compared to the median to the first quartile. The histogram, kernel density plot, box plot, and numerical summary from the EDA suggested that the data deviated from the normality assumption. The TTT plot exhibited a bathtub-shaped curve, which is a common term in reliability engineering and hazard modelling. From the exploratory data analysis, the shape of the data was positively skewed (Figure 3).

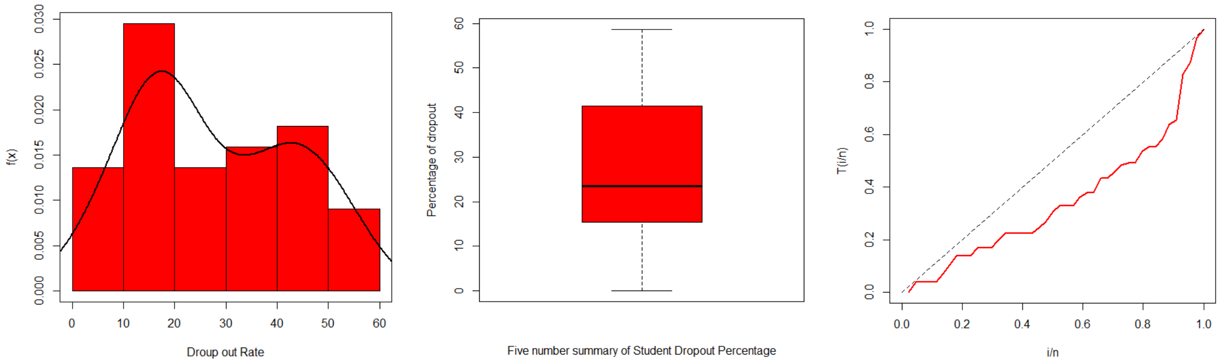


Fig. 3 Histogram with kernel density (left), box plot (middle) and TTT plot (right) panel

3.2.2 Parameter Estimation

The estimated parameters and their corresponding standard errors were obtained using the maximum likelihood estimation method with the "BFGS" method in the R in *maxlik* package (R core team, 2023). After, the estimation, we observed that both shape and scale parameters were statistically significant (p-value<0.001). In proposed Weibull distribution α is shape parameter and β is scale parameter. The coefficient of α =1.95 $(1<\alpha<2)$ indicate the increasing the hazard rate. This finding revealed that the dropout rate increased over time, suggesting that students were more likely to drop out as they progressed through their studies. Likewise, the coefficient of β =31.96 indicating that the dropout rate continued over a long time period. The higher the value of β suggested greater resilience among students initially, with an increased risk of dropout occurring only after a certain point (Table 2)(Lawless, 2003; Bishop & Dey, 2005).

Table 2 Estimated parameter value with standard error

Parameters	Estimated Value	Standard Error (SE)	p-value
α	1.95	0.24	<0.001
β	31.96	2.81	<0.001

3.2.3 Validation of Model

Furthermore, P-P and Q-Q plots demonstrated that the theoretical distribution and empirical distribution were closely aligned (Figure 4). Likewise, Kolmogorov-Smirnov (D = 0.10057, p-value = 0.419), Anderson-Darling test (An = 0.5475, p-value = 0.6979), Cramer-von Mises test (Ω_2 = 0.095922, p-value = 0.6074) of proposed distribution reveled that model was perfectly fitted in a given data set. Based on these findings, the Weibull distribution was recommended as the best model for predicting the dropout rate.

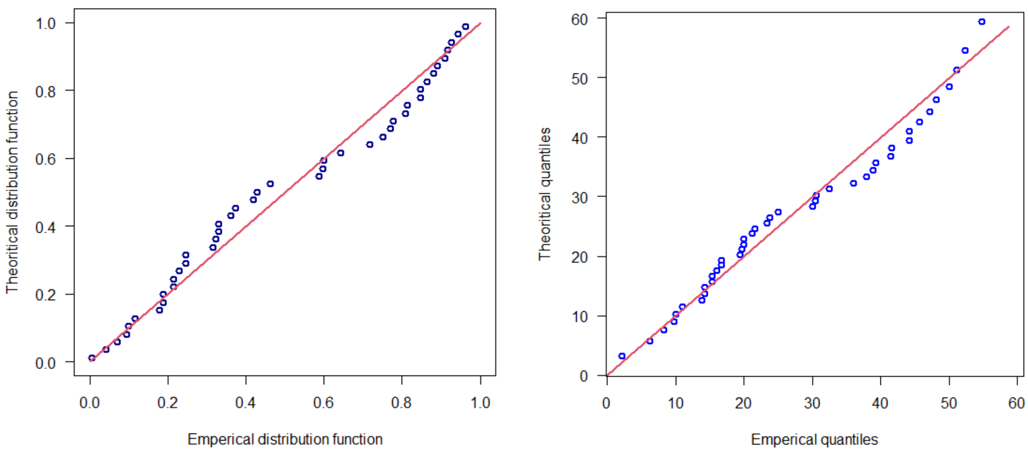


Fig. 4 P-P plot (left panel) and Q-Q plot (right panel) for model validation

3.3 Analysis of Bayesian Model of Weibull distribution

3.3.1 Model Analysis

The model was executed to generate two Markov chains, each chain consisting of 5,000 samples, with the first 1,000 samples used for warm up. To draw the posterior sample, firstly we take prior distribution $\alpha \sim G\,(2.0,0.1)$ and $\beta \sim G\,(2.0,0.1)$. Therefore, posterior sample drawn from first chain $\left(\alpha_1^{(j)},\beta_1^{(j)}\right), j=1,2,...5000$ to second chain $\left(\alpha_2^{(j)},\beta_2^{(j)}\right)$, $j=1,2,.....5000$. Initially, we extracted the posterior samples and then tested for convergence using trace plots, ergodic mean plots, and autocorrelation plots. Additionally, we measured the effective sample size (n_eff) and Rhat.

3.3.2 Convergence diagnostics

Trace Plot: Figure 5 illustrates the sequential manifestation of model parameters, revealing oscillations around a horizontal line without any trend. This observation suggests that the Markov chain is most likely to be sampling from the stationary distribution and exhibiting satisfactory mixing properties.

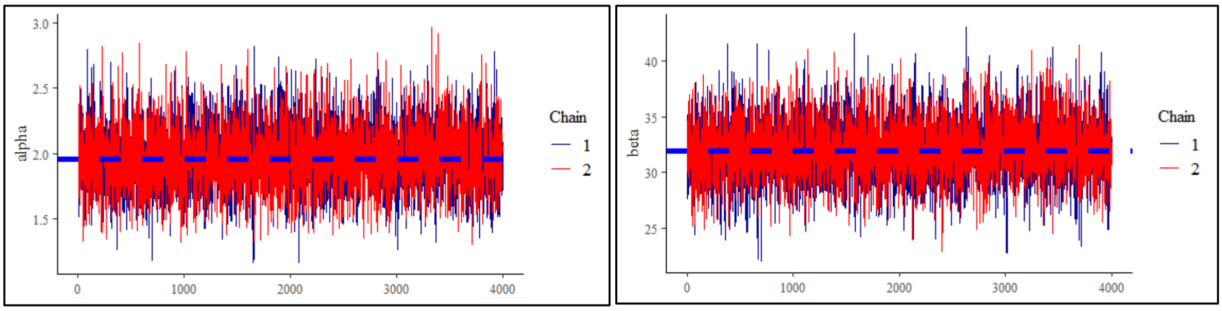


Fig. 5 Sequential realization of the parameter α and β

Running Mean (Ergodic mean) Plot: We create a time series plot, indexed by iteration number, illustrating the running mean for each parameter within the chain. The running mean is derived by computing the mean of all sampled values up to and including the iteration under consideration. The discerned convergence pattern, evident in the ergodic average, signifies the convergence of the Markov chain (Figure 6).

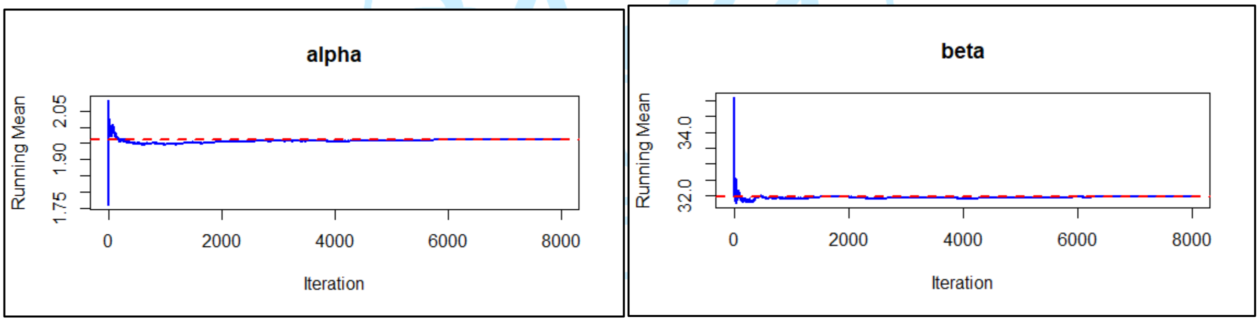


Fig. 6 The Ergodic mean plot of α (left panel), β (right panel)

Autocorrelation: The graph of autocorrelation is almost negligible; hence we consider an independent sample from the posterior distribution (Figure 7).

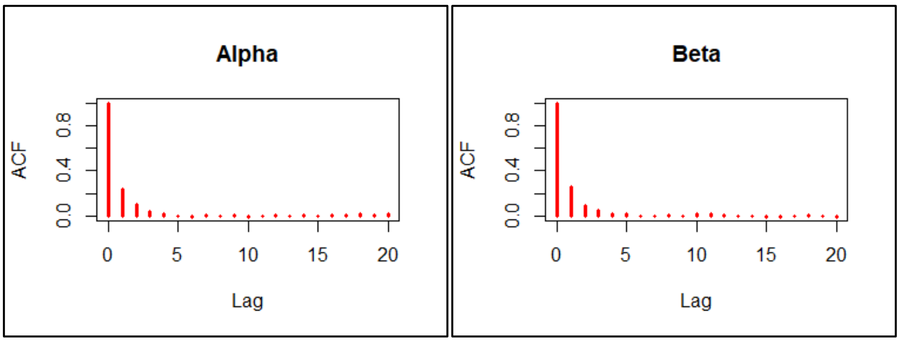
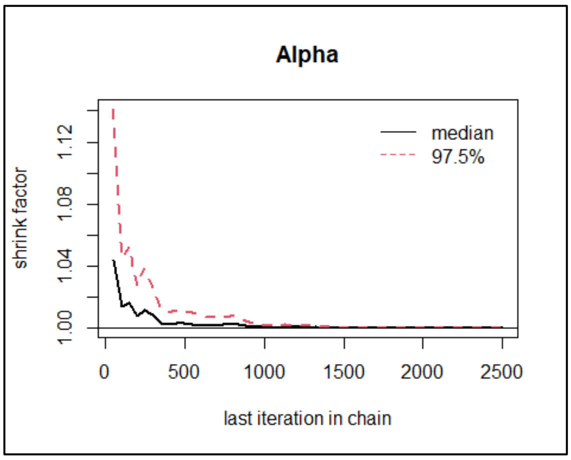


Fig. 7 Auto correlation of posterior sample of $\hat{\alpha}$ (left panel), $\hat{\beta}$ (right panel)

Brooks-Gelman-Rubin (BGR) diagnostic: The Brooks, Gelman, and Rubin convergence diagnostic is designed to assess the convergence of two or more parallel chains, each initialized with over dispersed starting values relative to the target distribution. The finding reveled that convergence is indicated when the line approaches 1. As shown in Figure 8, convergence is achieved, allowing us to proceed with obtaining the posterior summary statistics.



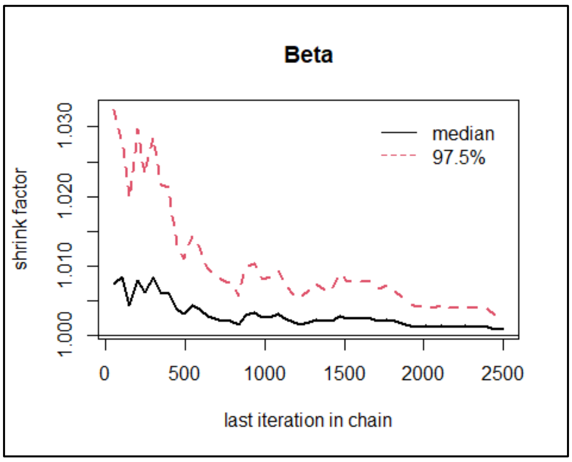


Fig. 8 BGR convergence diagnostic of $\hat{\alpha}$ (left panel), $\hat{\beta}$ (right panel)

3.3.3 Posterior Analysis

The numerical summary of $\left(\alpha_1^{(j)}, \beta_1^{(j)}\right)$;j=1,2,…,5000 from chain 1 and $\left(\alpha_2^{(j)}, \beta_2^{(j)}\right)$;j=1,2,…,5000 from chain 2 were presented in Table 3. The MCMC results for the proposed distribution included the posterior mean, standard deviation, first quartile, third quartile, 2.5th percentile, 97.5th percentile, number of effective sample size (n_eff), and the estimated potential scale reduction statistic (Rhat). The interpretation of n_eff indicated that all parameters had an effective sample size for estimating the posterior mean exceeding 10% of the total sample size, reflecting sampling efficiency. The analysis showed that the samples provided effective information for parameter estimation. Additionally, the Rhat values below 1.1 confirmed the convergence of the chains, validating the reliability of the MCMC results (Table 3).

Table 3 Numerical summaries based on MCMC sample of posterior characteristics

Characteristics	**Chain 1**		**Chain 2**	
	α	β	α	β
Mean	1.9585	32.0972	1.9651	31.9291
SD	0.2359	2.6596	0.2417	2.6721
2.5th Percentile	1.5219	27.2776	1.5261	26.8527
First Quartile	1.7919	30.3484	1.7927	30.0695
Median	1.9561	31.9426	1.9540	31.8436
Third Quartile	2.1127	33.8063	2.1233	33.6541
97.5th Percentile	2.4510	37.8194	2.4659	37.3712
n_eff	5593.7878	5314.6984	5593.7878	5314.6984
Rhat	0.99989	1.0005	0.99989	1.0005

<u>**Interpretation of Posterior Finding:**</u> In accordance with Chen and Shao, posterior parameter estimates were computed, yielding estimated values of $\hat{\alpha}$ =1.9634 and $\hat{\beta}$ = 32.0560. The findings revealed that the shape parameter of the Weibull distribution ($\hat{\alpha} > 1$) suggested an increasing hazard rate, indicating that the dropout rate increased over time. In other words, students were more likely to drop out as time progressed. Similarly, the posterior estimate ($\hat{\beta}$ = 32.0560) indicated that the dropout rate exhibited a relatively wide spread. Additionally, a 95% symmetrical confidence interval and a 95% Highest Posterior Density (HPD) interval were derived for each parameter (Chen & Shao, 1999). The calculated posterior likelihood associated with the model was -171.886325 (Table 4).

Table 4 Posterior Estimated parameter value with 95% symmetric and credible intervals

Posterior parameters	$\hat{\alpha}$	$\hat{\beta}$
Estimated Average value	1.9634	32.0560
95% Credible Interval	1.5109, 2.4629	26.9204, 37.5100
95%HPD Credible interval	1.4704, 2.4183	26.8625, 37.40411

Furthermore, a visual summary was created using density plots with histograms for each model parameter. This summary includes estimated parameter value and their associated 95% HPD intervals. The resulting visual summary serves as a diagnostic tool, providing insights into the shape and characteristics of the posterior distribution. This analysis aids in assessing the underlying assumptions of normality (Figure 9).

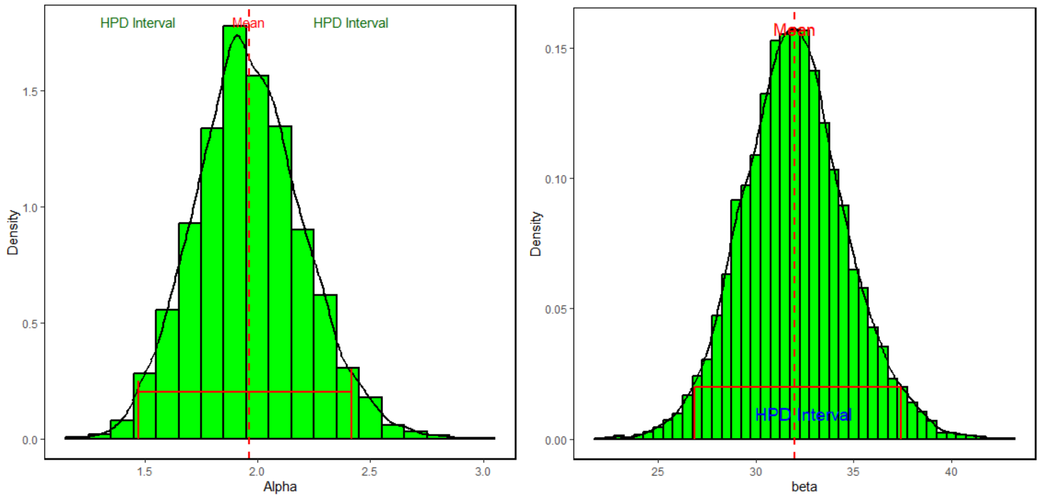


Fig. 9 Marginal posterior density, 95% HPD interval of $\hat{\alpha}$ (left panel), $\hat{\beta}$ (right panel)

We utilized a graphical approach to compare Bayesian estimates with MLEs. Figure 10 (left panel) presents the density functions $f(x/\hat{\alpha},\hat{\beta})$ derived from the MLEs and the Bayesian estimates with posterior means, $2.5^{th}, 25^{th}, 75^{th}$ and 97.5^{th} quantile using MCMC sampling of $\left(\alpha_1^{(j)},\beta_1^{(j)}\right)$;j=1,2,…,5000. From the figure, it is clear that the MLEs and Bayesian estimates align closely, both providing a good fit to the data. Similarly, we plot the reliability function of proposed distribution which has also provided a meaningful benchmark. The figure illustrates that the reliability estimates obtained via MCMC closely match the empirical reliability estimates, further validating the model's effectiveness (Figure 10, right panel).

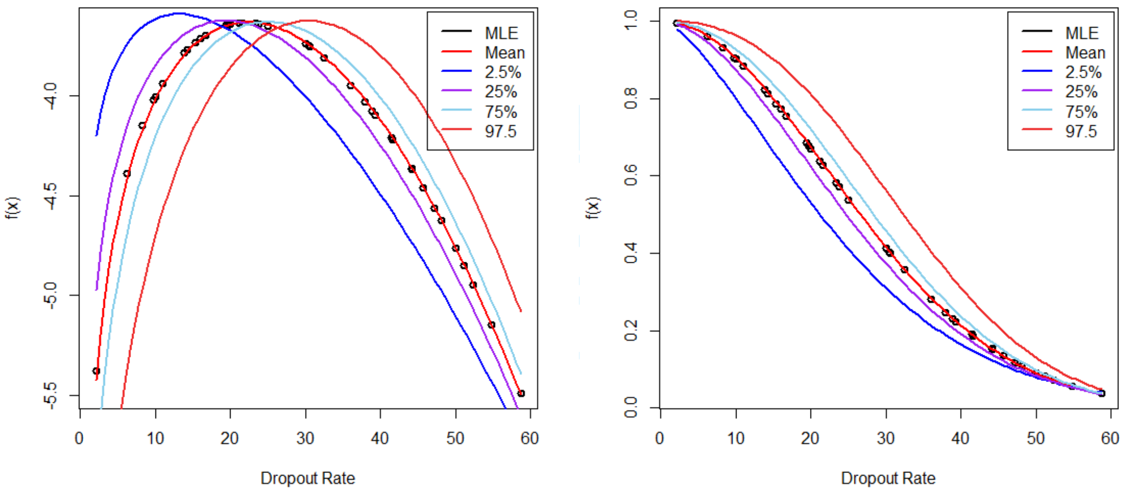


Fig. 10 Comparison with MLE and Bayesian estimates of posterior finding by density plot (left panel) and reliability plot (right panel)

3.4 Prediction of Student Dropout Rate

Based on the above-mentioned criteria, the Bayesian modelling of the Weibull distribution proves to be effective for predicting student dropout rates. To facilitate dropout rate predictions, we proposed generating random samples from the random deviate generation function of the specified distribution. The random deviate generation function for the proposed model is,

$$x = \beta\left[-\ln(1-u)\right]^{1/\alpha}; 0 < u < 1 \qquad (13)$$

We predict the dropout rate using the estimated parameter values applied in the random deviate generation function for the future of higher education institutions in Nepal from the MLE as well as Bayesian modelling. The average dropout rate is 27.63 % from the MLE (Min = 3.612, Q_1 = 16.087, Median=24.315, Q_3 = 38.076, Max = 66.774, SD = 15.049) % and 26.75% from the Bayesian modelling (Min = 3.379, Q_1 = 14.400, Median=25.074, Q_3 = 34.805, Max = 67.021, SD = 16.068) %. As a result, the predicted values from both techniques were closely aligned, with a difference of no more than 1%. The findings from the EMIS report for 2021/22 of the UGC indicated that approximately 579,488 students were enrolled in higher education institutions across Nepal (UGC, EMIS report, 2024). Our study predicts that the dropout rate is 26.75%, suggesting that around 1, 55, 013 students/year are discontinuing their studies before completion. This projection highlights that nearly one in four students/year leaves higher education prematurely. This significant dropout rate has serious implications for the country, particularly in terms of the loss of human capital. Therefore, to address this issue, it is crucial to understand the underlying reasons why students are leaving higher education prematurely. By identifying the factors contributing to the dropout rate, policymakers and educational institutions can develop targeted strategies to improve retention.

4. Discussion

This study utilizes Bayesian modelling of the Weibull distribution to predict student dropout rates. This approach aligns with existing research that has applied Bayesian methods for similar purposes, such as analyzing university dropout rates in Mexico (De La Cruz et al., 2024). Additionally, Bayesian analysis of the Exponentiated Weibull distribution has been employed to predict dropout rates in online class module (Alzahrani, 2024). Likewise, non-Bayesian models have also been used for dropout prediction. For instance, Rabelo & Zárate (2024) utilized an ensemble model combining Logistic Regression, Neural Networks, and Decision Tree methods to predict dropout rates among higher education students. Similarly, Niyogisubizo et al. (2022) proposed a novel stacking ensemble model that combines Random Forest (RF), Extreme Gradient Boosting (XGBoost), Gradient Boosting (GB), and Feed-forward Neural Networks (FNN) to predict student dropout rates in university classes. Hence, in available literature demonstrate that both Bayesian and non-Bayesian methods have been explored for dropout prediction. However, for accurate dropout rate predictions, the validity of the model is crucial. In this study, the proposed Weibull model was validated through several statistical tests and diagnostic checks. Goodness-of-fit tests such as the Kolmogorov-Smirnov test, Anderson-Darling test and Cramer-von Mises test were conducted, alongside visual inspections using P-P and Q-Q plots. Additionally, the Bayesian model was validated using diagnostic tools such as trace plots, running mean (ergodic mean) plots, autocorrelation plots, and Brooks-Gelman-Rubin (BGR) diagnostic plots. Metrics like the effective sample size (n_eff) and Rhat values were also checked, ensuring model convergence and reliability. The results of these diagnostics indicate a strong fit between the model and the data, affirming the proposed model's effectiveness in predicting student dropout rates. Thus, this study demonstrates that the validated Bayesian Weibull model provides an accurate and reliable framework for predicting student dropout rates.

The proposed model predicts that the dropout rate in higher education in Nepal is at least 26% i.e., one in every four students/year are dropout in higher education. This aligns with some studies, while other research indicates varying dropout rates across different countries and contexts. For instance, in Italy, the dropout rates were 12.2% for bachelor's, 6.2% for master's, and 7.4% for combined programs (Perchinunno et al., 2021). At the University of Agriculture in Latvia, dropout rates ranged from 24.3% to 51.6% across various faculties by the end of the first academic year (Paura&Arhipova, 2014). In London, dropout rates varied between 15% and 18% over several years (Bennett, 2003). Brazilian undergraduate courses had an average dropout rate of 17%, predicted using a unit ratio-extended Weibull family model(Peña-Ramírez et al., 2023). In Catalonia, dropout rates ranged from 28% to 33%, prompting efforts to improve the educational experience for both students and teachers (Gairín et al., 2014). Spain witnessed a 40% dropout rate among students (Lassibille& Navarro Gómez, 2008). In Germany, dropout rates vary by discipline, with 33% for bachelor's degrees, 30% for language and cultural studies, 27% for law, economics, business administration, and social sciences, 39% for mathematics and natural sciences, and 36% for engineering. Agricultural, forestry, and veterinary sciences had a dropout rate of 30%, while human medicine and health sciences faced a 28% dropout rate (Heublein, 2014). Similarly, Nilkantha Multiple Campus in Nepal reported a significant dropout rate of 50.26% at the bachelor's level (Subedi, 2022). Hence, the current study covers institutional dropout in higher education in Nepal. The factors behind the dropout such as the lack of part-time job opportunities for students, difficulty securing employment after graduation, job market discrimination, and an education system focused on theory rather than practical skills contribute to high dropout rates. Students often find it challenging to secure a sustainable job with a low salary after completing their degree. Additionally, many students choose to study abroad due to the perceived value of foreign degrees and the potential for higher income due to favorable exchange rates, contributing to the prevalent issue of brain drain in Nepalese society. Moreover, higher education institutions in Nepal face significant challenges with student dropout, often due to dissatisfaction with the education system or the pursuit of better opportunities abroad. Therefore, it effectively conveys the importance of government intervention to address the issue of dropout and suggests potential strategies to prevent it, such as identifying at-risk students and providing them with additional academic support, counseling services, or financial assistance to students.

5. Conclusion and Recommendation

To predict the dropout rate of higher education, Bayesian Modelling of Weibull is employed. The Weibull probability distribution utilized the maximum likelihood technique, while Bayesian analysis employed the MCMC technique with the NUTS algorithm. To perform Bayesian analysis by numerically and graphically used Gamma informative priors. Various posterior predictive diagnostic criteria were tested to accesses the predictability of the model. Hence, Bayesian modelling of Weibull distribution is an alternative model for prediction the educational data like as student's dropout rate.

The posterior distributions were then utilized for predicting the student's dropout. The study predicts the dropout rate of approximately 26% (one in every four student/year) among higher education in Nepal which is challenge to

Nepalese universities and the higher education system of Nepal. These findings explore the issue of dropout in higher education institutions in Nepal and suggest that the government and policymakers develop strategies to control dropout. Furthermore, the study recommends future research to explore the factors contributing to dropout in greater depth.

Acknowledgements
We would like to express our appreciation to Research Management Cell, Birendra Multiple Campus, Tribhuvan University for providing financial support through the Mini Research Grant program. Likewise, we would also like to extend my thanks to the department of administrative section of respective college who provide the data.

Conflicts of Interest
The authors declare that they have no conflicts of interest.

Data Availability Statement
All the data relevant to the study are presented in the manuscript and can be provided upon request from the corresponding author.

References

1. Alzahrani, M. R. (2024). *Bayesian Analysis of Exponentiated-Weibull Weibull Distribution for Modelling Student Dropout Times*. https://doi.org/10.20944/preprints202406.1660.v1
2. Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2017). *Predicting Student Dropout in Higher Education* (arXiv:1606.06364). arXiv. http://arxiv.org/abs/1606.06364
3. Bennett, R. (2003). Determinants of Undergraduate Student Drop Out Rates in a University Business Studies Department. *Journal of Further and Higher Education*, *27*(2), 123–141. https://doi.org/10.1080/030987703200065154
4. Bishop, M. A., & Dey, D. K. (2005). Bayesian Analysis of the Weibull Distribution. *Statistical Modelling*, *5*(4), 359-373.
5. Camizuli, E., & Carranza, E. J. (2018). Exploratory Data Analysis (EDA). In S. L. López Varela (Ed.), *The Encyclopedia of Archaeological Sciences* (1st ed., pp. 1–7). Wiley. https://doi.org/10.1002/9781119188230.saseas0271
6. Chauhan, C. P. S. (2008). Higher Education: Current Status and Future Possibilities in Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, and Sri Lanka. *Analytical Reports in International Education*, *2*(1), 29–48. https://doi.org/10.3890/arie.2.1.29
7. Chen, M.-H., & Shao, Q.-M. (1999). Monte Carlo Estimation of Bayesian Credible and HPD Intervals. *Journal of Computational and Graphical Statistics*, *8*(1), 69–92. https://doi.org/10.1080/10618600.1999.10474802
8. Crisp, G., Doran, E., & Salis Reyes, N. A. (2018). Predicting Graduation Rates at 4-year Broad Access Institutions Using a Bayesian Modeling Approach. *Research in Higher Education*, *59*(2), 133–155. https://doi.org/10.1007/s11162-017-9459-x
9. De La Cruz, J. R., Hernández-Romero, D. L., Arguijo, P., Herazo, L. C. S., & Meléndez-Armenta, R. Á. (2024). Analyzing University Dropout Rates Using Bayesian Methods: A Case Study in University Level in Mexico. In Á. Rocha, H. Adeli, G. Dzemyda, F. Moreira, & A. Poniszewska-Marańda (Eds.), *Good Practices and New Perspectives in Information Systems and Technologies* (Vol. 988, pp. 165–174). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-60224-5_18
10. Depaoli, S., Winter, S. D., & Visser, M. (2020). The Importance of Prior Sensitivity Analysis in Bayesian Statistics: Demonstrations Using an Interactive Shiny App. *Frontiers in Psychology*, *11*, 608045. https://doi.org/10.3389/fpsyg.2020.608045
11. Dhungana, G. P., & Kumar, V. (2022). *Modified Half Logistic Weibull Distribution with Statistical Properties and Applications*.
12. Gairín, J., Triado, X. M., Feixas, M., Figuera, P., Aparicio-Chueca, P., & Torrado, M. (2014). Student dropout rates in Catalan universities: Profile and motives for disengagement. *Quality in Higher Education*, *20*(2), 165–182. https://doi.org/10.1080/13538322.2014.925230
13. Garrido, A., Caro-Carretero, R., Jimenez-Octavio, J. R., Carnicero, A., & Such, M. (2021). A new approach to fitting the three-parameter Weibull distribution: An application to glass ceramics. *Communications in Statistics - Theory and Methods*, *50*(14), 3403–3420. https://doi.org/10.1080/03610926.2019.1702698
14. Heublein, U. (2014). Student Drop-out from G erman Higher Education Institutions. *European Journal of Education*, *49*(4), 497–513. https://doi.org/10.1111/ejed.12097
15. Hofman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res, 15, 1593–1623.
16. Kostopoulos, G., Kotsiantis, S., & Pintelas, P. (2015). Estimating student dropout in distance higher education using semi-supervised techniques. *Proceedings of the 19th Panhellenic Conference on Informatics*, 38–43. https://doi.org/10.1145/2801948.2802013

17. Lacave, C., Molina, A. I., & Cruz-Lemus, J. A. (2018). Learning Analytics to identify dropout factors of Computer Science studies through Bayesian networks. *Behaviour & Information Technology*, *37*(10–11), 993–1007. https://doi.org/10.1080/0144929X.2018.1485053
18. Lassibille, G., & Navarro Gómez, L. (2008). Why do higher education students drop out? Evidence from Spain. *Education Economics*, *16*(1), 89–105. https://doi.org/10.1080/09645290701523267
19. Lawless, J. F. (2011). *Statistical models and methods for lifetime data*. John Wiley & Sons.
20. Niyogisubizo, J., Liao, L., Nziyumva, E., Murwanashyaka, E., & Nshimyumukiza, P. C. (2022). Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization. *Computers and Education: Artificial Intelligence*, *3*, 100066. https://doi.org/10.1016/j.caeai.2022.100066
21. Paura, L., & Arhipova, I. (2014). Cause Analysis of Students' Dropout Rate in Higher Education Study Program. *Procedia - Social and Behavioral Sciences*, *109*, 1282–1286. https://doi.org/10.1016/j.sbspro.2013.12.625
22. Peña-Ramírez, F. A., Guerra, R. R., & Mafalda, C. P. (2023). The unit ratio-extended Weibull family and the dropout rate in Brazilian undergraduate courses. *PLOS ONE*, *18*(11), e0290885. https://doi.org/10.1371/journal.pone.0290885
23. Perchinunno, P., Bilancia, M., & Vitale, D. (2021). A Statistical Analysis of Factors Affecting Higher Education Dropouts. *Social Indicators Research*, *156*(2–3), 341–362. https://doi.org/10.1007/s11205-019-02249-y
24. R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URLhttps://www.R-project.org/.
25. Rabelo, A. M., & Zárate, L. E. (2024). A Model for Predicting Dropout of Higher Education Students. *Data Science and Management*, S2666764924000341. https://doi.org/10.1016/j.dsm.2024.07.001
26. Stan Development Team (2023) Stan modeling language users guide and reference manual, 2.32. https:// mc-stan.org/. Accessed 10 August 2023
27. Stăiculescu, C., & Elena Ramona, R. N. (2019). University dropout. Causes and solution. *Mental Health: Global Challenges Journal*, *1*(1), 71–75. https://doi.org/10.32437/mhgcj.v1i1.29
28. Subedi, Y. R. (2022). Factors Influencing Dropout Rate in Bachelor's Degree at Nilkantha Multiple Campus. *The EFFORTS, Journal of Education and Research*, *4*(1), 24–38. https://doi.org/10.3126/ejer.v4i1.44170
29. Udomvisawakul, A. (2021). *Applying Bayesian Growth Modeling in Machine Learning for*
30. *Longitudinal Data*. University of Northern Colorado.
31. UGC (2024) University Grants Commission Nepal (ugcnepal.edu.np)
32. University Grants Commission, Nepal. (2023). Education management information system: (EMIS) Report on higher education 2022/23. Sanothimi, Bhaktapur: The Author