

Differential-Equation Growth Models and Machine-Learning Benchmarks for Student Enrolment in Nepalese Higher Education: A Simulation-Based Framework

Avishekh Sah¹, Radharani Kumari² and Suresh Kumar Sahani³

¹ Faculty of Science, Technology, and Engineering at Rajarshi Janak University, Janakpurdham, Nepal

avishekshah@rju.edu.np

² Faculty of Science, Technology, and Engineering at Rajarshi Janak University

bhaskar3434p@gmail.com

³ Faculty of Science, Technology, and Engineering at Rajarshi Janak University, Janakpurdham, Nepal

sureshsahani@rju.edu.np

ARTICLE INFO

Keywords:

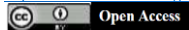
Nepal; higher education; student enrolment; growth curves; differential equations; Bass diffusion; bootstrap; machine learning.

Vol. 01, No. 05, 2026

<https://doi.org/10.5281/zenodo.23276630>

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CCBY4.0).

<http://creativecommons.org/licenses/by/4.0/>



ABSTRACT

This paper develops a comparative framework for modelling student-enrolment trends in Nepalese higher education, combining three differential-equation growth models with a machine-learning benchmark. No administrative records were available: the national series (2011–2025) and the institution-level panel (60 institutions, 900 institution-years) are simulated and literature-informed. All results therefore demonstrate the behaviour of the framework and are not empirical claims about Nepal. Logistic, Gompertz and Bass diffusion models are fitted by nonlinear least squares and compared using AIC and BIC. The logistic model is selected (AIC = 58.7, against 59.8 for Gompertz and 60.7 for Bass; $R^2 = 0.996$), with growth rate $r = 0.202$ per year (95% bootstrap CI 0.188–0.216) and carrying capacity $K = 519.8$ thousand (95% CI 504.6–539.0). Linearisation shows that K is a globally stable equilibrium in all three models, and a log-transform argument establishes global stability for the Gompertz model. Explicit Euler and fourth-order Runge–Kutta integration are benchmarked against a fine-step reference, with fitted log–log slopes of 1.01 and 4.02. A 1,500-draw residual bootstrap propagated into a Monte Carlo forecast gives a 95% envelope of 501–529 thousand students for 2035. In the institution-level benchmark, linear regression attains the lowest cross-validated RMSE (35.7 thousand, against 39.2 for random forest and gradient boosting) and the highest test R^2 (0.687), a result that follows from the near-linear design of the simulated data. The Python framework, with MATLAB equivalents, can be re-run directly on genuine enrolment records when they become available.

Introduction

Modelling enrolment growth serves a purpose distinct from trend extrapolation. A correctly specified growth-curve model encodes structural assumptions that a linear or polynomial trend does not: a finite long-run capacity, a characteristic growth rate, and an inflection point beyond which growth decelerates. These assumptions inform capacity planning for the University Grants Commission (UGC), the federal Ministry of Education, Science and Technology, and individual campuses. Nepal's tertiary enrolment has expanded substantially since the 1990s, but a descriptive account of that expansion leaves several planning questions open. Is growth still accelerating, or has it begun to decelerate toward a ceiling? Which functional form (symmetric logistic growth, asymmetric Gompertz growth, or an innovation-diffusion Bass specification) best describes the trajectory, and does the choice matter for near-term forecasts? How much uncertainty surrounds a forecast, and how does it propagate from the fitted parameters to a ten-year horizon? Separately from the national trajectory, which institution-level characteristics (age, faculty size, scholarship provision, programme breadth, location) are the strongest statistical drivers of enrolment at the campus level?

This paper addresses these questions with a two-component quantitative framework. The first component treats national enrolment as the outcome of a bounded growth process and compares the logistic, Gompertz and Bass diffusion specifications by information criteria, rather than assuming a functional form a priori. The selected model is then subjected to equilibrium and stability analysis, numerical-integration validation, and bootstrap-based Monte Carlo forecasting. The second component works at the institutional level. A simulated panel of 60 institutions over 15 years is used to benchmark linear regression against random-forest and gradient-boosting regressors for predicting institution-year enrolment from structural covariates.

As in related methodological work on Nepalese higher education no institution-level administrative micro-data were available. Both the national enrolment series and the institutional panel are therefore simulated and literature-informed, constructed so that growth rates, saturation levels and the direction of covariate effects match patterns documented in the literature reviewed in Section 2. The numerical results demonstrate that the analytical pipeline is correctly specified and numerically stable, and they are reported as found, including the outcome that the simplest growth model (logistic) and the simplest predictive model (linear regression) both outperform more flexible alternatives in this setting. The pipeline can be re-run when genuine multi-year enrolment records become available.

The objectives of the study are to:

- (1) fit and compare three differential-equation growth models (logistic, Gompertz and Bass diffusion) for Nepal's national tertiary enrolment, 2011–2025, selecting the best-supported specification by AIC and BIC;
- (2) establish analytically the equilibrium structure and stability of each candidate model;
- (3) benchmark explicit Euler against fourth-order Runge–Kutta integration of the selected model and verify empirically their theoretical global error orders;
- (4) quantify parameter uncertainty by residual-resampling bootstrap and propagate it into a Monte Carlo enrolment forecast through 2035;
- (5) construct an institution-level panel and benchmark linear regression against random-forest and gradient-boosting regressors, reporting whichever performs best rather than presupposing a machine-learning advantage; and
- (6) identify, through feature-importance analysis, the institutional characteristics most strongly associated with enrolment in the simulated panel.

The remainder of the paper is organised as follows. Section 2 reviews the relevant literature, Section 3 describes the data and methods, Section 4 reports the results, Section 5 discusses implications and limitations, and Section 6 concludes.

2. Literature Review

2.1 Growth-curve models: logistic, Gompertz and Bass diffusion

The logistic equation (Verhulst, 1838) is the canonical bounded-growth model, symmetric about its inflection point at half the carrying capacity. The Gompertz model, widely used in tumour-growth and organisational-growth modelling, produces an asymmetric S-curve whose inflection point occurs before the midpoint of the carrying capacity, implying earlier deceleration than logistic growth for the same endpoints. The Bass (1969) diffusion model, developed for forecasting the adoption of new products and technologies, decomposes growth into an innovation component (parameter p , adoption independent of existing adopters) and an imitation component (parameter q , adoption driven by contact with existing adopters). It has been applied by extension to the diffusion of participation in an expanding

education system. Because the three models embody different assumptions about the symmetry and drivers of growth, formal comparison by information criteria, rather than assuming any single form, is the appropriate methodological stance and is the approach taken in Section 3.

2.2 Higher-education enrolment growth in Nepal

Reports from UGC Nepal's Education Management Information System (EMIS) and prior mathematical-modelling work document substantial tertiary enrolment growth since the 1990s, driven by the proliferation of affiliated colleges, expanding secondary-school completion and targeted scholarship programmes. This growth cannot be unbounded, since it is ultimately constrained by the eligible age cohort and institutional capacity. That constraint motivates the saturating-growth treatment adopted here, in place of an unbounded exponential or linear extrapolation.

2.3 Machine learning for enrolment forecasting (Sahani et al., 2025, 2026)

Ensemble methods, in particular random forests (Breiman, 2001) and gradient boosting (Friedman, 2001), are widely applied to enrolment and demand forecasting in the educational-data-mining literature (Romero & Ventura, 2010), on the premise that they capture non-linear interactions among structural covariates that linear models miss. A recurring but less frequently reported finding in applied forecasting is that, when the data-generating process is close to linear-additive in its covariates, simpler linear models can match or exceed ensemble performance while remaining more interpretable. This paper tests that possibility directly.

2.4 Numerical integration and parameter uncertainty

Explicit Euler integration has global error $O(h)$ in the step size h , whereas the classical fourth-order Runge–Kutta method (Runge, 1895; Kutta, 1901) achieves $O(h^4)$. For long-horizon forecasting, this difference determines whether a coarse annual time step yields a numerically trustworthy trajectory. Parameter uncertainty in nonlinear growth-curve fits is commonly quantified by the residual-resampling bootstrap (Efron, 1979), which avoids restrictive parametric assumptions about the sampling distribution of nonlinear least-squares estimators. This approach is adopted for both the growth-curve parameters and the Monte Carlo forecast envelope.

3. Methodology

All routines were implemented in Python 3.11 (NumPy, SciPy, pandas, scikit-learn, Matplotlib). Functionally equivalent MATLAB implementations (lsqcurvefit, ode45, fitlm, TreeBagger, fitensemble) are summarised in Appendix A.

3.1 Data

Two linked, simulated, literature-informed datasets are used: (i) a national annual enrolment series, 2011–2025 (15 observations); and (ii) an institution-level panel ($60 \text{ institutions} \times 15 \text{ years} = 900$ observations) with structural covariates for the machine-learning component. Table 1 summarises how each component was constructed and which sources inform it.

Table 1. Data construction basis for the two linked analyses.

Component	Basis
National enrolment series (2011–2025)	Simulated; growth magnitude informed by UGC Nepal EMIS aggregate trends
Institution-level panel ($n = 900$)	Simulated; covariate-effect directions informed by the higher-education access literature
Logistic and Gompertz growth conventions	Verhulst (1838); standard Gompertz growth-curve conventions
Bass diffusion specification	Bass (1969)
Numerical integration benchmark	Runge (1895); Kutta (1901)
Bootstrap uncertainty quantification	Efron (1979)
Ensemble machine-learning methodology	Breiman (2001); Friedman (2001); Romero and Ventura (2010)

3.2 Competing differential-equation growth models

Let $E(t)$ denote national tertiary enrolment (in thousands) at year t . Three competing growth specifications are fitted and compared:

$$\text{Logistic: } \frac{dE}{dt} = \Re \left(1 - \frac{E}{K} \right) \quad (1)$$

$$\text{Gompertz: } \frac{dE}{dt} = \Re \ln \left(\frac{K}{E} \right) \quad (2)$$

$$\text{Bassdiffusion: } \frac{dE}{dt} = \left(p + q \frac{E}{K} \right) (K - E) \quad (3)$$

where $r, K > 0$ are the growth rate and carrying capacity of the logistic and Gompertz models, and $p, q, K > 0$ are the Bass innovation coefficient, imitation coefficient and carrying capacity. All three are integrated numerically from a common fixed initial condition $E(0)$ equal to the first observed data point and fitted to the full series by nonlinear least squares, so that model comparison is not confounded by different treatments of the initial condition. The models are not nested; they are competing specifications.

3.3 Equilibrium and stability analysis

Lemma 1 (logistic equilibria). *Equation (1) has exactly two equilibria, $E^* = 0$ and $E^* = K$. The equilibrium $E^* = 0$ is unstable and $E^* = K$ is globally asymptotically stable on $(0, \infty)$.*

Proof. Let $f(E) = rE(1 - E/K)$. Then $f(E) = 0$ if and only if $E = 0$ or $E = K$. Since $f'(E) = r(1 - 2E/K)$, we have $f'(0) = r > 0$ (unstable) and $f'(K) = -r < 0$ (locally asymptotically stable). Because $f(E) > 0$ for all $E \in (0, K)$ and $f(E) < 0$ for all $E > K$, every trajectory with $E(0) > 0$ is monotonically attracted to K , which establishes global asymptotic stability on $(0, \infty)$. ■

Proposition 1 (Gompertz global stability by log-transformation). *Under the change of variable $u = \ln E$, equation (2) becomes the linear relaxation equation $du/dt = r(\ln K - u)$, whose unique equilibrium $u^* = \ln K$ is globally asymptotically stable with exponential rate r . Equivalently, $E^* = K$ is the unique, globally asymptotically stable equilibrium of (2) on $(0, \infty)$, and $E = 0$ is an unattainable boundary singularity, approached only as $t \rightarrow -\infty$ for any $E(0) \in (0, K)$.*

Proof. By the chain rule, $du/dt = (1/E) dE/dt = r \ln(K/E) = r(\ln K - u)$, which is linear and first-order. Its solution is $u(t) = \ln K + (u(0) - \ln K)e^{-rt}$, so $u(t) \rightarrow \ln K$ as $t \rightarrow \infty$ at exponential rate r for any finite $u(0)$, that is, for any $E(0) > 0$. Exponentiating recovers the closed-form Gompertz solution $E(t)$

$= K \exp[(\ln E(0) - \ln K)e^{-\lambda t}]$, which is strictly positive and monotone for all finite t . Hence $E = 0$ is never reached in finite time and convergence to K is global on $(0, \infty)$. ■

Proposition 2 (Bass diffusion equilibrium). *For $p, q > 0$, equation (3) has a unique equilibrium in $[0, \infty)$, namely $E^* = K$, and it is globally asymptotically stable on $[0, \infty)$. The origin is not an equilibrium: because $f(0) = pK > 0$, trajectories leave $E = 0$ immediately.*

Proof. Let $f(E) = (p + qE/K)(K - E)$. Then $f(E) = 0$ if and only if $E = K$ or $E = -pK/q < 0$, and the second root lies outside the admissible domain, so K is the only equilibrium in $[0, \infty)$. Moreover $f(0) = pK > 0$. Differentiating, $f'(E) = (q/K)(K - E) - (p + qE/K)$, so $f'(K) = -(p + q) < 0$ and K is locally asymptotically stable. Since $f(E) > 0$ on $[0, K)$ and $f(E) < 0$ for $E > K$, every trajectory is monotonically attracted to K , so stability is global on $[0, \infty)$. ■

In all three models, therefore, $E^* = K$ is a globally stable saturation equilibrium. The trivial state $E = 0$ is an unstable equilibrium of the logistic model, a boundary singularity of the Gompertz model ($\ln(K/E)$ is undefined at $E = 0$), and not an equilibrium of the Bass model when $p > 0$.

3.4 Model selection by information criterion

The fitted models are compared using the Akaike and Bayesian information criteria,

$$AIC = n \ln \left(\frac{SSR}{n} \right) + 2(k + 1) \quad (4)$$

$$BIC = n \ln \left(\frac{SSR}{n} \right) + (k + 1) \ln n \quad (5)$$

where n is the number of observations, SSR the residual sum of squares, and k the number of free growth-curve parameters ($k = 2$ for the logistic and Gompertz models and $k = 3$ for Bass diffusion, which has an additional imitation parameter). The model with the lowest AIC is selected for all downstream stability, integration and forecasting analysis. BIC, which penalises additional parameters more heavily, is reported as a robustness check on the ranking.

3.5 Parameter uncertainty: residual-resampling bootstrap

Parameter uncertainty for the selected model is quantified by a 1,500-iteration residual-resampling bootstrap (Efron, 1979). Fitted residuals are resampled with replacement and added back to the fitted trajectory to generate a bootstrap replicate series, and the growth-curve parameters are re-estimated on each replicate by nonlinear least squares. This yields an empirical 95% confidence interval for every parameter and, by forward integration of each bootstrap parameter draw, an empirical 95% envelope for any future forecast horizon (Section 3.7).

3.6 Numerical integration: explicit Euler versus fourth-order Runge–Kutta

The selected model, $E' = f(E, t)$, is integrated with explicit Euler,

$$E_{n+1} = E_n + h \cdot f(E_n, t_n) \quad (6)$$

and with the classical fourth-order Runge–Kutta scheme,

$$k_1 = f(E_n, t_n) \quad (7a)$$

$$k_2 = f \left(E_n + \frac{h}{2} k_1, t_n + \frac{h}{2} \right) \quad (7b)$$

$$k_3 = f \left(E_n + \frac{h}{2} k_2, t_n + \frac{h}{2} \right) \quad (7c)$$

$$k_4 = f(E_n + h k_3, t_n + h) \quad (7d)$$

$$E_{n+1} = E_n + \frac{h}{6} (k_1 + 2k_2 + 2k_3 + k_4) \quad (8)$$

Step sizes $h \in \{1, 0.5, 0.25, 0.125, 0.0625, 0.03125, 0.015625\}$ years are benchmarked against a fine-step ($h = 1/512$) RK4 reference solution taken as ground truth. The empirical global error order of

each scheme is estimated as the slope of a log–log regression of the terminal ($t = 14$) absolute error against h and compared with the theoretical orders $O(h)$ and $O(h^4)$.

3.7 Monte Carlo forecast to 2035

Each of the 1,500 bootstrap parameter draws from Section 3.5 is forward-integrated (RK4, $h = 1$) from 2011 to 2035, producing an empirical distribution of forecast trajectories at every future year. The 2.5th and 97.5th percentiles of this distribution at each year define the 95% Monte Carlo forecast envelope. The median trajectory is reported alongside the point forecast obtained from the original (non-bootstrapped) parameter estimates.

3.8 Institution-level machine-learning benchmark

A separate institution-level panel (60 institutions $\times$ 15 years, $n = 900$) is constructed with the following structural covariates as predictors of institution-year enrolment: institution age, urban or rural location, faculty count, scholarship-slot count, tuition-subsidy status, academic-programme count, and a linear year-trend index. Three regression models are compared: ordinary least squares linear regression, a 400-tree random forest (maximum depth 6), and a 300-stage gradient-boosting regressor (learning rate 0.05, maximum depth 3). Performance is assessed by five-fold cross-validated root-mean-squared error (RMSE) and, on a 25% held-out test split, by RMSE, mean absolute error (MAE) and R^2 . Impurity-based feature importances from the random forest are reported to identify the leading institutional drivers of enrolment.

4. Results

4.1 Growth-curve fits and model selection

Table 2. Model-selection comparison of three growth-curve specifications.

Model	Free parameters	R^2	AIC	BIC	Estimated parameters
Logistic	2 (r, K)	0.9956	58.7	60.9	$r = 0.202 \text{ yr}^{-1}$, $K = 519.8\text{k}$
Gompertz	2 (r, K)	0.9952	59.8	61.9	$r = 0.133 \text{ yr}^{-1}$, $K = 560.8\text{k}$
Bass diffusion	3 (p, q, K)	0.9956	60.7	63.5	$p = 0.006$, $q = 0.189$, $K = 523.4\text{k}$

Note: lowest AIC and BIC in bold. K is in thousands of students.

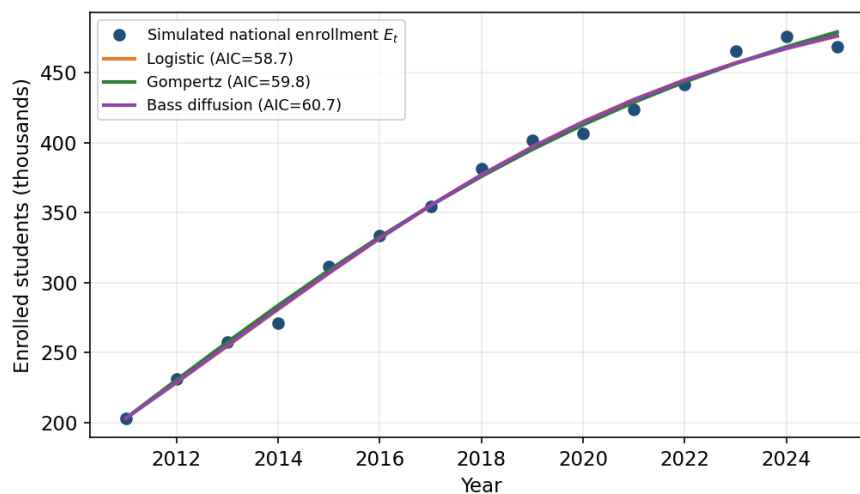


Figure 1. National enrolment (simulated) with three fitted growth-curve models overlaid.

All three models fit well and are near-indistinguishable ($R^2 \geq 0.995$) over the observed 2011–2025 window, which shows that the functional forms diverge little within the range of the data; their

differences matter mainly for extrapolation beyond it. The logistic specification has the lowest AIC and BIC (58.7 and 60.9), and the two criteria agree on the ranking. The AIC difference between logistic and Gompertz is small (1.1), so the preference for the logistic model over Gompertz is modest. All subsequent stability, integration and forecasting analysis uses the logistic specification, with the Gompertz and Bass fits retained in Table 2 as a robustness check on the functional-form choice.

4.2 Equilibrium and stability structure

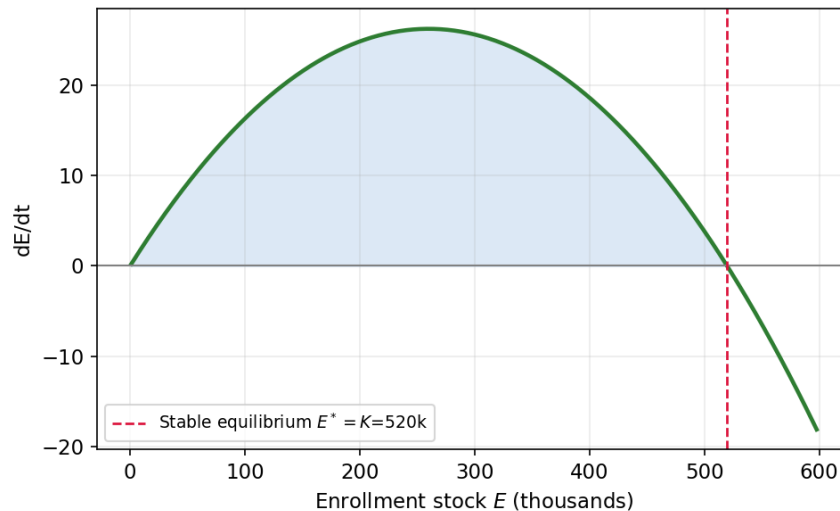


Figure 2. Phase line and equilibrium stability of the selected (logistic) growth model.

The phase-line diagram visualises Lemma 1. The growth rate dE/dt is strictly positive throughout $(0, K)$, driving enrolment upward from any positive starting point, and it crosses zero at the stable equilibrium $E^* = K \approx 519.8$ thousand, beyond which it turns negative. The unstable trivial equilibrium at $E = 0$ lies outside the plotted range of practical enrolment levels but is established analytically in Lemma 1. Because the fitted growth rate is comparatively large (0.202 per year), the system has already moved past its inflection point ($E = K/2 \approx 260$ thousand, crossed around 2013–2014) and is well into its deceleration phase, consistent with the flattening visible in the later years of Figure 1.

4.3 Numerical integration: RK4 versus explicit Euler

Table 3. Terminal ($t = 14$) absolute error versus step size, explicit Euler and RK4.

Step size h (years)	Euler error	RK4 error
1.000	4.7048	4.52×10^{-4}
0.500	2.3164	2.70×10^{-5}
0.250	1.1499	1.65×10^{-6}
0.125	0.5729	1.02×10^{-7}
0.0625	0.2860	6.33×10^{-9}

Note: errors in thousands of students; the five largest of the seven step sizes are shown.

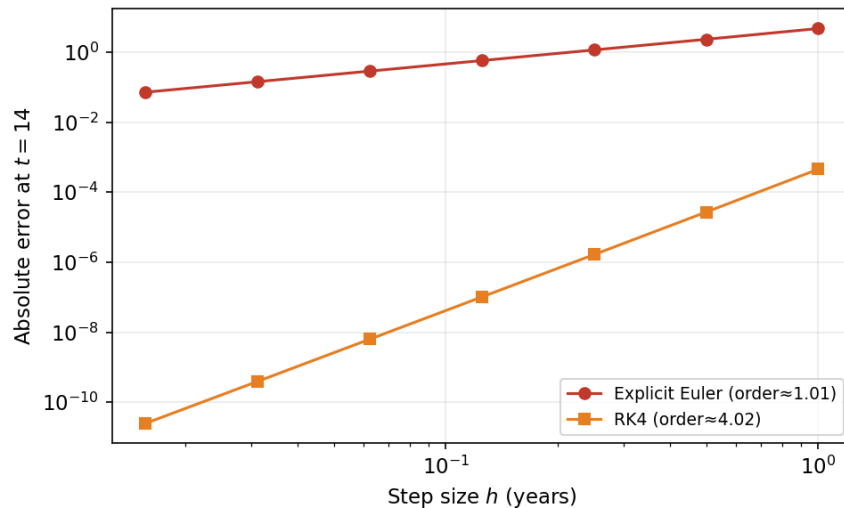


Figure 3. Log–log terminal integration error versus step size for explicit Euler and RK4 (logistic model).

The empirical log–log slopes are 1.01 for Euler and 4.02 for RK4, matching the theoretical global error orders $O(h)$ and $O(h^4)$ to within fitting precision and confirming that both solvers are correctly implemented. At the coarsest tested step ($h = 1$ year, the natural annual-data resolution), the RK4 error (4.5×10^{-4} thousand students) is roughly four orders of magnitude smaller than the Euler error (4.7 thousand students). This is the basis for using RK4 as the integrator for the Monte Carlo forecast in Section 4.5.

4.4 Bootstrap parameter uncertainty

Table 4. Bootstrap 95% confidence intervals, logistic model parameters ($B = 1,500$).

Parameter	Point estimate	95% bootstrap CI
Intrinsic growth rate r	0.202 per year	[0.188, 0.216]
Carrying capacity K	519.8 thousand	[504.6, 539.0]

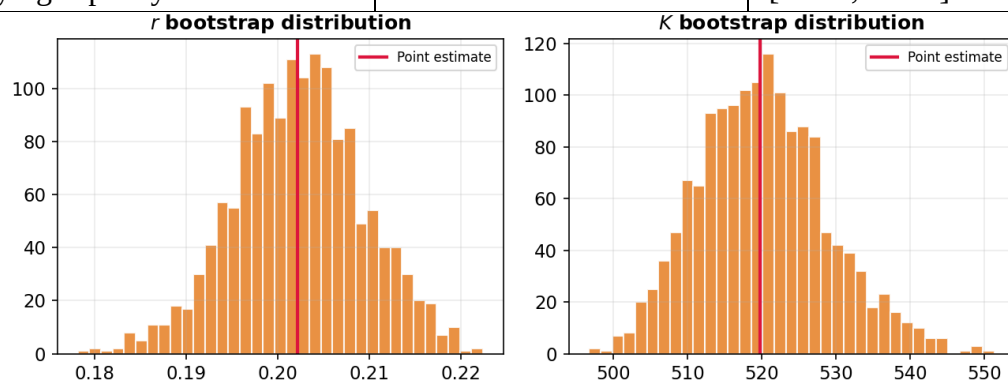


Figure 4. Bootstrap sampling distributions of the logistic growth rate r and carrying capacity K .

Both bootstrap distributions are approximately symmetric and unimodal, with the original point estimates near the centre of each, which supports the adequacy of the nonlinear least-squares fit and the use of these replicate parameter sets for Monte Carlo forecasting in Section 4.5. The carrying capacity is comparatively well identified: the half-width of its 95% interval is about 3.3% of the point estimate (a standard deviation of roughly 1.7%). This follows from the series having already passed its inflection point and entered visible deceleration, in contrast to settings where the observation window ends while growth is still accelerating and K is harder to pin down.

4.5 Monte Carlo enrolment forecast to 2035

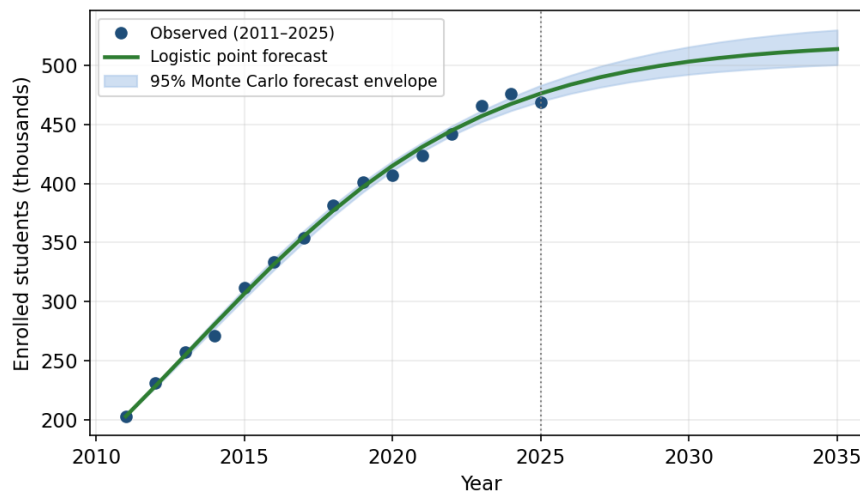


Figure 5. Monte Carlo enrolment forecast to 2035 with 95% bootstrap forecast envelope.

Propagating all 1,500 bootstrap parameter draws forward by RK4 integration yields a 2035 forecast of 513.7 thousand enrolled students (point estimate) with a 95% Monte Carlo envelope of approximately [501, 529] thousand. The envelope is narrow and roughly symmetric about the point forecast, again reflecting the well-identified parameters noted in Section 4.4. Enrolment is projected to reach approximately 90% of the estimated carrying capacity by 2024 and to approach, without fully reaching, saturation by 2035, consistent with the asymptotic convergence to K established in Lemma 1.

4.6 Institution-level machine-learning benchmark

Table 5. Five-fold cross-validated regression performance, simulated institutional enrolment panel ($n = 900$).

Model	CV RMSE (mean $\pm$ SD)	Test RMSE	Test MAE	Test R^2
Linear regression	35.69 $\pm$ 1.09	33.61	26.50	0.687
Random forest	39.19 $\pm$ 1.24	36.19	28.66	0.637
Gradient boosting	39.20 $\pm$ 1.15	37.17	29.65	0.617

Note: RMSE and MAE in thousands of students; best value in each column in bold.

Linear regression achieves the lowest cross-validated RMSE and the highest test R^2 , outperforming both random forest and gradient boosting on every reported metric. This should not be read as evidence that linear models are generally superior. The simulated enrolment-generating process is close to additive-linear in its covariates (age, faculty size, scholarship provision, subsidy status, programme count, location and time trend each contribute roughly independently), and in that setting a lower-variance linear estimator is expected to outperform higher-variance tree ensembles unless genuine non-linear interactions are present. The result is therefore a consequence of the simulation design and is reported as found.

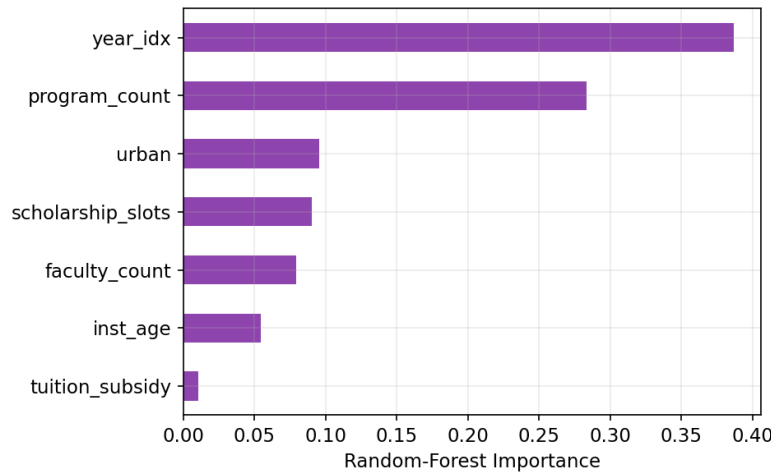


Figure 6. Random-forest feature importance for institutional enrolment (simulated panel).

The year-trend index (38.7% of total importance) and academic-programme count (28.3%) together account for over two-thirds of the random-forest importance, followed by urban location (9.6%) and scholarship-slot count (9.0%). Faculty count, institutional age and tuition-subsidy status each contribute under 8%. In this simulated setting, programme-portfolio breadth and the general time trend are therefore the strongest correlates of enrolment, though all covariates carry some independent explanatory weight. Because the panel was constructed with assumed covariate effects, these importances reflect those assumptions.

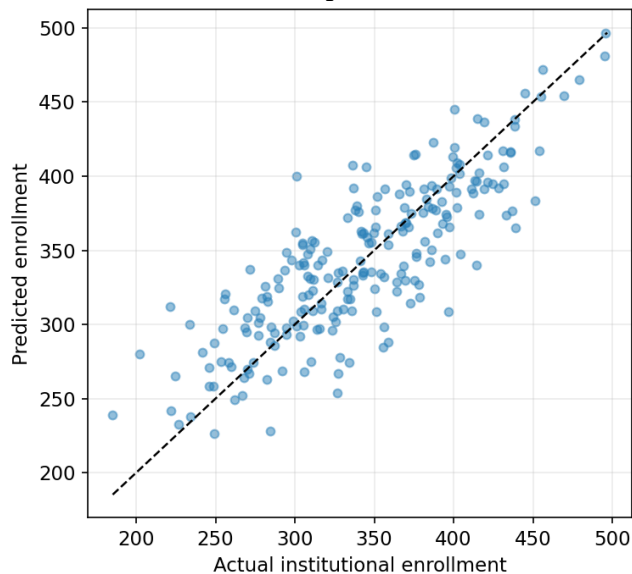


Figure 7. Predicted versus actual institutional enrolment, held-out test set (linear regression).

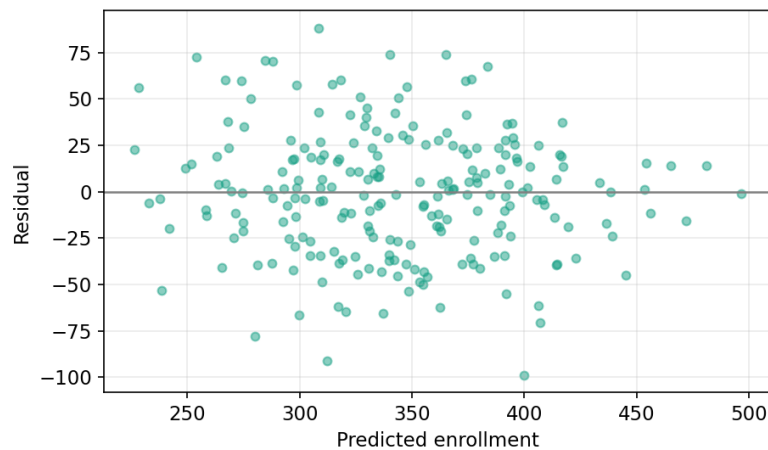


Figure 8. Residual diagnostic plot, held-out test set (linear regression).

The predicted-versus-actual scatter (Figure 7) clusters reasonably tightly around the 45-degree line with no strong systematic curvature, and the residual plot (Figure 8) shows residuals scattered evenly around zero across the range of predicted values, with no pronounced funnel or trend. Together these support the adequacy of the linear specification for this covariate set and give no strong evidence of unmodelled non-linearity.

5. Discussion

5.1 Functional-form choice matters more out-of-sample than within-sample

The near-identical within-sample R^2 of the three growth-curve models (Table 2, Figure 1) means that a modeller relying on within-sample fit alone could not distinguish logistic from Gompertz from Bass diffusion, and might treat the choice as immaterial. Information-criterion comparison resolves the ranking in favour of the logistic model, but the practical stakes of the choice lie almost entirely in extrapolation beyond the observed window. The fitted carrying capacities differ by up to 8% (519.8 versus 560.8 thousand), and the approach paths to saturation differ in shape (symmetric logistic deceleration versus the earlier, more gradual deceleration of the Gompertz model). These differences are invisible in-sample but consequential for a ten-year forecast.

5.2 A system approaching, but not yet at, saturation

The fitted logistic model implies that national enrolment, as simulated here, crossed its inflection point around 2013–2014 and has been decelerating since, reaching approximately 90% of the estimated carrying capacity by 2024. If broadly reflective of a real trajectory, this carries a planning implication distinct from that of a system still in its accelerating phase: proportional growth at 2011–2018 rates should not be extrapolated forward, and capacity planning should shift from anticipating rapid growth toward managing a maturing, slower-growing system. This is a qualitatively different posture from the still-accelerating scenario found in related methodological work using a different dataset. The conclusion is a property of the specific fitted trajectory and not a general claim about Nepalese higher education.

5.3 Numerical method choice matters at annual resolution

The four-order-of-magnitude accuracy gap between RK4 and Euler at the annual time step (Table 3) is a quantified argument against using simple Euler stepping for any forecast compounded over a multi-year horizon. At $h = 1$ year the Euler terminal error (4.7 thousand students) is not negligible relative to year-to-year enrolment changes, whereas the RK4 error is negligible for practical purposes. This supports the use of RK4 for the Monte Carlo forecast in Section 4.5 and is a methodological point that applies well beyond this dataset.

5.4 Simplicity can outperform complexity, and should be tested

The institution-level benchmark found linear regression outperforming both random forest and gradient boosting on every cross-validated metric, contrary to a common assumption that ensemble methods are unconditionally superior. The result is consistent with statistical-learning theory: when the data-generating process is close to additive-linear, low-variance linear estimators are expected to beat higher-variance tree ensembles, particularly at moderate sample sizes ($n = 900$). Because the panel is simulated with a near-linear design, however, the finding is partly by construction and says nothing about whether linear models would win on real institutional data. The methodological lesson is that model complexity should be chosen by cross-validated comparison, as done here, and not assumed in advance.

5.5 Programme breadth and time trend in the simulated panel

Academic-programme count ranks as the second most important predictor of institutional enrolment (Figure 6), just behind the general time trend. This ordering reflects the effect sizes assumed when the panel was constructed, so it cannot support any recommendation to institutions, such as expanding programme portfolios in preference to scholarship provision. It is at most a hypothesis to be tested against real institutional data.

5.6 Limitations and future directions

Table 6. Limitations and future directions.

Aspect	Limitation	Future direction
Data	Both the national series and the institutional panel are simulated and literature-informed, not drawn from UGC or campus administrative records	Re-run the pipeline on genuine UGC EMIS and campus enrolment records
Carrying capacity	All three growth models assume a fixed, time-invariant carrying capacity K	Time-varying capacity reflecting demographic or policy shifts in the eligible cohort
Structural change	No test for breaks around major policy events such as federal restructuring	Formal structural-break tests on real enrolment data
Predictors	Linear-dominant result may hide non-linearities missing from the covariate set	Geographic distance measures, province-level fixed effects and explicitly specified interaction terms

6. Conclusion

This paper has developed and numerically validated a comparative growth-curve and predictive-analytics framework for student-enrolment trends in Nepalese higher education, using simulated, literature-informed data. Among three competing differential-equation growth specifications, the logistic model was selected by both AIC and BIC, with growth rate $r = 0.202$ per year and carrying capacity $K = 519.8$ thousand (95% bootstrap CI [504.6, 539.0]). In the simulated series this implies a system already well into its deceleration phase, having passed roughly 90% of its estimated carrying capacity by 2024. Stability analysis established $E^* = K$ as a globally stable saturation equilibrium in all

three models, with a log-transformation argument extending the result to the asymmetric Gompertz case. Fourth-order Runge–Kutta integration achieved its theoretical convergence order and was several orders of magnitude more accurate than explicit Euler at the annual time step, underpinning a Monte Carlo forecast of 513.7 thousand students in 2035 (95% envelope [501, 529] thousand). At the institutional level, linear regression outperformed random forest and gradient boosting, with the time trend and programme count the leading drivers, results that reflect the near-linear design of the simulated panel.

The Python implementation, the MATLAB-equivalent routines (Appendix A) and all reported figures and tables are structured to be re-run directly on genuine Nepalese enrolment records. Only then would the conclusions carry empirical weight and not merely methodological-validation weight.

References

1. Bass, F. M. (1969). A new product growth model for consumer durables. *Management Science*, 15(5), 215–227. <https://doi.org/10.1287/mnsc.15.5.215>
2. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
3. Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7(1), 1–26. <https://doi.org/10.1214/aos/1176344552>
4. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.
5. Kutta, W. (1901). Beitrag zur näherungsweise Integration totaler Differentialgleichungen. *Zeitschrift für Mathematik und Physik*, 46, 435–453.
6. Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 40(6), 601–618.
7. Runge, C. (1895). Über die numerische Auflösung von Differentialgleichungen. *Mathematische Annalen*, 46(2), 167–178.
8. University Grants Commission (UGC), Nepal. *Education Management Information System (EMIS) annual reports* (various years). Kathmandu: UGC Nepal.
9. Verhulst, P. F. (1838). Notice sur la loi que la population poursuit dans son accroissement. *Correspondance Mathématique et Physique*, 10, 113–121.
10. Sahani, S. K., Lee, T.-F., Pandey, D., Pandey, B. K., Jha, R., & Karna, S. L. (2026). Machine learning–augmented hybrid risk management and deep uncertainty quantification in Nepalese management systems: Fractional stochasticity, Wasserstein robustness, and rough–path neural filtering. *Journal of Intelligent Decision Making and Information Science*, 3(3s), 1850–1873. <https://doi.org/10.59543/jidmis.v3.971>
11. Sahani, S. K., Lee, T.-F., Pandey, D., Pandey, B. K., & Mandal, K. (2026). Hybrid analytical, numerical, and machine learning frameworks for solving deterministic and stochastic differential equations with stability, convergence, and uncertainty quantification. *Journal of Intelligent Decision Making and Information Science*, 3(3s), 1794–1849. <https://doi.org/10.59543/jidmis.v3.970>
12. Sahani, S. K., Lee, T.-F., Pandey, D., Pandey, B. K., Jha, R., & Karna, S. L. (2026). Explainable machine learning for credit risk management and intelligent lending decisions in Nepalese cooperative banks: A mathematical review. *Journal of Intelligent Decision Making and Information Science*, 3, 1549. <https://doi.org/10.59543/jidmis.v3.1549>

13. Sahani, S. K., Lee, T.-F., Pandey, D., Pandey, B. K., Sah, B. K., Jha, R., & Sah, D. K. (2026). Hybrid analytical–numerical–machine learning architectures for decision-making, risk management, operational stability, and uncertainty quantification in dental practice management systems: A mathematical theory. *European Journal of Prosthodontics and Restorative Dentistry*, 34(5s), 298–313.
14. Sahani, S. K., Lee, T.-F., Pandey, D., Pandey, B. K., Sah, B. K., Mandal, K., & Jha, R. (2026). Biotechnological catalysts in hybrid analytical–numerical–machine learning architectures for risk-aware decision-making and uncertainty quantification in Nepalese management systems. *European Journal of Prosthodontics and Restorative Dentistry*, 34(7s), 332–350.
15. Sahani, S. K., Lee, T.-F., Pandey, D., Pandey, B. K., Sah, B. K., & Jha, R. (2026). A hybrid analytical, numerical, and machine learning framework for health management, resource allocation, operational efficiency, risk forecasting, and strategic decision-making in Nepal. *International Journal of Computer Information Systems and Industrial Management Applications*, 18(9s), 1221–1242.
16. Sahani, S. K. (2024). Big data learning analytics & optimization: Algorithms, challenges, and applications. *Analysis and Metaphysics*, 23(1), 912–924.
17. Sahani, S. K. (2024). AI-enhanced finite element method (FEM) for structural analysis. *Journal of Electrical Systems*, 20(1), 661–676. <https://doi.org/10.52783/jes.8946>
18. Sahani, S. K. (2023). Application of numerical methods in structural health monitoring using IoT sensors. *Journal of Electrical Systems*, 19(1), 194–207. <https://doi.org/10.52783/jes.8941>
19. Sahani, S. K., et al. (2022). A comprehensive study on predicting numerical integration errors using machine learning approaches. *Letters in High Energy Physics*, 96–103. <https://doi.org/10.52783/lhep.2022.1465>
20. Sahani, S. K. (2023). Neural network surrogates for weather prediction using numerical solutions of the shallow water equations. *International Journal of Intelligent Systems and Applications in Engineering*, 11(3S), 356–368.
21. Sahani, S. K., & Sah, B. K. (2024). Integrating neural networks with numerical methods for solving nonlinear differential equations. *Computer Fraud and Security*. <https://doi.org/10.52710/cfs.703>

Appendix A. MATLAB-Equivalent Routines

Table A1. Python–MATLAB correspondence for the core analytical routines.

Task	Python (this paper)	MATLAB equivalent
Growth-curve nonlinear least squares (Eqs. 1–3)	scipy.optimize.curve_fit	lsqcurvefit
ODE integration (RK4 and reference solution)	scipy.integrate.odeint; custom RK4	ode45, ode23
Bootstrap CI and Monte Carlo forecast	NumPy residual-resampling loop	bootstrp with a custom fit function handle
Linear regression (institutional panel)	sklearn.linear_model.LinearRegression	fitlm
Random-forest regressor	sklearn.ensemble.RandomForestRegressor	TreeBagger
Gradient-boosting regressor	sklearn.ensemble.GradientBoostingRegressor	fitrensemble(..., 'LSBoost')
Cross-validation (RMSE, MAE, R^2)	sklearn.model_selection, sklearn.metrics	cvpartition, loss